A SIMPLE DATA FUSION METHOD FOR INSTANTANEOUS TRAVEL TIME ESTIMATION

Michael Do, University of Twente, m.do@student.utwente.nl

Rattaphol Pueboobpaphan, University of Twente, r.pueboobpaphan@utwente.nl

Marc Miska, University of Tokyo, mmiska@iis.u-tokyo.ac.jp

Masao Kuwahara, University of Tokyo, kuwahara@iis.u-tokyo.ac.jp

Bart van Arem, Delft University of Technology, B.vanArem@tudelft.nl

ABSTRACT

Travel time is one of the most understandable parameters to describe traffic condition and an important input to many intelligent transportation systems applications. Direct measurement from Electronic Toll Collection (ETC) system is promising but the data arrives too late, only after the vehicles complete their trip. There are several existing models with varying degree of success to indirectly estimate travel time from loop detector. The performance of these models depends significantly on the variation of traffic condition. By closely looking at the time-series of the estimated travel time with the actual travel time, the error was found to follow a specific pattern for each traffic condition. The goal of this research is to develop a simple data fusion between loop detector data and ETC data to make more accurate estimation of instantaneous travel time on expressway corridor. With the error pattern for each traffic condition in mind, it is possible to develop a simple fusion method that can improve the accuracy of travel time estimate even under sparely distributed detectors.

Keywords: travel time estimation, instantaneous travel time, on-line estimation, data fusion, electronic toll collection, detector data

INTRODUCTION

With the emergence of Advanced Traveler Information Systems (ATIS), it is possible to provide various kinds of information to road users. Travel time is one of the most understandable measures for road users. By providing reliable travel time estimates it is possible to influence road users' route choice and travel behavior, hence improving the performance of traffic networks.

Because of the widespread deployment of loop detectors, most travel time estimation algorithms only have detector data as input. Although detectors continuously collect data, they do not provide an accurate image of the traffic conditions on the road. This is because detectors only collect data at point-locations and not over the entire road. Electronic Toll Collection (ETC) data on the other hand gives measured travel times over the entire road. But this data arrives too late. By the time travel time is measured, traffic conditions have most likely changed already.

There are several existing models to estimate travel time on-line based on detector data. Of interest in this paper are the two classes called "speed-based" and "flow-based" models (Turner et al, 1998; Li et al, 2006; Nam and Drew, 1996, 1998, 1999; Vanajakshi et al, 2009). The basic difference is on the usage of different types of data from detectors, i.e. the latter makes use of traffic count while the first makes use of point-speed. A number of studies have noted on the performance of the travel time estimates under different traffic conditions (Bovy and Thijs, 2000; Sun et al, 2008; Vanajakshi et al, 2009). In particular, different degrees of success are observed during transition flow and congestion conditions. Moreover, accuracy was found to degrade under larger detector spacing (Kothuri et al, 2006) and even more significant under congested condition and large spacing (Liu et al, 2006). For such situation, methods that rely on detector data alone might not give travel time estimates with sufficient accuracy.

A number of studies have explored the potential of using additional data sources, such as GPS-equipped probe vehicles and ETC, to improve travel time estimate from detector data (Chen and Chien, 2001; Nanthawichit et al, 2003; Chu et al, 2005; El Faouzi et al, 2009). These models are based on sophisticated methods such as Extended Kalman Filter and Dempster-Shafer theory but with varying degree of success. Among others, Li et al (2006) reported that speed-based models using only detector data tend to underestimate travel time with some lag. With a good knowledge of error patterns, it might be possible to use ETC data together with some simple rules to correct the travel time estimated from detector data

The goal of this research is to develop a simple data fusion between detector data and ETC data to make more accurate estimation of on-line travel time on expressway corridor with relatively long detector spacing. Unlike previous attempt of data fusion, this research does not use historical and statistical methods for data fusion. Rather some simple correction rules based on specific error patterns of the estimation model are considered.

The next section briefly reviews and describes some existing models that are considered in this paper. Next, study area and the available data are presented. This is followed by an analysis of the existing models under two detector spacing scenarios: (i) scenario with short spacing and (ii) scenario with relatively large spacing. Then the error from the existing models is analyzed and some specific error patterns are observed. Based on the error patterns, some simple correction rules using ETC data together with detector data are examined. Finally conclusions and further research recommendations are provided.

LITERATURE REVIEW

Two classes of travel time estimation models are considered in this paper, namely speed-based and flow-based models. They differ mainly based on the usage of different types of data from detector. In this paper we consider on-line estimation problem and therefore the instantaneous travel time estimation models are of main interest. However, for comparison purpose, an off-line travel time estimation model namely the time slice model is also discussed.

Speed-based model

Speed-based models are the most simple but widely adopted in practice (Turner et al, 1998). For this type of model, travel time for a segment is estimated based on speed data from upstream and/or downstream detectors of that segment. Mostly the speed within the segment is assumed to be constant and the vehicles change their speed abruptly when entering the next segment (Linveld et al, 2000). Some researchers also proposed a model that assumes vehicle speed changes linearly from upstream end to downstream end (Van Lint and Van der Zijpp, 2003). Li et al (2006) examined the performance of one on-line and three off-line speed-based models and found little difference in terms of estimation error.

In the speed-based model an entire road is divided into segments. At both ends of a segment there is a detector present (see Figure 1). The average speed for each segment is calculated using the following equation:

$$\mathbf{V}_{\mathbf{average}} = \frac{\mathbf{2}}{\left[\frac{\mathbf{1}}{\mathbf{V_1}}\right] + \left[\frac{\mathbf{1}}{\mathbf{V_2}}\right]} \tag{1}$$

V_1 is the measured speed at the beginning of the segment and V_2 is the measured speed at the ending of the segment. The travel time is calculated by dividing the segment's length by the average speed. By summing the separate segments, an instantaneous route travel time can be estimated for the entire route.

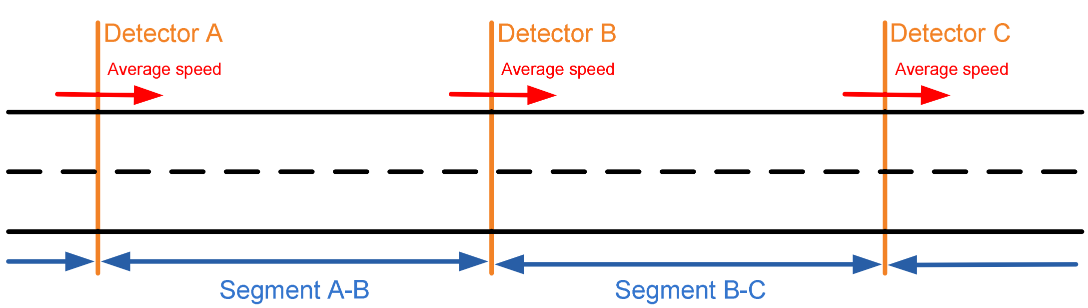


Figure 1 – Speed-based detector placement

Flow-based model

Another class of travel time estimation model considered in this paper is called “flow-based” model (Nam and Drew, 1996, 1998, 1999; Vanajakshi et al, 2009). This model makes use of the cumulative curves of traffic count at upstream and downstream detectors to capture the dynamic change of traffic in the segment. Travel time is calculated from the area under these two curves.

The original model from Nam and Drew suggested two different equations, one for free flow conditions and one for congested conditions. However, Vanajakshi et al (2009) found that the use of two different equations was unnecessary and by using only one equation for congested conditions resulted in a better travel time estimate for varying traffic flow conditions. Therefore we use only one equation for the congested conditions in this paper.

This model has the same segment placement as the speed-based model. But the detectors for this algorithm measure traffic flow instead of traffic speed. Based on the measured traffic flows the density on each segment is calculated with the following equitation:

$$k_{(t)} = \frac{Q_{in,(t)} - Q_{out,(t)}}{\Delta x}$$

In this equation $k_{(t)}$ is the density on a segment during interval t, $Q_{in,(t)}$ is the cumulative number of vehicles entering the segment during interval t, $Q_{out,(t)}$ the cumulative number of vehicles leaving the segment during the same interval, and Δx the length of the segment. With the measured traffic flow and estimated density the travel time for each segment is calculated with the following equation:

$$tt_{(t)} = \frac{\Delta x}{2} \cdot \frac{k_{(t)} + k_{(t-1)}}{q_{out,(t)}}$$

Here $tt_{(t)}$ is the estimated travel time for a segment at interval t, calculated with the density of the same interval ($k_{(t)}$), the density from the interval prior to the current interval ($k_{(t-1)}$), and the segment length (Δx).

Time slice model

Though our main interest is on on-line travel time estimation, the time slice model which is an off-line model is also considered for comparison purpose. The time slice model is more suited for historical travel time analyses. In case it is applied for real-time applications, a delay has to be taken into account. Unlike the above two models, the time slice model does not use all segment data from the same time-interval to estimate travel times. Instead, it determines when a vehicle enters each segment and uses the most up-to-date data available.

The equations used for calculating travel times for each segment can be the same as the previously mentioned models. The only difference is that this model uses data from different time-intervals to estimate travel times. For this research the equation from the speed-based model is used for the time slice model. For example a vehicle enters segment 1 at time t, the average speed on segment 1 is calculated as follow:

$$\mathbf{V}_{\mathbf{average}(1,t)} = \frac{2}{\left[\frac{1}{\mathbf{V}_{1(t)}}\right] + \left[\frac{1}{\mathbf{V}_{2(t)}}\right]}$$

Again V_1 is the measured speed at the beginning of the segment and V_2 is the measured speed at the ending of the segment. Time (t) determines data from which time-interval to be used. With the average speed the travel time for the first part of a vehicle's trajectory (segment 1) can be calculated, t(k). This travel time will be used to determine the average speed on segment 2:

$$\mathbf{V}_{\mathbf{average}(2,t)} = \frac{2}{\left[\frac{1}{\mathbf{V}_{1(t+t(k))}}\right] + \left[\frac{1}{\mathbf{V}_{2(t+t(k))}}\right]}.$$

This will continue on till the destination is reached. In real-time (on-line) situations the travel time for a vehicle entering at time t cannot be given since the data at moment t+t(k) is not available yet. In real time applications the delay of this model is equal to the travel time.

STUDY AREA AND DATA

This section describes about the study area and data used in this research.

Study area

The study area for this research is the metropolitan expressway (MEX), route #4, leading from Takaido towards the Tokyo ring (Miyake-zaka Junction). The length of the whole area is about 14 km, with two lanes in each direction. The Miyake-zaka Junction connects route #4 with the ring-road in Tokyo. During peak hours this ring is heavily congested. A simplified map of the study is given in Figure 2. The ETC-gates in the area are marked with their number in a circle. Length of each segment is given in meters and detectors are marked with blue line.

It can be seen that the existing detector placement on this freeway is very dense. Since travel time estimation problem under sparsely distributed detector is more difficult and therefore more challenging, in this paper we consider two scenarios of detector placement. The first one is exactly the same as the existing detector placement while in the second scenario about 70% of all the detectors have been dropped out from the research area (see Figure 3). Only the direction towards Tokyo was examined. More precisely traffic with the origin ETC-gate 251 towards either the destination ETC-gate 237 or 217. Maximum allowed speed on the MEX is 80 km/h.

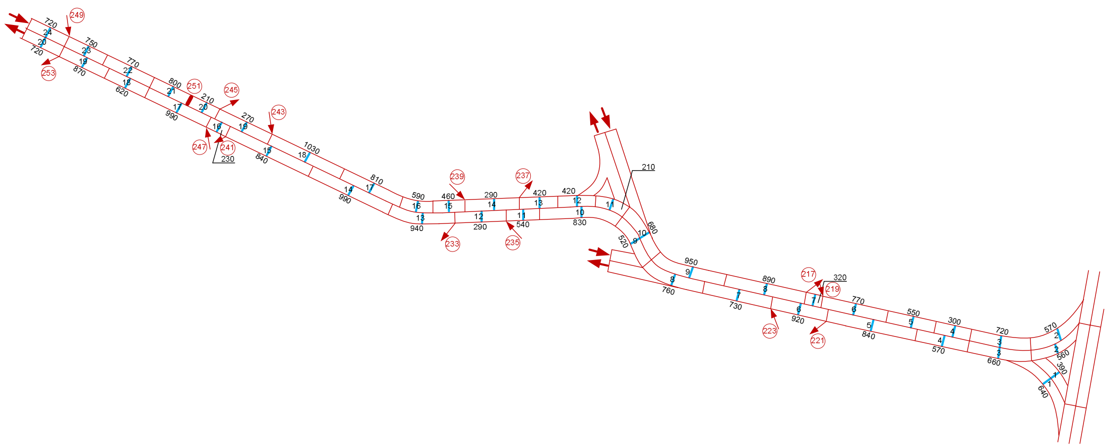


Figure 2 – Simplified map of study area with existing detector placement

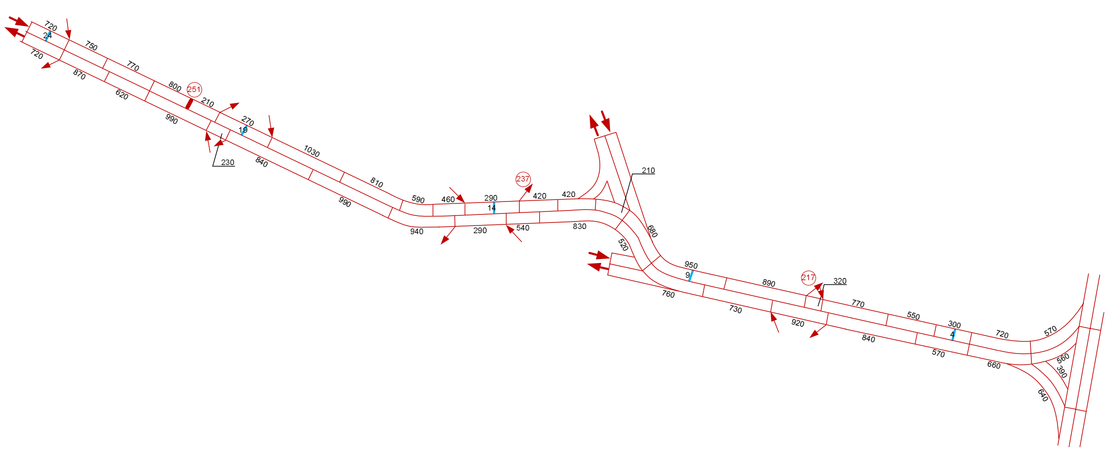


Figure 3 – Scenario with only about 30% of existing detector placement

Data

For this research two kinds of data were available, detector data and ETC data. In this study, detector data was collected by ultrasonic sensors. Furthermore all detectors are dual loops, which mean the measured speeds are more precise. For each segment aggregated loop data was available with a five minutes update interval.

The second data source for this research is ETC data. Both the ETC data and detector data are from the period July 1, 2006 till July 7, 2006. During this period the ETC market penetration was about 60%. From the ETC data it is possible to determine when and where each vehicle entered and left the research area. Based on the enter time and exit time the travel time of each vehicle can be obtained.

No errors are expected in the ETC data, although some vehicles showed an exceptional long travel time. Based on the average travel time each five minutes, these exceptional vehicles were filtered out. The travel times obtained from the ETC data are assumed to be the actual travel times. This will be the data to compare all estimates against. Although the ETC data will also be used for making travel time estimates, the data will still be a valid source for comparison. This is because there is a little delay between the data used for estimations and for comparison. This is illustrated in Figure 4.

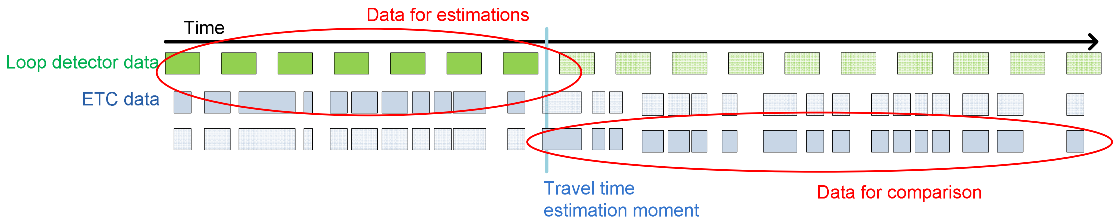


Figure 4 – Delay between estimation data and comparison data

Since in on-line travel time estimation vehicles get the travel time at the beginning of their journey, the actual travel time is yet to be realized. This means the ETC data used for comparison is not available for fusion, in Figure 4 on the left side of the "travel time estimation point". The right side is all the data that is available for comparison. So for each moment in time the data used for estimation is unrelated to the actual travel time (comparison) data.

TRAVEL TIME ESTIMATE FROM DETECTOR DATA

In this section, we investigate the performance of the existing models by comparing travel time estimated from these models with the actual travel time from ETC data. We first consider the case of existing detector placement followed by the scenario with relatively large detector spacing.

Performance under existing detector placement scenario

In this case, all detector data from the existing placement is used as input for the models. We first compare the performance of the instantaneous models: speed-based and flow-based models.

Performance of instantaneous models

For comparison of the accuracies of different models, an average overestimation and an average underestimation were determined for each model. For each interval it was determined whether the model overestimated or underestimated the travel time according to the ETC data. This way it was possible to keep overestimations and underestimations separate. By summing all overestimations and dividing it by the number of times travel time was overestimated, an average overestimation was calculated. The same procedure goes for the underestimation. Results of the comparison between the flow-based model and the speed-based model are shown in Table 1. Average absolute error is the average of the average overestimation and average underestimation. The first row in Table 1 is empty because there was an accident on that day (July 1, 2006) and the flow-based model seems to behave very strangely. By leaving out the result from this day, the comparison is fairer for the typical situation.

Table 1 – Error comparison between flow-based and speed-based models (from ETC-gate 251 to 217)

Day	**Flow-based model** (with all loop detectors)			**Speed-based model** (with all loop detectors)		
	Average overestimation	*Average underestimation*	*Average absolute error*	*Average overestimation*	*Average underestimation*	*Average absolute error*
1-Jul-06	-	-	-	-	-	-
2-Jul-06	85.30071	65.3429	75.32179	86.32575	47.437	66.8814
3-Jul-06	70.49105	47.9351	59.21305	50.41537	37.1489	43.78214
4-Jul-06	115.3225	57.393	86.35779	81.27327	50.0184	65.64581
5-Jul-06	138.3732	95.589	116.9811	109.3794	74.1254	91.75241
6-Jul-06	67.09878	51.0848	59.09177	54.148	35.9148	45.03138
7-Jul-06	124.8942	83.3378	104.116	93.97934	70.6825	82.33094
		$\sum =$	**501.0815**		$\sum =$	**395.4241**

Figure 5 shows an example of the travel time estimated from the flow-based (NamDrew) and the speed-based (Extrap) models shown in blue and pink lines respectively. The yellow dotted line represents the actual travel times obtained from ETC data.

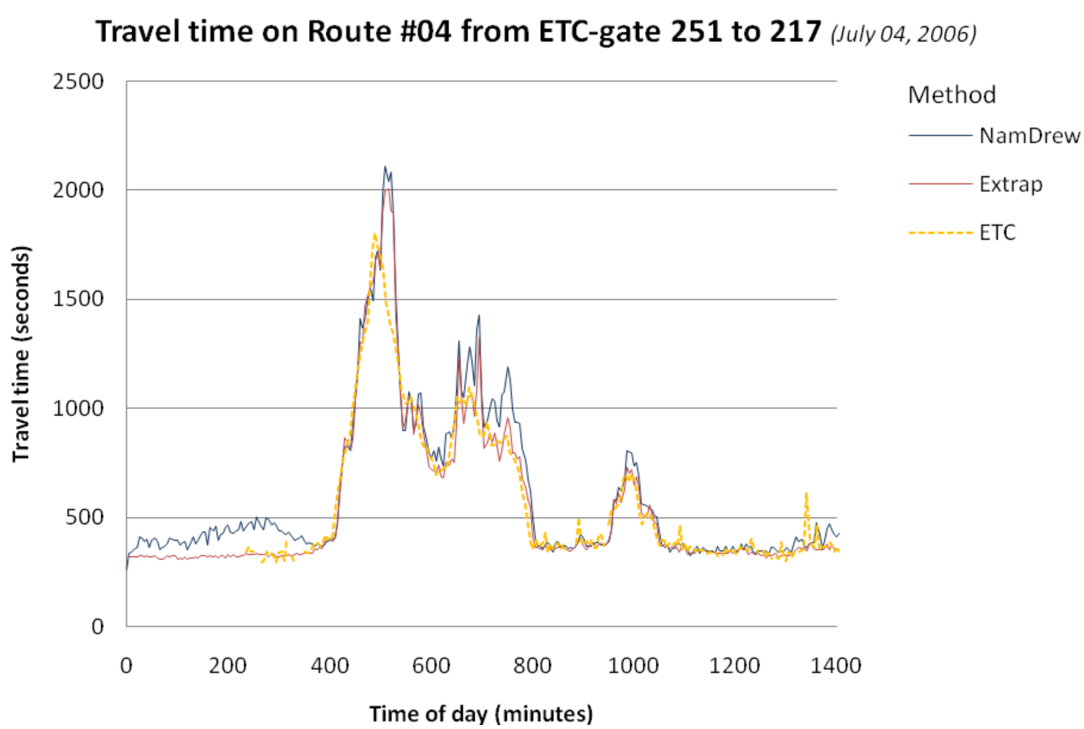


Figure 5 – Travel time estimates for the current situation

Based on the results shown above it is clear that the speed-based model performs better than the flow-based model. On all six days the average absolute error of the speed-based model is lower than that of the flow-based model. For the remaining of this paper, more emphasis is given to speed-based model during the development of the data fusion method.

Speed-based model vs. Time slice model

As mentioned before, the time slice model is considered here for comparison purpose. As the time slice is performed off-line, the data from all time intervals can be used during each estimation interval. Estimated travel times from this model are expected to be the most accurate of all existing models considered in this research. Example of the results is shown in Figure 6. The error results for the seven examined days are shown in Table 2. Compared to speed-based model, it is clear that the time slice model more accurately follow the actual travel times obtained from the ETC data.

Table 2 – Estimate error comparison between time slice model and speed-based model (from ETC-gate 251 to 217)

Day	**Time slice model** (with all loop detectors)			**Speed-based model** (with all loop detectors)		
	Average overestimation	*Average underestimation*	*Average absolute error*	*Average overestimation*	*Average underestimation*	*Average absolute error*
1-Jul-06	33.53924	39.7033	36.62129	129.1858	72.7797	100.9828
2-Jul-06	17.93999	30.5059	24.22292	86.32575	47.437	66.8814
3-Jul-06	27.17834	23.438	25.30816	50.41537	37.1489	43.78214
4-Jul-06	23.24303	34.0556	28.6493	81.27327	50.0184	65.64581
5-Jul-06	34.15193	43.3475	38.74971	109.3794	74.1254	91.75241
6-Jul-06	25.58925	25.0761	25.33269	54.148	35.9148	45.03138
7-Jul-06	26.54383	36.5633	31.55354	93.97934	70.6825	82.33094
		$\sum =$	**210.4376**		$\sum =$	**496.4069**

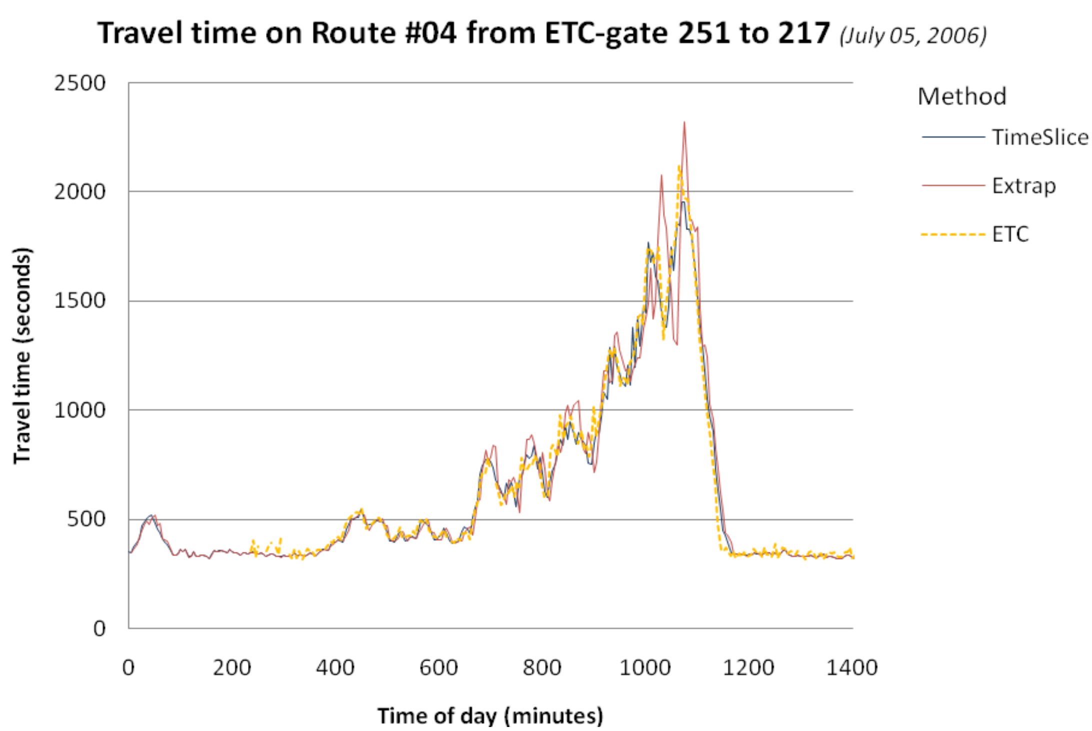


Figure 6 – Travel time estimates by the time slice model and the speed-based model

Performance under relatively large detector spacing scenario

As stated before, about 70% of all detectors were dropped in this scenario. Travel time estimates will be analyzed for the Speed-based model and the Time slice model. The results are shown in Table 3 and Table 4.

Table 3 – Travel time estimate accuracy for the Speed-based model with fewer detectors (from ETC-gate 251 to 217)

Day	Speed-based model (with 30% of all loop detectors)			Speed-based model (with all loop detectors)		
	Average overestimation	*Average underestimation*	*Average absolute error*	*Average overestimation*	*Average underestimation*	*Average absolute error*
1-Jul-06	104.9583	162.861	133.9099	129.1858	72.7797	100.9828
2-Jul-06	56.46005	70.3793	63.41968	86.32575	47.437	66.8814
3-Jul-06	73.84489	25.3089	49.57688	50.41537	37.1489	43.78214
4-Jul-06	87.79284	95.1859	91.48939	81.27327	50.0184	65.64581
5-Jul-06	110.9346	116.031	113.4826	109.3794	74.1254	91.75241
6-Jul-06	93.85418	52.5319	73.19304	54.148	35.9148	45.03138
7-Jul-06	117.525	117.347	117.4359	93.97934	70.6825	82.33094
		$\sum =$	**642.5074**			**496.4069**

Table 4 – Travel time estimate accuracy for the Time slice model with fewer detectors (from ETC-gate 251 to 217)

Day	Time slice model (with 30% of all loop detectors)			Time slice model (with all loop detectors)		
	Average overestimation	*Average underestimation*	*Average absolute error*	*Average overestimation*	*Average underestimation*	*Average absolute error*
1-Jul-06	45.34729	114.068	79.7075	33.53924	39.7033	36.62129
2-Jul-06	46.81081	63.4606	55.1357	17.93999	30.5059	24.22292
3-Jul-06	64.57156	21.8015	43.18654	27.17834	23.438	25.30816
4-Jul-06	48.63072	63.7232	56.17694	23.24303	34.0556	28.6493
5-Jul-06	75.51659	107.841	91.67896	34.15193	43.3475	38.74971
6-Jul-06	75.63694	40.695	58.16597	25.58925	25.0761	25.33269
7-Jul-06	81.09806	95.823	88.46053	26.54383	36.5633	31.55354
		$\sum =$	**472.5121**		$\sum =$	**210.4376**

As expected, both models perform less accurate with fewer detectors. Interesting to see is that the accuracy of the Time slice model with fewer detectors is better than the Speed-based model with the very dense detector placement.

Error patterns

This section describes the analysis of travel time estimation error over time. For each time-interval the estimates are compared to the actual travel time. The error can then be plotted into a graph. By keeping the actual travel time (the yellow dotted line) in the same graph it helps understand at what traffic conditions the models fail and how the errors are. Figure 7 shows an example of the plot of error on July 2, 2006. It is clear from this figure that when travel time increases the models underestimate travel times, while at decreasing travel times the models overestimate. Another interesting behavior is that during free-flow conditions, the actual travel time is usually between the estimates of the flow-based model and the speed-based model. At congested periods this behavior is not visible. So besides increasing and decreasing travel time situations, also free-flow traffic and congested traffic are situations that need to be treated differently. Note that similar error patterns were also observed in all examined days.

The figure clearly shows that there is a correlation between estimate error and traffic condition. This means estimate errors can be corrected with certain correction-rules for certain traffic conditions. Statistical correction method with assumption that the error is randomly distributed may not perform well in this case.

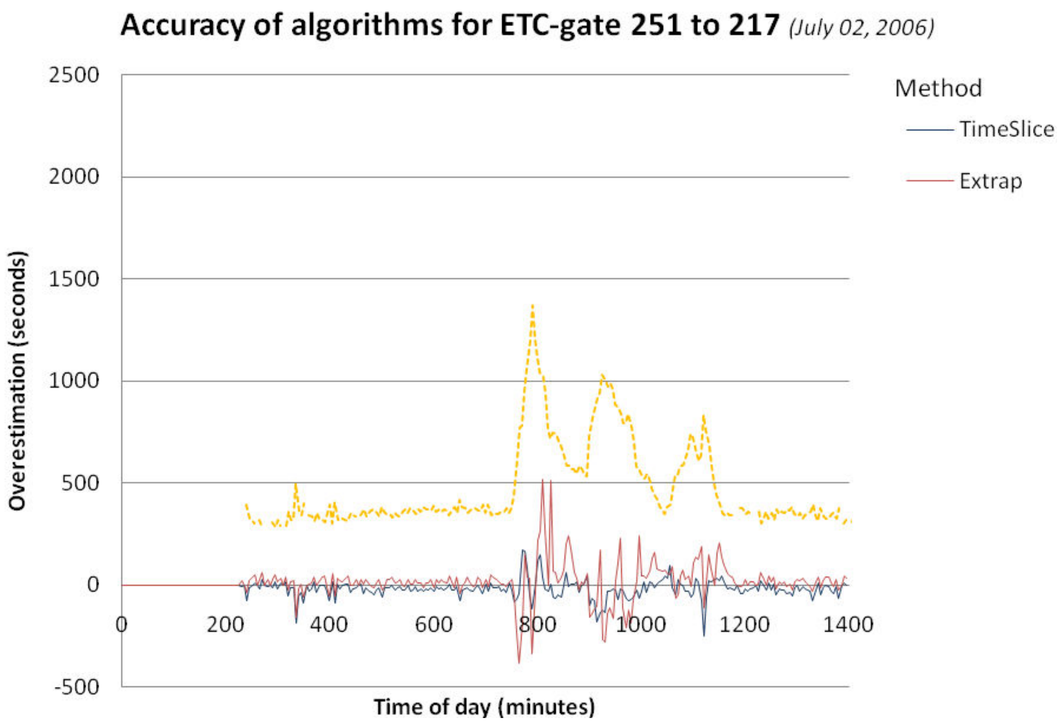


Figure 7 – Travel time estimate errors over time for situation with fewer detectors

TRAVEL TIME CORRECTION USING ETC DATA

The fusion model developed in this research uses an existing travel time estimation model as basis. By evaluating the estimates from previous intervals with the incoming ETC data, the current estimate by the existing model would be corrected to a more accurate estimate. Based on the observed error patterns, this paper considers three fusion concepts. The first one examined was a model running the speed-based and the flow-based models in parallel. The second concept uses only the speed-based model as basis. The last concept was to introduce the moving average and some boundaries to where correction would be applied.

Fusion 1: Corrections based on previous errors on two models running in parallel

The first fusion concept was based on the results of the performance analyses under existing detector placement. It seemed that the flow-based model overall overestimated the travel time, while the speed-based model overall underestimated the travel time (see Figure 5).

By averaging between these two models, travel time estimates were expected to improve. The correction rules and traffic condition identifications were as follow:

1. In case the last two evaluated intervals each time one model overestimated the travel time and one model underestimated the travel time. It is assumed that in the current interval the actual travel time is between the two estimates of the two models.

   By determining the difference between the two estimates of the last evaluated interval, let this be Δm. And by determining how much travel time was underestimated by the lowest travel time estimate, let this be Δu. A ratio ($\Delta u / \Delta m$) can be obtained to use to calculate the estimate for the current interval. For the current interval the difference between the two models is determined, let this be ΔM. And by taking the lowest travel time estimate and adding $\Delta M \times (\Delta u / \Delta m)$ a travel time output for the fusion model is obtained.

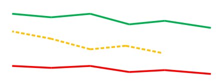
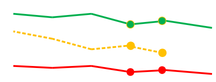
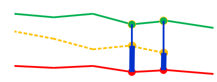
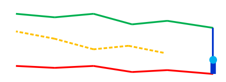
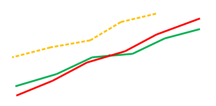
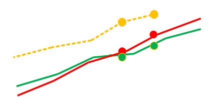
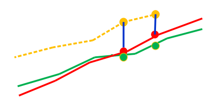
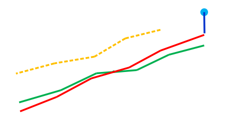
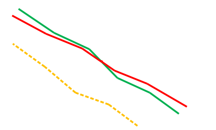
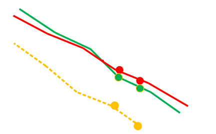
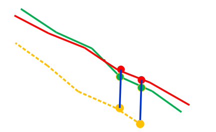
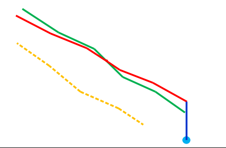
2. In case the last two evaluated intervals each time both models underestimated the travel time. It is assumed that the travel time in the current situation is increasing and thus the travel time is being underestimated. In this case the average error of the last two evaluated intervals is added to the current estimate of the speed-based model, thus the output of the fusion model.

3. In case the last two evaluated intervals each time both models overestimated the travel time. It is assumed that the travel time in the current situation is decreasing and thus the travel time is being overestimated. In this case the average error of the last two evaluated intervals is deducted from the current estimate of the speed-based model, thus the output of the fusion model.

4. For the rest of the situations the travel time estimates by the speed-based model are used as output of the fusion model.

By last two evaluated intervals, it means the last two intervals where ETC data is available. In real-time applications this is usually a few intervals back, depending on the delay of the ETC data. Illustrations of the correction methods are given in Table 5. The yellow dotted line is the ETC data, the green line is the flow-based model, the red line is the speed-based model, and the blue dot is the corrected estimation.

Table 5 – Illustrations of corrections for first fusion model

Correction rule	Situation	Recognize situation	Determine error	Correction
#1				
#2				
#3				

Unfortunately, the effect of this data fusion concept is minimal as shown in Table 6. Although travel time estimates improved in most situations (5 out of the 7 cases), it is not sufficient. To see more explicitly where travel time estimates were improved, errors over time were plotted (see Figure 8 for an example).

Table 6 – Travel time estimate accuracy for the First fusion model (from ETC-gate 251 to 217)

Day	First fusion model (with all loop detectors)			Speed-based model (with all loop detectors)		
	Average overestimation	*Average underestimation*	*Average absolute error*	*Average overestimation*	*Average underestimation*	*Average absolute error*
1-Jul-06	123.8824	74.502	99.19218	129.1858	72.7797	100.9828
2-Jul-06	84.58546	49.0246	66.80503	86.32575	47.437	66.8814
3-Jul-06	48.9654	37.5799	43.27266	50.41537	37.1489	43.78214
4-Jul-06	85.60895	48.2331	66.92104	81.27327	50.0184	65.64581
5-Jul-06	100.9127	76.4633	88.68801	109.3794	74.1254	91.75241
6-Jul-06	51.71613	35.6665	43.69129	54.148	35.9148	45.03138
7-Jul-06	93.04104	72.6042	82.8226	93.97934	70.6825	82.33094
		$\sum =$	**491.3928**		$\sum =$	**496.4069**

In Figure 8, the error of the first fusion model is plotted in blue and that from speed-based model is plotted in pink. Both graphs seem to be identical. This is because the defined condition identifications are too specific. Only a few times the estimates were corrected. This explains on such a small improvement.

In order to achieve better improvements, the identification rules need to be broader. With the current method of identifying traffic conditions, it is very difficult to make the rules broader. Not only two models are running parallel, also the data used for identification has a relatively long delay (equal to the travel time). For the next fusion concept only one model is used as basis and data with a smaller delay is used for identification of the traffic condition.

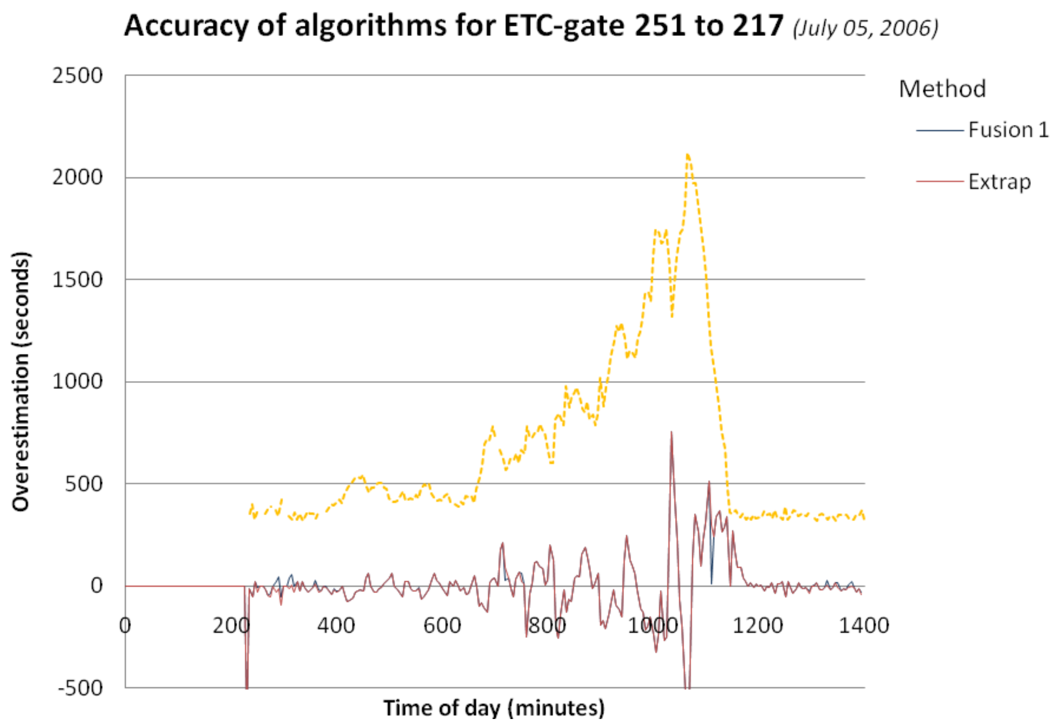


Figure 8 – Travel time estimate errors over time for first fusion model

Fusion 2: Corrections on one model based on current trend

Since the first fusion model is not very successful, this second model starts in a more simple way. For the identification of the traffic condition, only the last two estimates of the speed-based model are used. As for the correction of the estimates, the last two evaluated intervals are used to determine the error and this error is assumed to be the same in the current interval. This is performed under a scenario with relatively large detector spacing.

The correction rules and traffic condition identifications were as follow:

1. In case the last two estimates by the speed-based model are increasing by more than 20%, then it is assumed that the travel time in the current situation is increasing as well and thus the travel time is being underestimated.

   In this case the average error of the last two evaluated intervals is added to the current estimate of the speed-based model, thus the output of the fusion model.

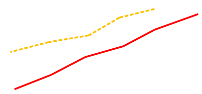
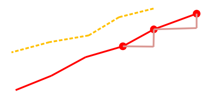
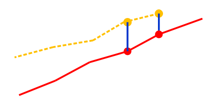
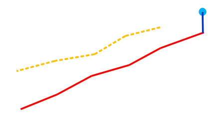
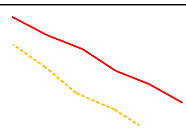
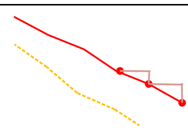
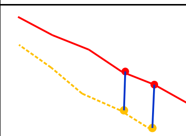
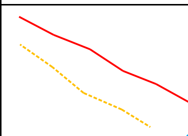
2. In case the last two estimates by the speed-based model are decreasing by more than 20%, then it is assumed that the travel time in the current situation is decreasing and thus the travel time is being overestimated.

   In this case the average error of the last two evaluated intervals is deducted from the current estimate of the speed-based model, thus the output of the fusion model.

3. For the rest of the situations no correction is performed.

Illustrations of the above described correction methods are shown in Table 7. The yellow dotted line is the ETC data, the red line is the speed-based model, and the blue dot is the corrected estimation.

Table 7 – Illustrations of corrections for second fusion model

Correction rule	Situation	Recognize situation	Determine error	Correction
#1				
#2				

Unfortunately, the results for this fusion concept turned out to be unsuccessful. In Table 8 the accuracies of this model are shown. Only one out of the seven cases the travel time estimate is improved. To investigate why the travel time estimates did not improve, the errors over time were plotted (see example in Figure 9).

Table 8 – Travel time estimate accuracy for the Second fusion model (from ETC-gate 251 to 217)

Day	Second fusion model (with 30% of all loop detectors)			Speed-based model (with 30% of all loop detectors)		
	Average overestimation	*Average underestimation*	*Average absolute error*	*Average overestimation*	*Average underestimation*	*Average absolute error*
1-Jul-06	105.1136	170.564	137.8386	104.9583	162.861	133.9099
2-Jul-06	60.69916	73.0866	66.8929	56.46005	70.3793	63.41968
3-Jul-06	73.54949	25.8932	49.72133	73.84489	25.3089	49.57688
4-Jul-06	82.2951	103.267	92.78099	87.79284	95.1859	91.48939
5-Jul-06	114.6733	118.513	116.5931	110.9346	116.031	113.4826
6-Jul-06	91.8775	51.5757	71.72658	93.85418	52.5319	73.19304
7-Jul-06	120.3491	119.775	120.0621	117.525	117.347	117.4359
		∑ =	**655.6156**		∑ =	**642.5074**

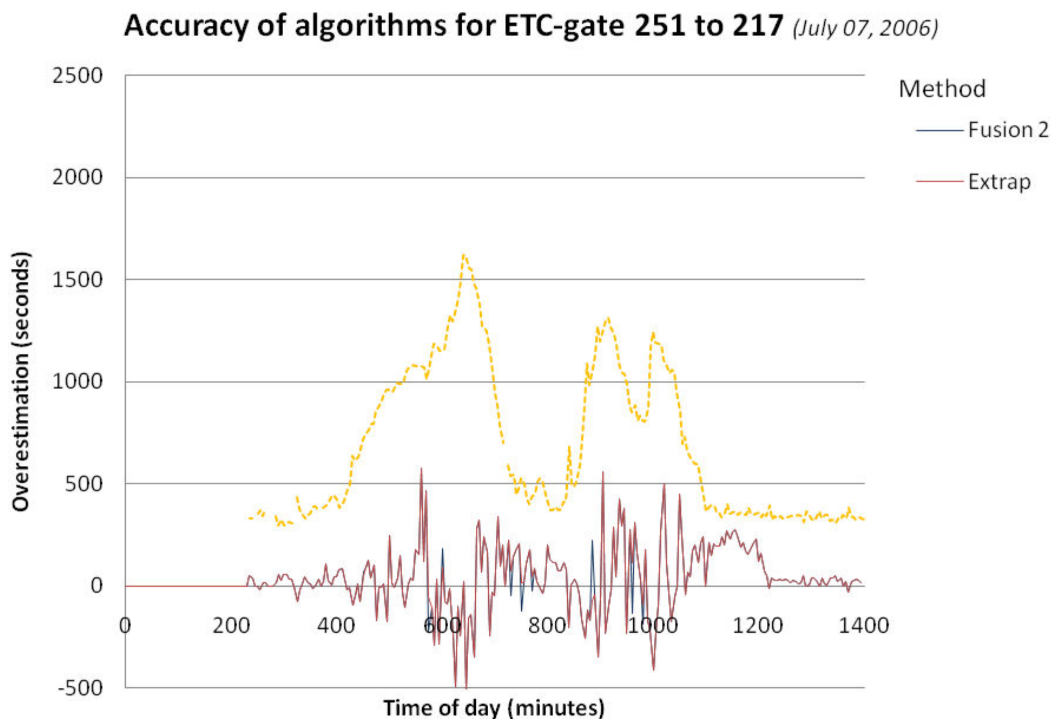


Figure 9 – Travel time estimate errors over time for second fusion model

As can be seen in Figure 9, in some situations the travel time does get improved. But there are also situations where travel time estimations get worse. In the traffic condition identification rules the increasing and decreasing criteria (above set to 20%) have been changed to investigate if it is possible to filter out the wrong situations. Even variations of the above mentioned rules were tested, they had very little effect on which situations would get "corrected".

As shown in Figure 10, the estimates provided by the Speed-based model under large spacing are very fluctuate. Such fluctuation could lead to a situation that the traffic condition identification rules are rarely activated. It is expected that introducing a moving average could lead to a better improvement.

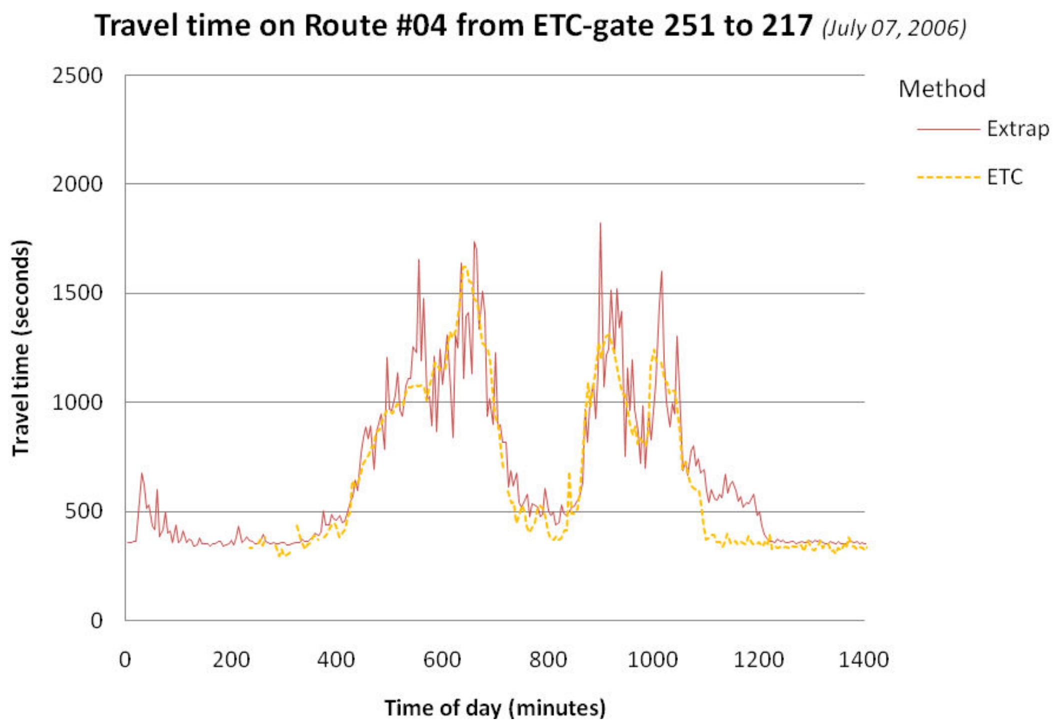


Figure 10 – Travel time estimates by the Speed-based model (in the situation of fewer detectors)

Fusion 3: Corrections based on current trend on one model with moving average

In order to make the loop data suitable for identifying traffic conditions, a moving average is introduced. By constantly averaging between values of the current interval and the prior interval, the travel time from speed-based model can be stabilized. Besides the output becoming stable, it also turned out that travel time estimates become more accurate.

For this research two boundaries have been set for applying the moving average. The first boundary is to only use data from the current interval and previous interval. Although averaging between more values resulted in an even more stable output, the delay also increased. Since data with as little delay as possible is preferred, only an average of two intervals is chosen.

The second boundary is to only apply the moving average when the traffic condition is around free-flow condition. This is because the speed-based model is already very accurate under free-flow condition (Li et al, 2006). This is realized by only averaging when the estimated travel time from speed-based model is larger than [1.2×Free-flow travel time]. By only applying the moving average to speed-based model without any other correction, the estimation accuracy was found to improve. Figure 11 shows a sample of the stabilized output.

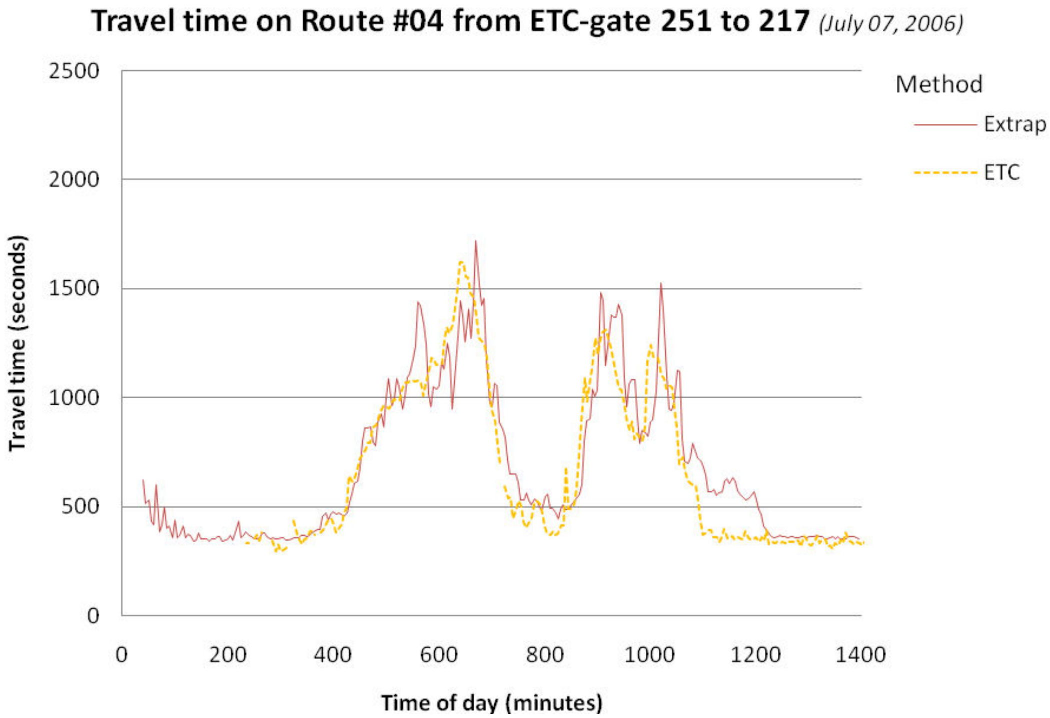


Figure 11 – Travel time estimates by the Speed-based model with averaging (in the situation of fewer detectors)

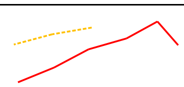
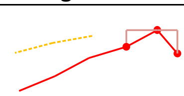
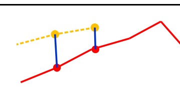
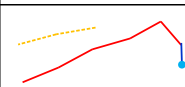
With the moving average applications it is expected that the speed-based model's output is stable enough for identifying traffic conditions. One condition and correction rule that was used for this third model is as follow:

1. In case the last two estimates by the speed-based model are first increasing and then decreasing by 20% or more compared to the previous interval. Then it is assumed that the travel time in the current situation is overestimated.

   In this case the average error of the last two evaluated intervals is deducted from the current estimate from the speed-based model, thus the output of the fusion model.

The above described correction method is illustrated in Table 9. The yellow dotted line is the ETC data, the red line is the speed-based model, and the blue dot is the corrected estimation.

Table 9 – Illustrations of correction for third fusion model

Correction rule	Situation	Recognize situation	Determine error	Correction
#1				

The reason for this rule is that all peaks of the travel time estimation from Speed-based model are always overestimated. And after the peaks there is usually a descending part. Accuracy results of this third model are shown in Table 10.

The green values in the tables are the values for the existing scenario with dense detector placement. The original target for the fusion model is to achieve the same level of accuracy as that under dense detector (but without fusion). Unfortunately, the improvement from this third model does not achieve this target yet. Anyway, as seen in Table 10, the average error of the third model decreases about one minute when comparing with the speed-based model with 30% of detector placement. This illustrates that with more conditions and correction rules, the higher accuracy could be achieved. Further research is required to investigate this.

Table 10 – Travel time estimate accuracy the Third fusion model (from ETC-gate 251 to 217)

Day	Third fusion model (with 30% of all loop detectors)			Speed-based model (with 30% of all loop detectors)		
	Average overestimation	*Average underestimation*	*Average absolute error*	*Average overestimation*	*Average underestimation*	*Average absolute error*
1-Jul-06	93.4612	131.599	112.5299	104.9583	162.861	133.9099
2-Jul-06	57.1456	64.0745	60.61005	56.46005	70.3793	63.41968
3-Jul-06	69.15854	29.6859	49.42222	73.84489	25.3089	49.57688
4-Jul-06	85.91212	75.7059	80.8090	87.79284	95.1859	91.48939
5-Jul-06	100.5495	105.675	103.1121	110.9346	116.031	113.4826
6-Jul-06	89.55234	44.1939	66.87313	93.85418	52.5319	73.19304
7-Jul-06	113.881	105.816	109.8487	117.525	117.347	117.4359
		∑ =	**583.2050**		∑ =	**642.5074**
			496.4069			

CONCLUSIONS AND RECOMMENDATIONS

In this paper, three fusion concepts have been examined. Unlike previous data fusion attempts, the concepts considered here are based on the analysis of error patterns. Though the improvement is not as much as expected, it illustrates the possibility to correct travel time estimate using the observed error patterns.

In the first fusion model, traffic conditions were identified by looking at time-intervals where ETC data was available. A small improvement was observed in this case. However, this method of identifying traffic conditions may not suit well under small traffic variations. This is because there is a relatively long delay before a condition has been recognized. For further research it might be interesting to run multiple models to identify very specific situations.

For the second fusion model, to minimize the delay in identifying traffic conditions, the last two estimates from the speed-based model were used for identification. This means that more up-to-date data would be used, but it turned out to be an unsuccessful fusion. As the spacing between detectors is relatively large, travel time estimated by the speed-based model were very fluctuate. By relying on such a fluctuating estimate, the traffic condition identification rule was rarely activated.

In the last model, moving average was introduced to the speed-based model. This is to stabilize the output and make it suitable for identifying traffic conditions. And in this model a criterion was introduced, so that free-flow condition remains untouched. This is because speed-based model was found to be accurate when traffic is around free-flow condition. Only one simple correction rule was made for the third model. Though the improvement was not as much as expected, the average improvement of this model is about one minute compared to the model without fusion method. Further development will most likely result in more accuracy.

For further research, more detailed condition and correction rules need to be developed, for example by using more variables to identify traffic conditions. Travel time estimations by instantaneous models depending on detector data clearly have systematic errors. Correcting these errors without statistical methods is possible. And this research has pointed out that identifying traffic conditions using detector data needs some averaging over time.

REFERENCE

Bovy, P. H. L. and R. Thijs (2000). Estimators of travel time for road networks: new developments, evaluation results, and applications. Delft University Press, Delft, The Netherlands.

Chen, M. and S. I. J. Chien (2001). Dynamic freeway travel time prediction using probe vehicle data: link-based vs. path-based. Transportation Research Record 1768, pp.157–161.

Chu, L., J. Oh and W. Recker (2005). Adaptive kalman filter based freeway travel time estimation. TRB annual meeting 2005.

El Faouzi, N.-E., L. A. Klein and O. D. Mouzon (2009). Improving Travel Time Estimates from Inductive Loop and Toll Collection Data with Dempster-Shafer Data Fusion, Transportation research record, 2129, 73-80.

Kothuri, S. M., K. Tufte, S. Ahn and R. L. Bertini (2006). Development of an ITS data archive application for improving freeway travel time estimation. Proceedings of the IEEE Intelligent Transportation Systems Conference, 2006, 17-20 September, 1263–1268.

Li, R., G. Rose and M. Sarvi (2006). Evaluation of speed-based travel time estimation models. Journal of Transportation Engineering, 132, 540-547.

Linveld, C. D. R., R. Thijs, P. H. Bovy and Van der Zijpp, N. J. (2000). Evaluation of online travel time estimators and predictors. Transportation Resource Record. 1719, 45–53.

Liu, Y., X. Lai and G. L. Chang (2006). Detector placement strategies for freeway travel time estimation. Proceedings of the IEEE Intelligent Transportation Systems Conference, 2006, 17-20 September, 499–504.

Nam, D. H. and D. R. Drew (1996). Traffic dynamics: methods for estimating freeway travel times in real time from flow measurement. Journal of Transportation Engineering, 122, 185-191.

Nam, D. H. and D. R. Drew (1998). Analyzing freeway traffic under congestion: traffic dynamics approach. Journal of Transportation Engineering, 124, 208-212.

Nam, D. H. and D. R. Drew (1999). Automatic measurement of traffic variables for intelligent transportation systems applications. Transportation Research Part B, 33, 437-457.

Nanthawichit, C., T. Nakatsuji and H. Suzuki (2003). Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway. Transportation research record , 1855, 49-59.

Sun, L., J. Yang and H. Mahmassani (2008). Travel time estimation based on piecewise truncated quadratic speed trajectory. Transportation Research Part A, 42, 173-186.

Turner, S. M., S. M. Eisele, R. J. Benz and D. J. Holdener (1998). Travel time data collection handbook, Report FHWA-PL-98-035, Federal Highway Administration, U.S. Department of Transportation.

Van Lint, J. W. C. and Van der Zijpp, N. J. (2003). Improving a travel time estimation algorithm by using dual loop detectors. Transportation Resource Record. 1855, 41-48.

Vanajakshi, L. D., B. M. Williams and L. R. Rilett (2009). Improved flow-based travel time estimation method from point detector data for freeways. Journal of Transportation Engineering, 135, 26-36.