JOURNAL OF ADVANCED TRANSPORTATION
J. Adv. Transp. 2013; **47**:720–736
Published online 30 December 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/atr.192

Distributions of travel time variability on urban roads

Susilawati Susilawati, Michael A. P. Taylor* and Sekhar V. C. Somenahalli

Barbara Hardy Institute, University of South Australia, Adelaide, Australia

SUMMARY

Reliability is an important factor in route, mode and also departure time choice analysis and is a key performance indicator for transport systems. However, the current metrics used to measure travel time variability may be not sufficient to fully represent reliability. Better understanding of the distributions of travel times is needed for the development of improved metrics for reliability. A comprehensive data analysis involving the assessment of longitudinal travel time data for two urban arterial road corridors in Adelaide, Australia, demonstrates that the observed distributions are more complex than previously assumed. The data sets demonstrate strong positive skew, very long upper tails, and sometimes bimodality. This paper proposes the use of alternative statistical distributions for travel time variability, with the Burr Type XII distribution emerging as an appropriate model for both links and routes. This statistical distribution has some attractive properties that make it suitable for explicit definition of many travel time reliability metrics. Copyright © 2011 John Wiley & Sons, Ltd.

KEY WORDS: highway and traffic engineering; intelligent transport systems; traffic systems; travel time; transport planning; transportation networks

1. INTRODUCTION

Many factors can adversely affect transport network performance. Different types of incidents, either short term (e.g. vehicle breakdowns) or long term (e.g. bridge collapse), or random (e.g. road crashes) or intentional (e.g. road works), can happen at any time and may lead to higher travel time variability and perhaps wider consequences for the community. In addition, the need for more reliable transportation systems and demands for 'just-in-time' services have generated new interest in transportation system reliability, which is thus a major research topic.

Travel time reliability is based on the concept of a travel time that meets travellers' expectations [1]. Travellers expect their travel times not to exceed a scheduled value, or average travel time plus some acceptable additional time, and hence, they can decide on a starting time for the journey. The concept of an acceptable additional time is subjective and will vary depending on perceptions and individual circumstances. Overly conservative travel time estimates may be unhelpful as these may cause travellers to arrive too early. This leads to the use of maximum utility models to jointly determine departure time and trip time, as discussed by Fosgereau and Karlstrom [2].

Acknowledging the appropriate travel time distribution and the probability of travel time 'failure'[1] is thus important for the development of travel time reliability metrics. This is consistent with practice in reliability engineering, which is concerned with measuring the consistency and the persistency of a product under different conditions over a period. On the basis of the following considerations,

*Correspondence to: Michael A. P. Taylor, Barbara Hardy Institute, University of South Australia, Adelaide, Australia. E-mail: map.taylor@unisa.edu.au
[1]Travel time failure is taken to be excess travel time incurred above some acceptable threshold.

- The current metrics used to measure travel time variability may not be sufficient to fully represent travel time reliability.
- There is known to be significant variability in individual travel times (e.g. [3]).
- There are growing demands for more reliable travel time measurements (e.g. see [4] and [5]).

A better understanding of the distributions of individual travel times is needed for the development of relevant metrics for assessing travel time reliability.

This paper focuses on the specification of appropriate day-to-day travel time variability distributions.[2] It first reviews previous research and then tests different statistical distributions using empirical data. Whereas previous travel time reliability studies have often focused on freeway travel times—usually because of the availability of suitable data sets involving observations of large numbers of individual travel times over a short period (hours of the day)—the present study investigates travel time reliability of two urban arterial road corridors. The study used continuous travel time data collected using GPS-equipped probe vehicles travelling along the routes, with repeated runs made over long periods (weeks and months) for individual journeys each starting at about the same time of day. This data collection replicates the experiences of an individual traveller making a routine trip, such as the journey to work.

2. TRAVEL TIME VARIABILITY DISTRIBUTIONS

Research on fitting continuous distributions to observed travel time data began many decades ago. Although initial belief was that the normal distribution was appropriate, Wardrop [6] first suggested that travel times followed a skewed distribution. Later, Herman and Lam [7] analysed urban arterial travel time data collected in Detroit in a longitudinal study of work trip journey times. They found significant skew in the observed times and proposed either the Gamma or lognormal distributions to represent travel time variability.

Richardson and Taylor [8] then collected and analysed longitudinal travel time data in Melbourne. They assessed the correlations between travel times on each section 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.

Using continuous travel time data collected in Chicago, Polus [9] found that the Gamma distribution was superior to normal or lognormal distributions. More recently Al-Deek and Emam [10] used the Weibull distribution to model travel time reliability.

3. THE BURR DISTRIBUTION

Previous studies have fitted travel time data to normal, lognormal, Gamma and Weibull distributions. However, these distributions do not seem to fit many empirical travel time data sets particularly well, as they are unable to model travel time distributions with strong positive skew and long upper tails. Similar problems have arisen in reliability engineering, where most life-test data is also distributed with positive skew and long tails. Study of the best-fit distributions for product lifetime data are thus of interest. Initial product reliability analyses assumed that the lognormal distribution could be appropriate for life-test data distributions. However, recent research has tended to reject this hypothesis, while the Weibull and Gamma distributions have also proved largely unsuccessful in fitting observed life-test data distributions. Zimmer *et al.* [11] noted the advantages of the Burr Type XII distribution (subsequently termed the Burr distribution) in modelling observed lifetime data. The Burr distribution is also well known in actuarial theory, where it has found a place in modelling distributions of insurance claims. It was developed by Burr [12] for the express purpose of fitting a cumulative distribution function (c.d.f.) to a diversity of frequency data forms. In its basic form, it has two parameters, c and k.

[2]This is in keeping with the broad planning definition of travel time reliability as the level of variation from day to day for a trip made by an individual starting at about the same time each day (e.g. see [22]).

DOI: 10.1002/atr

The probability density function (p.d.f.) f(x, c, k) of the Burr distribution is

$$\mathrm{f}(x, c, k) = ckx^{c-1}(1 + x^c)^{-(k+1)}$$

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

$$\mathrm{F}(x, c, k) = 1 - (1 + x^c)^{-k}$$

The distribution has some interesting statistical properties [13]. In the first instance, the rth moment of the distribution ($E(x^r)$) will only exist if $ck > r$, in which case,

$$E(x^r) = \mu'_r = \frac{k\Gamma\left(k - \frac{r}{c}\right)\Gamma\left(\frac{r}{c} + 1\right)}{\Gamma(k + 1)}$$

where $\Gamma(y)$ is the mathematical Gamma function. The product ck is thus an important factor for the visualisation of a specific Burr distribution fitted to observed data. In addition, the modal value x_m is given by

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

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

The Burr distribution thus has a flexible shape and is well behaved algebraically. A number of reliability engineering applications have utilised it to model the product life process [14]. The distribution has an algebraic tail that is useful in modelling less frequent failures [15]. As its c.d.f. can be written in closed form, its percentiles are easily computed. It allows a wide variety of shapes in its p.d.f. [11], making it useful for fitting many types of data and for approximating many different distributions (e.g. lognormal, log-logistic, Weibull and generalised extreme value).

4. EMPIRICAL TRAVEL TIME DATA

Our longitudinal journey to work travel time surveys are being conducted[3] on arterial road routes in the Adelaide metropolitan area by using GPS-equipped probe vehicles. The GPS provides a second-by-second data stream, including location and travel speed continuously recorded as the vehicle moves along the route. The routes, shown in Figure 1, are as follows:

(1) Glen Osmond Road, from the eastern suburbs of Adelaide into the central business district (CBD). This route comprises 16 links, with link lengths varying from 152 m to 1146 m, and posted speed limits of either 60 km/h or 50 km/h
(2) The South Road corridor, comprising 22 links. Link lengths vary from 135 m to 4007 m with posted speed limits between 80 and 60 km/h.

There are 180 runs for route 1 and 67 runs for route 2. Tables I and II show the mean, standard deviation and coefficient of variation of link travel times for each route.

4.1. Normal and lognormal distributions

Normal and lognormal distributions were first fitted to the observed data, using the Kolmogorov–Smirnov goodness-of-fit test. The results for the Glen Osmond data set are shown in Table III. Neither the normal nor lognormal distributions fitted any of link travel time data sets on this route. A slightly different

[3]The longitudinal surveys are ongoing, and the data presented in this paper represent the first 12 months of data collection on two specific routes. Other routes have recently been added to the study.

DOI: 10.1002/atr

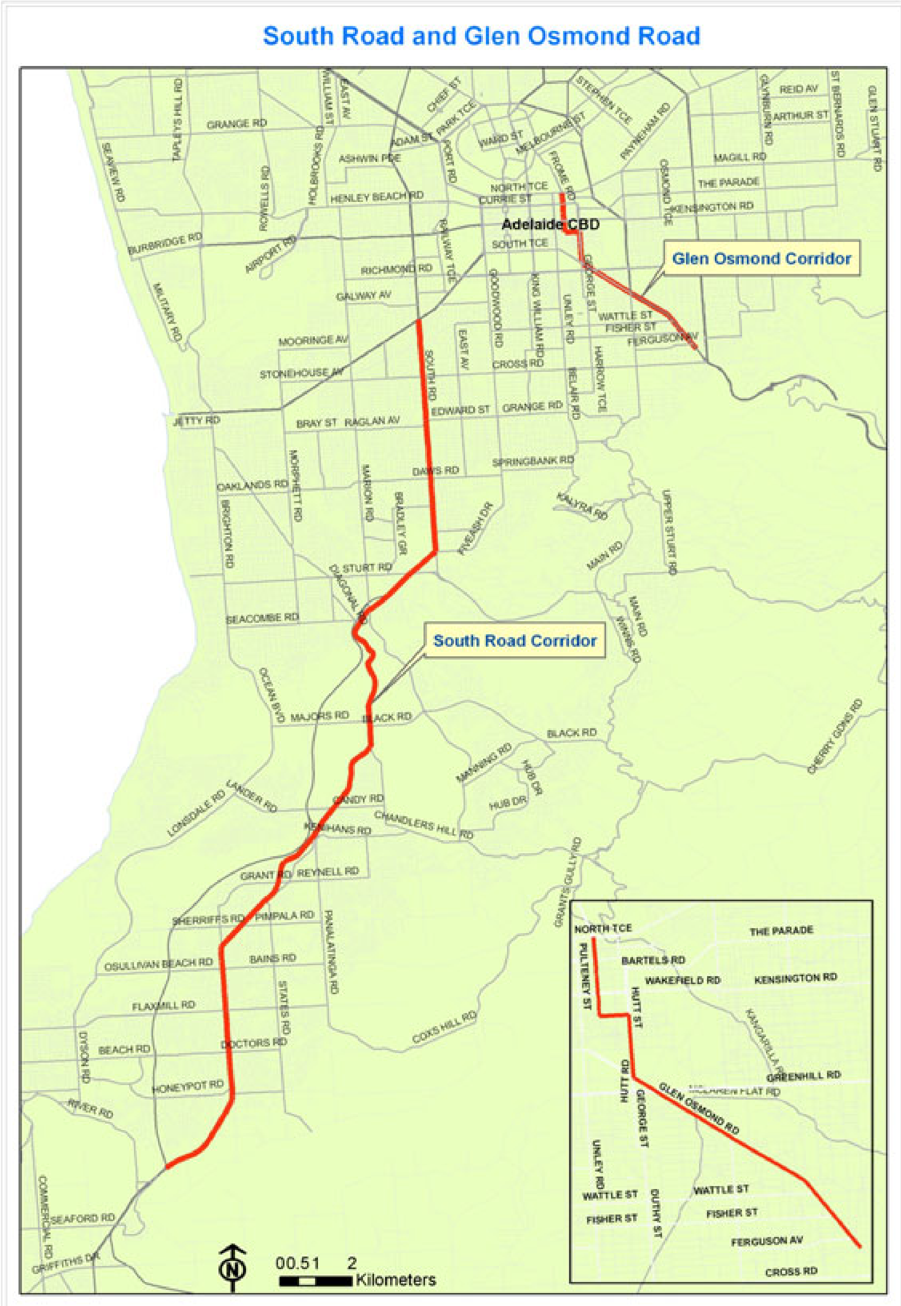

Figure 1. South Road and Glen Osmond Road study routes.

result was found for South Road, where the normal distribution fitted three of the 22 links at the 0.1 significance level (and six of the 22 links at the 0.05 significance level[4]), as shown in Table IV. Similar results were found for the lognormal distribution (see also Table IV).

The overall results confirm the inability of either the normal and lognormal distributions to adequately represent the observed data in most cases. Other distributions are required.

[4]That is, only three of the links had a better than 0.1 probability that the normal distribution fitted the observed data, whereas six of the links had a better than 0.05 probability of that. All other links had less than a 0.05 probability that the normal distribution represented their observed data.

Table I. Glen Osmond link travel time mean, standard deviation and coefficient of variation.

Link name	Link no.	Link length (m)	Mean (s)	Standard deviation (s)	Coefficient of variation
GOR: Queens Ln–Bevington Rd	1	1146	122.7	54.2	0.442
GOR: Bevington Rd–Fullarton Rd	2	1058	141.1	76.5	0.542
GOR: Fullarton Rd–Young St	3	458	38.5	21.2	0.552
GOR: Young St–Greenhill Rd	4	606	112.1	53.4	0.477
GOR: Greenhill Rd–Hutt Rd	5	331	33.0	21.1	0.640
Hutt Rd: GOR–South Tc	6	405	41.8	13.4	0.320
Hutt St: South Tc–Gilles St	7	165	23.2	13.5	0.583
Hutt St: Gilles St–Halifax St	8	150	16.5	7.3	0.446
Hutt St: Halifax St–Angas St	9	311	31.8	10.6	0.334
Angas St: Hutt St–Frome St	10	337	39.5	10.9	0.277
Frome St: Angas St–Wakefield St	11	165	45.7	26.4	0.578
Frome St: Wakefield St–Flinders St	12	165	26.7	17.2	0.644
Frome St: Flinders St–Pirie St	13	152	26.2	25.7	0.980
Frome St: Pirie St–Grenfell St	14	156	63.6	52.2	0.820
Frome St: Grenfell St–Rundle St	15	153	41.5	47.7	1.148
Frome St: Rundle St–North Tc	16	162	57.6	35.4	0.614

GOR, Glen Osmond Road.

Table II. South Road link travel time mean, standard deviation and coefficient of variation.

Link name	Link no.	Link length (m)	Mean (s)	Standard deviation (s)	Coefficient of variation
Southern Expressway II–Penney Hill Rd	1	3019	142.1	3.5	0.025
Penney Hill Rd–Honeypot Rd	2	213	14.5	5.1	0.354
Honeypot Rd–Doctors Rd	3	944	71.6	15.2	0.212
Doctors Rd–Flaxmill Rd	4	1177	100.1	26.7	0.266
Flaxmill Rd–Cannington Rd	5	710	45.2	3.9	0.086
Connington Rd–O'Sullivan Beach Rd	6	442	37.3	13.5	0.362
O'Sullivan Beach Rd–Sheriff Rd	7	1165	117.9	29.5	0.250
Sheriff Rd–Southern Expressway I	8	4008	183.8	20.2	0.110
Panalatinga Road–Lander Road	9	745	45.6	14.5	0.317
Lander Road–Chandlers Hill Road	10	1965	96.4	9.0	0.094
Chandlers Hill Road–Black Road	11	595	60.2	24.5	0.407
Black Road–Majors Road	12	135	4.6	2.9	0.625
Majors Road–Seacombe Road	13	3097	169.5	23.1	0.136
Seacombe Road–Marion Road	14	323	26.9	16.4	0.608
Marion Road–Southern Expressway	15	592	72.0	56.0	0.778
Southern Expressway–Flinders Drive	16	416	43.1	16.8	0.391
Flinders Drive–Sturt Road	17	393	53.5	18.1	0.339
Sturt Road–Ayliffes Road	18	837	69.2	41.7	0.602
Ayliffes Road–Daws Road	19	2037	300.8	158.1	0.525
Daws Road–Edward Street	20	1625	216.4	80.3	0.371
Edward Street–Cross Road	21	1209	105.8	54.1	0.512
Cross Road–Anzac Highway	22	1561	262.3	104.7	0.399

4.2. *Other distributions*

Because the travel time distributions were generally right-skewed with long upper tails, the next stage was to test other theoretical distributions that could better represent this phenomenon. This first required exploratory data analysis of the observed distributions, then comparisons with different distribution models. Thus the Burr, Generalised Pareto, Weibull and Gamma distributions were all used as candidate models.

In the first step, visualisation of the data was important. This was done by drawing histograms of observed link travel times and by superimposing a theoretical p.d.f. on the graph. Some example graphs are provided. Figure 2 shows the histogram for link 3 in the Glen Osmond data. Figure 3 shows

DOI: 10.1002/atr

Table III. Results for the goodness-of-fit test for Glen Osmond link travel time data—normal and lognormal distributions.

Link Number	Kolmogorov–Smirnov[a]			Kolmogorov–Smirnov[b] (log data)		
	Statistic	df	Sig.	Statistic	df	Sig.
1	0.222	176	0.000	0.129	176	0.000
2	0.172	176	0.000	0.108	176	0.000
3	0.329	176	0.000	0.260	176	0.000
4	0.157	176	0.000	0.162	176	0.000
5	0.395	176	0.000	0.355	176	0.000
6	0.159	176	0.000	0.169	176	0.000
7	0.227	176	0.000	0.201	176	0.000
8	0.248	176	0.000	0.207	176	0.000
9	0.262	176	0.000	0.230	176	0.000
10	0.142	176	0.000	0.145	176	0.000
11	0.115	176	0.000	0.136	176	0.000
12	0.181	176	0.000	0.154	176	0.000
13	0.264	176	0.000	0.201	176	0.000
14	0.157	176	0.000	0.198	176	0.000
15	0.273	176	0.000	0.188	176	0.000
16	0.103	176	0.000	0.105	176	0.000

[a]Fitting normal distribution to the link travel times.
[b]Fitting normal distribution to the logarithmic values of the link travel times.

Table IV. Results for the goodness-of-fit test for South Road link travel time data—normal and lognormal distributions.

Link Number	Kolmogorov–Smirnov[a]			Kolmogorov–Smirnov[b] (log data)		
	Statistic	df	Sig.	Statistic	df	Sig.
1	0.127	47	*0.057*	0.132	47	0.040
2	0.320	47	0.000	0.302	47	0.000
3	0.220	47	0.000	0.189	47	0.000
4	0.135	47	0.031	0.125	47	*0.063*
5	0.239	47	0.000	0.203	47	0.000
6	0.286	47	0.000	0.245	47	0.000
7	0.134	47	0.033	0.124	47	*0.067*
8	0.177	47	0.001	0.162	47	0.003
9	0.248	47	0.000	0.208	47	0.000
10	0.193	47	0.000	0.181	47	0.001
11	0.126	47	*0.061*	0.065	47	*0.200**
12	0.338	47	0.000	0.265	47	0.000
13	0.188	47	0.000	0.171	47	0.001
14	0.303	47	0.000	0.220	47	0.000
15	0.259	47	0.000	0.188	47	0.000
16	0.222	47	0.000	0.184	47	0.000
17	0.073	47	*0.200**	0.125	47	*0.062*
18	0.349	47	0.000	0.279	47	0.000
19	0.091	47	*0.200**	0.133	47	0.037
20	0.118	47	*0.102*	0.103	47	*0.200**
21	0.275	47	0.000	0.238	47	0.000
22	0.120	47	*0.090*	0.093	47	*0.200**

Significance values (Sig.) shown in italics are not statistically significant at the 5% level. Those marked with an asterisk (*) are clearly insignificant.
[a]Fitting normal distribution to the link travel times.
[b]Fitting normal distribution to the logarithmic values of the link travel times.

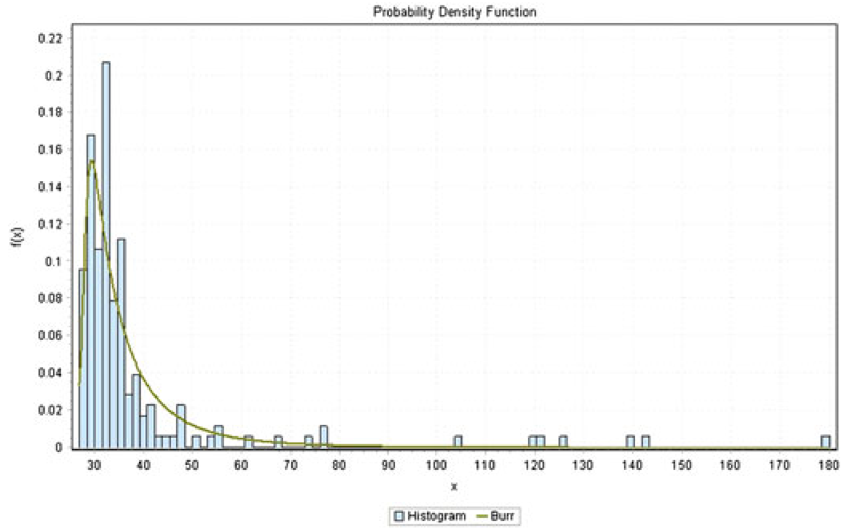

Figure 2. Histogram and fitted Burr distribution for link 3 Glen Osmond travel time data.

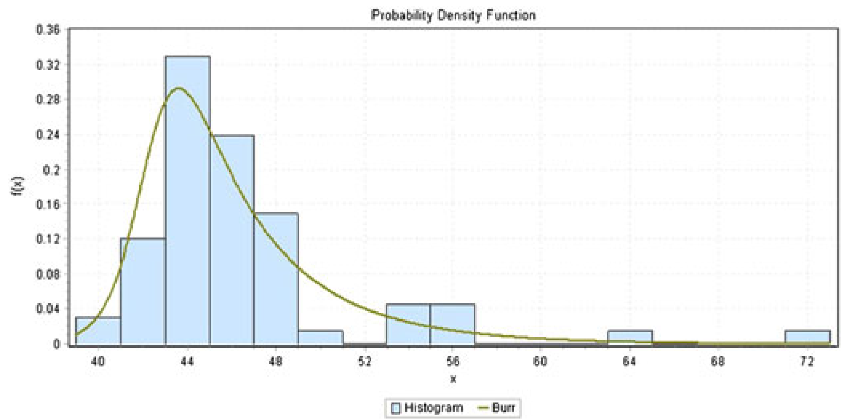

Figure 3. Histogram and fitted Burr distribution for link 5 South Road travel time data.

the histogram for link 5 in the South Road data. The theoretical curves in these figures are for the Burr distribution, fitted to each observed histogram using maximum likelihood estimation. The two plots show the inherent flexibility of the Burr distribution and its ability to replicate the long tails in the observed data.

Table V summarises the goodness-of-fit tests for the Glen Osmond data, for the Weibull distribution, Gamma distribution, Burr distribution and the Generalised Pareto distribution. The Gamma distribution was rejected for all links except link 16 (and then only accepted at the 0.01 significance level), whereas the Weibull distribution was rejected for all links except link 11 (again, accepted only at the 0.01 significance level). The Generalised Pareto distribution was rejected for a majority (10 of 16) of the links and was only accepted at 0.05 significance for two of the links. The Burr distribution was rejected for six of the 16 links and was accepted at 0.05 significance for three links and at 0.01 significance for the remaining seven links. Overall, it could therefore be seen as a plausible model for the link travel time data. A confounding factor is that several of the links on this route showed evidence of bimodality, including Glen Osmond links 4, 9, 10, 11, 12 and 14 for which the Burr distribution was rejected on statistical grounds (Table V). Figure 4 shows the observed histogram for Glen Osmond link number 10, showing indications of bimodality. The phenomenon of bimodality is explored in Section 4.3.

The results for the South Road data set were perhaps more conclusive (see Table VI). The Burr distribution fitted almost all of the links at the 0.05 significance level. Similar results were also found for

DOI: 10.1002/atr

Table V. Kolmogorov–Smirnov goodness-of-fit test results for the Gamma, Weibull, Burr and Generalised Pareto distributions fitted to the Glen Osmond link travel time data.

Link number	Glen Osmond			
	Gamma	Weibull	Burr	Generalised Pareto
1	Rejected	Rejected	Accepted	Rejected
2	Rejected	Rejected	Accepted at 0.01	Accepted at 0.01
3	Rejected	Rejected	Accepted at 0.01	Rejected
4	Rejected	Rejected	Rejected	Rejected
5	Rejected	Rejected	Accepted at 0.01	Rejected
6	Rejected	Rejected	Accepted	Rejected
7	Rejected	Rejected	Accepted at 0.01	Rejected
8	Rejected	Rejected	Accepted at 0.01	Rejected
9	Rejected	Rejected	Rejected	Rejected
10	Rejected	Rejected	Rejected	Accepted at 0.01
11	Rejected	Accepted at 0.01	Rejected	Accepted
12	Rejected	Rejected	Rejected	Accepted at 0.01
13	Rejected	Rejected	Accepted at 0.01	Accepted at 0.01
14	Rejected	Rejected	Rejected	Rejected
15	Rejected	Rejected	Accepted at 0.01	Rejected
16	Accepted at 0.01	Rejected	Accepted	Accepted

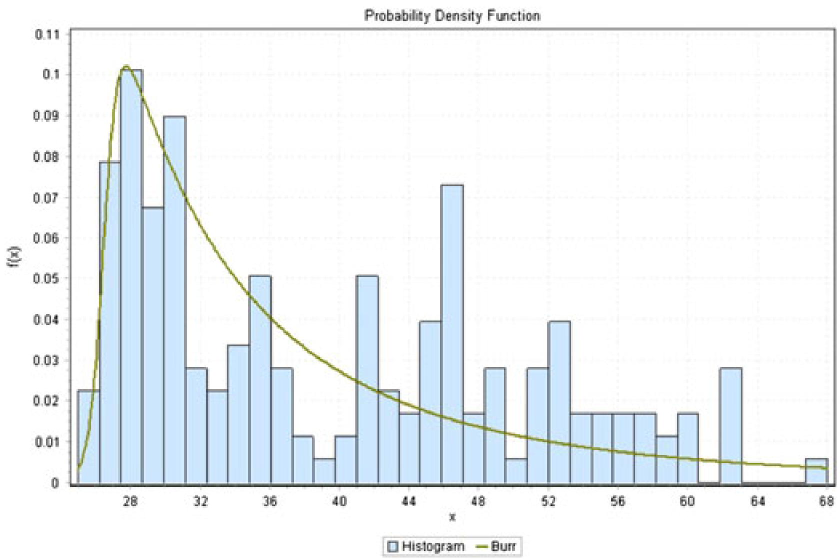

Figure 4. Histogram and fitted Burr distribution for link 10 Glen Osmond Road travel time data—indication of bimodality?

the Generalised Pareto distribution. The Gamma and Weibull distributions were each rejected for 15 of the 22 links and were only accepted at 0.01 significance for the other seven links.

Although previous studies [9,10] had suggested that the Gamma and Weibull distributions might fit the travel time distributions, our data sets gave different results. At the 0.05 significance level, the Gamma and Weibull distributions only fitted seven links of the South Road data set (see Table VI) and one link of the Glen Osmond data set (see Table V).

On the basis of the two data sets, it is reasonable to conclude that the Burr distribution can represent longitudinal travel time variability data and may be more useful in this regard than other distributions, such as the Generalised Pareto, and certainly better than the lognormal, normal, Weibull and Gamma distributions. The flexible form and attractive mathematical and computational characteristics of the Burr distribution enhance its suitability and therefore likely applications. For instance, as its c.d.f. is explicitly defined, percentile values can be computed directly.

DOI: 10.1002/atr

Table VI. Goodness-of-fit test results (Kolmogorov–Smirnov) for the Gamma, Weibull, Burr and Generalised Pareto distributions fitted to the South Road link travel time data.

Link number	South Road			
	Weibull	Gamma	Burr	Generalised Pareto
1	Accepted at 0.01	Accepted at 0.01	Accepted	Accepted
2	Rejected	Rejected	Rejected	Accepted at 0.01
3	Rejected	Rejected	Accepted	Accepted
4	Accepted at 0.01	Accepted at 0.01	Accepted	Accepted
5	Rejected	Rejected	Accepted	Accepted
6	Rejected	Rejected	Accepted	Accepted
7	Accepted at 0.01	Accepted at 0.01	Accepted	Accepted
8	Rejected	Rejected	Accepted	Accepted
9	Rejected	Rejected	Accepted	Accepted
10	Rejected	Rejected	Accepted	Accepted
11	Rejected	Rejected	Accepted	Accepted
12	Rejected	Rejected	Accepted at 0.01	Rejected
13	Rejected	Rejected	Accepted	Accepted
14	Rejected	Rejected	Accepted	Accepted
15	Rejected	Rejected	Accepted	Accepted
16	Rejected	Rejected	Accepted	Accepted
17	Accepted at 0.01	Accepted at 0.01	Accepted	Accepted
18	Rejected	Rejected	Accepted	Accepted
19	Accepted at 0.01	Accepted at 0.01	Accepted	Accepted
20	Accepted at 0.01	Accepted at 0.01	Accepted	Accepted
21	Rejected	Rejected	Accepted	Accepted
22	Accepted at 0.01	Accepted at 0.01	Accepted	Accepted

Tables VII and VIII show the estimated Burr distribution parameters for Glen Osmond and South Road link travel time data sets. The parameter estimates were obtained using maximum likelihood.

These tables clearly show that the range of fitted c parameters for both data sets is quite large. For instance the highest c value for the Glen Osmond data set is 79.05, whereas the lowest is 1.69. In this case, the lowest value occurs for an extremely high value of k, which is an outcome of the maximum likelihood computations and may therefore be seen as an outlier in that respect. The general range of the estimated c values is (16, 45). The range for the k parameter, ignoring the 'outlier' for link 16, is relatively small: (0.05, 0.12). Given the importance of the product ck in determining the statistical properties of the Burr distribution, it is interesting to consider this product (see right hand column in Table VII). Apart from the outlier, there is a high degree of consistency in the values of ck, with typical values of about 3.0.

The fitted values of c and k for the South Road data set are shown in Table VIII. This data set has two 'outliers' in terms of k values: links 17 and 20. Most of the fitted k values for the other links are in

Table VII. Parameters of the Burr distribution for Glen Osmond Road data, for those links fitted by the Burr distribution.

Link Number	Burr parameter		
	k	c	ck
1	0.05	8.96	0.45
2	0.11	16.13	1.77
3	0.09	44.71	4.02
5	0.06	52.81	3.17
6	0.04	79.05	3.16
7	0.06	27.74	1.66
8	0.12	25.15	3.02
13	0.05	29.16	1.46
15	0.06	20.93	1.26
16	42767.00	1.69	722276.23

DOI: 10.1002/atr

Table VIII. Parameters of the Burr distribution for South Road data, for those links fitted by the Burr distribution.

Link Number	Burr parameter		
	k	*c*	*ck*
1	2.470	58.249	143.88
3	0.099	46.133	4.57
4	0.068	48.187	3.27
5	0.258	49.391	12.74
6	0.191	16.979	3.24
7	0.035	94.399	3.30
8	0.062	109.490	6.79
9	0.097	51.734	5.02
10	0.097	81.603	7.92
11	1.066	5.644	6.02
12	0.251	13.191	3.31
13	0.044	161.350	7.10
14	0.171	16.699	2.86
15	0.083	16.752	1.39
16	0.054	42.015	2.27
17	49 424.000	3.061	151 286.86
18	0.068	70.856	4.82
19	4.828	2.595	12.53
20	13 791.000	3.050	42 062.55
21	0.036	74.759	2.69
22	1.656	4.852	8.03

the range (0.04, 0.1) although there are three values exceeding 1.0. The fitted c values are typically between 16.0 and 50.0, with the values of the product ck generally being between 3.0 and 8.0. Thus, there is some degree of consistency between the two data sets. It should be noted that the fitted values of c and k are affected by the choice of the measurement unit for the independent variable, being seconds for the observed link travel times in this case. As there is a wide range of link distances over the two data sets (see Tables I and II) and thus a wide range in mean travel times between links, a degree of variation in the distribution parameters is to be expected. In further research, we intend to attempt to relate the c and k parameter values to road, environment and traffic variables, to produce a method for selecting and applying suitable parameter values for applications of the Burr distribution to travel time reliability assessments.

The case of link 2 in the South Road data set is interesting. This was the only link in that data set that the Burr distribution did not fit, and the only distribution to fit the data from the link was the Generalised Pareto distribution (Table VI), and then only at 0.01 significance. The link is short in length (only 213 m, see Table II). The travel times for this link are very skewed, with a strong modal frequency at about 11 seconds and a very long tail with observations stretching out to a maximum of 30 seconds (mean observed travel time 14.5 seconds, standard deviation 5.1 seconds, sample size $n = 67$). The data are decidedly unimodal, but the length of the tail inhibits the degree of fit of the candidate distributions. Figure 5 shows the fitted distributions plotted with the observed histogram. Whilst the Burr distribution is not a good statistical fit, it is able to capture most of the shape of the observed histogram and indeed matches the modal frequency.

The analysis presented earlier is for individual links in the route. Similar results were also found in the assessment of travel time variability at the overall route level. In this case, the Weibull, Gamma, Burr and Generalised Pareto distributions were fitted to the overall route travel times. Goodness-of-fit tests were conducted, and the test results are shown in Table IX. These distributions were quite successful in representing the observed route travel time distributions particularly in relation to positive values and long tails. However, the goodness-of-fit tests indicated that only the Burr and Generalised Pareto distributions gave promising results. The Weibull and the Gamma distributions did not fit the Glen Osmond corridor at all. For South Road, however, the Weibull and Gamma distributions were accepted. The Burr distribution fits both data sets well, further supporting the notion that this distribution could be a useful model of travel time variability. Figure 6 shows the Burr c.d.f. and the observed

DOI: 10.1002/atr

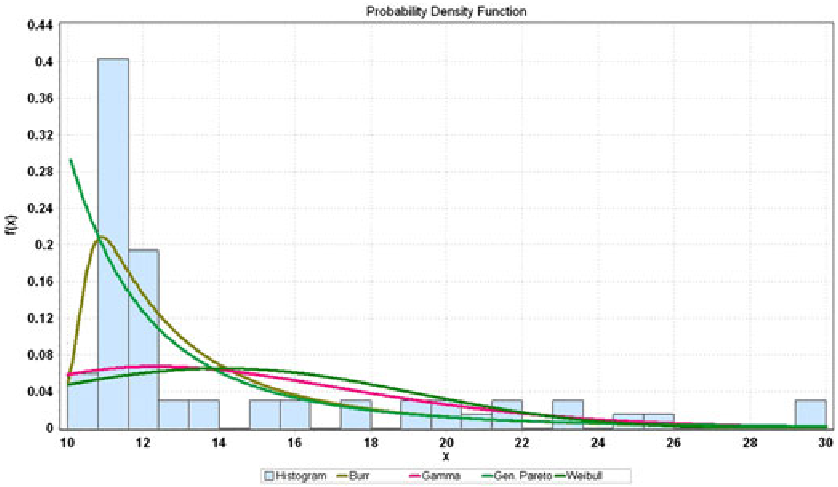

Figure 5. Fitted theoretical distributions (Burr, Gamma, Generalised Pareto and Weibull) for the observed travel time data for link 2 in the South Road data set.

Table IX. Goodness-of-fit tests for overall travel times on the two routes.

Route	Significance level	Significance value	Computed Kolmogorov–Smirnov statistic			
			Weibull	Gamma	Generalised Pareto	Burr
Glen Osmond	0.05	0.10150	0.16771	0.14573	0.09107	0.05778
	0.01	0.11346	0.16771	0.14573	0.09107	0.05778
			Rejected	Rejected	Accepted	Accepted
South Road	0.05	0.16322	0.06134	0.07707	0.09000	0.05666
	0.01	0.18252	0.06134	0.07707	0.09000	0.05666
			Accepted	Accepted	Accepted	Accepted

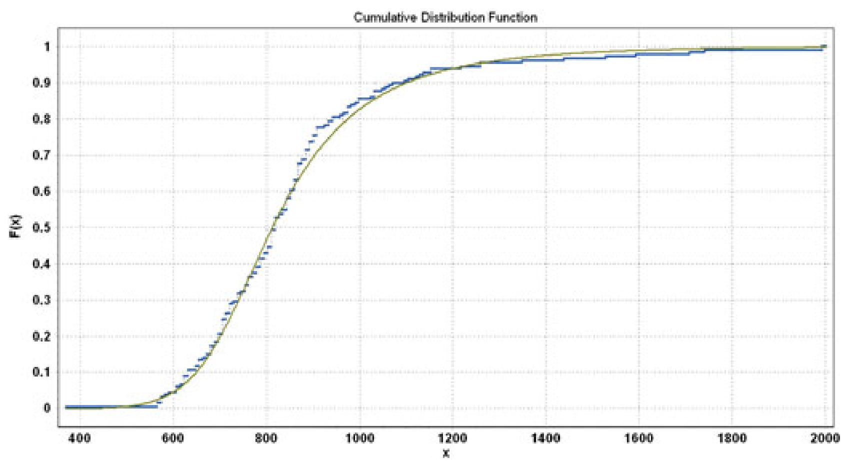

Figure 6. Burr distribution and observed cumulative density functions for the Glen Osmond route travel times.

c.d.f. for the Glen Osmond route. Figure 7 shows the corresponding plots for South Road. The ability of the Burr distribution to model the long upper tails of the observed distributions is evident in these graphs.

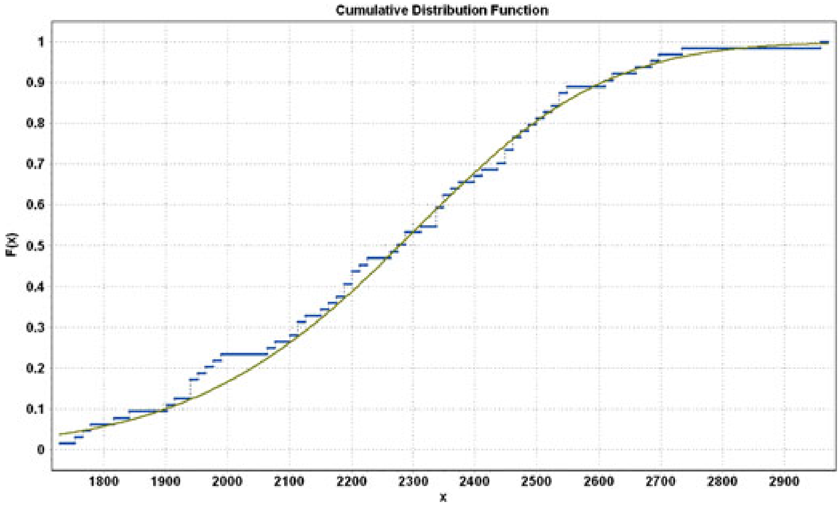

Figure 7. Burr distribution and observed cumulative density functions for the South Road route travel times.

4.3. Bimodality

The issue of bimodality as suggested in Figure 4 is of interest and was considered for all those links in the Glen Osmond data set, which could not be represented by a unimodal (Burr) distribution (as seen in Table V). If f(x) is the p.d.f. of a bimodal distribution comprising two component unimodal distributions $f_1(x)$ and $f_2(x)$, then it can be described mathematically as

$$f(x) = \phi_1 f_1(x) + (1 - \phi_1) f_2(x)$$

where ϕ_1 is the proportion of the overall distribution belonging to $f_1(x)$. The corresponding c.d.f. F(x) is given by

$$F(x) = \phi_1 F_1(x) + (1 - \phi_1) F_2(x)$$

Determination of the split of observed values of x between the two component populations, of the value of ϕ_1, and the resulting values of the parameters describing distributions $f_1(x)$ and $f_2(x)$ is a major issue. One approach to test for bimodality is the Hartigan DIP test [16]. The DIP statistic measures the maximum difference between an empirical c.d.f. and the unimodal c.d.f. that minimises that maximum difference. It produces a probability that the observed data could come from that unimodal distribution and can be used to test the null hypothesis that the data are unimodal. *The test was applied to those links in the Glen Osmond data set, which the Burr distribution did not fit*, with a bimodal normal distribution used as the theoretical model. It indicated that five of the 16 links (i.e. links 4, 9, 10, 11 and 12) in the Glen Osmond data showed statistical evidence of bimodality at the 0.1 significance level (see Table X). These tables show the probability of rejection of the null hypothesis, and the estimated means, standard deviations and proportions of the component distributions.[5]

Statistical evidence is useful, but an explanation of the phenomenon of bimodality in travel time variability distributions is also required. For the case of urban arterial roads, the influence of delays at traffic signal may provide an explanation. For instance, experiencing two or more red signal phases at an intersection may substantially increase link travel times. On already congested sections or routes, the queuing delay then experienced by drivers could be similar to or even exceed the running time needed to traverse the link, thus doubling or even tripling the total link travel time. On the other hand, experiencing less queuing at signalised intersections will substantially reduce the total travel time. This

[5]However, it must also be noted that neither the Burr distribution nor the Bimodal Normal distribution fitted the observed data for link 14 in the data set. Therefore, further investigation may be needed for that particular link.

DOI: 10.1002/atr

Table X. Bimodality results for Glen Osmond link travel times, including DIP statistic, means and standard deviations of component populations, and proportions.

Link number	DIP statistic	Probability	Component population 1			Component population 2		
			Mean1 (s)	SD 1 (s)	Proportion (ϕ_1)	Mean2 (s)	SD 2 (s)	Proportion ($\phi_2 = 1 - \phi_1$)
4	0.0693	0.999	102.9	35.7	0.92	219.6	93.8	0.08
9	0.0618	0.990	26.9	2.7	0.77	47.5	11.5	0.23
10	0.0393	0.950	28.5	1.4	0.33	44.9	9.4	0.67
11	0.0370	0.900	13.5	1.5	0.22	54.9	22.7	0.78
12	0.0450	0.900	13.4	1.5	0.39	35.1	17.3	0.61
14	0.0910	–	na	na	na	na	na	na

result was found by Davis and Xiong [17], who also observed bimodality in travel time distributions and were able to ascribe this to signal performance. We suspect that similar factors may apply in the Adelaide data sets, and this is the subject of ongoing research.

According to Ko and Guensler [18], the mixing parameter of the Bimodal Normal distribution in speed distribution analysis may be used to explain the state of traffic congestion. The first subgroup in the bimodal distribution might represent congested traffic, and the second is then uncongested traffic. A similar interpretation may also be applied to travel time analysis. The first subgroup of the distribution might represent freely flowing traffic or traffic near the free flow condition, whereas the second subgroup illustrates congested traffic with very long travel times. Therefore, a vital step is to determine the parameters for each data set. These parameters will describe the behaviour of the data and can also indicate additional factors that might affect the shape of the travel time distribution. As the bimodal travel time distributions mostly occurred in the Glen Osmond data set, the subsequent discussion focuses on that data set and in particular for those links for which the Burr distribution had been rejected. The results of the maximum likelihood parameter estimation for these links are given in Table X. In all cases, the first component group is that with the smaller mean value.

By examining the difference in the means and in the standard deviations of the two component populations in the bimodal speed distribution, Jun [19] postulated how these differences could be used to support the hypothesis of bimodality. This discussion could then lead to further tests to assess whether the bimodal speed, and travel time distributions are actually built up by two different populations—the short and long travel time population. For instance, Jun [19] suggested that the mean speed difference should be more than 32.2 km/hour (20 mph) in order to indicate a distribution based on two separate populations. Consequently, this current study needs to establish what the mean and the standard deviation differences would be for a fully bimodal distribution of travel times.

The means of the second component group seems to be double or even four times higher than those of the first group (see Table X). The standard deviations of the second component group are also much larger than the standard deviations of the first group.

Additionally, the mixing parameter for the Glen Osmond Road data varies between 0.2 and 0.9. One interesting finding is that the mixing parameter for link number 4 in Glen Osmond Road is 0.92, which means that, for this link, 92% of the travel time data belong to the first component group and only 8% belong to the second group (see Table X). For link 9, the second component group forms about a quarter (23%) of the total population, whereas for links 10, 11 and 12, the second component group forms a majority (between 61% and 78%) of the population.

The other factor that might induce bimodality in the urban road links is link length, as those links that have larger probabilities of bimodality tend to be the shorter links. For instance, in the Glen Osmond Road data set, links 9, 10 and 11, 12, 13 and 14 are short links with link lengths less than 320 m.

The short link phenomenon may be seen in parallel with our first assumption that the bimodal travel time might occur because some survey runs experienced two or more traffic signal cycles on the congested short link, whereas the normal experience might require only one cycle. Thus, we can say that the link length might be one contributing factor for the bimodal travel time distribution. The next section checks whether the coordinated signal setting in urban areas might contribute to these phenomena and discusses a possible way to overcome this.

The data collected in this study came from two urban arterial roads. Arterial roads are operationally and technically different to freeways, for which most previous data have been collected (e.g. [9] and [20]). Traffic movement on urban arterial roads is affected by the traffic signal settings, parking and bus stops. Additionally, each intersection has its own traffic signal settings based on road type, the intersecting volumes including turning movements, and physical road capacity. Therefore, another factor that might influence the bimodal distribution is traffic signal setting.

Our second assumption is that the bimodality occurred because of the large differences in travel times on short links depending on whether the test vehicle arrived during a green phase or during a red phase at the downstream signals. Therefore, a second test is to merge some consecutive short links into one longer link. We propose that the longer link will eliminate the observed effects of travel through a number of intersections along the link and could even out the overall impacts of delays at any one intersection.

Traffic signal settings, route data and link diagrams were obtained from the Transport SA Traffic Control Centre. From these, we selected two or more consecutive links in the Glen Osmond Road data set that travelled in the same direction without any right or left turn manoeuvres and then merged those links into single links. In addition, we also considered the type of signal setting priority for the roads and the link coordination system and checked whether those consecutive links were under the same control system; if so, then we grouped those links into one single link.

The selected links in Glen Osmond Road were merged into three single links based on the data of the coordinated signal setting. Table XI shows the intersection names and the controlling intersections along the Glen Osmond Road study route.

With the controlled intersection data, the new link configuration is shown in Table XII. New bimodality tests were conducted using the new link combination data set. A 0.90 probability of bimodality was found for link combination 1, whereas there was only a 0.01 probability for link combination 2 (see Table XII.)

The probability of 0.90 for bimodality means that the bimodal distribution was still apparent for link combination 1. For link combination 2, the 0.01 probability indicates that the bimodality was broken up. Additionally, Figure 8 illustrates the resulting histogram for link combination 2, which clearly shows the strong skew in the upper tail and is unimodal as observed for other sections of the travel time route. Thus, new tests were conducted to fit the Burr distribution to these link combinations. Thus, Table XII also shows the results of the goodness-of-fit tests. At the 0.05 significance level, link combination 1 did not fit the Burr distribution, as expected from the bimodality analysis. On the other hand, the Burr distribution provided a statistically significant fit for link combination 2.

Thus, the conclusion that can be drawn from this test is that the merged travel times might be helpful in the data analysis for short links, which were under the same controlled signalised intersection, but

Table XI. Intersection names and controlling intersection for the Glen Osmond route.

Intersection no.	Link no.	Intersection name	Controlling intersection	Controlled by
TS 0374	1	Bevington Rd/Glen Osmond Rd		TS 0093
TS 0093	2	Fullarton Rd/Glen Osmond Rd	CI	TS 0093
TS 0220	3	Young St/Glen Osmond Rd		TS 0093
TS 0069	4	Greenhill Rd/Glen Osmond Rd		TS 0093
TS 176	5	Hutt Rd/Glen Osmond Rd		TS 0093
TS 3059	6	South Tc/Hutt Rd		TS 3073
TS 3105	7	Gilles St/Hutt St		TS 3073
TS 3073	8	Halifax St/Hutt St	CI	TS 3073
TS 3081	9	Angas St/Hutt St		TS 3073
TS 3091	10	Frome St/Angas St	not coordinated	
TS 3069	11	Wakefield St/Frome St	CI	TS 3069
TS 3066	12	Flinders St/Frome St		TS 3069
TS 3062	13	Pirie St/Frome St		TS 3069
TS 3056	14	Grenfell St/Frome St		TS 3069
TS 3038	15	Rundle St/Frome St		TS 3069
TS 3037	16	North Tc/Frome St	CI	TS 3037

CI, controlled intersection.

DOI: 10.1002/atr

Table XII. The link combinations for the Glen Osmond data set, the DIP statistic and the result of Kolmogorov–Smirnov goodness-of-fit test.

Previous link no.	Link combination	DIP statistic	Probability	Goodness-of-fit test at 0.05 significant level (Burr)
3	1	0.044	0.90	Rejected
4				
11	2	0.017	0.01	Accepted
12				
13				
14				
15				

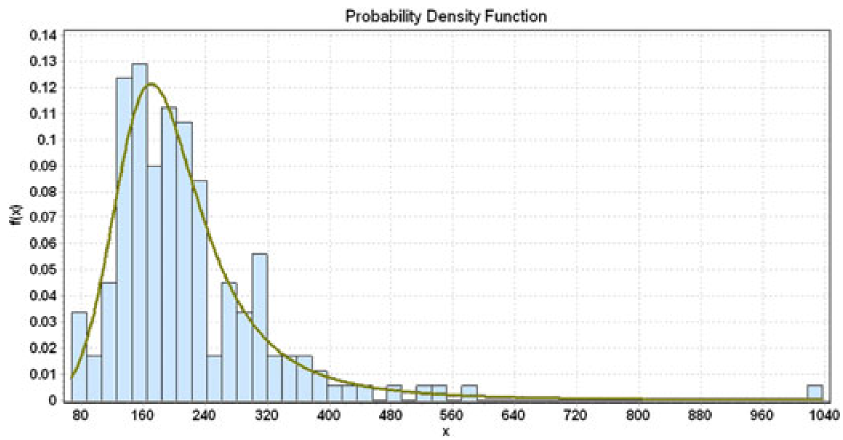

Figure 8. Histogram and fitted Burr distribution for link combination 2.

for some longer links such as link 3 and 4, the merger of two or more consecutive links did not change the occurrence of bimodality. Therefore, future research using new methods to interrogate historical data from urban traffic control systems [21] coupled with the continuous (time stamped) data from the GPS runs will address this issue.

5. CONCLUSIONS

This paper has approached the question of travel time reliability by considering two separate sets of longitudinal travel time data sets from arterial road routes. The search is for a tractable model that can reasonably represent observed variations in day-to-day travel times and thus can provide a statistical model for the analysis of travel time reliability. The observed travel time distributions are characterised by very long upper tails and strong positive skew. Analysis of these data sets led to the conclusion that the lognormal, Weibull and Gamma distributions, although having the characteristics of positive skew and reasonably long upper tails, were unable to fully represent the observed data. Therefore, the research focused on other continuous distributions that could accommodate those patterns. The Burr distribution was considered as a leading candidate for travel time variability, with some other distributions also suggested. The Burr and Generalised Pareto distributions emerged as reasonable models, for both links and routes. However, in terms of overall performance, the Generalised Pareto was less able to represent the characteristics of the observed travel time distributions. The Burr distribution was able to provide good overall representation of the observed data. Given the attractive features of this distribution in terms of its mathematical tractability and its flexibility, this distribution can be proposed as a useful model of variations in travel times. We believe that this finding has important implications for the development and use of metrics for travel time reliability, which are likely to

DOI: 10.1002/atr

be based on considerations of travel time distribution percentiles [22], because of the significant skew in the distributions. The Burr distribution is especially well able to represent the skew, and percentile values can be easily computed for it.

However, for some cases, the Burr distribution fails to fully portray the travel time variability data, especially for those links (such as found in the CBD), which are in grid systems and are short in length. Travel time data for short links tend to be bimodal. Therefore, this study tries to look for the appropriate bimodal distribution for those links and to examine potential factors that might affect the occurrence of bimodality. The initial results suggest the hypothesis that link length and coordinated signalised intersection may be the factors that contributes to bimodality.

The findings and the result from the data analysis show us that bimodality in the travel time distribution can be tested by using some existing procedures. Maximum likelihood estimation can be used to generate the bimodal normal distribution to help us in exploring the behaviour of each travel time distribution. What we had in this study is the case where the differences of two means and the differences of two standard deviations are quite large. For some links, the second mean of the link is twice of the first mean, and for some other links, the second mean could reach triple or even four times the value of the first.

The result show that the shorter links tend to have bimodal travel time distribution and these phenomena will be broke up when those short links have been merged into one single longer link based on the coordinated signal setting. The result is not end of the story as there is further scope to look at factors such as the effect of the SCATS (Sydney Coordinated Adaptive Traffic System) traffic control system on the travel time data. There will then be the issue of the development of suitable models for bimodal travel time variability distributions, most likely using mixture models.

Additionally, further research is also required to develop appropriate general Burr parameters that can characterise the variability of urban arterial road travel times and to relate those parameters to environmental and operational factors for road corridors.

REFERENCES

1. Small KA. The scheduling of consumer activities: work trips. *The American Economic Review* 1982; **72**(3): 885–915.
2. Fosgereau M, Karlstrom A. The value of reliability. *Transportation Research B* 2010; **44**(1):38–49.
3. Zhang Y, Liu YC. Analysis of peak and non-peak traffic forecasts using combined models. *Journal of Advanced Transportation* 2011; **45**(1):21–37.
4. Soriguera F, Robuste F. Highway travel time accurate measurement and short-term prediction using multiple data sources. *Transportmetrica* 2011; **7**(1):85–109.
5. Lam WHK, Chan KS, Li ZC, Bell MGH. A risk-averse user equilibrium model for route choice problem in signal-controlled networks. *Journal of Advanced Transportation* 2010; **44**(4):219–230.
6. Wardrop JG. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers* 1952; **1**(2):325–379.
7. Herman R, Lam T. Trip time characteristics of journeys to and from work. In Buckley, D.J. (ed.) *Transportation and Traffic Theory*, A.H. and A.W. Reed, Sydney, 57–85, 1974.
8. Richardson AJ, Taylor MAP. Travel time variability on commuter journeys. *High Speed Ground Transportation Journal* 1978; **6**:77–99.
9. Polus A. A study of travel time and reliability on arterial routes. *Transportation* 1979; **8**:141–151.
10. Al-Deek H, Emam EB. New methodology for estimating reliability in transportation networks with degraded link capacities. *Journal of Intelligent Transportation Systems* 2006; **10**:117–129.
11. Zimmer WJ, Keats JB, Wang FK. The Burr XII distribution in reliability analysis. *Journal of Quality Technology* 1998; **30**(4):386–394.
12. Burr IW. Cumulative frequency functions. *Annals of Mathematical Statistics* 1942; **13**:215–232.
13. Tadikamalla PR. A look at the Burr and related distributions. *International Statistical Review* 1980; **48**:337–344.
14. Abdel-Ghaly AA, Al-Dayian GR, Al-Kashkari FH. The use of Burr Type XII distribution on software reliability growth modelling. *Microelectronics and Reliability* 1997; **37**:305–313.
15. Soliman AA. Estimation of parameters of life from progressively censored data using Burr-XII model. *IEEE Transactions on Reliability* 2005; **54**:34–42.
16. Hartigan JA, Hartigan PM. The dip test of unimodality. *The Annals of Statistics* 1984; **13**(1):70–84.
17. Davis GA, Xiong H. *Access to destinations: travel time estimation on arterials*. Minnesota Department of Transportation, 2007.
18. Ko J, Guensler RL. *Characterization of congestion based on speed distribution: a statistical approach using Gaussian mixture model*. Transportation Research Board, National Research Council. Washington DC, 2005.

DOI: 10.1002/atr

19. Jun J. Understanding the variability of speed distributions under mixed traffic conditions caused by holiday traffic. *Transportation Research C* 2010; **18**(4):599–610.
20. Elliason J. *Forecasting travel time variability*, *Association of European Transport*, 2006.
21. Zhang K, Vogiatzis N, Taylor MAP. A new design for an intelligent event-responsive urban traffic management system. *Papers of the 30th Australasian Transport Research Forum*. Melbourne, September 2007 (http://www.patrec.org/atrf.aspx)
22. OECD. *The reliability of travel time. Research report*. Joint Transport Research programme, Organisation for Economic Cooperation and Development, Paris, 2010.

DOI: 10.1002/atr