Online Travel Time Estimation in Urban Areas Using the Occupancy of Long Loop Detectors

W. K. Mak, F. Viti, S. P. Hoogendoorn, A. Hegyi

Delft University of Technology, Transportation and Planning, P.O.Box 5048, 2600 GA Delft, The Netherlands, Tel. +31 15 2786909, email:{w.k.mak, f.viti, s.p.hoogendoorn, a.hegyi}@tudelft.nl

Abstract: Roads in the Netherlands are often heavily congested. Real-time travel time information can be a valuable instrument to reduce the impact of increasing traffic demand on travel time with advantages for traffic participants as well as for the traffic network managers.

For urban roads travel time estimation is a more complex problem than for freeways. In order to provide online, real-time and accurate travel times, in this paper a new estimation approach is developed based on measurements supplied by the (vehicle-dependent) traffic controllers. These measurements include the occupancy from the corresponding long loop, the flow from the short loop, and the green percentages. Based on these measurements the location and length of the queues is estimated.

Moreover, the approach transforms the data in a way that the underlying reason(s) of congestion becomes apparent (in terms of bottleneck location), which can be very useful for, e.g., advanced traffic information systems or decision support systems. Compared with the approach based on license plate recognition cameras, which is used in the Netherlands in many locations, this approach is expected to be more accurate in heavily congested situations, since for license plate recognition vehicles have to pass the end of the route before the traveled time can be measured.

The approach is tested with real traffic data for an urban route in the Netherlands. The test case presented in the paper shows that the use of the additional data from traffic controllers (occupancy, green percentage, waiting time of first vehicle in queue) is beneficial.

Keywords: Urban travel time estimation, long detector loops, travel information, traffic signal data.

1 INTRODUCTION

1.1 Traffic in the Netherlands

In the Netherlands, during the last 10 years, oversaturation of networks and traffic congestion have been accepted as the result from the persistent imbalance between traffic demand and traffic infrastructure supply. Advanced Traffic Information Systems (ATIS) provide opportunities to use the available network supply more efficiently (so-called demand-supply synchronization). Furthermore, providing travellers with reliable traffic information will reduce uncertainty and allow for better trip planning.

ATIS requires a data collection and state estimation system that provides reliable and accurate estimates of the prevailing traffic conditions on the freeway, rural, and urban network.

The state-of-practice of travel time estimation and prediction, and the distribution of this information on freeways in the Netherlands can be qualified high in comparison with urban roads, since most of the freeways are equipped with double loop detectors on every 500m. Van Lint showed [1] how these point measurements of traffic volumes and speeds can be used for travel time estimation and prediction on freeways.

However, these kinds of detectors are present on only a small part of the provincial and urban roads.

Estimation of travel times in urban areas is more complicated due to a variety of factors. An urban network is more complex than a freeway network, due to roads of different categories, more intersections with different control systems, and more route alternatives. Since delays at the intersections are a substantial part of the total travel time, the traffic dynamics has a stronger stochastic component in urban networks [2]. A detailed double loop detection system on every 500m for all urban roads in the network would be rather expensive, and would still ignore the waiting times at the intersections.

In this paper, a new and cost-efficient method is introduced for online travel time estimation. In comparison with other approaches, this method makes not only use of traffic volumes as online information from traffic controllers, but also of other data, from which the occupancy of long loop detectors is the most important. The model is developed and tested with use of real traffic data.

1.2 Travel time estimation in urban networks

In the past several approaches have been proposed for estimating travel times in urban areas. The approaches can be

distinguished according to the type of data used and the model type. The thesis of Liu contains an overview of most of

- Data-driven (black box) approaches,
- Model-based approaches,
- Hybrid approaches.

Data-driven approaches are not based on traffic theory or a traffic model. They either use a kind of measurement that directly relates to travel times, or estimate the relation between travel time and other measured variables, such as volumes, green times, etc, that indirectly relate to travel time. Examples of the direct approach are the use of data from mobile phones, use of GPS data from taxis and the use of License Plate Recognition cameras. An example of an indirect approach is the approach of Robinson[8] who estimated travel time in London using data from data short loop detectors using the k-NN (K-Nearest Neighbour) method. This method needs training data, a database consisting of measured travel times and loop detector data.

Model-based approaches make use of knowledge of traffic flow operations, formalized in either a static model or a dynamic model, which acts as base for estimation of travel times (for example, a standard BPR function, which is already included in the static model itself) [4]. The advantage of this approach is that travel time measured on a link can be added, which makes improvements by data-driven techniques possible.

In a hybrid approach, both approaches are combined.

While the data driven approaches may lead to accurate results but at a high cost (due to the required number of detectors), the model-based approaches are low(er)-cost, but less accurate methods. In this paper we aim at improving the accuracy of travel time estimation, using the available -- but until now unused – measurements from the traffic controllers.

In the Netherlands, typically loop detectors of traffic controllers provide information only on volumes of traffic over fixed time periods. The disadvantage of using volumes only as online data input is that no distinction can be made according to the cause of the changes in the traffic flows. Such changes may be a result of:

- varying traffic demands, or other stochastic disturbances,
- blocking back,
- saturation of an upstream intersection.

We argue that by using more data from traffic controllers than traffic volumes alone, a hybrid model could be more accurate and cost efficient. Liu [2] argues that this information is not available, but in practice, the additional costs of providing this data for online purposes are not very large.

A simple online approach of using volume data is, for instance, based on the cumulative count approach of Daganzo

these approaches [3]. In general, three types of approaches exist:

[6]. From the moment that a traffic controller measures a start of oversaturation, the change in queue length is estimated by the difference between entering and leaving vehicles of that link. This approach has some disadvantages and practical problems, as pointed out also by Liu [3], which may lead to inaccurate estimates:

- For one or more unmeasured intersections between two controlled intersections an estimation must replace measured data, which may lead to inaccuracies;
- Counting errors for the measured volumes tend to accumulate over time; if lanes are used for different directions, directions percentages have to be estimated and can change from reality;
- The method is vulnerable in situations where not all controllers provide information in time in an online situation (due to, e.g., communication errors), since data from 2 controllers for the algorithm are needed.

2 URBAN TRAVEL TIME ESTIMATION APPROACH

This section outlines a new hybrid approach to estimate travel times on urban networks using data from controlled intersections. The approach can be seen as an improvement of the so called 'Sandglass traffic model', using the image of sand through an hourglass for an oversaturated link [7]. The basic idea is to classify the traffic situation around a signalized intersection for each arm as either under- or oversaturated, and to use a simple calculation method for each class.

The approach first estimates the queue lengths at each intersection on the considered route, and next determines the travel time on each link of the route.

2.1 Queue length estimation

For the queue length estimation let's consider intersection i on the given route and refer to the incoming link to this intersection also by index i, and denote its length by l_i. In this approach it is assumed that link i is divided into two parts, the upstream part with length l_i^u and the downstream end with length l_i^d, with $l_i = l_i^u + l_i^d$.

In the Netherlands the typical loop configuration is to have a short loop at the stop line of the intersection and a long loop (with length of around 20m) upstream of it. The basic idea of the queue length estimation is that based on the occupancy measurement from the long loop and the flow measurement of the short loop of the intersections, the length of the queues on the upstream and downstream part of the link are estimated separately.

The estimation of the queue length on the downstream part of the link is based on the occupancy of the long loop. The

occupancy is the percentage of time that at least a part of a vehicle is above the detector. The long loop detector (20m) will be occupied for 100% if the distance between any two vehicles is less than 20m. So, if all vehicles have the same driving and car-following behaviour, oversaturation will theoretically lead to 100% occupancy. In case of under-saturation, after the green time, there will be gaps greater than 20m. So, under-saturation will lead to occupancy lower then 100%. However, in practice, even in the case of oversaturation, 100% occupancy during a time period is not always reached, since a truck or a slower vehicle may have a larger headway within a queue. In case of oversaturation we assume that there is a queue present at the downstream end of the link. The estimated queue length $l_i^{\mathrm{d,queue}}$ is given by

$$l_i^{\mathrm{d,queue}} = \beta(p_1, q_1, p_{\mathrm{occ}})l_i^{\mathrm{d}}, \qquad (1)$$

which is a scaling of the length of the length l_i^{d} by a factor given by the cumulative beta distribution, and p_{occ} is the measured occupancy. An example of the beta distribution is given in Figure 1. The tuning parameters p_1 and q_1 determine the shape and "smoothness" of the queue length as a function of the occupancy.

The next step is to estimate the queue length $l_{i+1}^{\mathrm{u,queue}}$ at the upstream part of the link leaving intersection i (which is therefore the upstream part of the incoming link of intersection $i+1$).

The reasoning is based on the observation that there may be basically two reasons for a high occupancy at intersection i:

- A lack of capacity at the intersection;

- A queue downstream of the intersection (blocking back).

If a high occupancy is measured (by the long loop detectors) and there is no blocking back, traffic volume (V) on a lane will be given by

$$V = p_g c, \qquad (2)$$

where c is the lane capacity, and p_g the green percentage within a cycle. However, in case of blocking back, V will be lower. Now, let us define the capacity factor p_{cp} given by

$$p_{cp} = \min\left(1; \frac{V}{p_g \cdot c}\right), \qquad (3)$$

where p_{cp} expresses undersaturation. Essential in (3) is the use of the measured value of the percentage of green time p_g and not an estimation of it, since the comparison between V and p_g must be made as accurate as possible.

With a similar reasoning a queue is assumed to be present at the upstream part of link $i+1$ (which is directly downstream of intersection i) only if (1) the measured occupancy is high and the queue discharge rate is lower than the capacity.
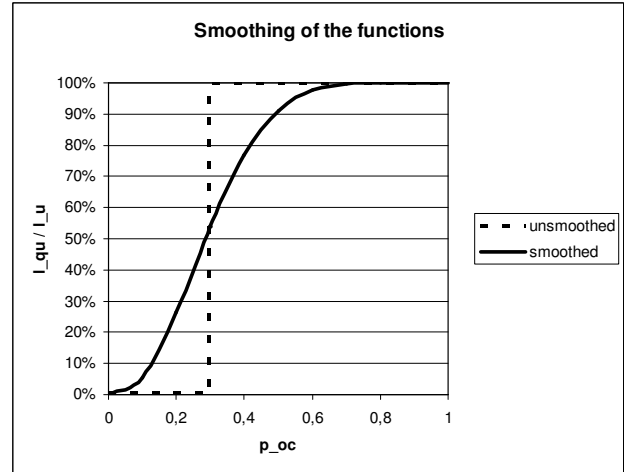

Figure 1 An example of the beta function.

The estimated queue length $l_{i+1}^{\mathrm{u,queue}}$ is expressed by

$$l_{i+1}^{\mathrm{u,queue}} = \beta(p_1, q_1, p_{\mathrm{occ}})\beta(p_2, q_2, 1 - p_{\mathrm{cp}})l_{i+1}^{\mathrm{u}}, \qquad (4)$$

where the first factor is the same as in (1), the second factor expresses the "degree" of flow saturation, and p_2 and q_2 are tuning parameters.

It is worth mentioning that in case of a growing or decreasing queue, the downstream queue length $l_i^{\mathrm{d,queue}}$ might be overestimated, and the upstream queue length $l_{i+1}^{\mathrm{u,queue}}$ underestimated, since the queue length is estimated based on a point measurement. When a queue starts building up, e.g., from a congested freeway to an urban road, the queue downstream of the last controller (the upstream queue) on the route will be ignored till the start of the queue has reached the intersection. At the moment the queue is somewhere upstream of the last controller, it is estimated halfway between the last two controllers, which is an overestimation of the queue length, if the distance between the two controllers is significant.

The beta parameters are assumed to be network parameters, which are only dependent of the length and position of the loop detector. Although traffic behaviour and vehicle composition will possibly influence the parameters p_1 en q_1, we don't expect this be too significant. For practical implementation a network estimation of p en q would be interesting.

2.2 Travel time calculation

Travel time in urban networks can be subdivided into three parts:

- Free driving time (t_0)

- Waiting time for signalized intersection, except queuing (or uncontrolled intersection) (t_w)

- Time spent in waiting queue (t_q)

The free driving time on link i can be approximated by:

$$t_i^0 = (l_i - (l_i^{d,queue} + l_i^{u,queue}))/v_0 \qquad (5)$$

Where $l_i\,l$ is the length of link i and $l_i^{d,queue} + l_i^{u,queue}$ is the total length of the queue, and v_0 is the maximum allowed speed.

In case of uniformly arriving vehicles, a fixed-time controller, the mean waiting time (until green) is the product of the probability of waiting and the half of the red time:

$$t_w = 0.5 \cdot t_r \cdot (1 - p_g) \qquad (6)$$

Where t_r is the red time and p_g is the percentage of green time. In our model, we replace t_r by the waiting time of the first vehicle in the queue, t_w^1:

$$t_w = 0.5 \cdot t_w^1 \cdot (1 - p_g) \qquad (7)$$

t_w^1 is measured by the traffic controller and is, in cases of low traffic and vehicle responsive control, more accurate than using the red time, since in that case the assumption of an uniform arrival time during red is not valid: the signal remains red until a first vehicle arrives.

The time spent in the waiting queue t_q is determined by the length l_q of the waiting queue, the flow V on one lane, and the headway distance d of the vehicles in the queue, according to

$$t_q = \frac{l_q}{d \cdot V}. \qquad (8)$$

In (8) V is measured by the short loop detectors of the traffic controller.

3 CASE STUDY: DATA DESCRIPTION

In the remainder of the paper, we will illustrate the functioning of the proposed estimation procedure by applying it to a case study with data from the Dutch city Zoetermeer.

3.1 Case Zoetermeer

Data was collected for a period of 3 months with accurate travel times from a License Plate Recognition (LPR) camera system and online data from 4 traffic controllers on a route in the city of Zoetermeer. A schematic representation of the considered route is given in Figure 2. The thick line indicates the measured route of the LPR system. The data is aggregated over 5 minute intervals, and contain high peaks in travel times. With use of these data, testing and calibration of the algorithm was in sight.
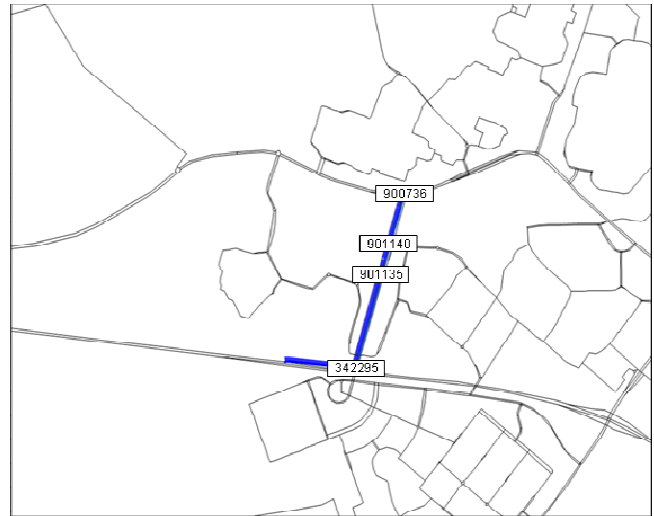

Figure 2 The investigated route in the city of Zoetermeer (the Netherlands)

The data is collected from September to December 2005. From these data a selection is made with days which consist of a travel time greater than 15 minutes: 7 days are selected.

The following data are collected by / from the traffic controllers:

- Traffic volume;

- Occupancy long detector loops;

- Occupancy short detector loops;

- Mean green time;

- Number of start green;

- Mean waiting time of first vehicle in waiting queue.

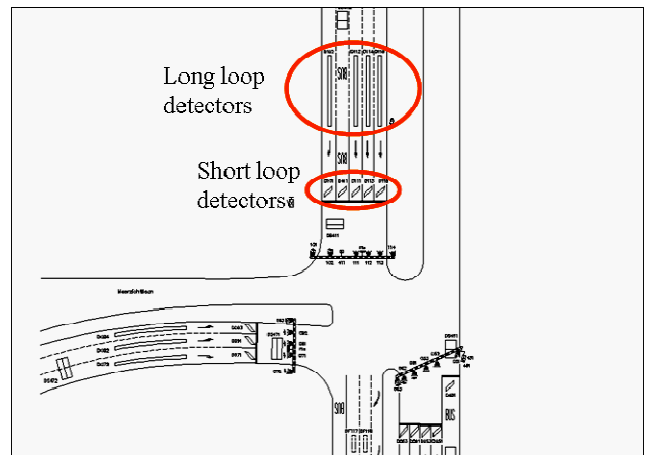Figure 3 shows an example of an intersection with position of long and short loop detectors.

Figure 3 Overview of typical loop detector infrastructure in the Netherlands

In the Netherlands, most detector infrastructure consists of a short detector for each signal group just before the stop line and a long loop detector (20m) at 40 m before the stop line, apart from bus, pedestrian and cycle detectors. These long loop detectors are used in order to optimize the end of green time.

3.2 Collected travel time data

From the LPR system, the following data are defined and collected:

- Departure travel time (DTT) = mean departure time of set of vehicles leaving from the first LPR camera within a certain time period

- Arrival travel time (ATT) = mean travel time of set of vehicles arriving at the second LPR camera within a certain time period

3.3 Level of aggregation

Travel times of shorter time periods consist of much heterogeneous variation (start and end of green times of controllers). Practical consequences are:

- For some periods no DTT or ATT exist: in the dataset these empty data are replaced by interpolation of previous and next data;

- The key data from the traffic controllers has some statistical variation, depending on the exact starting and ending times of a new phase cycle within the 5-minute periods.

- A shorter aggregation period would lead to more serious problems as mentioned above. A larger aggregation period will lead to an unnecessary loss of information.

3.4 Calculation of trajectories

In order to compare mean travel time, estimated by data from traffic controllers, and the measured travel times from the LPR, the DTT is chosen as travel time reference. The purpose is to estimate the DTT from data from traffic controllers. For practical reason, it is assumed that velocity is constant homogenous over the network within one period. In case of network travel times greater than the aggregation period, the part of the route travelled during the aggregation period is 5 minutes divided by the travel time of the aggregation time period. An example of this method is shown in TABLE 1.

TABLE 1 Example of estimating departure travel time from network travel time.

start time	end time	network travel time (s)	travel time on succeeding aggregation time periods (s)						propagation time (s)		estimated
			1	2	3	4	5	6	start time	end time	ddt (s)
7:50	7:55	538	300	300	94	0	0	0	694	955	824
7:55	8:00	812	300	300	300	55	0	0	955	1379	1167
8:00	8:05	1283	300	300	300	300	179	0	1379	1321	1350
8:05	8:10	808	300	300	300	300	121	0	1321	1272	1296
8:10	8:15	2154	300	300	300	300	72	0	1272	1037	1155
8:15	8:20	2347	300	300	300	137	0	0	1037	797	917
8:20	8:25	1396	300	300	197	0	0	0	797	598	697
8:25	8:30	824	300	298	0	0	0	0	598	584	591
8:30	8:35	468	300	284	0	0	0	0	584	739	662
8:35	8:40	791	300	300	139	0	0	0	739	642	690
8:40	8:45	720	300	300	42	0	0	0	642	463	552
8:45	8:50	682	300	163	0	0	0	0	463	290	376
8:50	8:55	290	290	0	0	0	0	0	290		

This method makes it possible to handle high values of network travel times in an easy, consistent and proper way. For example: if a driver starts at 7:50, your network travel time till 7:55 is 538 seconds, at 7:55 you have passed 300/538 = 56% of the route. The next period, the network travel time is 812 seconds, so you travel 300/812 = 37% of the route. The last 7.3% of the route takes 0.073*1283 = 94 seconds.

4 DATA ANALYSIS

In this section the values of the data are plotted as time series in order to know which approach can be suitable for estimation of travel time during congestion.

4.1 Travel times

In Figure 4 the departure travel time is shown in the morning peak for some selected days with congestion. The y-axis represents travel time in seconds, while the x-axis shows the time of the day (hours and minutes); the different lines are the different days. Shapes are different from day to day. It seems not easy to forecast travel times using historic data.
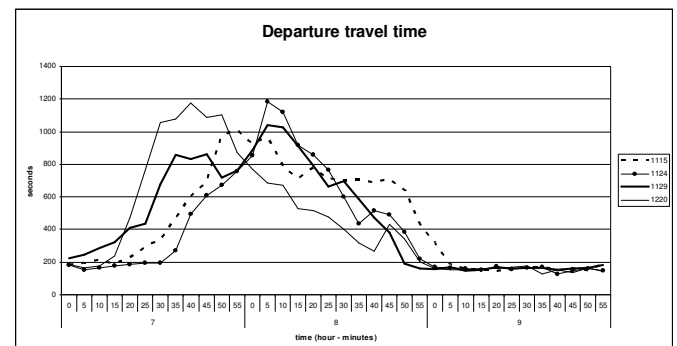

Figure 4 Departure travel time for selected days as function of aggregate time period

Figure 5 shows the arrival travel time on the y-axis, time on the x-axis for the same days. Peaks are reached later in time, which is logically correct by definition.
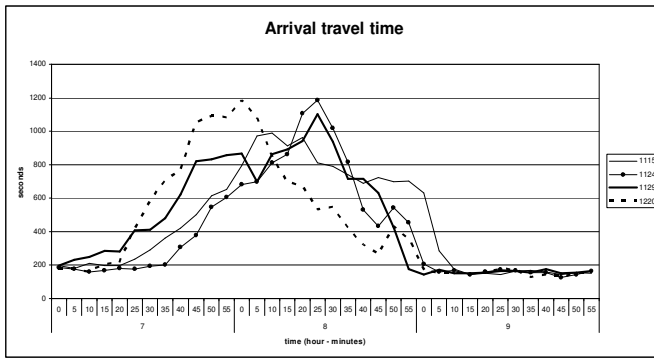
Figure 5 Arrival travel time for selected days as function of aggregate time period

4.2 Traffic volumes

Traffic volumes (mean value over traffic controllers and relevant lanes) are shown in Figure 6. During congestion, traffic volumes drop. By using only this information, it is not possible to estimate travel times, since the reason of a volume drop is not clear. However, in combination with occupancy this data can be very useful. The drop of volumes which explains downstream queuing is visible.
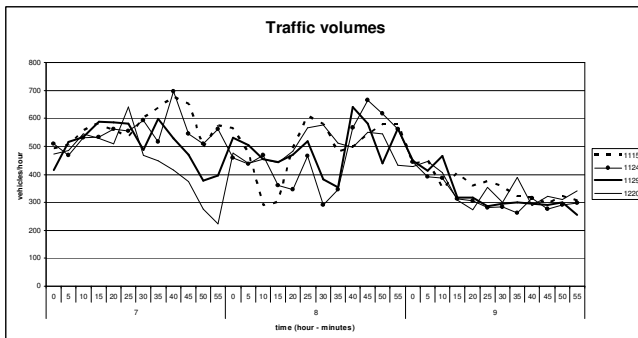

Figure 6 Mean traffic volumes for selected days as function of aggregate time period

4.3 Waiting time first vehicle in queue

Figure 7 shows on the y-axis the waiting time of the first vehicle in queue (mean value over controllers and relevant lanes) for the selected days. This data show an up going and down going line, showing the relationship between red time and traffic volume.
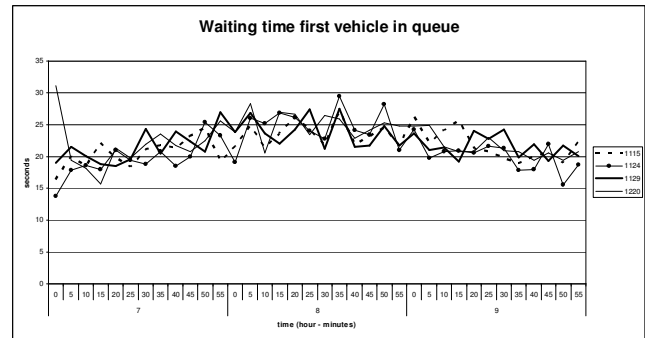

Figure 7 Waiting time first vehicle in queue for selected days as function of aggregate time period

4.4 Loop detectors

Figure 8 shows the occupancy of the short detectors (mean over controllers and relevant lanes). There is some correspondence between travel time and occupancy, but differences in occupancy are not very great.
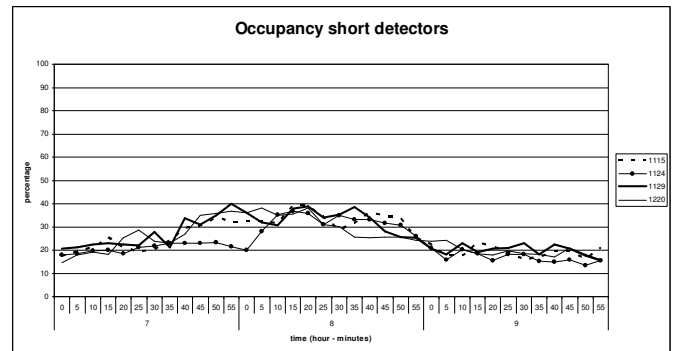

Figure 8 Occupancy short detectors for selected days as function of aggregate time period

Figure 9 shows the occupancy of the long detector, as mean over all controllers and lanes (figure 9a), and as a 1 day result (29th of November) separated by controller and relevant lane (figure 9b), where the thick lines represent the last, the thin lines the middle and the dotted lines the first controller. Queuing starts at the thick, the thin lines follows quick and the dotted comes later with two dips at 8:05 and 8:30. A high correspondence can be found with the travel time figures. Congestion starts at a saturation level of 70-80%.
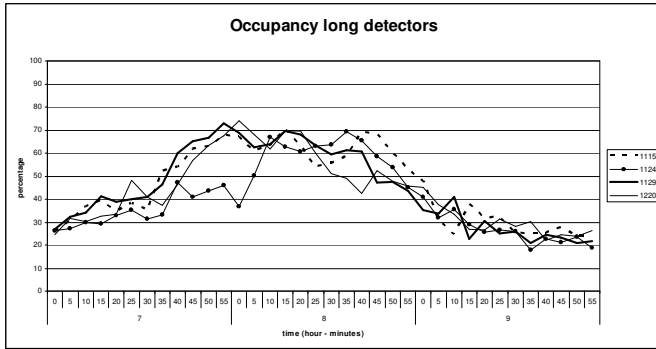
Figure 9a Occupancy long detectors for selected days as function of aggregate time period
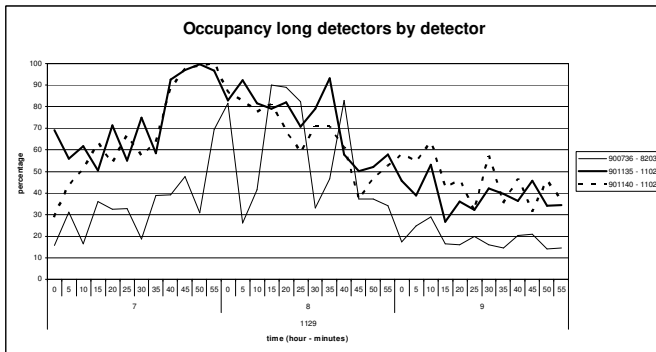

Figure 9b Occupancy long detectors on 1 day, for different positions on the route

5 RESULTS OF ESTIMATION

This section discusses the results of applying the estimation procedure presented in section 2 to the Zoetermeer case.

5.1 Data Correction

Controller 342295 provided counting data only (due to technical reasons). Therefore, it was decided to split the total route length of 2,249 m in the 6 parts upstream and downstream of the 3 remaining controllers. TABLE 2 depicts these choices.

TABLE 2 lengths of the upstream and downstream links of each controller

controller	upstream	downstream	total
900736	70	200	270
901140	239	100	339
901135	101	1,539	1,640

The longest section is the downstream section of controller 901135, about 70% of the whole route.

5.2 Parameter estimation

The model is estimated by means of successive manual optimization of the parameters, minimizing the RMSE (Root Mean Square Error). The estimated distance between two succeeding vehicles in a queue, $d = 6.5$ m. The estimated lane capacity = 1800 vehicle/hour, the estimated parameters of the cumulative beta distributions are: $p_1 = 12$, $q_1 = 4$, $p_2 = 1.5$, $q_2 = 5$. With this specification, 79% of the travel time variation can be explained (using a measure), where the ATT values as estimate of the DTT (LPR approach) come to 73%. The MARE is 18%. For measured travel times larger than 300 seconds, the MARE is 29%. Figure 10 shows the estimated travel times as dots against the line of measured values (10a) with a 25% lower and upper bound and the same data in a time series plot (10b). In situations of very high travel times, for some cases the fit is rather good, but in other cases, the measured travel time is underestimated by the model. This can be explained by the fact that the last controller did not provide good data. Therefore it takes some time before a queue is recognized (length of third downstream section = 1,539 m). During this time, travel time is underestimated. The result could be better if the last controller provided proper data.
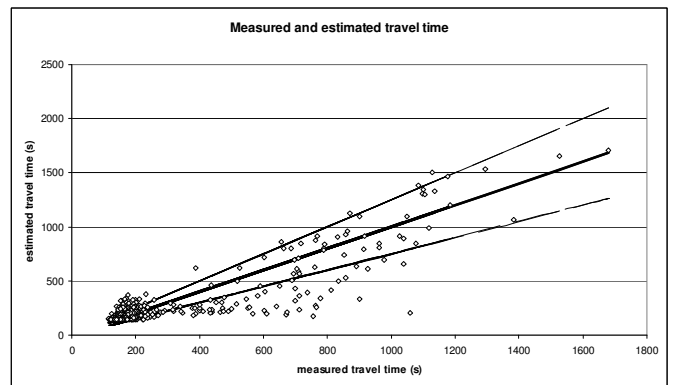

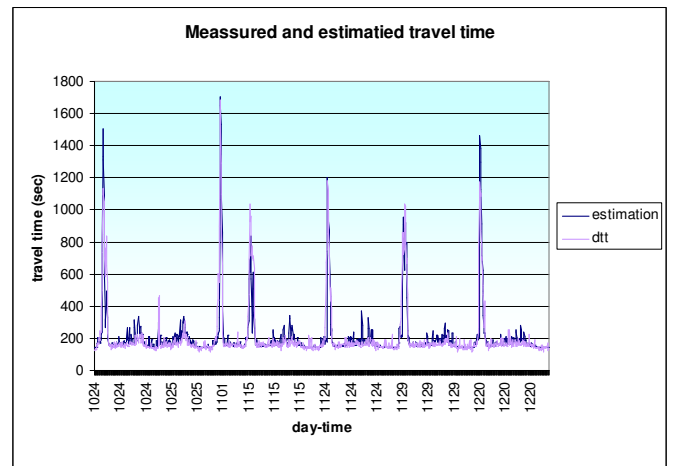**Figure 10a Results of travel time estimation**

Figure 10b Results of travel time estimation

6 CONCLUSIONS, FURTHER RESEARCH AND RECOMMENDATIONS

The use of on-line occupancy measurements from traffic controllers turns out to be a rather easy and valuable addition to online estimation of travel times. A rather simple model using this data turns out to have good results especially in case of serious congestion, compared with the travel time estimation from license plate recognition. The advantage of this approach is that it uses more recent information, compared to license plate recognition (which use arrival travel times), which is important in cases of serious congestion. In practice, for car drivers, robust information is more important than very accurate: it is more important to be sure about a route either it is congested or not, rather than to know if the congested travel time is 20 minutes or 25 minutes.

Furthermore, this approach shows the location of the congestion (starting and ending points in a network) and identifies the cause of congestion, which is very useful for dynamic traffic management. If the model will be used for online travel time estimation, another result can be a functional and systematic evaluation of the signalized intersections in the network: for every controller can be estimated the percent of time in which the controller is a bottle neck in the network and what the total lost of vehicle hours this means. Network owners can use this information for an efficient investment in traffic controllers and small infrastructure upgrades.

Further research needs to be done to test the model in a more extended network, in order to get insights into the sensitivity of the estimated parameters, with respect to changing traffic composition, starting and ending of incidents, routing, traffic volume, etc. The manual optimization can be replaced by an automatic method.

In the Netherlands, the typical detection infrastructure is the one presented in this paper. Most other countries use short detectors only. The data suggest that the occupancy of short detectors can also be used for this model, but that parameters will be different.

For ATIS it is recommended to collect more real-time data from traffic controllers (occupancy, green percentage, waiting time of first vehicle in queue) than just volumes.

ACKNOWLEDGEMENT

REFERENCES

[1] Lint, J.W.C. Van (2004), Reliable Travel Time Prediction for Freeways, PhD thesis, Trail research school, The Netherlands

[2] Viti F., (2006). The Dynamics and the Uncertainty of Delays at Signals. PhD Thesis, TRAIL Research School, Delft, The Netherlands.

[3] Liu, H. (2008), Travel Time Prediction for Urban Networks, PhD thesis, Trail research school, The Netherlands

[4] Bureau of Public Roads (1964), Traffic Assignment Manual, Prepared for the U.S. Department of Transport (DOT)

[5] Fellendorf M., Nokel K., Hanke N., (2000). Visum-Online Traffic Management for the EXPO 2000 based on a traffic model. Traffic Technology International, 2000.

[6] Daganzo, C.F. (1997), Fundamentals of Traffics and Traffic Operations, Elsevier Science Ltd, Oxford, UK

[7] Usami, T., K. Ikenoue & T. Miyashako (1986), Travel time prediction algorithm and signal operations at critical intersections for controlling travel time, in 'Second International Conference on Road Traffic Control', Institute of Electrical and Electronics Engineers.

[8] Robinson, S. & Polak, J.W. (2005), Modelling Urban link travel time with inductive loop detector data using the k-nn method, TRB Transportation Research Record No. 1935

[9] Mak, W.K., Reistijden schatten met Verkeersregelinstallaties (2007), Tech. Report, Transumo ATMO, Deliverable D5c-2, Vialis, In Dutch.