Australasian Transport Research Forum 2011 Proceedings
28 - 30 September 2011, Adelaide, Australia
Publication website: http://www.patrec.org/atrf.aspx

Modelling urban travel time variability with the Burr regression technique

Susilawati[1], Michael A P Taylor[1], Sekhar Somenahalli[1]

[1]Barbara Hardy Institute, University of South Australia

Email for correspondence: sussy002@mymail.unisa.edu.au

Abstract

The need for more reliable travel time in urban areas in order to provide better transport service to the community has attracted many studies to model travel time reliability and variability. The travel time distribution is basic knowledge for this modelling, and studies to fit continuous parametric distributions to travel time distribution have been conducted since the early 1950s. Two sets of empirical travel time data collected by GPS equipped vehicles in Adelaide indicate that travel time distributions are positively skewed and have long upper tails. The Burr distribution has been found to provide a good fit to the data. Utilising the Burr distribution properties and the Burr regression technique, this paper models the Adelaide urban arterial travel time variability by considering traffic variables such as link length, congestion index and degree of saturation. This study suggests how to fill some current gaps in travel time variability modelling, especially for urban arterial roads. It should also be useful for further travel time variability studies such as the valuation of travel time variability effects.

1. Introduction

Travel time is one of the most important variables in assessing traffic systems performance. Both travellers and traffic engineers would like to have high quality traffic performance including short travel times, as a result, there are many traffic demand management initiatives that have been implemented to achieve this goal. However, traffic systems are complex and have a stochastic nature, for example short and long incidents can disrupt performance and lead to high travel time. On a daily basis on urban arterial roads, the fluctuation of demand and supply will also greatly affect the traffic performance in general and the travel time in particular.

Given the need for more reliable travel time in urban areas in order to provide better transport service to the community, the variability of travel time by time of day, day of week or even due to the seasonal factors has attracted many studies. Those studies have modelled and assessed the travel time variability both in urban arterial roads and freeway by fitting continuous parametric distributions such as Normal and Log Normal distribution to travel time distribution (Faouzi and Maurin, 2007). Having the best fit of distribution, the variability of travel time can be measured by using its properties such as the standard deviation and the coefficient of variation.

Given the rapid development of traffic data collection techniques, the recent studies not only model the travel time variability but also the relationship between the travel time variability metrics, the traffic parameters such as link length, traffic flow, capacity, congestion index and also the occurrence of incidents using the regression technique (Li et al., 2006, Eliasson, 2006, 2007). Those studies have significantly increased the insight about the nature of the travel time and the possible factors that might influence its variability.

Travel time variability metrics that measure the dispersion of the actual travel time from the preferred travel time data have been introduced. However, there is a subjective judgment in selecting the preferred travel time. Some studies that assume travel time variability data are symmetric have used the mean travel time as the preferred travel time and the standard deviation as the travel time variability metric. Other studies have suggested that travel time data are skewed and have a long upper tail, have tended to use the median (50^{th} percentile) travel time and difference between the 90^{th} percentile and the 50^{th} percentile (quantile range) as preferred travel time and its variability respectively (Lint and Zuylen, 2006).

Most previous studies assumed that the Normal and Log Normal distributions are the best fit distribution to the travel time data, even though recent travel time studies have suggested that the travel time variability distributions are more likely to have long tails and very positive skew. Susilawati et al. (2010) confirmed that the Burr distribution has been found to provide a good fit to the empirical travel time data collected at two urban arterial roads in Adelaide.

The burr distribution has two parameter, c and k respectively. It has been widely used in reliability engineering and actuarial science. It was found the c parameter of the Burr distribution give shape of the distribution (the lower value of c the sharper fall of cdf curve). Previous studies also found that the Burr distribution is very sensitive to the c parameter, consequently this parameter plays a very important role in the distribution properties. The closed form and algebraic tractability of this distribution provide flexibility in shape of the pdf. The closed form makes the calculation of its percentiles easy to do. Once the c and k parameters have been estimated, the percentile can be estimated easily as well. Thus, this study investigates the travel time variability of two urban arterial roads in Adelaide by utilising the properties of Burr distribution.

In corresponding to the factors that might affect the travel time variability, it has been shown that the traffic variables play an important role in influencing the travel time variability. Recent studies by Eliasson (2006), Peer et al. (2009) and Black and Chin (2009) developed either linear or non linear travel time variability formulas which were formed by travel time variability metrics including the standard deviation and coefficient of variation and the traffic variables. Those studies suggested that the link length, free flow speed and capacity are the most important factors in travel time variability. Given the consideration that the c parameter shapes the Burr distribution density function and determine the Burr distribution's properties including the mean, median and percentiles this study therefore observe the influence of the traffic variables on the c parameter using the Burr regression technique. The Burr regression technique suggested that the c parameter allows to vary with the explanatory variables (e.g speed, volume, capacity, degree of saturation). This study therefore replaces the c parameter with the exponential function of degree of saturation data and used this function in parameter estimation. It was found that the proposed technique allows to gain insight the influence of degree of saturation associated with the shape of the burr distribution and can be also useful for further data analysis in corresponding to travel time variability and reliability valuing.

The next section reviews the past studies about the travel time variability modelling and how those studies estimated the influence of the traffic variables to the travel time variability. The third section discusses the Burr distribution and its properties and how to maximise these properties to develop the travel time variability metrics. This section also reviews the Burr regression technique and how to incorporate this technique to the travel time variability modelling. The fourth section is about the study area, travel time data collection and briefly discuss about the Sydney Coordinated Adaptive Traffic Systems (SCATS) data extraction. The fifth section is data analysis that covers the utilisation of SCATS data in this modelling and the discussion of the data analysis. The last section provides conclusions and directions for further research.

2. Travel time variability modelling

Travel time is the product of the complex traffic systems varies across time of day and day of week (Oh and Chung, 2006, Emam and Al-Deek, 2006, Zhang et al., 2007). Day by day travel time studies have confirmed that daily travel time generally has two distinct peaks which are morning peak (e.g. between 7 and 9AM) and afternoon peak, perhaps 4PM to 7PM (Zhang et al., 2007). Those studies also proven that Monday and Friday travel time pattern differ to other weekdays travel time pattern. Given the same period of time, the weekend travel time is much different to the weekdays travel time (Emam and Al-Deek, 2006). More interestingly, the holiday season travel time commonly also differs from the non holiday season travel time, suggesting that seasonal factors also have influence.

In recent times there have been several studies about travel time variability, which have quantified not only the variability but also considered the factors that might contribute to travel time variability. The impact of variability on traveller behaviour is also of interest. How variability affects traveller trip management such as departure time, route and mode choices, and how the variations may influence the quality of traffic performance can be explored using state preference (SP) and reveal preference (RP) surveys.

Initial research used the normal distribution properties such as mean and standard deviation to measure travel time variability. To investigate the role of traffic variables on travel time variability, previous studies develop models of the mean and standard deviation as functions of those traffic variables. For this modelling method, multiple linear regressions is the most commonly used technique because it is easy to apply yet can deliver good results. Back in late 1970s, the studies by Herman and Lam (1977) in Detroit, Richardson and Taylor (1978) in Melbourne Australia and Polus (1979) in Michigan can be considered as the pioneer studies on travel time variability. Those studies used similar techniques to collect travel time data and built simple linear regression models of the variability. Herman and Lam (1974) presented statistical analysis for to and from work trip times on 25 routes in Detroit, collected over 20 months. Each trip time was the sum of stop time and cruise time and it was found that the trip time was linear in both stop time and cruise time. This study found that on average from work trip time data are longer than to work trip time data. The to work trip time histogram tended to follow a normal distribution while from work trip time histogram was more like a uniform distribution. However, as this study did not conduct the goodness of fit test, for simplicity purposes, they assumed all the travel time distributions followed the normal distribution. Moreover, a model of the mean, standard deviation and coefficient of variation was been developed using the power and linear regression function. Interestingly, it was found that the linear function gave better results than power function.

With a similar objective to the Herman and Lam's work, Polus (1979) also collected the travel time data in Michigan. The main purpose was to indicate the degree of travel time reliability by observing the variabilities in travel times. Instead of directly using the standard deviation as the travel time variability metrics, he suggested that the travel time reliability function is the inverse of the standard deviation. In this way, the more reliable travel time routes would have a larger value of the reliability index,

Richardson and Taylor (1978) collected and analysed longitudinal travel time data in Melbourne. They assessed the correlations between travel times on successive sections of the study route, and developed relationships between the travel time variability and the level of congestion. They concluded that travel times on a link were independent of those on other links along the route, and that the observed travel time variability might be represented by a lognormal distribution

In parallel to rapid development of data collection techniques and the ongoing growth in computer capability to deal with huge data sets during the last decade, there are number of studies that have looked in more detail at travel time data. The automatic vehicle identification (AVI) technology allows the traffic engineer to record the time when a vehicle passes a start point and the time when the vehicle leaves the study area. The time differences between start and end times can be used as the travel time between these points. In-vehicle GPS systems have transformed travel time collection techniques. GPS equipped taxi and bus fleets have become major sources of travel time data. The travel time data allow the traffic engineer to obtain not only vehicle to vehicle travel time variability but also different routes and different time period travel time variabilities, due to the ability of this technique to cover wide areas in urban road networks. Vehicle to vehicle travel times vary according to the speed limit and driving style. In the urban arterial context, the vehicle to vehicle travel time differ greatly due to the actual times when the studied vehicles pass through the intersection. Vehicles that encounter delay (e.g. at traffic signals) will have higher travel times than those that do not encounter delay.

These types of travel time measurements have been used in recent studies. Eliasson (2007) forecasted the time of day and day to day travel time variability on a Stockholm bypass road by utilising travel time data collected thorough an automatic camera system. The travel time data collected from 6AM to 10 PM, and were then analysed in 15min interval. This study used a nonlinear regression model to seek the relationship between the travel time, free flow travel time, speed and Standard Deviation of vehicle to vehicle travel time data across the day. This study used the following equation based on a volume delay function:

$$\boldsymbol{\sigma = \lambda * \lambda_{TOD} * \lambda_{speed} * L^{\alpha} + t^{\gamma} * \left(\frac{t}{t_0} - 1\right)^{\omega}}$$ **Equation 1**

where
t=actual travel time
t_0 = free flow travel time
L = link length

λ_{TOD} and λ_{SPD} are dummy variables representing time of day and the speed limit,
λ is a constant, and α, γ and ω are estimated parameters.

Eliasson (2007) also discussed the travel time pattern in the study area. He found that travel time is less skewed under higher congestion conditions and the travel time covariance is relatively small. It is then possible to directly sum link travel time variances in order to get the route travel time variance.

A similar technique has been used for analysing the travel time variability in Peer et al. (2009). This study aimed to forecast travel time variability for use in cost benefit analysis of a transportation project. Peer et al. (2009) also assessed 15 minute interval travel time that was derived from speed data collected by loop detector. The speed data then can be converted to travel time data used piecewise-linear-speed based (PLSB) method. This study used 255 working days travel time data in its analysis. Similar to Eliasson's model, Peer at al also used a regression technique. However, instead of modelling the relationship between standard deviation and speed and travel time, this study focused on the link between the standard deviation, delay and the V/C ratio. The fitted model was:

$$\sigma = \beta_1 * delay * VCR^{\beta 3}$$ **Equation 2**

$$\boldsymbol{\sigma = L_1 * \left(\beta_{11} * delay^{\beta 21} * VCR^{\beta 31}\right) + L_2 * \left(\beta_{12} * delay^{\beta 22} * VCR^{\beta 32}\right)}$$ **Equation 3**

where
VCR = flow capacity ratio
L=link length
β=estimated parameter

In the United Kingdom, a model of link and corridor travel time variability, using GPS data collected by individual vehicles for 34 routes in ten of the largest urban areas in England, was developed by Black and Chin (2007). The travel time data from each vehicle were then grouped to 15-30 min interval data. Similar to previous methods that look at the relationship between the travel time variability and other traffic parameters, this study related the coefficient of variation of travel time variability (CV_t) to the congestion level in the study area

$$CV_t = \alpha CI_t^{\beta} \qquad \textbf{Equation 4}$$

where CI_t is a congestion index, defined as $CI_t = t / t_0$, where t is actual travel time t_0 is free flow travel time and α and β are estimated parameters. They first considered travel time variability at the link level, and then used standardised link travel times to develop a corridor travel time reliability model:

$$CV_t = 0.16 CI_t^{1.02} L^{-0.39} \qquad \textbf{Equation 5}$$

where L (km) is the link length and -0.39 is an estimated parameter (the elasticity of CV_t with respect to distance).

A similar model was developed by Richardson and Taylor (1978), who showed that under certain restrictive conditions the theoretical value of β would be 0.5.

Given the assumption that the day by day travel time variability can be treated as the time series data, Sohn and Kim (2009) used the autoregressive moving average-generalized autoregressive conditional heteroscedasticity (ARMA-GARCH) method to investigate the travel time pattern for road networks in South Korea. This study also tested the impact of the different combination of traffic and geometric condition that might influence the travel time reliability. It was found that the link width and congestion level play an important role in travel time variability.

The question of serial correlation between travel times on individual links along a route is of interest, and there are a number of different results in the reported studies. Richardson and Taylor (1978) found no significant statistical eveidence of such correlation, and as indicated above, neither did Eliasson (2007). Faouzi and Maurin (2007) suggested how to treat serial correlation between link travel times if those times were log normally distributed. Nicholson and Nicholson and Munakata (2009) found evidence of strong serial correlation in a study of travel times on the Tokyo expressway network.

2.1 Travel time variability metrics

Given the definition of the travel time variability – the dispersion of the actual travel time to the preferred travel time – the mean is the commonly used preferred travel time while standard deviation, coefficient of variation and a quantile range (difference between the 95th or 90th and 50th percentile) are suggested variability metrics. The use of the 90th percentile is notionally justified by the implication that it allows an employee to possibly arrive late in the office twice in 20 working days.

Some alternative travel time variability metrics have been proposed. Federal Highway Administration (FHWA) (2006) introduced a buffer time (BT_t) to represent the additional time above the average travel time ($\bar{t}$) required for on-time arrival. The buffer time is the difference between the 95th percentile travel time (t_{95}) and the mean travel time:

$$BT_t = t_{95} - \bar{t} \qquad \textbf{Equation 6}$$

FHWA (2006) also established a travel time reliability index (Planning Index, PI_t), which is the ratio of the 95th percentile travel time to the 'ideal' travel time, taken to be the free flow travel time (t_0):

$$PI_t = t_{95} / t_0$$ **Equation 7**

Additionally, van Lint and van Zuylen (2005) proposed the so-called skew-width methods. The skewness of the travel time λ^{skew} is defined as the ratio of difference of the 90th percentile (t_{90}) and 50th percentile (t_{50}) travel times and the difference between 50th percentile and 10th percentile (t_{10}) travel times. The width of travel time λ^{var} is defined as the ratio of the differences of 90th percentile and 50th percentile travel time and the 10th percentile travel time. The equation of the skewness and the width travel time metrics are as follows;

$$\lambda^{skew} = \frac{t_{90} - t_{50}}{t_{90} - t_{10}}$$ **Equation 8**

$$\lambda^{var} = \frac{t_{90} - t_{50}}{t_{50}}$$ **Equation 9**

From the review, it can be seen that the existing travel time variability model use different approaches as well as different shapes of a parametric distribution. There is commonality in the research in the use of the standard deviation, congestion index, link length and the mean of travel time to measure variability. Despite the fact that previous studies have attempted to model travel time variability and examine the contribution of each traffic variables to the travel time variability, most of this research has simply used the normal and log normal distributions and their properties in modelling travel time variability. There is still left room to study the best fit travel time distributions for travel time variability studies. Therefore, the next section discusses the use of the Burr distribution as a proposed travel time variability model.

3. The Burr distribution and Burr regression technique

Recent collected empirical travel time data exhibit positive skew and a long tail, and the normal distribution therefore does not really fit these data. Through exhaustive goodness of fit testing, using the eight years of continuous travel time data from the Adelaide database, Susilawati et al (2010) found that neither the Normal nor Log Normal distributions could fit the observed travel time distributions. Susilawati et al (2010) found that the Burr distribution fitted the empirical travel time data from Adelaide urban arterial roads and concluded that the Burr distribution could be used to represent the observed data. The next section reviews the Burr distribution as a proposed travel time distribution and the application of this distribution in reliability engineering and other reliability studies.

The Burr distribution was developed by Burr (1942) for the express purpose of fitting a cumulative distribution function (cdf) to a diversity of frequency data forms covering a wide range of distribution shapes. It can also be used to model an unspecified distribution.

In its basic form it has two parameters, c and k. The probability density function (pdf) f(x c,k) of the Burr distribution is

$$f(x,c,k) = ckx^{c-1}(1 + x^c)^{-(k+1)}$$ **Equation 10**

where $x > 0$, $c > 0$ and $k > 0$. The cdf $F(x, c, k)$ is given by

$$F(x,c,k) = 1 - (1 + x^c)^{-k}$$ **Equation 11**

The distribution has some interesting statistical properties (Tadikamalla, 1980). In the first instance the *r*th moment of the distribution ($E(x^r)$) will only exist if $ck > r$, in which case

$$E(x^r) = \mu_r' = \frac{k\Gamma(k - \frac{r}{c})\Gamma(\frac{r}{c} + 1)}{\Gamma(k+1)}$$ **Equation 12**

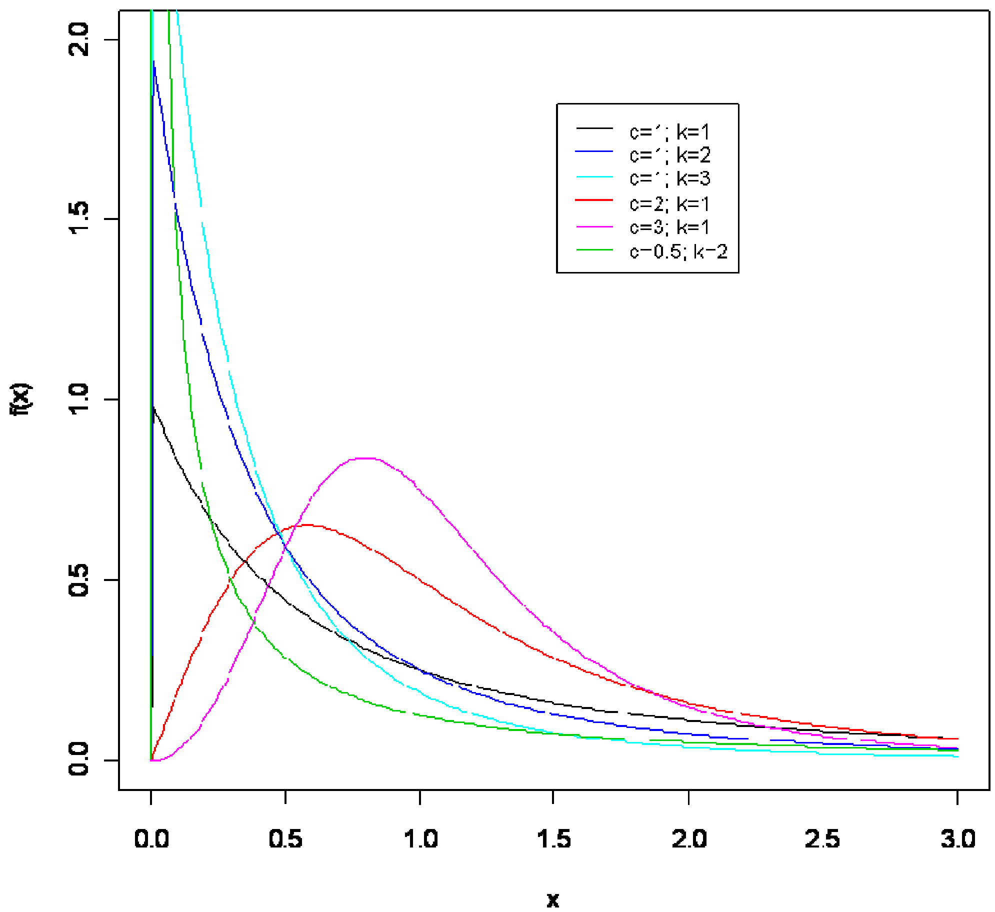
where $\Gamma(y)$ is the mathematical Gamma function. In addition, the modal value x_m is given by

$$x_m = \left[\frac{c-1}{ck+1}\right]^{1/c}$$ **Equation 13**

but x_m will only exist if $c > 1$. [If $c \leq 1$, then the distribution is L-shaped.]

Based on previous research, the value of c plays a very important role in determining the shape of the cdf curve -the lower the value of c, the sharper the fall of cdf curve (c = 1 the sharper fall of cdf curve, c = 2 shallower fall of curve) while the lower the k parameter, the sharper the initial rise of the curve (k=2, sharper the initial rise of cdf curve, k = 3, shallower initial rise of curve) as shown in figure 1. Interestingly, the k parameter can also define the skewness, for k more than one means that the distribution are right skew and less than one is left skewed. See Figure 1 for some alternative Burr distribution shapes.

Figure 1: Burr distribution cdf for different c and k parameter



Interestingly, the wide range of possible shapes of the Burr distribution and its well behaved algebraic properties make it is useful for fitting many types of data and for approximating many different distributions (Zimmer et al, 1998). For instance, the Weibull distribution is specific type of Burr distribution where one of the parameters is gamma distributed. When the k parameter is one then the Burr distribution is a special case of log logistic distribution. The versatility and flexibility of the Burn distribution makes it quite attractive as a tentative model for data whose underlying distribution is unknown

The Burr distribution is very popular distribution and has gained a strong interest in reliability engineering for modelling lifetime data and monotone failure rates. A number of reliability engineering applications have utilised it to model the product life process (Abdel-Ghaly et al., 1997). Also, Shao et al. (2004) studied the models for extended three parameter of Burr distribution and used this distribution to model extreme event with application to flood frequency. The three parameter Burr distribution or Singh Maddala distribution is the extended version of the Burr distribution which has a scale parameter for the independent variable *x*. The scale parameter defines the point where the data is most concentrated. For purposes of simplicity, the scale parameter can be used to estimate the median of the data.

The distribution has an algebraic tail that is useful in modelling less frequent failures (Soliman, 2005). As its cdf can be written in closed form, its percentiles are easily computed. To compute percentiles for the Burr distribution, use the following approach

Given the CDF defined by equation (10), for a given value of *F(x, c, k)*, the percentile *P*, i.e.

$$P = 1 - (1 + x^c)^{-k} \quad \textbf{Equation 14}$$

$$\therefore \quad (1 + x^c)^k = \frac{1}{1 - P}$$

$$\therefore \quad x^c = (1 - P)^{-1/k} - 1$$

$$\therefore \quad x_P = \sqrt[c]{(1 - P)^{-1/k} - 1}$$

the median is then

$$x_{50} = \sqrt[c]{2^{1/k} - 1} \quad \textbf{Equation 15}$$

90th percentile is

$$x_{90} = \sqrt[c]{10^{1/k} - 1} \quad \textbf{Equation 16}$$

10th percentile is

$$x_{10} = \sqrt[c]{\left(\frac{10}{9}\right)^{1/k} - 1} \quad \textbf{Equation 17}$$

On the basis of equations 15, 16 and 17, the next section indicates how to compute current travel time variability metrics directly from the Burr distribution parameters.

4. Proposed travel time variability metrics

Preceding section reviewed the travel time variability modelling and metric development and also the properties of the Burr distribution. Equation 4 and 5 show that the 50th percentile and the 90 percentile are easy to obtained once the parameter c and k are known. Since the travel time variability can also be measured as the dispersion of the preference travel time and the actual travel time, this study used the similar method used by Lint and Zuylen (2005) to develop the travel time variability metrics. According to Lint and Zuylen (2005), λ^{var} is the ratio between the difference of the 90th percentile and the 50th percentile and the 50 percentile. This metric is useful to illustrate the travel time dispersion. Replacing the 90th percentile and the 50th percentile in the Lint and Zuylen (2005) metrics, skewness of travel time' (λ^{skew}) and 'width of travel time' (λ^{var}) in terms of Burr parameters is possible, given that these two metrics are both defined in terms of percentiles. Therefore we obtain refined travel time variability metrics as below.

$$\lambda^{skew} = \frac{t_{90} - t_{50}}{t_{90} - t_{10}}$$

$$\lambda^{\mathrm{var}} = \frac{t_{90} - t_{50}}{t_{50}}$$

Given x_{50} and x_{90} as defined by equation 15, 16 and 17, it follows that

$$\lambda^{skew} = \frac{\sqrt[c]{10^{1/k} - 1} - \sqrt[c]{2^{1/k} - 1}}{\sqrt[c]{10^{1/k} - 1} - \sqrt[c]{\left(\frac{10}{9}\right)^{1/k} - 1}}$$ **Equation 18**

and

$$\lambda^{\mathrm{var}} = \frac{\sqrt[c]{10^{1/k} - 1} - \sqrt[c]{2^{1/k} - 1}}{\sqrt[c]{2^{1/k} - 1}}$$ **Equation 19**

5. The travel time data database

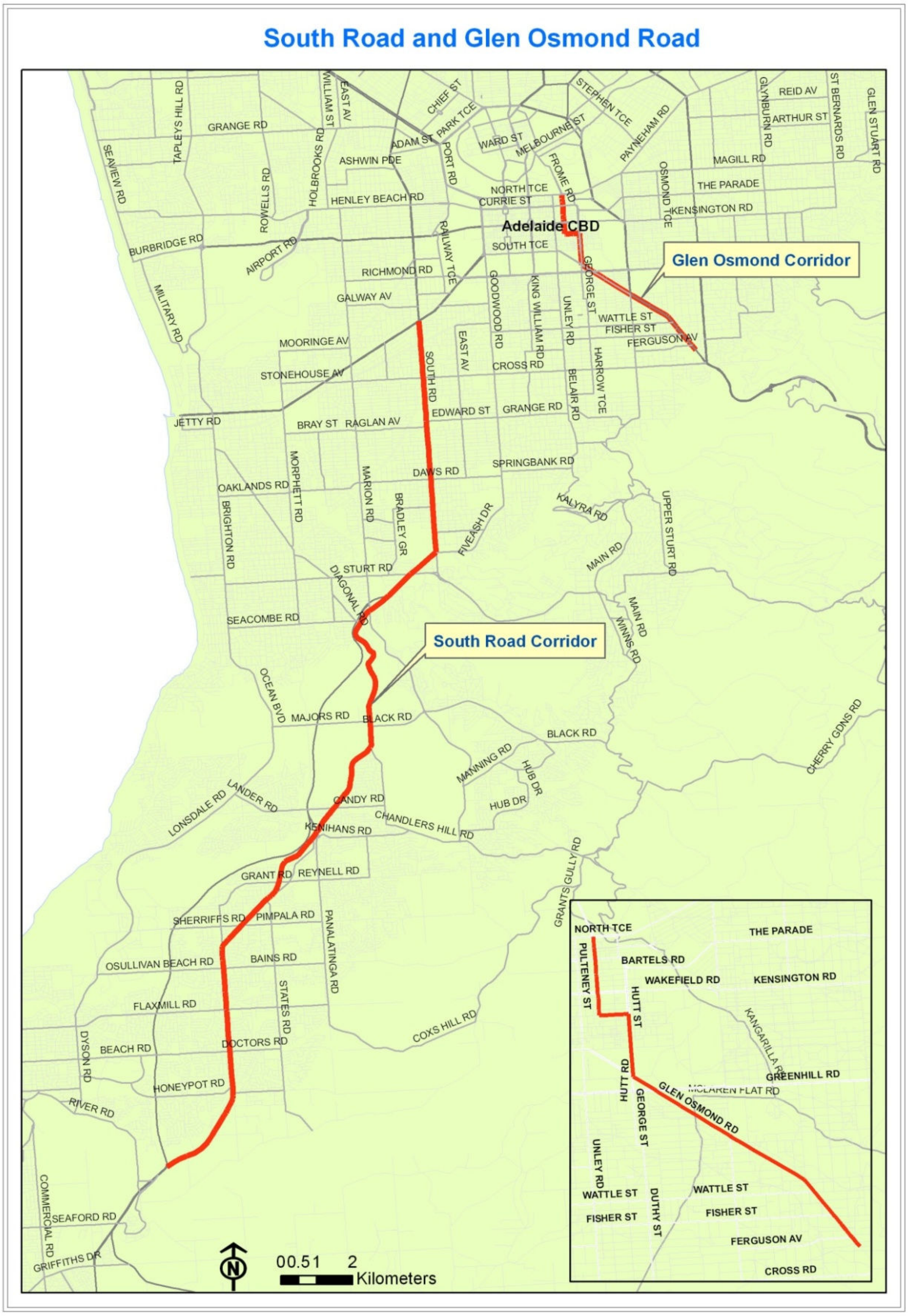
This study used two of the sets of longitudinal travel time data in the database. The data was collected by using GPS during regular journey to work trips for different commuters, each of which started at about the same time of day and with similar driving behaviour.

The two sets of travel time data are

(1) the Glen Osmond Road (GOR) data set, including 16 successive links from Glen Osmond to the Adelaide CBD. The length of links vary between 150 m to 600 m with speed limits of either 50km/h or 60km/h, and

(2) the South Road (SR) data set, which is the collection of travel time from a major freight corridor in Adelaide. This route serves the southern suburbs of Adelaide metropolitan area. The links length of this route also varies from 165 m to more than 4000 m and the speed limit varies between 40 km/h (because of road works) and 80km/h.

Figure 2: shows each of these routes on a map of the Adelaide metropolitan area.



Using maximum likelihood estimation, the three parameters (c, k and scale) of the Burr distribution were obtained. These parameters therefore can be used for calculating th 90th and 50th percentile and the λ^{var} . The three parameters and the variability of each links are shown in Table 1 (South Road) and Table 2 (Glen Osmond Road).

Table 1: The Burr parameters and the link variability of South Road data set

Link No	Length (m)	Parameter			Percentile		λvar	free flow t_0(s)	CI
		k	c	scale	50th	90th			
1	2782	1.17	63.86	142.94	142.51	146.78	0.03	125.19	1.138
2	213	0.10	44.03	10.34	12.14	17.63	0.45	10	1.214
3	944	0.02	222.22	58.55	68.24	97.27	0.43	56.64	1.205
4	1177	0.07	48.40	70.93	87.97	145.05	0.65	70.62	1.246
5	710	0.09	76.58	41.66	45.90	57.51	0.25	42.60	1.077
6	442	0.05	62.53	26.91	33.89	57.89	0.71	26.52	1.278
7	1340	0.07	47.36	83.66	104.08	172.84	0.66	80.40	1.295
8	3243	0.06	103.87	163.10	181.35	231.99	0.28	145.94	1.243
9	745	0.07	56.33	35.17	41.84	62.62	0.50	33.53	1.248
10	1965	0.04	223.44	89.02	97.05	118.59	0.22	88.43	1.098
11	595	1.54	5.59	61.79	55.86	77.16	0.38	26.78	2.086
12	3232	0.09	96.22	149.33	161.76	194.77	0.20	145.44	1.112
13	322	0.00	23414	14.54	N/A	N/A	N/A	14.49	N/A
14	592	0.03	47.03	30.28	51.02	171.59	2.36	30.45	1.676
15	416	0.04	48.25	25.59	35.44	75.51	1.13	21.39	1.657
16	393	52408	3.02	2124.2	51.33	76.42	0.49	20.21	2.540
17	841	0.08	66.62	65.89	75.47	103.44	0.37	50.46	1.496
18	2032	1.24	3.39	244.57	224.39	401.98	0.79	121.92	1.840
19	871	N/A	N/A	N/A	N/A	N/A	N/A	52.26	N/A
20	761	4.34	3.37	190.73	113.34	171.58	0.51	45.66	2.482
21	1209	0.02	124.19	75.22	94.64	161.32	0.70	72.54	1.305
22	1590	0.78	5.35	264.35	282.74	454.65	0.61	95.40	2.964

N/A : not available data due to the burr distribution did not fit the travel time data

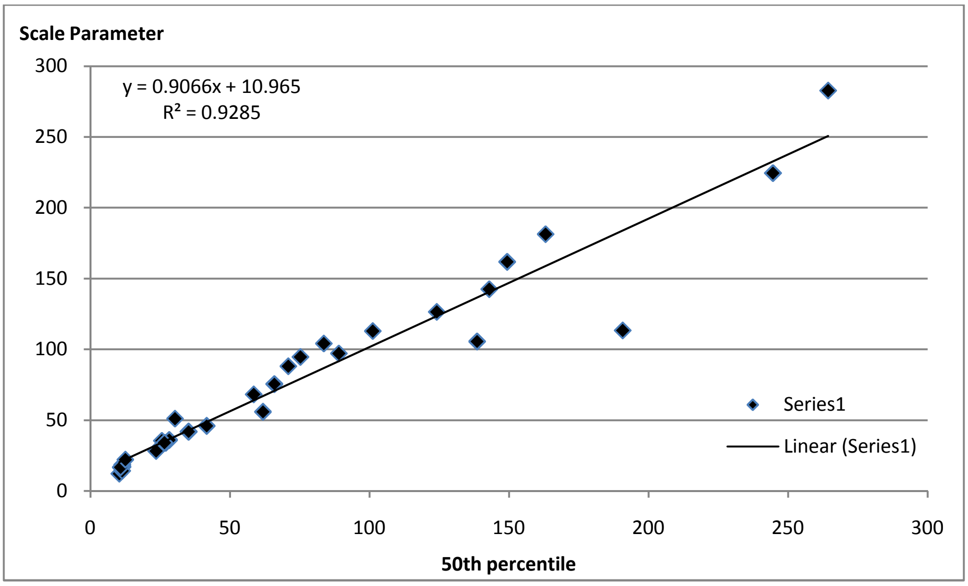
Table 2 : The Burr parameters and the link variability of Glen Osmond Road data set

Link No	link Length (m)	Parameter			Percentile		λ^{var}	CI
		c	k	scale	50	90		
1	1146	8.96	0.53	101.15	112.84	163.45	0.45	1.763
2	1058	4.17	0.94	124.10	126.48	217.71	0.72	2.008
3	458	N/A	N/A	N/A	N/A	N/A	N/A	N/A
4	606	3.32	2.04	138.59	105.52	173.22	0.64	2.931
5	331	N/A	N/A	N/A	N/A	N/A	N/A	N/A
6	405	92.98	0.03	28.25	35.96	62.97	0.75	1.332
7	165	27.74	0.06	11.44	16.99	42.57	1.51	1.545
8	150	25.15	0.12	11.34	14.15	23.67	0.67	1.415
9	311	53.94	0.07	23.50	28.16	42.82	0.52	1.280
10	337	62.88	0.04	26.54	34.03	60.61	0.78	1.361
11	165	N/A	N/A	N/A	N/A	N/A	N/A	N/A
12	165	33.50	0.04	11.42	18.40	55.68	2.03	1.673
13	152	29.16	0.05	10.84	16.73	45.85	1.74	1.673
14	156	N/A	N/A	N/A	N/A	N/A	N/A	N/A
15	153	19.05	0.06	12.48	22.13	83.64	2.78	2.012
16	162	N/A	N/A	N/A	N/A	N/A	N/A	N/A

N/A : not available data due to the burr distribution did not fit the travel time data

It can be seen that the travel time variability metrics for both data sets do vary. The highest travel time variability is for SR data set is link 14 (2.36) while the lowest is for link 1 (0.03). On average the travel time variability for SR data set is about 0.6 which means that the dispersion of the travel time from the median 0.6 times median. On the other hand, the travel time variability for GOR data set is much more diverse than the SR data set. The highest is for link 15 with λ^{var} equalling 2.78. This means that the travel time variation is almost triple the median travel time. Link 1 has the lowest travel time variability which is only 0.45 of the median.

Figure 3 the scatter plot of Scale parameter and the median of the SR and GOR data



Using the results in Tables 1 and 2, Figure 3 shows the scatter plot of the scale and the median of the SR and GOR data sets. It clearly shows that there is a linear relationship between these two variables (with R^2 =0.93). It can be suggested here that for purposes of simplicity the scale parameter of Burr distribution can be estimated the median of travel time.

These tables also show that the travel times do vary link by link. Since the travel time data used in this study is the link travel time where the link lengths also vary, further data analysis to assess the role of traffic variables on travel time variability is conducted. There are two question arose. The first question is how the link travel time will likely affect the travel time variability. In order to answer the first question, a similar formula to that used by Black and Chin (2009) – Coefficient of Variation (CV) as a function of link length and congestion Index (CI) is adopted. Given that formula (see equation 5), the λ^{var} is a function of the link length and the congestion index as a refined travel time variability formula is developed. To measure the congestion index, instead of using the mean as the denominator, the median (50^{th} percentile) was using.

Adopting the new proposed travel time variability metric, the link by link travel time variability and the congestion index modelling is tabulated in table 3.

Table 3: Estimated Parameter for travel time variability modelling

Parameter	Description	Value
α	constant	28.0913
β_1	estimated parameter for Congestion Index	0.7462
β_2	estimated parameter for link length	-0.6373

It shows that both link lengths give the negative value of the power function, while congestion index give the positive value to the λ^{var}. This result is quite different to Black and Chin's research, but this may reflect the different natures of the two data sets

The proposed method can be extended to investigate the influence of other traffic variables for further data analysis. However, for this model, the data analysis limits for just testing the affect of the link length on the λ^{var} .

A second objective of this research is to seek the role of the degree of saturation in influencing the c parameter. Before conducting further data analysis, there is a judgement required in selecting the degree of saturation as the main factor in this study. Based on the literature, it is found to be an important variable in determining urban traffic system performance. The degree of saturation gives the ratio of the capacity and flow. These variables then can be input into the travel time variability modelling.

This paper does not discuss the method to obtain the degree of saturation and how to relate the SCATS degree of saturation data to the theoretical degree of saturation. The discussion about that issue can be found in Suksri e et al (2010). The degree of saturation data obtained from SCATS data of each link has been used as the covariates factors. Next section reviews the Burr regression technique and how to utilise this technique in the travel time variability modelling.

6. Burr regression

The Burr regression technique was firstly introduced by Beirlant et al. (1998) in related to the special case of the Burr distribution. It has been known for some time that the log logistic distribution is a special case of the Burr. Adopting the similar parameterisation as used in the log logistic regression, Beirlant et al. (1998) proposed Burr distribution parameterisations.

Parametrization:

The shape parameter *(c)* is allowed to vary with $\boldsymbol{y}$ where $\boldsymbol{y}$ is covariate (dimensional vector)(Beirlant et al., 1998)

Given the cdf of the Burr distribution

$$f(x,c,k) = ckx^{c-1}(1+x^{c})^{-(k+1)}$$ **Equation 20**

then based on the first parameterization, the new cdf of Burr distribution is as follows

$$x \mid \boldsymbol{y} \sim Burr(k, c(\boldsymbol{y})).$$

Since $c(\boldsymbol{y}) = \exp(\theta'\boldsymbol{y})$ can be termed as an exponential function of θ' and y ,then replacing the c with the *c(**y**)* the cdf becomes

$$\boldsymbol{x_i \mid y_i \sim Burr\ (k, c_i),\qquad c_{i=}\ exp(\theta' y_i)}$$

$$\boldsymbol{f(x,\theta',y,k) = \exp(\theta' * y) * k * x^{\exp(\theta' * y)} * \left(1 + x^{\exp(\theta' * y)}\right)^{-(k+1)}}$$ **Equation 21**

Where,

k=shape parameter
c=shape parameter
x= travel time
y = Degree of saturation

Maximum likelihood estimation can be used for Burr regression parameter estimation by replacing the c with the $\exp(\theta'\boldsymbol{y})$ expression. The maximum likelihood analysis can be done as follows:

The likelihood function is

$$\boldsymbol{L(k,\theta|x) = k^{n} exp\left(\sum_{i=1}^{n}\theta' y_i\right)\prod_{i=1}^{n}\frac{x^{exp(\theta' y_i)-1}}{\left(1+x_i^{exp(\theta' y_i)}\right)^{k+1}}}$$ **Equation 22**

$$\boldsymbol{lnL(k,\theta|x) = nln(k) + \sum_{i=1}^{n}\theta' y_i + \sum_{i=1}^{n}[exp\ (\theta' y_i) - 1]\ ln(x_i) - (k+1)\sum_{i=1}^{n} ln\ \left[1 + x_i^{exp(\theta' y_i)}\right]}$$ **Equation 23**

Then

$$\boldsymbol{\frac{\vartheta ln\ L\ (k,\theta|x)}{\vartheta k} = \frac{n}{k} - \sum_{i=1}^{n} ln\ \left[1 + x_i^{exp(\theta' y_i)}\right]},$$ **Equation 24**

$$\boldsymbol{\frac{\vartheta ln\ L\ (k,\theta|x)}{\vartheta \theta_j} = \sum_{i=1}^{n} y_{ij} + \sum_{i=1}^{n} exp\ (\theta' y_i) y_{ij}\ ln(x_i) - (k+1)\sum_{i=1}^{n}\frac{x_i^{exp(\theta' y_i)}\ ln(x_i)}{1+x_i^{exp(\theta' y_i)}}\ exp(\theta' y_i)\, y_{ij}}$$ **Equation 25**

For illustrative purposes, this study merely applied the first parameterisation of Burr distribution for selected GOR and SR links. The SCATS degree of saturation data has been used as the covariate matrix (*y*) - representing the traffic variables. The Burr regression modelling result is given in Table 3

Table 4: Burr regression modelling for Glen Osmond data set

Link No	Parameter Burr Distribution			Parameter Burr regression technique			Log Likelihood test ratio cv*=3.8412
	c	k	scale	k	θ	scale	
6 (GOR)	20.9046	0.0570	12.3585	0.1329	2.4435	11.4297	Accepted (-2651.8)
12 (GOR)	35.0420	0.0406	11.3570	0.1361	3.1132	10.3965	Accepted (-2474.8)
15 (SR)	44.5127	0.0560	25.6863	0.0347	0.0459	25.3426	Accepted (-49.6138)
20 (SR)	3.3591	4.7948	200.6249	0.0237	4.1639	54.2633	Accepted (-1123.2)
3(SR)	200.8478	0.0220	58.5894	0.1202	4.6466	59.8820	accepted (-15366.1)
22(SR)	5.4379	0.7124	261.3926	0.0356	3.0636	148.8896	accepted (-1709.3)

cv*=critical value at the 95^{th} confidence level

The result of the general form Burr distribution and Burr regression technique parameter estimation are shown in Table 4. To test the similarity of these two techniques the log likelihood ratio test was conducted. This result is promising as the value of k and scale parameters from both techniques are similar. From the log likelihood test, it was found that the Burr regression technique for selected links of GOR and SR perform well, as indicated by the p value of the log likelihood test. The p values are less than the critical value at 95^{th} confidence level.

The θ as the estimated parameter for degree of saturation data set is showing that for different link with different traffic condition gave different estimated variables. By assuming that the link which have higher variability index, higher congestion index might have higher degree of saturation, then this table can describe that behaviour. For example link 12 has higher variability and also has higher θ than link 6.

According to the variability and congestion index values, the GOR data set may well represent oversaturated conditions. For undersaturated condition, the similar technique was applied for selected SR links.

7. Conclusion

The need for more reliable travel time has gained strong interest in international research. There are numerous studies that have suggested metrics to quantify the variability of travel time and have looked into the factors that might contribute to the higher travel time variability. Those studies suggested that the basic traffic variables were playing an important role in modelling the travel time variability and used the normal distribution properties as the travel time variability metrics. However, the recent empirical travel time data exhibit positive skew and long tails, and so are not well represented by standard pdfs such as the Normal distribution. This study therefore suggested refined travel time variability metrics by utilising the Burr distribution properties. From two sets of link travel time data collected in Adelaide, it was found that the link travel time variability does vary. The first hypothesis needed to test is whether the link length is the main contributor to the travel time variability, the second is what congestion index role in modelling the travel time variability. The results confirmed the positive effect of the link length and the power effect congestion index, respectively, on the link travel time variability. This result is similar to the earlier research conducted by Black and Chin (2007).

This paper gives an insight of the use of the Burr distribution properties to refine the current travel time variability metrics. This technique has been tested for selected links of the GOR and SR data sets. Under the 95th confidence level, the result of this proposed modelling technique is similar to the result of the general form of Burr regression technique. Interestingly, this proposed technique can be extended to investigate more traffic variables in related to the Burr distribution parameter. This new approach can be also used for further travel time variability modelling.

Appendix

Notation	Description
σ	standard deviation
λ	constant
λ_{TOD}	dummy variables representing time of day
λ_{speed}	dummy variables representing the speed limit
L	link length
t	actual travel time
t_0	free flow travel time
α	estimated parameter
γ	estimated parameter
ω	estimated parameter
β	estimated parameter
β_{ij}	estimated parameter
VCR	flow capacity ratio
CI	congestion Index
CV_t	coefficient of Variation
BT_t	buffer time
PI_t	planning Index
t_{95}	95th percentile travel time
t_{90}	90th percentile travel time
t_{50}	50th percentile travel time
t_{10}	10th percentile travel time
$\bar{t}$	mean travel time
λ^{skew}	the skewness of the travel time
λ^{var}	the width of travel time
f	pdf
F	cdf
c	shape parameter Burr distribution
k	shape parameter Burr distribution
$\Gamma(y)$	Gamma function
x_m	mode
P	percentile
x_{90}	90th percentile Burr distribution
x_{50}	50th percentile Burr distribution
x_{10}	10th percentile Burr distribution
y	degree of saturation
θ	estimated parameter

References

Abdel-Ghaly, A.A., Al-Dayian, G.R. & Al-Kashkari, F.H. (1997) The use of Burr Type XII distribution on software reliability growth modelling. *Microelectronics and Reliability,* 37, 305-313.

Beirlant, J., Goegebeur, Y., Verlaak, R. & Vynckier, P. (1998) Burr regression and portfolio segmentation. *Insurance: Mathematics and Economics,* 23**,** 231-250.

Black, I. & Chin, T. K. (2007) Forecasting Travel Time Variability in Urban Areas. *Deliverable D1: Data Analysis and Model Development.* Hyder Consulting (UK) Limited.

Burr, I.W. (1942) Cumulative frequency functions. *The Annals of Mathematical Statistics,* 13**,** 215-232

Eliasson, J. (2006) Forecasting travel time variability. *Proceedings of the European Transport Conference.*

Eliasson, J. (2007) The relationship between travel time variability and congestion. *11th World Conference on Transport Research.* Berkeley CA.

Emam, E. B. & Al-Deek, H. (2006) Using real-life dual-loop detector data to develop new methodology for estimating freeway travel time reliability. *Journal of the Transportation Research Board,* 1959.

Faouzi, N.-E. E. & Maurin, M. (2007) Reliability Metrics for path travel time under log-normal distribution. *3rd International Symposium on Transportation Network Reliability.* The Netherlands.

Herman, R. & Lam, T. 1974, 'Trip time characteristics of journeys to and from work', In Douglas J. Buckley (ed.) *Transportation and Traffic Theory*, A.H. and A.W. Reed, Sydney ,57-85.

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

Lint, J. W. C. v. & Zuylen, H. J. v. (2006) Monitoring and predicting travel time reliability: Using width and skew of day-to-day travel time distribution. *Journal of the Transportation Research Board,* 1917**,** 54-62.

Nicholson, A. & Munakata, K. (2009) Estimating the benefits of trip time reliability. *32nd Australasian Transport Research Forum 2009.* Auckland, New Zealand.

Oh, J.-S. & Chung, Y. (2006) Calculation of travel time variability from loop detector data. *Journal of the Transportation Research Board,* 1945**,** 12-23.

Peer, S., koopmaans, C. & verhoef, E. (2009) Predicting travel time variability for cost benefit analysis. *ERSA Conference Lodz.* Poland.

Polus, A. 1979, 'A study of travel time and reliability on arterial routes', *Transportation,* vol.8**,** 141-151.

Richardson, A. J. & Taylor, M. A. P. 1978, 'Travel time variability on commuter journeys', *High Speed Ground Transportation Journal,* vol.6**,** 77-99.

Shao, Q., Wong, H., Xia, J. & Ip, W.-C. (2004) Models for extremes using the extended three parameter Burr XII systems with application to flood frequency analysis. *Hydrological Sciences - Journal,* 49**,** 685-701.

Shao, Q., Wong, H., Xia, J. & Ip, W.-C. (2004) Models for extremes using the extended three parameter Burr XII systems with application to flood frequency analysis. *Hydrological Sciences - Journal,* 49**,** 685-701.

Soliman, A. A. (2005) Estimation of parameters of life from progressively censored data using Burr-XII model. *IEEE Transactions on Reliability*, 54, 34-42.

Suksri, C., Taylor, M. A. P. & Yue, W. L. (2011) Traffic analysis of a signalised intersection under high throughputs. To be presented at *The Eastern Asia Society for Transportation Studies.* Jeju, South Korea. 20-23 June

Susilawati, Taylor, M. A. P. & Somenahalli, S. (2010) Distribution of variability in travel times on urban roads - a longitudinal study. *12th World Conference of Transport Research.* Lisbon, Portugal.

Tadikamalla, P.R. (1980). A look at the Burr and related distributions. *International Statistical Review*, 48, 337-344

Zhang, W., Medina, A. & Rakha, H. (2007) Statistical analysis of spatiotemporal link and path flow variability. *Proceedings of the 2007 IEEE, Intelligent Transportation Systems Conference.* Seattle, WA, USA.

Zimmer, W. J., Keats, J. B. & Wang, F. K. (1998) The Burr XII distribution in reliability analysis. *Journal of Quality Technology*, 30 (4), 386-394.