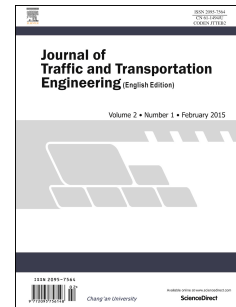


# Accepted Manuscript

"Weather" transit is reliable? Using AVL data to explore tram performance in Melbourne, Australia

Mahmoud Mesbah, PhD, Lecturer, Johnny Lin, Graham Currie



PII: S2095-7564(15)00023-9

DOI: [10.1016/j.jtte.2015.03.001](https://doi.org/10.1016/j.jtte.2015.03.001)

Reference: JTTE 13

To appear in: *Journal of Traffic and Transportation Engineering (English Edition)*

Please cite this article as: Mesbah, M., Lin, J., Currie, G., "Weather" transit is reliable? Using AVL data to explore tram performance in Melbourne, Australia, *Journal of Traffic and Transportation Engineering (English Edition)* (2015), doi: 10.1016/j.jtte.2015.03.001.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Original research paper

# "Weather" transit is reliable? Using AVL data to explore tram performance in Melbourne, Australia

Mahmoud Mesbah<sup>a,\*</sup>, Johnny Lin<sup>a</sup>, Graham Currie<sup>b</sup>

<sup>a</sup> School of Civil Engineering, The University of Queensland, Brisbane, QLD 4072, Australia

<sup>b</sup> Department of Civil Engineering, Monash University, Melbourne, VIC 3800, Australia

## Abstract

This paper uses automatic vehicle location (AVL) records to investigate the effect of weather conditions on the travel time reliability of on-road rail transit, through a case study of the Melbourne streetcar (tram) network. The datasets available were an extensive historical AVL dataset as well as weather observations. The sample size used in the analysis included all trips made over a period of five years (2006–2010 inclusive), during the morning peak (7 am–9 am) for fifteen randomly selected radial tram routes, all traveling to the Melbourne CBD. Ordinary least square (OLS) regression analysis was conducted to create a linear model, with tram travel time being the dependent variable. An alternative formulation of the model is also compared. Travel time was regressed on various weather effects including precipitation, air temperature, sea level pressure and wind speed; as well as indicator variables for weekends, public holidays and route numbers to investigate a correlation between weather condition and the on-time performance of the trams. The results indicate that only precipitation and air temperature are significant in their effect on tram travel time. The model demonstrates that on average, an additional millimeter of precipitation during the peak period adversely affects the average travel time during that period by approximately 8 s, that is, rainfall tends to increase the travel time. The effect of air temperature is less intuitive, with the model indicating that trams adhere more closely to schedule when the temperature is different in absolute terms to the mean operating conditions (taken as 15 °C).

## Keywords

automatic vehicle location; transit performance; weather condition; regression analysis

\*Corresponding author. Tel.: +61-7-3365-1569; fax: +61-7-3365-4599.

E-mail address: mahmoud.mesbah@uq.edu.au (Mahmoud Mesbah, PhD, Lecturer).

30

31 **1 Introduction**

32 An ever-growing population and an increasing push toward sustainability are among the major causes for a need to  
33 better understand the public transport systems. Accordingly, the quality of public transport and transit services, as  
34 determined by associated performance measures, has been subject to extensive research (Berkow et al., 2009;  
35 Böcker et al., 2013). In particular, much focus has been placed on transit operational performance and the key  
36 drivers which impact service reliability (Carrion and Levinson, 2012). This paper aims to determine and quantify the  
37 effect that weather conditions may have on the travel time reliability of on-road transit at a network level. A large  
38 scale dataset of the Melbourne streetcar (tram) network is used to generate reliable results over an extended period  
39 of five years.

40

41 Although it is, for now, still impossible to control the weather, the research has been conducted in the  
42 understanding that recognition of weather impacts on transit performances will provide authorities with useful  
43 knowledge when considering improvements to or investments in transit infrastructure (Guo et al., 2007). Additionally,  
44 it may also allow these authorities to further optimize or create robust travel schedules that are considerate of  
45 weather effects.

46

47 **2 Research context**

48 Melbourne is the only Australian city with a major streetcar (tram) network. The iconic trams provide significant

49 coverage as part of the Melbourne public transport system. The network is the largest and the oldest of its kind in the  
 50 world which has over 250 km of double track, 1763 tram stops, and serviced a total of 182.7 million passenger trips  
 51 in the 2010-2011 (Yarra Trams, 2012). Fig. 1 shows the tram network and its catchment area.

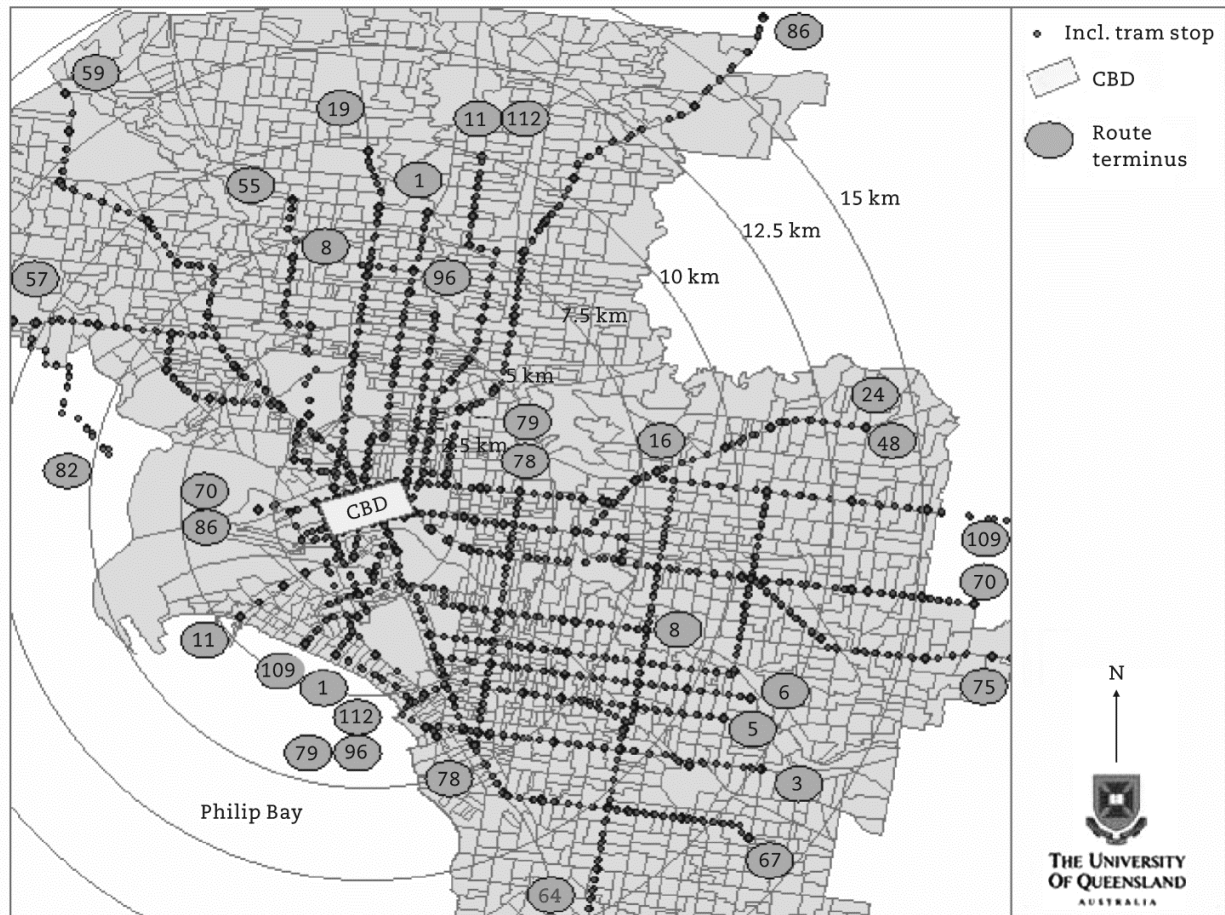


Fig. 1 Melbourne tram network

54 The ability to benchmark the network's performance stems from the operator's use of automatic vehicle location  
 55 (AVL) technology. Automatic vehicle monitoring has been in use since 1985 (Yarra Trams, 2012), with timing points  
 56 along each tram route providing trip data regarding schedule adherence. Previous research suggests that despite  
 57 their significance, Melbourne trams experience relatively poor reliability. Currie and Shalaby (2007) found that

approximately 27.4% of all services were 'not on time'; defined as being more than 1 min early or more than 6 min late. Public transport victoria (PTV) dataset shows that on average, over the previous five quarters (April 2011 to June 2012), operators have delivered 99.12% of all scheduled tram services, with 28.62% of these services being 'not on time' (defined by PTV as being more than 1 min early or more than 5 min late).

As travel time variability is an important operational metric, it is imperative to determine its underlying drivers. A comprehensive historical AVL dataset with recorded scheduled arrival times, actual arrival times, times of departure and more was obtained from the operator for analysis. The dataset includes all services across all routes, 24 hours a day, seven days a week for a period of ten years since March 2001. This data, along with coincident recorded weather observations obtained from the Australian Government Bureau of Meteorology (BoM) from three weather stations in the Greater Melbourne Region, have been used to benchmark performance and investigate any correlation between weather effects and travel time reliability. This study demonstrates the power of analyzing an extremely large intelligent transport system and investigates the temporal and spatial consistency of the intervening weather parameters.

### **3 Research background**

A majority of existing literature on travel time reliability has been focused on road users, such as private passenger vehicles or bus networks. However, given that approximately 80% of trams in the Melbourne streetcar network share

76 road space with car users (Yarra Trams, 2012), this literature is still relevant in determining the factors which cause  
 77 travel time variability. Further, research has suggested that some of these factors such as congestion effects may  
 78 have an even higher impact on tram reliability than road traffic (Currie and Shalaby, 2007).

### 79 3.1 Operational performance measures

80 Rietveld et al. (2001) attest that reliability (or unreliability) of scheduled services can be analyzed in terms of the  
 81 distribution of departure times  $T_D$ , travel times  $T_R$  and arrival times  $T_A$ , where, obviously

$$82 \quad T_A = T_D + T_R \quad (1)$$

83  
 84 For consistency, this study will adopt the measure of travel time reliability when referring to 'reliability'. The  
 85 scheduled travel time is to be defined as the difference between arrival times between each station or timing point.  
 86 That is, any dwell time at timing points are incorporated as part of the travel time.

87  
 88 Utilizing the distribution approach, the standard deviation of actual running times has been previously prescribed  
 89 for urban bus performance (Hofmann and O'Mahony, 2005). Another measure presented in the literature is the  
 90 coefficient of variation (CV), expressed in Eq. (2)

$$91 \quad CV = \frac{SD}{\bar{x}} \quad (2)$$

where SD represents the standard deviation;  $\bar{x}$  represents the mean travel time.

92

93 CV is often favored and beneficial due to its dimensionless quality. This allows for a comparison of different  
94 datasets to represent the variation between them (Washington et al., 2003). Currie and Mesbah (2011) used this  
95 approach in their work on visualizing GIS and AVL data to explore transit performance; citing its utility in allowing for  
96 comparing route sections of alternative length.

97

98 Further metrics used in measuring travel time reliability and variability include quantile range (such as the  
99 difference between the 95<sup>th</sup> or 90<sup>th</sup> and 50<sup>th</sup> quantile), buffer time (defined as the additional time above average travel  
100 time required for on-time arrival) as well as planning index (Federal Highway Administration, 2009).

101

102 Authors who use the approaches outlined above tend to be focused on measuring transit reliability, or drawing  
103 comparisons between the reliability of two or more different services. Research in this area generally uses this  
104 information as a means to model or simulate travel time on transit routes, determine the effects travel time reliability  
105 has on passenger travel demand, or evaluate travel time reliability itself (Richardson and Taylor, 1978; Tseng,  
106 2008). However, there is a lack of long term empirical evidence on the underlying causes of travel time reliability.

107

108 The focus of this paper is on determining the significance of weather effects on travel time. As such, travel time  
109 variability is measured by using the difference between actual travel time and scheduled travel time where it is

available. Kwon et al. (2011) used this approach in decomposing travel time reliability into various sources. Aside from its simplicity, the main benefit of using this methodology is that under basic regression analysis, the result is clearly intuitive—some linear combination of factors results in a service being either late or early compared to its scheduled operations.

### 3.2. Quantifying drivers and weather effects

It is of particular importance to determine the sources and drivers of travel time variability so that they can be addressed by policymakers. This has been explored in previous research, with Cambridge Systematics and Kwon et al. (2011) providing seven such sources (or drivers) of travel variability, traffic incidents, work zone activity, weather, fluctuations in demand, special events, traffic-control devices, and inadequate base capacity.

Kwon et al. (2011) further present a statistical framework which fits travel time variability, such as the 95<sup>th</sup> percentile, on individual sources using empirical data through a linear regression model. The model minimizes the sum of observed error terms, and the technique can be extended to identify the contribution of each factor to the variability of travel time in the form of an 'unreliability pie'. Kwon et al. (2011) found that during the morning period, the majority of unreliability was a result of congestion. On average, weather contributed up to 5% to travel time unreliability.



128 Mazloui et al. (2010) also used a minimum least-squares linear regression technique which allowed for the  
129 determination of the statistical significance of independent variables. A backward stepwise selection method was  
130 used to select the significant variables. The model regressed the log-transformed standard deviation of travel as the  
131 dependent variable, and results indicated that rainfall was only a significant factor during the AM (morning) peak,  
132 where congestion was a factor. In other periods, the rainfall effect was not significant.

134 A basic extension of the standard linear regression model is the inclusion of dummy variables. Whereas a standard  
135 linear regression model considers the extent of variability at a continuous scale of variable dependence, dummy  
136 variables (also known as indicator variables) are discrete in nature and can capture impacts of the presence (or  
137 absence) of some categorical effect. A dummy variable for 'heavy rain' was implemented by Guo et al. (2007) to  
138 capture the effect of significant precipitation. The authors' decision to include such a variable was intuitive.

140 In terms of determining the drivers of travel variability, it is important to distinguish the differences in models which  
141 predict travel time, and models that predict travel time deviation. Those models travel time deviation may experience  
142 extremely poor  $R^2$  values. This is to be expected, as travel time deviation is often measured in comparatively large  
143 increments (even 1 s is significantly different to 5 s), and is, intuitively, a function of many unobservable unknowns.  
144 Past run time deviation studies have provided models with  $R^2$  values as low as 0.07. Despite this, the causes of  
145 deviation in run time along the route (drivers of travel time variability) and their significance have still been considered

admissible (El-Geneidy et al., 2011).

Existing research generally concludes that adverse weather conditions do have significant effects on ridership (Arana et al., 2014) and performance measures for transit (Hofmann and O'Mahony, 2005). Even for instances where weather was found to contribute relatively little to travel time unreliability, the weather factor terms were still statistically significant (Kwon et al., 2011). The intention of this study is to further extend these studies by considering a large scale case study of the Melbourne tramway network. A spatially and temporally large network was investigated including 15 tram routes over a period of five years. Therefore the outcomes of the models should be reliable for long term planning and operation.

#### **4 Data preparation**

The Melbourne tram network utilizes an AVL system to record data such as scheduled and actual time of arrival (and departure) at various timing points for all services across all routes. Data was extracted for fifteen randomly selected routes (ten routes for calibration (Routes 1, 3, 8, 19, 59, 64, 86, 96, 109 and 112) and five for validation (Routes 5, 48, 67, 72 and 75)) for a period of five years, from Jan. 2006 to Dec. 2010 (inclusive). The only condition for selection was that routes were required to run radially to the Melbourne CBD such that any congestion or transfer effects would be present across all routes. More details on the AVL data of Melbourne trams were available in Mesbah et al. (2012).

Under the selected formatting options, the extracted data contained information including: date and time, route number, timing point, scheduled time of arrival, recorded time of arrival, and recorded time of departure. The morning peak (7 am-9 am) was focused in this study.

#### 4.1 AVL data cleansing

The following steps were performed to clean the data:

- split the data records by direction of travel (upstream and downstream direction);
- remove trips that did not start and complete strictly within the peak period;
- remove “non-normal” trips—trips flagged as short start/end, cancelled, or non-AVM;
- remove incomplete entries—trips which did not have records for all timing points (despite being flagged as normal);
- calculate the travel time (both actual and scheduled) between each timing point; where travel time has been defined as time of arrival at the previous point to time of arrival at the current point.

##### 4.1.1 Missing data correction

The filtering process above ensured consistency and allowed dependable temporal comparison across years and routes. The percentage of entries filtered due missing data (averaged across both directions of travel) is shown in Table 1.

Table 1 Percentage of missing data by route

Route	Percentage of missing data(%)					Average value(%)
	2006	2007	2008	2009	2010	
1	0.88	1.71	3.46	46.18	1.49	10.74
3	1.55	3.27	3.63	9.96	0.94	3.87
8	1.26	1.16	3.58	1.26	1.11	1.67
19	0.86	0.50	0.21	0.08	0.37	0.40
59	1.10	1.02	0.72	0.20	0.25	0.66
64	1.31	4.07	3.50	2.33	1.72	2.59
86	3.24	2.45	2.55	1.44	3.95	2.73
96	1.22	1.82	4.59	42.5	0.65	10.16
109	1.02	0.44	0.50	30.2	0.6	6.55
112	0.61	1.73	0.40	0.19	1.00	0.79

As seen in Table 1, the amount of data filtered is generally low, at about 3% with a few exceptions. It should be noted that the instances where the amount of data discarded was more than 10% are related to the restructure of the routes and the timing points. Incomplete entries were marked as such if the number of timing points with data for that record was less than the maximum number of timing points recorded along that route for a single trip during that calendar year. For example, for Route 1 in 2009, the number of timing points decreased from 17 to 16 in July. As a result, all records from that date onwards were discarded as being incomplete. Similar restructuring happened for Routes 96 and 109 in 2009.

#### 4.1.2 Travel time aggregation

Using the remaining 'complete' records, a representative total travel time was calculated. This was obtained by aggregating travel times (in seconds) between timing points from the second to second-last points in each direction.

Fig. 2 demonstrates this process for an example with Route 19. The timing points are represented as a four letter

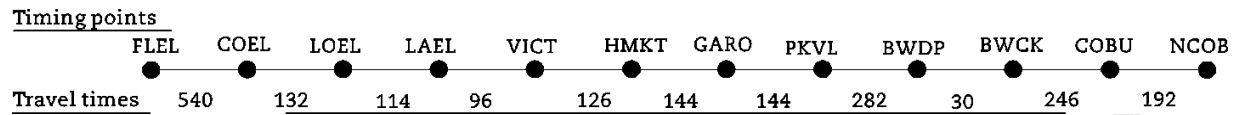


Fig. 2 Travel time aggregation

The decision to aggregate only the second to second-last timing points was due to the fact that travel time at the first and last stops (based on arrival-arrival times) were consistently significantly longer than scheduled. This skew is likely to be a result of drivers arriving at the stops early to ensure that they would be ready to commence the service (or following service) as scheduled. This aggregation process was similarly applied to scheduled times to obtain a scheduled total travel time.

#### 4.1.3 Outlier removal

Once representative travel times were calculated for all trips, any trips which were more than 3 standard deviations from the mean for their respective route and year were deemed to be outliers and removed from the dataset. This is in accordance with the three sigma rule (Nikulin, 2011), which assumes that for a normal distribution, nearly all values lie within three standard deviations for the mean. The percentage of entries filtered as outliers is shown in Table 2.

Table 2 Percentage of outliers by route

Route	Percentage of outliers(%)					Average value(%)
	2006	2007	2008	2009	2010	
1	1.06	0.84	1.13	0.80	1.07	0.98
3	0.72	0.78	0.67	1.08	0.56	0.76
8	0.97	0.90	0.92	0.83	1.32	0.99
19	0.96	0.81	0.69	0.67	0.79	0.78
59	1.24	1.13	1.24	0.95	1.39	1.19
64	0.65	0.66	0.84	0.81	0.66	0.72
86	1.10	0.89	1.23	1.31	1.16	1.14
96	1.00	1.01	0.69	0.89	0.62	0.84
109	1.33	0.98	0.95	1.26	1.40	1.18
112	0.96	0.83	0.92	1.21	1.36	1.06

Even though different routes have a different number of trips run during the same period, the number of outliers identified was fairly consistent across routes. This is expected, as while shorter routes can be potentially run with higher frequency, the number of 'problem trips' (such as cases of congestion caused by road accidents, operator or track failure) would remain consistent on a percentage basis. Given that the average percentage of trips discarded due to the outlier criterion was at around 1%, the three sigma rule was deemed appropriate. Although the trip data was not strictly normally distributed, a normal distribution is considered to be an adequate assumption. At this point, the necessary AVL data preparation was completed, and needed to be consolidated with weather data from the Australian Government Bureau of Meteorology prior to analysis.

#### 4.2 Weather observation data

The Australian Government Bureau of Meteorology has approximately twenty weather observation stations within the Greater Melbourne Region; however, the data quality of these stations varies significantly. Stations used in this study are:

- Melbourne Airport (Station 086282);
- Moorabbin Airport (Station 086077);
- Melbourne CBD Office (Station 086071).

Many weather variables were available for each minute of the day. To aggregate the observations across the three

stations, the arithmetic average of weather conditions at 8 am was taken into account to represent the morning peak (7 am-9 am). The selected methodology was used to determine data for air temperature, wet bulb temperature, dew point temperature, relative humidity (percentage), wind speed, maximum wind gust in last 10 min, and mean sea level pressure.

Precipitation was determined by aggregating by-the-minute precipitation for the two hours of between 7 am and 9 am at each station, and then averaging across the three stations to determine an average (total) precipitation value. Where data was missing at any minute during the period, the entire record was marked as invalid for that date at that station. Due to this stringent requirement, a significant amount of data was inadmissible. The amount of invalid data discarded under these requirements for each weather station is shown in Table 3.

Table 3 Days of precipitation and percentage of data discarded

Year	Data discarded(%)			Days with precipitation	Days without valid data
	Melbourne Airport	Moorabbin Airport	Melbourne CBD Office		
2006	3.56	73.42	80.56	40	7
2007	2.47	54.79	78.90	53	3
2008	2.19	31.42	67.49	54	2
2009	1.64	46.03	81.92	48	4
2010	13.42	12.60	33.97	70	2

Table 3 shows the variability in data quality from different weather stations. The amount of precipitation was taken as the average across valid records when valid data was available. Days with no data were excluded from the analysis (Table 3).

## 5 Econometric framework and analysis

An ordinary least square (OLS) linear regression is adopted in this study which is consistent with previous methods in the literature (Guo et al., 2007; Mazloumi et al., 2010). Linear regression has the advantage of being computationally non-intensive, as well as producing results which are intuitive in interpretation.

A backward stepwise selection method is used to select the significant variables; starting by including all variables and then eliminating insignificant variables from the model using an iterative process. The available relevant variables which are candidates for inclusion are:

- observed run time (the regressand);
- scheduled travel time;
- direction of travel (dummy variable);
- route identifier (dummy variable);
- year identifier (dummy variable);
- day identifier (dummy variable);
- public holiday identifier (dummy variable);
- school holiday identifier (dummy variable);
- air temperature;
- dew point temperature;
- wet bulb temperature;



- humidity;
- wind speed;
- maximum 10 min wind gust;
- wind direction;
- mean sea level pressure

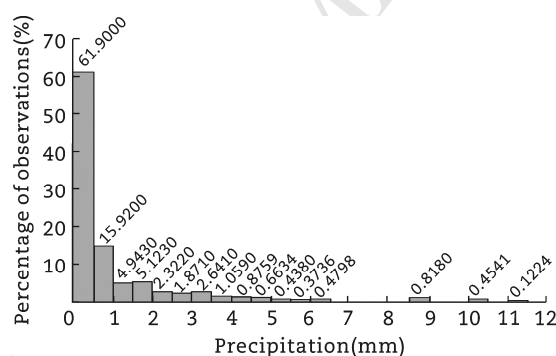
Since air temperature, dew point temperature and wet bulb temperature are closely correlated, only air temperature was used for simplicity. Using the air temperature observation directly would not have a clear interpretation in this regression—*ceteris paribus* (A Latin phrase commonly used in economics; literally translated as “all other variables held constant” (Schlicht, 1985). In this case, the assumption rules out all effects other than that of temperature), and the model would be predicting travel time based on a temperature of 0 °C.

A more useful variable would be the absolute difference in temperature from a ‘standard operating condition’. Services are expected to operate ‘normally’ (on schedule) under such conditions, and be either delayed or early under adverse weather. The average or expected temperature that people experience during a year is assumed as this ‘standard operating condition’. The mean temperature across all observations was found to be 13 °C (This is the result of a direct arithmetic average across the years 2006-2010 (inclusive), and does not consider seasonal effects). With this in mind, the ‘standard operating temperature’ has been defined as 15 °C, which is the expected temperature of ‘cool morning weather’ in Melbourne.

Wind speed is used as the wind descriptor while the 10 min wind gust variable was dropped for simplicity. Wind direction was excluded from the model, as the direction was measured in terms of true north, rather than direction of travel and different routes travelling in different directions. Similarly, the direction of travel identifier was also considered to be inadequate and potentially spurious if included in the model.

287

288 In determining precipitation effects, consideration was made as to whether or not significant precipitation (i.e. heavy  
 289 rain) had the same effect as mild precipitation. For this purpose, a 'heavy rain' dummy variable is utilized. The key  
 290 issue is determining what level of precipitation constitutes 'heavy'. Guo et al. (2007) considered the 80<sup>th</sup> percentile for  
 291 all rainy days as 'heavy'. However, the amount of rainfall observed during the AM period was small both in magnitude  
 292 and range. Fig. 3 shows the days where precipitation was greater than zero. The data Included only observations  
 293 where precipitation was greater than zero during the peak period. The maximum and average rainfall recorded were  
 294 11.4 mm and 0.91 mm over the two hours, respectively. Over 60% of non-zero precipitation observations did not  
 295 exceed 0.5 mm. Given the small range of precipitation, no dummy variable was defined for 'heavy rain'.



296

297 **Fig. 3 Percentage distribution of precipitation observations**

298 An initial model is developed using all the remaining variables as listed in Table 4. It is found that at the five  
 299 percent level, the wind speed and humidity were not significant. As such, these variables can be removed and the  
 300 model re-estimated. 'Saturday', 'Route 19', and 'Year 2006' dummies were omitted (based on lowest significance in  
 301 their respective variable class) due to collinearity in order to satisfy the Gauss-Markov assumptions (Wooldridge,

2009).

Table 4 Backward stepwise variable selection

Regressor	Initial model	First pass	Second pass
Scheduled travel time	Y	Y	Y
Precipitation	Y	Y	Y
Public holiday dummy	Y	Y	Y
School holiday dummy	Y	Y	Y
Weekday dummy	Y	Y	Y
Saturday dummy	Y	-	-
Sunday dummy	Y	Y	Y
Wind speed	Y	N	-
Abs. temperature difference	Y	Y	Y
Humidity	Y	N	-
Pressure	Y	Y	N
Route 1 dummy	Y	Y	-
Route 3 dummy	Y	Y	-
Route 8 dummy	Y	Y	-
Route 19 dummy	Y	-	-
Route 59 dummy	Y	Y	-
Route 64 dummy	Y	Y	-
Route 86 dummy	Y	Y	-
Route 96 dummy	Y	Y	-
Route 109 dummy	Y	Y	-
Route 112 dummy	Y	Y	-
Year 2006 dummy	Y	-	-
Year 2007 dummy	Y	Y	-
Year 2008 dummy	Y	Y	-
Year 2009 dummy	Y	Y	-
Year 2010 dummy	Y	Y	-
Constant term	Y	Y	Y
Regression adjusted $R^2$	0.956	0.956	0.954

Note: Y: significant and included; N : not significant and removed; - : excluded

To further generalize the model, route and year dummies were stripped from the model. Removal of the route and year dummies meant that it would be able to examine the effects of the regressors without this information. That is, the resulting model could also be generalized for routes or years outside of the sample. Removal of these variables also did not significantly decrease the explanatory power of the model, with the adjusted  $R^2$  value (The adjusted  $R^2$  was preferred as an indicator of explanatory power over the unadjusted  $R^2$  value, as  $R^2$  will generally increase when more independent variables are added irrespective of the new variables' explanatory power. Adjusted  $R^2$  adjusts for the number of variables, imposing a penalty for superfluous regression Quantitative Micro Software (2009)) only

changing from 0.9556 to 0.9537. The size of the  $F$  value (also referred to as the  $F$ -statistic) also remained large, indicative of overall model significance. However, removing these indicator variables did result in the variable for mean sea level pressure being no longer significant at the five percent level and thus, it was also removed. The backward stepwise regression process is shown in Table 4.

In order to ensure that the variables in the selected model were independent, a correlation matrix for all variables was generated. There was no significant correlation between the independent variables above. Variance inflation factors (VIFs) were calculated, as summarized in Table 5. As the VIF does not exceed 10 for any of the variables (that is, tolerance, represented as  $1/\text{VIF}$  is not lower than 0.10), there does not appear to be any multi-collinearity in the model and no further investigation is required.

Table 5 Variance inflation factors

Regressor	VIF	1/VIF
Scheduled travel time	1.09	0.918654
Precipitation	1	0.998257
Public holiday dummy	1.01	0.991561
School holiday dummy	1	0.996016
Weekday dummy	1.41	0.710401
Sunday dummy	1.34	0.748293
Abs. temperature difference	1	0.997077
Mean VIF	1.14	

## 6 Model forms

### 6.1 Primary model

Eq. (3) is obtained from the backward stepwise selection process

$$t_1 = -47.392 + 1.0240t_2 + 7.8316\nu - 0.9566B - 41.413P_1 - 5.9074P_2 + 21.413P_3 + 26.282P_4 \quad (3)$$

where  $t_1$  represents the travel run time;  $t_2$  represents the scheduled travel time;  $\nu$  is the average amount of rainfall

observed during the peak morning period;  $B$  is the absolute difference in temperature from 15 °C at 8 am;  $P_1$  is a dummy variable (1, if the trip is conducted on a public holiday and 0, otherwise);  $P_2$  is a dummy variable (1, if the trip is conducted on a school holiday and 0, otherwise);  $P_3$  is a dummy variable (1, if the trip is on a weekday and 0, otherwise);  $P_4$  is a dummy variable (1, if the trip is conducted on a Sunday and 0, otherwise).

Eq. (3) is estimated using the full combined data set, with a total sample size of 217,782 trips across the ten routes over the five year period during the 7 am to 9 am peak. The adjusted  $R^2$  value was 0.9537 and all parameters were significant at 0.95 level.

Since the model meets OLS assumptions (Table 5), the weather coefficients in Eq. (3) provide useful interpretations. The interpretation of weather coefficients above is that on average, *ceteris paribus*:

- an additional millimeter of rainfall during the peak morning period will increase the actual travel run time by approximately 8 s beyond the scheduled travel time;
- every 1 °C temperature different from 15 °C will decrease travel run time by approximately 1 s less than the scheduled travel time.

The interpretation of the public holiday, school holiday, weekday, and Sunday coefficients is more ambiguous.

Direct interpretation of the coefficients would indicate that compared to Saturday, *ceteris paribus*:

- travel run time is approximately 42 s less on a public holiday;

- travel run time is approximately 6 s less on a school holiday;
- travel run time is approximately 22 s more on a weekday;
- travel run time is approximately 26 s more on a Sunday.

## 6.2 Alternative models

### 6.2.1 ALT1-weather variables

Given that the focus of this study is on the effects of weather on travel time, ALT1 model is developed by excluding the day type dummy variables:  $P_1$ ,  $P_3$ , and  $P_4$ . Eq. (4) is obtained using the full dataset with a total sample size of 217,782 trips across all specified routes and years.

$$t_1 = -40.2351 + 1.0264t_2 + 7.8115v - 0.8847B \quad (4)$$

All the variables are significant at the five percent level. Compared to Eq. (3), the new model with only the scheduled travel time and weather-specific variables exhibits:

- marginally lower adjusted  $R^2$  value, at 0.9535 compared to 0.9537;
- same precipitation and temperature effects when rounded to the nearest second.

To investigate whether the model was robust regardless of year and route, Eq. (4) was also re-estimated using year and route specific data only.

### 6.2.2 ALT2-year specific

The model was estimated individually by year for each of the five years from 2006 to 2010. Results are summarized

in Table 6. A graphical representation of the precipitation and temperature coefficients has been plotted in Fig. 4 below.

Estimation by year shows that the model remained significant in explanatory power irrespective of year, with an adjusted  $R^2$  value of more than 0.95. The precipitation and temperature variables are statistically significant in all years based on the  $t$ -student test. The sign of precipitation remains unchanged although the coefficient values vary. However, the absolute temperature coefficient in 2008 is positive while this coefficient is negative in all other models. This exception shows that the temperature effects should be used with caution. Identification of such an exception is only possible when analyzing a large dataset of several years which is one of the advantages of this study. Furthermore, the precipitation coefficient varies considerably from one year to another. This is evidence that models developed just based on one short period of time are not reliable as the coefficient may vary considerably.

Table 6 Year-specific model estimation

Model coefficient	Year				
	2006	2007	2008	2009	2010
$t_2$	1.0205	1.0319	1.0305	1.0303	1.0182
$\nu$	10.0361	4.2751	3.9870	11.4031	11.1831
$B$	-2.2814	-1.3390	0.9125	-0.9337	-0.6538
Constant	-36.107	-59.122	-49.008	-47.921	-8.266
Model details					
Observation No.	43,165	46,157	45,495	39,206	43,759
Adjusted $R^2$	0.9573	0.9519	0.9507	0.9553	0.9543
$F$ statistic	3.2e+05	3.0e+05	2.9e+05	2.8e+05	3.0e+05

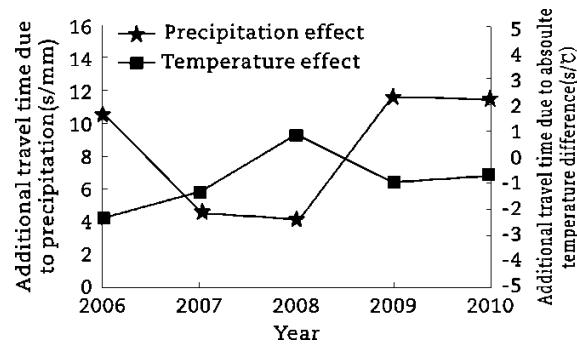


Fig. 4 Coefficient plot for year-specific models

### 6.2.3 ALT3-route specific

The model was estimated individually by route for each of the ten routes across all five years. Model estimates are shown in Table 7. A graphical representation of the precipitation and temperature coefficients has been plotted in Fig. 5.

The delay suggested by the coefficient of the precipitation term is consistent in sign, and ranges from approximately 4 s/mm of rainfall for Route 96 to 13 s/mm of rainfall for Route 59. It should be noted that the adjusted  $R^2$  values for the route-specific models were consistently lower than explanatory power of the aggregated full sample model. This is to be expected due to the significantly smaller sample sizes, and not of concern as this paper is interested in the effect of weather, rather than actually modeling expected travel run time (El-Geneidy et al., 2011).

In Table 7, the absolute temperature differential term (the effect that a one degree difference from 15 °C has on travel run time) is negative for all routes. However, it is not significant at the five percent level for four out of the ten routes. This indicates that the temperature effect is marginal, and removal of the variable of the term from the model

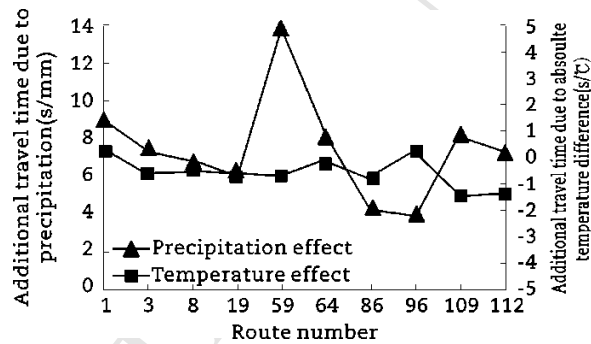


may be warranted.

**Table 7 Route-specific model estimation**

Model coefficient	Route									
	1	3	8	19	59	64	86	96	109	112
$t_2$	1.0115	0.9979	1.0569	0.8022	1.0960	0.9910	1.0950	0.9133	0.9864	1.0207
$v$	9.0763	7.4318	6.7552	6.0507	13.8201	8.1939	4.2924	3.7923	8.1406	7.1984
$B$	-0.0589*	-0.9720*	-0.8327*	-0.8674	-1.0170	-0.4622*	-1.0952	-0.0904*	-1.6907	-1.6005
Constant	-20.8050	60.2771	-154.4644	387.6638	-185.8559	36.5480	-314.1944	281.1045	98.5102	-17.1113*
Model Details										
Sample No.	19,674	12,502	20,608	40,425	30,614	13,980	16,519	22,724	18,819	21,917
Percentage of obs. (%)	9.03	5.74	9.46	18.56	14.06	6.42	7.59	10.43	8.64	10.06
Adjusted $R^2$	0.7668	0.7920	0.7990	0.6822	0.6994	0.8739	0.8560	0.5099	0.8909	0.8631
$F$ statistic	21,560	15,866	27,302	28,922	23,742	32,282	32,723	7882	51,235	46,077

Note: \*Not significant at the five percent level.



**Fig. 5 Coefficient plot for route-specific models**

#### 6.2.4 ALT4-schedule deviation

An alternate model specification which provides increased emphasis on travel time variability is to model the

difference between scheduled travel time and actual travel time directly. This model is expressed in Eq.(5) below

$$t_3 = 8.182v - 0.498B - 50.00P_1 - 2.786P_2 + 42.65P_3 + 32.81P_4 \quad (5)$$

where  $t_3$  represents the difference between travel run time and scheduled travel time.

402

403 In this case, the constant term has been removed. Theory prescribes that the expected value for the  $t_3$  term should  
 404 be zero—schedules are, obviously, designed to be adhered to (Eisenhauer, 2003). ‘Sunday’, ‘weekday’ and ‘public  
 405 holiday’ dummy variables have been maintained to observe schedule adherence on these days. ‘Saturday’ was  
 406 omitted to avoid collinearity. All variables are significant at the five percent level based on a  $t$ -student statistic.

407

408 The model is significant based on an  $F$ -test with 6 degrees of freedom. However, the adjusted  $R^2$  term in this  
 409 model is low, at around 0.05. This is in line with calculated  $R^2$  values in the literature which represent deviation (or  
 410 difference) of run time, rather than run time itself (El-Geneidy et al., 2011). Again, since this paper is focused on the  
 411 weather causes of deviation from run time, the low adjusted  $R^2$  value is not an issue of concern. For more discussion  
 412 on  $R^2$  see the research of Richardson and Taylor (1978).

413

414 The results from the ALT4 model indicate that schedule adherence is generally better on Saturdays compared to  
 415 weekdays or Sundays. Ceteris paribus, services on public and school holidays tend to arrive earlier than scheduled.  
 416 The coefficients for the precipitation and absolute temperature differential terms are very similar to those of the  
 417 primary model (Eq. (3)) and ALT1 model (Eq. (4)). This reconfirms that precipitation and temperature affect travel  
 418 time variability.

### 6.3 *Model interpretation*

In interpreting the effect of weather parameters, a number of reasons may be provided. Since more than 80% of the Melbourne tram network operates in mixed traffic, the interaction of general traffic and trams is paramount. A change in the rainfall or temperature may encourage some passengers to use their personal car which could lead to an increased congestion and therefore higher travel times for both car and tram trips. For example, more passengers may choose to drive if the weather is forecasted to be rainy. On the supply side, lower operational speed during rain and local drain blockages could be among the reasons for a longer tram travel time in an adverse weather. Regarding temperature, the problem for Melbourne is not so much the freezing winter but the very hot summers which acts to limit walking and make short tram trips preferable.

One of the advantages of analyzing such a large data set, with an extremely wide spatial and temporal spread, is to demonstrate how robust and consistent these models are. In Tables 6 and 7, although the value of coefficient for precipitation varies, the results show that precipitation is a statistically significant variable in all years and all routes. However, Table 7 shows that temperature is only significant for some tram routes. This highlights the importance of analyzing a large data set versus models focusing on a route or a limited time period.

## 7 **Model validation**

The fundamental mark of a good model is whether or not it produces useful estimates in prediction. Although the primary focus in this study is the effects of weather, one method of checking the estimation reasonable is to test the

model's overall accuracy in prediction.

The methodology used in this section involves predicting the travel run time for trips which have already been conducted, and then comparing these predicted travel times to the actual travel run time values. If the developed model is reasonable, then the predicted (estimates) should be relatively similar to the actual travel run time values experienced.

#### 7.1 Validation dataset

A separate dataset is extracted for the purpose of testing the validity of the model. These data include five randomly selected routes; one for each year within the five year period. This represents approximately 10% of the data used in the model development (calibration) stage. The extracted data is for:

- Route 5, Year 2010;
- Route 48, Year 2009;
- Route 67, Year 2008;
- Route 72, Year 2007;
- Route 75, Year 2006.

Again, the validation data is randomly selected, with the only criterion for route selection being that it had to run

radially to the Melbourne CBD. This is consistent with the data selection process used for the development of the model. A summary of the validation data is presented in Table 8.

Table 8 Validation data

Validation data	Year				
	2006	2007	2008	2009	2010
Route selected	67	5	72	75	48
Observation number	2955	3162	1866	2920	4241
Percentage of obs. (%)	19.51	20.88	12.32	19.28	28.00

## 7.2 Model validity

All the independent variables including the scheduled travel times for the selected routes as well as coincident weather observations are input into Eq. (4) as a model to predict travel run time. Although this section focuses on Eq. (4), the process for the other model alternatives is similar. The results are plotted against the actual travel run time in Fig. 6 below.

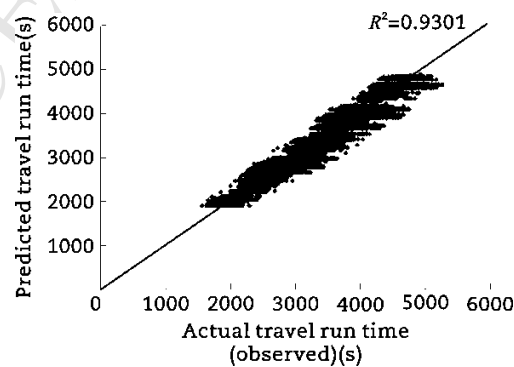


Fig. 6 Actual run time (observed) vs. predicted travel run time

468

469 As seen in the figure, a majority of the data lies near or along the 45 degree line, indicating a reasonable fit for the  
470 model. The  $R^2$  value of fit is estimated to be approximately 0.9301.

471

472 It should be noted that such a close fit with the observed travel times is expected since the predicted travel time  
473 and actual travel time are both heavily based upon the same schedule. The above plot graphically verifies the  
474 intuition that the model is valid.

475

476 The coefficient of variation of the root mean square errors ( $CV_{RMSE}$ ) is another common benchmark for forecast  
477 models, and is calculated using Eq. (6). It measures the accuracy of model predictions by indicating uncertainty in  
478 the model, and is used to compare different forecasting errors. The fact that  $CV_{RMSE}$  is a normalized (scale  
479 independent) measure means that it can be used to compare different models and datasets

480

$$CV_{RMSE} = \frac{\sqrt{E((\hat{\theta} - \theta)^2)}}{\bar{x}} \quad (6)$$

where  $\hat{\theta}$  represents the value predicted by the model;  $\theta$  represents the observed (or actual) value.

481

482 The value of the  $CV_{RMSE}$  for Eq. (4) is found to be 0.044, or approximately 5%. Although there is no consensus on  
483 an acceptable level, a lower value is suggestive of a good fit. Maximum allowable percent deviation generally ranges

from less than 10% to less than 5%.

As another check, the residual plot in Fig. 7 demonstrates that there is a random pattern in the residuals. The existence of a random pattern suggests that the linear model structure fits the data well, and is further evidence for model validity.

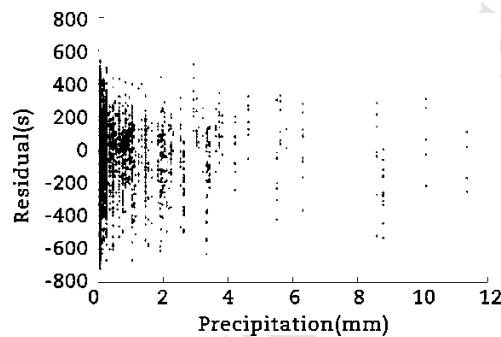


Fig. 7 Residual plot (vs. precipitation)

## 8 Conclusions

The presented paper not only confirms the general understanding that adverse weather conditions affect travel time reliability, but more importantly quantifies this effect on an extensive historical dataset. There has been no study in the literature that tested weather variables in such a wide spatial and temporal range. More than 200,000 trips extended over a period of five years were analyzed. The results demonstrate weather variables which are consistently significant in travel time. The sensitivity of the results over 10 different routes (spatial variation) and 5 years (temporal variation) is presented. Moreover, it has been shown that the value of model coefficients should be

used with caution since they might vary from one route to another or from one year to another. The ranges of these variations are also quantified.

The OLS linear regression models indicate that, of the available weather data, only precipitation and air temperature have a significant effect on travel time reliability. Both the proposed travel run time and deviation of travel run time models provide similar conclusions:

- every additional one millimeter of precipitation during the morning peak period will increase travel time by approximately 8 s for each trip conducted during the period;
- every 1 °C change in air temperature from 15 °C will decrease travel run time by 1 s for each trip during the morning peak period.

The air temperature effect may be considered marginal and excluded when producing a route-specific model, whereas the precipitation effect is present and consistent in all model formulations.

The model proposed in this paper provides a useful interpretation of the effect of weather on transit travel reliability. However, like all regression models, the model is only valid in the range it has been calibrated for. In this case, precipitation experienced over the five year period is heavily skewed toward zero, and did not exceed 11.4 mm in the two hour peak period. Thus, the linear model in its current form may not be suitable to predicting weather



effects for higher or extreme conditions.

The results of this study pave the way for further research in the future. Investigation could be conducted placing different weight on delay and early arrival. Although this paper models travel run time and variability, it does not distinguish the differences in trips being early or trips being late. Furthermore, since passenger mode choice is unlikely to be determined at the time of travel, it may be interesting to conduct an investigation in travel time reliability based on lagged weather effects (such as the precipitation from the previous evening). Also, further research could be developed in how travel schedules can be further optimized to be robust knowing that the weather effects exist.

## Acknowledgments

The authors would like to acknowledge Yarra Trams for providing the data. This research was partially supported by the Australian Research Council (No. DE130100205).

## REFERENCES

- Arana, P., Cabezudo, S., Peñalba, M., 2014. Influence of weather conditions on transit ridership: a statistical study using data from smartcards. *Transportation Research Part A: Policy and Practice* 59, 1-12.
- Berkow, M., El-Geneidy, A. M., Bertini, R. L., et al., 2009. Beyond generating transit performance measures: visualizations and statistical analysis with historical data. *Transportation Research Record*, 2111, 158-168.

- 538 Böcker, L., Dijst, M., Prillwitz, J., 2013. Impact of everyday weather on individual daily travel behaviours in perspective: a  
539 literature review. *Transport Reviews*, 33(1), 71-91.
- 540 Carrion, C., Levinson, D., 2012. Value of travel time reliability: a review of current evidence. *Transportation Research Part A:*  
541 *Policy and Practice* 46(4), 720-741.
- 542 Currie, G., Mesbah, M., 2011. Visualization of geographical information system and automatic vehicle location data to explore  
543 transit performance. *Transportation Research Record*, 2216, 59-66.
- 544 Currie, G., Shalaby, A.S., 2007. Success and challenges in modernizing streetcar systems-experiences in Melbourne, Australia,  
545 and Toronto, Canada. *Transportation Research Record*, 2006, 31-39.
- 546 Eisenhauer, J. G., 2003. Regression through the origin. *Teaching Statistics*, 25(3), 76-80.
- 547 El-Geneidy, A. M., Horning, J., Krizek, K. J., 2011. Analyzing transit service reliability using detailed data from automatic  
548 vehicular locator systems. *Journal of Advanced Transportation*, 45(1), 66-79.
- 549 Federal Highway Administration, 2009. Travel time reliability-making it there on time, all the time. Available:  
550 [http://www.ops.fhwa.dot.gov/publications/tt\\_reliability/TTR\\_Report.htm](http://www.ops.fhwa.dot.gov/publications/tt_reliability/TTR_Report.htm) [Accessed 29 August 2012].
- 551 Guo, Z., Wilson, N. H. M., Rahbee, A., 2007. Impact of weather on transit ridership in Chicago, Illinois. *Transportation Research*  
552 *Record*, 2034, 3-10.
- 553 Hofmann, M., O'mahony, M., 2005. The impact of adverse weather conditions on urban bus performance measures. In:  
554 *Proceedings of 8th International IEEE Conference on Intelligent Transportation Systems*, Vienna, 2005.
- 555 Kwon, J., Barkley, T., Hranac, R., et al., 2011. Decomposition of travel time reliability into various sources-incidents, weather,

work zones, special events, and base capacity. *Transportation Research Record*, 2229, 28-33.

Mazloumi, E., Currie, G., Rose, G., 2010. Using GPS data to gain insight into public transport travel time reliability. *Journal of Transportation Engineering*, 136(7), 623-631.

Mesbah, M., Currie, G., Lennon, C., 2012. Spatial and temporal visualization of transit operations performance data at a network level. *Journal of Transport Geography*, 25, 15-26.

Nikulin, M. S., 2011. Three-sigma rule. Available: [http://www.encyclopediaofmath.org/index.php/Three-sigma\\_rule](http://www.encyclopediaofmath.org/index.php/Three-sigma_rule) [Accessed 2 September 2012].

Quantitative Micro Software, 2009. *EViews 7 User's Guide II*. QMS LLC, Irvine.

Richardson, A. J., Taylor, M. A. P., 1978. Time travel variability on commuter journeys. *High Speed Ground Transportation Journal*, 12(1), 77-99.

Rietveld, P., Bruinsma, F. R., Van Vuuren, D. J., 2001. Coping with unreliability in public transport chains: a case study for Netherlands. *Transportation Research Part A: Policy and Practice*, 35(6), 539-559.

Schlicht, E., 1985. *Isolation and Aggregation in Economics*. Springer-Verlag, Berlin.

Tseng, Y.Y., 2008. *Valuation of Travel Time Reliability in Passenger Transport*. Vrije University, Amsterdam.

Washington, S.P., Karlaftis, M. G., Mannering, F. L., 2003. *Statistical and Econometric Methods for Transportation Data Analysis*. CRC Press, Boca Raton.

Wooldridge, J. M., 2009. *Introductory Econometrics: A Modern Approach-International Student Edition*. South-Western Cengage Learning, Mason.

574 Yarra Trams, 2012. Facts & Figures. Available: <http://www.yarratrams.com.au/about-us/who-we-are/facts-figures/> [Accessed 11  
575 April 2012].

576

577

578

579

580

581

582

583

584

585

586

587

588

589