ELSEVIER

International Journal of Forecasting 13 (1997) 87–103

international journal of forecasting

A dynamic traffic forecasting application on the Amsterdam beltway

Nanne J. Van Der Zijpp[a], Erik De Romph[b]

[a]Delft University of Technology, Faculty of Civil Engineering, P.O. Box 5048, 2600 GA Delft, Netherlands
[b]Goudappel Coffeng, Parkweg 4, 7411 SH Deventer, Netherlands

Abstract

This article presents some theoretical developments that have resulted in a dynamic traffic forecasting procedure that is feasible both from a technical viewpoint (data availability) and from a practical viewpoint (data preparation). The procedure consists of a dynamic origin–destination (OD) matrix estimation module and a dynamic traffic assignment module. The OD-estimation module is an extension of traditional (static) OD-estimation methods, i.e. production–attraction models combined with the use of a deterrence function. To make the procedure computationally feasible an efficient parameter estimation method has been provided.

To test the combined OD-estimation/dynamic assignment model, data were collected continuously during 3 weeks at 141 locations on the beltway of Amsterdam. As an alternative to the proposed procedure, historical averages have been compiled from all observed data. Comparisons between true, predicted and averaged data show that a lot of effort must be invested in specifying OD-demand and network characteristics in order for the new method to be competitive with historic averages for traffic forecasting. Exceptional circumstances such as severe incidents, however, are reported better with the dynamic forecasting procedure.

Keywords: Dynamic traffic assignment; Origin–destination estimation; Travel demand modelling

1. Introduction

One of the requirements of dynamic traffic management is to provide predictions of future traffic conditions. Based on these predictions a traffic operator can make decisions concerning rerouting, travel information and advice, and ramp metering. For on-line application, a prediction system must respond to accidents, weather conditions and road works. To perform this task, the system must be provided with data from various sources.

In practice a traffic prediction system must work within the limitations of data availability. This article presents a traffic prediction system designed to operate with data available in The Netherlands. The main components of this system are a module for the estimation of dynamic origin–destination (OD) matrices and a dynamic assignment module. The OD-estimation module was developed especially to cope with imperfections in the input data, such as missing values, inaccuracies, mis-specifications and so on. A large part of the article is devoted to a descrip-

PII S0169-2070(96)00703-0

tion of the actions that were taken to arrive at a dynamic OD matrix. The dynamic assignment module is a result of an on-going research project at Delft University, and the model was used without modification from De Romph (1994).

The structure of the article largely corresponds to Fig. 1, which gives an overview of the prediction system. This framework consists of seven data sets (rectangles) and five actions (circles). Each data set and action is described briefly.

Fig. 1 contains the following data sets:

- Network: A specification of the topology of the network. For each link the maximum speed, the maximum density and the road type are given. The road type of a link determines the speed–density function to be used. The network used for this pilot study is described in Section 2.1.
- Induction loop data: Every minute the flow and the speed are measured at several locations in the network. The Dutch Motorway Traffic Management (MTM) system is used to collect the data. MTM is a system designed primarily for monitoring and operational control of the Dutch motorway network. A description of MTM and the properties of the data that were derived from it is given in Section 2.2.
- 30-min OD matrix: A dynamic OD matrix specified with an aggregation level of 30 min. The origins and destinations are the on- and the off-ramps in the network.
- Dynamic OD matrix: A dynamic OD matrix specified with a 5-min interval. This matrix is the result of a 'dynamizing' procedure and is used as input to the dynamic assignment model.
- On-ramp volumes: Time series of flows with an interval length of 5 min for all the on-ramps in the network.
- Speed–density functions: For each road type in the network a different relation between speed and density is used.
- Prediction: As a result of the dynamic assignment, estimates of the flow, the density, the speed and the travel time for each link in the network and for each period of the total time span are given.

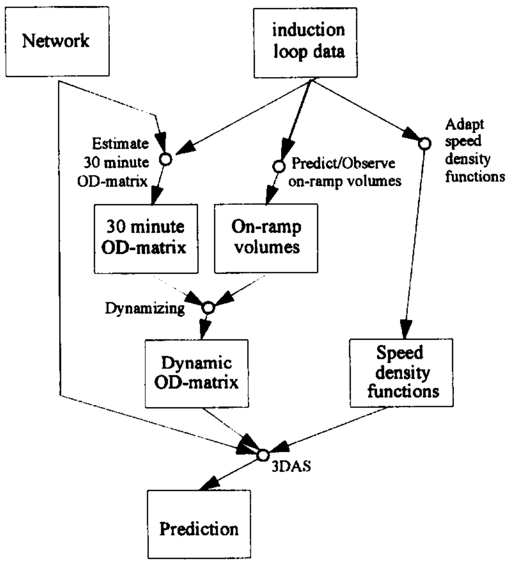

Fig. 1. Framework of the prediction system.

Fig. 1 contains the following actions:

- Predict/observe on-ramp volumes: For each on-ramp in the network the volumes entering the network for future periods are estimated from historical data. Current volumes may be matched to historical patterns using a least square procedure[1] (see De Romph, 1994).
- Estimate 30-min matrix: A new time-dependent model of travel demand is used to generate synthetic OD matrices. The method is described in detail in Section 4.
- Dynamizing: Based on the 30-min OD matrix and the on-ramp volumes, the interval of the OD matrix is reduced to 5 min. This procedure, called *dynamizing*, is described in detail in Section 5.
- Adapt speed–density functions: Based on historical induction loop data a speed density function is estimated for each road type in the

[1] In this specific pilot this has not been done, the actual measured on-ramp volumes were used.

network. The measured data are fitted to the function using least squares. This procedure is described in detail in De Romph (1994).
- 3DAS: The 3DAS model is a dynamic assignment model. The model calculates the traffic flows, densities, speeds and travel times for each link in the network and for each period of time. The 3DAS model is described briefly in Section 6 (and in detail in De Romph, 1994).

The above description accounts for Sections 4, 5 and 6. The other sections contain a description of the input data (Section 2), the research approach (Section 3), the data preparation (Section 7), the results (Section 8) and the conclusions (Section 9).

2. Input data

2.1. Network

The study area is the Amsterdam beltway (Fig. 2). The beltway-network representation consists of a list of nodes and a list of directed links, extracted from "basisnetwerk Nederland" (Dienst Verkeerskunde, 1974). The total length of the beltway is approximately 30 km. The network has two tunnels and four major motorway intersections. The network contains 286 nodes and 430 links of which 76 are on- or off-ramps. A small part of the arterial network is present near on- and off-ramps, making it possible to combine clockwise and anti-clockwise on-ramps or off-ramps in one origin or destination node. In total there are 21 origins and 21 destinations.

2.2. Induction loop data, the MTM system

This section summarizes the properties of the input data that are available at this moment as a basis for dynamic traffic prediction in The Netherlands. The limitations of these data are

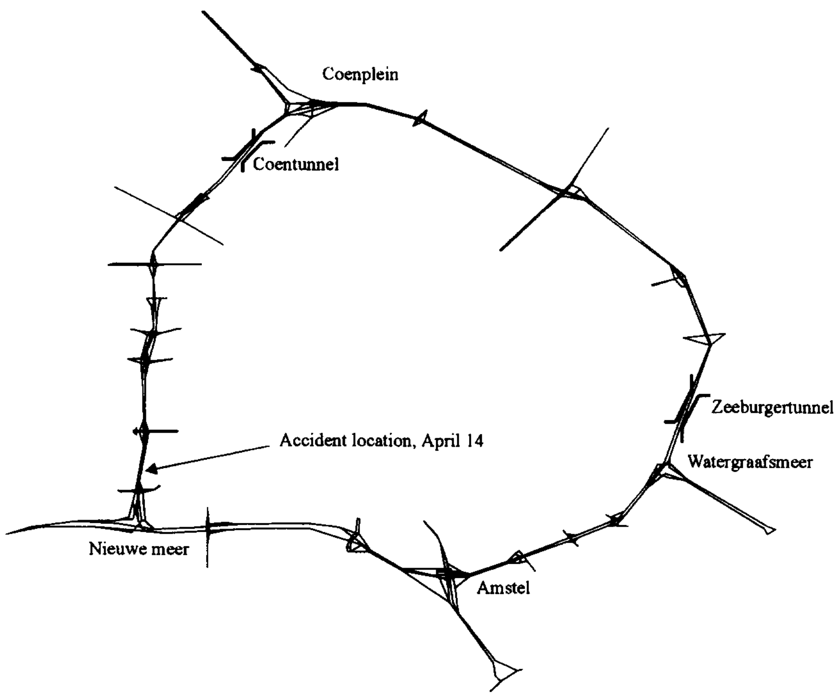

Fig. 2. Map of the study area (Amsterdam beltway).

critical to the choices that need to be made for the OD-prediction module.

Data were obtained from MARE, a research facility connected to the national MTM system. For 3 weeks data were collected at 141 locations on the Amsterdam beltway.

Motorway sections in which MTM is installed are equipped with dual induction loops on all lanes at intervals of approximately 500 m. One of the primary objectives of MTM is upstream warning for congestion and slow traffic. It is claimed that such warnings lead to a significant decrease in the risk of pile-ups and incidents (−50%), an increase of flow (+5%) and a decrease of travel time (−15%) (source: Adviesdiemst Verkeer & Vervoer, Transport Research Centre, 1994).

Not all collected data are available at a central computer. Data are aggregated and smoothed in road-side processors (substations) before transmission to the central computer. Therefore, dynamic traffic management applications using MTM collected data intended for use in the near future are subject to certain input data limitations, of which the most important are:

- Data are aggregated to 1-min periods.
- Arithmetic averages of observed speeds are stored whereas harmonic averages are needed to compute average travel times.
- Missing observations of entry and exit volumes. In general, motorway on- and off-ramps are not explicitly monitored. The detectors of the monitoring system are located on the through lanes of the motorway. At an aggregate level on- and off-ramp volumes can be reconstructed from the observations at the adjacent links, if present. Applying this technique to 1-min data, however, introduces significant errors, regularly leading to negative flows. A factor that makes the problem worse is that the accuracy of induction loop detectors decreases in areas with many lane changes, which is typical for on- and off-ramps.
- The possibility of faulty detectors. The probability of all detectors working at a given time is very small.
- Large observation errors when the traffic is slow or when traffic is characterized by a high frequency of lane changes.
- Unreliable counts of low traffic volumes. A minimum of 4 vehicles min^{-1} is 'observed' in every substation. For some technical reason the predecessor of the MTM system generated dummy vehicles every 15 s in the absence of real vehicles. This property is preserved in the present version of MTM and is a source of error if the real volume is below 4 vehicles min^{-1}, which is frequently the case at night and on less frequently used links, such as on-ramps.
- Poor positioning of induction loops. Some loops are located halfway up an on-ramp or off-ramp. The exact number of vehicles using the ramps cannot be determined in this case, as the point where vehicles change between ramps and main lanes varies between motorists, and might be up- or downstream of the detector.

3. Research approach

The *primary* objective of this pilot study was to find answers to the following two questions:

- Can the system give a sufficiently accurate prediction of the future traffic conditions?
- Can the system react to changing conditions, such as accidents and weather changes?

The Amsterdam beltway was selected as the pilot area. This network is well equipped with induction loops, has a manageable size, and several dynamic traffic management instruments are in operation or planned for operation in the near future.

To answer the questions given above, data were collected for 3 weeks at 141 locations on the Amsterdam network, using induction loop detectors. The data were screened for errors and 2 days, April 14th and April 28th, were selected to be reproduced by the prediction system. A historical data base was set up using data of 5 other days.

It was decided to consider three different scenarios to answer the questions above:

I. The first scenario tried to reproduce the morning rush-hour of Thursday, April 28th. The data on this day have few errors and no major accidents or other disturbances. It was decided to reproduce the rush-hour from 06:00 in the morning until 09:00. The period length used in the calculation is 5 min. This period length was chosen because a total of 36 periods could be managed by the software, and resulted in a feasible representation of the dynamics in the network. A shorter period length would suggest an accuracy that cannot be reached, and with a longer period length, the interaction between time periods diminishes. An impression of the capabilities of the system can be achieved by comparing the predicted and actual measured flows.
II. The second scenario tried to reproduce the morning rush-hour of Thursday, April 14th. On this day an accident occurred on one of the freeways which blocked the freeway for 20 min. The same time span and period length as in the first scenario were used. The predictions are again compared with the actual measured flows for this scenario.
III. To get an impression of the relevance of a system for making predictions, a third scenario was considered that does not use the system to make a prediction, but used the collected induction loop data to make a prediction. The measurements were summed and averaged for 5 measured days with similar conditions. This scenario represents the *historical average traffic pattern*.

The first research question was answered by the first scenario (see Section 8.1) and the second question was answered by the second scenario (see Section 8.2). Section 8.3 gives the results of a comparison between the first two scenarios and the historical average.

During this research, various computers, including a parallel computer (see Van Grol, 1992) were used to make the calculations. Most modules of the system can work independently. The OD-estimation model and the dynamic assignment model are the most computationally intensive modules.

4. OD estimation

The estimation of origin–destination tables is a classical subject in transportation engineering. The main characteristic of the problem is that many OD tables satisfy the constraints posed by observations taken from a transport system, usually volume counts. Therefore a unique solution to this assignment problem is impossible, unless one uses a ranking of the underlying OD matrices. Such a ranking can be based on different principles, such as compliance with a model of travel demand, closeness to a prior matrix, maximization of the number of micro states or minimization of the total travel time in a system (for relevant publications, see, among others Van Zuylen and Willumsen, 1980; Maher, 1983; Cascetta and Nguyen, 1988; Hamerslag and Immers, 1988).

Most of the work to date concentrates on *static* OD tables; that is, all traffic volumes are considered at an aggregate level. For dynamic traffic predictions, travel demand should be specified in a time dependent manner. For this purpose we define a *dynamic* OD table as a series of OD tables, ordered with respect to departure time. Because of differences in spatial aggregation and time dependence, these methods deviate significantly from those developed for static OD estimation.

The great majority of dynamic OD estimators is based on the prediction error minimization principle: inferences about OD patterns are made on the basis of similarities between the up- and downstream traffic flows. Examples of methods based on this principle are described in Cremer and Keller (1981), Nihan and Davis (1987), Keller and Ploss (1987), Bell et al. (1991), Cascetta et al. (1993) and Van Der Zijpp and Hamerslag (1994) and Van Der Zÿpp (1996). Platoon dispersion and congestion have an adverse effect on methods based on prediction error minimization, and limit the applicability to simplified networks like intersections and motorway corridors.

These theoretical considerations, together with practical considerations such as limitations posed by input-data and a limited time-budget for data preparation necessitate other methods for large or complex networks. Considerations related to input-data are the existence of accurate observations on *all* on-ramps, knowledge of the exact location of induction loops, synchronization, accuracy and so on. See Section 2.2 for more details on data availability on the Dutch motorway system.

In this article the generation of *synthetic* time-dependent OD matrices is addressed. The problem of determining an OD matrix from an under-specified set of traffic counts is solved by the use of a model of travel demand. In the literature the field of synthetic dynamic OD estimation is relatively unexplored. This is probably due to the disaggregate nature of the problem and the fact that a motorway network rather than a complete transport network is considered. These considerations invalidate most of the models of travel demand.

At least two references to synthetic time differentiated OD estimation can be found. Willumsen (1984) describes an extension of ME2. ME2 is a method to estimate static OD tables based on the maximum entropy model (see Van Zuylen and Willumsen, 1980). Van-Aerde et al. (1993) describe a procedure referred to as QUEENSOD. Relative to previous research this article presents three novelties:

- A derivation of a static model of travel demand motorway networks.
- A time-dependent extension of this model.
- An efficient procedure for solving the unknown variables.

These elements are treated in Sections 4.1, 4.2 and 4.3, respectively. The output of the procedure described in the following sections is a time dependent OD table with time intervals of 30 min. A typical period length for dynamic traffic prediction is 5 min. Therefore the 30-min estimates are converted to 5-min intervals that match the observed on-ramp volumes. This process, referred to as 'dynamizing', is described in Section 5.

4.1. A model of travel demand for subnetworks

This section derives a static model of travel demand for motorway networks on the basis of a generally accepted model of travel demand for the surrounding network. In this context, a motorway network is referred to as a subnetwork. As a point of departure the well-known gravity model with an exponential deterrence function is used (see e.g. Ortúzar and Willumsen, 1990, for more background on this model). If T_{rs} denotes the number of trips from 'true' origin to 'true' destination then this model prescribes the following relation:

$$T_{rs} \equiv A_r B_s \exp(-\beta c_{rs}) \tag{1}$$

where

A_r production ability for zone r
B_s attraction ability for zone s
β parameter in deterrence function
c_{rs} generalized travel costs from zone r to zone s

Define the assignment map τ with:

$$\begin{aligned} \tau_{rsij} &= 1 \quad \text{if flow } rs \text{ contributes to flow } ij \text{ on the subnetwork} \\ \tau_{rsij} &= 0 \quad \text{otherwise} \end{aligned} \tag{2}$$

In addition to the validity of (1), assume that m disjunct sets of origins $\boldsymbol{O}(i)$, $i = 1,2\ldots m$ exist that jointly cover all origins in the surrounding network, and that similar sets of destinations $\boldsymbol{D}(j)$, $j = 1,2\ldots n$ exist, and that for these sets the assignment map satisfies:

$$\tau_{rsij} = \tau_{ri}\tau_{sj}$$

where:

$$\begin{aligned} \tau_{ri} &= 1 \quad \text{if} \quad r \in \boldsymbol{O}(i)\,, \quad \tau_{ri} = 0 \quad \text{otherwise} \\ \tau_{sj} &= 1 \quad \text{if} \quad s \in \boldsymbol{D}(j)\,, \quad \tau_{sj} = 0 \quad \text{otherwise} \end{aligned} \tag{3}$$

Condition (3) is satisfied if the entry where an arbitrary OD path enters the subnetwork depends only on the trip-origin and the exit depends only on the trip-destination. This is the case for example if the subnetwork under consideration is embedded in a network that may be represented by a directed tree, see Fig. 3.

The last assumption is that the generalized travel costs are *additive*, i.e.:

$$c_{rs} = c_{ri} + c_{ij} + c_{js} \tag{4}$$

This is in accordance with the usual assumptions.

Combining (2) and (3) it follows that the subnetwork entry-exit-flows are given by:

$$f_{ij} = \sum_{r \in O(i)} \tau_{ri} \sum_{s \in D(j)} \tau_{sj} T_{rs} \tag{5}$$

Combining (5) with (1) provides the simplification:

$$f_{ij} = a_i b_j \exp(-\beta c_{ij}) \tag{6}$$

with

a_i entry production ability,

$$a_i \equiv \Sigma_{r \in O(i)} A_r \exp(-\beta c_{ri}) \tau_{ri}$$

$\mathrm{b_j}$ exit attraction ability,

$$b_j \equiv \Sigma_{s \in D(j)} B_s \exp(-\beta c_{sj}) \tau_{sj}.$$

The above derivation shows that under certain conditions usage of model (1) on a subnetwork is justified.

4.2. Time-dependent extension of the static model

In this section the time-dependent case is considered. As a time-dependent extension of model (6), we propose:

$$f_{ijt} = a_i(t) b_j(t + c_{ij}) \mathrm{f}(c_{ij}) \tag{7}$$

with

f_{ijt} flow from entrance i to exit j, departing period t

c_{ij} average travel-time from entrance i to exit j

$a_i(t)$ production ability related to entrance i in period t

$b_j(t + c_{ij})$ attraction ability related to exit j in period $t + c_{ij}$

$\mathrm{f}(c_{ij})$ deterrence function, distributed in classes; $\mathrm{f}(c_{ij}) = \Sigma_k F^k \delta^k_{ij}$, $\delta^k_{ij} = 1$ if c_{ij} is in class k, and zero elsewhere

Note that the interpretation of c_{ij} has changed from generalized travel costs to average travel time, and the continuous deterrence function is replaced by a discretized deterrence function. The latter replacement was done to facilitate computational procedures.

A second element of a time-dependent model description relates to the trip execution. Since travel time can no longer be neglected flows with a certain departure period contribute to link flows in different periods. For this purpose we introduce a time-dependent assignment map: π_{ijtap}, proportion of flow f_{ijt} travelling over link a in period p.

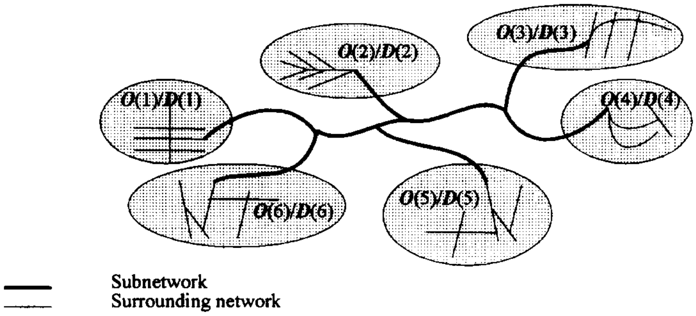

Fig. 3. Hierarchical network structure.

As a result, to the observed link flows applies:

$$y_{ap} = \sum_{i,j,t} f_{ijt}\pi_{ijtap} \tag{8}$$

with: y_{ap}, observed flow on link a in period p.

4.3. Solution procedure

The model described above converts the under-specified problem into an over-specified problem; that is, it will generally not be possible to compute a set of production and attraction abilities in such a way that all observations are exactly matched by the model predictions. To obtain an optimal fit between model and observed link volumes a weighted least squares error approach is proposed. In this approach the errors are weighted with their approximated variances, which, in turn, are approximated by observed values. This results in the problem of minimizing:

$$\sum_a \sum_p \left(y_{ap} - \sum_{i,j,t} f_{ijt}\pi_{ijtap}\right)^2 \Big/ y_{ap} \tag{9}$$

using:

$$f_{ijt} = a_i(t)b_j(t + c_{ij})\left(\sum_k F^k \delta_{ij}^k\right) \tag{10}$$

The solution procedure is expressed in the following vector notation. Denote

- m number of entrances to the subnetwork
- n number of exits of the subnetwork
- A number of links on which the volumes are observed
- T number of departure periods
- P number of observation periods

and

- $\bar{y}$ vector of observed flows for all periods (length AP)
- $\bar{f}$ vector of subnetwork flows for all periods (length mnT)
- $\boldsymbol{U}$ dynamic subnetwork assignment map (height mnT, width AP). $\boldsymbol{U}_{rs}$ is the proportion of the rth element of flow vector $\bar{f}$ that contributes to the sth component of the observation vector $\bar{y}$
- $\boldsymbol{Y}$ a diagonal matrix containing the elements of $\bar{y}$ (height AP)

The equivalent of (9) in vector notation is:

$$\text{minimize:} \quad j(\bar{a}, \bar{b}, \bar{F}) = (\bar{y} - \boldsymbol{U}'\bar{f})'\boldsymbol{Y}^{-1}(\bar{y} - \boldsymbol{U}'\bar{f})$$
$$\bar{a} > 0, \bar{b} > 0, \bar{F} > 0 \tag{11}$$

with

- $\bar{a}$ vector of production abilities for all periods, $\bar{a}_{(t-1)J+i} = a_i(t)$
- $\bar{b}$ vector of attraction abilities, $\bar{b}_{(t-1)I+j} = b_j(t)$
- $\bar{F}$ vector of deterrence values, $\bar{F}_k = F^k$

The minimization of this expression is a non-linear problem, subject to non-negativity constraints. In a first attempt to solve this problem a conjugate gradient method was used (see Bazaraa et al., 1993). Although this method converged, computation times were too high to yield results for problems of realistic size within reasonable time. To overcome this problem an alternative solution algorithm is proposed. This algorithm sequentially solves the production abilities, $\bar{a}$, the attraction abilities, $\bar{b}$, and the coefficients in the deterrence function, $\bar{F}$. Each time two sets of parameters remain constant while the third set is solved, resulting in the following iterative procedure:

$$\bar{a}^{n+1} = \underset{\bar{a}}{\text{argmin}}\ j(\bar{a}, \bar{b}^n, \bar{F}^n) \tag{12a}$$

$$\bar{b}_{n+1} = \underset{\bar{b}}{\text{argmin}}\ j(\bar{a}^{n+1}, \bar{b}, \bar{F}^n) \tag{12b}$$

$$\bar{F}^{n+1} = \underset{\bar{F}}{\text{argmin}}\ j(\bar{a}^{n+1}, \bar{b}^{n+1}, \bar{F}) \tag{12c}$$

Provided that the function j is convex, this procedure can be shown to yield the solution to the minimization problem (11). As each of the subproblems (12a–c) is quadratic, the solution to these subproblems can be determined in one step. For example if in (12c) the vectors $\bar{a}^{n+1}$ and $\bar{b}^{n+1}$ are fixed diagonal matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$ exist such that:

$$\bar{f} = AB\bar{F} \tag{13}$$

And expression (11) is converted into:

$$\text{minimize:} \quad j(\bar{F}^{n+1}) = (\bar{y} - U'AB\bar{F}^{n+1})' \times Y^{-1}(\bar{y} - U'AB\bar{F}^{n+1}) \tag{14}$$

The value of $\bar{F}^{n+1}$ that minimizes expression (14) for fixed values of $\bar{a}^{n+1}$ and $\bar{b}^{n+1}$ can be determined directly using:

$$\bar{F}^{n+1} = (BAUY^{-1}U'AB)^{-1}BAUY^{-1}\bar{y} \tag{15}$$

An identical approach can be used to solve (12a) and (12b). This procedure of solving sequentially for $\bar{F}$, $\bar{a}$ and $\bar{b}$ turns out to be much more CPU-time efficient for this particular problem than the conjugate gradient method.

5. Dynamizing

As there is a gap between the ability of the OD-prediction module (30-min aggregation) and the need of a dynamic prediction system (5-min aggregation), additional steps are necessary. Reducing the aggregation interval of the OD-prediction module would invalidate the underlying model. Therefore a heuristic procedure is followed to convert the 30-min estimates into 5-min estimates. This procedure, referred to as 'dynamizing', is based on the assumption that processes resulting into travel demand have a constant or slowly changing influence.

From an observational point of view, the traffic patterns are random, and local and temporary deviations from the general traffic pattern may occur. These deviations may emerge as peaks or drops in on-ramp volumes, but also as deviations in the distribution of trips for the off-ramps. The first phenomenon can be observed directly and prior to the execution of trips, but the second can only be observed indirectly and posterior to the execution of trips.

As traffic predictions need to be made prior to the execution of trips, it is not realistic to assume knowledge on the second phenomenon. To obtain a 'best' guess of the OD matrix at the time an on-ramp volume is observed, the on-ramp volumes are multiplied by split proportions that are derived from the 30-min aggregated matrices.

6. Dynamic assignment

The dynamic assignment model, 3DAS, is based on work initiated by Hamerslag (1989). The basic feature of a dynamic assignment model is the partitioning of time into small time slices, usually referred to as periods. Over the last 2 years the model has been improved, particularly in its dynamic aspects. The 3DAS model is described in detail in De Romph (1994).

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

The basic iteration in Fig. 4 is essentially the same as that used for *static* assignment models. The difference lies in the 'all-or-nothing-in-time' module. In this module an extra iteration over the departure period is needed; finding the shortest path and performing the assignment have to be done *in real time*.

The paths are defined using the travel time 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 present on a link only for a part of the period, 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

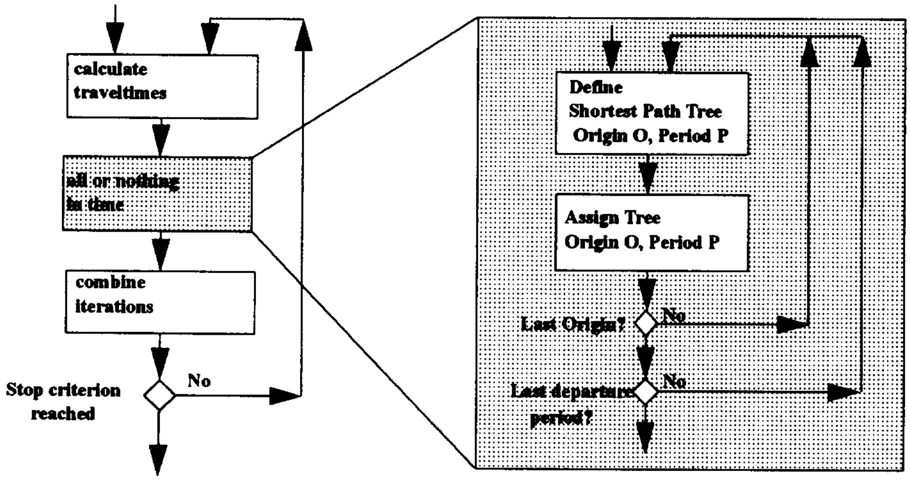

Fig. 4. The iteration scheme.

periods and one should be assigned entirely for each individual period.

At the start of each iteration the travel times on the links are derived from the load of the previous iteration. For each link the travel time is calculated with a speed–density function. Speed–density functions are used instead of traditional travel time intensify functions so as to model the presence decreasing flow of congestion. The conservation of traffic and the continuity of flow is maintained; overflows are 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 works (De Romph et al., 1992). 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 of the following form is used:

$$\nu(\rho) = \begin{cases} \nu^{\max} \cdot \left(1 - \dfrac{\alpha\rho}{\rho^{\max}}\right) & 0 < \rho < \rho^{\mathrm{crit}} \\ \phi \cdot \left(\dfrac{1}{\rho} - \dfrac{1}{\rho^{\max}}\right)^{\beta} & \rho^{\mathrm{crit}} \leqslant \rho < \rho^{\max} \end{cases} \tag{16}$$

In which $\nu^{\max}$ is the free-flow speed, ρ^{crit} is the critical density and $\rho^{\max}$ is the maximum density. The maximum density represents a no-motion traffic-jam. The value of ϕ is chosen to make the function continuous at ρ^{crit}. The parameter α influences the steepness of the linear first part of the function, the parameter β influences the curvature of the second part of the function. For each link in the network different parameters are possible.

7. Data preparation

The data sets that are available for implementation of the traffic prediction system in the pilot area were described in Section 2.

For the proper functioning of the OD-estimation module several additional items need to be specified:

- A list of link-number/loop-number pairs, specifying the correspondence between MTM loop-numbers and beltway-network link-numbers. This list was made on the basis of detailed maps of the area (Rijkswaterstaat, 1993).
- A list of origin nodes and a list of destination nodes.

- An 'origin/on-ramp' table, specifying the feeder links that correspond to the origin zones. The specified links are used in the 'dynamizing' procedure, see Section 5.
- A 'missing ramps' table specifying which monitored link volumes should be added or subtracted to obtain a certain non-monitored link volume. The use of this table is twofold: firstly the table is needed to compute on-ramp volumes that are required by the 'dynamizing' procedure, see Section 5, secondly the table can be used to replace two largely overlapping observations by two observations with a lower degree of redundancy. For example, two inner link volumes are replaced by one inner link and one exit volume.

After specification of these items, a number of automated actions is performed:

- Preparation of the MARE data. The MARE data consist of a series of 1-min values. Occasionally data are missing or erroneous, which can be detected via the time-stamp and the error-status attributes. Missing data are replaced by averages of previous periods.
- Calculation of the assignment map based on shortest paths for use in the OD-estimation module.
- Merging redundant data. If one link has several induction loops, these loops generate largely redundant data. This fact can be established by inspection of the assignment map. If certain pairs of induction loops have only a combined appearance in the assignment map, these loops are merged by averaging.

Another preparatory step involved the visual inspection of network and input data. All elements relevant to the OD-estimation and dynamic assignment procedures have been made available in a graphical environment that contains visualizations of network, origins, destinations, routes, loop-data, distributions functions, assigned link volumes and predicted speeds.

Visual inspection of the network has revealed a number of shortcomings in the original input data. Induction loops producing improbable values were dropped. Routes containing U-turns were detected using a selected link analysis. Routes containing U-turns are regarded as a highly unlikely phenomenon on a motorway network, which can only occur because of omission of the secondary network. These routes were prevented from being optimal by the addition of dummy links.

8. Results

The results of the three scenarios proposed in Section 3 are described in the next three sections.

8.1. Morning rush-hour on Thursday, April 28th

A dynamic assignment was carried out based on the dynamic OD matrix and the speed–density functions. Calculation started at 06:00. At this time the network was almost empty. During calibration numerous modifications were made to the network. A wrong number of lanes at one location can disturb traffic flow over a large part of the network. Several on-ramps were not well represented, and at some locations individual speed–density relationships were required. Detailed maps were required to figure out the exact lay-out of the network. Finally a fairly good reproduction of the actual traffic flow was achieved. Two different locations in the network are shown to give an impression of the results. Fig. 5 gives flow and speed at two locations near the "Coentunnel", referring to the network shown in Fig. 2.

The left side of Fig. 5 shows a location with traffic queuing before entering the "Coentunnel". The flow pattern is reproduced fairly well with the combined dynamic OD estimation/traffic assignment, except in the last six periods, where model predicted speed increases, while in reality speed remains low. The right side of Fig. 5 shows a location in the other direction with traffic leaving the "Coentunnel" and travelling north. This location shows a free flow situation and is a good reproduction.

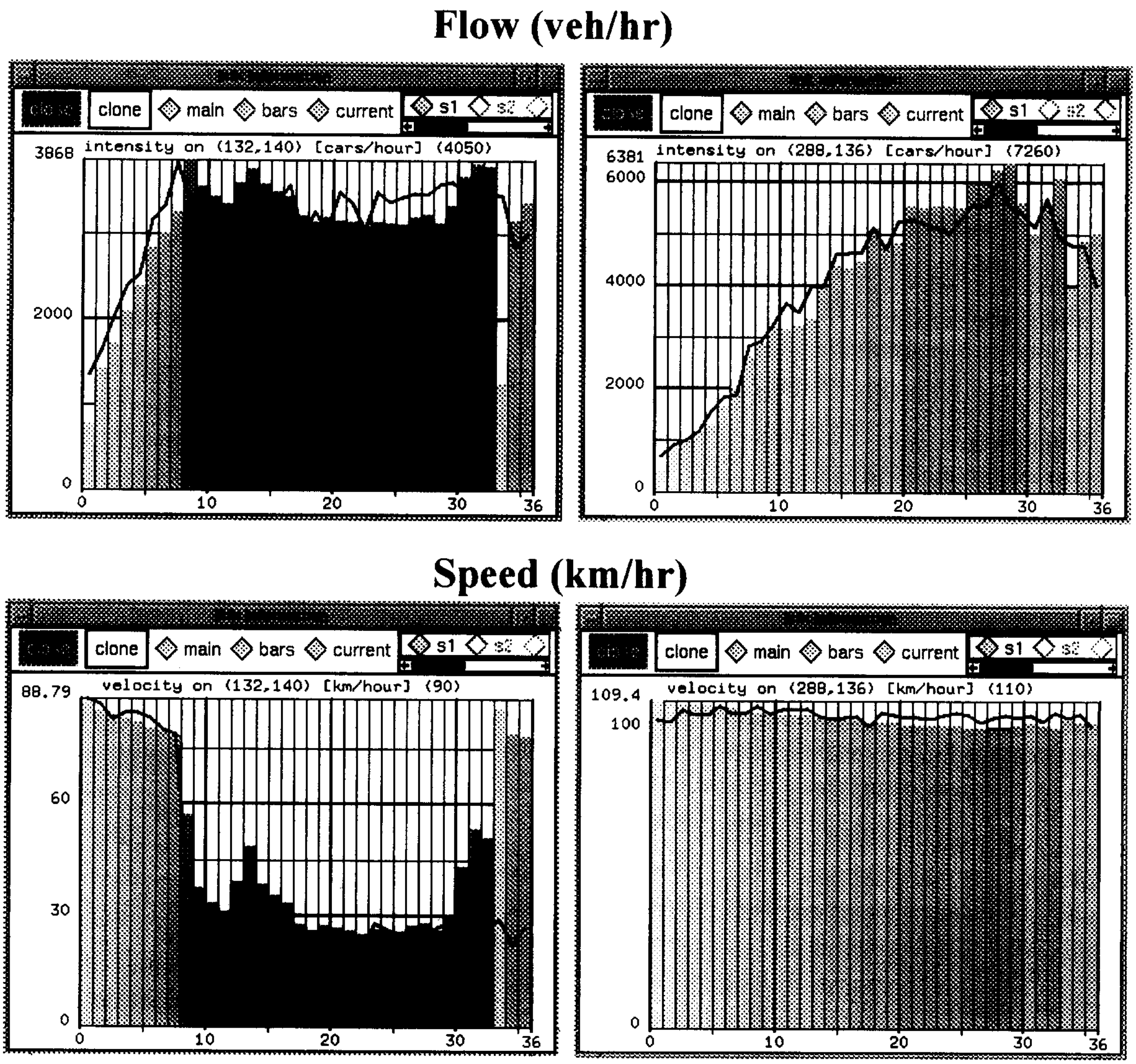

Fig. 5. Model results (bars), compared with the actual measured traffic flow (solid line) at two locations near the "Coentunnel".

8.2. Morning rush-hour on Thursday, April 14th

A dynamic assignment was made for the morning rush-hour of April 14th based on the dynamic OD matrix of April 28th. Calculation started at 06:00. At 06:50 ($t=10$) an accident occurred north of the "Nieuwe Meer" intersection northbound, see Fig. 2. This accident shows clearly in the induction loop data. Downstream flow is almost zero and a large traffic jam has built up in several directions. At 07:35 ($t=19$), a larger downstream flow was again measured and the traffic jam builds down from the start point, but it still builds up from the tail of the traffic jam. At 08:00 ($t=24$) the tail of the traffic jam reached the "Watergraafsmeer" intersection. The accident was introduced into the model by setting the maximum density to 1 vehicle per km during the accident for the appropriate section.

The traffic flow was reproduced fairly well, but not as well as for April 28th. Two different locations near the accident are shown in Fig. 6 to give an impression of the results near the traffic jam.

The left side of Fig. 6 shows a location downstream of the accident, and shows a restored flow and speed. Flow is low for the first 20 periods. Around the 20th period (07:35), the freeway is clear again and flow increases. The right side of Fig. 6 shows a location upstream of the accident. The flow pattern is quite complicated here. The first decrease in speed around the 18th period is the result of the accident. The second decrease in speed, around the 27th period is the result of a traffic jam that appears east of "Nieuwe Meer", for trips going east and has nothing to do with the accident. This traffic

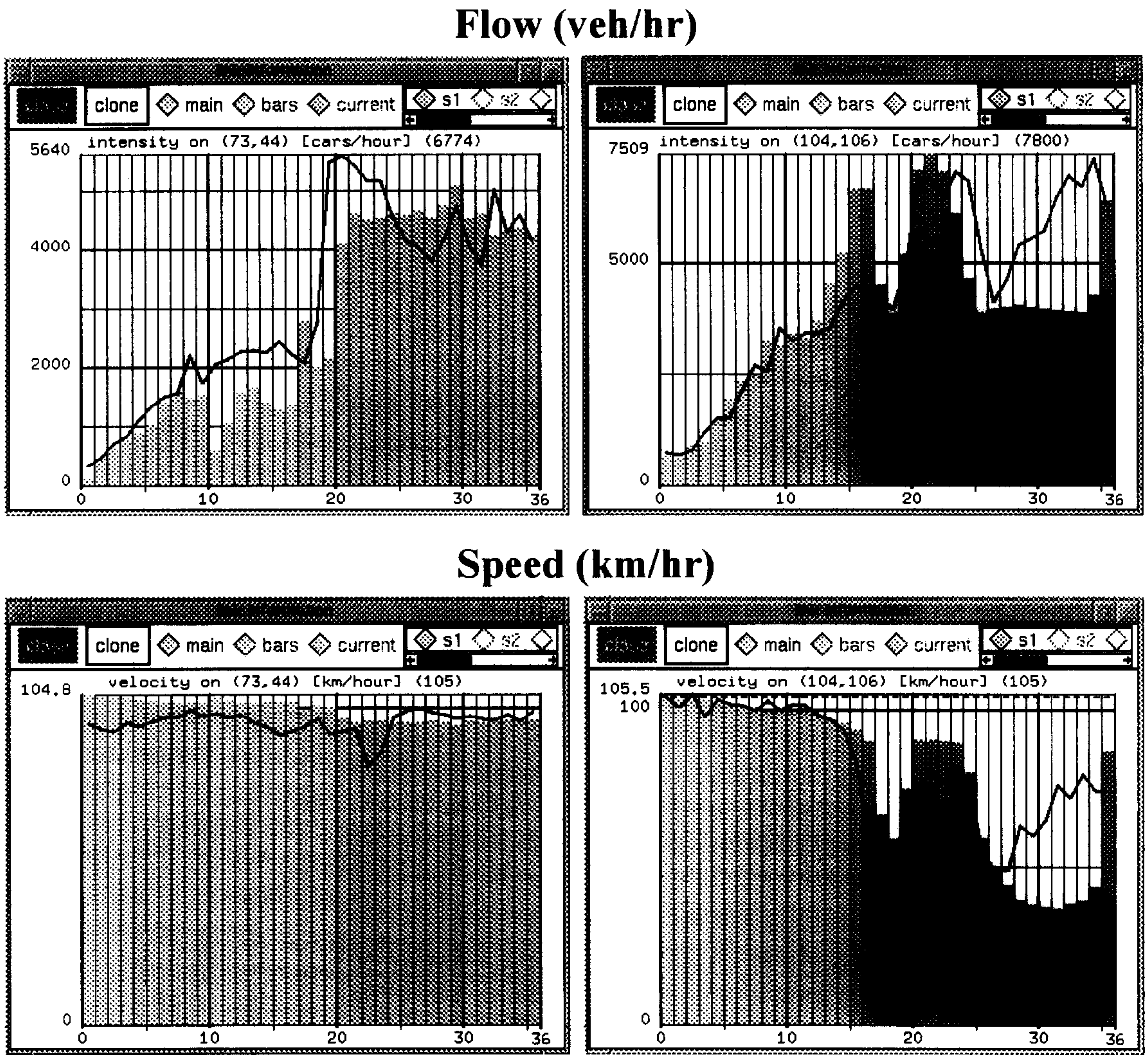

Fig. 6. Model results (bars), compared with actual measured traffic flow (solid line) at two locations near an accident that occurred at $f = 0$; the figures on the left refer to downstream of the accident, travelling north and, the right figures to upstream of the accident travelling east.

jam was not reproduced exactly by 3DAS, and for this reason speed is somewhat underestimated around the 30th period.

8.3. *A prediction based on the historical average*

The two scenarios described were validated using data measured for 3 weeks by induction loop detectors at 141 locations. Five days were selected from this data base which contained no major disturbances. The traffic patterns for these days were summed and averaged. The resulting data set represents the 'average traffic pattern' for the morning rush-hour from 06:00 until 09:00 for a 'normal' weekday.

Including the data of the two scenarios described earlier, there are now five different sets of speeds and intensities for the morning rush-hour:

I. The first set represents the actual measured situation from 06:00 until 09:00 with 36 periods of 5 min on Thursday, April 28th.
II. The second set represents the predicted results for Thursday, April 28th.
III. The third set represents the actual patterns from 06:00 until 09:00 with 36 periods of 5 min on Thursday, April 14th.
IV. The fourth set represents the prediction results for Thursday, April 14th.
V. The fifth set represents the average traffic pattern for a morning rush-hour from 06:00 until 09.00 in 36 periods of 5 min.

A comparison can be made between the re-

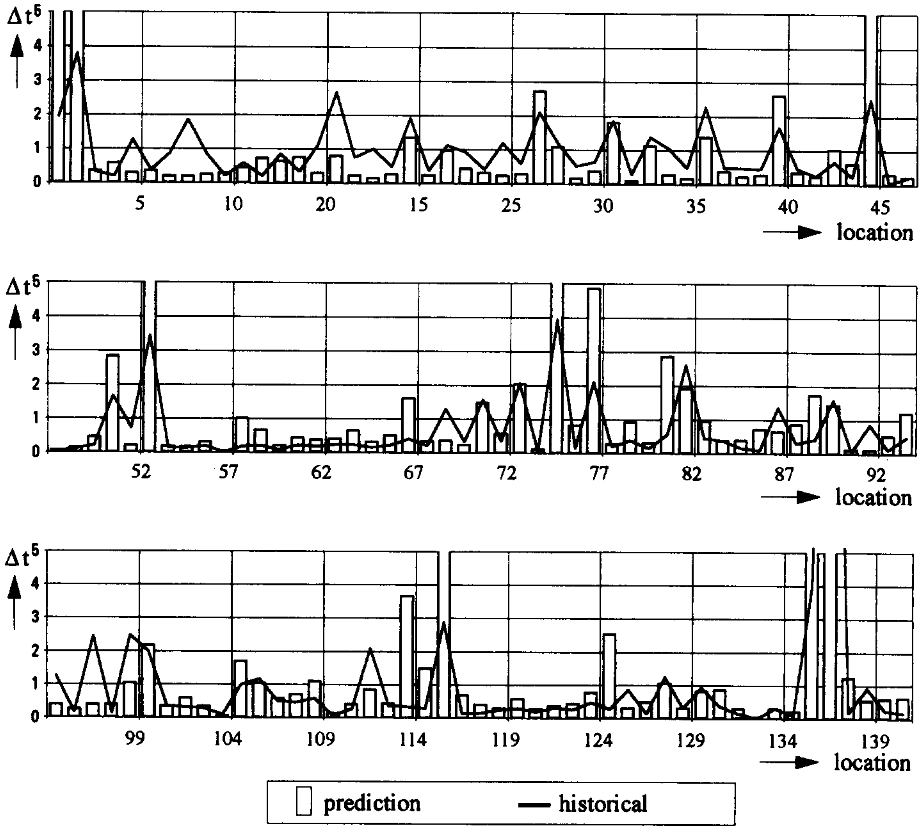

Fig. 7. The absolute error in minutes for the prediction system (set II) in bars, and for the historical average (set V) as a solid line. The observed data of April 28th (set I) are used as a reference.

sults of the prediction system and the historical average for a normal day and for a day with an accident.

As a measure of effectiveness we use the discrepancy between the predicted or averaged travel times and the observed travel times, i.e. comparisons take place between the sets I and II, I and V (normal day), III and IV, and III and V (accident scenario).

This discrepancy is expressed for each location/link separately. The absolute error for a specific link between scenario I and scenario II may be computed as:

$$\sum_{p} |t_{I}^{p} - t_{II}^{p}| \tag{17}$$

where p is the period number and t_i^p the travel time in period p for scenario i.

Fig. 7 shows these absolute errors for each location and each period. The absolute error for the predicted values II is represented by a bar, and in the same figure, the absolute error for the historical average V is shown as a solid line. Fig. 7 shows that the prediction system is capable of reproducing the actual situation of April 28th, as for most links the absolute error in travel time summed over the entire time span is only a few minutes. The same figure, however, shows that the historical average is at least as good. To make a prediction for the morning rush-hour of April 28th, it might be better to use the historical average of several similar days. Fig. 8 shows the same entities as Fig. 7, but now for April 14th. The absolute error for the prediction IV is represented by a bar for each location, and in the same figure, the absolute error for the

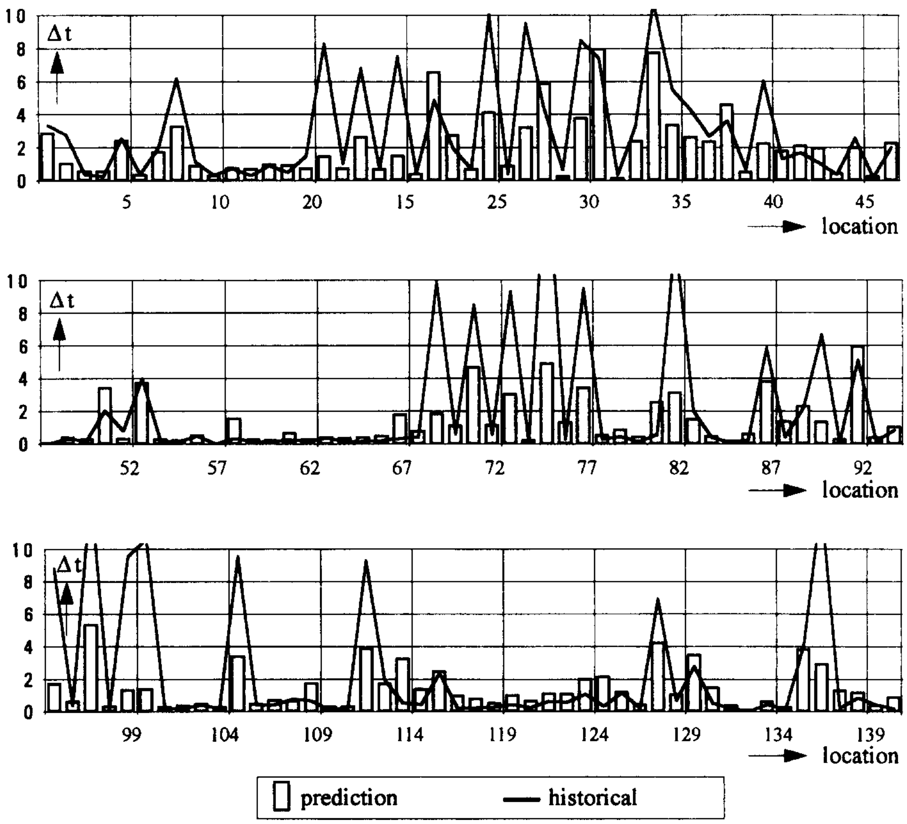

Fig. 8. The absolute error in minutes for the prediction system (set IV) in bars, and for the historical average (set V) as a solid line. The observed data of April 14th (set III) are used as a reference.

historical average V is shown as a solid line. Fig. 8 shows that the prediction system is also capable of reproducing the actual situation on April 14th, while the historical average gives poor estimates for several locations.

9. Conclusion

This article illustrates that the proposed traffic forecasting system can be used to make predictions of traffic flows in a network. OD estimation and the 3DAS model are capable of reproducing a traffic pattern on a specific day, including accidents.

Correct representation of the network and speed–density functions is very important. Specifying the wrong number of lanes at a road section can have serious consequences for the entire network. These errors can be located easily and corrected using the available graphic software.

The model was tested on 2 different days. One day represented a 'normal' morning rush-hour, the other day represented a morning rush-hour with an accident which blocked all lanes for 45 min.

Comparing the traffic flow pattern of a specific day with the historical average pattern, it can be concluded that for 'normal' days the system can reproduce the situation, but that the historical pattern is at least as good. This is explained by the fact that predictions obtained from historical data are not hindered by mis-specification of OD-demand and network characteristics.

The day with the accident however, was better reproduced by the prediction system than by the historical data. These results suggest that the ideal forecasting procedure is a combination of both approaches.

Acknowledgements

Part of the research upon which the article is based was supported by EC grant no. ERBCHGCT930412. This grant supports a 1-year fellowship for Nanne van der Zijpp in the framework of the Human Capital and Mobility Programme.

References

Adviesdiemst Verkeer & Vervoer, Transport Research Centre, 1994, *Motorway Signalling, AVV-Brochure* (Directorate-General for Public Works and Water Management, The Netherlands).

Bazaraa, M.S., H.D. Sherali and C.M. Shetty, 1993, *Non-linear Programming, Theory and Algorithms*, 2nd edn. (Wiley).

Bell, M.G.H., D. Inaudi, J. Lange and M. Maher, 1991, Techniques for the dynamic estimation of OD-matrices in traffic networks, in: European Commission, *Proceedings of the Drive Conference*, 4–6 February, 1991 (Elsevier Science B.V., Amsterdam, The Netherlands).

Cascetta, E. and S. Nguyen, 1988, A unified framework for estimating or updating Origin–Destination matrices from traffic counts, *Transportation Research-B*, 22B(6), 437–455.

Cascetta, E., D. Inaudi and G. Marquis, 1993, Dynamic estimators of Origin–Destination matrices using traffic counts, *Transportation Science*, 27(4), 363–373.

Cremer, M. and H. Keller, 1981, Dynamic identification of flows from traffic counts at complex intersections, in: V.F. Hurdle et al., *Proceedings 8th Int. Symposium on Transportation and Traffic Theory* (University of Toronto Press, Toronto, Canada).

De Romph, E., 1994, *Dynamic traffic assignment, theory and applications, Ph.D. Thesis* (Delft University of Technology).

De Romph, E., H.J.M. Van Grol and R. Hamerslag, 1992, A dynamic assignment model for short term prediction, in: N.H. Gartner and C. Improta, eds., Urban traffic networks: flow model and control, *Second International Capri Seminar on Urban Traffic Networks*, 5–8 July 1992, Italy (Springer-Verlag, Berlin).

Dienst Verkeerskunde, 1974, *Basisnetwerk Programmatuur, Reistijden* (Rijkswaterstaat, Centrum voor Vervoersplanen).

Hamerslag, R., 1989, Dynamic assignment in the three dimensional timespace, in: *Transportation Research Record 1220* (National Research Council).

Hamerslag, R. and L.H. Immers, 1988, Estimation of trip matrices: shortcomings and possibilities for improvement, in: *Transportation Research Record 1203* (National Research Council) 27–39.

Keller, H. and G. Ploss, 1987, Real-time identification of O–D network flows from counts for urban traffic control, in: N.H. Gartner and N.H.M. Wilson, eds., *Proceedings 10th Symposium on Traffic Theory* (Elsevier Science B.V., Amsterdam, The Netherlands).

Maher, M.J., 1983, Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach, *Transportation Research-B*, 17B(6), 435–447.

Nihan, N.L. and G.A. Davis, 1987, Recursive estimation of Origin–Destination matrices from Input/Output counts, *Transportation Research-B*, 21B(2), 149–163.

Ortúzar, J.D. and L.G. Willumsen, 1990, *Modelling Transport*, 2nd edn. (John Wiley and Sons, Chichester, UK).

Rijkswaterstaat, 1993, *Portalen MCSS* (Rijkswaterstaat Directie Noord Holland).

VanAerde, M., B. Hellinga and G. MacKinnon, 1993, QUEENSOD: a method for estimating time varying Origin–Destination demands for freeway corridors/Networks, presented at TRB annual meeting, Washington, DC.

Van Der Zijpp, N.J., 1996, A Bayesian framework for dynamic Origin–Destination matrix estimation on motorway networks, *Ph.D. Thesis* (Delft University of Technology).

Van Der Zijpp, N.J. and R. Hamerslag, 1994, An improved Kalman filtering approach to estimate Origin–Destination matrices for freeway corridors, in: *Transportation Research Records 1443*, 54–64 (National Research Council).

Van Grol, H.J.M., 1992, Traffic assignment problems solved by special purpose hardware with emphasis on real time applications, *Ph.D. Thesis* (Delft University of Technology).

Van Zuylen, H.J. and L.G. Willumsen, 1980, The most likely trip matrix estimated from traffic counts, *Transportation Research-B*, 14B, 281–293.

Willumsen, L.G., 1984, Estimating time-dependent trip matrices from traffic counts, in: B. Volmuller and R. Hamerslag, eds., *9th International Symposium on Transportation and Traffic Theory* (VNU Science Press), 397–411.

Biographies: Nanne J. VAN DER ZIJPP works as a researcher at the Civil Engineering Department at Delft University of Technology. He received his Ph.D. for a thesis on dynamic OD-estimation in 1996, and recently completed a one-year fellowship at the Centre for Transport Studies, London. His major research interests are OD-matrix estimation, dynamic assignment and traveller information systems.

Erik DE ROMPH is working as a project leader at "Goudappel Coffeng", a traffic and transportation consultancy. He completed his masters in Computer Science in 1989 and received his Ph.D. for a thesis on dynamic traffic assignment in 1994 from Delft University of Technology. His major research interests are dynamic traffic management and dynamic traffic assignment.