Australasian Transport Research Forum 2023 Proceedings
29 November – 1 December, Perth, Australia
Publication website: http://www.atrf.info

Case Study: Validating standard deviation as a reliability measure for high frequency buses

Mathew Hounsell (he/him)

School of Civil and Environmental Engineering at the University of Technology Sydney.

Abstract

This paper examines empirical data from public bus operations in Sydney, to assess if the mean end-to-end runtime, and its standard deviation, are useful measurements to monitor the reliability and efficiency of the services delivered. This paper shows that the mean plus two standard deviations for end-to-end runtimes was an effective estimator of the 97th percentile of runtimes for the route studied during both headway-based and timetable-based operations.

1. Introduction

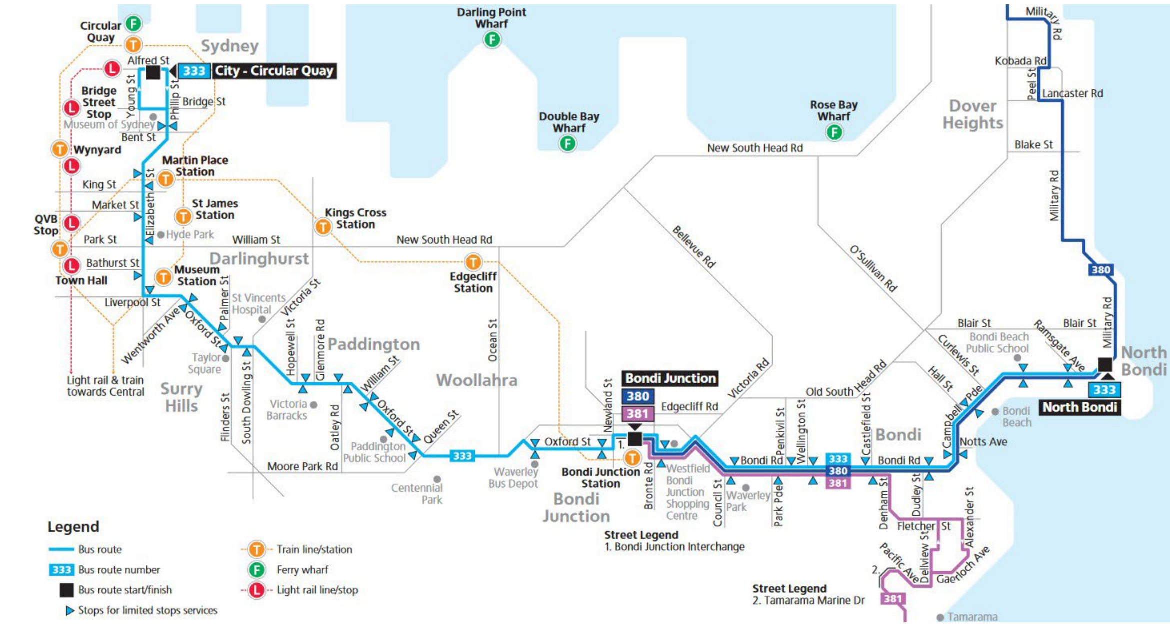
This paper is an empirical examination to demonstrate that the Run Time Variability (RTV) of high-frequency public transport routes provided by buses in mixed traffic can be planned and assessed using the 97th percentile, mean, and standard deviation measurements. In order, to allow partners to deploy approaches like Statistical Process Control (SPC) to monitor and control transit operations more efficiently than is possible using On-Time Running, this paper builds on previous work, by demonstrating those techniques apply to regular buses and not just trains operated in dedicated carriageways or trams operating in segregated lanes.

Hounsell (2022b) and Hounsell (2022a) are a pair covering the empirical examination of the unique case study context with the more general theoretical examination — addressing the case study duality criterion (Gammelgaard, 2017; Jacoby, 1976, 1978). Hounsell (2022b, p. 12) concluded that for Sydney's Metro and Light Rail the '*the mean end-to-end runtime* ($\bar{\mathbf{e}}$), *and the standard deviation* ($\boldsymbol{\sigma_e}$) *are useful measurements and provide reasonable estimators to measure the reliability and efficiency of the public transport services delivered.*' Efficiency being the ratio of runs delivered to the deployed fleet size. Those papers built on an empirical Big Data examination of end-to-end runtimes on Sydney's 333 bus route in (Hounsell, 2018b) and the theoretical discussion of standard deviation usage in Hounsell (2018a).

This paper returns to examine Sydney's 333 bus route to demonstrate that the same approach can be applied to buses. Transport for New South Wales (TfNSW) 333 bus route runs inbound the 9km from the Campbell Parade tram terminus (Stop #200634) at North Bondi via Bondi Junction station Interchange to Elizabeth St near Martin Place Station (#2000421) in the Sydney Central Business District (CBD) and then to Circular Quay before returning outbound to Bondi. In November 2016, the 333 was one of the most popular bus routes in New South Wales (NSW) with the corridor having 25,000 embarkations on an average workday (Hounsell, 2018b, pp. 1-2; 2020, § 1.5.2). The 333 is scheduled as a high-frequency service across the route throughout the day. TfNSW currently withholds individual bus route patronage from the taxpayers, despite that data being available for other modes.[i]

Kimpel et al. (2008) states Portland's Tri-Met aims to maintain bus headways. '*When [planned] headways are <10 min: buses travel at maximum speed, provided that maximum headways are not violated and no passengers are left at stops*'; then '*in off-peak periods, or when [planned] headways* *are >10 min, schedules are strictly maintained and synchronized at major transfer points.*' That paper describes many useful operator-focused metrics. It notes that '*It has taken approximately two decades for [Portland] to fully transition from a data poor to a data rich environment*' using Automatic Vehicle Location (AVL) and Automatic Passenger Counting (APC) technologies. Most of the day, Sydney's 333 operates with planned headways less than 10 minutes and now uses a headway-adherence Key Performance Indicator (KPI). As such, the 333 operator aims to flexibly run buses at their maximum speed (so as to minimise the deployed fleet size), rather than aiming to deliver an exact timetable. Thus, the running times of this service behave naturally and may fit a normal distribution.

Figure 1: Diagrammatic map of the 333 bus-route from the Sydney CBD to Bondi Beach



Source: TfNSW (2023-03-24) Timetable: 333 North Bondi to City Circular Quay via Bondi Junction. Transport for NSW, State of NSW, https://transportnsw.info/documents/timetables/30-333-North-Bondi-to-City-Circular-Quay-via-Bondi-Junction-20230130.pdf

When measured, key parameters of a natural system will often have a normal distribution; therefore, measuring and controlling variance has proven to be a practical approach for achieving economy and efficiency in soft systems. Since the early twentieth century, the business community has developed the principles of Statistical Process Control (SPC); e.g., the work of W.E. Deming (Deming, 1982, 2018). SPC developed into Six Sigma (Basu, 2009; Brussee, 2006; Sameni, 2012; Vaidya, 2018), which is often paired with the Lean waste reduction philosophy (Shah & Ward, 2007; Smith, 2014; Suárez-Barraza et al., 2012).

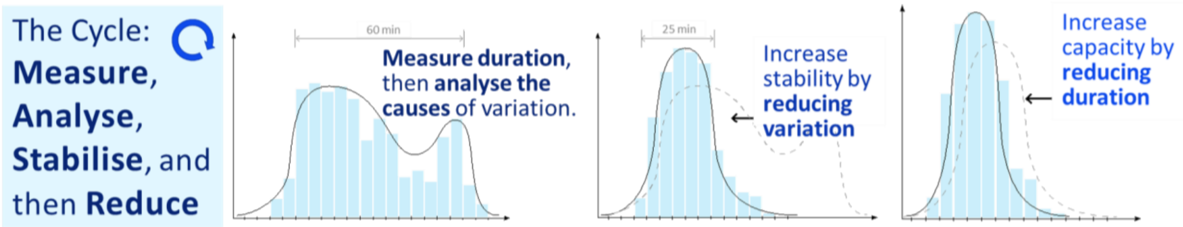
The SPC approach monitors the mean ($\bar{e}$) and the standard deviation (SD) ($\boldsymbol{\sigma_e}$) of the normally distributed parameters, to ensure the output of the service remains within the quality tolerances of the process 84%, 97%, or 99% of the time (1, 2, or 3 SD). SPC requires the monitoring of inputs, operations, and outputs. A comprehensive view allows the process to be observed and the root-cause of problems to be determined (Barr, 2017, 2019).

If the below hypothesis for real world transit operations from Hounsell (2022b, p. 2) applies to high frequency bus routes, then SPC approaches (such as controlling variation) may assist transit service providers in improving the service quality delivered (CEN, 2002, p. 6):

> the mean end-to-end runtime ($\bar{e}$) plus two standard deviations ($\boldsymbol{2\sigma_e}$) **provides a reasonable estimate** for the 97th percentile of the observed end-to-end runtimes (**e**).

This papers builds on the continuing research programme where UTS has demonstrated that transit efficiency can be improved through the application of the "Measure, Analyse, Stabilise, and Reduce" (MASR)[ii] framework shown in Figure 2 (Martin, 2018; Mileusnic, 2017; Samra, 2017; Smalley, 2017; Zeibots, 2016).

Figure 2: MASR — Runtime variability impacts services targets and service delivery



1.1 Literature review

Travel time estimation and prediction using vehicle-based data can be classified into data-driven & model-based methods, as in Ramakrishnan et al. (2018) and Hounsell (2020, p. 7).

Noland and Polak (2002) provides a literature review; they note variability can be considered a problem of increased costs, of reduced certainty, or as a problem of innate discomfort.

The primary research paradigm assumed that travellers make mode, route, and departure-time choices based on perceptions & expectations of travel time & travel time reliability aiming to minimise the risk of late arrival at their destination — '*A transit user's goal is to arrive at his or her trip destination (e.g., work and school) at a target time.*' (Gittens & Shalaby, 2015, p. 94). No papers reviewed considered the perspective of recreational or hedonistic travellers.

Harsha et al. (2020) states '*Passengers lose their confidence in [their choices of] departure time, routes, modes of transport, etc., due to day to day travel time variability in transport systems [causing] unpredictable waiting time, in-vehicle time, transfer time, etc.*' Gittens and Shalaby (2015) states that the passengers' '*recollection of experiences tends to be based on extreme feelings and how the experiences end … the "peak-end" phenomenon. The implication is that the traveler's perception of reliability would be better expressed through extreme values rather than average values, which raises the question, how extreme? The 95th percentile has been proposed [by several sources, after] assuming that a commuter could accept being late for work one day per month, or 5% of the time.*' They state that '*The literature shows that on-time arrival of passengers at travel destinations, short wait time at the origin stop, and low variability in wait and travel times have the greatest influence on travelers' perceptions of reliability.*'. They too propose another new measure: the interesting Journey Time Buffer Index (JTBI).

Historically, the ability to measure and monitor operational performance has been limited by data availability and the cost of manual supervision (Lam, 2000; Levinson, 1991). However, the deployment of vehicle loop detectors, electronic tolling, number plate recognition, Electronic Ticketing System (ETS), and Global Position System (GPS) based AVL has significantly reduced the cost of data acquisition leading to the creation of large rich dataset, which have expanded the research possibilities. Previously researchers examined the distribution of car driver and transit passenger travel times in the ordered traffic of the United States and the United Kingdom (respectively); however, recent research has been undertaken on the different traffic in developing countries such as China, as well as on the laissez-faire traffic of India (Chepuri et al., 2020; Harsha et al., 2020).

Johnson (1966) is a historically precedential paper with a model for the utility derivable from time usage (especially travel). That is expanded by Small (1982), with a model to cover the scheduling of activities (especially travel). Noland and Small (1995) is a well written paper that proposes a "mean-variance" cost model of the utility of travel time variability; the paper focuses on motorways using a uniform model for arrivals and an exponential model for accidents. Elíasson (2007) and (Eliasson, 2009) then uses Noland & Small's model on motorways and conclude that highly congested links are normally distributed, but occasionally congested links are right skewed. Carrion and Levinson (2012) is a review focusing on the economic models of motorway travel time; they categorise travel time as *free flow travel time (FFTT)* and *additional time* (subdivided into *predictable* and *unpredictable variation)*. They conduct an unfortunately inconclusive meta-analysis to determine the Value of Reliability (Carrion-Madera et al., 2012; Carrion-Madera et al., 2011). Taylor (2013) is a useful and comprehensive literature review for reliability of motorway traffic; they note a taxonomy of variations in travel times consistent with SPC: *random variations*, *regular condition-dependent variations*, and *irregular condition-dependent variations*.

Li et al. (2016) support conventional route choice models by developing economic models based on utility maximisation and Stated Preference (SP) surveys. To examine the value of reliability on motorways, Bates et al. (2001) derive a utility function from a lognormal distribution (to reduce computational complexity). For analysing variation, they prefer the median to the mean and the 90th percentile to the SD. Gan and Bai (2014) and (Batley et al., 2019) investigated SP and abstract away the specifics of the distribution. Tilahun & Levinson (2010) provides several useful economic choice models of transit lateness and reliability; but begin by assuming the timetable is natural. The economic models assumed all travellers had a fixed-arrival time, like commuters. Other papers focused on practically applicable models.

van Lint and van Zuylen (2005) considers two measures λ^{skew} and λ^{var} based on percentiles rather than mean and variance. Their results are then used to develop mesoscopic models using their methods of estimating skew and standard deviation, see also Zheng et al. (2018). Yazici et al. (2012) applied their methods in a reproduction for New York City.

Noland and Polak (2002) provides a literature review; they note variability can be considered a problem of increased costs, of reduced certainty, or as an innate problem. Some papers are descriptive, such as Lyman and Bertini (2008) who focus on skewed motorway reliability; they discuss metrics such as 95th Percentile Travel Time, Travel Time Index, Buffer Time Index (BTI), Planning Time Index (PTI), and Frequency. Lomax et al. (2003) is an easy-to-read summary of more performance measures, such as Travel Time Window, Variable Index, Travel Rate Envelope, Florida Reliability Method, On-Time Arrival, and the Misery Index; they consider normal distributions but prefers a log-normal distribution. Other papers discuss several models for highly skewed road distributions, including normal, log-normal, normal-mixture, exponential, Gamma, Gamma Mixture, Weibull, Generalised Pareto, Generalized Exponential, and Singh-Maddala (Aron et al., 2014; Clark & Watling, 2005; Fosgerau & Karlström, 2010; Guessous et al., 2014; Hollander & Liu, 2008; Lam, 2000; Li et al., 2015; Rakha et al., 2010; Uno et al., 2009). Many of these papers use the Coefficient of Variation (percent-variation). Puvvala (2014) examines reliability using the Cronbach's Alpha (α). Some papers even consider the application of Chaos Theory (Cartwright, 1991; Frazier & Kockelman, 2004). Vlahogianni and Karlaftis (2011) address the fact that many transport papers ignore time-series issues of stationarity and volatility (change in variability over time).

Chen et al. (2018) proposed that the predictable motorway travel time variations (both regular and irregular condition-dependent variations) could be modelled as different states at various locations. They found Gamma and Weibull distributions fitted some states and locations the best, but a lognormal distribution fitted more states and locations overall. Li et al. (2013) found similarly; except they demonstrated that with sufficient data, the estimates for average road travel times from a normal distribution were reasonable. Chen and Fan (2020) undertook additional empirical analysis using Weibull, Gamma, Burr, and lognormal distributions for Charlotte, North Carolina with respect to time-of-day, day-of-the-week, and the weather. Kieu et al. (2015) tests normal, lognormal, gamma, and Burr distributions for major bus routes in Brisbane, preferring the median not the mean. Susilawati et al. (2013) is an important paper focused on Adelaide. It argues for the use of Burr distributions for motorways, and discusses the bimodality of arterials, especially those with signals. The results of this paper indicate that no distribution truly matches the real world. Ma et al. (2016) tests normal, Weibull, Logistic, Gamma, Lognormal, Loglogistic, Burr, and Gaussian mixture models (GMM) for fit with passenger travel times for two bus lines in Brisbane. El Faouzi and Maurin (2023) expands on lognormal distributions in transport models. Pu (2011) uses the assumptions of a log-normal distribution to derive the standard forms of common reliability measurements, such as Buffer Index and Planning Time.

Zheng and van Zuylen (2015) argue that arterial times could be accurately modelled, but only if the models analyse the probabilistically spillback from downstream intersections (Zheng et al., 2017). Moghaddam et al. (2011) provides a model to estimate the average and standard deviation of origin-destination travel times between two segments of a route. Al-Deek and Emam (2006) propose a reliability model based on volume over capacity (V/C) analysis for network links. Some papers compute the roadways Level of Service (LOS) using the travel times distribution (Chen & Fan, 2020; Guessous et al., 2014; Shao et al., 2006). Hall (1983) is an interesting attempt to consider reliability, accessibility, and transfers in a spatiotemporal analysis of travel time, including possible sanctions for passenger and their resultant margins.

Mahmassani et al. (2012) suggest that for choice modelling at multiple scales, the SD is the representative statistic for travel time variability. They suggest due to high correlation, the actual SD[iii] can be best estimated from a linear model using the travel time per unit; however, heteroscedasticity requires the use of a weighted least squares method. Guo et al. (2010) suggests that there are multiple applicable distributions (not just different parameters) depending on the state of the network; and each of those states have a different likelihood of occurring. Their paper also uses a multi-component model to match a bimodal distribution. That approach was expanded in Park et al. (2010). Then in Park et al. (2011), they test the number of component Gaussians models to mix to achieve the best fit, from two to seven.

Recent papers have challenged the earlier assumption of statistical independence across different links and times of day (Bates et al., 2001; Mahmassani et al., 2012). Kim and Mahmassani (2015) propose a combined Gamma-Gamma distribution to account for vehicle-to-vehicle day-to-day variations in travel delay (relative to motorway free-flow) under congestion. They suggest a double stochastic model is necessary to model the random congestion levels in the temporal (day) and spatial (vehicles) dimensions. They chose Gamma distributions as those are mathematically tractable and can model travel delays.

Harsha et al. (2020) notes that most travel time studies were in the homogenous controlled orderly USA, not in the heterogeneous lane-less disorderly traffic of India; see also Harsha and Mulangi (2022). They determined that Generalised Extreme Value (GEV) distributions had the highest accuracy and robustness for corridors; however, normal distributions were a reasonable fit, especially for individual segments. Chepuri et al. (2018) determined that normal distributions applied to peak hours and GEV distributions applied to off-peak hours. They calculated other statistics including the average time, BTI, & PTI; then using buses as probe vehicles they calculated the average speed, V/C ratio, and LOS. Chepuri et al. (2018) and Chepuri et al. (2020) then developed a Reliable Buffer Index (RBI) measure.

Rahman et al. (2018) created a predictive model for Calgary bus arrival times using horizons; they suggest that lognormal (or gamma or log-logistics) distributions are better up to 7km, and beyond 7km then normal distributions are better, probably due to the Central Limit Theorem. Strathman et al. (2002) notes '*The [driver] effects on running time appear to be normally distributed, which has important implications.*' Ramakrishnan et al. (2018) states that '*The studies showed that historical trajectory-based travel time prediction is highly accurate when a large amount of historical data is available. The historical data inherits the stochastic features of traffic for each hour of the day, thus enabling the model to accommodate for most possible changes in the traffic conditions.*'

Many papers aggregated travel times in 5, 15, 30, 60-minute Departure Time Windows (DTW). Mazloumi et al. (2008) examined travel time variability as a metric in Melbourne (percentile range over median for a 15-minute DTW). Their analysis showed land use, link length, time of day, number of signals, number of stops, and timetables impacted reliability (in that order). Mazloumi et al. (2009) expanded with an analysis of an orbital route, which determined that running times appeared normal and travel times appeared lognormal. However, the shape of the distribution was affected by the DTW, with shorter and longer windows both appearing normal — <u>an effect that was observed in many of these papers</u>. Noland and Polak (2002) notes '*Herman and Lam (1974)*[iv] *... found that shorter trips tended to be normally distributed while longer trips followed a log-normal distribution.*'

1.1.1 Theoretical context

Jacoby (1976) critiqued Consumer Research and lamented that most papers proposed new measures and models, leading to limited theoretical development, few reproduction studies, and limited analysis of real-world applications (Churchill, 1979; Hansson, 2017; Jacoby, 1978; Martin, 2019; Parasuraman et al., 1988; Peter, 1979; Thornton, 2014). Unfortunately, that pattern continues in multiple fields, including transportation. TfNSW uses several LOS to measure their road quality including average travel speed, V/C ratio, density, degree of saturation, average delay, queue lengths, and cycle length (Hounsell, 2021; RTA, 2002, §4.2) — not to mention the plethora of transit LOS in (TCRP, 2003). Despite that, many of the papers reviewed explicitly aimed to find a single probability model and a single measure to illuminate all aspects of, and situations for, travel time reliability. However, it could be concluded from the literature review, that a single model and measure would not be sufficient. Ramakrishnan et al. (2018) states '*the studies are concentrated locally on the application of each method on a particular system ... results observed are constrained to a particular set of data. Thus, the comparison of each method would not reflect its true potential.*'

The relationship between the quantitative measurement 'score', instruments, & errors is often expressed as $observed\_score = true\_score + systemic\_variation + random\_error$ (Bateson et al., 1985; Churchill, 1979). That model is inadequate because it doesn't capture the variations in the subject, nor in the instrument, nor in the environment. The Generalizability Analysis model below is better but still limited (Alkharusi, 2012, p. 189).

Equation 1: The linear model of Generalizability Analysis

$$X_{pi} = \mu + \mu_p + \mu_i + \mu_{pi} + e$$

X_{pi} is the observed score obtained by measuring an entity with a given instrument,
μ is the overall mean score of the population of entities and universe of instruments,
μ_p is the score effect attributable to that entity at the time of measurement,
μ_i is the score effect attributable to measurement instrument at that time,
μ_{pi} is the score effect attributable to the interaction of that entity and that instrument, and
e is the effect of random sources of error

Churchill (1979) identifies some systematic sources of error in measurements of a trait:
- situational/environmental factors during measurement (e.g., traffic),
- individual factors for the subject being measured (e.g., bus / driver),
- the experimenter's use or misuse of the instruments,
- ambiguity intrinsic to the instruments,
- experimental impacts from sampling (e.g., DTW),
- instrument (mechanical) misfunction, *and finally*
- true differences in other stable traits which affect the measurements of the examined trait.

Many of the papers lacked consideration of measurement theory (Alkharusi, 2012; Bryan et al., 2000; Campbell & Fiske, 1959; Cho, 2016; Cronbach, 1951, 1960; Cronbach & Shavelson, 2004; Goforth, 2015; Huebner & Lucht, 2019). Strathman et al. (2002) was one of the few papers that built variables explicitly representing drivers in their running time model.

Taylor (1982) observe that link travel times for buses appear normal while metro appear lognormal, '… *the mean bus travel times were far enough removed from the free travel times (i.e. the physical limits on travel times) so that variations both above and below the mean could be reasonably be expected, whereas the metro travel times ... were not sufficiently removed from their lower limits to permit the observations of times well below the means.*' On an empty motorway, the posted speed limit will set the free flow travel time, creating a lower bound and right skewing the travel times. However, on congested arterials, vehicles are unable to reach, let alone sustain, the posted speed limits, preventing skew. Returning to the theory of Statistical Process Control, a process is not under control if the process/service is aborted, deviating wildly, or skewed — the root cause of those situations have to be identified and any faults eliminated (Ang & Tang, 2007; Montgomery, 2013). Many of the reviewed papers treated all bus system delays and failures as innate; however, as Amsterdam, Madrid, Oslo, Paris, and Strasbourg have demonstrated prioritising private cars in cities is a choice, and as such, a long tail of slow bus travel times is a political effect on travel times not a natural one.

The papers reviewed noted that experiments showed that travellers can place a greater emphasis on reducing travel time variability than on the mean travel time; and some papers suggested that unreliable travel time might be innately displeasing. The papers note that the transit would fail to attract potential passengers when the service provided was unreliable. As such, in order to deliver a customer-centric transit system the operators must measure, analyse, and eliminate the causes of travel time variability, then work to reduce travel times. Levinson (1991) suggests that bus timetable '*running time is set slightly shorter than the observed average, ignoring extreme observations*' to prevent drivers from running slowly.

The advantage of the normal distribution is that it does not require fitting, is widely understood, and is suitable for SPC. The review indicates that the normal distribution could be applicable to frequent routes on congested roads with sufficient data. This case study seeks to determine, if the normal distribution is reasonable for planning models (Hounsell, 2022b).

2. Method

Strathman et al. (2002) states '*Designing and delivering high quality transit is an information-intensive undertaking. ... Detailed operations data are needed to develop schedules that respond to the demand for service, while at the same time conserve scarce agency resources.*' and '*In a data-poor environment, planners and schedulers must depend on ad hoc feedback from [drivers] and passengers to adjust routings, running times, and service frequencies.*'

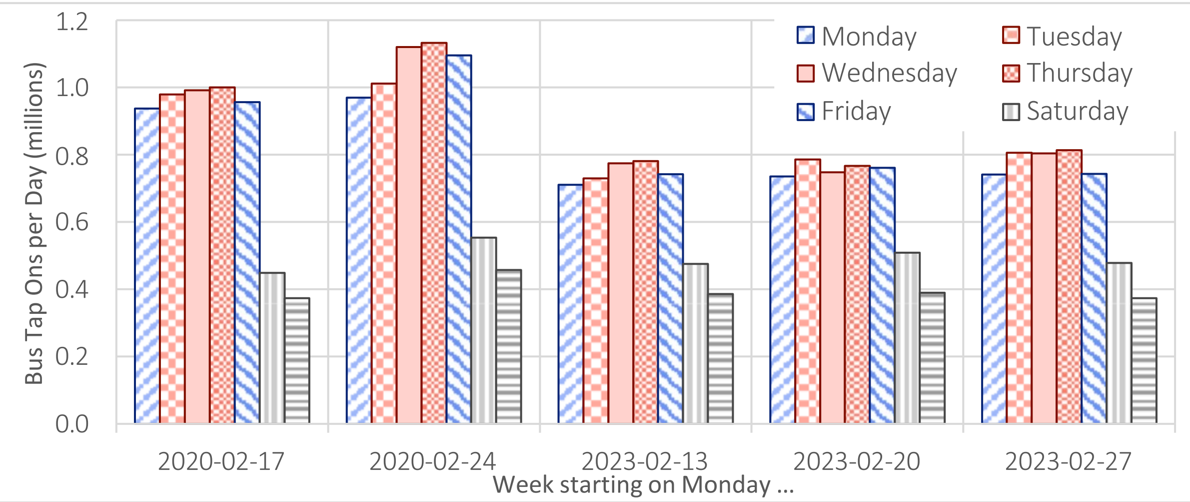
The GPS receivers in the NSW Opal ETS are used by TfNSW to provide real-time location information for each bus in metropolitan NSW utilising 3G mobile internet for ship-to-shore communication. The Bus Opal Assignment Model (BOAM) provides historical datasets of the bus location and estimated loads for metropolitan NSW. This study makes use of the open-data versions released as TfNSW (2017) and TfNSW (2020a, Version 2020-03-19). These datasets consist of very large, delimited text files. For example, the file for Thursday 2023-11-17 alone was over 457 MB[v]; and the files for all the dates studied took over 41 GB. Practical analysis of Big Data requires multi-stage processing to limit the time and memory required for each calculation. The first stage used Microsoft Excel's Power Query to read every BOAM file to extract, transform, and filter every row where the ROUTE column contained "333". Over one million records were loaded during this first stage.

Service optimisations and changes in the route over time resulted in some stops being unsuitable for use in this study. For example, for a few weeks the 333 used a temporary stop on Oxford St after York Rd (#2021106) while the stop before York Rd (#202260) near the depot was out of commission during installation of a bike lane. As such, the second stage counted the number of records per stop and plotted those on a map of Sydney (colour coded by direction) to allow manual identification of stops suitable for use as timing points. This study was able to continue examining the inbound end-to-end runtimes using Campbell Parade, Bondi (#200634) to Martin Place, CBD (#2000421) as in Hounsell (2018b). Those stops were selected because their position stayed the same from 2016 to 2023, and they were served by the majority of inbound runs (unlike the stops at Circular Quay).

Each BOAM row provides the scheduled and actual arrival times for one bus run at one stop, as well as the start date of that bus run and a bus run identifier. The third stage used Excel to extract and match each row referring to Campbell Parade to the row for Martin Place (using the bus run date and identifier) and then the calculated the end-to-end runtime duration. This study examines these actual runtimes and not the estimates in the GTFS timetable.[vi]

In NSW the transit usage on Monday and Friday for a "typical" working week is less than the usage on Tuesday, Wednesday, and Thursday (TWT) with the weekends having the lowest transport demand (Hounsell, 2023). Those assumptions appear to be correct for NSW buses managed by the Opal ETS, as shown in Figure 3 below. Figure 3 shows the number of Tap Ons (embarkations) in millions on the y-axis and the day of the week on the x-axis. Figure 3 covers two "typical" school weeks in 2020 before the first COVID-19 pandemic lockdown (Botha et al., 2023, p. 3), and three typical weeks in 2023 as identified by (Hounsell, 2023).

Figure 3: Patronage (millions) for all NSW buses with Opal for the day in the week starting on Monday…



Data Source: (TfNSW, 2020b) where [mode_name] = "Bus" and [ti_region] = "All - NSW" [E4]

Figure 3 shows that patronage on the weekend (Sat-Sun) of each week was substantially lower than the patronage of that week's workdays (Mon-Fri). The plot also shows that bus patronage on Monday and Friday is often lower than the bus patronage on TWT.

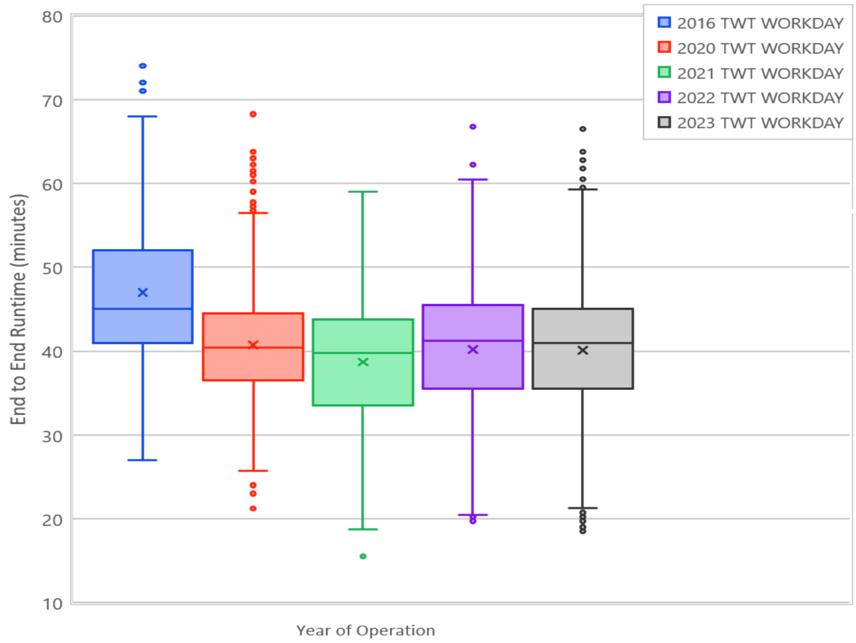
The level of private traffic in the eastern suburbs[E1] where the 333 operates is known to vary across the days of the week impacting bus runtimes. In addition, TfNSW contracts[vii] specify the service targets, and TfNSW only pays bus operators for a significantly reduced level of service on weekends and public holidays, with modest reductions during school holidays — this reduced supply will impact demand. Therefore, this paper analysed TWT school days to limit environmental differences, and because more services provided more samples; however, the greater patronage will have increased variability in dwell times and stopping patterns.

3. Results

Filtering was applied to eliminate deviant records, such as those without actual arrival times or a negative duration. Next, statistical summaries including the Quartiles, the Inter Quartile Range (IQR), and the Quantile/Galton/Bowley skewness were calculated. Then, the probable number of outliers were estimated using the standard 1.5 IQR rule (NIST, 2018).

The box plot in Figure 4 shows the end-to-end runtime (minutes, y-axis) of 333 buses running inbound from Campbell Pde to Martin Pl on TWT broken down by the year of the samples (x-axis). The plot shows that half of the runtimes in March 2023 were between 35.5 and 45.0 minutes, with an IQR of 9.5 minutes indicating substantial Running Time Variability (RTV). Statistically expected range of running times was between 21.3 and 59.3 minutes, with several runtimes exceeded even that broad range, indicating that the runtimes are unstable. Figure 4 raises several questions for the operators. Firstly, why are the recent runtimes so unstable, and secondly, why are the runtimes more unstable than 2020?

Figure 4: Box plot of the end-to-end runtime (minutes) for TWT days per year



Data Source: TfNSW (2017) and TfNSW (2020a, Version 2020-03-19) where [route] = "333" [E3]

There was only one week of data publicly available for 2016, while none was available for 2017 to 2019. The global COVID 19[viii] pandemic began in December 2019 in the city of Wuhan in Hubei Province in the People's Republic of China before spreading to Italy. For April to May 2020, NSW was in lockdown to prevent infections increasing. NSW was in lockdown again in December 2020 and January 2021, as well as from June to October 2021 (Botha et al., 2023; Stobart & Duckett, 2022). The weeks in 2020 and 2021 are atypical, however, they provide a useful period with lower vehicle traffic and patronage.

Table 1: First day of the week (Monday) starting the weeks used for this analysis.

2016	2020	2021	2022	2023
2016-11-21	2020-11-02	2021-11-01	2022-10-31	2023-03-06
	2020-11-09	2021-11-08	2022-11-07	2023-03-13
	2020-11-16	2021-11-15	2022-11-14	
	2020-11-23	2021-11-22	2022-11-21	

Traffic and patronage vary across the day in Sydney, being more intense during the day than in the evening and night; as in Table 2. Therefore, **in this study the hours of the day were classified into two groups: *night-time*** between 20:00 and 05:59 **and *worktime*** between 07:00 to 18:59. These periods were based on the number of services (as in Table 2); as well as results from Hounsell (2018b) and other research. The transitional hours of 06:00 and 19:00 were excluded to minimise environmental disruption. For your reference, the number of samples for each period of the TWT days is listed in Table 3 below on page 16.

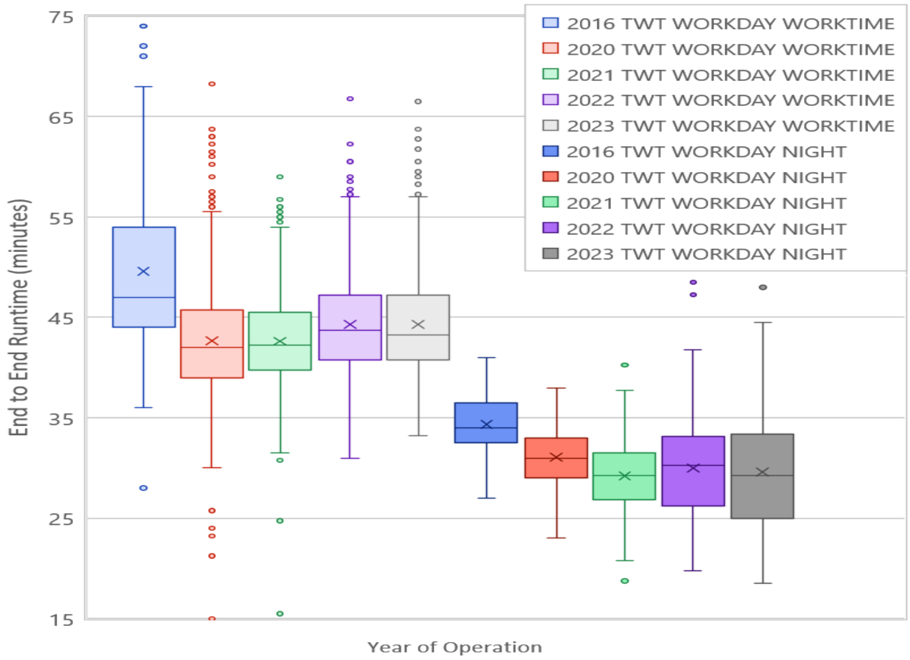
Table 2: Bus services per hour on an average TWT for the three weeks beginning 2023-02-13

0	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	1	11	32	80	86	42	33	33
12	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**
34	37	50	76	64	71	48	24	17	14	9	4

Data Source: (TfNSW, 2020b) where [mode_name] = "Bus" and [ti_region] = "All - NSW"

Figure 5 show the runtimes (y-axis) on the 333 route not only vary across the years (x-axis) but also across the time of the day (x-axis) with the median runtimes during the night in 2023 being 29.5 minutes and during the day being 43.5 minutes.

Figure 5: Box plot of the end-to-end runtime (minutes) for TWT days per year per time-of-day

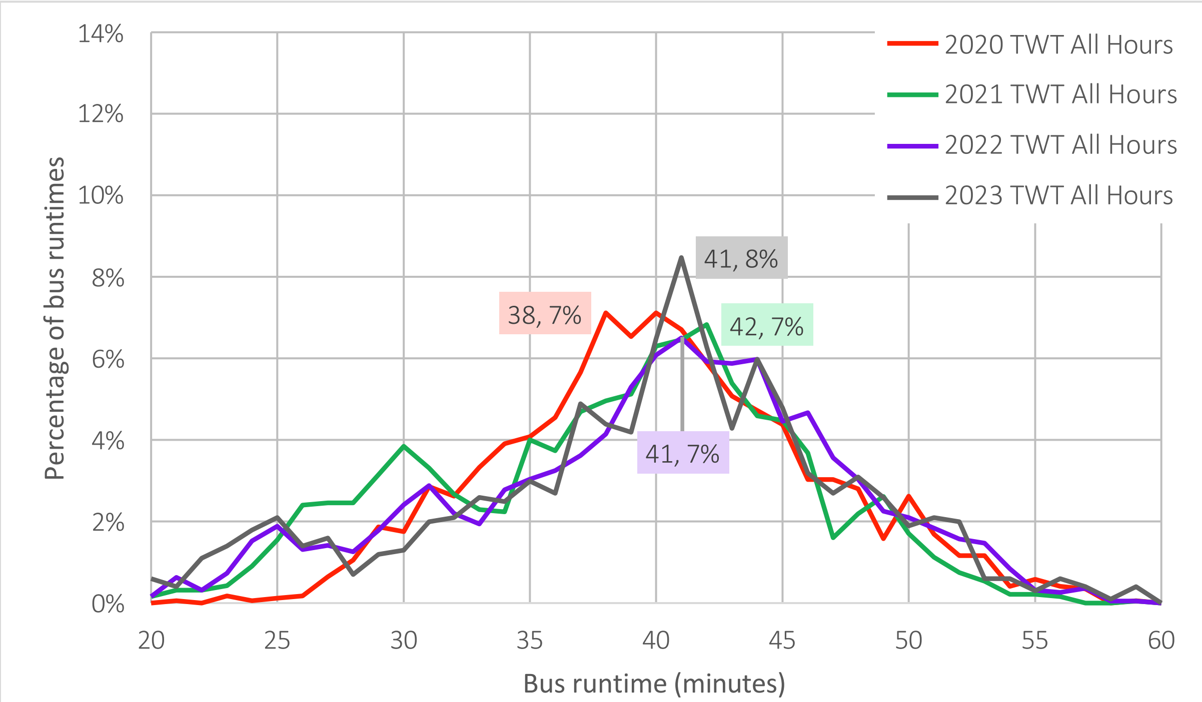


Data Source: TfNSW (2017) and TfNSW (2020a, Version 2020-03-19) where [route] = “333” [E3]

Figure 5 shows that runtimes are lower on the route during the night during than the day, however, the runtimes have become less stable since November 2020. Half of the runtimes during the 2023 night were between 25 and 33.5, while during the day half of the runtimes were between 41 and 47.5 minutes. Thus, the IQR for the night in 2023 was 8.3 minutes (over 241 samples) versus 6.5 minutes (over 667 samples). Note that in the latest contract period the 333 is operating to a headway-adherence Key Performance Indicators (Excess Wait Time (EWT)) instead of an On-Time Running (OTR) KPI, like TfNSW (2018)

Box plots are useful because they plot the key information — the quartiles, median, means, and outlier limits (1.5 IQR); however, they don't communicate the shape of the distribution. Figure 6 shows the data in Figure 4 but as a frequency distribution with the bus runtimes on the x-axis (minutes) and the percentage of each runtime on the y-axis. For clarity, the plot does not show 2016 because its offset quartiles create an interfering and confusing pattern. The plot shows that the 2023 runtime regularly ranged over 30 minutes, from 22 to 52 minutes, with the mode of 8.5% at 41 minutes.

Figure 6: Frequency of runtimes (percentage) by runtime (minute) per year

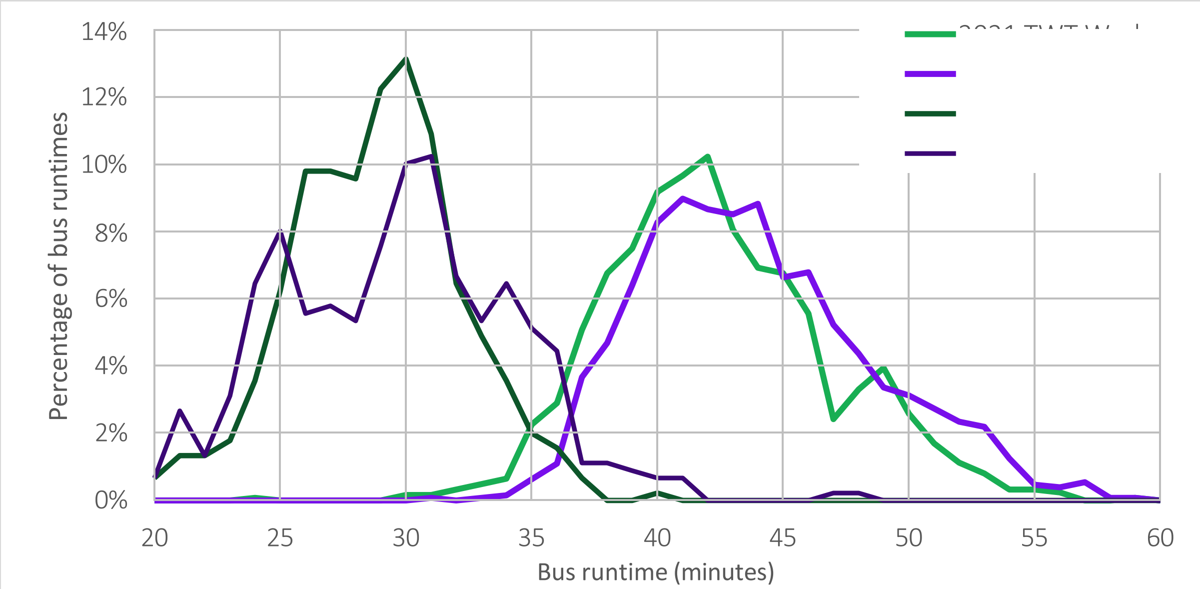


Data Source: TfNSW (2020a, Version 2020-03-19) where [route] = "333" [E3]

Figure 7 below shows that the distribution of runtimes (percentage, y-axis) compared to the runtimes (minutes, x-axis) broken down by year (2021 vs 2022) and time of the day (night vs worktime). This plot shows that most of the daytime inbound runs are more than 35 minutes while most night-time inbound runs are less than 35 minutes. This plot indicates a left skew to the daytime runtimes suggesting drivers are waiting to keep to a "timetable", whereas the night runtimes are less controlled (Hounsell, 2022a). These distributions indicate that the runtimes are being slowed by the traffic in Oxford St and Bondi Road corridors, as well as the traffic lights for traffic crossing that corridor (TfNSW, 2012; Unsworth, 2004). Further, research is suggested to determine if this applies for other routes on other corridors. The plot only shows 2021 and 2022 as adding the other six distributions created confusing clutter.

Figure 6 above shows a range of runtimes of over 40 minutes. In contrast, Hounsell (2022a) showed that City to South-East Light Rail (CSELR) runtimes had a range of 13 minutes. Figure 7 shows that the range of 333 runtimes during the daytime and at night-time is "only" 20 minutes. Together, these plots clearly demonstrate that careful analysis of Sydney buses requires identification of the different operating environments and possible methodological errors[ix][E5] before application. The runtime distribution in Figure 6 is clearly distorted by the aggregation of different hourly distributions within a single dataset, and therefore, the full day estimates for runtime and RTV will be suboptimal for buses running in mixed traffic.

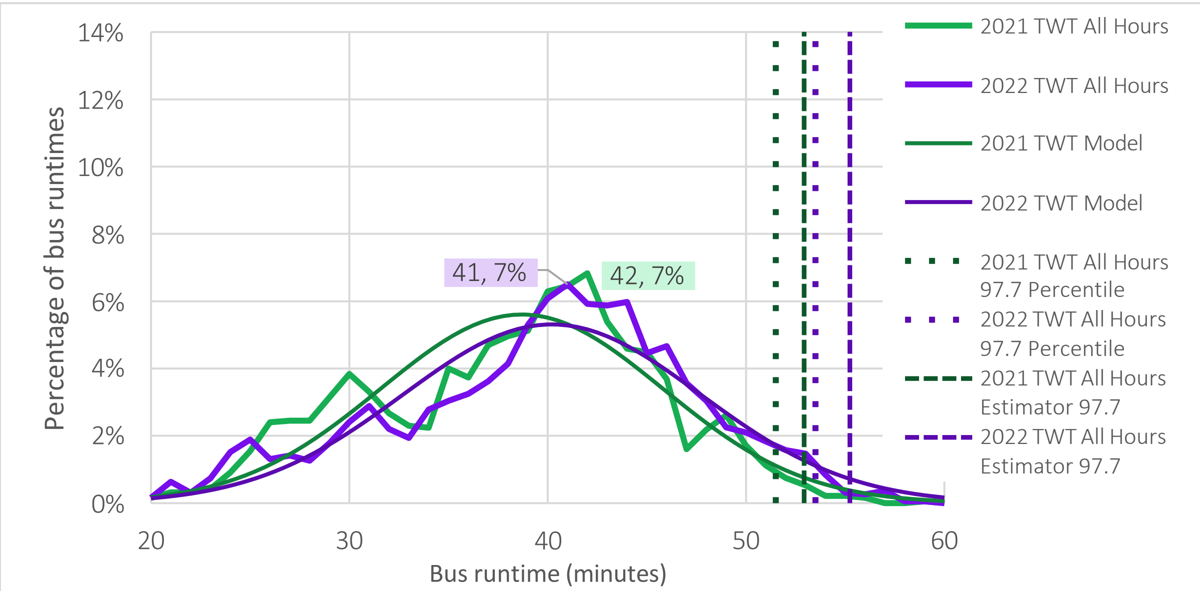
Figure 7: Frequency of runtimes (percentage) by runtime (minute) per year per time-of-day



Data Source: TfNSW (2020a, Version 2020-03-19) where [route] = "333" [E3]

This paper aims to test if the hypothesis that the mean and the standard deviation can be used to construct running time estimators for high-frequency bus routes in mixed-traffic such as the 333. Figure 8 shows the distribution of the observed 333 runtimes for all hours with the percentage of each runtime on the (y-axis) and the runtimes on the x-axis (minutes) for 2021 and 2022. There is a second thin line for each year that shows the normal distribution estimated using the trimmed mean and standard deviations. The plot also has for each year a line for the observed 97th percentile and the estimated 97th percentile ($\bar{e} + 2\sigma_e$).

Figure 8: Frequency of runtimes (percentage) by runtime (minute) all day compared to model



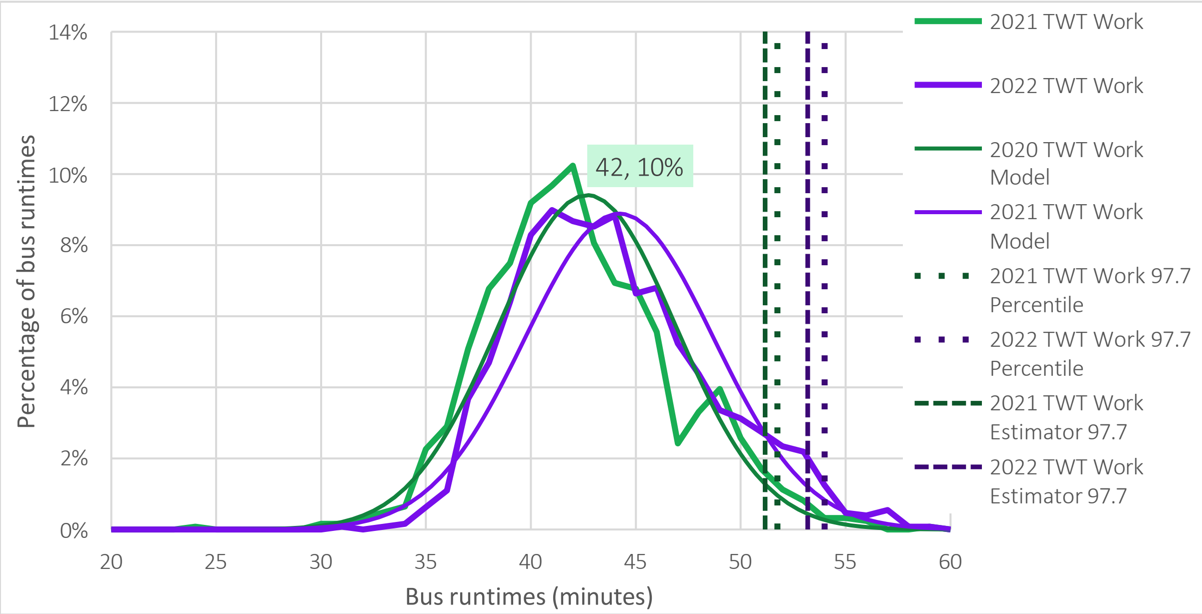
Data Source: TfNSW (2020a, Version 2020-03-19) where [route] = "333" [E3]

Figure 8 shows that the estimators using all hours are within the right region for the 97th percentile, but the model is not as accurate as it could be. The plot also shows that 97th

percentile for 2022 was 53.5 minutes which is substantially higher than the mode of 41 minutes; demonstrating that the 333 operations are not as efficient as possible.

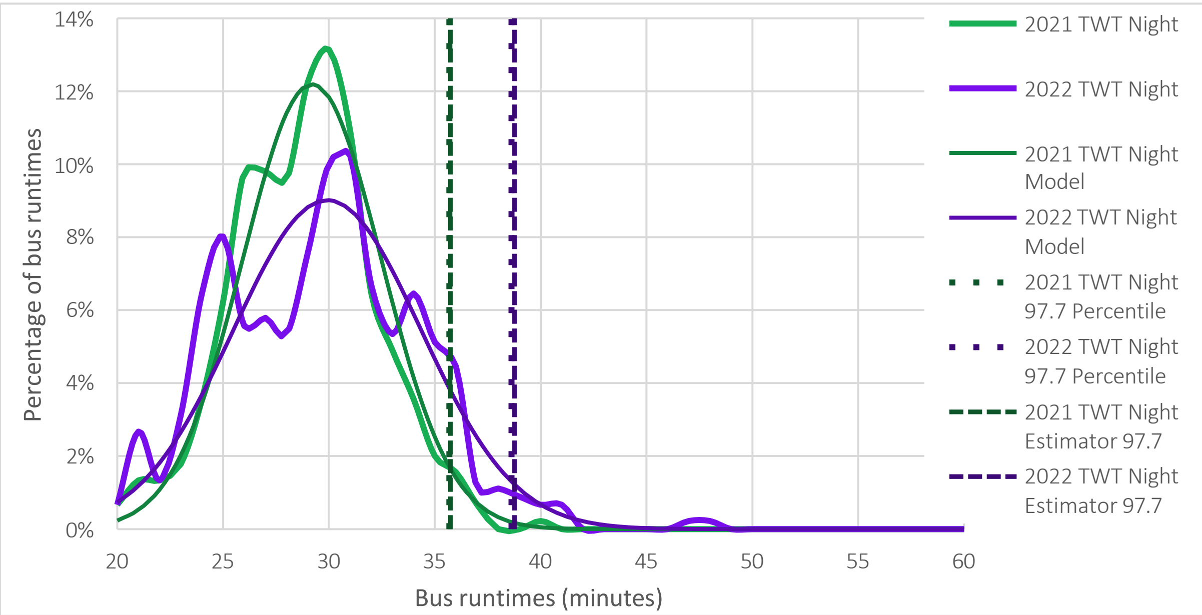
In response to the significant difference in the distribution of runtimes between the day and the night time shown in Figure 7, a separate estimator model was developed for the daytime and night time operations as shown in Figure 9. Figure 9 again shows the bus runtimes on the x-axis (minutes) and the percentage for each runtime on the (y-axis) for 2022 and 2023, *but this time only for the working hours*. The plot also shows the model as normal distribution, the observed 97th percentile, and the estimated 97th percentile. Compared to Figure 8, Figure 9 shows that a model of the daytime runtimes is similar to a normal distribution and that the estimator for the 97th percentile ($\bar{e} + 2\sigma_e$) was close to the observed percentile. Whereas, the model for night time as shown in Figure 10 has more accurate 97th percentile estimates.

Figure 9: Frequency of runtimes (percentage) during worktime compared to model



Data Source: TfNSW (2020a, Version 2020-03-19) where [route] = "333" [E3]

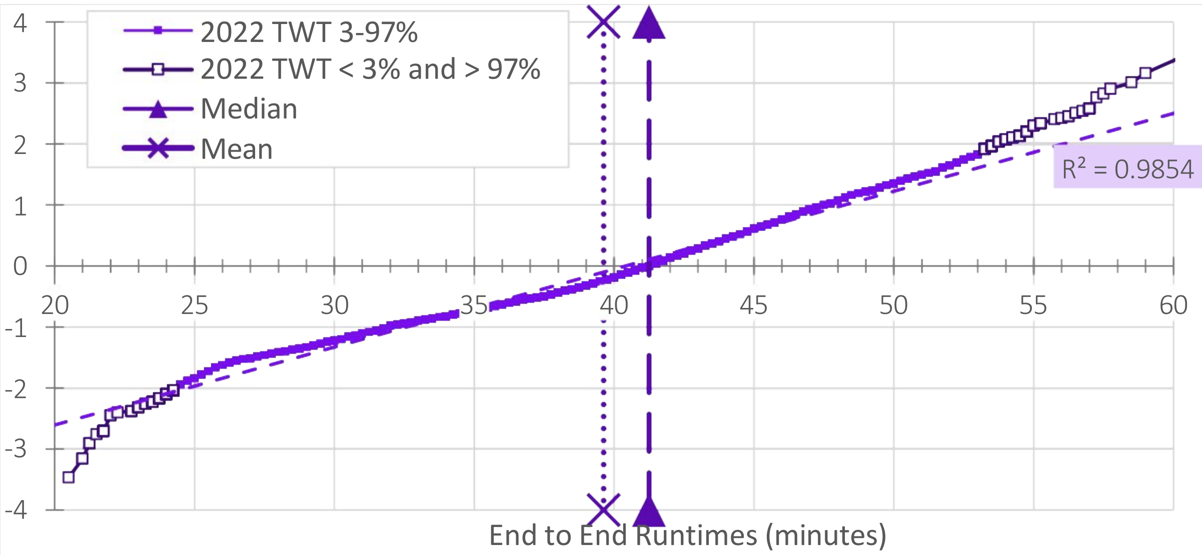
Figure 10: Frequency of runtimes (percentage) during night time compared to model



Data Source: TfNSW (2020a, Version 2020-03-19) where [route] = "333" [E3]

A normal probability plot is another approach to assessing whether the mean and the standard deviation are useful in constructing models and estimators for a given dataset. A normal probability plot uses cumulative probability, z-scores, and a linear regression to assess whether a dataset has a nearly normal distribution (NIST, 2018). Figure 11 show the 333 TWT runtimes (all hours) are nearly normally distributed with tails (R^2 is 0.9854).

Figure 11: Normal probability plot of end-to-end 2022 TWT full day runtimes



Data Source: TfNSW (2020a, Version 2020-03-19) where [route] = "333" [E3]

4. Discussion and Conclusion

Hounsell (2022b) postulated that the efficiency of transit operations might be improved through the application of common Statistical Process Control (SPC) techniques if the service runtimes were normally distributed, or the normal estimators could be used reliably. These papers examine end-to-end runtimes because those determine the deployed fleet size which determines how efficiently the service delivers the targeted headway. This paper examined the hypothesis that for high-frequency bus operations in mixed-traffic, the normal-distribution model, the mean end-to-end runtime ($\bar{e}$), and the mean plus two SD ($\bar{e} + 2\sigma_e$) **provide reasonable estimators** for the distribution, middle, and 97th percentile of the observed end-to-end runtimes respectively.

Continuing the research programme, this paper demonstrates that bus running times naturally have a normal distribution, like trams and metros in Hounsell (2022a). The charts and tables show that when buses operate with natural runtimes (especially when focused on delivering a vehicle-headway) then the normal model, mean, and standard deviation are useful runtime estimators. Strathman et al. (2002) states '*Schedule design is based on running times that are sufficient for an average [driver] to complete a trip under normal conditions.*' They suggest schedulers set '*running times low enough to avoid having [drivers] kill time, while setting generous recovery times to avoid late departures on subsequent trips.*' Also, that '*recovery time requirements tend to be larger when clock face schedules are employed. While [clock-face] schedules are convenient for passengers, their usefulness is diminished when service is frequent [as] passengers do not consult schedules.*' They note that running time reductions ensure '*adjustments could be made … with sufficient running and recovery time to maintain reliable service … that would effectively "add a [service]" … at no additional cost.*' This paper suggests that normal based estimators are adequate to assist planners and schedulers analysing the usual operation of buses on frequent routes on busy roads. This paper suggests common process management techniques could be used to monitor and improve the regular operations of high-frequency bus routes (Barr, 2017, 2019). However, care needs to be taken to remove outliers caused by uncontrolled process deviations within the transport system.

Note that the literature indicates, that if a route is sustaining long periods at the speed limit, then another distribution (e.g., lognormal) or a data transformation may be required.

Although not shown here, analysis was undertaken on running times on the subsections before, and after Bondi Junction, and that those were reasonably estimated by a normal distribution. The timing point at Bondi Junction Interchange, and a driver transfer point opposite the Bondi Junction deport, complicates analysis of times through Bondi Junction, as those represented uncontrolled process deviations. Further analysis supported the literatures' conclusions that when cars are prioritised through a lack of public transport signal priority, then stop-to-stop runtimes are abnormal, and can be multi-modal (Susilawati et al., 2013)

Further research could evaluate the distributions of observed runtimes in peak and inter-peak periods; however, specific care will need to be taken due to the smaller sample sizes. Further research investigating the entire year is possible, however specific care must be taken to account for the changed state caused by holidays, different patronage, and different weather.

5. Acknowledgement

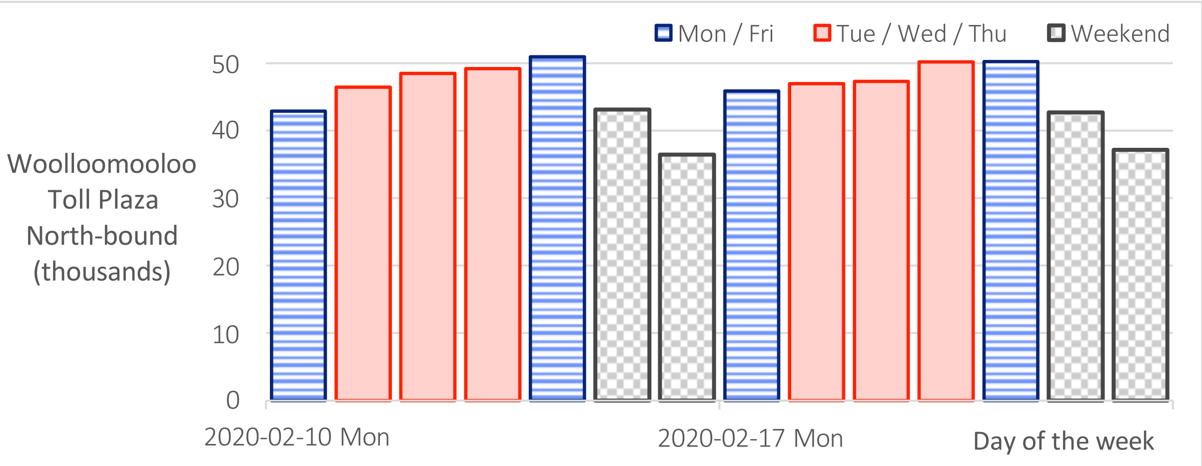
This research is funded by Transdev Australasia Pty Ltd and iMOVE Australia Ltd and supported by the Cooperative Research Centres program, an Australian Government initiative.

6. Endnotes

E0 – **Supplementary material are available**. The slides, public data, and refences (with links) are available on ResearchGate. Watch the presentation on YouTube. Further references and citations can be found via Scopus, Google Scholar, and orcid.org/0000-0003-0448-0374.

E1 – Figure 12 plots the number of vehicles (y-axis) that paid tolls for the Eastern Distributor at Woolloomooloo heading north-bound on each of the days of the week between Monday 10th February 2020 and Sunday 23rd February 2020. This graph shows that the traffic in the eastern suburb varies between the workdays and the weekends; and that the traffic varies between the days of the week. Unfortunately, the TfNSW Traffic Volume Viewer indicates the state does not maintain a fixed traffic counter within the Oxford St corridor, although it does maintain cameras to monitor the traffic flow.

Figure 12: Number of vehicles (thousands) passing the toll plaza per day



Source: RSM (2023). *NSW Toll Road Data*. Transurban Ltd (ACN 098 143 410), and RSM Australia Pty Ltd (ACCC approved Independent Auditor). https://nswtollroaddata.com/data-download/

E3 – The TfNSW (2020a) dataset has multiple columns including [route]; such as, "333".

E4 – TfNSW (2020b) dataset has a [ti_region] column, such as, Chatswood & Parramatta.

E5 – Table 3 below contains the count of samples for the periods of the days studied.

Table 3: Number of samples for each period of the TWT day by the start of the sample period

Year	Start of Week	Work Time	Night-Time	Skipped	Full Day
2016	Tue 2016-11-22	191	29	19	239
2020	Tue 2020-11-03	328	55	38	421
	Tue 2020-11-10	334	57	41	432
	Tue 2020-11-17	337	53	41	431
	Tue 2020-11-24	341	59	40	440
	Total	1,340	224	160	1,724
2021	Tue 2021-11-02	318	112	47	477
	Tue 2021-11-09	305	113	48	466
	Tue 2021-11-16	301	109	45	455
	Tue 2021-11-23	317	115	48	480
	Total	1,241	449	188	1,878
2022	Tue 2022-11-01	333	121	48	502
	Tue 2022-11-08	326	107	42	475
	Tue 2022-11-15	330	119	46	495
	Tue 2022-11-22	291	102	43	436
	Total	1,280	449	179	1,908
2023	Tue 2023-03-07	328	119	46	493
	Tue 2023-03-14	339	122	50	511
	Total	667	241	96	1,004

---

[i] A request was lodged with TfNSW requesting the number of embarkations on the 333 on a typical week, such as the week of Monday the 6th to Sunday the 12th of March 2023.

[ii] Measure Stabilise Reduce was coined by Dr Michelle Zeibots. After considerable usage, the author has found "Measure, Analyse, Stabilise, and Reduce" (MASR) to be a more useful and andragogic description of the process; as shown in the adapted mnemonic image in Figure 2. In addition, when conceptualised as MASR, the framework can more easily be related to the Six Sigma DMAIC cycle — Brussee, W. (2006). *All about six sigma : the easy way to get started*. McGraw-Hill.

[iii] Remember the SD estimated from the samples will differ from the actual population SD

[iv] This paper was not available: Herman, R., & Lam, T. (1974). Trip time characteristics of journeys to and from work. *Transportation and traffic theory*, 6, 57-86.

[v] This paper uses the modern International System of Units (SI) convention of the megabyte (MB) being 10^6; whereas the old mebibyte (MiB) is 2^{20} and 1024^2. https://en.wikipedia.org/wiki/Megabyte

[vi] GTFS stands for the General Transit Feed Specification (Hounsell 2020, §4.3, §4.3.1)

[vii] In response to a query TfNSW stated '*Specific contracts are commercial in confidence and as such are not made publicly available. As of 3 April 2022, Region 9 has been operating as per Greater Sydney Bus Contract 9 operated by Transdev John Holland Buses. As of 17 March 2023, a model version of all Greater Sydney Bus Contracts is publicly available at* https://www.transport.nsw.gov.au/operations/buses-and-coaches/bus-contracts'

[viii] '*COVID-19 is the disease caused by the coronavirus (CoV), [Severe Acute Respiratory Syndrome] SARS-CoV-2. Coronaviruses are a large family of viruses that cause respiratory infections. These can range from the common cold to more serious diseases.*' DHAC (2022, 2022-12-09). *COVID-19 disease and symptoms*. Department of Health and Aged Care, Commonwealth of Australia (CoA). Accessed 2023-02-10. https://www.health.gov.au/health-alerts/covid-19/symptoms

[ix] Churchill identifies some systematic sources of error in measurements of traits.

7. References

Abkowitz, M. & Engelstein, I. 1984, 'Methods for maintaining transit service regularity', Transportation Research Record, vol. 961, pp. 1-8.

Al-Deek, H., & Emam, E. B. (2006). New Methodology for Estimating Reliability in Transportation Networks with Degraded Link Capacities. *Journal of Intelligent Transportation Systems*, *10*(3), 117-129.

Alkharusi, H. (2012). Generalizability Theory: An Analysis of Variance Approach to Measurement Problems in Educational Assessment. *Journal of Studies in Education*, *2*, 184-196.

Ang, A. H.-S., & Tang, W. H. (2007). *Probability concepts in engineering : emphasis on applications in civil & environmental engineering* (2nd ed. ed.). Wiley.

Aron, M., Bhouri, N., & Guessous, Y. (2014). *Estimating travel time distribution for reliability analysis.* Transportation Research Arena 2014, Paris.

Barr, S. (2017). *Prove it! : how to create a high-performance culture and measurable success*. Wiley.

Barr, S. (2019). How to Measure What Matters: How to end your KPI and performance measurement struggles and reach your goals sooner and with less effort. Stacey Barr.

Basu, R. (2009). Implementing Six Sigma and Lean : a practical guide to tools and techniques. Elsevier Butterworth-Heinemann.

Bates, J., Polak, J., Jones, P., & Cook, A. (2001). The valuation of reliability for personal travel. *Transportation Research Part E: Logistics and Transportation Review*, *37*(2-3), 191-229.

Bateson, J. E., Reibstein, D. J., & Boulding, W. (1985). *Conjoint analysis reliability and validity: A framework for future research*. Graduate School of Business, Stanford University.

Batley, R., Bates, J., Bliemer, M., Börjesson, M., Bourdon, J., Cabral, M. O., Chintakayala, P. K., Choudhury, C., Daly, A., Dekker, T., Drivyla, E., Fowkes, T., Hess, S., Heywood, C., Johnson, D., Laird, J., Mackie, P., Parkin, J., Sanders, S., Sheldon, R., Wardman, M., & Worsley, T. (2019). New appraisal values of travel time saving and reliability in Great Britain. *Transportation*, *46*(3), 583-621.

Botha, F., Morris, R. W., Butterworth, P., & Glozier, N. (2023). Trajectories of psychological distress over multiple COVID-19 lockdowns in Australia. *SSM - Population Health*, *21*, 101315.

Brussee, W. (2006). All about six sigma : the easy way to get started. McGraw-Hill.

Bryan, S., Gold, L., Sheldon, R., & Buxton, M. (2000). Preference measurement using conjoint methods: an empirical investigation of reliability. *Health Economics*, *9*(5), 385-395.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81-105.

Carrion-Madera, C., Levinson, D., & Harder, K. (2012). *Value of travel-time reliability : commuters' route-choice behavior in the Twin Cities, Phase 2*. Transportation Research and Education Center (TREC).

Carrion-Madera, C., Levinson, D., Harder, K., & Figliozzi, M. A. (2011). *Value of travel-time reliability: Commuters' route-choice behavior in the twin cities*. Transportation Research and Education Center (TREC).

Carrion, C., & Levinson, D. (2012). Value of travel time reliability: A review of current evidence. *Transportation Research Part A: Policy and Practice*, *46*(4), 720-741.

Cartwright, T. J. (1991). Planning and Chaos Theory. *Journal of the American Planning Association*, *57*(1), 44-56.

CEN. (2002). Transportation - Logistics and services - Public passenger transport - Service quality definition, targeting and measurement. European Committee for Standardization. European Standard EN 13816

Chen, P., Tong, R., Lu, G., & Wang, Y. (2018). Exploring Travel Time Distribution and Variability Patterns Using Probe Vehicle Data: Case Study in Beijing. *Journal of Advanced Transportation*, *2018*, 3747632.

Chen, Z., & Fan, W. D. (2020). Analyzing travel time distribution based on different travel time reliability patterns using probe vehicle data. *International Journal of Transportation Science and Technology*, *9*(1), 64-75.

Chepuri, A., Joshi, S., Arkatkar, S., Joshi, G., & Bhaskar, A. (2020). Development of new reliability measure for bus routes using trajectory data. *Transportation Letters*, *12*(6), 363-374.

Chepuri, A., Ramakrishnan, J., Arkatkar, S., Joshi, G., & Pulugurtha, S. S. (2018). Examining Travel Time Reliability-Based Performance Indicators for Bus Routes Using GPS-Based Bus Trajectory Data in India. *Journal of Transportation Engineering, Part A: Systems*, *144*(5), 04018012.

Cho, E. (2016). Making Reliability Reliable:A Systematic Approach to Reliability Coefficients. *Organizational Research Methods*, *19*(4), 651-682.

Churchill, G. A. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research*, *16*(1), 64-73.

Clark, S., & Watling, D. (2005). Modelling network travel time reliability under stochastic demand. *Transportation Research Part B: Methodological*, *39*(2), 119-140.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.
Cronbach, L. J. (1960). *Essentials of psychological testing* (2nd ed. ed.). Harper.
Cronbach, L. J., & Shavelson, R. J. (2004). My Current Thoughts on Coefficient Alpha and Successor Procedures. *Educational and Psychological Measurement*, *64*(3), 391-418.
Deming, W. E. (1982). *Out of the crisis*. MIT Press.
Deming, W. E. (2018). *The new economics : for industry, government, education* (Third edition. ed.). MIT Press.
El Faouzi, N.-E., & Maurin, M. (2023). Reliability of travel time under log-normal distribution: Methodological issues and path travel time confidence derivation.
Eliasson, J. (2009, 2019-09-30). *Forecasting travel time variability.* 2009 European Transport Conference,
Elíasson, J. (2007, 2007-6-24). *The relationship between travel time variability and road congestion.* 11th World Conference on Transport Research, Berkeley CA, United States.
Fosgerau & Karlström (2010). The value of reliability. *Transportation Research Part B: Methodological*, *44*(1), 38
Frazier, C., & Kockelman, K. M. (2004). Chaos Theory and Transportation Systems: Instructive Example. *Transportation Research Record*, *1897*(1), 9-17.
Gammelgaard, B. (2017). Editorial: the qualitative case study. *The International Journal of Logistics Management*, *28*(4), 910-913.
Gan, H., & Bai, Y. (2014). The effect of travel time variability on route choice decision: a generalized linear mixed model based analysis. *Transportation*, *41*(2), 339-350.
Gittens, A., & Shalaby, A. (2015). Evaluation of Bus Reliability Measures and Development of a New Composite Indicator. *Transportation Research Record*, *2533*(1), 91-99.
Goforth, C. (2015, 2015-11-16). *Using and Interpreting Cronbach's Alpha*. University of Virginia Library.
Guessous, Y., Aron, M., Bhouri, N., & Cohen, S. (2014). Estimating Travel Time Distribution under Different Traffic Conditions. *Transportation Research Procedia*, *3*, 339-348.
Guo, F., Rakha, H., & Park, S. (2010). Multistate Model for Travel Time Reliability. *Transportation Research Record*, *2188*(1), 46-54.
Hall, R. W. (1983). Travel outcome and performance: The effect of uncertainty on accessibility. *Transportation Research Part B: Methodological*, *17*(4), 275-290.
Hansson, S. O. (2017, 2019-04-17). *Science and Pseudo-Science*, Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.
Harsha, M. M., & Mulangi, R. H. (2022). Probability distributions analysis of travel time variability for the public transit system. *International Journal of Transportation Science and Technology*, *11*(4), 790-803.
Harsha, M. M., Mulangi, R. H., & Kumar, H. D. D. (2020). Analysis of Bus Travel Time Variability using Automatic Vehicle Location Data. *Transportation Research Procedia*, *48*, 3283-3298.
Hollander, Y., & Liu, R. (2008). Estimation of the distribution of travel times by repeated simulation. *Transportation Research Part C: Emerging Technologies*, *16*(2), 212-231.
Hounsell, M. (2018a). *Better service through runtime savings – Inner West Light Rail case study.* Australasian Transport Research Forum (ATRF), Darwin, Northern Territory, Australia. On researchgate.net
Hounsell, M. (2018b). *Implications of running time variability for passengers and operators–a case study using Bondi's 333 buses.* Australasian Transport Research Forum (ATRF), Darwin, NT, Australia On researchgate.net
Hounsell, M. (2020). Using TOTOR datasets in transport operations - Facilitating an empirically-driven continuous-optimisation approach to sustainable transport operations using TOTOR datasets [Conventional Thesis, University of Technology Sydney]. http://hdl.handle.net/10453/144073
Hounsell, M. (2021). A Literature Review for a customer centric framework for empirically driven corridor improvement strategies that use big data to increase the market share for sustainable transport. School of Civil and Environmental Engineering, UTS. http://dx.doi.org/10.13140/RG.2.2.16902.50247/1
Hounsell, M. (2022a). *Case Study: Standard deviation as a reliability measure for public transport managed by headways.* Australasian Transport Research Forum (ATRF), Adelaide, South Australia. On researchgate.net
Hounsell, M. (2022b). An examination of an approach to assist transit in maintaining political legitimacy by achieving reliability for passengers and efficiency for the public. ATRF, Adelaide, SA. On researchgate.net
Hounsell, M. (2023). *Case Study of NSW train patronage in 2023 vs 2020* [Public Analysis]. University of Technology Sydney On researchgate.net
Huebner & Lucht (2019). Generalizability theory in R. *Practical Assessment, Research, and Evaluation*, *24*(1), 5.
Jacoby, J. (1976). Association for Consumer Research (ACR) Presidential Address — Consumer Research: Telling it like it is. Advances in Consumer Research - North American Advances,
Jacoby, J. (1978). Consumer Research: How valid and useful are all our consumer behavior research findings?: A State of the Art Review. *Journal of marketing*, *42*(2), 87-96.
Johnson, M. B. (1966). Travel Time and the Price of Leisure. *Economic Inquiry*, *4*(2), 135-145.

Kieu, L.-M., Bhaskar, A., & Chung, E. (2015). Public Transport Travel-Time Variability Definitions and Monitoring. *Journal of Transportation Engineering*, *141*(1), 04014068.

Kim, J., & Mahmassani, H. S. (2015). Compound Gamma representation for modeling travel time variability in a traffic network. *Transportation Research Part B: Methodological*, *80*, 40-63.

Kimpel, T. J., Strathman, J. G., & Callas, S. (2008). Improving Scheduling Through Performance Monitoring. In Hickman, Mirchandani, & Voß, *Computer-aided Systems in Public Transport* Berlin, Heidelberg.

Lam, T. C. (2000). *The effect of variability of travel time on route and time -of -day choice* (Publication Number 9980901) [Ph.D., University of California, Irvine].

Levinson, H. S. (1991). *Synthesis of Transit Practice 15: Supervision strategies for improved reliability of bus routes*. National Cooperative Transit Research and Development Program, Transportation Research Board

Li, Tu, & Hensher (2016). Integrating the mean–variance and scheduling approaches to allow for schedule delay and trip time variability under uncertainty. *Transportation Research Part A: Policy and Practice*, *89*, 151-163.

Li, R., Bradley, M., Jones, M., & Moloney, S. (2015, 2015-09-30). *Quality investigation and variability analysis of GPS travel time data in Sydney.* Australasian Transport Research Forum (ATRF), Sydney, NSW, Australia.

Li, R., Chai, H., & Tang, J. (2013). Empirical Study of Travel Time Estimation and Reliability. *Mathematical Problems in Engineering*, *2013*, 504579.

Lomax, T., Schrank, D., Turner, S., & Margiotta, R. (2003). *Selecting travel time reliability measures*. Texas Transportation Institute and Cambridge Systematics Inc.

Lyman, K., & Bertini, R. L. (2008). Using Travel Time Reliability Measures to Improve Regional Transportation Planning and Operations. *Transportation Research Record*, *2046*(1), 1-10.

Ma, Z., Ferreira, L., Mesbah, M., & Zhu, S. (2016). Modeling distributions of travel time variability for bus operations. *Journal of Advanced Transportation*, *50*(1), 6-24.

Mahmassani, H. S., Hou, T., & Dong, J. (2012). Characterizing Travel Time Variability in Vehicular Traffic Networks:Deriving a Robust Relation for Reliability Analysis. *Transportation Research Record*, *2315*(1), 141

Martin, E. C. (2019). *Science and Ideology*, The Internet Encyclopedia of Philosophy. University of Tennessee

Martin, S. P. (2018). *Analysis of operational procedure within the inner west light rail to reduce end-to-end run time* (Publication Number A18-139) [Capstone, University of Technology Sydney]. Ultimo, NSW.

Mazloumi, E., Currie, G., & Rose, G. (2008). *Causes of travel time unreliability – a Melbourne case study.* Australasian Transport Research Forum (ATRF), Gold Coast, Queensland, Australia.

Mazloumi, E., Currie, G., & Rose, G. (2009). Using GPS data to gain insight into public transport travel time variability. *Journal of Transportation Engineering*, *136*(7), 623-631.

Mileusnic, I. (2017). Reducing the impact of passenger behaviour on dwell times along the Inner West Light Rail through an improved passenger guidance program (Publication Number A17-308) [Capstone, UTS]. Ultimo.

Moghaddam, S. S., Noroozi, R., Casello, J. M., & Hellinga, B. (2011). Predicting the Mean and Variance of Transit Segment and Route Travel Times. *Transportation Research Record*, *2217*(1), 30-37.

Montgomery, D. C. (2013). *Introduction to statistical quality control* (7th ed. ed.). Wiley.

NIST. (2018). *Engineering Statistics Handbook* (SEMATECH, Ed.). National Institute of Standards and Technology, United States of America (USA).

Noland, R. B., & Polak, J. W. (2002). Travel time variability: A review of theoretical and empirical issues. *Transport Reviews*, *22*(1), 39-54.

Noland, R. B., & Small, K. A. (1995). Travel-Time Uncertainty, Departure Time Choice, and the Cost of the Morning Commute.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing*, *64*(1), 12.

Park, S., Rakha, H., & Guo, F. (2010). Calibration Issues for Multistate Model of Travel Time Reliability. *Transportation Research Record*, *2188*(1), 74-84.

Park, S., Rakha, H., & Guo, F. (2011, 5-7 Oct. 2011). Multi-state travel time reliability model: Impact of incidents on travel time reliability. 2011 14th International IEEE Conference on Intelligent Transportation Systems,

Peter, J. P. (1979). Reliability: A review of psychometric basics and recent marketing practices. *JMR, Journal of Marketing Research (pre-1986)*, *16*(000001), 6.

Pu, W. (2011). Analytic Relationships between Travel Time Reliability Measures. *Transportation Research Record*, *2254*(1), 122-130.

Puvvala, R. K. (2014). *Cronbach's coefficient as a performance measure to assess link-level reliability* (Publication Number 1585386) [M.S., The University of North Carolina at Charlotte].

Rahman, M. M., Wirasinghe, S. C., & Kattan, L. (2018). Analysis of bus travel time distributions for varying horizons and real-time applications. *Transportation Research Part C: Emerging Technologies*, *86*, 453-466.

Rakha, H., El-Shawarby, I., & Arafeh, M. (2010). Trip Travel-Time Reliability: Issues and Proposed Solutions. *Journal of Intelligent Transportation Systems*, *14*(4), 232-250.
Ramakrishnan, J., Kumar, B. A., Arkatkar, S. S., & Vanajakshi, L. (2018). Performance Comparison of Bus Travel Time Prediction Models across Indian Cities. *Transportation Research Record*, *2672*(31), 87-98.
RTA. (2002). Guide to Traffic Generating Developments - October 2002. Roads and Traffic Authority, State of New South Wales. Version 2.2
Sameni, M. K. (2012). *Railway track capacity: measuring and managing* [Doctoral, University of Southampton].
Samra, O. (2017). *Investigating operating procedures on the Inner West Light Rail to Improve Level Of Service* (Publication Number A17-196) [Capstone, University of Technology Sydney]. Ultimo,NSW
Shah, R., & Ward, P. T. (2007). Defining and developing measures of lean production. *Journal of Operations Management*, *25*(4), 785-805.
Shao, H., Lam, W. H. K., Meng, Q., & Tam, M. L. (2006). Demand-Driven Traffic Assignment Problem Based on Travel Time Reliability. *Transportation Research Record*, *1985*(1), 220-230.
Small (1982). The Scheduling of Consumer Activities: Work Trips. *The American Economic Review*, *72*(3), 467
Smalley, J. (2017). *Improving the level of service of the Sydney Inner West Light Rail in the Sydney city district* (Publication Number S17-220) [Capstone, University of Technology Sydney]. Ultimo, NSW
Smith, S. (2014). Muda, Muri and Mura. *Lean & Six Sigma Review*, *13*(2), 36.
Strathman, J. G., Kimpel, T. J., Dueker, K. J., Gerhart, R. L., & Callas, S. (2002). Evaluation of transit operations: data applications of Tri-Met's automated Bus Dispatching System. *Transportation*, *29*(3), 321-345.
Suárez-Barraza, M. F., Smith, T., & Dahlgaard-Park, S. M. (2012). Lean Service: A literature analysis and classification. *Total Quality Management & Business Excellence*, *23*(3-4), 359-380.
Susilawati, S., Taylor, M. A. P., & Somenahalli, S. V. C. (2013). Distributions of travel time variability on urban roads. *Journal of Advanced Transportation*, *47*(8), 720-736.
Taylor. (1982). Travel Time Variability—The Case of Two Public Modes. *Transportation Science*, *16*(4), 507-521.
Taylor, M. A. P. (2013). Travel through time: the story of research on travel time reliability. *Transportmetrica B: Transport Dynamics*, *1*(3), 174-194.
TCRP. (2003). *Transit capacity and quality of service manual* (2nd ed). Transportation Research Board, Transit Development Corporation, Transit Cooperative Research Program, Federal Transit Administration.
TfNSW. (2012). NSW Long Term Transport Master Plan. Transport for NSW, New South Wales, Australia.
TfNSW. (2017). *Bus Occupancy - Aug 2016 to Jan 2017* (f9393f9c-f8ad-4628-bec2-542dbc8ad9de; Version 2017-07-28) Transport for NSW (TfNSW), State of New South Wales.
TfNSW. (2018). Sydney Bus Service Contract (SBSC) - R6BSP. Transport for NSW, State of NSW.
TfNSW. (2020a). *BOAM - Bus Opal Assignment Model.* (d1a0828f6b-7ad2-4f0c-890b-d0a5483a0485) TPA, NSW.
TfNSW. (2020b). *Opal Patronage.* (5786d677-c319-48f1-b399-a17c2dc08ca0) TPA, NSW.
Thornton, S. (2014, 2014-12-31). *Karl Popper*, Stanford Encyclopedia of Philosophy.
Tilahun, N. Y., & Levinson, D. M. (2010). A Moment of Time: Reliability in Route Choice Using Stated Preference. *Journal of Intelligent Transportation Systems*, *14*(3), 179-187.
Uno, N., Kurauchi, F., Tamura, H., & Iida, Y. (2009). Using Bus Probe Data for Analysis of Travel Time Variability. *Journal of Intelligent Transportation Systems*, *13*(1), 2-15.
Unsworth, B. (2004). Review of bus services in New South Wales. State of NSW.
Vaidya, O. S. (2018). A six sigma based approach to evaluate the on time performance of Indian railways. *International Journal of Quality & Reliability Management*, *35*(10), 2212-2226.
van Lint, J. W. C., & van Zuylen, H. J. (2005). Monitoring and Predicting Freeway Travel Time Reliability:Using Width and Skew of Day-to-Day Travel Time Distribution. *Transportation Research Record*, *1917*(1), 54-62.
Vlahogianni, E., & Karlaftis, M. (2011). Temporal aggregation in traffic data: implications for statistical characteristics and model choice. *Transportation Letters*, *3*(1), 37-49.
Yazici, M. A., Kamga, C., & Mouskos, K. C. (2012). Analysis of Travel Time Reliability in New York City Based on Day-of-Week and Time-of-Day Periods. *Transportation Research Record*, *2308*(1), 83-95.
Zeibots, M. E. (2016). *Lecture Notes - 48370 Road & Transportation Engineering* [Lecture Notes]. UTS.
Zheng, F., Li, J., van Zuylen, H., Liu, X., & Yang, H. (2018). Urban travel time reliability at different traffic conditions. *Journal of Intelligent Transportation Systems*, *22*(2), 106-120.
Zheng, F., Liu, X., & van Zuylen, H. (2017). A Methodological Framework of Travel Time Distribution Estimation for Urban Signalized Arterial Roads [Article]. *Transportation Science*, *51*(3), 893-917.
Zheng, F., & van Zuylen, H. (2015, 2015-01-11). *Travel Time Distribution Estimation for Urban Signalized Arterials Considering Spillback.* 94th meeting of the Transportation Research Board, Washington D.C.