Evaluation of Speed-Based Travel Time Estimation Models

Ruimin Li[1]; Geoffrey Rose[2]; and Majid Sarvi[3]

Abstract: Travel time estimation models, which rely on speed data provided from point detectors (usually inductive loops), have application in travel time prediction and network performance monitoring. Unfortunately there are limited, and at times counterintuitive, results in the literature about their performance. This paper focuses on the field evaluation of four speed-based travel time estimation models, namely, the instantaneous model, the time slice model, the dynamic time slice model, and the linear model. Those models are evaluated using data from two operational motorways in Melbourne, Australia. Travel time estimation errors are quantified against actual travel times measured using a timed number plate survey and time-stamped toll tag data. There was little difference in the travel time estimation error across the models and they were all found to underestimate actual travel times. Errors ranged from about 7% in the off peak up to 15% in the peak. Marginal improvements in model performance were achieved through careful selection of which detectors provide input for each section (upstream, downstream, or the average of those values) and by conversion of the inputs from time mean speed to an estimate of space mean speed.

DOI: 10.1061/(ASCE)0733-947X(2006)132:7(540)

CE Database subject headings: Travel time; Estimation; Algorithms; Evaluation; Traffic management.

Introduction

Travel time information is of interest to both road users and network managers. Travel time data are rarely directly collected in real time except where technology such as global positioning system (GPS) enabled probe vehicles, toll tag readers, or automatic licence plate recognition systems are deployed. The most common approach is to estimate travel times from speed data provided by inductive loop detectors or other point-based sensors.

This paper is concerned with the models used to estimate motorway travel times from speed data. Speed-based travel time estimation models have been widely used because of their simplicity (Turner et al. 1998; Smith et al. 2004). These models can be used in two application contexts: on-line, where the focus is on real-time application, and off-line where historic data analysis is of interest.

In an on-line context, travel time prediction models provide the key input needed for communicating travel time information to road users through roadside signs, the internet, radio, mobile phones, or in-vehicle devices. In the real-time context, both the accuracy and timeliness of the travel time estimates are of interest.

[1]Ph.D. Student, Dept. of Civil Engineering, Institute of Transportation Research, Monash Univ., Victoria 3800, Australia (corresponding author). E-mail: ruimin.li@eng.monash.edu.au

[2]Associate Professor, Dept. of Civil Engineering, Institute of Transportation Research, Monash Univ., Victoria 3800, Australia. E-mail: geoff.rose@eng.monash.edu.au

[3]Lecturer, Dept. of Civil Engineering, Institute of Transportation Research, Monash Univ., Victoria 3800, Australia. E-mail: majid.sarvi@eng.monash.edu.au

Note. Discussion open until December 1, 2006. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on April 12, 2005; approved on December 9, 2005. This paper is part of the ***Journal of Transportation Engineering***, Vol. 132, No. 7, July 1, 2006. ©ASCE, ISSN 0733-947X/2006/7-540–547/$25.00.

Travel time estimation models also have application in an off-line context where they can be viewed as a way of reconstructing travel times from a set of historic speed data. In this context, timeliness loses its importance. Road system managers can use estimated travel times to monitor the performance of road network over time. The estimated travel times are also commonly used by model developers to provide a basis for assessing the accuracy of proposed predictive models (Lindveld et al. 2000; Bajwa et al. 2003). We characterize a predictive model as one which uses only data up to and including time k to predict the travel time for a vehicle starting its journey at time k. When developing a predictive model, the "true" values of the dependent variable (T for travel time) are needed to assess model accuracy, however, they are rarely directly measured in the field. Model developers rely on historical data and often use another model to produce an estimate (or reconstruct a value) of the travel time for a vehicle starting its journey at time k (T_{est}) on the basis of data from time period k and later time periods. The assessment of the accuracy of proposed predictive models are then based on how well the predictions match T_{est}. The implicit assumption is that $T_{\text{est}} \approx T$. The research reported in this paper uses field data to assess the validity of that assumption by focusing on how well four different models reconstruct actual travel times.

The emergence of new speed-based travel time estimation models and refinements to existing models has heightened the need for comprehensive evaluation of model performance. In the research reported here, measured travel times from two field sites are used to quantify the accuracy of different travel time estimation models. The results provide valuable insight into the field performance of these models and have relevance to model developers/researchers as well as practitioners.

In this work, the alternative formulations for the four speed-based travel time estimation models and validation results reported in the literature are summarized and reviewed first. We then describe the field data used in the empirical work undertaken in this research. Validation results are reported in the following section along with the results obtained from model refinements. Finally, conclusions and research directions are provided.

Speed-Based Travel Time Estimation Models

Four alternative speed-based travel time estimation model formulations have been proposed in the literature. Those models are described and then validation results which have been reported in the literature are reviewed.

Model Formulations

The travel time estimation models considered here rely on average speed data collected from point detectors (e.g., inductive loops) which are commonly installed about every 500 m on motorways. Those detectors report average speed data for periods which are typically of 20 or 30 s duration. The historical speed data can be used to reconstruct the trajectory of a hypothetical vehicle which starts its journey at the beginning of the road section of interest at a known time. The difficulty is that the speeds vary over time and depend on when a vehicle is traversing a particular segment/link. The key difference between the models described below is the extent to which they account for those variations in speed over time.

In each of the models considered here, the freeway is represented as a series of sections (or links) each bounded by two consecutive loop detectors (denoted here as points a and b). Under the notation system used here, each of the models aim to estimate the travel time of a vehicle departing at time k and traveling along n sections of the motorway.

Instantaneous Model

The instantaneous model uses the speeds collected from each link at time k. The travel time for each link is calculated as the link length divided by the average of the upstream and downstream speeds

$$t(i,k) = \frac{2l_i}{v(i_a,k) + v(i_b,k)} \tag{1}$$

where $v(i_a,k)$ and $v(i_b,k)$=measured speeds at the extremities of link i at time k; and $t(i,k)$=estimated travel time for a vehicle departing at time k on link i. Note that speeds from only two points on each link are used to estimate link travel times. Speed variations within a link, which may be more likely on longer links, would result in errors in the travel time estimates.

The total travel time $[T(k)]$, for a vehicle beginning its journey at time k, is then calculated by summing the estimates from the n segments

$$T(k) = \sum_{i=1}^{n} t(i,k) \tag{2}$$

Since the model relies only on speeds at time k, it assumes that these speeds will not change dramatically over the time it takes the vehicle to traverse the network. Given the dynamic characteristic of traffic, this model is a rough approximation to the actual travel time. However during uncongested, incident free conditions when speeds may vary little over time, or during congestion when speeds are relatively stable, it could be expected to provide a reliable estimate of travel time.

One advantage of the instantaneous model is that it can be used on-line to provide real time travel time information since it relies only on current traffic measurements (i.e., measurements at time k). The VicRoads Drive Time system which calculates freeway time travels for real-time display on roadside signs in Melbourne, Australia (Kloot 1999) is essentially based on the instantaneous model as are many operational systems in the United States (Oz Engineering and Motive Maps 2004).

Time Slice Model

The time slice method attempts to account for the variation in speeds over time by constructing the vehicle trajectory using downstream speed values which correspond to the time period when the vehicle would be traversing each segment. Thus for the first link, the travel time is estimated as for the instantaneous model [i.e., Eq. (1)].

The vehicle will then arrive at Link 2 at time $k+t(1,k)$ (i.e., the entry time plus the time to traverse the link), so the speed for that link is taken at that time

$$t(2,t_2) = \frac{2l_2}{v[2_a,k+t(1,k)] + v[2_b,k+t(1,k)]} \tag{3}$$

In general terms, the model can be written

$$t(n,t_n) = \frac{2l_n}{v(n_a,t_n) + v(n_b,t_n)} \tag{4}$$

where $t_n = k + t(1,k) + \sum_{i=2}^{n-1} t(i,t_i)$. As in the instantaneous model, the total travel time is then calculated by summing the corresponding link travel times.

It would be expected that this method would be able to produce an estimate superior to the instantaneous model, as evidenced by its capability to consider traffic evolution along the temporal axis. If speed is invariant over time, then the time slice method collapses to the instantaneous method.

To produce travel time estimates, the time slice method requires more extensive data, covering more time periods, than the instantaneous model and is suited to off-line rather than on-line applications. In an on-line context, the model could provide an estimate of the travel time of a vehicle departing at time k by the reconstructed travel time for a vehicle which arrived at the end of the motorway at time k. That estimate would be affected by changes over time in the speeds particularly when the speeds used for early links in the route would be quite "old" (i.e., come from much earlier time periods).

Dynamic Time Slice Model

Cortes et al. (2002) focused on the instantaneous calculation of $t(i,t_i)$ at the end points of each link to enhance the time slice method at the section level by using a recursive formulation. Their model acknowledges the finite length of each section and distinguishes the speed at the upstream node (a) when the vehicle enters the section $[v(i_a,t_i)]$ from the speed at the downstream section at the time the vehicle exits the section $\{v[i_b,t_i+t(i,t_i)]\}$. The link travel time is then calculated as

$$t(i,t_i) = \frac{2l_i}{v(i_a,t_i) + v[i_b,t_i + t(i,t_i)]} \tag{5}$$

where $t(i,t_i)$=travel time along section i for a vehicle entering that link at time t_i. Unlike the time slice method, the speed used by the dynamic time slice model for the exit of the link depends on the time to traverse the link. In this way, it is a natural extension of the preceding two methods.

Cortes et al. (2002) use an iterative algorithm to obtain the value of $t(i,t_i)$. The algorithm requires the input of initial value of

Table 1. Model Comparison

Models	Input speed data at link level: Upstream boundary	Downstream boundary	Link speed	Real time application
Instantaneous	At time k^a	At time k	Average of upstream and downstream	Yes
Time slice	At time vehicle enters the link	At time vehicle enters the link	Average of upstream and downstream	Yes but with delay
Dynamic time slice	At time vehicle enters the link	At time vehicle leaves the link	Average of upstream and downstream	Yes but with delay
Linear	At time vehicle enters the link	At time vehicle leaves the link	Linear function of position along the link	Yes but with delay

[a] k=time of entry into the first link in the motorway.

$t(i,t_i)$ and threshold precision before calculation. Since the section level travel time is calculated by iteration, the number of iterations determines the calculation speed and accuracy.

The dynamic time slice model follows a similar approach to the time slice method to compute network level travel time—that is adding up the corresponding link travel times. The dynamic time slice model has similar data requirements to the time slice model and has similar shortcomings for on-line applications.

When the link length is larger and spot speeds are aggregated over shorter time intervals, it would be expected that the dynamic time slice model would give improved estimates of travel time over the preceding models. However, this may not be the case in the field where for loop detector stations are commonly spaced approximately 500 m apart and speed data are aggregated every 20 s.

Linear Model

Each of the three preceding methods assumes the (average) section speeds are constant along each link when calculating travel times at the section level. As a consequence, a discontinuity of speed occurs as a vehicle leaves one link and enters the downstream link. Van Lint and Van der Zijpp (2003) proposed a linear method in which vehicles change their speeds gradually within the link. The speed of the vehicle at time t in section i; is not only a function of the speeds collected from the upstream and downstream detectors, as in the preceding methods, but also a function of distance of the vehicle to the end points of the section

$$v(i,t) = v(i_a,t) + \frac{x(i,t) - X(i_a)}{l_i}[v(i_b,t) - v(i_a,t)] \qquad (6)$$

where $x(i,t)$=location of vehicle at time t in section i; and $X(i_a)$=location of the upstream detector for section i. Van Lint and Van der Zijpp (2003) highlighted that the above equation is a typical first-order ordinary differential equation, which when combined with the boundary condition $x(i,t)$, can be written as an exponential function of t.

In this model, speed points are interpolated within the section (Van Lint and Van der Zijpp 2003). If the vehicle has not been able to leave the current section within the time interval in which it enters (for example, if the spot speeds are averaged every 20 s), the distance it can travel within the current time interval is calculated. The speed used then is approximated by the data collected from detectors during the next 20-s time interval. Correspondingly, the traveled distance in the next 20-s period is calculated and that process repeats itself until the vehicle exits the section.

Compared to the aforementioned methods, the main advantage of linear model is that the sudden change of speeds is avoided. It therefore provides an enhanced approximation to the speed profile likely to be experienced in the field.

Summary of Models

Table 1 summarizes the key features of the different models. Each of the models essentially calculates travel times for each link based on an (average) speed for that link. Model sophistication is therefore reflected in how the (average) link speed is calculated. Only the instantaneous model is suited to on-line application. Fig. 1 provides a simplified visual representation of how each model calculates the link level speed for a vehicle which starts its journey at the beginning of the road section of interest at time k.

The models should produce different travel time estimates if speeds vary over time. At times where speeds are relatively constant, e.g., light traffic flowing at free speed or congested traffic flowing at constant speed, the models should give similar predictions—collapsing in the ultimate case to the instantaneous model. For situations where the detectors are spaced further apart, resulting in longer sections, it would be expected that the dynamic time slice and linear models would yield superior estimates because those models try to account for conditions at the time when a vehicle would exit each link.

Model Validation Reported in Literature

Validation results reported in the literature are based on either simulations or field studies. Simulation provides an ideal test bed because it offers a controllable experimental environment where conditions can be varied. However, there is a need for reassurance that the simulation model reflects real world conditions. In contrast, the field study setting is inherently a reflection of real world conditions but it is usually associated with more data noise and is

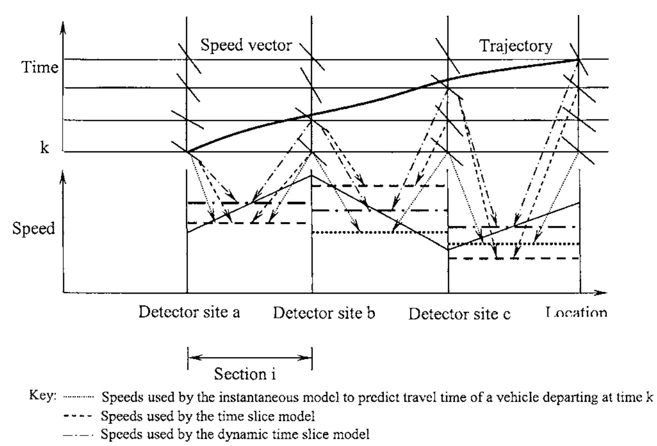

Fig. 1. Time space comparisons of four speed-based travel time estimation models

restricted in the sense that the study can consider only the conditions at that site and in so doing may produce conclusions which are only valid at that site.

The instantaneous model has been subjected to the most extensive testing (Haj 1998; Lindveld et al. 2000; Paterson 2000). From that testing it is clear that while it is an easy solution to the travel time prediction problem its results are poor unless traffic conditions are uncongested (Oz Engineering and MotionMaps 2004). Bovy and Thijs (2000) used a microscopic traffic simulator [freeway operations simulation (FOSIM)] to test the factors that impact on the accuracy of the instantaneous and time slice models. They concluded that the performance of these two methods varies with congestion level, percentage of slow traffic (specifically trucks), the length of road section, and the time interval over which the speed data are aggregated. The effect of section length and time interval length were found to be slight compared with that of congestion level and the proportion of trucks.

Cortes et al. (2002) validated their proposed dynamic time slice method using Paramics. Evaluations were conducted on the basis of two scenarios: moderate congestion and severe congestion. Travel times from the dynamic time slice model were compared to benchmark travel times estimated from a 100% probe fleet. The speeds used were spot speeds collected in continuous time and 30-s aggregated speeds. The percentage error in the dynamic time slice estimates were around 5% when 30-s aggregated speeds were used while those errors were reduced by about one third by using continuous spot speeds as input. The performance of the dynamic time slice model using individual speeds was comparable to that of a 5% probe rate. However, individual speeds are not available in practice. Model performance deteriorated with increasing levels of congestion, as could be expected, with the mean absolute percentage error (MAPE) under normal and congested traffic being 4.6 and 7%, respectively.

Van Lint and Van der Zijpp (2003) validated their proposed linear method, and compared the results to those obtained with the time slice model, using data generated from a microscopic traffic simulation model. Their results indicated that the time slice method significantly overestimated the actual travel time by 6%, whereas there was less than 1% error associated with the linear model. The root-mean-squared errors were 60.6 s for the time slice model compared to 32.3 s for the linear method. Therefore, they concluded that the linear method dramatically improved the accuracy of speed-based travel time estimation model. Based on these results, the linear method was then used to reconstruct travel times from speed data. Those estimated travel times were then used in calibrating travel time prediction models and for analysis of the distribution of travel time (Van Lint 2004; Van Lint et al. 2004).

Limited field test based evaluations are reported in the literature. The DACCORD project (Haj 1998; Lindveld et al. 2000) reported the accuracy of the instantaneous model and the time slice model, respectively, using travel time data collected from three European Test sites—Amsterdam, The Netherlands, Paris, and near Padua, Italy. Travel time data were collected via number plate surveys (Amsterdam), floating cars (Paris), and time stamped toll tickets (Padua). A range of factors resulted in limited insight from the Padua test site. There were serious reservations about the quality of the Italian travel time data due to toll barrier delays and clock synchronization problems. In addition the loop spacing on the Italian site was between 2 and 4 km—much greater than is common on motorways. The errors for the Paris test declined with increasing congestion level (from about 60 to 20%). While this result is consistent with that found using simulation (Bovy et al. 2000) it is still counterintuitive since lower errors would be expected under free flow conditions. In this respect, results from the Amsterdam site were consistent with expectations with the errors increasing with increasing levels of congestion. However the magnitude of the errors was considerable, increasing from about 12 to 65% under high congestion. The rapid decline in performance of the models on the Amsterdam site was attributed in part to a difficulty with the loops which appeared unable to record speeds below 18 km/h. While the DACCORD project provides an indication of the level of errors which can be expected in practice, it also highlights the instrumentation challenges presented by field studies.

Bajwa et al. (2003) used the time slice method to provide benchmark travel times for validating the performance of a nonparametric travel time prediction model they proposed. However, no results are reported to evaluate the accuracy of this benchmark. Zhang and Rice (2003) also used travel times estimated from a modified time slice model (which used only upstream speed as an input) as a key input for a proposed linear model for prediction of travel time on freeways.

A number of key issues emerge from the validation studies which are reported in the literature:

1. Only limited validation results are reported in the literature for the other models and some models (e.g., the linear model) only have validation results reported on a simulation test bed rather than in a field setting;
2. There is evidence of inconsistent or counterintuitive results whereby the level of error declines with increasing levels of congestion; and
3. While acknowledging the small number of reported validations, it would appear that the levels of error reported in simulation validations are in some cases up to one order of magnitude lower than results obtained in the field.

These issues point to the need for further field validation work. In the following sections, we compare the performance of the four speed-based travel time estimation models in a rigorous field study.

Field Study Approach

The field study focused on two operational motorways in Melbourne, Australia, where measured travel times were available. The errors in the travel times estimated by each of the four models described in the previous sections were quantified in relation to the measured travel times on each motorway. While it would have been desirable to test the models under a wider range of field conditions, data availability was a constraint. The data that were available do, however, provide a much larger sample of measured travel times than other field studies reported in the literature. This section begins by describing the two field test sites. The quantitative error measures used in the study are then identified before the travel time estimation errors for the four models are presented.

Field Data

The first site is the South Eastern Freeway, a radial freeway which provides a commuting route into the central business district (CBD) for residents of the growing south eastern suburbs of the metropolitan area. The data for this facility were collected in 1998 as part of a Ph.D. research project (Paterson 2000). At that time there was no freeway connection at the CBD end of the facility, however, that connection is today provided by the CityLink Toll-

Table 2. Key Features of Field Test Sites

Description	South Eastern Freeway	CityLink Tollway
Data collected	September 9, 1998	October 31, 2003
Section length	6.5 km	14 km
Speed data source	Double inductive loops at an average spacing of 500 m	(Image processing from incident detection cameras to produce speeds with cameras 200 m apart in a 3.2 km tunnel and about 1.5 km apart elsewhere)+(dual inductive loops over a 3 km section, not covered by the cameras, at an average spacing of 600 m)
Speed data collection interval	20 s	20 s
Travel time data source	Number plate survey conducted 6:30 a.m. to 10:45 a.m.	Calculated from the AVI data obtained from the tolling system which provides the time when the transponder passed under a tolling gantry
Travel time sample size	126 number plate matches	7,600 toll tag matches

way. That tollway, which opened in 2000 as one of the first in the world to operate fully electronic tolling, provided the second site for the field test. CityLink extends from the end of the former South Eastern Freeway, around the outskirts of the city (via a tunnel and above ground sections) and then provides a radial link to the north of the CBD.

Table 2 summarizes key features of the two test sites. Speed data for the South Eastern Freeway site were provided by inductive loops while a combination of cameras and inductive loops provided speed data for CityLink. A traditional number plate survey provided travel time data for the South Eastern Freeway site. In contrast, the tolling system on the CityLink facility, which is based on automatic vehicle identification (AVI) technology, provided much richer travel time data because of the ability to compare the times at which the tags are observed at different toll gantries. It is appropriate to note that the travel time data for the South Eastern Freeway cover a single morning peak (approximately 4 h of travel time data) while the AVI system on CityLink provides data throughout the day.

In both cases, the speed data to be input into the travel time estimation models was provided at 20 s intervals. The high variability in the raw speed data over the 20 s collection interval translates into fluctuations in the estimated travel times unless smoothing of the speed data is undertaken. In this research, the speed data were smoothed using wavelet techniques (Burrus et al. 1998) because of that procedure's capability to reduce data noise.

Performance of Speed-Based Travel Time Estimation Models

A range of conventional measures is used to evaluate the accuracy of the estimation models. By defining the error as $e_i = T_i - \hat{T}_i$, where T_i is the actual travel time (measured from either the number plate survey or the AVI system), and $\hat{T}_i$ is the estimated travel time (from the instantaneous, time slice, dynamic time slice, or linear models), three summary measures are then used

$$\text{Mean absolute error (MAE)} = \frac{\sum_{i=1}^{n} |e_i|}{n} \tag{7}$$

$$\text{Root mean square error (RMSE)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \tag{8}$$

$$\text{Mean absolute relative error (MARE)} = \frac{1}{n}\sum_{i=1}^{n} \frac{|e_i|}{T_i} \tag{9}$$

Comparative Performance

Table 3 summarizes the performance of speed-based travel time estimation models on the South Eastern Freeway. There is less than 1% difference in the results across the four models (MARE ranges between 11.7 and 12.3%) while average errors are on the order of 2 min. Validation based on the CityLink data set (Table 4) also shows similar results. In this case, the average errors are on the order of 1 min, MARE values are of a similar order of magnitude at about 9%, and the variation in performance is less than 1% across the four models. The results indicate that the four models produce estimates with similar levels of error, although the errors for the South Eastern Freeway are higher reflecting the peak period nature of that data. The dynamic time slice model does not exhibit superiority over the time slice model in the two field cases considered here. The performance of the linear model is the best across the two data sets used for validation, although it produces only marginally lower errors than the other models.

Fig. 2 shows the visual comparison of the model's predictions with the actual travel times aggregated to 5 min average values

Table 3. Measured Travel Time Estimation Error on South Eastern Freeway

Error measure	Instantaneous model	Time slice model	Dynamic time slice model	Linear model
MAE (s)	115	112	111	109
RMSE (s)	139	146	143	141
MARE (%)	12.3	11.9	11.9	11.7

Table 4. Measured Travel Time Estimation Error on CityLink

Error measure	Instantaneous model	Time slice model	Dynamic time slice model	Linear model
MAE (s)	71	68	68	66
RMSE (s)	106	103	104	100
MARE (%)	8.9	8.5	8.5	8.3

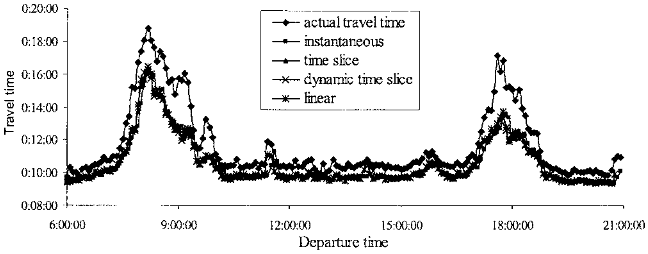

Fig. 2. Visual comparisons of speed-based methods with actual travel times (5 min aggregation using CityLink data set)

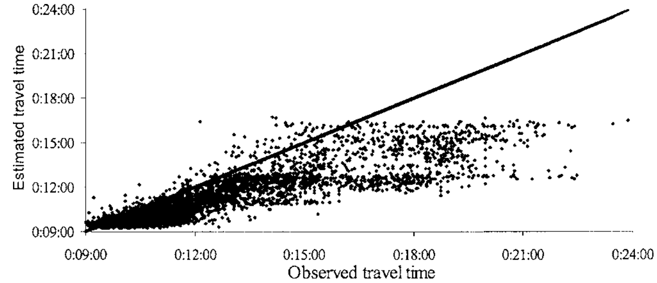

Fig. 3. Relationship between observed and estimated travel times for time slice model (CityLink data set)

for the CityLink data set. Data aggregation to 5 min periods reduced the MARE values by about 0.2–0.5% compared to the values shown in Table 4. The overall tendency is for the models to underestimate and slightly lag the actual travel times. It is interesting to note that with the exception of Van Lint and Van der Zijpp (2003), who found evidence of overestimation, most studies reported in the literature have found speed-based travel time estimation models to underestimate the actual travel times (Lindveld and Thijs 1999; Paterson 2000; Smith et al. 2004). Lindveld and Thijs (1999) made an attempt to improve the time slice model by using an approximate conversion of the speed data collected from loop detectors into space-mean speed values. Space-mean speed is lower than time mean speed where there is a variation in vehicle speed (May 1990). We consider the impact of using space mean speed in the following section.

Fig. 3 shows a plot of the estimated versus observed travel times for the time slice model using the CityLink data set. The pattern in these data is typical of the results obtained for all of the models considered here. This figure highlights the general underestimation of travel times and the higher level of underestimation at higher travel times (i.e., congested conditions). All of the models predicted a similar maximum travel time of about 16 min while measured travel times were much higher during the peak period.

Model accuracy was also quantitatively assessed under different flow conditions on the CityLink data set. As shown by Fig. 2, morning peak congestion tends to be longer and more severe than the afternoon peak. The performance of the models under different flow conditions is shown in Table 5. It is obvious that all four methods perform better under free flow conditions than under congestion. MARE values of around 7, 15, and 12% were found under free flow, morning peak and afternoon peak conditions, respectively. This is consistent with Bovy et al.'s (2000) simulation results which highlighted that congestion is the most important factor determining the error in the travel time estimates. Importantly, congestion does not appear to impact on the comparative performance of four models validated here. Under any of the flow conditions there is little to separate the models on the basis of their performance although the linear model does produce the lowest errors.

Refinement in Representation of Mean Speed

In this section two model refinements are considered in an effort to reduce the model prediction error. The first refinement involves reviewing the selection of the detector location used to provide an indication of the speed along each section. The second refinement attempts to overcome the inevitable underestimation in travel times arising from the use of time mean speed values by calculating an approximation to the space mean speed.

In the original model formulations, the average speed from two adjacent detectors is assumed to represent the traffic situation over the link of interest [i.e., Eq. (1)]. An alternative approach is to use the spot speed collected from either the upstream or downstream sensors to represent the whole link's average speed so that the section travel time is calculated as either

$$t(i) = \frac{l_i}{\nu(i_a)} \quad \text{or } t(i) = \frac{l_i}{\nu(i_b)} \tag{10}$$

where l_i=link length; and $\nu(i_a)$ and $\nu(i_b)$=speeds measured from upstream detector and downstream detector, respectively. Zhang and Rice (2003) used the speed from the upstream sensor to reconstruct historical travel times from point speed data. This approach is only applicable to the instantaneous and the time slice models.

Tables 6 and 7 show the model errors for the instantaneous and the time slice models using different representations of the speed along the whole section. Using speeds from either the upstream or downstream sensors offers the same or better accuracy compared with the average of the two. It is interesting to note that the upstream and downstream approaches perform differently on the two validation data sets. For example, travel times estimated from upstream speed provide the best accuracy on the South Eastern Freeway while the downstream speed approach performs best on CityLink. This may be due to the different positions of bottlenecks in two networks. A severe bottleneck occurs around the middle links of the CityLink network, whereas congestion is more prevalent in the beginning and ending sections of the South Eastern Freeway. This result is also consistent with a finding from Smith et al.'s (2004) simulation study that detector location

Table 5. Measured Travel Time Estimation Errors under Different Traffic Flow Conditions (CityLink Tollway)

	Instantaneous		Time slice		Dynamic time slice		Linear	
Error	Off peak	Morning/afternoon peak	Off peak	Morning/afternoon peak	Off peak	Morning/afternoon peak	Off peak	Morning/afternoon peak
MAE (s)	44	154/112	42	149/110	42	149/111	42	141/106
RMSE (s)	57	193/153	53	186/151	53	187/152	53	179/148
MARE (%)	6.8	15.1/12.5	6.4	14.5/12	6.4	14.6/12.1	6.3	13.8/11.8

Table 6. Measured Travel Time Estimation Errors Using Different Representative Speeds (South Eastern Freeway)

	Instantaneous			Time slice		
Error	Average	Upstream	Downstream	Average	Upstream	Downstream
MAE (s)	115	101	106	112	99	109
RMSE (s)	139	125	129	146	126	145
MARE (%)	12.3	11	11.7	11.9	10.8	11.7

Table 7. Measured Travel Time Estimation Errors Using Different Representative Speeds (CityLink)

	Instantaneous			Time slice		
Error	Average	Upstream	Downstream	Average	Upstream	Downstream
MAE (s)	71	70	60	68	67	56
RMSE (s)	106	102	90	103	98	85
MARE (%)	8.9	8.8	7.6	8.5	8.4	7.1

Table 8. Measured Travel Time Estimation Errors Using Approximate Harmonic Mean Speeds under Different Traffic Flow Conditions (CityLink)

	Instantaneous		Time slice		Dynamic time slice		Linear	
Error	Off peak	Morning/afternoon peak	Off peak	Morning/afternoon peak	Off peak	Morning/afternoon peak	Off peak	Morning/afternoon peak
MAE (s)	41	131/99	38	122/93	38	125/94	38	117/88
RMSE (s)	54	169/138	50	158/134	50	160/136	49	153/129
MARE (%)	6.3	12.9/11.1	5.9	12.1/10.3	5.9	12.2/10.4	5.8	11.6/9.8

played a key role under congested conditions. It would appear that the selection of the "representative" speed for a section should be based on the application scenario. Under a situation where actual travel times are not available, the earlier results would suggest that the approach that provides longer travel times should be adopted since that will reduce errors.

The underestimation of travel times can at least partially be attributed to use of the time mean speed values which are produced by the point-based sensors. It would be expected that using the harmonic mean of individual vehicle speeds (i.e., the space mean speed) would improve the performance of the travel time estimation models since the space mean speed is lower than the time mean speed and the models underestimate travel times. However, while data were available for each lane, individual vehicle speeds were not available from the field equipment at either test site. Therefore an approximation to the harmonic mean speeds is used in an effort to reduce the error of speed-based models. We assume that all the vehicles in each lane in a 20 s period travel at the same speed. An approximate harmonic mean speed ($v_{\text{harm.}}$) is then calculated across lanes as follows:

$$\frac{1}{v_{\text{harm.}}} = \frac{\sum_{i=1}^{n} \frac{N_i}{v_i}}{\sum_{i=1}^{n} N_i} \tag{11}$$

where n=number of lanes; N_i=number of vehicles recorded in lane i during the 20 s period; and v_i=speed recorded for lane i.

Table 8 shows the results obtained under different traffic conditions when the approximate harmonic mean speeds (an estimate of space mean speed) are used as input. Compared with results obtained with the conventional average spot (time mean) speeds (Table 5), the approximated space mean speed has resulted in a reduction in the relative errors of up to 2%. While this improvement is only modest, it can still be achieved even when the raw data do not allow more precise conversion to space mean speed.

Conclusions and Research Directions

This paper has focused on the evaluation of four speed-based travel time estimation models: the instantaneous model, the time slice model, the dynamic time slice model, and the linear model. The literature provides conflicting insight into the performance of these models with counterintuitive cases reported where error levels under congestion were less than under free flow conditions. There was also evidence that the error levels estimated when the models are run on simulation test beds are in some cases one order of magnitude lower than the error levels which have been found in field studies.

Data sets from two operational motorways (South Eastern Freeway and CityLink) in Melbourne, were used to undertake a field validation of the four travel time estimation models. The AVI technology deployed on CityLink for electronic tolling provided a rich set of measure travel times while a traditional timed number plate survey provided the ground truth data for the South Eastern Freeway case.

Key findings from the field validations were as follows:

1. There was little difference in the relative performance of the four models across the two data sets although the linear model was found to perform marginally better than the others. The similarities in the results may be due to the limi-

tation of using the speeds collected from two points to represent the traffic condition along each section. That form of input is inherent in the all speed-based travel time estimation models considered here.
2. All models were found to underestimate and slightly lag the actual travel times. The instantaneous model has the largest lag among the four speed-based models. The underestimation is to be expected when time mean speeds are used as input.
3. The results were sensitive to the level of congestion, and consistent with expectations, errors in the off-peak were about half those observed in the peak. During the peak, average errors were on the order of 2 and 1/2 min and percentage errors were around 12–15%. Underestimates of 4–6 min were observed in the peak periods.
4. Using data from either the upstream or downstream detectors, rather than the average of those values, was found to reduce the errors in some cases. This does suggest that errors may be able to be minimized through local "tuning" of the model inputs to produce longer travel times.
5. Converting spot speeds across lanes to harmonic mean speeds, as an approximation to space mean speed, produced a marginal reduction in error of up to 2%. While the improvement was only modest, it can still be achieved even when the raw data do not allow more precise conversion to space mean speed. Road authorities employing the instantaneous model for travel time estimation should implement the revised calculation of average speed to reduce the travel time prediction error.

These results have important implications for practitioners and model developers. Practitioners need to be aware of the level of error particularly where travel time information is communicated to the public on the basis of models such as the instantaneous model or where these models are used to monitor network performance and may therefore underestimate actual levels of congestion experienced by road users. During peak period conditions, road users are likely to place greatest value on real time information, and the results reported here suggest there is the potential for sizeable underestimation of travel times. Model developers also need to be aware of these results particularly where they use travel time estimation models to provide "quasi-ground-truth" values for calibrating predictive models. Prediction errors will be higher in the field because of the errors identified here in that quasi-ground-truth data.

As noted above, there was a uniform underestimation of travel times with each of the models considered here even when an approximation for space mean speed was used as input. It is expected that the accuracy could be further improved by basing the calculation on a more accurate estimation of space mean speed. Further exploration of the approach proposed by Lindveld and Thijs (1999) for converting time mean speed to space mean speed would therefore appear to have merit as a part of future research efforts.

Acknowledgments

The writers gratefully acknowledge Transurban and VicRoads for supplying the data used in this research.

References

Bajwa, S. U. I., Chung, E., and Kuwahara, M. (2003). "A travel time prediction method based on pattern matching technique." *Proc., 21st ARRB Transport Research Conf.* (CD-ROM), ARRB, Cairns, Queensland, Australia.

Bovy, P. H. L., and Thijs, R. (2000). "Estimators of travel time for road networks." ⟨http://www.trail.tudelft.nl/verkeerskunde/⟩ (July 20, 2003).

Burrus, S. C., Gopinath, R. A., and Guo, H. T. (1998). *Introduction to wavelets and wavelet transform: A primer*, Prentice-Hall, Upper Saddle River, NJ.

Cortes, C. E., Lavanya, R. O., Jun-Seok, and Jayakrishnan, R. (2002). "General-purpose methodology for estimating link travel time with multiple-point detection of traffic." *Transportation Research Record. 1802*, Transportation Research Board, Washington, D.C., 181–189.

Haj, S. H. (1998). "Appendix F: France test site: Travel time estimation and forecasting: Evaluation results." *Daccord Deliverable 10.1*, Hague Consulting Group, Hague, The Netherlands.

Kloot, G. (1999). "Melbourne's arterial travel time system." *Proc., 6th World Congress on ITS* (CD-ROM), ITSA, Toronto.

Lindveld, C. D. R., and Thijs, R. (1999). "On-line travel time estimation using inductive loop data: the effect of instrumentation peculiarities." *Proc., 6th World Congress on ITS* (CD-ROM), ITSA, Toronto.

Lindveld, C. D. R., Thijs, R., Bovy, P. H. L., and Van der Zijpp, N. J. (2000). "Evaluation of online travel time estimators and predictors." *Transportation Research Record. 1719*, Transportation Research Board, Washington, D.C., 45–53.

May, A. D. (1990). *Traffic flow fundamentals*, Prentice-Hall, Englewood Cliffs, N.J.

OZ Engineering and MotionMaps. (2004). "Travel time estimates, displays and forecasts–Final report." *AZTECh Phase III Program for Advanced Traveller Information Services for the Deployment of Service D: Technical Rep. No. 2*, Maricopa County Dept. of Transportation, Ariz.

Paterson, D. W. (2000). "The real time prediction of freeway travel times." Ph.D. thesis, Monash Univ., Melbourne, Australia.

Smith, B. L., Holt, R. B., and Park, B. (2004). "Travel time estimation for urban freeway performance measurement: understanding and improving upon the extrapolation method." *Proc., 83rd Transportation Research Board Annual Meeting* (CD-ROM), Transportation Research Board, Washington D.C.

Turner, S. M., Eisele, S. M., Benz, R. J., and Holdener, D. J. (1998). "Travel time data collection handbook." *Rep. No. FHWA-PL-98-035*, Texas Transportation Institute, Texas A&M Univ., College Station, Tex.

Van Lint, J. W. C. (2004). "Quantifying uncertainty in real-time neural network based freeway travel prediction." *Proc., 83rd Transportation Research Board Annual Meeting* (CD-ROM), Transportation Research Board, Washington D.C.

Van Lint, J. W. C., and Van der Zijpp, N. J. (2003). "An improved travel-time estimation algorithm using dual loop detectors." *Proc., 82nd Transportation Research Board Annual Meeting* (CD-ROM), Transportation Research Board, Washington D.C.

Van Lint, J. W. C., Tu, H., and Van Zuylen, H. J. (2004). "Travel time reliability on freeways." *Proc., 10th World Conf. on Transport Research (WCTR)* (CD-ROM), WCTRS, Istanbul, Turkey.

Zhang, X., and Rice, J. A. (2003). "Short-term travel time prediction." *Transp. Res., Part C: Emerg. Technol.*, 11C (3-4), 187–210.