



CAITR 2004

Examining travel time
variability using AVI
data

by

Ruimin li
ruimin.li@eng.monash.edu.au

8 - 10 December 2004

INSTITUTE OF TRANSPORT STUDIES
The Australian Key Centre in Transport Management
Department of Civil Engineering
Building 60
Monash University Vic 3800

Established under the Australian Research Council's Key Centre Program

Abstract

Knowledge of travel time variability is valuable for improving the reliability of traffic information services and increasing the accuracy of travel time predictions. To identify the source of travel time variability, information on the travel time distribution properties is needed. Most investigations of travel time distributions rely on data from probe vehicles and consequently have limited sample size. Based on extensive Automatic vehicle identification (AVI) data collected from the CityLink Tollway in Melbourne, a comprehensive investigation of travel time distributions was conducted in terms of various time windows. Given the number of factors affecting travel time variability and their interaction effects, multiple regression with two-way interaction terms was used to quantify the contribution of the various sources to the variability in travel time. The application of the methodology to two groups of data, namely travel times in morning peak and afternoon peak, demonstrates that they have distinctive sources of variability. Morning peak travel times vary mostly because of demand related factors, while 25% variability of travel times in afternoon peak is related to capacity related factors.

Key words: travel time, distribution, variability, AVI

1. Introduction

A reliable traffic system indicates that travellers can anticipate their travel times accurately before their trip, based on the experience gained from the past trips. In that sense, the measurement of reliability is how stable rather than severe the congestion is from day to day. Therefore, travel time variability can be used for evaluating the reliability of transportation systems. A high degree of variability indicates that the travel time would be unpredictable and the traffic service is less reliable (Turochy and Smith, 2002). From the traveller's perspective, a decrease in travel time variability reduces the uncertainty in decision-making about departure time and route choice as well as the anxiety and stress caused by such uncertainty (Sun, Arr et al., 2003).

Previous qualitative studies found that users value the reliability of a transport system more highly than other features (Lam and Small, 2001, Bates et al., 2001). Bates, Polka et al. (2001) pointed out that a reduction in variability is as valuable as the reduction of mean travel time or even more valuable in some situations. They further proposed that the median and the distribution of travel times are better measures than the mean travel time. However, the extent of research into the travel time variability is rather small compared to the efforts that have been done in analysing mean travel time and measuring behavioural reactions to the change of travel time variability. This is due, at least partly, to the fact that there is limited knowledge regarding the distribution of travel time because of the lack of a large enough sample of real travel time data.

The objective of the research reported in this paper is to analyse travel time distribution properties from various time windows and identify the contribution of different factors to travel time. This paper first introduces the Automatic Vehicle Identification (AVI) data used for travel time measurement in this study and that is followed by an exploration of the distribution of travel time. The multiple regression methodology used to identify the sources of travel time variability, along with their interaction effects, is then described. The conclusion and implication of the study are summarized in the final section.

2. AVI for travel time measurement

The AVI data used in this study were collected from the CityLink tollway in Melbourne Australia. In order to protect privacy of individual drivers, each record had been encrypted before the data was supplied to the research team. Most vehicles on the tollway are equipped with an E-tag, which communicates with overhead gantries to pay the tolls electronically. Video cameras are mounted on the overhead gantries as well so that automatic licence plate recognition can be used to identify vehicles that are not fitted with an E-tag. A video recording of each number plate at one gantry is matched against those at other gantries to provide information such as location and time. Thus, all vehicles, no matter whether fitted with an E-tag or not, can be identified by one of those two systems which ensures a large sample size of data for this study. The raw data in its original form comprises observation recorded at each tolling point gantry of a vehicle identifier (number plate or E-tag ID), observation time and vehicle type indicator (car or commercial vehicle). By matching the encrypted vehicle ID's across toll points, the travel times could be calculated by subtracting the observation times.

The AVI data used in this project was for the entire month of October 2003 and covered a 14 kilometres long section of the tollway, including a 3.2- kilometre long tunnel. Four gantries were positioned along this section. In addition to a route travel time between gantry one and gantry four, each pair of gantries provided a basis for separately measuring three link specific travel times. In addition, the information on vehicle type and the proportion of trucks in the traffic stream makes it possible to analyse the variability between cars and commercial vehicles. An effective method was used to detect travel time outliers in this study and the detailed introduction is in Li et al. (2004). Roadside accident detection cameras along the tollway recorded the information of accidents. The observed weather condition, primarily comprises of rain information, was obtained through the Bureau of Meteorology, Victoria. Together three data sources provide a sound basis for conducting travel time variability analysis.

3. Travel time distributions

Travel time distribution properties provide a valuable basis for travel time prediction and estimation models based on statistical techniques, as well as traffic and driver behaviour simulation. The study of travel time distributions has attracted research interest for many years. The early research mainly suggested two kinds of travel time distributions; namely, the normal distribution (Smeed and Jeffcoate, 1971) and the lognormal distribution (Richardson and Taylor, 1978). However, due to the limited capacity to collect travel time data at the time of these earlier studies, the data used to derive the travel time distribution was collected over different routes. The inclusion of that cross route variability, which was not separately measurable, complicates the interpretation of the final distributions.

Recently, Kwon et al. (2000) analysed the daily pattern of travel times observed from probe vehicles. Their analysis revealed that the distribution of travel time was skewed to the right. One drawback in their research approach is that the sample size of travel time collected from probe vehicles was so limited that they had to interpolate the values of travel time between two consecutive runs. Work performed by Sun et al. (2003) investigated the distribution of travel time in a one and a half hour travel time window (8-9.30am) on one day by analysing 500 samples from two video detectors spaced 130m apart. The study found that the travel time distribution was not symmetrical, indicating that the mean and median value would not be same.

Figure 1 shows the distribution of car travel times over the 14-kilometre section of CityLink for all weekdays in October 2003. Observation between 6 am and 9 pm are included in this analysis. It is clear that this distribution is highly skewed with a flat and long right tail. Figure 2 shows the distribution of car travel times for different time periods. It can be seen that in a comparative sense the distribution of travel times under free flow conditions (off-peak) has the shortest right tail, whereas the travel times in the afternoon peak presents the most skewed distribution. This highlights that the afternoon peak travel times show the highest variability for this location. Figure 3 and Figure 4 show the distributions of travel time in a typical double peak workday and a typical single peak workday. These figures also compare the distributions of car and commercial vehicle travel times. Compared with the distribution of travel times over all weekdays, the distribution of travel times over a single day is less skewed but still not normally distributed. In addition, it is clear from these two figures that the mean travel time of commercial vehicles is higher than that of the cars. In practice,

the travel times of cars and commercial vehicles can reflect differences in their performance characteristics particularly when compounded by adverse geometric conditions such as sustained grades.

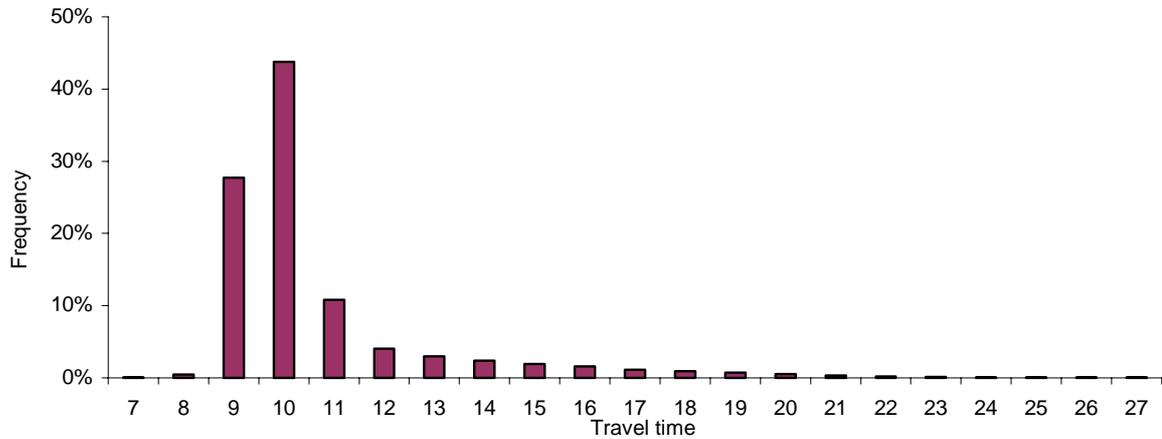
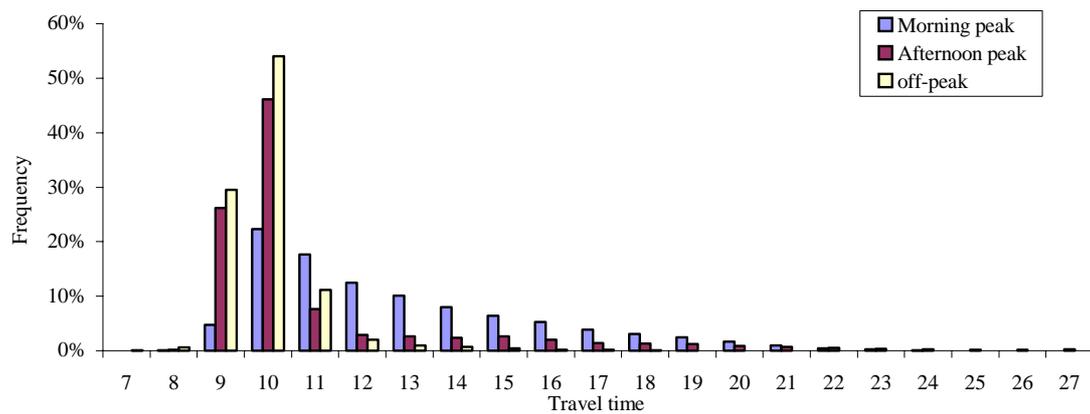


FIGURE 1 Distribution of car travel times: all weekdays over a one-month period



Key: Morning peak: 7:00-9:30
 Afternoon peak: 16:00-19:00
 Off-peak: 6:00 to 7:00, 9:30 to 16:00, and 19:00 to 21:00

FIGURE 2 Distributions of car travel times for different time windows flow: all weekdays over a one-month period

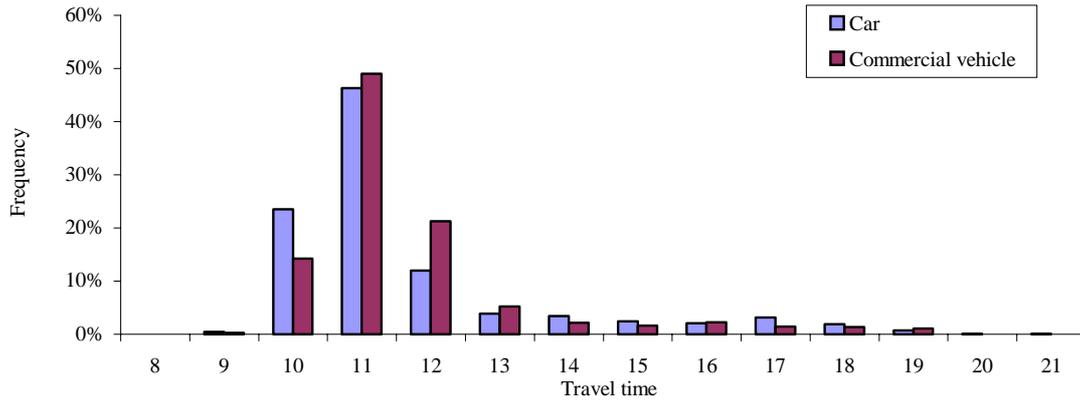


FIGURE 3 Distribution of travel times in a typical double peak workday (morning and afternoon peak)

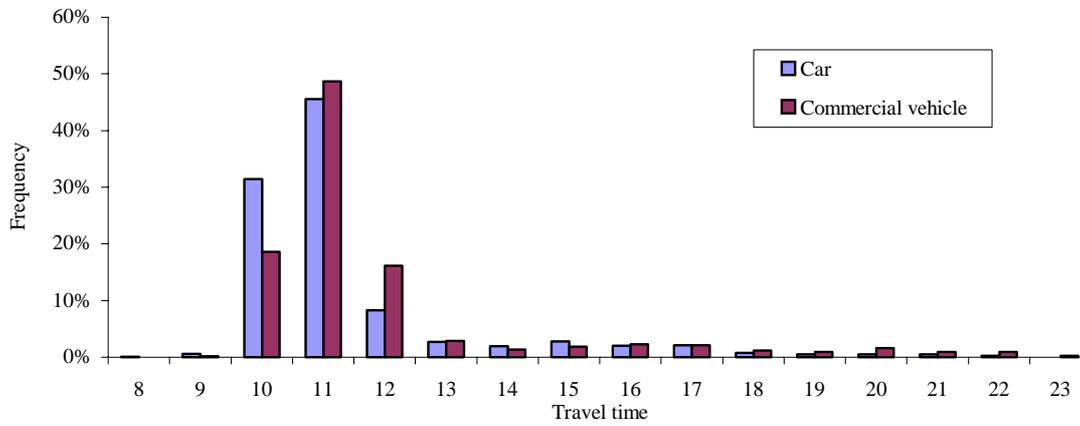


FIGURE 4 Distribution of travel times in a typical single peak workday (morning peak only)

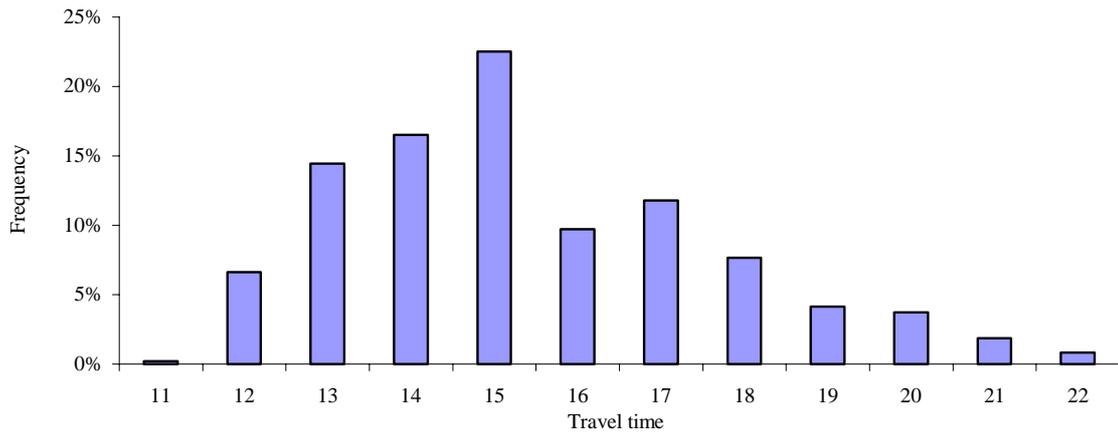


FIGURE 5 Distribution of car travel times in a one-hour time window (peak hour)

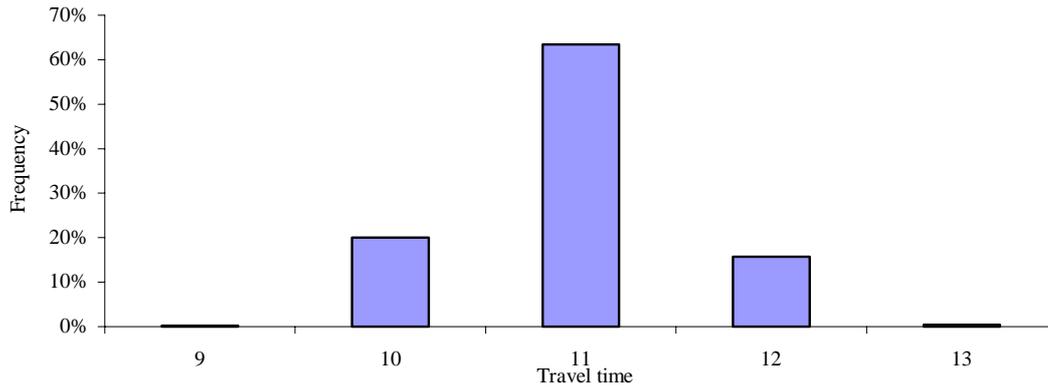


FIGURE 6 Distribution of car travel times in a one-hour time window (free flow)

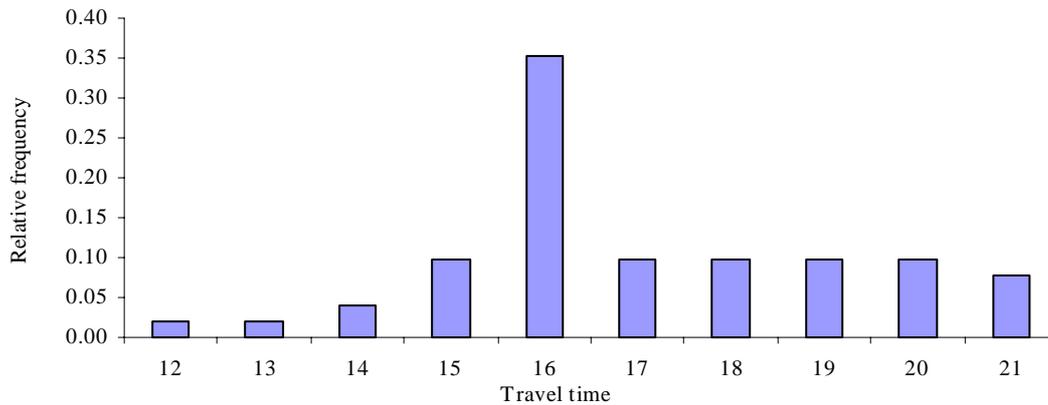


FIGURE 7 Distribution of car travel times in a 5-minute time window

Additional insight can be obtained by reducing the time window associated with the travel time distribution, to say, one hour. Figures 5 and 6 show the distributions of travel times in two one-hour time windows, one associated with congestion and the second reflecting free flow conditions. It can be seen that the travel times in a one-hour time window is very close to a normal distribution for both flow conditions. Figure 7 shows the distribution of travel times in a 5-minute time window. Again, it can be seen that this distribution can be approximately characterized by a normal distribution. Table 1 shows the descriptive statistic of travel time distribution in different time windows. The high value of skewness indicates that the distribution is asymmetrical. The kurtosis value is used to measure the weight of tails relative to the rest of a distribution. The kurtosis will increase as the tails of a distribution become fatter and will decrease as the tails become thinner. A normal distribution has zero skewness and kurtosis of 3. It can be seen from the table that the skewness value decreased as the time window is reduced. It indicates that the distribution become symmetrical.

These results suggest that the travel time distribution analysis in different scale time window indicates that the distribution tends towards a normal distribution as the time window decreases. Van lint (2004a) analysed the distribution of travel times in a small time window (1minute) and also found that it is approximately normal. This finding is relevant to applications in traffic simulation and travel time estimation and prediction.

Table 1 Travel time distribution in different time windows (in seconds)

Distribution	Mean (Sec.)	Median (Sec.)	Variance (Sec ²)	Skewness	Kurtosis
Car (double peak weekday)	655	628	14719	2.051	3.712
Commercial vehicle (double peak weekday)	657	641	10706	2.358	5.769
Car (single peak weekday)	639	616	14890	2.631	7.308
Commercial vehicle (single peak weekday)	664	634	23145	2.479	5.891
One hour (peak)	891	868	19514	0.805	0.239
One hour (off peak)	627	624	1134	0.273	0.584
5 minutes	993	946	17765	0.032	-0.217

4. Travel time variability

Many measures of travel time variability have been proposed in the literature. For example, Lomax et al (2003) used a buffer time index, defined as the difference between the 95th percentile travel rate (in minutes per mile) and the average travel rate (in minutes per mile) divided by average travel rate. Turochy and Smith (2002) developed a variability index, which measured the size of the confidence region based on multivariate statistical quality control. Van Lint (2004b) proposed two measures. First, the ratio of the distance between the 90th and 50th percentile to the distance between the 50th and 10th percentile, and second the ratio of the distance between 90th and 10th percentile to the median travel time. The first measurement implies the relative width of travel time distribution while the second indicates the skewness of the distribution. In this paper, traditional statistical measures are used, specifically the coefficient of variation and the 90th percentile travel time.

Noland and Polak (2002) characterized travel time variability as comprising three distinct components: variance of travel time from day-to-day, from time-of-day and from vehicle-to-vehicle. The vehicle-to-vehicle travel time difference explains the travel time variability in a small time window on a certain day of the week. This variability is introduced by driver behaviour difference, such as aggressiveness and lane choice decisions. Here we assume a sample of size n travel time observations which can be divided into k groups according to criteria such as day of week, or time of day and so on. Group i has k_i travel time points with the corresponding mean being \bar{t}_i . Then the mean over all observations is:

$$\bar{T} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{k_i} t_{ij} \quad (1)$$

Where t_{ij} indicates travel time observation j in group i .

The variance of all travel time observations can be written as:

$$V = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{k_i} (t_{ij} - \bar{t}_i)^2 + \frac{1}{n} \sum_{i=1}^k k_i (\bar{t}_i - T)^2 \quad (2)$$

The first item in the above equation is driver-to-driver or vehicle-to-vehicle travel time variability, whereas the second item is variation due to external factors.

We analyse travel times in 5 minute intervals and assume that the variance of travel times of vehicles departing within a 5 minute period in a certain day is mainly comprised of vehicle-to-vehicle variability. Chen et al. (2003) stated that this travel time variability is proportional to the mean travel time. Figure 8 shows the relationship between the coefficient of variation and mean travel time for two different days. It can be seen that vehicle-to-vehicle variability increases with mean travel time up to a threshold mean travel time and beyond that. It may decrease or at least remain a constant after that threshold. According to (Li et al, 2004), vehicle-to-vehicle travel time variability is primarily a function of lane choice. When volume approach capacity, lane changing is restricted and the vehicle-to-vehicle variability which is measured may be primarily a reflection of lane-to-lane variability.

Having observed more scatter plots of mean travel time and vehicle-to-vehicle variability in different days, it is found that approximately each mean travel time level has a certain variability associated with it, independent of the time of day, day of week or weather conditions. In other words, vehicle-to-vehicle variability can be treated as a function of mean travel time only. The slope of the fitted line might vary from freeway to freeway because of the different geometric condition. This finding indicates that when predicting travel time variability, the contribution from the individual driver differences can be estimated from the historical data according to the predicted mean travel time. For freeways where a large sample size of travel times is not available, speed-based travel time estimation methods can be used (Chen et al., 2003, Texas Transportation Institute, 2003, Van lint, 2004a).

The second term of equation 2 is of primary interest. It consists of demand and capacity related variations. The variation of demand results in the day-to-day, time to time flow fluctuation, whereas the variation of capacity results from the bad weather, incidents, and roadwork. Figure 9 shows the mean travel time and variability as a function of time of day using data from all weekdays. The mean and variability of travel times are calculated by observations in the same 5-minute time window across all weekdays. It can be seen that the travel times in the afternoon peak present smaller mean travel time but higher variability than in the morning peak.

The flow analysis in Figures 10 and 11 show that this high variability corresponds to high flow variability in afternoon peak. The values of hourly flow were obtained by summing up the flow of main entrances to the study section. During the off peak period, Figure 9 shows that the mean travel time is approximately around 10 minutes and the coefficient of variation is around 10%. From Figure 8, we see that when the mean travel time is 10 minutes, the corresponding driver-to-driver variability is around 5%. Therefore, we can say that the half of the travel time variability in off-peak period is due to the vehicle-to-vehicle variability. A similar situation occurs to travel times variability in morning peak. However, in afternoon peak period, the contribution from vehicle-to-vehicle variability to overall variance is reduced while the proportion from other source like flow fluctuation (Figure 10), weather, and incidents is substantially higher.

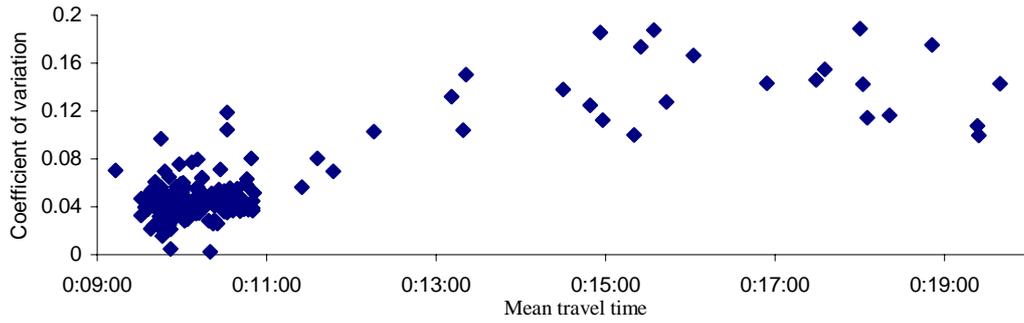
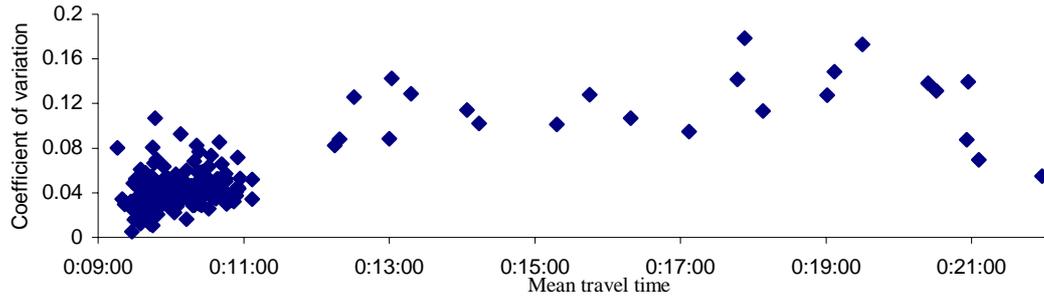


FIGURE 8 Scatter plot of mean travel time against vehicle-to-vehicle variability on different days

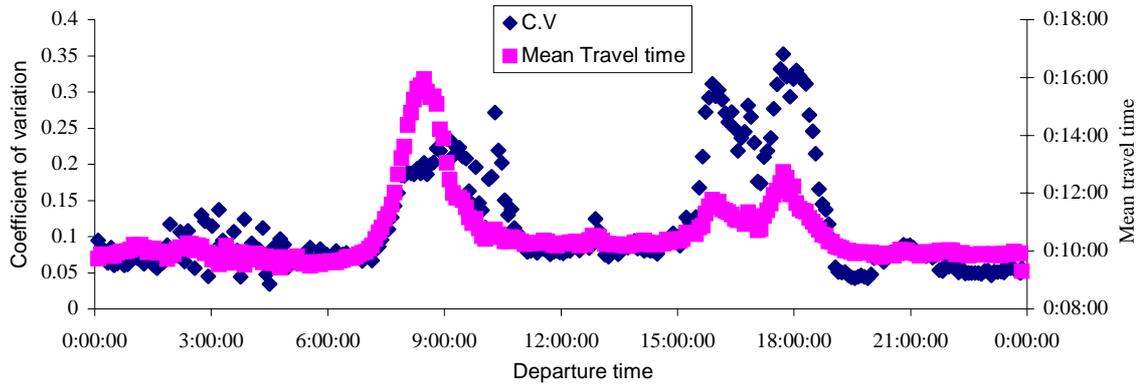


FIGURE 9 Mean travel time and variability according to the time of day over all weekdays

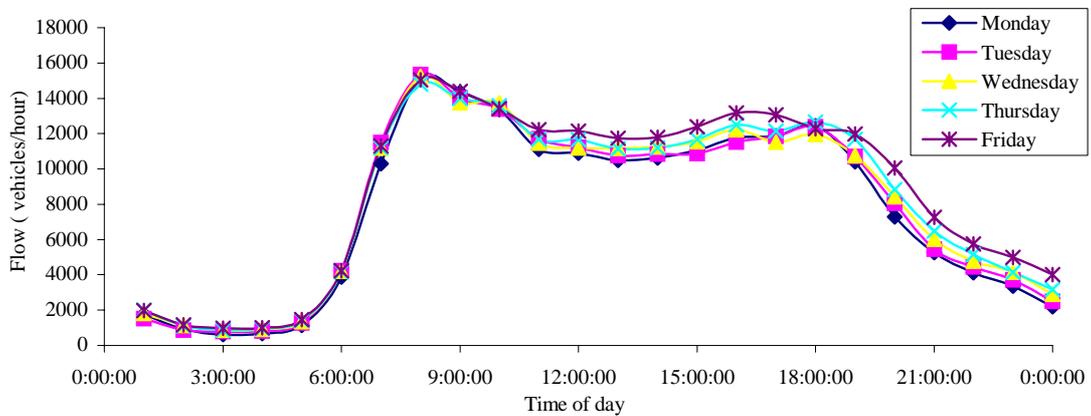


FIGURE 10 Flow variability by day of week

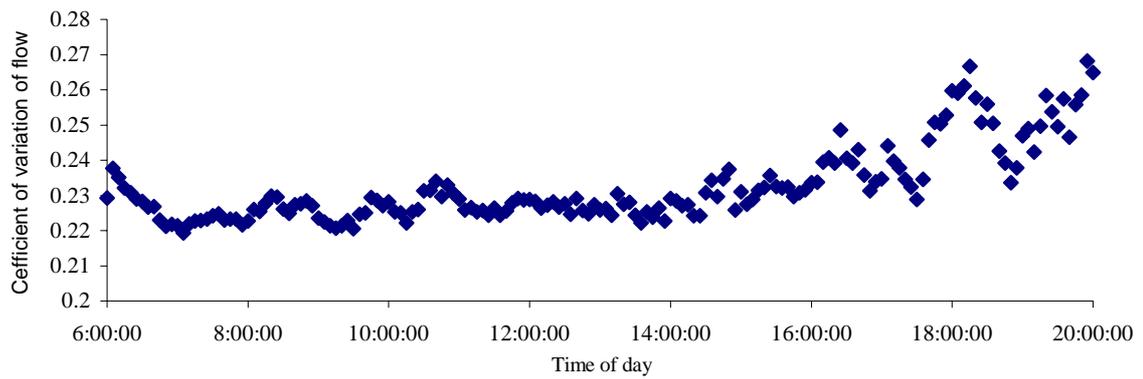


FIGURE 11 Flow variability by time of day

It is important to consider the interaction between various factors when analysing travel time variability. For example, wet pavement conditions in afternoon peak could cause more serious congestion than under free flow conditions. Figure 12 visually illustrates this effect. Heavy rains occurred during the day on Tuesday 21 October 2003 while the weather was fine on another Tuesday, 14 October 2003. There were no accidents on either day. The difference between the 90th percentile travel times is as high as 8 minutes in the peak period but is small under free flow conditions.

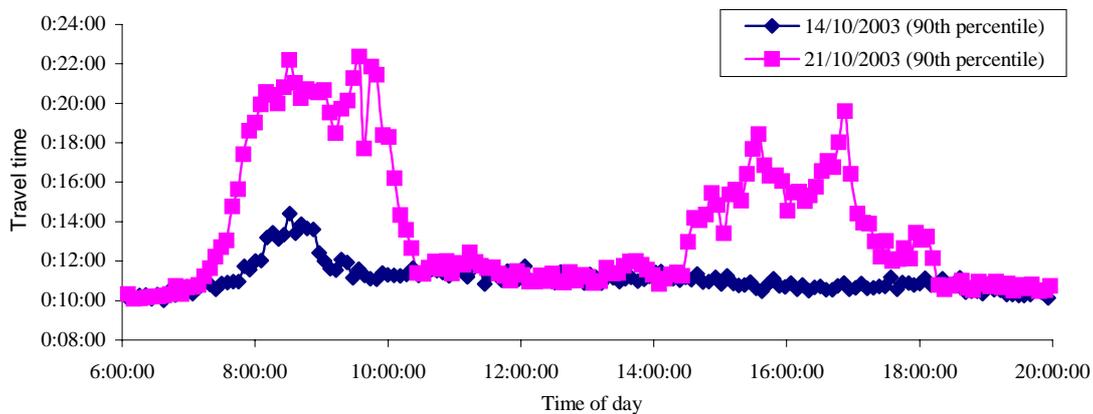


FIGURE 12 Comparison of 90th percentile travel time on 14/10/2003 and 21/10/2003

As noted earlier, the four gantries along the study site provide not only route travel times but also three link travel times. An examination of the variance of link travel times and route travel times found that the travel times were spatially correlated. Figure 13 shows the difference between the variance obtained from the observed route travel times and that derived from the link travel times by assuming independence. When link travel times are assumed to be independent, the route travel time variance is simply the sum of the travel time variance of links composed the route. It is clear that the variance from observed route travel times is greater than that estimated by the variance from the link travel times based on an assumption of independence, which indicates that there are covariance between link travel times. In other words, travel times are spatially correlated.

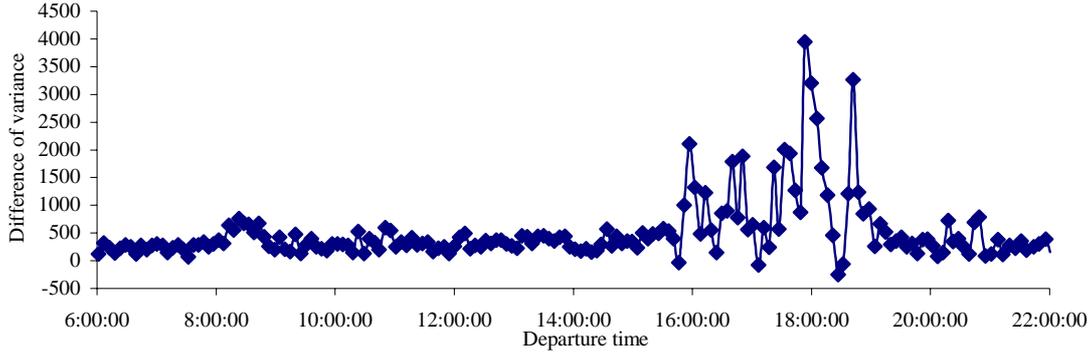


FIGURE 13 Variance difference of the observed route travel times and derived from link travel times by assuming independence

5. Travel time variability analysis by multiple regression

The final component of the analysis focuses on identifying the sources of travel time variability. Multiple regression was used to represent the variation in the dependent variable (travel time) as a function of the explanatory variables and their interactions. The travel times were aggregated over 5 minutes to eliminate the effect of driver-to-driver variability. Because there were no road works or major events during the analysis period, four sources of variability are considered in this study: time of day, day of week, weather condition and presence of an incident. The impact of weather is measured by the amount of rain. The presence of a hazardous incident (causing lane closing) located downstream during the period when the vehicle was traversing the study section is indicated by a 1, otherwise a 0. The natural log travel time is used to normalize the distribution of travel times.

The two-way interaction effects were included but not higher order interaction effects as they are not common for travel times over one-month period. Equation 3 relates the travel time to each of the explanatory variables and therefore provides a basis to test the contribution of each source to the travel time variability:

$$\ln t = b_0 + b_1 a_1 + b_2 a_2 + b_3 a_3 + b_4 a_4 + b_5 a_1 a_3 + b_6 a_1 a_4 + b_7 a_2 a_3 + b_7 a_2 a_4 + \varepsilon \quad (3)$$

Where: t = Average travel time for a 5 minute time window

$\ln t$ = Natural log travel time

a_1 = Day of week, coded by a set of dummy variables

a_2 = Time of day, indication for each 5 minute period of the day

a_3 = Weather condition

a_4 = Incident condition

$a_1 a_3$ = Interaction effect of a_1 and a_3

$a_1 a_4$ = Interaction effect of a_1 and a_4

$a_2 a_3$ = Interaction effect of a_2 and a_3

$a_2 a_4$ = Interaction effect of a_2 and a_4

ε = Error term

The error term represents the deviation of the observed individual value from that obtained from model. It is assumed that the errors have a normal distribution with constant variance (σ^2).

A set of dummy variables was used to represent each day of week. First, travel times collected from 7:30-8:30 am, and 5:00-6:00 pm are compared in terms of day of week. Table 2 shows that the mean and variance of the morning peak travel times on Mondays are larger than the rest of the week. During the afternoon period, the travel times on Wednesdays have the second largest mean and the highest variability, while the travel times on Fridays show the largest mean and the second highest variability. There were two occurrences of heavy rain during the afternoon peak on two of five Wednesdays, which caused the increased mean travel time and high variability. Whereas the high demand occurring in Friday afternoon resulted in the higher mean travel time.

Table 2: Comparison of mean and variability of travel times in morning peak and afternoon peak in term of day of week

	Day	Mean travel time in seconds*	N	Std. Deviation
Morning Peak 7:30-8:30	Monday	902	48	165
	Tuesday	808	48	146
	Wednesday	830	60	126
	Thursday	845	60	138
	Friday	835	48	120
Afternoon Peak 5:00-6:00	Monday	607	48	7
	Tuesday	690	48	114
	Wednesday	852	60	317
	Thursday	678	60	107
	Friday	900	48	176

Key: average of travel times in 5-minute periods across an hour for the same type of days

There are two groups of data used in the multiple regression. The first group are comprised of travel times collected from the morning peak. The second group of data were collected from the afternoon peak as the higher afternoon travel time variability across different days is of interest. Using multiple regression of the morning peak data against the factors listed in Equation 3, it is found that 56% of the variability can be explained (See Table 3). Further regressions were undertaken and it is found that time of day and day of week explain 53% of the variability while another 2% of the variability is explained by rain and incidents. Importantly, the interaction effects only explain 1% of the travel time variability in the morning peaks across different days of the week. Analysis of the afternoon data shows that the independent variables excluding interaction terms explain 43% travel time variability in afternoon peak crossing different day of weeks. The interaction effects contribute 7% to the variability in the afternoon peak. Again, time of day and day of week have more significant influence on travel times than other factors. Around 25% variability is contributed by these two factors.

Table 3: Multiple regression by different explanatory variables

	Independent Variables	Adjusted R Square
Morning peak	Day of week, time of day, weather, incident and all interaction terms	0.56
	Time of day, day of week, weather, incident	0.55
	Time of day, day of week	0.53
Afternoon peak	Day of week, time of day, weather, incident and all interaction items of them	0.50
	Time of day, day of week, weather, incident	0.43
	Time of day, day of week	0.25

It can be seen from the above findings that day of week and time of day effects have more significant influence on morning peak travel times than on afternoon peak travel times. Whereas effects like rain and incidents, contribute more variability to the afternoon travel times than the morning travel times. Therefore, almost half travel time variability in morning peak is caused by demand related variations, while the 25% of the travel time variability in afternoon peak results from the capacity related variations. This is consistent with the traffic situation of the study site. The tunnel part of the route is vulnerable to the incidents occurring inside and traffic condition on the freeway downstream of it. Once the incidents happen or the congestion on the downstream freeway tails back to the tunnel, one lane is closed. That causes a dramatic decline in the tollway capacity. The unpredictable occurrence of these conditions is reflected by the high variability as shown in Figure 10. It is expected that including of a traffic flow variable in the multiple regression would help to explain more variability in the travel times. However, the variation of flow is also caused by the variables listed in Equation 3 so that it cannot be treated as an independent variable.

6. Conclusions and recommendation

The analysis undertaken in this study was conducted on a specific tollway using travel time data derived from AVI equipment. Apart from providing a large quantity of data, the AVI technology also guarantees accuracy in the travel time measurements. Thus it supplies a sound and robust database for performing travel time variability analysis. First, travel time distribution properties were investigated in term of various time windows. We observed that the travel time distribution tends towards a normal distribution as the time window is reduced. This finding provides a constructive theoretical foundation for the modelling of travel times.

Secondly, the overall variance of travel times has been decomposed into vehicle-to-vehicle travel time variability and between group travel time variability. The latter was further decomposed into demand and capacity related variability. The relationship between mean travel times and vehicle-to-vehicle variability has been examined and it was established that the vehicle-to-vehicle travel time variability in a

small time interval (5 minutes) is related to the magnitude of mean travel time. In addition, the vehicle-to-vehicle variability explained about half travel time variability in the off-peak and morning peak period, but contributed comparatively little to afternoon peak travel time variance. The influences of demand and capacity related factors and interaction factors on travel time variability have also been illustrated graphically and interpreted.

Multiple regression methodology was used to identify the contribution of each source to the travel time variability of two groups of data. The travel times were averaged over a 5 minute period in morning peak and afternoon peak period, respectively. The variability of two groups data present distinctive properties. It indicated that the source of variability might be different even for the same road with respect to travel times for different times of day. Though the research is based on a specific site, we believe that the analysis and methodology are applicable to other locations.

As noted earlier the research interest lies more in the day-to-day instead of time-to-time travel time variability. Stated another way, the day-to-day travel time variability measures the change of travel times for the same trip at the same time from day-to-day, while time-to-time travel time variability emphasizes the change of travel times of the same trip on the same day from time to time. For the latter, the variation of demand associated with the commuter-working pattern is the main cause of that variability. The day-to-day travel time variability is thus the authentic measurement of the reliability of traffic service. To further explore that construct, travel time data across a longer time period than the one-month used in this study are needed to separate out the effect of time of day and day of week. This appears to be a promising future direction for research.

Acknowledgement

The authors would like to thank Transurban for supplying the data used in this research.

References:

- Bates, J., Polka, J., Jones, P., and Cook, A. (2001). The valuation of reliability of personal travel. *Transportation Research E* 37 (2-3), 191-229.
- Kwon, J., Coifman, B., and Bickel, P. (2000). Day-to-Day travel time trends and travel-time prediction from loop-detector data, *Transportation Research Record* (1717), 120-129.
- Jaccard, J., Turrisi, R., and Wan, C. K. (1990). *Interaction effects in multiple regression*, Newbury Park
- Lam, T. C., and Small, K. A. (2001). The value of time and reliability: measurement from a value pricing experiment. *Transportation Research E* 37 (2-3), 231-251.
- Li, R. M., Rose, G., and Sarvi, M. (2004). Evaluation of speed-based travel time estimation methods. Personal communication.
- Lomax, T., Turner, S., and Margiotta, R. (2003). Monitoring urban roadways in 2001:examining reliability and mobility with archived data. Report FHWA-OP-03-141. Texas Transportation Institute, The Texas A&M University System, USA.

- Noland, R.B., and Polak, J. W. (2002). Travel time variability: a review of theoretical and empirical issues. *Transportation Reviews* 122 (1), 39-54.
- Richardson, A.J., and Taylor, M.A.P. (1978). Travel time variability on commuter journeys. *High Speed Ground Transportation Journal* 12 (1), 77-99.
- Smeed, R.J., and Jeffcoate, G.O. (1971). The variability of car journey times on a particular route. *Traffic Engineering Control* (12).
- Sun, C., Arr, G., and Ramachandran, R. P. (2003). Vehicle reidentification as method for deriving travel time and travel time distribution. *Transportation Research Record* (1826), 25-31.
- Turochy, R.E., and Smith, B. L. (2002). Measuring variability in traffic conditions by using archived traffic data. *Transportation Research Record* (1804), 168-172.
- Van Lint, J. W. C., Tu, H., and Van Zuylen, H. J. (2004). Travel time reliability on freeways, *Proceedings of 10th world conference on transport research (WCTR)* [CD Rom], Istanbul, Turkey.