ELSEVIER

International Journal of Forecasting 13 (1997) 73–85

international journal of forecasting

Travel time estimation in the GERDIEN project

Bart Van Arem[a,*], Martie J.M. Van Der Vlist[a], M. (Rik) Muste[b], Stef A. Smulders[b]

[a]*TNO Institute of Infrastructure, Transport and Regional Development, P.O. Box 6041, 2600 AA Delft, Netherlands*

[b]*Transport Research Centre AVV, Directorate-General for Public Works and Water Management, Ministry for Transport, Public Works and Water Management, Rotterdam, Netherlands*

Abstract

In order to produce traffic predictions up to 1 h ahead for a motorway network, it is essential to produce reliable estimates for the travel time over motorway sections. This article describes a model for estimating travel times on motorway sections 3–5 km in length. The model uses measurements from inductive loop detectors and is based on a linear input–output ARMA model representation. The model is evaluated using results from a field study.

The evaluation data set covers traffic situations with both 'normal' congestion and congestion due to accidents or incidents. The observed and estimated travel times are very close in cases of normal congestion. Travel time estimates in cases concerning accidents showed large deviations, but in these cases there were no travel time observations to verify their accuracy.

Keywords: Short-term traffic prediction; ARMA model; Evaluation; Travel times

1. Travel time estimation and traffic forecasting

This article considers the issue of the estimation of the travel time of drivers entering a section of a motorway. The current practice of measuring traffic flows is based on inductive loop detectors embedded in the road surface. This technology enables the measurement of passage time instants and speeds of passing vehicles on each lane of a motorway. The travel time between two such measurement points is one of the most basic quantities available to describe a traffic flow that cannot be measured directly, but has to be estimated using a model.

The estimation of the travel time of a driver entering a motorway section is closely related to other traffic prediction methods. Firstly, as indicated above, it is one of the most basic traffic predictions linking different points in a road network. Thus, a model for travel time estimation may provide a benchmark for the evaluation of a model predicting traffic with a longer time horizon or for larger road network configurations (Ben-Akiva et al., 1992). Secondly, travel time estimates are also used as input in a prediction method operating on a wider scale. Therefore, they are necessary to provide the basic input

* Corresponding author.

PII S0169-2070(96)00702-9

before models on a wider scale can be made operational. The accuracy of travel time estimates also restricts the possible accuracy of a prediction method operating on a larger scale.

The work described here was conducted as part of the GERDIEN project in the DRIVE-II programme (Van Arem et al., 1993, Blonk et al., 1994). One of the functions developed and implemented in GERDIEN is the monitoring of the current and expected traffic situations. The objective is to provide a traffic operator with information to support decisions concerning output warnings to drivers, travellers' information or to take controlling actions. The Network State Monitoring and Prediction system (NSMP) has been designed and built to achieve this objective.

The NSMP system is an on-line system. Measurement data are fed in continuously. As input, the system uses speed, traffic volume per vehicle length class and occupancy on a 1-min basis per lane produced by loop detectors. Every minute, the most recent data are fed into three submodels which produce calculation results. The three submodels implemented at present are models for estimating current roadway capacity, for detecting congestion and estimating travel times and for predicting the near-future traffic situation. The article focuses on the second aspect.

The available literature on travel time estimation is extensive (see e.g. Cremer and Schütt, 1990; Stephanedes and Chassiakos, 1993). Most of these models have, however, been developed and tested for relatively short road sections (up to 5 km). In the GERDIEN context, a specific requirement was that the model should be applicable to road sections up to 8 km. Therefore, several models from the literature were investigated with respect to their suitability for the specific requirements imposed. This analysis resulted in the combination of an ARMA model for detecting traffic congestion with a counting mechanism to estimate the travel time and the time delay. This model serves as the starting point for the work described in this article.

The method used for estimating travel times is described in Section 2. Section 3 will describe the evaluation data. The results are given in Section 4. Section 5 contains conclusions.

2. The GERDIEN model for estimating travel times on motorway sections

2.1. General description

The model for estimating travel times is based on results of earlier research by the Transport Research Centre of the Dutch "Rijkswaterstaat". The model is applied to motorway sections with or without on- and off-ramps. Fig. 1 gives an example of a three-lane road section with single-lane on- and off-ramps. In the GERDIEN pilot the length of a motorway section is 3–5 km. In the original application area of the model by "Rijkswaterstaat" the section length may extend to 10 km.

As input, the model uses speed and traffic volume per minute of all entry and exit points to a section, including the ramps. It is applied every minute, and consists of two steps. The first step concerns the detection of any deviation from normal traffic flow. The second step concerns the estimation of travel time. If the first step has indicated that the traffic flow is normal, the travel time is estimated by a default free-flow travel time. If, however, the first step indicates that the traffic flow is disturbed, the second step starts by estimating the number of vehicles on the section in a cumulative way by a simple counting procedure. This estimate is used for estimating the travel time. In Sections 2.2 and 2.3 these two steps are described in more detail. Section 2.4 describes how the model was implemented.

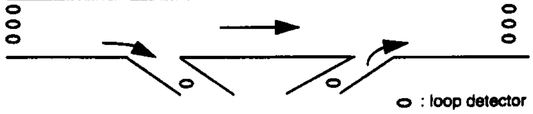

Fig. 1. Example of a road section.

2.2. The model for detecting congestion (model description)

Suppose that minutes are numbered $m = 1, 2, 3, \ldots$, and that the model is invoked at the completion of minute m. The model combines a speed criterion with an ARMA model criterion. It gives an output signal $I_c(m) = 0$ or 1 if the traffic is free-flowing or congested, respectively. These criteria are described in more detail below.

The average speed of vehicles entering (A) or leaving (B) the section on the main carriageway during minute m is denoted by $V_A(m)$ and $V_B(m)$, respectively. If

$$\min(V_A(t), V_B(t)) < V_{\max} \tag{1}$$

with $V_{\max}$ a suitably chosen threshold value, traffic is assumed to be congested at the entry and/or exit point, and $I_c(m)$ is assigned the value 1. If this is not the case, traffic is considered to be flowing freely at both the entry and exit point. However, that does not exclude the possibility of congested traffic between the entry and exit point. The second part of the congestion detection module is designed to detect this type of congestion.

The second part of the model assumes that measurement data are available for minutes $m - M_1 + 1, \ldots, m$, with M_1 a fixed constant. The choice of M_1 depends on two criteria. Firstly, the model needs to respond quickly to congestion, so M_1 should not be too large. Secondly, the model should be based on a sufficient amount of data in order to represent traffic in a realistic manner (M_1 not too small). Explorative experiments revealed that M_1 equal to 30 resulted in good compromise. These data are used to estimate an ARMA model (Chatfield, 1989). The model estimated has the following form:

$$Y(t) = \sum_{i=1}^{n_a} - a_i Y(t-i) + \sum_{j=1}^{n_b} b_j X(t-j-\Delta_1) + E(t)\,,$$
$$t = m, m-1, \ldots, m - M_{\text{eff}} + 1 \tag{2}$$

where

$X(t)$	random variable denoting the total traffic volume into the section in minute t
$Y(t)$	random variable denoting the total traffic volume out of the section in minute t
$E(t)$	random variable denoting the model error in minute t
n_a	number of autoregressive parameters
a_i, $i = 1, \ldots, n_a$	autoregressive parameters
n_b	number of moving average parameters
b_j, $j = 1, \ldots, n_b$	regression parameters
Δ_1	built-in time delay between inflow and outflow
M_{eff}	effective number of minutes for which the measurement data are available, $M_{\text{eff}} = \min(M_1 - n_a, M_1 - n_b - \Delta_1)$. It is assumed that $M_{\text{eff}} > 0$

In words: the total outflow in minute t is a function of the past total outflow and the past total inflow with a built-in time delay. The parameters of the mode are estimated in such a way that (2) gives a good description of the traffic in the most recent M_1 minutes. Eq. (2) assumes that in the given minutes, outflow and inflow are in balance. Therefore it does not contain a constant term. For the estimation of the parameters of this model several approaches are available (Chatfield, 1989). In this case, a practical approach is used. Suppose that $x(t)$ and $y(t)$ are the observed traffic volumes in minute t into and out of the section, then the observed error is

$$e(t) = y(t) + \sum_{i=1}^{n_a} a_i y(t-i) - \sum_{j=1}^{n_b} b_j x(t-j-\Delta_1)\,,$$
$$t = m, m-1, \ldots, m - M_{\text{eff}} + 1 \tag{3}$$

Assuming the parameters n_a, n_b and Δ_1 to be fixed, the remaining parameters a_i, $i = 1, \ldots, n_a$, b_j, $j = 1, \ldots, n_b$ are computed by minimizing the total squared observed error, which, using (3) leads to the problem of minimizing

$$\sum_{t=m-M_{\mathrm{eff}}+1}^{m}\left[y(t)+\sum_{i=1}^{n_a}a_i x(t-i) - \sum_{j=1}^{n_b}b_j x(t-j-\Delta_1)\right]^2 \tag{4}$$

That is, standard least squares estimators are used.

If the model has been successfully estimated, it is used to check what type of traffic it represents. This is done by computing the impulse response function (Box and Jenkins, 1970). The impulse response function describes how, according to the model, the outflow reacts to a unit pulse in the input. It may be compared to a description of how a compact platoon of vehicles that enter a road section leave the section in a dispersed way, see Fig. 2.

For the model given by (2) the response is computed (Box and Jenkins, 1970, p. 354) under the following assumptions:

- The input is given by $x(u)=1$, $x(t)=0$ for $t\neq u$,
- $e(t)=0$ for all t,
- the system is initially empty, so there is no output before $t=u$, i.e. $y(t)=0$ for $t<u$.

The height of the impulse response function after k, $k=0,1,2,\ldots$ minutes is denoted by h_k. Evidently, $h_k=y(u+k)$, and $y(u+k)$ is computed iteratively from (2) under the above assumptions and rewritten as:

$$y(u+k)=\sum_{i=1}^{n_a}-a_i y(u+k-i) + \sum_{j=1}^{n_b}b_j x(u+k-j-\Delta_1) \tag{5}$$

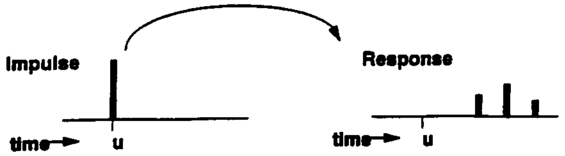

Fig. 2. The impulse response function.

The impulse response function is used in two ways to verify whether the traffic is congested.

Firstly, the sum of the responses over a finite number of minutes $k=0,\ldots,F$ is computed. If the traffic is free flowing then the total response within a reasonable number of minutes should approximately be equal to one. If there is congestion, more vehicles are going in than are going out and a value smaller than 1 is expected. In the case of resolving congestion a value more than 1 is expected. Therefore, if

$$\sum_{k=0}^{F}h_k<1-\epsilon \tag{6}$$

with ϵ a suitable threshold constant, then congestion is suspected.

Secondly, the average response time is computed. The average response function can be used as an estimate for the average travel time on the road section under consideration. Denote by $T_{\max}$ a suitable threshold for the travel time to distinguish free-flowing traffic from disturbed traffic. If

$$\frac{\sum_{k=0}^{\infty}kh_k}{\sum_{k=0}^{\infty}h_k}>T_{\max} \tag{7}$$

congestion is suspected. In order to evaluate (7) it is required that both denominator and numerator converge. In the application of (7), both denominator and numerator were evaluated for a finite number of minutes.

If both (6) and (7) are true, $I_c(m)$ is assigned the value 1.

2.3. The model for estimating the travel time

2.3.1. Basic principle

Denote by $S(m)$ the travel time encountered by an arbitrary vehicle entering the section under consideration in minute m. It is assumed that the travel time can be represented by:

$$S(m)=T_f+W(m) \tag{8}$$

where T_f is a constant representing the free-flowing travel time on the section and $W(m)$ denotes the time delay encountered by an arbitrary vehicle entering the section in minute m. If the model for detecting congestion produces an output $I_c(m) = 0$, which means no congestion, it is assumed that $W(m) = 0$, and the travel time is estimated by $S(m) = T_f$.

If the model for detecting congestion has resulted in suspected congestion $I_c(m) = 1$, the time delay $W(m)$ is estimated by computing the number of 'excess' vehicles and dividing it by the traffic volume leaving the section. Denote by $N(m)$ the number of 'excess' vehicles on the section. During free-flowing traffic ($I_c(m) = 0$), it is assumed that $N(m) = 0$. During congested traffic, $N(m)$ is computed from the recursive relation:

$$N(m) = N(m-1) + X(m) - Y(m - \Delta_2) \tag{9}$$

with Δ_2 an integer constant approximately equal to the free-flow travel time T_f. With respect to (9) the following two remarks are in order. Firstly, the recursive relation requires a starting value on the (theoretical) case of $I_c(1) = 1$. In practice $N(0) = 0$ suffices. Secondly, the constant Δ_2 is chosen in such a way, that during free-flowing traffic $X(m) - Y(m - \Delta_2)$ is approximately equal to zero. As soon as the traffic flow is disrupted, this is no longer the case and $X(m) - Y(m - \Delta_2)$ expresses the magnitude of the disruption. The time delay is computed by dividing the number of 'excess' vehicles by the outflow during the last M_2 minutes, as follows:

$$W(m) = \frac{N(m)}{\Sigma_{t=m-M_2+1}^{m} Y^*(t)} \tag{10}$$

where $Y^*(t)$ denotes the traffic volume out of the section in vehicles per minute on the main carriageway exit point during minute t. The constant M_2 is introduced in order to stabilize the estimate, since 1-min traffic volumes on dense motorways are in general too noisy. In cases where the denominator in the right-hand side of (10) is zero, a large default value $W(m) = T_L$ (e.g. 60 min) is taken. Estimate (10) assumes a homogeneous outflow. The variance in the outflow (e.g. in headways), which normally has a positive correlation with the average time delay, is implicitly neglected.

2.3.2. *Incorporation of drift correction*

The number of 'excess' vehicles $N(m)$ during minute m is a key quantity for estimating the travel time on a section. During congested conditions it is estimated from counts of in- and outflowing traffic. These counts may however be hampered by measurements errors. Experience with paired loop detectors has shown that these errors may be systematic. In addition, one could have sections where not all on- and off-ramps are equipped with measurement systems. Owing to its recursive nature the estimate (9) for $N(m)$ is sensitive to these errors which cause the estimate to 'drift' away. The method below can be used to correct for this drift.

Denote by $F(m)$ the set of the M_3 most recent minutes of free-flowing traffic on the section before minute m. Denote by $C(m)$ a correction function estimating the current ratio between in- and outflowing measured traffic. It is computed from:

$$C(m) = \frac{\Sigma_{t \in F(m)} X(t - \Delta_2)}{\Sigma_{t \in F(m)} Y(t)} \tag{11}$$

For minutes for which insufficient data of preceding free-flowing minutes are available, or for which the denominator of (11) is equal to zero, $C(m)$ is assigned a constant default value, e.g. 1.

If the correction function $C(m)$ is less than 1, probably not all inflowing traffic is measured. In this case, the inflow may be corrected, as it is known to be probably higher. If the correction function is larger than 1, the outflow has to be corrected. Denote by $X_c(m)$ and $Y_c(m)$ the corrected inflow and outflow during minute m. The corrected traffic volumes are given by:

$$X_c(m) = \frac{X(m)}{C(m)}, \; Y_c(m) = Y(m) \tag{12}$$

for $C(m) \leq 1$, and for $C(m) > 1$

$$X_c(m) = X(m), \; Y_c(m) = C(m)Y(m) \tag{13}$$

Finally, the number of excess vehicles is computed analogously to (9).

2.4. Implementation

2.4.1. Implementation environment

As mentioned in Section 1, the model for estimating travel times is part of the GERDIEN Network State Monitoring and Prediction (NSMP) system. It was implemented as an application to the GERDIEN Generic Road Management System (GRMS) (Van Arem et al., 1993; Blonk et al., 1994). GRMS is a generic term for a family of systems with common characteristics in the GERDIEN network. In the GERDIEN network, road-side data are distributed to multiple end-user systems. The GRMS can subscribe itself to available information streams from road-side systems in the network. It provides an open environment for the integration of applications, such as the NSMP system. It also provides basic functionality for the presentation of dynamic data on a graphics-map basis and for a standardized data exchange.

The traffic data collected consist of records per lane and per minute, giving speed, traffic volume, traffic composition and occupancy as well as status messages on detection errors and the time during which the loops were functioning properly. For the travel time estimation model, the Network State Monitoring and Prediction system maintains a data base with the data in a suitable format and aggregation level. If the incoming data are incomplete because of malfunctioning road-side equipment the data are filled in. If data are present for 55 s instead of 60 s, the observations are extrapolated. If the data are not present or unreliable, it is filled in by data from the previous minute.

2.4.2. The model for detecting congestion

The speed threshold for application of (1) was chosen to equal $V_{\max} = 70\,\text{km-h}^{-1}$. To evaluate the speed criterion at a location, it is necessary that both loops of a paired loop detector are in operation. If this is not the case, congestion detection is left to the ARMA model criterion. The ARMA model criterion only uses traffic volumes. For measuring traffic volume it is sufficient that only one of the loops is in operation. The numbers of autoregressive and moving average parameters n_a and n_b were assumed to be fixed system parameters. Preliminary experiments revealed that small values are satisfactory. They were chosen equal to $n_a = 1$ and $n_b = 2$. The minimization of (4) was done by solving the minimization with Δ_1 varying in a range of possible values.

Preliminary experiments indicated that the ARMA model criterion is less suitable for application in low traffic volumes. Small deviations at low traffic volumes are relatively large. They appear to result quite frequently in 'strange' estimates, resulting in false alarms. Therefore, the ARMA model criterion is applied only if the traffic volume at the entry carriageway per lane is larger than 3 vehicles $\min^{-1}$. For the computation of the sum of the impulse response according to (6), values $F = 8$ and $\epsilon = 0.1$ were found to perform adequately. For the evaluation of (7) both denominator and numerator in the left-hand side of (7) were evaluated for a finite horizon of 16 min. The parameter $T_{\max}$ was chosen equal to:

$$T_{\max} = 1.4T_f \tag{14}$$

The combination of both criteria was implemented as follows (see Fig. 3). If the speed criterion was evaluated and resulted in suspected congestion, then the model returns $I_c(m) = 1$. If the speed criterion was evaluated but (1) was found to be untrue, then $I_c(m) = 0$ and the ARMA model criterion is applied, which may overrule this value. If the speed criterion could not be evaluated, $I_c(m)$ is assigned a value -1. The ARMA model criterion is only applied if the

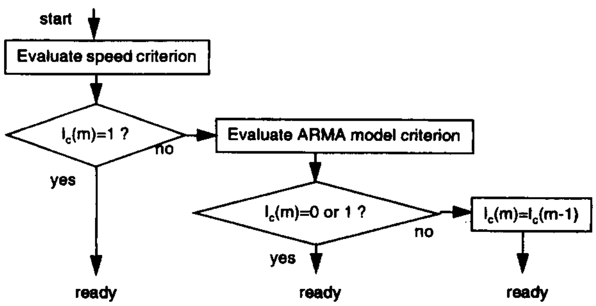

Fig. 3. Implementation of the congestion detection model.

speed criterion has returned a value -1 or 0. The ARMA model criterion is evaluated successfully if the traffic volume was large enough (see above) and if the parameters a_i, $i=1,\ldots,n_a$, b_j, $j=1,\ldots,n_b$ and Δ_1 have been estimated successfully. If the ARMA model criterion subsequently assigns $I_c(m)=1$, the value produced by the speed criterion is overruled, and a value 1 is returned.

If both criteria $I_c(m)=-1$, it is overruled by the value of the previous minute.

2.4.3. The model for estimating travel times

The free-flow travel time is section specific. A speed limit of 120 km h^{-1} is effective in the pilot area (see the section below). The free-flow travel time was based on a speed of 110 km h^{-1} and the section length. Further, the parameter Δ_2 was chosen per section equal to the smallest integer number larger than T_f. The outflow used in (10) was averaged over $M_2=15$ min.

The correction function is assumed to ensure that $N(t)$ does not drift away from zero during free-flowing traffic. For the incorporation of the drift correction, it has to be ensured that (11) is not disrupted by congested traffic. During congested traffic, the ratio between inflowing and outflowing traffic may be subject to large variations. Therefore, the last 10 min before suspected congestion and the 15 min afterwards were excluded from the set of free-flowing minutes $F(m)$. The number M_3 of minutes in $F(m)$ was chosen equal to 30.

A number of preliminary experiments was carried out to assess the suitability of the correction function using data from the GERDIEN pilot area, in which all entry and exit points of a section are equipped with loop detectors. It turned out that deviations in the correction function just before congestion can result in unrealistic estimates for the number of vehicles on the section during the congested period. If for instance $C(t)=1.05$, the measured outflowing traffic is 5% more than the measured inflowing traffic. The use of this correction factor during congestion resulted in negative estimates for the number of excess vehicles during obviously congested conditions. A correction function always equal to one was found to perform well as an alternative. It is expected that in conditions in which data are missing due to detector failures, the correction function may be more suitable. However, this was not investigated during the study described here.

3. Evaluation data for the model for estimating travel times

The model for estimating travel times was applied to a part of the pilot area in the GERDIEN project in The Netherlands. The application area was a part of the A12 motorway, comprising about 16 km in length in both directions. Paired inductive loops are present on all on-ramps, off-ramps and at a number of points on every lane of the main carriageway. Fig. 4 gives an overview of names, measurement points, sections and distance markers (on- and off-ramps are not shown).

Two sets of data recordings were obtained from the pilot area, which are described in the following two sections.

3.1. The tuning data

The first set of data was collected by "Rijkswaterstaat" during 41 days in late 1993 and made available for use in the GERDIEN project. The data comprised averages and standard deviations of speed, traffic volume, traffic composition and occupancy for every minute and every lane, 24 h

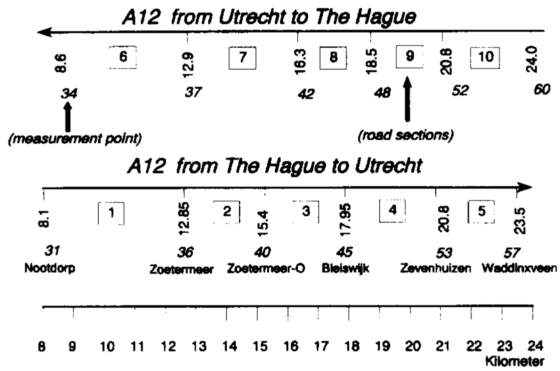

Fig. 4. Measurement points, sections and distance markers.

day^{-1}. The first set of data was supplemented by congestion reports of the Support Division of the Dutch National Police Force (KLPD). These reports contain congestion observations with an indication of time of day, location, length, development and cause. The reports are based on subjective observation. They were considered suitable for preliminary testing of the model as a general indication of congestion. The model was tested using data from 12 combinations of a day and a section. This data set contained combinations with recurrent and non-recurrent congestion as well as free-flowing traffic at high traffic volumes.

3.2. The evaluation data

The second set of data recordings was collected to enable a detailed evaluation of the models. The data set was collected for 21, 22, 25, 26 and 27 April, 1994. These days were all working days. The data set was again supplemented by congestion reports of KLPD and similar reports of the Dutch Motorist Organization ANWB.

Field observations (Van Arem et al., 1994a) were conducted during the morning and evening peak travel times and speeds during congested conditions. The field study was conducted from six observation cars, each with a driver and an observer. The cars drove in 10-min intervals on those sections of the area which are particularly prone to congestion. The drivers were instructed to drive at 120 km h^{-1} (which is the speed limit in the pilot area) as much as possible. To the drivers, congestion was defined as the situation in which drivers were forced to drive at less than 100 km h^{-1} and they could observe the total traffic flow slowing down. All drivers registered their speed every minute during congestion. When they encountered congestion, the distance markers at the road side and their speed were registered.

For the evaluation of the congestion detection model, a 'reference' congestion indicator was constructed defining for which days, sections and time instants congestion actually did take place. The drivers' observations of the congestion were used as a basis, using only observations below 80 km h^{-1}. This indicator was supplemented by the speed criterion (1) of the model. Finally, the congestion reports by KLPD and ANWB were used to fill in a small number of gaps in the reference congestion.

A correction was applied for differences in the orientation points for the field observations and the location of the loop detectors for the evaluation of the travel time estimates. The correction modified the free-flow travel time of the model in such a way that it fitted the observation during free-flowing traffic. The evaluation was conducted for a total of ten combinations of a day and a section, which had ample traffic and which were distributed along both directions of the motorway stretch. All congestion encountered in the pilot was recurrent.

4. Evaluation of the model for estimating travel times

4.1. Results from the tuning phase

The first results of the congestion detection model revealed that all congestion reported by KLPD was detected within several minutes. However, the number of alarms in which no congestion was reported was too high, especially at night. A more detailed analysis revealed that these probable 'false alarms' had two main causes.

Firstly, false alarms can be caused by slowly moving vehicles (less than the threshold set at 70 km h^{-1}). Single, slowly moving vehicles were observed at night in particular. This cannot be said to be a congested situation. As such a situation is indeed unusual, the threshold was not changed.

Secondly, a false alarm may be caused by the estimation of the ARMA model parameters. A good model estimation during free-flowing traffic is possible as long as the inflow resembles the outflow process with a time delay of several minutes. If this resemblance is not present, a sensible ARMA model estimate is not possible. The resulting estimate generally results in a false alarm. It appeared that this phenomenon is likely to occur at low traffic volumes in particular when small deviations are relatively large. Other false alarms were produced in rare situations in which the least square estimates were suddenly in a completely different range than the 'usual' estimates produced during free-flowing traffic, e.g. becoming negative instead of positive. There was no information on any special conditions at the time instants of these estimates. It is suspected that these false alarms may have been caused by coincidental disappearance of the resemblance in the input and output process.

The travel time estimate is based on the number of vehicles too many and the outflow from a section. A clear difference was observed between cases with recurrent congestion and accidents. In the case of recurrent congestion, the volume of the outflowing traffic remains high. As a consequence a relatively low time delay is estimated. In the case of accidents, the volume of outflowing traffic decreases drastically, resulting in a very high travel time estimate, as one would expect. The numerical validity of this estimate could not be verified. In cases where the number of excess vehicles on the section is negative, the estimated travel time is evidently not reliable.

A more extensive report of the results from the tuning phase can be found in the project report (Van Arem et al., 1994b).

4.2. Detailed evaluation of the travel time estimates

The ten combinations of a day and a section were subjected to a detailed evaluation. The reference congestion was compared to the congestion trigger $I_c(m)$ from the model, as shown in Fig. 5. The figure depicts both the reference congestion and the detection signal. A vertical bar indicates congestion or a detection of it.

In the case of Fig. 5, the first detection took place 11 min after the first minute in which congestion was reported. At night, the congestion trigger C was generated for a short period of time. This was probably caused by one or several

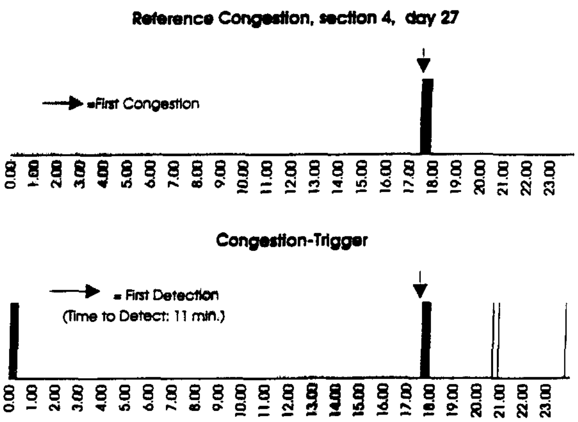

Fig. 5. Congestion detection Bleiswijk to Zevenhuizen, April 27, 1994.

slowly moving vehicles. The time-to-detect (TTD) was computed for all selected combinations by manually selecting periods of reference congestion and the corresponding period of detection and then computing the difference between the corresponding first time instants. Very often, congestion was detected within 1 min. In these cases both the first reference congestion and first detection were generated by the fact that the speed was below the 70 km h^{-1} threshold. In other situations, not reported here, a TTD up to 31 min was observed. In addition to the comparison of the congestion trigger and field observations, a number of descriptive indicators can be calculated according to the definition in DRIVE-II Task Force Automatic Incident Detection (1994):

- False alarm frequency (FAF), calculated as the number of minutes with a congestion detection and no reference congestion divided by the total number of minutes in a day (1440).
- False alarm rate (FAR), the number of minutes with a congestion detection and no reference congestion divided by the number of minutes with a congestion detection.
- Detection rate (DR), computed as the number of minutes with both a congestion detection and reference congestion divided by the number of minutes with reference congestion.

Table 1
Descriptive measures for congestion detection

Section, day	# of actual congestion	# of alarms	False alarm frequency (FAF)	False alarm rate (FAR)	Detection rate (DR)
1, 26/4	14	10	0.003	0.50	0.36
3, 25/4	97	148	0.040	0.39	0.94
4, 26/4	122	83	0.017	0.30	0.48
4, 27/4	27	36	0.015	0.61	0.52
5, 22/4	20	37	0.017	0.68	0.60
6, 21/4	87	90	0.007	0.11	0.92
6, 22/4	1	11	0.008	1.00	0.00
7, 22/4	5	1	0.001	1.00	0.00
10, 25/4	130	142	0.008	0.08	1.00
10, 27/4	171	170	0.000	0.00	0.99
Total	674	728	0.012	0.23	0.83

Table 1 contains these measures for the ten evaluated combinations of a section and a day.

The table shows that as the number of measured and reported congestion increases, the detection rate and the false alarm frequency improve. For the combinations with only a small number of observations and/or measured congestion the results are somewhat less good. A special case appears to be Section 4 on Day 26/4. There are 122 min of congestion, yet the false alarm rate is quite high and the detection rate quite low. A closer look at the data set for this case revealed that the congestion occurred during 2 h of the evening peak time. During this period, average speeds only dropped below the threshold now and then, while recovering at the entry and exit points between that time. Drivers reported congestion on the section between the entry and exit points, but they did not experience significant delays. This congestion was detected by the model, although not as a continuous period and with some delays. After modification of the free-flow travel time as described in Section 3.3, travel time estimates using (8) were compared to the field observations. The travel time of a vehicle entering a section was predicted using data of the last 30 min. Two examples are given in Figs. 6 and 7. Travel time estimates are represented by a solid line and the field observations by small circles.

Fig. 6 shows that the estimates correspond well with the measured travel times. Congestion occurred between 7.30 am and 9.00 am resulting in a higher estimated travel time. In Fig. 7 the match is less spectacular, but still fairly good.

As descriptive indicator of the goodness-of-fit

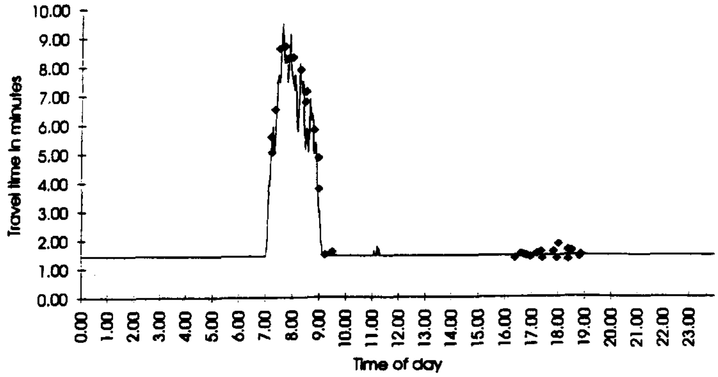

Fig. 6. Travel times Moordrecht to Zevenhuizen, April 25, 1994.

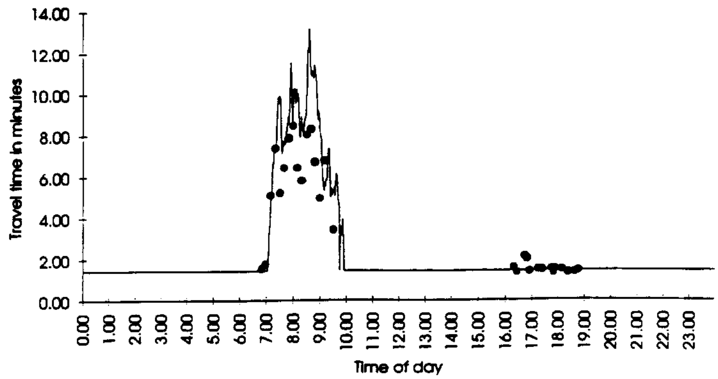

Fig. 7. Travel times Moordrecht to Zevenhuizen, April 27, 1994.

the root mean square error proportion (RMSEP) was calculated, defined as (Van Arem, 1994b):

$$\mathrm{RMSEP} = \frac{\sqrt{n \sum_{i=1}^{n} (x_i^{\text{predicted}} - x_i^{\text{observed}})^2}}{\sum_{i=1}^{n} x_i^{\text{observed}}} \quad (15)$$

A low RMSEP means that the calculated travel times match the measured travel times from the field observation well. Table 2 presents the resulting RMSEP values for all the selected combinations. The RMSEP is given for situations in which there actually was congestion (reference congestion) and for free-flowing conditions. A – indicates that there were no travel time observations.

In order to put the results into context, it should be noted that the RMSEP values for free-flowing conditions (in general below 0.1) can be regarded as a lower bound to the possible forecasting accuracy. That is, such values apply to drivers that had received the same instructions, had similar cars but had encountered slightly different conditions.

The results for April 22, Section 5 were the worst. The discrepancies were caused by both false alarms and undetected congestion. During congestion, the RMSEP remains below 0.25. An exception is Section 10 on April 27 (Fig. 7), which already shows considerable variations in the observed travel times.

Finally, an important observation is that the travel time estimates appear quite insensitive to false alarms. The RMSEP increases but remains at an acceptable level (0.40 for Fig. 7). This effect can be explained by the fact that the travel time estimate is based indirectly on the congestion detection model, viz. via the number of excess vehicles and the outflowing traffic. In

Table 2
RMSEP for all ten selected combinations

Section	Day in April 1994	RMSEP free flow	RMSEP congestion	Times congestion
1	26	0.08	0.13	14
3	25	0.09	0.25	97
4	26	0.07	0.28	122
4	27	0.14	0.25	27
5	22	0.25	0.28	20
6	21	0.08	0.24	87
6	22	0.07	–	1
7	22	0.08	–	5
10	25	0.09	0.16	130
10	27	0.14	0.40	171

cases of a false alarm, the estimates for the latter two quantities are usually acceptable. It has been noted in Table 1 that the congestion detection results for Section 4 on Day 26/4 were somewhat less good. From Table 2 it appears that the RMSEP value for this combination is very acceptable. This emphasizes the insensitivity of the model for estimating travel times to congestion detection errors. An extensive report of the evaluations is included in Van Arem et al. (1994b).

5. Conclusion

This article has described the travel time estimation model employed in the GERDIEN project and has given detailed evaluation results using field observations.

The primary evaluation quantity was the travel time over a section. For cases with free-flowing traffic or recurrent congestion the travel time estimates were found to fit the observations well. For non-recurrent congestion, the travel time estimates behaved in a reasonable way, but there were no observations to verify the order of magnitude.

The secondary evaluation quantity was the detection of congestion. In a number of cases, the congestion detection rate was significantly lower than 1, and the false alarm rate was significantly different from zero. However, as an input to the travel time estimation model, the congestion detection model performs satisfactorily. The travel time estimation model appeared to be reasonably insensitive to false alarms triggered by the congestion detection model.

The results are also relevant to forecasting models, which estimate future travel times on section or parts of the motorway network. The RMSEP values obtained in free-flowing conditions (up to 0.1) and during recurrent congestion (around 0.25, up to 0.40) give an indication as to what could be possible (and perhaps impossible) in travel time estimates using longer prediction horizons or being applied during non-recurrent congestion.

Acknowledgements

The present article reflects work conducted as part of the DRIVE-II project GERDIEN (V2044). The work was conducted by the TNO Institute of Infrastructure, Transport and Regional Development and the Transport Research Centre (AVV) of "Rijkswaterstaat". The data recordings and field observations used in the evaluation reported in this document were made available by AVV. Congestion reports were provided by the Support Division of the Dutch National Police Force (KLPD) and the Dutch Motorist Organization ANWB. Coding of the software modules was done by A.J. van der Meer of the TNO Institute of Applied Physics. Data processing and statistical analyses were done by T.J. Leering of "FOCUS Onderzoek & Advies".

References

Ben-Akiva, M., E. Kroes, J. De Ruiter, S. Smulders, J. Whittaker and K. Lindveld, 1992, *Real-time prediction of traffic congestion*, in: *Proc. World Conference on Transportation Research*, 1992, Lyon, France.

Blonk, J.C., J. Giezen, L. Van Den Berg and R. Ziegler, 1994, *Development of a General European Road Data Information Exchange Network (GERDIEN), An architectural development approach*, in: *Proc. First World Congress on ATT & IVHS*, 1994, Paris (Artec House, London).

Box, G.E.P. and G.M. Jenkins, 1970, *Time Series Analysis – Forecasting and Control* (Holden-Day, San Francisco).

Chatfield, C., 1989, *The Analysis of Time Series – An Introduction*, 4th Edn. (Chapman and Hall, London).

Cremer, M. and H. Schütt, 1990, *A comprehensive concept for simultaneous state observation, parameter estimation and incident detection*, in: M. Koshi, ed., *12th International Symposium on Transportation and Traffic Theory* (Elsevier Science Publishers).

Dailey, D.J., 1993, Travel time estimation using cross-correlation techniques, *Transportation Research B*, 27(2).

DRIVE-II Task Force Automatic Incident Detection, 1994, *Draft guidelines*, May.

Stephanedes, Y.J. and A.P. Chassiakos, 1993, Application of filtering techniques for incident detection, *Journal of Transportation Engineering*, 119(1), 13–26.

Van Arem B., 1994, *Short-term forecasting in inter-urban areas in the DRIVE-II programme*, *Report 1994-02* (INRO Centre for Infrastructure, Transport and Regional Development, TNO Institute for Policy Studies, Delft, The Netherlands).

Van Arem, B., M.J.M. Van Der Vlist, J.C. Blonk and L. Van Den Berg, 1993, Demonstration of a General European Road Data Information Exchange Network (GERDIEN), in: D.H.M. Reekie, ed., *Proc. Vehicle Navigation and Information Systems VNIS'93*, October 1993, Ottawa (IEEE Conference Services, Piscataway, NJ) 163–168.

Van Arem, B., A.R. Van Dijk and E.B.T. Rooijmans, 1994a, *Veldwaarnemingen binnen GERDIEN, Report INRO-VVG 1994-10* (INRO Centre for Infrastructure, Transport and Regional Development, TNO Institute for Policy Studies, Delft, The Netherlands) (in Dutch).

Van Arem, B., M.J.M. Van Der Vlist, J.C.C. De Ruiter, M. Muste and S.A. Smulders, 1994b, *Detailed evaluation of the models for current capacity estimation and travel time and congestion monitoring, GERDIEN Deliverable 16, Technical Annex 1* (INRO Centre for Infrastructure, Transport and Regional Development, TNO Institute for Policy Studies, Delft, The Netherlands).

Biographies: Bart VAN AREM obtained his M.Sc. (1986) and Ph.D. (1990) degrees in Applied Mathematics at the University of Twente, The Netherlands. After having worked at the Dutch Ministry for Transport, Public Works and Water Management, he joined the TNO Netherlands Organization for Applied Scientific Research in 1991. His main subject is dynamic traffic management. He was "traffic monitoring" team leader in the GERDIEN project in the DRIVE-II programme of the Commission of the European Communities. He was "short term traffic forecasting" coordinator in the evaluation task force of the DRIVE-II programme.

Martie J.M. VAN DER VLIST obtained his M.Sc. (1975) degree in Applied Mathematics at the Technical University of Delft, The Netherlands. He first worked at the information centre of the Dutch Tax Administration. During the period between 1979 and 1987, he worked as a researcher at the NEA Transport Research and Education Institute in the field of information systems and transportation models. Since 1987, he has been with the TNO Organization. As senior project manager he has participated in many projects in the fields of individual, public and freight transport. His special interests are projects that focus on intelligent highway systems and dynamic traffic management.

M. (Rik) MUSTE obtained his Ing. (1971) in electrical engineering at the Hogere Technische School of Amsterdam, the Netherlands. Since 1975 he is affiliated with the Transport Research Centre of "Rijkswaterstaat" of the Dutch Ministry of Transport, Public Works and Water Management. His field of interest is dynamic traffic management.

Stef A. SMULDERS obtained his M.Sc. (1984) and Ph.D. (1989) degree in Applied Mathematics at the University of Twente, The Netherlands. The subject of his thesis was control of freeway traffic flow and was carried out at the Centre for Mathematics and Computer Science in Amsterdam, The Netherlands. He has been affiliated with the Transport Research Centre of "Rijkswaterstaat" of the Dutch Ministry of Transport, Public Works and Water Management since 1989. His main fields of interest are traffic theory and dynamic traffic management. He was the representative of the Dutch Ministry of Transport in the DRIVE-II project DYNA, concerned with the real-time prediction of traffic flows on motorway networks. He is also involved in experiments with speed control and ramp metering on motorways and in studies into the effectiveness of advanced vehicle control systems.