Measurement and Classification of Transit Delays Using GTFS-RT Data

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at:
http://dx.doi.org/10.1007/s12469-022-00291-7

Zack Aemmer
Department of Civil and Environmental Engineering
University of Washington, Seattle, WA, 98195
Email: zae5op@uw.edu

Andisheh Ranjbari
Department of Civil and Environmental Engineering
Penn State University, University Park, PA 16802
Email: ranjbari@psu.edu

Don MacKenzie
Department of Civil and Environmental Engineering
University of Washington, Seattle, WA, 98195
Email: dwhm@uw.edu

Abstract:

This paper presents a method for extracting transit performance metrics from a General Transit Feed Specification's Real-Time (GTFS-RT) component and aggregating them to roadway segments. A framework is then used to analyze this data in terms of consistent, predictable delays (systematic delays) and random variation on a segment-by-segment basis (stochastic delays). All methods and datasets used are generalizable to transit systems which report vehicle locations in terms of GTFS-RT parameters. This provides a network-wide screening tool that can be used to determine locations where reactive treatments (e.g., schedule padding) or proactive infrastructural changes (e.g., bus-only lanes, transit signal priority) may be effective at improving efficiency and reliability. To demonstrate this framework, a case study is performed on one year of GTFS-RT data scraped from the King County Metro bus network in Seattle, Washington. Stochastic and systematic delays were calculated and assigned to segments in the network, providing insight to spatial trends in reliability and efficiency. Findings for the study network suggest that high-pace segments create opportunity for large, stochastic speedups, while the network as a whole may carry excessive schedule padding. In addition to the static analysis discussed in this paper, an online interactive visualization tool was developed to display ongoing performance measures in the case study region. All code is open-source to encourage additional generalizable work on the GTFS-RT standard.

Keywords: Public transit, GTFS Real Time, Performance

Acknowledgements: This study was supported by Amazon, Uber, King County Metro, Sound Transit, the Seattle Department of Transportation, Challenge Seattle, and the Mobility Innovation Center at University of Washington CoMotion.

Data Availability Statement: The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

1. INTRODUCTION

The flexibility afforded by sharing existing roadway infrastructure has allowed bus transit systems to flourish across a spectrum of transportation settings and cultures. However, one inherent disadvantage to sharing existing infrastructure is that buses are susceptible to delays caused by congestion, roadway controls, and shared curb use in addition to regular passenger boarding and alighting; all of which can be highly unpredictable in nature. If these delays are left unchecked, a bus system may become unreliable, and the utility it provides to its riders will diminish rapidly. This might cause ridership to drop, bringing down fare recovery, and eventually leading to service cuts compounding these issues.

To avoid this, transit agencies and planners account for delays when scheduling buses, or simply fix their root causes. When determining where and how to target such treatments, agencies and planners must often make broad strokes to account for cumulative delays across entire routes. One of the primary tools used is schedule padding, which adds buffer time to transit segments that are consistent sources of delay. Schedule padding can improve travel time reliability by accounting for delays but can also hinder it in locations where travel time is highly variable, and it does not fundamentally improve the efficiency of the transit system. Furthermore, the iterative process of assigning buffer time to transit segments is informed using automatic vehicle location (AVL) systems, for which analytical tools must be built on a system-by-system basis.

However, the General Transit Feed Specification (GTFS) has provided a fully generalizable framework by which agencies can collect and share transit scheduling data. More importantly, it has led to the development and initial proliferation of the GTFS-RT (real-time) standard by which *actual* bus locations and stop times can be shared in a standardized format, once collected by various onboard systems (Antrim and Barbeau, 2017). To date, analytical tools and support systems are relatively scarce for GTFS-RT. This may be because there is little incentive to create them, as agencies collecting this data likely have AVL-based performance measurement tools in place. Although powerful, each AVL-based tool must be individually maintained by its respective agency, and the lack of standardization prevents clear comparison between transit systems.

In this paper, we examine how the GTFS-RT standard may be used to quantify transit performance in a way fully generalizable to any bus transit network. We group bus delays into two categories: stochastic and systematic, which are quantified through several metrics. These metrics are then calculated at the segment-level across all routes and vehicles in the entire transit network. This enables prioritization for specific locations where infrastructural treatments (e.g., dedicated bus lanes, transit signal priority) or operational treatments (e.g., schedule padding) will be most effective in improving transit performance. Because these measures are quantified from aggregated data across all buses utilizing a given segment, treatments can be identified where they stand to create the largest benefits across all vehicles, rather than individual routes.

Following this introduction, we discuss prior work conducted in this area. The data collection and analysis are presented in the Methods section, where we propose a framework to identify individual instances of delay and examine temporal and spatial trends in delay from the GTFS-RT data. In the Case Study section, we introduce the data sources used for an exploratory analysis of transit delays in the GTFS-RT system for King County Metro (KCM) in Seattle, Washington. This is followed by Results and Discussion on what locations exhibit the most delay, how delay varies spatially, and how it relates to segment pace. Last, the Conclusion section summarizes the findings, states the limitations of the study, and presents suggestions for future research.

2. LITERATURE REVIEW

2.1 Transit Performance and Delays

Transit performance is somewhat of an amorphous concept. Many works have aggregated, surveyed, and proposed classification hierarchies for a multitude of established metrics (Danaher et al., 2020; Gleason and Barnum, 1982). One of its most commonly emphasized components is travel time reliability; quantified through specific metrics such as expected waiting time, variance of travel time, or on-time performance (Danaher et al., 2020). In other words, metrics of reliability measure how consistent the system is. One way reliability is frequently reported is through schedule adherence, which relies on measures of individual delays or the amount of time a vehicle has deviated from schedule in a given time period (Strathman et al., 1999). As delays accumulate, buses deviate further from their schedule. Reliability is one of the most important factors in the utility provided by transit systems; shown to be valued higher than comfort (Nurul Habib et al., 2011) or frequency (Bowman and Turnquist, 1981; de Oña et al., 2012). Work has explicitly tested the user cost of stop arrival time variability (Bowman and Turnquist, 1981) and found it highly impactful. Other work has documented how the utility of reliability extends to all modes of transportation, not only transit (Bates et al., 2001). Thus, if bus transit systems are to compete with other modes their reliability must be at least comparable to cars, trains, or other modes.

One of the fundamental treatments for unreliability in transit service is “schedule padding”. This is an iterative process of schedule design which allocates additional scheduled travel time to roadway segments tending towards highly variable travel times (Ceder, 2002). While this does not eliminate the delay, it does improve the reliability of the transit service by allowing extra time for when delays occur, and the vehicle falls behind schedule. Unfortunately, when the expected delays do not occur, this can lead to buses arriving at stops ahead of time, which also reduces the overall reliability of the system (Furth et al., 2006). In many cases, drivers are encouraged to slow down when ahead of schedule, either by a central dispatcher or onboard clock, thus maintaining equal headways between vehicles (Lin and Bertini, 2004). Schedule padding is a relatively simple and inexpensive treatment to apply on a system-wide scale but does not address the actual cause of delays.

Another component of transit performance is transit efficiency, or productivity (Gleason and Barnum, 1982). This component plays a smaller role in utility provided to individuals using the transit system but has ramifications for cost of operation and net emissions network-wide. Metrics of transit efficiency are expressed as ratios, and might include farebox recovery ratio (operating cost divided by revenue) (Danaher et al., 2020), or volume to capacity ratio (bus flow divided by lane capacity) (Godavarthi et al., 2014). Example treatments to improve efficiency might include addition of transit signal priority, queue-jump lanes, or route changes, among others (Danaher et al., 2020). Schedule padding on its own is not sufficient to improve efficiency, as it simply adjusts bus schedules, without necessarily improving their movement. In fact, if excessive schedule padding is used and a bus slows down to remain on-time, there may be a reduction in efficiency-related metrics for that trip.

The term stochastic delay has been utilized in a transportation context by a handful of previous works. Most recently, it appears in work using GTFS data to determine locations where there is high schedule padding in a network (Wessel and Widener, 2017). They used stochastic delay to describe all deviations from the GTFS schedule and suggest roadway treatments. In train networks, stochastic *propagation* of delay was used in a method for modeling how an single delaying event can progress through an entire system (Berger et al., 2011). In other work, the quantity of schedule deviation due to individual, delay-causing incidents was estimated through a model using a combination of deterministic and stochastic delay elements (Fu and Rilett, 1997). In general, stochastic delays are characterized as occurring due to "non-recurring" traffic conditions, which are thus difficult to predict (Sheu et al., 2004). These occurrences rely on probabilistic methods to predict accurately (He et al., 2014), and real-time monitoring systems to accommodate (Fu and Rilett, 1997; Yu and Yang, 2009).

2.2 Transit Data

A burgeoning wealth of transit data sources such as automated fare collection (AFC), automated passenger count (APC), AVL data, smart cards, phones, and social media websites are documented in recent literature (Ge et al., 2021; Li et al., 2018; Lu et al., 2021). In particular, AFC, APC, and AVL data provide high resolution, targeted insights for measuring transit performance. For example, recent work has outlined use cases for AVL data as a measure of on-time performance outcomes during adverse weather conditions (Mesbah et al., 2015), and AFC/smart card data as input data for modeling passenger waiting time at rail stops (Tavassoli et al., 2018). Other work has found that APC data may be more effective as a measure of demand response to reliability (Bills and Carrel, 2021), or a measure of the relative impact of delays for routes with varying ridership (Arias et al., 2021). Together, these data provide a fairly complete picture of transit performance. However, other emerging data sources such as cellular and social media data can provide unique insights, though they require separate collection efforts and processing. Several works have distinguished which problems these datasets may support for the collecting agency, that are not better served by typical automated data (Diaz et al., 2021; Jevinger and Persson, 2019). For example, using social media posts as a measure of communication between transit agencies and the public, or using cellular routing information to understand passenger destinations. Last, although many automated collection methods have replaced the need for traditional surveys, there may still be value in using surveys as a complement to these "big data" sources (Ge et al., 2021). In particular, census data is essential for observing socio-demographic trends. In the remainder of this section, we discuss solutions to performance measurement built on AVL data, which is most similar to GTFS-RT data in regards to its structure. We then review and discuss literature on the benefits of generalization for AVL data through the GTFS-RT standard.

2.3 Automated Data Collection in Bus Performance Management

To determine performance metrics, and ultimately inform system treatments, data must be collected at stops or corridors of interest. This can be expensive, time consuming and provides only a narrow window of perspective into the nature of delays. AVL technology offers an automated means to record and archive individual bus movements and can thus help illuminate of the locations of delays for each vehicle in the system. They may also communicate with a central dispatch to provide real-time updates to a controlling agency for oversight and performance monitoring purposes.

In prior work, AVL data has been used to quantify bus reliability and efficiency at the individual bus level (Bertini and El-Geneidy, 2003; Cham, 2006). Aggregation then allows for metrics to be calculated at the stop, corridor, route, or system levels (Bertini and El-Geneidy, 2003; Duddu et al., 2019). Stop-to-stop and route level analysis in particular has been supported by AVL data, with performance measures calculated from schedule adherence being used to determine reliability throughout such corridors in the system (El-Geneidy et al., 2011). Additionally, AVL data has been used to analyze secondary interactions which are not directly reported by the system such as bus bunching (Feng and Figliozzi, 2012a), travel time for specific roadway segments (Berkow et al., 2007), and even in aggregate as probes to measure the performance of *all* roadway users, not just buses (Bertini and Tantiyanugulchai, 2004). Among these and other works, some classifications for delays based on their cause, or location reported in AVL data have been proposed. For example, classifying the cause of delays based on their coordinates and surrounding infrastructure (Coghlan et al., 2019) or whether they occurred due to a scheduled stop (Pi et al., 2018).

Recent work has also emphasized the use of archived AVL data to establish reliability and classify causes of unreliability for transit segments or routes in a network. For example, one study developed a "time budget" for each trip in a network broken up by dwell time, free-flow time, etc. (Lind and Reid, 2021). Speed was calculated from GPS traces and tracked locations were classified depending on additional characteristics such as whether it was less than a certain threshold, or close to a stop. Aggregation across routes determined which locations in the network cost certain portions of the overall time budget. A general set of performance metrics specifically relating to bus service reliability, which can be informed by AVL data, is also documented (Albadvi et al., 2018). Another work examined waiting times along a chosen BRT line and constructed a model to estimate waiting time based on average vehicle delay informed by AVL data (Webb et al., 2020). AVL data has also been used in combination with AFC data as a measure of reliability for intermodal trips crossing metro, tram, and bus transit (Dixit et al., 2019). Similarly AVL, GTFS, and AFC-based methods of calculating travel time reliability were compared in another study which considered walking and transit modes (Li et al., 2021).

Although powerful in analyzing performance, AVL and APC data is not often made publicly available, lacks a standardized format, and leads to individual "home-grown" analysis tools that are not reusable by other agencies or researchers (Furth et al., 2006). Previous academic works have developed tools that suggest automatically generated performance metrics (Mandelzys and Hellinga, 2010), or visualize transit performance through an analysis interface (Currie and Mesbah, 2011; Feng and Figliozzi, 2012b). While effective and useful, these precipitate a burden on individual agencies to maintain supporting software and databases for each technology and its respective analysis tools. It also creates opportunities for discrepancies in the calculation methods for metrics between systems.

2.4 GTFS and GTFS-RT Standards

Public-facing GTFS-RT feeds provide a standardized format for repackaging AVL data, from which schedule padding and other system characteristics can be reliably ascertained (Ge et al., 2021; Wessel and Widener, 2017). Many agencies currently provide this data through application programming interfaces

(APIs), from which users may request and receive unified results. Given that the GTFS standard has been adopted by more than 1,000 transit agencies globally, the GTFS-RT standard may be poised to follow. Using this specification, generalizable tools for analysis and visualization of transit data can be applied to draw comparisons between networks with ease.

The proliferation of the GTFS format has supported generalizable analysis and comparison of scheduled coverage and service quality between networks. For example, one team measured gaps in transit supply and demand using GTFS schedules and Canadian census data (Kaeoruean et al., 2020), and another estimated the travel time savings potential for transit priority treatments (Arias et al., 2021). However, as pointed out in several studies, actual service may deviate significantly from scheduled service. This is where AVL, GTFS-RT (which is essentially repackaged AVL), smart card, and other observed data sources must be used to determine actual transit performance, and several studies have been performed which compare or combine the observed results of these data against GTFS schedules (Kumar et al., 2018; Li et al., 2021). One study explored the relationship between scheduled and actual data through a comparison of travel times along shortest paths between origin-destination pairs in a network, finding that the actual times were right-skewed and overall higher than those in the schedule (Li et al., 2021). Another observed service coverage informed by GTFS feeds in relation to the relative demand of different destinations in Calgary, Canada across both spatial and temporal boundaries (Kaeoruean et al., 2020). However, due to its closed-source nature, AVL data do not have the same benefits of generalizability that GTFS-RT data do. This is where GTFS-RT creates opportunity for unified analysis of actual bus performance in the same way that GTFS unifies scheduled analysis of bus performance.

There is currently much less literature on calculating performance metrics from GTFS-RT data than AVL/APC data. This is likely due to the greater establishment of AVL/APC data, and its offering of additional variables and finer temporal detail. In prior work on GTFS-RT, projects have developed simple real-time observation tools to display immediate bus locations to an end-user such as in one academic work (Yue et al., 2017). Some studies have gone a step further and applied real-time prediction methods to determine delays in active systems and used that information to provide better scheduling information to travelers: This was found to lead to lower overall wait times for travelers, and a more positive perception of transit reliability (Watkins et al., 2011). Sophisticated methods for archiving GTFS-RT data have also been developed, which is the first step for unified analysis (Barbeau et al., 2020). Preliminary work has developed and tested a framework for analyzing bus corridors using archived GTFS-RT data (Caros et al., 2021). Their work first developed a method for combining routes into consistent stop-to-stop corridors, then calculated aggregate performance metrics for each corridor.

2.5 Challenges for GTFS-RT Development

Overall, GTFS-RT is in a nascent state of adoption, with few open-source analysis tools available for performance analysis. One of the largest challenges currently facing GTFS-RT data is quality control (Steiner et al., 2015). Many GTFS-RT feeds are not aligned with the standard, and thus fully generalizable tools for differing feeds require additional consideration. As more agencies make bus position data available through GTFS-RT, and additional tools are constructed on the standard, there will be more incentive to fix the integrity of these feeds to benefit from open-source work. There is currently ongoing work to develop tools specifically for validating GTFS-RT feeds (Barbeau, 2018).

This paper contributes to the state of the art through a case study of bus transit data collected from a web API-based GTFS-RT system and disaggregate analysis of segment-level reliability and efficiency metrics informed by that data. We build on prior work classifying stochastic delays with the additional distinction of systematic delays, accruing from consistent, predictable reductions in transit free-flow speed due to congestion, signals, and transit stops. This allows for analysis of the transit network on a segment-by-segment basis, and classification of performance losses between inefficiency and unreliability. Combined, these can be useful for informing transit priority treatments or schedule adjustments in any network.

3. METHODS

This section details the process by which GTFS-RT data was collected into a PostGIS database and analyzed using a combination of built-in functions and QGIS software for visualization, following the general steps outlined in **Fig. 1**.

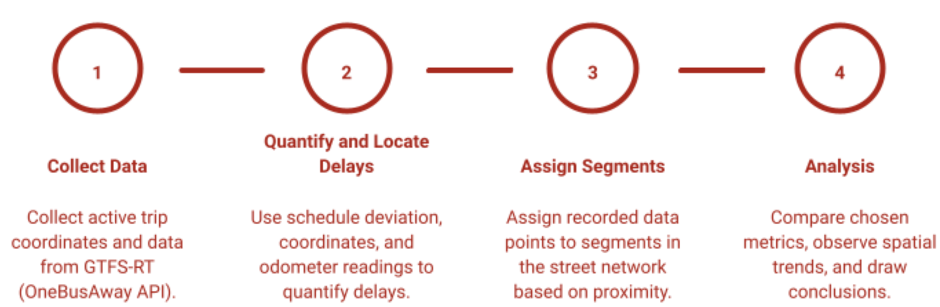

Fig. 1 Overview of the steps used to measure and classify transit performance

3.1 Collect Data

The data source for this study was a programming API providing access to underlying GTFS-RT data for the study region (see Case Study section for additional details). As a developing standard, there is ongoing debate as to which parameters should be required in GTFS-RT, and which should be optional (Barbeau, 2018). The list below documents each parameter used in this analysis framework using the nomenclature of the API from which they were collected. At a minimum, these are the required parameters to recreate this analysis:

activeTripId:

Unique identifier for each vehicle, traversing a single route, on a single day (i.e., a single scheduled run of a route).

vehicleId:

Unique identifier for each vehicle, consistent across days and routes. It is physically painted on each vehicle.

scheduleDeviation:

Difference in seconds between a trip's known position, and where it is scheduled to be as calculated by the AVL system. Can be positive (behind schedule) or negative (ahead of schedule).

lastKnownDistanceAlongTrip:

The distance that the current trip has traveled since its inception. Resets to zero for the next trip.

lastKnownLocation:

The last updated position for a given trip.

lastLocationUpdateTime:

The timestamp corresponding to when a trip's last known location was recorded.

Queries were made at the highest resolution feasible; different buses may update the API at different frequencies, and the system itself may update at its own frequency. The higher the frequency that the data is collected, the more precisely delays can be quantified and located. This comes at heavier computational costs during the analysis and generates more calls to the API. Additionally, it becomes more computationally challenging to assign segment delays to their respective road segments as the size of the location and roadway segment datasets increase. Although the archiving framework used here has suited the needs of this project, there is room for improvement in efficiency and modularization, as detailed by recent work in the field of GTFS-RT "big-data" management (Barbeau et al., 2020).

3.2 Quantify and Locate Delays

In this paper, we present several metrics calculated from onboard vehicle data feeds provided in the GTFS-RT format, which fall separately under reliability and efficiency metrics. These are based on time-series data and are calculated between each vehicle update interval. Thus, the unit of analysis used is an individual instance of "delay" or step in time in which a bus performs better or worse relative to either its schedule, or its free flow speed. Metrics were chosen to capture both reliability and efficiency components of the system and are classified for treatment analysis as either stochastic or systematic delays. We assume that the current route schedules are appropriate for the expected quantity of delay on each given segment, and thus any schedule deviation is the result of unpredictable variation, i.e., stochastic delay.

Metrics for quantifying delay between consecutive bus locations *i-1* and *i* were calculated according to **Equations 1-4**:

Vehicle pace at time *i* is calculated using the *lastKnownDistanceAlongTrip* (TD) and *lastLocationUpdateTime* (LT) parameters as indicated in **Equation 1**. Total delay (combined systematic and stochastic delay) between location *i* and *i-1* is then calculated as a product of the difference between the measured pace and the 95th percentile pace for the segment, and the distance traveled between those two locations (**Equation 2**).

$$\frac{LT_i - LT_{i-1}}{TD_i - TD_{i-1}} = Pace_{\ i}\ (sec/mi) \qquad (1)$$

$$(Pace_{95} - Pace_i) * (TD_i - TD_{i-1}) = Total\ Delay_i\ (sec) \qquad (2)$$

Stochastic delay (**Equation 3**) is estimated in terms of the *scheduleDeviation* (SD) GTFS-RT parameter. In cases where delays occur due to planned or predictable changes to vehicle scheduling, such as road work or system maintenance, scheduled travel times will account for these delays (so long as these delays are accounted for at the scheduling stage). Therefore, changes in schedule deviation indicate an instance of unexpected delay such as a collision, blocked stop pull-in, delayed re-entry, or an unusually high number of boardings. Schedule deviation is available for all buses broadcasting real-time arrival information and is continuously calculated as a cumulative measure of the delays and speedups experienced by a vehicle during a given trip. Any decrease in schedule deviation is treated as a negative delay, i.e., the bus sped up due to lack of congestion or other factors. Any increase in schedule deviation is treated as a positive delay, i.e., the bus slowed down due to delaying factors (**Fig. 2**).

$$SD_i - SD_{i-1} = Stochastic\ Delay_i\ (sec) \qquad (3)$$

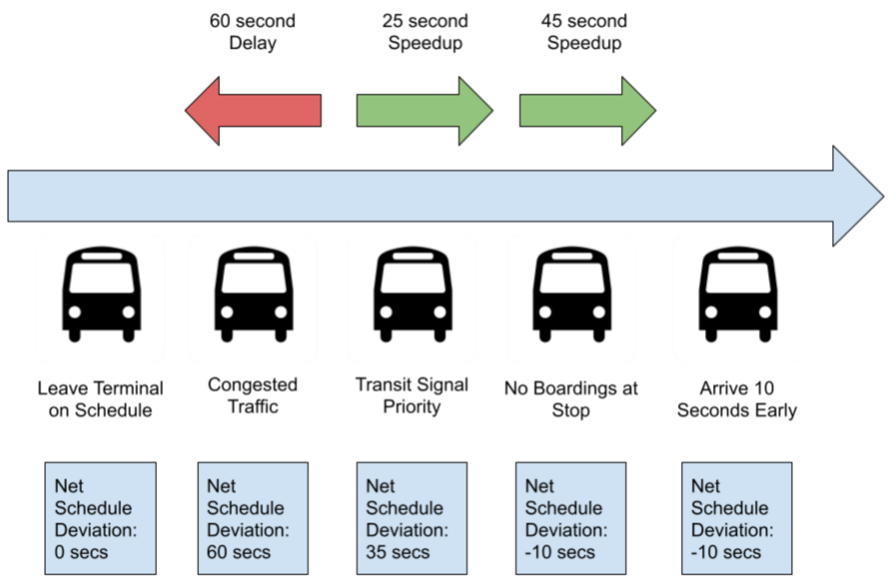

Fig. 2 The process of determining individual instances of delay from the cumulative schedule deviation parameter

Systematic delay (**Equation 4**) is calculated from previously estimated stochastic and total delays. Systematic delay measures the predictable delays that accrue due to congestion and other interference.

$$Systematic\ Delay_i = Total\ Delay_i - Stochastic\ Delay_i\ (sec) \qquad (4)$$

In summary, systematic delays account for scheduled delays accommodated in the timetable. Stochastic delays account for unscheduled delays which have caused a vehicle to deviate from its schedule. This distinction is achieved through the use of two different approaches to calculating bus delay from the GTFS-RT parameters available.

3.3 Assign Segments

After the data is processed to determine all instances of delay during the study period, and each delay has been assigned a location based on the bus location of its second timestamp, each delay location is then assigned to its closest roadway segment. Contemporary practice is to use a map-matching algorithm to assign locations to the nearest geographical feature (Quddus et al., 2007). In our study, delay locations are matched to street segments based on their *lastKnownLocation* and spatially indexed coordinates of each roadway segment using nearest neighbors. The atomic units of aggregation used for calculating performance metrics from each observed bus coordinate in this method are street segments of approximately one block length.

4. CASE STUDY

4.1 Case Study Network

To demonstrate the collection and analysis process of reliability and efficiency performance metrics for GTFS-RT data, a case study was performed. Data was collected from the KCM bus network in Seattle, Washington, over 12 months from August 2020 – August 2021 (**Fig. 3**). This section details the data sources used for this purpose.

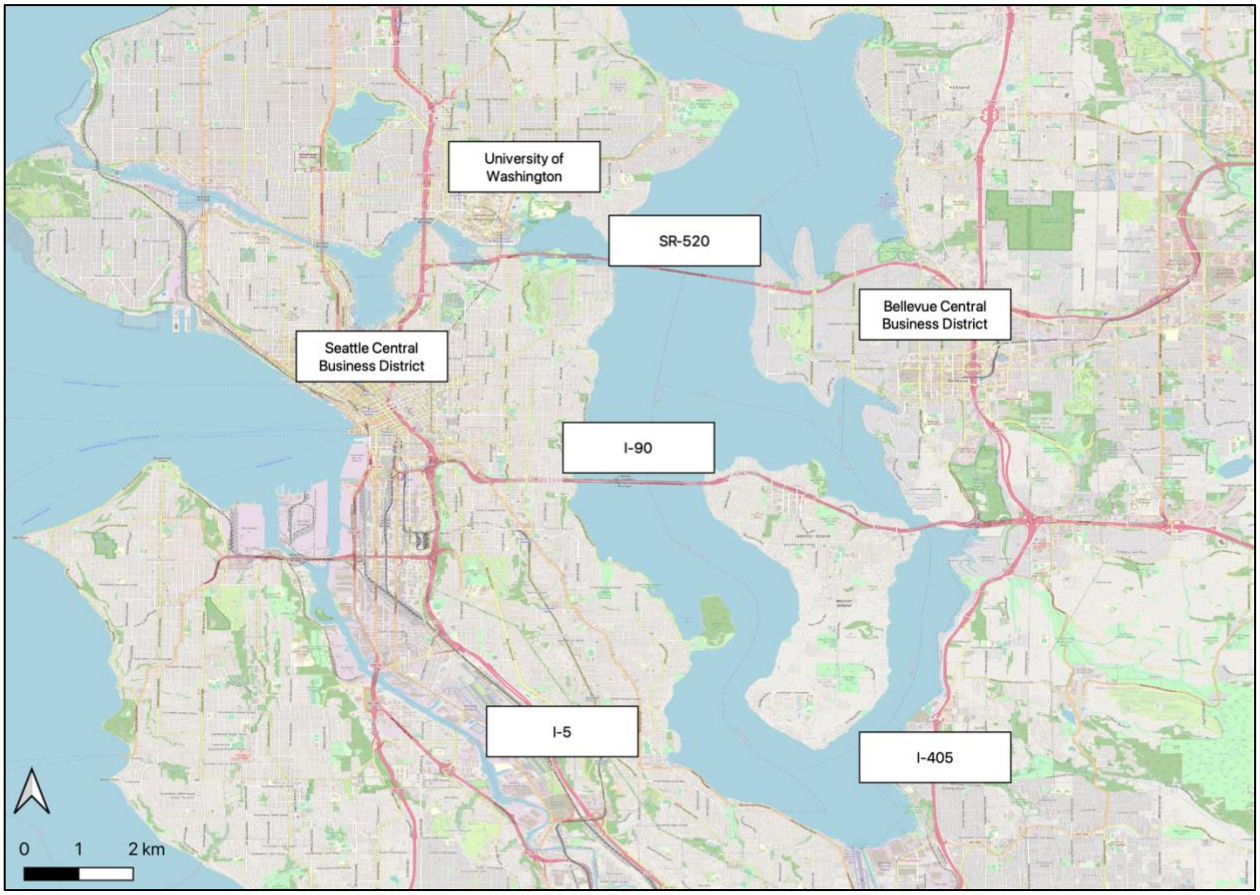

Fig. 3 Map of case study region with major highways and central business districts labeled

4.2 Data Sources

Several data sources were utilized to record and analyze system-wide bus delays in the KCM network. Bus delays, locations, and other GTFS-RT elements were drawn from the OneBusAway API, which provides access to data for all major transit agencies in the Puget Sound region, and is managed by Sound Transit (Ferris et al., 2010). The API provides various function calls to access this data, although some underlying variables such as AVL system location accuracy and refresh rate are not documented and can be dependent on the agency or vehicle in question.

Analysis is performed at the segment-level across the KCM network, necessitating separate shapefiles for street segments and active bus routes. The street segments are drawn from the American Community Survey (ACS) TIGER shapefiles (US Census Bureau, 2021). This dataset includes all street segments regardless of whether they are utilized by the transit network and is available for all roadways in the United States. To improve computational efficiency in assigning GPS coordinates for active routes to the closest street segment, these segments are filtered to only include those traversed by buses in the KCM network. This is accomplished using a second shapefile; the bus route shapefile provided by the KCM GTFS data. These routes are buffered, and street segments contained within them are kept in the analysis. This reduces the overall number of potential segments to search when assigning bus coordinates to the nearest roadway and ensures that bus coordinates are only assigned to segments traversed by their respective route. Segments were first determined based on intersecting roadways, then long segments were broken up if longer than a specified length.

Data was collected on both weekdays and weekends. API queries were made starting at 6AM and ending at 7PM. To gather data for all active transit routes, the OneBusAway API was queried once every ten seconds using a single API call to the "vehicles-for-agency" API endpoint, which provides a list of vehicle IDs, and information about the trip status for each vehicle in the response. To obtain information on only the active vehicles, trip statuses listed as canceled, or trips with a null identifier were removed. Additionally, any active trips without posted GPS coordinates were removed, on the assumption that these vehicles were not equipped with real-time tracking technology. These responses were then timestamped and inserted into a relational database.

5. RESULTS AND DISCUSSION

5.1 Summary

After cleaning, the final dataset consisted of approximately 27,000,000 tracked bus locations belonging to 6,000 unique transit trips. Median stochastic delay was -19 seconds, median systematic delay was 95 seconds, and median total delay was 72 seconds. **Figs. 4-6** show the distribution of a sample of each type of delay across the dataset.

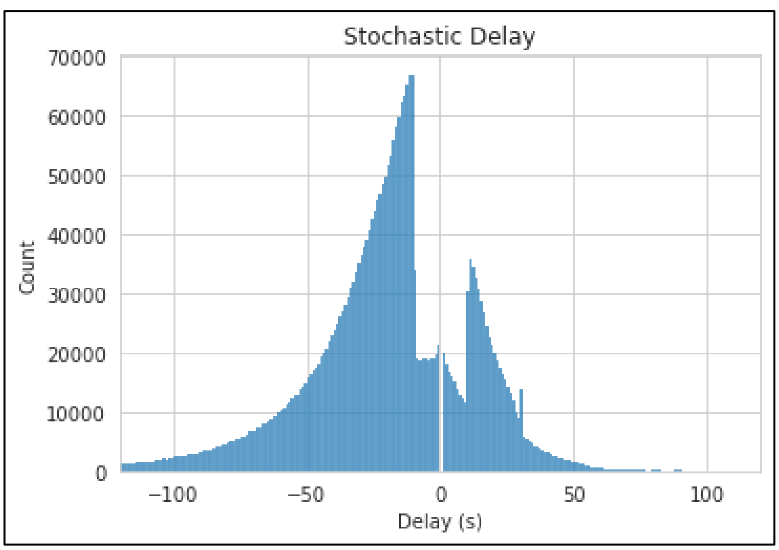

Fig. 4 Distribution of stochastic delays in the bus locations dataset

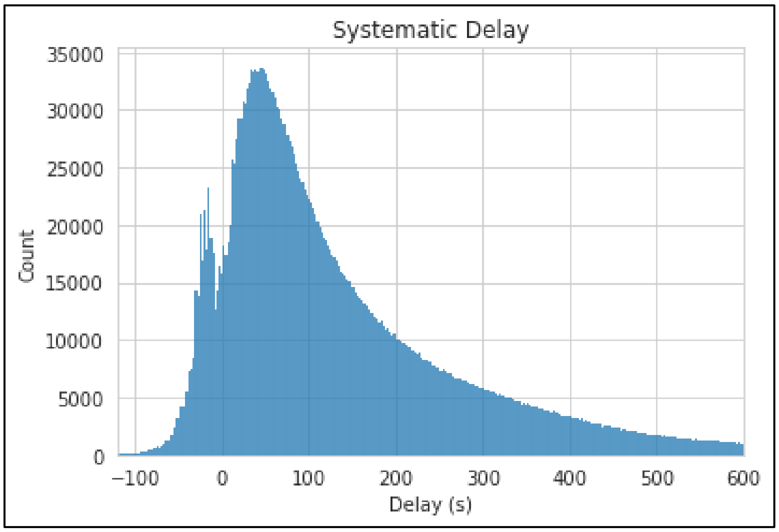

Fig. 5 Distribution of systematic delays in the bus locations dataset

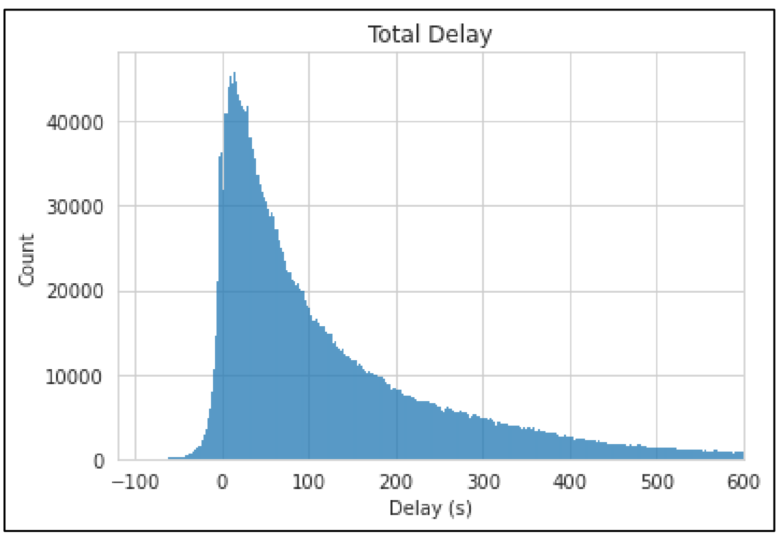

Fig. 6 Distribution of total delays in the bus locations dataset

Stochastic delay is approximately normal and centered around approximately -19 seconds. However, as shown in **Fig. 4**, instances of schedule deviation reported by the system appear to be rounded down to 0 when less than 15 seconds in magnitude, leading to a discontinuity in the center of the distribution at +/- 15 seconds. Because the distribution is non-zero centered, there could be systematic delays that are unaccounted for, and potential reliability improvements gained by adjusting bus schedules. In this case, the distribution of stochastic delays is skewed negative; therefore, buses are traveling ahead of schedule more often than behind or on-time. One cause of this may be over-padded bus schedules, as buses are consistently traveling faster than intended. The net effect may be a less reliable system, as passengers arrive at bus stops only to find that their bus has already passed them by. Treatments might include training for drivers to wait at hold positions when ahead of schedule on segments with high stochastic delay, or overall schedule adjustment to decrease allocated travel time. Stochastic delays greater than 120 seconds in magnitude were quite rare, with the combined set of all stochastic delays having a standard deviation of 34 seconds. The standard deviation of stochastic delay may be used to compare the relative unpredictability of segments within and across networks.

Total delay is skewed right and positive. This is because free-flow pace has been defined in this study as the 95th percentile, and total delay was calculated using the difference between actual and free-flow pace. Systematic delay is also skewed right, making it mostly positive. Due to its derivation from total and stochastic delays, there is a small discontinuity near -15 seconds. This may again be the result of schedule deviation being rounded down to zero by the system when smaller than 15 seconds in magnitude.

5.2 Segment Results

The analysis lends itself well to identifying hot spots in the network which carry high-magnitude delays. **Figs. 7-12** show the quantitative and spatial distributions of each classification of mean delay experienced by buses traversing a given segment in the network.

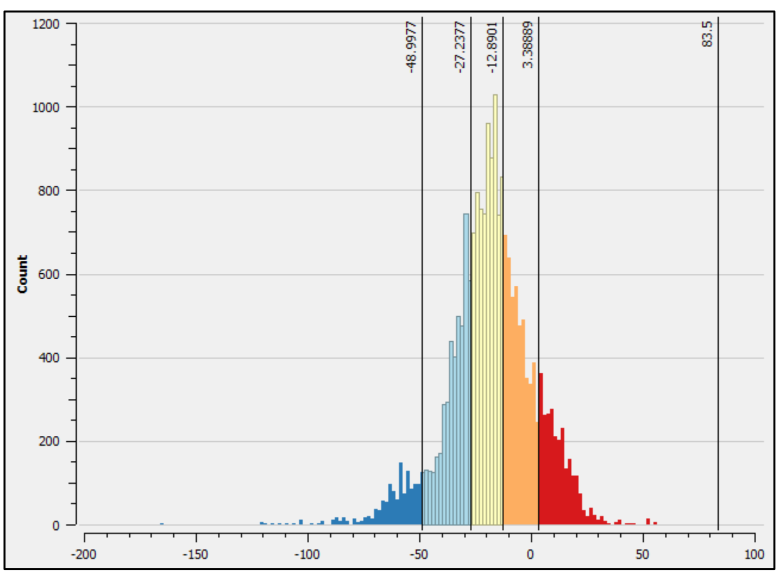

Fig. 7 Distribution of mean stochastic delay for all segments in the study network

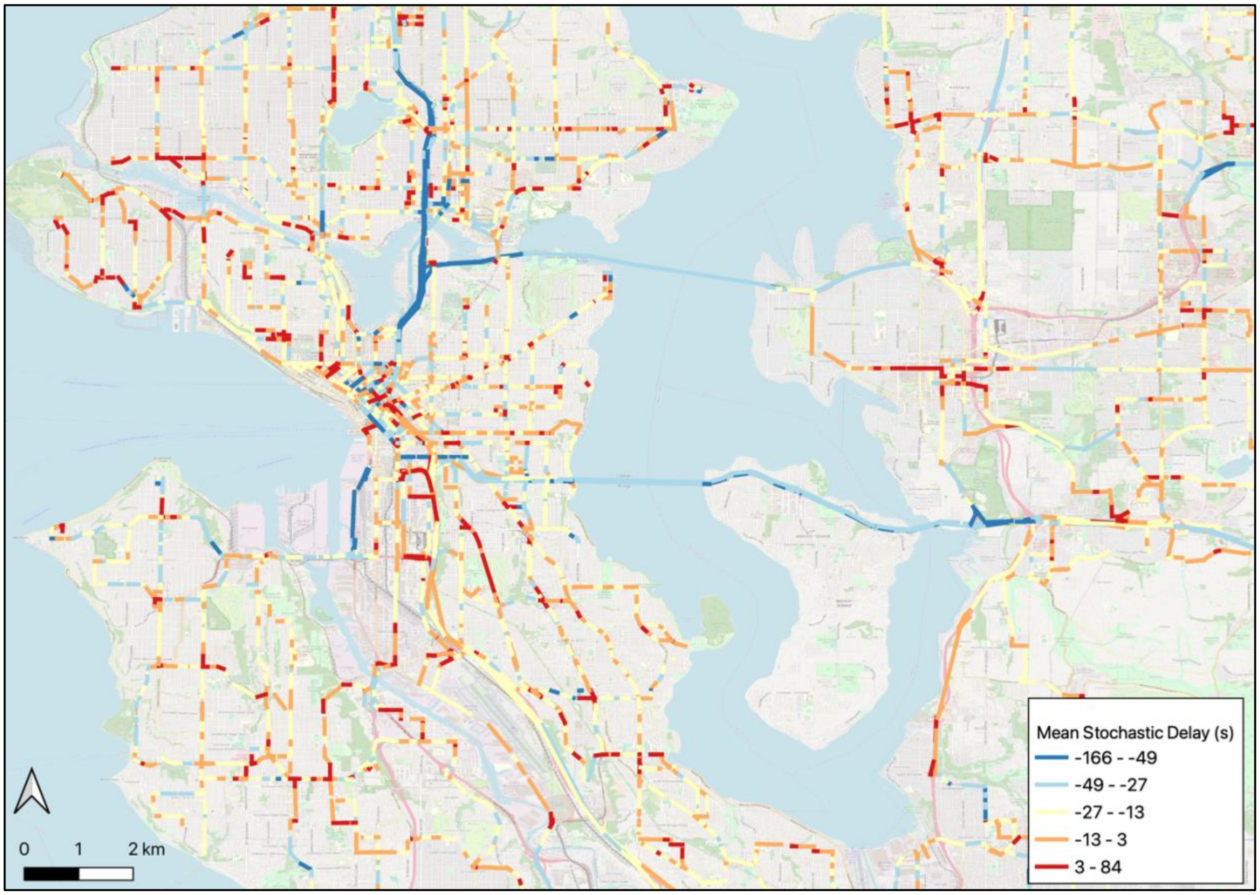

Fig. 8 Spatial distribution of stochastic delays in the study network

One of the surprising findings of this case study was the presence of very high stochastic speedups on some highway segments relative to local streets. This is shown most clearly on the north-south traversing I-5 segments north of Seattle which each carry average speedups of about 1-2 minutes. This may be because bus schedules accounting for time between stops group these segments with local streets. So, when a bus leaves the highway, and continues on local streets to its next stop, it accumulates delays which balance out the highway speedups. Alternatively, this may be explained by large amounts of schedule padding for highway segments (which can have highly variable speeds depending on congestion conditions) with the understanding that drivers may wait at control points if not delayed.

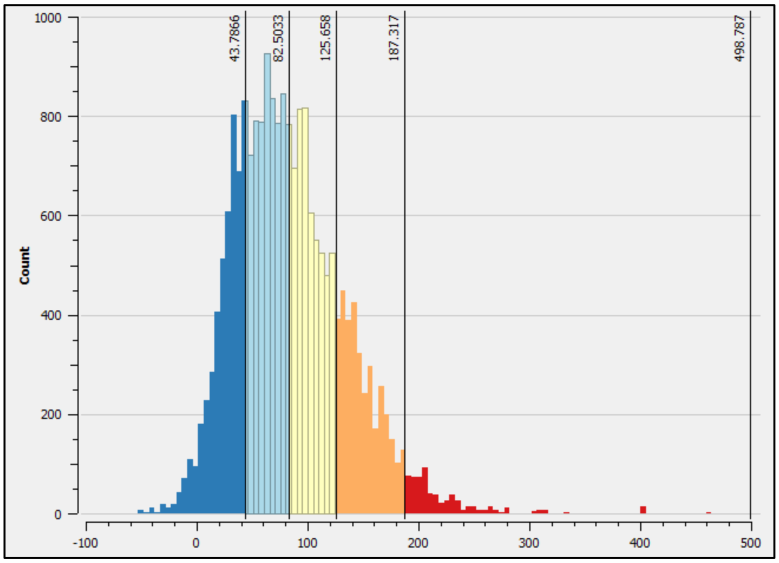

Fig. 9 Distribution of mean systematic delay for all segments in the network

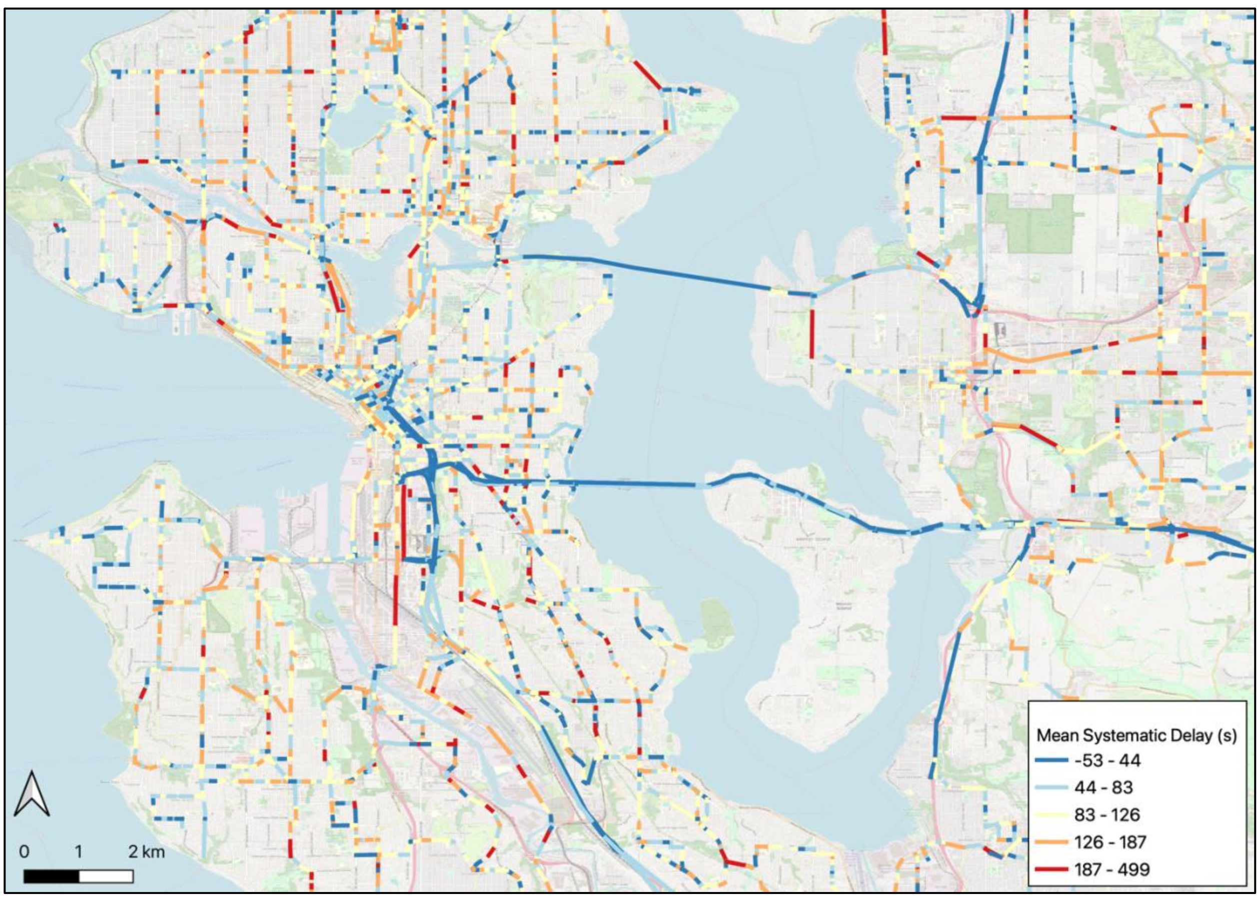

Fig. 10 Spatial distribution of systematic delays

In comparison to the stochastic delay results, some aspects of systematic delays were similar. Again, certain highway segments such as the east-west I-90 and SR-520 segments carried relatively low average delay. However, the distribution of mean segment delay reveals a shift from speedups to slowdowns, with most segments carrying a positive average delay. With the exception of a handful of segments, the magnitude of stochastic delay is relatively small compared so systematic delay. Thus, the trends of total delay (**Figs. 11, 12**) tend to mirror those of systematic delay.

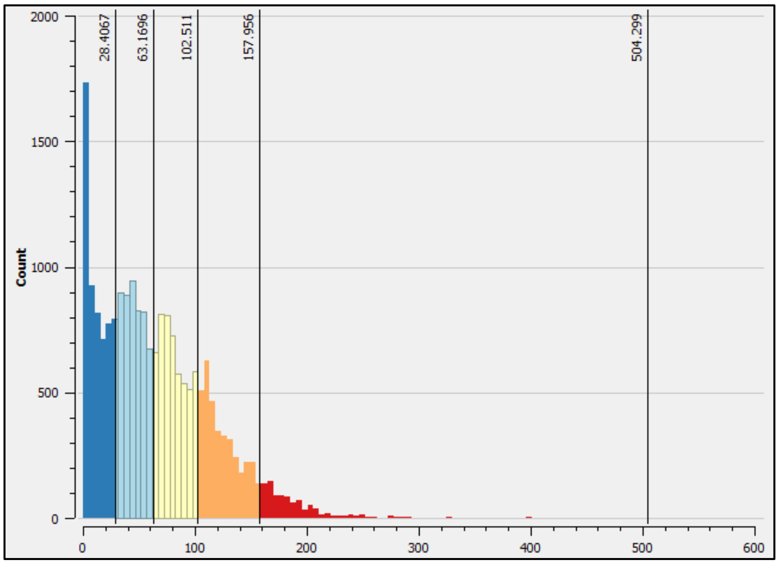

Fig. 11 Distribution of mean total delay for all segments in the network

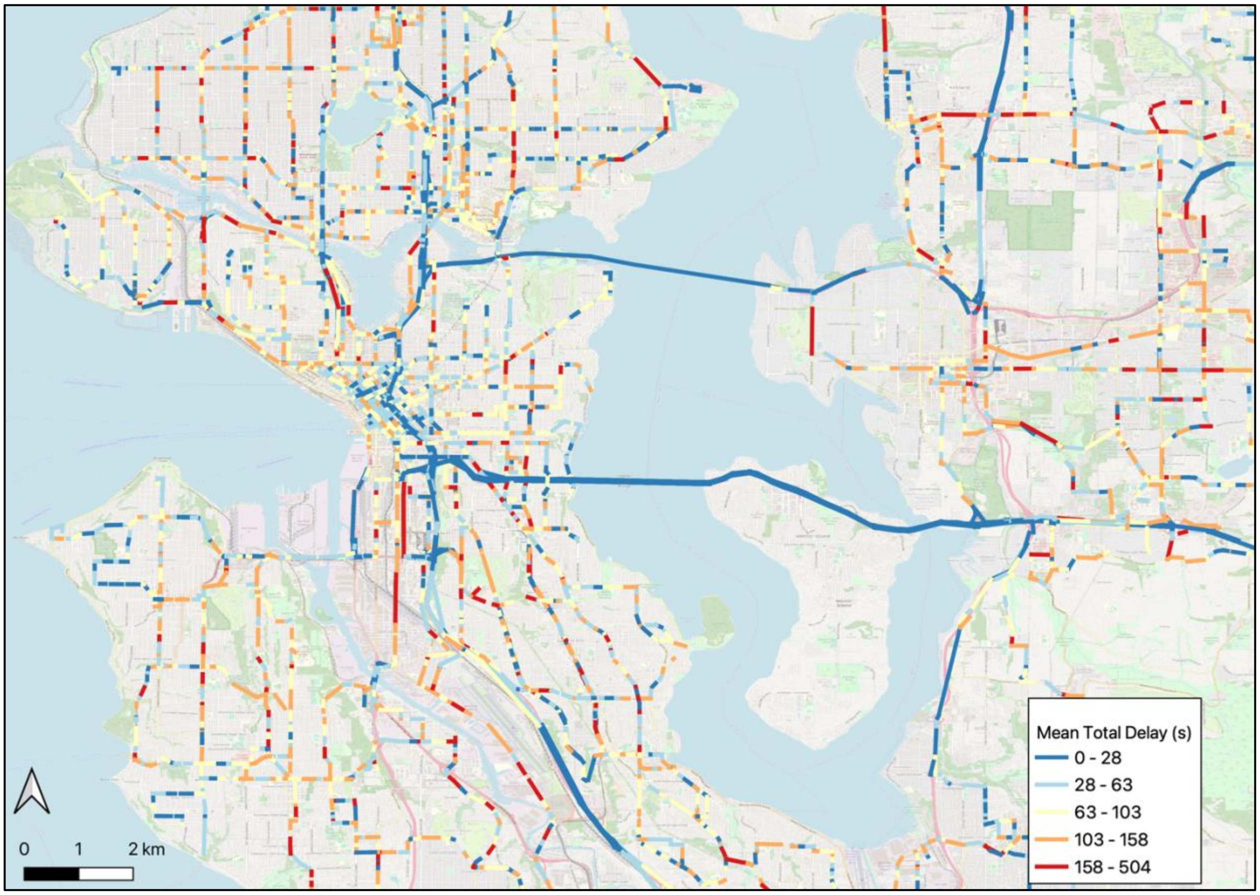

Fig. 12 Spatial distribution of total delay

In addition to mean delay, the mean pace is calculated for each segment in the network across all observations in the dataset. The spatial distribution is shown in **Fig. 13**. This provides perhaps the clearest picture of efficiency in the network derived from GTFS-RT, despite its lack of passenger data. The largest cluster of slower segments is found in the central business district (CBD), with additional zones to the east (CBD of neighboring city Bellevue) and north (University of Washington). As would be expected, highway segments are consistently higher pace than the rest of the network. Of course, this does not necessarily make them more efficient than the rest of the network, as they have fewer stops and higher speed limits.

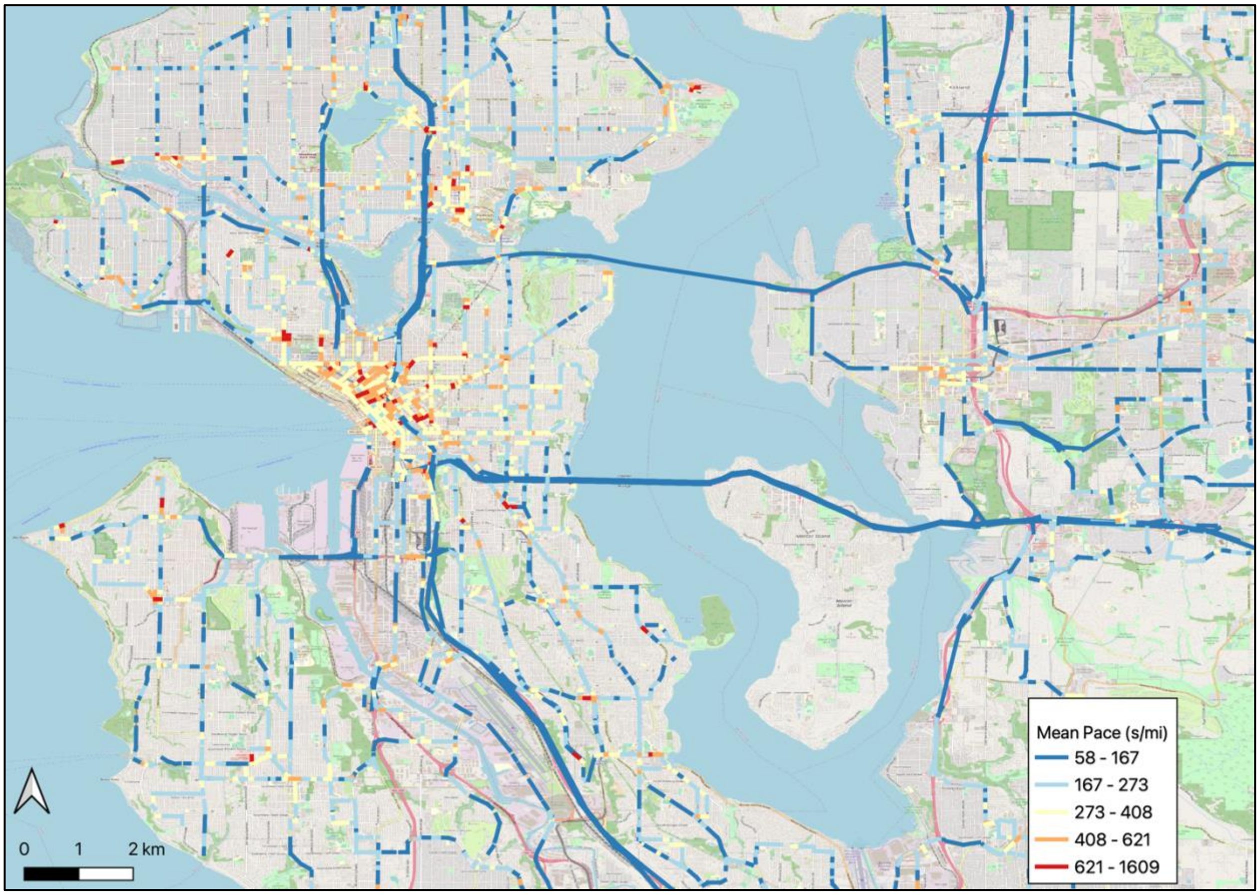

Fig. 13 Spatial distribution of pace

These results present a static analysis of 12 months of transit location data; however, the primary strength of GTFS-RT is that it is real time and can provide continuous performance insights as buses traverse the network. Thus, an interactive transit network performance visualization tool "TransitVis" was developed for the data used in this paper and is available online, along with open-source Python code for creating the necessary databases and performing the analysis (Aemmer, 2022). This tool uses the same analysis framework as discussed in this paper but provides additional performance measures and a higher temporal resolution (**Fig. 14**).

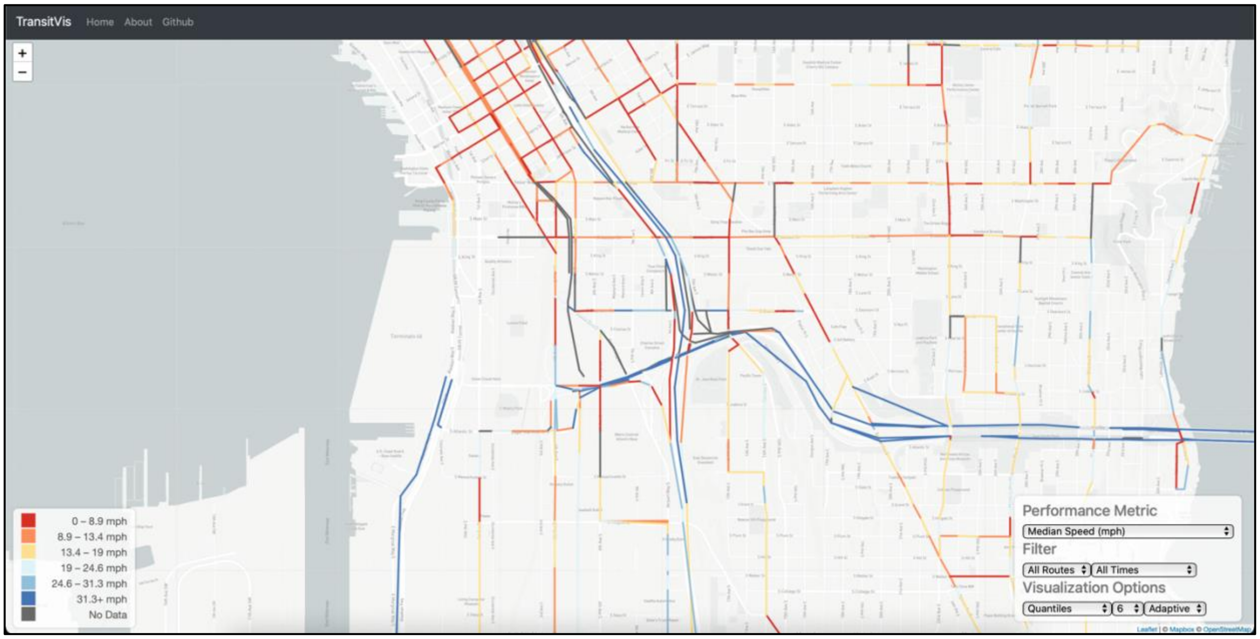

Fig. 14 User interface for interactive visualization tool

6. CONCLUSIONS

In this paper, we developed a framework for utilizing GTFS-RT data for locating and characterizing service performance metrics. We outlined a framework for collecting and analyzing transit delay data on a system-wide scale, using a case study of the KCM network in Seattle, Washington.

We developed two metrics that GTFS-RT data was capable of informing. The first, stochastic delay, was based on the schedule deviation parameter, which is an indication of delays that are not accounted for by schedule padding or intentional driver checkpoint slowdowns. In the case study, stochastic delay was distributed equally throughout the network, with the exception of highway segments which experienced unscheduled speedups. The second metric, systematic delay, was based on odometer (distance) and timestamp (time) sensors reported in the system. This allowed for the calculation of average speed and pace at each tracked location. 95th percentile pace was taken as the free-flow, non-delayed pace, and delay was calculated based on the GTFS-RT measurements. In theory, this should account for all systematic and stochastic delay that occurred between tracked locations. This assumes that the current schedule properly accounts for all systematic delay. However, we observed that the distribution of stochastic delays in the system was centered slightly towards speedups, indicating that perhaps on a system-wide level the current schedule is slightly over padded.

This framework allowed for segment-based analysis of transit hotspots and differentiated between types of delay which may require different types of treatment. For example, less reliable areas with negative stochastic delay may be treated by removing schedule padding. Whereas segments with high systematic delay may be inefficient and could be improved with transit signal priority or bus lanes. Any corridor identified by either of the two metrics may also become candidate for follow-on observation to identify the exact causes of delay. In fact, because GTFS-RT has less detail than combined AVL/APC data, the main use-case for this class of analysis may be as a preliminary screening tool for locations in the network with performance anomalies, which can then be used to target additional monitoring resources more efficiently.

This approach also does not measure other important components of transit performance, such as accessibility, which may rely on factors of the built environment to quantify, for which there is currently no standardized data source (Kim and Song, 2018). Additionally, many measures of efficiency rely on data gathered from APC systems, which record the number of passengers boarding and alighting at each transit stop. Bus occupancy is currently an experimental feature of GTFS-RT. This data is not currently available in the feeds used for this case study but would greatly improve the number of performance metrics that can be informed by GTFS-RT.

Last, this work focused on a segment-based analysis of transit performance. However, the GTFS-RT standard includes additional information on the current stop of each active vehicle. This could be used to further separate analysis based on delay accrued at segments, and delay accrued at individual stops. This would provide a more nuanced analysis of why certain segments or corridors decrease performance relative to others and open the way for predictive delay models based on GTFS-RT data. Future work might build on the methods presented here through the inclusion of field observation, video, or AVL data as ground truth validation for information provided in a GTFS-RT feed. Field observations in particular might complement this work by providing labeled delay data, which could enable classification models for the cause of transit delays, and census survey data could explain who is affected by delays and how. Upscaling methods for GTFS-RT trained on AVL data might also bring more detailed performance insights. Overall, synthesizing these different data sources would allow for the development of more sophisticated metrics related to bus movement such as acceleration, average queuing distance for individual intersections, or average stop dwell time. Deriving these secondary features from those already available in GTFS-RT could then inform targeted treatments for specific causes of delay.

7. REFERENCES

Aemmer, Z., 2022. Transit Vis - Real-Time Seattle Transit Performance Metrics [WWW Document]. TransitVis. URL https://www.transitvis.com/ (accessed 1.20.22).

Albadvi, A., Hashemi, H., Amin-Naseri, M., Teimourpour, B., 2018. A Statistical Analysis Framework for Bus Reliability Evaluation Based on AVL Data: A Case Study of Qazvin, Iran. Journal of Industrial Engineering and Management 5(1), 85–105. https://doi.org/10.22116/JIEMS.2018.66506

Antrim, A., Barbeau, S.J., 2017. Opening the Door to Multimodal Applications: Creation, Maintenance and Application of GTFS Data. Presented at the Transportation Research Board 96th Annual Meeting, Washington DC, United States.

Arias, D., Todd, K., Krieger, J., Maddox, S., Haley, P., Watkins, K.E., Berrebi, S., 2021. Using GTFS to Calculate Travel Time Savings Potential of Bus Preferential Treatments. Transportation Research Record 2675(9), 1643–1654. https://doi.org/10.1177/03611981211009226

Barbeau, S.J., 2018. Quality Control - Lessons Learned from the Deployment and Evaluation of GTFS-Realtime Feeds. Presented at the Transportation Research Board 97th Annual MeetingTransportation Research Board, Washington DC, United States.

Barbeau, S.J., Pham, M., Nieves, J.A., Bertini, R.L., 2020. Improving Transportation Performance Measurement via Open “Big Data” Systems – Phase 1 Transit. Center for Transportation, Equity, Decisions & Dollars (CTEDD), Arlington, Texas.

Bates, J., Polak, J., Jones, P., Cook, A., 2001. The Valuation of Reliability for Personal Travel. Transportation Research Part E: Logistics and Transportation Review, Advances in the Valuation of Travel Time Savings 37(2), 191–229. https://doi.org/10.1016/S1366-5545(00)00011-9

Berger, A., Gebhardt, A., Müller-Hannemann, M., Ostrowski, M., 2011. Stochastic Delay Prediction in Large Train Networks. Presented at the 11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Saarbrücken, Germany, pp. 100–111. https://doi.org/10.4230/OASIcs.ATMOS.2011.100

Berkow, M., Chee, J., Bertini, R., Monsere, C., 2007. Transit Performance Measurement and Arterial Travel Time Estimation Using Archived AVL Data, in: ITE District 6 Annual Meeting. Presented at the ITE Annual Meeting and Exhibition, Institute of Transportation Engineers, Washington DC, United States, p. 6.

Bertini, R.L., El-Geneidy, A., 2003. Generating Transit Performance Measures with Archived Data. Transportation Research Record 1841(1), 109–119. https://doi.org/10.3141/1841-12

Bertini, R.L., Tantiyanugulchai, S., 2004. Transit Buses as Traffic Probes: Use of Geolocation Data for Empirical Evaluation. Transportation Research Record 1870(1), 35–45. https://doi.org/10.3141/1870-05

Bills, T.S., Carrel, A.L., 2021. Transit Accessibility Measurement Considering Behavioral Adaptations to Reliability. Transportation Research Record 2675(5), 265–278. https://doi.org/10.1177/0361198120986567

Bowman, L.A., Turnquist, M.A., 1981. Service Frequency, Schedule Reliability and Passenger Wait Times at Transit Stops. Transportation Research Part A: General 15(6), 465–471. https://doi.org/10.1016/0191-2607(81)90114-X

Caros, N., Guo, X., Stewart, A., Attanucci, J., 2021. Systematic Bus Corridor Identification and Performance Measure Aggregation: Method and Case Studies.

Ceder, A., 2002. Urban Transit Scheduling: Framework, Review and Examples. Journal of Urban Planning and Development 128(4), 225–244. https://doi.org/10.1061/(ASCE)0733-9488(2002)128:4(225)

Cham, L.C., 2006. Understanding bus service reliability : a practical framework using AVL/APC data (Thesis). Massachusetts Institute of Technology.

Coghlan, C., Dabiri, S., Mayer, B., Wagner, M., Williamson, E., Eichler, M., Ramakrishnan, N., 2019. Assigning Bus Delay and Predicting Travel Times using Automated Vehicle Location Data. Transportation Research Record 2673(3), 624–636. https://doi.org/10.1177/0361198119832866

Currie, G., Mesbah, M., 2011. Visualization of Geographical Information System and Automatic Vehicle Location Data to Explore Transit Performance. Transportation Research Record 2216(1), 59–66. https://doi.org/10.3141/2216-07

Danaher, A., Wensley, J., Dunham, A., Orosz, T., Avery, R., Cobb, K., Watkins, K., Queen, C., Berrebi, S., Connor, M., McLaughlin, J., 2020. Minutes Matter: A Bus Transit Service Reliability Guidebook (Transit Cooperative Research Program No. 215), Transportation Research Board; Transit Cooperative Research Program. National Academies of Sciences, Engineering, and Medicine. https://doi.org/10.17226/25727

de Oña, J., de Oña, R., Calvo, F., 2012. A Classification Tree Approach to Identify Key Factors of Transit Service Quality. Expert Systems with Applications 39(12), 11164–11171. https://doi.org/10.1016/j.eswa.2012.03.037

Diaz, F., Abbasi, S.J., Fuller, D., Diab, E., 2021. Canadian Transit Agencies Response to COVID-19: Understanding Strategies, Information Accessibility and the Use of Social Media. Transportation Research Interdisciplinary Perspectives 12(), 100465. https://doi.org/10.1016/j.trip.2021.100465

Dixit, M., Brands, T., van Oort, N., Cats, O., Hoogendoorn, S., 2019. Passenger Travel Time Reliability for Multimodal Public Transport Journeys. Transportation Research Record 2673(2), 149–160. https://doi.org/10.1177/0361198118825459

Duddu, V.R., Pulugurtha, S.S., Najaf, P., 2019. Using Link-Level Archived Automatic Vehicle Location Data to Assess Transit System LOS at Bus-Stop Level. Transportation in Developing Economies 5(2), 11. https://doi.org/10.1007/s40890-019-0079-3

El-Geneidy, A.M., Horning, J., Krizek, K.J., 2011. Analyzing transit service reliability using detailed data from automatic vehicular locator systems. Journal of Advanced Transportation 45(1), 66–79. https://doi.org/10.1002/atr.134

Feng, W., Figliozzi, M., 2012a. Empirical Findings of Bus Bunching Distributions and Attributes Using Archived AVL/APC Bus Data. Presented at the 11th International Conference of Chinese Transportation Professionals, American Society of Civil Engineers, pp. 4330–4341. https://doi.org/10.1061/41186(421)427

Feng, W., Figliozzi, M., 2012b. Developing a Bus Service Reliability Evaluation and Visualization Framework Using Archived AVL / APC Data. Presented at the Western ITE Conference, Institute of Transportation Engineers, Santa Barbara, California, United States.

Ferris, B., Watkins, K., Borning, A., 2010. OneBusAway: A Transit Traveler Information System, in: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering Book Series. Presented at the International Conference on Mobile Computing, Applications, and Services (MobiCASE), LNICST, pp. 92–106. https://doi.org/10.1007/978-3-642-12607-9_

Fu, L., Rilett, L.R., 1997. Real-Time Estimation of Incident Delay in Dynamic and Stochastic Networks: Transportation Research Record 1603(1), 99–105. https://doi.org/10.3141/1603-13

Furth, P., Hemily, B., Muller, T., Strathman, J., 2006. Using Archived AVL-APC Data to Improve Transit Performance and Management. World Transit Research.

Ge, L., Sarhani, M., Voß, S., Xie, L., 2021. Review of Transit Data Sources: Potentials, Challenges and Complementarity. Sustainability 13(20), 11450. https://doi.org/10.3390/su132011450

Gleason, J.M., Barnum, D.T., 1982. Toward Valid Measures of Public Sector Productivity: Performance Measures in Urban Transit. Management Science 28(4), 379–386. https://doi.org/10.1287/mnsc.28.4.379

Godavarthi, G.R., Chalumuri, R.S., Velmurugun, S., 2014. Measuring the Performance of Bus Rapid-Transit Corridors Based on Volume by Capacity Ratio. Journal of Transportation Engineering 140(10), 04014049. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000698

He, Y., Blandin, S., Wynter, L., Trager, B., 2014. Analysis and Real-Time Prediction of Local Incident Impact on Transportation Networks. Presented at the IEEE International Conference on Data Mining Workshop, IEEE, Shenzhen, China, pp. 158–166. https://doi.org/10.1109/ICDMW.2014.183

Jevinger, Å., Persson, J.A., 2019. Exploring the Potential of Using Real-Time Traveler Data in Public Transport Disturbance Management. Public Transport 11(2), 413–441. https://doi.org/10.1007/s12469-019-00209-w

Kaeoruean, K., Phithakkitnukoon, S., Demissie, M.G., Kattan, L., Ratti, C., 2020. Analysis of Demand–Supply Gaps in Public Transit Systems Based on Census and GTFS Data: A Case Study of Calgary, Canada. Public Transp 12(3), 483–516. https://doi.org/10.1007/s12469-020-00252-y

Kim, H., Song, Y., 2018. An Integrated Measure of Accessibility and Reliability of Mass Transit Systems. Transportation 45(4), 1075–1100. https://doi.org/10.1007/s11116-018-9866-7

Kumar, P., Khani, A., He, Q., 2018. A Robust Method for Estimating Transit Passenger Trajectories Using Automated Data. Transportation Research Part C: Emerging Technologies 95(), 731–747. https://doi.org/10.1016/j.trc.2018.08.006

Li, T., Meredith-Karam, P., Kong, H., Stewart, A., Attanucci, J.P., Zhao, J., 2021. Comparison of Door-to-Door Transit Travel Time Estimation Using Schedules, Real-Time Vehicle Arrivals, and Smartcard Inference Methods. Transportation Research Record 2675(11), 1003–1014. https://doi.org/10.1177/03611981211023768

Li, T., Sun, D., Jing, P., Yang, K., 2018. Smart Card Data Mining of Public Transport Destination: A Literature Review. Transportation Research Part C: Emerging Technologies 9(1), 18. https://doi.org/10.3390/info9010018

Lin, W.-H., Bertini, R.L., 2004. Modeling schedule recovery processes in transit operations for bus arrival time prediction. Journal of Advanced Transportation 38(3), 347–365. https://doi.org/10.1002/atr.5670380306

Lind, E., Reid, J., 2021. Diagnosing Obstacles to Speed and Reliability with High-Resolution Automatic Vehicle Locator Data: Bus Time Budgets. Transportation Research Record 2675(12), 464–474. https://doi.org/10.1177/03611981211029923

Lu, K., Liu, J., Zhou, X., Han, B., 2021. A Review of Big Data Applications in Urban Transit Systems. IEEE Transactions on Intelligent Transportation Systems 22(5), 2535–2552. https://doi.org/10.1109/TITS.2020.2973365

Mandelzys, M., Hellinga, B., 2010. Identifying Causes of Performance Issues in Bus Schedule Adherence with Automatic Vehicle Location and Passenger Count Data: Transportation Research Record 2143(1), 9–15. https://doi.org/10.3141/2143-02

Mesbah, M., Lin, J., Currie, G., 2015. "Weather" Transit is Reliable? Using AVL Data to Explore Tram Performance in Melbourne, Australia. Journal of Traffic and Transportation Engineering (English Edition) 2(3), 125–135. https://doi.org/10.1016/j.jtte.2015.03.001

Nurul Habib, K.M., Kattan, L., Tazul Islam, M., 2011. Model of Personal Attitudes Towards Transit Service Quality. Journal of Advanced Transportation 45(Modeling and Optimization of Transportation Systems), 271–285. https://doi.org/10.1002/atr.106

Pi, X., Egge, M., Whitmore, J., Silbermann, A., Qian, Z.S., 2018. Understanding Transit System Performance Using AVL-APC Data: An Analytics Platform with Case Studies for the Pittsburgh Region. Journal of Public Transportation 21(2), 19–40. https://doi.org/10.5038/2375-0901.21.2.2

Quddus, M.A., Ochieng, W.Y., Noland, R.B., 2007. Current Map-Matching Algorithms for Transport Applications: State-of-the Art and Future Research Directions. Transportation Research Part C: Emerging Technologies 15(5), 312–328. https://doi.org/10.1016/j.trc.2007.05.002

Sheu, J.-B., Chou, Y.-H., Chen, A., 2004. Stochastic Modeling and Real-Time Prediction of Incident Effects on Surface Street Traffic Congestion. Applied Mathematical Modelling 28(5), 445–468. https://doi.org/https:/doi.org/10.1016/j.apm.2003.10.004

Steiner, D., Hochmair, H., Paulus, G., 2015. Quality Assessment of Open Realtime Data for Public Transportation in the Netherlands. Journal for Geographic Information Science, Geospatial Minds for Society 3(1), 579–588. https://doi.org/10.1553/giscience2015s579

Strathman, J.G., Dueker, K.J., Kimpel, T., Gerhart, R., Turner, K., Taylor, P., Callas, S., Griffin, D., Hopper, J., 1999. Automated Bus Dispatching, Operations Control, and Service Reliability: Baseline Analysis: Transportation Research Record 1666(1), 28–36. https://doi.org/10.3141/1666-04

Tavassoli, A., Mesbah, M., Shobeirinejad, A., 2018. Modelling Passenger Waiting Time Using Large-Scale Automatic Fare Collection Data: An Australian Case Study. Transportation Research Part F: Traffic Psychology and Behaviour 58(), 500–510. https://doi.org/10.1016/j.trf.2018.06.037

US Census Bureau, 2021. TIGER/Line Shapefiles [WWW Document]. The United States Census Bureau. URL https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html (accessed 9.8.21).

Watkins, K.E., Ferris, B., Borning, A., Rutherford, G.S., Layton, D., 2011. Where Is My Bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. Transportation Research Part A: Policy and Practice 45(8), 839–848. https://doi.org/10.1016/j.tra.2011.06.010

Webb, A., Kumar, P., Khani, A., 2020. Estimation of Passenger Waiting Time Using Automatically Collected Transit Data. Public Transport 12(2), 299–311. https://doi.org/10.1007/s12469-020-00229-x

Wessel, N., Widener, M.J., 2017. Discovering the Space–Time Dimensions of Schedule Padding and Delay from GTFS and Real-Time Transit Data. Journal of Geographical Systems 19(1), 93–107. https://doi.org/10.1007/s10109-016-0244-8

Yu, B., Yang, Z., 2009. A Dynamic Holding Strategy in Public Transit Systems with Real-Time Information. Applied Intelligence 31(1), 69–80. https://doi.org/10.1007/s10489-007-0112-9

Yue, W.S., Chye, K.K., Hoy, C.W., 2017. Towards Smart Mobility in Urban Spaces: Bus Tracking and Information Application, in: AIP Conference Proceedings. Presented at the American Institute of Physics Conference, AIP, p. 020145. https://doi.org/10.1063/1.5005478