Network Transmission Model: a dynamic traffic model at network level

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

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

November 7, 2013

Word count:
- nr of words in abstract: 157
- nr of words (including abstract): 4135
- nr of figures & tables: 8 * 250 = 2000
- total: 6135

Presented at the 93rd Annual Meeting of the Transportation Research Board
ABSTRACT

The Macroscopic or Network Fundamental Diagram (NFD) describes the traffic flow in an area as a function of the number of vehicles in that area. In recent years, the NFD has been experimentally validated. Gating has been mentioned as most promising as the main application of the NFD. However, a state description in only a few parameters also gives advantages for setting up a dynamic traffic simulation program.

This paper uses this advantage and proposes the Network Transmission Model, a dynamic simulation program based on the NFD. The network is split up into subnetworks (cells), for all of which an NFD is defined. Based on the accumulation, the flows between the cells are determined. Contrary to earlier approaches, the method is applicable to a many-subnetwork system and accounts for the limited capacity from one subnetwork to the next. The model is applied to a network showing the calculations of various control schemes, including routing and gating.
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 needs to be tested, possibly on-line, for which traffic simulation programs are being used. We argued earlier (1) that for the optimization of a larger area, for instance a network, the look-ahead period also should be larger. Moreover, the number of vehicles or links in a larger network increases if one wants to optimize the control on the larger network. That means that model predictive control using traditional simulation programs will be too time consuming. Hence, there is a need for a quick model which can simulate a wide area dynamically in a fast computation time.

It has been shown that on an aggregate level there is a relation between the number of vehicles and their speeds (2, 3). 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 equilibrium conditions (3). Some dynamics have been described (4) or explored (5). The models describing the flows from one area to the next are often only considering two regions Geroliminis et al. (6) or for models describing more regions e.g., (7), no physical effects of the limited boundary capacity is taken into account. 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 into account. Such a model is useful for on-line optimization of traffic measures. The paper also shows the application of the model in relation to simple NFD-based control concepts.

The remainder of the paper is set up as follows. In the next section, the literature on the NFD and traffic flow simulations is discussed. Section 3 describes the newly developed model. Then, section 4 describes control scenarios. The control scenarios are implemented in a case study, as is presented in section 5. Finally, section 7 presents the conclusions.

2 LITERATURE REVIEW

This section describes the literature in traffic flow dynamics, as well as the control concepts related to the NFD.

2.1 Control using NFDs

Already several decades ago, the concept of an area-wide variation of the fundamental diagram was proposed (8). This has been tested (9). After Daganzo reintroduced the concept (2) in relation to gating, the topic has gained more attention. The paper by Geroliminis and Daganzo (3) has shown empirically that the NFD holds. Other researchers have further studied the impact of inhomogeneity. For instance, there are simulation observations that productions decrease with inhomogeneity (10), empirical observations during strikes and hence large inhomogeneities (11). Daamen et al. (12) explained the network dynamics in a simulation network and Daganzo et al. (5) looked at the dynamics of a simplified system. Knoop et al. (4) gave an equation to incorporate the 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. (13) or Geroliminis et al. (6). The basic idea is to keep the number of vehicles in an area under the critical level. 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 (14), also leading to considerable travel time gains. Both of these control concepts will be tested here.

2.2 Simulation of traffic flow dynamics

In traffic flow theory, several macroscopic models are available. One of the most intuitively understandable is the cell transmission model (15). Consistent with the ideas of a demand and a supply (see also (16)), this model describes the flow 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 (17), which is adapted in the Link Transmission Model (18). Similarly, there are hybrid approaches or solutions in Lagrangian coordinates (19). 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, it will give difficulties with multiple directions within each cell.

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. (20).

Concluding, we combine the concepts of the cell transmission model and good node models in a Network Transmission Model.

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 Traffic coding

The basis of the model are subnetworks, called cells in the description of the computational methodology. 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 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 $w$

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

(1)
The accumulation is now calculated as

\[ K = \sum_{Z \in A} k Z L Z \frac{l Z}{w A} \]  

(2)

For each cell, it is registered which fraction of the vehicles (and thus accumulation) is heading towards which destination \( s \); this is called \( \zeta_s \). The routing from cell \( A \) to the destination is coded by the next neighboring cell \( B \) in so called destination-specific split fractions \( \eta_{s,A,B} \). All neighboring cells of \( A \) are indicated by the set \( B \). The fraction lies between 0 and 1, \( 0 \leq \eta_{s,A,B} \leq 1 \) and all vehicles should be heading somewhere: \( \sum_{B \in B} \eta_{s,A,B} = 1 \). In our formulation, vehicles are assumed to have arrived their destination once they arrive somewhere in the cell. This could be changed in a future version.

### 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.

The dynamics of traffic are simulated in these subnetwork, using properties of the NFD in each subnetwork. For these cells the NFD is assumed to be known. 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 \( B \), \( C^B_A \); this is determined exogenously
2. The demand from cell \( A \) to cell \( B \), \( D^B_A \)
3. The supply in cell \( B \), related to the total demand to cell \( B \)
The demand from A to B is determined based on the NFD, the function which relates production $P$ to the accumulation $K$: $P = P(K)$. This NFD has to be determined exogenously, for which are several methods, empirically (3) or theoretically (21). In fact, we can construct a demand and supply scheme similar to the cell transmission model Daganzo (15). 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 NFD for higher accumulations:

$$S = \begin{cases} 
P_{\text{crit}} & \text{if } K \leq K_{\text{crit}} \\
P(K) & \text{if } K > K_{\text{crit}}
\end{cases}$$  \hspace{1cm} (3)

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

$$D = P(K)$$  \hspace{1cm} (4)

This is graphically shown in figure 2. 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, $D_B^A$ is only a part of the total demand in cell A, $D_A$. In fact, we consider the destinations separately. Hence, the demand in A for each of the destinations is

$$D_{A,s} = \zeta_s D_A$$  \hspace{1cm} (5)

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

$$D_A^B = \sum_{\text{all destinations }} \eta_{A,s}^B D_{A,s}$$  \hspace{1cm} (6)

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

$$\tilde{D}_A^B = \min \{ D_A^B, C_A^B \}$$  \hspace{1cm} (7)

The fraction of traffic allowed over the boundary between A and B is indicated by $\theta_A^B$

$$\theta_A^B = \min \left\{ \frac{\tilde{D}_A^B}{D_A^B}, 1 \right\}$$  \hspace{1cm} (8)
As an intermediate step, we now have the effective demand from cell A to destination \( s \) via cell B:

\[
\tilde{D}_{A,s}^B = D_{A,s} \theta_{A,s}^B \tag{9}
\]

The total demand towards cell \( B \) is determined by adding all effective demands towards cell \( B \):

\[
D^B = \sum_{A \in \text{neighboring cells of } B} \tilde{D}_{A,s}^B \tag{10}
\]

This is compared with the supply in cell \( B \). 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 \( B \) \( \psi^B \) is calculated:

\[
\psi^B = \min \left\{ \frac{S^B}{D^B}, 1 \right\} \tag{11}
\]

All cells \( B \), neighbours of \( A \), which have effective demand \( \tilde{D}_{A,s}^B \) larger than zero are combined in set \( B \). It is now calculated what is the lowest of these outflow fractions. This will be the restricting factor for the flow from cell \( A \): \( \Psi_A \):

\[
\Psi_A = \min_{B \in B} \{ \psi^B \} \tag{12}
\]

If the supply restricts the flow, the actual flow to cell \( j \) is proportional to the demands to cell. Now, the flow is set as the minimum of demand and supply. 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 \( s \), indicated \( q_{B,s}^A \) and the flow in the opposite direction, \( q_{B,s}^B \):

\[
K_s^A(t + \tau) = K_s^A(t) + \left( \sum_{B \in B} q_{B,s}^A - \sum_{B \in B} q_{B,s}^B \right) \frac{\tau}{w_A} \tag{13}
\]

In this equation, \( \tau \) 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 (2), is to limit the inflow in an area. For the study at hand, we choose limit the inflow such that the accumulation will not exceed the critical accumulation \( (K_{\text{crit}}) \), i.e. the accumulation from which the performance is decreasing. The number of vehicles that can be added to the cell is:

\[
(K_{\text{crit}} - K) w \tag{14}
\]
Under the control we limit the supply (modify eq. 3) such that this is not exceeded

$$S = \max \{(K_{\text{crit}} - K) \frac{w}{\tau}, 0\}$$  \hspace{1cm} (15)$$

The consequence is that the vehicles are waiting in the neighboring cells. In this paper, no advanced predictive control strategies are developed. In the simple control scheme applied here, we simply reduce the supply (eq. 3). We vary the cells on which this inflow limitation is being applied – the details of the settings are specified in section 5.1.3.

4.2 Routing

4.2.1 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, 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.

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 $v = \frac{P}{K}$.

From distance and speed the time to go from one neighboring 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 neighboring 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 neighbor $B$. This fraction, called $\eta_{s,A,B}$ in section 3 is applied in the simulation.
4.2.2 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 Network Transmission Model 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 $\tau$:

$$\tau = \frac{N_B}{q_B}$$  \hspace{1cm} (16)

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

The direction-specify NFD routing algorithms uses this expected travel time $T$ as basis for the route choice. In the routing, the cost of travelling from $A$ to $B$ has to be determined. For the trip from $A$ to $B$ for traffic towards cell $s$ we add half $\tau_A$ to half $\tau_B$, thus representing 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 4.2.2, 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 in which fraction of the disturbed travel times the fastest route from cell $A$ to $D$ goes towards cell $D$. This is determined for each neighbouring cell of $A$. These numbers give the split rate of traffic for traffic to $s$ in cell $A$.

5 CASE STUDY

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

5.1 Setup

5.1.1 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 NFD 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.
5.1.2 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 base demand 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. To this base demand, an extra, time varying demand is added. Figure 4b shows this demand for the direction right-left, the highest demand. The other demands get an additional demand which is proportional to this. Note that the profile simulates the loading onto the network, and then a decreased demand. 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.

5.1.3 Control

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

5.2 Results

Figure 5a shows a snapshot of the model during the simulation. 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 low, 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. It is remarkable that if gating is applied to more cells, the network is more prone to gridlock for longer update times. 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.
FIGURE 5 Snap shot of traffic flow operations – color indicates speed, and bar height accumulation.

FIGURE 6 The delays for different routing strategies

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.

Earlier it has been found (22) 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). The G-NFD can be plotted as color graph from the top of the plane; figure 7b is an example of an experimental version of this G-NFD is taken from (22) based on 10 months data from the A10 urban freeway. It shows that the performance is indeed concave with the accumulation with no spread in density, like a fundamental diagram. The performance decreases with increasing standard deviation of density. The white areas indicate no data was found.
To produce a similar graph, we calculate for the standard deviation of the accumulations of the neighboring 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. (23). The simulation also gives the accumulation and the flow. Figure 7a show this for the simulation (for a different scale). The figure shows that performance indeed decreases with increasing inhomogeneity. This is because of the following. If a neighboring 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. Of course, comparing figure 7a with figure 7b, it is clear that our model only has part of the plane filled. This is because our simulation setting is limited. Also, the performance values, and the typical accumulation values differ. This is all 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 paper is a first step towards an aggregated traffic description. Some remarks need to be made, which is done in this section.

A first test for a new model is to see how it copes with known situations. The most simple case for a traffic model is a line of areas and a capacity decrease from halfway the line (see a snapshot in figure 5b), representing for instance a freeway with a lane drop. In this case, the Network Transmission Model simplifies to a Cell Transmission Model and predict the same traffic operations (see figures 8a and 8b). Nevertheless, the case of all cells in a line this is not the right model, since it aims at areas. In fact, for a temporal decrease of capacity (a freeway with an accident), the model will in fact fail, see figures 8c and 8d. The density in the upstream cell will increase to jam density. Once the accident has been cleared, the upstream cell is still in jam density, the demand will be zero, and there will be no outflow. This is not a defect of the model, but the wrong application of the model. This one-directional motion is not the application area for the Network Fundamental Diagram, and hence not for the Network Transmission Model.
Extensions or alternative models can be thought of to handle these cases. 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.

Secondly, it has been shown that the spread of congestion has an effect on density. This can be analysed Mazloumian et al. (10), explained Daganzo et al. (5) and described in a equation Knoop et al. (4). 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 7a. It is unclear whether this captures all these effects. A comparison with a real network, homogeneously and inhomogeneously loaded, should show so.

The third remark is related. The Network Transmission Model 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 a road stretch. 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.

7 CONCLUSIONS

This paper introduced the Network Transmission Model describing the traffic flow dynamics on an aggregate level. The network was splitted into cells and for the traffic flow dynamics a numerical approach based on the MFD was introduced. 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.

The model was applied in a test case in which we analysed to which extent the model could be used to predict the impact of traffic control. In cases without adaptive routing, the model tended to lead to gridlock results. This seems not realistic, and might be due to an unrealistic assumption of static, non-equilibrium routes. 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.

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

REFERENCES


Knoop and Hoogendoorn


