An Automated Quality Assurance Procedure for Archived Transit Data from APC and AVL Systems

by

Marian Saavedra

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Civil Engineering

Waterloo, Ontario, Canada, 2010

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Automatic Vehicle Location (AVL) and Automatic Passenger Counting (APC) systems can be powerful tools for transit agencies to archive large, detailed quantities of transit operations data. Managing data quality is an important first step for exploiting these rich datasets.

This thesis presents an automated quality assurance (QA) methodology that identifies unreliable archived AVL/APC data. The approach is based on expected travel and passenger activity patterns derived from the data. It is assumed that standard passenger balancing and schedule matching algorithms are applied to the raw AVL/APC data along with any existing automatic validation programs. The proposed QA methodology is intended to provide transit agencies with a supplementary tool to manage data quality that complements, but does not replace, conventional processing routines (that can be vendor-specific and less transparent).

The proposed QA methodology endeavours to flag invalid data as "suspect" and valid data as "non-suspect". There are three stages: i) the first stage screens data that demonstrate a violation of physical constraints; ii) the second stage looks for data that represent outliers; and iii) the third stage evaluates whether the outlier data can be accounted for with valid or invalid pattern. Stop-level tests are mathematically defined for each stage; however data is filtered at the trip-level. Data that do not violate any physical constraints and do not represent any outliers are considered valid trip data. Outlier trips that may be accounted for with a valid outlier pattern are also considered valid. The remaining trip data is considered suspect.

The methodology is applied to a sample set of AVL/APC data from Grand River Transit in the Region of Waterloo, Ontario, Canada. The sample data consist of 4-month's data from September to December of 2008; it is comprised of 612,000 stop-level records representing 25,012 trips. The results show 14% of the trip-level data is flagged as suspect for the sample dataset. The output is further dissected by: reviewing which tests most contribute to the set of suspect trips; confirming the pattern assumptions for the valid outlier cases; and comparing the sample data by various traits before and after the QA methodology is applied. The latter task is meant to recognize characteristics that may contribute to higher or lower quality data. Analysis shows that the largest portion of suspect trips, for this sample set, suggests the need for improved passenger balancing algorithms or greater accuracy of the APC equipment. The assumptions for valid outlier case patterns were confirmed to be reasonable. It was found that poor schedule data contributes to poorer quality in AVL-APC data. An examination of data distribution by vehicle showed that usage and the portion of suspect data varied substantially between vehicles. This information can be useful in the development of maintenance plans and sampling plans (when combined with information of data distribution by route).

A sensitivity analysis was conducted along with an impact analysis on downstream data uses. The model was found to be sensitive to three of the ten user-defined parameters. The impact of the QA procedure on network-level measures of performance (MOPs) was not found to be significant, however the impact was shown to be more substantial for route-specific MOPs.

Acknowledgements

I want to express my profound thanks to all the people whom helped me prepare this thesis.

The warmest gratitude is reserved for my advisor, Dr. Bruce Hellinga, for his patience and for always asking the right questions. I sincerely appreciate his support, both academically and financially.

I would also like to thank my co-supervisor, Dr. Jeffrey Casello, for his valuable contributions to my thesis as well as his financial support. I am grateful for his time and expertise.

This project was funded through a research partnership with Grand River Transit, the Region of Waterloo and the University of Waterloo. Thank you to the transit planning team for their valuable feedback and financial support. Special thanks to Reid Fulton for providing the necessary data and for answering numerous questions.

I owe my sanity to Fei Yang for his assistance with all things code. He has graciously consented to endless requests for programming help. I am greatly indebted to him.

For all the input from every weekly meeting, thank you to Mike Mandelzys and Akram Nour.

Thank you to all my transportation colleagues, especially the UW-ITE student executives and fellow WPTI researchers, for a rewarding place to work and learn. To Stephanie and Aaron, I am grateful for our many therapeutic discussions about life as a graduate student.

To my parents, thank you for endless love and support. To my brothers, thank you for the get-away from work and school.

And to my biggest fan of all, Al, a very special thank you for always reminding me what I am capable of, including this.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

There is a mutual relationship between the ability of a region to move goods and people and the region's economic well-being. As cities grow, travel demand increases from the need to connect workers to their workplaces, suppliers to customers, and trade within and between regions. A comprehensive transportation network is needed to meet this demand and governments are recognizing the need to invest in sustainable transportation infrastructure. An efficient public transportation system is a necessary part of a comprehensive network; it is a practical solution for urban mobility needs because transit makes more effective use of the limited public space in urban areas.

Efficient transit operation can be achieved when transit agencies can monitor their operations, report performance and plan for future demand. Some transit agencies have adopted Automatic Vehicle Location (AVL) and Automated Passenger Counting (APC) systems as useful tools to achieve these tasks.

1.1 Introduction to AVL and APC Systems

AVL technologies allow transit agencies to monitor vehicle movements through the transmission of geographic location data to a central controller. There are three primary methods of tracking vehicles: signpost technique, LORAN C technology, and global positioning system (GPS). Signposts determine position via a fixed installation of electronic beacons located at various bus stops or traffic signals. LORAN C is land-based technology that consists of radio transmissions relayed through land connections (Perk & Kamp, 2003). Modern AVL systems use global positioning systems (GPS) technology, which rely on satellite tracking, for time and location stamps. AVL systems have historically been developed for real-time applications, such as support tools for dispatchers (Furth et al., 2004) and these applications continue to be the primary use of AVL systems (Parker, 2008). Other real-time uses include communication to traffic control systems for transit signal priority and integration to traveller information systems, such as next bus notification.

APCs automatically count the number of boarding and alighting passengers by door and stop. They can be useful for estimating ridership, passenger miles and peak loads without the need of farebox data or manual count surveys. Legacy APC systems are comprised of treadle mats to count passengers as they step onto or off of transit vehicles, however modern APCs use infrared sensors mounted at each door. The configuration and placement of multiple detectors allow the APC units to determine the number of people and direction of movement. The counts are usually stored to an on-board computer and later downloaded at a garage for off-line analysis. The need to associate passenger activity to designated stops means that most standalone APCs also have independent location referencing; stand-alone APC units were more expensive and were generally less popular during their earlier deployments (Furth et al., 2006).

The recognition of offline applications for AVL data led transit agencies to merge the two technologies into a hybrid AVL/APC system. Another advantage of a hybrid system is reduced marginal costs for APC installation by relying on the AVL component for location referencing. Hybrid systems commonly resulted from the upgrade of an AVL system to include APC features or vice versa, or as part of a broader Intelligent Transportation System (ITS) deployment. ITS refers to the use of information technology to advance and improve transportation systems. Therefore transit agencies and the literature sometimes refer to a hybrid AVL/APC system simply as an AVL or APC system. Such hybrid systems typically are comprised of both location data transmitted by radio for real-time applications and on-board event recording for data archives. Real-time AVL data is generally polled on a time recursive cycle (typically every 40 to 120s) also known as "*location-at-time*" and events records are generally triggered only at scheduled and unscheduled stops also known as "*time-at-location"* (Furth et al, 2006).

Archived AVL/APC data is most commonly of interest to transit planning and operation groups; rich datasets of vehicle movements and passenger activity information leads to more opportunities to monitor performance, analyse service deficiencies and plan routes. As AVL/APC systems advance, more uses for these archived data are being recognized by other transit business units. Some examples are the investigation of rider complaints by the customer service department and tracking of vehicle mileage and/or equipment malfunctions by maintenance crews. The transition from a data-poor to data-rich environment is transforming how transit agencies are monitoring their operations; on-going research is very active for the most effective methods to exploit archived AVL/APC data. For transit directors, data quality management is a key requisite for sound business decisions supported from these data.

1.1.1. Data collection process

Figure 1 depicts a modern AVL/APC system configuration. Modern AVL/APC systems generally employ GPS technology to track time and location. Recurring *location-at-time* data are transferred to a central computer via radio transmission for real-time applications.

Mounted on each door is an infrared sensor that counts the number of passengers boarding and alighting the vehicle. Counting sensors are generally triggered when the bus stops and doors open. The on-board computer processes information from the APC sensors and AVL equipment to generate a stop event record. Basic information contained in a stop record are the arrival and departure times, number of passengers boarding and alighting for each door (if the doors opened), the odometer distance travelled and GPS time and location coordinates.

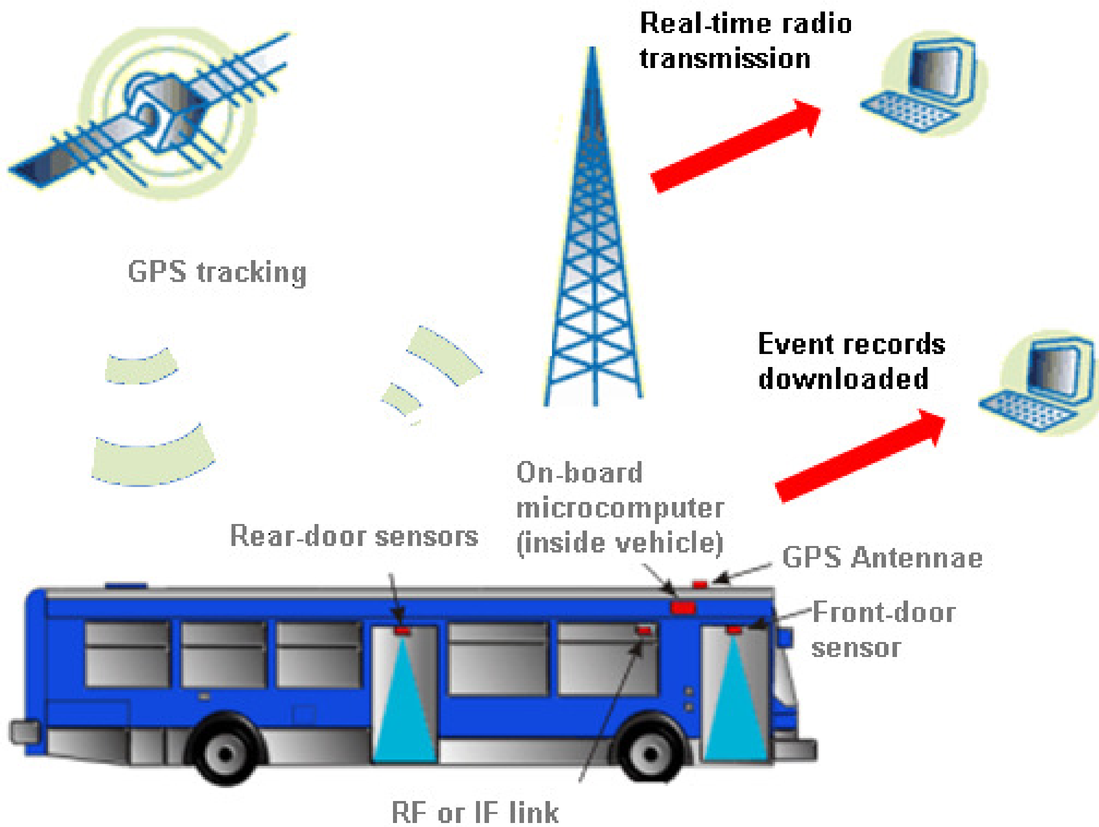

Figure 1 Physical components of an AVL-APC system (Source: Infodev)

Stop event records (*time-at-location* data) are stored on-board and later downloaded at the garage. Data stored on-board may be downloaded automatically via infrared (IF) or radio frequency (RF) modems or manually with a handheld data collector.

Once the data are downloaded, matching algorithms relate stop event records to schedule data. Schedule data may typically contain route identifiers and information about designated stops such as stop name, location, the distance between designated stop and an expected arrival and/or departure time. Not all stops are associated with a scheduled time; schedule planners often do not design routes at this level of detail. Therefore, a designated stop associated with a schedule time is called a time point.

To demonstrate the type of data collected by AVL/APC systems, Figure 2 is an example bus route and Table 1 demonstrates how the data might look within an archived database. The round points in Figure 2 represent terminal stops and the square points represent designated bus stops along the route. Time points are depicted by larger square points. Scheduled departure time is shown at Terminal A and scheduled arrival time is shown at Terminal B. Most schedule data will have the same arrival and departure time for stops without any planned dwell time.

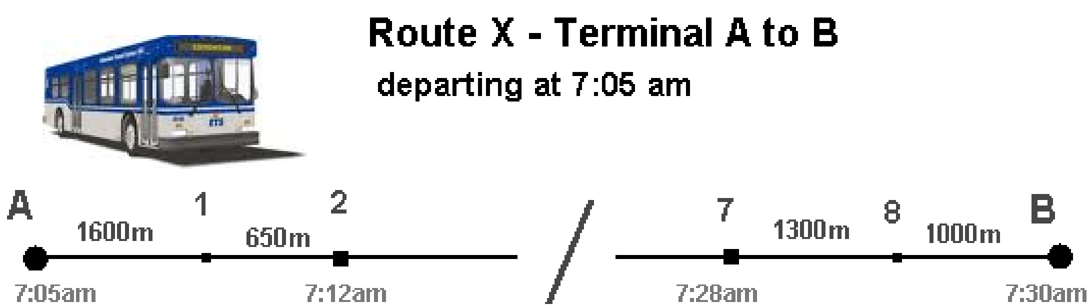

Figure 2 Example trip route

Table 1 Example stop-level view of AVL/APC data

EventID	TripID	Stop Name	Act Arr	Act Dep	Odom	Sch Arr	Sch Dep	Sch Dist	Board	Alight	Load
12345	2222	Terminal A	07:01:00	07:05:23	2391	07:05:00	07:05:00	0	14	0	14
12346	2222		07:07:48	07:07:54	2393						
12347	2222	Stop 1	07:10:30	07:10:54	4091			1600	8	7	15
12348	2222	Stop 2	07:11:58	07:12:12	4791	07:10:00	07:10:00	2250	16	4	23
12349	2222		07:13:02	07:13:10	4791						
12350	2222		07:14:06	07:14:12	4791						
-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-
12360	2222	Stop 7	07:33:24	07:33:30	11591	07:28:00	07:28:00	8900	5	11	10
12361	2222		07:34:36	07:34:40	11591						
12362	2222		07:34:02	07:35:00	11591						
12363	2222	Stop 8	07:35:26	07:37:42	12991			10200	3	7	14
12364	2222		07:38:45	07:38:51	12992						
12365	2222	Terminal B	07:40:12	07:40:36	14091	07:30:00	07:30:00	11200	0	14	0

Table 1 shows the AVL/APC data at the stop-level detail after passenger counts are balanced and the stop events records are matched to the schedule. Stop event records are not always associated with a route-designated stop (see EventID 12346, 12349, 12350 etc.). This type of stop event record is sometimes identified as a disturbance stop or an interstop record related to intersection or traffic delay.

Aggregated information and route attributes may also be stored at the trip-level and linked to the stop-level data in a relational database (Table 2).

Table 2 Example trip-level view of AVL/APC data

TripID	RouteID	Direction	VehID	Act Start	Act End	Dist	Sch Arr	Sch Dep	Sch Dist	Board	Alight
2222	X	Outbound	332	07:01:00	07:40:12	12362	07:05:00	07:30:00	11200	71	71
2223	Y	Inbound	225	08:07:48	09:02:54	24092	08:10:00	09:00:00	23900	38	38
2224	X	Outbound	231	10:30:12	11:24:34	11243	10:30:00	11:20:00	11200	28	28
2225	X	Inbound	432	18:33:00	19:12:56	10923	18:30:00	19:05:00	11200	55	55
2226	Z	Inbound	342	14:11:20	14:35:02	10232	14:10:00	14:35:00	10000	21	21
2227	Y	Inbound	231	13:21:00	14:35:00	23952	13:20:00	14:40:00	23900	12	12

Note that Tables 1 and 2 are example views of archived AVL/APC data. Additional columns may display the time in other formats to facilitate analysis. Both raw and balanced passenger counts can be included. (More information on passenger balancing algorithms is available in

Section 2.1). Load can be derived from the boarding and alighting counts, therefore some database designs would omit load. Odometer readings may be processed into travel distances referenced from the first or previous stop. GPS coordinates may be included for mapping capabilities.

Differences in ITS architecture, hardware and software vendors and in-house IT (information technology) support result in variations to the data collection and processing routines. For example, some AVL/APC systems store only a series of sensor signals on-board and then aggregated into stop event records offline. Some AVL/APC systems record stop events only at designated stops or only at time points. Advanced systems are capable of identifying the route and run in real-time; an example is buses with next stop announcement or estimated arrival times at stations. There are several approaches to schedule matching that may rely on operator sign-in, vehicle to run assignment files and/or spatial analysis in geographic information systems (GIS). Built-in quality checks may occur before or after stop events processing and schedule matching; these processes can also fluctuate between vendors.

1.1.2. Data quality issues

Different data quality needs for different purposes

As transit agencies began to incorporate AVL data for offline analyses, practitioners recognized the contrasting data quality needs from real-time applications. In real-time, erroneous data may appear as just a small blip in a continual data stream transmitted to the controller; no lasting impact is made to the immediate operations. However in off-line analysis, errors in data archives could results in poor performance analysis results (Kemp, 2002). An example of this impact may be demonstrated with schedule adherence.

Schedule adherence measures state how often buses arrive within a given threshold from the scheduled stop time (i.e. 85% of the time, buses arrive within 5mins of the schedule). Temporary malfunctioning clocks will minimally disturb real-time applications but can render large schedule deviations during offline analysis. Even where recorded time and location are correct, matching algorithms may associate a stop event with another designed stop or route, rendering the schedule adherence measures invalid.

Among offline uses, differences in data quality requirements also vary based on the intended purpose. When adjusting the schedule based on expected travel times, planners may not be interested in data resulting from severe weather events that can skew a runtime analysis. (Runtime analyses examines the travel times of in-service transit vehicles between two designated stops). However data from exceptional events would still be considered valid for performance reporting. Along the same note, applications of archived AVL/APC data also often rely on extreme values such as: identifying routes with the highest or lowest ridership, analysing run times of routes with poor schedule adherence (i.e. large schedule deviations). Business decisions allocating limited resources become more prone to invalid extreme data.

More opportunities to introduce error

Automatic data collection is computationally more complex than traditional methods. In a manual survey, a ride checker can easily identify the route, direction, stop location, time and number of boarding and alighting passenger by visual inspection. Error may be introduced by poor visual inspection and improper documentation due to human limitation. In an AVL/APC system, passengers are detected first by breaks in the APC sensor beams. Active sensors and passive sensors are two types of infrared sensors, they each work differently. Active sensors rely on the reflection of passing objects dark colors do not reflect well; passive sensors are based on detecting change in body heat. Some manufacturers combine the two types for improved accuracy (Perk & Kamp, 2003). Systematic under or over-counting (bias) and random counting error can present themselves in APC collected data; bias is considered a more serious measurement error (Furth, Strathman, & Hemily, 2005).

Next, an APC analyzer must process these beam breaks into a count and direction. The on-board computer must aggregate the stream of stop, door open/close and APC sensors signals into a single stop event record. At the garage, the on-board data is downloaded to an offline computer. Matching algorithms parse the data into separate trips, identify the route and link event record to designated stops. While some studies have shown raw counts from APC equipment to be more accurate that ride checkers (Kemp, 2002), more opportunities inherently exist to introduce error in the final archived AVL/APC database. Multiple routines are required to process the data before it is in a useable format. (Figure 3)

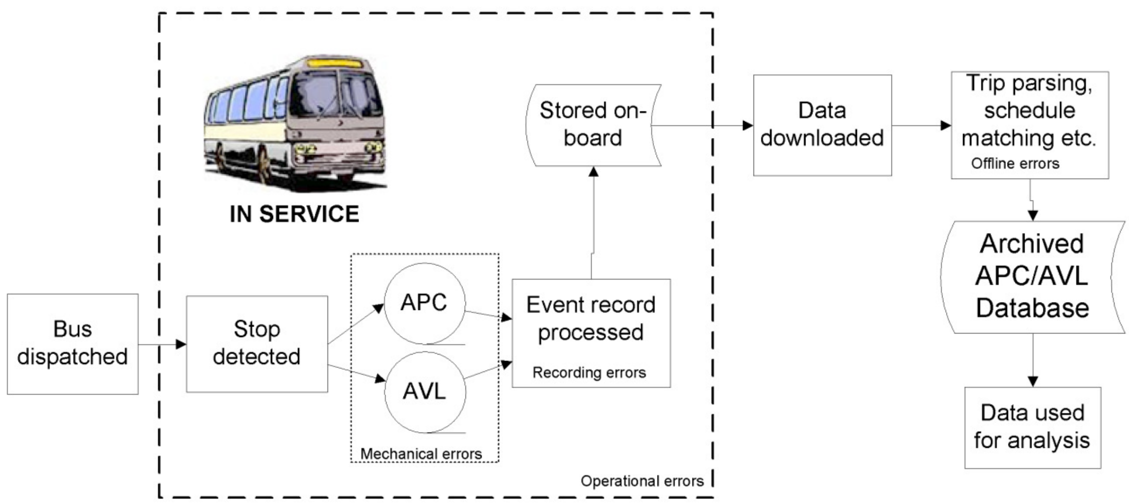

Figure 3 Data collection flow and potential for error

Error introduction opportunities are shown in Figure 3. Mechanical errors such as malfunctioning equipment can lead to the miscounting of passengers. Recording errors such as the loss of a door open signal can deactivate the APC sensor. Operational errors may cause incorrect segmentation of the data (identification of start of new trips) such as the replacement of an AVL/APC equipped bus with a non-equipped bus while in service. Poor matching algorithms can classify the wrong schedule or designated stop when the matching process is conducted offline. These algorithms further rely on good schedule and bus stop data; bad vehicle-to-route assignment files or incorrect operator sign-in can affect the results. Chapter 2 provides more details on current data management practices for their types of errors.

Larger scope of data

Another barrier to validating the APC/AVL data is that a larger scope of data is now available through automation. Transit agencies must adjust their analysis methods, which were originally designed for a data-poor environment. Oftentimes, there is a large learning curve and more training is needed (Boyle, 2008, Stratham et al, 2008, Parker, 2009). This adjustment period can require further data management expertise that can exceed agency resources. The tools for this data management can be developed in-house. However many agencies rely on vendor-supplied software because it is less onerous than exhausting agency resources (Kimpel et al, 2003).

1.2 Motivation

The popularity of archived AVL/APC data is growing, yet some transit agencies still hesitate to readily accept these technologies (Boyle, 2008). Transit agencies often face the challenge of getting staff to readily accept automatically collected data as "valid enough" for their uses (Parker, 2008). For example, some transit agencies still frequently rely on manual surveys or revenue data for ridership estimates even when an APC system is employed. The general problem appears to be that the level of routine data processing has not advanced to the point that AVL/APC data can be used with confidence, without an analyst carefully checking and adjusting it (Furth et al., 2005) and without the frequent need for external data.

Transit systems that have strived for data quality have been most successful in the implementation of APC/AVL systems (Boyle, 2008). Implementing quality assurance processes is an important first step towards maximizing the utility of archived data. However, most research of archived AVL/APC systems focuses on the development of tools for service analysis such as determining schedule adherence, run times, ridership and peak loads. TRCP Synthesis 77 *Passenger Counting Systems* calls for further research on the evaluation of techniques for data cleaning and validation. This study is intended to contribute to AVL/APC data quality research by addressing some common problems:

Though internal validation studies are common in transit agencies with AVL/APC deployment (Boyle, 2008), they are often unpublished. Transit agencies still frequently rely on manual surveys or external data sources to validate their AVL/APC data and focus more on the accuracy of raw measurements. The larger scope of data also makes validation difficult. An automated validation program is a powerful tool for data quality management. However many programs are developed by vendors or third-party contractors and these tests are not always transparent nor understood by the user. **Through the development of an automated quality assurance (QA) procedure, this research intends to improve the availability of resources for data quality management of archived AVL/APC system.**

1.3 Research goals and objective

The goal of this research is to develop a framework for identifying unreliable data collected from AVL/APC systems based solely on information contained within the archived database. This thesis has the following objectives:

1. Define an analytical methodology to apply the quality assurance (QA) framework and mathematically define the process steps;
2. Perform the QA methodology on a sample of archived APC/AVL dataset from Grand River Transit in the Region of Waterloo, Ontario, Canada;
3. Evaluate the output of the QA methodology and model robustness; and
4. Assess the impact of this QA methodology on further applications of the archived AVL/APC data.

All of these objectives help service providers and planners apply a quality assurance framework within their respective transit agencies and enhance the quality of their archived AVL/APC databases for other uses such as performance assessment, monitoring and future planning. It is important to note that the proposed procedure is meant to complement, not replace, current data processing and quality control routines.

1.4 Thesis Outline

Chapter 2 is a literature review of the current approaches to quality assurance of archived AVL/APC data. Chapter 3 describes and defines the proposed QA methodology. Chapter 4 describes the application of the methodology to a sample AVL/APC database. Chapter 5 evaluates the results of the quality assurance analysis and discusses study limitations. Chapter 6 summarizes the conclusions of this study and identifies potential future extensions to the work.

Chapter 2

Literature Review

Section 1.2 discusses data quality issues related to archived data from AVL/APC systems. This chapter discusses some of the current practices associated with the management of data quality. The first section is an overview of the current quality control and quality assurance practices by transit agencies. The second section investigates data quality management practices in other ITS examples. A summary of the findings is provided at the end with a discussion of the limitations to current practices.

2.1 Current quality management practices

Methods described below are several data quality assurance practices available to manage archived AVL/APC data quality.

Appropriate System Design

Following the recognition of different quality needs of archived transit data, new data collection processes and databases structures were proposed for AVL systems to integrate them with APC systems. Furth et al. (2004) categorize different uses for archived AVL/APC data and the necessary data requirements (Tables 3 and 4).

The quality of downstream analysis from AVL/APC data can be improved and facilitated through a proper system design, thus providing transit data with sufficient detail for practical database structure. Data detail levels are: (A) round robin polling, (B) timepoint records, (C) stop records, (D) level segment performance summary and (E) finest level of detail . Level A refers to collection at recursive time intervals (usually 40 to 120s). Level C may refer to a record at designated transit stops.

Table 3 Decision tools and data needs (Source: Furth et. al. 2005)

Function	Tool/Analysis and [Usage Code*]	Detail Level Needed	Additional Items Needed	External Data Needed
General service monitoring, including contract compliance	• Missed trips [1] • Schedule adherence [4]	A or B	*Incident codes, control messages*	Schedule
Targeted Investigations • Customer service (complaints) • Security/legal (incidents, accidents) • Operator performance	Trip investigation at gross level (was it there? was it off route?) [4]	A	*Off route, incident codes, control messages*	*Schedule, payroll*
	Trip investigation: early, late, overcrowded? [3]	C		
	Trip investigation: speed, acceleration [2]	D or E	Maximum speed; *records every 2 s or more to measure accel., decel. rates; GPS altitude*	
Scheduling and Monitoring Running Time	Route and segment running time analysis (mean and distribution) [4]	B		
	Suggesting running time based on percentiles [3]	B		
	Selecting homogeneous running time periods [3]	B		
	Suggesting half-cycle time based on percentiles [2]	B		
	Running time analysis net of holding time [2]	C	*Incident codes, control messages*	Schedule
	Speed and traffic delay [2]	D		*Schedule*
	Unsafe operations monitoring [0]	D or E	Maximum speed; *records every 2 s or more to measure accel., decel. rates*	
	Relating running time to weather, roadway incidents, and special events [1]	B		Weather, roadway incident data, special event data
Schedule Adherence and Connection Protection **(service and operational quality)**	Percent early, late by location [4]	B (timepoint-level) or C (stop level)		Schedule
	Distribution of schedule deviation at a timepoint [3]	B or C		
	Graphical display of schedule deviation distribution along a route [2]	B or C		
	Experienced lateness and earliness [1]	C		*Farebox transactions with linked trip data*
	Connection protection [1]	C	Control messages	*Farebox transactions with linked trip data*
Headway Analysis **(service and operational quality)**	Headway deviations (mean and distribution by timepoint) [3]	B (timepoint-level) or C (stop level); all buses reporting	*Incident codes, control messages*	Schedule
	Impact of headway variability on passenger waiting time for random passenger arrivals [1]	C	Incident codes	*Farebox*
	Plot successive trajectories (bunching analysis) [2]	C	*Incident codes, control messages*	Schedule
Demand Analysis	Load profile (mean ons, offs, and load by stop along a route; also passenger-miles) [4]	C		
	Load variations [3]	C		
	Analysis of trip maximum loads and maximum load points [1]	C		
	Time-dependent demand and load analysis, and suggesting trip start times to achieve load targets [1]	C		
	Analyze overload, lift, bicycle, and other events by stop and time [3]	C	Incident codes	
	Transfer and linked trip analysis [1]	C	Farebox transactions with card IDs	Farebox

NOTE: Italics indicate optional items or data. GPS = Global Positioning System. GIS = geographic information system.
*Usage codes: 4 = used commonly by agencies with AVL-APC data; 3 = used by some agencies with AVL-APC data; 2 = used by only a few agencies with AVL-APC data; 1 = used experimentally or ad hoc; 0 = not used.

(continued on next page)

Table 4 Decision tools and data needs continued

Function	Tool/Analysis and [Usage Code*]	Detail Level Needed	Additional Items Needed	External Data Needed
Geographic and Planning Analysis	Geocoding stops and other points of interest [2]	C		GIS
	Mapping bus path through shopping centers, new subdivisions, etc. [3]	E		
	Comparing measured vs. nominal stop locations [1]	C		
	Relate on and off data to demand rates in traffic analysis zones and to geographic database [1]	C		GIS, regional travel demand model database
	Relate service quality data to geographic database [1]	B or C		GIS, schedule
Utilities	Monitoring system failures [4]	A	System diagnostics	
Other Operations Analysis	Operator performance (schedule adherence, on-time start, running time, headway maintenance) [1]	B (timepoint level) or C (stop level)	*Incident codes, control messages*	Schedule, *farebox data*
	Dwell time analysis [2]	C	Passenger entry-exit moment, *farebox transactions, incident codes*	
	Layover and pull-in/pull-out analysis [0]	B	*Incident codes, control messages, off route*	Schedule
	Control effectiveness: any service quality monitoring or service analysis, related to control messages	*As required for each analysis*	Incident codes and control messages	
	• Before/after study • Special event/weather analysis	*As required by the type of analysis*	*As required by the type of analysis*	*As required by the type of analysis*
Passenger Information Monitoring	Prediction accuracy (match announced stop or predicted arrival time with actual) [1]	C	Annunciator	Schedule, GIS
	Accuracy of route data in destination sign and farebox [0]	A	Destination sign, farebox	Schedule
Payroll	Verify sign-in data [2]	A		Schedule, payroll
	Examine operator's duty when there's an overtime claim [2]	A	Off route, *incident codes, control messages*	Schedule, payroll
Maintenance Management	Analyze maintenance incidents [0]	D	Incident codes, *control messages, on and off counts, GPS altitude, vehicle health indicators*	*Maintenance, altitude*
	Monitoring vehicle demands [0]	D	*On-off counts, GPS altitude, vehicle health indicators*	Maintenance, *GIS*
	Analyze failure trends [0]	D	Incident codes, *control messages, on and off counts, GPS altitude, vehicle health indicators*	Maintenance, *GIS*
Strategic Planning	Trends analysis [2]	*As required by the type of analysis*	*As required by the type of analysis*	*As required by the type of analysis*

NOTE: Italics indicate optional items or data. GPS = Global Positioning System. GIS = geographic information system.
*Usage codes: 4 = used commonly by agencies with AVL-APC data; 3 = used by some agencies with AVL-APC data; 2 = used by only a few agencies with AVL-APC data; 1 = used experimentally or ad hoc; 0 = not used.

Timepoints (B) are a subset of designated stops (or it can also be an alternate location) associated with a scheduled arrival and departure time. Performance summaries in Level D refer to the inclusion of additional travel information in the preceding segment to a stop; time spent below crawl speed is an example performance summary. Lastly, Level E refers to a finer disaggregation of events below a stop record such as door open/close events or wheelchair lift use. The sample AVL/APC data shown in Table 1 may be considered between Level C and D, instead of assigning segment performance characteristics to the subsequent stop record, non-designated stops are recorded separately.

Manual Surveys and comparison to external data sources

Many transit agencies base their data validation on ensuring the accuracy of the AVL/APC equipment hardware and software. The general technique to address this quality perspective is to verify APC counts with another source. Comparison to manual counts remains the most common method of validation. Even where APC/AVL systems are already deployed, manual surveys are still predominantly used to supplement ridership estimates (Boyle, 2008). However manual surveys are time and labour-intensive and are often limited in sample size. Results may also be readily influenced by extreme events such as traffic accidents or severe weather conditions. The customary assumption that manual counts are 100% accurate has been challenged by some APC vendors, which insist on video surveillance verification for passenger counts (Furth et al., 2006).

Another external data source is ridership generated from revenue data (farebox), however TCRP Synthesis 34 concluded that farebox counts are shown to be less accurate than conventional APC systems. With the gradual proliferation of special fare programs using various fare media, boarding estimates using revenue-based models are becoming increasingly suspect (Kimpel et al, 2003).

Test criteria for APC equipment often fail to distinguish between random error and bias (systematic error). Bias is more serious than random measurement error, and accuracy tests should specifically check for bias. However few agencies can afford the research needed to establish the level of systematic over- or undercount (Furth et al., 2005). Instead most agencies rely on choosing a vendor with high accuracy rates or apply vendor-supplied correction factors. State-of-the-art APC systems guarantee count accuracy in the two to three percent range (Boyle, 2008).

The reliance on external data sources for validation ultimately requires additional resources and leads to increased workload for the transit agency. For example, annual validation by manual counts is required by the Federal Transit Agency (FTA) for ridership submissions based on APC/AVL systems. For a smaller to medium sized transit agency that already collects annual ridership data by manual methods, further deployment of AVL/APC systems for ridership data collection may be discouraged and instead viewed as an additional expense. Techniques of data validation that preclude external data collection efforts are more easily adopted by resource constrained agencies. Some cost data is available for manual versus APC/AVL methods, however it is difficult to compare these numbers because APC/AVL systems are often incorporated in to a larger ITS package. Boyle (2008) estimates a median capital cost of $6,638 per APC unit based on a survey of 26 transit agencies. TRCP 29 reports a median operating cost of $650,000 annually for manual methods and $90,000 annually for APC methods based on incomplete data and does not fully reflect differences among agencies in terms of size and/or varying labour cost (Boyle, 1998)

Passenger Balancing Algorithms

Passenger balancing algorithms are another technique to correct of counting error. Raw APC counts are subjected to a set of rules; counts are adjusted where the rules identify inconsistent data. A most basic test is to check for a discrepancy between the total number of boarding and alighting passengers at the trip level and block level. A block is a set of trips that is assigned to a transit vehicle. The algorithm may also apply rules to the derived load values. For example, load balancing adjusts raw counts to correct to a zero or positive load where negative loads are detected. Commonly used load balancing approaches can result in error propagation when applied at the block level. Furth et al. (2005) provide a good overview of the errors associated with APC systems and a detailed discussion about balancing algorithms. Included in their report are some key management decisions related to the storage of raw values, whether load is derived on-the-fly or on-board the transit vehicle, and when an algorithm is applied.

Each transit agency may apply a different set of rules based on assumptions about vehicle behaviour at the end of the line, zero load locations and the tolerance level for on-off discrepancies. Even if the same basic rules apply, the methods to adjust the raw values can be different between transit agencies. TRCP Report 113 provides some suggested correction methods including assigning the correction to the end (or start) of the trip, to the largest count or to distribute the correction proportionally among all stops (Furth et al., 2006)

The majority of agencies rely on the system vendor for data processing (Boyle, 2008), which frequently include passenger counting algorithms. While some more advanced agencies have developed their methods in-house , often passenger balancing algorithms are proprietary to an APC vendor. More information about transit agencies with advanced APC/AVL systems are provided at the end of this section.

Statistical analysis

Statistical summaries of ridership by route, street, stop, trip, time of day, timepoint arrival and municipalities is an advanced feature in some AVL/APC systems (Hwang et al., 2006). These summaries are sometimes useful for a data analyst to quickly identify possible errors by flagging unexpected values. For example at STM, in Montreal, each Operations Chief has the responsibility of verifying collected data. Passenger load and running time summary reports are processed within 48 hours of data download; the chiefs have the ability to temporarily set aside data that he/she believes is invalid for reasons that must be justified (Furth et al, 2003).

Where sampling plans exist, missing data can be identified by comparing planned and actual percentage of runs for data collection. Missing data is often simply omitted from the database; however Furth (2006) discusses imputed values as an alternate approach. Imputed values may be based on historical averages or on values from “similar” trips, allowing analyst to not have to deal with missing data or varying sampling rates. While imputation is practiced in other ITS databases (e.g. automatically collected traffic data), not much literature discusses its application to AVL/APC data.

Database management systems and enterprise data

AVL/APC systems for archived transit data is one example of an ITS application for public transportation; other ITS applications among the multiple business units of a transit agency are web-based trip planners and automated timesheets for human resources. These different applications typically need to receive and share data with other transit agency information technology (IT) and ITS systems and databases. Duplicative data maintenance efforts and data inconsistencies reduce efficiencies within a transit agency.

By integrating and maintaining a set of core service and operation data at the enterprise-wide perspective, a transit agency can more cost effectively realize the benefits of ITS investments (Hwang et al., 2007). Enterprise data is the name of that core service and operational data; examples of enterprise data components are schedule and stop inventory information. Proper maintenance of enterprise data is expected to increase data quality in connected ITS systems such as AVL/APC systems. For example, higher quality schedule data would result in more effective matching algorithms. Higher quality stop data would reduce errors associated with poor location attribution. The concept of enterprise data is to progress transit ITS architecture from application-centric to data-centric depicted in Figure 4 (Hwang et al., 2007). The integration of various data sources into a single transit data warehouse complements data quality management by streamlining cross-validation and facilitating the development of new quality control techniques. Data ownership for each data source is recommended to the business units most interested in its accuracy.

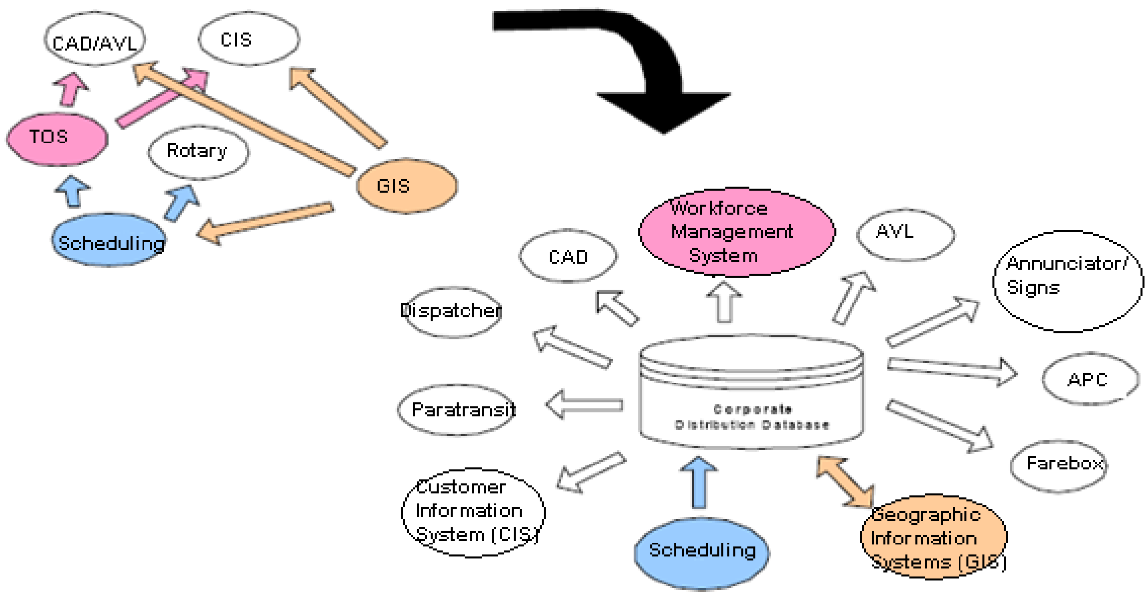

Figure 4 Evolution of transit data from application-centric to data-centric (Source: APTA)

The FTA recommends that enterprise data in transit needs to have the key technology elements in place, including: core data sets that are shared; a transportation network; commonly shared database management and reporting tools to minimize multiple learning curves and maintenance needs; a distributed logical data model; and well communicated policies and procedures (Hwang et al, 2007).

Luao & Liu (2010) develop a data processing framework specifically for transit performance analysis. The framework proposes processing APC/AVL and electronic fare payment data to exclude outliers and then cross-validating both data sources into an integrated data warehouse (Figure 5). Improved data quality is implied through use of data mining and fusion; though no details are provided on the cross-validation methods.

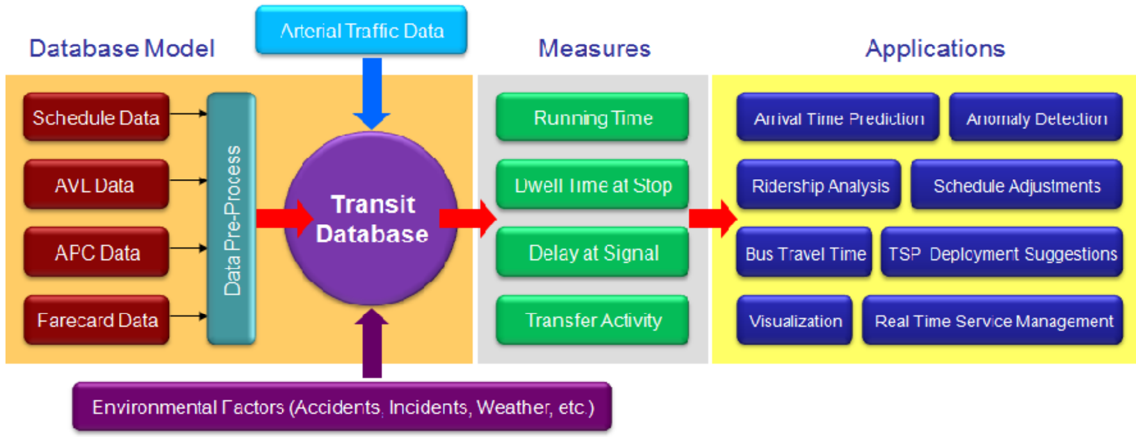

Figure 5 Data Processing Framework for Transit performance (Source: Luao & Liu, 2010)

An example of the influence enterprise data management (i.e. controlling quality of core data sets) has on archived AVL/APC data is shown for NJ Transit; it was found that if stops were within 300 ft of actual stop (determined with high quality aerial maps in GIS), then the ability to match trips increased to 81% from 65% (Furth et. al 2006). High quality GIS maps may also be a transportation network component for enterprise data. Stop inventory maintenance is considered a state-of-the-art quality management technique for transit planning and operations data. Robinson (2009) demonstrates a method to validate bus stop data for London Bus before the implementation of the current AVL system. GIS maps are used to determine bus headings and the travel distances between stops (required tracking information by the AVL system) by snapping bus stops to route on the road network. Inconsistencies between GIS-derived route information to previous schedule data are detected and corrected before AVL implementation.

Management of enterprise data demonstrates improved ITS data quality downstream; however maintenance efforts still rely on cross-validation and data ownership by individual units to ensure high quality data. Automated validation programs are an example of the processing modules applied to archived AVL/APC data to improve data quality for other planning and/or operations applications.

Automated validation programs

By looking for internal inconsistencies, validation of the archived AVL/APC data can be automated. Automated validation programs are similar to passenger balancing algorithms; a set of rules are applied to the AVL/APC data to detect suspect data. In fact, vendor-supplied automatic validation programs often include passenger balancing algorithms within their software package. In a 2008 study, 52% of surveyed transit agencies reported using an automated program to analyze APC data. Agencies reported various thresholds for determining validity; example rules are shown in Table 5 (Boyle, 2008).

The most common test shown in Table 5 is to compare total boardings against total alightings by the block or trip level. Note that if the threshold is crossed, discarding the data is recommended. However adjustments can be made to the passenger counts if below the threshold. Adjustments resulting from the first and second test are examples of how passenger balancing algorithms are typically incorporated into automated validation programs, especially when it is vendor-provided. Unfortunately, vendor software is not always transparent to the user, and it is important to understand how the validation checks work.

Table 5 Examples of automated validation program rules

Test	Threshold	Action
Boardings vs. alightings by block or by trip	5% 10% 20% 30%	Discard block or trip data if exceed threshold
Loads	Less than 0	Adjust boardings/alightings at heaviest use stops
Bus stop location	within 200 feet of actual bus stop	Flag stop data if exceed threshold
Actual vs. scheduled block miles/kilometers	10%	Discard data exceed threshold
Actual vs. scheduled block pullout/pull-in times	30 minutes	
Actual vs. scheduled trip start/end times	20 minutes "significantly off-schedule"	
Observed vs. "expected" results at the route, block, trip, and stop levels	Not specified	Assign quality code to data
Geographic information vs. computerized scheduling software data	Look for match	Assign probable route/block
Block data	No data	Discard block data

Another feature of the automated validation programs is that most tests shown in Table 5 are based on block- or trip-level summaries. Kimpel et al. (2003) recommend that data quality flags be applied to estimates of passenger activity at each summary level, not just the route by trip summary level. Schedule matching is sometimes considered a form of validation; some matching algorithms are based on identifying probable route based on spatial comparison of the raw AVL/APC data to schedule data.

In addition to matching and balancing algorithms, automated validation programs may also integrate other advanced AVL/APC features to authenticate the data. For example, Tri-Met reports exception events (i.e. when a trip deviates significantly from schedule) with real-time messages that are eventually stored. Metro transit has an automated sign-in protocol set up in their AVL/APC system, which requests manual verification of the route and run by the operator before service (Furth et al., 2006). Kings County Metro (Seattle) continuously monitors data quality as it is collected and does not store data that fall outside of defined parameters (FWHA, 2007). More examples of transit agencies with advanced AVL/APC systems are shown in the next section.

Transit Agencies with advanced AVL/APC systems

The data quality control practices of transit agencies with advanced AVL/APC systems are highlighted in TRCP Web document 23 and are summarized below. Canadian examples are Societe du Transport Montreal (STM) and OC Transpo in Ottawa. Other leading transit agencies are NJ Transit in New Jersey and TriMet in Portland, Oregon.

OC Transpo

OC Transpo is one of the earliest adopters of APC technology, first employing the technology in 1975. In 1987, a major custom-designed upgrade of the APC system with AVL technology resulted in a whole set of new tools and procedures for data matching and enhanced quality. These tools include: hardware diagnostics, automatic generation of bus-to-run assignment at start-of-day, and automatic nightly processing and checks.

Each bus is equipped with: infrared light technology; an on-board microprocessor to store passenger activity and other progression logs; a microwave receiver; and a radio control module to transfer data to the central computer. On-street, 35 microwave signposts help locate the buses accurately. The system includes 80 instrumented buses (reported in 2006). The data are transferred at night automatically from the buses to a central computer through the radio system, and a number of automatic procedures take place to sort and validate the downloaded data overnight.

Bus-to-run assignment files are used to split the data into individual trips. Nightly processing procedures check the following items against pre-set user-defined tolerance criteria a) actual versus scheduled pull-out time b) actual versus scheduled pull-in time c) total versus scheduled run length d) number and sequence of signposts passed e) difference between total ons and offs. The nightly processing also identifies suspected malfunction in a diagnostics report. More information is available in Appendix F of TRCP web document 23.

STM

The STM has a long history of using sophisticated methodologies and computerized tools for activities like scheduling and planning. The STM first became interested in using APC technology to gather service planning-related data in the early 1990s. The APC system, named SCAD, converted to infrared technology in 1996 after the legacy treadle mats were found to result in insufficient accuracy rates for low floor buses.

Automated data collection at STM has been institutionalized over many years. Sampling plans, prepared by Service Planning department are carefully reviewed by Operations Management. Processed data are posted within 48 hours to database in the form of the Passenger Load / Running Time Reports which include scheduled and real departure times, arrival times and passenger loads by timepoint. Each Operations Chief has the responsibility of verifying the collected data, and has the ability to temporarily set aside data that he/she believes are invalid for reasons that must be justified.

Data collected are downloaded every night and a first set of validation procedures are conducted automatically. These include the following routines: vehicle assignment (initial match of raw data to real runs), validation of signpost sequence, matching to stop inventory data and passenger balancing. One of STM's tests is that at the trip level, the average absolute deviation between automated and manual counts of boardings should be less than 5% of average trip boardings. Manual count surveys by ride checks are conducted once a year. Because it uses absolute deviations, this test masks systematic error. However, the strict criterion of 5% effectively forces both random and systematic error to be small. STM defines limits to the maximum measurement of count errors for a trip to be valid. Other criteria are used to reject suspect data including: +/- 10% for inter-signpost distance; over +/- 20 minutes from scheduled departure from garage; over +/- 10 passengers at terminals. However rejected data are not discarded and are later analyzed to test new balancing and matching algorithms. More information is available in Appendix E of TCRP Web document 23.

NJ Transit

NJ Transit is widely considered to have the most advanced APC system in the US (Furth et al. 2006); in addition to passenger counting, travel time analysis is enabled by inter stop events recorded at a set polling rate as well as expandable "smart-bus" on-board architecture. This design allows for future integration of other on-board technologies with a vehicle area network communication. Unlike other transit agencies, its AVL/APC system does not currently include real-time radio communication. An on-board computer serves as the event recorder and stamps stop records with time and location through its GPS receiver. The on-board computer is also connected to the APC analyser, odometer and speedometer and wireless LAN transmitter to upload data to computers at the garage.

Data are automatically uploaded nightly, matched to the schedule, and loaded into a database. Instead of operator sign-in or vehicle assignment files, schedule matching relies on externally developed APC software called "correlator" which identifies potential routes/runs by pull-in and pull-out time and the number of stops; the correlator interprets their sequence to assign trips to the appropriate route. The use of spatial analysis for matching highlights the need for good stop and schedule data.

NJ Transit applies the enterprise data concept by adopting its APC applications in a transit data warehouse. The AVL/APC data is "cleansed" by over 70 quality checks related to business rules, missing data and problem data are flagged before they are loaded into the warehouse. Though the details of the 70 quality checks are not published, it is likely that most quality checks involve cross-validation between the many related datasets within the transit data warehouse or the tally of quality checks include schedule matching and passenger balancing routines

Hardware problems are detected by reviewing the time it takes to upload and process data. When the processing time is longer or shorter than usual, a message is automatically sent to the project manager, vendor and maintenance contractor. Imbalances between on-offs of more than 5 or 10% of the total trip are automatically screened and balancing algorithms are applied. Other tests include comparing GPS displacement and odometer distances and discrepancies between GPS and clock data. More information is available in Appendix D of TRCP web document 23.

TriMet

Tri-Met has an extensive history with APC technology and successful deployment of AVL technology; many reporting and data-processing programs have been developed in-house. APC units were first installed on Tri-Met buses in 1982. In the late 1990s, AVL components were combined with a bus dispatching system upgrade. A hybrid AVL/APC system was implemented with on-board data storage of stop-referenced data. The APC component uses infrared beam technology; the AVL component uses satellite GPS units to identify location and stop-generated records for on-board storage. Schedule deviation is monitored in real-time, with an exception report automatically transmitted whenever the bus deviates from the route and when the bus is behind or ahead of schedule based on a predetermined value. At the garage, data is transferred from memory cards to an archived AVL/APC database. TriMet screens its bus APC data and deletes trip block records where the aggregate difference between boardings and alightings exceeds 10%, which should improve accuracy. For the data that pass through this initial screen, another postrecovery data processing activity involves load balancing. Load balancing corrects for the remaining differences in boarding and alighting counts to "zero out" passenger loads, usually at the trip or block level.

Partnerships with Portland State University led to numerous research efforts using APC data. Kimpel et al. (2003) assessed the validity of APC boarding, alighting and load count related to data collected from on-board cameras. The study uncovered count biases introduced by the APC equipment; a correction factor to raw counts is used to adjust for this bias.

Table 6 is a summary of the transit agencies presented in this section. More information for each transit agency is available in the Appendices of TRCP web document 23.

2.2 Other QA practices for automated data collection

While data quality issues related to AVL/APC systems are well cited, literature regarding the development of quality management plans, tools and procedures is limited. Researchers can look towards other realms of automated data collection with more advanced quality control practices for guidance. Two examples of automated data collection are ITS-generated traffic data and pavement distress data. This section describes these examples in more detail.

Quality control procedures for traffic operations data

Traffic operation systems automatically collect traffic data from network ITS sensors along high volume roads. Similar to AVL systems, ITS sensors throughout the transportation network were initially installed for monitoring or "real-time" purposes. The recognition of uses for the data in an archived format led to the development of archived data management systems (ADMS) for traffic data.

Table 6 Summary of transit agency with advanced AVL/APC systems

Agency	Technology	Data collected and frequency	Purpose	Quality Criteria and Control Practices
OC Transpo	APC - infrared sensors AVL - microwave signposts Radio control data transfer Automatic nightly download	Level segment performance Passenger log - counts by door plus time at stop, time with doors Idle log - stop even more than 45s with more passenger activity Stop and Go log - identifies segments of low average speed 9% of fleet equipped Sampling plan implemented	Multiple in planning, operation, schedule and management Service performance by route (on-time and revenue generation) Schedule adherence and high load Annual service planning Identifying locations for shelter installation Complaint program in custormer	Actual vs. Schedule pull-out time Total vs. Scheduled run length Number and sequence of signposts passed Difference between total on/offs Diagnostics report on hardware malfunctions Staff verifies cause of poor data and
STM	APC - infrared sensors AVL - radio signposts, GPS Automatic download nightly	Stop-level detail (C) with passenger counts "Idle" operation 12-15% of fleet equipped Sampling plan implemented	Increase data aceess Reduce data collection time Passenger load/run-time reports Schedule adherence Ridership profiles Service adjustment proposals	Automated validation include: vehicle assignment (trip parsing)/stop matching checking signpost sequence Passenger balancing algorithms Annual manual count survey Actual vs. Schedule pull-out time Difference between total on and offs
NJ Transit	APC - infrared sensors AVL - GPS for time/location coordinates Expandable "smart-bus" architecture No real-time data collection	Finest detail level (E), each event records include time/location stamps: Door open/close, passenger counts, Stop events and and "crawl" speed Polling between stop events Change in direction 10% of fleet equipped	Reduce data collection costs for NTD passenger miles Ridership - identify overcrowding/underutilized routes Operations - running time and on-time performance analyses Potential vehicle maintenance tool	Archived data management system Automated validation with 70 business rules: Spatial analysis for schedule matching Difference between total on/offs Comparing GPS vs. Odometer distances Marking unusual processing time
Tri-Met	APC - infrared sensor AVL - GPS satelittle Manual download nightly Integrated with Bus Dispatch System	Stop-level detail (D) with passenger counts Fleet-wide deployment 500,000 stop and event records/day	Quarterly performance reports Planning - ridership and underutilized routes Schedulers - running time analyses Operation - schedule adherence Customer service - investigate complaints	Schedule deviation monitored in real-time Exceptional events recorded Difference between total on/offs passenger balancing algorithms Equipment bias adjustment

In 2007, the Federal Highway Administration (FWHA) published a synthesis of practices and recommendation for the quality control procedures of archived traffic data (Turner, 2007). The report summarizes quality control procedures suggested in the literature as well as those that are used in numerous ADMSs. It was found that most validity criteria can be broken down into three main groups: uni- and multivariate range checks, spatial or temporal consistency and detailed diagnostics. The synthesis suggests a basic set of validity tests and the use of flags or codes to indicate failed criteria. A key recommendation from that study is to provide metadata on quality control procedures and results; the ASTM standard E2468 (Standard Practice for Metadata to Support Archived Data Management Systems) is cited as reference material. Additional related standards are ASTM E2259-03, a guide for Archiving and Retrieving ITS-Generated Data, and ASTM WK7604, specifications for a data dictionary of archived traffic data. The guide stresses thorough practices for the development information systems and maintenance of data quality through mechanisms such as retaining original source data, correcting data at the source, and constructing an audit trail. Other recommendations included: metadata on traffic sensor configuration and/or historical status; further development of spatial and temporal consistency within the ADMS; and the ability to visualize data post-validation.

Additional support tools for ADMSs from the FWHA are the publication of cross-cutting findings of several case studies and lessons learned. Quality assurance strategies from several ADMS deployments are discussed. Various quality assurance practices are represented in the study: rejection of out-of-range data for storage; imputation of missing values; detector diagnostics to flag suspect data; the ability to choose whether or not to include corrected data in user queries; and using ownership policies to manage data accuracy from multiple sources. One example of quality control criteria for a traffic ADMS is demonstrated for Kentucky.

Kentucky ADMS and quality control criteria

The Kentucky Archived Data Management Service (ADMS) was developed under the framework of the archived data user service (ADUS). The Kentucky ADMS disseminates traffic data from two earlier ITS deployments, ARTIMIS and TRIMARC. The ADMS applies a set of quality control criteria to the archived data; records that fail the criteria are flagged (Table 7).

Quality control criteria are based on logical rules such as physical constraints on the roadway and duplicate records. Imputation is performed to correct missing or erroneous values flagged by the quality control criteria. Several methods of imputation are adopted by the Kentucky ADMS: historical average, temporal interpolation, spatial interpolation, hybrid algorithm and artificial neural networks. A decision workflow was developed to help select appropriate imputation methods based on data characteristics. An additional feature of the Kentucky ADMS is that users can select whether corrected or original data is included in their query (Chen & Xia, 2007).

Table 7 Example quality control criteria used to screen traffic data (Source: Chen, 2007)

Rules	Sample Code with Threshold Values	Action
Logical consistency tests • Typically used for date, time and location. • Caused by various types of failures.	If DATE={valid date value} If TIME={valid time value} If DET_ID={valid detector location value}	• Write to off-line database and/or remove records with invalid date, time or location values.
Controller error codes • Special numeric codes that indicate that controller or system software has detected an error or a function has been disabled.	If VOLUME={code} or OCC={code} or SPEED={code} where {code} typically equals "-1" or "255"	• Set values with error codes to missing/null, assign missing value flag/code.
No vehicles present • Speed values of zero when no vehicles present • Indicates that no vehicles passed the detection zone during the detection time period.	If SPEED=0 and VOLUME=0 (and OCC=0)	• Set SPEED to missing/null, assign missing value code • No vehicles passed the detection zone during the time period.
Consistency of elapsed time between records • Polling period length may drift or controllers may accumulate data if polling cycle is missed. • Data collection server may not have stable or fixed communication time with field controllers.	Elapsed time between consecutive records exceeds a predefined limit or is not consistent	• Use flow rate rather than actual counts for VOLUME.
Duplicate records • Caused by errors in data archiving logic or software process.	Detector and date/time stamp are identical	• Remove/delete duplicate records.
Maximum volume • Traffic flow theory suggests a maximum traffic capacity.	If VOLUME > 750 per lane	• Assign QC flag to VOLUME, write failed record to off-line database, set VOLUME to missing/null.
Consecutive identical volume values • Research and statistical probability indicates that consecutive runs of identical data values are suspect. • Typically caused by hardware failures.	No more than 8 consecutive identical volume values	• Assign QC flag to VOLUME, OCCUPANCY and SPEED; write failed record to off-line database; set VOLUME, OCCUPANCY and SPEED to missing/null
Maximum occupancy • Empirical evidence suggests that all data values at high occupancy levels are suspect. • Caused by detectors that may be "stuck on."	If OCC > 80%	• Assign QC flag to VOLUME, OCCUPANCY and SPEED; write failed record to off-line database; set VOLUME, OCCUPANCY and SPEED to missing/null
Minimum speed • Empirical evidence suggests that actual speed values at low speed levels are inaccurate.	If SPEED < 5 mph	• Assign QC flag to SPEED, write failed record to off-line database, set SPEED value to missing/null
Maximum speed	If SPEED > 80 mph	• Assign QC flag to SPEED, write failed record

Quality assurance programs for pavement distress data

Essentially all North American highway agencies are collecting and using pavement condition data through some automated means. For example, digital image technology is used to conduct surface distress surveys and electronic sensors collect longitudinal and transverse profile, roughness indices, rut-depth and joint-faulting measurements. Automation in the context of pavement cracking data involves the use of digital recognition software capable of recognizing and quantifying variations in grayscale that relate to striations (sometimes cracks) on a pavement surface (McGhee, 2004).

The main techniques used for pavement data quality management are: calibration of equipment before data collection; control site testing before data collection; and software routines for checking the reasonableness, consistency, and completeness of the data (Flintsch & McGhee, 2009). However due to the temporal nature AVL/APC data, calibration and control testing before each data collection session is infeasible; this technique is more appropriate for the entity-based pavement infrastructure.

Instead quality assurance practices with respect to software processing errors are relevant to AVL/APC systems. Some automated validation programs exist for pavement condition data among several state DOTs. Software programs used for quality management usually search for data that are missing, misidentified, incorrect with respect to segment size, improperly formatted, and/or outside of expected ranges (Flintsch & McGhee, 2009). Wolters, McGovern & Hoerner (2006) discuss the development of an automated quality assurance (QA) tool for pavement condition data in the Oklahoma Department of Transportation. As with many state DOTs, collection of pavement condition data is outsourced to vendors with automated data collection technology. The QA tool helps identify potential data quality problems to the vendor before the data is accepted within the pavement management database.

Another important lesson that can be learned from the pavement data example is the progressive approach to developing formal data quality management plans. Based on a NCHRP Synthesis 401 Quality Management of Pavement, the majority of highway agencies (62%) in the US have a formal data quality management plan. In addition to providing a list of quality control tools/techniques, a plan describes the quality policies and procedures; areas of application; and roles, responsibilities, and authorities.

2.3 Summary

Hybrid AVL/APC systems came about when transit agencies began to recognize different offline uses for AVL systems combined with APC technology. Along with the merger of these two technologies came the realization of different data quality needs. Design guidelines for archived AVL/APC systems were proposed to address some of these issues.

The transition from a data-poor to a data-rich environment prompts many public transportation groups to adjust their current operations and planning practices. However the main problem seems to be getting agency staff to accept automatically collected data as “valid enough” for their purposes (Parker, 2008). Many transit agencies tend to focus on assessing the accuracy of AVL/APC equipment as a form of validation. Unfortunately, this approach requires an external data source to verify AVL/APC collected data; manual surveys are the most common validation technique. Though calibration or periodic assessment of AVL/APC equipment is recommended, manual data collection is resource intensive and undermines the value provided by AVL/APC systems. Alternative methods exist such as: passenger balancing algorithms and cross-validation through route matching. Both processes are commonly incorporated parts of automatic validation programs along with other logical tests. One limitation of automatic validation programs is that the routines are often proprietary to suppliers. Vendor software is not always transparent to the user, and it is important to understand how the validation checks work (Boyle, 2008).

Another higher-level approach to improving data quality of AVL/APC systems is its integration within an archived data management system (ADMS) and the development of enterprise data. Enterprise data consist of a shared set of core transit data such as schedule and stop inventory data. Proper maintenance of enterprise data leads to improved data quality of all applications within the ADMS, including AVL/APC data. Other features of enterprise data include a transportation network; management and reporting tools; a logical data model; and related policies and procedures (Hwang et al., 2007). In this perspective, automated validation programs are part of the AMDS as a data quality management tool.

Advancement of data quality concepts and quality control practices may be guided by examining other examples of automated data collection. Two examples are discussed: archived ITS-generated traffic data and automated pavement condition data. From the traffic data example, test criteria for automated validation programs are well developed and basic tests are suggested by a FWHA synthesis of quality control procedures. Other key recommendations are metadata on quality control processes. Standards and guidelines have already been developed for the archived of ITS-generated traffic data. Although imputation is a common data correction practice, its application for archived AVL/APC data is limited and not included in this research.

From the pavement data example, proper documentation is emphasized for the development of automated quality assurance procedures. Turner (2007) points out that automated validation (QA checks) are just one component to a comprehensive quality plan.

This research aims to develop an automated quality assurance procedure to flag suspect data from archived AVL/APC systems without the need of external data. The procedure is meant to complement, not replace, automated validation programs provided by the vendors (or developed in-house) by providing users with auxiliary quality tests based on expected data patterns. The rationale and mathematical definitions of each test are to be documented clearly and transparently. The user has indirect control over the resulting data by specifying the parameter values or by disregarding certain tests.

Chapter 3

Quality Assurance Methodology

The purpose of this chapter is to describe the quality assurance (QA) methodology of archived AVL/APC data. This methodology is constructed on the following principles:

1. Decrease dependence on vendor automatic validation programs through supplementary QA quality tests;
2. Increase transparency to users by providing more detailed documentation of QA tests;
3. Reduce the need for external data sources for data validation by building a method based solely on data contained within a standard AVL/APC archive; and
4. Maintain a general QA framework by designing the method based on universal or common database attributes of archived APC/AVL systems.

This methodology is intended to complement, not replace, existing validation methods. Schedule matching and passenger balancing algorithms should already be applied to AVL/APC data before this QA procedure.

3.1 Approach and rationale

Many data quality issues may arise from the transformation of raw sensor data to stop-level and trip-level data records. The proposed method was developed by evaluating how errors in the data collection process would present themselves in the archived AVL/APC database. These errors are assumed to propagate in the data collection process and manifest themselves in three forms:

1. data is missing or unavailable;
2. attributes appear as outliers; or
3. erroneous data remains undetectable in current database structure.

Missing data

Without a sampling plan, missing data at the block or trip level is not easily detectable within a standard APC/AVL setup. Identifying missed trip records would require the scheduled trip plan of equipped vehicles, then matching it to imported data. Most systems conduct schedule matching in the reverse direction; recorded data are linked to a schedule after it is downloaded from equipped vehicles and imported to the archived database. Therefore, block or trip-level missing data are not identified by this methodology.

Stop-level missing data are sometimes conveyed in a matching quality attribute; the number of recorded scheduled stops is compared to the designated number of stops on a route. However, this trip-level attribute does not convey a corrective action. This methodology intends to classify data as suspect or non-suspect only at the trip-level; the distinction of data quality as either good or bad suggests a clear instructive on which data to include or reject for analytic purposes.

Outlier data

This methodology concentrates on identifying erroneous data by screening for outliers because they are more readily visible in the database. The concept is that if errors exist in the data; then these errors are anticipated to result in outliers of key attributes of the database.

Errors in the recorded travel time (arrival or departure) and distance values are anticipated to result in outliers of travel pattern characteristics. Passenger count errors are anticipated to result in outliers of passenger activity or count corrections from balancing algorithms. (Balancing algorithms are assumed to be applied to passenger count data during standard AVL/APC data processing software). Poor schedule matching results may stem from incorrect attribution of stop type, designated stop or specified route. Validation of these data attributes is based on analyzing schedule deviation outliers.

Based on the errors discussed above, the methodology focuses on validating these general stop-level data attributes:

1. Type of stop (scheduled stop versus non-scheduled)
2. Time of stop (arrival times and departure times)
3. Distance travelled (derived from odometer readings)
4. Passenger count (e.g. boarding, alighting and load)
5. Location (identified scheduled stop)

Instead of simply removing outlier data, further pattern identification is used to determine if a valid explanation exists to justify an outlier. Another distinguishing feature of the proposed methodology is the use of stop-level tests to screen for outliers; suspect data are then flagged at the trip level. Existing automated validation programs more often use trip or block-level thresholds.

Non-outlier erroneous data

There is likely no way to identify the data as erroneous when outliers cannot be detected without the need for external data. However if the erroneous data do not result in outliers, there is little consequence to including those erroneous data within the AVL/APC database. Most service analysis depends on identifying exceptional activity in transit operations data (high ridership and poor performing routes).

Even in analyses that do not depend on exceptional data, the inclusion of non-outlier erroneous data is not expected to greatly impact the analysis results. Or, the impact of the erroneous data is not considered too severe (e.g. run time analysis: average travel time calculations are not expected to change significantly when a portion of the sampled trips have non-extreme, but incorrect travel time data). Therefore, the methodology can only recognize erroneous data with extreme values.

User input

The approach of this methodology is to identify outliers and assume these data are erroneous unless a valid explanation can be found to explain the data. The degree to which this approach is conservative is determined by parameter values; these values can be modified by the user. There are four potential outcomes of applying the QA procedure (Figure 6).

The objective of the QA procedure is to classify erroneous data as "suspect" and non-erroneous data as "non-suspect" (i.e. maximize the trips associated with Case A and D); thus suspect data can be rejected for analytical purposes. Type I errors occur when non-erroneous data are considered suspect (Case B), also known as a false positive. Type II errors occur when erroneous data are considered non-suspect (Case C); also known as false negative.

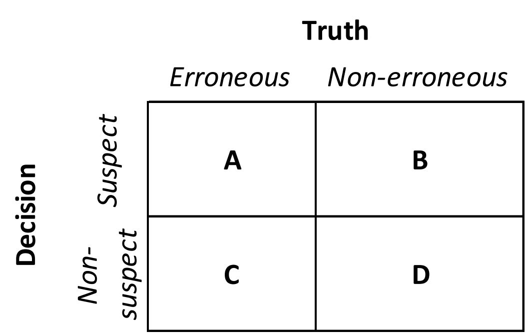

Figure 6 Potential outcomes of QA procedure

Conservative users are typically more concerned about excluding Type II errors within their sample data; these users can apply more stringent parameters for identifying outliers. Users whom are aggressive for a larger data sample tend to focus on avoiding Type I errors; less stringent parameters may be applied to satisfy these needs.

3.2 Outline of the methodology

While individual tests are conducted on a stop-level, data are flagged as suspect at the trip-level. Recall Section 1.1.1 where the data collection process is described. Table 1 and Table 2 are examples of stop-level and trip-level data, respectively. Many automated validation programs apply tests at the trip-level (as shown in Table 5). A trip-level test is applied to data at the trip-level. For example, the the recorded start and end times may be compared to the schedule and the trip-level test can screen those with large discrepancies. Stop-level data may also be aggregated over the trip (e.g. total boardings and total alightings) and the trip-level summaries may be examined in a trip-level test.

Stop-level tests are applied to individual records such as the passenger count, distance travelled or arrival and departure times for each stop event. One limitation of this methodology is that the behaviour of previous and subsequent trips is not incorporated in the tests because they cannot be easily identified in a standard AVL/APC database.

The QA methodology is broken down into three stages shown in Figure 7:

i) Base Checks
ii) Outlier Identification and
iii) Valid Outlier Identification.

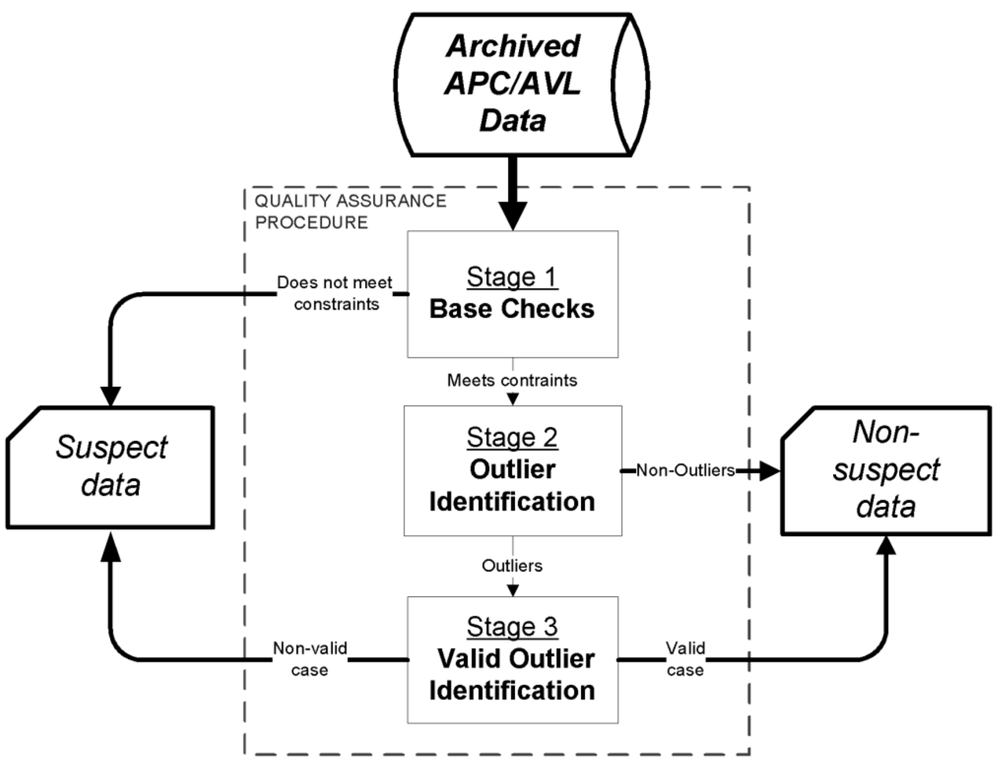

Figure 7 High-level schematic of the QA procedure

Base checks tests for physical constraints in the travel characteristics of a trip. If a trip fails any test in the Base Check (BC) stage, it is immediately flagged as a suspect trip. Outlier Identification (OI) screens key attributes in the APC/AVL database for outliers. If no stop-level outliers are found then the QA procedure is complete for the given trip. Trips identified in the OI stage continue to the third step: Valid Outlier Identification (VOI). In this step, the stop-level travel patterns are examined for valid case options. If a valid case is found, then a screened trip may be considered “non-suspect”. However, if no valid case is found for a screened trip, then its status is set to “suspect”.

This methodology identifies and investigates outliers associated with a divergence from expected travel or passenger activity patterns. The rationale and definition for each individual QA test is described in Sections 3.2.2 to 3.2.4. Detailed data definitions are shown in the next section.

As mentioned, suspect data is flagged at the trip level although the QA tests are conducted at the stop-level. A binary variable may be defined to represent the individual tests at the stop-level and trip-level, respectively:

$$\delta_{i,j}^{k} = \begin{Bmatrix} 0 \ \text{fail} \\ 1 \ \text{pass} \end{Bmatrix} \tag{1}$$

$$\delta_{j}^{k} = \begin{Bmatrix} 1 & \text{if } \delta_{i,j}^{k} = 1 \ \forall \ i \\ 0 & \text{otherwise} \end{Bmatrix} \tag{2}$$

where δ represents the binary variable (0 means fail and 1 means pass);
k represents the test (e.g. BC1, OI3, VOI5 etc.);
i represents a stop record starting from 1 to n_j (number of stops for a given trip j); and
j represents a trip from 1 to $N^k{}_T$ (total number of trips being tested in a given test).

The set of trips that pass or fail a given test may be defined by the following:

$$S_P^k = \{\text{set of j where } \delta_j^k = 1\} \tag{3}$$

$$S_F^k = \{\text{set of j where } \delta_j^k = 0\} \tag{4}$$

S_P^k and S_F^k are mutually exclusive and therefore the relationship between the number of trips subject to a given test and the number that passes or fails is given by:

$$N_T^k = N_P^k + N_F^k \tag{5}$$

$$N_P^k = \sum_{j=1}^{N_T^k} \delta_j^k \tag{6}$$

where $N^k{}_T$ is the number of trips that are subject to test k;
$N^k{}_P$ is the number of passing trips; and
$N^k{}_F$ is the number of failing trips.

3.2.1. Data Definition

The passenger count components of AVL/APC data are defined by the following variables:

$P_{Bi,j}$ = number of passengers boarding at stop i, trip j;
$P_{Ai,j}$ = number of passengers alighting at stop i, trip j; and
$P_{Li,j}$ = load derived from balanced passenger counts for stop i, trip j.

Figure 8 depicts the passenger data components in a passenger profile.

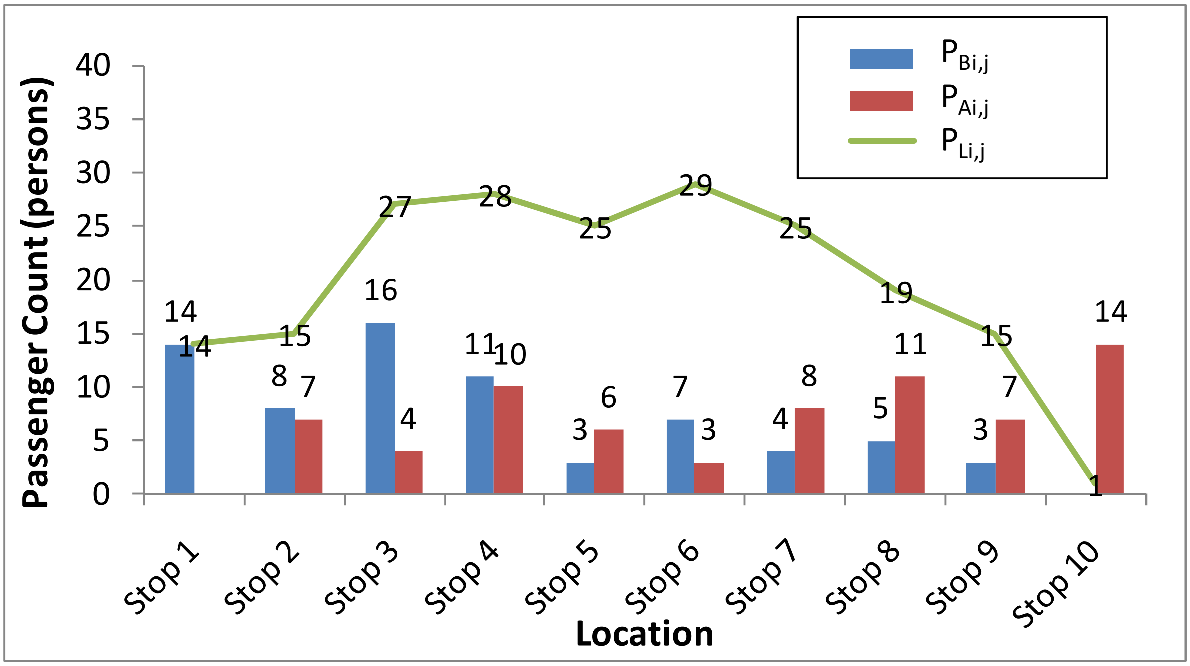

Figure 8 Passenger profile depicting AVL/APC data

It is assumed that standard data processing routines include balancing algorithms; available raw count data (i.e. not subject to balancing algorithms) is defined by:

$P'_{Bi,j}$ = raw number of passenger boarding for stop i, trip j; and
$P'_{Ai,j}$ = raw number of passengers alighting for stop i, trip j j.

The travel characteristic components of the AVL/APC data are defined by the following variables:

$A_{i,j}$ = actual arrival time at stop i , trip j, measured in seconds from midnight of the start date;
$D_{i,j}$ = actual departure time at stop i , trip j, measured in seconds from midnight of the start date; and
$Dist_{i,j}$ = actual cumulative distance traveled at stop i , trip j, in metres from first stop.

An example of arrival time of a transit vehicle arriving at 1:32:24 pm is 13h × 3600s + 32 × 60s + 24s = 48,744s. For the case when trips occur through midnight, time is associated relative to the start date. Therefore a trip ending at 12:02:32AM the following day would have arrival time = 24h× 3600s + 2 × 60s + 32s = 86,552s. Cumulative distance for each stop can be derived by odometer readings by subtracting the value from the first stop.

Expected travel activity is depicted generally by a time-space diagram; Figure 9 depicts the travel component data for a trip in red along with its associated schedule in black.

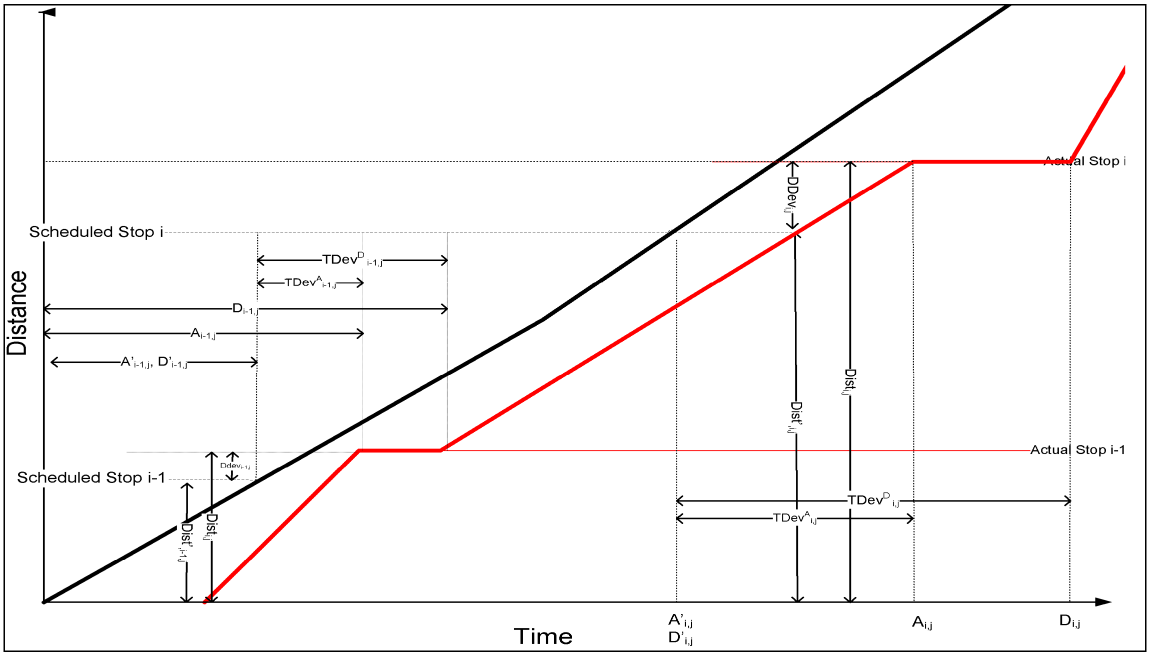

Figure 9 Time-space diagram depicting AVL/APC data

The AVL/APC data attributes shown in Figure 9 can in defined for the schedule associated with a trip:

$A'_{i,j}$ = scheduled arrival time at stop i , trip j, in total seconds of the day;
$D'_{i,j}$ = scheduled departure time at stop i , trip j, in total seconds of the day;
$Dist'_{i,j}$ = scheduled cumulative distance travelled at stop i , trip j, in metres;
$TDev^{A}_{i,j} = A_{i,j} - A'_{i,j}$, scheduled arrival time deviation, in seconds;
$TDev^{D}_{i,j} = D_{i,j} - D'_{i,j}$, scheduled departure time deviation, in seconds; and
$DDev_{i,j} = Dist_{i,j} - Dist'_{i,j}$, scheduled distance deviation, in metres.

Except for terminal and layover stops, schedule data often list the arrival and departure as the same time due to the level of detail involved in planning route schedules. These attributes are shown at same time in Figure 9, however they are defined separately. Appendix A contains time-space diagrams of the expected travel patterns for trips that fail BC tests and for trips that are screened for outliers in OI tests. Expected travel patterns are also shown for valid cases as tested in the VOI stage.

3.2.2. Base Checks (BC)

Tests incorporated into the Base Checks (BC) stage involve examination of the travel characteristics for each trip **(Figure 10)**. There are four tests:

1. BC1 tests recorded time values for increments in a positive direction;
2. BC2 tests recorded distance values for increments in a positive direction;

3. BC3 checks each that the time increment and distance increment between subsequent stops do not exceed physical constraints; and
4. BC4 checks that the transit vehicle does not exceed speed constraints.

All four tests are applied to the entire data set because if a trip fails any of the four base checks, then it is considered suspect. Each test is described in more detail below.

In Figure 10, the rectangles represent each test and the cards represent the results of each test. Bolded cards represent the set of trips that pass (S^k_P); recall k signifies the specified test (BC1, BC2, BC3 or BC4). Dashed cards represent the set of trips that fail (S^k_F). Each test in the BC stage is applied to all the trips in the database (i.e. $N_T^{BC} = \{|S|\}$). The mathematical expressions for each individual BC tests are discussed below.

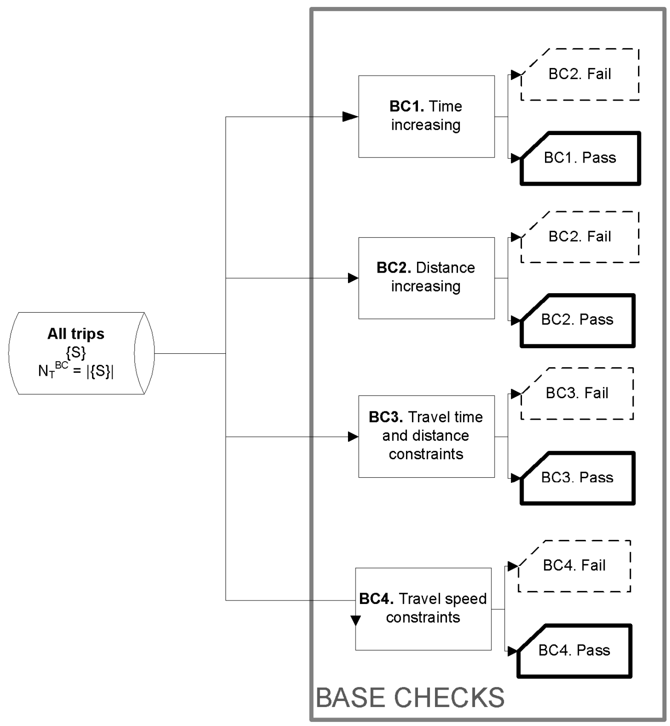

Figure 10 Schematic of Base Checks

BC1

Test BC1 checks that subsequent time records increase. Figure A1 in Appendix A demonstrates a trip failing BC1. The binary test variable for BC1 is:

$$\delta_{i,j}^{BC1} = \begin{cases} 0 & \text{if } \left(A_{i,j} - D_{i-1,j}\right) < 0 \; AND \; 1 < i \le n_j \\ & \text{OR } \left(D_{i,j} - A_{i,j}\right) < 0 \; AND \; 1 < i \le n_j \\ & \text{OR } \left(A_{i,j} - A_{i-1,j}\right) < 0 \; AND \; 1 < i \le n_j \\ & \text{OR } \left(D_{i,j} - D_{i-1,j}\right) < 0 \; AND \; 1 < i \le n_j \\ 1 & \text{otherwise} \end{cases} \tag{7}$$

Equation 7 implies that subsequent recorded arrival and departure times should be larger (later) than previous arrival and departure times for the same trip; any stop record that shows a backwards (or negative) increment is invalid. The first stop (i.e. i = 1) is excluded from the test because no previous stop record exists.

BC2

Test BC2 checks that recorded distances increase in following records. Figure A2 in Appendix A shows an example of a failed trip for test BC2. The test variable for BC-2 is defined as:

$$\delta_{\mathrm{i,j}}^{\mathrm{BC2}} = \begin{Bmatrix} 0 \text{ if } \left(\mathrm{Dist}_{\mathrm{i,j}} - \mathrm{Dist}_{\mathrm{i-1,j}}\right) < 0 \; AND \; 1 < i \leq \mathrm{n_j} \\ 1 \qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{Bmatrix} \tag{8}$$

Equation 8 implies that the recorded distance at a stop should be larger (farther) than previous distance stop records for the same trip; any stop record that shows a decrease in distance travelled is invalid.

BC3

Test BC3 applies constraints to the travel time and distance between consecutive stops. (See Figure A3 in Appendix A). The BC3 test variable is:

$$\delta_{\mathrm{i,j}}^{\mathrm{BC3}} = \begin{Bmatrix} 0 \;\text{ if } \qquad \left(\mathrm{A}_{\mathrm{i,j}} - \mathrm{D}_{\mathrm{i-1,j}}\right) \geq \mathbf{P1} \text{ AND } 1 < \mathrm{i} \leq \mathrm{n_j} \\ \text{OR } \left(\mathrm{Dist}_{\mathrm{i,j}} - \mathrm{Dist}_{\mathrm{i-1,j}}\right) \geq \mathbf{P2} \text{ AND } 1 < i \leq \mathrm{n_j} \\ 1 \qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{Bmatrix} \tag{9}$$

where P1 = Max Time Increment, maximum reasonable travel time between stops* (seconds); and

P2 = Max Distance Increment, maximum reasonable travel distance between stops* (metres).

*Note that between stops refers to two consecutive stop events for scheduled stops on the route; the scheduled stop events include passing a scheduled stop without physically stopping. Values must be specified for parameters P1 and P2; the selection for values of the parameters is discussed in the Chapter 4. In general, it would be reasonable to assume that the upper bound for travel time between consecutive stops (P1) should be less than the one-way cycle time for a route because for each route, trips are separated for each direction (previous or subsequent trips cannot be easily identified as mentioned in Section 3.2).

Though travel time between stops can greatly vary, the recorded travel distance between two stops is not expected to change significantly even via alternate paths to the next designated stop. The AVL/APC database is expected to contain an event record for each designated stop of a route even if the bus skips stops. Therefore, trips failing the distance component of test BC3 would suggest a missed event record. Missed event records constitute incomplete trip data that would be unsuitable for analysis (e.g. missing stop records can alter passenger load, which is derived from the on-off differences at each stop).

Therefore, the upper bound for P2 is not set to the one-way route distance of the route. It is instead set to the longest distance travelled between any two stops on the route. If a travel time or distance is greater than the given parameter thresholds, it is considered invalid.

BC4

Test BC4 checks the travel speeds between stops and compares them to a given threshold (Figure A4 in Appendix A):

$$\delta_{\mathrm{i,j}}^{\mathrm{BC4}} = \begin{cases} 0 & \text{if } \frac{(\mathrm{Dist_{i,j}} - \mathrm{Dist_{i-1,j}})}{(\mathrm{A_{i,j}} - \mathrm{D_{i-1,j}})} \geq \mathbf{P3} \text{ AND } (\mathrm{A_{i,j}} - \mathrm{D_{i-1,j}}) > 0 \text{ AND } 1 < i \leq n_j \\ 1 & \text{otherwise} \end{cases} \quad (10)$$

where $\frac{(\mathrm{Dist_{i,j}} - \mathrm{Dist_{i-1,j}})}{(\mathrm{A_{i,j}} - \mathrm{D_{i-1,j}})}$ = travel speed for the segment previous to stop i;

$(\mathrm{A_{i,j}} - \mathrm{D_{i-1,j}}) > 0$ screens for undefined values of travel speed; and

P3 = Maximum Speed, the limiting speed of the transit vehicle (m/s).

Equation 10 shows that the speed of a transit vehicle is physically constrained; records are invalid if it shows a travel speed greater than the given threshold.

BC Output

In combination with Equation 2, a trip may fail any of the BC tests if at least one stop-level test fails according to Equations 7-10; the result is the set of trips that fail the BC stage:

$$\mathrm{S_F^{BC}} = \{\text{set of trips j where } \prod_{\mathrm{k=BC1}}^{\mathrm{BC4}} \delta_{\mathrm{j}}^{\mathrm{k}} = 0 \ \forall \ \mathrm{j}\} \quad (11)$$

where $\mathrm{S_F}^{\mathrm{BC}}$ = the set of trips which fail the BC stage. Alternately, for trips to pass the BC stage, they must pass all BC tests. This rule can be represented by:

$$\mathrm{S_P^{BC}} = \{\text{set of trips j where } \prod_{\mathrm{k=BC1}}^{\mathrm{BC4}} \delta_{\mathrm{j}}^{\mathrm{k}} = 1 \ \forall \ \mathrm{j}\} \quad (12)$$

where $\mathrm{S_P^{BC}}$ = the set of trips that pass the BC stage. This set of passing trips at the end of the BC stage is the set used as the input to the next stage (OI) and the set of failing trips is included into the set of suspect trips.

3.2.3. Outlier Identification (OI)

Outlier identification (OI) focuses on 4 attributes of travel and passenger activity of the transit vehicles: passenger counts, schedule time deviation, schedule distance deviation and passenger count correction. These attributes are derived from data fields: passenger boardings, passenger alightings, load, recorded arrival and departure times, scheduled arrival and departure times, and raw boarding and alighting counts. Figure 11 is a schematic of the OI sub-procedure.

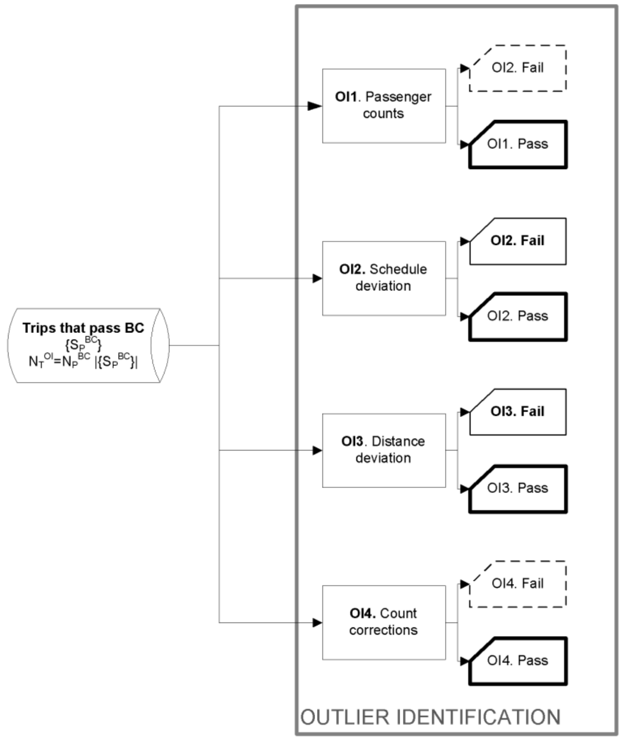

Figure 11 Schematic of Outlier Identification

Figure 11 shows that inputs of data into the OI stage are the set of trips that pass the BC stage (S_P^{BC}). Each trip in this set is screened for outliers according to the following four tests:

1. OI1 filters trips where the passenger counts are greater than the bus capacity;
2. OI2 filters trips that observe a large schedule deviation in time;
3. OI3 filters trips that observe a large schedule deviation in distance travelled; and
4. OI4 filters trips with large corrections to the raw passenger count value.

OI1

Test OI1 examines passenger counts (Figure A5 in Appendix A) and is defined by the variable:

$$\delta_{i,j}^{OI1} = \begin{cases} 0 & \text{if } P_{B_{i,j}} \geq \mathbf{P4} \\ & \text{OR } P_{A_{i,j}} \geq \mathbf{P4} \\ & \text{OR } P_{L_{i,j}} \geq \mathbf{P4} \\ 1 & \text{otherwise} \end{cases} \quad (13)$$

where

P4 = Maximum Passenger Count, maximum value for a reasonable passenger count (persons).

P4 is synonymous with the maximum bus capacity; therefore the selection of P4 is associated with maximum loads. However it is also not expected that the number of passengers that board or alight the transit vehicle should exceed the bus capacity (i.e. it would require that some boarding passengers then alight due to space constraints or that more passengers alight then can be situated on the bus); it is considered an unusual and highly unlikely event. Therefore boarding and alighting counts are also incorporated in test OI1. Any stop with passenger counts larger than P4 is considered invalid.

OI2

Figure A6 in Appendix A demonstrates how outliers in schedule time deviation are detected. Test OI2 looks at the schedule time deviations of both arrival and departure times:

$$\delta_{i,j}^{OI2} = \begin{Bmatrix} 0 \text{ if } \left|\mathrm{TDev^A}_{i,j}\right| \geq \mathbf{P5} \text{ AND } 1 < i \leq n_j \text{ AND } i \in \{i_R^t\} \\ \text{OR} \left|\mathrm{TDev^D}_{i,j}\right| \geq \mathbf{P5} \text{ AND } 1 \leq i < n_j \text{ AND } i \in \{i_R^t\} \\ 1 \qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{Bmatrix} \tag{14}$$

where $\mathrm{TDev^A}_{i,j}$, $\mathrm{Dev^D}_{i,j}$ = arrival and departure time deviations, respectively, as previously defined in Section 3.2.1 at stop *i* and trip *j*;
P5 = Max Time Deviation, maximum reasonable schedule time deviation (metres); and
$i^{t\text{-}}_R$ = set of all time points on bus route R.

Since schedule time deviation can only be calculated where scheduled times are available, this test can only be applied to time points signified by i^t. Schedule information is considered irrelevant for the arrival time at the first stop and the departure time at the last stop; these cases are excluded from test OI2.

OI3

Figure A7 in Appendix A demonstrates how outliers in schedule distance deviation are detected. The test for OI3 is defined by this variable:

$$\delta_{i,j}^{OI3} = \begin{Bmatrix} 0 \text{ if } \left|\mathrm{DDev}_{i,j}\right| \geq \mathbf{P6}) \\ 1 \qquad\qquad \text{otherwise} \end{Bmatrix} \tag{15}$$

where P6 = Max Distance Deviation, maximum reasonable schedule distance deviation (metres).

OI4

The last OI test, OI4, screens stop records where “large” corrections were applied during passenger count balancing:

$$\delta_{i,j}^{OI4} = \begin{Bmatrix} 0 \quad \text{if } \left|P_{B_{i,j}} - P'_{B_{i,j}}\right| \geq \mathbf{P7} \\ \text{OR} \left|P_{A_{i,j}} - P'_{A_{i,j}}\right| \geq \mathbf{P7} \\ 1 \qquad\qquad\qquad \text{otherwise} \end{Bmatrix} \tag{16}$$

where P7 = Max Count Correction, maximum reasonable passenger count correction (persons).

It is considered suspicious when there is a substantial difference between the raw and balanced passenger counts. It is important to apply this test at the stop-level because large stop-level corrections can be masked by trip-level aggregation. Load values were not tested in OI4 because they are calculated and not part of the raw data set. Secondly, differences between the balanced load and the load derived from raw count may be artificially larger as a result of additive corrections in the boarding and alighting count.

OI Output

Similar to the BC stage, the set of trips that pass the outlier identification stage is:

$$S_P^{OI} = \{\text{set of trips j where } \prod_{k=OI1}^{OI4} \delta_j^k = 1 \ \forall \ j\} \qquad (17)$$

If a trip passed all tests in the BC and OI stages, no other tests are applied and they are included in the set of non-suspect trips. Those set of trips that fail an OI test (screened due to a stop-level outlier) continue to the third VOI stage.

3.2.4. Valid Outlier Identification (VOI)

The VOI stage investigates specific valid case options based on the type of outlier identified. This sub-procedure targets those trips that contain schedule deviation outliers by time (test OI2) and/or distance travelled (test OI3). Currently the methodology does not account for potential valid cases of outlier passenger activity (test OI1) or passenger count correction (test OI4). No valid patterns were recognized for these two outlier types during the development of the QA procedure. Therefore, the set of trips that fail tests OI1 or OI4 are immediately incorporated into the set of suspect trips.

For trips identified with outliers in scheduled time deviations (i.e. fail test OI2), valid cases are: 1. congestion or operational delay occurring over the entire trip; 2. congestion or operational delay over a portion of the trip; and 3. “incidents” involving the transit vehicle such as traffic accident or break-downs.

For trips identified to have distance deviation outliers (i.e. fail test OI3), there is one valid case: detours. The VOI procedure also checks for 3 invalid data patterns: trips mis-matched to the wrong schedule; trips that contain mis-matched stop locations; and trips that contain a single stop event record and results in a schedule deviation. The first option is associated with time deviation outliers, the second option is associated with distance deviation outliers and the last option may occur for both. Figure 12 is a schematic of the VOI sub-procedure.

In Figure 12, bolded boxes represent the tests that identify valid case options. The dashed line represent suspect trips: those in white are suspect due to a specific invalid case pattern and those highlighted in grey are suspect due to the lack of valid case identification. The tests included in the VOI procedure are generally numbered in sequence of test order. The tests applied to trips with time deviation outliers are: VOI0 to VOI5 inclusive. The tests applied to trips with distance deviation outliers are: VOI6 and VOI7. The following sections describe the VOI tests and their mathematical form.

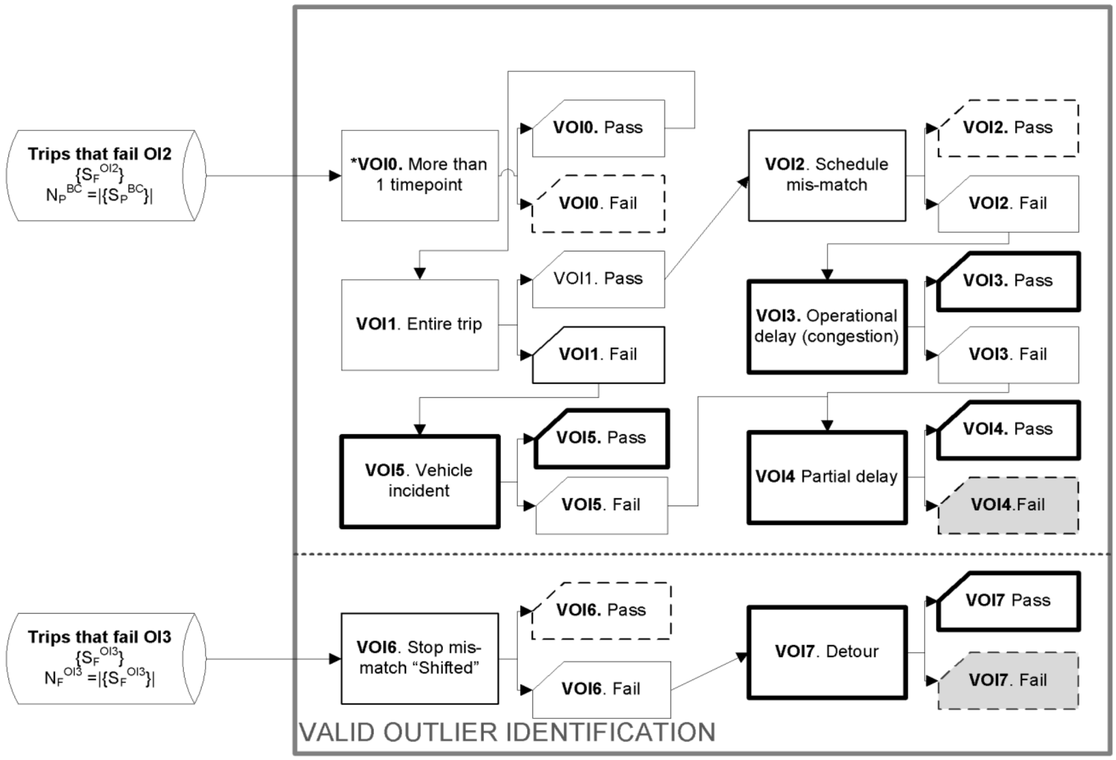

Figure 12 Schematic of the VOI Procedure

Time deviation outliers

There are six tests available for time deviation outlier trips. Each test involves analysing the pattern of schedule time deviations over the progression of the trip. Therefore, these tests can only be applied at the timepoint-level. Test VOI0 first checks for trips with a single timepoint. Unlike all other test, this can test can be applied directly at the trip-level instead of applying Equation 2:

$$\delta_j^{VOI0} = \begin{Bmatrix} 0 & \text{if } n_j^t = 1 \\ 1 & \text{otherwise} \end{Bmatrix} \qquad (18)$$

where n^t_{j} = the number of time points for trip j.

It is assumed that a trip with only 1 timepoint that is also a time deviation outlier is suspect because no further tests can be applied to determine if it is a valid case.

Figure 13 depicts some of the patterns associated with potential valid and invalid cases of time deviation outliers. VOI2 checks for a possible mis-match of the recorded data to the wrong schedule (the dashed line signifies that this type of pattern represents an invalid case and the solid line represents a valid case). VOI3 examines the trip pattern to identify congestion or operational delay over the entire trip. VOI4 is similar to VOI3 except it focuses on trips where time deviations only occur during a portion of the trip; and VOI5 examines the trips pattern for vehicle incidents.

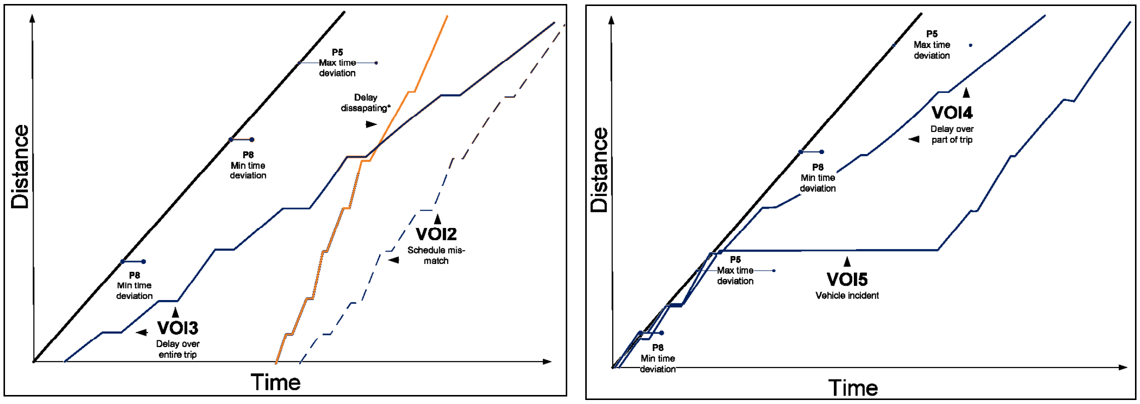

Figure 13 Time deviation outlier types

Figure 13 shows that VOI2-VOI5 can be separated into 2 classes: trips with time deviations that occur over the entire trip or trips with time deviations that occur over a portion of the trip. The plot on the left shows the former associated with tests VOI2 and VOI3. The plot on the right shows the latter associated with tests VOI4 and VOI5. The orange curve on the left hand side plot depicts a valid case of delay where it is dissipating over the course of the trip, where the previous trip likely experienced dealy. This case is not included in the VOI methodology because it requires confirming that the previous trip experienced a delay pattern. This trip chaining process requires that the AVL/APC database contains an identifier that links the sequence of trips; however this data is not always available in a conventional system.

VOI1

Test VOI1 is used to determine which valid case tests can be applied downstream. Recalling Equation 2, trips can be separated in the classes described in Figure 13 with the VOI1 test variable:

$$\delta_{i,j}^{VOI1} = \begin{cases} 0 & \text{if } \mathrm{TDev^A}_{i,j} \le \mathbf{P8} \text{ AND } 1 < i \le n_j \text{ AND } i \in \{i_R^t\} \\ & \text{OR } \mathrm{TDev^D}_{i,j} \le \mathbf{P8} \text{ AND} 1 \le i < n_j \text{ AND } i \in \{i_R^t\} \\ 1 & \text{otherwise} \end{cases} \qquad (19)$$

where i^t_R = the subset timepoint stops for route R; and
P8 = Min Time Deviation, user-defined minimum time deviation value (seconds).

P8 is a lower bound for the schedule time deviation to be considered a real difference between the recorded and scheduled time. Trips that fail VOI1, meaning that a real time deviations occurs for only a portion of the trip, are next checked by test VOI5. Trips that pass VOI1, meaning that a real time deviation occurs throughout all time points in the trip, are sent to VOI2 to be tested. The arrival time at the first stop and departure time at the last stop are excluded from the test.

VOI2

Test VOI-2 checks if the trip pattern suggests a mis-match to the wrong schedule**.** It is assumed that a mis-match would results in a pattern of relatively uniform time deviations through-out the entire trip (Figure A8 in Appendix A). A direct comparison of time deviation, which is measured in seconds, would be difficult to implement; a new attribute is defined based on the growth or increase of the time deviation between stops. The test variable for VOI2 is:

$$\delta_{\mathrm{i,j}}^{\mathrm{VOI2}} = \begin{cases} 0 \ \text{ if } \left|\mathrm{TIncr^A}_{\mathrm{i,j}}\right| > \mathbf{P9} \text{ AND } 2 < i \leq n_j \text{ AND i} \in \{\mathrm{i_R^t}\} \\ \quad \text{OR}\left|\mathrm{TIncr^D}_{\mathrm{i,j}}\right| > \mathbf{P9} \text{ AND } 1 \leq i < n_j \text{ AND i} \in \{\mathrm{i_R^t}\} \\ 1 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \tag{20}$$

where $\mathrm{TIncr^A}_{\mathrm{i.j}} = \frac{\mathrm{TDev^A}_{\mathrm{i,j}} - \mathrm{TDev^A}_{\mathrm{i-1,j}}}{\mathrm{TDev^A}_{\mathrm{i-1,j}}} \times 100\%$, increase in schedule arrival time deviation from previous time point (%); and

$\mathrm{TIncr^D}_{\mathrm{i.j}} = \frac{\mathrm{TDev^D}_{\mathrm{i,j}} - \mathrm{TDev^D}_{\mathrm{i-1,j}}}{\mathrm{TDev^D}_{\mathrm{i-1,j}}} \times 100\%$, increase in schedule departure time deviation increase from previous time point (%);

P9 = Max Time Increase, user-defined threshold to define growth of schedule time deviation (%).

The uniformity of the time deviation is tested by comparing the growth (or decline) of the time deviation value to a given threshold, P9. The increase for arrival time deviation is tested from the third stop onwards; the attribute is based on growth from the previous time point and arrival time at the first stop is considered irrelevant. If the trip fails VOI2 (i.e. it is not a schedule mis-match) then it may be tested for congestion or operational delay via test VOI3.

VOI3

Test VOI looks for a valid case pattern of congestion or operation delay; the expected pattern is increasing time deviation values in consecutive time points (Figure A9 in Appendix A). Test VOI3 variable checks that the time deviation between time points increases or remains the same; the fail condition is triggered when time deviations decline between time points as detected by a given threshold:

$$\delta_{\mathrm{i,j}}^{\mathrm{VOI3}} = \begin{cases} 0 \quad \text{if}\left(\mathrm{TIncr^A}_{\mathrm{i,j}}\right) < \mathbf{P10} \text{ AND } 2 < i \leq \mathrm{n_j i} \text{ AND i} \in \{\mathrm{i_R^t}\} \\ \quad\ \ \text{OR}\left(\mathrm{TIncr^D}_{\mathrm{i,j}}\right) < \mathbf{P10} \text{ AND } 1 \leq \mathrm{i} < \mathrm{n_j i} \text{ AND i} \in \{\mathrm{i_R^t}\} \\ 1 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \tag{21}$$

Where P10 = Max Time Decrease, the user-defined threshold to detect decline in time deviation (%). Note the absence of absolute brackets because positive growth is expected with congestion and negative growth (decline) is expected to trigger a fail condition. The schedule time deviation may increase or remain the same for trips with a congestion or delay pattern; P10 is recommended to be a negative value. Trips that pass VOI3 are identified as valid cases for a detected time deviation outlier. Trips that fail continue to test VOI4.

VOI4

It is recognized that congestion or operational delay may not occur over the entire trip, VOI4 tests for congestion or operational delay occurring during a portion of the trip. (Figure A10 in Appendix A). The test variable for VOI4 is:

$$\delta_{i,j}^{VOI4} = \begin{cases} 0 & \text{if } TIncr_{i,j}^{A} < \mathbf{P10} \text{ AND } i^* < i \leq n_j \text{ AND } i \in \{i_R^t\} \\ & \text{OR } TIncr_{i,j}^{D} < \mathbf{P10}\ AND\ i^* < i < n_j \text{ AND } i \in \{i_R^t\} \\ 1 & \text{otherwise} \end{cases} \quad (22)$$

where i* = the earliest timepoint where a time deviation outlier is identified.

P10 is defined in test VOI3, a similar criteria as test VOI3 is used except it is only applied to time points after i*.

VOI5

For the trips that failed VOI1, they are tested for the vehicle incident valid case with test VOI5. The expected pattern for a vehicle incident is split into 2 segments: before and after time point i* (See Figure A11 in Appendix A). The test variable is:

$$\boldsymbol{\delta}_{\mathbf{i,j}}^{\mathbf{VOI5}} = \begin{cases} 0 & \text{if } TDev_{i,j}^{A} \geq \mathbf{P5} \text{ AND } i < i^* \text{ AND } i \in \{i_R^t\} \\ & \text{OR } TDev_{i,j}^{D} \geq \mathbf{P5} \text{ AND } i < i^* \text{ AND } i \in \{i_R^t\} \\ & \text{OR } \left|TIncr_{i,j}^{A}\right| \geq \mathbf{P9} \text{ AND } i^* + 1 < i \leq n_j \text{ AND } i \in \{i_R^t\} \\ & \text{OR } \left|TIncr_{i,j}^{AD}\right| \geq \mathbf{P9} \text{ AND } i^* < i < n_j \text{ AND } i \in \{i_R^t\} \\ 1 & \text{otherwise} \end{cases} \quad (23)$$

For the segment before the outlier stop, no large time or distance deviations are expected. Therefore a fail condition is triggered if a time or distance deviation outlier is found before the timepoint i*. After the outlier stop, the schedule deviation pattern should remain uniform (similar to VOI2). Therefore the parameters P5 and P9 are previously defined in test OI2 and VOI2, respectively. Trips that pass VOI5 are considered valid cases of a time deviation outlier trip. Trips that fail this test are sent to VOI4 to check for congestion or operational delay patterns.

Distance deviation outliers

The next two tests are applied to distance outliers as identified by the set of trips failing test OI3. Figures A12 and A13 in Appendix A demonstrate the expected data patterns for test VOI6 and VOI7, respectively. VOI6 examines the trip patterns to identify mis-matched to schedule stops or a "shift" in the distance deviation value; and VOI7 examines trip patterns to identify detours.

VOI6

Test VOI6 is similar to test VOI2; however the parameter is applies to the distance deviation. The following variable checks for an expected pattern of uniform distance deviations:

$$\boldsymbol{\delta}_{\mathbf{i,j}}^{\mathbf{VOI6}} = \begin{cases} 0 \text{ if } \left|\mathrm{DIncr}_{\mathrm{i,j}}\right| \geq \mathbf{P11} \text{ AND } 1 < i \leq \mathrm{n_j} \text{ AND } \mathrm{i} \in \{\mathrm{i_R^t}\} \\ 1 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \tag{24}$$

where DIncr = $\frac{\mathrm{DDev_{i,j}} - \mathrm{DDev_{i-1,j}}}{\mathrm{DDev_{i-1,j}}}$, increase or growth in distance deviation in the segment before time point i; and

P11 = Max Dist Increment, user-defined threshold to detect increase of growth of distance deviation (%).

Test VOI7 follow similar logic as test VOI2 for uniformity except applied to distance deviations. A side effect of the VOI6 test design is that trips with one record, which show a distance deviation, will pass as a stop mis-match. The first record will not have a DIncr value because it requires a previous designated stop, the default value is zero. Trips that pass VOI6 are considered suspect, trips that fail are sent to VOI7.

VOI7

Similar to the vehicle incident pattern in test VOI5, detours will show a shift of distance deviation for all stops after the first distance deviation outlier point, i*. Figure A13 in Appendix A demonstrates this pattern. The test variable for VOI7 is:

$$\boldsymbol{\delta}_{\mathbf{i,j}}^{\mathbf{VOI7}} = \begin{cases} 0 \text{ if } \left|\mathrm{DDev_{i,j}}\right| > \boldsymbol{P6} \text{ AND } \qquad i < \mathrm{i}^* \\ \quad \text{OR } \left|\mathrm{DIncr_{i,j}}\right| \geq \mathbf{P11} \text{ AND } \mathrm{i}^* < i \leq \mathrm{n_j} \\ 1 \qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \tag{25}$$

All stops before the outlier stop are expected to remain within a reasonable range P6, as defined in test OI3. After the deviation point the increase in distance deviation between subsequent stops is expected to remain under a given growth threshold, P11, as defined in VOI6.

VOI Output

The set of trips that failed the VOI stage move into the set of suspect trip:

$$\mathrm{S_{SUSPECT}} = \mathrm{S_F^{BC}} \cup \mathrm{S_F^{VOI}} \tag{26}$$

where the VOI-stage failed set of trips (S_F^{VOI}) is defined as a union of:

- the outliers without a valid case (i.e. fail OI1, OI4),
- the set of trips identified as an invalid case (i.e. pass VOI0, VOI2, VOI7); and
- the sets of trips where no valid case option is identified (i.e. fail VOI4 and VOI7).

$$\mathrm{S_F^{VOI}} = \left[\mathrm{S_F^{OI1}} \cup \mathrm{S_F^{OI4}}\right] \cup \left[\mathrm{S_P^{VOI0}} \cup \mathrm{S_P^{VOI2}} \cup \mathrm{S_P^{VOI6}}\right] \cup \left[\mathrm{S_F^{VOI4}} \cup \mathrm{S_F^{VOI7}}\right] \tag{27}$$

The remainder trips in the database that have not been flagged during the QA procedure are considered non-suspect:

$$\mathrm{S_{NON-SUSPECT}} = \mathrm{S} - \mathrm{S_{SUSPECT}} \tag{28}$$

where S = the initial set of trips in the APC/AVL database that is inputted into the QA procedure. An alternate way to express the non-suspect trips is:

$$S_{NON-SUSPECT} = S_P^{BC+OI} + S_{Valid}^{VOI} - S_{Invalid}^{VOI} \qquad (29)$$

where the set of non-suspect trip consist of set of trip which passed all BC and OI tests (Equation 31), the set of trips with valid case outliers (Equation 32) minus the set with an invalid case outliers (Equation 28):

$$S_P^{BC+OI} = S_P^{BC} \cap S_P^{OI} \qquad (30)$$

$$S_{Valid}^{VOI} = \left[\left(S_F^{OI2}\right) \cap \left(S_P^{VOI3} \cup S_P^{VOI5} \cup S_P^{VOI12}\right)\right] \cup \left[\left(S_F^{OI2}\right) \cap \left(\left(S_P^{OI2}\right)\right] \qquad (31)$$

Valid case outlier trips are not mutually exclusive of invalid case outliers because there are multiple outlier types. For example, a trip with a time deviation outlier may be identified with a valid case option of experiencing congestion; however the same trip may have distance deviation outliers without a valid explanation.

3.3 Summary of QA methodology

The summaries of tests at each stage are shown in Tables 8 to 10. Table 11 is a summary of the parameter definitions.

Figure 14 is a detailed summary of the QA procedure. Tests are labelled in rectangles Each test output (denoted by the cards) is associated with a set variable, S_P^k and S_F^k, respectively for passing and failing sets for test k. Parameters are bolded and shown in brackets with their associated tests. The decisions to identify suspect and non-suspect trips are shown in the diamonds.

Table 8 Summary of BC tests

Name	Test Question	Parameter (s)
BC1	*Does time increment?*	None
BC2	*Does distance increment?*	None
BC3	*Are time and distance steps within reasonable thresholds?*	P1, P2
BC4	*Is the travel speed within a reasonble threshold?*	P3

Table 9 Summary of OI tests

Name	Test Question	Parameter (s)
OI1	*Are the passenger counts within a reasonble range?*	P4
OI2	*Is the schedule time deviation within a reasonble range?*	P5
OI3	*Is the schedule distance deviation within a reasonble range?*	P6
OI4	*Are the raw count corrections within a reasonable range?*	P7

Table 10 Summary of VOI tests

Name	Test Question	Parameter (s)
VOI0	*Is there more than 1 time points?*	None
VOI1	*Does a time deviation occur at every time point?*	P8
VOI2	*Does the trip pattern imply a mis-match to time schedule?*	P9
VOI3	*Does the trip pattern imply congestion or operational delay through entire trip?*	P10
VOI4	*Does the trip pattern imply congestion or operational delay for part of trip?*	P10
VOI5	*Does the trip pattern imply a vehicle incident?*	P5, P9
VOI6	*Does the trip pattern imply a mis-match to stop locations?*	P11
VOI7	*Does the trip pattern imply a detour?*	P11

Table 11 Summary of Parameter

Ref.	Name	Definition	Unit	Tests
P1	*Max Time Increment*	Threshold of reasonable time step between subsequent stops	seconds (s)	BC3
P2	*Max Distance Increment*	Threshold of reasonable distance step between subsequent stops	metres (m)	BC3
P3	*Max Travel Speed*	Threshold of reasonable travel speed between subsequent stops	metres per second m/s	BC4
P4	*Max Passenger Count*	Threshold to identify an outlier passenger count (board, alight or load)	persons (prs)	OI1
P5	*Max Time Deviation*	Threshold to identify an outlier schedule time deviation	seconds (s)	OI2, VOI5
P6	*Max Distance Deviation*	Threshold to identify an outlier schedule time deviation	metres (m)	OI3, VOI7
P7	*Max Count Correction*	Threshold to identify an outlier correction to raw count (board or alight)	persons (prs)	OI4
P8	*Min TIme Deviation*	Minimum time difference from schedule to be considered a deviation	seconds (s)	VOI1
P9	*Max Time Increase*	Threshold of time deviation growth between subsequent stop to be considered uniform	Percent (%)	VOI2, VOI5
P10	*Max Time Decrease*	Threshold of time deviation decline between subsequent stop to be considered decreasing	Percent (%)	VOI3, VOI4
P11	*Max Distance Increase*	Threshold of distance deviation growth between subsequent stop to be considered uniform	Percent (%)	VOI6, VOI7

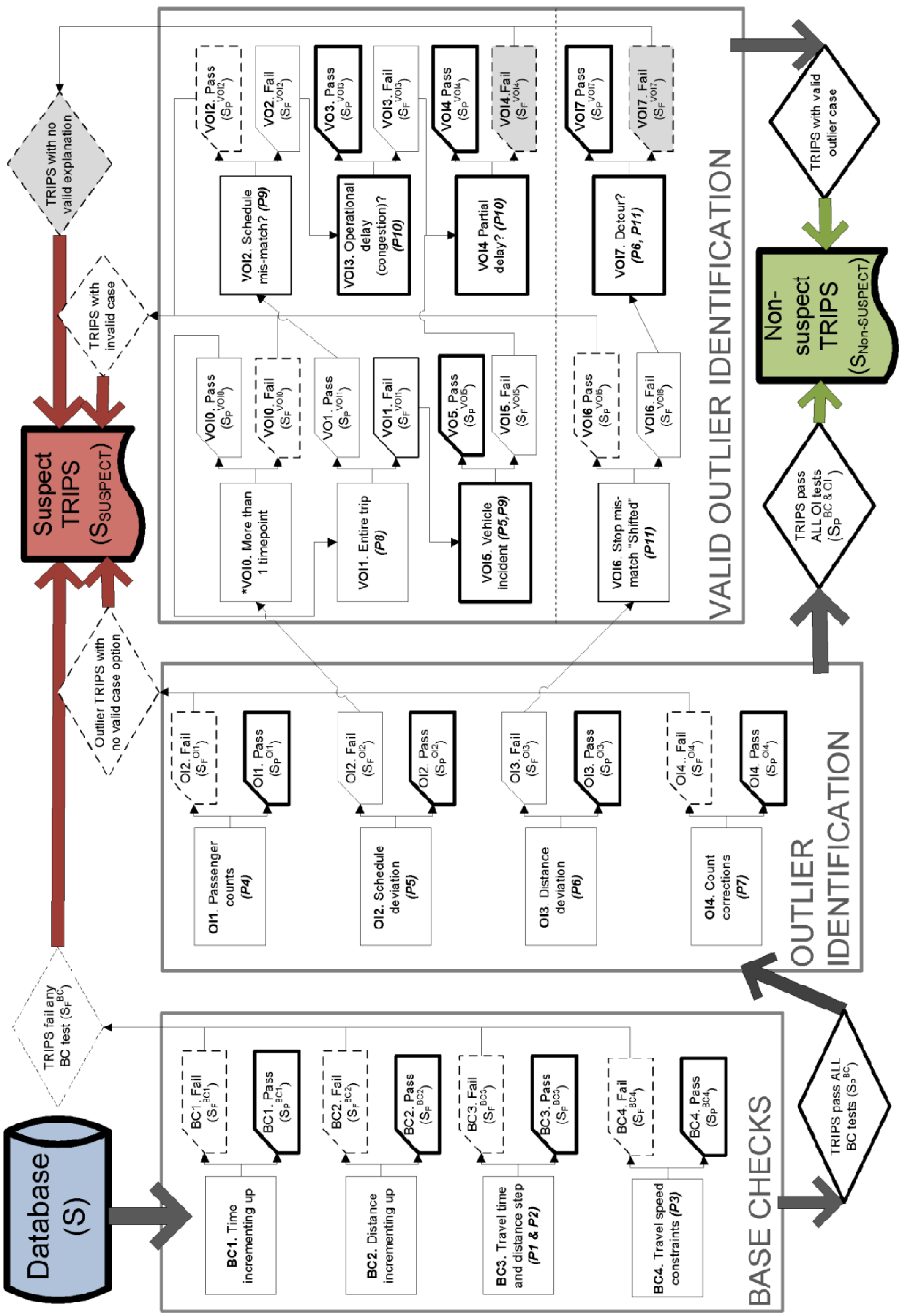

Figure 14 Detailed Summary of the QA Procedure

As mentioned in the methodology outline (Section 3.2), the QA procedure flags suspect data at the trip-level while the individual tests are applied at the stop-level. The binary test variables together with Equations 2 and 7 can be used to derive the number of trips failing each stage of the QA procedure:

$$N_F^{BC} = N - N_P^{BC} \quad (32)$$

$$N_F^{OI} = N_P^{BC} - N_P^{OI} - N_{F\prime}^{OI} \quad (33)$$

$$N_F^{VOI} = N_{F\prime}^{OI} - N_P^{VOI} \quad (34)$$

$$N_{SUSPECT} = N_F^{BC} + N_F^{OI} + N_F^{VOI} \quad (35)$$

$$N_{Non-SUSPECT} = N - N_{SUSPECT} \quad (36)$$

where N represents the total number of trips; the superscripts ^{BC}, ^{OI} and ^{VOI} signify the associated QA stage subset; the subscripts P, F and F` signify pass, fail and conditional fail, respectively.

Chapter 4

Case Study: Grand River Transit

The QA procedure was applied to a sample of archived APC/AVL data for Grand River Transit (GRT), based in the Region of Waterloo (the Region). The data spanned four months, or one service period, and was collected between September 1, 2008 and December 31, 2008. This chapter describes the Region, the transit agency characteristic and its APC/AVL database, the calibration of the QA parameters for the GRT data and a sample manual survey comparison of the archived data.

4.1 The Region of Waterloo and GRT

The Region of Waterloo is a regional municipality located in southwest Ontario, Canada, approximately 100 km southwest from Toronto, 150 km northwest from the United States border-crossing at Niagara, and 300 km northeast of the Detroit-Windsor border crossing (Figure 15). The Region has a population of roughly 534,900 people (2008 census) and is comprised of three municipalities and four rural townships: the cities of Cambridge, Kitchener and Waterloo (also known as the Tri-City) as well as the townships of Wilmot, Woolwich, Wellesley and North Dumphries. Grand River Transit services the Tri-City area, where approximately 90% of the population live. The annual GRT ridership in 2008 was 15.8 million trips over 60 regular routes with a fleet of 208 buses (Beniston, 2009).

4.2 The APC and AVL System

In 2005, the Region of Waterloo began service of the iXpress route, a Federal Transit Showcase Program for a central transit corridor express service. The central transit corridor denotes key passage areas of the transportation network that represents a large portion of trips; therefore, strategic transit service development is focused in these areas.

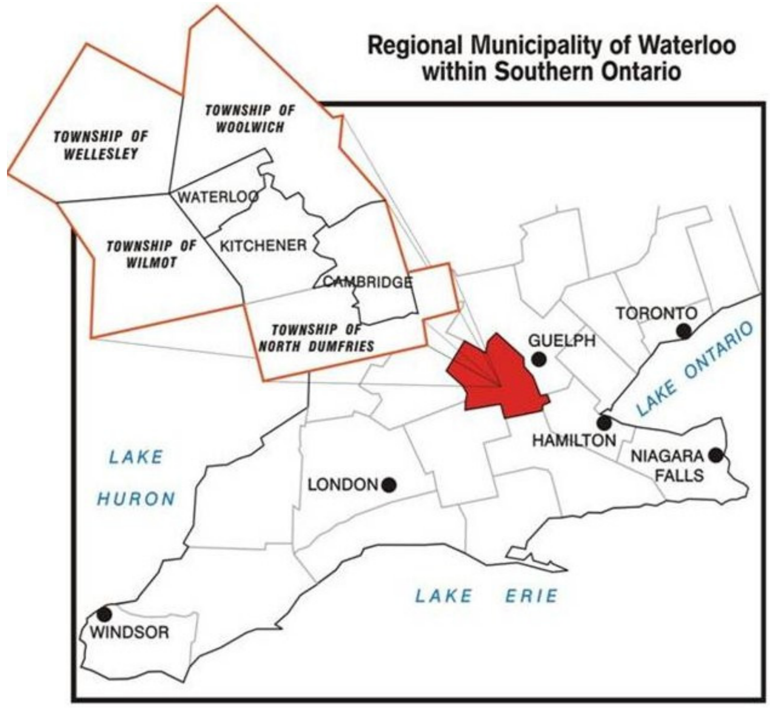

Figure 15 Map of the Region of Waterloo (Source: Regional Municipality of Waterloo, 2009)

As part of the showcase program, APC and AVL technology was featured and integrated as part of a larger Intelligent Transportation System (ITS) package that included: transit signal priority, dispatch and controller support, and real-time passenger information. The APC and AVL system components became fully operational on GRT vehicles in 2007.

Equipped buses are mounted with infrared passenger counters on each door and a GPS antennae to record stop times and locations; these instruments are linked to an on-board computer that stores the data while the transit vehicle is in service and communicates the information to controllers. Of the 208 buses in the fleet, 34 vehicles are equipped with APC/AVL technology during the data collection period for this study; 15 are fully dedicated to the iXpress route and the 19 remaining buses are circulated among the other routes.

One key objective of the ITS deployment was to enhance data collection and information processing; these goals help improve service planning and market research initiatives by archiving APC and AVL data. The database structure and data collection processes are explained in the next section, followed by a description of the sample data.

4.2.1. Data collection and storage

The APC/AVL system logs different events such as when the bus stops, opens or closes its doors, passenger boarding and alighting at each door, time of arrival and departure at the stop and location. An on-board computer processes this information into stop-level records. There are six events that trigger a stop record in the GRT system (Table 12).

Table 12 Events that generate a record in the GRT APC/AVL system

Event	Transit stop?	Transit timepoint?	Doors opened?	Event Trigger	Possible Cause
Stop with schedule time	✓	✓	✓	when the vehicle stops at a designated stop location with a scheduled timepoint and the doors are opened	passenger boarding or alighting requested
Drive through with schedule time	✓	✓	X	when the vehicle passes a designated stop location with a scheduled timepoint and the doors do not opened	no passenger boarding or alighting requested
Stop without schedule time	✓	X	✓	when the vehicle stops at a designated stop location without a scheduled timepoint and the doors are open	passenger boarding or alighting requested
Drive through without schedule time	✓	X	X	when the vehicle passes a designated stop location without a scheduled timepoint and the doors do not open	no passenger boarding or alighting requested
Stop without doors	X	X	X	when the vehicle stops at an un-designated stop location without a scheduled timepoint and the doors do not open	vehicle delayed by traffic signals, congestion, yielding right-of-way etc.
Stop with doors	X	X	✓	when the vehicle stops at an un-designated stop location without a scheduled timepoint and the doors are opened	passenger request due to safety, accessibility etc.

(Source: Mandelzys, M., 2010)

When a transit vehicle reaches the garage, a wireless link enables download of the on-board data to an ORACLE database where matching algorithms link the stop-records to schedule data. For some vehicles, the data must be uploaded manually. Some data are lost in the matching phase due to lack of necessary information (i.e. missing data) to match to the schedule or due to ambiguous matching rules. The QA procedure is applied after schedule matching has occurred. Due to confidentiality issues, a limited view of the relevant components in the ORACLE database is shown in Figure 16; a full description of the ORACLE structure would require proprietary information from the system vendor.

Each rectangle in Figure 16 represents a separate table in the ORACLE database, the solid lines signify the direct relationships between each table and the dashed line represents a pseudo-linkage through other relationships.

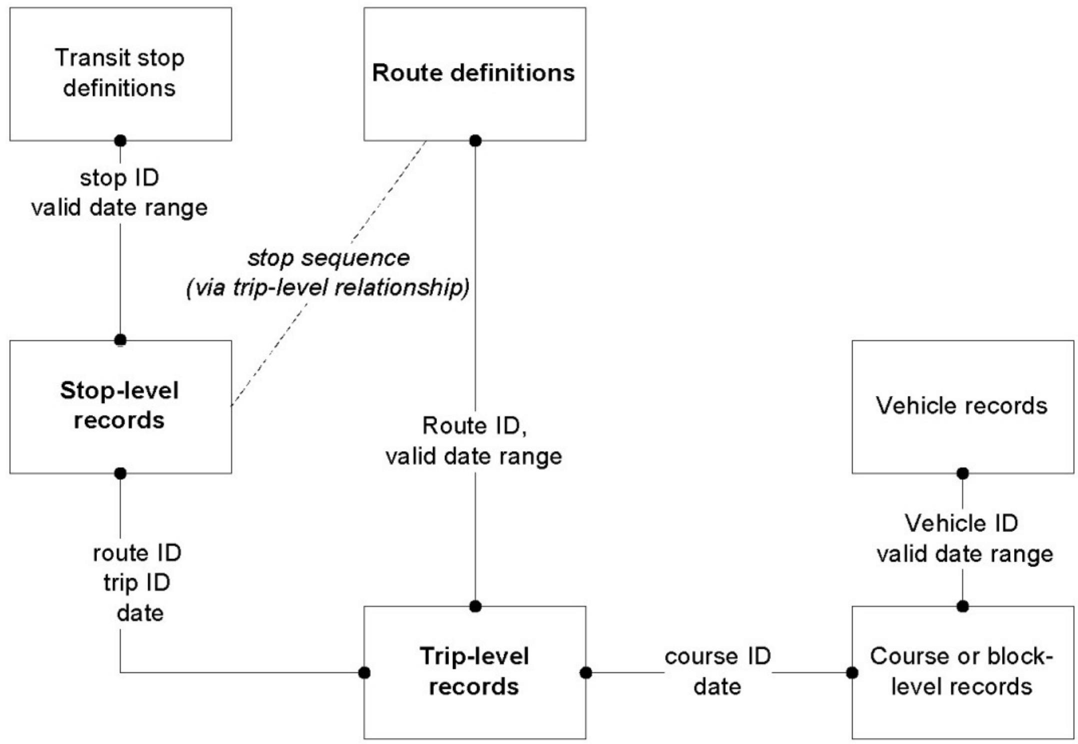

Figure 16 Snapshot of the ORACLE database structure

The methodology described in Chapter 3 is based on the analysis of data at the stop-level and the identification of suspect data at the trip-level. These tables are imported into an MS Access database and a combination of queries and VBA code are used to implement the QA tests. Tables 13 to 15 show the fields within the selected tables; fields associated with the QA tests are highlighted.

Table 13 Relevant fields in the Trip-level records

Field	Type	Description
Trip ID	NUMBER	Index, (primary key) related to stop-level records
Operation Date	DATE	Date of record
Line No	NUMBER	Route name
Route ID	NUMBER	Reference to specific route pattern in route definition
Route direction	NUMBER	Reference to route direction
Vehicle ID	NUMBER	Reference to transit vehicle
Actual start time	NUMBER	Actual trip start time in seconds past midnight
Actual end time	NUMBER	Actual trip end time in seconds past midnight
Scheduled start time	NUMBER	Scheduled trip start time in seconds past midnight
Scheduled end time	NUMBER	Scheduled trip end time in seconds past midnight
Stop sequence	STRING	Order of stops by stop no. according to reference route
Odometer	NUMBER	Odometer reading at start of trip

The events listed in Table 13 describe the records in the trip-level table. Table 14 describe stop-level data and Table 15 show relevant fields from route definitions.

Table 14 Relevant fields in the stop-level records

Field	Type	Description	Variables
Event ID	NUMBER	Index (primary key)	used to assigni, i =1...n_j, for a given j
Trip ID	NUMBER	Reference to trip-level record	used to assign j
Operation Date	DATE	Date of record	
Vehicle No	NUMBER	Reference to transit vehicle	
Stop No.	NUMBER	Reference to transit stop definition	
Stop Name	STRING	Name of designated stop location	
Stop Type	NUMBER	Type of event triggering stop (refer to Table 5)	
Previous Event ID	NUMBER	Reference to previous event ID	i-1, for given j
Stop Index	NUMBER	Reference to stop pattern in route definition	
Scheduled arrival time	NUMBER	Scheduled arrival time in seconds past midnight	$A'_{i,j}$ where $i \in i^t_R$
Scheduled departure time	NUMBER	Scheduled departure time in seconds past midnight	$D'_{i,j}$ where $i \in i^t_R$
Actual arrival time	NUMBER	Actual arrival time in seconds past midnight	$A_{i,j}$
Actual departure time	NUMBER	Actual departure time in seconds past midnight	$D_{i,j}$
Odometer	NUMBER	Odometer reading	used to derive $Dist_{i,j}$
Boardings	NUMBER	Number of passengers boarding, balanced	$P_{B\ i,j}$
Alighting	NUMBER	Number of passengers alighting, balanced	$P_{A\ i,j}$
Load	NUMBER	Number of passengers, balanced	$P_{L\ i,j}$
Raw Boardings	NUMBER	Number of passengers boarding, raw	$P'_{B\ i,j}$
Raw Alighting	NUMBER	Number of passengers alighting, raw	$P'_{A\ i,j}$

Table 15 Relevant fields from schedule definitions

Field	Type	Description	Variable
Route ID	NUMBER	Index (primary key)	i =1...n_j, for a given j
Line No.	NUMBER	Reference to trip-level record	j
Index No.	DATE	Date of record	
Stop ID	NUMBER	Reference to transit vehicle	
Distance to start	NUMBER	Reference to transit stop definition	$Dist'_{i,j}$
Distance to next	STRING	Name of designated stop location	
Stop Name	NUMBER	Type of event triggering stop (refer to Table 5)	

4.2.2. Sample Data

The sample data were collected between September 1, 2008 and December 31, 2008; this time span represents one of three annual service periods for the GRT. The sample data are comprised of 612,400 stop-level records, representing 25,021 trips. The distributions of data at the trip-level are shown by route-type, month, day of the week and start time Tables 16 and 17 and Figures 16 and 17, respectively.

Table 16 Distribution of trips in sample data by route type

Route Type	No. of Trips	Percentage
Regular	15,124	60.4%
iXpress	9,608	38.4%
Special	289	1.2%
Total	**25,021**	**100.0%**

Since iXpress trips are fully operated by equipped vehicles, they represent a substantial portion of the sample data. Regular routes refer to all other regular stop, full-day service routes and special routes refer to school specials or custom routes for special events.

Table 17 Distribution of trips in sample data by month

Month	Ixpress	Regular	Special	All trips	Percentage
September	2478	3769	73	6320	25%
October	2735	5148	79	7962	32%
November	2372	3453	85	5910	24%
December	2023	2754	52	4829	19%
Total	**9608**	**15124**	**289**	**25021**	**100%**

The larger portion of sample trips in October is somewhat unexpected; this trend is most apparent for regular routes. iXpress trips show the most even distribution of trips by month. December service may be less frequent due to holiday schedules.

Figure 17 shows a large majority of the sample data representing weekday trips. Upon closer inspection of the distribution by route-type, only iXpress trips contain weekend data.

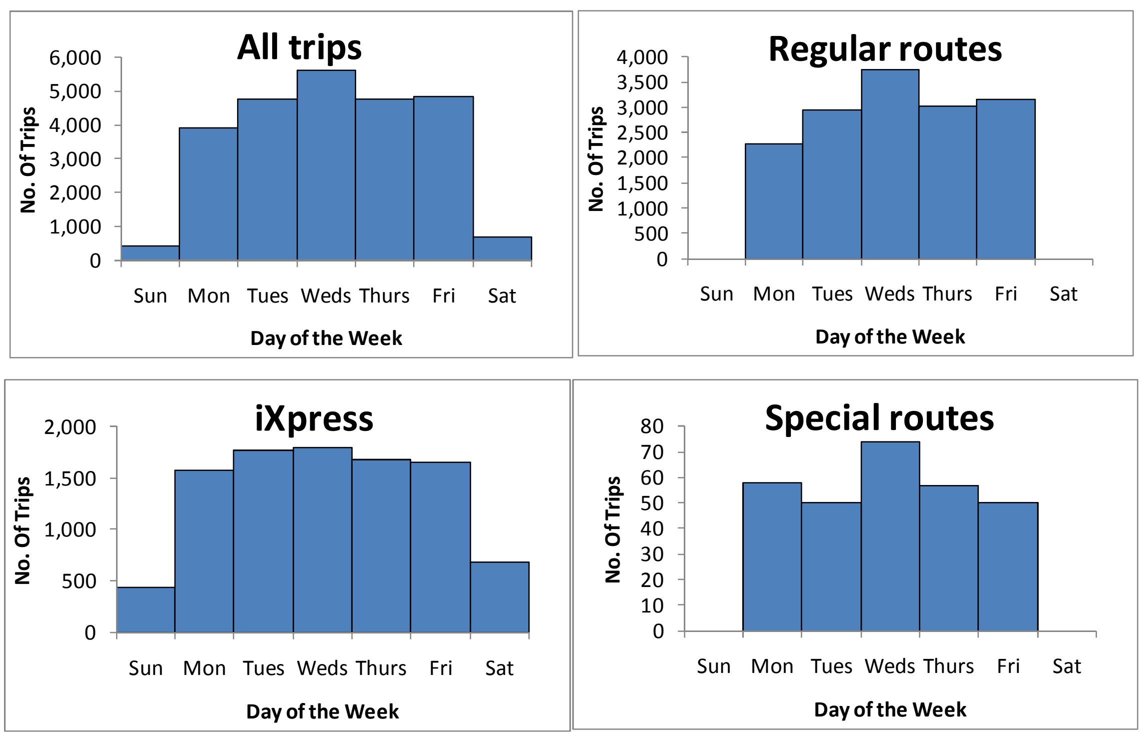

Figure 17 Distribution of trips in sample data by day of the week and by route

Figure 18 shows that both AM and PM peak periods (6-9AM and 3-6PM, respectively) are slightly more visible in the sample. It is likely that more frequent service during the weekday and peak periods results in a greater representation in the sample data. This trend is exaggerated in regular routes. Special routes are comprised mostly of school specials, early and late night trips. iXpress service frequency is constant for most of the day; therefore the sample data is evenly distributed except after 6pm, when service frequency decreases. There is also no weekend iXpress service after 6pm.

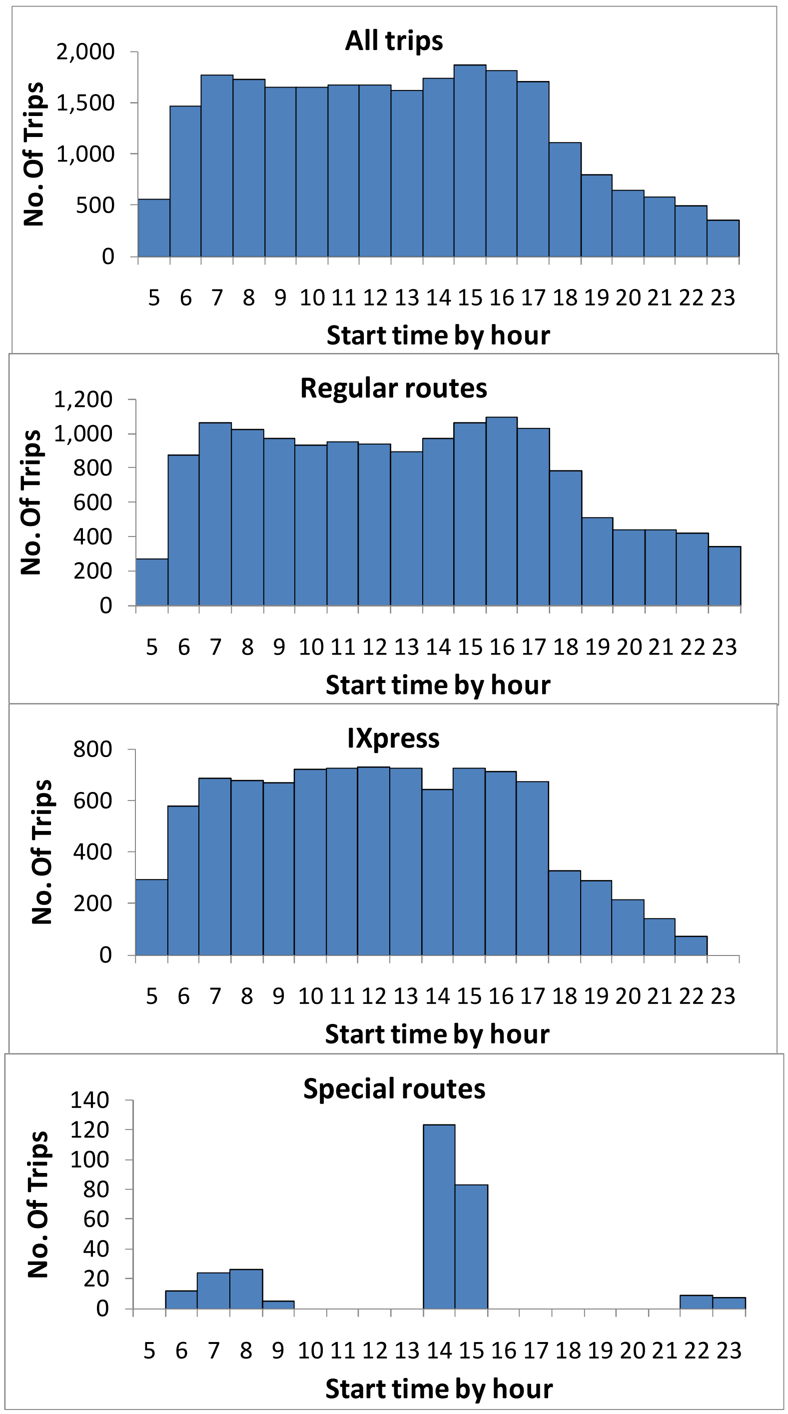

Figure 18 Distribution of trips in the sample data by start time and by route

4.3 Calibration of QA parameters for GRT

Before the methodology can be applied to the sample data, the parameters in Table 10 need to be calibrated for the GRT system. For example, the vehicle capacity (P4) and maximum vehicle speed (P3) would change as a result of the type of transit vehicle. The maximum time (P1) and distance increments (P3) would be a function of the network characteristics. Thresholds for schedule deviations in time (P4) and distance (P6) may depend on historical performance. The maximum acceptable count correction can depend on proven accuracy levels of the specific APC equipment.

To determine the appropriate parameter values, the stop-level attribute associated with each parameter is evaluated for each stop record in the database. For parameters related to a maximum threshold, the largest stop-level attribute was assigned to each trip; for minimum thresholds, the smallest stop-level attribute was assigned to each trip. The parameter is selected based on trip-level distribution of the assigned stop-level attribute. The cumulative distribution represents the portion of trips that would pass the given test for a range of parameter values.

An example of the parameter selection process is depicted for P1 (maximum time increment) for test BC3. For each stop record, the time increment from the previous stop is calculated; each trip is associated with the largest time increment among its stops. Figure 19 shows the distribution of trips by largest time increment, which is compared to P1 in test BC3.

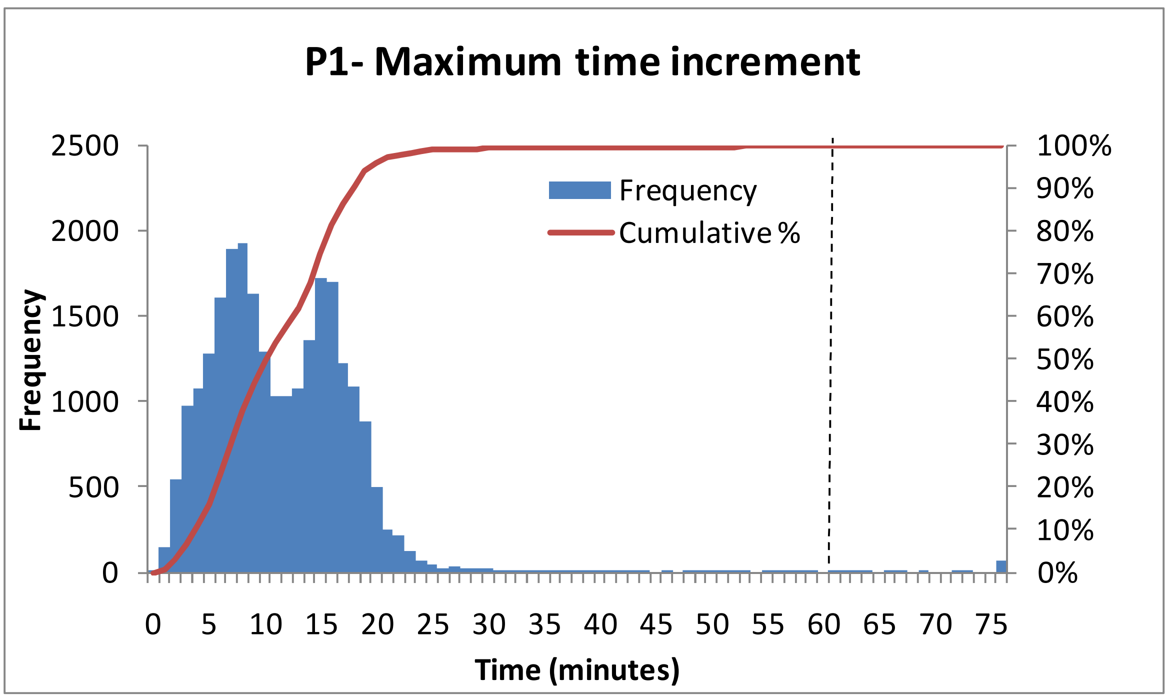

Figure 19 Trip distribution by largest stop-level time increment (Selection of P1)

Figure 19 shows a distribution with two peaks: one peak ranges from 5-10 minutes and the second ranges from 15-20 minutes. The second peak likely represents a majority of iXpress trips where limited stop service results in longer travel times between stops. Test BC3 checks that the time increments remain within a reasonable time range; P1 represents the threshold for a reasonable time step. It is assumed that a reasonable upper bound for the time increment is the maximum one-way cycle time for a route. The two peaks suggest that this parameter should ideally be set by the route or route-type. For the proposed QA procedure, only one network-level value is chosen to simplify the application. The selected value for P1 is depicted by the dashed line: 60min (1hr). The cumulative percentage on the right side of P1 shows the portion of trips that have a maximum time increment less than P1; that portion represents more than 99.6% of trips.

Appendix B shows the distributions and cumulative percentage plots for all the parameters. The calibrated values were chosen based on the shape of the distribution, engineering judgement and local knowledge of the GRT route network as demonstrated for the selection of P1 above. Table 18 is a summary of the calibrated parameters.

Table 18 Summary of QA parameters calibrated for the GRT system

Ref.	Name	Calibrated Value	Tests
P1	*Max Time Increment*	1hr (3600s)	BC3
P2	*Max Distance Increment*	15km (15,000m)	BC3
P3	*Max Travel Speed*	100 km/hr (27.8m/s)	BC4
P4	*Max Passenger Count*	80 persons	OI1
P5	*Max Time Deviation*	20min (1200s)	OI2, VOI5
P6	*Max Distance Deviation*	2km (2,000m)	OI3, VOI7
P7	*Max Count Correction*	6 persons	OI4
P8	*Min TIme Deviation*	1 min (60s)	VOI1
P9	*Max Time Increase*	10%	VOI2, VOI5
P10	*Max Time Decrease*	-5%	VOI3, VOI4
P11	*Max Distance Increase*	5%	VOI6, VOI7

It is important to note that the parameters were chosen with the quality assurance of service journey trips in mind. Deadhead runs to the garage may result in transit vehicles travelling greater than 100km/hr and with distance steps much greater than 15km.

4.2.1. User preferences

As noted in section 3.1, the selection of the methodology parameters can be tailored to the user goals. An aggressive user may calibrate the QA procedure with the intention of maintaining a larger dataset that presents less risk of throwing out valid data. Conservative users will generally be more concerned with including erroneous data, thus setting more stringent parameters.

A similar concept can be applied to inclusion of tests outputs from the VOI stage within the set of suspect or non-suspect data. The transit agency can decide not to include certain valid outlier cases within their set of useable or "non-suspect" data. For example, GRT expressed concerns regarding the inclusion of vehicle incidents and detours within the input data for service planning purposes. Although these data are still considered valid and useful for performance reporting, these data should not be included in the input for travel time analyses (for example) of future route planning because these routes should not be designed to anticipate accidents and detours. This preference can be corrected by simply including those output of test VOI5 (vehicle incidents) and VOI7 (detour) within the set of suspect trips.

For the application demonstrated in this thesis, performance monitoring is the focus for downstream data uses. Therefore, vehicle incidents and detours are still included as valid trip outliers.

4.4 Manual survey comparison

Although the QA methodology was designed to preclude the need for external data, manual surveys were previously conducted by GRT staff in October 2008 to assess passenger counter accuracy. The author also conducted a separate survey of a two-way cycle of an iXpress trip on November 13, 2008 to observe and assess potential error sources to the data collection process of the GRT APC/AVL system. The results of these efforts are presented in this section.

The passenger count accuracy was calculated using four error definitions for the passenger count, boarding count, alighting count and unbalanced count (Equations 37 to 41). These equations are provided by the system vendor, INIT Inc (Goetz, 2006):

$$\varepsilon = \frac{P^R - P^M}{P^M} \times 100\% \qquad (37)$$

where ε = total passenger count error; and

P = total passenger count (from all doors for each stop over all trips).

Superscripts *R* and *M* denote the raw APC counts and manual counts, respectively. The total passenger count is determined by:

$$P = \frac{\sum_{i=1}^{N}\sum_{d=1}^{n_d} P_{B_{i,d}} + \sum_{i=1}^{N}\sum_{d=1}^{n_d} P_{A_{i,d}}}{2} \qquad (38)$$

where N = number of stop observations when passengers are counted (i = 1 to N);

n_d = number of doors (d=1 to n_d); and

P_B, P_A = boarding and alighting passengers, respectively.

INIT stipulates that at least 700 passengers must be counted before the error equations apply. The total number of stop observations, *N*, is the sum of stops over all trips required to reach the 700 passenger count threshold (i.e. error is not calculated for each trip, one error value for the raw

APC counts is calculated for all observations by stop and door in the survey). The passenger boarding and passenger alighting errors may be calculated separately:

$$\varepsilon_B = \frac{\sum_{i=1}^{N}\sum_{d=1}^{n_d} P_{B_{i.d}}^{R} - \sum_{i=1}^{N}\sum_{d=1}^{n_d} P_{B_{i.d}}^{M}}{\sum_{i=1}^{N}\sum_{d=1}^{n_d} Z_{B_{i.d}}^{M}} \times 100\% \qquad (39)$$

$$\varepsilon_A = \frac{\sum_{i=1}^{N}\sum_{d=1}^{n_d} P_{A_{i.d}}^{R} - \sum_{i=1}^{N}\sum_{d=1}^{n_d} P_{A_{i.d}}^{M}}{\sum_{i=1}^{N}\sum_{d=1}^{n_d} P_{A_{i.d}}^{M}} \times 100\% \qquad (40)$$

where ε_B and ε_A are the boarding and alighting errors, respectively. Lastly the unbalanced error describes the number of absolute errors in the counts:

$$\varepsilon' = \frac{\sum_{i=1}^{N}\left|\sum_{d=1}^{n_d}\left(P_{B_{i,d}}^{R} - P_{B_{i.d}}^{M}\right)\right| + \sum_{i=1}^{N}\left|\sum_{d=1}^{n_d}\left(P_{A_{i,d}}^{R} - P_{A_{i,d}}^{M}\right)\right|}{\sum_{i=1}^{N}\sum_{d=1}^{n_d}\left(P_{A_{i,d}}^{M} + P_{B_{i.d}}^{M}\right)} \times 100\% \qquad (41)$$

where ε' represents the unbalanced error for the passenger count. Table 19 represents the results of the GRT survey. Table 20 displays the results of the individual survey. Typical values represent average results from other transit agencies with the same APC technology; these values are provided by the vendor (Goetz, 2006).

Table 19 Results of GRT Manual Survey

Error	Survey Results	Typical Values	Vendor Guarantee
Balanced Total Passenger Count (ε)	7.7%	+/-4 %	≤ 5%
Balanced Total Boarding Count (ε_B)	8.9%	+/-5 %	≤ 10%
Balanced Total Alighting Count (ε_A)	6.4%	+/-5%	≤ 10%
Unbalanced (absolute) Total Passenger Count (ε')	11.7%	15%	-

Table 20 Results of Individual Manual Survey

Error	Survey Results	Typical Values	Vendor Guarantee
Balanced Total Passenger Count (ε)	-5.3%	+/-4 %	≤ 5%
Balanced Total Boarding Count (ε_B)	-5.9%	+/-5 %	≤ 10%
Balanced Total Alighting Count (ε_A)	-4.8%	+/-5%	≤ 10%
Unbalanced (absolute) Total Passenger Count (ε')	8.9%	15%	-

The results in Table 19 are based on a sample size of 1099 observations; the GRT survey observed two doors for 551 stops over eight two-cycle trips to reach the 700 passenger count target. The results in Table 20 are based on a small sample size of 52 observations: two doors for 13 stops over one two-way trip cycle. Despite the difference between the two surveys, both boarding and alighting count errors remain below the +/-10% vendor guarantee.

Raw APC counts are often adjusted as a result of passenger balancing algorithms. Based on the accuracy assessment, it is reasonable to assume that most count corrections fall within 10% of the count value. Considering bus capacity, the results are useful in the selection of P7 (maximum count correction). GRT estimates that the bus capacity is 50 persons. Although the error value represents a global mean, above-average errors would likely result in relatively low count corrections because most passenger counts are expected to be below bus capacity. For example, count corrections greater than 5 would imply that a boarding or alighting passenger count is both at capacity (50 persons) and error is greater than or equal to 10% or both the passenger count is below capacity and error is much greater 10%.

Another exercise during the second survey was to track the bus trajectory using external GPS units; this trajectory is compared to the trajectory produced from AVL stop records to assess the system capability for identifying appropriate stops and schedule matching. Figure 20 is a sample trajectory comparison between two consecutive route-designated stops on an iXpress route.

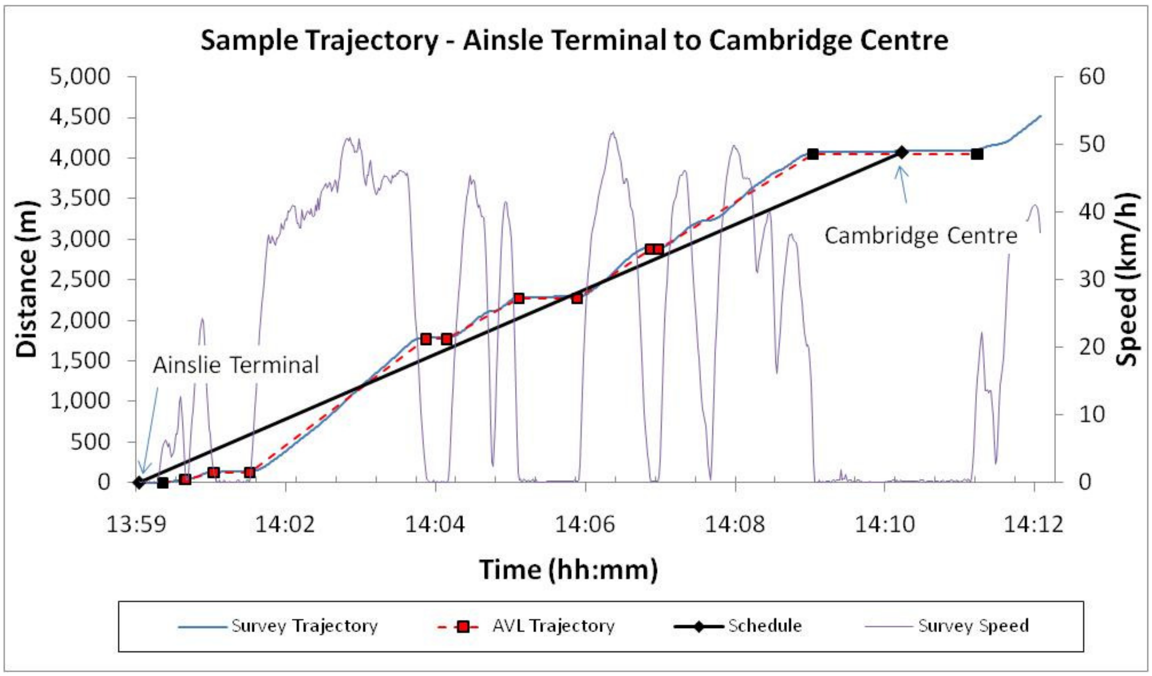

Figure 20 Sample Trajectory Comparison

From Figure 20, the survey bus speeds verify that AVL stop events are triggered when the bus speed is zero; low or crawl speeds do not trigger event records as shown at 14:05 and 14:08. Despite the lack of records for these events, the AVL trajectory, which is a straight line interpolation (i.e. constant speed), sufficiently agrees with the survey trajectory. The black points on the AVL trajectory represent scheduled stops; comparisons between the schedule and AVL odometer distances demonstrate adequate matching capabilities of the system.

Chapter 5

Results and Discussion

This chapter presents the results in three parts; the first section examines the outputs of the QA procedure; the second section evaluates the sensitivity of the QA output to different parameter values and changes in the QA structure; and the third section discusses the impact of the QA procedure on transit performance measures.

5.1 Analysis of Output

The QA procedure presented in Chapter 3 and using the parameter values in Table 18 flagged 3,583 suspect trips in the four months of GRT AVL/APC data; Table 21 shows the impact of the suspect data on the available data for service monitoring and performance analysis.

Table 21 Number of non-suspect records before and after QA

Records	Without QA	With QA	% Suspect
Trip-level	25,021	21,438	14.3%
Stop-level	612,400	509,938	16.7%

The next sections further examine the results by: reviewing the causes for flagging data as suspect; re-evaluating the original pattern assumptions developed in the VOI stage of the methodology; and observing the characteristics of the remaining non-suspect data.

5.1.1. Analysis of suspect trips

The purpose of this section is to determine the extent of which each test contributes to the identification of suspect data and to confirm that suspect trips do represent unreliable data.

Suspect data by test

Table 22 is a summary of the reasons why trips are flagged as suspect (Note: it is possible for a trip to be flagged for multiple reasons).

Table 22 Summary of suspect reasons

Reason to suspect data	Test	No. Trips	% Total
Time does not increment forward	BC1 Fail	92	2.6%
Distance does not increment forward	BC2 Fail	1	0.0%
Unreasonable time or distance step	BC3 Fail	1326	37.0%
Unreasonable travel speed	BC4 Fail	96	2.7%
Passenger count greater than bus capacity	OI1 Fail	18	0.5%
Raw passenger count over-corrected	OI4 Fail	1511	42.2%
Single outlier timepoint deviation	VOI0 Pass	4	0.1%
Suspected mis-match in schedule	VOI2 Pass	10	0.3%
Suspected mis-match in stop locations	VOI6 Pass	133	3.7%
Unknown reason for large schedule time deviation	VOI4 Fail	11	0.3%
Unknown reason for large schedule distance deviation	VOI7 Fail	577	16.1%
Total number of suspect trips		***3,583***	

The top three reasons to flag a trip as suspect are: unreasonable time or distance step (BC3 fail), raw passenger count over-corrected (OI4 fail) and unknown reason for large distance deviation (VOI7 fail). Unfortunately, test BC3 checks time and distance steps simultaneously and therefore, it is not possible to segregate the two causes. Trips failing test OI1 and OI4 reflect trips with unreliable passenger counts. The greater portion failing due to raw count over-correction may suggest that the passenger balancing algorithm may be improved.

Trips failing test BC3 in effect flag trips with unreliable arrival and departure times or distance travelled. There are several reasons why a large time or distance step (fail BC3) might be recorded: malfunction in the on-board computer may miss a stop-record, assignment the bus stop-time and start-time events to another record, or improperly processing the sequence of stop events. Incorrect time and distance values might also be logged due to malfunctioning instruments.

The only valid case for distance deviation outliers is detours, which is not expected to be a common occurrence in the sample data. Therefore, the majority of distance outliers are flagged as suspect. Trips in VOI6 pass and VOI7 fail represent trips with unreliable stop-type identification or poor matching to designated stops. In contrast there are several valid cases identified for time deviation outliers. Table 22 shows that there are fewer suspect trips identified due to time deviation outliers (VOI0 pass, VOI2 pass and VOI4 fail), suggesting that the schedule matching algorithm works adequately. It is also possible that the smaller subset of timepoints subject to stop-level VOI tests and more valid case options results in more positive outcomes for non-suspect trips.

Confirming suspect data patterns

The first seven tests in Table 22 use straightforward thresholds to identify suspect trips. For example Figure 21 is an example trip for which passenger counts exceed P4 (maximum passenger count). Boarding and alighting counts are depicted by the blue and red bars, load is represented by the solid green line and P4 is shown in the dashed purple line. Test OI1 stipulates that all passengers count values (boarding, alighting and load) should be below the value of P4, which represents a maximum capacity.

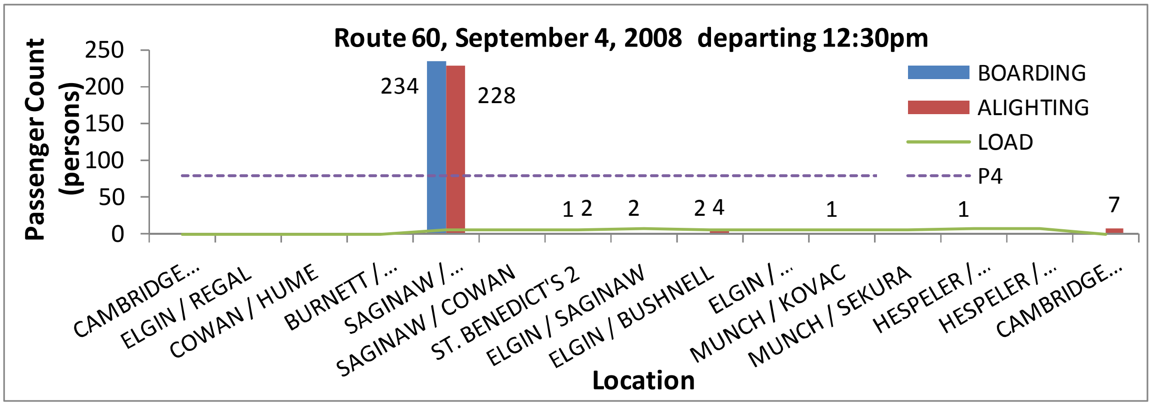

Figure 21 Example suspect trip due to passenger counts over bus capacity

Figure 21 shows that boarding and alighting counts at the Saginaw stop were recorded at 234 and 228, respectively. Although the load count remains below P4, this data is suspicious because it highly unlikely to observe these boarding and alighting counts during a service journey. Ridership statistics derived from boarding counts would be impacted from this data.

Data flagged with BC and OI tests are similar to the example in Figure 21 in that a simple threshold is applied: these tests are straightforward to visualize. Instead, the analysis focuses on suspect data associated with an invalid pattern; to confirm pattern assumptions in the VOI tests. VOI tests identify suspect trips in four ways: schedule mis-match (VOI2 pass), stop mis-match or unreliable odometer values (VOI6 pass), unknown time deviation pattern (VOI4 fail) or unknown distance deviation pattern (VOI7 fail). The first two represent assumed patterns for invalid data and the latter represent unexplained patterns for outliers.

Known error patterns (VOI2 pass and VOI6 pass)

Figures 22 and 23 demonstrate some suspect trips with assumed error patterns. In the following four figures, the black line represents the scheduled trip with labels on the right side identifying timepoints locations. The red line represents the recorded trips with labels on the left side identifying the locations of all designated stops. Time is shown relative to the scheduled trip start time.

Figure 22 shows a suspect trip flagged in test VOI2, mis-matched schedule. According to the APC/AVL database, this trip was scheduled to start at 2:28pm but it is recorded to have started 20 minutes early. However, it is highly unlikely that the operator would start a trip 20 minutes early; if the bus was at the terminal early, it likely waits at the terminal until the scheduled departure time. Furthermore, this route has a 15 minute headway before 2:28pm and a 30 minute headway after 2:28pm adding to the likelihood that the trip recorded in the database has been matched to the 2:28 trip and should have been matched to an earlier scheduled trip.

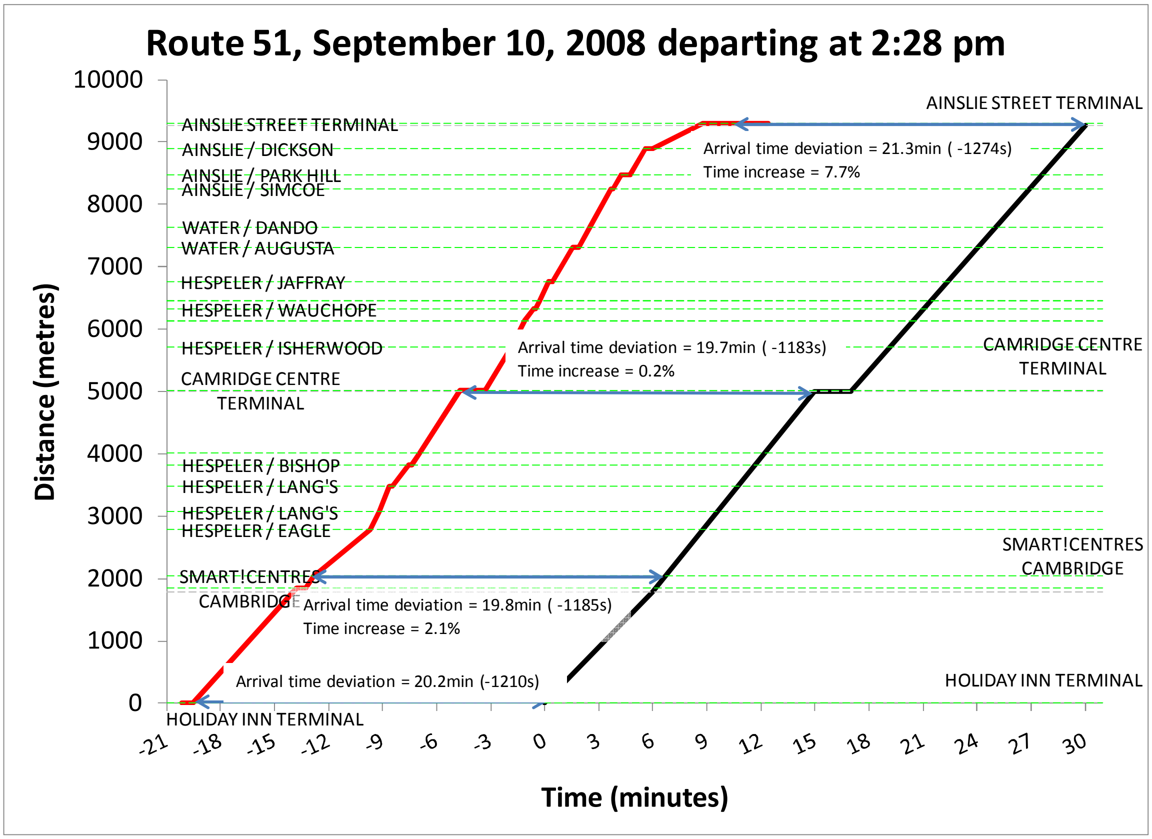

Figure 22 Example suspect trip flagged as schedule mis-match

The pattern used to identify this trip as a mis-match is a uniform schedule time deviation at each timepoint. Figure 22 shows a time deviation of approximately 20 min at each timepoint on the route. The consistency of the time deviations at each timepoint is measured by the percent change in the time deviation value from the previous timepoint. Sample calculations for the arrival time deviation are shown on Figure 22, however departure time deviation is also incorporated into test VOI2. Recall the maximum time deviation increase (P9) is set to 10%.

Figure 23 shows the patterns characteristics of a suspect trip flagged for stop mis-match.

According to the APC/AVL data, the trip completed Route 8 in half the scheduled trip distance (the schedule route is 13km long). The error is visible at the second stop where the archived data suggests that only 300m is travelled to reach Charles Street terminal from Fairview Mall; the actual distance is about 6.5km. As a result, there is a uniform distance deviation among all stops of approximately 6.5km. Figure 23 displays the distance deviation and percent increase only timepoints to avoid overcrowding the plot; however data for all designated stops are subject to test VOI7 because a schedule distance is available for all stops. Both examples in Figure 22 and 23 demonstrate how known error patterns can screen out invalid data.

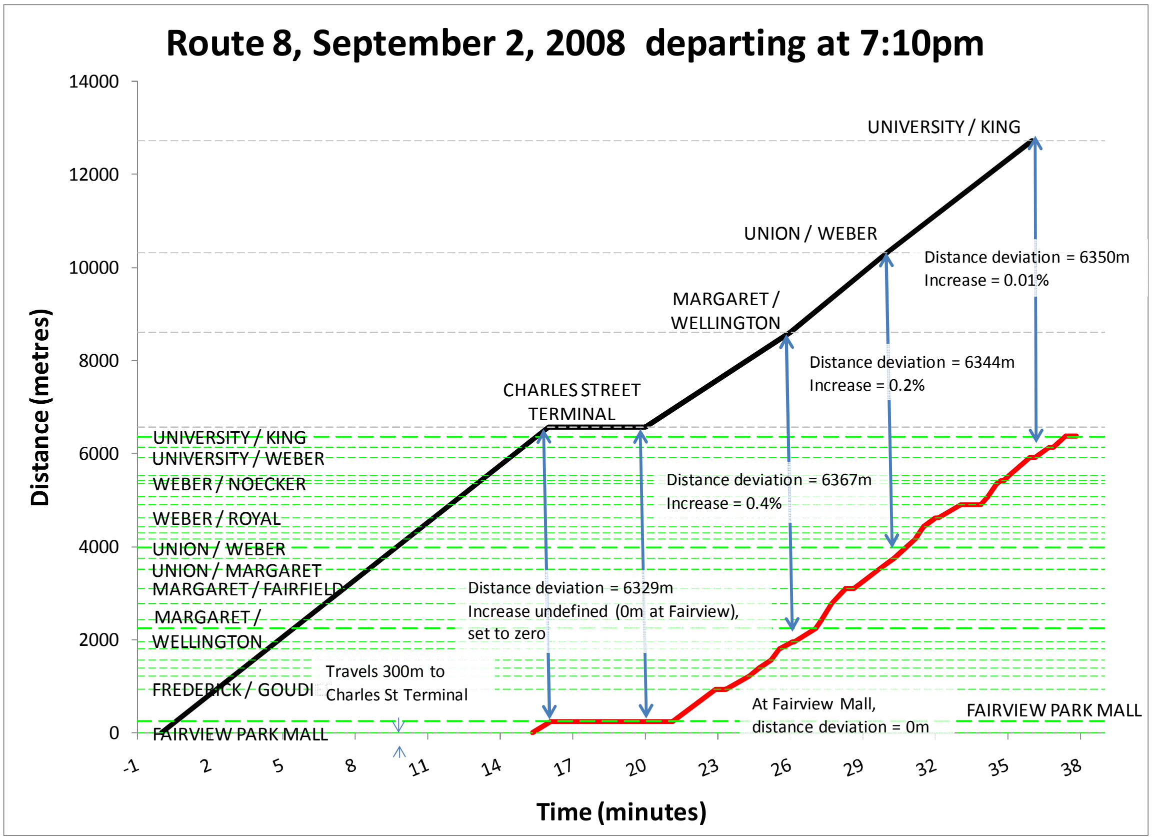

Figure 23 Example trip with a mis-match to stop locations

Unknown error pattern

When an outlier trip fails all tests associated with a valid case option, it is considered to have an unknown error pattern. The proposed QA procedure identifies these trips in VOI4 fail and VOI8 fail. Figure 24 shows a trip that results in VOI4 fail, unknown time deviation.

At first glance, the trip in Figure 24 appears to be a valid case of congestion or operational delay that should be recognized by test VOI3. The test assumes that the time deviation either increases or remains the same and uses a maximum time decrease (P10= -5%) threshold to identify when the time deviation gets smaller. For this trip, the time deviation decreases beyond the time decrease threshold at Shantz Hill/Preston from Sportsworld. Although a change in P10 may include this trip, not many trips have an unexplained time deviation outlier as shown in Table 21. The trade-off of a more relaxed parameter value is the potential to allow more invalid data to pass as valid case trips. This trip is an example of the limitations of the proposed QA procedure for recognizing valid from invalid data; ambiguous trip patterns make it difficult to assess whether some trips constitute a schedule mis-match or valid case of congestion/operational delay.

Figure 25 shows a suspect trip with an unknown explanation for distance deviation outliers. The first distance deviation outlier occurs at Charles Street terminal where the recorded distance lags the scheduled distance by 2.9km; however this trip does cannot represent a detour where excess kilometres is coupled with uniform deviation.

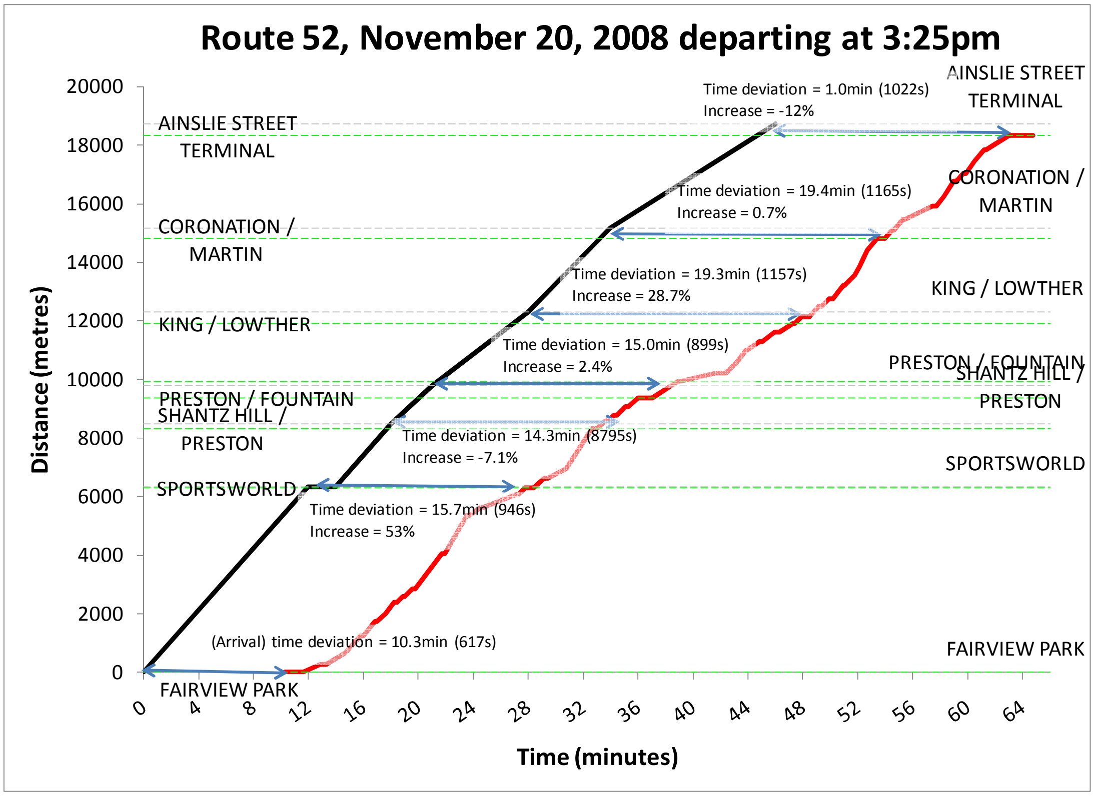

Figure 24 Example suspect trip flagged due to unknown time deviation

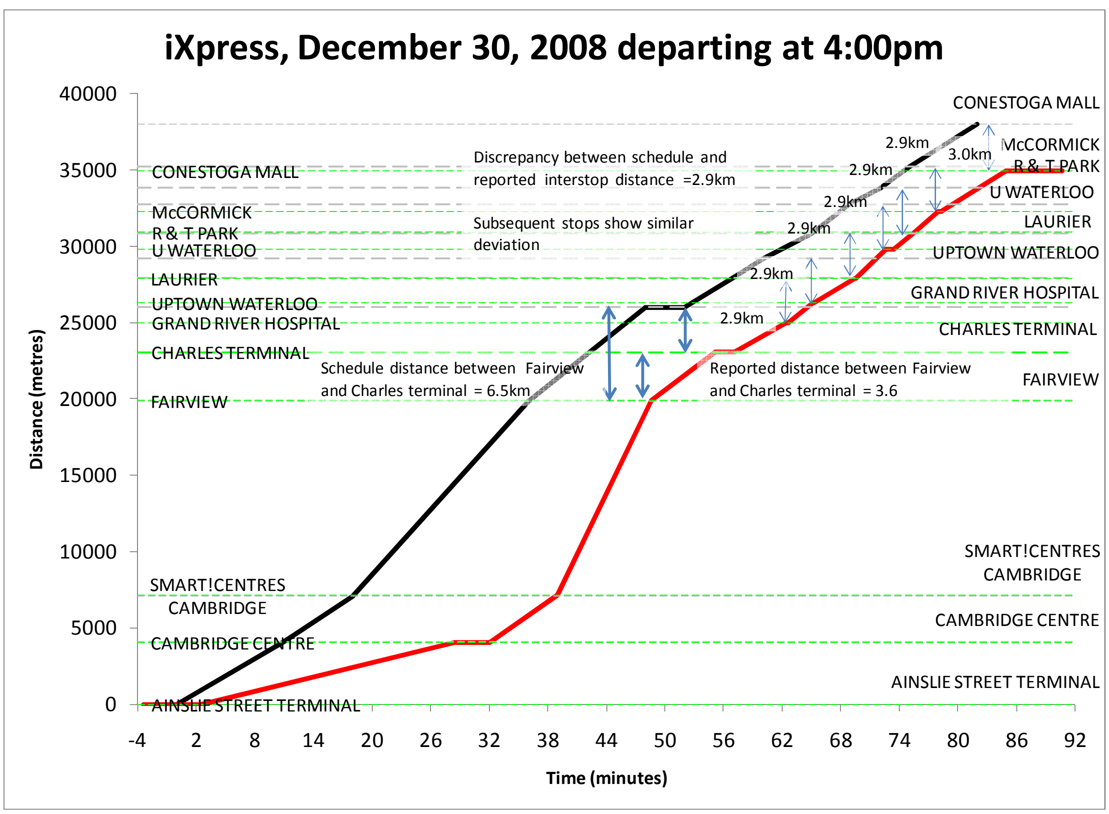

Figure 25 Example trip without explanation for distance deviation outlier

The trip in Figure 25 shows the vehicle arriving at Charles Street terminal from Fairview within 3km; the scheduled distance between these two stops is about 6.5km. Even if the vehicle travelled along a shortcut the minimum travel distance required is 6km. Therefore this trip is a good example of how unexplained patterns for distance deviations do suggest unreliable data.

5.1.2. Analysis of Valid Case Trips

Table 23 is a summary of trips identified as having a valid case. Without the VOI stage, these trips would be considered suspect.

Table 23 Summary of valid case outliers

Valid case	Test	No. Trips	% Total
Total number of time deviation outliers	*OI-2 Fail*	*252*	
Vehicle incident	VOI-5 Pass	10	4.0%
Congestion or operational delay throughout trip	VOI-3 Pass	141	56.0%
Congestion or operational delay during part of trip	VOI-12 Pass	83	32.9%
		232	92.1%
Total number of distance deviation outliers	*OI-3 Fail*	*747*	
Detour	VOI-8 Pass	64	1.8%

The valid case options explain the majority (92.1%) of time deviation outliers however, not many (1.8%) distance deviation outliers are explained by valid cases. One potential reason is that the stop identification algorithm is less accurate than the schedule matching algorithm. Another explanation is that the time deviation tests are applied to a smaller sample of timepoints; distance deviation tests are conducted at each designated stop because scheduled distances are available for each stop. Figures 26 to 30 show example trips of valid case outliers.

The iXpress trip in Figure 26 is supposed to travel from Conestoga Mall to Ainslie Terminal, however no records appear after Ottawa. Before the deviation point, all previous stops are within schedule time and distance deviation thresholds. The trajectory shows the bus remaining at Ottawa for 22 minutes. This dwell time is a result of the given departure time for the Ottawa stop event record, which was likely assigned when the AVL/APC system was shut off or departed for the garage. The GRT keeps track of “change-offs”; these are instances of when in-service vehicles are replaced with spare vehicles. Supervisors write reports about notable operational events. According to the change-off records, the vehicle servicing this trip was changed-off at Ottawa with a non-equipped bus after steering problems were identified.

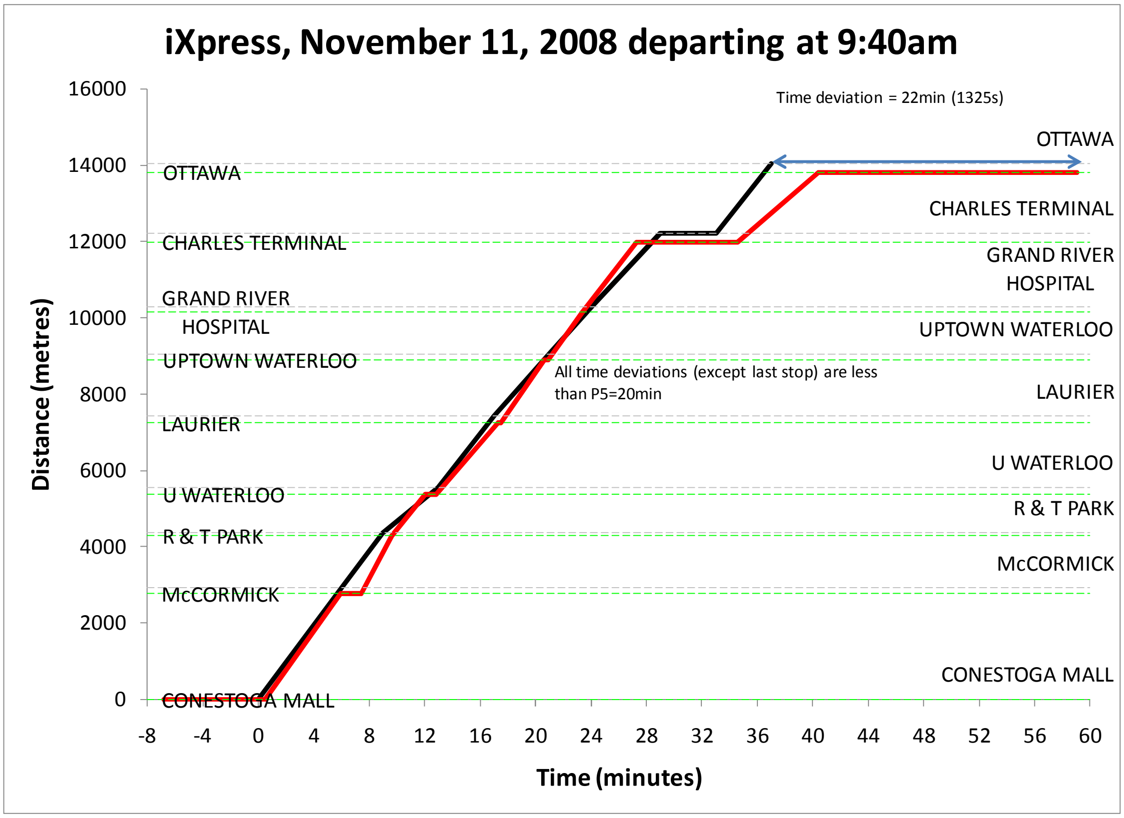

Figure 26 Example trip of a vehicle incident

Figure 27 is another example of an iXpress trip travelling in the other direction; a change-off was also recorded at Charles Street terminal. After the deviation point, the vehicle incident test (VOI5) looks for a continued uniform time deviation. However, most trips identified as a vehicle incident show the deviation point at the last recorded stop; this pattern is likely due to change-offs of in-service equipped vehicles with non-equipped vehicles. Since change-off records are recorded internally but not integrated with the AVL/APC database, future development of the QA procedure could include them as a confirmation of VOI5 results.

Figures 28 and 29 show trips identified as congestion or operation delay. The former figure demonstrates this valid case for the entire trip (pass VOI3) and the latter figure shows delay for only part of the trip (pass VOI4). Since there are no available records of traffic conditions experienced by transit vehicles, it is difficult to confirm whether these trips actually encountered congestion. Instead, the patterns in Figures 28 and 29 can be compared to the intended patterns that tests VOI3 and VOI4 are meant to capture.

Figure 28 follows the congestion pattern assumptions for test VOI3, the time deviation either increases or remains the same. The smaller slopes represent lower average speeds between stops and imply more delays. External weather data on December 19, 2008 reveal exceptional weather events; there was 8.5 cm of snow (16.6 mm precipitation equivalent) reported at the University of Waterloo weather station. Additional investigation of the weather data shows that most of the snowfall occurred from 9am to 1pm. This weather data further suggests that this trip likely experienced congestion or operational delay on the given date and time.

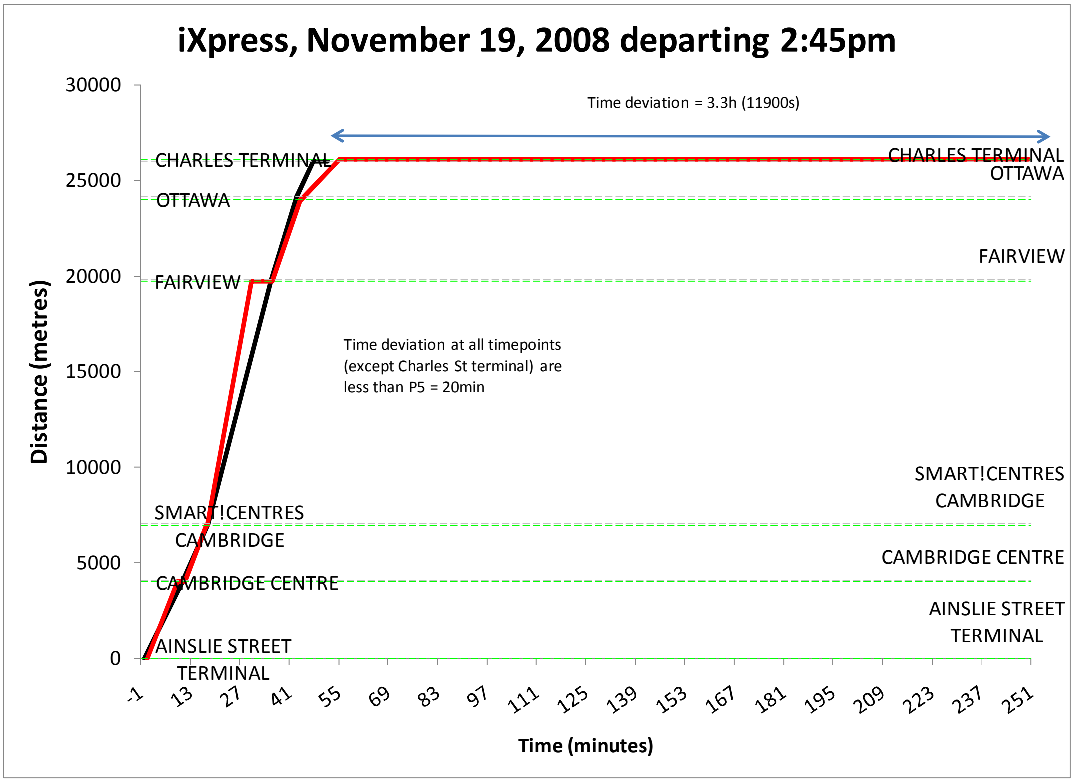

Figure 27 Second example trip of vehicle incident

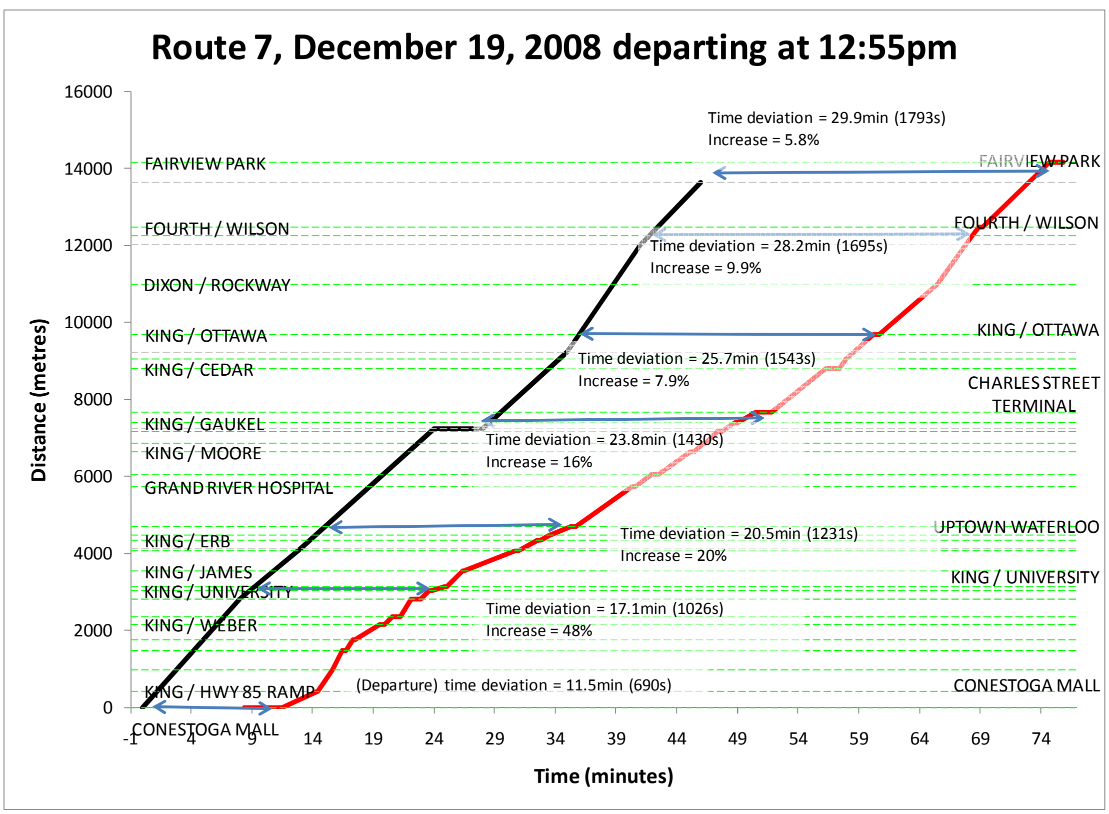

Figure 28 Example valid trip due to congestion/operational delay

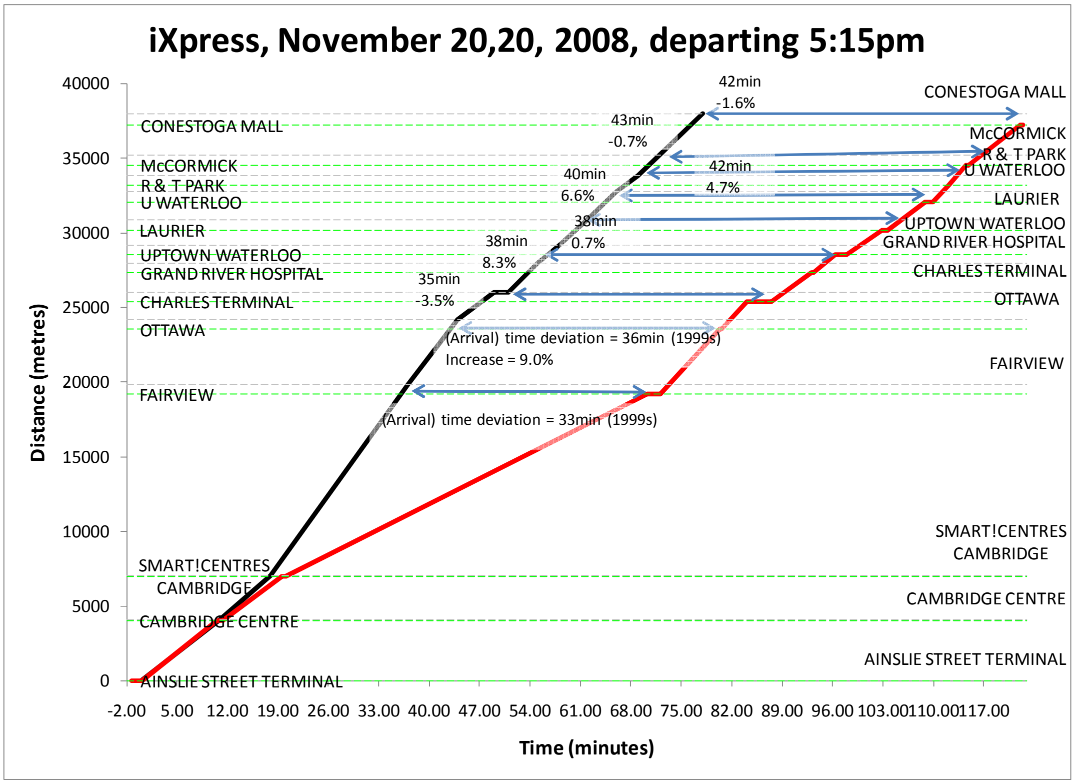

Figure 29 Example valid trip due to partial congestion/operational delay

Figure 29 demonstrates a similar pattern as Figure 28 after the deviation point. This trip was flagged by test VOI4. After the deviation point (Fairview), time deviations at the following stops either increase or remain the same (i.e. no decline is experienced below P10 = -5%). The segment from Smart Centre to Fairview Mall requires the vehicle to travel on Highway 401 and Highway 8; those highways are prone to congestion during peak AM and PM periods. The fact that this trip started at 5:15pm (i.e. conventionally considered rush hour) further suggests that this trip is indeed a valid case of congestion or operational delay.

Figure 30 shows an example valid trip identified as a detour. Similar to the congestion or operational delay cases, there are no records for transit vehicles when they detour. Again, the intended pattern is instead reviewed. Detours are generally expected to travel a longer distance than the scheduled distance, therefore excess kilometres is expected in a detour pattern. In the trip shown in Figure 30, the last stop is the deviation point and all previous stops are within time and distance deviation threshold.

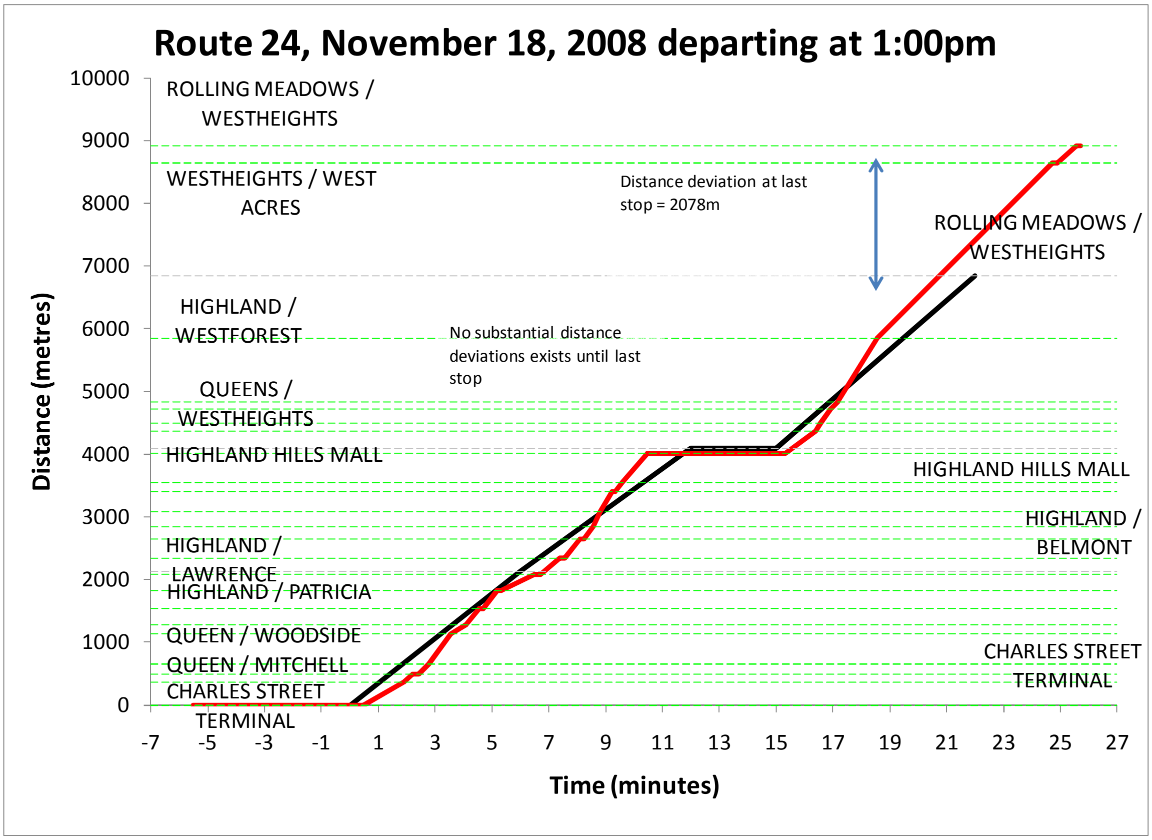

Figure 30 Example trips of a detour

Confirmation of the example trip in Figure 30 as a detour can be seen by plotting the GPS coordinates on a map (Figure 31).

In Figure 31, the red line represents the scheduled route path and the points represent recorded stops (orange signify stops in the schedule and green represent non-scheduled stops). Note that the coordinates for the segment from Queens/Westheights to Rolling Meadows/Westheights go off the route path; the route path shows that the bus should travel north on Westheights Dr past Blackwell Dr to west on Highland Rd W, south again on Westhights Dr, followed by a counter-clockwise loop on Rolling Hills and Driftwood Dr. However, the coordinates for the Westheights/Blackwell stop shows that the bus was actually south on Elm Ridge Dr.

Based on the distance travelled at subsequent recorded stops, the bus likely detoured via south on Elm Ridge Dr to east on McCarry Dr to north on Westheights Dr and continued it scheduled path north of Queen Blvd (coordinates are missing for the Highland/Westforest stop). This detour is approximated 2.2km and may be the result of an obstruction along Queen Blvd from Elm Ridge Dr to Westheights Dr.

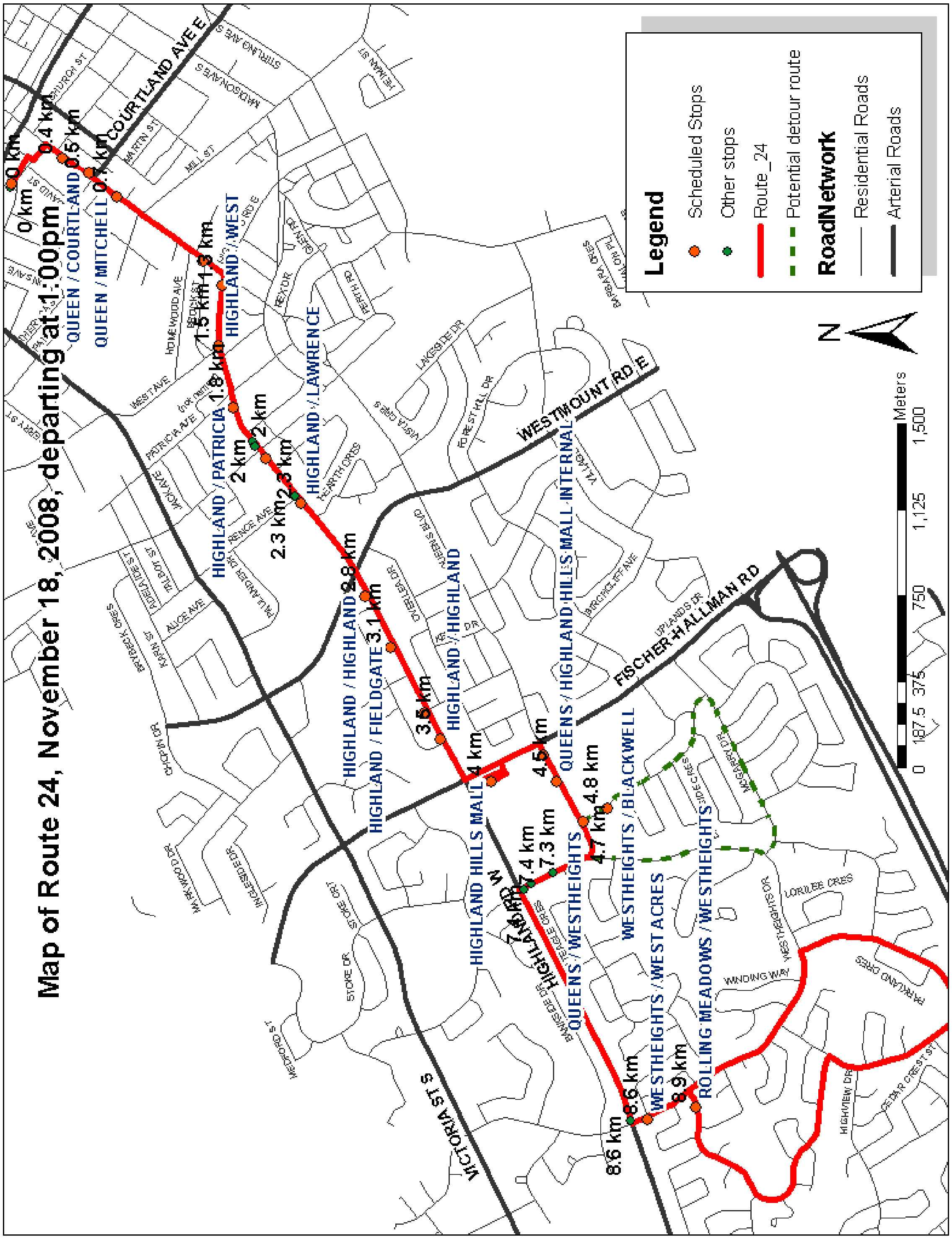

Figure 31 Route path of detoured trip

5.1.3. Analysis of non-suspect trips

Table 24 breaks down the trip-level data availability by route-type.

Table 24 Impact of QA procedure on data availability by route-type

	Without QA	With QA	% Suspect
Total number of trip records			
Regular	15,124	12,309	18.6%
iXpress	9,608	8,949	6.9%
Special	289	180	37.7%
Average number of trips per route, both directions			
Regular	297	241	18.8%
iXpress	9,608	8,949	6.9%
Special	7	4	38.1%

Table 24 shows that iXpress routes have a better penetration rate through the QA procedure than regular and special routes. Table 25 is a breakdown of the suspect trips by route type.

Table 25 Summary of suspect trips by route type

Reason to suspect data	Test	Regular	iXpress	Special	All trips
Time does not increment forward	BC1 Fail	91	0	1	92
Distance does not increment forward	BC2 Fail	1	0	0	1
Unreasonable time or distance step	BC3 Fail	1108	206	12	1326
Unreasonable travel speed	BC4 Fail	94	2	0	96
Passenger count greater than bus capacity	OI1 Fail	2	12	4	18
Raw passenger count over-corrected	OI4 Fail	1102	335	74	1511
Single outlier timepoint deviation	VOI0 Pass	1	3	0	4
Suspected mis-match in schedule	VOI2 Pass	4	5	1	10
Suspected mis-match in stop locations	VOI6 Pass	117	13	3	133
Unknown reason for large schedule time deviation	VOI4 Fail	9	2	0	11
Unknown reason for large schedule distance deviation	VOI7 Fail	407	136	34	577

Recall Table 16 in Section 4.2.2; the distribution of trips in the database is 60/39/1 for regular, iXpress and special routes. A similar distribution should be expected for the suspect trips if the data for each route-type has an equal penetration rate through the QA procedure. Table 25 shows that regular route trips appear to contribute a larger share of suspect trips for most tests (BC1, BC3, BC4, OI4, VOI7 and VOI8). Special routes show a greater than 1% share in tests OI1 and VOI8. Overall the distribution of suspect trip by route-type seems to suggest that iXpress trips produce higher quality data.

One potential explanation for the higher iXpress penetration rate is the quality of schedule data. The AVL/APC system was initially designed to be implemented for the iXpress route as part of a larger ITS package. Limited stops also allow for less stop records per trip to be tested and all designated stops are also time points; this detail makes stop and schedule matching simpler and could result in a more successful matching algorithm. Alternately, the low penetration rate for special route trips is likely due to poor schedule data quality in special routes. Maintenance of special route schedule data is usually less thorough due to low priority. One management strategy to increase post-QA sample size by route is to improve the quality of schedule data.

Regular and special routes are already at a disadvantage with respect to sample size due to a lack of equipped vehicles servicing those routes; however service planners are generally more interested in regular route trips for performance monitoring and future service planning. Performance of iXpress trips is already well-documented by GRT staff. Therefore, the lower penetration rate for regular routes is of particular relevance for GRT management (Figures 32 and 33 show the sample sizes for each regular route).

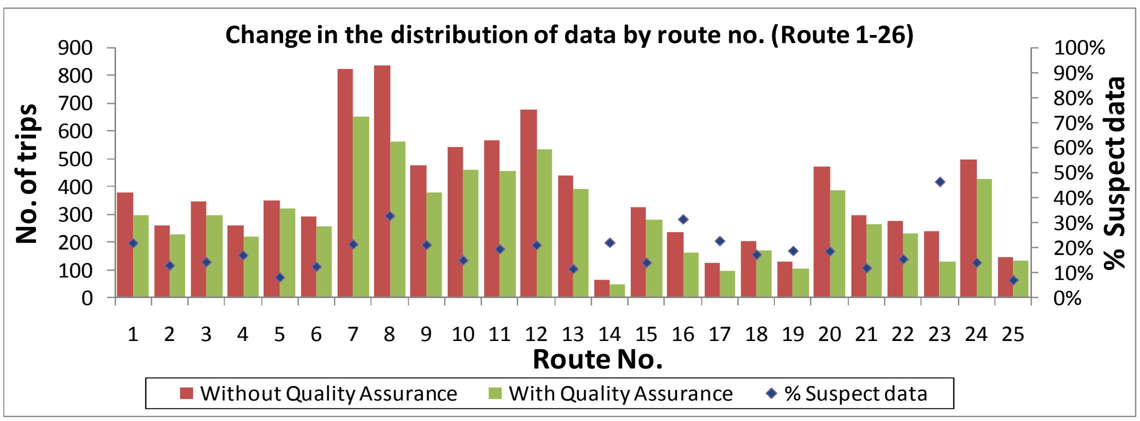

Figure 32 Sample size by route (Routes 1-26)

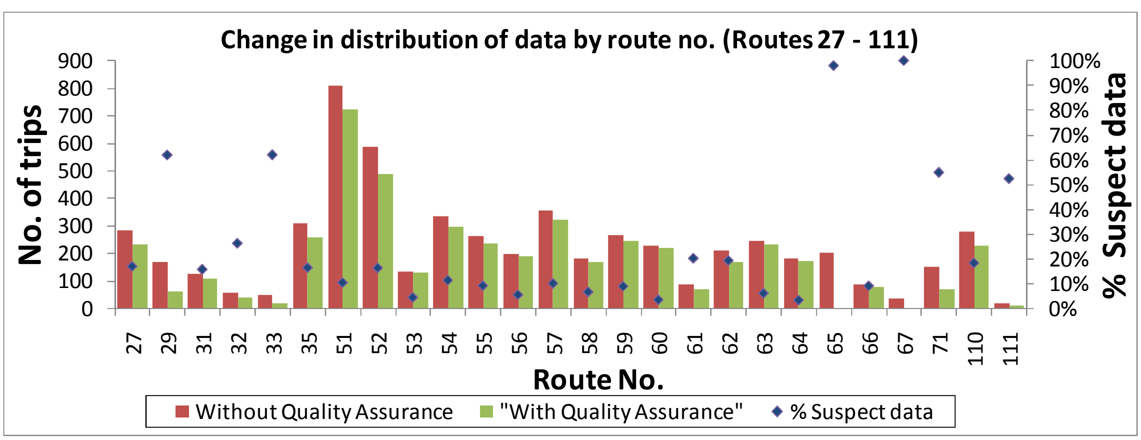

Figure 33 Sample size by route (Routes 27-111)

It is apparent from Figures 32 and 33 that the sample size and penetration rate varies considerably by route; higher frequency routes tend to have larger sample sizes and two routes in particular have unusually high suspect data rates (Route 65 and 67). Figure 35 demonstrates why most Route 67 trips are flagged by the QA procedure (Figure 34 shows the route configuration).

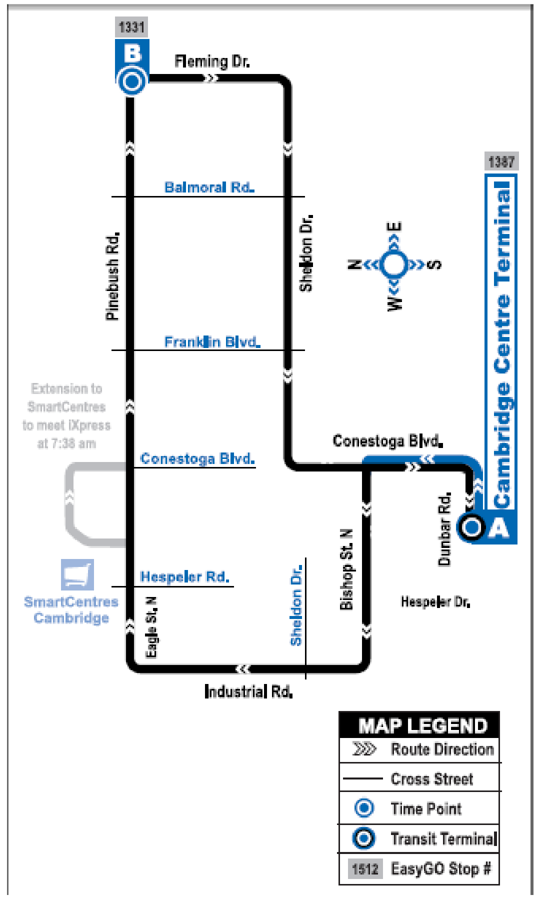

Figure 34 Configuration of Route 67 (Source: Grand River Transit, 2010)

On closer analysis, the single direction loop route matches the last stop to the first stop in the schedule data. In a loop route, the last stop has the same location as the first; however the schedule data should still have separate stop definition for the last stop of this route with a different schedule distance. Instead the last stop is matched with the first stop definition as shown in Figure 35 (the black line represents the scheduled trip and the red line represents the recorded trip). The matching of the last stop to the wrong stop definition is associated with a problem in the schedule or stop location data; schedule matching processes occur before the QA procedure is applied. Higher quality schedule data could avoid this problem.

Another sample management strategy is better coordination with operations to increase use of equipped vehicles in service. Figures 36 and 37 show the distributions of trips by vehicle.

Figure 36 shows an even distribution of equipped buses among iXpress routes; however Figure 37 shows that some equipped vehicles are under-utilized for data collection.

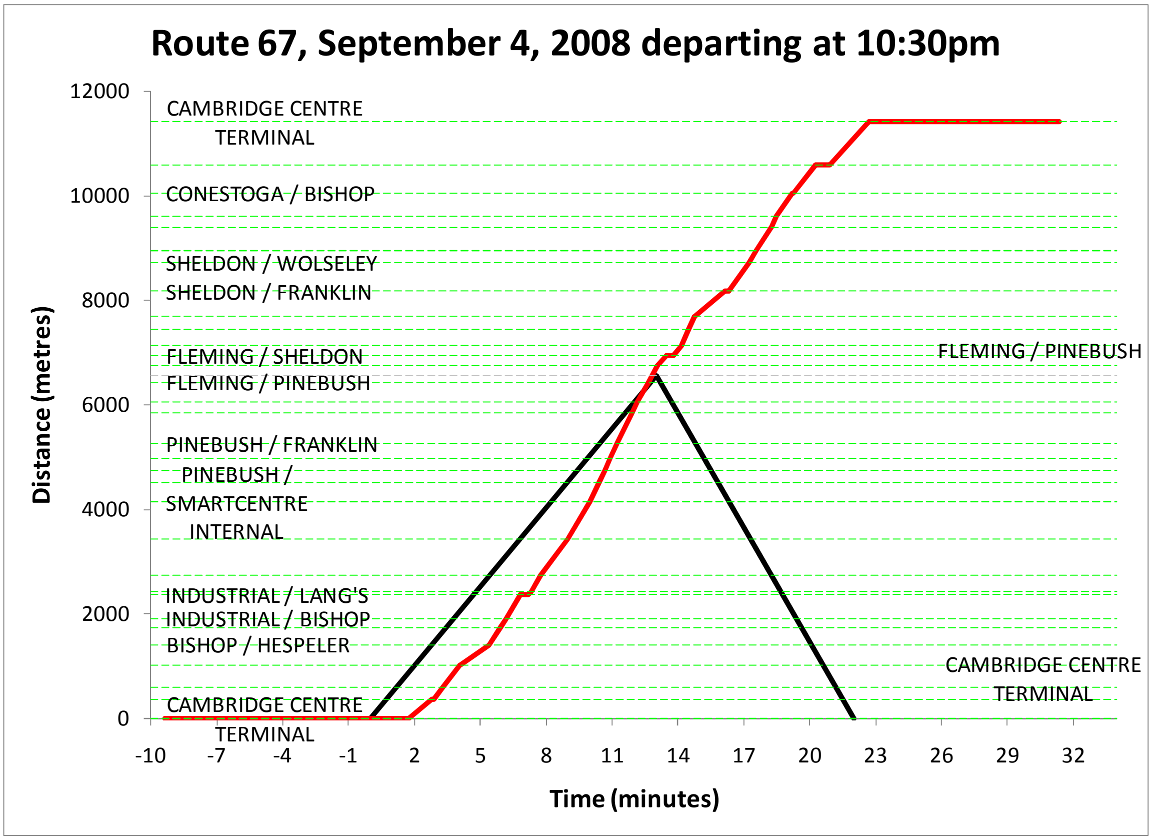

Figure 35 Example Route 67 trip

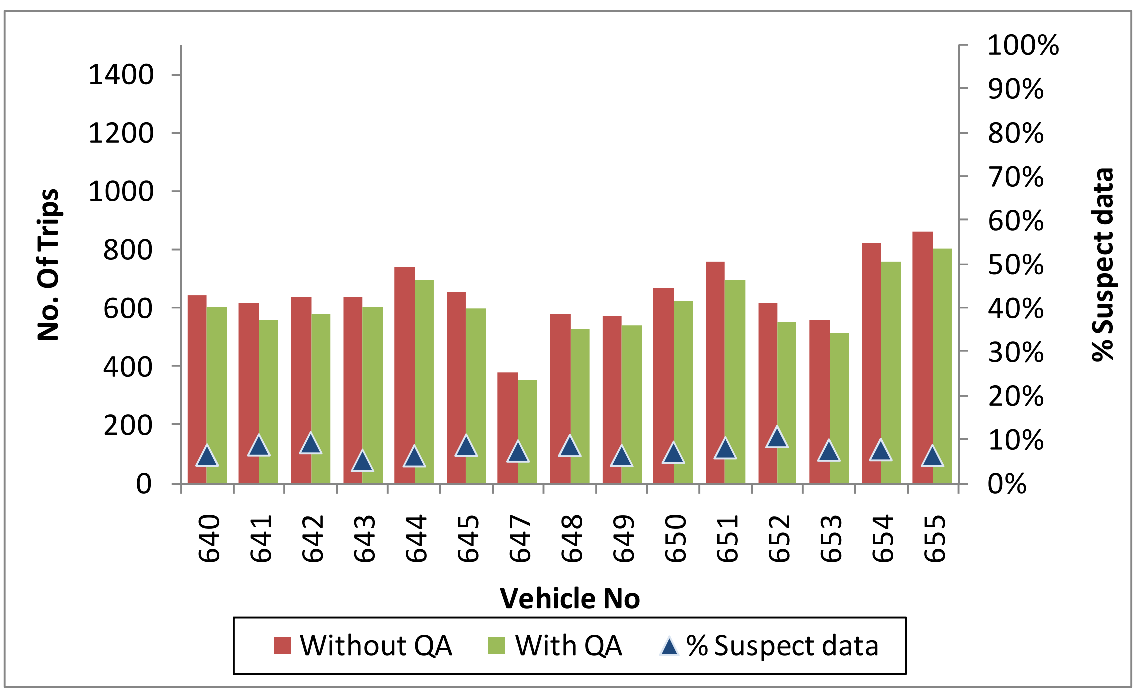

Figure 36 Distribution of trip sample by vehicle for iXpress buses

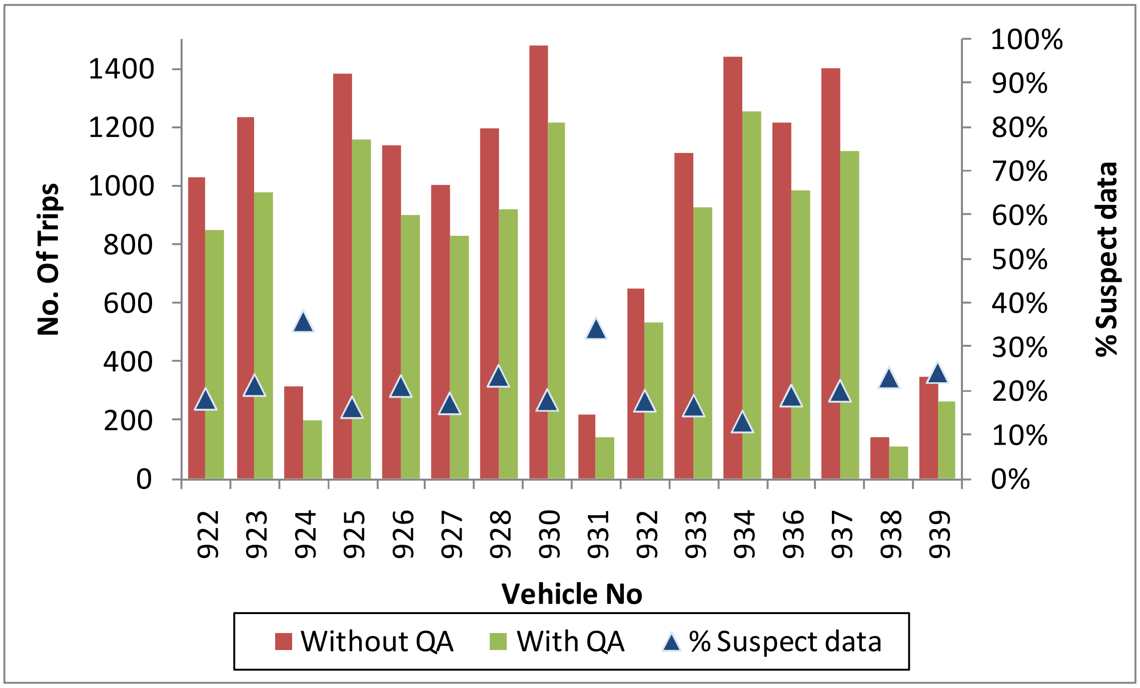

Figure 37 Distribution of trip sample by vehicle for regular route buses

Sampling plans may be useful to more effectively utilize APC/AVL equipped buses. In fact, information from Figures 36 and 37 in conjunction with the information from Figure 32 and 33 (sample size by route) may be useful in formulating a sampling plan.

Figure 37 shows a larger percentage of suspect data for vehicles No. 924 and No. 931. Figure 38 is a distribution by route for the sample data from these vehicles. The figure shows that the vehicles usage was distributed among various regular and special (9000-series) routes, therefore the higher rate of suspect data cannot be attributed to a specific route with poor schedule data. The dashed line represents the vehicle-average percentage of suspect data from Figure 37. The vehicle-specific suspect data rate in Figure 38 varies by route and is higher than the rate shown in Figure 36 (average from all vehicles).

For example, for vehicle No. 924, the largest portion of the data comes from routes 7 and 12 and these trips show a suspect rate of 54% and 50%, respectively. Figure 36 shows a lower 21% suspect data for both these routes from all vehicles. Similarly in vehicle No. 931, the largest portion of trips comes from routes 62 and 71 and the suspect rates shown in Figure 38 are 78% and 62%, respectively. However the route-average from all vehicles is 19% and 55%, respectively. Therefore, poor quality AVL/APC data (shown by a higher percent of suspect data) is more likely contributed by the vehicle and not specific to the route. The recognition of higher suspect data rates for vehicles No. 924 and No. 931 (as shown in Figure 37) is an example of how this information can be used to identify maintenance needs of equipped vehicles; these vehicles are likely contributing to poor data quality due to poor calibration or defective equipment.

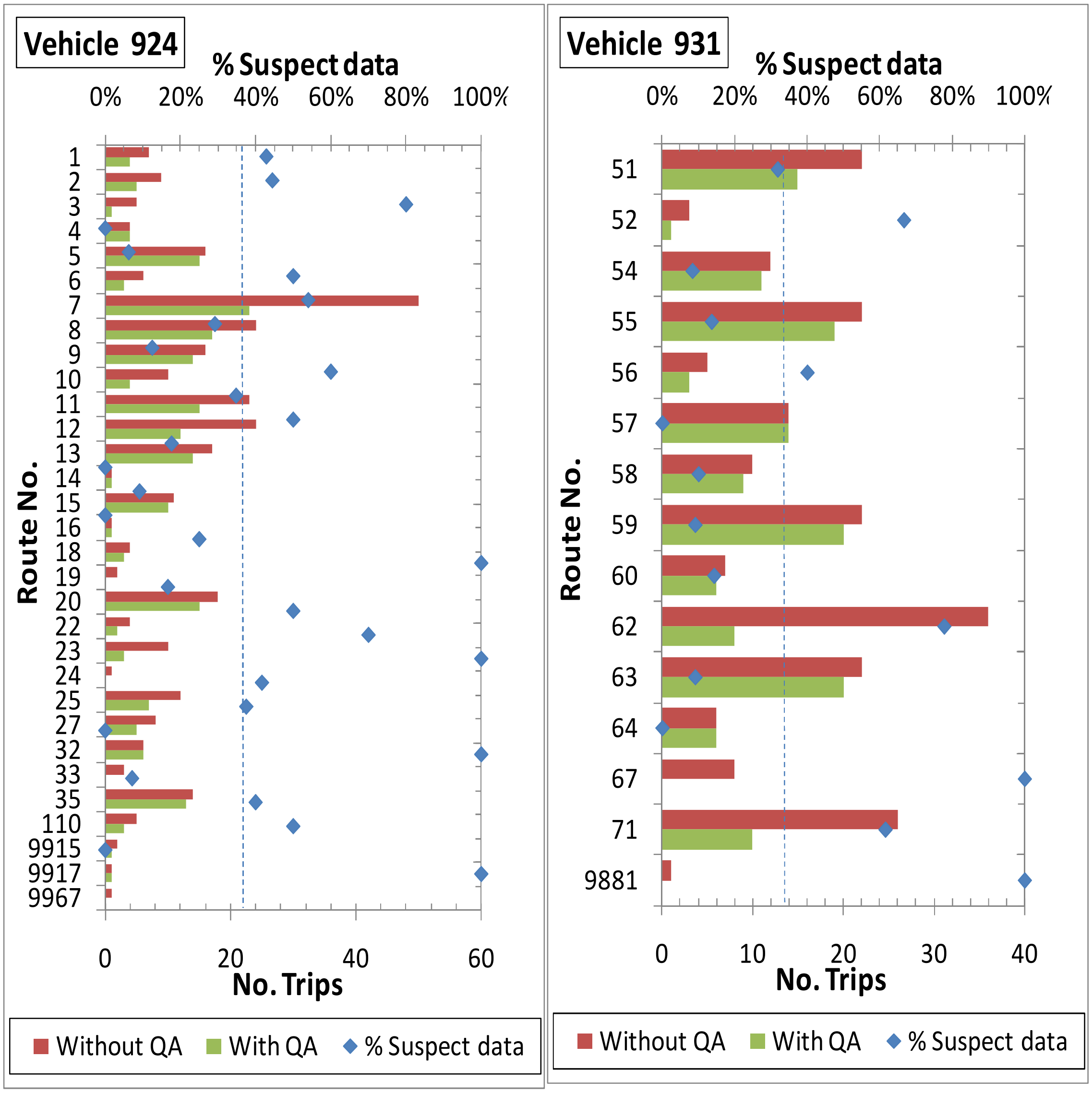

Figure 38 Sample size by route for vehicles No. 924 and 931

Lastly, the distribution of trips by time period and day of the week are reviewed. Figure 39 is the trip distribution by time period and Figure 40 shows the distribution by day of the week. There does not seem to be any discrepancies in the percent of suspect data by time. The 10pm & later time period may show a slightly lower suspect rate, but there are less routes operating during this time period possibly leading to a better schedule matching result.

Figure 38 shows higher suspect data rates for weekday trips. Since regular and special routes are only sampled on the weekday, it is likely due to the contribution of suspect data from regular and special route, which average at 20% and 29%, respectively. The average weekend failure rate is close to the iXpress rate of 7%.

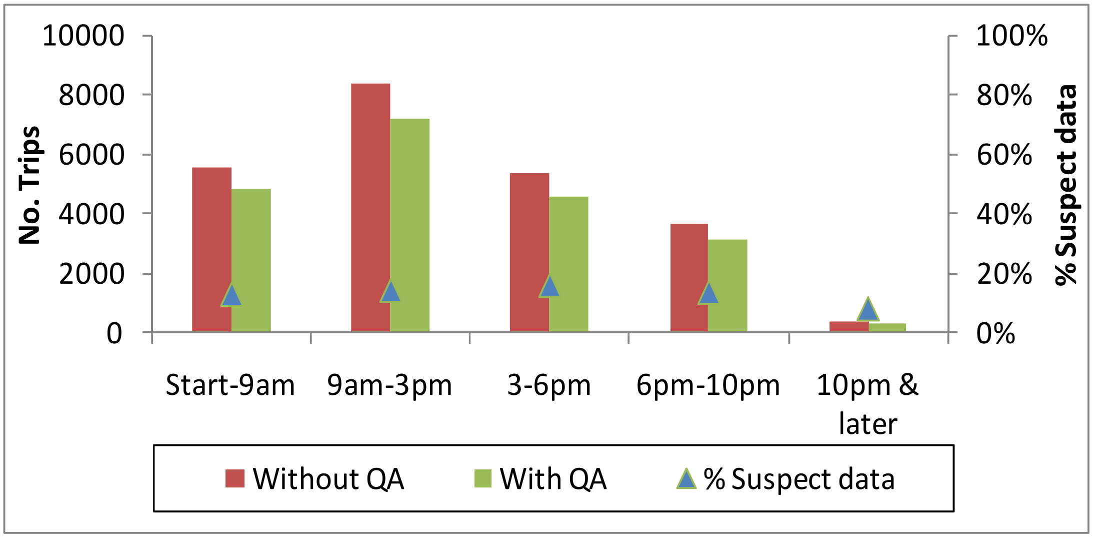

Figure 39 Distribution of trips by time of day before and after QA

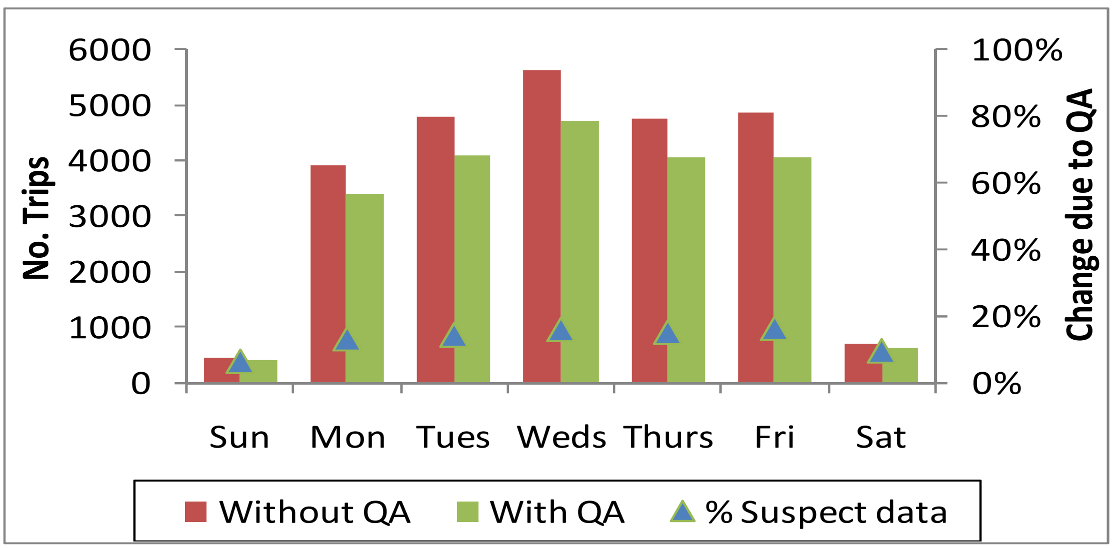

Figure 40 Distribution of trips by day of week before and after QA

5.2 Sensitivity Analysis

A sensitivity analysis was performed to determine the impact of changing parameter values and changes in the QA structure. The analysis was conducted by comparing the QA procedure output in terms of the number of suspect trips identified versus the change in parameter value. Additionally, a no-test value was generated to determine the impact on the QA output if the given test was not included in the QA procedure. Since most tests are based on a maximum threshold, the common pattern for the sensitivity plots are an increase in the number of suspect trips for smaller (more stringent) parameter values and a decrease in the number of suspect trip for larger (less stringent) parameter values.

Figure 41 demonstrates common sensitivity plot pattern for P2, maximum distance increment.

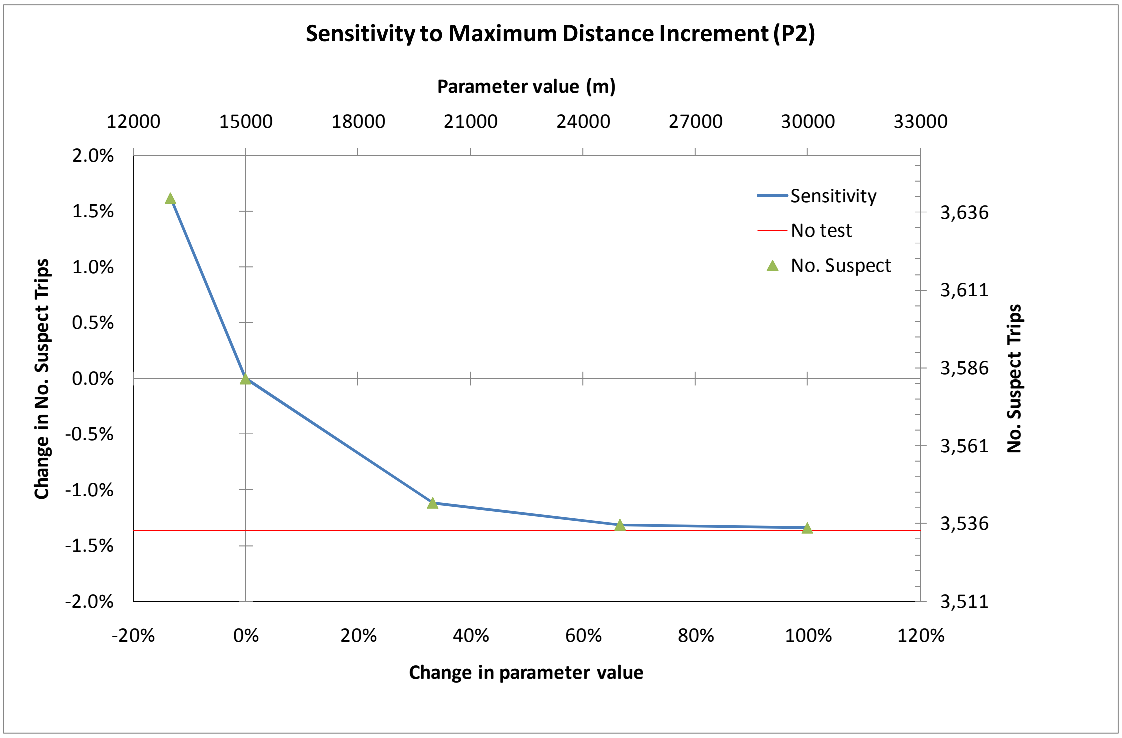

Figure 41 Example sensitivity plot for P2, maximum distance increment

Although the slope between the tests points appear large in the plot, the range for percentage change of suspect trips remains between +/- 5% for most parameters. Figures 42 to 44 show parameters with greater ranges suggesting increased sensitivity. The no-test curve in Figure 41 represents the number of suspect trips that would be otherwise identified if the QA procedure did not test for large distance increments. Another interpretation of the no test curve is that the parameter value is set high (or low) enough such that no trip would fail the test.

For most tests, the no-test curve is below the sensitivity plot because a removal of a QA test would generally result in fewer suspect trips identified. For parameters related to valid case tests, the no-test curve is above the sensitivity curve because those valid cases outliers would otherwise be considered suspect. Figure C9 in Appendix C demonstrates this pattern for P10, maximum distance increase, which is used to identify detours. One parameter that did not follow this trend for the no-test curve (i.e. to be above or below the sensitivity curve) is P9, maximum time increase. P9 is used in both tests VOI2 and VOI5, where the former represents an invalid case (i.e. schedule mis-match) and the latter represents a valid case (i.e. vehicle incident). Ten trips were identified for each these tests; therefore the no-test curve is zero because the exclusion of both tests results cancel each other. However, relative changes the results of test VOI2 and VOI5 with respect to P9 is associated with the sensitivity curve of P9 (Figure C8 in Appendix C). Although P10 also represents both a valid and invalid case (i.e. VOI6 and VOI7), the removal of these tests would results in all trips with a distance outlier to be considered suspect.

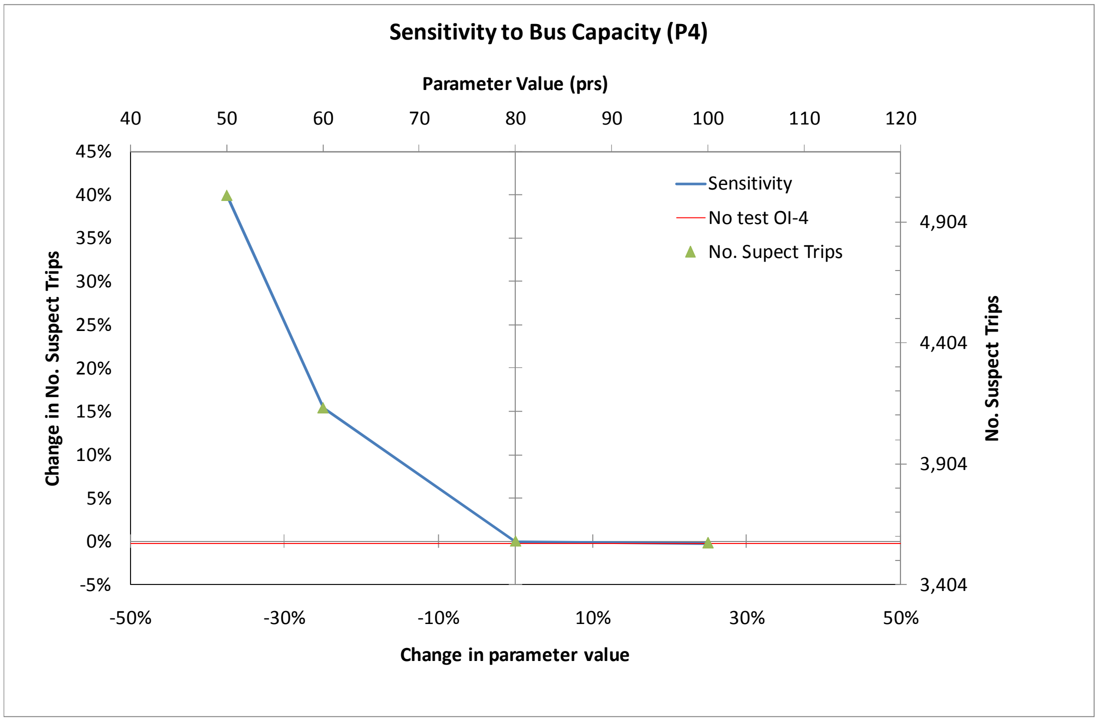

Figure 42 Sensitivity plot for P4, maximum passenger count

Figure 42 shows that the methodology is highly sensitive when P4 is less that the chosen value of 80 persons, but not for larger values. As mentioned, Figure 42 adopts a pattern related to the parameter selection distributions in Appendix B. Figure B3, which is the trip distribution of the largest passenger count, is closely related to the sensitivity plot in Figure 42. The difference between the two plots is that Figure 42 incorporates the impact of test sequencing; BC tests remove some data during the QA procedure before OI tests are applied. Therefore the resulting number of suspect trips may be different from those identified in Figure B3 in Appendix B. The heightened sensitivity for lower P4 values means that there are many trips with maximum trip-level passenger counts between 50 to 80 persons *and* these trips would otherwise not be identified as suspect if not for test OI1.

Figure 43 shows a heightened sensitivity to the value of parameter P6. If the value of P6 is increased from 2km to 5km, there would be an almost 15% decrease in suspect trips.

In Figure B6 in Appendix B, the proportion of trips with largest distance deviation greater than 2km was relatively small (approximately 5%). Therefore, it is unexpected that the QA procedure is sensitive to increasing the value of P6 from 2km to 5km. However, it was also found that most distance deviation outliers were not found to have a valid case during the analysis of suspect trips. Therefore, the sensitivity to P6 is probably due to the fact that most trips with distance deviation outliers end up identified as suspect.

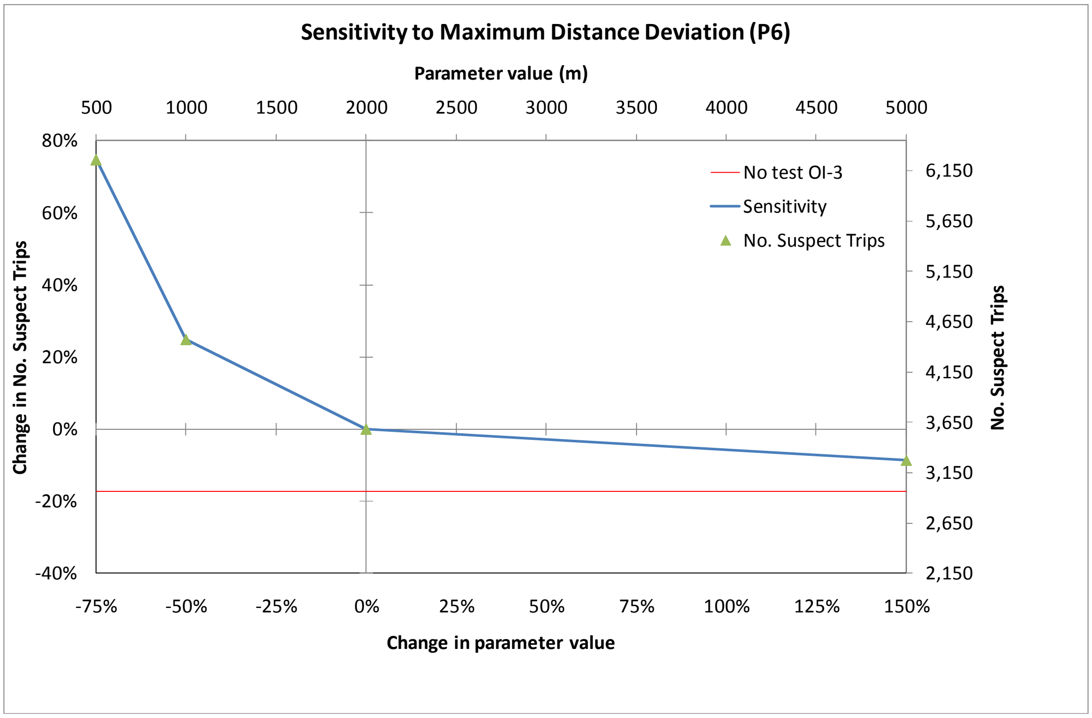

Figure 43 Sensitivity plot for P6, maximum distance deviation

Figure 44 shows that the QA procedure is most sensitive to P7, maximum correction count.

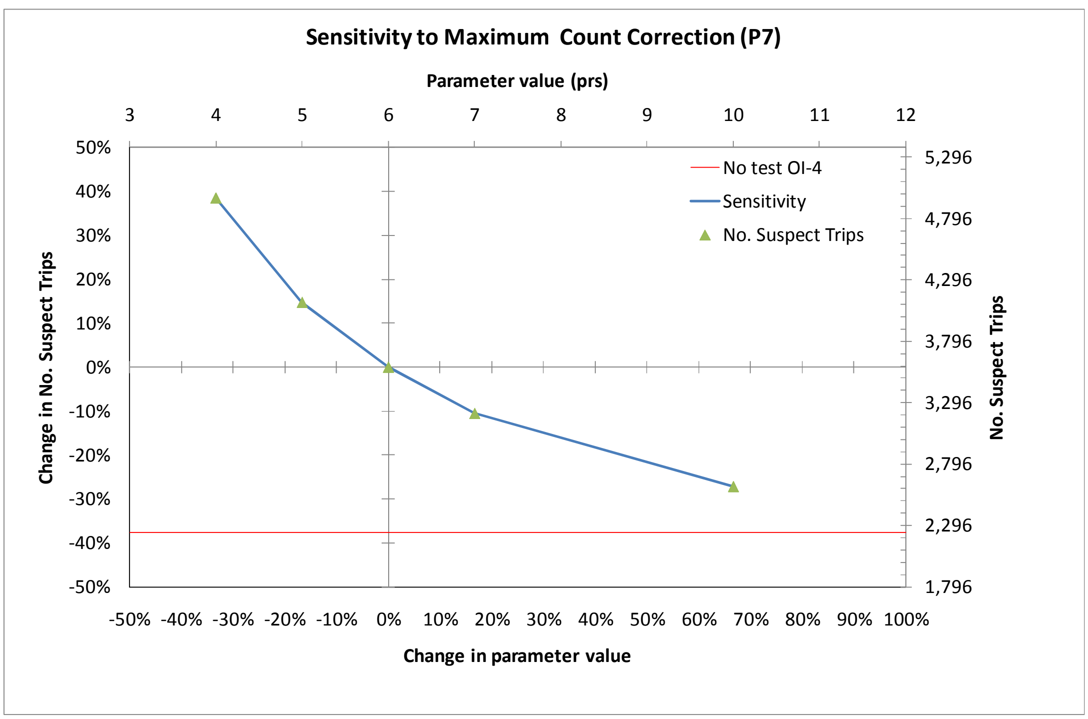

Figure 44 Sensitivity plot for P7, maximum count correction

The right-hand tail of the trip distribution of largest count correction (Figure B7 in Appendix B) spreads further into higher parameter values. As a result, the shape of the sensitivity plot in

Figure 44 is more linear and less comparable to a negative exponential shape. Since unacceptable count correction was the top reason to flag a suspect trip (and there are no valid case options for this outlier), it is intuitive that the number of suspect data changes nearly proportionally to P7. Although Figure B7 in Appendix B shows that only 4.9% of *all* trips have would fail if P7 is greater than or equal to 6 persons, the Figure 44 represents the impact of the change in P7 to the suspect trips. If P7 is changed to 5 persons, Figure B7 shows 6.5% of all trips fail; the difference of 1.6% translates to 414 trips (1.6% × 25,051trips) as shown on the right-hand vertical axis.

Sensitivity plots for each parameter is available in Appendix C. Overall, most sensitivity plots demonstrated a negative exponential shape where the rate of change for suspect trips increases from smaller parameter values and decreases for larger values. The pattern is related to a maximum threshold; most parameter values represent a maximum threshold. Despite the sensitivity to lower parameter values, most sensitivity plots shows change for the number of suspect trip within ±5%.

The QA procedure is most sensitive to changes in P7 (maximum correction count) followed by P6 (maximum distance deviation) and lower values for P4 (crush-load capacity). It is noticed that the test associated with these parameters tend to have no or few valid cases for outlier data. The sensitivity to these parameters can potentially be reduced if more valid case tests are developed.

5.3 Impact on Performance Measures

To assess the cumulative impact of the QA procedure on service monitoring, some performance measures were calculated for the dataset before and after QA is applied. Several parameter sets were created to represent different data consumers, the scenarios range from very data aggressive to very data conservative. Table 26 shows the parameter values associated with each scenario.

Table 26 Parameter values sets for various QA scenarios

Parameter values	**No QA Applied**	Very Aggressive	Moderately Aggressive	**Control**	Moderately Conservative	Very Conservative
P1. Max Time Increment		7200	5400	3600	2700	1800
P2. Max Distance Increment		30000	20000	15000	15000	15000
P3. Max Travel Speed		36.1	33.3	27.8	27.8	27.8
P4. Max Bus Capacity		100	90	80	70	60
P5. Max Time Deviation		2400	1800	1200	900	600
P6. Max Distance Deviation		5000	4000	2000	2500	1000
P7. Max Count Correction		12	8	6	5	4
P8. Min Time Deviation		60	60	60	60	60
P9. Max Time Increase		5%	10%	10%	10%	10%
P9b Max Time Decrease		-10%	-5%	-5%	-5%	-3%
P10 Max Distance Increase		10%	5%	5%	5%	20%

Aggressive refers to a more relaxed set of parameters values; this scenario would reflect a data consumer whom is more concerned with obtaining larger sample size of usable APC/AVL data. Conservative refers to a more stringent set of parameter values; this scenario would reflect a consumer whom is more concerned with removing invalid data and is more willing to discard extreme, yet perhaps valid data. Table 27 shows the output of the QA procedure for each scenario.

Table 27 QA output for parameter sets

Parameter values	No QA Applied	Very Aggressive	Moderately Aggressive	Control	Moderately Conservative	Very Conservative
Totaltrips	25,021	25021	25021	25021	25021	25021
Suspect	0	1937	2508	3583	4054	5395
Non-suspect	*25,021*	*23,084*	*22,513*	*21,438*	*20967*	*19626*
Failure rate		**7.7%**	**10.0%**	**14.3%**	**16.2%**	**21.6%**

As expected, the data aggressive scenarios identify less suspect data due to more relaxed parameter values and data conservative scenarios flag more suspect data.

Three performance measures were evaluated for the sample data set: % time bus is not "on-time", the % sampled trips "under-capacity" and the % sampled trips "over-capacity". The not "on-time" percentage is based on the fraction of not on-time timepoint observations over the total number of timepoint observations. The GRT definition of "on-time" is within zero minutes early and three minutes late. Since the standard is not defined in seconds, an early threshold of 30 seconds is used. The under-capacity percentage is the fraction of trips where the stop-level load was observed to be less than 25% of the seated capacity (9 persons) for 75% of the time before 6pm or less than 10% of the seated capacity (4 persons) for 75% of the time after 6pm. The over-capacity percentage is the fraction of trips where the stop-level loads was observed to be greater than the bus capacity (50 people) for more than 5% of the time. Table 28 is a summary of the network performance for each parameter set.

Table 28 Overall network performance based on parameter sets

Impact on Performance Measures	No QA Applied	Very Aggressive	Moderately Aggressive	Control	Moderately Conservative	Very Conservative
Schedule Performance Measures						
% not "on-time"(bus perspective)	26.7%	26.32%	26.11%	25.7%	25.56%	25.01%
Δ On-time measure (bus perspective)	**0.0%**	**-1.5%**	**-2.2%**	**-3.9%**	**-4.3%**	**-6.4%**
Passenger Performance Measures						
% sampled trips "under capacity"	15.5%	14.3%	14.2%	13.9%	13.9%	13.8%
% sampled trips "over capacity"	5.7%	13.5%	13.6%	13.7%	13.3%	3.6%
Δ Under capacity measure	**0.0%**	**-8.0%**	**-8.8%**	**-10.5%**	**-10.5%**	**-11.0%**
Δ Over capacity measure	**0.0%**	**138.8%**	**139.6%**	**141.9%**	**135.2%**	**-36.6%**

On the network level, the change to the schedule adherence measure is not very large. However, there is more variation when viewing the route-level changes. Figure 45 is a scatter plot of the schedule adherence values by route before and after the QA procedure. The difference in the % not "on-time" measures in Figure 45 before and after QA is small for most routes, however there are substantial differences in some other routes. For example in Route 33, there is a 6% difference between the before and after cases for one direction and a 31% difference in the other direction. Without the QA procedure, this route might be overlooked at the monitoring stage and not considered for service improvements at the planning stage.

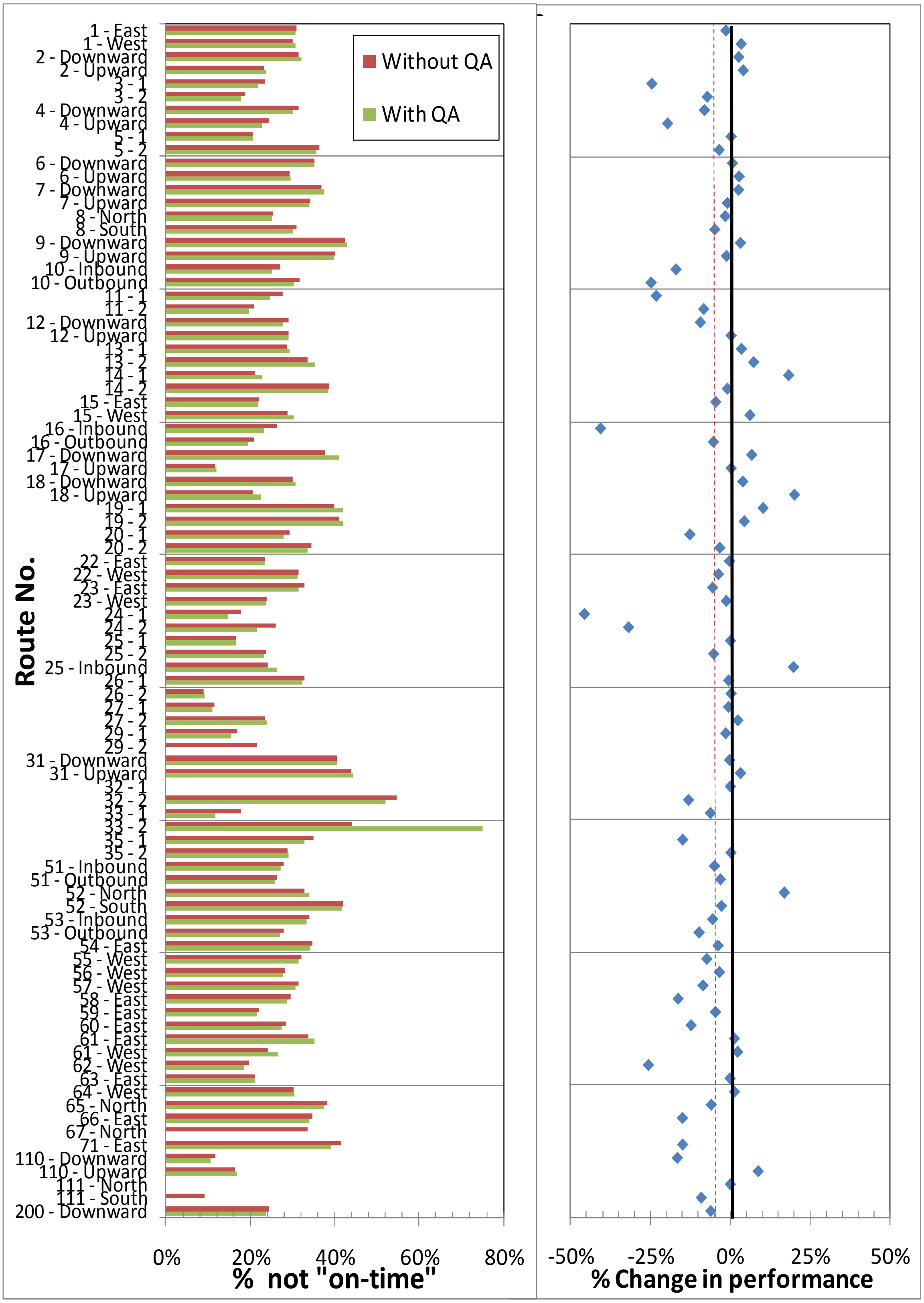

Figure 45 Route-level impact of schedule adherence measure

The impacts to passenger activity measures are more apparent. Figures 46 and 47 show the change in the under-capacity and over-capacity measures, respectively. For most routes, the under-capacity measure was changed moderately for most routes. Some route observed a reduction to zero percent under-capacity after the QA. In some cases, the QA procedure removed all samples data for a given route (e.g. Route 29 inbound, Route 66 northbound, Route 111 southbound and some special routes).

Figure 45 shows increased observations of over-capacity trips (except for one observation) after the QA procedure is applied. It is possible that the balancing algorithm tends to decrease boarding counts and increase alighting counts when there is a discrepancy in the counts. This removal of these over-corrected counts would results in higher average loads, thus identifying more over-capacity trips. Though network-level performance measures tend not to change substantially after QA is applied; route specific measures are impacted.

Passenger kilometres are another measure that can be used to assess the impact of the QA procedure on performance analysis. Since the sample data represents (almost) all iXpress trips and only a portion of the regular and special route trips, a method is needed to expand the sample data for the entire network. Currently, no such method exists that is easily implemented; therefore this measure is not evaluated. However, it is expected that this measure would change significantly for quality assured data due to the large contribution of schedule distance deviation and unreliable passenger counts to the suspect dataset.

5.4 Limitations

Some limitations to the QA methodology are outlined below:

- Several components of the QA procedure require schedule matching and passenger balancing algorithms to be included in the standard AVL/APC data processing. This feature is a result of the intention to complement, but not to replace, current validation techniques.
- The proposed methodology relies on erroneous data to generate outliers in the passenger activity or travel activity. Erroneous data that does not result in an outlier cannot be detected by the proposed QA procedure.
- Missing data is not directly addressed by the proposed methodology, however the impact of missing stop-level attributes are indirectly detected through the outlier identification structure.
- The methodology relies on the expected patterns to categorize the data at the trip-level into suspect or non-suspect. Ambiguous trip patterns are difficult to classify as valid or invalid.
- The application of the QA procedure to GRT is based on network-level parameters. However route-level parameters are more suitable for some tests.

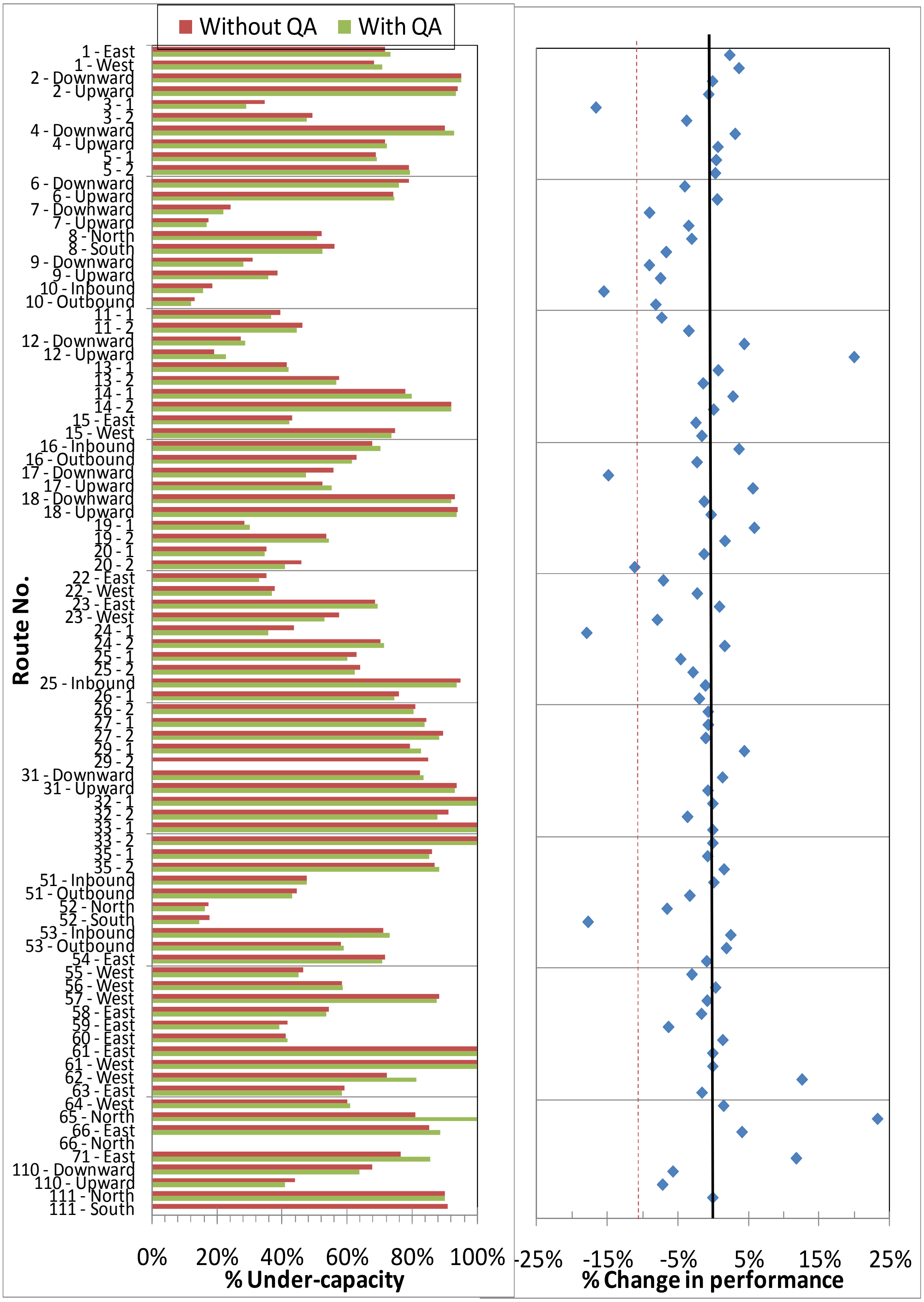

Figure 46 Impact of QA on under-capacity monitoring

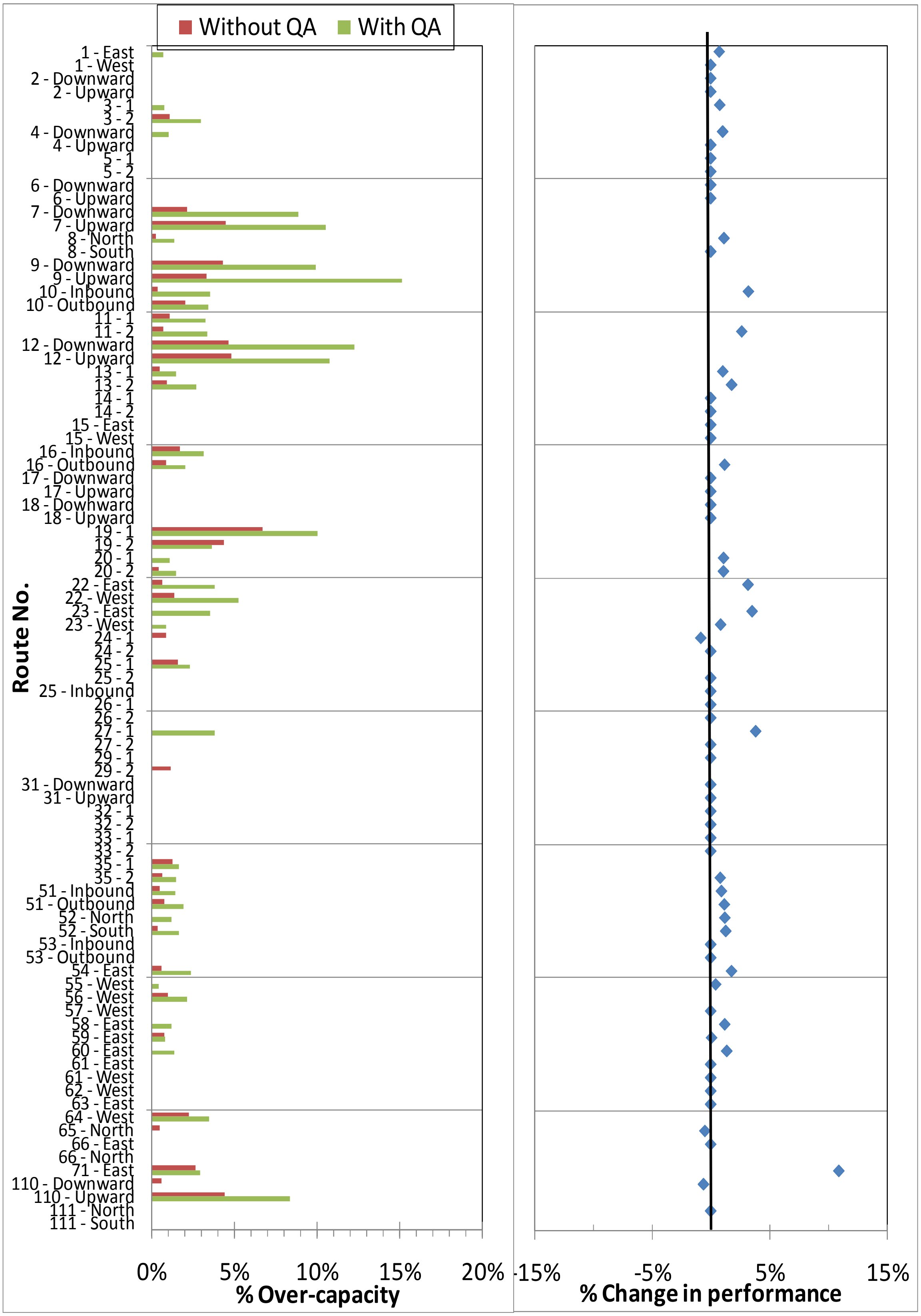

Figure 47 Impact of QA on over-capacity monitoring

- No methods are suggested to "correct" suspect data; suspect data is rejected from the database at the trip-level for tests associated with both passenger activity and travel patterns. The intent of this approach is to provide a clear directive on how to manage suspect data; however it may also lead to a smaller size of sample data for analysis. For records where only the passenger-related test screened the data as suspect, useful travel pattern data may be discarded and vice versa.
- The proposed methodology does not consider previous trips in the assessment of valid case patterns. For example, schedule mis-match trips are potentially a valid trip following a trip that experiences congestion. The structure of the AVL/APC database makes this case difficult to assess and previous trip is not always unavailable when only a portion of the fleet is equipped.

Chapter 6

Conclusions

The availability of archived AVL/APC data generates multiple opportunities to enhance transit operations and planning activities; however quality assurance is an important prerequisite for business decisions supported by these data. An automated quality assurance (QA) procedure was developed to improve the reliability of archived AVL/APC data. The procedure is intended to complement current quality control techniques.

The proposed methodology is described fully in this thesis. Calibration methods are discussed for a sample application to Grand River Transit in Waterloo Region, Ontario. The output of the QA procedure is examined and a sensitivity analyses is conducted to assess the impact of changes to the user-defined parameters on the output of the procedure. Further impacts to downstream applications of the archived data are also examined.

The development and testing of this methodology led to the following findings:

- The use of expected pattern analysis proved useful in identifying both valid and invalid trip data.
- The inclusion of valid case outliers can "save" AVL/APC data that would otherwise be considered suspect. The development of more valid cases can help improve the penetration rate of data through the QA procedure and the sensitivity to key parameters.
- Analysis of the vehicle usage and route-sample distributions can provide useful information for management such as the preparation of sampling plans and vehicle maintenance programs. The lower penetration rate for regular route is a problem for service planners attempting to amass a significant sample size; for the GRT system, fewer equipped vehicles are available for regular routes than for iXpress routes.
- Quality assured data can change the results of performance analyses. Although the impact is less apparent at the network-level, route-level performance measures are necessary to target poor performing routes. The over-capacity performance measure for GRT is most impacted by the application of the QA procedure.
- Limited-stop express route data appear to have a better penetration rates through the QA procedure than regular service and special routes. The quality of the schedule data and the availability of timepoint data at each designated stop seem to improve the results of schedule matching algorithms.

It is recognized that the quality of schedule data, passenger balancing algorithms and database structure impact the outcome of the QA procedure. Therefore the development of an automated validation program should be considered just one component to a proper data quality management plan. Other considerations for the development of a comprehensive data quality plan are provided in the next section.

6.1 Future Work and Recommendations

Several limitations suggest the need for future work following this study. Regarding further improvement of the proposed QA procedure, the following works are recommended:

- Develop more valid case options for outliers to reduce the possibility of losing valid, yet exceptional data. For example:
  - passenger count outliers might be valid for surge boardings and a valid test can be to check counts by key stops;
  - higher count corrections might be expected on higher load trips; a valid test can be to compare the count correction as a percentage of the highest trip load; and
  - large schedule deviations may result from a previously delayed trip and a valid test can check for dissipating time deviations over the trip;
- Include available external data to confirm valid case options (e.g. weather data for trips with congestion and delay patterns and change-off records for vehicle incident trips)
- Calibrate the QA parameters at the route-level where applicable; and
- Separate the identification of suspect data by data type (e.g. suspect passenger count data, suspect time and distance values).

Other data quality management considerations related to the QA procedure are:

- Sample size for regular routes is smaller than for iXpress due to the less available equipped buses and lower penetration rate through the QA procedure. Limited sample size restricts the utility of AVL/APC systems for operations and planning. Sampling plans can be developed from route sample distributions and vehicle usage statistics.
- Improvements to the database structure can facilitate analysis of the AVL/APC data (e.g. separate identifiers to sequence interstop, designated stop and timepoints would be useful for the application of this QA procedure as well as other service analyses.) A thorough investigation of database structural concerns can help develop recommendations for improvement.
- The application of the QA procedure has identified upstream contributors to poor data quality (e.g. poor schedule data is found to result in poor penetration rates in the QA procedure). Examination of the specific design elements that constitute high quality schedule data can help improve schedule matching results.

References

Beniston, Gethyn (2009). *Report P-09-051: Proposed 2009 Transit Service Changes*. Retreived August 20, 2010 from Report to the Planning and Work Committee, Regional Municipality of Waterloo: http://www.region.waterloo.on.ca/web/region.nsf/0/F9E1FB94C3DFA5D6852575C5004DE58F/$file/P-09-051.pdf?openelement

Boyle, D. (1998). *Synthesis of Practice 29: Passenger Counting Technologies and Procedures.* Washington: Transportation Research Board.

Boyle, D. (2008). *TCRP Synthesis 77: Passenger Counting Systems.* Washington, DC: Transportation Research Board.

Chen, M., & Xia, J. (2007). *Archived Data Management System in Kentucky.* Lexington, KT: Kentucky Transportation Centre.

Federal Highway Administration. (2007). *Archived Data Management Systems: A Cross-Cutting Study.* Retrieved April 13, 2010, from RITA: National Transportation Library: ntl.bts.gov/lib/jpodocs/repts_te/14128/14128.pdf

Flintsch, G., & McGhee, K. K. (2009). *NCHRP Synthesis 401: Quality Management of Pavement Condition Data Collections.* Washington, DC: Transportation Research Board.

Furth, P. G., Hemily, B., Muller, T. H., & Strathman, J. G. (2006). *TRCP Report 113: Using Archived AVL-APC Data to Improve Transit Performance and Monitoring.* Washington, DC: Transportation Research Board.

Furth, P. G., Hemily, B., Muller, T. M., & Strathman, J. G. (2003). *TCRP Web document 23: Uses of Archived AVL-APC Data to Improve Transit Performance and Management - Review and Potential.* Washington, DC: Transportation Research Board.

Furth, P. G., Muller, T. H., Strathman, J. G., & Hemily, B. (2004). Designing Automated Vehicle Location Systems for Archived Data Analysis. *Transportation Research Record; Journal of the Transportation Research Board , No. 1887*, 62-70.

Furth, P. G., Strathman, J. G., & Hemily, B. (2005). Making APC Mainstream. *Transportation Research Record: Journal of the Transportation Research Board , No. 1927*, 207-216.

Goetz, J. (2006). *System Design Document: Chapter 2.2.4 Passenger Information System IRMA.* INIT Inc.

Grand River Transit. (n.d.). *Route 67: Lovell Industrial.* Retrieved July 9, 2010, from Routes and Schedules of GRT:

http://grt.ca/web/transit.nsf/DocID/17B830751F69A6DB8525746C0067D743/$file/RT%2067.pdf?openelement

Hwang, M., Kemp, J., Lerner-Lam, E., Neuerberg, N., & Okunieff, P. (2006). *Advanced Public Transportation Systems: State of the Art Update 2006.* Washington, DC: Federal Transit Administration.

Kimpel, T. J., Strathman, J. G., Griffen, D., Callas, S., & Gerhart, R. L. (2003). Automatic Passenger Counter Evaluation; Implications for the National Transit Database Reporting. *Transportation Research Record; Journal of the Transportation Research Board , No. 1835* (03-2089), 93-101.

Liao, C.-F., & Liu, H. X. (2010). Development of a Data Processing Framework for Transit Performance Analysis. Washington, DC: Transportation Research Board.

Mandelzys, M. (2010). *Using Transit AVL/APC System Data to Monitor and Improve Schedule Adherence.* Waterloo, ON: University of Waterloo.

McGhee, K. H. (2004). *NCHRP Synthesis 334: Automated Pavement Distress Collection Techniques.* Washington, DC: Transportation Research Board.

Parker, D. (2008). *TRCP Synthesis 73: AVL Systems for Bus Transit.* Washington, DC: Transportation Research Board.

Perk, V., & Kamp, N. (2003). *Handbook of Automated Data Collection Methods for the National Transit Database.* Tallahasee, FL: Florida Department of Transportation.

Regional Municipality of Waterloo. (2009). *iXpress: Central Transit Corridor Express Project Urban Transportation Showcase Program Final Report.* Transport Canada.

Turner, S. (2007). *Quality Control Procedures for Archived Operations Traffic Data: Sythesis of Practice and Recommendations.* College Station, TX: FHWA.

Wolters, A. S., McGovern, G., & Hoerner, T. (2006). Development of a Tool to Assess the Quality of Collected Pavement Management Data. *Transportation Research Record: Journal of the Transportation Research Board , No. 1974*, 37-46.

Appendix A

Expected data patterns

Note that the following plots are based on hypothetical trips

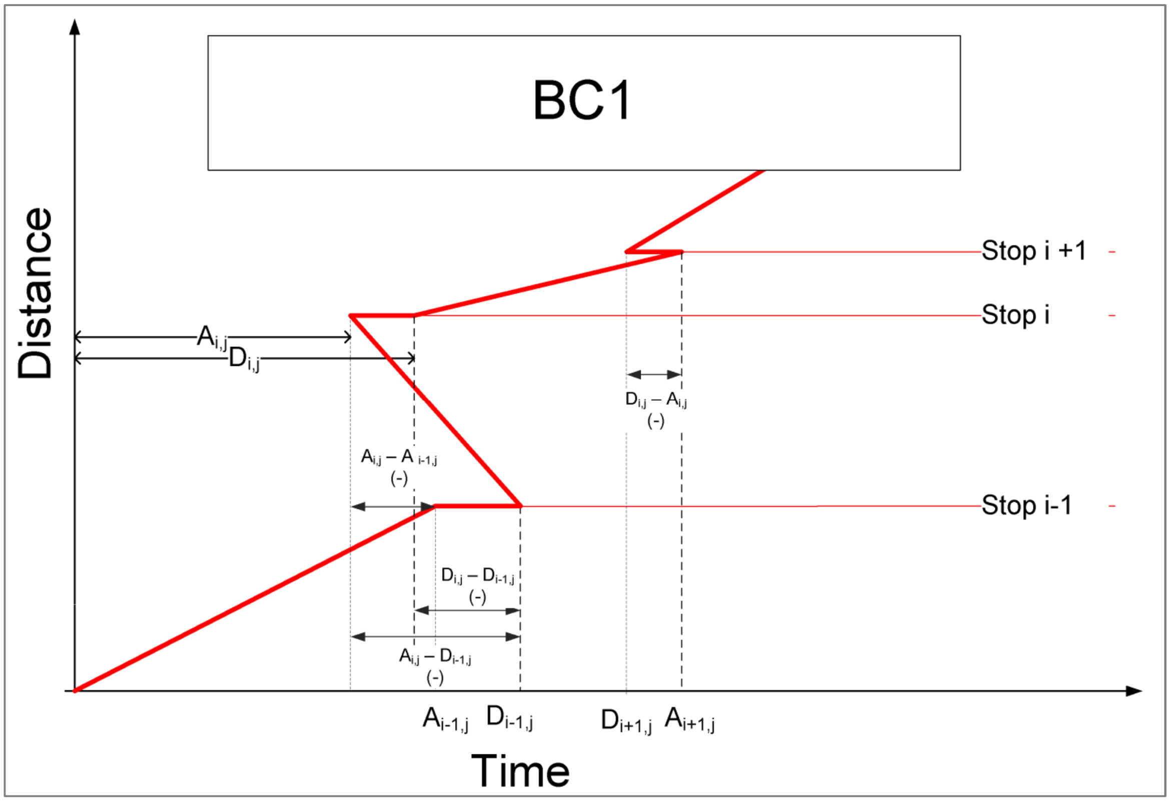

Figure A 1 Trip failing BC1 (Time increasing)

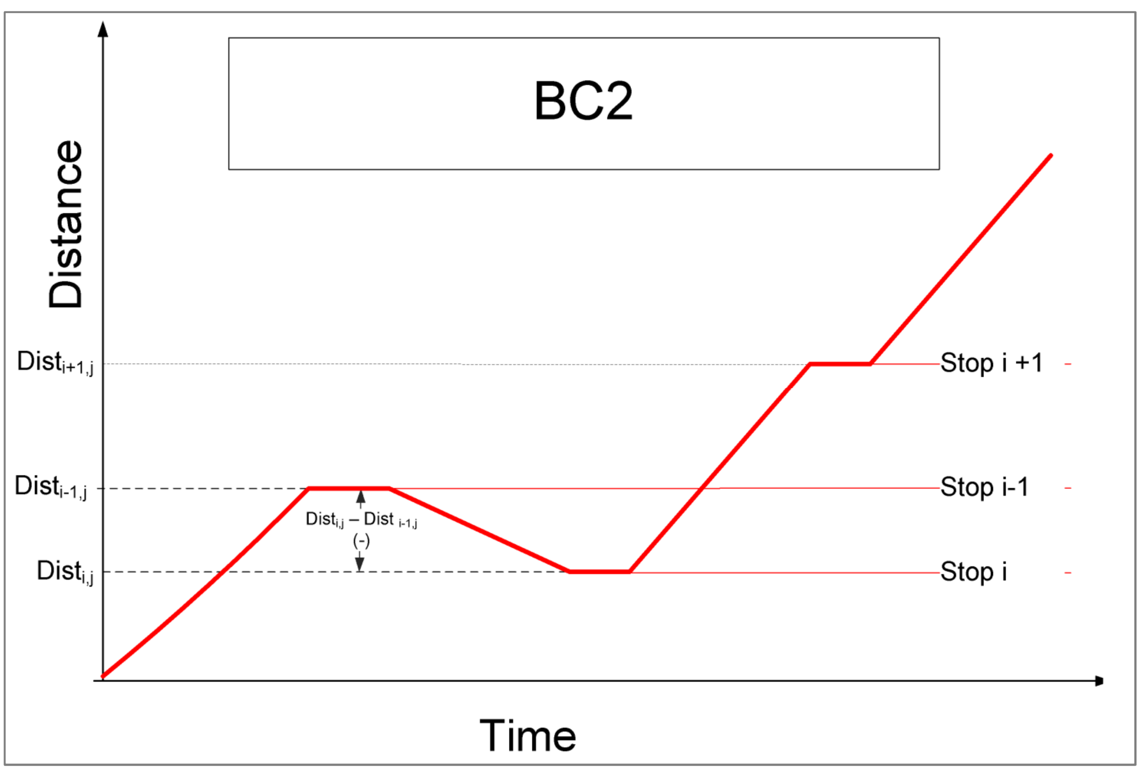

Figure A 2 Trip failing BC2 (Distance increasing)

Note that the following plots are based on hypothetical trips

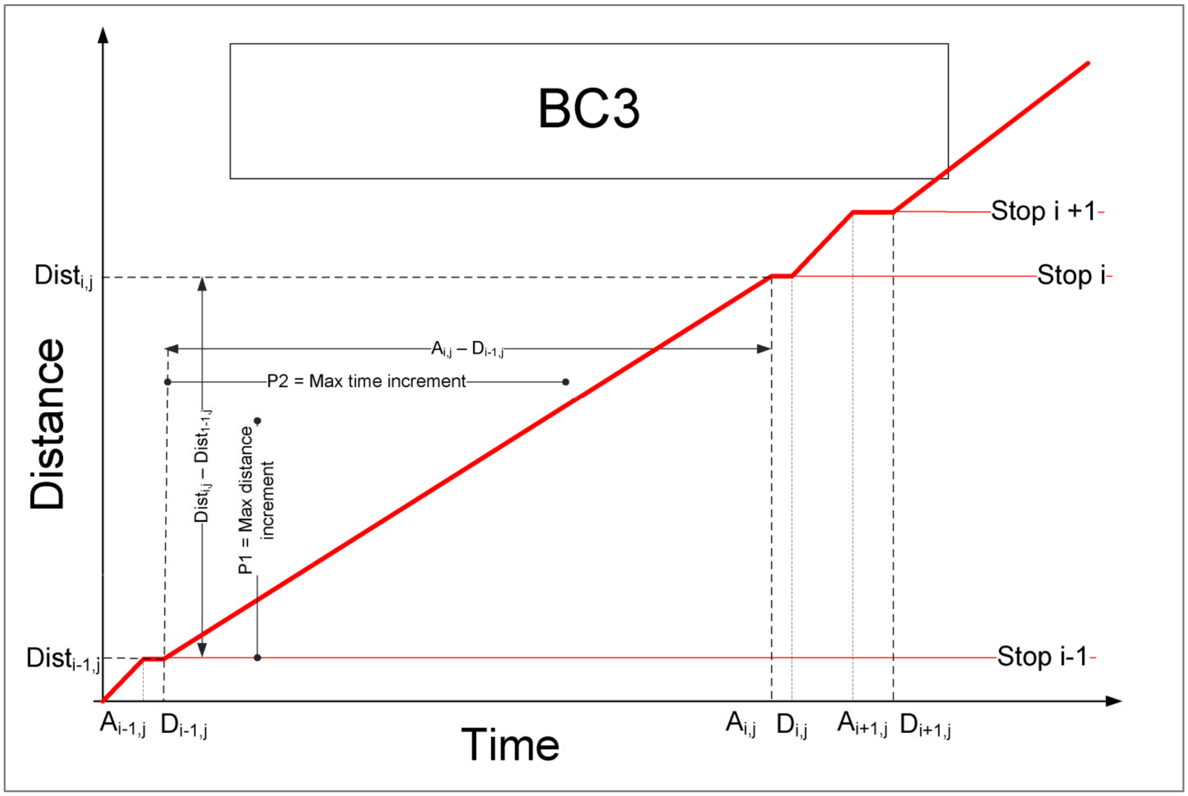

Figure A 3 Trip failing BC3 (Time and distance increment constraints)

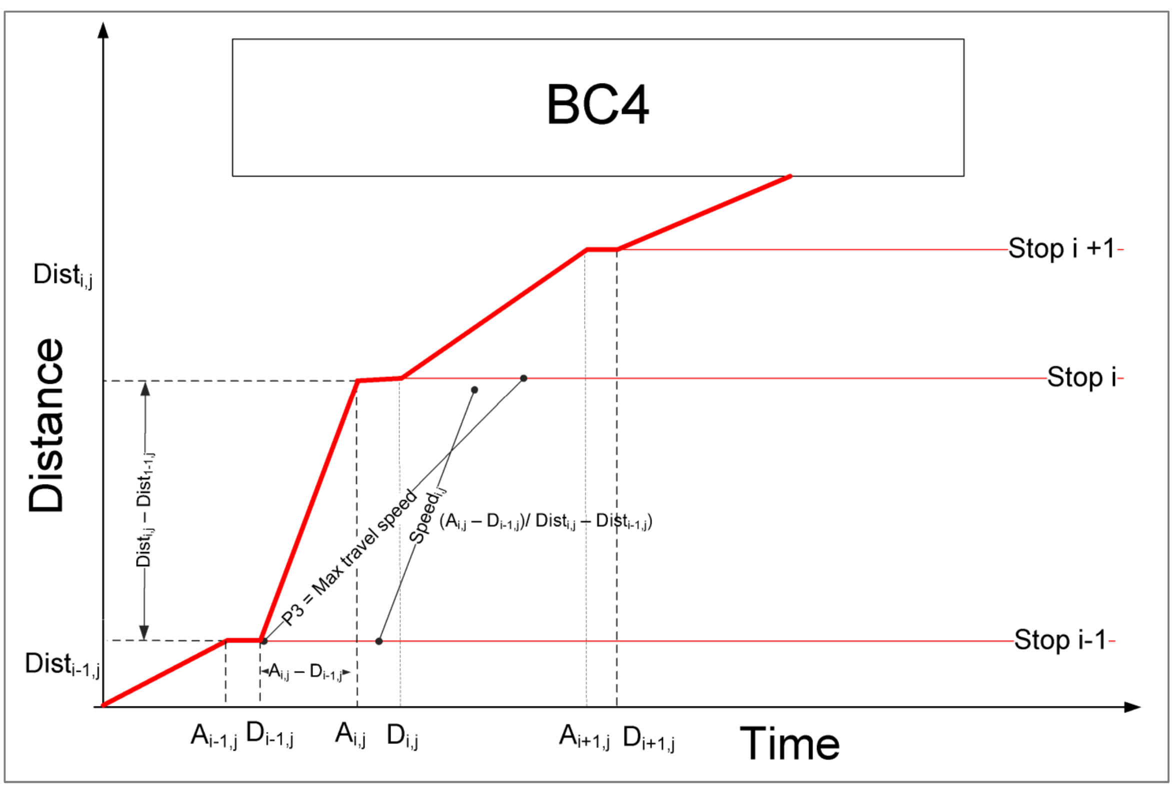

Figure A 4 Trip failing BC4 (Travel speed constraint)

Note that the following plots are based on hypothetical trips

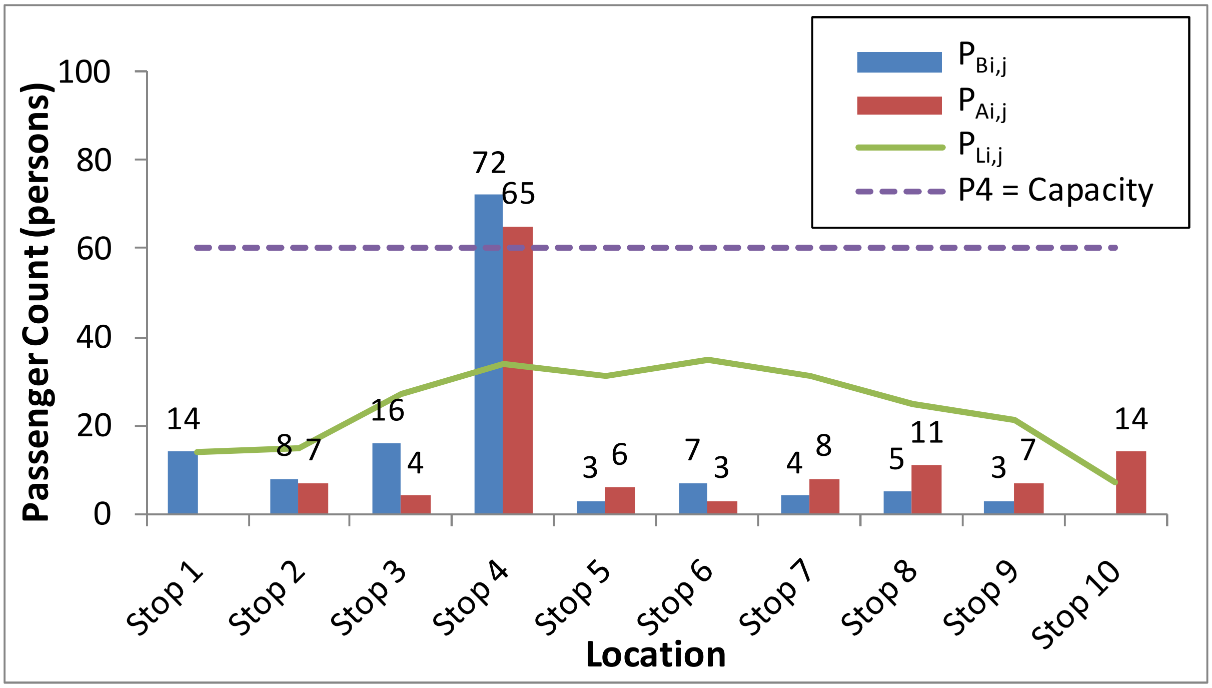

Figure A 5 Trips failing OI1 (Passenger count outlier)

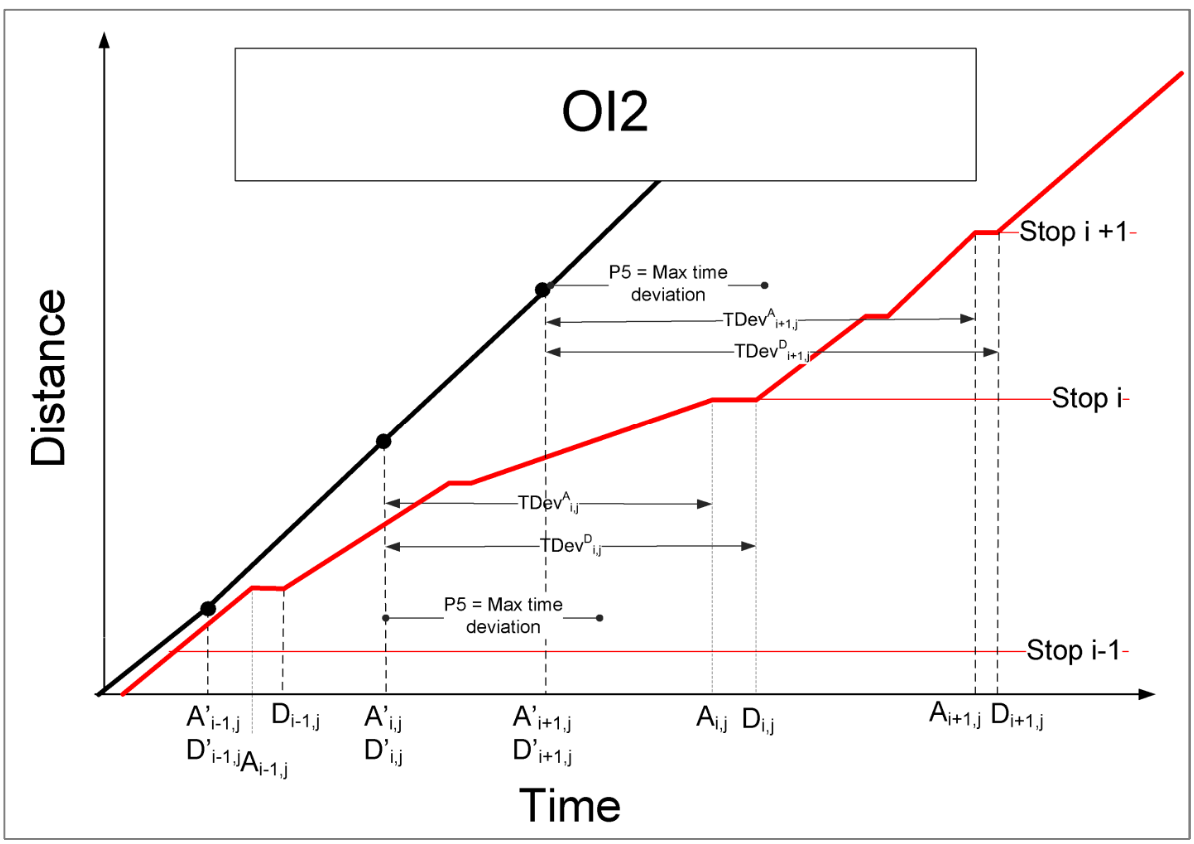

Figure A 6 Trip failing OI2 (Time deviation outlier)

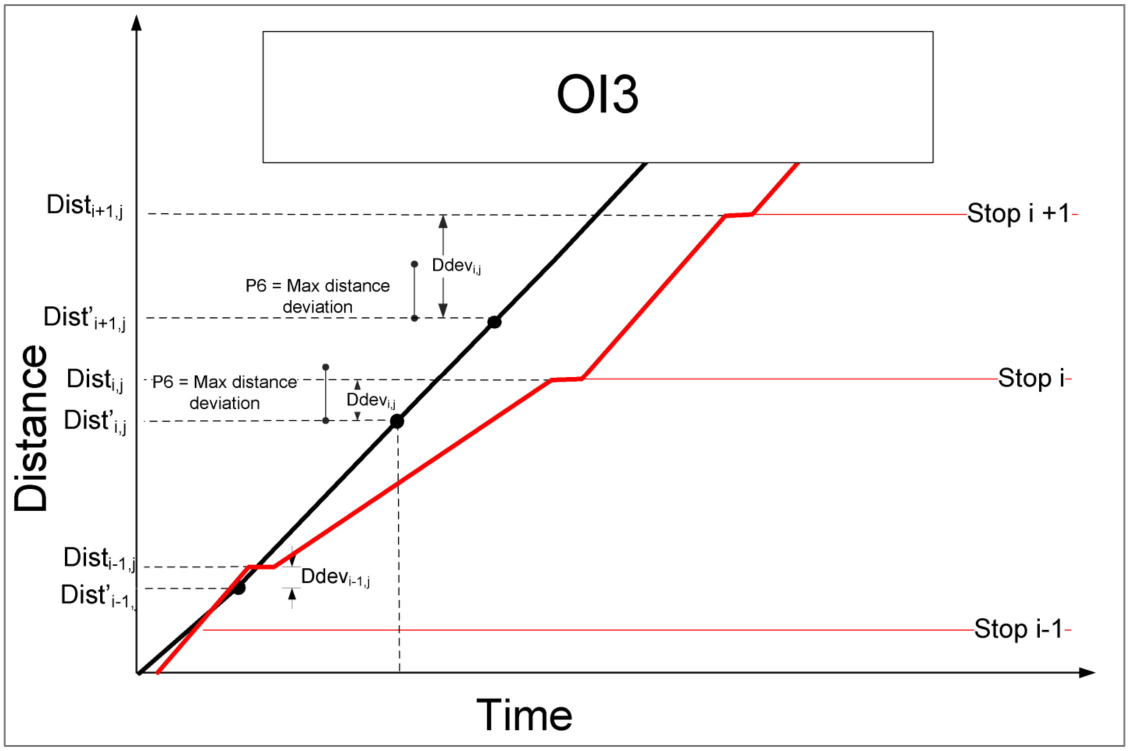

Figure A 7 Trip failing OI3 (Distance deviation outlier)

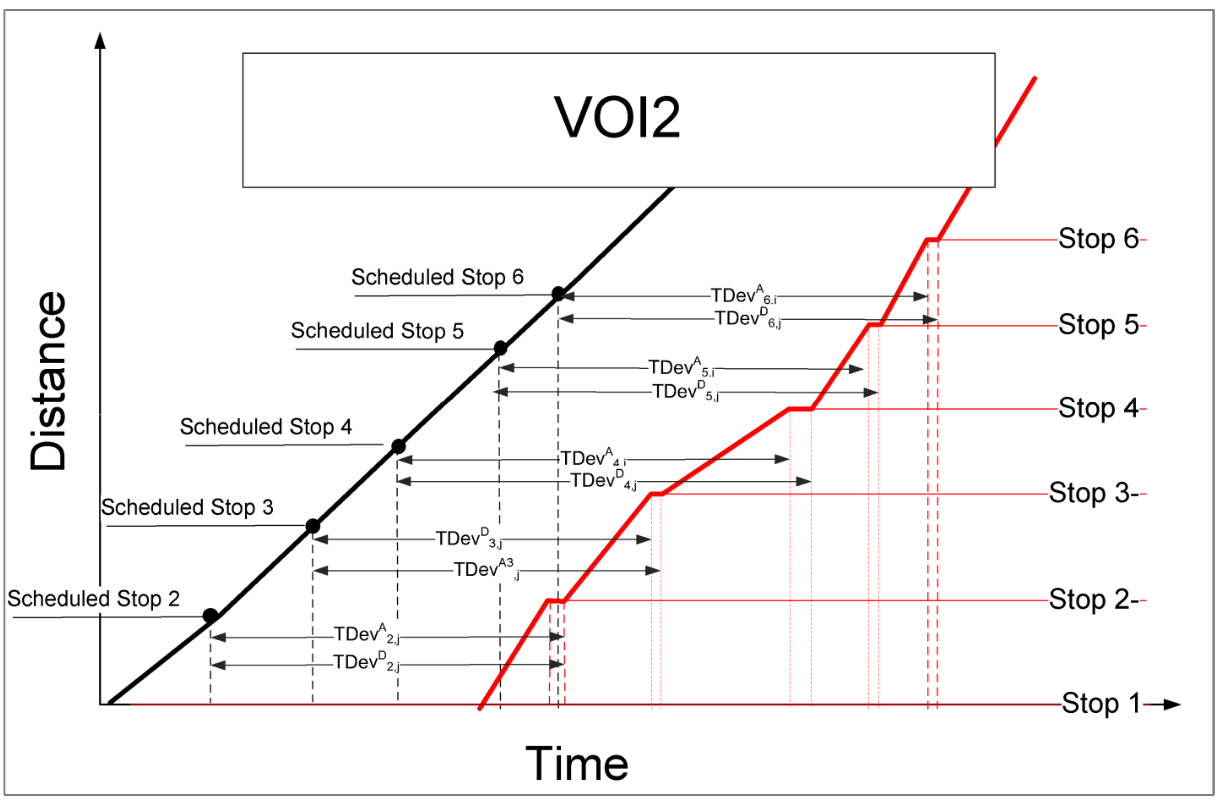

Figure A 8 Trip passing VOI2 (Suspected mis-match to schedule)

Note that the following plots are based on hypothetical trips

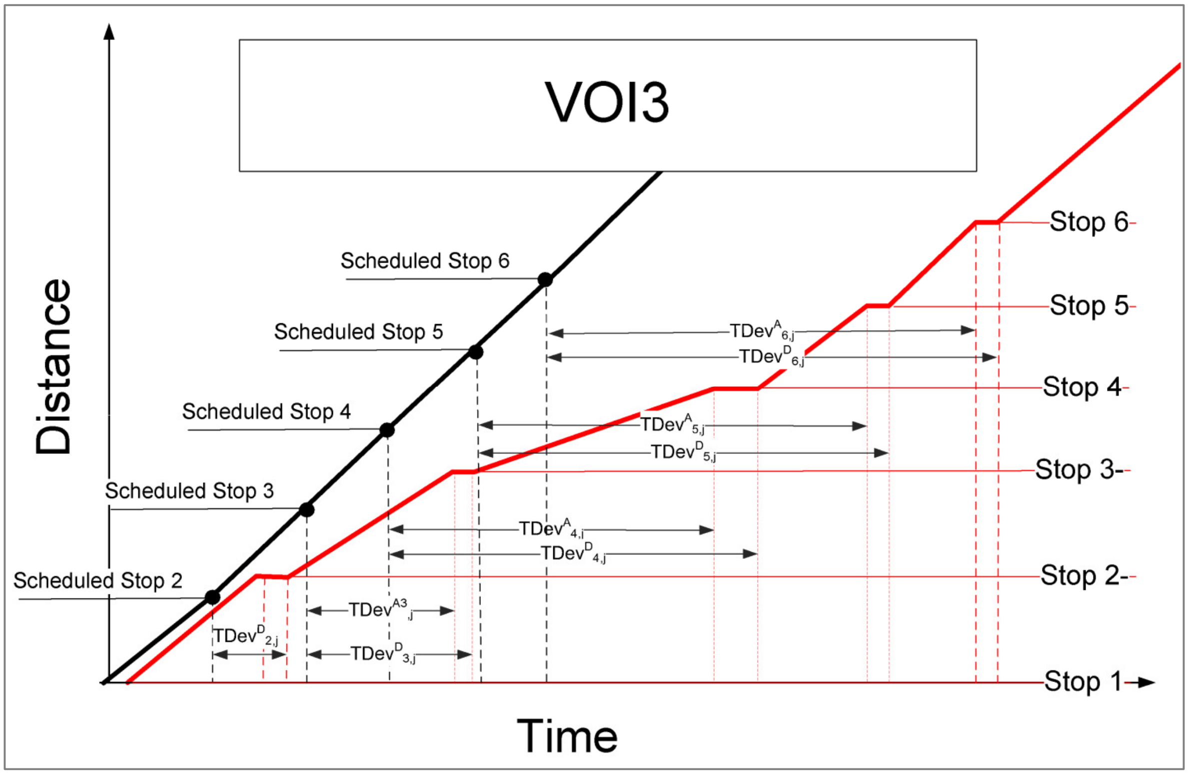

Figure A 9 Trip passing VOI3 (Valid time deviation outlier by congestion or operational delay)

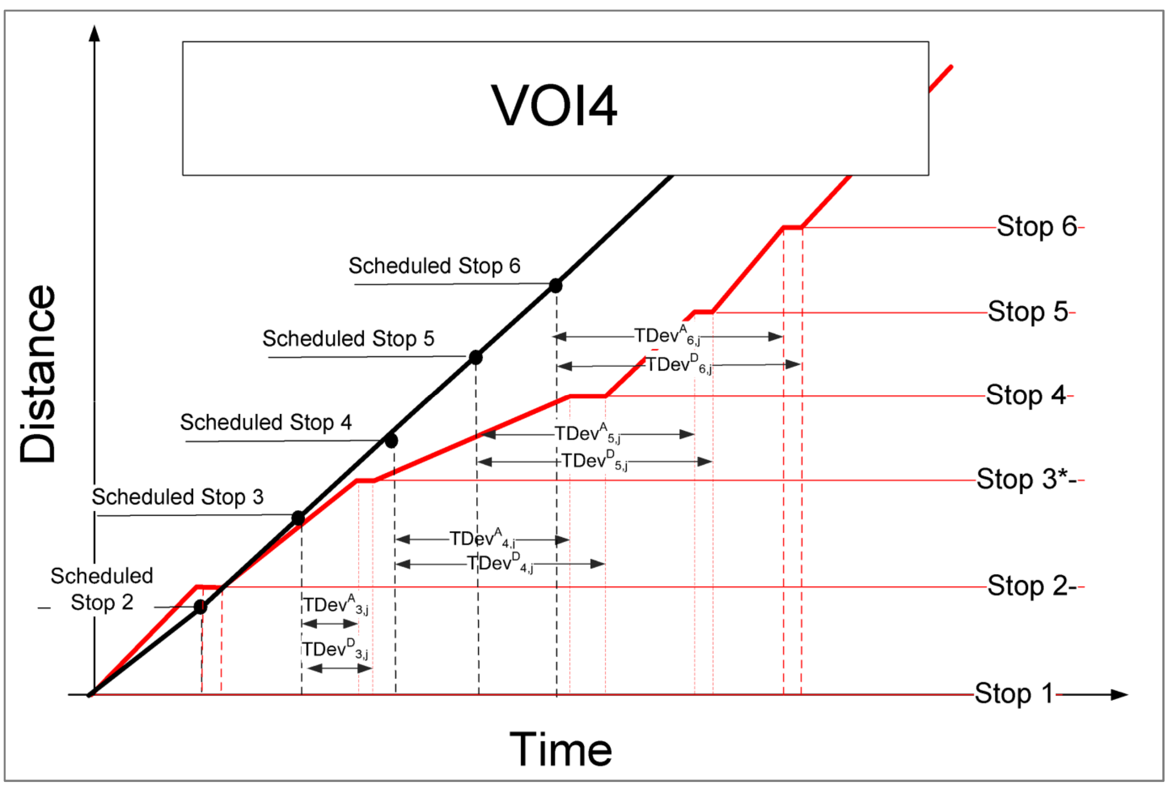

Figure A 10 Trip passing VOI4 (Valid time deviation outlier by partial congestion or operational delay)

Note that the following plots are based on hypothetical trips

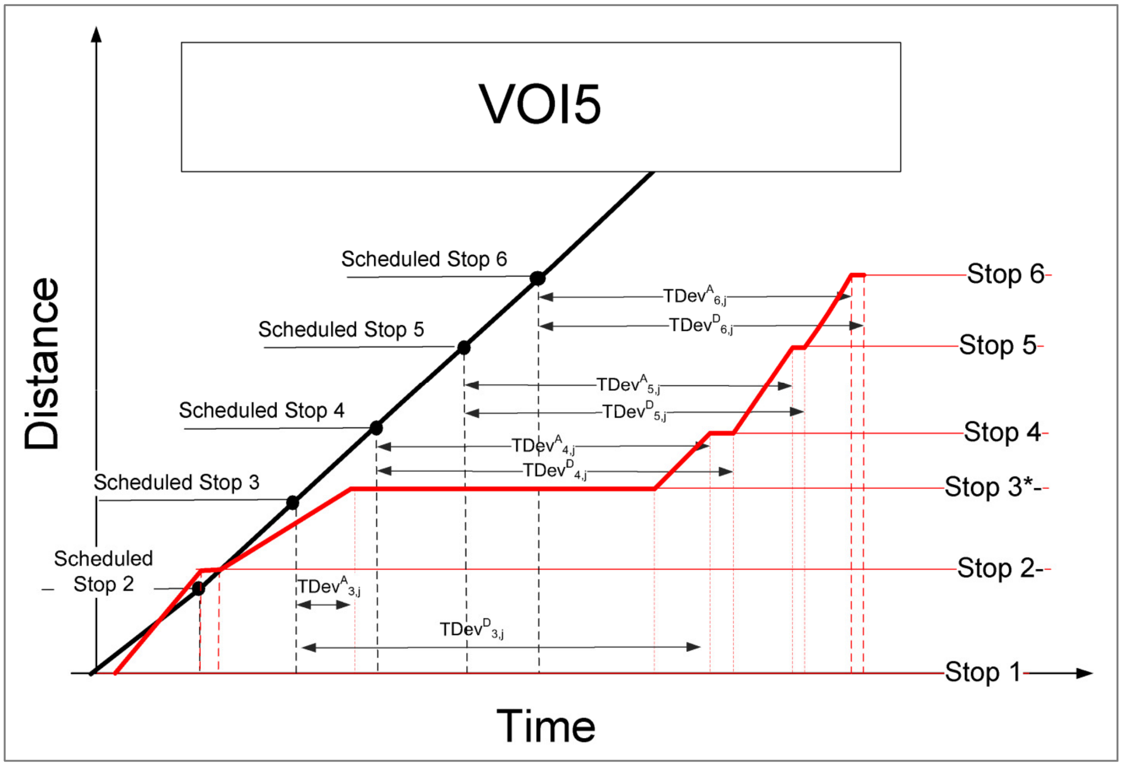

Figure A 11 Trip passing VOI5 (Valid time deviation outlier by transit vehicle-related incident)

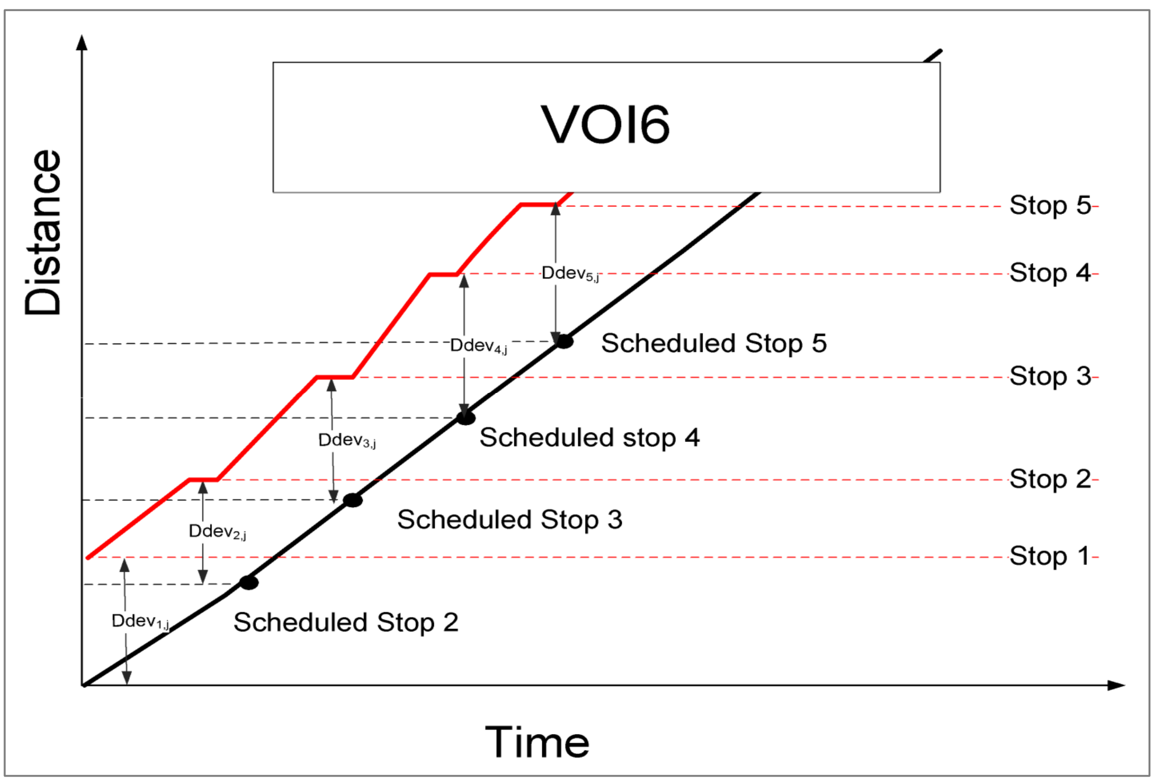

Figure A 12 Trip passing VOI6 (Suspected shift in stop matching)

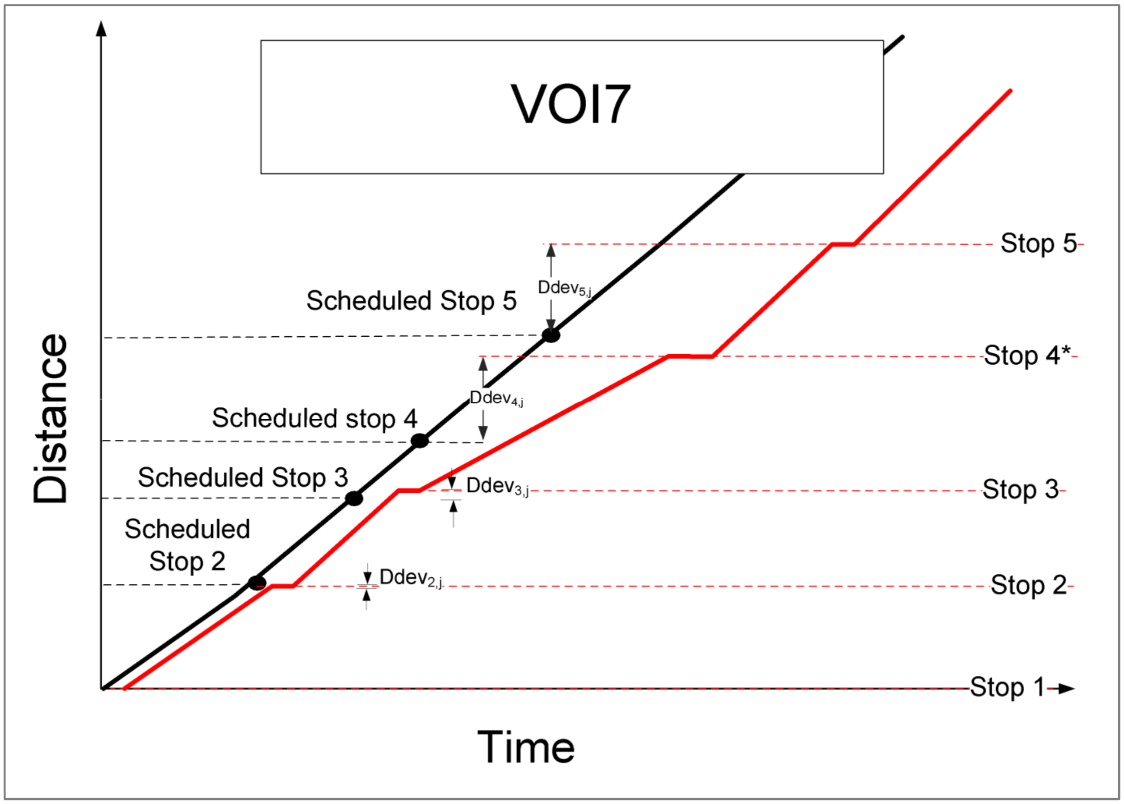

Figure A 13 Trips passing VOI7 (Valid distance deviation by detour)

Appendix B

Parameter selection

Section 4.3 discusses the approach to calibrating parameters for the proposed QA procedure. This appendix provides distribution plots and a description for the selection of P2 through P11.

The second parameter for BC3 is P2, maximum distance increment. The trip distribution of the largest distance step is shown in **Figure B1.** Similar to the trip distribution for largest time increment, there are two peaks for the distribution of largest distance increments. The second peak likely represents mostly iXpress trips because the longest inter-stop distance on the iXpress route is 12.6km from Cambridge Centre to Fairview Mall. P2 should be assigned by route or route type. Unlike the P1, a reasonable upper bound for the distance increment is not the longest one-way cycle, but the longest distance between any two stops for the route. The AVL/APC system is expected to generate an event record for all stops on a route even when it is skipped. Missing stop-level event records constitute incomplete data for a trip-based analysis such as the derivation of load values from stop-level on-off differences. Like P1, one network-level parameter is selected instead of several for each route or route type. P2 is set to 15km; more than 95% of trips are shown to be below this threshold.

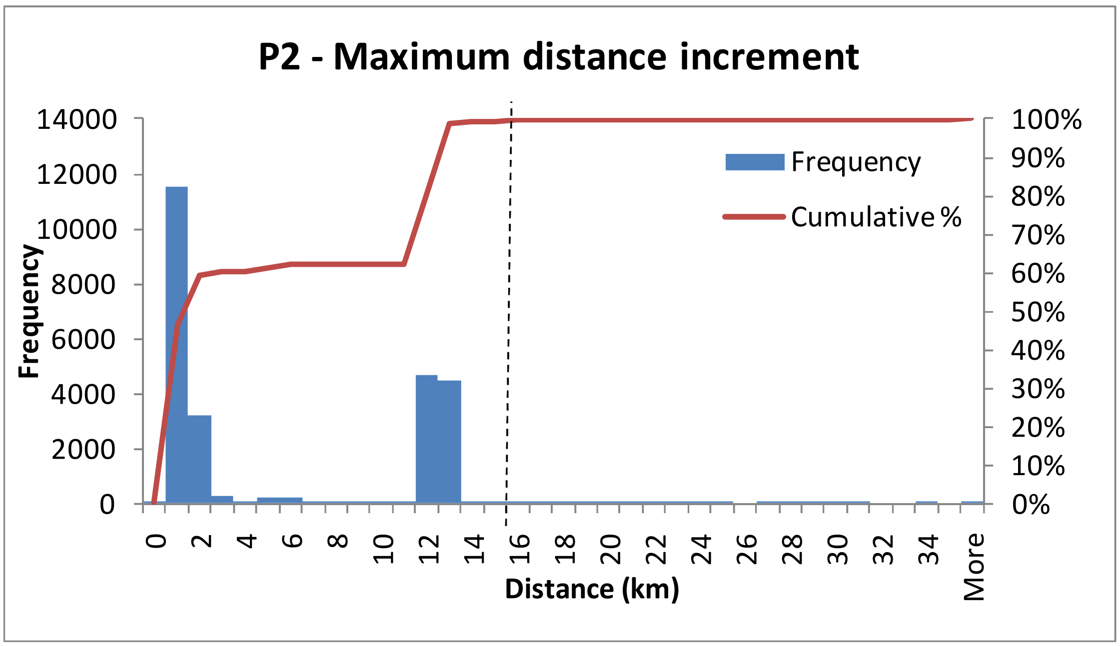

Figure B 1 Selection of P2

Figure B2 shows the selection of P3 in test B4. The distribution in **Figure B2** represents the trip-level maximum calculated speed between two stops. Again, there appears to be two peaks. However the selection of this parameter should not be route-based. The speed constraint in test BC4 is based on the physical limitations of the transit vehicle and operational limits. Though some GRT routes travel along the highway, transit buses are expected to oblige all posted speed limits. The highest posted speed limit in the Region of Waterloo is 100km/hr on the 401 highway. Therefore, P3 is set to 100km/hr. The cumulative percentage plot shows that 99.7% of the trips are below this threshold.

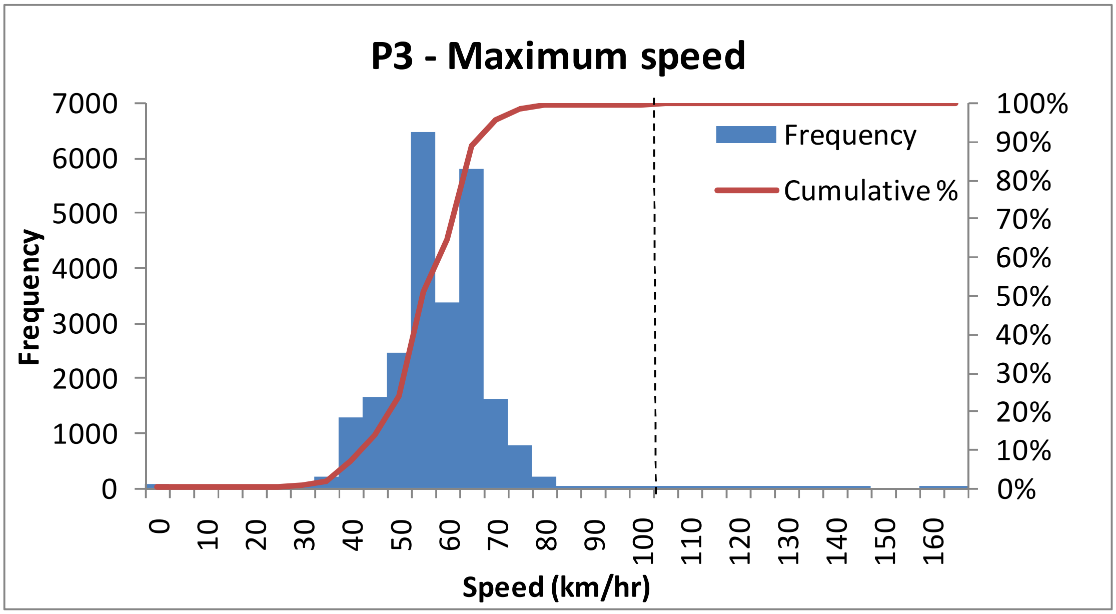

Figure B 2 Selection of P3

Figure B3 shows the selection of P4 in test OI1. Similar to P3, the maximum passenger count (P4) in test OI1 is meant to test a physical constraint: the space limitations on a bus. The number of passengers boarding, passengers alighting and load is capped by the capacity of the bus. While it is possible that there can be an infinite number of the passengers boarding so long as an equivalent number of passengers are alighting; however this occurrence is an unexpected pattern. The test is included as part of the Outlier Identification stage because it is based on expected patterns in the data. P4 is set to 80 passengers; the cumulative percentage plot shows that 99.1% of trips are below this range. P4 not only represents a vehicle capacity, but a crush load capacity should the vehicle be over-capacity.

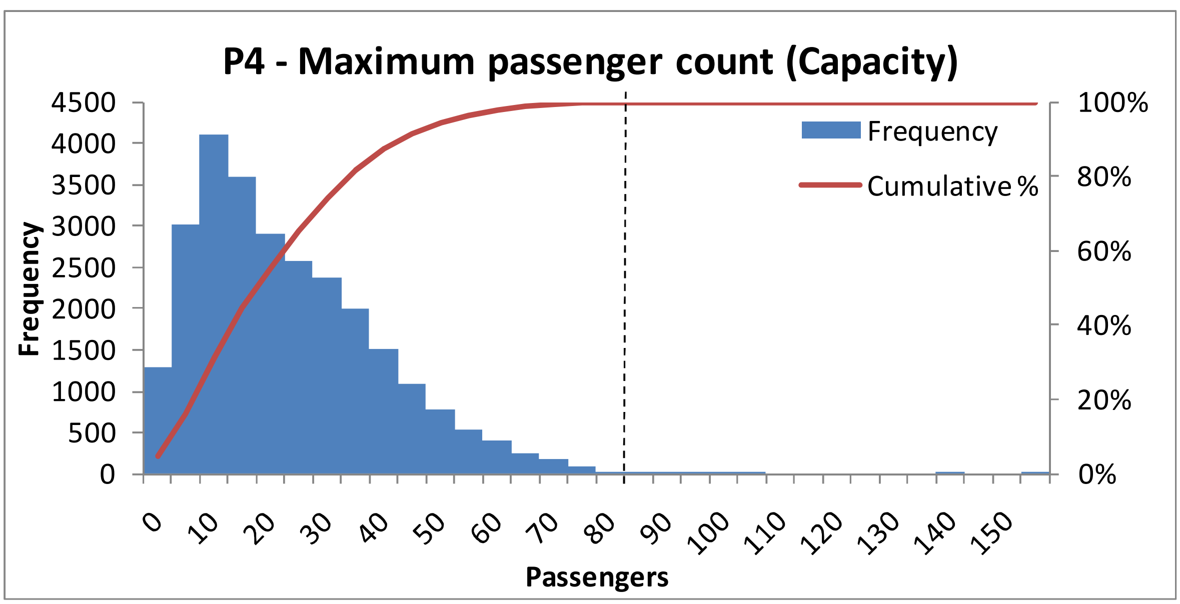

Figure B 3 Selection of P4

Figure B4 shows the selection for P5 in test OI2. The distribution shown in **Figure B4** demonstrates the largest stop-level arrival or departure time deviation for each trip. Outlier values are screened when they exceed the 95th percentile; the cumulative distribution graph shows that a P5 value of 20 minutes is required to identify these outliers.

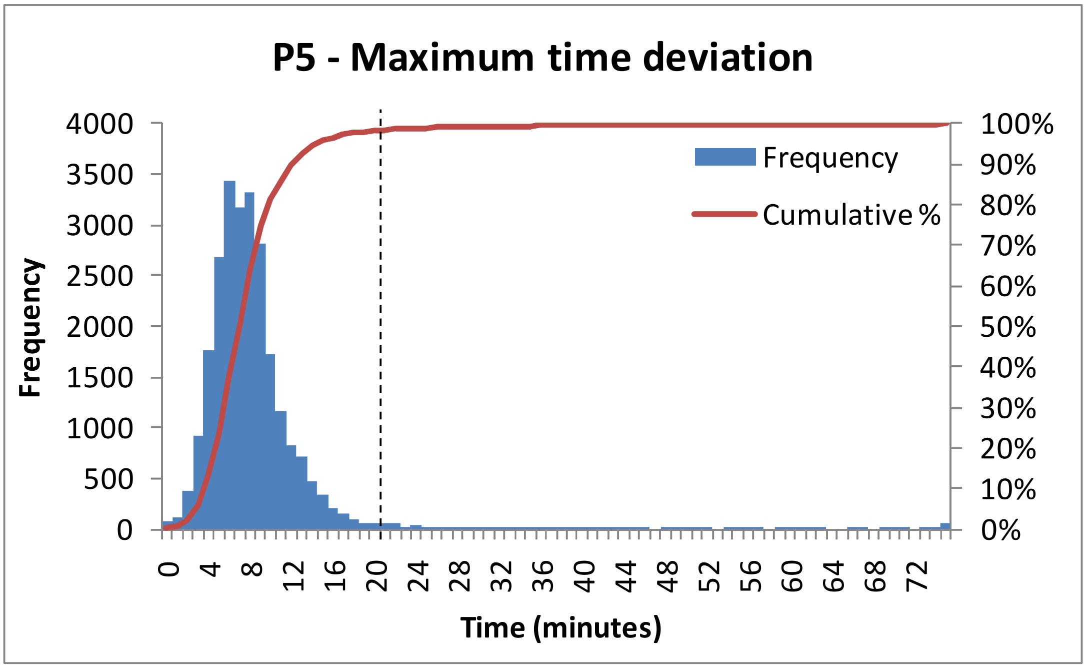

Figure B 4 Selection of P5

Figure B6 shows the selection for P6 in test OI3. Similar to the logic used to select P5, P6 is set to 2km and this represent threshold represents the approximately the 95th percentile.

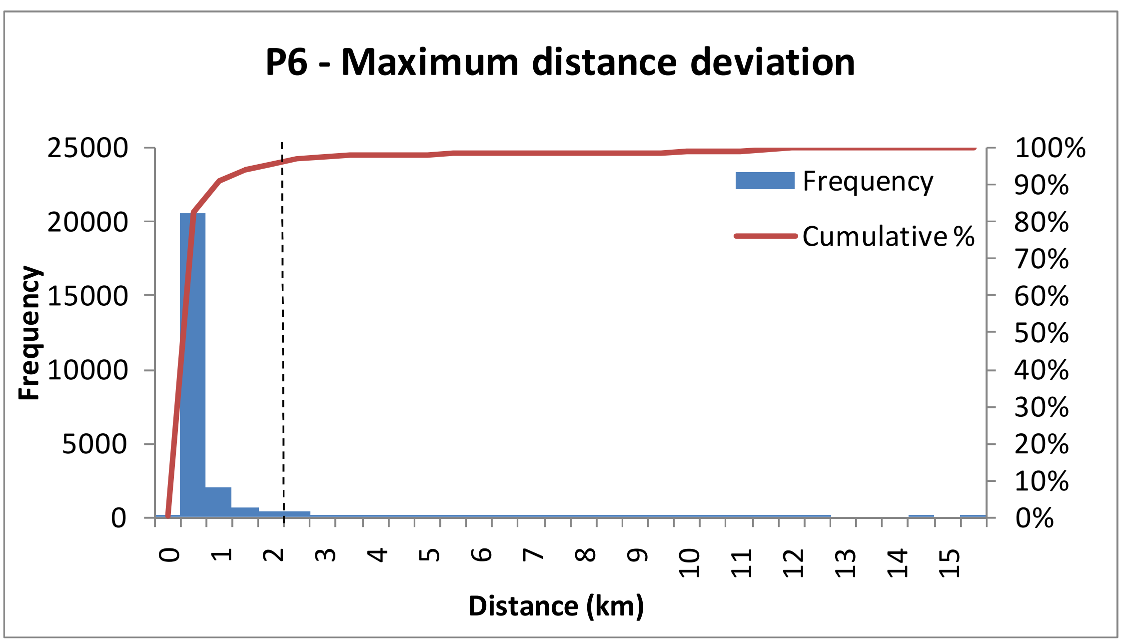

Figure B 5 Selection of P6

Figure B6 shows the selection for P7 in test OI4. Again similar logic is used for the selection of P7 as for the selection of P5 and P6. However additional consideration is taken from the results of the manual count surveys in Section 4.4. The surveys show an error of 8.9% and 6.4% for the total boarding counts and alighting counts. The vendor guarantees these errors to be less than or equal to 10%. In other words, the discrepancy between the raw APC count and the manual count (i.e. assumed to be the true value) should be less than or equal to 10% of the manual count. Typical bus capacity is 50 passengers. Therefore a 10% 'maximum' discrepancy, as backed by vendor guarantee, and demonstrated by the manual survey, should be roughly 5 passengers. The 10% guarantee might not be defendable for count corrections on buses with crush load (i.e. near 80 passengers). Literature notes that low floor buses are subject to greater error related to bus configuration and proximity of passengers near the doorways (Kimpel et al, 2003). Therefore, P7 is set to 6 passengers assuming typical bus capacity is 50 passengers (a value of 6 passengers provides a buffer to an assumed 10% error).

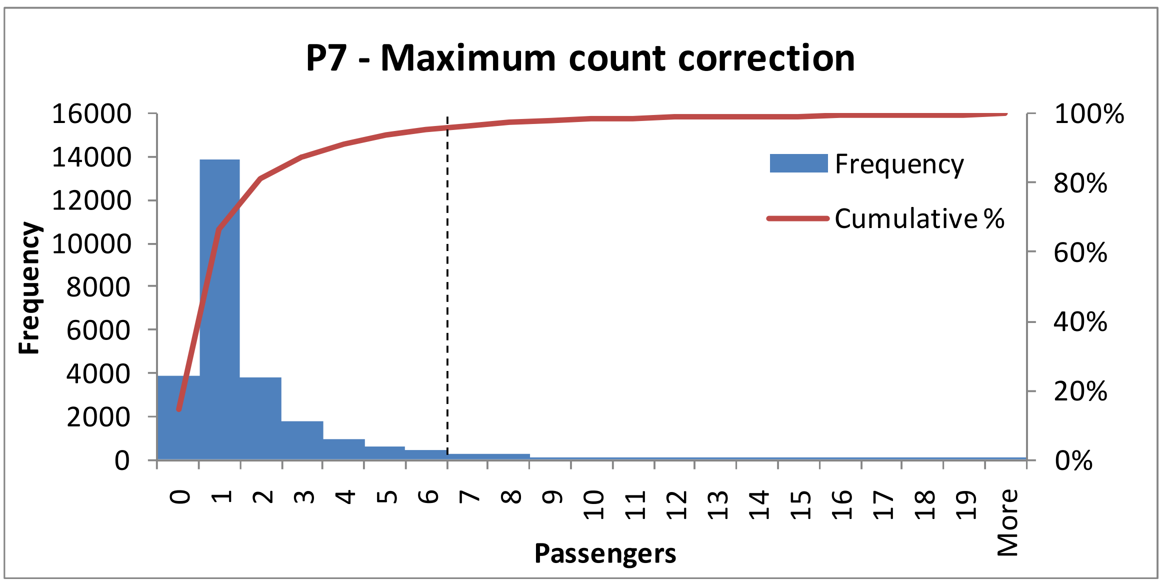

Figure B 6 Selection of P7

Figure B7 shows the selection of P8, the minimum time deviation required on all stops to classify between trip with time deviations occurring over the entire trip (group A) or trips with time deviations occurring over only a portion of the trip (group B). The distribution of the minimum time deviation is affected by the threshold to identify time deviation outliers in test OI2 (P5). The trips in group A are further tested by test VOI3, congestion or operational delay of the entire trip and the trips in group B are further tested for vehicle incident patterns (VOI5). The selection of this parameter may be slightly arbitrary because trips that fail VOI3 and VOI5 are further tested for congestion or operational delay over a portion of the trip (VOI4).

The value of this parameter impacts the sequence of tests that are applied to a given trip; the resulting outcome decides which valid or invalid pattern should be tested. P8 is set to 60s (1min).

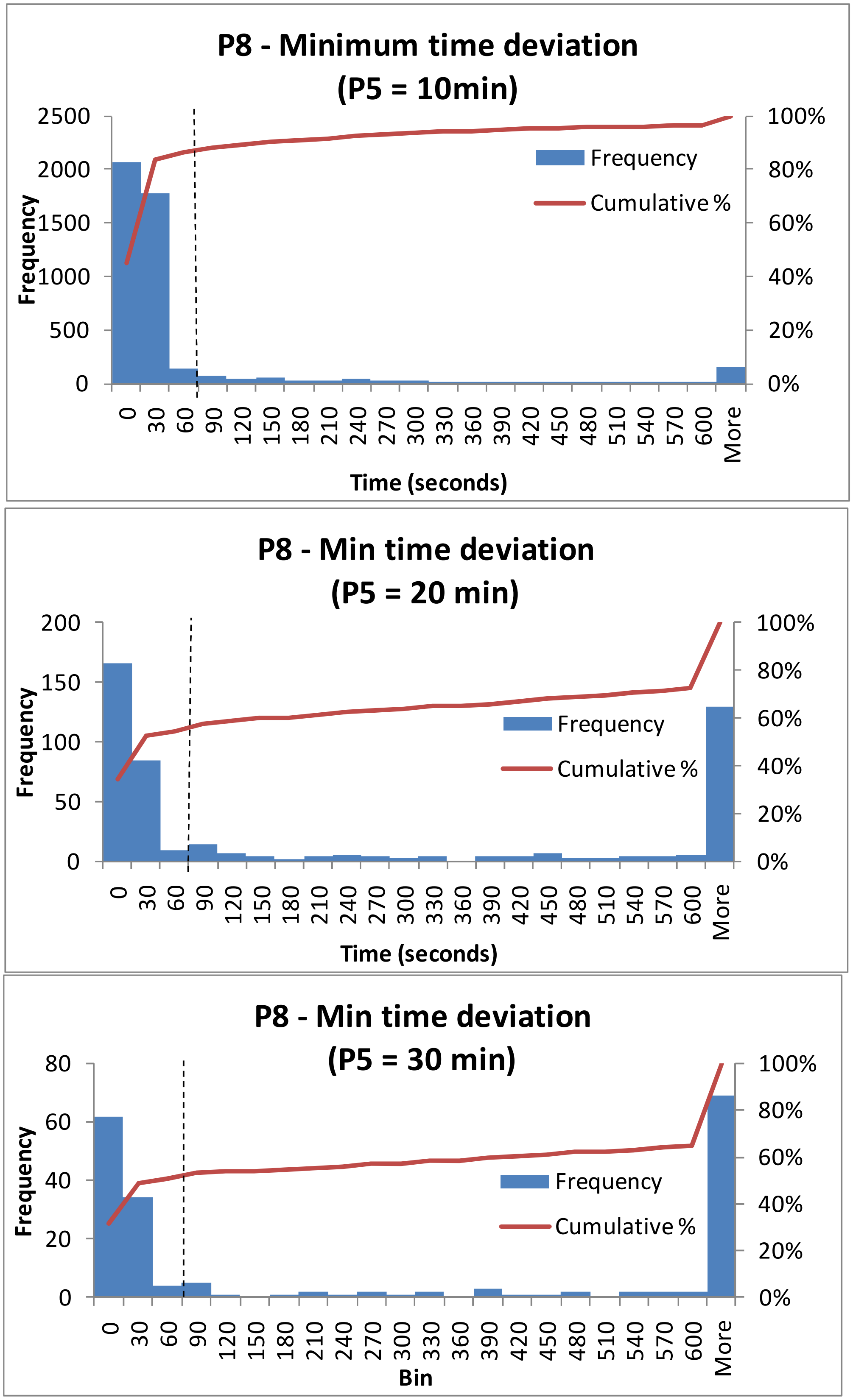

Figure B 7 Selection of P8

Figure B8 shows the selection of P9, the maximum time deviation increase. This parameter is used in test VOI2 to find schedule mis-matches. Schedule mis-matches are considered to be a rare occurrence by the GRT staff. Test VOI2 is designed to find trips with an outlier time deviation and a pattern of consistent time deviation values throughout the trip. This consistent pattern is identified by monitoring the time deviation increase between stops. A threshold (P9) is set to flag if the deviation increase constitutes a growing pattern of time deviation values. Because the expected pattern is comprised of both a large time deviation and a consistent pattern, the test relies on the outcome of previous test OI2 (P5) to screen trips with time deviation outlier.

P9 is set to 10%, this value represents a small portion of trips identified with both a outlier time deviation and a significant deviation value at each time point. So if a deviation increases by more than 10%, the trip is not a mis-match. The value also coincides with the inflection point on the cumulative percentage plot.

Figure B9 demonstrates the selection of P10, which is incorporated in test VOI3 and VOI4 (valid delays trips). Just like for P9, the outcome of test OI2 (P5) impacts the distribution shown in **Figure B9**, a higher outlier threshold results in fewer trips in the distribution. The difficulty with the distribution in Figure B10 is that is includes time deviation increases from the first arrival and last departure time deviation, which are generally considered irrelevant. The distribution to the right of the dashed line represents trips that may constitute a congestion or operational delay pattern, if they are not already identified as a mis-match pattern. Since P10 is also used in test VOI4 (congestion or operational delay pattern over a portion of the trip), minimum time deviation declines are not relevant before the turning time point i*.

Although P10 impacts which trips with distance deviations are considered valid or invalid, the selection of P10 is somewhat arbitrary because it is difficult to estimate what portion of trips with time deviation outliers are expected to experience delay. The tests VOI3 and VOI4 expect the time deviation to increase between time points (time deviation increase should stay positive) or at the very least, not decrease (only a small negative decrease can be tolerated as a slight variation). Therefore P10 is set to -5%.

Lastly, **Figure B10** shows the selection of P11, the maximum distance deviation increase. P11 is used as the threshold in test VOI7 and VOI8. In test VOI7, a uniform pattern of distance deviation among all stops represent a “shift” in the matched location or improper resetting of the odometer; thus unreliable distance values. P11 is used to cap the distance increase to identifiy a uniform pattern. In test VOI8, a uniform pattern of distance is expected after the detour segment for a trip.

The distribution in **Figure B10** shows that more that 50% of the trips tend to have very large distance deviation increases. Many of these trips results from a small absolute change in the distance deviation from zero, but results in a large percentage change. However, these trips are not the target for test VOI7. Instead, a “shift” in distance values likely result from a mis identification of the first stop. The selection of P11 is based on the population on the left side of the plot. P11 is set to 5%.

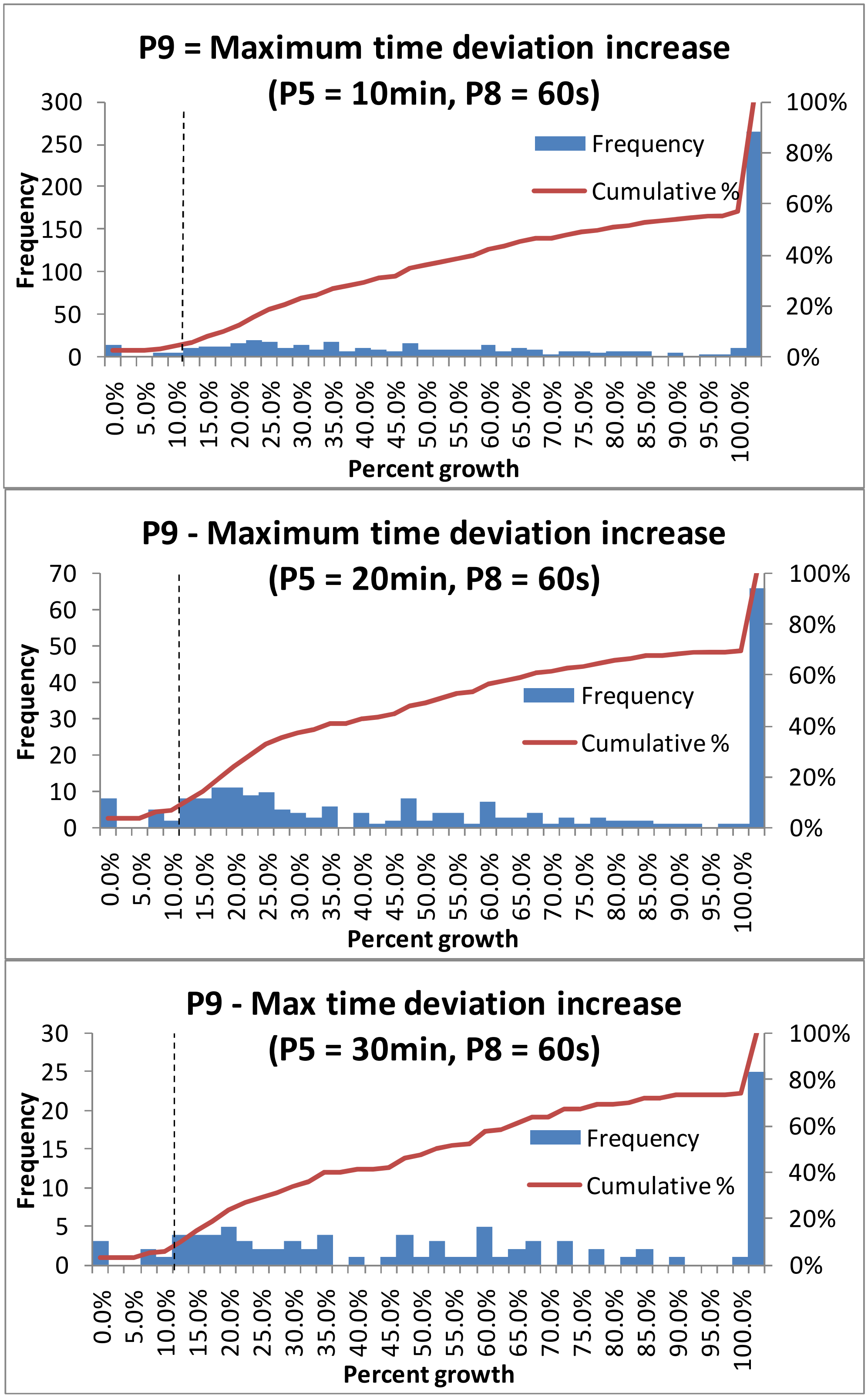

Figure B 8 Selection of P9

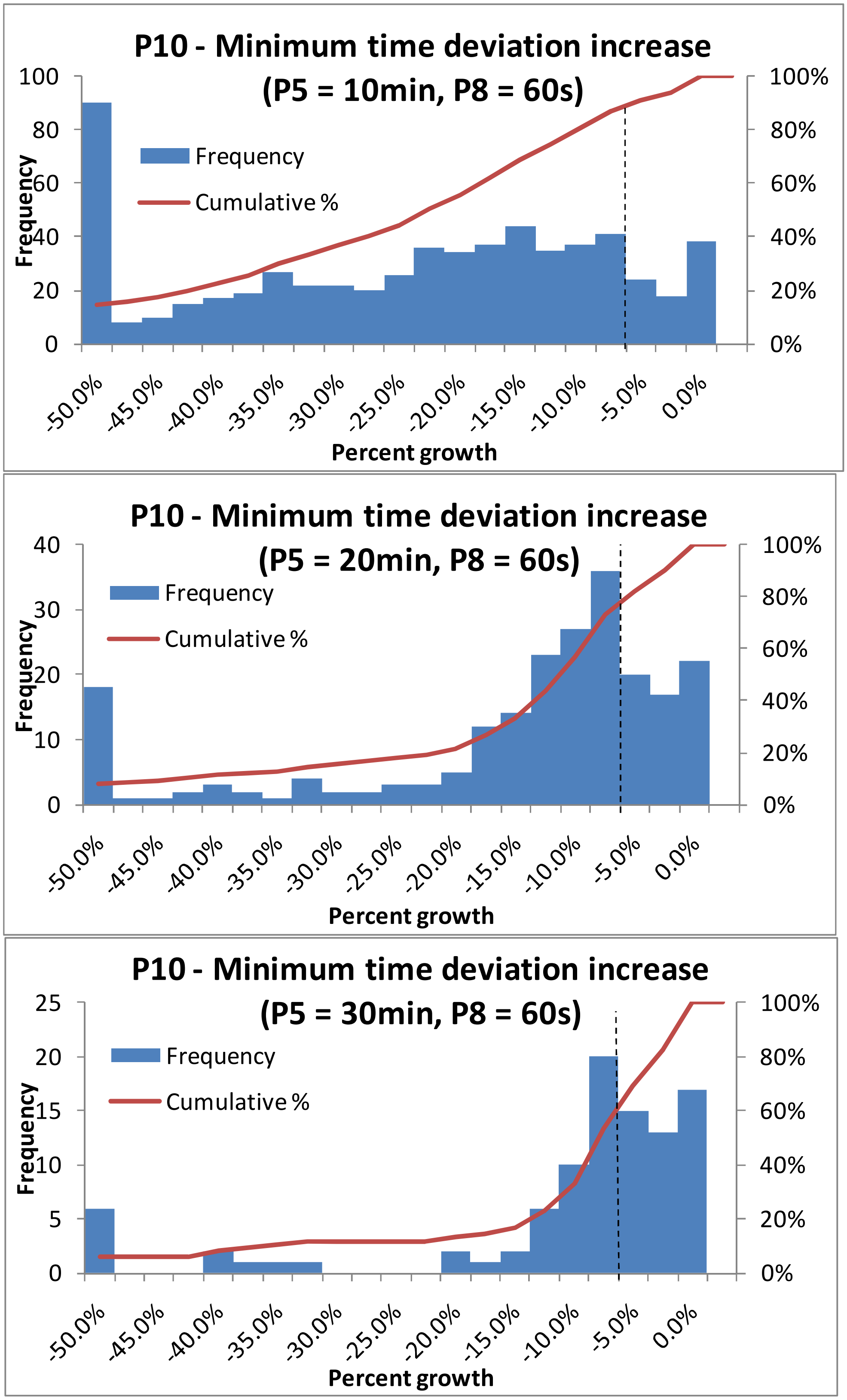

Figure B 9 Selection of P10

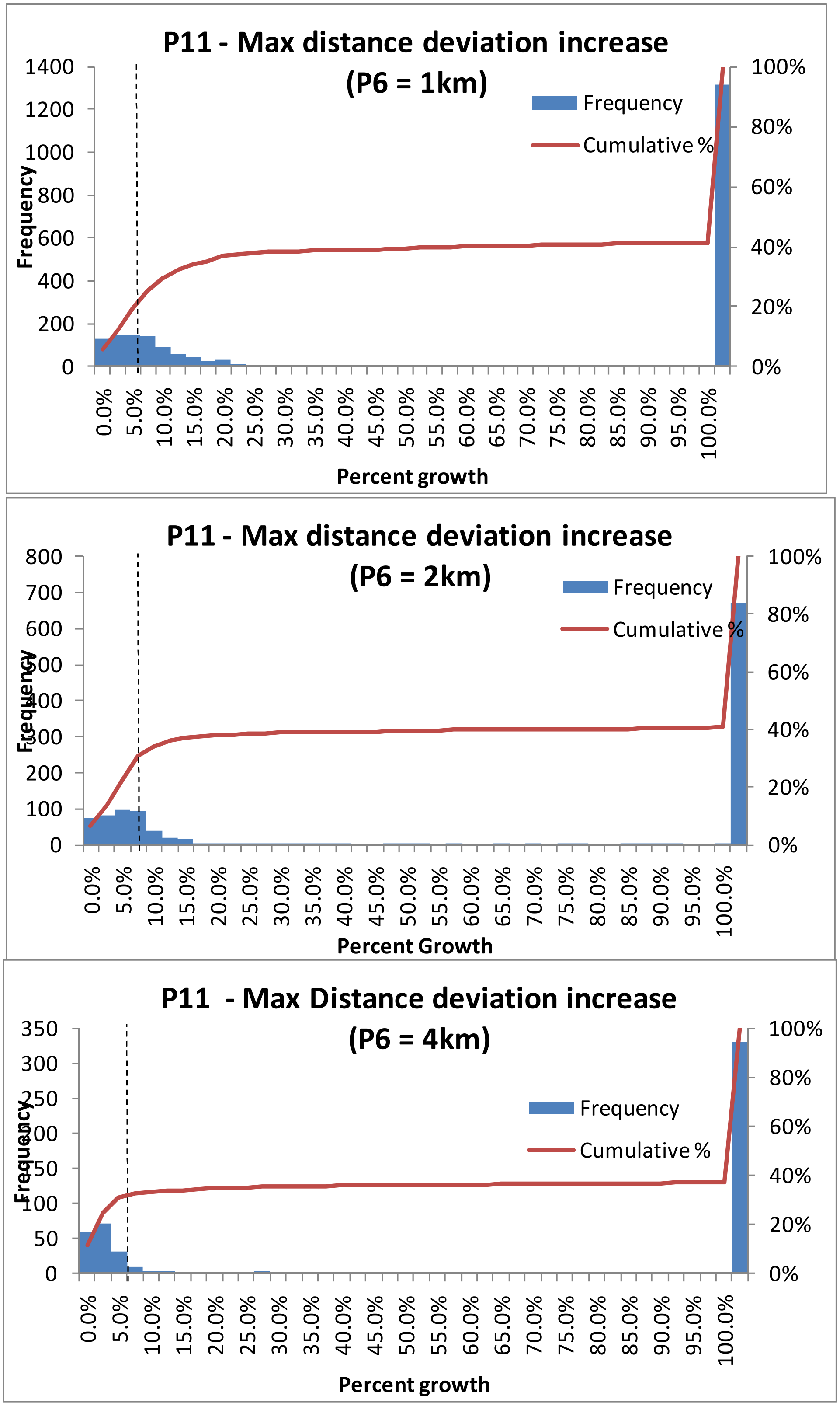

Figure B 10 Selection of P11

Appendix C

Sensitivity plots

Appendix C – Sensitivity Plots

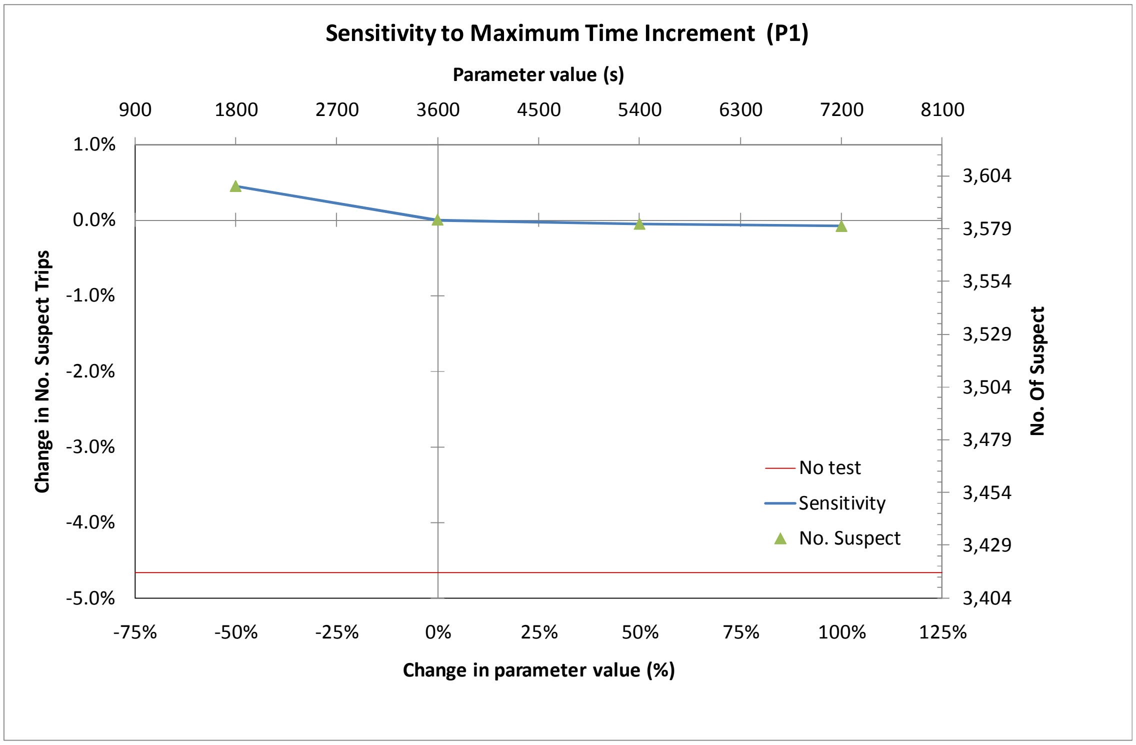

Figure C 1 Sensitivity to P1

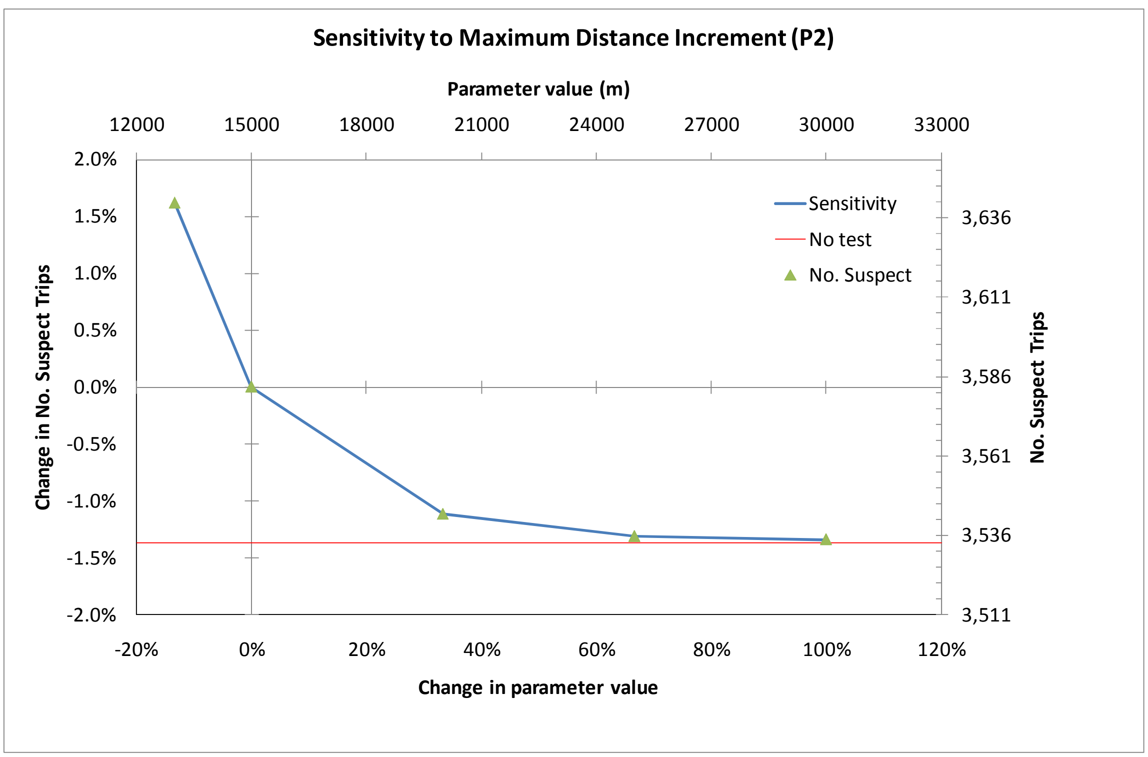

Figure C 2 Sensitivity to P2

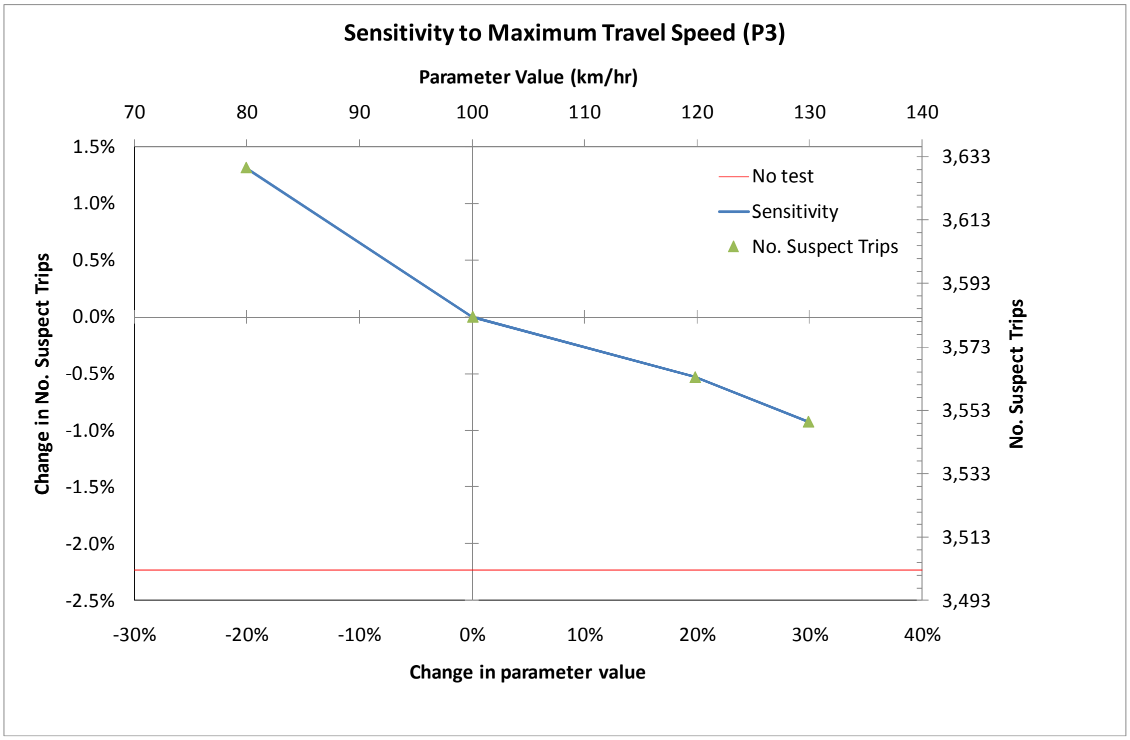

Figure C 3 Sensitivity to P3

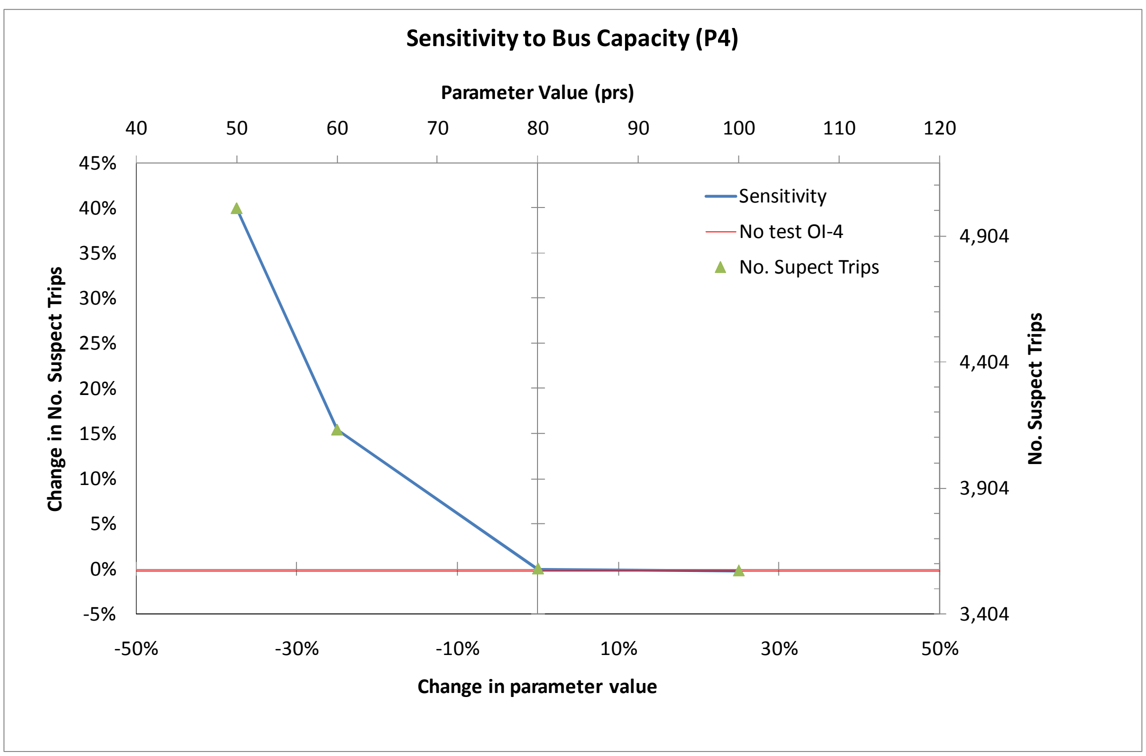

Figure C 4 Sensitivity to P4

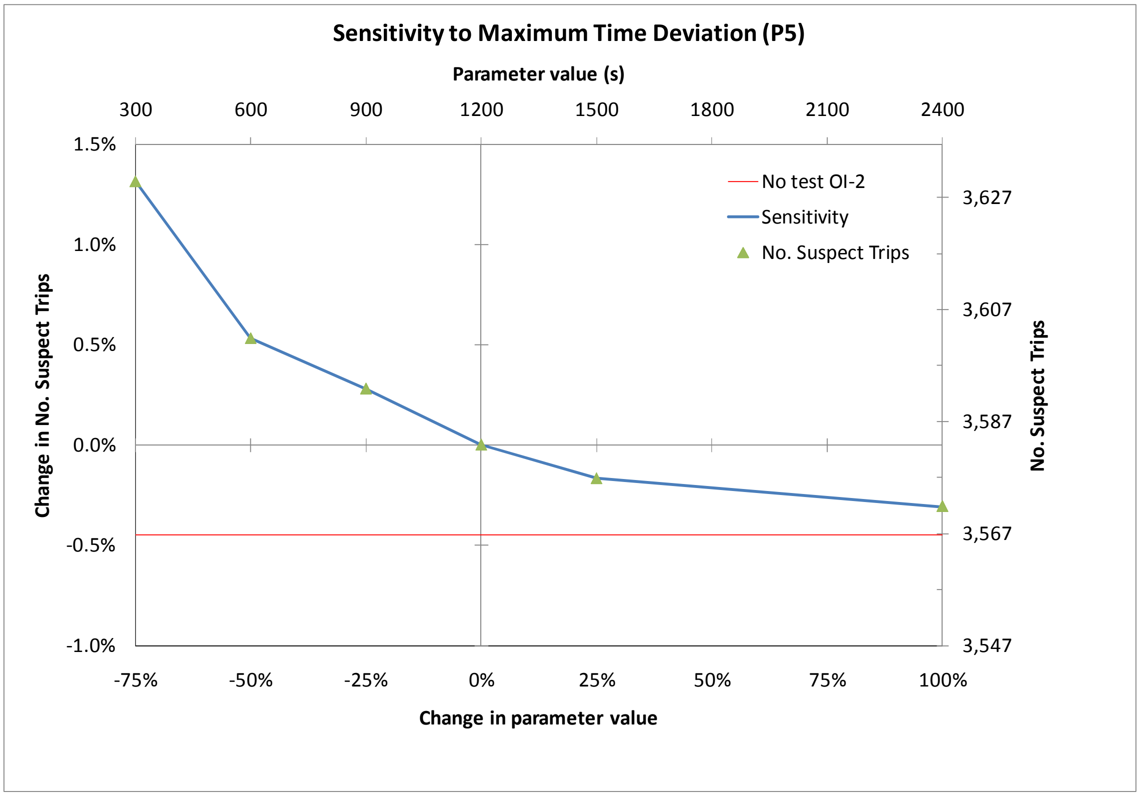

Figure C 5 Sensitivity to P5

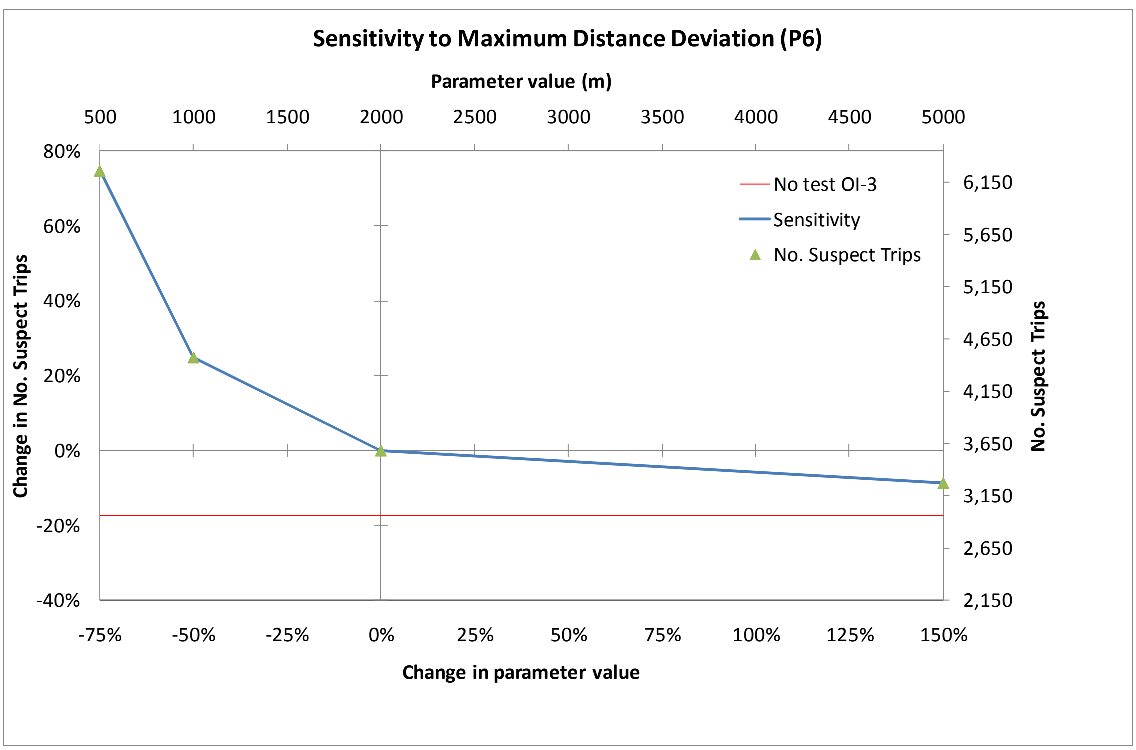

Figure C 6 Sensitivity to P6

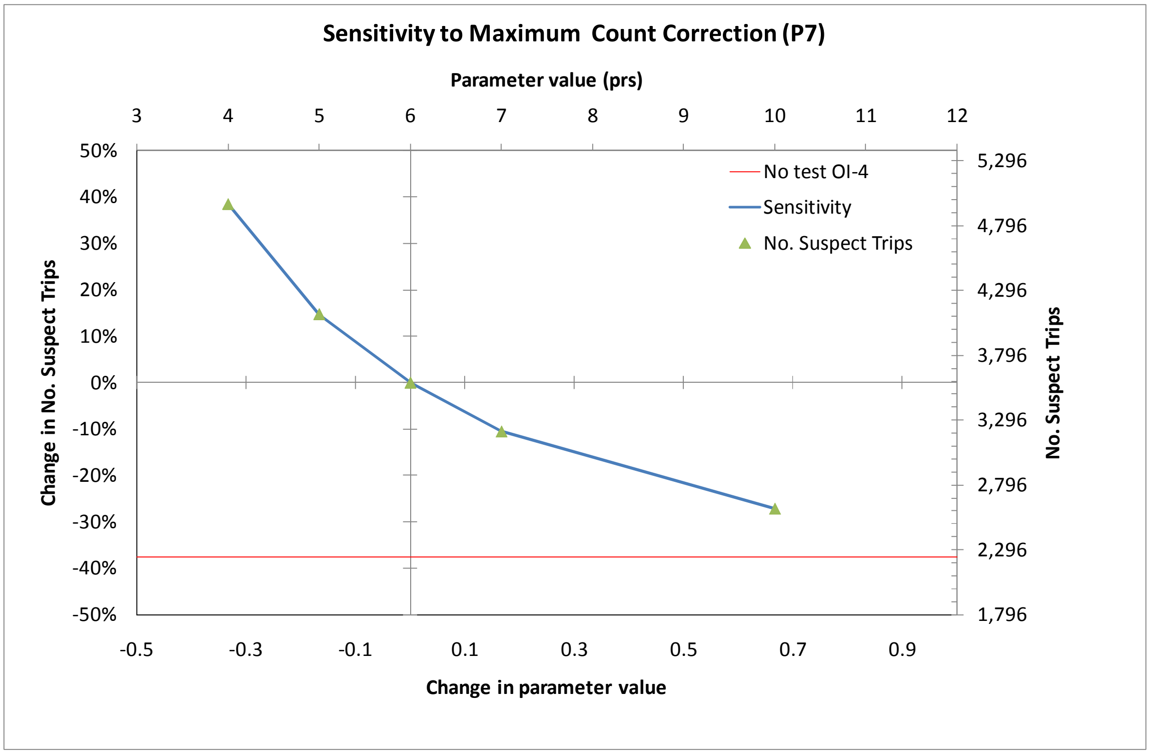

Figure C 7 Sensitivity to P7

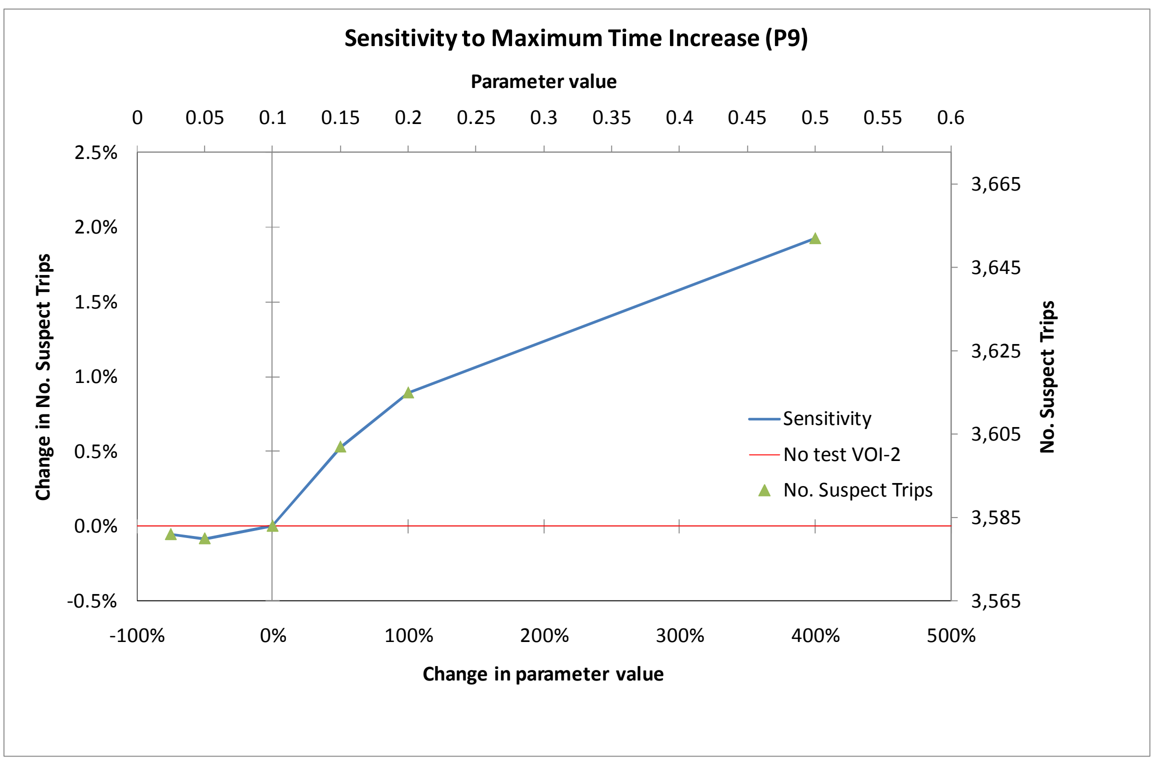

Figure C 8 Sensitivity to P9

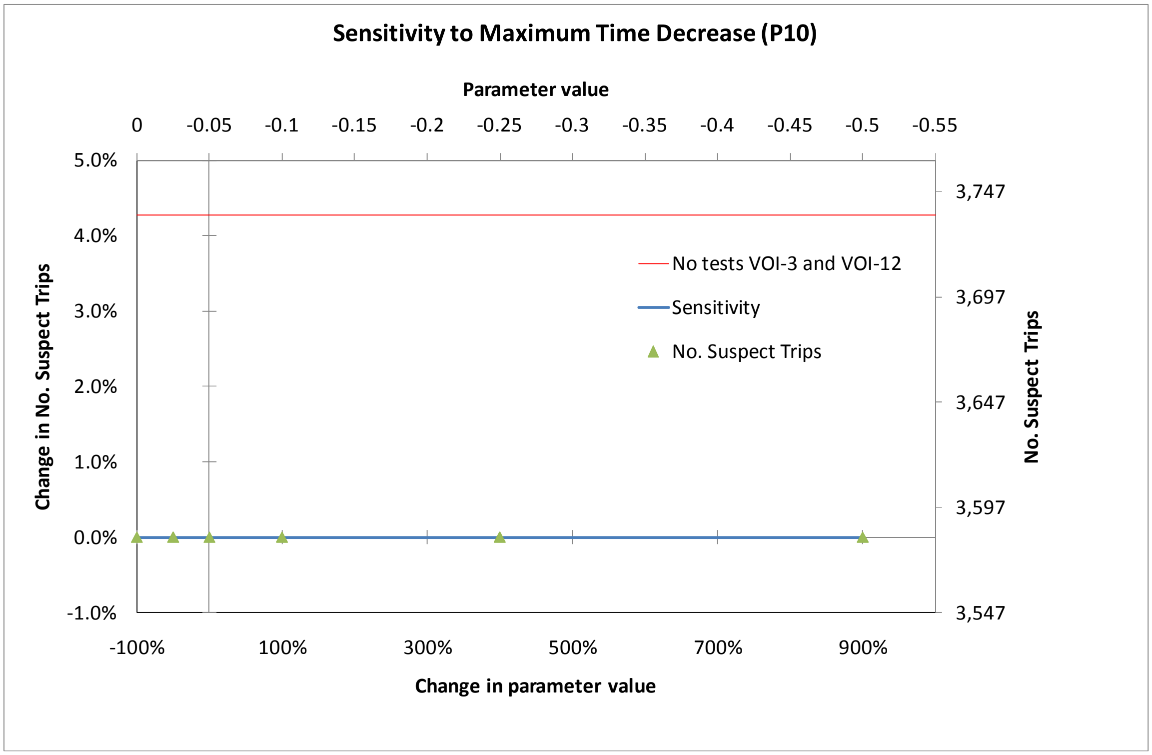

Figure C 9 Sensitivity to P10

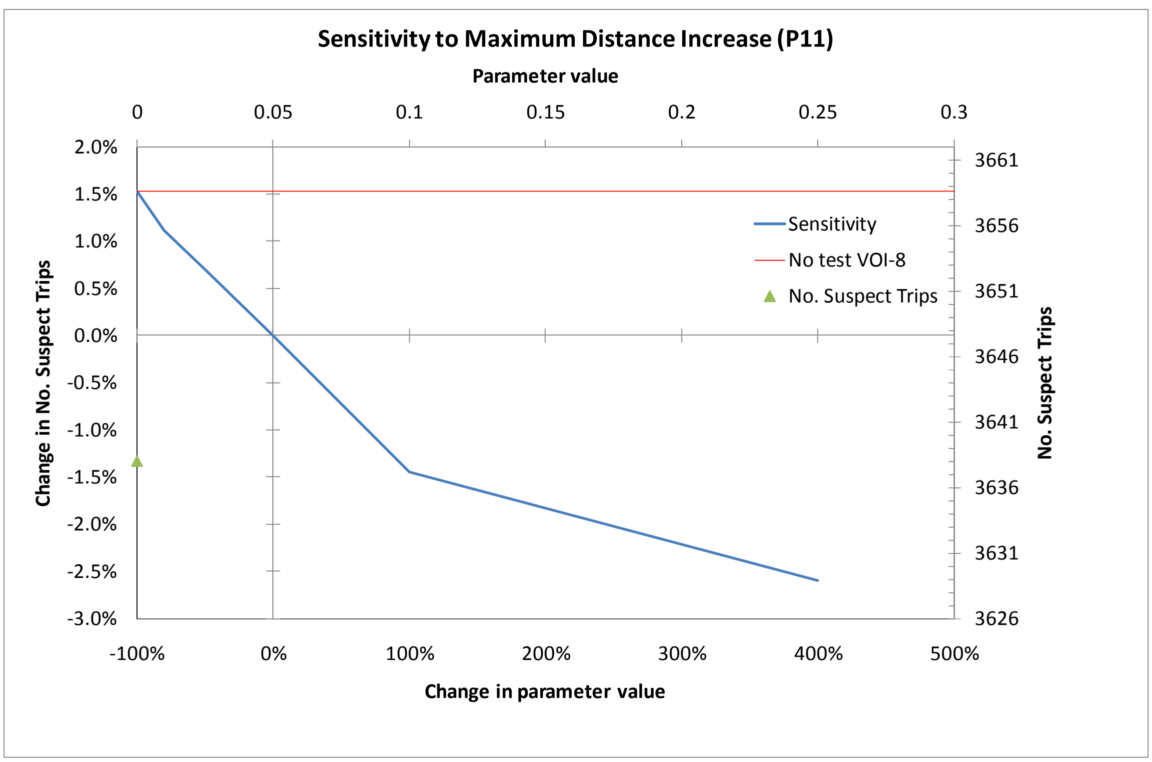

Figure C 10 Sensitivity to P11