Control strategies for a highway network

A joint research project of SWOV, the Technical University Delft and the Institute for Perception TNO sponsored by the Dutch Ministry of Transport and Watermanagement

PART I

Preface

E. de Romph; H.J.M. van Grol & R. Hamerslag. *3DAS (Three Dimensional Assignment): A Dynamic Assignment Model for Advanced Traffic Management and Driver Information Systems.*

E. de Romph; H.J.M. van Grol & R. Hamerslag. *Application of 3DAS (Three Dimensional Assignment) at the Washington Metropolitan Area.*

H.J.M. van Grol. *Dedicated hardware for an assignment model.*

R-94-34 I
Leidschendam, 1994
SWOV Institute for Road Safety Research, The Netherlands

SWOV Institute for Road Safety Research
P.O. Box 170
2260 AD Leidschendam
The Netherlands
Telephone 31703209323
Telefax 31703201261

Contents

PART I

1. *Preface*
1.1. Introduction
1.2. Synopsis of the results

2. *The reports*

E. de Romph; H.J.M. van Grol & R. Hamerslag. *3DAS (Three Dimensional Assignment): A Dynamic Assignment Model for Advanced Traffic Management and Driver Information Systems.*

E. de Romph; H.J.M. van Grol & R. Hamerslag. *Application of 3DAS (Three Dimensional Assignment) at the Washington Metropolitan Area.*

H.J.M. van Grol. *Dedicated hardware for an assignment model.*

PART II

Dr. P.H. Polak & T. Heijer. *Control Strategies for a Highway Network.*

E. Wiersma & T. Heijer. *Safety of the Traffic Process on Highways.*

PART III

W.B. Verwey & T. Heijer. *Linking Driver Visual Workload on Near-congested Highways to Inductive Loop Measurement.*

C.M. Gundy & T. Heijer. *A Traffic Model based upon Neural Networks.*

1. Preface

1.1. Introduction

The research reported in these combined papers constitutes a further step on the way towards the realization of an extended concept of freeway traffic control. This concept, which is described in previous reports, is characterized by an emphasis on maintaining safe traffic conditions using a hierarchical structure in which several forms of traffic control are combined:

- dynamic traffic assignment by means of route guidance and traffic information, based upon a prediction model; this model takes into account the dynamically changing characteristics of all parts of the network and generates predictions over a time interval 30-60 minutes ahead
- ramp metering and speed control, organized in local units
- safety monitors, generating special instructions like: keep your lane instructions, separate speed regimes per lane, overtaking limitations for heavy vehicles etc. The monitors are incorporated in the local controllers.

These local units, also called subsystems, need not cover the whole of the network and can operate as stand-alone controllers if located far apart; in case the units are closer together their interactions may be coordinated by an additional coordinating unit. Also, the units are always linked to the assignment module in order to adapt their operations smoothly to the predicted changes in traffic state.

Apart from the development of the strategy, some work has been done to develop suitable software to test and calibrate controller algorithms by way of computer simulations.

1.2. Synopsis of the results

The aims in the current stage of the project were set on three main items: the development and calibration of a first version of an assignment model with stationary parameters, the development of a local control unit together with a software testing tool and fundamental theoretical and experimental development of practical safety criteria. Although it must be conceded that these aims could not be completely achieved, it is still felt that this stage of the project has produced significant steps in the construction of a comprehensive control strategy for the following reasons:
- the assignment model in the current version produces realistic predictions in a variety of traffic conditions in a small network; in any case, the model seems to work well in relatively stationary conditions that is, without accidents or sudden drastic changes and these are precisely the conditions that we want to maintain (incident detection and management, while doubtlessly vital to a complete strategy, are not a part of the *current* stage of the project)

- although the local control unit could not be developed completely, a crucial part of this unit, the reference model, has been derived;it is indicated that further development into a local control unit for ramp metering and speed "control" is not very complicated. Furthermore, it is also indicated that simple forms of coordination between local controllers and between local controllers and assignment model are quite feasible
- a sound theoretical basis has been developed to relate parameters of individual behaviour to more general stream characteristics, which provides to quantifiable safety criteria
- although the experimental research into the relation between driver task load (an important individual safety indicator) and measurable safety parameters (on the basis of induction loop data) has yielded no conclusive results (partly due to failure of the data collection system of the induction loops) there are still indications that such relations may be found. This is largely a problem of extending the database. Moreover, apart from the Time To Collision based indicator used now, there are several other types of indicators that can be derived from measurable parameters that have not been tried yet.

Possible continuation of this project in the short term would focus on the following items:
- construction of operational traffic model on the bases of the neural network model
- completing and testing of a local controller
- further investigation into safety criteria consisting of:
    * further development of criteria for the stationary stream
    * development of a theoretical basis for safety criteria in non-stationary conditions

2. The reports

The research for these various parts has been carried out by several contractors, hence the segmentation of this report.
The first part covers the developments of the assignment prediction model, in particular the calibration of the first implementation based on static parameters. This research has been reported by the Traffic group of the faculty of Civil Engineering of the Technical University Delft.

The second part regards the development of a reference model of a local traffic stream, based specifically upon parameters that are deemed relevant for assessment and control of safe traffic conditions. This model constitutes the linchpin of the control algorithm in the local control units, as was described in previous reports. The research involves an extension of current traffic flow theory to account for the influence of parameters directly related to (safety of) individual behaviour and has been carried out by the Traffic Safety Group of the faculty of Civil Engineering, Technical University Delft in cooperation with the Institute for Road Safety Research SWOV.

The third and fourth parts of the report contain results of a joint effort by the Safety Science Group of the Technical University Delft, SWOV and the Institute for Perception TNO. These investigations aim to quantify safety and stability of traffic processes. To this end it was attempted to find a relationship between local traffic parameters (as reported in the second part) and the average task load for drivers in this traffic.

In this way, it was attempted to identify those traffic conditions that evoke high task loads (as a correspondingly increased probability of errors) and that therefore should be omitted or counteracted by traffic control.
To this end, video observations and measurement of mental load characteristics of individual drivers were used together with detailed induction loop data and expert opinions on the safety of various traffic conditions. The preliminary results have also been used to define relevant parameters for the reference model of the local controller.

So far, the results indicate that quantifiable characteristics relevant to safety of individual behaviour can indeed be extracted from induction loop data. Definite conclusions about functional criteria have not been reached however, partly due to a rather extensive failure of the data acquisition system associated with the induction loops. This will have to be subject of further investigation.

Finally, the fifth part reports the results of the application of neural networks to modelling a local traffic stream, specifically in a situation containing on-ramp and intended for use in calibration and testing of control algorithms. The general idea here is the following: build a relatively simple neural network and have it "learn" to reconstruct traffic data as measured by induction loops in a number of successive cross sections. Since one of these cross sections contains an on-ramp, the network thus constructed can be used to predict the effects of ramp metering actions in conjunction with speed control on the input (upstream) side of the model. In this way, we may simply construct a practical model from actual loop data.

This investigation looked at several concepts for such a network and indeed produced a viable model. A comparison with simpler linear modelling techniques however, has shown that the improvement gained by neural networks over these linear models is only small.

3DAS

Three Dimensional Assignment

E. de Romph[1]
H.J.M. van Grol[2]
R. Hamerslag[1]

9 November 1993

1. Delft University of Technology, Faculty of Civil Engineering, Department of Transportation Planning and Highway Engineering.
2. Delft University of Technology, Faculty of Applied Physics, Department of Physics Informatics / Computational Physics.

1. Introduction

To describe the traffic conditions in a network the dynamic assignment model —3DAS— has been developed. It determines the time-varying traffic conditions, intensity and traveltime, for each link during a certain simulation period. Like most other dynamic assignment models, the simulation period is divided into intervals of equal duration, to which we will refer as periods. The link parameters may be defined separately for each period. The properties of the network and the travel demand are presumed given.
Two commonly used assumptions are made concerning route choice:
- all travellers are completely informed and are taking into account the future congestions.
- all travellers choose their minimum cost path.

The assumptions suggest a user optimized approach. In a realistic situation however these assumptions are not valid. For example in case of an accident the traffic will queue at the accident and not diverse to different routes automatically. Only if a certain route remains congested for a long time, or in case of rerouting information, travellers will start to take different routes. The model is primarily constructed as a user-optimal model. In paragraph 7. however a strategy to model a more realistic route choice is proposed.

Due to the discretization of time and for simplicity the two following assumptions are made concerning the links:
- the traffic conditions are homogeneous along a link and constant for the duration of a period.
- The variables of a link —intensity (q), density (ρ) and speed (v)— are subject to a fixed relation described by:

$$q(t) = \rho(t) \cdot v(t) \qquad \text{(eq 1)}$$

This report will first give an outline of the model and describe the iteration process, then the input requirements are described. All the different parts of the model are explained in detail in paragraph 4 through 6. In paragraph 8 the importance of the developed software is shown. Finally in paragraph 9 several parts of the model are calibrated. This is done by comparing the model with the micro-simulation FOSIM, and with data from the Amsterdam freeway network.

2. Outline of the model

The (user-optimal) model determines, in an iterative process, the conditions on the network for every link in every period. In every iteration the route for every OD-relation in each departure period is calculated. According to these routes the OD-matrix is loaded on the network. For each iteration new link-traveltimes are calculated based on the conditions created in the previous iteration. The iteration process stops when the stop criterion is reached, i.e.when the loading of the network remains unchanged over the iterations.

The iteration process is basically the same as used for static equilibrium assignment methods, and is shown as a flow chart in the next figure.

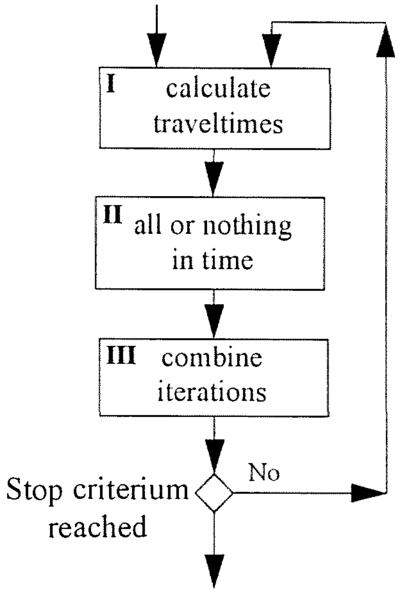


Figure 1: Iteration process flow-chart

Each block in the flow chart is referred to by a roman number.

I. For each link and for all periods the traveltime to traverse each link is calculated. The traveltime is based on the conditions in the network as a result from the previous iteration. In the first iteration, when the network is empty, the traveltimes are free flow travel times. The relation between the load on the network and the traveltime is determined with traveltime functions and discussed in section 4.

II. An "All or Nothing Assignment in Time" is performed to load all OD-pairs for all departure periods on the network. For every OD-relation and every departure period the shortest path is calculated based on the traveltimes calculated in (**I**). According these shortest paths, the traffic is assigned to the links in the path. This part is further explained with the flow-diagram in figure 2.

III. The end-result of the current iteration is obtained by combining the results of the last iteration with the results of the previous iteration. See section 6

The "All or Nothing Assignment in Time" (**II**) is explained in the flow-diagram of figure 2:

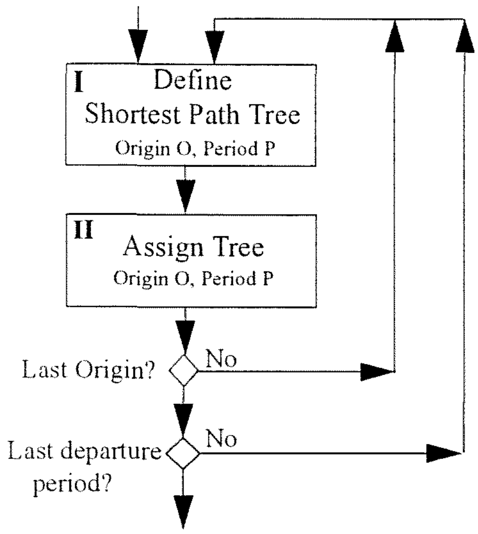


Figure 2: All-or-Nothing-in-Time flow chart

I. Starting with the first origin and the first *departure* period the shortest path tree to all the destinations is calculated.

II. For the current origin and departure period the value of the OD-relation is loaded on the network, using the calculated shortest path tree.

This process is repeated until all origins and all departure periods are loaded on the network. The pathfinding is further explained in section 5.1, the assignment in section 5.2

3. Input Requirements

The basic input requirements of the model are a network and an OD-matrix. Special settings to simulate ATMS measures are described in section 7.

3.1 Network

The network is given in the form of a directed graph, with links to represent the road segments and nodes to connect the links. The links have been given parameters to represent network properties in general and road characteristics in particular. Some link parameters are constant, e.g. the link length, others may change, e.g. the number of lanes, due to an accident or road-works. Several link parameters can therefore be separately defined for each period. However all parameters are constant over the length of the link and for the duration of the period.
In principle there is no minimum or maximum link length, neither any requirements concerning the difference in link length. However long links are less sensitive to a fluctuating traffic demand and therefore to congestions.

3.2 OD-Matrix

The travel demand is specified in an OD-matrix. Because the model is considering time-variance, a time varying OD-matrix is required for the model. For each departure period an OD-matrix is needed similar to an OD-matrix used for static assignment models. This results in a three-dimensional OD-matrix, OD_{odp}, which denotes the amount of travellers travelling from o to d in period p.

In general a three-dimensional matrix in not available, and a traditional matrix is used which is
defined for a larger time-span and is divided over the periods with a *departure time function*. The departure time function specifies for each period the percentage of the OD-relation that departs during that period. There are several methods to determine the departure time function. The simplest method is using the same function for all OD-relations. A more complicated method takes the traveltime to the destination into account and calculates an individual function for every OD-relation. The process is illustrated in figure 3.

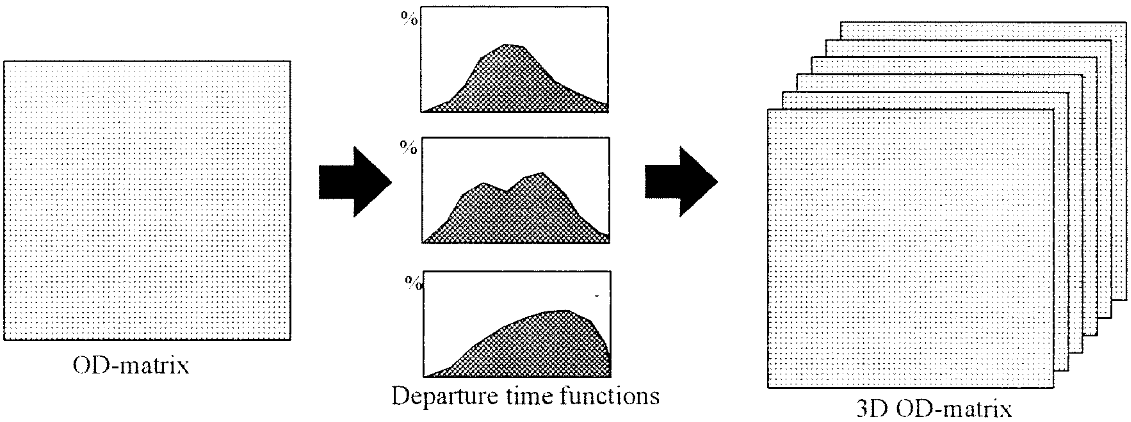


Figure 3: Determining a 3-dimensional matrix using departure time functions

4. Traveltime Functions

As follows from section 1, the traveltime for each link in each period is estimated by an iteration process, based on the traffic conditions on a link. Usually the traveltime is determined with a function describing the relation between intensity and traveltime (e.g. BPR function).
In a dynamic assignment model however a traveltime-intensity function causes some inconsistencies. This function shows an increasing traveltime with growing intensity. However when we observe the intensity, density and speed on a road section in reality, we find a different relation. In non-congested situations traffic will flow freely and with growing density and intensity the traveltime will indeed increase. In congested situations however the speed will drop rapidly and the intensity will **decrease** while the traveltime still increases.

Therefore, because the intensity does not decrease, a traveltime-intensity function is **not** suitable to describe a congested situation. A true traveltime-intensity function would show a duality, because with one intensity two traveltimes are possible (see figure 4).

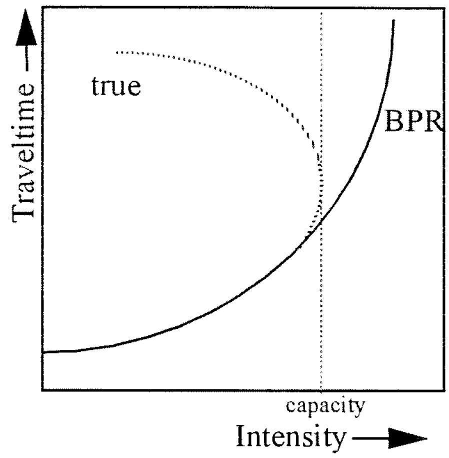


Figure 4: Traveltime-Intensity diagram

The relation between **density** and traveltime does not show this duality and is therefore more suitable for dynamic assignment. Instead of using the intensity as explanatory variable the density is used.

In a density-traveltime function the traveltime monotonously increases. The true form of the traveltime-density function is subject to further research. The function used in the model during the development is a speed-density function used by Smulders [30]. The speed v is given by two functions of the density ρ. See equation 2.

$$v(\rho) = \begin{cases} v^{max} \cdot \left(1 - \dfrac{\rho}{\rho^{max}}\right) & 0 < \rho < \rho^{crit} \\ \phi \cdot \left(\dfrac{1}{\rho} - \dfrac{1}{\rho^{max}}\right) & \rho^{crit} < \rho < \rho^{max} \end{cases} \qquad \text{(eq 2)}$$

In which v^{max} is the free-flow speed, ρ^{crit} is the critical density and ρ^{max} is the maximum density. At the critical density where free flow converts to congested flow, ϕ is chosen to make the function continuous at ρ^{crit}. The maximum density represents a no-motion traffic-jam,
In figure 5 this speed-density relation is plotted with the matching intensity-density relation. When the critical density is reached the speeds drops more rapidly and the corresponding

intensity indeed decreases.

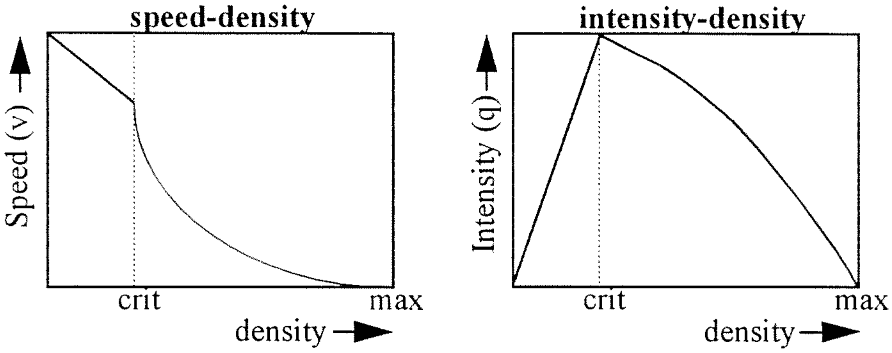


Figure 5: Speed-Density function and Intensity-Density function

In accordance with the density on a link, the matching speed (traveltime) is found using this **speed-density** function.

5. All or Nothing Assignment in Time

This paragraph describes the "All or Nothing Assignment in Time", which consists of the pathfinding method and the assignment method. Four path-finding methods are described and two assignment methods.

5.1 Path Finding methods

The paths are determined by the arrival time in each node of the path. As in static assignment models the arrival time in a node is calculated with the consecutive traveltimes of the preceding links in the path. In dynamic assignment models a difference arises because the traveltime on a link can vary over periods.

The pathfinding is illustrated by the following example in figure 6:

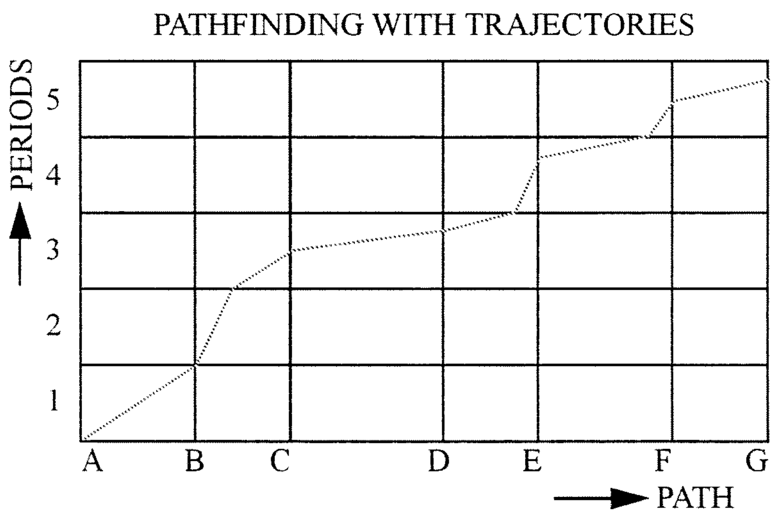


Figure 6: Pathfinding with trajectories

The path in time is drawn as a trajectory (distance-time) graph. The x-axis shows the path. The path starts in node A (Origin) and ends in node G (Destination). The y-axis shows the time. The time is divided in 5 periods.
The traveltime for link A-B is (in this case) equal to one period. So the traffic leaving node A in the first period, arrives in node B at the beginning of the second period. At link B-C in period 2 a different traveltime is encountered and the trajectory is adapted. Halfway link B-C the third period is entered with a different traveltime, so the trajectory is adapted to the new traveltime. Consequently link B-C is traversed in ±1½ period and the arrival time in node C is halfway during the third period. The traveltime to D is ±¼ period. So they arrive in node D at ±¾ of the third period, etc.

The arrival times in the nodes are used to define the shortest path tree in the same way as other (static) shortest path finding algorithms.

The described method to calculate the arrival times in the nodes is the most accurate method, considering that the traffic conditions are homogeneous along a link and constant for the duration of a period. During the research three more methods were developed to calculate the arrival times in the nodes. The first method is the method used by Hamerslag in 15. The second and the third method are improvements of this method. The method described above will be referred to as the fourth method.

I. The first method is the simplest method. The traveltime to traverse a link is the traveltime encountered in the period when entering the link. The traveltime is constant until the end of the link is reached. In case the next period is reached there is no adaptation of the trajectory. The method is illustrated in figure 7a. The method is computational very cheap, but problems with this method occur in case the traveltime in the next period differs very much from the current period. It is possible for traffic that departs a period later to arrive earlier. In figure 7a this can be observed. The traffic that depart in period *p2* will follow the dotted trajectory which arrives earlier in node *j* then the traffic departing in period *p1* which follows the solid trajectory.
II. The second method tries to deal with radical changes in successive periods by checking whether skipping to the next period will result in a shorter traveltime. The method is illustrated in figure 7b. With this method a better adaptation to sudden changes is achieved but it is still possible to depart later and arrive earlier.
III. The third method is almost equal to the second method. Instead of skipping a full period, the method skips to the beginning of the next period. The method is illustrated in figure 7c. With this method it is no longer possible to depart late and arrive early.
IV. The fourth method adapts the traveltime every time a new period is entered. This method is described at the beginning of this paragraph and is also illustrated in figure 7d. With this method it is not possible to depart later and arrive earlier.

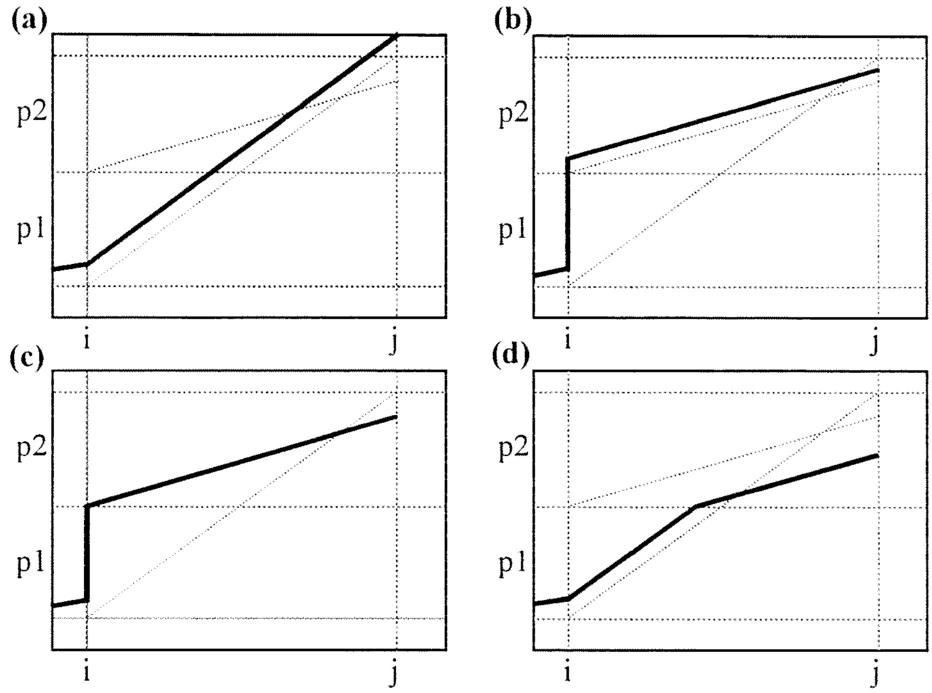


Figure 7: Four different methods, a,b,c and d, to calculate the arrival times in the nodes. The travel-times valid in period p1 and p2 for link i,j are given as dotted lines. The dashed lines are period and link borders. The solid lines represent the chosen trajectory.

The first pathfinding method is the "cheapest" for implementation purposes, the fourth pathfinding method is theoretically the most accurate, but involves more computational effort. The different methods are tested in paragraph 9.2.

5.2 Assignment method

The OD-matrix is assigned for each origin and each departure period according to the calculated trajectories. To calculate the contribution of a traveller to the density on link a in period p, the duration of presence on link a in period p is calculated.

If we focus on one traveller then two situations can occur:
1. *Several links are covered in one period.* In this case the traveller is only for a part of the period present on the link, and therefore should only be assigned for this part.
2. *One link is covered in several periods.* The traveller is present on the link during multiple periods and should be assigned entirely for each individual period.

Two methods were used during the development of the model. The first method will be addressed as the "*trajectory method*" and the second method as the "*surface method*".

The "trajectory assignment" method assumes that all traffic leaves at the beginning of the period. This method is computationally less demanding then the surface method.

The method is illustrated with the example in figure 8:

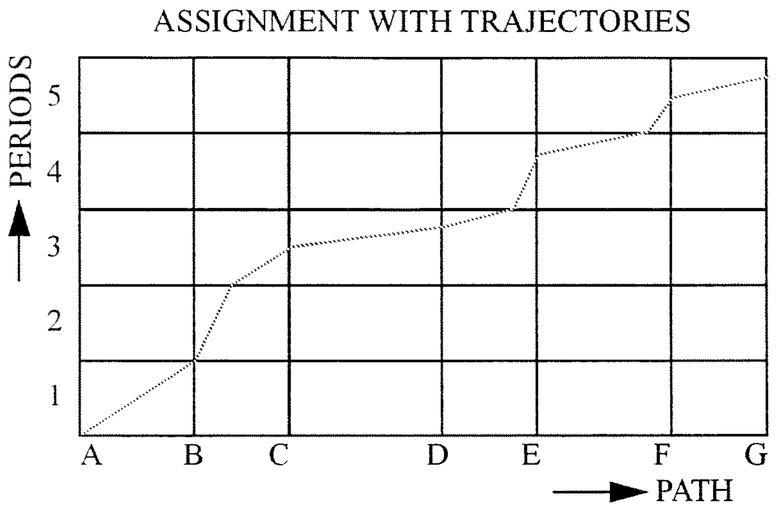


Figure 8: Assignment with trajectories

The traffic departing from node A in the first period with destination G (e.g.: $OD_{AG1}=100$), follow the trajectory calculated by the pathfinding. During the first period they pass the link A-B. In the first period all the 100 cars are assigned to the link A-B. In the second period they are all present on link B-C. In the third period they are partly present on link B-C and partly on C-D and D-E. So in the third period they are assigned according to the duration they are present on each link. Resulting in ±50% to B-C, ±30% to C-D, and ±20% to D-E. Etc.

Usually more then one traveller departs. In that case it is assumed that all travellers, departing in one period, leave the origin uniformly spread over the duration of the period. The travellers are assigned according to the *surfaces* between the trajectory the first car follows and the trajectory the last car follows.

The assignment method is illustrated with the following example in figure 9:

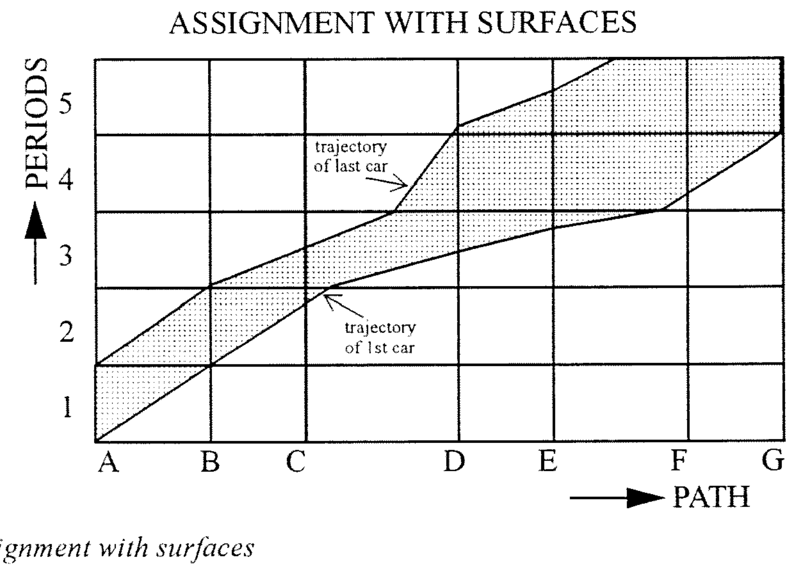


Figure 9: Assignment with surfaces

The first car leaving from node A in period 1 arrives at node B at the end of period 1. The last car leaving in period 1 arrives at the end of period 2 at node B. All the remaining cars in this period are leaving evenly spread between these two trajectories. So the traffic is assigned according to the surface between the two trajectories.
In the first period 50% is assigned to link A-B. In the second period ±50% is assigned to A-B and ±45% to B-C, and ±5% to C-D. In the third period the traffic is assigned to B-C, C-D, D-E and E-F. Since a density is assigned, the total amount of cars present in each period should be 100%, except for the period of departure and the period of arrival.
The link C-D shows that there can be a significant difference in traveltime between the last car and the first car. So the distance between cars can increase or decrease.

The final density on link a in period p is a summation of the contribution of all OD-pairs from all departure periods which traverse link a in period p.

Using densities instead of intensities offers, next to a more realistic representation of the delay function, the possibility to check the conservation of traffic present in the system. The amount of traffic present on the network should be equal for each period. Except in the period of departure from the origin and in the period the traffic reaches its destination.

The two described assignment methods are evaluated in paragraph 9.2.

6. Iteration Process

One iteration could be called an "All or Nothing Assignment" because all the traffic has been assigned to one physical route. However because the assignment is dynamic one iteration has no practical meaning due to temporal discontinuities (as Janson calls them [18]). Even if there is only one route several iterations are needed to eliminate the discontinuity. It is eliminated when the densities (and with densities the traveltimes) do not change between subsequent iterations. Only then a stable solution has been reached and the iteration process is stopped.

The subsequent iterations are weighted using:

$$\rho_{a,p} = (1-\lambda) \cdot \rho_{a,p}^{old} + \lambda \cdot \rho_{a,p}^{new} \qquad \text{(eq 3)}$$

Where $\rho_{a,p}$ represents the density on link a in period p, $\rho_{a,p}^{ol}$ is the result from all previous iterations and $\rho_{a,p}^{ne}$ the result of the last iteration. Until now λ is fixed (e.g. 0.2) or $\lambda = 1/(i+0.5)$ where i is the iteration number.

The solution reached represents a user-equilibrium. No mathematical proof of uniqueness or convergence is given. However several experiments with different networks and OD-matrices have shown a convergence in all cases after 20 to 40 iterations. The largest number of iterations were needed for networks with heavy congestion. We believe the iteration process can still be improved considerably, although it should be realised that the number of iterations also determines the maximum number of alternative routes.

7. Additional Features of the Model

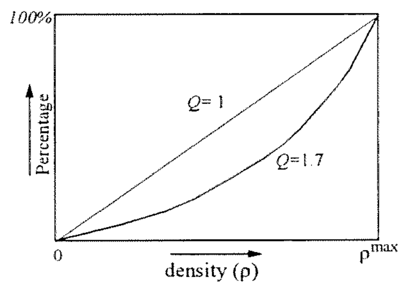
For a realistic simulation of traffic some extra features are needed in the model, that are now discussed briefly.

Jam building upstream

Two different methods for jam building upstream were developed:

- In the model and in reality the density on a link has a maximum. Maximum density represents a no-motion traffic-jam. When during the calculation the maximum density has been reached on a link, no more traffic can be assigned to it. Within one iteration however more than the maximum amount of traffic is allowed on a link. This results in a higher density than the maximum density. Because the traveltime on the link will now increase considerably, the chosen routes in the next iteration will most likely not include the overloaded link. When the same route *is* chosen, the traffic is assigned to the preceding link in the path in the same period. Because after each iteration the previous results are combined with the current, the density of the overloaded link will decrease again. Eventually no overflow will occur. Upstream of the overloaded arc, due to the reassigned traffic, the density will increase. Due to the increasing density the traveltime on the upstream link will increase. Depending on the demand this effect will proceed upstream and results in a queuing effect, which simulates the jam-building upstream.
- The second method does not wait for the density to reach maximum density, and starts to queue an increasing percentage of the traffic with increasing densities. The method is similar to the first method, but instead of queuing 100% of the traffic to the preceding link when maximum density is reached, this method determines a certain amount of traffic to be assigned to the preceding link with a function of the density. The function determines the percentage to be assigned to the preceding link in the next iteration according to equation 4.

$$F(\rho) = \left(\frac{\rho}{\rho^{max}}\right)^{Q} \quad \text{(eq 4)}$$



The parameter Q determines the "speed" with which the queue builds upstream. With increasing density on a link, more traffic gets assigned to the preceding link. When traffic gets assigned to a link for several iterations the density will still reach maximum density. Then the function F will return 100% so the maximum density will not be exceeded, because the next iteration *all* the traffic is assigned to the preceding link.

Due to the assignment of traffic to a preceding link a discontinuity in flow will appear. The discontinuity involves only the links that are involved in the queuing process. Up- and downstream that section no discontinuity will appear. Moreover, because the influence the queuing has on the traveltimes, the blocking-back in the final iterations is minimal, resulting in only minor discontinuity.

More Realistic Pathfinding

In every iteration the shortest paths are calculated, based on the assumption that the drivers are completely informed. In case of an accident, applied as a decreased capacity, the traffic will reroute automatically. In reality this will not happen and instead the traffic will queue on the route with the accident. Only when information of the accident is available to the drivers or when the duration of the congestion lasts very long, the traffic will start to take alternative routes. For the model, this means that route choice should not be based on the traveltimes resulting from the previous iteration, but rather on the traveltimes before the accident occurred. This suggests a delayed updating of traveltimes when it concerns route choice. To implement this two traveltimes are used for each link. A *historic* traveltime and an *actual* traveltime. For route choice the historic traveltimes are used and for assignment the actual traveltimes are used. Since in reality after some time travellers are informed (by radio) of the accident, the historic travel times are updated with the actual traveltimes if there is a significant difference for a long time. In order to maintain a diversion of traffic over different routes the traveltimes are only corrected slightly towards the actual traveltime. On links where rerouting information is available the historic traveltimes are more progressive updated by the actual traveltimes.

Rerouting

In case it is necessary to force a certain percentage to apply to rerouting measures, it is possible to enter fixed routes between two nodes for certain groups of OD-pairs. The percentage of the traffic to be rerouted can be specified.

On-ramp link dependency

At sections of the freeway with on-ramps a less efficient throughput takes place. The more cars enter the freeway at the on-ramp, the more delay occurs. The delay on the freeway is influenced by the ratio of the traffic on the on-ramp and the traffic on the main lanes. If the merging process is without delay the capacity of the freeway would not be influenced. However when the intensity on the on-ramp gets larger the capacity of the freeway decreases, due to an inefficient merging process.
By making the capacity of "the section of the freeway with the on-ramp" dependent of this ratio a more realistic behaviour is achieved.

The relation used in the model to achieve the dependency is given in equation 5.

$$\rho_m^{max} = \rho^{max} - M\left(\sqrt{\frac{\rho_r \rho_m}{(\rho^{max})^2}}\right)\rho^{max} \qquad \text{(eq 5)}$$

In which M is a parameter to adjust the impact of the dependency, ρ_m^{ma} is the new maximum density valid for the main section. ρ^{max} is the normal maximum density of the main section, ρ_r is the current density at the ramp, and ρ_m is the current density at the main section.

M is a value between 0 and 1 and determines the influence the ramp has. The dependency is visualized in figure 10 for two different values for M, as a function of the density on the main section and the density on the ramp

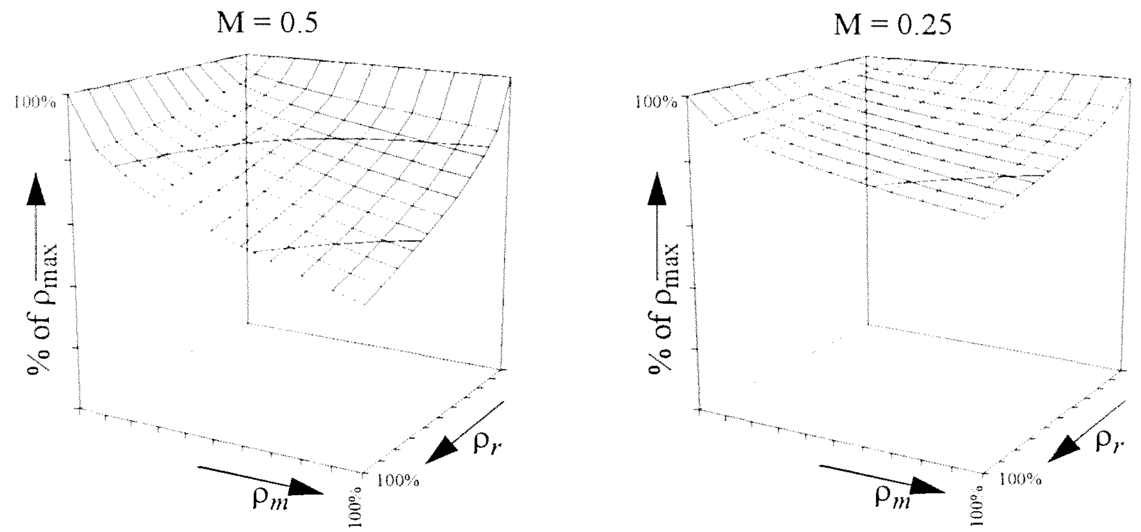


Figure 10: On-ramp link-dependency

If the demand on the main road is close to 100% and the demand on the on-ramp close to 0%, the influence on the maximum allowed density is minimal. However when there is 100% demand on the on-ramp and 100% demand on the main road the maximum allowed density of the section is lowest. The true form of the dependency relation is probably not symmetric and needs further research.

Ramp-metering

Ramp-metering is implemented as a limited capacity of an on-ramp. Since ramp-metering shows a more efficient merging process, the on-ramp link dependency explained above can be made less strict.

Speed control measures

The speed on certain links in the network can be controlled per period. Also the relation between speed and density can be changed.

Tidal flow

By changing link-capacities (to zero) for certain periods, the effects of tidal flow measures can be investigated.

8. Software

The model described in the previous chapter has been implemented on a graphical workstation. The software makes testing and validation of the model possible. The program is implemented on UNIX[1] and the X-windows graphical environment, and makes extensive use of graphical presentations. This has not been done to present "nice pictures" but provides very important benefits. The graphical presentation of the flow in time is a basic requirement to

1. UNIX is a trademark of AT&T Bell Laboratories.

analyse the results of the assignment. The graphics give a lot of insight in the model and not rarely unexpected results became clear because of the visualization of the results. This paragraph is therefore not meant to describe the developed software, but to stress the importance of visualization in dynamic traffic assignment and to present some ideas of visualization.

The software package is window based and every visualization is presented in a separate window. Six aspect of the model are visualized in the next six windows:

- The *main window*. This window shows the network as nodes and links. Different types of links are presented in different colours. Links can be selected and attributes can be changes. Zooming and panning is provided. The main window comes with several small pop-up windows to select several graphical settings and to select setting of the model. For example the pathfinding method and the assignment method. Several previous calculations can be saved in memory and/or to file. It is possible to compare different calculations.
- The *departure function window*. This window shows the departure function for the selected origin. The function can be edited interactively. A new function can be loaded from file and saved to file.
- The *traveltime functions window*. This window shows the traveltime function for the selected link. The parameters of the function can be changed interactively, and different types of function can be chosen.
- The *link information window*. This window shows the attributes of the selected link in time.For all the periods the speed, flow, density or traveltime is displayed. It is possible to show previous saved calculations in the same window simultaneously.
- The *trajectory window*. This window show the trajectories of a previously specified path for several departure periods. For each iteration the trajectories are saved and the convergence towards a certain trajectory can be observed.
- The *film window*. This window present the flow in the network in a film-like manner. In different colours the density on a link is shown. Usually a colour palette is used which presents a low density as light blue. With increasing density the blue darkens, until critical density has been reached. Critical density occurs at maximum flow. Then the colour changes to light orange, which darkens to dark red with increasing density. Light orange presents the start of congestion and dark red presents maximum density with a very low speed.

9. Calibration of the model

In this paragraph several parameters of the model are calibrated. Firstly the influence of period length and link length is investigated. Secondly the influence of speed-density functions is illustrated. Thirdly the model is calibrated by comparing the model with the microsimulation model FOSIM for two basic situations. FOSIM is a micro simulation model which is under development at the Delft University of Technology, which made the model easily available for these tests. The pathfinding method, the assignment method, the queuing algorithm and the merging algorithm are calibrated.

9.1 Comparing 3DAS with the micro simulation model FOSIM

The best way to validate a model is to test it in a real-life situation. However the model can be tested in some simple situations first. If the model does not perform well in these simple situ-

ations, simulations of large networks are useless.
Some simple basic tests have been done to validate the traveltime estimation of the model, which is important for a good behaviour of the model. No tests on route choice are included due to lack of comparison material.

Results of the following tests are reported:
- a 10 km roadway with a bottleneck
- a 10 km roadway with an on-ramp

The model is in principle not intended to reproduce each kilometre roadway in detail, but it should be able to simulate a correct in- and outflow and traveltime for the entire roadway.

These tests are chosen for two reasons:

I. They can be found in any network and are crucial for the traffic behaviour in a network. In case a bottleneck situation is included in a larger network the model is capable to simulate the delay of traffic due to this bottleneck and the effects up- and downstream in the network.

II. Test results of these situations are well known, which makes them easy to validate.

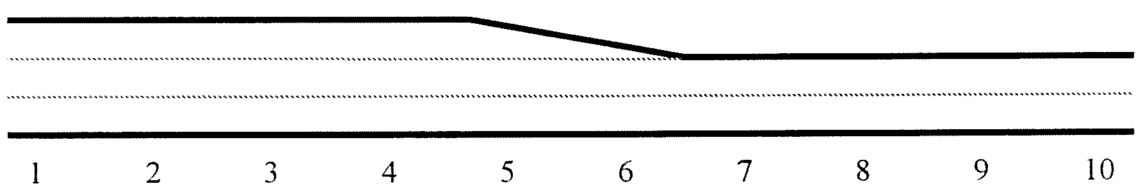
During the tests the different methods for pathfinding and assignment are tested and the parameter for queuing is calibrated. For the on-ramp test the parameter for the merging algorithm is tested. Firstly the results on the comparison is reported, in paragraph 9 the results of the calibration are reported.

The tests show the capability of the model to simulate some simple basic test situations. The model however is intended for **large networks** and should not be confused with micro-simulation models and its purposes. Whether the model behaves correct for large networks is tested in chapter 6, where the model is applied on the Amsterdam freeway system and in chapter 7 where the model is applied at a part of the Washington road system.

9.1.1 A bottleneck

A ten kilometre road with three lanes running into two lanes. The road is made with ten links of one kilometre. The simulated time is twenty minutes, with 40 periods of a ½minute.

The situation:



The traffic is travelling from left to right. Every minute a number of cars are entering the section according to table 1:

Table 1: Departure values

period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
#cars	40	56	72	88	104	104	104	104	104	88	72	56	40	20	8	0	0	0	0	0

The next graphs in figure 11 show the intensity (veh/hr) on a link in time on six of the ten links. Along the x-axis the periods in time are displayed. The fifth graph is the first link with two lanes. The solid line displays FOSIM results and the dashed (grey) line displays the 3DAS results.

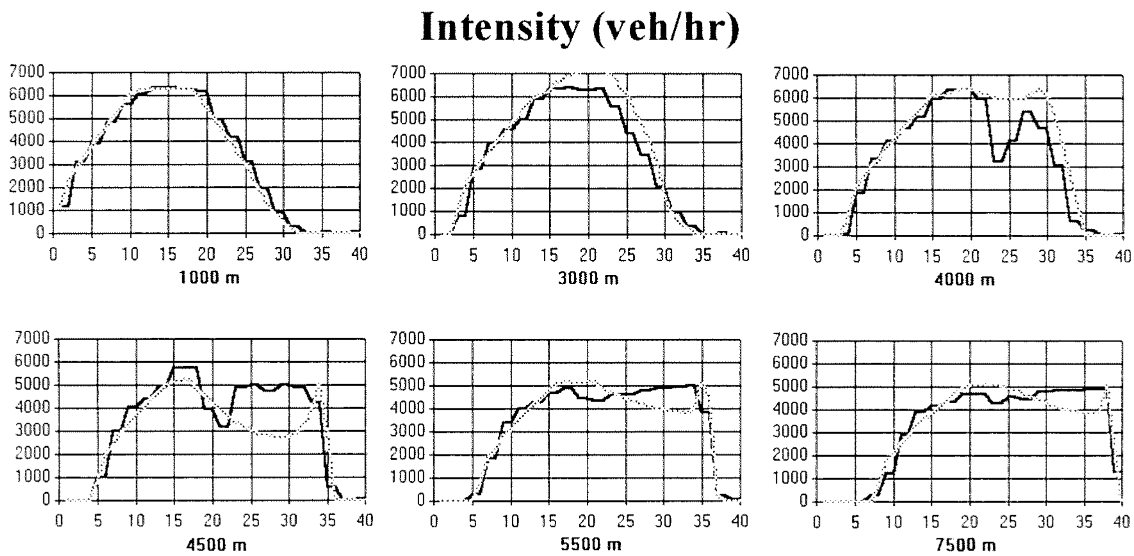


Figure 11: Intensity at six links for FOSIM and for 3DAS

The next six graphs in figure 12 show the matching speeds (km\hr) on the same links.

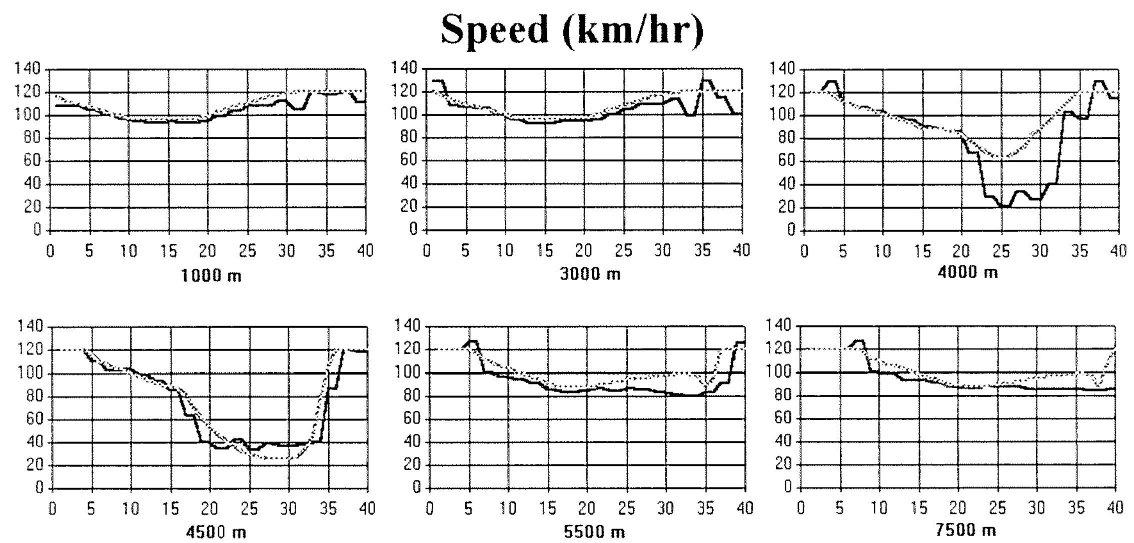


Figure 12: Speed at six links for FOSIM and for 3DAS

Comments:

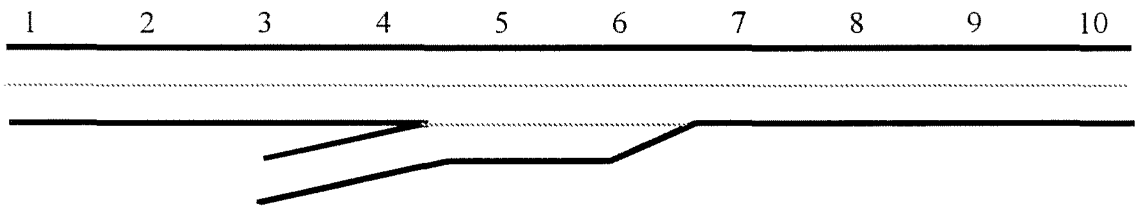
At 1000m the graphs show the intensity and the speed of the traffic entering the network. At 3000m the same shape shifted in time is observed. Due to a queuing in the bottleneck (at 5000m) the intensity at 4000m drops at the 12th minute. The matching speed decreases to 20 km/hr. Up to the 14th minute the number of cars entering the bottleneck decreases and the intensity increases again. After the 14th minute the number of cars entering the bottleneck decreases further and the intensity decreases too, while the speed further increases. At 4500m the same process shifted in time is observed. Downstream the bottleneck a steady intensity of approximately 5000 veh/hr is observed. This intensity is shifted in time on the last graph at 7500m.

The behaviour at the bottleneck is not reproduced in detail by the model, but downstream the bottleneck the results matches good. So the delay due to the bottleneck is reproduced. The queuing effects at 4000 m are with FOSIM larger as with 3DAS.

9.1.2 An on-ramp

A ten kilometre road with two lanes with an on-ramp at the 5th kilometre. The road is made of ten links of one kilometre, the on-ramp starts at the 3rd kilometre and merges at the fifth kilometre. The simulated time is one hour, with 60 periods of a ½ minute.

Situation:



The traffic is going from left to right. Every minute a number of cars are entering the section according to table 2:

Table 2: Departure values on-ramp

periods	1		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
main	43	43	43	43	43	43	43	43	43	43	43	43	43	43	43	43	43	43	43	43
ramp	12	12	12	12	12	12	12	12	12	18	26	34	39	38	34	29	23	23	14	2

Different methods of pathfinding and different methods of assignment were used. The parameter for queuing was calibrated by matching the results of 3DAS with FOSIM.

The next graphs in figure 13 shows the intensity (veh/hr) on a link in time on 6 of the ten links. At 3500 meter and 4500 meter the intensity at the on ramp is in the second lower line.

The solid line displays FOSIM results and the dashed (grey) line displays the 3DAS results.

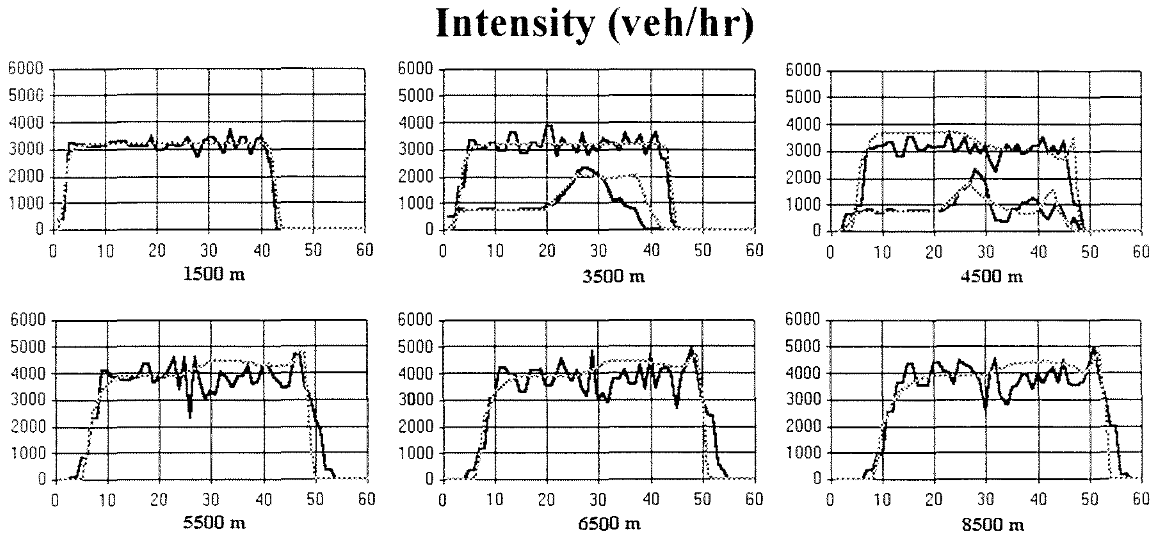


Figure 13: Intensity at six link for FOSIM and 3DAS

The next six graphs in figure 14 show the matching speeds (km/hr) on the same links.

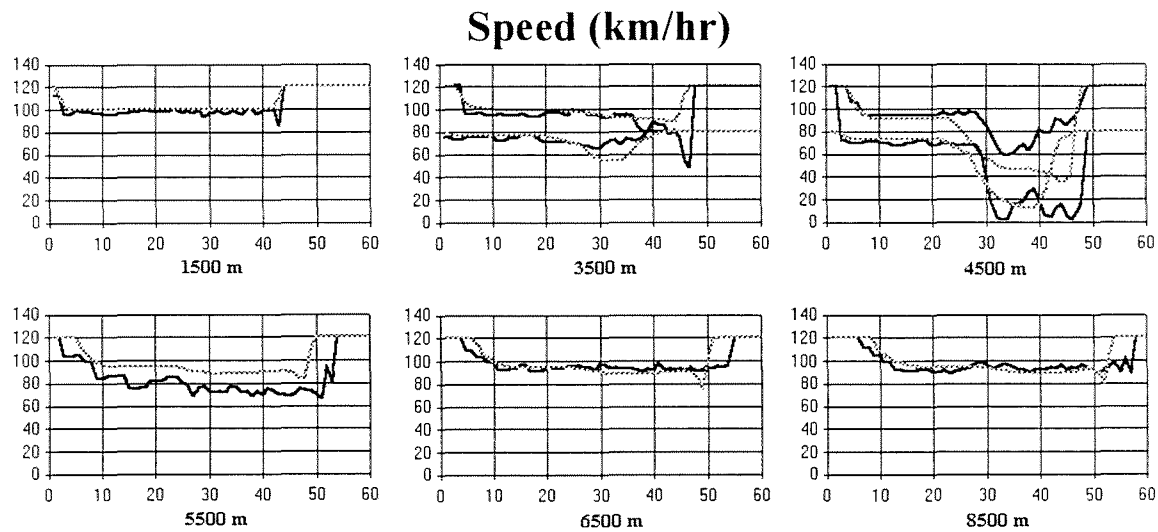


Figure 14: Speed at six links for FOSIM and 3DAS

Comments:

The first graph at 1500m shows the traffic entering the network. For the main road a steady intensity of 3000 veh/hr is entering the network for a total time of 20 minutes. At 3500m the same intensity is observed shifted in time. The traffic entering at the on-ramp is shown in the second line in these graphs. A steady intensity of 800 veh/hr enters the on-ramp for 10 minutes, the next 4 minutes the intensity increases to 2200 veh/hr and then it decreases to zero. At 4500m the traffic does not merge yet and the same intensity shifted in time on the main road can be observed. At the on-ramp however a queuing is observed at 4500m. Speeds are very low. At 5500m the traffic has merged and a intensity of approximately 4000 veh/hr is observed. The same intensity, shifted in time, with a slowly increasing speed is observed at 6500m and 8500m.

The results at the on-ramp are not reproduced in detail. FOSIM has more queuing at the on-ramp then at the main road. 3DAS model queues the same on the main road as at the on-ramp. This can be good observed at 4500 m. The total effect of an on-ramp is fairly good described.

9.2 Calibration results by comparing with FOSIM.

During the comparison with FOSIM the different methods for pathfinding, assignment are tested and the parameters for queuing and merging are calibrated.

9.2.1 Pathfinding

The four methods described in paragraph 5.1 are tested with various networks. The only two methods suitable for further validation are method three and four. With these methods later departing traffic cannot arrive earlier. Method four is expected to give the best performance. During the validation however some unexpected problems with method four were encountered. In case of a bottleneck at a certain link the trajectories departing from different departure periods tend to converge to the same trajectory, resulting in a high concentrated volume. The observation is illustrated with figure 15.

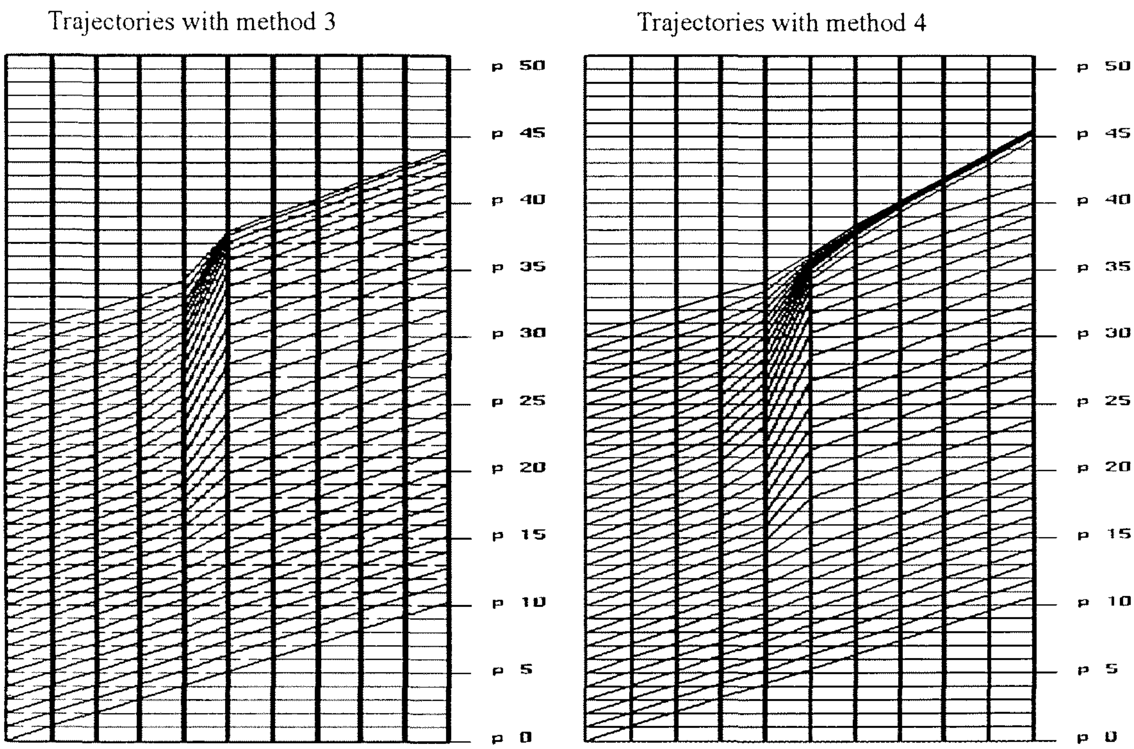


Figure 15: Trajectories (final iteration) with network "bottleneck" for method 3 and method 4.

Due to this concentration of traffic with certain conditions, method 3 gives better results then method 4.

9.2.2 Assignment

The two assignment methods for assignment described in paragraph 5.2 are tested with various networks. Of the two described methods, the "surface" method is expected to give the best results. The results of the trajectory method depends very much on the network and the chosen period length. A situation in which the trajectory method performs very poorly occurs with the bottle network. Figure 16 shows the load on the seventh link in the network. The link is situated downstream the bottleneck.

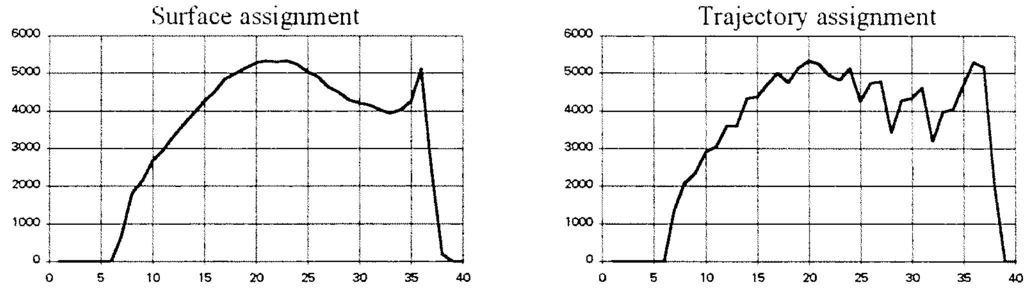


Figure 16: Flow on the seventh link of the network "bottleneck" assigned with the "surface" method and with the "trajectory" method.

The very fluctuating pattern with the trajectory method can be explained by the fact that a trajectory can skip a link in a certain period. The situation can occur in which a trajectory traverses link *a* in period *p2* and the next trajectory traverses link *a* in period *p4*. This results in no load in period *p3*. The situation is demonstrated in figure 17.

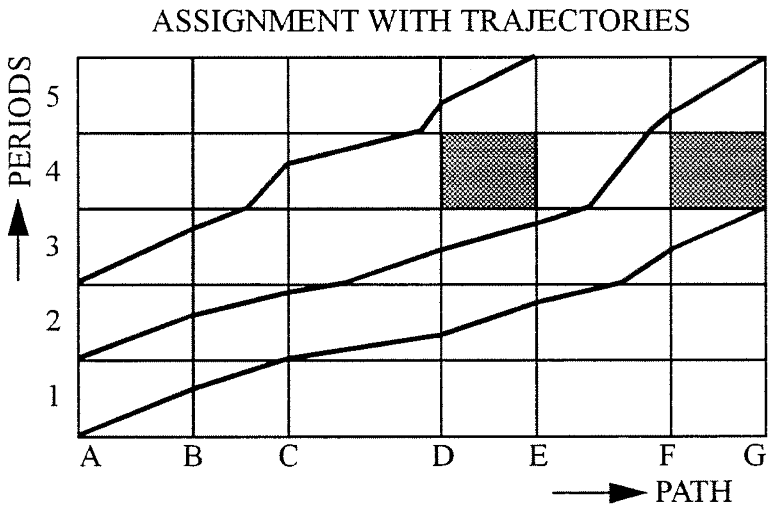


Figure 17: A trajectory in which certain periods are skipped. In this example nothing is assigned in period 4 to link E and to link G.

With the "surface assignment method" the traffic is assigned between the trajectories. This results in a much smoother (better) result.

9.2.3 Queuing

The queuing algorithm in 3DAS is explained in paragraph 7. Two different methods to achieve a blocking back mechanism and to prevent to density on a link to exceed maximum density are described. Both methods are tested with the "bottleneck" network. Based on those results only the second method is tested with the "on-ramp" network.

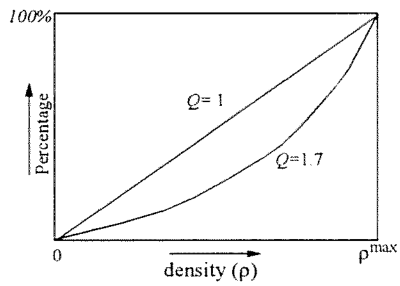
Queuing method 1

Queuing method 1 starts to block-back when maximum density is exceeded on a link. In the next iteration all the traffic will be assigned to the preceding link in the path. When using the Smulders speed-density function, this means that at maximum density the speed is equal to zero, and that the traveltime is very high. This results in an almost zero flow at the congested link, resulting in an almost zero flow at all the downstream links. Since this situation is not observed in FOSIM nor in reality, a certain outflow has to be maintained. Two different solutions are possible:

- Change the traveltime function, and maintain a certain minimal speed at maximum density.
- Do not allow the traffic to reach maximum density, but start to block back before maximum density is reached.

As described in paragraph 7 the second option is chosen. The FOSIM results show that there is blocking back to preceding links before the maximum density is reached on the congested link. This method prevents the congested link to reach maximum density and a consistent outflow is maintained. The outflow is controlled by the shape of the speed-density function and the parameter Q in equation 4. The equation and the shape of the corresponding function are repeated in equation 6.

$$F(\rho) = \left(\frac{\rho}{\rho^{max}}\right)^{Q} \qquad \text{(eq 6)}$$



Different values for the parameter Q are tested. With a value of 1.7 the best reproduction of the FOSIM results was achieved. The influence of the parameter Q is very intuitive. A higher value ($Q > 2$) results in queuing with higher densities, which results in higher densities, and a lower outflow, while a lower value ($Q < 1$) results in more queuing, lower densities and a higher outflow.

The next three graphs in figure 18 show the results with the network "bottleneck" for the link preceding the bottleneck, the link with the bottleneck and the link after the bottleneck. The intensity and the speed is shown for three different values of Q.

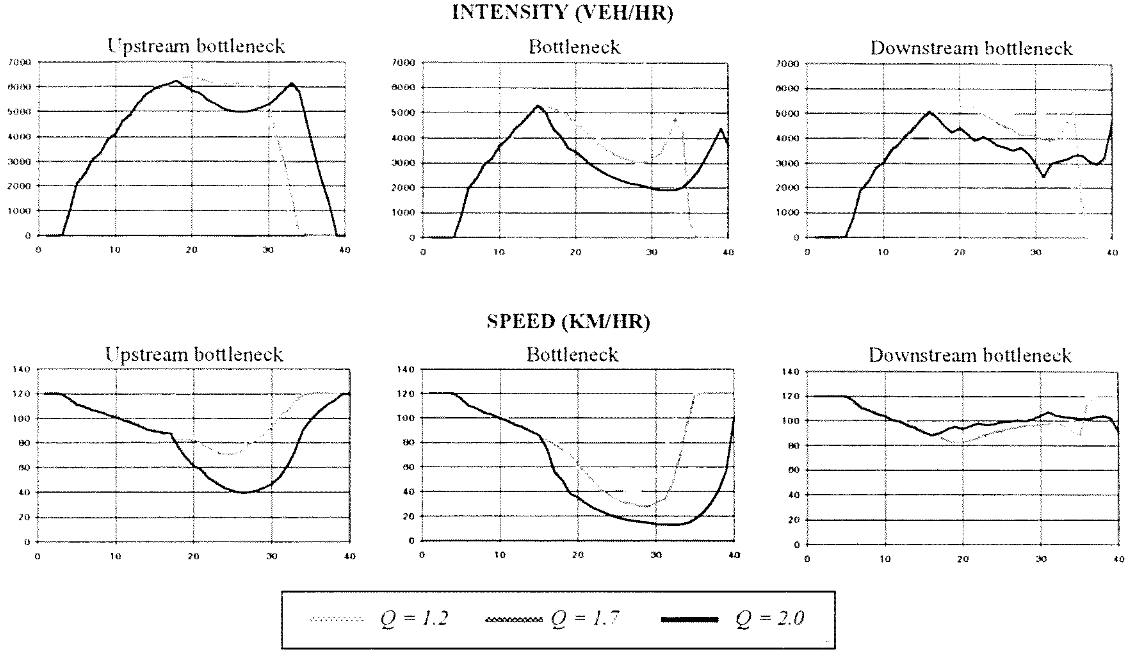


Figure 18: Three successive links with the load-patterns in time for three different values of Q.

Figure 18 shows with $Q = 2.0$ the lowest speed in the bottleneck, resulting in a longer queue, hence lower speeds upstream, the outflow of the bottleneck is therefore low. With $Q = 1.2$ the density does not become to high in the bottleneck, and the speed in the bottleneck and the outflow remain higher.

9.2.4 Merging

The merging algorithm is tested with the on-ramp network of FOSIM. As explained in paragraph 9.2.4 the parameter M determines the influence on the maximum density of the on-ramp. The best results were achieved with a value for M of 0.25. The influence of the parameter is very intuitive. With higher values of M more influence of the on-ramp is observed, with lower values less influence is observed. The merging process needs more study. Data from the field is very sparse, and the influence of the on-ramp much more complex.

9.3 Calibrating Speed-density functions with Amsterdam data

Each type of link in the network has its own parameters for its speed-density function. With the parameters of the speed-density functions the capacity, the speed and the throughput of a link are influenced. So the effects of accidents, weather changes, roadworks, etc. are incorporated in these functions.
The correct shape of a speed-density function is subject to many discussions. It is not possible to give an average speed-density function which is applicable for all types of roads. Even for one link a speed-density function can change due to weather changes, percentage of trucks, time of day, etc...
Since the speed-density functions are very important for the behaviour of the model, some research has been done towards the shape of the function, the influence of weather on these functions, and how the function changes during the day.

Data available of the Amsterdam freeway system has been used for this research. For one hundred induction loops, one minute data was available including the weather conditions reported at the Schiphol International Airport.
This paragraph will first give a short review of travel time functions, secondly the influences of location, weather and time of day are illustrated with the Amsterdam data. Thirdly two possible functions are compared for their use as a suitable function for real-time dynamic assignment. Finally a method is described to perform the adaptation of the function.

9.3.1 Travel time Functions: A short review

Travel time functions or delay functions have always been an important part of assignment models. They are used to express the travel time on a link as a function of the traffic volume, and they specify the effect of road capacity on travel times and route choice. Many different types of delay functions have been used in the past, and it would appear that for almost every major transportation study, a new and different delay function has been proposed. For a review see Branston 1976 [4] and Suh 1990 [31].
Two main approaches can be classified in the development of traveltime functions; the *mathematical* approach and the *theoretical* approach:

- In the *mathematical* approach a simple mathematical function is devised to replicate the observed data. The function usually does *not* contain parameters which are directly based on network and/or link characteristics, such as signal settings.
- In the *theoretical* approach these characteristics are usually well represented and the function is usually based on queuing theory.

The most widely used function based on the *mathematical* approach is the BPR-function (Bureau of Public Roads 1964 [2]). Its polynomial form is easy in application to computational procedures. The function is originally written as a traveltime-intensity function and is given in equation 7:

$$t = t_0 \cdot \left(1 + \beta \cdot \left(\frac{q}{c}\right)^n\right) \qquad \text{(eq 7)}$$

In which t is the traveltime on a link, t_0 the free-flow traveltime, and q is the volume, and c is the capacity. β and n are parameters of the function. Different combinations of β and n are used. The US department of transportation uses $\beta = 0.474$ and $n = 4$.

Since 3DAS uses density as explanatory parameter, the function has to be rewritten as speed-density function. Unfortunately it is impossible to derive a direct analytical transformation of the BPR function in a speed-density function. However it is possible [24] to derive an inverse speed-density function by using the fundamental hydrodynamic relation describing the throughput of a flow q in equation 8:

$$q = \rho \cdot v \qquad \text{(eq 8)}$$

In which q is the flow, ρ is the density and v the speed. The resulting density-speed function is given in equation 9:

$$\rho = \frac{c}{v} \cdot \left(\frac{1}{\beta} \cdot \frac{v_0 - v}{v}\right)^{\frac{1}{n}} \qquad \text{(eq 9)}$$

This function makes it possible to plot the function with its related functions (figure 19).

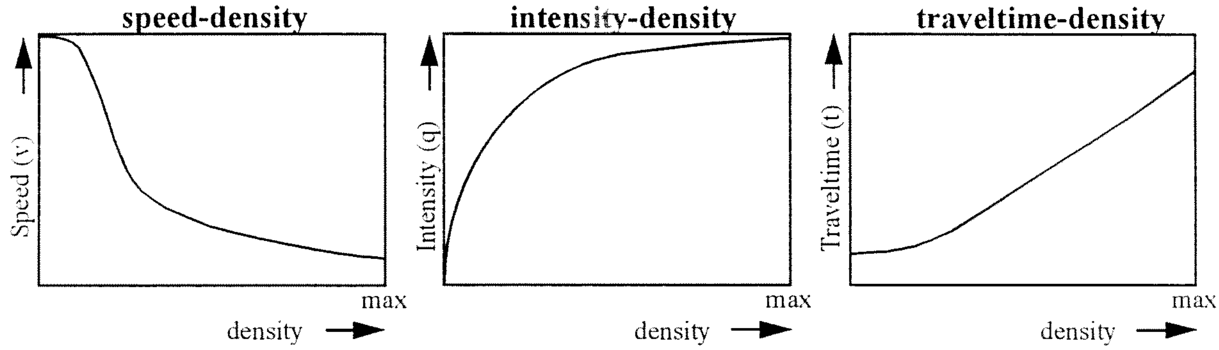


Figure 19: The BPR function with its related functions

The BPR function has four parameters: c, v, β and v_0.

A well known function based on the *theoretical* approach is the Davidson function [10]. This function introduced a parameter for type of road. The function is given in equation 10:

$$t = t_0 \cdot \frac{1 - mx}{1 - x} \qquad \text{(eq 10)}$$

In which t_0 represents the free-flow traveltime on a link, t is the average traveltime on a link, m a model parameter expressing 'quality of service' (value between 0.8 an d 1.0), and x is the volume capacity ratio: q/c, in which q is the flow, and c is the capacity.

By using the hydrodynamic relation of equation 8, equation 10 can be rewritten as a speed-density function [24]. This results in equation 11:

$$v = \frac{c + (\rho v_0) - \sqrt{c + (\rho \cdot v)^2 - 4m\rho v_0 c}}{2m\rho} \qquad \text{(eq 11)}$$

In which v_0 is the free-flow speed and ρ is the density.

The Davidson function has three parameters: v_0, m and c. The shape of the Davidson function and some related functions are given in figure 20.

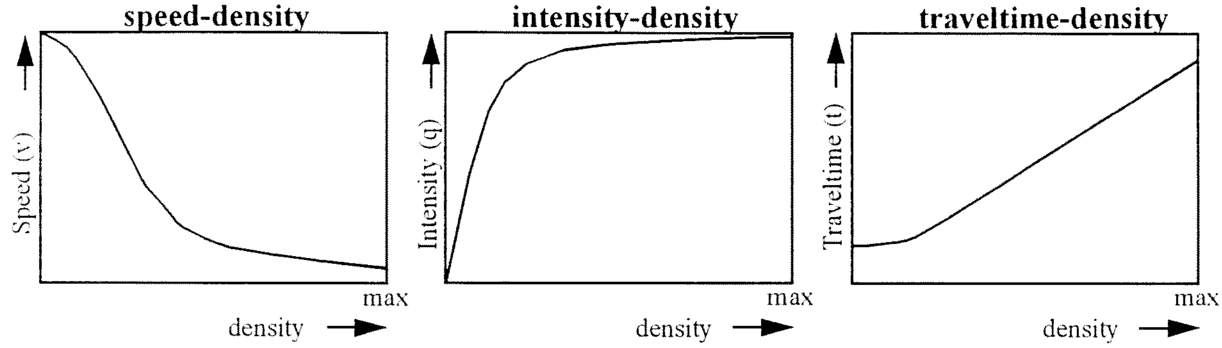


Figure 20: Davidson speed-density function and related functions

Almost all the delay functions that are used for transportation studies satisfy the following condition:

f(x) = strictly increasing.

A necessary condition for the assignment to converge to a unique solution.

It is a well-known fact however that the relation between volume and traveltime is *not* a strictly increasing function but a dual function. As the travel time increases the volume increases, but at a certain (critical) point the volume *decreases* while the travel time still increases. This effect is also observed in the intensity-density relation in figure 19 and 20. The intensity increases with increasing density. In reality the intensity should decrease after a certain level of density is reached, because the maximum density represent a no-motion traffic-jam with intensity zero.

The reason for using a strictly increasing function is, apart for convergence reasons, given by Rothrock and Keefer in 1957 [27]. They pointed out that for data collected with a 6 minute interval there is indeed a dual relation between volume and traveltime. With an increasing sample interval however this effect becomes smaller, and with a sixty minute interval it almost vanishes completely. Because static assignment models usually consider a total time interval of one hour to a complete day, the usage of a strictly increasing function seems plausible.

This experiment has been replicated with the available data from Amsterdam and the first three graphs in figure 21 show a similar observation, although not as clearly as in the original tests done in 1957.

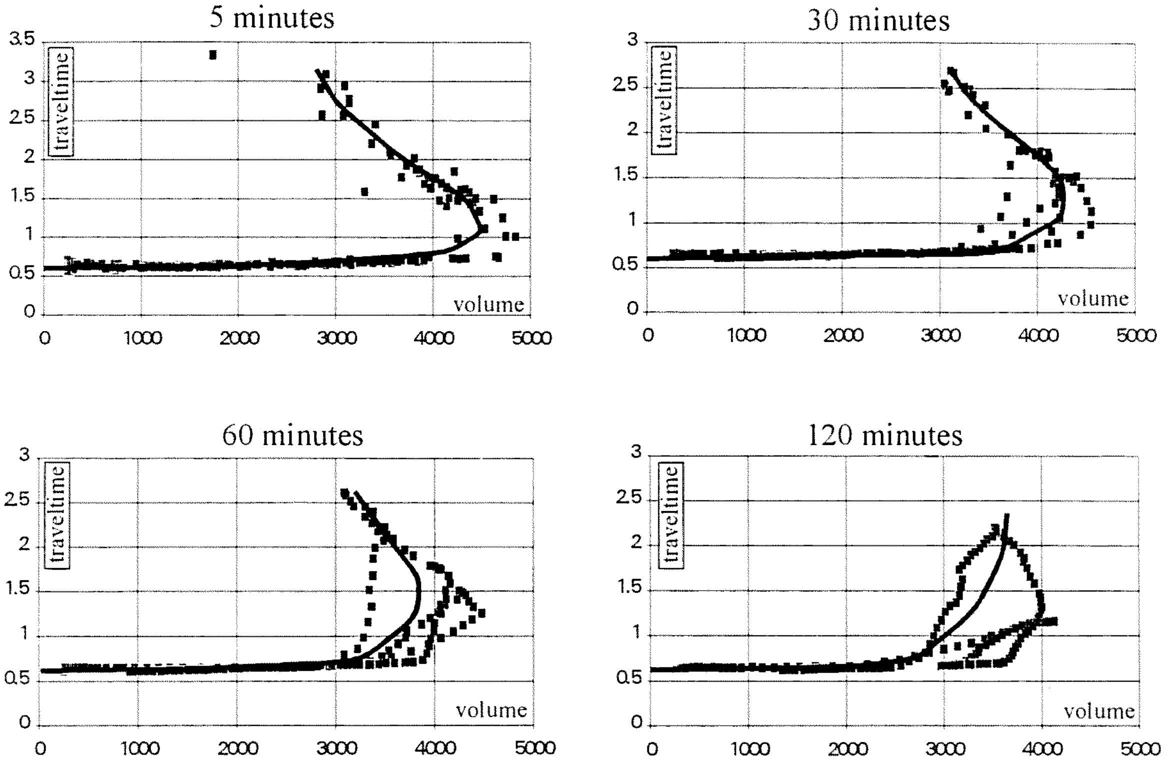


Figure 21: Relation between volume and traveltime for the same dataset with increasing sample interval (as moving average).

The reason that the "turning back" of the relation is still there with a sample interval of 60 minutes could easily be explained. Nowadays the rush-hour lasts for more then one hour. The speed drops at certain locations to less then 50 km/hr for a duration of two or three hours. In case the sample interval is further increased to 120 minutes the "turning back" is no longer observed, as can be seen in the last graph of figure 21.

With the introduction of dynamic assignment models however the justification of a long time-span is no longer valid because dynamic models divide the total time interval in equal periods of typically one to ten minutes.

Another justification for using a strictly increasing function is given by Rose et. al. [26]. He distinguishes interrupted and uninterrupted traffic flow conditions. Uninterrupted flow conditions exist where vehicles traversing a roadway are not impeded by any causes external to the traffic stream, such as signs or signals, although vehicles may be stopped by causes internal to the traffic stream. Interrupted flow conditions exist where the flow *is* impeded by external causes [3]. Therefore, freeway system provide uninterrupted flow conditions, while arterials are characterized as providing interrupted flow conditions. The relationship between travel time and volume take a different form when we must account for the effect of flow interruptions. These relations are shown in figure 22.

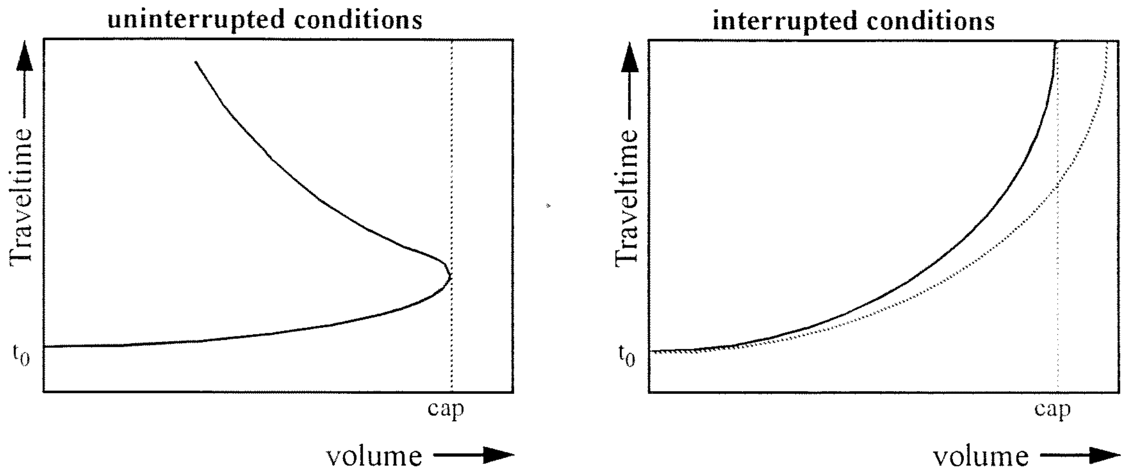


Figure 22: Volume-Travel time relation for interrupted and uninterrupted conditions (the dashed line is usually used to overcome problems with infinite travel times)

Only when the network is entirely constructed with arterials, the usage of a strictly increasing travel-time function is acceptable. For assignment studies with time-spans shorter than one hour and dealing with freeways a strictly increasing traveltime function is not explainable.

Many dynamic assignment models have recognized the problems with the traditional travel time functions and have adopted different travel time functions. By using the density on a link as explanatory parameter instead of volume the problem with the duality is easily solved. The (observed) relation between density and traveltime is a strictly increasing relation. By using this function the flow and traveltime increase with increasing density, and at a certain (critical) density the volume decreases again while the traveltime still increases. The basic relations between density, volume and traveltime are all satisfied.

For the 3DAS model the Smulders function [30] has been used during the development of the model. This function is developed according to the *mathematical* approach. The formulation is written as a speed-density function and given in equation 12.

The speed v is given by two functions of the density ρ.

$$v(\rho) = \begin{cases} v^{max} \cdot \left(1 - \dfrac{\rho}{\rho^{max}}\right) & 0 < \rho < \rho^{crit} \\ \phi \cdot \left(\dfrac{1}{\rho} - \dfrac{1}{\rho^{max}}\right) & \rho^{crit} < \rho < \rho^{max} \end{cases} \qquad \text{(eq 12)}$$

In which $v^{\max}$ is the free-flow speed, ρ^{crit} is the critical density and $\rho^{\max}$ is the maximum density. At the critical density where free flow converts to congested flow, ϕ is chosen to make the function continuous at ρ^{crit}. The maximum density represents a no-motion traffic-jam.
When the critical density is reached the speeds drops more rapidly and the corresponding

intensity decreases.

The Smulders function and some related functions are displayed in figure 23.

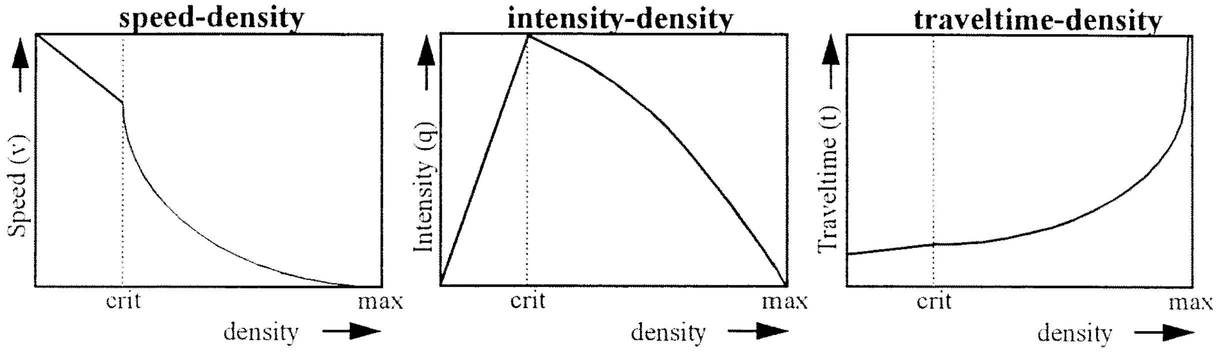


Figure 23: Smulders Speed-Density function and related functions

The difference with the BPR and Davidson functions shows most clearly in the intensity-density graph of figure 23. After a certain critical density has been reached the intensity drops, representing the start of congestion with decreasing throughput.

To get some more influence on the shape of this function two extra parameters are introduced in 3DAS for this function. The parameter *alpha* which influences the steepness of the first linear part of the function, and a parameter *beta* which influences the curve of the second part of the function. The equation is given in equation 12:

$$v(\rho) = \begin{cases} v^{max} \cdot \left(1 - \dfrac{\alpha\rho}{\rho^{max}}\right) & 0 < \rho < \rho^{crit} \\ \phi \cdot \left(\dfrac{1}{\rho} - \dfrac{1}{\rho^{max}}\right)^{\beta} & \rho^{crit} < \rho < \rho^{max} \end{cases} \qquad \text{(eq 13)}$$

In total the Smulders function has five parameters: α, β, ρ^{max} ρ^{crit} and v^{ma} .

Bin Ran et. al. [6] developed a traveltime function according to the *theoretical* approach. Two different sets of functions are recommended. One function for longer time intervals (5-30 minutes) and one for shorter time intervals (1-5 minutes). For both functions, the function is expressed as the sum of two components. The link is for that purpose divided in two parts, a flowing part and a queuing part. So the traveltime is the sum of a flow-dependent cruising time D_1 for the first part of the link and a queuing delay D_2 for the second part of the link. See equation 14, in which τ represents the traveltime for a link.

$$\tau = D_1 + D_2 \qquad \text{(eq 14)}$$

The function to express D_1 is a linear speed-density relation, derived from Greenshields formula (1933). The function is rewritten coping with the length of the queue which might be present on the second part of the link. See equation 15.

$$D_1 = 3600\left(\frac{l - (x_2)/e_m}{w}\right) \qquad \text{(eq 15)}$$

In which l is the length of the link and x_2 represents the number of vehicles present on the second part of the link, e_m is the maximum density and w is the cruising speed for inflow on the link.

The second part of equation 14, the term D_2, is expressed as the sum of two delay functions. See equation 16.

$$D_2 = d_1 + d_2 \qquad \text{(eq 16)}$$

In which d_1 expresses a non-random delay due to signal cycle effects. See equation 17.

$$d_1 = \frac{0.5c\left[1 - \frac{g}{c}\right]^2}{1 - \rho\frac{g}{c}} \qquad \text{(eq 17)}$$

The term d_2 expresses an overflow effect due to random arrival effects and oversaturation delays. For d_2 a difference is made for long-term and short term applications.
For long-term the following expressions for d_2 is used. See equation 18.

$$d_2 = 3600\frac{x_2}{\mu} + 900\Delta k[\rho]^n\left([\rho-1] + \sqrt{[\rho_a-1]^2 + \frac{m\rho}{\mu\Delta k}}\right) \qquad \text{(eq 18)}$$

In which g is the effective green time in seconds during the time interval, c is the signal cycle time in seconds, and p is the degree of saturation at the exit from link a during the time interval. In the second part of equation 18, x_2 denotes the number of vehicles present on the second part of the link, μ is the capacity at the exit of the link, Δk is the period length, and m and n are calibration parameters.

For short-term applications the following expression for d_2 is used. See equation 19.

$$d_2 = 1800\,\frac{(u_2 - v_2)\,\Delta k + 2x_2}{\mu} \qquad \text{(eq 19)}$$

For further details is referred to Ran 1992 [6], since a discussion on queuing theory and signalled intersections is out of the scope of this dissertation.

The first part of Ran's traveltime formula is visualized as a speed-density function in figure

24.

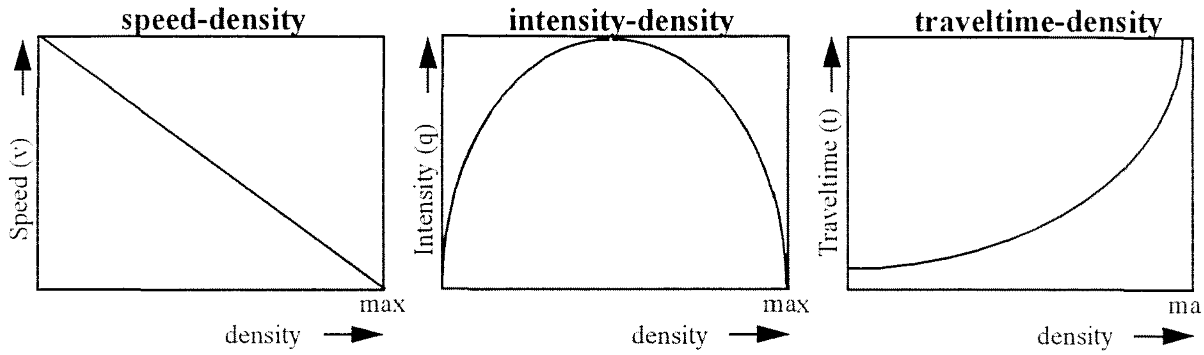


Figure 24: The flow-dependent part of Ran's travel time function

Ran's function is primarily meant for arterial networks with signalled intersections.

9.3.2 Influences of location, weather, and time of day.

Four days in December 1992 are chosen from our data collection. These days are chosen because the data sets contain few errors and there are several traffic-jams during those days. The weather conditions for these days are listed in table 3.

Table 3: Weather conditions

Day	Wind (knots)	Rain (time)	Visibility	Sky
3 Dec. 1992	18	16:00 - 18:00 22:00 - 24:00	good > 10 km	cloudy
7 Dec. 1992	15	01:00 - 12:00 15:00 - 16:00	good > 10 km	half cloudy
8 Dec. 1992	5	-	04:00 - 07:00 fog 11:00 - 21:00 fog visibility 100-200m	cloudy
9 Dec. 1992	6	-	good > 10 km	cloudy

From each day ten locations are chosen. The locations are selected on the variability of measurements during the day. In order to investigate the functions, measurements in all stages of traffic flow are the most useful. A location with only free-flow traffic is not useful to investigate a speed-density function.

The locations are chosen in different situations, i.e. merge sections, off-ramps and tunnels. For each day the same locations are used. The chosen locations are listed in table 4.

Table 4: Induction loop locations

nr	loopnr	location	lanes	remarks
1	75	A16 Utrecht -Amsterdam	3	
2	232	A4 The Haque - Amsterdam	2	
3	108a	A8 Alkmaar - Amsterdam	2	entering the Coen tunnel
4	105	A8 Alkmaar - Amsterdam	3	
5	160	A4 NieuweMeer - Amstel	3	near on-ramp
6	157	A4 NieuweMeer - Amstel	4	merging area
7	204	A19 NieuweMeer	3	fly-over
8	200a	A19 Coenplein - NieuweMeer	2	exit ramp
9	196a	A19 Coenplein - NieuweMeer	1	exit ramp
10	195	A19 Coenplein - NieuweMeer	3	

This paragraph illustrates the influence of several external factors, as location, weather, and time of day. The purpose is to show that it seems useless to define a general applicable function. The influence of weather for instance is very hard to quantify. All kind of rainfall have to be distinguished, several levels of visibility, etc.

Influence of location

To illustrate the influence of location, four speed-density graphs are shown in figure 25 at four very different locations. In each graph the location is numbered according to table 4. The measurements are taken with a 5 minute interval for one complete day (24 hours).

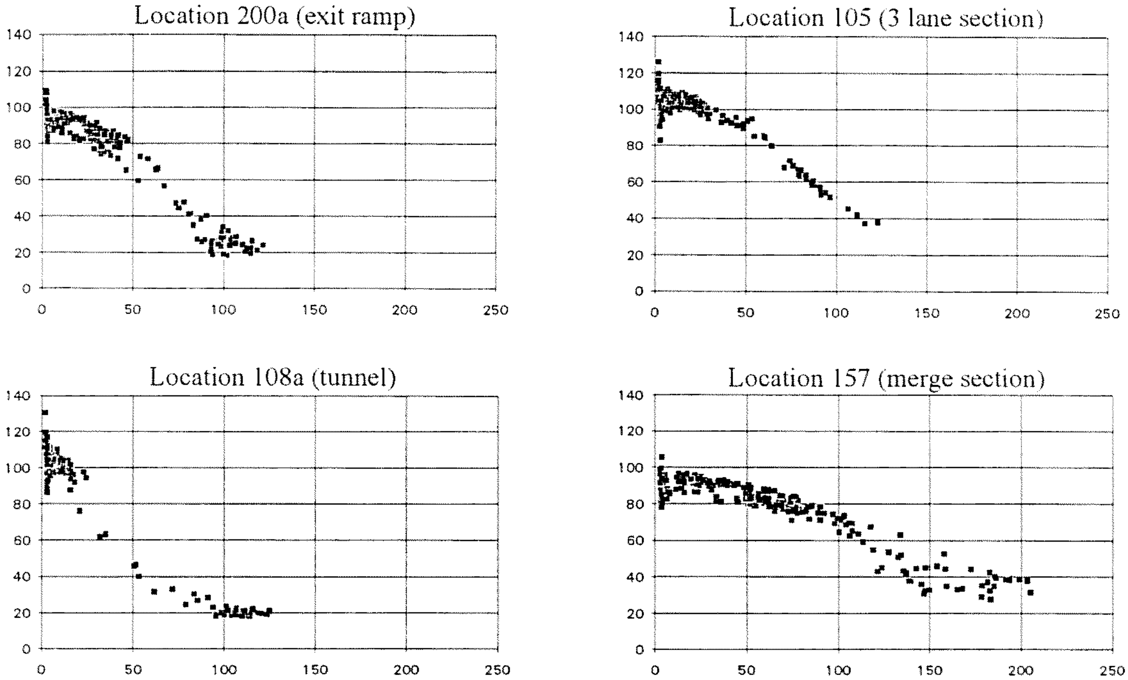


Figure 25: Speed-density graphs at the same day for four different locations.

The graphs in figure 25 show that the location has significant influence on the relation between speed and density. Location 200a is a two lane exit which has a steep relation between speed and density. Location 105 is a 3 lane freeway section which has a normal relation. Location 108a is a two lane section entering a tunnel. The speed drops here very soon with increasing density. Location 157 is a four lane merging section. Two freeways merge here in to one. The maximum speed is low and the relation is very flat.
Each day roughly the same patterns are observed at all locations.

Influence of weather

To illustrate the influence of weather, the Smulders function is fitted to the observed data with a least squares method. The resulting parameters of the calibrated functions are displayed in figure 34. For each location the maximum speed, the critical density, the beta parameter, and the resulting capacity are displayed The beta parameter is displayed in hundreds and the critical density as percentage of the maximum density which is assumed to be 125 vehicles per lane.

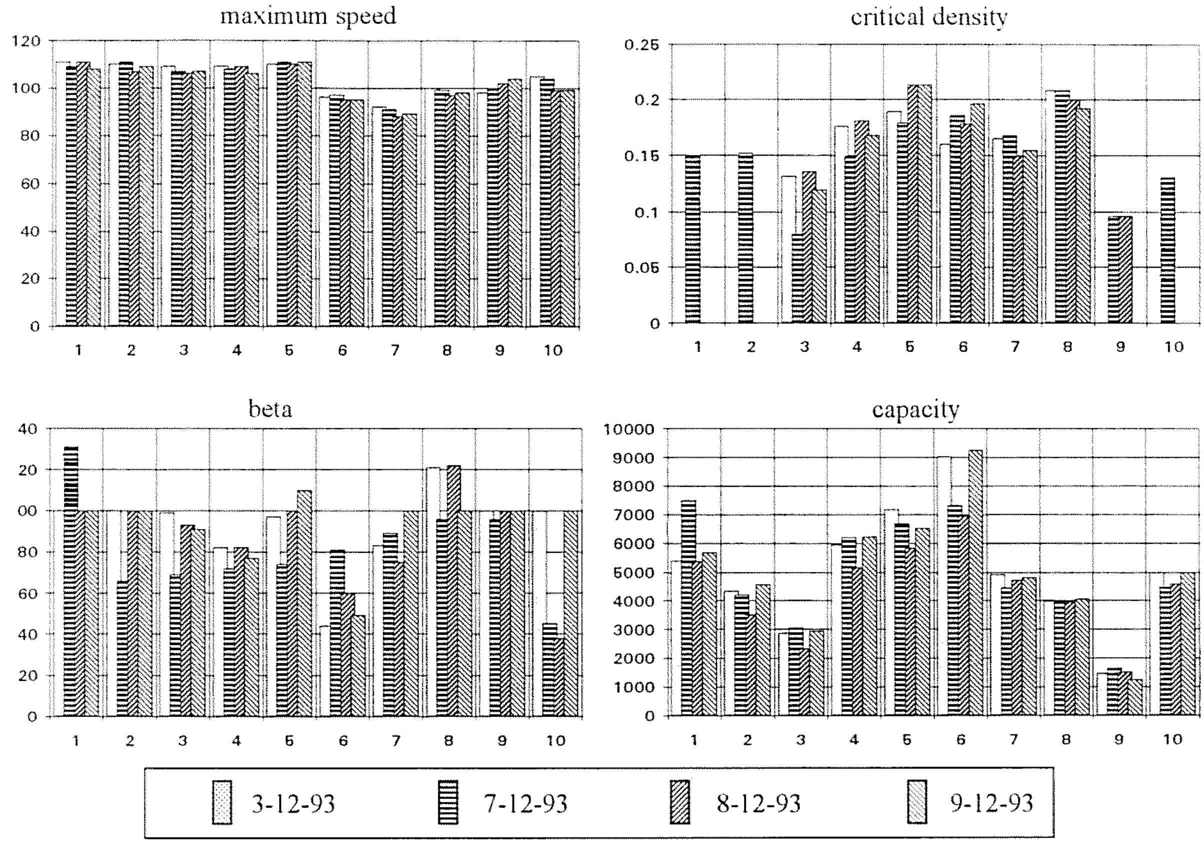


Figure 26: Maximum speed, critical density, beta and the capacity for each location at different days. Locations 1,2,9 and 10 not always contained enough "low speed" data to give a reliable estimate of beta and the critical density. For those days the beta value remained at the default value of 100, and the critical density is omitted.

The graphs show little variation in maximum speed. Especially the beta parameter which represents the relation in a congested state shows a lot of variation. The capacity is a resulting parameter of all the other parameters. The last graph in figure 26 shows that at some locations there is a significant difference in capacity. In general the data of 8-12-93 results in the lowest performance of the links. Since this is the day with the fog, this result is explainable.

Changes during the day

Until now the relation between speed and density is compared between different *days*. For real-time prediction it is also important to study the changes of the function *during* the day. These changes are probably related to weather circumstances and time of day (darkness).

The influence of weather is illustrated with the dataset of the 8th of December 1993. It was a day with a lot of fog and bad visibility. Only between 7:00 and 11:00 hours in the morning there was a good visibility.

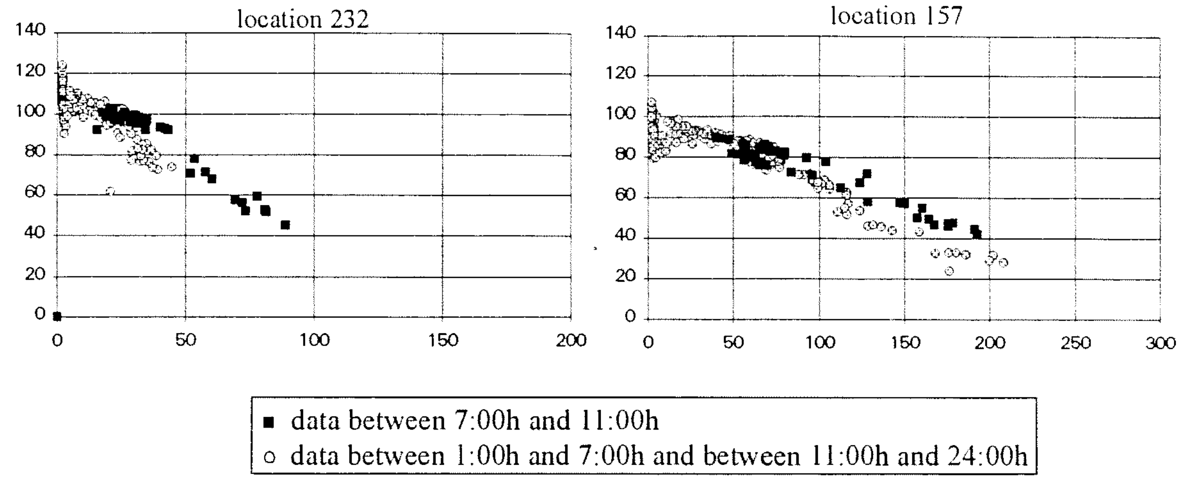


Figure 27: Within day influences for speed-density relations

The graphs in figure 27 show the data measured during the period with good visibility as black squares. For both locations higher speeds are observed.

When the dataset is smoothed the changes in time of the speed-density relation are more clear. In figure 28 the data is smoothed for 15 minutes.

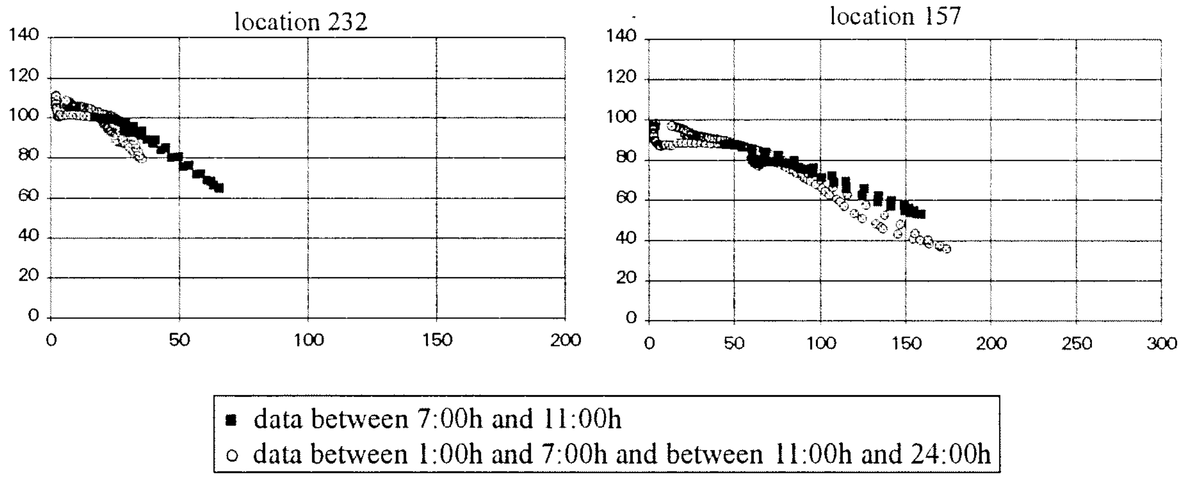


Figure 28: Within day influences for speed-density functions with aggregated data. (moving average for 15 minutes)

For both locations in figure 28 the patterns shows a different relation during the fog than during the clear period.
The 3rd and the 7th of December were days with rain. The influence of rain seems to be not as clear as the influence of fog. Only very slight changes in the speed-density relation during the day were observed for those days. The influence depends of course on the amount of rainfall, which was not very high during those days.

9.3.3 A Traveltime function for Real-time dynamic assignment

In case of a *real-time* situation all sorts of influences such as weather conditions, time of day and percentage of trucks, have their influence on the traffic situation. All these influences can be incorporated in the delay function. The former paragraphs showed that a general applicable function will not satisfy, and that a good strategy might be to just adapt the function to the current measured situation. Each link in the network can have its own individual function, which needs to be adapted to the actual conditions. From the functions introduced in paragraph 9.3.1 the only function which is suitable for this purpose is the Smulders function. A disadvantage of the Smulders function is the discontinuity. This makes it difficult to fit the function to the data. Therefore in case of *real-time* dynamic assignment a simple speed-density function is required which is easily adaptable with real-time data to different situations.

In this section such a simple function is introduced. This function is introduced to provide a function which is computational 'cheap'. Since every link in the network can have its own function, the adaptation of these function to different conditions will take some computational effort. The "*3DAS*" function is given is equation 20:

$$v = \frac{\beta \cdot v_{max}}{\rho^{\alpha} + \beta} \qquad \text{(eq 20)}$$

In which v is the speed on a link, v_{max} is the free-flow speed, ρ is the density. α and β are two parameters. The function and its related functions are plotted in figure29

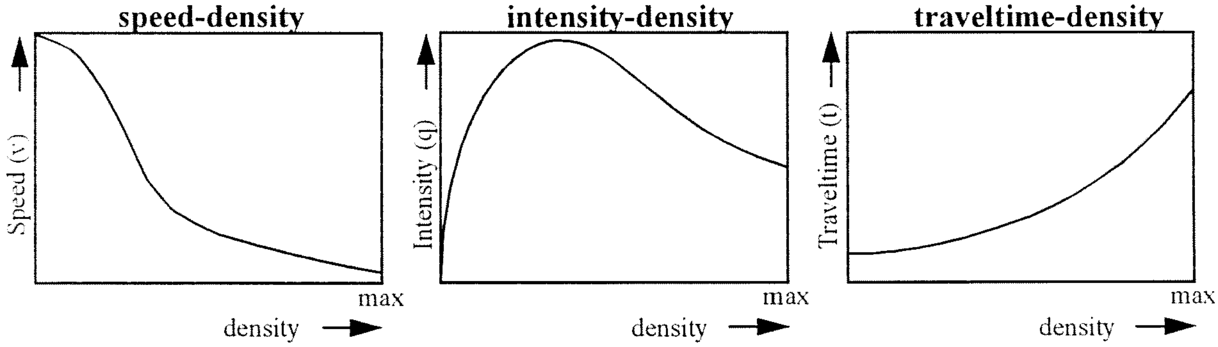


Figure 29: The 3DAS function and its related functions

The 3DAS function has three parameters: α, β and v_{max}.

For each location in table 4, and each day in table 3 the 3DAS function is calibrated using a least squares method. The fit is compared with the Smulders function. In figure 30 the mean square error is shown for both functions for each day at the ten locations. The locations are set along the x-axis and the mean square error are along the y-axis. The solid line represents the Smulders function, the dashed line represents the 3DAS function.

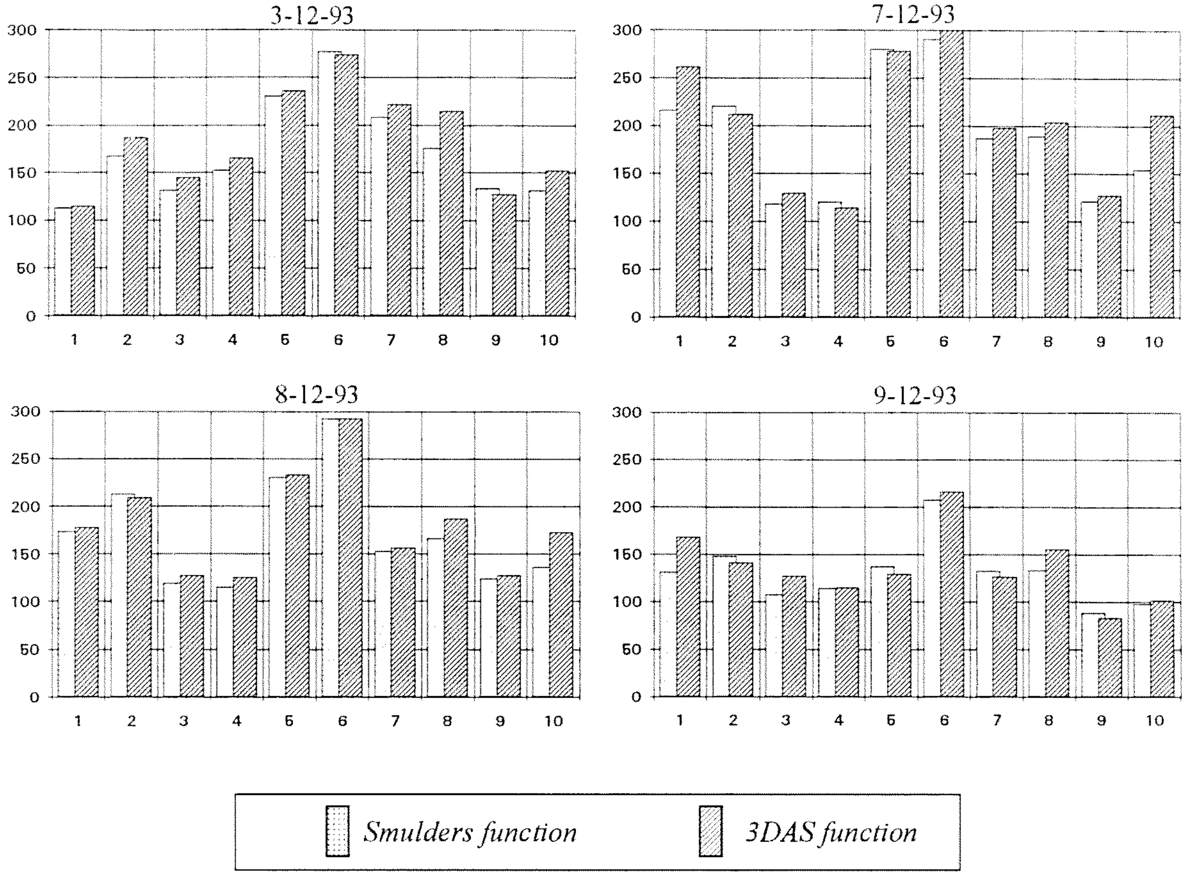


Figure 30: Mean Square Error for ten locations at four different days.

The Smulders function results in most cases in a slightly better match then the 3DAS function. The 3DAS function is however easier to calibrate, and doesn't have the discontinuity. The discontinuity can give problems with the least squares method, because the minimum is not uniquely defined.

9.3.4 Method for adapting the speed-density function

For each link in the network a speed-density function can be derived using historical data. This function is the default function for that link. By using the real-time data from induction-loops the relation between speed and density is measured. This data is used to adapt the function to current conditions. In principle it should be possible to incorporate incidents and weather changes in the model by adapting the speed-density functions with the real-time data. In theory it is possible that a sudden change at a certain location is detected, and the speed-density function is adapted, without any knowledge of the cause of the disturbance (a sudden weather change or an incident).

The function is adapted with a least square method. For every new prediction the function is updated. The last measurement has more influence than the older measurements. The starting point for each minimization is the default (historical) function. The following function F (eq

21) is minimized with a steepest descent method. (Conjugate gradients)

$$Min\ F = \sum_{t=p}^{P} w_t q_t [f(\rho_t) - v_t]^2 \qquad \text{(eq 21)}$$

In which w_t represents the weight for the measurement at time t. Recent data has more weight than older data. q_t represents the flow at time t, $f(\rho_t)$ represent the speed-density function value at ρ_t. The matching measured speed at time t is represented by v_t. Summarized is for period p until the current period P, for a total of 12 periods ($p = P$-12). The parameters to be solved are the parameters of the 3DAS function, α, β and v_{max}.

10. CONCLUSION

The 3DAS model can predict a realistic flow in a network. There are no constraints on the network definition, or the time period length. The continuity of flow and the conservation of traffic is maintained.
The model can simulate traffic jams and the effects up- and downstream. The relation between link occupancy and traveltime is modelled with speed-density functions, instead with intensity- traveltime functions. A situation with congestion —low intensity and high traveltime— is included this way.
The effects of Dynamic Traffic Management instruments, such as: "rerouting and (network coordinated) ramp-metering, speed control and tidal-flow can be investigated.
The model has been tested by comparing small networks with the results of a micro simulation model. Several parameters have been calibrated.
To incorporate all kind of external influences in the model the speed-density functions should be calibrated. It is possible to calibrate these functions with data from induction-loops. Fro real-time applications an easily adaptable function is required which is introduced as the 3DAS function.
Furthermore it is shown that there are significant differences for speed-density relations at different locations. A merging section shows a different relation than an exit-ramp. For each type of link a different speed-density function is required. A list of parameters for different locations seems in demand.

REFERENCES

[1] Updates to COG/TPB Travel Forecasting Procedures, Highway Applications, 1990 Validation and 2010 Simulation, DRAFT, March 1993.

[2] Bureau of Public Roads, "Traffic Assignment Manual", U.S. Department of Commerce, Urban Planning Division, Washington D.C.

[3] Highway Research Board, "Highway Capacity Manual, *Special Report 87 (1965).*

[4] D. Branston, "Link Capacity Functions: A Review", *Transportation Research, volume 10, pp 223-236, 1976*

[5] B. Ran, D.E. Boyce, L.J. LeBlanc,"An Instantaneous Dynamic User-Optimal Traffic Assignment Model and Solution Algorithm", submitted to *Operations research,* 1990.

[6] B. Ran, N. Rauphail, A. Tarko & D.E. Boyce, "Towards a set of Dynamic Link Traveltime Functions for Dynamic Traffic Assignment", *39th North American Meeting of the Regional Science Association Inter-*

national, 13-15 November 1992, Chicago Illinois USA

[7] D.E. Boyce, Bin Ran, L.J. LeBlanc, "Dynamic User-Optimal Traffic Assignment Model: A New Model and Solution Technique", submitted to *Transportation Science,* 1991.

[8] Byung-Wook Wie, "Dynamic Analysis of User Optimized Network Flows with Elastic Travel Demand", *Presented at the 70th Annual Meeting of the Transportation Research Board, Washington DC., USA 13-17 January.*

[9] E. Cascetta, G.E.Cantarella, M. Di Gangi, "Evaluation of Control Strategies trough a Doubly Dynamic Assignment Model", *Presented at the 70th Annual Meeting of the Transportation Research Board, Washington DC., USA 13-17 January.*

[10] K.B. Davidson, "A Flow Traveltime Relationship for use in Transportation Planning", *Proceedings 3rd Australian Road research Board Conference, Part 1, 1966, pp 183-194.*

[11] D.C. Gazis, "Car following theory of Steady State Traffic Flow", *Operations Research, No 7. pp 499-505, 1961.*

[12] H.J.M. van Grol, A.F. Bakker, "Shortest path finding: from algorithm to special purpose hardware", *Proceedings of Seminar C, "Transportation Planning Methods" at the 17th PTRC Transport & Planning Summer Annual Meeting, University of Sussex, England, 11-15 September 1989.*

[13] H.J.M. van Grol, A.F. Bakker, "A Special-Purpose Parallel Computer for Traffic Simulation", *Transportation Research Record 1306, pp 40-48, 1991.*

[14] H.J.M. van Grol, "Traffic Assignment Problems solved by Special Purpose Hardware with Emphasis on Real Time Applications", *Ph.D. thesis, ISBN:90-9005441-3, 1992.*

[15] R. Hamerslag, P.C.H. Opstal, "A Three-dimensional assignment method in Timespace", *Rapport 87-94 Faculty of Mathematics and Informatics, Delft University of Technology, The Netherlands, 1987.*

[16] R. Hamerslag, "Dynamic Assignment in the Three Dimensional Timespace, Mathematical Specification and Algorithm", *USA-Italy seminar on transportation networks, Napoli & Capri, 1989.*

[17] R.Hamerslag, "Dynamic Assignment in the Three Dimensional Timespace", *Transportation Research Record 1220, 1989.*

[18] B.N. Janson, "A Convergent Algorithm for Dynamic Traffic Assignment", *Presented at the 70th Annual Meeting of the Transportation Research Board, Washington DC., USA 13-17 January.*

[19] D.R. Leonard, P. Gower, N.B. Taylor, "CONTRAM Structure of the model", *Research Report 178, TRRL Crowthorne 1989.*

[20] M.J. Lighthill, G.B. Witham, "On Kinematics Waves II. A theory of traffic flow on long crowded roads", *Proc. Roy. Soc., Series A, No 1178, Vol 229, pp 317-345, 1955, London.*

[21] P.G. Michalopoulos & D.E. Beskos, "Improved continuum models of freeway flow", *Ninth International Symposium on Transportation and Traffic theory, VNU Science Press, pp 89-111, 1984.*

[22] M. Papageorgiou, J-M Blosseville, H. Hadj-Salem, "Macroscopic Modelling of Traffic Flow on the Southern Part of Boulevard Peripherique in Paris", *Transportation Research B, Vol 23B, No 1, pp 29-47, 1989.*

[23] M. Papageorgiou, J-M Blosseville, H. Hadj-Salem, "Modelling and Real-time Control of Traffic flow on the Southern Part of Boulevard Peripherique in Paris. Part I: Modelling", *Transportation Research A, Vol 23A, No 5, pp 345-359, 1990.*

[24] B. Przebinda, "Comparison of a view Speed-Density functions in Application to Dynamic Traffic Management.", *A study report from Training in Placement at Delft University of technology under TEMPUS "TREP" Joint European Project. Delft September 1991.*

[25] E. de Romph, H.J.M. van Grol & R. Hamerslag, "A Dynamic Assignment Model for Short Term Prediction", *Capri Seminar on Urban Traffic Networks, 5-8 July 1992, Italy.*

[26] E. de Romph, H.J.M. van Grol & R. Hamerslag, "3DAS -3Dimensional ASsignment- A dynamic assignment model for short term predictions", *39th North American Meeting of the Regional Science Association International, 13-15 November 1992, Chicago Illinois USA.*

[27] G. Rose, M.A.P. Taylor & P. Tisato, "Estimating Travel Time Functions for Urban Roads: Options and Issues", *Transportation Planning and Technology, 1989, Vol 14,pp 63-82.*

[28] C.A. Rothrock & L.A.Keefer, "Measurement of Urban Traffic Congestion", *Highway Research Board, Bulletin 156, pp 1-13, 1957.*

[29] S. Schwerdtfeger, "DYNEMO: A model for the simulation of traffic flow in motorway networks", *Ninth International Symposium on Transportation and Traffic theory, VNU Science Press, pp 89-111, 1984.*

[30] Y. Sheffy, "Urban Transportation Networks", Englewood Cliffs, NJ,: *Prentice Hall, 1985.*

[31] S.Smulders, "Modelling and Filtering of Freeway Traffic Flow", *Report OS-R8706, Centre for Mathematics and Computer Science, November 1988.*

[32] S. Suh, C. Park & T.J. Kim, "A Highway capacity Function in Korea: Measurement and Calibration", *Transportation Research, volume 24A, No 3, pp 177-186, 1990*

[33] P.C. Vythoulkas, "A Dynamic Stochastic Assignment Model for the Analysis of General networks", *USA-Italy seminar, "Urban Traffic Networks: Dynamic Control and Flow Equilibrium", Napoli & Capri, 1989.*

[34] M. Westerman, R.Hamerslag, "Collecting the Dynamics of Metropolitan Freeway Systems: A Simulation Study", *Delft University of Technology 1993.*

APPLICATION OF 3DAS
(3-Dimensional ASsignment)
AT THE WASHINGTON METROPOLITAN AREA

E. de Romph[1]
H.J.M. van Grol[2]
R. Hamerslag[1]

11 November 1993

1. Delft University of Technology, Faculty of Civil Engineering, Department of Transportation Planning and Highway Engineering.
2. Delft University of Technology, Faculty of Applied Physics, Department of Physics Informatics / Computational Physics.

ABSTRACT

This paper describes a study in which the dynamic assignment model -3DAS- is used as a planning tool. As study-area the Virginia part of the Washington Metropolitan area has been chosen. This area offers a heavily congested urban network with several rerouting possibilities. Based on the data available it is decided to calculate a morning rush-hour from 5:00am until 11:00am, in 24 periods of 15 minutes.

The study has two objectives. The primary objective is to find answers to the following three questions:

i. Can *dynamic* assignment be used for planning purposes?
ii. Is there an advantage of *dynamic* assignment above *static* assignment?
iii. Is *dynamic* assignment a useful tool to investigate the effects of Advanced Traffic Management Systems (ATMS)?

The secondary objective is to gain insight in the possibilities and problems associated with the application of 3DAS on larger networks.

To answer these questions, four scenario's are considered. The first scenario tries to achieve a reasonable reproduction of the morning rush hour. The second scenario is a static assignment, to find an answer to the second question. Several inconsistencies of static assignment are solved. The third scenario introduces rampmetering at several on-ramps, and the fourth scenario introduces an accident at one of the freeways. For the latter scenario two situations are considered. One using an equilibrium approach, assuming total knowledge of the drivers. The second situation uses initial traveltimes for the section with the accident, resulting in no knowledge of the accident for any driver.

The results show that using dynamic assignment for planning purposes can be very helpful. The results give more detailed information about the occurrences of traffic jams then static assignment methods can give. A more precise location or the cause of congestion can be identified. Introducing ATMS measures to alleviate the congestion can be simulated, and all kinds of evaluations are possible. Influences on traveltime and jam-length, effects of rampmetering and rerouting can be investigated.
The accuracy of the data is however more demanding, and the computers need more calculation time. Also very important is the ability to visualize the results. With dynamic assignment the results are flows *in time*. The best way to analyse the results is as a *film*. In order to do that a graphically powerful computer is needed.

In order for Advanced Traffic Management Systems to be successful there is a large demand for more data and better (3D) OD-matrices. Hopefully new methods for OD-estimation and more data from induction-loops and probe-vehicles will improve the reliability of the results in the future.

1. Introduction

A *dynamic* assignment model can be useful for several applications: Traffic Control Systems, Advanced Traffic Management Systems, Advanced Drivers Information Systems, and Planning Systems. Each application field has different demands concerning data requirements and detail. Planning Systems, which is a typically long-term application, has the least high requirements concerning accuracy of the data and detail of the study area. Traffic Control Systems have the highest data requirements [7].

This paper describes a study in which the dynamic assignment model -3DAS- is used as a planning tool. The study has two objectives. The primary objective is to find answers to the following three questions:

i. Can *dynamic* assignment be used for planning purposes?
ii. Is there an advantage of *dynamic* assignment above *static* assignment?
iii. Is *dynamic* assignment a useful tool to investigate the effects of Advanced Traffic Management Systems (ATMS)?

The secondary objective is to gain insight in the possibilities and problems associated with the application of 3DAS on large networks.

The model is applied on the south-western part of Washington DC in the USA. This area is chosen because it offers a heavily congested urban network, with several rerouting possibilities. Nowadays there are several ramp-metering installations in operation, and parts of the freeways are monitored.
The data used for this research is obtained from Virginia Department of Transportation (VDOT), and Metropolitan Washington Council of Governments (COG). A small portion of the study-area is monitored by induction loops. One minute data from these induction loops is used to derive the departure functions and to validate the calculation results.

The research is conducted during a four-month visit at the "Center for Transportation Research" at "Virginia Polytechnic Institute & State University (Virginia Tech)". As follows from the objectives, this study is only meant as an example of using dynamic assignment as a planning tool. Due to a lack of data, and the short study time, the results are not suitable for any serious planning decisions. The results are only suitable for decisions concerning the *usefulness* of 3DAS for planning purposes.

The paper will briefly discuss the 3DAS model, the research approach and how the data is derived. Apart from a static assignment, three different scenario's are calculated; a morning rush-hour, a scenario with several ramp-metering installations, and a scenario with an incident. The results of the model for these scenario's are reported.

2. The 3DAS model

The 3DAS model is based on the work carried out by Hamerslag as described in: "Dynamic Assignment in the Three Dimensional Timespace" [1,2,3]. The basic feature of a dynamic assignment model is the partitioning of time in small time-slices, usually referred to as periods.
Over the last two years the model has been improved. In particular towards its dynamic

aspects. The 3DAS model has been described in two papers by E. de Romph & H.J.M. van Grol [4,5], and in the dissertation of H.J.M. van Grol [6].

The model determines the flow distribution on the network in an iterative process. In each iteration the shortest-paths in the network are calculated for all OD-pairs and every departure period. The link parameters are defined separately for each period. The properties of the network and the travel demand are presumed given.

The basic iteration scheme in figure 2 is in principal the same as for *static* assignment models. The difference is the "all-or-nothing-in-time" module. In this module an extra iteration over the departure period is needed, and the shortest pathfinding and the assignment have to be performed *in time*.

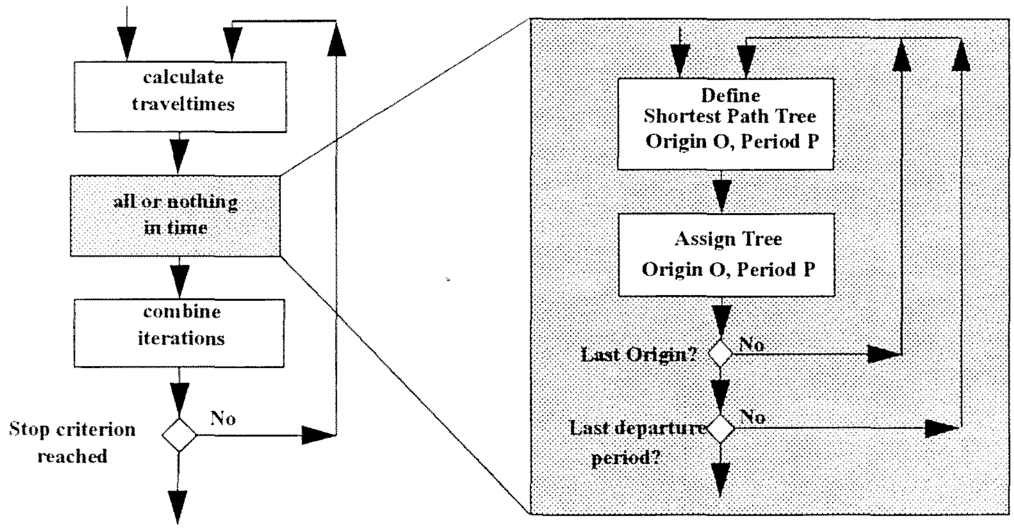


Figure 1: The iteration scheme

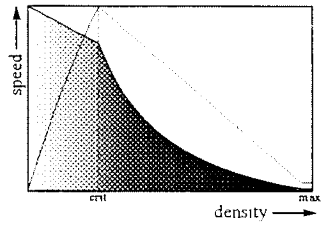
The paths are defined using the traveltime on a link in the period in which the traffic actually traverses the link, i.e. the trajectory the traffic follows in time is calculated. The network is loaded, based on these trajectories. During the assignment the contribution of a traveller to the traffic-load on a link in a certain period is determined by calculating the duration of presence on that link in that period. If we focus on one traveller then two situations can occur:

1. *Several links are covered in one period.* In this case the traveller is only for a part of the period present on the link, and therefore should only be assigned for this part.
2. *One link is covered in several periods.* The traveller is present on the link during multiple periods and should be assigned entirely for each individual period.

At the start of each iteration the traveltimes on the links are derived from the load of the previous iteration. For each link the traveltime is calculated with a speed-density function. Speed-density functions are used instead of traditional traveltime-intensity functions to be able to model a decreasing flow in case of congestion. The conservation of traffic and the continuity of flow is maintained. In case of overflow, the overflow is assigned to the preceding link of the path in the same period. The stop criterion is reached when there is no difference between two successive iterations.

The 3DAS model has been tested on several small networks [4]. Several parameters of the model were calibrated using these networks. The initial settings of these parameters followed from these tests and were not changed for this study. A speed-density function is used with the following form:

$$\rho) = \begin{cases} v^{max} \cdot \left(1 - \frac{\rho}{\rho^{max}}\right) & 0 < \rho < \rho^{crit} \\ v^{max} \rho^{crit} \cdot \left(\frac{1}{\rho} - \frac{1}{\rho^{max}}\right) & \rho^{crit} < \rho < \rho^{max} \end{cases} \qquad \text{(eq 1)}$$



In which v^{max} is the free-flow speed, ρ^{crit} is the critical density and ρ^{max} is the maximum density. The maximum density represents a no-motion traffic-jam,

3. Research Approach

In order to satisfy the objectives stated in the introduction the following research approach has been set up. The first objective consists of three questions:

Can dynamic assignment be used for planning?

Dynamic assignment can be used for planning if, given a network and a traffic demand, it can predict a correct distribution of traffic flow. Since for long term purposes the traffic demand will represent the average demand, the expected traffic *distribution* will also be average. This is in contrast with *real-time* applications when the results should be based on the actual situation of that moment. Including temporary disturbances such as incidents or roadworks.

To validate the model, the average traffic demand is required and a measured traffic distribution averaged over a longer period.

Is there an advantage of *dynamic* assignment above *static* assignment?

There are several (well-known) problems with static assignment models. A static assignment model:

- can give wrong results in case of congestion. Because traffic is assigned along the complete route, a car can contribute to more than one congestion at the same time.
- cannot correctly show the effects of a variable traffic demand.
- cannot correctly show the effects of temporal disturbances, such as road works or accidents.
- cannot predict queue-lengths, and how a growing queue can limit the capacity of upstream junctions.

Studied is whether dynamic assignment can solve these inconsistencies, and how it will improve the decision making for planning.

Is dynamic assignment a useful tool to investigate the effects of Advanced Traffic Management Systems?

The model has been extended to model several ATMS instruments, such as ramp-metering, rerouting and tidal-flow. To answer the question two tests are executed. The first scenario considers several ramp metering installations, and the second scenario considers an accident at one of the freeways. For the second scenario the effects of diversion measures are reported.

Since the network used for this study is fairly large, the secondary objective of this research, which is to gain insight in the possibilities and the problems associated with dynamic assignment applied on larger networks, is also satisfied with the above research approach.

4. The data

The study area covers the eastern part (Virginia part) of the beltway at Washington DC. The major interstates are the I-95, I-395, I-66 and the I-495, also a large part of the arterial network is included.

4.1 The Network

Figure 2 represents the network used for this study. The network consists of 857 nodes and 2086 links. There are 180 zones. Most freeway intersections are represented fairly detailed. An example of two of these intersections is given in detail. Each line in the figure depicts a separate one-directional road consisting of one up to four lanes.

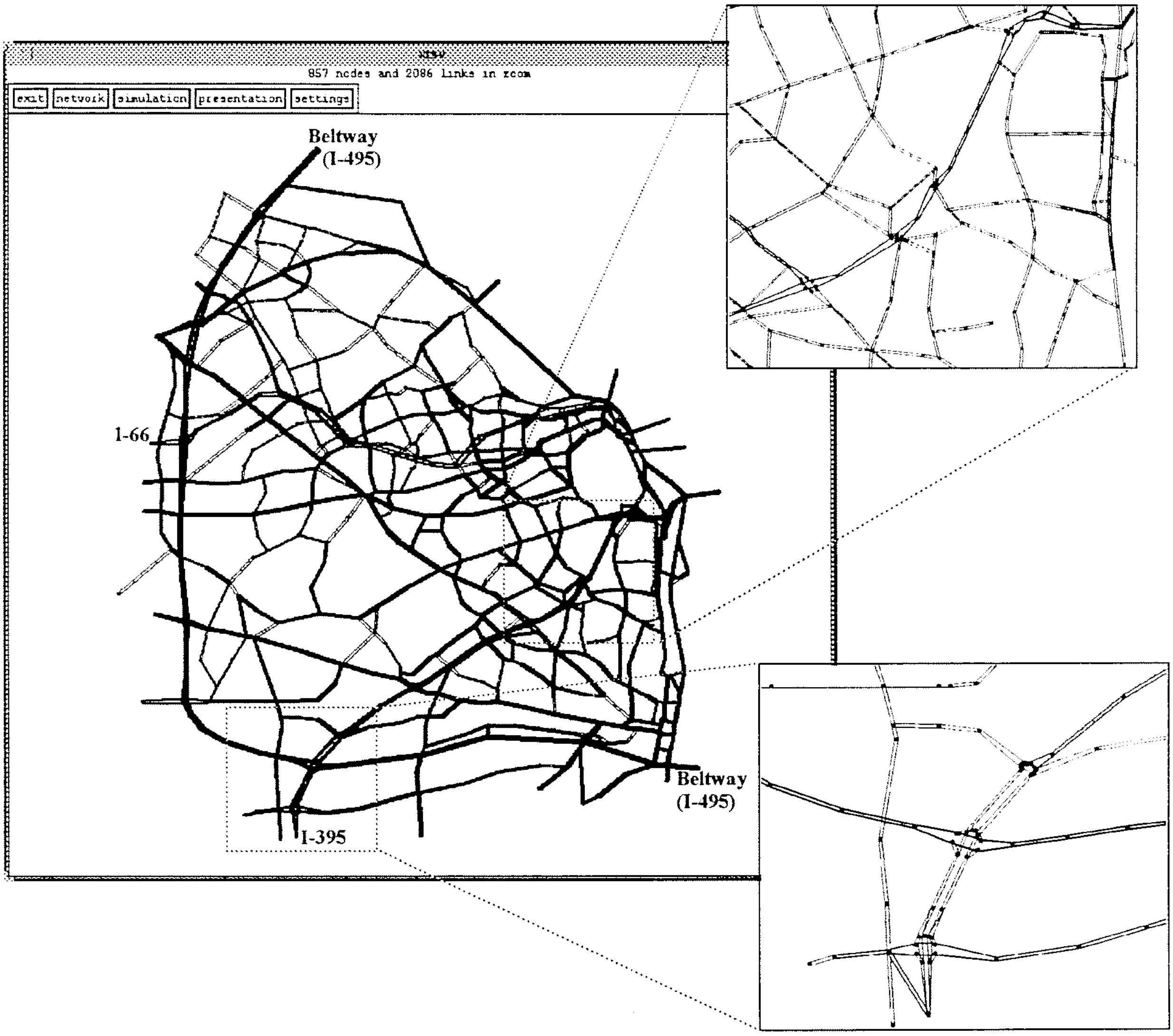


Figure 2: The study network

The 2086 links are divided in 13 types. Each type represents a certain road type. All the links in one type have the same attributes, which are given according to Table 1. The attribute for the capacity is an attribute which is not given but derived from the maximum density, the maximum speed, and the speed-density function.

Table 1: Network Characteristics

Type	Description	Max Speed (km/hr)	Max Dens (veh/km)	Capacity (veh/hr)
1	Beltway, 4 lanes	100	440	7250
2	3 lanes part of the I-395	100	330	5800
3	2 lanes part of I-395 and 2 lanes part of I-66	100	224	4390
4	3 lanes part of the I-66	100	330	5800
5	4 lanes part of the I-395	100	440	7250
6	freeway intersection (right turn)	70	110	1460
7	freeway intersection (left turn = circle)	40	115	850
8	bridges in centre over the river Potomac	50	224	2200
9	on- and off ramps	50	115	1060
10	Unused road type	n.a.	n.a.	n.a.
11	1 lane arterials	35	112	730
12	2 lanes arterials	60	224	1750
13	3 lanes arterials	65	336	2740
14	4 lanes arterials	65	448	4000

4.2 The OD-matrix

The network is not accompanied by a matching *dynamic* OD-matrix. This OD-matrix has to be constructed from other data sources.

The best OD-matrix available is a (static) 24-hour matrix which covers a much bigger area. The OD-matrix is the result of a study done by the *Metropolitan Washington Council of Governments* (COG). The OD-matrix for the study area has to be extracted from this OD-matrix. To make the OD-matrix dynamic, departure functions are used. A departure function gives for each OD-pair the portions of the amount of traffic departing in each period. These departure functions can be estimated and calibrated with link volume data.

The big network covered 5983 nodes and 18104 links.

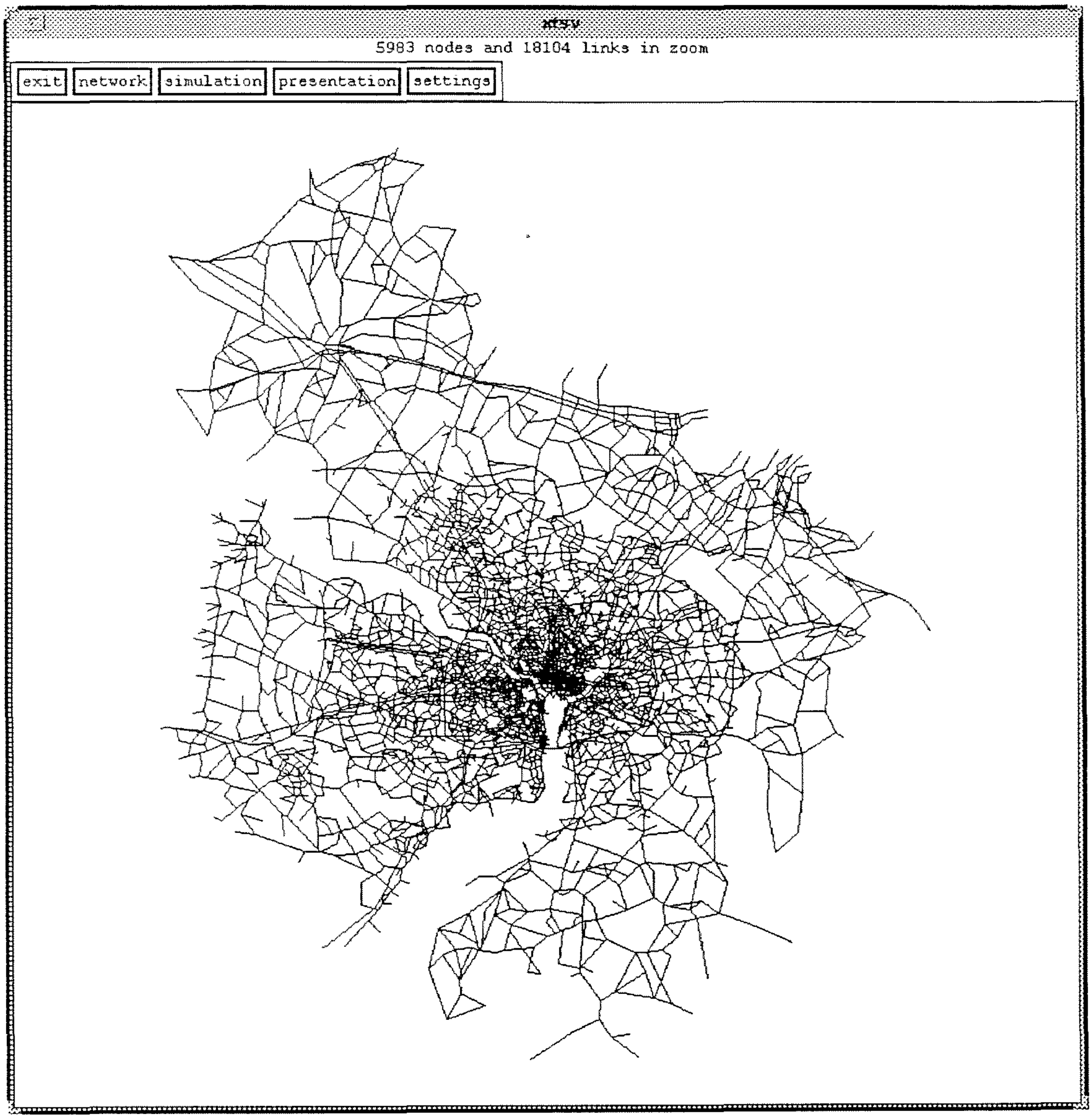


Figure 3: The total network

The COG study [8] was done with 1990 as a base year, and comprised of 293 districts (1478 zones) which cover the following jurisdictions:

- District of Colombia, (DC)
- Montgomery County, MD
- Prince George's County, VA
- Arlington County, VA
- Alexandria, VA
- Fairfax County, VA
- Loudoun County, VA
- Prince William County, VA
- Frederick County, MD
- Howard County, MD
- Ann Arundel County, MD
- Charles County, MD

VA = Virginia, MD = Maryland.

The model used by COG for the trip generation, distribution and mode choice was a gravity model and calculated at district level. The districts were then "split" to zones via land-use factors. For productions these land-use factors were based on households and groups-quarter population. For attractions they were based on office, retail, industrial and other employment. The resulting OD-matrix had 1478 zones.

The network used for this study (figure 2) is only a small part of this large network, so the OD-matrix for the smaller network (180 zones) has to be derived from the big OD-matrix (1478 zones). When we zoom in, the study-network becomes recognizable. All trips made *within* the study-network are easily derived. All trips entering, leaving or passing through the study-network are derived with a selected link analysis. All the links marked with a dot are selected links.

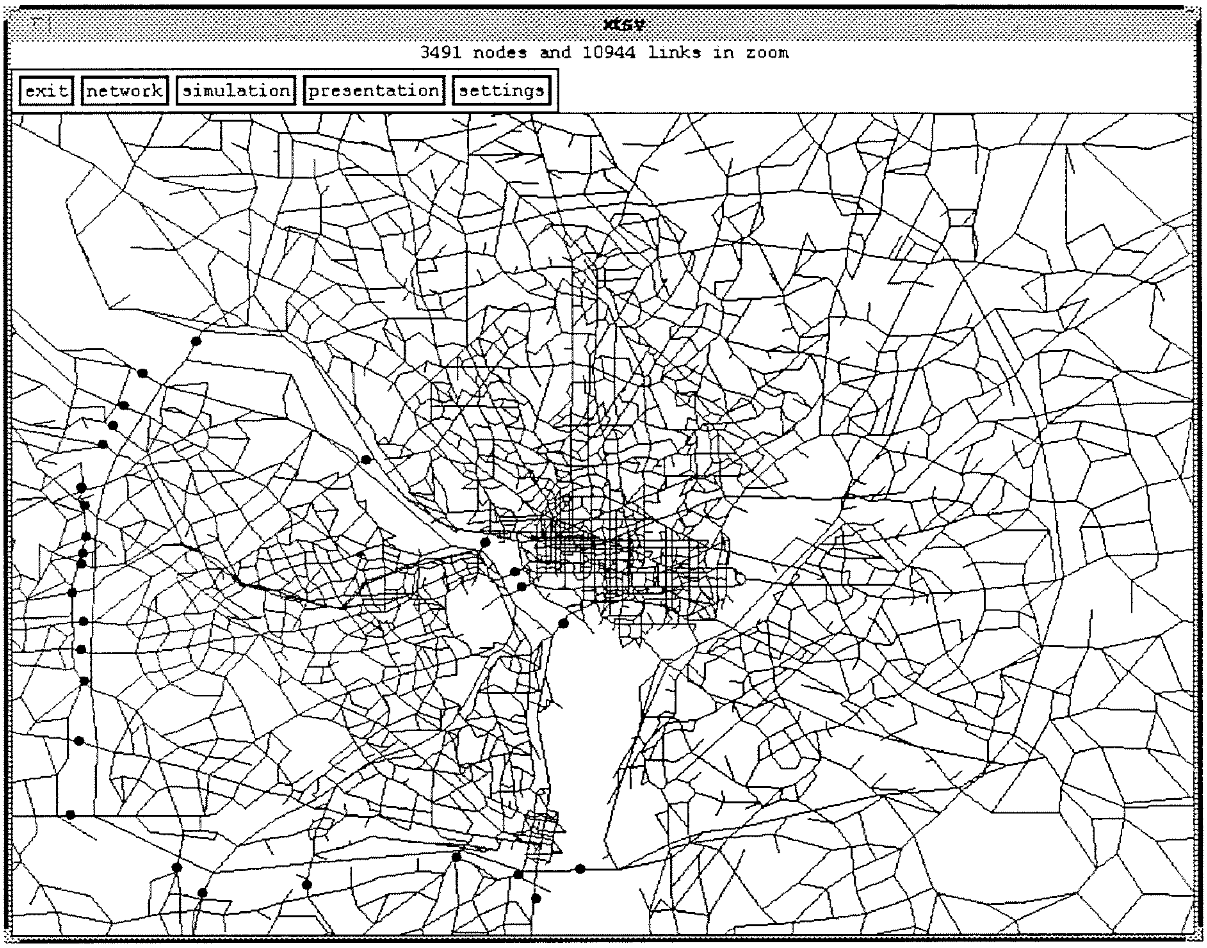


Figure 4: Zoomed in part of the big network

To perform the selected link analysis, the OD-matrix is assigned to the network with a static "All-or-Nothing" assignment. The shortest path-finding is done based on the *actual* speeds in the network. These *actual* speeds are derived from the static assignment done by the Washington Council of Governments. For all OD-pairs crossing the selected links, the origin and the destination are stored. The entering and exiting links become new origins and destinations, and the trips are summed. With this method all the entering and exiting traffic is aggregated to the links in which they exit or enter the subnetwork.
Deriving an OD-matrix for the subnetwork using this method has one major drawback. Because an All-or-Nothing assignment is used, no alternative routes are chosen for OD-pairs. In order to minimize the effects of this problem some links are added to the subnetwork to allow a diversion for some origins to different links to enter the network.

4.3 Induction-Loop data

The Northern Virginia Traffic Control Centre controls a part of the freeway system in Northern Virginia. The covered freeways are the I-66 and the I-395. These freeways are equipped with several hundreds of induction loops. One minute data of a fixed portion of these induction loops can be downloaded on a datatape. Unfortunately the Traffic Control Center is not yet fully equipped, and the downloading of data from induction loops is therefore not easy. Only one tape (one day) was available for this research. Although the traffic-patterns of this one day are not sufficient to derive any statistical information, it was the best data available.
The tape used for this study was measured on Monday the 7th of December 1992 from 4.00 pm until 11.00am the next day. Every minute the number of passed vehicles are registered and downloaded. This tape was available for research purposes. Unfortunately the tape contains many errors and at several locations only one of the three or four lanes is read. Nevertheless the data of several locations are useful for this research.
The following graphs give an impression of the traffic pattern at several locations on the I-66 and the I-395. The x-axis shows the time in hours. The registration started at 18:00 hours and lasted until 11:00 hours the next day. The y-axis shows the flow in vehicles per hour. For each freeway and each direction two graphs are shown. The first graph is located somewhere at the beginning of the freeway, the second graph is located somewhere near the end of the freeway. The locations of the induction loops are displayed in figure 9 for the I-66 and in figure 10 for the I-395.

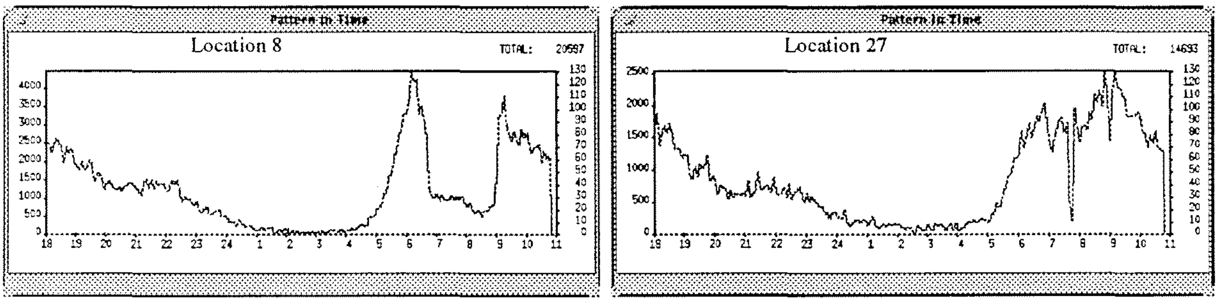


Figure 5: The traffic flow from 18:00h until 11:00h (next day) for locations 8 and 27 on the I-66 Eastbound.

The graphs in figure 5 clearly show that the rush-hour starts at ±5.00am. At location 8 the flow increases in approximately one hour to 4500 veh/hr. At ±6:00am some kind of congestion occurs and the flow drops rapidly (possibly an incident). After ±9:00am the flow increases again. The end of the rush-hour is at approximately 11:00am. At location 27, which is further downstream the I-66, the flow increases to ±2500 veh/hr.

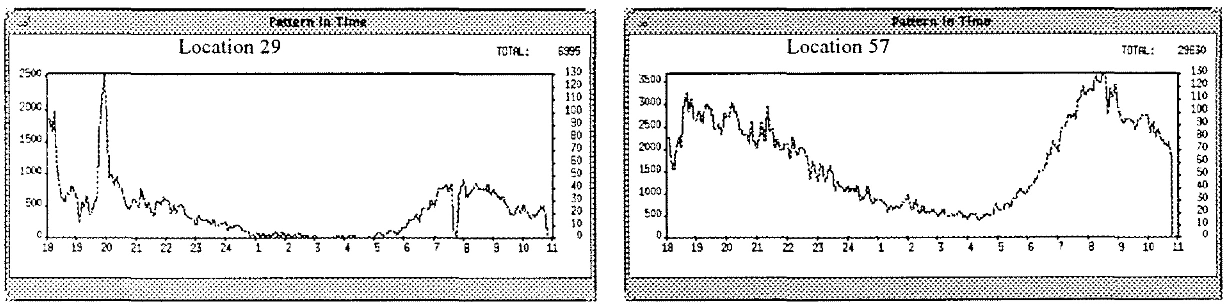


Figure 6: The traffic flow from 18:00h until 11:00h (next day) for locations 29 and 51 on the I-66 Westbound

The two graphs in figure 6 show that the flow on the I-66 westbound is significantly lower, and that no congestion occurs in this direction. At location 29 the flow increases to ±2000 veh/hr. At location 51 the flow increases to ±3500 veh/hr.

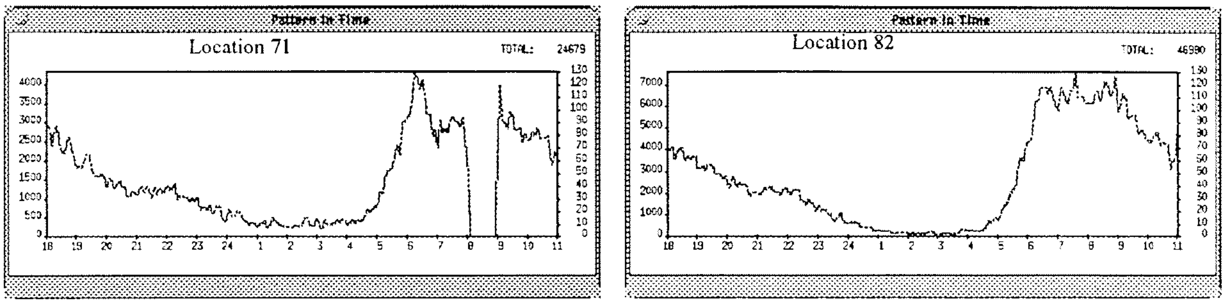


Figure 7: The traffic flow from 18:00h until 11:00h (next day) for locations 71 and 82 on the I-395 Northbound

On the I-395 (figure 7) roughly the same pattern is found as on the I-66. The rush-hour starts at 5.00am. At location 71 the flow decreases after ±6:00am due to congestion. Between 8:00am and 9:00am there is probably an error in detection. Further downstream the I-395 at location 82 no congestion occurs. The flow increases to 7000 veh/hr (4 lane section).

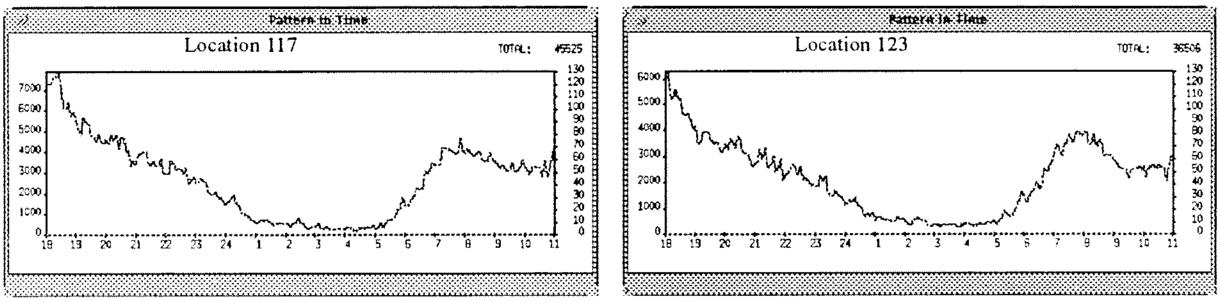


Figure 8: The traffic flow from 18:00h until 11:00h (next day) for locations 117 and 123 on the I-395 Southbound

In this direction on the I-395 there is no congestion. At location 117 the flow increases to ±4000 veh/hr, and at location 123 to ±4000 veh/hr as well.

The graphs in figure 7 shows higher flows in the morning rush-hour than in the evening rush-hour, figure 8 shows the opposite, higher flows in the evening rush-hour than in the morning rush-hour. This is explainable due to commuting traffic.Commuting traffic enters the city in the morning and leaves in the evening. On the I-66 (figure 5 & 6) show the same pattern, although not as clear.

Based on these data, it is decided to calculate a morning rush hour from 5.00am to 11.00am. This time-period captures the total morning rush-hour, and the graphs show that before 5:00am the network is still reasonably empty. This gives the advantage that the calculations can be started with an empty network.

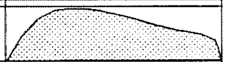
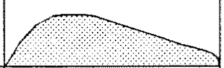
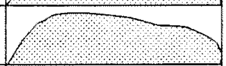
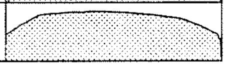
4.4 Departure functions

To use a static OD-matrix as a substitute for a dynamic OD-matrix, departure functions are required. A departure function is a discrete function, which determines for each period how much percent of the OD-value departs during that period. To derive these departure functions,

induction-loop data can be used.
One departure function for all OD-pairs will not give a realistic representation of the dynamic OD-matrix for the rush-hour. The departure functions of individual OD-pairs can be quite different. The graphs in paragraph 4.3 show that traffic on the I-66 travelling westward (figure 6), and traffic on the I-395 travelling southward (figure 8), have a lower volume in the morning rush-hour and a higher volume in the evening rush-hour, and depart according a different departure function. This requires at least different departure function for traffic going in to Washington and traffic leaving Washington.
For this study the OD matrix is split into four major trip types. For each type a different departure function is used. The different types are described in Table 2.

In total the 24-hour OD-matrix contains 293930 trips. From the induction loop data it can be derived that during the rush-hour, from 5:00am until 11:00am, approximately 35 percent of this matrix departs. The total surface of the departure function determines the percentage of the OD-matrix. So for each type the total percentage of travellers that departs for that type is influenced by the total surface of the function.
Table 2 gives the total amount (trips) and the surface of the departure function (%) and the form of the function.

Table 2: Departure Tables

Type	Description	Trips	%	Form
1	Entering Washington or passing through	88625	36.9	
2	Entering the Washington on the freeways (I-66 from the east and I-395 from the south)	88527	13.9	
3	Exiting Washington on the freeways (I-66, I-395)	19625	33.4	
4	Internal traffic within the network	97153	35.0	

For type 1 the total number of cars entering Washington from the east and the south (or passing through) is 88625 cars (for the total 24-hour period). From this total 36.9% is used to assign during the rush-hour period. They depart their origin according to the departure function in the table.
The table shows a low percentage for the cars entering the network on the I-66 (east) and the I-395 (south). The reason for this percentage is a large overflow on these freeways. With a percentage as low as 13.9% a reasonable reproduction of the induction loop data is achieved. This overspecification of the flow in the OD-matrix is partly due to the fact that an All-or-Nothing assignment has been used to derive the OD-matrix, and partly to a possible overspecification in the original matrix.

Using only four different departure functions will give a rather rough reproduction of the traffic patterns. This choice has been made due to lack of data to derive more departure functions and to lack of data to evaluate the results.

For estimation of departure patterns more data and maybe some new approaches are desirable.

5. Modelling and Calibration

Based on the data available it is decided to calculate a morning rush-hour from 5:00am until 11:00am, in 24 periods of 15 minutes. The total time-span of six hours captures the total morning rush-hour. The period-lenght of 15 minutes is chosen for practical reasons to keep the calculation time in bounds. A period length shorter than 5 minutes is not recommended because it suggest a detail which cannot be reached with the available data. A period-length longer than 20 minutes dissipates the dynamics in the assignment of traffic to much.

To satisfy the objectives of this research four scenario's are considered.

I. The first scenario is meant to achieve a reasonable reproduction of the morning rush hour. The departure functions are calibrated using induction-loop data, and the resulting flows are compared with the induction-loop data. By adapting the departure functions it is possible to reproduce the induction loop data at the beginning of a route. In case the flow pattern further down-stream that route still matches the induction-loop data, this is considered a good result. The flow-pattern can be tested at several locations on: "the form in time, the average height of the flow, and the moments of sudden changes in the flow". Since we have only one day of induction-loop data and no information about weather or incidents, this data does not represent an average flow pattern. Only a rough reproduction of volume patterns can be expected.

II. The second scenario is a static equilibrium assignment, comparing the results with dynamic assignment. The advantages and disadvantages of time variation are studied.

III. The third scenario introduced rampmetering at all ramps on the I-66 Eastbound and all ramps at the I-395 Northbound. The influences on queue-length, traveltime, and diversion behaviour are investigated.

IV. The fourth scenario introduced an accident at one of the freeways (I-66). For this scenario two different situations are calculated. In the first situation the drivers are unaware of the accident, by using initial traveltimes for the section with the accident. In the second situation the drivers are assumed to be fully informed. Here an equilibrium assignment is used.

The third and the fourth scenario investigate the possibilities of dynamic assignment for Advanced Traffic Management Systems. The input data used for these scenarios is the same as for the morning rush-hour scenario (I). The departure functions and the OD-matrix are unchanged.

6. Hardware & Software

For the calculation the model is implemented as an X-windows program under the UNIX operating system. Several different computers were used to run the program, whichever computer was available. Either a Silicon Graphics 320VGX or INDIGO, an IBM RS6000, or SUN Sparc2. For the research the computers at the "Laboratory for Scientific Visualization" at Virginia Tech were available.

On the Silicon Graphics 320VGX one iteration took approximately 5 minutes. In one iteration all the OD pairs are assigned to their time dependent shortest paths. For this study with 180 zones and 24 departure periods, this results in 28835 OD-relations per period ($180^2 = 32400$; not all OD-pairs exist!). For 24 periods these are 692040 OD-relations.

Since it is impossible to study numbers on paper, the visualisation of the results is very important. The software (3DAS) is a program that displays several results in graphics. For each link the pattern in time can be investigated, and to get an overall impression of the traffic flows, build up of traffic jams, etc., a presentation as a "film" is implemented. Errors in the input or other anomalities are easily found using a good graphics representation.

7. The results

7.1 The Morning Rush-hour scenario

Based on the OD-matrix, the departure functions and the network attributes, a dynamic assignment is done. Heavy congestion is found on the I-66 and the I-395 going into Washington. Small congestions are found at several locations on the beltway, and at certain arterials. The "film" representation shows quite clear where the congestion starts, and how it evolves. To give an impression of the results the flow patterns at two locations along the I-66 (figure 9) and along the I-395 (figure 10) are shown.

The graphs in figure 9 and figure 10 show the *intensity* (flow) at four different locations on the I-66 and the I-395. The x-axis shows the time and the y-axis shows the intensity. Each bar represents a time-period of 15 minutes. The height of the bars shows the intensity, while the colour of the bars shows the density. Light grey represents a low density and dark grey represents a high density. Using this representation the difference between a low intensity caused by a high density (dark grey) and a low intensity caused by a low density (light grey) can be discriminated.

The graphs in figure 9 represent the same locations on the I-66 as the induction loop graphs in figures 5 and 6 in paragraph 4.3. For figure 10, which represents the I-395, the matching induction loop graphs are figure 7 and 8.

Comparing the graphs in figure 9 and 10 with the induction loop graphs (figure 5,6,7 and 8) shows that a reasonable reproduction of the traffic distribution is possible. On the I-66 eastbound however the induction loop data shows heavy congestion, with a low flow (almost zero). Based on the low flow downstream, we may assume that there was probably some kind of incident during that day. In the simulation a much higher flow downstream is found. If there really was an incident the differences between the model and the induction loop data are explainable. The flow-patterns on the freeways are compared with the induction loop data at more places then shown in figure 9 and 10.
Since there is no induction loop data available of the arterial system, no validation for these sections is possible.

7.2 The Static Assignment Scenario

The 3DAS model is compared with a static user-equilibrium assignment. For the static assignment the results show heavy overload on the entire I-66 and I-395. The results do not show where the congestion starts, i.e. where the bottleneck is located. Especially the down-stream occurrences of overflow do not exist due to upstream congestion. As an example on the I-66 the *dynamic* assignment shows congestion halfway this freeway, downstream a free-flow situation exists. The *static* assignment shows an (incorrect) congestion situation downstream as well. This might results in wrong planning decisions.
Obviously the static assignment does not give any insight in the *development* of the rush hour

as it evolves, and it does not give information on the length of traffic jams and how they evolve.
The results of a static assignment are on the other hand much easier to interpret. The amount of data produced by a static assignment is much less. It takes a shorter time to evaluate the results of a static assignment.

I-66 Westbound

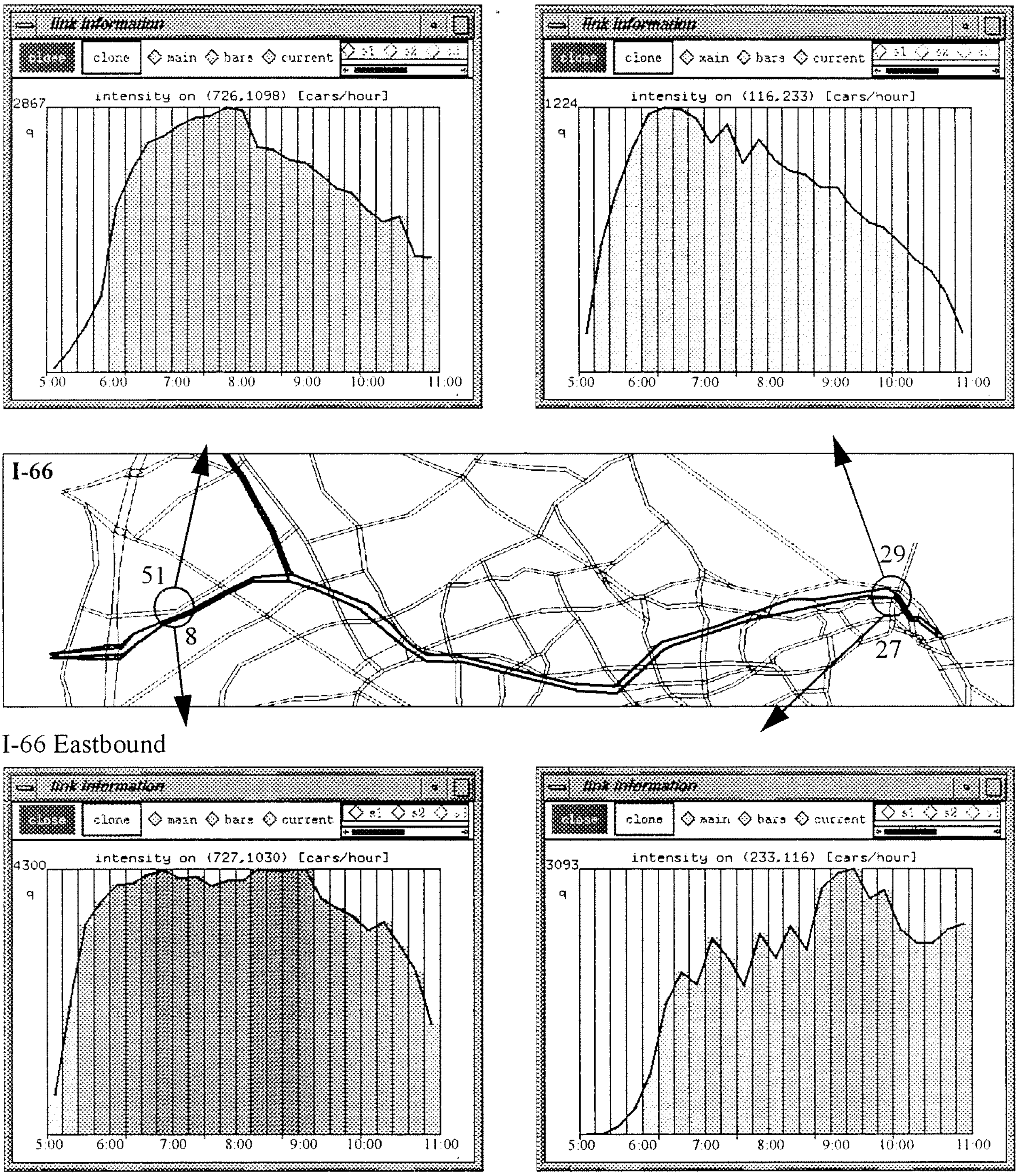


Figure 9: The intensity (flow) calculated by 3DAS for the rush hour from 5:00am to 11:00am at 4 locations (51,8,29,27) on the I-66.

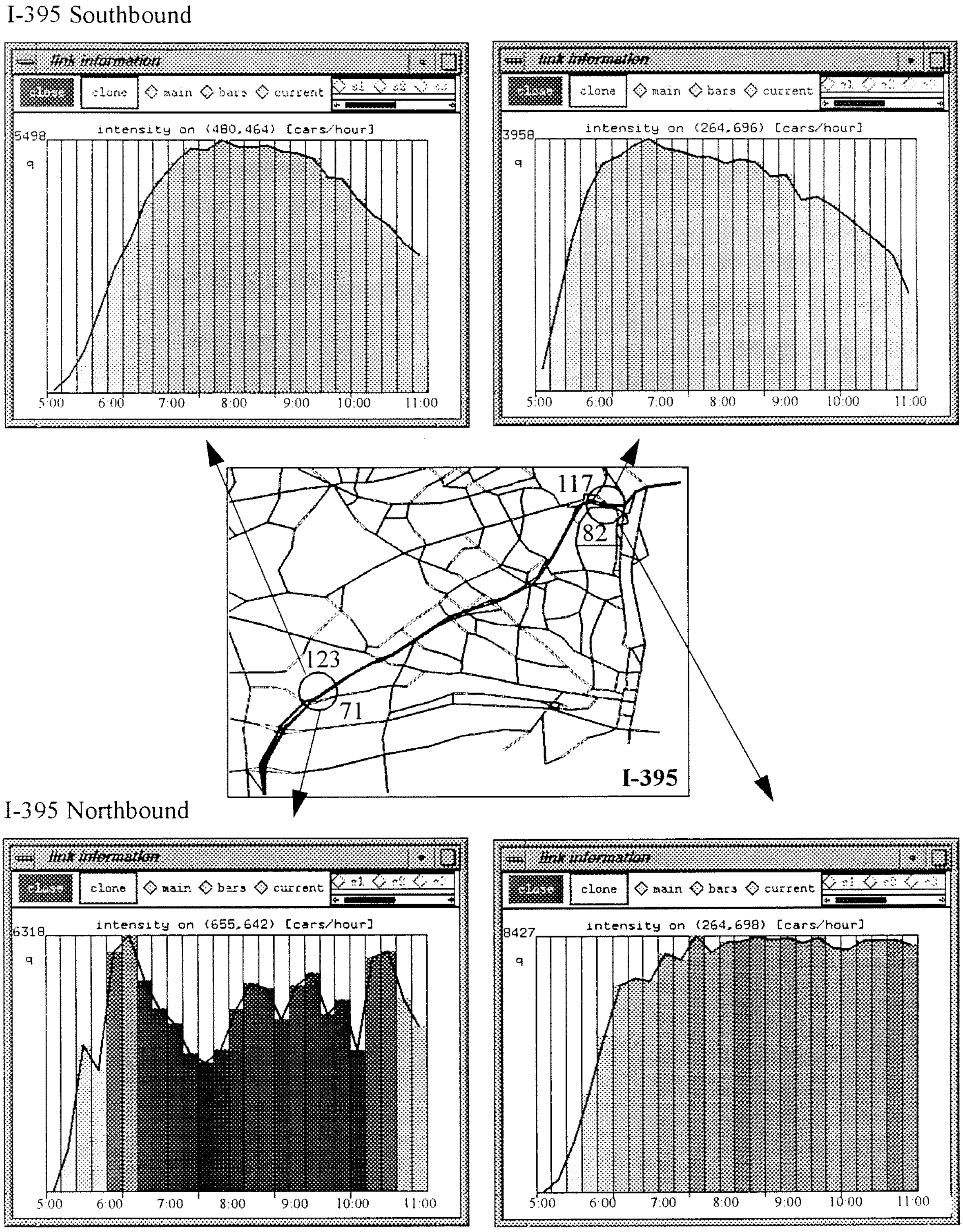


Figure 10: The intensity (flow) calculated by 3DAS for the rush hour from 5:00am to 11:00am at 4 locations (123,71,117,82) at the I-395:

7.3 The Advanced Traffic Management Scenario

To test whether ATMS strategies can be investigated with 3DAS, a scenario with rampmetering is created. Rampmetering is introduced on all on-ramps on the I-66 eastbound, and all on-ramps on the I-395 northbound. The ramp-metering is implemented as a simple maximum flow limit for all on-ramps. Since the shape of the speed-density functions dictates the maximum flow (i.e. capacity), the maximum flow limit is achieved by using different speed-density functions.

The graphs in figure 11 and 12 compare the ramp metering scenario with the normal scenario. The graphs show the *speed* as height of the bars and the density as colour of the bar for the ramp metering scenario. The speed results for the normal rush-hour scenario are displayed as a solid line in the same graph. The x-axis shows the time.

On the I-66 a location upstream and a location more downstream are selected to display the results:

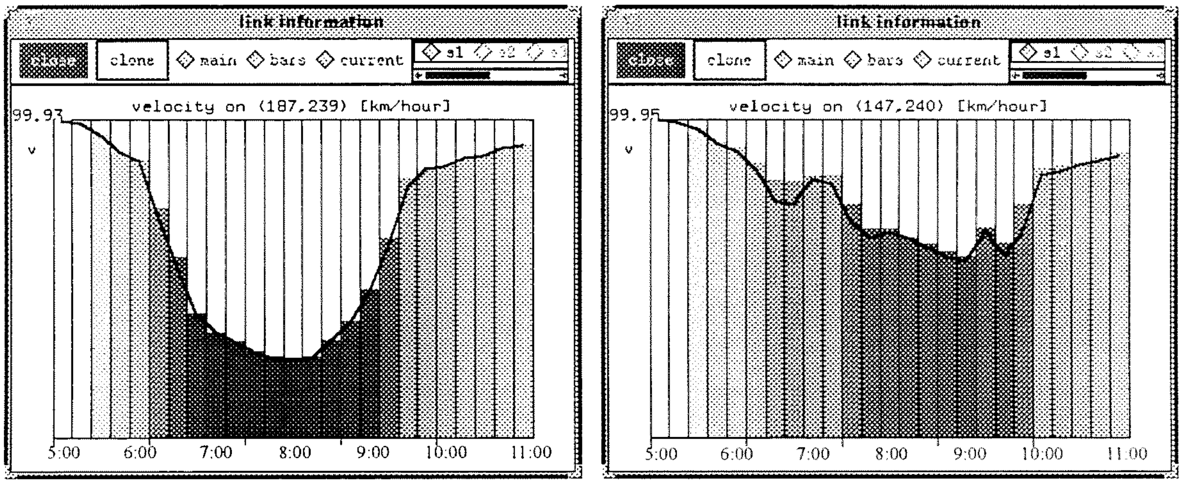


Figure 11: The velocity (km/hr) calculated by 3DAS at two locations on the I-66. In bars the rampmetering scenario is displayed. The solid line displays the velocity during the normal rush-hour.

The graphs in figure 11 show that the impact of ramp-metering is hardly noticeable on the I-66. The speed in the normal rush-hour (solid-line) is the same as during the metering scenario (bars). This is partly due to the originally low ramp intensities.

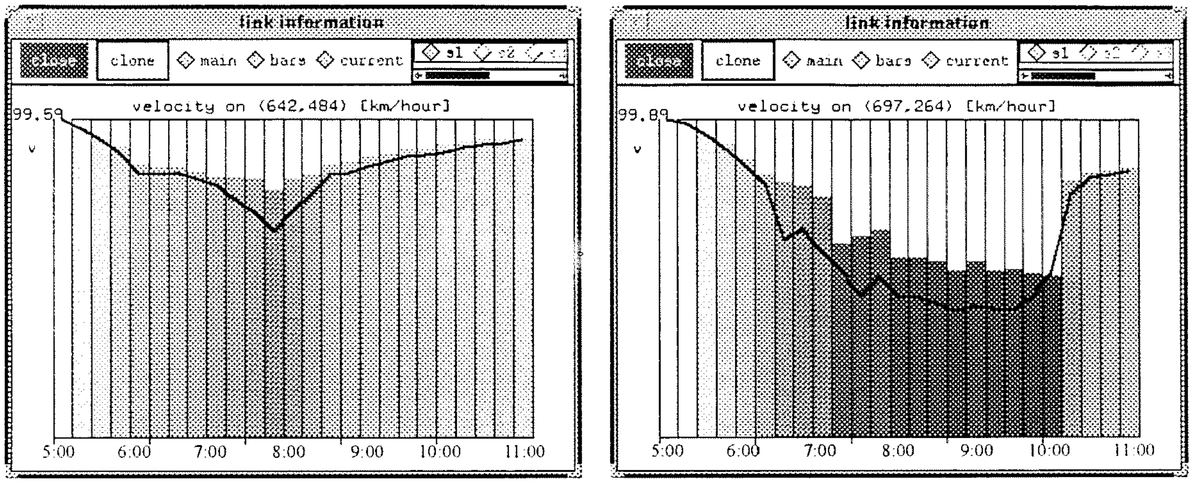


Figure 12: The velocity (km/hr) calculated by 3DAS at two locations on the I-395. In bars the rampmetering scenario is displayed. The solid line displays the velocity during the normal rush-hour.

Figure 12 shows a location somewhere halfway at the I-395, and one downstream the I-395. The two graphs show that on the I-395 there is a noticeable impact on the I-395. Both the locations show a slight improvement in speed. In the left graph the temporal decrease in speed at 8:00 in the normal rush-hour (solid line) is no longer there. At the other location, shown in the right graph, there is an improvement in speed almost over the total duration with approximately 10%.

Figure 13 shows the impact ramp-metering has on the arterial network. The figure shows a location at the end of the I-395.

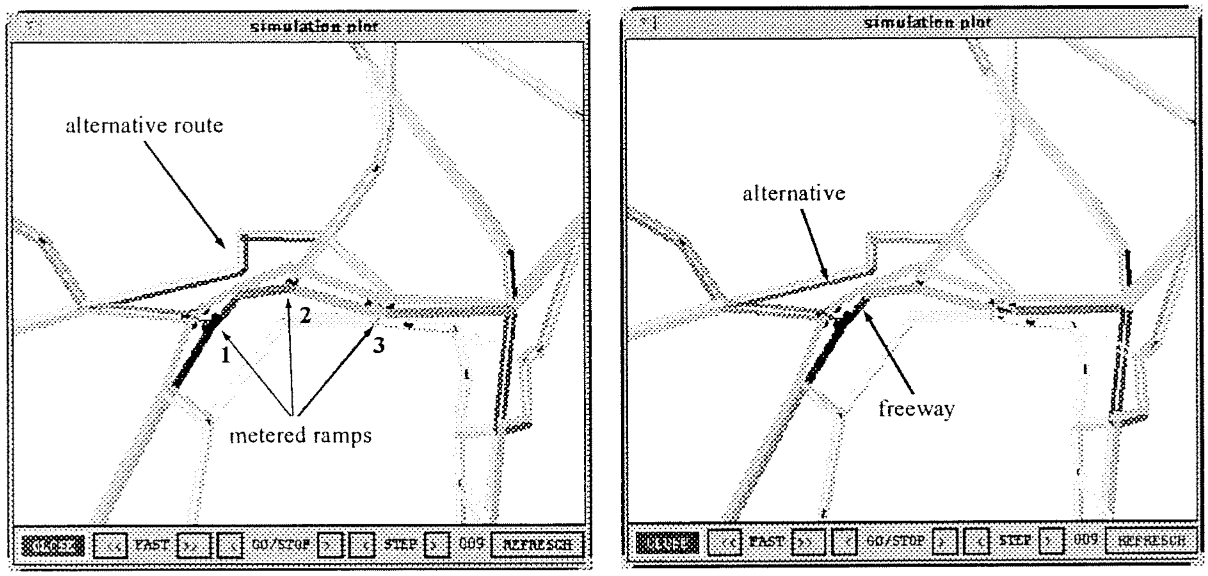


Figure 13: Rerouting behaviour. The density in the 13th period (9:00 - 9:15). The left figure displays the ramp metering scenario. The right figure displays the normal rush-hour scenario.

The figures in figure 13 are chosen to illustrate that, due to ramp-metering, alternative routes parallel to the freeway could be chosen. The left figure shows a slightly darker grey (higher density) at the alternative route then the right figure. At the freeway a slightly lower flow is detected.

Table 3: Freeway versus Alternative route at 9:00am

Road	Scenario	Density (veh/km)	Flow (veh.hr)	Speed (km/hr)
Freeway	Normal	127	5800	45
	ATMS	110	6000	50
Alternative	Normal	71	2000	30
	ATMS	87	1800	25

Table 3 shows the values for several parameters of one link in the freeway and one link in the alternative route at 9:00am. The links are marked with arrows in the right figure of figure 13. In the table the normal morning rush-hour is compared with the ATMS rampmetering scenario. The values show that traffic is avoiding ramp 1 and that a higher density is found at the alternative route. (The flow at the alternative route has not increased because 2250 is the capacity of that road. The density has increased but the troughput, due to slight congestion, has decreased.)

7.4 The Accident Scenario

To test whether the effects of incidents can be investigated with 3DAS an accident is simulated on the I-66. The accident is introduced by decreasing the capacity for the link with the accident by 60%. The OD-matrix and the departure functions are unchanged.
Two different route choice strategies are used. One strategy using the same routes that were chosen during a normal morning rush hour, the other route choice strategy is according an equilibrium assignment.
The first strategy represents a situation in which the accident is unknown to the travellers, while the second scenario is a situation in which each traveller is optimally diverted.

In the first scenario (no diversion) there is a traffic-jam at the I-66 which grows further upstream then in the normal morning rush-hour. The average speed of the congested links is very low. The two graphs in figure 14 show the situation at the I-66. The graphs show the middle section of the I-66. In grey the density at each link during the period is displayed. The darker the grey the higher the density and the lower the speed. The situation in the 5th and in the 10th period is displayed.

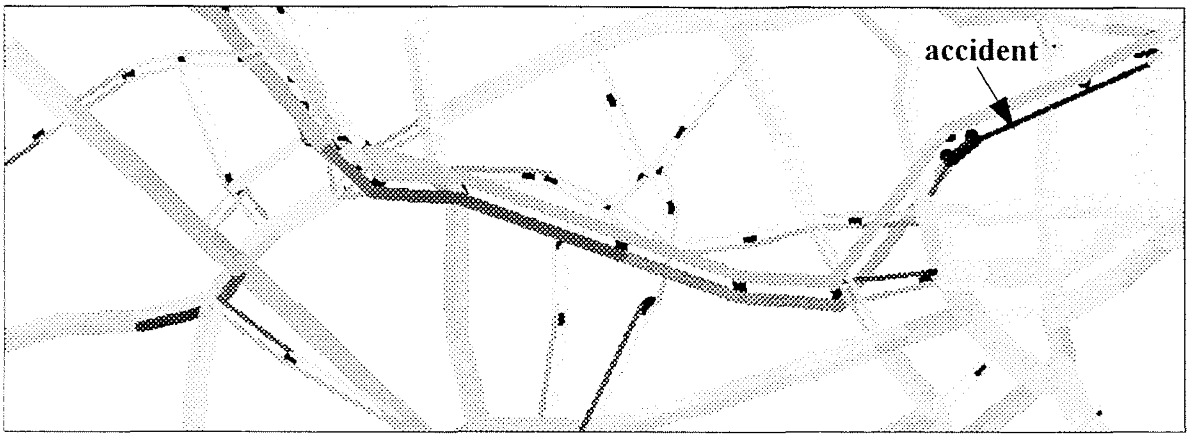


Figure 14: Accident at I-66; 5th period; 6:15am. Traffic jam starts

Figure 14 shows that at the link of the accident a queue starts to build. Further upstream is another bottleneck, because the traffic queues up there as well.

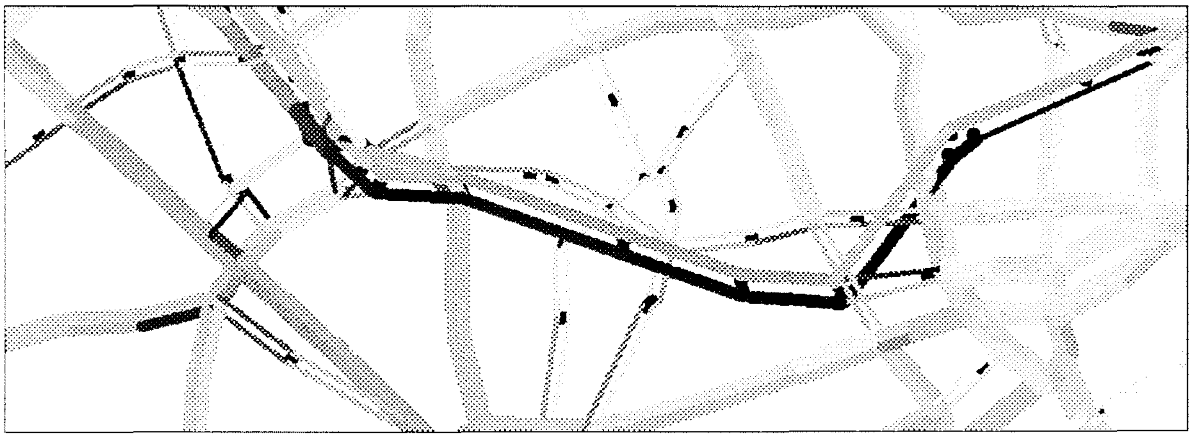


Figure 15: Accident at I-66; 10th period; 7:30am. Traffic jam along entire section

Figure 15 shows the same section five periods (1¼ hour) later. In this figure the congestion has grown upstream and almost the entire section is jammed.

In this scenario the drivers did not divert to a different route because they are not aware of the accident. In the next scenario an equilibrium assignment has been used. This means that all travellers are informed about the accident and will choose their route accordingly.

The equilibrium assignment gives some remarkable results. The total length of the traffic-jam that started due to the incident didn't grow further upstream than in the *normal* morning rush-hour. Comparison with the normal rush hour shows that the length of the queue is in fact shorter but the average speed is lower. Arterials around the location of the accident are all used more. Figure 16 and 17 show the I-66 again at the 5th and the 10th period for this scenario.

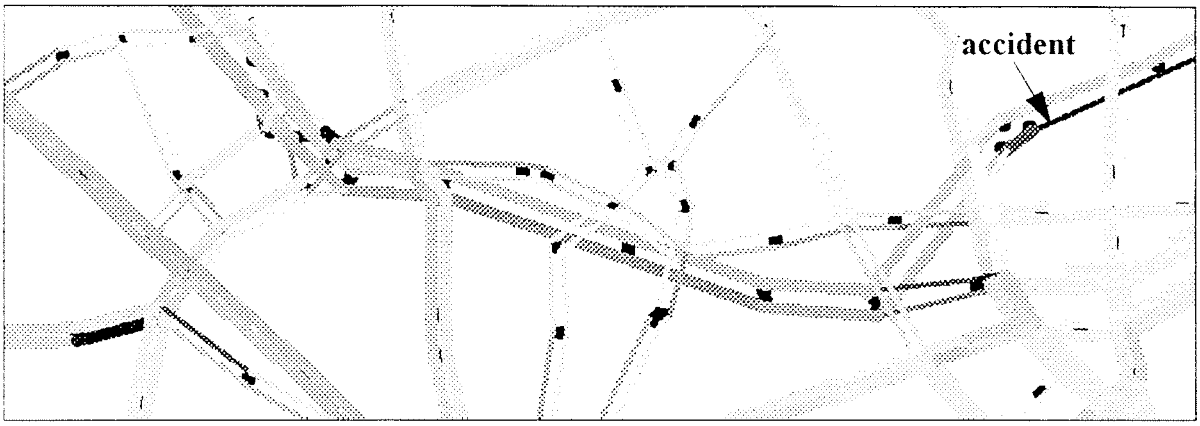


Figure 16: Accident at I-66; 5th period; 6:15am. Diversion scenario

Figure 16 shows that again in the 5th period the queue starts at the location of the accident. Further upstream the density is lower because travellers start to divert already. At the parallel route a higher density is noticed.

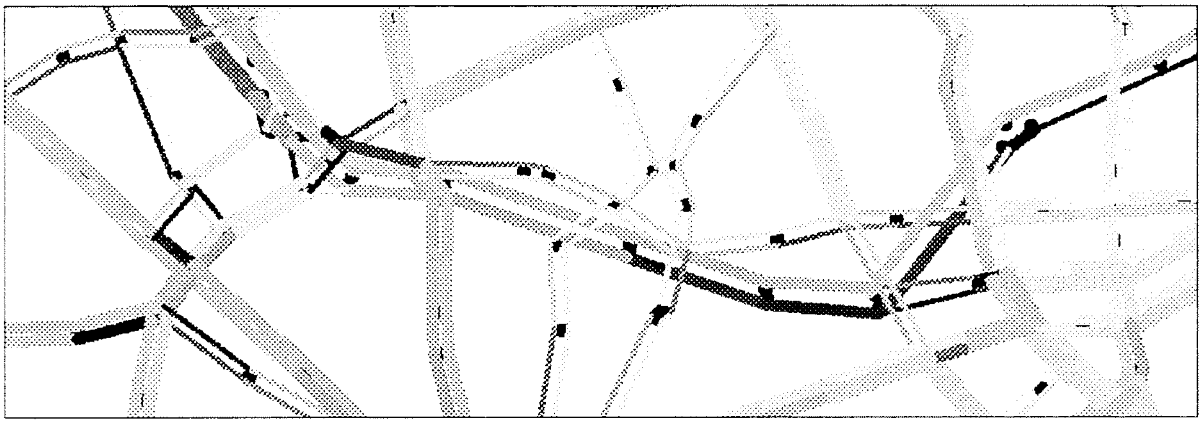


Figure 17: Accident at I-66; 10th period; 7:30am. Diversion scenario.

Figure 17 shows the same section five periods (1¼ hour) later. Here the congestion has grown upstream but not as far as in the other scenario. All the parallel routes show higher densities.

When the traveltimes to traverse the entire I-66 are compared, there is a significant difference between the two accident scenarios. In figure 18 the *normal* rush hour traveltime is compared with the accident scenario and with the accident with diversion scenario.

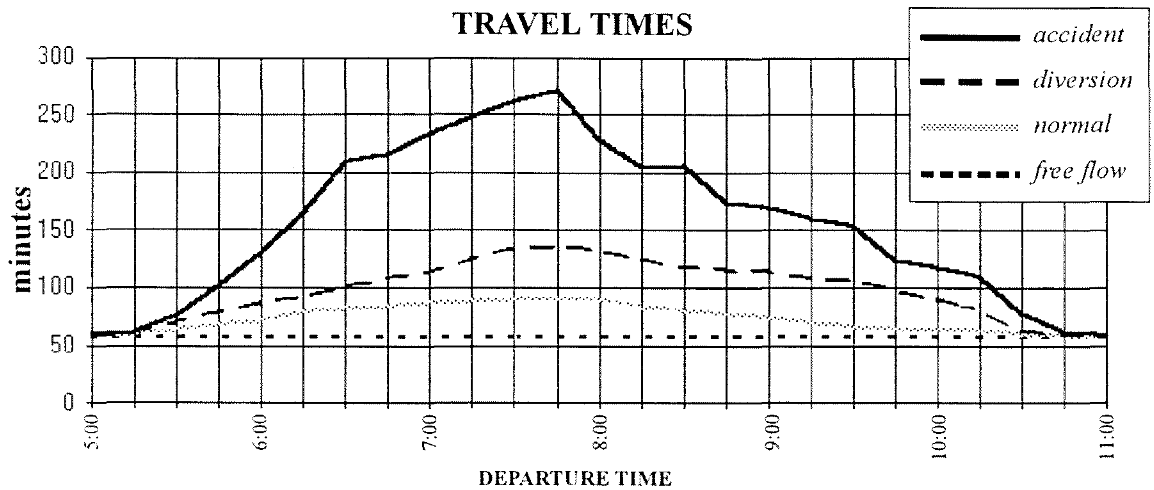


Figure 18: Traveltimes to traverse the I-66 for four scenarios: The free flow situation, the normal rush hour, the accident scenario, and the accident with diversion scenario.

The free-flow traveltime on the I-66 is 59 minutes. For the normal morning rush-hour it takes approximately 93 minutes to traverse the I-66 for the traffic that departs at 7:30am. In case of the accident, in which the traffic is rerouted, the traveltime has increased significantly, although the total length of the traffic-jam has not grown. When the traffic is not rerouted, the traveltime increases to more then four hours to traverse the I-66 for traffic that departs at 7:45.

Figure 18 shows that the traveltime during a normal rush-hour is (obviously) the shortest. The scenario with the accident has a traveltime which is approximately 3 times as high. In case we allow diversion the traveltimes are approximately 1½ times as high. This case shows an improvement of traveltime with approximately 50%. Of course this is an extreme case. The worst case is compared with the optimal case, and there seems to be enough capacity on the arterial network.

8. Conclusion

This study shows that a dynamic assignment model can be very useful for planning applications. A number of clear advantages for using 3DAS instead of static assignment are shown. The results give more detailed information about the occurrences of traffic jams, and the location or the cause of congestion can be identified more precise. To alleviate congestion, ATMS measures can be simulated, and all kinds of evaluations are possible. Influences on traveltime and jam-length, effects of rampmetering and rerouting can be investigated.
Dynamic assignment also has the advantage that all kind of *temporary* disturbances, such as accidents or road-works, are possible to simulate, and the duration of delays can be derived. The study also showed that 3DAS can be successfully used with larger networks

It must be stressed however that the data-requirements are much more stringent. Since by using 3DAS the level of detail is higher, the data must support this level of detail also. The accuracy of time variance is directly dependent on the accuracy of the time variance of the OD-matrix. For the amount of data that 3DAS requires and produces, it is essential to find a good system of organizing and maintaining this large amount of data. In the beginning this costs a great effort, but with increasing experience using 3DAS this disadvantage will most probably disappear.

The calculation time required for 3DAS is longer than required for static assignment. For planning purposes however, calculation time is not the main concern. Much more important is visualization of the results. With dynamic assignment the results are flows *in time*. The best way to analyse the results is as a *film*. In order to do that a graphically powerful workstation can be used. This is one of the main reasons graphical workstations are used for this research. When the model is used for traffic control and real-time management a faster computation time is needed. This can be achieved by reducing the problem size (smaller network). When this is not possible a faster computer could be used. The research, described by H.J.M. van Grol in [6], towards the development of special purpose hardware for assignment calculations gives a cost-effective solution.

In this specific study the amount and the quality of the data is very poor, and there are limited possibilities to verify the data. Since there isn't any insight in the local traffic patterns, we cannot judge the quality of the OD-matrix. The time spent on this research was too short to make any serious planning decisions. The study is therefore primarily meant to investigate the (un)usefulness of dynamic assignment for planning purposes. For *real* planning decisions a more elaborate study is required.

In order for Advanced Traffic Management Systems to be successful there is a large demand for more data and better (3D) OD-matrices. Hopefully new methods for OD-estimation and more data from induction-loops and probe-vehicles will give better results in the future.

With more time available for this study, more local knowledge of the study area, and more induction loop data, the model has the potential to give reliable information for *real* planning strategies and driver information systems.

9. Acknowledgements

We are thankful to *NWO* (Netherlands Organization for Scientific Research) and *University-funds Delft* for there financial support to make the four month visit at Virginia Tech possible. *Virginia Polytechnic & State University (Virginia Tech)* for the possibility to use all the facilities at the University, the *Center for Transportation Research* at Virginia Tech for their hospitality and their scientific support, the *Laboratory for Scientific Visualization* at Virginia Tech for the availability of there hardware and their enthusiasm for visualisation in traffic. *Virginia Department of Transportation* (VDOT) and *Washington Metropolitan Council of Governments* (COG) who provided us with the necessary data, and *Peter Maas* (Applied Physics student) who worked hard to get all the necessary graphical software running.

REFERENCES

[1] R. Hamerslag, P.C.H. Opstal, "A Three-dimensional assignment method in Timespace", *Rapport 87-94 Faculty of Mathematics and Informatics, Delft University of Technology, The Netherlands, 1987.*

[2] R. Hamerslag, "Dynamic Assignment in the Three Dimensional Timespace, Mathematical Specification and Algorithm", *USA-Italy seminar on transportation networks, Napoli & Capri, 1989.*

[3] R.Hamerslag, "Dynamic Assignment in the Three Dimensional Timespace", *Transportation Research Record 1220, 1989.*

[4] E. de Romph, H.J.M. van Grol & R. Hamerslag, "A Dynamic Assignment Model for Short Term Prediction", *presented at Capri Seminar on Urban Traffic Networks, 5-8 July 1992, Italy.*

[5] E. de Romph, H.J.M. van Grol & R. Hamerslag, "3DAS -3Dimensional ASsignment- A dynamic assignment model for short term predictions", *39th North American Meeting of the Regional Science Association International, 13-15 November 1992, Chicago Illinois USA.*

[6] H.J.M. van Grol, "Traffic Assignment Problems solved by Special Purpose Hardware with Emphasis on Real Time Applications", *Ph.D. thesis, ISBN:90-9005441-3, 1992.*

[7] M. Westerman, R.Hamerslag, "Collecting the Dynamics of Metropolitan Freeway Systems: A Simulation Study", *Delft University of Technology 1993.*

[8] Updates to COG/TPB Travel Forecasting Procedures, Highway Applications, 1990 Validation and 2010 Simulation, DRAFT, March 1993.

Dedicated hardware for an assignment model

H.J.M.van Grol

11 November 1993

Delft University of Technology, Faculty of Applied Physics, Department of physics Informatics/ Computational Physics

Introduction

Real time calculation of assignment algorithms requires large computational capacity that can only be realized with parallel computers. The development of such a dedicated processor array has been the subject of a PhD thesis by L.H.M. van Grol.
His work consisted of both the analysis op assignment problems and the design and implementation of the machine. Special attention has been paid to the minimization of costs.

The resulting computer design is suitable for a wide variety of assignment problems, but within the framework of the joint project emphasis has been on accellerating the 3DAS model.
The machine is a linear processor array (LPA) that is interfaced to a host computer (figure 1). The design is based on independant calculation of Origin-Destination pairs. Furthermore, the unavoidable communication overhead has been minimized by means of an optimal configuration of the connection between processors:the pipeline. As a consequence the speed of calculations increases linearly with the number of processors added. A typical machine at this moment consists of 16 processors but the number can easisly be expanded as needed.
A number of processor boards has been constructed and tested. Right now (november 1993) a larger number of these boards is being put together in order to make several operational systems: the first of those systems has been finished in nvember '93.
As an indicator of possible computational speed the following estimate can serve:
real-time predictions using the 3DAS model for a net of 1500 nodes, 500 origins (zones) 12 periods and 10 iterations can be produced in less than 4 minutes.
For a more elaborate description of this part of the project, see the following summary of the PhD thesis of H.J.M van Grol.

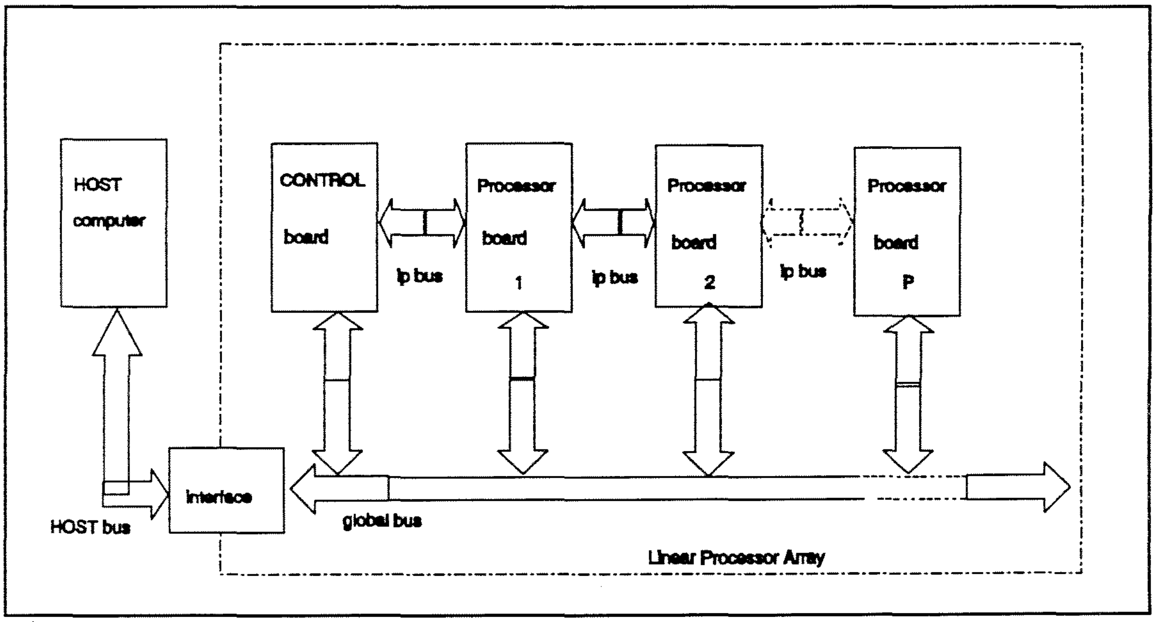


Figure 1: Layout of LPA.
ip bus is data pipeline

SUMMARY

In this thesis, an answer is given to the question, whether special purpose hardware can be used to reduce the computing times of the time-consuming traffic assignment models. This question was put forward by the Department of Transportation Planning and Highway Engineering at the Faculty of Civil Engineering at the Delft University of Technology. The investigations, to answer the question, were carried out in the Computational Physics group at the Faculty of Applied Physics; this group is specialized in the design and construction of special purpose hardware for computationally intensive tasks.

Research goal

The current and expected growth of traffic, the limited available road-space and the pressure from society to solve the congestion and environmental problems lead towards the development of more accurate and detailed planning methods to *improve* the infrastructure, and towards the development of methods and instruments to improve the *performance* of the traffic on the existing infrastructure.

Transportation planning methods are used to make prognosis for the traffic conditions in the future and to determine the effect of changes to the infrastructure on the performance of the traffic in a network. For these long term predictions *static* assignment models are used. These computer-models are time-consuming, and with the need for more detailed and accurate results the computing time becomes a restricting factor.

Improving the performance of traffic on the existing infrastructure is part of dynamic traffic management. Developments in electronics, telecommunications, etc., have led to instruments and techniques that improve the throughput of traffic in a network. To control these instruments efficiently and safely, the short term development of the traffic must be known. To provide these short term predictions, the use of dynamic assignment models is considered in this thesis. These models are more complex than the static models. Moreover, the time-restrictions for these models are more stringent, since the prediction must be made in 'real'-time. A short-term prediction is useless when the computing time exceeds the required prediction-time.

Reducing the computing time is the main theme of this thesis. At least four ways have been identified to tackle this problem.
These four ways are:

- to redefine the problem
- to improve the model
- to improve the algorithm and the implementation
- to use faster hardware

A redefinition of the problem can mean that the problem is rescaled, solved in a completely different way, etc. Improvement of the model can be invoked by new insights in the problem or requirements that are made less stringent. The quality of the algorithms and their implementation generally depends on the ability of the programmer and are often a source for improvement. Obviously, the use of faster (special purpose) hardware should only be considered when all possibilities to optimize performance have been exploited.

However, it is mainly this last point which is addressed in this thesis. Faster hardware leads towards general purpose *super*computers and special purpose hardware. General purpose supercomputers are commercially available, but are expensive. Special purpose hardware, designed for a specific task, can form a cost-efficient alternative, provided that the problem is suitable for such an approach.

The main research issue formulated is:

> *"Is special purpose hardware a **cost-efficient** way to speed-up the traffic assignment calculations ?"*

Three additional questions have been derived.

- have the alternative ways, to reduce computing time, been examined thoroughly ?
- is the traffic assignment problem suitable for special purpose implementations ?
- can a special purpose computer for traffic assignment compete with commercially available computers?

Answering these questions largely covers answering the main research issue. The questions are answered in five distinguished parts in this thesis.

- **Part I**: Part I gives a general overview, formulates the above research questions and discusses the research approach.

- **Part II**: In Part II the traffic assignment problem is analyzed. An overview of static assignment models is given and the overall structure is derived. This structure is the repeated determination and assignment of shortest path trees. The most time-consuming part is the determination of the shortest path tree. The algorithms for shortest path finding have been studied thoroughly in the past. A few of the most efficient and robust algorithms are a threshold-method, T-calc (address calculation) and a label-correcting method, C-dlque (double linked queue). There are no major improvements to be expected from the algorithm side. The run-time complexity of the assignment problem is $O(n^2)$ where n is the number of nodes in the network.

- **Part III**: In Part III the hardware options are discussed. After an overview of computer architectures the use of a special purpose computer is compared with general purpose supercomputers. Clearly, the decision to construct or purchase a special purpose computer depends on the suitability of the problem, the available budget, the life-cycle of the problem and the cost-efficiency of the computer.

Further analysis of the traffic assignment algorithms shows that there are no efficient *single*-purpose solutions. However, a *multi*-purpose solution is possible by exploiting the inherent parallelism in determining and assigning shortest path trees. It has been shown that a parallel architecture of p processing units with the possibilities of fast summation and broadcasting —independent from p— can be an efficient way to speed-up the assignment computations.

Based on these requirements a special purpose computer, called the Linear Processor Array (LPA), is proposed. The LPA consists of P processor boards, that are linearly inter-connected. Each board contains a i860 RISC microprocessor, 512 Kbyte SRAM and

32-128 Mbyte DRAM. The peak-performance from one board of 40 MIPS and simultaneously 80 Mflops is possible due to the so called 'dual instruction' and 'dual operation' modes. The communication between two boards takes place with FIFO's. These devices have the functionality of a buffer and allow asynchronous communication with a speed of 160 Mbytes per second. Each board can independently transfer information to its neighbors. The linear interconnections provide the means for fast summation and broadcasting. The LPA is connected to a host computer which provides a user friendly work environment.

An adapted operating system provides a solid programming environment on each processor board and allows the use of a high-level programming language (e.g. C). The user is supported by code-optimizers and by (parallel) programming tools in the form of a LPA-simulator package. This package allows the development of LPA-programs on general purpose computers and takes care of startup-procedures, etc.

Early test-results compared with the performance on a Cray Y-MP have shown that a LPA with P=16 boards will be about as fast as a Cray Y-MP8. Depending on the amount of memory per board the price of a 16 board LPA varies between \$160.000 and \$260.000. When the price-difference with the Cray Y-MP8 is considered (\$20 million), the cost-performance ratio is improved by at least a factor 40. The large differences in the cost of maintenance, energy and personnel provide an additional major saving in favor of special purpose computers.

- **Part IV**: Part IV deals with the possibilities of real-time assignment. An overview of dynamic assignment models is given and the use of short-term models for Dynamic Traffic Management is briefly discussed. The requirements for such a model are formulated. The model must be able to predict a realistic development of the traffic flow distribution in the coming hour, it must be able to simulate traffic control instruments and it must be sufficiently fast with large networks (300-3000 nodes) and short periods (1-10 minutes).

A new dynamic assignment model is introduced. This model is based on a model developed by Hamerslag at the Department of Transportation Planning and Highway Engineering. This model uses existing network definitions, unlike some models, that require a completely new definition. The model has been improved towards its dynamic aspects; a better representation of the traffic conditions on the roads by using densities rather than intensities.

To test the model, to provide a user-friendly environment and for demonstration purposes, the model has been implemented on a graphics workstation in a window-environment.

Because of the close resemblance of the dynamic assignment model with static assignment models the performance-improvement on the LPA will be about the same. Real-time predictions for a network of 1500 nodes, 500 origins in 12 periods of 5 minutes and 10 iterations, are estimated to be done in less than 4 minutes.

- **Part V**: In Part V the future developments are briefly discussed and the conclusions in this thesis are summarized.

The dynamic assignment model will be tested on the ring-road around Amsterdam.

A few suggestions are made to improve the cost-performance of the LPA further. The LPA can be improved be using new, faster processors. A reduction in cost can be obtained by placing up to 4 processors on each board (more processors would deteriorate the efficiency for assignment algorithms). When computer-manufactures would recognize the power of the simple but fast inter-processor connections and introduce them onto their processor boards, a building block can be created that facilitates a straightforward construction of an LPA (possibly at even lower cost).