Research Article

Measuring Service Reliability Using Automatic Vehicle Location Data

Zhenliang Ma, Luis Ferreira, and Mahmoud Mesbah

School of Civil Engineering, The University of Queensland, Brisbane, QLD 4072, Australia

Correspondence should be addressed to Zhenliang Ma; zhenliang.ma@uq.edu.au

Received 20 December 2013; Accepted 3 April 2014; Published 24 April 2014

Academic Editor: Paolo Lonetti

Copyright © 2014 Zhenliang Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Bus service reliability has become a major concern for both operators and passengers. Buffer time measures are believed to be appropriate to approximate passengers' experienced reliability in the context of departure planning. Two issues with regard to buffer time estimation are addressed, namely, performance disaggregation and capturing passengers' perspectives on reliability. A Gaussian mixture models based method is applied to disaggregate the performance data. Based on the mixture models distribution, a reliability buffer time (RBT) measure is proposed from passengers' perspective. A set of expected reliability buffer time measures is developed for operators by using different spatial-temporal levels combinations of RBTs. The average and the latest trip duration measures are proposed for passengers that can be used to choose a service mode and determine the departure time. Using empirical data from the automatic vehicle location system in Brisbane, Australia, the existence of mixture service states is verified and the advantage of mixture distribution model in fitting travel time profile is demonstrated. Numerical experiments validate that the proposed reliability measure is capable of quantifying service reliability consistently, while the conventional ones may provide inconsistent results. Potential applications for operators and passengers are also illustrated, including reliability improvement and trip planning.

1. Introduction

Considerable effort has been made by transport planners to implement strategies related to advanced technologies capable of improving transport service reliability. Improvements in public transport service reliability will produce benefits for both passengers and operators [1]. Routes characterized by unreliable service may have difficulty in attracting potential riders and suffer patronage declines over time. Increased perceived burdens of waiting at stops may ultimately impact mode choice decisions. Transit systems with poor reliability performance require extra fiscal resources due to higher operation costs. A survey study suggests that it is twice as important from a passenger's perspective to improve reliability as to increase the frequency of the service [2].

For transport planners, the workable and consistent reliability measurement improves several different aspects central to the provision of transport system service [3],

including identifying and understanding problems in reliability, identifying and measuring actual improvements in reliability, relating such improvements to particular strategies, and modifying strategies to obtain greater reliability improvements.

To a large extent, the implementations of these and other applications depended on the data that are available for service reliability quantification. Previously, manually surveyed data at service terminals allows transport planners to gain a snapshot of service quality as a result of high collection cost and limited area coverage. The emergence of automatic vehicle location (AVL) technology has led to great improvements in this area, producing a wealth of accurate, continuous, and automated point-to-point data on individual vehicle movements that could be used to assess service reliability more cost-effectively [4]. Most importantly, the large scale of spatial-temporal data coverage enables analysing reliability at a high-level detail that would provide

deep insights into the nature of service attributes and thus help planners to make efficient strategies to improve service reliability.

By taking advantage of the availability of the AVL data, this paper aims to propose a set of service reliability measures that can capture passengers' perspectives on reliability and develop practical applications for use by operators and passengers, respectively. The remainder of the paper is organized as follows. Section 2 provides a brief overview of service reliability. Buffer time measures are specially discussed in Section 3 and two issues with regard to buffer time estimation are identified. Section 4 explains the methodology for performance disaggregation and develops reliability measures for operators and passengers. Section 5 presents case studies to investigate travel time distributions using empirical AVL data from Brisbane and validate the proposed measures effectiveness using a numerical example. Potential applications of the proposed measures to operators and passengers are discussed as well. Finally, Section 6 provides the conclusion of the study and future work.

2. Overview of Service Reliability

Abkowitz et al. [3] defined reliability as the invariability of service attributes that could influence the decisions of passengers and operators. The distinction made between the perspective of passengers and operators is important when developing reliability measures because each is influenced differently by service unreliability [5].

2.1. Operator-Oriented Measures. On-time performance and headway regularity are the two mostly used operator-oriented service reliability measures. For routes characterized by low frequency services, on-time performance plays the most significant role, since passengers plan their arrivals to coordinate with the scheduled departures to minimize waiting time at stops with a tolerance probability of missing the expected trips [6]. On-time performance is defined as the percentage of trips that depart up to m minutes late and n minutes early from the scheduled departure time [7]. For routes characterized by high frequency services, headway regularity becomes important [8]. In these circumstances, passengers tend to arrive at stops randomly, and the aggregated waiting time is minimized when services are evenly spaced [9]. Although the operator-oriented measures often help to illustrate the level of service provided for passengers, they do not completely match their actual experienced service reliability. For instance, by altering the on-time tolerance interval from 5 minutes to 10 minutes, the measured reliability improves without any changes experienced by passengers [10]. Also, driving ahead or being late would have totally different impacts on passengers.

2.2. Passengers' Perspective on Reliability. For a complete journey, excluding access time from the origination and egress time to the destination, a passenger is concerned with waiting time at the first stop, in-vehicle time during the trip, and transfer time between different trips [11]. Generally,

the unreliability can impact the duration and predictability of travel time which ultimately influence passengers' trip planning behaviours [12]. Due to the vehicle travel time variability, passengers may experience longer or shorter journey times which lead to early or late arrivals at their destination. These can be quantified using a measure of variability, such as standard deviation of travel time. In addition, unreliable service brings uncertainty to travel time which hinders passengers' ability to make optimal travel decisions to minimize disutility [5]. For an infrequent service, passengers tend to arrive as close to their desired service departure time without missing the expected vehicle at the first stop.

For a frequent service, passengers would be more interested in choosing a departure time that can minimize their late arrivals. If a passenger with a desired arrival time T_{des} travels in an ideal transport system without any variability, the departure time should be exactly the desired arrival time T_{des} minus the expected travel time of the trip TT_{exp} . However, in reality, a passenger would experience a stochastic arrival time distribution with nonzero probability of a late arrival for each departure time as shown in Figure 1.

If a passenger valued an on-time arrival more than in-vehicle travel time, he/she should shift the departure time earlier to reduce the probability of a late arrival. For example, to guarantee a late arrival probability no more than 5%, the passenger should leave before $T_{dep,early}$ which is the difference between the desired arrival time T_{des} and 95th percentile travel time of a trip TT_{95prc} . This additional time budgeted by a traveller to increase the probability of an on-time arrival can be regarded as buffer time. Intuitively, as the service reliability decreases, a passenger needs to budget a larger buffer time to avoid a late arrival.

2.3. Passenger-Oriented Measures. Generally, there are two different groups of waiting time reliability measures, namely, average waiting time and potential waiting time [13, 14]. The former one quantifies the expected waiting time that a passenger would experience at a stop while the latter one indicates the additional time that a passenger budgets for his/her arrival at a stop to avoid missing an expected bus [15]. Lomax et al. identified three types of travel time reliability measures: measure of statistical range, measure of additional budgeted travel time, and measure of tardy trips [16]. The first type of measure typically serves as an approximate estimate of the range of in-vehicle time experienced by passengers, calculated based on standard deviation statistics. The second type of measure represents the additional time that a passenger budgets to reduce the probability of late arrival at the destination. The third type of measure relates to the probability that a passenger may encounter an extremely long travel time and is calculated by setting an unacceptable threshold value in the form of additional time plus expected time. Transfer time can be straightforwardly calculated from scheduled stops using smart card data [17, 18]. Therefore, statistical indicators can be applied to measure transfer time reliability, such as the coefficient of variation of transfer delays [19]. However, day-to-day arrival time variations make

FIGURE 1: Journey departure decision and arrival time distribution.

the measurement rather difficult [7]. Transfer waiting time usually serves as a transfer time reliability measure [11, 20–22].

3. Buffer Time Concept and Estimation

Although there is no consensus on which attribute is capable of appropriately characterizing service reliability due to the heterogeneity of stakeholders' preferences and perceptions, buffer time measure is believed to be more appealing than its alternatives [8]. Mathematically, percentile-based buffer time is an indicator of the compactness of travel time distribution, while conceptually it can capture the influence of service variability on passengers travel decisions [3]. Analytical and empirical studies have confirmed buffer time as a powerful tool in indicating service reliability [23].

Generally, buffer time is defined as the 95th percentile travel time minus average or median travel time which indicates the additional time that a passenger should budget to guarantee an on-time arrival under a given probability [16]. Although the term “buffer time” usually denotes buffer travel time, it could be recognized as a concept of extreme-value based reliability evaluation concept, which can be applied manifoldly to the following: (a) buffer waiting time to indicate excess waiting time needed to catch an expected bus [24]; (b) buffer transfer time to indicate additional time required to avoid missing connections [20]; and (c) buffer travel time to indicate extra time necessary for an on-time arrival [25].

However, two causes may reduce the usefulness of the existing buffer time measures in the context of performance evaluation when directly applying it to mixture travel time distributions. One reason is that two different mixture travel time distributions could have the same buffer time value as shown in Figure 2. It shows that groups A and B have different probability density functions (PDFs), thus different reliability performance. However, they have exactly the same buffer time value (4.7 minutes) calculated using cumulative density functions (CDFs). This is conceptually unreasonable in reality. Further, any travel time samples with the same 95th percentile and median travel times would have the same buffer time value. It indicates that applying the buffer time measure directly on the source travel time profile could also lead to an inconsistent reliability assessment when mixture distributions exist.

The other reason is that, by considering the travel time distribution as a whole, the buffer time measure could hide

FIGURE 2: Travel time samples with mixture distributions.

the sources of observed reliability changes, thus making it hard for the identification of unreliability factors [10]. For example, many studies have claimed that travel time distribution can be classified into a recurrent and nonrecurrent (incident-influenced) states for a given time period [26–28]. Suppose the nonrecurrent state (uncontrollable) occurrence probability is 10%; then no matter how the recurrent state travel time performance changes (e.g., decrease standard deviation of the recurrent travel time distribution), the buffer time value will keep constant since the 95th percentile total travel time will remain the same (assume the median travel time is unchanged).

In addition, passengers would experience different arrival time distributions and thus have different departure decisions under different occasions. Conceptually, bus service state can be classified into three types, namely, a fast service state, a slow service state, and a nonrecurrent state. The former two states can also be aggregated together as the recurrent state. It should be noted that the service state in this paper is related to travel time. Bus travel time under different states can have different characteristics. In a recurrent state, travel time is largely determined by traffic flow fluctuations and passenger demand characteristics. The difference between the fast and slow service states is mainly caused by stop delays (e.g., red light and queuing) and intersection delays (e.g., serving passengers, bus bunching, and merging to the traffic flow). A vehicle in a fast service state may experience less intersection delays and stop delays than one in a slow service state. In this case, the passenger taking a fast service would plan less “buffer time” than one taking a slow service or even plan no “buffer time” if the 95th percentile arrival time $TT_{arrival,95prc}$ under a fast service state is already smaller than the desired arrival time T_{des} (Figure 1). In the nonrecurrent state, the corresponding travel time unreliability will be higher than that in the recurrent state. In addition, the nonrecurrent state can be further broken down to a more refined subset influenced by different factors, such as incidents, weather, and extreme events [27]. In this case, passengers would not

consider the nonrecurrent state travel time in trip planning since the nonrecurrent state cannot be repeated.

In conclusion, buffer time measure can evaluate the reliability experienced by passengers in the context of departure planning using operational data. However, directly applying buffer time measure to a whole travel time distribution may give inconsistent reliability assessments, may hide unreliability causes, and could not effectively capture passengers' departure behaviours. The AVL system provides analysts with a wealth of high-level detailed spatial-temporal operational data. It is reasonable to develop a "buffer time" concept based measure that can assess service reliability under different states separately rather than together using AVL data. Two issues related to reliability measure development will be addressed in this paper, namely, performance disaggregation and capturing passengers' perspectives on reliability.

4. Methodology

Based on the discussions before, the primary task is to disaggregate the overall travel time performance for a trip origination-destination (OD) pair in a specific time period across different days into different states (or categories). A Gaussian mixture models (GMM) approach is applied to disaggregate the performance data that can identify the underlying statistics of sample data. After that, a reliability buffer time measure is proposed to capture passengers' perspective on reliability and different set of measures are developed for use by operators and passengers, respectively.

4.1. Performance Disaggregation. Mixture models are a type of density models which comprise a number of component functions, usually Gaussian. These component functions are combined to provide a multimodal density. Mixture models provide a great flexibility and precision in modelling the underlying characteristics of performance data and they are able to smooth over gaps from sparse sample data. In addition, they can directly output the distribution functions which are the prerequisite for the analysis and calculation of buffer time under different states.

From the mathematical perspective, the mixture models are a mixture of finite number of component distributions. Each component represents the travel time performance under its corresponding state. The mixture coefficients of component distributions indicate the occurrence probabilities of different service states. A mixture model for travel time with finite K components has the following probability density function (PDF) [26]:

$$f(\mathbf{t} | \mathbf{w}, \boldsymbol{\theta}) = \sum_{k=1}^K w_k f_k(\mathbf{t} | \boldsymbol{\theta}_k), \quad (1)$$

where $\mathbf{t} = D$ -dimensional continuous-valued travel time data, $\mathbf{w} = (w_1, \dots, w_K)$ vector of mixture coefficients such that $w_k \geq 0$ and $\sum w_k = 1$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K)$ matrix of model parameters for each component, $\boldsymbol{\theta}_k = (\theta_{k1}, \dots, \theta_{kl})$ vector of model parameters for component k , and $f_k(\mathbf{t} | \boldsymbol{\theta}_k) =$ probability density for component k .

By changing the component distributions (e.g., normal, log-normal, or gamma) and the mixture coefficients, a mixture model is flexible to approximate a large range of different travel time distributions. The parameters of the mixture models can be estimated using an expectation and maximization (EM) algorithm based on maximum likelihood estimation criteria [26]. The basic idea of EM algorithm is beginning with an initial model f to estimate a new model $\bar{f}$, such that $\text{prob}(\mathbf{t} | f) \geq \text{prob}(\mathbf{t} | \bar{f})$. The new model then becomes the initial model for the next iteration and the process is repeated until the desired convergence threshold is reached. Based on the posterior probability of each data point belonging to each cluster, the performance data is disaggregated automatically.

A special case is 2-component Gaussian mixture models (GMM2) which is a weighted sum of two component Gaussian densities as given by [29]

$$g(t | \lambda, \mu_1, \sigma_1, \mu_2, \sigma_2) = \lambda g_1(t | \mu_1, \sigma_1) + (1 - \lambda) g_2(t | \mu_2, \sigma_2), \quad (2)$$

where λ = mixture coefficient for the first component, $g_1(t | \mu_1, \sigma_1)$ = density for the first component following a normal distribution with mean μ_1 and variance σ_1^2 , and $g_2(t | \mu_2, \sigma_2)$ = density for the second component following a normal distribution with mean μ_2 and variance σ_2^2 . Each component density is a one-variate Gaussian function having the following form:

$$g_i(t | \mu_i, \sigma_i) = \frac{1}{(2\pi)^{1/2} \sigma_i} \exp \left\{ -\frac{1}{2\sigma_i^2} (t - \mu_i)^2 \right\}. \quad (3)$$

The mixture models connect the parameters of mixture distributions with the underlying service states. In particular, the mixture coefficient w_k in (1) can be interpreted as the probability that a travel time that follows the travel time state k (fast service state, low service state, and nonrecurrent service state) and the component distribution $f_k(\mathbf{t} | \boldsymbol{\theta}_k)$ indicates the distribution of travel time under such state. These connections provide an opportunity for analysing passengers' experienced reliability under different states separately and then aggregating together to get a complete picture of the expected travel time reliability that a passenger would experience for a trip.

4.2. Reliability Measurement Development. The concepts of service variability and reliability are different. Service variability is defined as the distribution of output values for the supply side of public transport, such as vehicle trip time, departure time, and headways. It indicates the objective service performance provided by operators. Service reliability is defined as the degree of matching between the supplied service and the expected service. A service with high variability does not necessarily lead to poor reliability experienced by passengers. Given an expected trip travel time (e.g., 40 min), a passenger would perceive an early arrival time with large variability (e.g., range 34–39 min) as more reliable than a late arrival time with small variability (e.g.,

range 42–45 min). It is reasonable to capture passengers' different perspectives on reliability under different conditions given a certain expectation.

4.2.1. Reliability Buffer Time. Denote by $F_k(t)$ the cumulative density function for the probability density function $f_k(t)$ under service state k in (1). Further denote by $F_k^{-1}(x)$ the inverse distribution function of $F_k(t)$. Then, the x th percentile travel time TT_k^x under service state k can be calculated as follows:

$$TT_k^x = F_k^{-1}(x). \quad (4)$$

Along the same route for the definition of traditional buffer time, the reliability buffer time RBT_k for service state k is defined as the difference between M th percentile travel time TT_k^M under service state k and N th percentile travel time TT_{typical}^N under a typical condition. By the nature of the buffer time concept (additional time budgeted for a trip), the value of buffer time should be no less than zero. The RBT can be formulated as follows:

$$RBT_k = \begin{cases} TT_k^M - TT_{\text{typical}}^N, & \text{if } TT_k^M \geq TT_{\text{typical}}^N \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

The recurrent service state is chosen as the typical condition instead of using the whole service states, because in public transport, the direct expectation of a trip travel time comes from the time-table published by operators which is usually designed based on the average travel time under recurrent service state. In addition, it is meaningless to incorporate the unpredictable incident-influenced nonrecurrent state in modelling a service expectation from the perspective of passengers, even though they might experience long travel times. The selection of M and N depends on the usage purpose and transit passengers' preferences. Wakabayashi and Matsumoto [30] presented a detailed performance analysis of different percentile-based reliability measures and their relationships. Usually, M and N are chosen as 95 and 50, respectively. The 95th percentile refers to a traveller who can be late for a work one time a month without getting in too much trouble [16]. The 50th percentile relates to the typical travel time for a trip under a certain condition.

For a given trip OD pair in a specific time period, TT_k^M represents the service variability performance provided by the operators under state k , and TT_{typical}^N represents the travel time expectation for such service by passengers. Figure 3 illustrates the possible service states for a trip and the calculation of RBTs under different service states. The definition of RBT for different service states could be regarded as an approximation of passengers' experienced buffer times under different situations. It could be interpreted like this, if a passenger experiences a slow service state for his/her trip, the budgeted buffer time required to guarantee a late arrival possibility less than 5% is the difference between 95th percentile travel time (TT_{slow}^{95}) and his/her expectation of the service ($TT_{\text{recurrent}}^{50}$). However, if a passenger experiences a fast service state that 95th percentile travel time (TT_{fast}^{95}) is

FIGURE 3: Illustration of possible service states and reliability buffer time.

still in his/her expectation ($TT_{\text{recurrent}}^{50}$), there is no need to budget any buffer time no matter how variable the service performance is under such state.

4.2.2. Expected Reliability Buffer Time (Operators). As a service industry, public transport operators are concerned with providing satisfactory service to their passengers. Reliability measures are required to indicate current service reliability, identify causes of unreliability, assess different strategies effect on reliability, and modify strategies to improve it. Based on different application objectives, a set of expected reliability buffer time measures (ERBT) are developed for operators by using different combinations of RBTs in (5).

The OD-level ERBT ($ERBT_{\text{od}}$) is defined as the occurrence probabilities weighted RBT_k under different states for an OD trip pair, within a specific day time period across different days:

$$ERBT_{\text{od}} = \left(\sum_{k=1}^K p_k \times RBT_k \right)_{\text{time period, days}}, \quad (6)$$

where p_k is the occurrence probability of the k th service state. The subscripts indicate different dimensions of aggregation including OD pairs, time period (e.g., morning peak, off peak, or afternoon peak), and a span of time over many days (e.g., a 12-weekday sample). Changing the latter two dimensions allows for different levels of temporal analysis.

The line-level ERBT ($ERBT_{\text{line}}$) spatially aggregates $ERBT_{\text{od}}$ weighted by the passenger demand of an OD pair along a certain line during a studied time period [31]:

$$ERBT_{\text{line}} = \left(\frac{\sum_{\text{od} \in \text{line}} (d_{\text{od}} \times ERBT_{\text{od}})}{\sum_{\text{od} \in \text{line}} d_{\text{od}}} \right)_{\text{time period, days}}, \quad (7)$$

where d_{od} is the passenger demand of an OD pair along the concerned line.

The $ERBT_{\text{od}}$ and $ERBT_{\text{line}}$ measures indicate the average overall reliability performance for a given service for different temporal-spatial scales. The passengers' experienced RBTs for different states are aggregated together based on their contributions to the overall performance instead of treating them equally. For example, the nonrecurrent travel times would cause the highest RBT but if they rarely happen, the contribution of them to the overall service performance

should not be so much. In addition, as different factors contributing to different states, the RBTs in (5) could be used to separate different factors influenced reliability apart which can make causes identification and strategies assessment more effectively. To make a fair reliability compare between two different services, the ERBT index (ERBTI) can be applied which is defined as the expected reliability buffer time divided by the expected travel time (e.g., travel time under recurrent state for OD level).

4.2.3. Trip Planning Time (Passengers). Although the passengers would experience different service states in their daily travels, it is hard for them to get an accurate and complete picture of the operational performance by themselves as shown in Figure 3. However, online applications of trip planner on websites and mobile phones provide a good way to convey such information to the public. Current trip planners provide a departure time calculated using average trip duration based on which a passenger could expect a low chance of on-time arrival. For a bus trip planning, passengers are concerned with deciding departure time to avoid late arrivals at their destinations and thus they are more interested in travel time reliability than travel time itself. Considering travel mode choice and departure time planning, two types of times should be interesting to a passenger, namely, the average trip duration and the latest trip duration. The former one could be used to make a service mode choice while the latter one could be used to determine the departure time for a trip.

The average trip duration (ATD) for a trip is defined as the 50th percentile travel time under the recurrent service state instead of the whole service states for a specific time period over different days. Nonrecurrent service state is excluded because it is rare and unpredictable in reality, and including it would increase the ATD to a much higher value, which is meaningless for a trip planning:

$$\text{ATD} = \left(\text{TT}_{\text{recurrent}}^{50} \right)_{\text{trip, time period, days}}, \quad (8)$$

where the subscript trip indicates a passenger travel from a boarding stop to an alighting stop along the same service route.

The latest trip duration (LTD) for a trip is defined as the 95th percentile travel time under the slow service state for a specific time period over different days. It indicates that, in 95% occasions, the vehicle would arrive at the destination using less than the LTD time. In other words, if a passenger plans a trip according to the LTD, he/she would encounter late arrival only once in a month (5% late arrival). Consider

$$\text{LTD} = \left(\text{TT}_{\text{slow}}^{95} \right)_{\text{trip, time period, days}}, \quad (9)$$

One of the issues in practical application is how to determine the minimum number of observations to ensure that the estimated travel time distribution can represent the trip travel time with reasonable accuracy. According to the NCHRP

report [32], the minimum number of travel time observations is

$$N = 4 \times \left[t_{(1-\alpha/2), N-1} \times \frac{S}{\text{CI}_{(1-\alpha)\%}} \right]^2, \quad (10)$$

where $\text{CI}_{(1-\alpha)\%}$ = confidence interval for the true mean with probability of $(1 - \alpha)\%$, $t_{(1-\alpha/2), N-1}$ = t -statistic for the probability of two-sided error summing to α with $N - 1$ degree of freedom, and S = standard deviation in the measured travel times.

5. Case Studies and Potential Applications

AVL data were collected on a bus way section operating in Brisbane, Australia, from 5:30 am to 11:30 pm for a two-week time period. The selected route was approximately 20 km long with 10 stops. Figure 4 illustrates the studied transit service route. In total, 1794 trips are identified for travel time samples from Cultural Centre to Springwood.

The archived data were screened to minimize the possibility of erroneous data. Two filters were used, namely, erroneous trips and outliers. The erroneous trip filter excludes false trip records caused by incomplete trips, abnormal stops, and hardware failures, and, overall, 3.7% erroneous trips were identified. The outlier filter screens the abnormal records with extremely long travel time caused by false recording, and a total of 3.2% outlier trips are identified. The median absolute deviation (MAD) technique is applied for outlier identification. An item sample is considered as an outlier if it is outside the range of the lower bound value (LBV) and upper bound value (UBV) determined by the MAD 3-delta criteria [33, 34]. Travel time samples for weekday inbound (WD-IN) service during the morning (AM) peak period (7:00 am–9:00 am) are used in this study. A total of 85 trips are considered for this scenario. The morning peak period is chosen since the reliability performance is rather complex and different states of service can normally be observed during this period.

5.1. Case Study 1: Travel Time Distribution. Modelling travel time profile to a theoretical distribution is the prerequisite for the calculation of RBT and it can also provide the maximum information for reliability evaluation [35]. Most importantly, assigning travel time data points to different clusters based on distribution densities can disaggregate travel time data at a high-level of detail. Single and mixture distribution models are tested to verify whether mixture states exist for the WD-IN service during AM peak period.

5.1.1. Single Model Distribution. The single distribution model assumes the travel time samples come from a single travel time state during a given time period. The Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) tests were used to test the hypothesis that the AM-peak WD-IN travel time follows the potential theoretical distributions, including Burr, exponential, extreme value, gamma, log-normal, logistic, log-logistic, normal, and Weibull [36]. The hypothesis test results

TABLE 1: Hypothesis tests for potential travel time distributions.

Single models	K-S test (P -value)*	A-D test (P -value)*	Parameters [#]		Goodness-of-fit AIC
			Shape	Scale	
Burr	0.111 (0.089)	1.574 (0.160)	15.86/1.59	30.00	-590
Exponential	0.563 (<0.05)	46.657 (<0.05)	28.89	N/A	-1071
Extreme value	0.107 (0.110)	2.160 (0.075)	30.25	2.75	-612
Gamma	0.111 (0.088)	1.364 (0.212)	118.19	0.24	-585
Log-normal	0.110 (0.097)	1.339 (0.220)	3.36	0.09	-584
Logistic	0.110 (0.096)	1.643 (0.146)	28.83	1.59	-595
Log-logistic	0.111 (0.091)	1.625 (0.149)	3.359	0.055	-593
Normal	0.112 (0.087)	1.372 (0.210)	28.89	2.68	-587
Weibull	0.091 (0.241)	1.679 (0.139)	11.17	30.13	-602

* P value < 0.05 rejects the null hypothesis that the data come from the distribution.

[#]The scale parameter indicates the degree of the spread for travel time distribution.

The shape parameter indicates the shape and location of the travel time distribution.

FIGURE 4: Studied bus route service operation map.

are provided in Table 1. Parameters for the fitted travel time distribution are also provided and the goodness-of-fit is measured using Akaike information criterion (AIC) [37]. The test results illustrate that the travel time distribution can come from any individual theoretical distribution presented here, except the exponential model. However, the P values for the accepted distributions are rather low and the largest P value is only 0.241 (Weibull). These indicate the limited ability of single distribution models in modelling the AM-peak WD-IN travel time data profile.

An interesting phenomenon observed in Table 1 is that the K-S test and A-D test give inconsistent results. While the K-S test identifies the Weibull distribution as the best fitting model, A-D test shows the Log-normal distribution to be the best one. The seemingly inconsistent results are caused by the

TABLE 2: Parameters for the fitted single model and mixture models distributions.

	Parameters			AIC
	Mixture coefficient	Shape	Scale	
Observed	N/A	N/A	N/A	N/A
GMM2	0.21	26.1	0.164	-542
Log-normal	N/A	3.360	0.090	-584

fact that the K-S test is more sensitive to a distribution centre while A-D test is more sensitive to a distribution tail. The log-normal model has a relatively stronger ability for tail-fitting but a weaker ability for centre fitting when compared with the Weibull model as shown in Figure 5.

The AIC value indicates that the best distribution fitting model is log-normal. And many studies have also claimed log-normal as an appropriate travel time distribution model which can be justified from an equivalent theorem derived from central limit theorem [38]. Therefore, the log-normal model is selected as a representative of single model distribution for comparison purposes with the mixture models distribution.

5.1.2. Mixture Models Distribution. The source travel time profile is fitted using GMM2 model in (2). The parameters of the fitted distributions are shown in Table 2. From Figure 5, it can be seen that the mixture models are promising in capturing the bimodal characteristics of travel time distribution. The single model seems to have limited ability in tackling bimodal distribution. The DIP test confirms the existence of bimodal phenomenon in such distribution with P value less than 0.05 [39]. The goodness-of-fit AIC value in Table 2 also verifies the superiority of the GMM2 model when compared with the log-normal model in modelling travel time distribution for the AM-peak, WD-IN services.

The first component of the GMM model in Figure 5 can be regarded as the fast service state that encounters short stop delays and intersection delays and the second one is

FIGURE 5: Travel time distribution fitting result using single model and mixture models.

FIGURE 6: CDFs and PDFs for different groups of GMM2 travel time samples.

the slow service state that experiences long stop delays and intersection delays. According to the discussion in Section 2, there should be a third component as well, namely, nonrecurrent service state, that influenced by incidents (accident, road closure, or extreme events). However, as constrained by the data collection temporal scale, no incident occurred during the studied period.

5.2. Case Study 2: Effectiveness of ERBT. The proposed ERBT is validated by comparing with the existing reliability measures using numerical travel time samples, including planning time index, buffer time index, and reliability time index [36]. The planning time index (PTI) is calculated as the 95th percentile travel time TT^{95} divided by average travel time TT^{avg} :

$$PTI = \frac{TT^{95}}{TT^{avg}} \times 100\%. \quad (11)$$

The buffer time index (BTI) is calculated as the difference between the 95th percentile travel time and average travel time divided by average travel time:

$$BTI = \frac{TT^{95} - TT^{avg}}{TT^{avg}} \times 100\%. \quad (12)$$

The reliability time index (RTI) is calculated as the difference between the 95th percentile travel time and median travel time TT^{50} divided by median travel time:

$$RTI = \frac{TT^{95} - TT^{50}}{TT^{50}} \times 100\%. \quad (13)$$

Considering the fact that the BTI may be too conservative to incorporate random travel time fluctuations, the RTI is estimated by using median travel time instead of average travel time. To illustrate the current and proposed reliability measures of performance, different groups of travel time samples are generated using GMM2 models, with different parameters combination. To keep the central tendency the same for comparison, the parameters of means μ_1 and μ_2 for different groups are set equal to those of the empirical travel time samples. The parameters of proportions p_1 and p_2 are set randomly in range $[0, 1]$ with the constraint that $p_1 + p_2 = 1$. To illustrate the current measures limitations under some specific occasions, the parameters of sigmas σ_1 and σ_2 are calculated by solving the equation that $TT^{95} - TT^{50} = 4.4$ min. It indicates the fact that even with the same assessment result by the current measures, different shape of distributions can still exist, thus different actual reliability performance. Without loss of generality, five groups of travel time samples are presented to investigate different measures assessment ability. Table 3 shows the parameters for different groups travel time samples. The overall mean times, median times, and planning times (95th percentile travel time) are also calculated.

Figure 6 displays the CDFs and PDFs of different groups travel time samples. The order of the reliability performance can be identified by comparing the compactness of the distribution. That is, if the service travel times have a highly compact distribution, the service is perceived to be very reliable. From Figure 6, the reliability performance order (from the best to the worst) is identified as Group C, Group D, Group A, Empirical, Group B, and Group E.

Table 4 provides the reliability assessment results using PTI, BTI, RTI, and ERBTI measures. It can be seen that the conventional PTI, BTI, and RTI measures give inconsistent indications of reliability performances with the real ones

identified in Figure 6. For example, the BTI measure indicates that Group D (0.125) has a better reliability than Group C (0.139), but oppositely, the latter group (blue line) should have a better reliability since its distribution is more compact than that of the former one (dark green line). In addition, the RT measure gives equally the same reliability time values (4.4 minutes) estimations for different groups travel time samples. These indicate the limitations of conventional PTI, BTI, and RTI measures in reliability assessment under such situations when multimode service states exist.

The proposed ERBTI measure can give a consistent reliability assessment with the reliability order identified using ERBTI being the same as that identified from Figure 6. Furthermore, the ERBTI measure can provide a significant identification of reliability differences for different groups. For example, it can be observed that Groups C (blue line) and E (red line) have a much different reliability performance from Figure 6. The ERBTI values for such two groups are 0.083 and 0.535, respectively, which clearly indicates a much different reliability performance between the two groups.

5.3. Potential Applications. In public transport, different stakeholders have different requirements. Operators are responsible for providing a reliable service to the public. They are concerned with reliability assessment to gain a deep insight into casual relationships between service inputs (service strategies) and outputs (reliability performance). Passengers are the recipient of bus services. They are concerned with deciding departure time to avoid late arrivals at their destinations [40]. Potential applications for fulfilling different stakeholders' requirements are analysed.

5.3.1. Strategy Assessment (Operators). Diab and El-Geneidy [41] studied the impacts of various improvement strategies on service reliability and concluded that exclusive bus way strategy could decrease the standard deviation of travel time. Assuming the current service travel time distribution follows a 3-component GMM model (GMM3), the parameters for the GMM3 model are set as $\mu = [25, 30, 45]$, $p = [0.1, 0.8, 0.1]$, and $\sigma = [1, \sigma_2, 15]$. Different values of σ_2 indicate reliability performance changes after strategies are applied. Further assuming the exclusive bus way strategy could decrease the standard deviation σ_2 from 6 minutes to 5 minutes and then to 4 minutes, different reliability measures are applied to indicate such improvement.

From Table 5, it can be seen that the proposed ERBTI measure can accurately reflect the service changes after applying a strategy, while the conventional reliability measures values counterintuitively stay unchanged. Such phenomenon could be caused by the fact that the conventional measures are largely impacted by travel times under a nonrecurrent state. If a nonrecurrent state occupies more than 5% of the entire travel time profile, the 95th percentile travel time will remain constant no matter what improvements are made in other states. Furthermore, by considering service reliability under different states separately, different contributions of causes of service reliability can be distinguished based on

which efficient strategies can be made to improve service performance.

5.3.2. Trip Planning (Passengers). As passengers are more concerned with travel time reliability than travel time itself in mode choice and departure planning, a new trip planner design is presented to convey such information to passengers as shown in Table 6. The new trip planner provides passengers with a trip summary and different departure options. Under a specific departure option, the trip travel information is presented using two different sections, namely, SCHEDULED and EXPECTED. The SCHEDULED section is a brief summary of the scheduled travel information for a trip published by operators, including scheduled departure time, scheduled arrival time, and scheduled total time. Usually, the scheduled timetable for the duration of a trip is not necessarily equal to the actual operational travel time. The EXPECTED section displays the information of actual travel time and travel time reliability of a trip, including the expected arrival time, the latest arrival time, and total expected travel time and total latest travel time based on service reliability.

The information shown in Table 6 is given as an example. The total expected travel time and total latest travel time in the EXPECTED section are calculated using (8) and (9). Three departure options are provided for different risk-aversion passengers for different trip purposes. For a passenger who needs high reliability, he/she might choose OP1, since the expected arrival and latest arrival time both occur before 8:45 am. For a passenger who has less need for reliability, he/she might choose OP1 or OP2 since the expected arrival times are before 8:45 am and the latest arrival time is within a tolerable time range.

6. Conclusions

The concern with the impacts of reliability on operating efficiency for operators as well as service effectiveness for passengers brings about the need to identify and develop meaningful and consistent measures of reliability in public transport. Buffer time measures are believed to be appropriate to quantify reliability experienced by passengers in the context of departure planning using operational data. Conceptually, the extreme-value based buffer time measures can separate the unreliability impact on incremental operations from the impact on passenger planning. Two issues related to buffer time estimation under multimode service states are addressed in this paper, namely, performance disaggregation and capturing passengers' perspectives on reliability.

A GMM model based approach is applied to disaggregate the performance data which provides a great flexibility and precision in modelling the underlying characteristics of travel time. Based on the mixture models distribution, a buffer time concept based RBT measure is proposed to approximate passengers' experienced reliability by considering passengers' different perspectives on reliability under different operational service states. A set of ERBT measures is developed for operators by using different spatial-temporal levels combinations of RBTs. Average trip duration and the

TABLE 3: Parameters for different groups travel time samples.

Groups	Mu		Sigma		Proportion		Mean time*	Median time*	Planning time*
	μ_1	μ_2	σ_1	σ_2	p_1	p_2			
Empirical	26.1	29.6	0.16	6.40	0.21	0.79	28.9	28.8	33.2
Group A	26.1	29.6	1.80	5.20	0.35	0.65	28.4	28.2	32.9
Group B	26.1	29.6	7.20	7.40	0.15	0.85	29.1	29.2	33.9
Group C	26.1	29.6	0.10	2.00	0.70	0.30	27.2	26.3	31.0
Group D	26.1	29.6	0.10	2.00	0.50	0.50	27.9	26.8	31.4
Group E	26.1	29.6	0.10	8.00	0.01	0.99	29.6	29.6	34.3

*Time unit is minutes.

TABLE 4: Assessment results for different groups travel time samples.

Groups	PTI (PT*)	BTI (BT*)	RTI (RT*)	ERBTI (ERBT)
Empirical	1.148 (33.2)	0.148 (4.30)	0.153 (4.40)	0.373 (10.7)
Group A	1.158 (32.9)	0.158 (4.50)	0.156 (4.40)	0.291 (8.20)
Group B	1.164 (33.9)	0.165 (4.80)	0.151 (4.40)	0.500 (14.6)
Group C	1.139 (31.0)	0.139 (3.80)	0.165 (4.40)	0.083 (2.19)
Group D	1.125 (31.4)	0.125 (3.50)	0.164 (4.40)	0.127 (3.40)
Group E	1.159 (34.3)	0.159 (4.70)	0.149 (4.40)	0.535 (15.8)

*PT: planning time; BT: buffer time; RT: reliability time.

TABLE 5: Assessment of service performance changes using different reliability measures.

σ_2	PTI (PT*)	BTI (BT*)	RTI (RT*)	ERBTI (ERBT)
4	1.520 (45.0)	0.452 (14.0)	0.520 (15.4)	0.304 (9.0)
5	1.520 (45.0)	0.452 (14.0)	0.520 (15.4)	0.351 (10.4)
6	1.520 (45.0)	0.452 (14.0)	0.520 (15.4)	0.399 (11.8)

*PT: planning time; BT: buffer time; RT: reliability time.

TABLE 6: New designed trip planner for passengers.

Trip planner						
Arrival before: 8:45 am on Friday 5 April 2013						
From: origination						
To: destination						
Requirements: bus, train, or ferry						
Options	Scheduled			Expected		
	Depart	Arrival	Total	Arrival	Latest	Total
OP1: XXX	8:10	8:36	00:26	8:39	8:41	00:29 (00:31)
OP2: XXX	8:15	8:41	00:26	8:44	8:46	00:29 (00:31)
OP3: XXX	8:20	8:46	00:26	8:49	8:51	00:29 (00:31)

latest trip duration measures are proposed for passengers that could be used to make a service mode choice and determine the departure time for a trip. Case studies verify the existence of multimode service states during a given time period. The proposed ERBT measures can provide consistent reliability assessment with a high-level detail, while the conventional reliability measures may give inconsistent assessment results. In addition, by considering different passengers' experienced

reliability under different service states, different contributions of causes can be evaluated based on which effective and efficient improvement strategies can be made. A new trip planner design is presented to convey reliability information to passengers. Different options for a trip are also provided in the trip planner based on which a passenger can easily make service a choice and plan a departure time according to different requirements.

As the proposed measure is based on the mixture distribution models, the critical point is to ensure that the statistical results can be explained with reference to physical reality. In addition, the mixture models using skewed component distribution instead of a symmetric Gaussian distribution should be investigated further.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to acknowledge the TransLink Division of Queensland DTMR (Department of Transport and Main Roads) for assistance with data collection. This work is supported by China Scholarship Council.

References

- [1] T. J. Kimpel, *Time point-level analysis of transit service reliability and passenger demand [Doctor of Philosophy, Urban Studies]*, Portland State University, Portland, Ore, USA, 2001.

- [2] R. Balcombe, R. Mackett, N. Paulley et al., "The demand for public transport: a practical guide," TRL Report 593, 2004.
- [3] M. Abkowitz, H. Slavin, R. Waksman, L. English, and N. H. M. Wilson, *Transit Service Reliability*, U.S. Department of Transportation, Cambridge, UK, 1978.
- [4] M. Mesbah, G. Currie, C. Lennon, and T. Northcott, "Spatial and temporal visualization of transit operations performance data at a network level," *Journal of Transport Geography*, vol. 25, pp. 15–26, 2012.
- [5] D. Uniman, J. Attanucci, R. Mishalani, and N. Wilson, "Service reliability measurement using automated fare card data," *Transportation Research Record*, no. 2143, pp. 92–99, 2010.
- [6] P. G. Furth, B. Hemily, T. H. J. Muller, and J. G. Strathman, "Using archived AVL-APC data to improve transit performance and management," 2006.
- [7] Kittelson & Associates, Urbitrans, LKC Consulting Services, MORPACE International, Q. U. o. Technology, and Y. Nakanishi, *A Guidebook for Developing a Transit Performance-Measurement System*, Transport Research Board, Washington, DC, USA, 2003.
- [8] G. Currie, N. J. Douglas, and I. Kearns, "An assessment of alternative bus reliability indicators," in *Proceedings of the Australasian Transport Research Forum Proceedings*, Perth, Australia, 2012.
- [9] E. E. Osuna and G. F. Newell, "Control strategies for an idealized public transportation system," *Transportation Science*, vol. 6, pp. 52–72, 1972.
- [10] D. L. Uniman, *Service reliability measurement framework using smart card data: application to the london underground [Master of Science, Civil and Environmental Engineering]*, Massachusetts Institute of Technology, 2009.
- [11] A. A. Ceder, *Public Transit Planning and Operation: Theory, Modelling and Practice*, Elsevier, London, UK, 2007.
- [12] N. van Oort, *Service reliability and urban public transport design [Doctor of Philosophy]*, Netherlands TRAIL Research School, 2011.
- [13] N. van Oort and R. van Nes, "Service regularity analysis for urban transit network design," in *Proceedings of the Transportation Research Board 83rd Annual Meeting*, 2004.
- [14] A. Yoh, H. Iseki, M. Smart, and B. D. Taylor, "Hate to wait: effects of wait time on public transit travelers' perceptions," *Transportation Research Record*, vol. 2216, pp. 116–124, 2011.
- [15] P. G. Furth and T. H. J. Muller, "Service reliability and hidden waiting time insights from automatic vehicle location data," *Transportation Research Record*, vol. 1955, pp. 79–87, 2006.
- [16] T. Lomax, D. Schrank, S. Turner, and R. Margiotta, *Selecting Travel Reliability Measures*, Texas Transportation Institute Monograph, 2003.
- [17] J. Y. Park, D.-J. Kim, and Y. Lim, "Use of smart card data to define public transit use in Seoul, South Korea," *Transportation Research Record*, vol. 2063, pp. 3–9, 2008.
- [18] W. Jang, "Travel time and transfer analysis using transit smart card data," *Transportation Research Record*, no. 2144, pp. 142–149, 2010.
- [19] M. A. Turnquist and L. A. Bowman, "The effects of network structure on reliability of transit service," *Transportation Research B*, vol. 14, no. 1-2, pp. 79–86, 1980.
- [20] R. M. P. Goverde, *Improving punctuality and transfer reliability by railway timetable optimization [Doctor of Philosophy]*, Delft University of Technology, 1999.
- [21] P. Knoppers and T. Muller, "Optimized transfer opportunities in public transport," *Transportation Science*, vol. 29, no. 1, pp. 101–105, 1995.
- [22] R. C. W. Wong, T. W. Y. Yuen, K. W. Fung, and J. M. Y. Leung, "Optimizing timetable synchronization for rail mass transit," *Transportation Science*, vol. 42, no. 1, pp. 57–69, 2008.
- [23] W. Pu, "Analytic relationships between travel time reliability measures," *Transportation Research Record*, vol. 2254, pp. 122–130, 2011.
- [24] P. G. Furth and T. H. J. Muller, "Service reliability and optimal running time schedules," *Transportation Research Record*, vol. 2034, pp. 55–61, 2007.
- [25] T. Lomax, S. Turner, and R. Margiotta, *Monitoring Urban Roadways in 2000: Using Archived Operations Data For Reliability and Mobility Measurement*, Federal Highway Administration, Washington, DC, USA, 2001.
- [26] F. Guo, H. Rakha, and S. Park, "A multi-state travel time reliability model," in *Proceedings of the Transportation Research Board 89th Annual Meeting*, Washington, DC, USA, 2010.
- [27] T. Barkley, R. Hranac, and K. Petty, "Relating travel time reliability and nonrecurrent congestion with multistate models," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2278, pp. 13–20, 2012.
- [28] S. Susilawati, M. A. P. Taylor, and S. V. C. Somenahalli, "Distributions of travel time variability on urban roads," *Journal of Advanced Transportation*, vol. 47, no. 8, pp. 720–736, 2012.
- [29] D. A. Reynolds, *Gaussian Mixture Models*, 2009.
- [30] H. Wakabayashi and Y. Matsumoto, "Comparative study on travel time reliability indexes for highway users and operators," *Journal of Advanced Transportation*, vol. 46, pp. 318–339, 2012.
- [31] J. Chan, *Rail transit OD matrix estimation and journey time reliability metrics using automated fare data [Master of Science]*, Massachusetts Institute of Technology, 2007.
- [32] NCHRP, "National Cooperative Highway Research Program (NCHRP) Report 599: Default Values for Highway Capacity and Level of Service Analyses," Transportation Research Board, 2008.
- [33] R. K. Pearson, "Outliers in process modeling and identification," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 1, pp. 55–63, 2002.
- [34] L. Kieu, A. Bhaskar, and E. Chung, "Bus and car travel time on urban networks: integrating bluetooth and bus vehicle identification data," in *Proceedings of the 25th ARRB Conference: Shaping the future: Linking Policy, Research and Outcomes*, Perth, Australia, 2012.
- [35] S. Clark and D. Watling, "Modelling network travel time reliability under stochastic demand," *Transportation Research B: Methodological*, vol. 39, no. 2, pp. 119–140, 2005.
- [36] H. C. Chu, "Estimating travel time reliability on freight corridors," in *Proceedings of the Transportation Research Board 89th Annual Meeting*, Washington, DC, USA, 2010.
- [37] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, pp. 716–723, 1974.
- [38] N. E. E. Faouzi and M. Maurin, "Reliability of travel time under lognormal distribution," in *Proceedings of the Transport Research Board 86th Annual Meeting*, Washington, DC, USA, 2007.
- [39] J. A. Hartigan and P. M. Hartigan, "The dip test of unimodality," *The Annals of Statistics*, vol. 13, no. 1, pp. 70–84, 1985.

- [40] B. Kuhn, L. Higgins, A. Nelson et al., “Effectiveness of Different Approaches to Disseminating Traveler Information on Travel Time Reliability,” 2013.
- [41] E. Diab and A. El-Geneidy, “Variation in bus transit service: understanding the impacts of various improvement strategies on transit service reliability,” *Public Transport*, vol. 4, pp. 209–231, 2013.