Australasian Transport Research Forum 2015 Proceedings
30 September – 2 October 2015, Sydney, Australia
Publication website: http://www.atrf.info/papers/index.aspx

Quality investigation and variability analysis of GPS travel time data in Sydney

Ruimin Li[1], Malcolm Bradley[1], Matthew Jones[1], Steven Moloney[2]

[1] Bureau of Statistics and Analytics, Transport for New South Wales, NSW 2008

[2] Intelematics, Richmond, Vic 3121

Email for correspondence: Ruimin.li@transport.nsw.gov.au

Reliable and accurate travel time data can provide valuable performance measures to support operational applications in areas of congestion management and routing analysis. The travel time information is also essential to the calibration and validation of travel demand models in order to better model current and future travel for congested area.

Historically, the ability to collect travel time data in sufficient quantity to provide reliable, robust evidence has been severely limited, almost to the point of being unattainable. Indeed, to take it further, any aspirations of addressing travel time variability are almost inconceivable.

Recent developments with Global Positioning Systems (GPS) data have shed new light in this field. The ability to collect large volumes of data from GPS devices has provided a wealth of data for use in this area. Such data collected and processed by Intelematics for a large proportion of the Greater Metropolitan strategic road network has been reviewed and analysed to examine time of day and day of week variations in travel times by traffic conditions for individual sections of road. Findings from this research could significantly influence the guidelines, processes and procedures for future model validation.

Keywords: GPS, Travel time variability, Travel demand model

1. Introduction

Reliably measured travel times can play an important role in assisting road authorities in controlling congestion and managing road operations, as well as assisting drivers to select optimal routes. It can be used to well represent traffic conditions and is more explicit than traffic flow for road users. Travel time is also a key input for travel demand forecasting model calibration and validation, especially in the development of parameters of Volume Delay Functions (VDF). To develop a reliable travel demand forecasting model, adequate samples of travel time data are essential for model calibration and validation. Historically, the most common approach has been to reconstruct the travel times from speed data provided by inductive loop detectors or other point-based sensors as loop detector is the most widely available traffic data source (Van Lint, 2004, Bhaskar, A. and Qu, M., 2015). Traditional techniques for measuring travel times include the floating car method and number plate matching. These techniques don't require any instrumentation or infrastructure to be constructed on the road network. However, the application of such techniques on the wide area network is not cost-effective and reliable. With the developments of new technologies, travel time data can be collected directly by deploying technology such as GPS enabled probe vehicles, toll tag readers or automatic licence plate recognition. There were a number of studies which focused on traffic analysis using bus, truck or Taxi GPS data (Uno et al, 2009, Mazloumi et al 2010, Lu and Li 2014, Bernardin et al 2015, Flaskou et al 2015).

The quality investigation and variability analysis of GPS based travel time data are essential for managing urban road network traffic and understanding the reliability of congested traffic system. Over the past decade, GPS travel time data has been widely used in the transport field (Shen and Peter, 2014). However, the quality of GPS data is not assessed comprehensively, especially in Australia. Another key hurdle is of how to accurately understand and handle the travel time variability, which has drawn attention and incurred preliminary researches in the past few years (Sohn and Kim, 2009, Susilawati et al, 2011).

In our research, we paid particular attention to different aspects of travel time variability based on GPS data collected in Sydney. A better understanding and investigation on GPS based travel data analysis will have beneficial impacts on many challenges in the transport field. One of them is to better understand the reliability and performance of the Sydney traffic system. Another is to investigate the impact of such variability on the calibration and validation of the travel forecasting model.

This paper first evaluates the reliability of GPS-based data using travel time data collected from traditional floating car and number plate survey methods. We then try to gain an insight on the day-to-day variability using the large amount of GPS data collected from 5 Sydney major motorways and roads, which exhibit different traffic patterns. A detailed variability analysis is also conducted to reflect the challenges experienced to model the averaged traffic conditions in the field of travel demand forecasting.

2. Data source

Adequate samples of travel time data are essential to provide not only the mean travel time but also the measurements of travel time variability. The use of traditional floating car methodologies for estimating travel times is now being superseded by the proliferation of GPS devices that are able to be sourced and utilised as a comprehensive set of probe data. This probe data will cover a greater area more frequently throughout the day than is practical through the floating car service. Data can be utilised in a near-real time or aggregated dimension for historic and predictive purposes.

The GPS data used in this study were supplied by Intelematics Australia. Intelematics is an Australian pioneer of connected mobility services and has been operating telematics programs in Australia since 1999. Intelematics began collecting probe data in 2010 and today receives billions of samples per year across Australia and New Zealand.

Each GPS-equipped probe vehicle delivers periodic data samples that typically include as a minimum:

- Date/Time
- Device ID
- Latitude/Longitude
- Speed
- Heading

The data are generated by a combination of private, commercial vehicles, truck and taxi.

The private sources for Intelematics data include portable navigation devices, navigation apps and connected vehicles. The growing volume of probe data is expected to escalate further as the market for in-vehicle traffic advisory and navigation services shifts from portable navigation/traffic advisory services to services embedded in vehicles. It is important to note that the integrated nature of these services means an anticipated higher probe data generation rates than would be achieved from aftermarket systems in the same vehicle (most of which are used only when travelling to an unfamiliar destination). All consumer-generated probe data is encrypted prior to being obtained by Intelematics and only aggregate information is provided to third parties. Use of customer data for this purpose is fully disclosed.

Travel speed data is created using a range of mathematical and geo-spatial techniques. Intelematics has developed proprietary algorithms to filter and process probe data in order to calibrate the products to the average driver prior to it being delivered. In the matching process, any probe data that is exhibiting unexpected behaviour in regards to driving or data that contains too much uncertainty is removed. Each probe is then routed to get their path of travel and calculate speed based on this path. The overall speeds are obtained using a weighted harmonic mean to correctly account for distance and time of each probe on each link.

Probe speed is determined based on a range of techniques including instantaneous speeds collected along road segments and traversal-based calculations based on a probe's journey. Traversal probe calculations are based on the time it takes a vehicle to travel between the entry and exit of a link. This type of calculation can provide more accurate travel time data that takes more account of intersection delay and time spent stationary in traffic. This can then be converted back into an average speed travelled by vehicles on an individual segment of road. There are restrictions on when probe traversals are used to eliminate the error caused by probe drift or vehicles parking, getting drive through, etc. This can result in a smaller sample of probe data; however, the remainder is of higher accuracy.

Intelematics has provided Bureau of Transport Statistics (BTS) two sets of data:

- Half year samples for 2014 by day of week with 15 minute aggregation across the Sydney region
- One year samples for 2014 by date with 1 minute aggregation for M5 and Parramatta road

In total, the size of the sample exceeds 50 million observations. The first dataset is mainly used for quality comparison and day-to-day travel time variability analysis, while the second set is used for travel time variability analysis due to the finer time aggregation window. The study sites are the major Motorways and arterial roads in Sydney, such as M5, M4, M2, Parramatta Road, Military Road and Victoria Road (Figure 1).

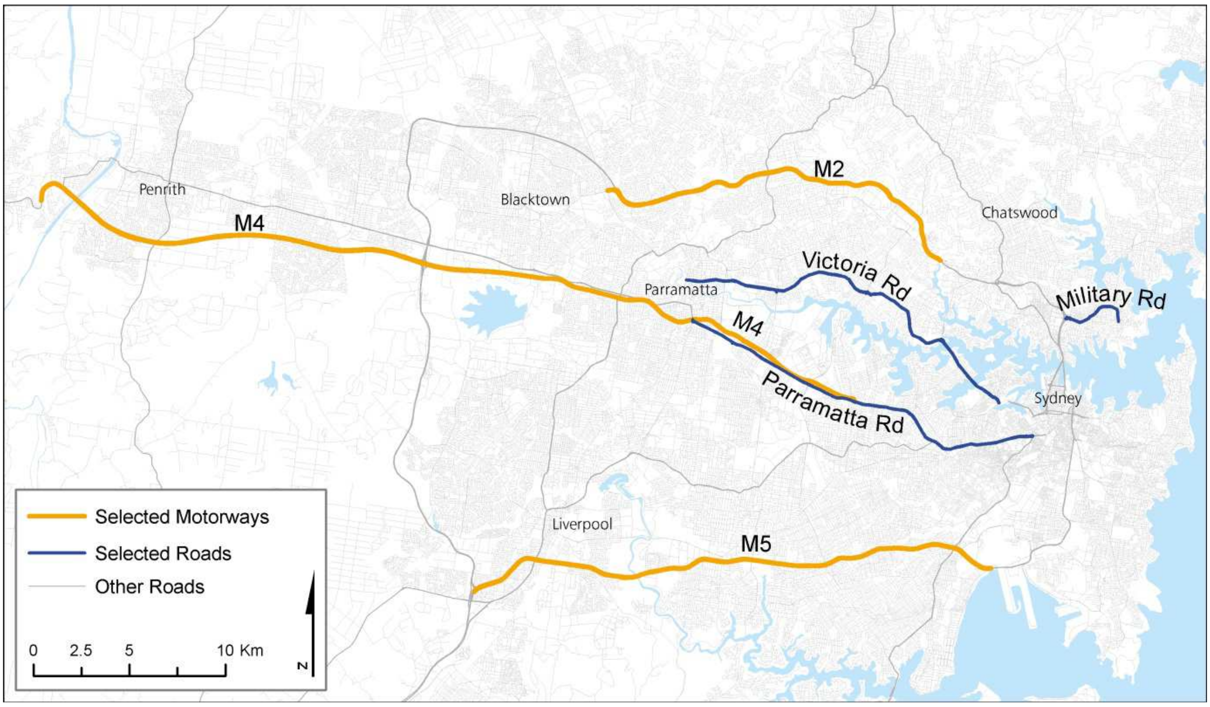


Figure 1: Study sites

The focus of the analysis within this paper is on the M5 and Parramatta Road, given the congestion conditions on these two roads.

3. Data quality

The accuracy of GPS data depends on many factors, including the quality of the GPS receiver; the position of the GPS satellites at the time that the recording was made; and the characteristics of the surrounding landscape. Over the past decade, GPS travel time data has been widely used in the transport field. However, the quality of GPS data is not assessed comprehensively, especially in Australia. In this study, Intelematics data has been compared with floating car survey data and a recent Number Plate survey. For each floating car run, the corresponding Intelematics mean travel time is calculated for that 15 minute departure time window. Outliers for the 15 min aggregated data sets are identified and removed according to the following criteria:

- Calculate mean travel time for each time window
- Calculate standard deviation for each time window
- If a record falls out of range [mean-2*standard deviation, mean+2*standard deviation}, it is treated as outlier

3.1 Comparison of floating car data and Intelematics data

BTS collected floating car travel time data along the M5 and Parramatta Road on a typical Tuesday and Wednesday in November 2012, respectively. Whilst somewhat dated, this survey data set is still considered sufficient to enable a valid comparison. The comparisons of the floating car travel time with the corresponding Intelematics travel time in AM and PM peak periods in the peak directions are shown in Figure2 and 3.

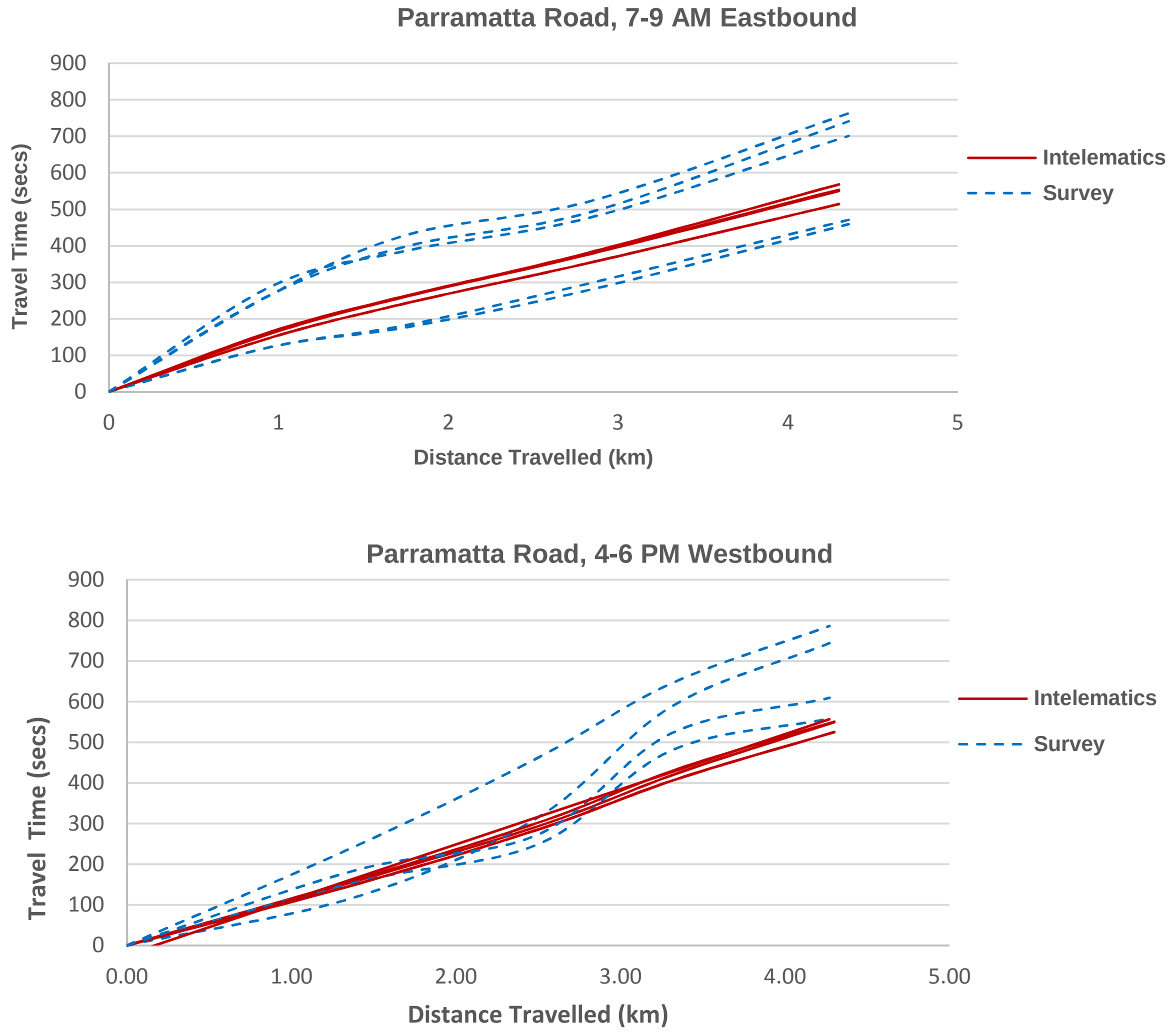
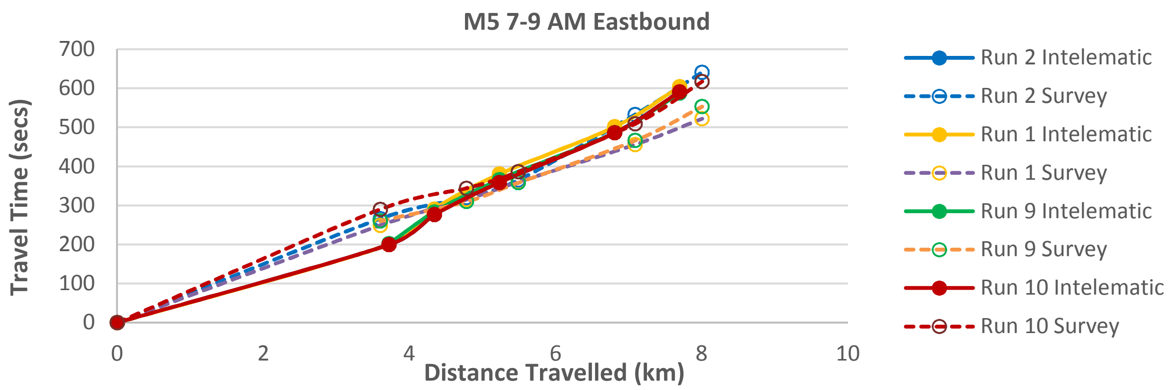


Figure 2: Intelematics and floating car survey travel time comparison, Parramatta Road at West Street to Cleveland St to Botany Rd and O'Riordan Street



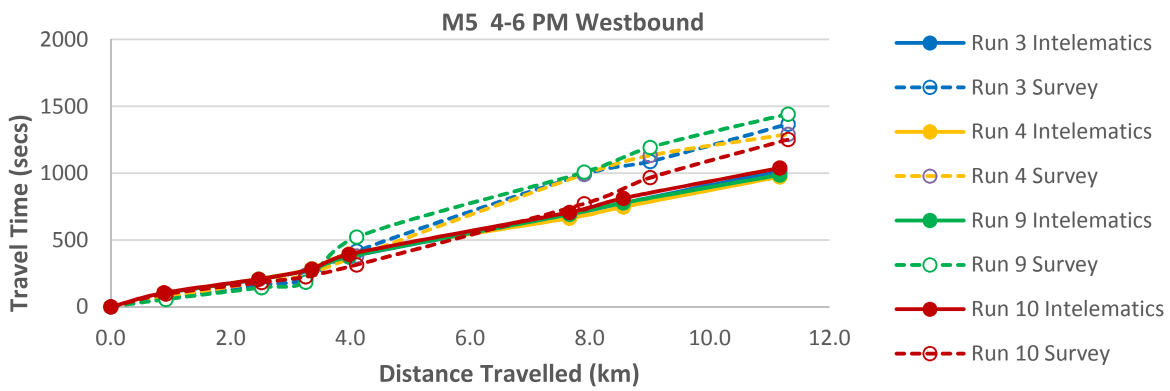


Figure 3: Intelematics and floating car survey travel time comparison, M5 at King George Road to Marsh Street, Airport Drive and Qantas Drive

It can be seen from the results that the travel time is consistently higher from the floating car survey compared to the Intelematics data for the M5 in the westbound direction. The Intelematics data was collected in 2014 and the floating car data was collected in 2012. There were some major road works on the M5 Motorway starting from August 2012 that may explain the higher travel time in November 2012. There is little difference between the two data sets on Parramatta Road.

3.2 Comparison of Number Plate survey and Intelematics data

A number plate matching survey was conducted for the same section of Parramatta Road for 4 hours over the AM and PM peak periods in October, 2014. Figure 4 shows the comparisons of the range of travel times from the number plate survey and the Intelematics travel time data, which corresponds to each floating car run.

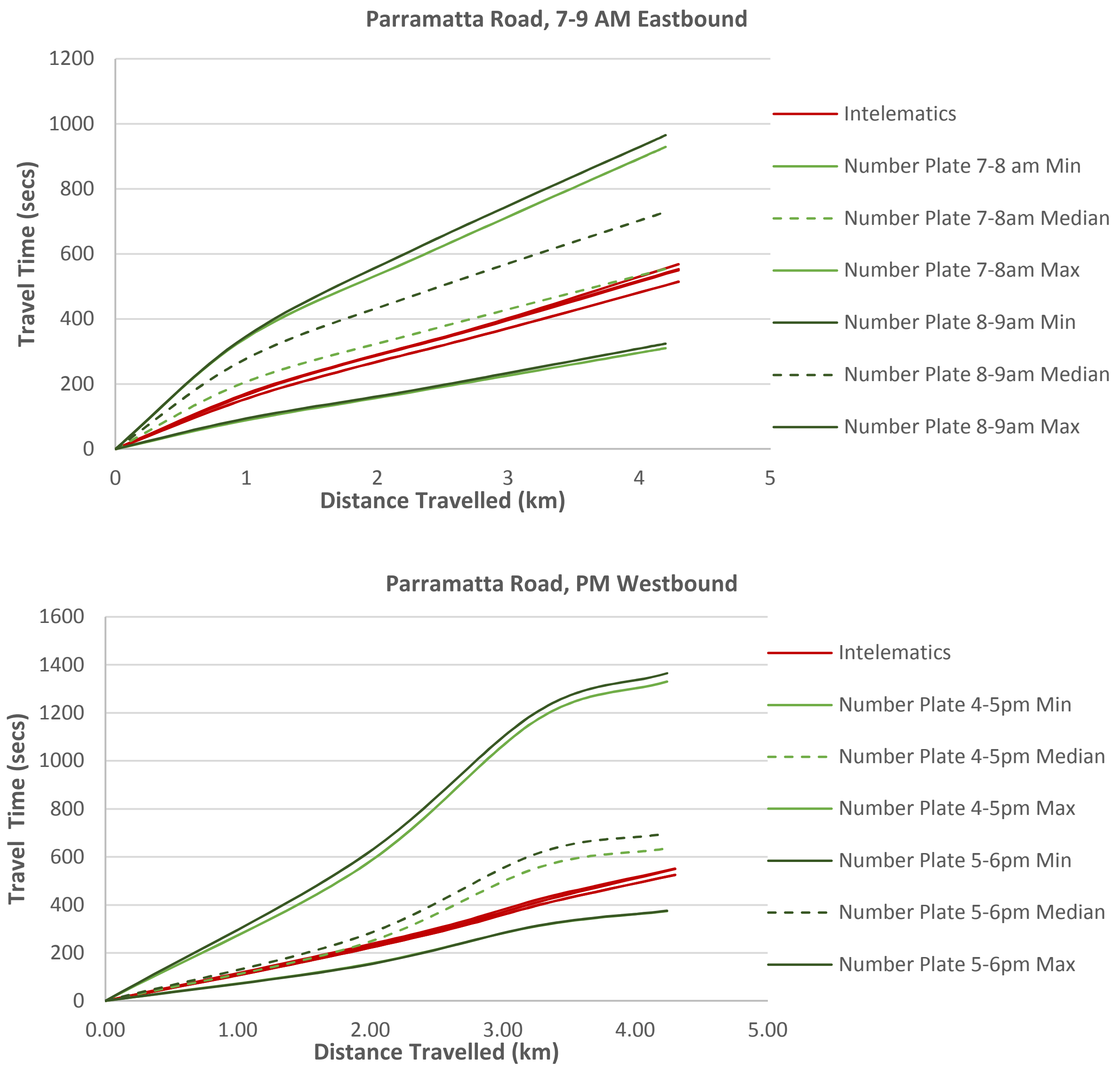


Figure 4: Intelematics and Number Plate survey travel time comparison, Parramatta Road at West Street to Cleveland St to Botany Rd and O'Riordan Street

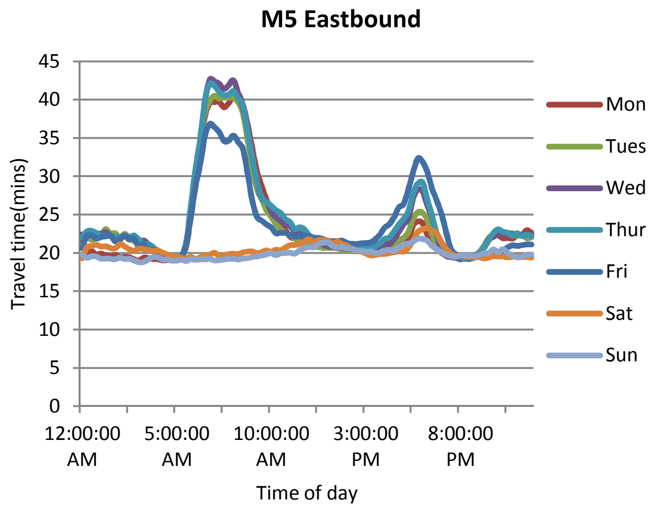
The results show that the Intelematics travel time data fall comfortably within the range of number plate survey travel times. The comparison of floating car data, Intelematics data and number plate survey indicates that GPS data is reliable, and offers the potential for the greatest scope in the collection of "big traffic data".

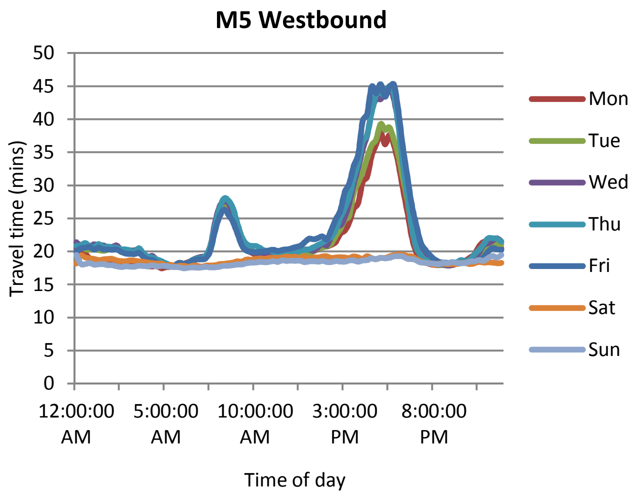
4. Day to day travel time variability

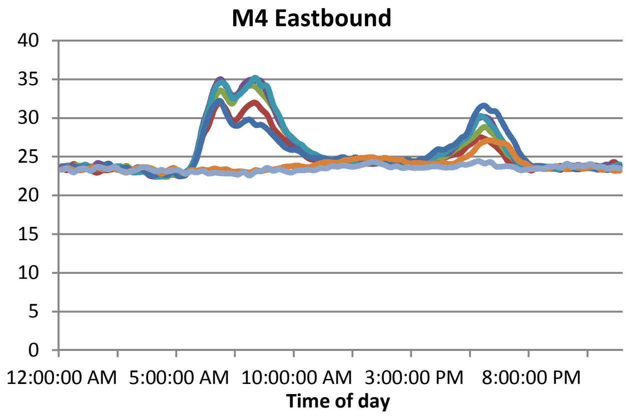
Day to day travel time variability reflects the change of travel times for a trip taken at the same time on different days. It is a direct reflection of the performance and reliability of the traffic system. Strategic transport models are generally calibrated to replicate the traffic conditions on an average weekday or working day (Hidas and Milthorpe (2009)). It is expected that there is underlying natural variances in travel times from one day to the next. This section aims to examine the travel time variability across the different days of week for major Motorways and arterial roads in Sydney. Saturdays and Sundays are also included to show the difference in travel times between motorways and arterial roads in weekends.

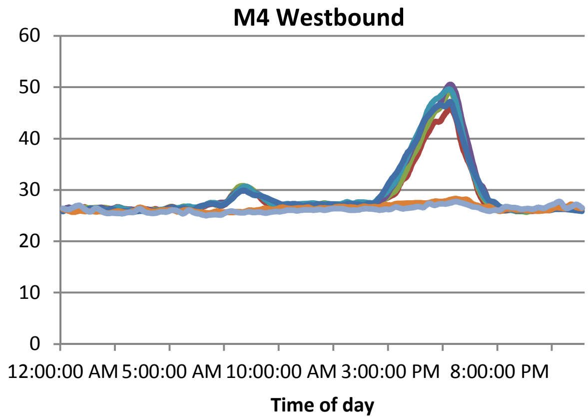
Figure 5 shows the averaged travel time in each 15-minute time window on the same day of week over the half year period. In the M5 peak directions, the congestion is longer and more severe than other motorways. Interestingly, the average travel times on Fridays are lowest in the morning peak but highest in the PM peak in the eastbound direction. The eastbound counter-peak direction also shows a similar order of magnitude of variability as the peak direction, though the average travel time is lower. Travel times on the M4 display similar patterns in terms of the mean across the day in the westbound direction, whereas mean travel times in the eastbound direction are significantly different in both AM and PM peak periods. The travel times on the M2 are characterised by high day-to-day variability along the peak direction in the PM peak period. There is almost no congestion in the counter-peak direction.

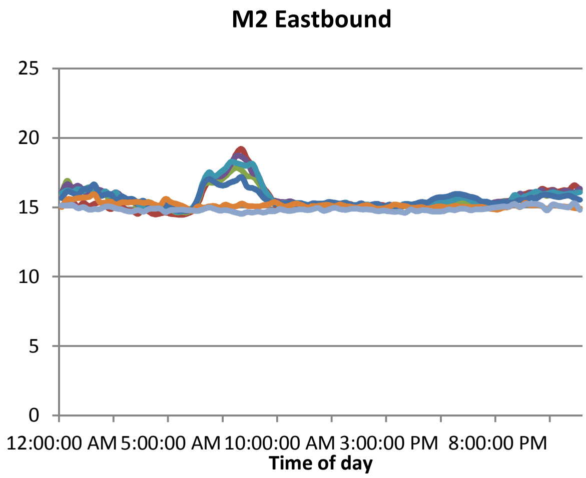
It also can be seen from Figure 5 that the identified arterial roads display different congestion patterns to the motorways. Parramatta road and Military Road also exhibit high variability in mean travel times across the different day of week in the inter-peak period in both directions. On average, Wednesdays appears to display the most congested traffic condition in the peak directions for both motorways and arterial roads. There is almost no congestion or slight congestion at the weekend on the motorways, while congestion is still evident on the major arterial roads in both directions. The congestion broadly starts to be shown from 9am and can last till the end of PM peak. The travel times on the M5 and Parramatta road are shown to be less reliable than other roads. They are further analysed in the following section using the data aggregated in 1-minute window.

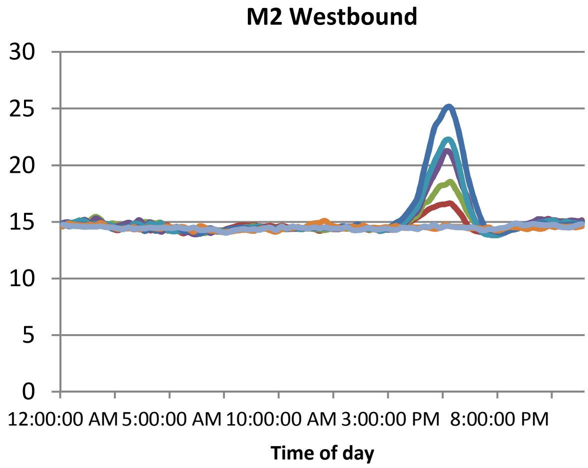


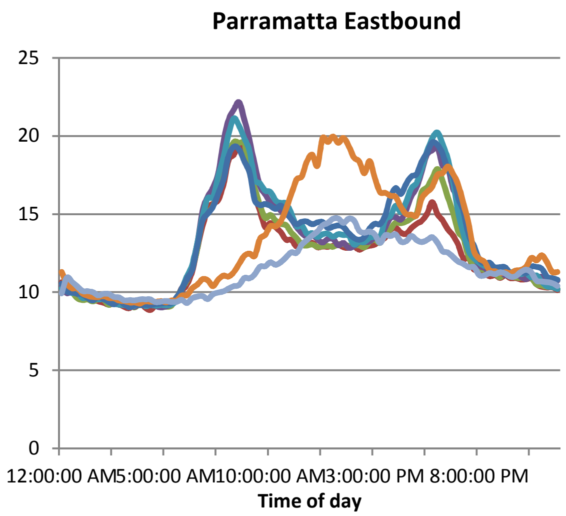


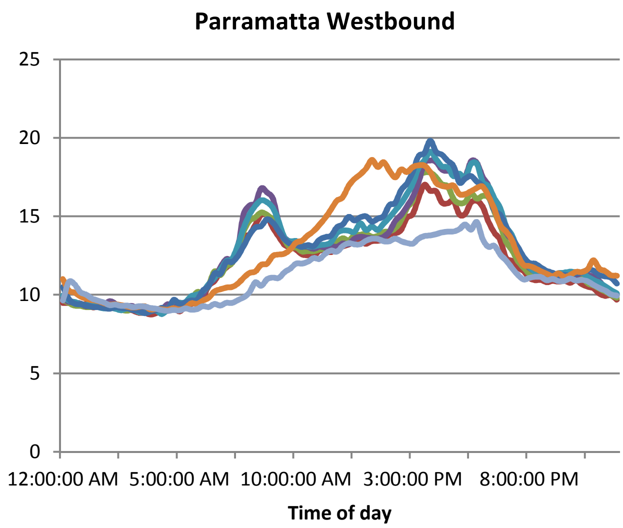

M4 Eastbound
40
35
30
25
20
15
10
5
0
12:00:00 AM
5:00:00 AM
10:00:00 AM
3:00:00 PM
8:00:00 PM
Time of day


M4 Westbound
60
50
40
30
20
10
0
12:00:00 AM
5:00:00 AM
10:00:00 AM
3:00:00 PM
8:00:00 PM
Time of day


M2 Eastbound
25
20
15
10
5
0
12:00:00 AM
5:00:00 AM
10:00:00 AM
3:00:00 PM
8:00:00 PM
Time of day


M2 Westbound
30
25
20
15
10
5
0
12:00:00 AM
5:00:00 AM
10:00:00 AM
3:00:00 PM
8:00:00 PM
Time of day


Parramatta Eastbound
25
20
15
10
5
0
12:00:00 AM
5:00:00 AM
10:00:00 AM
3:00:00 PM
8:00:00 PM
Time of day


Parramatta Westbound
25
20
15
10
5
0
12:00:00 AM
5:00:00 AM
10:00:00 AM
3:00:00 PM
8:00:00 PM
Time of day

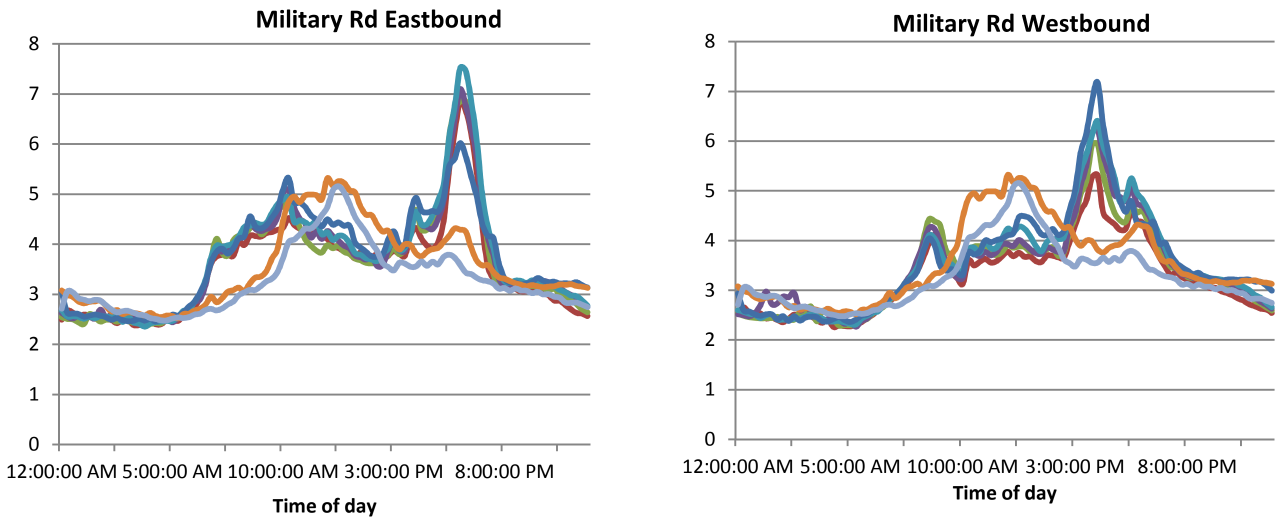


Figure 5: Day-to-day travel time variability analysis

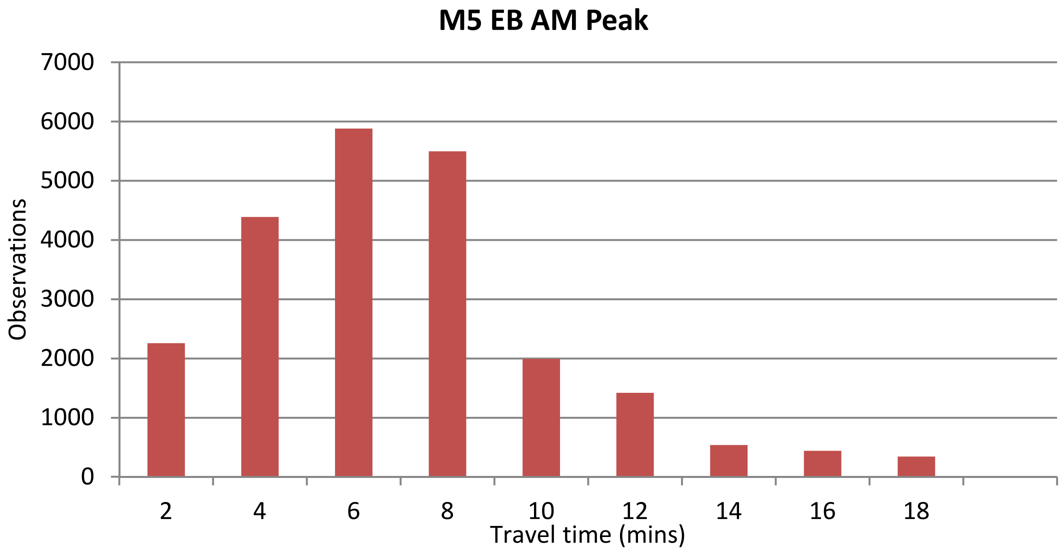
5. Time of day travel time variability and distribution analysis

Travel demand models usually predict the temporal changes of congestion by aggregated time periods. For example, STM3 produces vehicle matrices for the following four periods:

- 2 hour AM peak (7-9 am)
- 6 hour interpeak (9am-3pm)
- 3 hour PM peak (3-6pm)
- 13 hour off-peak (6pm-7am)

It is assumed that the average demand and travel time for each time period is representative of the travel patterns for the time period. In other words, there is constant travel time within a study period.

The entire 2014 data set of 1-minute aggregated travel time information on a 1.8 km long section on the M5 is analysed to evaluate and assess the travel time distribution in both the AM and PM peak periods (more than 50,000 measurements). The section starts from King Georges Rd in the city direction, representing the most congested section of the M5. Figure 6 shows the distribution of 1-minute aggregated travel times for the AM and PM peak periods on all workdays in 2014 in the peak direction.



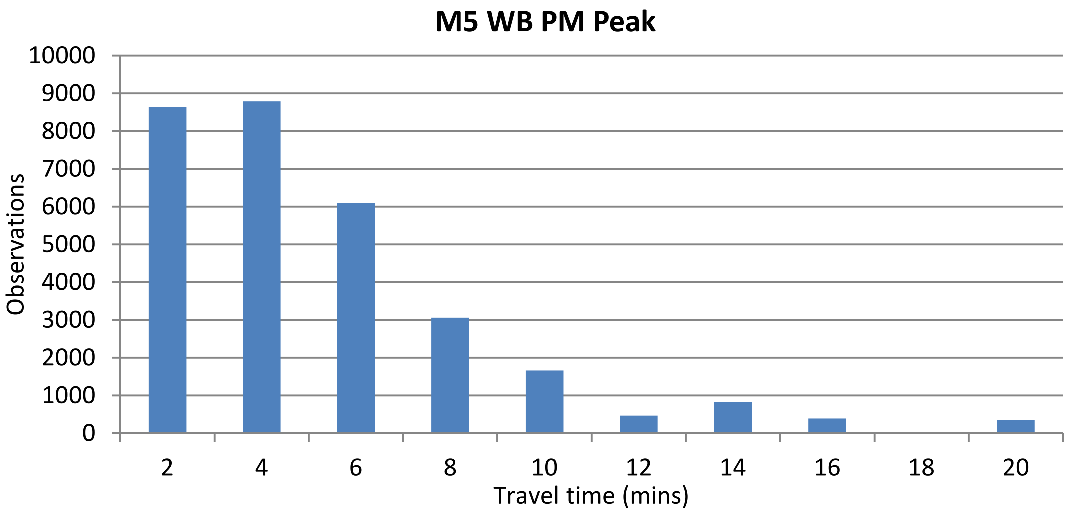


Figure 6: Distribution of travel time: all weekdays in 2014

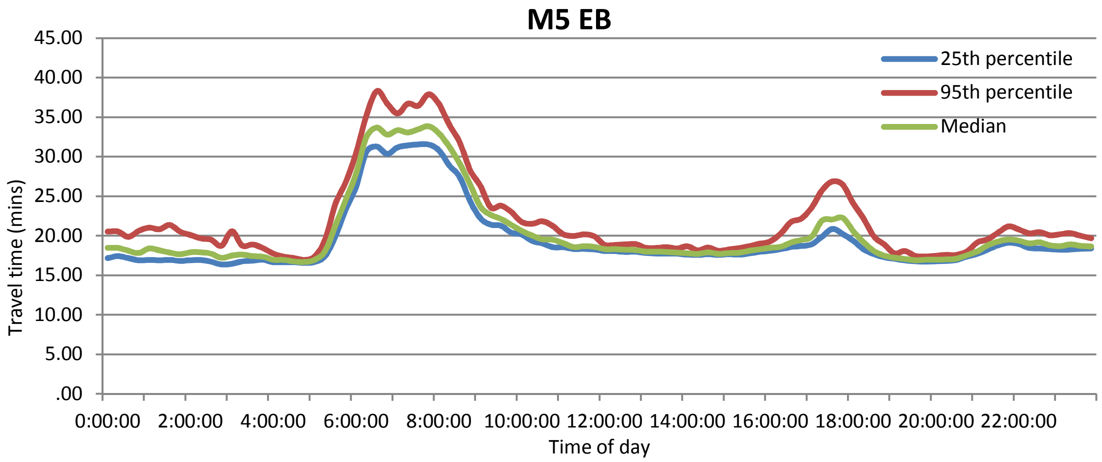
Table 1: Statistical measures of travel time distribution

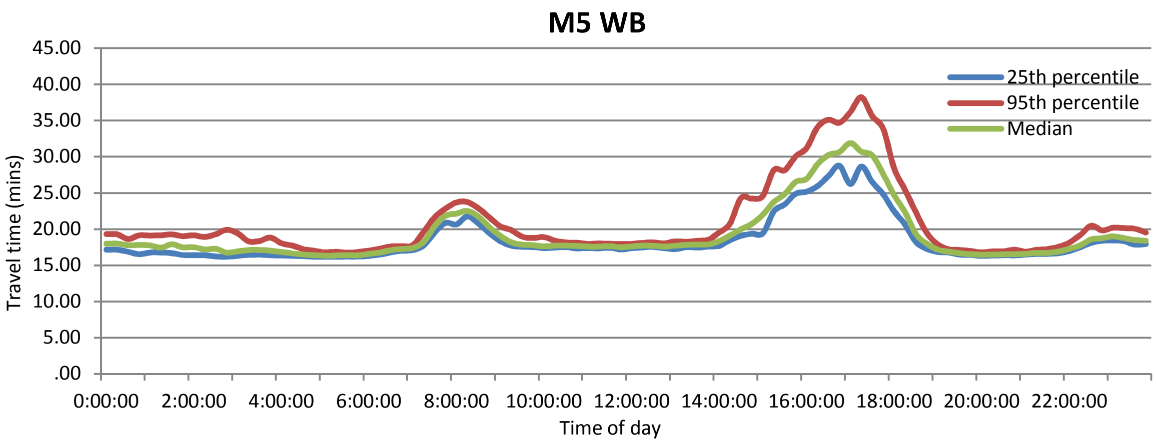
	Mean (mins)	Median (mins)	Skewness	Kolmogorov-Smirnov test Sig.
M5 EB AM	6.8	5.8	7.5	.000
M5 WB PM	5.3	3.6	7.2	.000

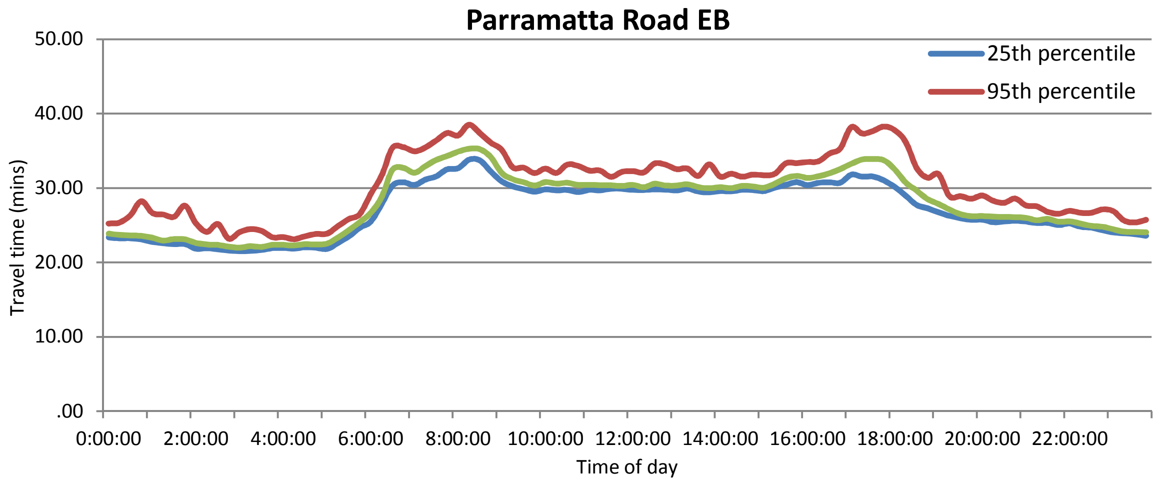
Table 1 reports statistical measures for the travel times collected the section of M5 for all workdays in 2014 by the AM and PM time periods. Given $p < 0.05$, the H_0 normality assumption is rejected. The distribution is thereby not "Normal". Higher value of skewness indicates that the distribution is asymmetrical. It is also supported by the large difference between the mean and median travel times. This result is consistent with the findings from Li et al. (2006) that travel times on a toll way in Melbourne tends towards a normal distribution only as the time window decreases.

To further examine the shape of travel time distributions across the day, 1-min travel time data are aggregated in different 15 minute time window. Figure 7 shows the distributions of 25th percentile (T25), median (T50) and 95th percentile (T95) travel times by time of day for the M5 and Parramatta Road. The reason to use the 25th percentile instead of 5th percentile is that T25 is less affected by outliers. M5 displays the similar order of magnitude of congestion in both AM and PM peak periods. The median travel time reaches 30 minutes in the worst congested time period. The AM peak congestion on the M5 can start as early as 5:30am, and not settled until 9am. The differences between T95 and T25, which indicates the variability of travel times in each 15 min time interval, are larger in the peaks. The afternoon peak on the M5 also starts early (at 3pm). In the most congested time interval (5:15-5:30 pm), the 95th percentile minus 25th percentile value is 10 minutes.

Drivers would experience longer travel times on Parramatta Road even in the interpeak period along both directions. The measure of T95-T25 is relatively consistent from the AM peak period to PM peak period, indicating eastbound direction is congested but reasonably reliable. The westbound direction of Parramatta road displays different patterns in terms of variability in travel times. In the afternoon peak, the difference between T95 and T25 is 15 minutes. This result indicates that drivers are unable to predict their travel time when traveling along Parramatta Road westbound direction, especially in the peak period. Overall, the peaks of T95 are wider than T50 and T25 for both study sites.







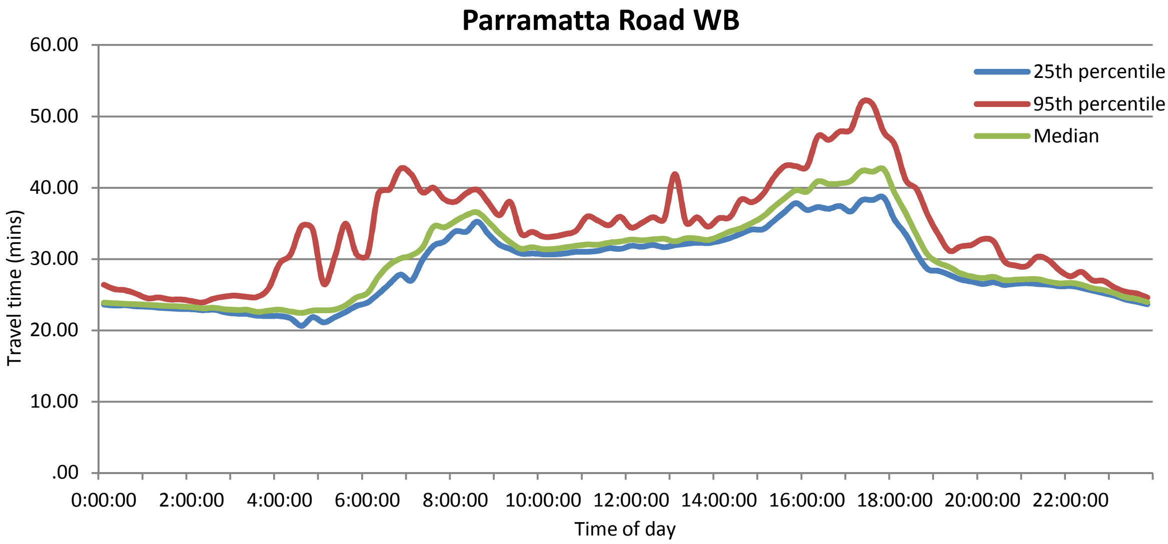


Figure 7: Travel time variability across the day

6. Average weekday vs. average schoolday

Strategic transport models often reproduce travel patterns on an average weekday or average schoolday for the study area. The Sydney Strategic Travel Model Version 3 (STM3) models the traffic condition on average schoolday in the Sydney region. It has been reported that there are significant differences in traffic counts between the average of weekdays and average of schooldays (Hidas and Milthorpe (2009)). In this study, we used the 1 minute aggregated data to investigate if the average travel time of weekdays is significantly different from the average travel time of schooldays. The one-minute travel time data in AM peak on each links of M5 and Parramatta Road are aggregated into a 15 minute time window on the peak direction.

Paired-samples T test was conducted to compare the means of travel time in each 15 minute time window in AM peak on each link of M5 and Parramatta Road for weekdays and schooldays, respectively. It is concluded from the preceding analysis that the travel time data is not normally distributed. However, the Central Limit Theorem lets us use the t-test even if the population is not normally distributed as long as the sample sizes are large enough (Hays 1981).

The null hypothesis is:

H_0 : There is no difference in mean travel times average weekday and average schoolday

Table 2 shows the T test results for M5 and Parramatta Road travel time data.

Table 2: T test statistics

	Mean (mins)	Std. Deviation	t	df	Sig. (2-tailed)
M5 Schoolday-Weekday	0.01445	0.02589	9.071	263	.000
Parramatta Schoolday-Weekday	0.0046	.00977	10.725	519	.000

It can be seen from the t statistics that the null hypothesis is rejected since $p < 0.05$ (in fact p =.000). There is strong evidence that the travel times on an average schoolday and average weekday are statistically different in the AM peak. This result indicates that it is probably insufficient to model the traffic pattern on an average weekday given schooldays can have more severe congestions, particularly in the AM Peak.

7. Conclusions

It is a challenging task to establish reliable traffic forecasting models in congested urban networks. Practitioners suffer from a lack of large sample size and multi-modal travel times data collected from undesignated vehicles. As a result, investigations and analysis on the available data was limited to enable full picture of the real road conditions. This paper provides an analysis of a large amount of GPS data collected from the Sydney region. It focuses on 5 Sydney major motorways and arterial roads to gain greater insights into day-to-day variability and traffic patterns. The M5 and Parramatta roads are analysed further given their widely accepted congestion and less reliable performance. Some general findings from this study are:

- GPS travel time data appear to be reliable when compared with travel time data collected from floating car and number plate matching techniques.
- Drivers usually experience the worst congestions on Wednesdays on both motorways and arterial roads. Generally, the travel times on motorways are relatively more predictable than the arterial roads, particularly in the interpeak period.
- To some extent, the weekend congestions on some arterial roads can be as severe as weekdays, and it lasts much longer than the weekday peaks.

- Strategic models assume stable traffic conditions for the time periods modelled, usually four distinct periods. It is found that travel times are not normally distributed in the peak periods.
- Travel time variability is generally proportional to the congestion level. Travel times on the same road but different directions can display very different variability patterns.
- Travel times on an average weekday are statistically different from the average schoolday travel times in the AM peak.

Moving forward, beyond the quality investigation and variability analysis conducted in this paper, we will take advantage of the big GPS data streamed in to further observe the dynamics of the overall traffic patterns. The relationship between congestion and travel time variability will be investigated by road types together with the traffic counts data that BTS collected. The theoretical relationship would be beneficial for travel demand forecasting model development where the validation data sample size is limited. Advanced machine learning techniques, as popularised in the boom of 'Big Data' (Vij and Shankari, 2015), will be applied to analyse such voluminous and fast-changing data collected by GPS devices.

Acknowledgment

The authors would like to thank the BTS spatial team for assistance in pre-processing the GPS data used in this study.

References

Bernardin, V., Trevino, S. and Short, J. B. (2015), Expanding Truck GPS-based Passive Origin-Destination in Iowa and Tennessee, Proceedings of Transportation Research Board 94th Annual Meeting.

Bhaskar, A. and Qu, M. (2015), A Hybrid Model for Motorway Travel Time Estimation-Considering Increased Detector Spacing, Proceedings of Transportation Research Board 93th Annual Meeting.

Carrion, C., Levinson, D. (2013), Valuation of travel time reliability from a GPS-based experimental design, Transportation Research Part C: Emerging Technologies, Volume 35, October 2013, Pages 305-323

Hays, W.L. (1981). *Statistics* (3rd. Ed.). New York: Holt, Rinehart & Winston. (Chapter 6, "Normal Population and Sampling Distributions").

Hidas P and Milthorpe F (2009), Traffic Counts for Strategic Transport Model Validation: What Counts? Proceedings of the 32rd Australasian Transport Research Forum (ATRF), Auckland: ATRF. (http://www.bts.nsw.gov.au/ArticleDocuments/82/CP2009-5-Traffic-Counts-for-Strategic-Transport-Model-Validation.pdf.aspx, accessed in July, 2015)

Li, R., Rose, G., and Sarvi, M. (2006). Using automatic vehicle identification data to gain insight into travel time variability and its causes. Transportation Research Record, 1945, 24–32.

Lu, Y. and Li, S. (2014), An Empirical Study of With-in Day OD Prediction Using Taxi GPS Data in Singapore, Proceedings of Transportation Research Board 93th Annual Meeting.

Flaskou, M., Dulebenets, M.A., Golias, M.M., Mishra, S. and Rock, R. M. (2015), Analysis of freight corridors using truck GPS data, Proceedings of Transportation Research Board 94th Annual Meeting.

Mazloumi, E., Currie, G., and Rose, G. (2010). "Using GPS Data to Gain Insight into Public Transport Travel Time Variability." Journal of Transportation Engineering, 136(7), 623–631.

Shen, L., and Stopher, P. R. (2014), Review of GPS travel survey and GPS data-processing methods, Transport Reviews, Vol. 34, No. 3, 316-334.

Sohn, K. and Kim, D. (2009). Statistical Model for forecasting link travel time variability. Journal of Transportation Engineering, 135(7), 440–453.

Susilawati, S., Taylor, M. and Somenahalli, S. (2011), Distributions of travel time variability on urban roads J. Adv. Transp. 2013; 47:720–736,

Uno, N., Kurauchi, F., Tamura, H. & Iida, Y. (2009) Using Bus Probe Data for Analysis of Travel Time Variability, *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 13:1, 2-15

Van Lint, J.W. C. (2004). Quantifying uncertainty in real-time neural network based freeway travel prediction. In Proceedings of 83rd Transportation Research Board Annual Meeting.

Vij, A. and Shankari, K. (2015), When is Big Data Big Enough? Implications of Using GPS-Based Surveys for Travel Demand Analysis, Technical Report No. UCB/EECS-2014-141 (http://www.eecs.berkeley.edu/Pubs/TechRpts/2014/EECS-2014-141.html)