Experience in Data Quality Assessment

on Archived Historical Freeway Traffic Data

by

Jothan P. Samuelson

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2011 by the
Graduate Supervisory Committee:

Ram Pendyala, Chair
Soyoung Ahn
Wang Zhang

ARIZONA STATE UNIVERSITY

May 2011

ABSTRACT

Concern regarding the quality of traffic data exists among engineers and planners tasked with obtaining and using the data for various transportation applications. While data quality issues are often understood by analysts doing the hands on work, rarely are the quality characteristics of the data effectively communicated beyond the analyst. This research is an exercise in measuring and reporting data quality. The assessment was conducted to support the performance measurement program at the Maricopa Association of Governments in Phoenix, Arizona, and investigates the traffic data from 228 continuous monitoring freeway sensors in the metropolitan region.

Results of the assessment provide an example of describing the quality of the traffic data with each of six data quality measures suggested in the literature, which are accuracy, completeness, validity, timeliness, coverage and accessibility. An important contribution is made in the use of data quality visualization tools. These visualization tools are used in evaluating the validity of the traffic data beyond pass/fail criteria commonly used. More significantly, they serve to educate an intuitive sense or understanding of the underlying characteristics of the data considered valid. Recommendations from the experience gained in this assessment include that data quality visualization tools be developed and used in the processing and quality control of traffic data, and that these visualization tools, along with other information on the quality control effort, be stored as metadata with the processed data.

TABLE OF CONTENTS

Page

LIST OF TABLES iv

LIST OF FIGURES v

CHAPTER

1 INTRODUCTION 1

Purpose of Report 2

Overview 3

Description of Data Set 5

2 LITERATURE REVIEW 7

Defining Traffic Data Quality 7

Measuring Traffic Data Quality 7

Quality Control Procedures 9

Research and Evaluations on the ADOT FMS 11

Summary of Recommendations 19

3 METHODOLOGY 21

Microscopic Quality Control 21

Data Quality Assessment Tools 24

Macroscopic Quality Control 34

Data Quality Measure Targets and Calculations 36

4 RESULTS 42

Accuracy 42

Completeness 45

Validity 46

Timeliness 48

CHAPTER Page

Coverage 48

Accessibility 50

Summary of Findings 51

Application of Recommendations 54

5 DISCUSSION 56

Educated Questions 56

Lessons Learned 63

6 CONCLUSION 67

Recommendations for Future Work 69

REFERENCES 72

APPENDIX

A EXAMPLE CHARTS 74

Normal Detector Locations 75

Poorly Calibrated Detector Locations 78

Missing Data 82

Information from Combination of Charts 84

B SPATIAL COMPARISON OF ADJACENT DETECTOR CHARTS 88

LIST OF TABLES

Table Page

2.1 Count Accuracy Results 14

2.2 Speed Accuracy Results 14

2.3 Truck Classification Count Accuracy Results 15

3.1 Initial Targets for Six Data Quality Measures 36

4.1 Percent of Data Valid 47

4.2 Coverage of Regional Freeway during Past Three Years 49

4.3 Result Summary for Data Quality Measures 52

5.1 Priority Maintained Percent Validity 62

LIST OF FIGURES

Figure Page

1.1 FMS detector Locations by Type 6

3.1 Annual Hourly Average Speed 26

3.2 Annual Hourly Average Throughput per lane 27

3.3 Annual Hourly Average Occupancy Percent 27

3.4 Annual Average By-lane Profile for Speed, Volume and Occupancy 28

3.5 Distribution by Annual Date for Data Passing Validity Criteria 29

3.6 Distribution by Weekday for Data Passing Validity Criteria 30

3.7 Count of Annual Quality Control Flags by Hour of the Day 31

3.8 Flow–Density Relationship Data Points 32

3.9 Speed–Density Relationship Data Points 33

3.10 Speed–Flow Relationship Data Points 33

4.1 By-lane Profile at “poorly calibrated” Detector 44

4.2 Distribution of Error Flags by Peak Period 46

4.3 Frequency Distribution for Detector Spacing 49

4.4 Pie Chart of the Relative Weight of Measures on Data Quality 53

5.1 Distribution of Hourly Error Flags by Time of Day 57

5.2 2009 Distribution of Detector Level Validity 60

5.3 2008 Distribution of Detector Level Valid 61

CHAPTER 1

INTRODUCTION

Traffic data is used for several applications by agencies responsible for planning, building, and managing transportation facilities. The quality of traffic data and the information produced from the data are critical factors that affect the ability of agencies to effectively manage transportation resources. Analysis and research on continuous monitoring permanent detector data have identified several issues regarding quality, including missing values and significant accuracy error. Even where problems are not clearly defined, many users have developed skepticism, leading to duplicate data collection efforts, and vast amounts of underutilized traffic data. Without accurate and reliable detector data, transportation management decisions based on real-time or historical data are compromised.

Quality control on detector data is an essential component of any traffic analysis. Quality control efforts and validity filtering criteria in the various applications of traffic data are not new, and yet these are probably underutilized in practice. Developing a system for assessing and controlling traffic data quality is an extensive effort that improves the data but cannot fully do so to the point of arriving at a perfect representation of real traffic conditions. The cost and difficulty of obtaining high quality traffic data often requires an analysis to settle for data that is in one way or another less than ideal. The relevant question in doing so becomes “how good does the data have to be to be useful for its intended purpose?” Of particular concern, is the potential to lose any real sense of the data quality even when preliminary screening efforts have taken place. The truth in a traffic analysis result can only really be understood where error rates, sample bias and size are carried through each step of the analysis allowing for a “confidence level" to be attached to each result. Experienced analysts who work hands on with the data usually develop an intuitive sense for data quality, but the state of the practice has not included effective communication of any sort of confidence level associated with analyses results. Effective quality control, therefore, is needed not only in filter out "bad" data but in providing an understanding of the inherent quality of the "good" data.

Purpose of this Report

This research was conducted to support the performance measurement program at the Maricopa Association of Governments (MAG) and the analysis taking place of continuous monitoring archived freeway detector data in the Phoenix Arizona metropolitan region. The interest at MAG has been in developing freeway performance measures reporting throughput, speed, travel time variability, and various other congestion measures. Trending in performance measures between analysis years is desired. Concern, however, exists regarding the extent to which data quality issues play into and affect facility performance measures and trends. The analyses described in this research are steps toward improved use of continuous monitoring archived traffic detector data and calculation and measurement of annual freeway performance for the MAG region. Quality control procedures have been applied in filtering out erroneous data and in identifying traffic sensor locations and data that most effectively contribute to useful performance measurement.

More importantly than the application of this research specifically to performance measurement for the MAG region is the experience gained in assessing and communicating traffic data quality. Lessons learned in this assessment can be applied to other sources of traffic data wherever an understanding of quality, and the implications of that quality, is needed in improving an analysis task.

This report makes at least two contributions to the analyses of traffic data. First is as a case study in the applications of the data quality measurement framework presented in the literature based on six fundamental data quality measures of accuracy, completeness, validity, timeliness, coverage, and accessibility. The guidelines in this framework set a standard for a consistent starting point for analyzing traffic data quality. Second, several data quality visualization charts have been created as a standard output in the data processing program created. These visualization tools, stored with the processed data for each traffic sensor, both serve to aid in filtering out bad data and to communicate important quality characteristics of the data considered to be valid. A confidence level derived in the use of these visualization tools is gained primarily as an intuitive sense from experience seeing the data. This experience, coupled with future analytical work, potentially contributes to the needed development of a statistical framework for measuring a confidence level for traffic data based on quality attributes.

Overview

This exercise takes place using continuous monitoring archived detector data from the Arizona Department of Transportation (ADOT) Freeway Management System (FMS). The later portion of this introduction provides some background and description of the data set.

Chapter two of this report will provide an overview of research on traffic data quality, including recommendations and practices used by researchers, agencies and State DOTs. Evaluations and research relating to the ADOT FMS data set used here will be presented.

Chapter three describes the methodology applied in identifying and controlling faulty data during the processing and aggregating of the archived traffic data. The analysis looks both at error on microscopic data elements in individual 5-minute data rows as well as macroscopic criteria and comparisons between spatially adjacent detectors along a given corridor. Ten visualization charts utilized in the quality assessment and control of the data are described. Definitions and calculation methods are also described for the six recommended fundamental measures of traffic data quality upon which the data quality measurement framework suggested in the literature is based.

Chapter four describes the quality of the traffic data for the ADOT FMS using each of six measures or categories contributing to the overall quality of the data. While the finding reported from this exercise may be helpful, they are not specifically conclusive, as the emphasis in this report is on the various elements of the assessment and how they contribute to understanding and measuring quality. Effort has been in summarizing data quality results for the region wide system and not for individual detector locations. Also described in this chapter is the extent to which the various recommendations from the literature have been applied in the analysis taking place with this project.

Chapter five provides some discussion regarding the application of the framework for data quality measurement and the use of data quality visualization tools. Questions highlighted in the quality assessment that took place, as well as some of the qualitative lessons learned, are further discussed.

The report concludes in Chapter six with a summary of the data quality assessment and major points and contributions provided. Some recommendations for future work are suggested.

Description of Data Set

Permanent traffic detectors cover roughly 250 miles of the Phoenix metro area freeway. These sensors continuously collect average speed, volume, and the duration of time in which a detector is occupied (occupancy), which is really a measure of density. These averages are communicated on a 20 second basis to the traffic operations center year round. Speed information from these detectors feed the ADOT real-time speed map at the http://az511.com website. All traffic data is then archived and data aggregated to 5-minute averages and higher are placed on the ADOT ftp.az511.com file transfer protocol site. This data, along with documentation describing the data, can be downloaded via internet by any potential user.

For this research, 5-minute data for all traffic detectors has been analyzed. The 5-minute data files contain lane specific information, and as such, are quite large. An annual raw data file for each detector location will for example exist as a matrix of 180 columns by just over 105,000 rows. Because complete 2010 data has not yet been made available, the performance measurement analysis takes place for the 2008 and 2009 years only. Results from the data quality assessment reported here are in most cases reported for 2009 data only.

Installation of traffic sensors is ongoing and currently there are 286 active detector locations. The number of active detectors available during previous years is 228 for 2009 and 199 for 2008. These sensors exist primarily as embedded inductive loops and pole mounted passive acoustic detectors (PADs). Of the current detectors, 66% are loops and 34% PADs. Of the 228 detectors locations in 2009, 58% are loops and 42% are PADs. Figure 1 on the next page is a map of the Phoenix metro region showing all detector locations by technology type.

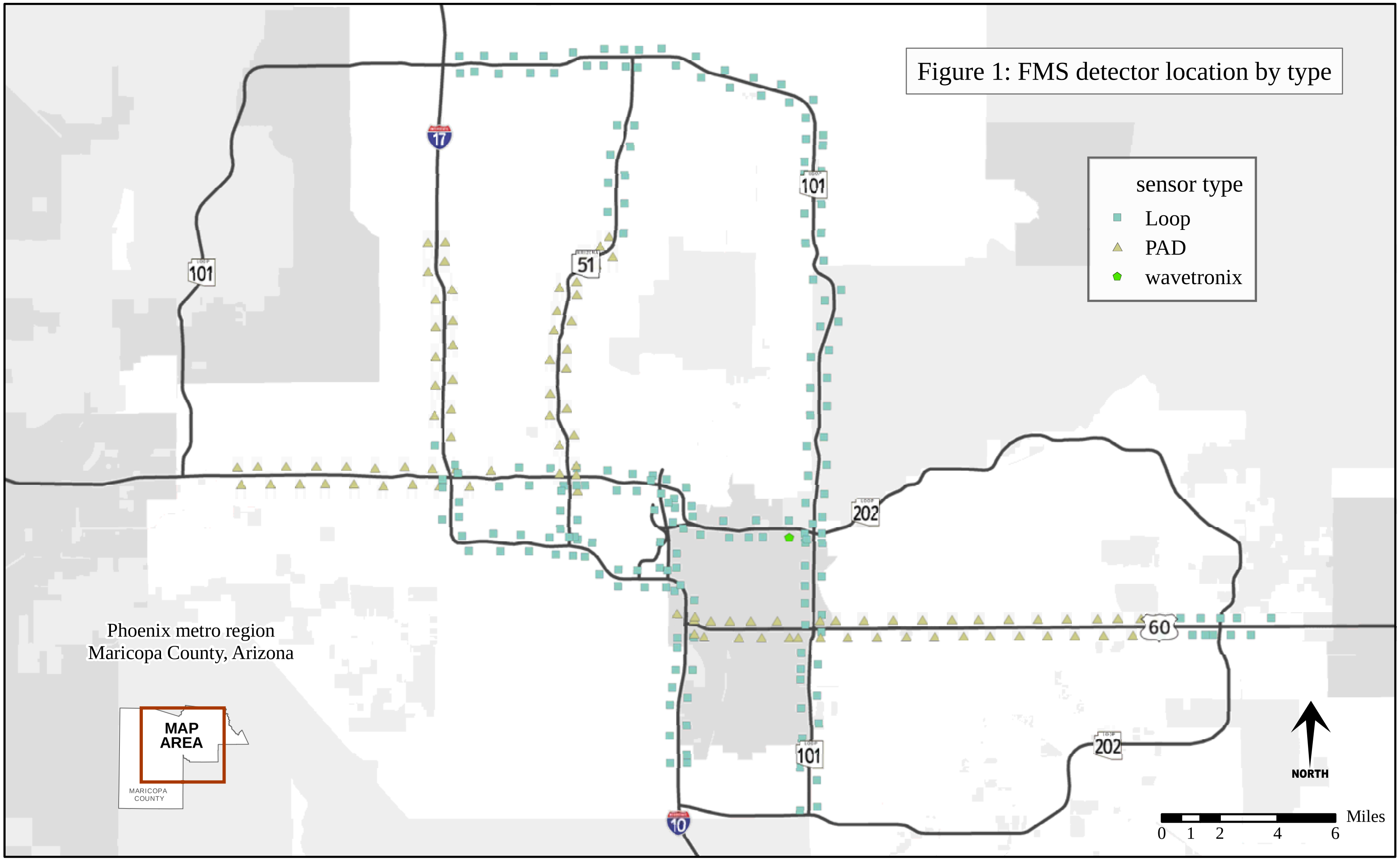


Figure 1: FMS detector location by type

CHAPTER 2

LITERATURE REVIEW

In 2003, the Federal Highway Administration (FHWA) sponsored two regional workshops on traffic data quality. The objective of the workshops was to stimulate discussions and obtain input from the workshop participants in developing an action plan that addresses traffic data quality issues. The resulting workshop proceedings defined an action plan that builds upon the findings of three white papers prepared for and used in the workshops, as well as input obtained from workshop participants (Fekpe and Gopalakrishna 2003), (Turner 2004), (Margiotta 2002), (Middleton et. al. 2003).

Defining Traffic Data Quality

The first of the three traffic data quality white papers, Turner (2004), provides the following definition of traffic data quality:

"*Data quality is the fitness of data for all purposes that require it. Measuring data quality requires an understanding of all intended purposes for that data.*"

The definition of data quality is therefore a relative concept that can have different meanings for different users. Data quality is also a composite measure of several elements or characteristics contributing to the overall quality of the data. The white paper recommends that goals, and target values for data quality measures, be established at the jurisdiction or program level based on a clear understanding of intended uses.

Measuring Traffic Data Quality

The foremost action item from the FHWA sponsored workshops was a call for "guidelines and standards for calculating data quality measures". In response, a report was published called *Traffic Data Quality Measures* that presents a framework with methodologies, guidelines and standards for calculating data quality (Battelle 2004). The framework developed is based on six fundamental characteristics or measures of traffic data quality proposed in the Turner (2004) white paper. These characteristics are defined as follows:

1. Accuracy – The measure or degree of agreement between a data value or set of values and a source assumed to be correct. It is also defined as a qualitative assessment of freedom from error, with a high assessment corresponding to a small error.
2. Completeness (also referred to as availability) – The degree to which data values are present in the attributes that require them. In other words, completeness is the degree to which data is not missing and is typically described in terms of percentages or number of data values.
3. Validity – The degree to which data values satisfy acceptance requirements of the validation criteria or fall within the respective domain of acceptable values. Validity is commonly reported as the percentage of data values that either pass or fail data validity checks.
4. Timeliness – The degree to which data values or a set of values are provided at the time required or specified. Timeliness can be expressed in absolute or relative terms.
5. Coverage – The degree to which data values in a sample accurately represent the whole of that which is to be measured. As with other measures, coverage can be expressed in absolute or relative units.
6. Accessibility (also referred to as usability) – The relative ease with which data can be retrieved and manipulated by data consumers to meet their needs. Accessibility can be expressed in qualitative or quantitative terms.

The Battelle (2004) report on data quality measures includes guidance on data quality targets for different transportation applications as well as the level of effort required to develop a data quality assessment system. Transportation planning applications, for example, might target accuracy and validity measures at less than 10% error, and expect to spend close to 160 hours developing a system for data quality assessment and reporting. Guidelines are also provided for sharing data and in creating and using metadata in reporting data quality. Limited beta testing provided the opportunity to validate the concepts and methodologies of the framework and also to validate some draft estimates for data quality targets and estimates of level of effort.

Recommended in the report is that data quality targets and estimates be tested and validated based on actual experience in using the framework and guidelines. Three case studies, using mostly real data, were contrived for the purpose of illustrating the application of the framework for three different primary groups of data consumers.

Ahn, Rakha, and Hill (2008) explore the relevance of these six data quality measures for real-time travel time, speed and weather information, and provide recommendations for quality levels for each of the six measures. An overview is given of utilization in public and private sectors of the data quality measures specific to real-time transportation applications. Further development and more extensive validation of standards for evaluation of data quality are recommended.

Quality Control Procedures

The 2003 data quality action plan also called for a "synthesis of validation procedure and rules used by various states and other agencies for traffic monitoring devices". Turner (2007) addresses this action item with a report called *Quality Control Procedures for Archived Operations Traffic Data: Synthesis of Practice and Recommendations*. In the report, three categories of data quality procedures are presented.

1. Univariate and multivariate range checks – These criteria typically correspond to the minimum, maximum, or range of expected values for a single variable or a combination of variables.
2. Spatial and temporal consistency – These criteria evaluate the consistency of the traffic data across lanes or as compared to nearby locations or time periods.
3. Detailed diagnostics – These refer to the set of validity criteria in the literature that rely on detailed sensor output not typically available in the archived traffic data.

In the review of current practices provided in the Turner (2007) synthesis report the quality control rules utilized by the following nine agencies are specified: Virginia

DOT, Caltrans, Maryland State Highway Administration, Florida DOT, FHWA (for the Mobility Monitoring Program), Kentucky Transportation Cabinet, Maricopa County DOT, Oregon DOT, and Wisconsin DOT. The report contains individual descriptions and a combined summaries of the criteria used in the nine data archives. One of the findings identified was that the validity criteria were similar across the different data archives. For many of the archives, validity rules and criteria were modeled after or consistent with the criteria used in the FHWA Mobility Monitoring Program. The Mobility Monitoring Program used the following quality control checks to identify invalid detector data from archived data sets for nearly 30 cities (Turner et al. 2004):

1. Controller error codes: remove numeric error code values (typically "-1" or "255")
2. No vehicles present: replace zero speed values with null/missing speed values when Volume=Occupancy=Speed=0
3. Check for consistency of elapsed time between data polls
4. Check for and remove duplicate records (location identifier, date and time stamp are identical)
5. Check for and remove date, time, and location identifier values that are not in the valid domain range
6. Maximum volume (varies based on time interval, invalid if Volume>3000 vphpl as max rate)
7. Maximum occupancy (invalid if Occupancy>95% for 20-30 second periods, Occupancy>80% for 1-5 minute periods)
8. Minimum speed (invalid if Speed<5 mph)
9. Maximum speed (invalid if Speed>100 mph for 20-30 second periods, Speed>80 mph for 1-5 minute periods
10. Multivariate consistency (invalid if Speed=0 and Volume>0 [and Occupancy>0], or if Volume=0 and Speed>0, or if Occupancy>0 and Speed=0 and Volume=0)
11. Truncated occupancy values of zero (invalid if Occupancy=0 and Volume>[(2.932×Speed×Elapsed Time)/600]
12. Maximum estimated density (invalid if density > 220 where density=[Volume×(3600/Elapsed Time)/Speed]
13. Consecutive identical volume, occupancy and speed values (invalid if more than 8 consecutive volume and occupancy and speed values are identical, including zero values).

The systematic use of visualization tools for quality control in filtering bad data is not common practice. Tufte et al. (2007), however, presents an example of visualizations being used to communicate data quality for the Portland, Oregon regional transportation archive listing in addition to presenting validity criteria in detecting malfunctioning detector data. The Caltrans PeMS traffic data archive website similarly reports detector health statistics including a pie chart showing the distribution of suspected error types for each detector (Caltrans PeMS).

Another area of data quality procedures is the imputation of missing data. Smith et al. (2003) provides an overview and methods for evaluating imputation techniques. Techniques range from simple and historical average to data augmentation and linear regression between adjacent lanes and time periods. Fernandez-Moctezuma et al. (2009) provides a thorough and recent review of imputation techniques and procedures reported in the literature.

Research and Evaluations on the ADOT FMS

Jonas (2001) provides a history of early implementation of the Arizona Department of Transportation (ADOT) freeway management system (FMS) traffic detectors and evaluation of passive acoustic technologies that were used along with loop detector technology. The first phase of the traffic detector implementation began operation in 1995 and ADOT utilized speed-trap pairs of magnetic loop detectors installed at 1/3 mile spacing. Due to concerns that saw cut loops could significantly weaken the pavement, non-intrusive technologies were selected for the next phase where loops had not already been imbedded in the pavement. Instrumentation taking place in 1998 started with installation of Smartsonic passive acoustic detectors (PADs) mounted above the roadway shoulder on existing polls. It became evident however that the Smartsonic sensors did not properly indicate speeds lower than about 30 mph, and when volumes were above 1,900 vehicles per hour per lane, under reporting took place to as little as less than half the actual volume. Maximum hourly lane capacities can exist above 2,000 vehicles and validity criteria do not typically reject a data element as invalid until hourly rates are at or above 3,000 vehicles. Inaccuracies in the data were therefore biased to those locations where congestion was most significant. Later the same year, SmarTek SAS-1 passive acoustic detectors were proposed along with a test to demonstrate how well the SmartTek PADs functioned in comparison to embedded magnetic loops. The primary criteria defined for agreement between magnetic loops and the PADs was +-10% agreement for volume, and +- 7 mph agreement for speed greater than 15 mph. The first two demonstration tests failed to yield satisfactory results. After significant effort was given to calibration of the SAS-1 PADs, a third test took place that included comparison of PAD data to both loop detector data and counts taken from video footage. Results showed good agreement with speed and volume during important times of the day. Charts included in the evaluation do however show significantly lower speeds from PAD sensors during periods of low traffic volumes, and slightly lower volumes during periods of high volume.

With full-scale deployment of the PAD detectors, further evaluation took place comparing adjacent detectors upstream and downstream from one another. Where these detectors exist between the same cross-street entrance and exit ramps, volume counts should be identical. Discrepancies in volumes on the comparison charts therefore became the basis for further investigation of PAD calibration problems. Results of the evaluation showed that SAS-1 PADs, when properly calibrated, count up to 8% less traffic than loops during high-volume congested traffic, and up to 5% more traffic during low volume periods. Speeds detected were within +-5% of loop speeds, except during very low volume where PADs can indicate significantly lower speeds than loops. PAD performance was shown to be affected significantly during rain, where counts were significantly higher than loops and speeds where 25 mph or lower. Just after rain, volume detected by PADs was shown to fall significantly below loop volume for eight hours or more, perhaps due to wet pavement.

Through the phases of implementation and expansion of the FMS, passive acoustic detectors and dual loops have both been installed, with PADs being installed as recently as May, 2008. Currently, the list of active detector locations is just under 34% as SmarTek SAS-1 or SAS-2 PADs. Figure 1 in first chapter shows active detector locations by technology type.

In 2005 and 2006 the Maricopa Association of Governments (MAG) contracted with the Texas Transportation Institute (TTI) to conduct an *Accuracy Evaluation of Arizona DOT Freeway Management System Detectors* against reference data with a known accuracy (Turner 2008). ADOT and MAG have designated 58 of the freeway detector locations as priority maintenance locations for regional traffic monitoring. 30 of the 58 locations are loop detectors and the remaining 28 are SmartTek PADs. The evaluation took counts from recorded video footage during both high traffic peak-flow conditions and off-peak light traffic conditions at all the priority maintained loop detectors and at 10 of the 28 PAD detectors. Vehicles classification counts took place at a subsample of these locations into three classification categories consistent with the categories reported in the archived FMS data (Passenger car, Truck length 30-55 feet, and Truck length > 55 feet). At 21 of detector locations, speeds were periodically measured on a single lane for at least 5-minutes at a time using a LIDAR gun. Based on poor count accuracy during the first year of the evaluation, TTI worked with ADOT maintenance staff to calibrate two of the PAD locations, thus allowing for a before-and-after-calibration evaluation at these locations.

Results showed nearly all loop detector with an average count error less than or equal to 10%. PADs however, appeared to undercount total traffic with an average peak traffic error of 34%, and all PADs had average errors greater than 10%. Table 2.1 below shows evaluation results for traffic counts.

Table 2.1: Count Accuracy Results

type	det count	time period	absolute error range	mean absolute error
loops	24	peak traffic	1% - 20%	6%
loops	29	off-peak	1% - 10%	3%
PADs	10	peak traffic	13% - 91%	34%
PADs	10	off-peak	4% - 72%	24%

After calibration at two PAD locations, evaluation results show substantial improvement to traffic count accuracy from 27% and 18% error during the first year to 7% error at both locations in the second year.

Speed error results showed nearly all loop detectors with an average speed measurement of less than or equal to 10%. Most of the PAD detectors, however, showed error in excess of 10%. Table 2.2 on the next page shows evaluation results for speed.

Table 2.2: Speed Accuracy Results

type	det count	time period	absolute error range	mean absolute error
loops	13	peak traffic	3% - 11%	7%
loops	19	off-peak	2% - 12%	6%
PADs	8	peak traffic	9% - 68%	26%
PADs	7	off-peak	5% - 17%	17%

At loop detectors locations speeds are calculated as the known distance divided by the travel time between the paired loops. Speed evaluation results for loop detectors with absolute error greater than 5% were provided to ADOT to allow for estimated field lengths between paired loops to be adjusted, thus improving accuracy of estimated speeds at these locations.

Truck classification performed poorly at nearly all locations, consistently undercounting the total number of trucks at both loop detectors and PADs. Table 2.3 below shows evaluation results for truck classification counts. A much longer duration of data collection may in fact be necessary before directly comparing truck count error rates with the error rates for total traffic because the actual number of trucks represents only a small portion of the total traffic.

Table 2.3: Truck Classification Count Accuracy Results

type	det count	time period	absolute error range	mean absolute error
loops	13	peak traffic	0% - 93%	46%
loops	13	off-peak	3% - 100%	80%
PADs	2	peak traffic	29% - 100%	73%
PADs	2	off-peak	72% - 100%	89%

Calibration of the two PAD detectors did not improve truck count accuracy to an acceptable error.

Evaluations of detector accuracy like that performed by MAG and TTI are rare. For obvious reasons, accuracy evaluations that do get conducted by managing agencies or private sector technology providers are not usually published.

A study, *Enhancing Arizona Department of Transportation's Traffic Data* Resources, was undertaken and published in 2001, to assess ADOT's traffic data needs and identify sources of traffic data (Sterling et al. 2001). The primary objective was to identify impediments to efficient collection, management, and dissemination of the data. The study also included evaluation of the technologies used by ADOT to process, store, manage, and disseminate data across the agency. The study reported several important challenges and outlined a clear and comprehensive Implementation Plan. Cited challenges facing the ADOT freeway data included reduced operation and consistency of the traffic sensors, and reactive rather than regular maintenance and calibration efforts. The availability of data was consequently not so great a concern as the validity and quality of the data provided. Distrust has led to underutilization of data and redundant collection efforts by various agencies and groups even within ADOT. The study identified personnel and budget resource limitations devoted to the processing and management of traffic data available. The Implementation Plan for enhancing traffic data resources presented recommendations for ADOT to address the identified needs and issues. This plan calls for actions summarizes as follows:

1. *Establishment of a Traffic Data Working Group* – Enhanced communication and coordination between divisions, and groups involved in collecting, processing, or using traffic data will allow ADOT to improve administration of funding, set priorities and monitor progress.
2. *Adopting and publishing of procedures, standards and guidelines for data collection, processing, metadata, and data documentation* – Standards should be consistent between all groups (at least within ADOT) performing data collection and processing. More importantly, these standards need to be stated and included with data being used by other agencies and divisions. "Truth-in-data" information might include description of data collection equipment, time period of data collection,

analysis methods used, accuracy of the estimate, and any factoring or adjustments made.

3. *Development of a Traffic Data Clearinghouse and Traffic Data Warehouse* – A clearing house would store information about the existence of traffic data and ideally incorporate a GIS system such that traffic data are spatially referenced. A Warehouse would be implemented in a secondary phase to store all actual data in a standard format.

Recommended action items were addressed in part with a research program in early 2002 to consolidate ADOT's varying independent applications monitoring traffic activity and roadway and weather conditions. The resulting report, *ITS Traffic Data Consolidation System* (Guerra and Satoyoshi 2005) summarizes the integration of data applications into a single user friendly graphical user interface. The Highway Condition Reporting System (HCRS) is a GIS-based and web-enabled application, accessible to ADOT staff through ADOT's own intranet, containing various real-time data sources, including congestion and travel times. The website az511.com provides public access to the real-time information hosted in the HCRS. Another application currently employed by ADOT to maintain traffic counts is the Transportation Data Management System by Midwestern Software Solutions, accessible at the website adot.ms2soft.com. This application system has been used to store and provide access to traffic counts collected from various ADOT sources and has recently begun as well to host hourly average counts from actual FMS detectors for 2009 and 2010.

In 2006 ADOT approved a three phase research project for a *State of the Art Evaluation of Traffic Detection and Monitoring System* as is outlined in the State Planning and Research (SPR) program for project 627 (ADOT 2010). The first two phases of the project, completed by the Texas Transportation Institute (TTI), included a state-of-the-practice review to identify the most appropriate detection technology to meet local needs, and the design for a detector testbed facility (Middleton et al. 2007). The research conducted by TTI outlined some of the specific concerns that exist with both PAD and loop detector technologies. Potential issues with inductive loops include: broken or shorted loops, shifted or incorrectly placed loops, false actuations due to similar frequencies, malfunctioning detector cards in the roadside cabinet where 20-second data is aggregated and transferred to the traffic operations center, and sensitivity settings. The nominal spacing for loop detectors is 18 feet, but actual spacing varies from 17 to 24 feet. Checks on calculated speed based on assumed loop spacing indicate that loop detector speeds are off by as much as 15 mph. Issues with PADs could include pavement texture, reflection echoes, detector alignment, various software settings, and communication errors between the cabinet and the traffic operations center. The state-of-the-practice review cites several specific studies where various traffic detection technologies are evaluated. Results indicate that under all environments, loop detectors appear to be more accurate than all the newer non-intrusive detector types. Newer technologies that have the most promise are microwave radar, video imaging, and magnetic detectors, where traffic counts can be consistently within 5% of real counts.

In the ADOT published research program it states that all funding for construction of the detector testbed facility outlined for Phase 3 of the SPR-627 research project had been withdrawn and committed to another project (ADOT 2010). The systems engineer at the ADOT traffic operations center, has indicated that ADOT is no longer installing new detector locations with PADs and intends to phase out existing PAD locations sometime in the future in favor of loop detectors. Supporting this claim are the 56 new detector locations added in the past two years as embedded induction loop sensors. No indication has been found that resource allocation for improved maintenance and calibration of existing detectors has yet changed.

Summary of Recommendations

Consistent themes in the recommendations provided in the research have emerged. An important distinction in the recommendations exists between “quality assurance” and “quality control”. In the synthesis on practice and recommendations for quality control procedures (Turner 2007), cited previously, quality control is the identification, review, and treatment of data not meeting validity criteria and can only take place after data has been collected. Quality assurance is a much broader term encompassing actions taken throughout the entire traffic monitoring cycle, before and after data collection, and would include things like regular maintenance and calibration efforts. Action steps restricted to simply fixing data are less effective in influencing long term changes to the causes for poor data quality, and yet the first steps in moving toward a position that influences how traffic data are collected and archived include user end improvements in how data is processed and understood. The following is a list summary of user end recommendations felt to be most important or relevant for application of archived detector data for annual performance measure reporting.

1. Implementation of at least basic foundational data validity criteria in the automated processing of archived detector data.
2. Development of additional validity criteria relevant to the data and intended purpose for the data, and the use of visual reviews when feasible.
3. Use of consistent standards and guidelines for calculating traffic data quality measures.

4. Metadata to document quality control procedures and results along with relevant information about the data.
5. Data sharing efforts that avoid duplication in data collection and underutilization of data. Primary in this is the need for a centralized clearinghouse where information about data sources and characteristics of the data are recorded.
6. Sensitivity studies that demonstrate the value of data and highlight the effects of poor quality data on various applications. These studies allow for refinement of targets for data accuracy and other quality measures.

The guidance provided in the literature on data quality in general, and the research conducted specific to the ADOT FMS, provide direction in the assessment of data quality described and discussed in the remainder of this thesis report. The user end recommendations just outlined have can be applied fully or partially, as has been the intent with the assessment taking place.

CHAPTER 3

METHODOLOGY

This chapter will describe the process by which the quality of the data has been evaluated. The specific methodology used in the processing and aggregation of the archived detector data are discussed only briefly in relation to the quality control procedures. The quality control effort has taken place at two distinct levels. The first is at the microscopic level in the application of pass/fail validity criteria to all individual 5-minute data rows. The validity filters used at this stage are not new and therefore receive little discussion. The second level is a macroscopic analysis focused on evaluating the validity of each detector in whole. The macroscopic rules serve as heuristics and rely heavily on a set of quality assessment visualization tools. The principle goal is in utilizing the best available data in producing the needed performance measures. Beyond the accept/reject decision steps in screening the data at each of the two levels, a quantitative assessment of the data quality feeding the performance measurement is desired. The assessment utilizes the six data quality measures described in the literature, which are: accuracy, completeness, validity, timeliness, coverage, and accessibility.

The order in this chapter follows the order in which the different stages of the analysis take place. First is described the detector level aggregation process and microscopic validity criteria, then the data quality assessment tools, third the macroscopic corridor level quality control analysis, and finally a description of the method by which the six data quality measures are calculated.

Microscopic Quality Control

Preliminary sorting of raw data files obtained from the ADOT ftp site has resulted in a single text file for each detector containing all 5-minute data for the analysis year. Values for speed, volume and percent occupancy are provided in these annual files for all lanes individually. Occupancy relates to the percent of time in which the vehicle detection zone at a sensor is occupied and is given as a number between one and one hundred. Car pool or high occupancy (hov) lanes exist for most, but not all, freeway corridors in the Phoenix region. Development of performance measures requires speed, volume, and occupancy data for general purpose lanes together and for hov lanes. A calculation template has been built to aggregate the by-lane data values to the average for all lanes together, general purpose lanes only, and the hov lane only. The analysis does not distinguish whether or not a hov lane exists, but performs the analysis assuming the last column of the lane data represents traffic for a hov lane. For a four lane location, for example, unprocessed data would be copied into the aggregation template and would result in average values for all four lanes, the first three lanes and the fourth lane alone. Variations of the calculation template are used for the few locations where the configuration of the lane data in the raw data files deviate from this most common order. This aggregation process has been automated with the use of a program.

Nine validity criteria or rules have been built into the aggregation template and applied to individual data elements and rows. Where an individual data element fails to meet the required criteria, the data row is flagged and removed from the output. The criteria used at this level are similar to those cited in the literature review from the analysis for the FHWA Mobility Monitoring Program. The nine criteria are categorized under five different error flags and these error flags are stored and reported in an output file with the processed data. The error flags and validity criteria are defined as follows:

A. Speed error flag
    1. Speed > 85 mph in any lane
    2. Speed < 5 mph but > 0 in any lane

3. $Speed_{n+1} < (0.45 \times speed_n)$ but $speed_{n+1} > 0$

B. Volume error flag (applied to GP, HOV, or All lanes together rather than each individual lane)

4. Volume > 3,000 vph per lane
5. Density > 220 vehicles per mile per lane

C. Occupancy error flag

6. Occupancy > 80% in any lane

D. Difference error flag

7. Speed, volume, or occupancy = 0 where the sum of all is > 0

E. Zero error flag (missing data)

8. all lanes = 0
9. any individual lane = 0 continuously for 20-minutes or more

F. Row errors

Count if one or more of the five error flags exist

G. Rows valid

Count of rows with no error flags

Data processed in the aggregation template is saved to an output file containing all valid data rows with speed, volume and percent occupancy for the hov lane, general purpose lanes, and all lanes together. An error flag on any element of a data row disqualifies all the data for that 5-minute period. For example, a row would be flagged with a speed error where a single lane showed a speed at say 100 mph thereby disqualifying all speed, volume and percent occupancy data for that 5-minute data row. The belief is that a detector malfunction or communication error leading to an erroneous speed value would likely lead to faulty volume and occupancy data as well. In many places this is empirically apparent but not always. Resulting processed data and error flags are saved to an annual result excel file for each given detector.

Data Quality Assessment Tools

The output excel file from the analysis just described is unique in the sense that it contains a quality control panel for reporting detailed information from the microscopic filtering and resulting error flags. Two types of tools have been provided on the quality control panel to be used and stored with the valid data for the given detector and year analyzed. The first is a numerical measure showing the percent of data passing the validity criteria just described. The second is visualization tools used in the macroscopic criteria applied to reject detectors from the aggregation of the data in the calculation of corridor level performance measures.

The percentage of valid data reported here is similar to statistic reported in other programs such as the Mobility Monitoring Program, the Caltrans PeMS and others. The microscopic data validity criteria have been designed to most effectively validate traffic patterns taking place during regular or heavy traffic. At some locations, night time traffic can be light enough that the validity criteria, appropriately utilized for daytime traffic, can be less effective. For this reason, two “percent valid” statistics are produced to represent both the percent valid of data from all times and dates, as well as the percent valid of data where weekends, holidays and all nights are omitted. As all performance measures, with the exception of AADT, are calculated from weekday traffic, the second “percent valid” statistic is believed to be more useful and used here in reporting validity. The result of how validity criteria affect night time traffic differently will be demonstrated and discussed with Figure 4.2 in the next chapter.

Visualization tools to evaluate data quality are not widely used and standard in most applications that process and use traffic data. More commonly, quality assessment visualizations of the traffic are used only in spot checks or during specific evaluations of data quality. In this analysis, data quality charts have been utilized for visualization of all processed data. The output file, resulting from the detector level data processing, has been built to include 10 charts that address and summarize as many of the important characteristics of the data as possible. While additional effort is needed in programing, and the result files containing extensive graphics require more storage space, the benefit provided in seeing the data seems essential. Technological advances in computing virtually demand that this additional effort take place. The 10 visualization tools developed are listed below following which each of the individual charts are shown and described individually.

Chart 1: Annual Hourly Average Speed

Chart 2: Annual Hourly Average Throughput per Lane

Chart 3: Annual Hourly Average Occupancy Percent

Chart 4: Annual Average By-lane Profile of Speed, Volume and Occupancy

Chart 5: Distribution by Annual Date of Data Passing Validity Criteria

Chart 6: Distribution by Weekday of Data Passing Validity Criteria

Chart 7: Count of Annual Quality Control Flags by Hour of the Day

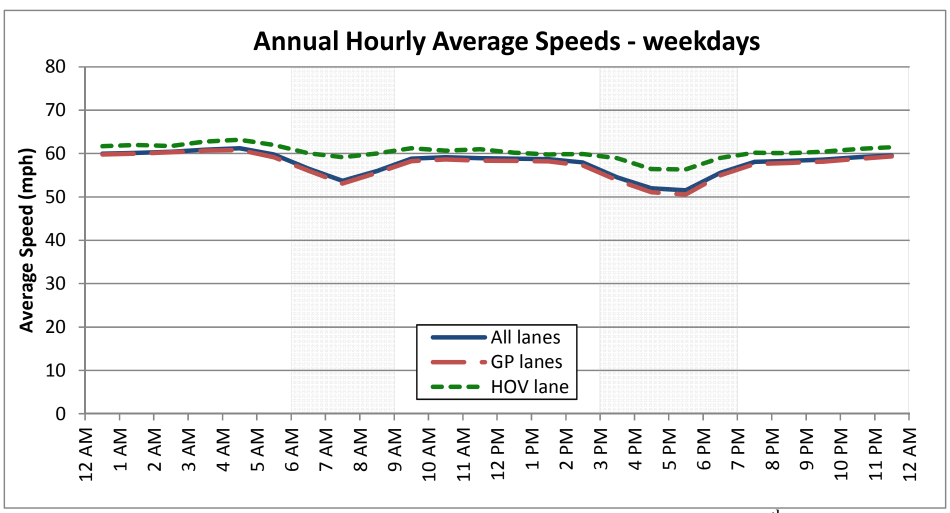
Chart 8: Flow–Density Relationship Data Points

Chart 9: Speed–Density Relationship Data Points

Chart 10: Speed–Flow Relationship Data Points

Charts 1-3 and 7-10 utilize data from the 250 qualified non-holiday weekdays only. The remaining charts, 4-6, include data from all 365 days of the year. In the development of performance measures, speed, volume, and occupancy data are aggregated across lanes resulting in a single average for each 5-minute data row at a location. Charts 1-3 below show the annual averages of these values for general purpose lanes, hov lanes, and the all lanes together. Where hov lanes do not exist, the average from all lanes together should be used while disregarding or deleting the other two results. The calculation program regards both hov and non-hov locations the same calculating all performance measures at each location for all three potentially valid lane aggregations.

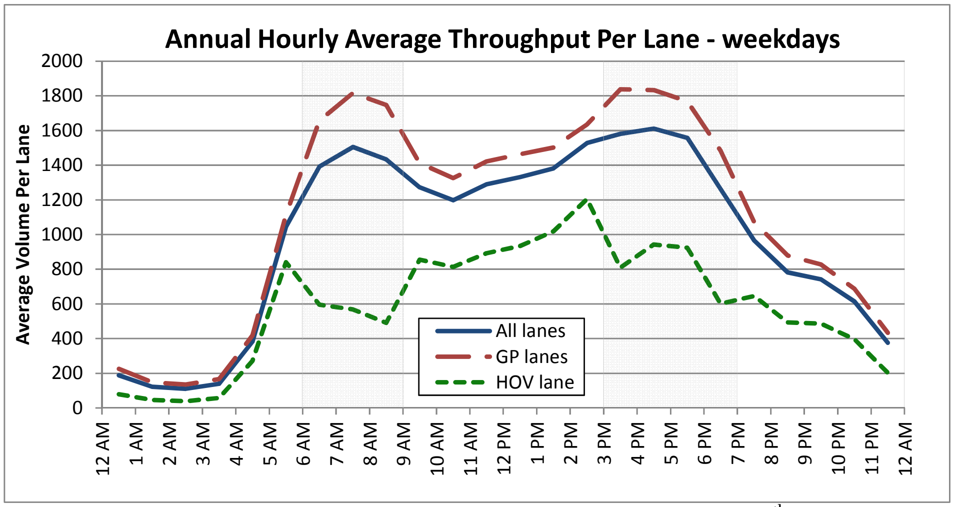
The first chart, Figure 3.1 below, shows the hourly annual average speed during non-holiday weekdays. During peak period congestion, this profile can show a characteristic reduction in speed and a difference in the average hov and general purpose speeds.



(Loop 202 EB at 24th Street, 2009)

Figure 3.1: Annual Hourly Average Speed

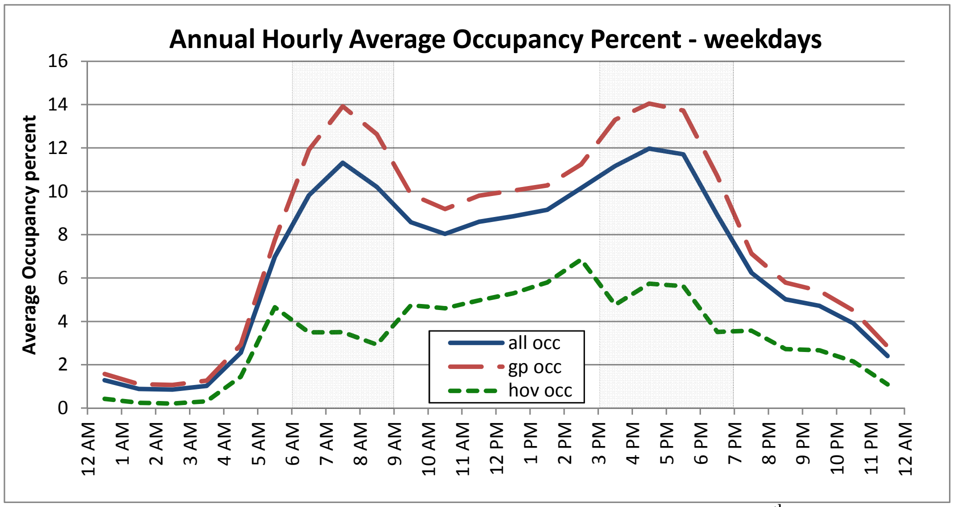
The second chart, shown in Figure 3.2, is the profile for hourly average volume by lane. Like the speed profile, this chart will usually show the characteristic behavior in the hov lane near the peak hours. Errors typically revealed with this chart include suspiciously high or low traffic as well as unreasonable relationships between hov and general purpose lanes.



(Loop 202 EB at 24th Street, 2009)

Figure 3.2: Annual Hourly Average Throughput per lane

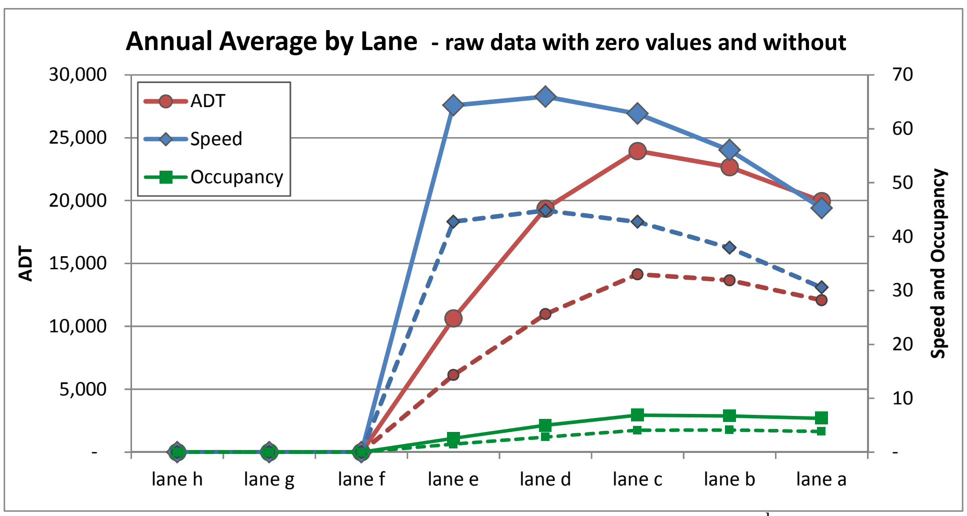
The third chart, shown next in Figure 3.3, is the annual average distribution of detector percent occupancy. The highs and lows on this profile correlate closely to the volume chart in Figure 3.2 during uncongested flow. During severe congestion however, occupancy continues to increase while throughput is reduced.



(Loop 202 EB at 24th Street, 2009)

Figure 3.3: Annual Hourly Average Occupancy Percent

Chart four, shown in Figure 3.4, is the by-lane distribution of raw data values, and serves as a very useful tool in identifying potential calibration issues. A data quality evaluation should look for consistency between adjacent detectors locations, and this study does that as is described later in this chapter. The evaluation can, however, also look for consistency between data in adjacent lanes. Changes in speed, volume, or percent occupancy are expected to transition gradually from one lane to another. Sharp contrast between adjacent lane data can be representative of an individual 5-minute period during a single day, but unreasonable contrasts in the total annual average of all data rows most likely represents systematic error at a detector needing recalibration.

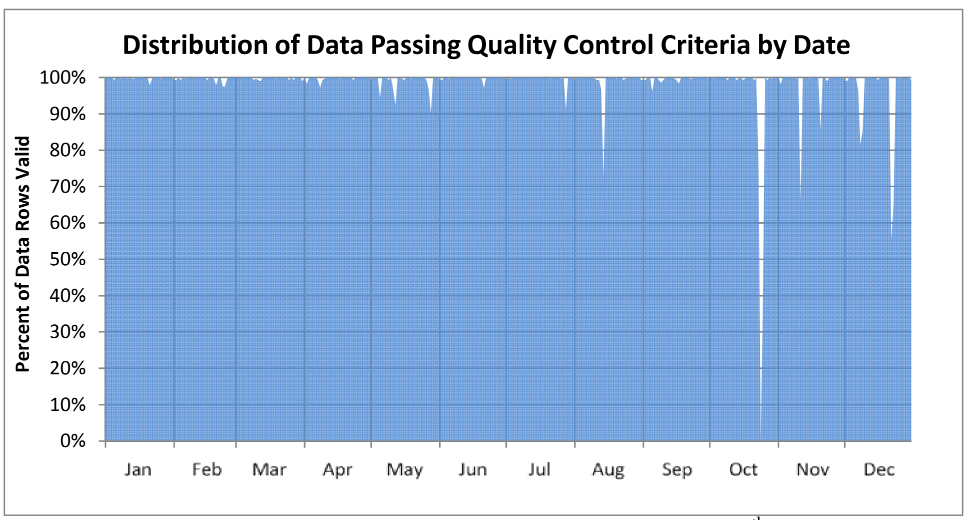


(I-10 EB at 42nd Avenue, 2009)

Figure 3.4: Annual Average By-lane Profile for Speed, Volume and Occupancy

Two profiles are shown in this chart for speed, volume, and occupancy. The dotted profile is a straight average of all raw unfiltered data rows for the entire analysis year. The solid line profile is the same average with all zeros values removed. Zero values exist wherever collection or transmission issues occur and the archiving agencies aggregation scripts identify the data as faulty or missing. Zero values can also represent actual traffic conditions. The differences between the dotted and solid lines provide a rough sense of how much of the raw unprocessed data exist as zeros. The only control that has been applied to the data at this point is the exclusion of out of range high speeds, as many of these exist as extreme outliers several times higher than what would be reasonable. The solid line more closely represents actual traffic but neither should be regarded as such since validity criteria have not been applied. The point of interest in these charts is the relationships between adjacent lanes.

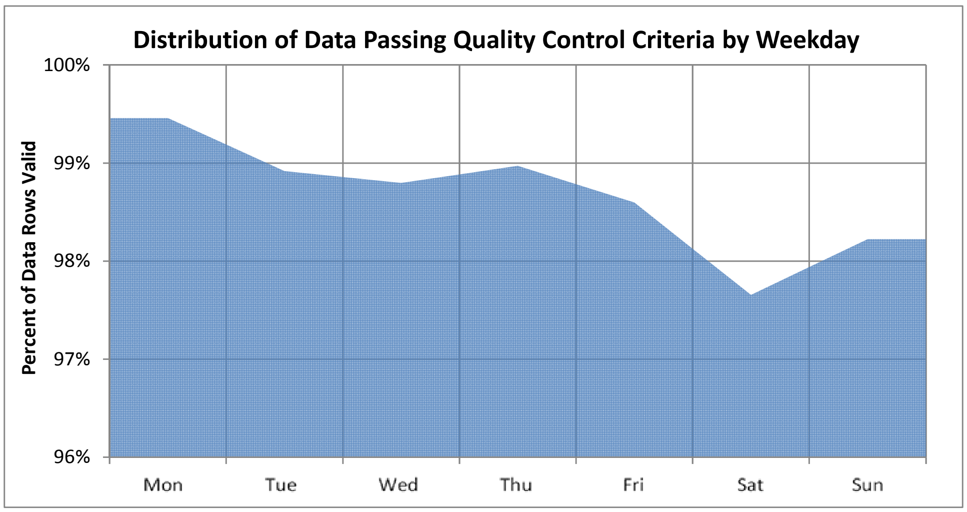
Chart five, shown below in Figure 3.5, is a plot of the percent valid of all data by each day of the year. The most common error reducing the percent of valid data are periodic communication errors leading to blocks of missing data. These blocks can exist for a small portion of one day or for continuous months. This chart provides a clear visual of when and how long such blocks of missing data occur. Where data has been communicated, systematic errors can sometimes be seen distributed across all days or for months at a time.



(I-10 EB at 65th Avenue, 2009)

Figure 3.5: Distribution by Annual Date for Data Passing Validity Criteria

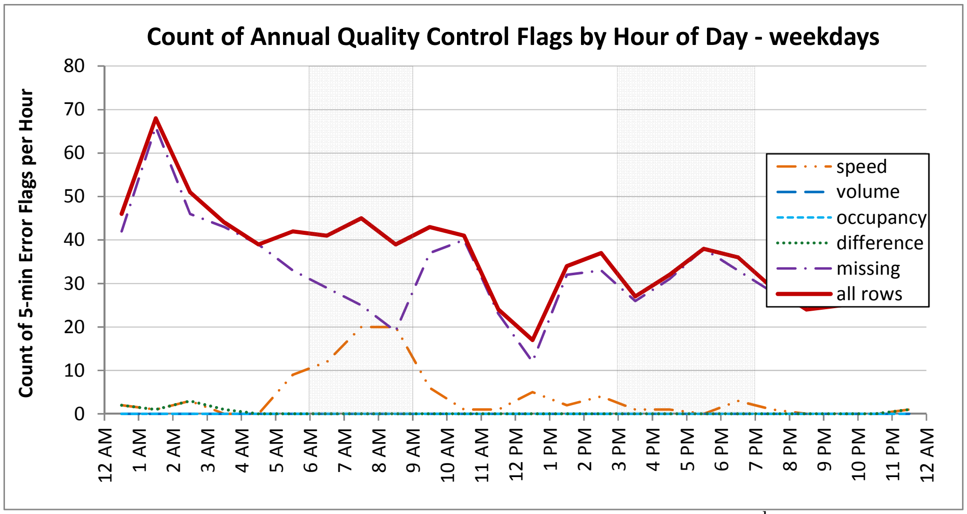
The sixth chart, Figure 3.6, is similar to the annual distribution above but shows the distribution of valid data by day of the week. This chart would reveal any systematic error affecting one day or portion of the week more than any other. A difference in validity of several percent appears typical between various days, but the kind of error potentially revealed by this chart almost never appears to exist. This is probably the least useful visualization described here.



(I-10 Papago WB at I-17, 2009)

Figure 3.6: Distribution by Weekday for Data Passing Validity Criteria

Chart seven, shown in Figure 3.7 reports the hourly count of each of the five error flags collected and total rows flagged annually. Because multiple error flags can be applied to a single data row, the total of “all rows” flagged and filtered out is not necessarily the sum of all error flags. This distribution provides a visual representation of which error flags exist and what part of the days are most affected by errors. The total number of rows annually for each hour is 12 five-minute rows per hour x 250 qualified weekdays = 3,000 rows. The hourly error count can, therefore, also be seen as a percent of the total number of rows.



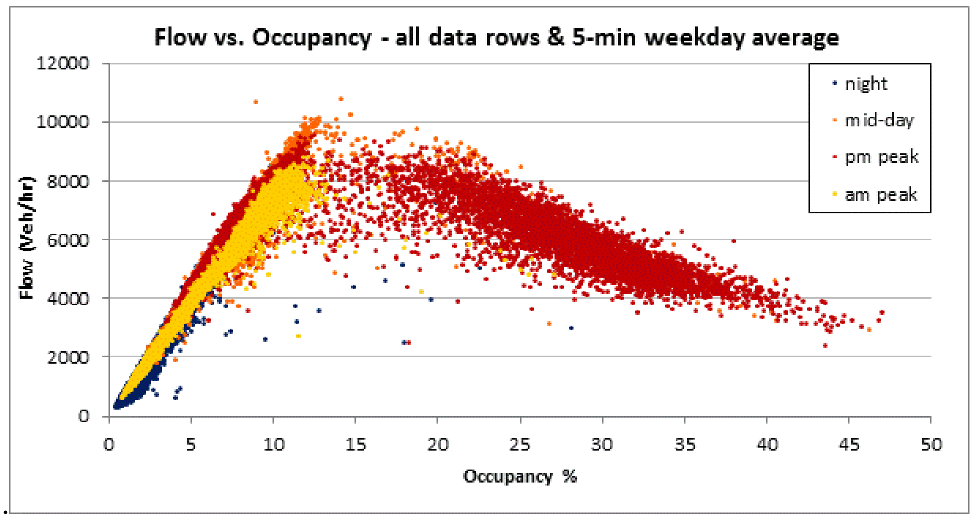
(I-10 EB at 43rd Avenue, 2009)

Figure 3.7: Count of Annual Quality Control Flags by Hour of the Day

The final three charts are a scatter plot for every 5-minute data row showing the important relationships between speed, flow and density during both congestion traffic patterns and free flow uncongested traffic. Percent occupancy is again defined as the percent of time in which a traffic sensor detection zone is occupied, and correlates directly to the density of traffic. These charts represent data passing microscopic validity criteria only, and hov and general purpose lanes have been grouped together where a single average value is plotted for all lanes. The color of the dots distinguish between a.m., mid-day, p.m. and night times allowing the deviations from free flow traffic patterns to be identified by the time period in which they occur.

The scatter plots below are helpful in a number of ways. As error to data values are often not catastrophic, meaning they do not necessarily fall beyond maximum or range levels defined by the validity criteria, the scatter charts demonstrated more clearly the actual condition of the data and often show highly unusual patterns of at detectors where the data technically passes the microscopic validity checks. Insight is also gained in seeing how much and during what time of the day congested traffic occurs, as well as in estimating traffic flow parameters such as flow capacity, optimum speed, free flow speed, optimum density, and jam density.

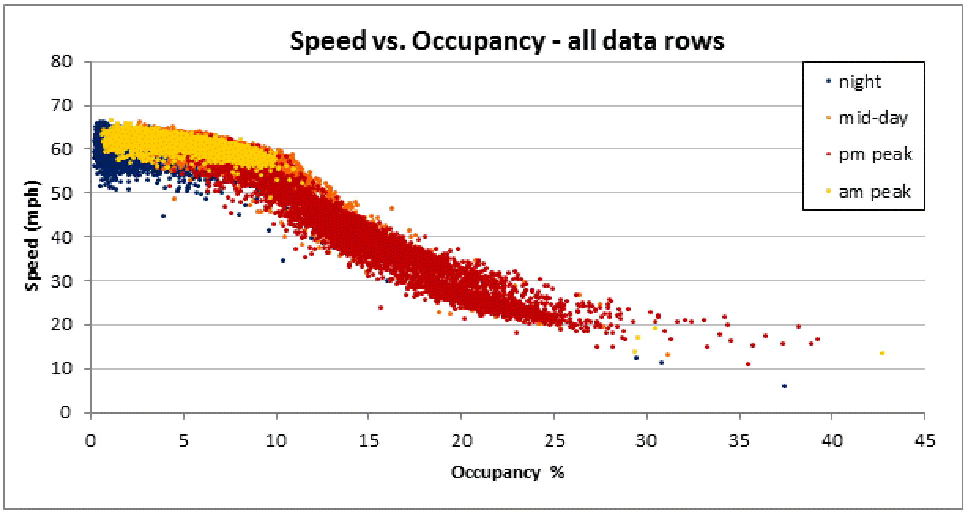
Chart eight, in Figure 3.8 below, shows the relationship between total flow and percent occupancy or density. Uncongested traffic is characterized by a linear relationship between flow and density but where volume approaches and exceeds free flow capacity, congestion ensues, reducing throughput flow while substantially increasing density.



(I-10 EB at 48th Street, 2009)

Figure 3.8: Flow–Density Relationship Data Points

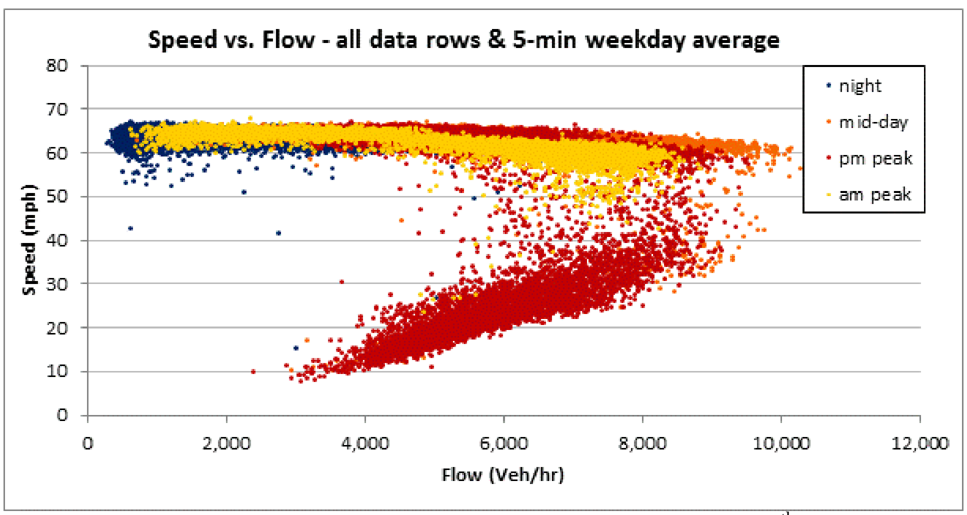
Chart nine, in Figure 3.9, shows the relationship between speed and percent occupancy or density. Speeds are relatively steady during free flow conditions. When volume approaches and exceeds capacity speeds reduce significantly while density further increases.



(Loop 101 SB at Broadway Rd, 2009)

Figure 3.9: Speed–Density Relationship Data Points

The final visualization, chart 10 in Figure 3.10 below, shows the distribution of speed versus flow. Speeds are relatively stable and constant until volume reaches a critical point where flow is reduced in congested traffic.



(I-10 EB at 48th Street, 2009)

Figure 3.10: Speed–Flow Relationship Data Points

Macroscopic Quality Control

In order to obtain corridor level statistics measuring freeway performance, results at individual detector locations are expanded to represent a link portion of the corridor, and multiple link level results are then combined forming the corridor results. The geographic length of each link is calculated as half the distance to the nearest valid detector location, both upstream and downstream. Corridor lengths are most often about one mile, but can be much longer when adjacent detectors are rejected from the corridor level aggregations.

The evaluation process for rejecting detector locations is quite different from the microscopic validity criteria applied in detector level quality control. The microscopic validity criteria look at individual data elements and rows eliminating major deviations from expected values only. The corridor level macroscopic criteria look at the data at a detector as a whole, from the perspective that a detector is either functioning correctly or not. The macroscopic criteria are also not applied as pass/fail checks in an automated analysis. The levels of quality in data on different corridors vary greatly making the use of an absolute standard ineffective. The goal is in identifying and using the best detector locations available while maintaining a sufficient number of locations to calculate the desired corridor level performance measures. The criteria, or decision rules, serve as guidance in an individual assessment of each detector location on each corridor. Guidance on rejecting detector locations fit into the following three rules:

1. *Insufficient Data* – If a significant portion of the data has been filtered out in the microscopic validity checks, the detector location is rejected. Usually 50% or more of the data is considered significant. Consideration is given for whether or not missing or rejected data is distributed throughout the year or encompasses a portion of the

year continuously. Detectors with less than 50% valid data are utilized along corridors where missing data is extensive, such as the Loop 202 Red Mountain freeway in 2009 where construction activities affected data for nearly all detector locations for most of that year.

2. *Unreasonable Values* – Detector locations are rejected where speed, volume or occupancy trend lines appear unreasonably high or low, as demonstrated from the first three visualization charts for each detector location. It is believed that very inaccurate data can exist where significant portions of the data still fall within valid ranges. Visualization chart seven, showing the annual hourly count for each of the five error flags, is utilized to correlate speed, volume, or occupancy errors by time of day with high and low points in the profiles from the first three charts. The speed, volume, and occupancy (or density) relationships, shown in charts eight through ten, also add insight into when a detector may be have too many unreasonable values, usually characterized by unusual dispersion in the scatter plots. In some cases a clear boundary will be visible beyond which the data is filtered out by the microscopic validity filters.
3. *Spatial Comparison* – Detector locations are rejected where spatial comparisons with adjacent detector locations show unreasonable trends. The first two visualization charts showing the speed and volume profiles are principally used. Additionally, chart 4, showing the by-lane distribution, and charts 8, 9, and 10 showing the scatter plot relationships between speed, volume and density, are used extensively in comparing detectors side by side.

Data Quality Measure Targets and Calculations

Beyond the process of identifying and selecting valid detector data and locations, the data used in the analysis for calculating performance measures is evaluated and described using the six data quality measures presented in the research. The research cited has recommended that goals and specific target values for data quality be established at the jurisdiction or program level based on a clear understanding of intended uses. A starting place for target values on each of these measures have however been suggested for various transportation applications in the literature and specifically in the FHWA guidance document on *Measuring Traffic Data Quality* (Battelle, 2004). The analysis contained here has, where possible, defined target values for each of the six measures, in most cases corresponding to the suggested values. In Table 3.1 is a summary

Table 3.1: Initial Targets for Six Data Quality Measures

Data Quality Measure		Target Values
Accuracy	speed	5% MAPE
	counts	10% MAPE
Completeness		< 20% missing data
Validity	from all detectors	85% valid
	from valid detectors only	90% valid
Timeliness	accuracy	< 5-minute lag
	availability	3 years of data
Coverage	density	< 1.5 mile spacing
	extent	60% of facility
Accessibility	time required	unspecified
	qualitative-features	

of preliminary data quality targets. Several of the quality measures can have more than one meaning, or in other words, measure two very different things. This section describes each of the six measures more closely and provides the calculation procedures or guidance for determining whether or not each measure target has been met.

Accuracy is "the measure of or degree of agreement between a data value or set of values and a source assumed to be correct." Accuracy can be measured as the mean absolute percent error (MAPE), as is used in the MAG accuracy evaluation of FMS detectors described in Chapter two. Signed percent error and root mean squared error are also measures of accuracy, but for simplicity only one measure will be discussed or used in this exercise.

$$Mean\ Absolute\ Percent\ Error\ (\%) = \left(\frac{1}{n}\right) \times \left(\sum_{i=1}^{n} \left|\frac{x_i - x_{reference}}{x_{reference}}\right|\right)$$

where: x_i = the observed data

$x_{reference}$ = the reference value (usually ground truth)

n = the total number of observed data values

An appropriate target for accuracy is suggested as less than 5% error for corridor level speeds and less than 5-10% error for average annual daily traffic volume. Implied in this is the possibility of having individual detectors and values exceeding the targeted error rate while corridor level annual average results fall within the desired targets. Appropriate accuracy targets for individual locations and data samples depend on whether the error is systematic or random. Where accuracy error is random, spatial and temporal aggregation tends to normalize and reduce the error. For this exercise, accuracy targets will be established at an MAPE of less than 5% for speed and less than 10% for volume for each detector individually.

Completeness is defined as "the degree to which data values are present in the attributes that require them." That is, the degree to which data is not missing, which can include data flagged as invalid for one or more of the microscopic validity criteria. Completeness, which is a measure of temporal availability of the data, should not be confused with coverage, which is a spatial measure of completeness, and will be discussed below. Completeness can be calculated as a percentage of all potential data rows as shown in the following equation.

$$Percent\ Complete\ (\%) = \frac{n_{available\ values}}{n_{total\ expected}} \times 100$$

where: $n_{available\ values}$ = the number of rows with all available values present

$n_{total\ expected}$ = the total number of records or rows expected

Note that in the filtering criteria described, the "zeroes error flag" is attached to a row where either speed, volume or occupancy information is missing for any lane. Percent complete in this analysis is therefore a measure of the percentage of rows that have all three traffic elements defined for all lanes of traffic. Common are situations where malfunctions affect some but not all data. An example would be missing speed data but not volume or occupancy, or missing data in one lane but not the remaining lanes.

An appropriate target for completeness is suggested at 80% of the data. The significance of missing data varies, however, depending on whether or not it is missing as a continuous period of time or in smaller intervals throughout the year.

Validity is defined as "the degree to which data values satisfy acceptance requirements of the validation criteria, or fall within the respective domain of acceptable values.” Validity can be calculated as a percentage of all potential data rows as shown in the following equation.

$$Percent\ Valid\ (\%) = \frac{n_{valid}}{n_{total}} \times 100$$

where: n_{valid} = the number of records or rows with values meeting validity critera

n_{total} = total number of records or rows subjected to validity criteria

Technically, missing data, just discussed under *completeness*, is considered *invalid* and has been defined by one of the five error flags described in the microscopic validity criteria earlier in this chapter. Evidence, in both this research and other, suggest that missing data will compose the majority of what gets filtered out as invalid. For the purpose of reporting the quality of the data validity will now be seen only as the data that passes the microscopic validity criteria described under the other three error flags of speed errors, volume errors, occupancy errors. Detector locations where large portions of the data are rejected as invalid are often the same detectors that get rejected completely in the corridor level macroscopic quality control. Validity can therefore be seen both as a percentage of data from all detectors and the percentage of data only from detectors that are at least partially valid. Validity from detectors being used in the actual corridor level aggregation, where the worst detectors are taken out, will be somewhat higher than the validity from all detectors. Guidance suggests that targeted validity should be at about 90% of the data. In this assessment targeted validity is set at 85% from all detectors and 90% from at least partially valid detectors only.

Timeliness is defined as “the degree to which data values or a set of values are provided at the time required or specified.” Two considerations are contained in this. First, *specified* refers to how closely the data matches the indicated time of collection.

Timeliness in this sense is most relevant to real-time applications but a significant lag between the indicated time of collection and the actual time could also compromise the analysis for annual performance measures. Second, *required* refers to how timely the required data is made available to potential users. Guidance suggests that a potential lag between reported and real time of collection within 5-minutes would more than satisfy the requirements for annual performance measurement. Regarding the availability of the data, the performance measurement program will require data to be available for at least the most recent three years.

Coverage is defined as “the degree to which data values in a sample accurately represents the whole of that which is to be measured.” Continuous data collection potentially provides a complete sample at a given location, but the geographic sample across all locations is not so continuous. In this way, *completeness*, discussed previously, is a temporally measure, while *coverage* is spatial.

In this analysis, coverage is regarded both in terms of the extent of the detector instrumentation on the study area as well as the spacing of valid detector locations on instrumented segments. In the analysis for corridor level performance measures, detectors at each point location are expended to represent a link with length equal to half the distance to the next adjacent detector on either side. Macroscopic quality control performed in this analysis at the corridor level includes the rejection of entire detector locations where too much the data is missing, unreasonable, or too inconsistent with spatial trends along a corridor. With the rejection of a detector location the zone of influence, or link length, for adjacent detectors on either side are expanded, sometimes to two or three times the average segment length. Rejection of too many detectors can lead to conditions where remaining “valid” detectors insufficiently represent the corridor as a whole. The extent of coverage, therefore, includes freeway corridors that have been instrumented with traffic sensors and is expressed as the percent of the total roadway, while coverage spacing is expressed as the distance between valid detectors.

Guidance suggests that an extent of coverage of 55-60% of the freeway would be sufficient for highway performance monitoring traffic volumes. For congestion management, 100% of the study area is needed. For the purpose of this assessment the target values for the coverage will be 60% of the facility and spacing of less than 1.5 miles between detectors.

Accessibility is defined as “the relative ease with which data can be retrieved and manipulated by data consumers to meet their needs.” Accessibility is measured both by the time required to retrieve and manipulate the data as well as the qualitative descriptions of features like user-friendliness of interface or options for online queries. The quality of annual performance measurements is not necessarily sensitive to poor accessibility of data. Practically speaking, however, an analysis process relying on difficult to use data will invariably cost in terms of resource availability. Accessibility of the data before and after the analysis that took place her will be discussed in the next chapter on results. Targets for accessibility have not, however, been provided in the guidance literature.

CHAPTER 4

RESULTS

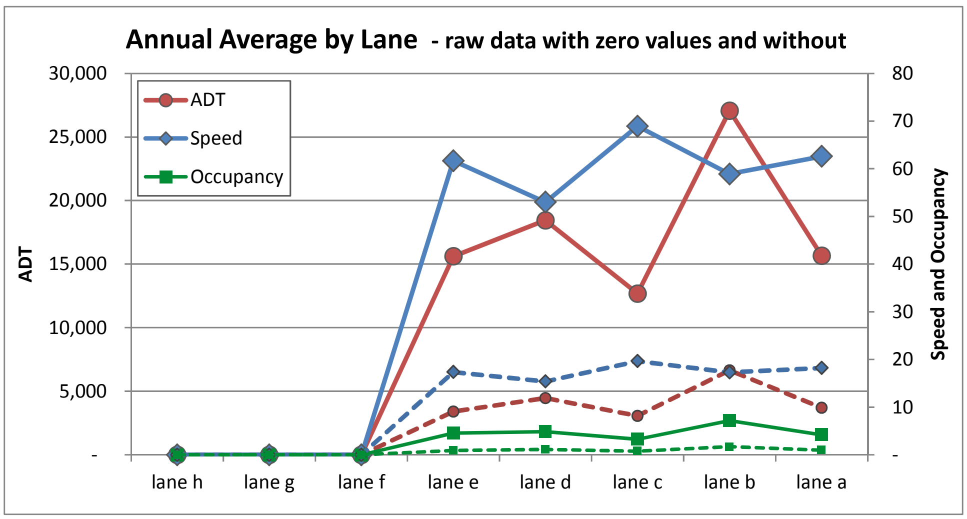
The research discussed in chapter two described six traffic data quality characteristics or measures contributing to the overall quality of the data. These are *accuracy*, *completeness*, *validity*, *timeliness*, *coverage*, and *accessibility*. Further description and calculation methodology for each of these measures was discussed in chapter three. In this chapter, the detector data used in the analysis of regional performance measures are described for each of these six characteristic or measures of quality and, where possible, an actual score is given. A summary of the overall data quality assessment is provided, and an attempt is made to assign a percentage to each of the six data quality measures based on how much they contribute to the overall quality of the data. Some of the assessments made are specific to the analysis taking place here for regional freeway performance measures, and would likely differ in an assessment made for other uses. More meaningful than the specific results obtained here is the experience provided in this exercise in defining and reporting traffic data quality. This chapter concludes by summarizing the extent to which recommendations from the literature listed in chapter two have been applied in this assessment.

Accuracy

This investigation of data quality provides significant insight into the nature and characteristics of the data, but without the ability to make direct comparison with authoritative ground truth data, a defensible assessment of accuracy is not possible. Where this authoritative assessment cannot be made, information is looked for that at least suggests the possibility of significant accuracy or error. Guidance, both from evaluations conducted in previous years and the data quality assessment conducted here through the visualization tools used can inform some educated decisions regarding current accuracy levels.

Two evaluations of ADOT FMS detectors have been cited in chapter two, the first of which compared passive acoustic detectors (PADs) to loop detectors, and the second which compared both detector types to a ground truth benchmark. In the first evaluation, Jonas (2001) demonstrated that passive acoustic detectors can be calibrated to within an error threshold to values reported by loop detectors. The evaluation conducted by TTI (Turner, 2008) similarly showed the potential for calibration of passive acoustic detectors. Both evaluations however showed significant and consistent error with PADs where sufficient calibration efforts had not taken place. The second evaluation also showed error rates in excess of 5% at most of the loop detectors evaluated. Additional research cited in chapter two has suggested ADOT maintenance and calibration efforts to be under funded. There may be, therefore, little historic reason to assume current accuracy levels would not be consistent with the 2005, and 2006 evaluation results recorded by TTI in the second evaluation cited.

Two tools have been provided in this assessment, independent of previous evaluation studies, allowing some inference to be made regarding the accuracy of the FMS detector data. The first is the by-lane profile of annual average speed, volume and occupancy shown as the fourth visualization chart described in the previous chapter. Where a detector has been properly calibrated, a smooth transition of annual average values between adjacent lanes will be visible, and the values themselves will appear reasonable. However, at a large portion (the majority) of detector locations, the by-lane profiles show an irregular transition in values between adjacent lanes and in some cases very unreasonable values. These by-lane profiles serve to flag poorly calibrated detectors, and poor calibration is believed to directly correlate to poor accuracy. In the following figure a detector that is likely to be poorly calibrated is shown. Figure 3.4, in chapter three is an example of by-lane profiles where transitions between speeds, counts, and occupancy in the respective lanes are smooth. Appendix A shows more example charts with sections devoted both to “normal” detectors and “poorly calibrated” detectors. Note that the irregular distribution such as that shown here in Figure 4.1 is much more common with passive acoustic detectors than it is with loop detectors. The figure here however is actually of a loop detectors location, which can also be very poorly calibrated.



(Loop 101 SB approaching Southern Avenue, 2009)

Figure 4.1: By-lane Profile at a “poorly calibrated” Detector

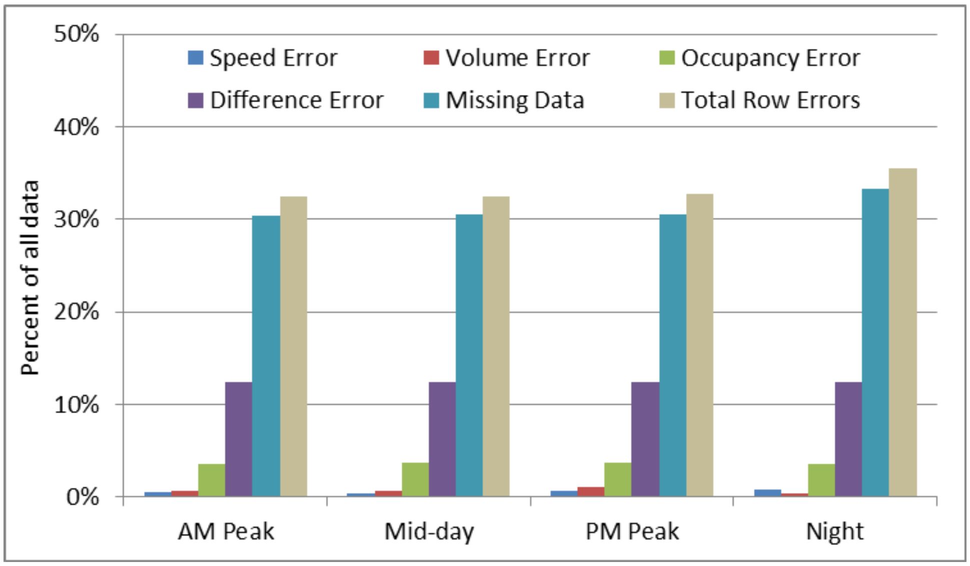
The second tool used in identifying potential accuracy issues is corridor level comparisons between multiple adjacent detectors. Several individual detectors have, for example, been rejected as faulty and omitted from the calculation of corridor level performance measures based solely on the inconsistency of charts in comparison with other adjacent detectors. Appendix B demonstrates one such example along the I-17 corridor where detector 364 has been rejected. The comparisons that take place heavily utilized chart types 1 and 2 showing the average speed and throughput profile, specifically noting the relationship between HOV and GP lane traffic patterns. Consistencies are also looked for in the by-lane profile shown in chart type 4, as well as consistencies in chart types 7, 8, 9, and 10. The comparisons between adjacent detectors allow inference both in determining which detectors are most likely to contain significant error and whether that error is likely to be in the negative or positive directions.

In the macroscopic or corridor level data quality analysis, described in the previous chapter, roughly 25 of the 228 detector locations have been rejected largely due to either excessive concern regarding the by-lane profiles or significant discrepancies with adjacent detectors. These rejected detectors represent only the most severe concerns where removal of the concerning detector allowed sufficient remaining data for calculation of performance measures. These guidelines used in determining what should be allowed as accurate are therefore applied slightly different to each individual corridor. The goal has been in obtaining the most representative results possible and in providing some understanding of the inherent accuracy.

Completeness

Completeness has been calculated as the percentage of all potential data rows for all detectors where values have been provided. Data at all 228 detector locations during 2009 have shown a total of 32% of all 5-minute data rows where speed, volume or occupancy is missing on one or more lanes of traffic. In some cases, missing data exists affecting only one of the three data elements and often this occurs in one or two lanes only. Figure 4.2 on the next page shows the distribution of all five error flags defined in the last chapter. The “difference error” flag exists wherever zero values are being reported for speed, volume, or occupancy but at least one of the three variables have values. Difference errors have been shown to exist in more than 12% of all data rows.

Figure 4.2 demonstrates clearly that more than one third of the 32% of rows with missing values (12% of all rows) exist where speed, volume, or occupancy are not missing. This gives some indication of some data that might be gained in defining the rejection criteria to apply to speed, volume or occupancy independently rather than together. The other error flags shown in Figure 4.2 are discussed under validity in the next section.



(all detectors, 2009)

Figure 4.2: Distribution of Error Flags by Peak Period

All but two malfunctioning detectors have provided at least some data for speed, volume or occupancy. For detectors that were at least partially functional, completeness is 68%. The quality target of 80% completeness has therefore not been met.

Validity

Detector level validity consists of the rows that are not missing where all elements of the data pass the validity criteria described in the microscopic quality control defining the speed, volume and occupancy error flags. Technically this does not included data filtered out of the analysis as missing. Figure 4.2 demonstrates that 32% of all rows filtered out are missing data. It's important to note with Figure 4.2 that the error flags are not necessarily additive, as multiple error flags can potentially be applied to the same individual rows. Validity at the detector level has been calculated at more than 95% of all data rows.

In addition to detector level checks applied at individual data rows, are the macroscopic validity checks applied at the corridor level. These macroscopic validity checks have led to the rejection of additional detector data passing the initial microscopic validity criteria. At the detector level 36% of all data rows have been filtered out as missing or otherwise invalid. At the corridor level, 44 detectors have been rejected, where at least some data passing the microscopic validity criteria existed. Table 4.1 below is a matrix summarizing validity. The columns represent the percentage of valid data before and after corridor level validity checks where the worst detectors are rejected completely. The rows represent validity strictly as the percent of available data or data that is not missing, and the percent of all data that could be valid. Seen alongside Figure 4.2, the 95% in the first cell of the matrix represents the data not filtered out with a speed, volume, or occupancy error. 64% represents the data not filtered out with any of the microscopic validity checks, including difference and missing data errors. The second column results are calculated after macroscopic validity checks, and the values are higher because they include only those detectors that have not been rejected completely, which are consequently the better detectors. Independent of completeness, the validity of the data is higher than the target value of 90%.

Table 4.1: Percent of Data Valid

	all detectors	valid detectors only
from available data	95%	96.5%
from all data	64%	74%

Timeliness

Traffic data can be timely in two very different ways. First is when the time stamp on the data is accurate as specified, and second is when the data is available for the time periods for which it is needed. The MAG and TTI evaluation discussed in chapter two analyzed video footage to measure the accuracy of select FMS detectors. In preparation, a verification of FMS data latency took place to establish any difference between the time stamp in the FMS data and the NIST internet time service. The results indicated the latency was negligible (Turner, 2008). In June of 2010 the ADOT FMS communications system crashed, halting the preliminary data processing and archiving that takes place at the traffic operations center. Rather than restoring the previous system, ADOT has utilized this as an opportunity to transition to a new data management system. While traffic data continues to be collected, the internal system for processing and making available this data has been out of commission now for more than ten months. The performance measure analysis being conducted now, in the early portion of 2011, includes only 2008 and 2009 data, while the analysis of 2010 data will need to take place at a later date. To some degree, the value of the analysis conducted now is reduced with the exclusion of this most recent year. The requirement for the recent three years of available data has not been met. Timeliness has therefore been met as it relates to the accuracy of the time signature in the data but not in relation to the availability of all recent data.

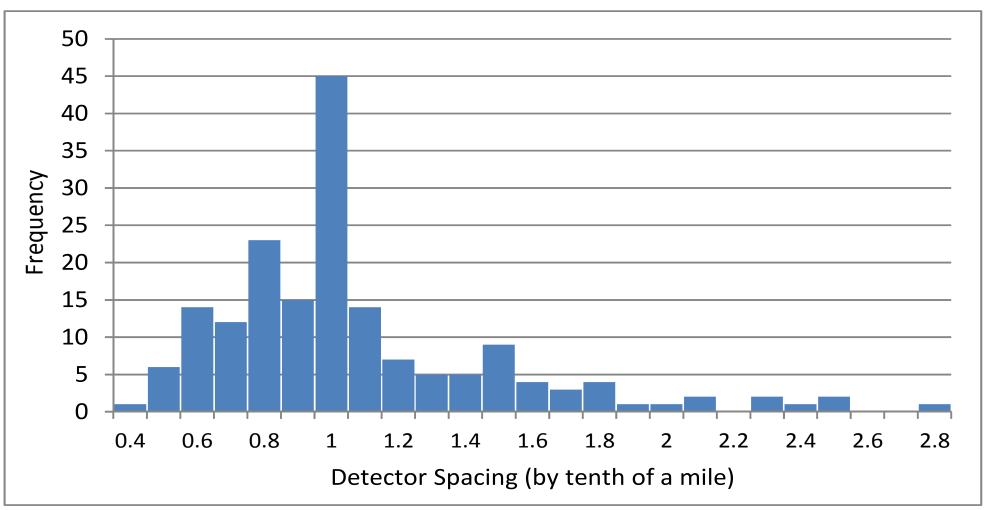
Coverage

The Phoenix metropolitan contains roughly 235 miles of freeway, 127 miles of which have currently been instrumented with permanent count traffic detectors. By the beginning of 2009, 238 traffic detectors had been installed representing 100 miles of freeway. The extent of coverage for the analysis conducted in 2009 is therefore 42% of the regional freeway. Figure 1 in the introduction chapter, shows a map where the extent and spacing of all active detectors can be seen. Table 4.2 below, summarizes the coverage for the past three years.

Table 4.2: Coverage of Regional Freeway during Past Three Years

coverage year	detector count	covered miles	percent coverage
2008	199	87	37%
2009	238	100	42%
2010	286	127	54%

Figure 4.4 below shows the frequency distribution of the detector spacing for 182 valid detector locations after 46 detectors have been rejected completely from the analysis. The median spacing is one mile and the maximum spacing is at 2.75 miles. A total of 21 detectors have a link level zone of influence larger than the goal of 1.5 miles for spacing, 9 detectors at which the link length exceeds 2 miles. The coverage target has therefore not been met either in terms of the extent of coverage or coverage spacing.



(all valid detectors, 2009)

Figure 4.3: Frequency Distribution for Detector Spacing

Accessibility

FMS detector data can be accessed and downloaded via the internet from the ADOT file transfer protocol site at ftp.az511.com. Data is available in various aggregation levels from 5-minute to 24-hour. The 5-minute data used in this analysis provides detailed lane-by-lane values for speed, volume and occupancy. Compressed files around 12 megabytes in size are provided for each annual day containing 5-minute data for all FMS detector locations, half of which are inactive detectors decommissioned in 2005. Extracted, these files are about 110 megabytes in size. Before beginning actual analysis work, data from the 365 daily files for each year have been resorted to create a single text file for each detector for the entire year. The annual data files for each location are then analyzed using the procedure briefly described in chapter three.

Downloading the data for 365 days took place in bursts over a day or so. The sorting of data into individual detector files for the analysis year took place using a simple script in VBA Excel. Other software exists that no doubt could perform this task more quickly and with less computer memory. The process with VBA was however simple enough, but certainly not quick at about 30 minutes for each detector location. The sorting program was for the most part left to run unattended, usually at night, and after a week or two all 228 detector raw data files were created for 2009. A similar amount of time was used to develop 2008 raw data detector files. The processing of individual files to obtain the detector level performance measure results and detector level quality assessment reports also took place using VBA Excel, requiring about 10 minutes for each detector. Most of these were also run unattended at night but some monitoring, checking and rerunning was necessary. The output files created are large in and of themselves, but, in addition to the quality control panel and the performance measure summaries, they serve as a database for future use with all 5-minute data by lane type (general purpose, hov, or all lanes). Data processing and calculations from the detector result files is relatively smooth.

Relative inexperience in programing data processing and analysis scripts have no doubt contributed to the total length of time needed in the analysis of the data for this project. Run time can also vary a lot depending on computer processor speed. It may therefore be unrepresentative to report the total length of time devoted to working with the data, but the short story is that months were devoted. Initially efforts targeted performance measure calculations, and then increasing levels of effort were put into data quality assessment and control as challenges in the analysis due to poor quality data were identified. The time required to access and manipulate the data has been extensive. The second element of accessibility is in the qualitative features such as user-friendliness of interface. No qualitative features were used in this analysis. The overall accessibility of the data is described as difficult, and time consuming.

Summary of Findings

Table 4.3 provides a summary of the data quality assessment results for each of the six measures discussed in the previous sections. The data quality targets have been met in some but not all or even most of measures calculated. The next chapter includes some discussion regarding the merit of the data quality targets as they have been defined. The targets shown in this table are not authoritative criteria defining whether or not data is acceptable for this particular application, but serve only as a reference point in evaluating the quality.

Table 4.3: Result Summary for Data Quality Measures

Data Quality Measure		Target Values	Result	Target Met?
Accuracy	speed	5% MAPE	current reference data not available	suspect not
	counts	10% MAPE		
Completeness		80% or < 20% missing data	32% missing	no
Validity	from all detectors	85% valid	95%	yes
	from valid detectors	90% valid	96.5%	
Timeliness	accuracy	< 5-minute lag	negligible lag	yes
	availability	3 years of data	2010 not available	no
Coverage	density	< 1.5 mile spacing	82% of detectors < 1.5 mile	partially
	extent	60% of facility	42%	no
Accessibility	time required	unspecified	extensive	-
	qualitative-features		none used	

An important consideration is in the fact that various data quality attributes will have different value in an overall assessment of quality. For example, traffic sensor coverage on only 70% of a network may be far more acceptable than an accuracy level of only 70%. Figure 4.5 shows a pie chart of different data quality attributes and the relative weight they might have in assessing the overall quality of the data for performance measurement. The distribution in the pie chart is somewhat subjective and would again vary depending on the intended application of the data. For regional freeway performance measurement and reporting however, the analysis here might consider the accuracy and validity of the data to be most significant in influencing the usefulness of the data. Completeness, coverage, and accessibility make valuable contributions, and timeliness probably has the least influence. The importance of these measures also varies depending on the quality level of the individual attribute. Coverage, for example, matters a lot more where network coverage is limited, say less than 20%. At that point, resources might best be allocated to adding detector locations even where accuracy levels are less than what might be standardly acceptable. Where the network coverage exceeds 50% or so, less is gained in adding detectors, especially where accuracy levels exist below about 10%. The pie chart shown here might also be thought of as the relative distribution of where the overall data quality can be improved.

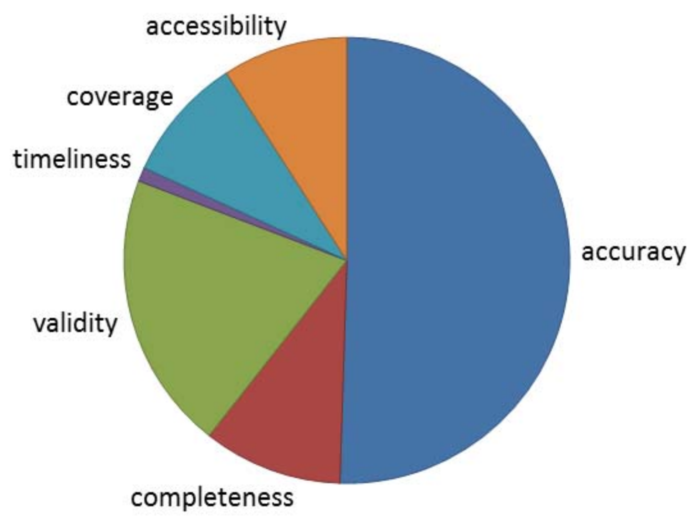


Figure 4.4: Pie Chart of the Relative Weight of Measures on Data Quality

Some discussion has taken place within the literature regarding a “composite data quality score” representing two or more data quality attributes (Battelle 2004, IDNC Newsletter 2004). A single number, or score, might be calculated as the average of several individual data attribute scores, or an average weighted by the relative importance the individual attribute or measure.

Application of Recommendations

Beyond the potential usefulness of describing this data set, has been the experience gained in the attempted application of recommendations from the literature reviewed. The experience of applying the recommended framework, and the use of the six data quality measures, is in fact the principle result from this assessment. The following is the list of recommendations that concluded chapter two along with a description of the extent to which the recommendations have been applied in the analysis taking place in this project.

1. Implementation of at least basic foundational data validity criteria in the automated processing of archived detector data.
   - This has been applied in the application of microscopic, detector level, validity criteria described in Chapter 3 of this report.
2. Development of additional validity criteria relevant to the data and intended purpose for the data, and the use of visual reviews when feasible.
   - This has been applied in the macroscopic, corridor level, validity checks and in the use of the ten visualization tools.
3. Use of consistent standards and guidelines for calculating traffic data quality measures.
   - This has been applied in the use of the six data quality measures presented in the literature as standard traffic data quality measures. Some liberties have been taken in extending the definitions of some measures as seemed appropriate to this study.
4. Metadata to document quality control procedures and results along with relevant information about the data.

- This has been applied with the data quality panel, created in the processing of the archived data, which contains with the visualization tools, some information about the data, the link to the original data source, and the link to metadata files available with the original data.

5. Data sharing efforts that avoid duplication in data collection and underutilization of data. Primary in this is the need for a centralized clearinghouse where information about data sources and characteristics of the data are recorded.
   - This can partially be applied in the utilization of the data files created here for other applications and tasks. The inclusion of the data quality panel, with visualization charts to describe the data, make it possible for future users to have a more appropriate level of confidence in using the processed data.
6. Sensitivity studies that demonstrate the value of data and highlight the effects of poor quality data on various applications. These studies allow for refinement of targets for data accuracy and other quality measures.
   - This is applied to a degree with the assessment that has taken place here and can be applied in future work building upon this assessment. Insight for further refinement of data quality goals and targets are provided from this research.

Results in this chapter have included measurement and description of the FMS detector data using each of six data quality measures. The extent to which the data quality targets have been met for each of measures has also been described. The more significant aspect of this assessment has been the resulting experienced gained in applying data quality recommendations suggested in the literature as described.

CHAPTER 5

DISCUSSION

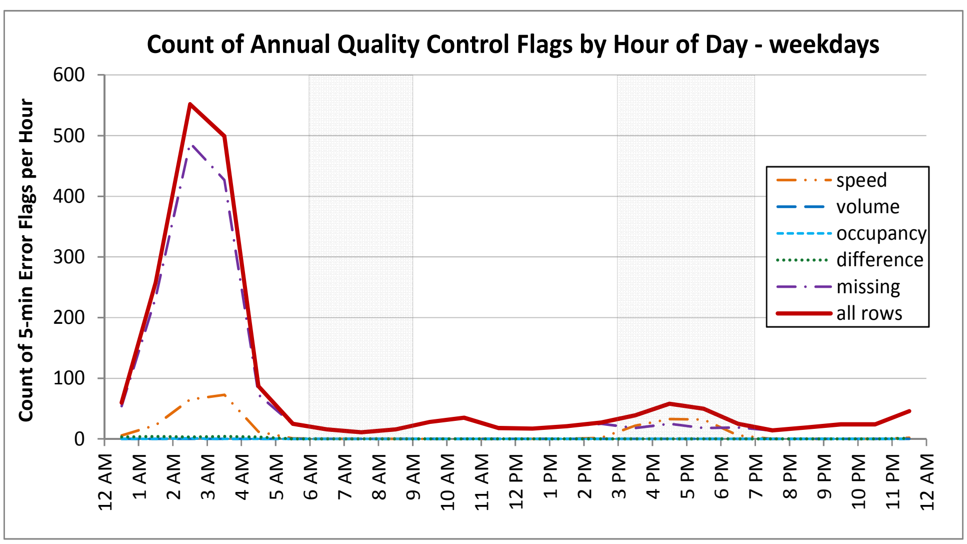
The use of data quality visualization charts at each freeway traffic detector has provided important insight that cannot be communicated easily through numeric statistics. With available data quality information at each detector, summary visualization charts can also be created to describe corridor level or system wide data quality. Much of the mystery in the data has in this way been exposed for better understanding and further questioning. The potential in the data quality assessment taking place here is specifically in the questions it educates. This chapter discusses some of the additional questions that can be answered or further investigated after insight provided from this data quality assessment. This chapter also discusses some of the qualitative lessons learned from this exercise.

Educated Questions

Several questions about the quality of the data and the quality control and data aggregation process are exposed in this assessment for further investigation. Questions exist in things like the value or appropriateness of each of the validity criteria, and in things like investigating the application of principles of traffic flow theory to assist in estimating the accuracy of detector data. The following is a list of some potential issues or questions to be investigated. Some questions are easily answered but most are only highlighted for further study.

1. How appropriate are each of the automated microscopic validity criteria?

Not all faulty data are captured with error criteria being used in the quality control, and some of the data being filtered out are in fact valid. The rules applied in the microscopic detector level validity checks are taken directly from recommendations in the research and have been applied in as simplistic a manner as possible. The merits of the specific validity criteria are beyond the scope of this research taking place here, and yet the assessment makes it clear there is room for improvement in the filtering effort. One of the ten visualization charts discussed in chapter three, showing the distribution of hourly error flags by time of day, is shown in Figure 5.1 below. This chart demonstrates two examples during the middle of the night, where 5-minute average speeds probably are at times higher than the threshold of 85 mph, and individual lanes almost certainly do have no traffic for 20-minutes at a time or more. These “real zeros” being filtered out as missing data in the middle of the night are especially common on suburban freeway corridors that see little long distance traffic. Both the speed and missing data errors here are cases where the microscopic validity criteria most appropriate for daytime traffic are shown to be less effective for low volume night time conditions.



(I-10 WB at Loop 202/SR 51 interchange, 2009)

Figure 5.1: Distribution of Hourly Error Flags by Time of Day

Deficiencies in the validity criteria during moderate or heavy traffic would likely be much more difficult to see than the rejection of valid data at night, as shown in the previous figure. Apparent rejection of real zeroes and speeds in the middle of the night encourages experimentation with missing data threshold higher than 20-minutes for a given lane and a speed threshold higher than 85 mph. It is unclear however how such changes to the criteria would change filtering during congested times of day, which are consequently the most important in affecting performance measure results. Validity criteria should be directed first to the most important traffic conditions.

2. Should validity criteria be applied differently to different times of the day or traffic conditions?

With some additional programing effort the validity criteria can be could be defined as a function of volume or time of day. Defining validity criteria appropriate to specific traffic conditions would allow rules that better address the data quality challenges most typical in the data without erroneously filtering out large amounts of valid data.

3. How significant is the effect of potentially faulty data on end performance measures?

Some performance measures are highly sensitive to changes in the data while others are not. The effect of faulty data therefore depends on the performance measure or the application of the data. Where information about the detector data is available, and there are reasons to suspect data to be in error, sensitivity analyses can be done in estimating the importance of having high quality data.

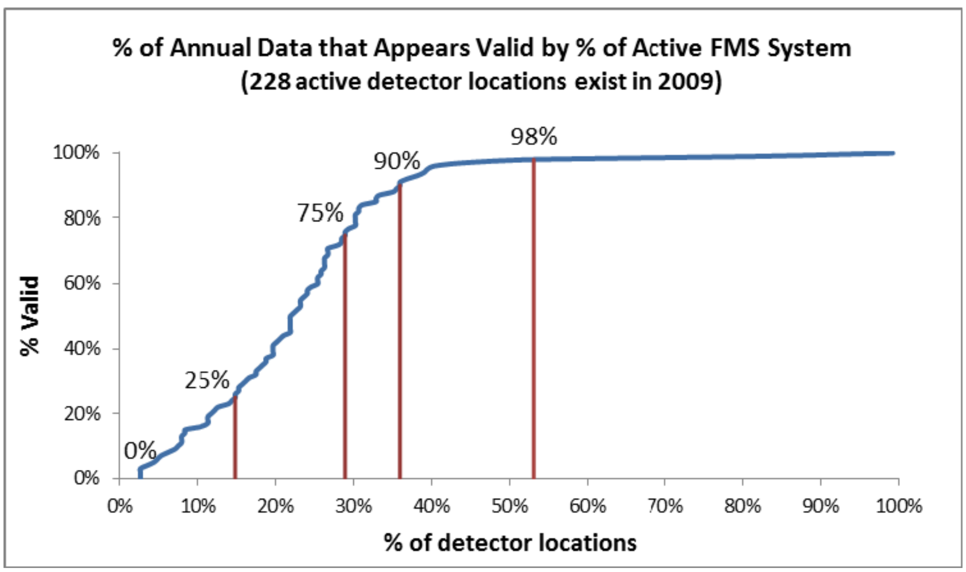
An example of investigating the effect poor quality processed data on performance measures is given in the case of real zeroes being filtered out during low traffic night time hours as shown in Figure 5.1. In this instance, all performance measures derived from speed data are completely unaffected, as speeds are weighted by volume and zero traffic data rows do not influence average speeds regardless of whether or not they are filtered out as missing. Volume statistics are however affect where, for example, an hourly average count with real zero values filtered out would appear higher than they should if the zero count rows were included in the average. Volume derived performance measure are however almost always calculated from daytime weekday hours. The only performance measures where night time traffic values are used in the analysis performed with this data is in the calculation of annual average daily traffic (AADT) and average annual weekday traffic (AAWT). A sensitivity analysis was therefore performed where several detector locations were selected and the calculated AADT was estimated as if the zero value rows were included. Because night time traffic represents such a small portion of the overall traffic, the difference in AADT was limited to a few percentage point change in the cases investigated. Regardless of the significance of the error, the information available makes it possible to identify and correct the count estimations.

4. Should the rejection of speed data preclude using volume or occupancy data?

In the current analysis all data elements are rejected for a given 5-minute data row wherever any of the elements fail to meet even one of the validity criteria. The assumption is that if a detector malfunctions with regard to speed data, values shown for volume and occupancy should be suspect as well. While this assumption is usually reasonable and makes the analysis a little more simple, it may not be the most effective way to utilize the available data. The most significant cause of data being rejected from the analysis is in missing data in one or more of the data columns for a given row. Figures 4.2, in the previous chapter, shows that more than 12% of the data rows are flagged with a “difference error” flag, meaning data is missing in at least one lane for speed, volume and/or occupancy but not for all three of these data elements. The rejection criteria could therefore be changed to allow, for example, speed data where volume data is missing, or any other variation of partially missing data between speed, volume and occupancy. Most appropriate may be an order of priority where some data error flags preclude use of all data elements while others do not.

5. How does the data quality change from year to year and with the addition of new detector locations?

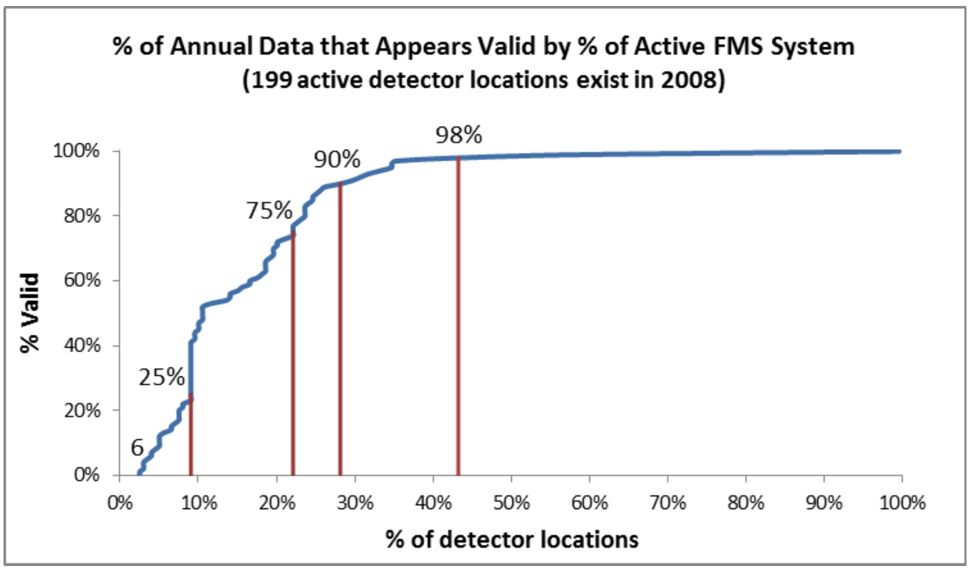
Another question that can be answered is regarding the change in the system wide quality of the data after 29 new detector locations have been installed between the beginning of 2008 and the beginning of 2009. In the following figure, the total number of detectors is shown on the x-axis as a percentile, while the total percent of the data considered both valid and not missing is shown on the y-axis.



(all detectors, 2009)

Figure 5.2: Year 2009 Distribution of Detector Level Validity

The chart shows that about 15% of the detectors have 25% or less of the data as valid and not missing and about 29% of the detectors have 75% or less of the data as valid. It could also be stated that 64% of the detectors have 90% or more of data passing validity criteria. The same chart for 2008, Figure 5.3, shows that less than 10% of the detector locations have 25% or less valid data, 22% have less than 75% valid, and about 72% of the locations have at least 90% of the data as valid. It would appear that with the addition of the 29 new detector locations, the overall percent valid of all detector locations has actually been reduced. Investigation on individual detector results actually shows that most of the new detector locations are in the better half of detectors. The reduction in system wide validity in 2009 is largely attributed to construction activities and related lane closures in a single corridor where data is missing at 14 detector locations for most of the year.



(all detectors, 2008)

Figure 5.3: Year 2008 Distribution of Detector Level Valid

6. Are priority maintained detectors any better than others?

58 detector locations have been designated by ADOT as priority maintained detectors. The aggregation and quality control result files created in the analysis

performed in this project have been investigated for these 58 detectors. Visualization charts for individual detectors continue to show many of the concerns with missing data and irregular transitions in the annual average values between adjacent lanes. The overall validity at these detectors is however somewhat better than the validity at all other detectors. The count of the detectors rejected in the macroscopic corridor level validity checks is eight (or 14%) from the priority maintained detectors, and 38 (or 32%) from non-priority maintained detectors. The table below summarizes the percent of the data considered valid after detector level microscopic validity checks and after corridor level macroscopic validity checks. The corridor level percent valid is higher in both instances as the worst detector locations, rejected in the corridor level quality control, are not included in the average.

Table 5.1: Priority Maintained Percent Validity (including missing data criteria)

	detector level	corridor level
from 58 priority maintained detectors	85%	90.7%
from remaining 170 detectors	78.9%	86.6%

7. To what extent can principles of traffic flow theory be used to validate, reject or even adjust suspect data?

The last three of the ten visualization charts, presented in chapter three, portray the well-studied relationships between speed, flow, and density. Macroscopic corridor level quality control in this analysis uses these charts to identify major deviations from the expected patterns, potentially justifying rejection of a detector location even where most of the data does pass the threshold validity criteria defined in the microscopic detector level quality control. These visual inspections, as well as other validity criteria cited in the Turner (2007) synthesis of quality control procedures, target malfunctioning detector data where the relationships between these three fundamental data elements show gross deviations from expected patterns. What is not understood is the extent of accuracy error due to poor detector calibration. A number of traffic stream models have been proposed over the years, the earliest of which dates back to 1934 with the Greenshield's model (May 1990). More recent, multi-regime traffic stream models can provide a framework for identifying systematic deviation at individual sensors from the theoretical speed volumes and occupancy values. In this way an estimation of accuracy, and a corresponding adjustment factor, could potentially be developed to improve data quality.

These are only some of the questions and issues that can be investigated as a result of the data quality assessment that has taken place. Information and visualizations on data quality makes it possible to identify and investigate questions relevant to not only a performance measurement analysis but any other transportation data analysis task.

Lessons Learned

In additions to the questions just discussed, and many others not discussed, are some important qualitative lessons learned from the experience in this assessment. Three important statement or conclusions are highlighted.

1. Data quality targets are relative, and the quality control process is more than just the application of Pass/Fail decision rules.

The quality targets defined for this analysis, and in the guidance literature, serve only as a starting point in assessing where the data quality stands, from which some insight may be provided on where the quality targets should be set. The suitability of data is, therefore, not a discrete pass/fail assessment. Often the quality of the data itself dictates the quality level that will be accepted and utilized in an analysis. Where data is consistent, complete and accurate, more scrutiny may be allowed in rejecting data raising even mild suspicions. Where incomplete, inconsistent, and inaccurate data are wide spread however, questionable data may be used in place of no data. The critical element is in carrying some understanding of the inherent quality of the data into the analysis. Where an understanding of the data is carried through to the end performance measures, analysts, managers, and other decision makers can have a sense of the confidence level allowed, even in using data that falls below preferred data quality targets.

2. Analysts stand to benefit most from tools that educate an intuitive sense of data quality – visualization tools.

This second statement may not always be true in the future, but currently, data quality measurement standards, and traffic data standards in general, have probably not progressed to a level allowed by current technology. Data quality control and assessment in its current state appears to be a matter of engineering judgment as much as anything else. As such, educating the analyst becomes a primary, or the primary, goal in the data quality assessment and control.

An example is seen in the performance measurement analysis just performed that demonstrates just how subjective can be the quality control effort and resulting performance measurement. The current MAG freeway performance measurement analysis is a second generation effort of the analysis that took place previously using 2006 and 2007 traffic data from the same source. After performing the current analysis for 2008 and 2009, the data for 2007 was analyzed again, presumably using the same methodology as the first generation effort. Comparisons between new and old 2007 performance measure results showed more than 20% difference in corridor level hourly speed and volume results on some of the less healthy corridors. Detector level microscopic data filtering are probably not the problem, but the corridor level decisions are, where rejecting one or some detector locations can have huge effects on corridor level performance measure results. The corridor level decision step, as defined for this analysis, is largely subject to the analyst's discretion as to what data most effectively contributes to the desired corridor results. Future refinement to quality control standards, especially at the corridor level, as well as improvements to traffic data collection in general, promise to add some stability to traffic analysis tasks.

Another example that characterizes the current dependence on engineering judgment relates to the issue of detector spacing. Bertini and Lovell (2009) have taken an analytical approach applying first principles of traffic flow theory for establishing optimal sensor density. The question actually becomes somewhat complicated as the optimal density changes with the various traffic states. More significant than density to the accurate estimation of travel time along a corridor is perhaps the location of the detectors. Sensors are most often placed just prior to a freeway onramp in a position to enable operation of ramp metering, which is often a bottleneck location, and the location where counts will be lowest, between freeway ramps. Fujito et al., (2007) performed an empirical evaluation of several detector spacing's from 0.3 to 4 miles and found that different spacing led to changes in the over- or under-estimation of the travel time index. No evidence was found, however, that the travel time index actually got worse with fewer detectors. The study also suggested that the location of the detectors is important. In summary, travel time estimates from the current traffic data quality levels are not so precise as to benefit much from analytical methods developed. The priority in improving performance measure estimates remains in the common sense decisions regarding detector location and corridor level validity checks.

3. Analytical methods using advanced statistics and/or imputation methods mean little if anything when dealing with poor accuracy data.

The over/under variation in the travel time index shown in the Fujito (2007) detector spacing experiment is likely due to random calibration and accuracy error in individual detector locations. Analytical approaches to optimal spacing, or other data quality questions, are far more relevant when excessive variation due to poor data accuracy is not so large a factor.

Imputation of missing data is another area where analysts can apply very sophisticated approaches. Chen et al. (2003), for example, proposed using linear regression between adjacent lanes and time periods for imputation of missing data values. While this degree of specificity may improve estimations in highly accurate data, it makes little sense in context to a detector chart like that shown in Figure 4.1 from the previous chapter, showing the by-lane profile at a “poorly calibrated” detector.

Failure to understand the data quality may still be a weak link in many traffic engineering and planning fields. Improvements in data collection will continue to contribute to better traffic analyses, especially as consistent quality control and assessment standards become more common practice. Critical in contributing to improved use of traffic data are the use of visualization tools that educate data users to the characteristics and quality of the data as well as standardized application of data quality procedures.

CHAPTER 6

CONCLUSIONS

This research has taken continuous monitoring freeway detector data, at 228 detector locations on 15 corridors in the Phoenix metro region, and performed a data quality assessment to determine suitability for analysis of freeway performance measures. Results from the assessment and research are three fold. First is in identifying the most valid data and in filtering out data considered to be invalid. This has taken place through the use of evaluation criteria applied at the microscopic level, in the rejection of individual data rows, and at the macroscopic corridor level, in the rejection of detector locations in whole. Evaluation criteria used at the microscopic level have been taken directly from recommendations in the literature and serve as pass/fail validity checks. Macroscopic validity checks, though guided by clearly defined principles, serve as heuristics and differ in their application from one corridor to another as required by the characteristics and availability of the data for each corridor. Important in conducting the corridor level validity checks has been the use of ten visualization tools or charts created to describe all analyzed data for each detector location.

The second result achieved in this assessment is in the communication of data quality. Six traffic data quality measures have been defined in the literature as accuracy, completeness, validity, timeliness, coverage, and accessibility; and detector data has been described in terms of each of these six quality measure, and in some cases, subcategories of these measures. With some guidance from the literature, data quality targets have been defined for the quality measures described, and where possible, results have been calculated to evaluate the quality of the data against these target values.

Data quality is also communicated in the analysis result files that have been created for each detector individually. In these files processed data and detector level performance measures are recorded. Each of these detector result files contains a data quality panel where quality statistics and the ten visualization charts described in chapter three are stored. This data quality panel also contains reference to metadata where detector specific information can be found, such as the type of sensor technology used. The data quality panel also reports the specific validity criteria that were used in the processing of the raw data. In this way the detector result files can serve as a database for future use in other traffic analysis tasks where significant information is provided regarding the quality of the data, and transparency exists in how the data was processed.

Corridor result data files have also been created to summarize and aggregate detector information to the corridor level. These corridor result files contain all visualization charts and validity statistics for each detector along the corridor, and a summary of the validity statistics for the corridor as a whole. The visualization charts in the corridor files are used specifically in performing the spatial comparisons of the data between adjacent detector locations, as well as performing the other macroscopic validity checks. These files allow data quality information to be communicated along with corridor level performance measures.

The third result achieved in this research, and the chief emphasis in this report, has been the experience gained in the assessment of data quality and the attempted application of recommendations from the literature reviewed. This research highlights several questions regarding the validity criteria used at both the detector and corridor levels. These and other questions are defined in the extensive use of visualization tools. Experience would therefore recommend that data quality visualization tools be developed and used in the processing and quality control of traffic data wherever possible. Also recommended is that information on quality control efforts, and the measures of data quality, be stored as metadata with processed traffic data.

The chief conclusion arising from this research is that data quality assessment and control are essential elements to any traffic data analysis, but that quality control itself is insufficient. Needed are guidelines and tools that educate an underlying sense of the data. An important contribution is made in this research in the use of data quality visualization tools in assessing the validity of the traffic data beyond pass/fail criteria commonly used. More significantly, these tools they serve to educate an intuitive sense or understanding of the underlying characteristics and quality of the data considered valid.

Recommendations for Future Work

With the rapid advancement of technology, increasing opportunities exist to advance the transportation engineering and planning fields through better utilization of high quality traffic data. Data quality research conducted in the past decade or more is really only the start of what it needed to bring analysis methods closer to what current technology will allow. The analysis conducted for this study underscores some important elements of the data quality assessment that have not been satisfactorily addressed in the research and in current practice. Future work needed in addressing a few of these deficiencies are as follows:

1. Documented sensitivity testing to further validate and refine microscopic data filtering rules and criteria.

Validity checks cited in the research and from practice do not usually differ greatly from one source to another, and in cases are adapted directly from other research, as has been the case with most of the validity filters in this analysis. Potential concerns with some validity criteria have been discussed in the previous chapter and evidence exists to suggest that both invalid data passing validity checks and valid data being erroneously filtered out can have a significant effect on resulting processed data. Further research is therefore needed in refining these validity criteria.

2. Case studies that specifically utilized macroscopic validity rules in filtering data evidenced to be in error.

The validity rules used in evaluating and rejecting detector locations from corridor level aggregation have received less attention in the literature than the microscopic validity criteria. The characteristics between corridors can vary such that it becomes difficult to apply standard rejection criteria. Published data quality case studies are however needed to add insight and to help narrow in on useful standards for spatial consistency and other macroscopic validity rules.

3. Development and systematic use of visualization tools in corridor level validity checks.

The ten visualization charts, presented in chapter three of this report, vary in their respective value they have in communicating relevant or important information on the data. Further experience can refine this set of visualization tools to include other more relevant or useful charts. Of particular importance the effective utilization of visualizations charts in the corridor level quality control process.

4. The analytical framework for calculating confidence intervals based on the quality of traffic data.

Important work can be done in merging analytical methods with empirical research, ultimately providing the means for estimating confidence intervals in the data. An important distinction is made between the measured variance in the data and the more difficult to measure bias that exists where detectors systematically report inaccurate data. Specific guidance can be defined in the application of traffic flow theory in using relationships between speed, flow, and density in performing validity checks as well as providing reasonable estimates on the potential error in the data.

REFERENCES

ADOT, Estimated State Planning & Research Program Part II – Research. Arizona Transportation Research Center, Arizona Department of Transportation, 2010.

Ahn, K. and H. Rakha, Data Quality White Paper. 2008, Prepared for Federal Highway Administration: Washington D.C.

Bertini, R. L., and D. J. Lovell., Impacts of Sensor Spacing on Accurate Freeway Travel Time Estimation for Traveler Information. Journal of Intelligent Transportation Systems, Vol. 13, 2009, pp. 97–110.

Battelle, Traffic Data Quality Measurement. 2004, Prepared for Federal Highway Administration: Washington D.C.

Chen, C., Kwon, J., Rice, J., Skabardonis, A., Varaiya, P. Detecting Errors and Imputing Missing Data for Single Loop Surveillance Systems. In Transportation Research Record: Journal of the Transportation Research Board, Washington, D.C., 2003.

Fekpe, E. and D. Gopalakrishna, Traffic Data Quality Workshop Proceedings and Action Plan. 2003, Prepared for Federal Highway Administration: Washington, D.C.

Fernandez-Moctezuma, R. J., Bertini, R. L., Maier, D., Tufte, K. A. Toward Improved and Transparent Imputation Techniques for Online Traffic Data Streams and Archiving Applications. Transportation Research Board 88th Annual Meeting, 2009, 16p

Fujito, I., R. Margiotta, W. Huang, and W. Perez, Effect of Sensor Spacing on Performance Measure Calculations. Transportation Research Record, 1945, 2006, pp. 1–11.

Guerra, T., and K. Satoyoshi, ITS Traffic Data Consolidation System. 2005, Prepared for Arizona Department of Transportation: Phoenix, Arizona

ICDN. "Insights from the ITS Data Quality Workshop." Newsletter of the ITS Cooperative Deployment Network. http://www.ntoctalks.com/icdn/dq_workshop.html (accessed November 22, 2010).

Jonas, G., Evaluation of SmarTek SAS-1 Passive Acoustic Detectors on Interstate 17. Arizona Department of Transportation, Phoenix, Arizona, 2001.

Margiotta, R., State of the Practice for Traffic Data Quality: White Paper. 2002, Prepared for Federal Highway Administration: Washington D.C.

May, A. D., Traffic Flow Fundamentals. Upper Saddle River, NJ; Prentice Hall, 1990

Middleton, D., D. Gopalakrishna, and M. Raman, Advances in Traffic Data Collection and Management: White Paper. 2003, Prepared for Federal Highway Administration: Washington D.C.

Midwestern Software Solutions. ADOT Transportation Database Management System. http://adot.ms2soft.com/tcds/tsearch.asp?loc=Adot&mod= (accessed March 9, 2011).

Smith, B. L., W. T. Scherer, and J. H. Conklin. Exploring Imputation Techniques for Missing Data in Transportation Management Systems. In Transportation Research Record 1836: Journal of the Transportation Research Board, Washington, D.C. , 2003.

Caltrans PeMS. http://pems.dot.ca.gov/?dnode=State&content=detector_health&tab=dh_summay (accessed March 24, 2011).

Sterling, J., S. Hossak, and T. Bills, Enhancing Arizona Department of Transportation's Traffic Data Resource. 2001, Prepared for Arizona Department of Transportation: Phoenix, Arizona

Tufte, Kristin A; Ahn, Soyoung; Bertini, Robert Lawrence; Auffray, Benjamin; Rucker, James. Toward Systematic Improvement of Data Quality in Portland, Oregon, Regional Transportation Archive Listing. Transportation Research Board 86th Annual Meeting; 2007, 11p

Turner, S., Defining and Measuring Traffic Data Quality: White Paper on Recommended Approaches. In Transportation Research Record 1870, Transportation Research Board, 2004, pp. 62-69.

Turner, S., R. Margiotta, and T. Lomax., Monitoring Urban Freeways in 2003: Current Conditions and Trends from Archived Operations Data. Texas Transportation Institute and Cambridge Systematics, 2004.

Turner, S., Quality Control Procedures for Archived Operations Traffic Data: Synthesis of Practice and Recommendations. 2007, Prepared for Federal Highway Administration: Washington D.C.

Turner, S., Year Two Report on Accuracy Evaluation of Arizona DOT Freeway Management System Detectors. 2008, Prepared for Maricopa Association of Governments: Phoenix AZ

APPENDIX A

EXAMPLE CHARTS

Normal Detector Locations:

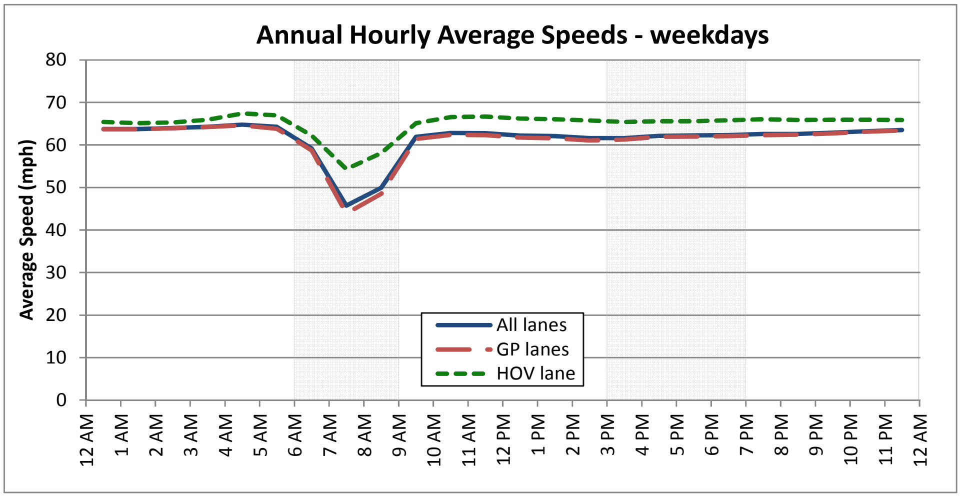


Chart 1, detector 417, 2009

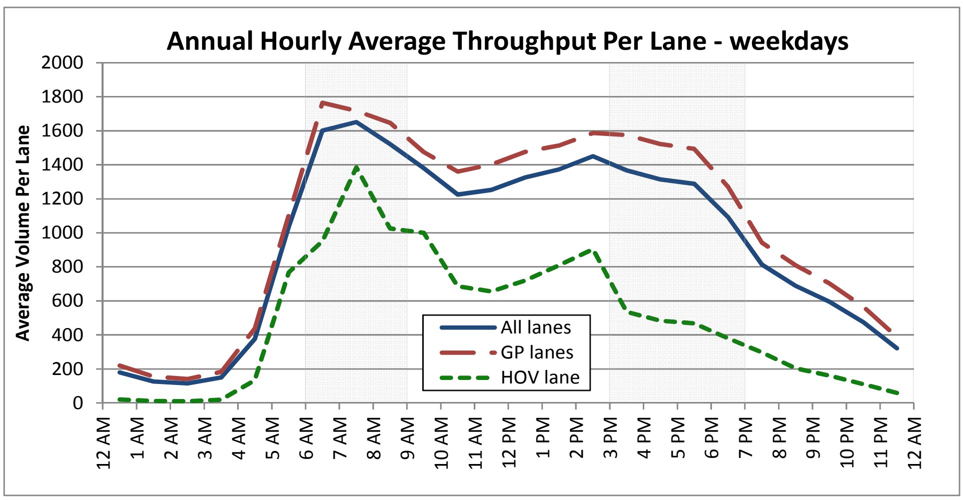


Chart 1, detector 417, 2009

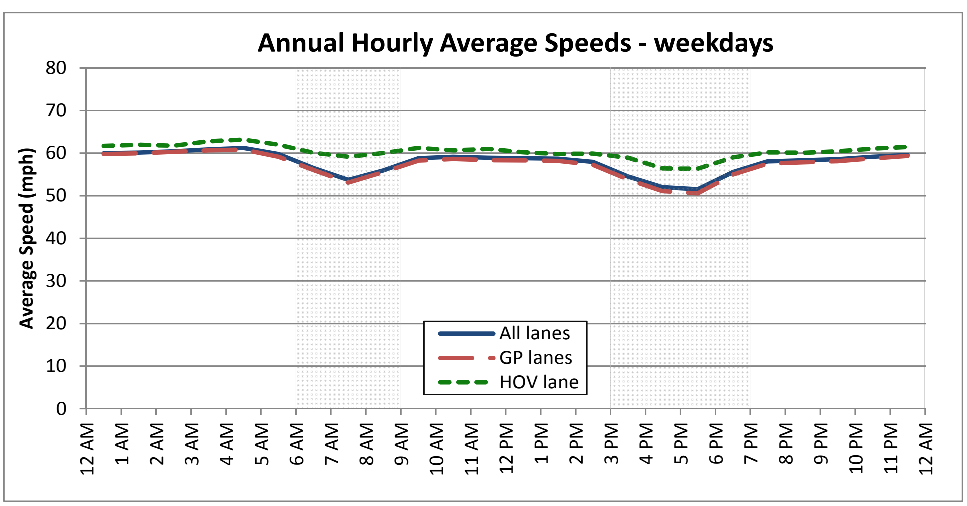


Chart 1, detector 230, 2009

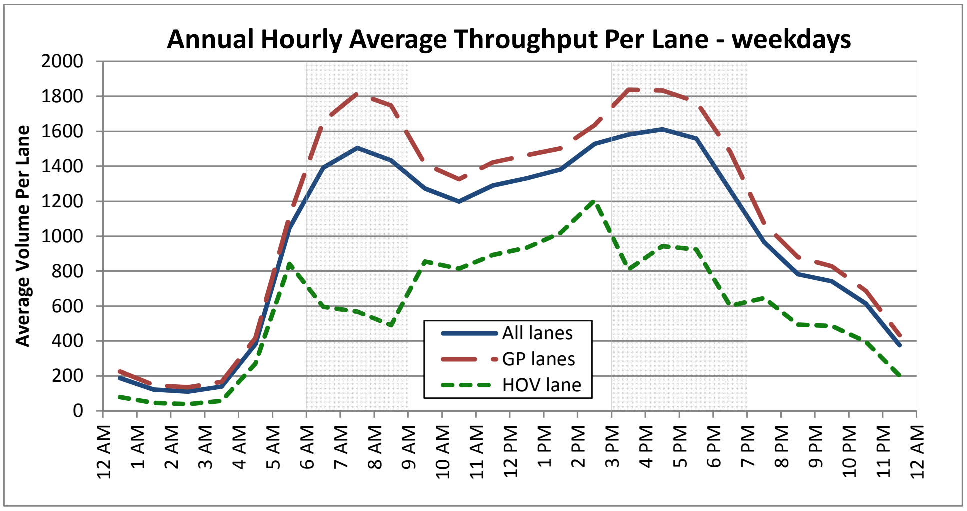


Chart 2, detector 230, 2009

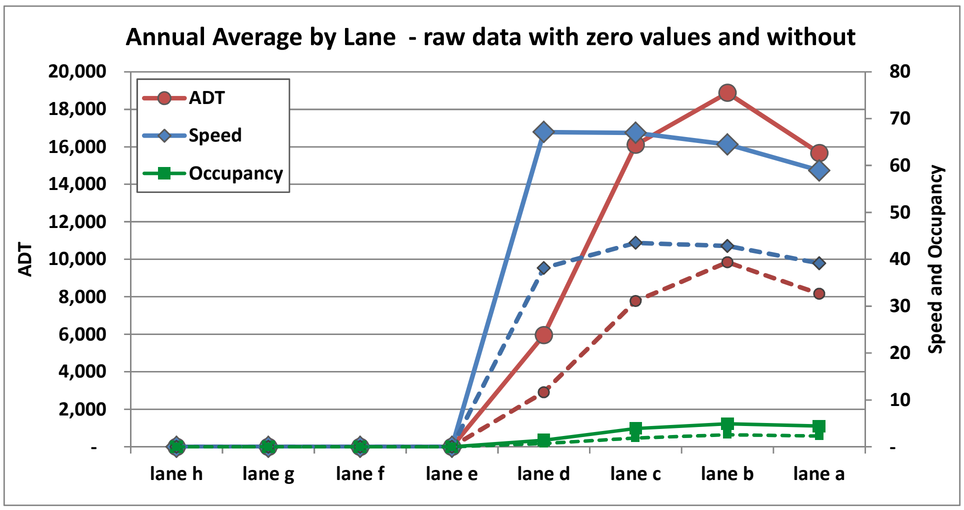


Chart 4, detector 300, 2009

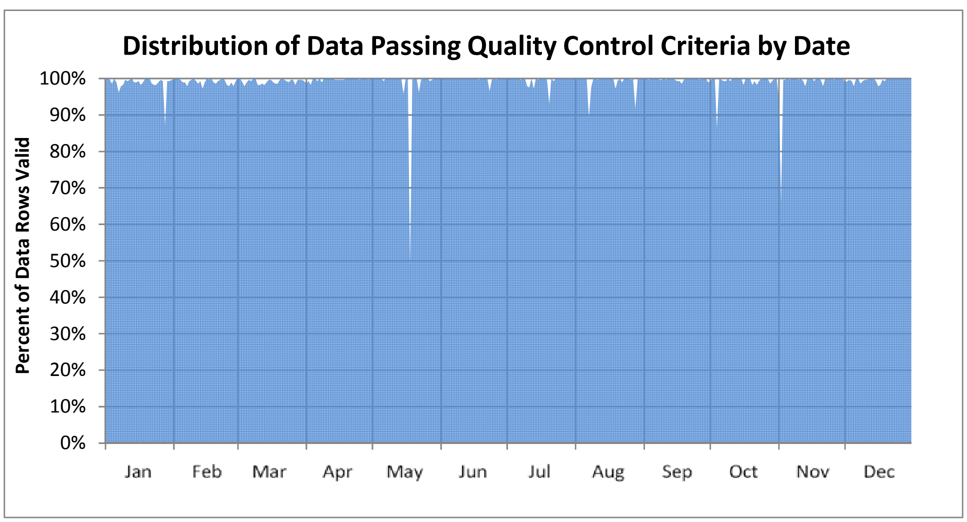


Chart 5, detector 14, 2008

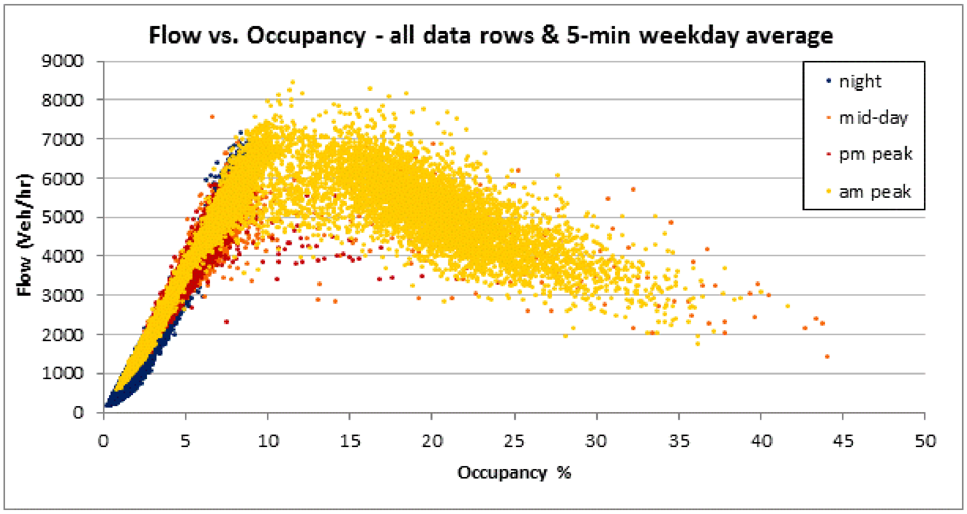


Chart 8, detector 411, 2009

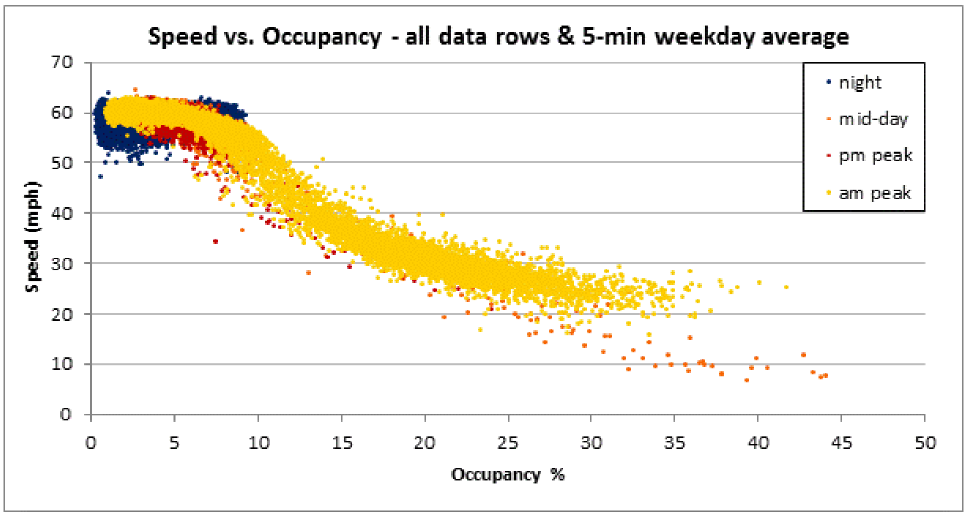


Chart 9, detector 411, 2009

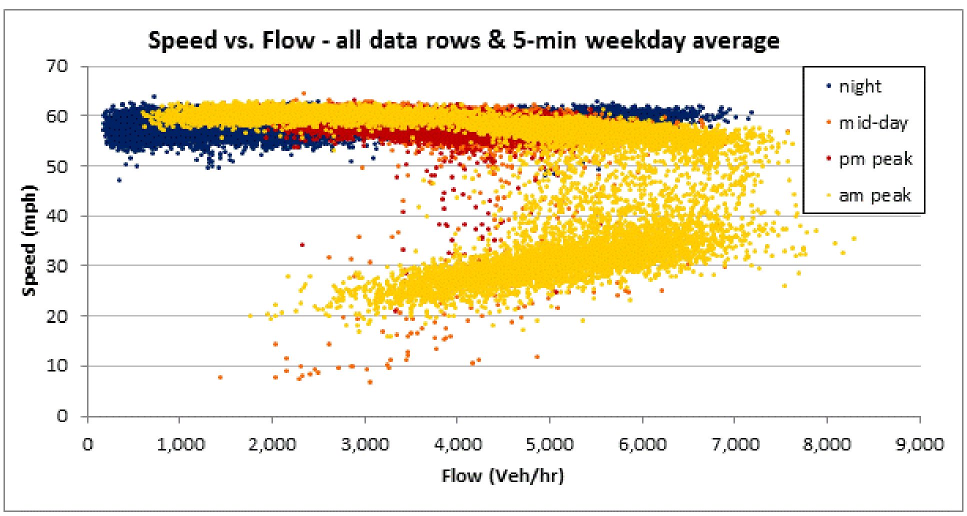


Chart 10, detector 411, 2009

Poorly Calibrated Detector Locations:

Detector 318 shows a the annual average speed in the hove lane to be more than 15 mph lower than the average speed for the other four main line general purpose lanes. While hov speed may drop below mainline lane speeds during individual 5-minute period, they do not typically do so in an annual average of all data periods.

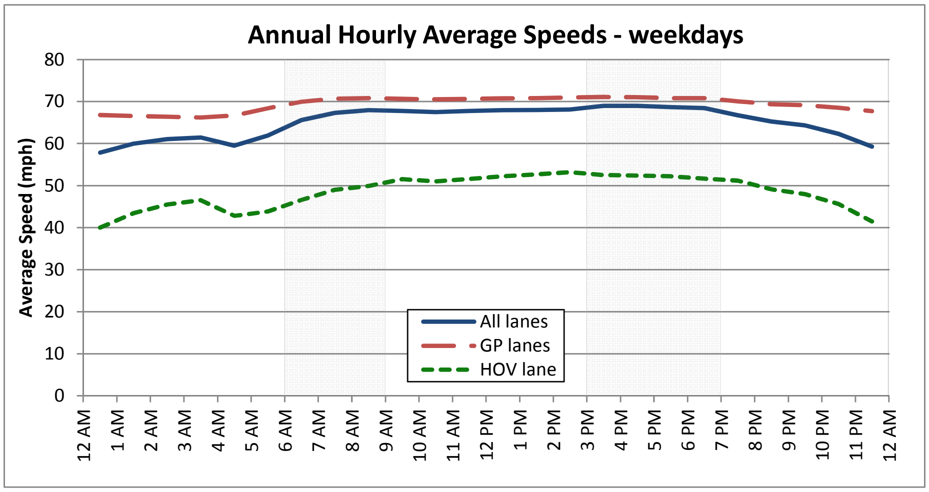


Chart 1, detector 318, 2009

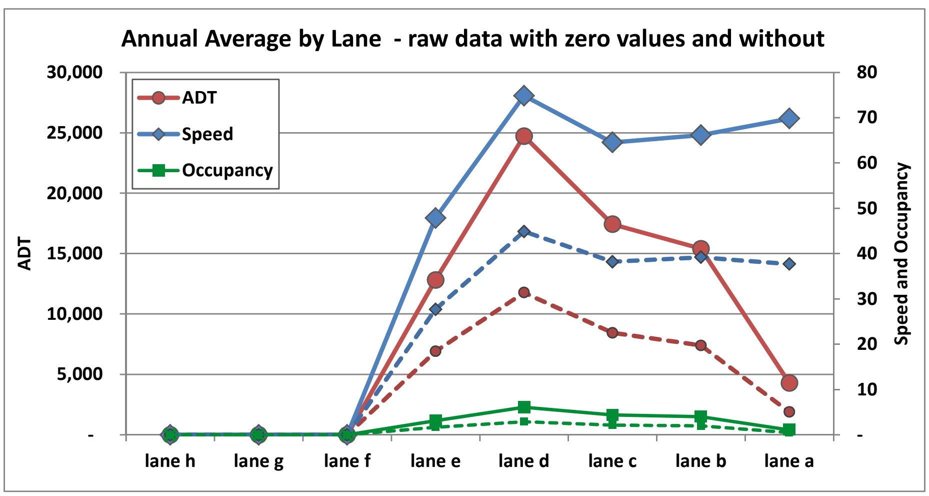


Chart 4, detector 318, 2009

Detector 502 is another instance where chart 4 shows large and irregular variation in the annual average values between adjacent lanes.

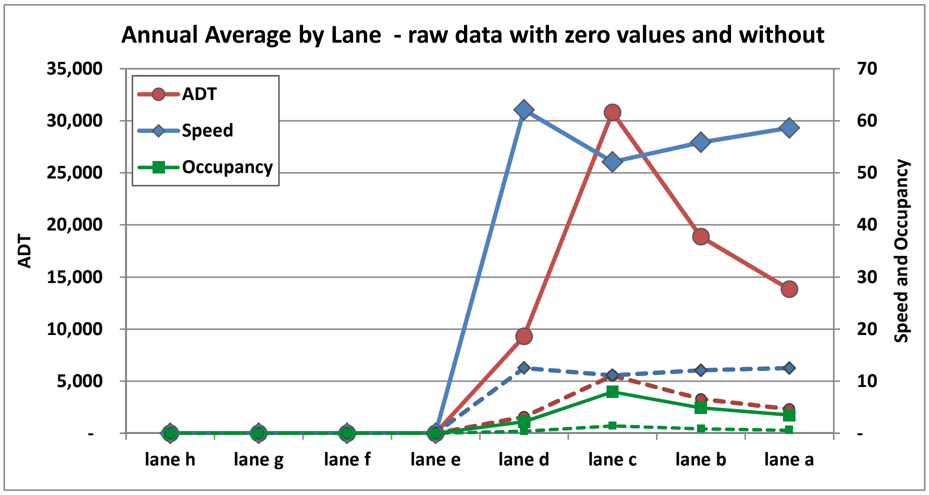


Chart 4, detector 502, 2009

Detector 2, shown in the next several charts, is of a loop detector location where data errors exist do not necessarily exist in all lanes or during the all of the year.

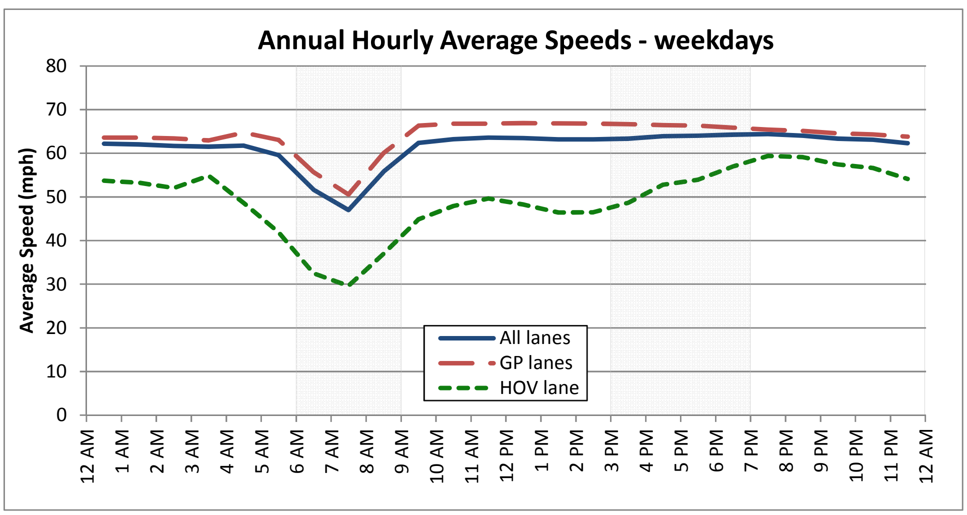


Chart 1, detector 2, 2008

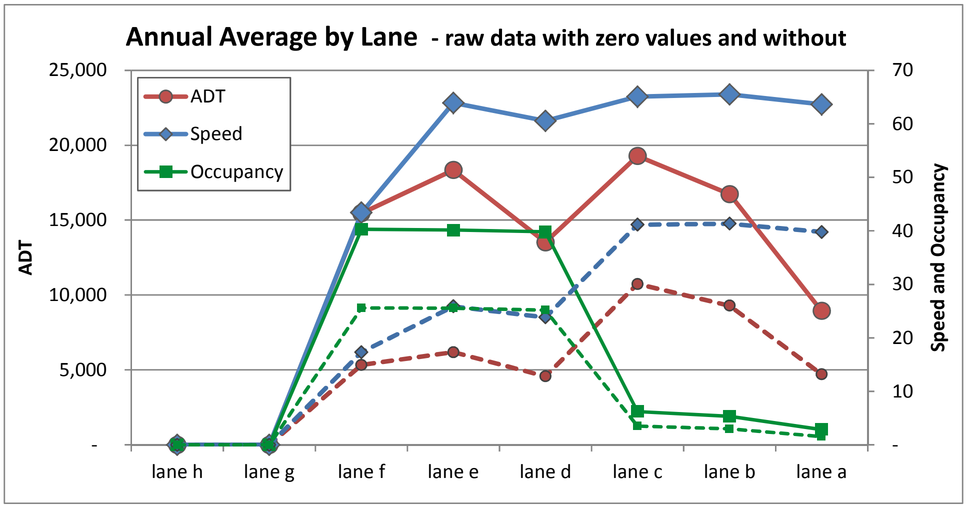


Chart 4, detector 2, 2008

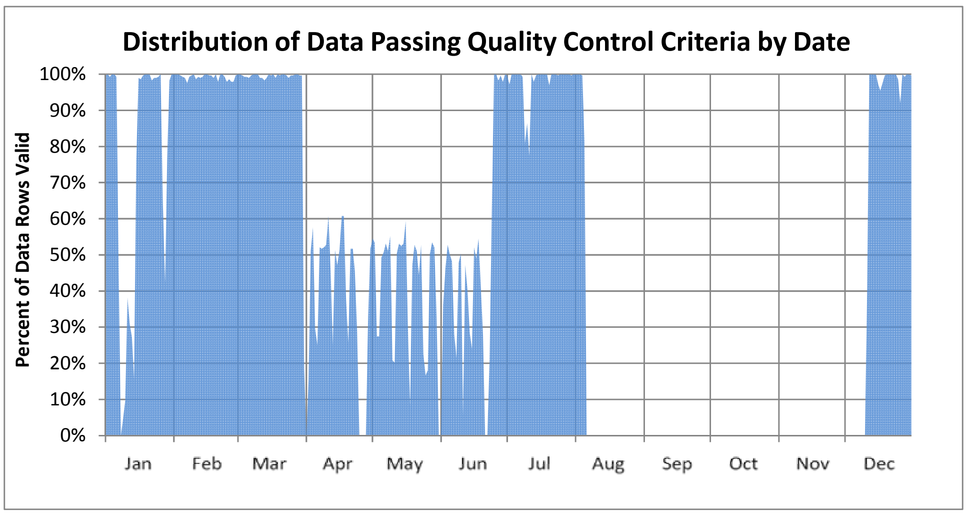


Chart 5, detector 2, 2008

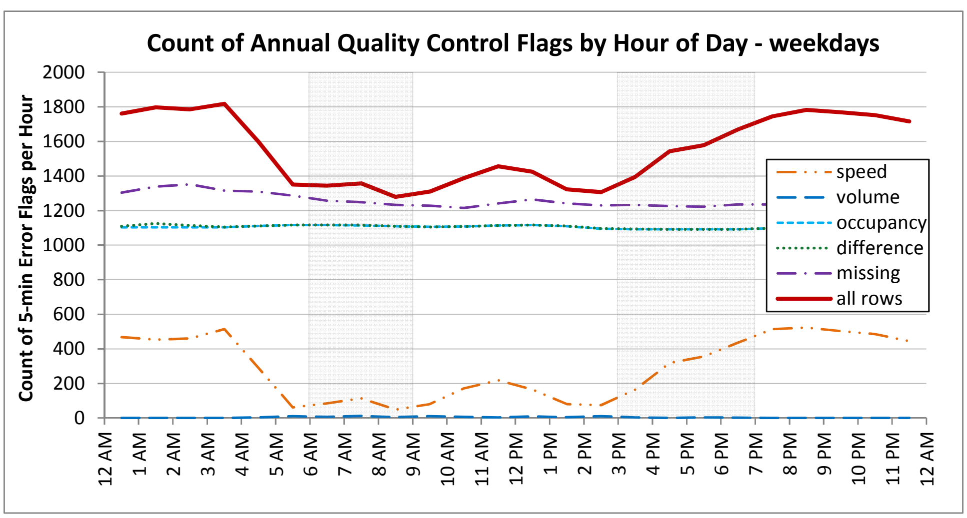


Chart 7, detector 2, 2008

Detector 255 shows a location where malfuctioning detectors leard to unreasonable traffic flow relationships.

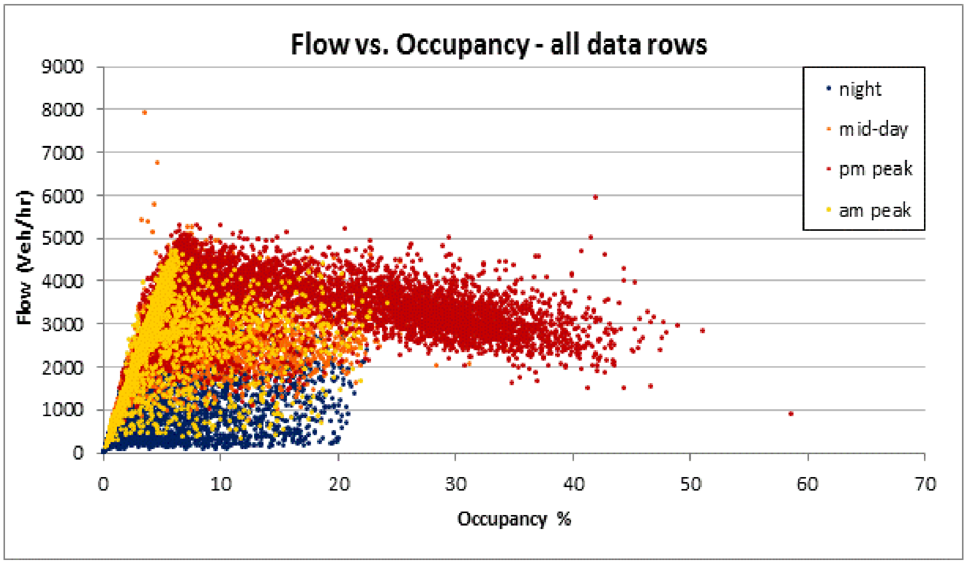


Chart 8, detector 255, 2009

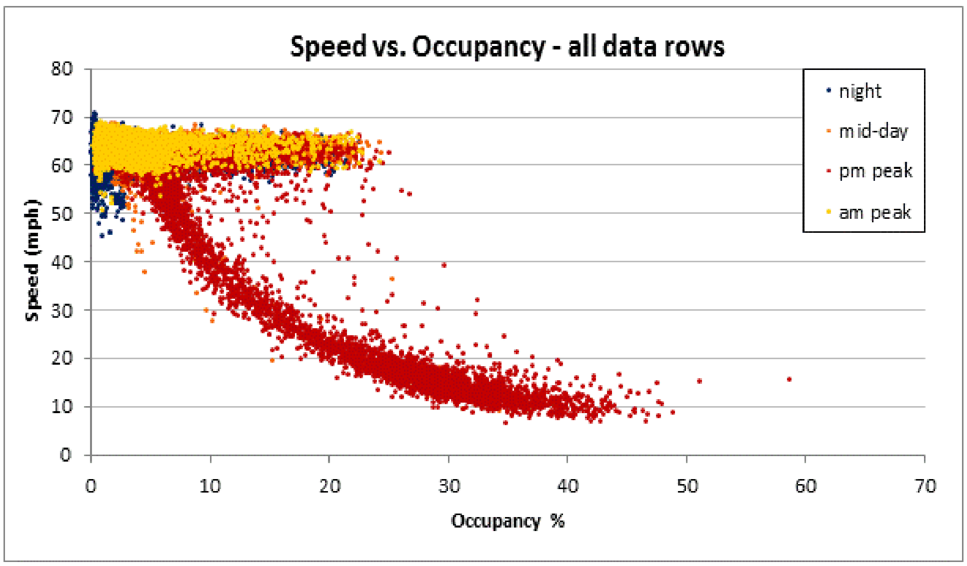


Chart 9, detector 255, 2009

Missing Data:

The following charts show locations where data is missing for only several weeks and another location where data is missing for most of the year.

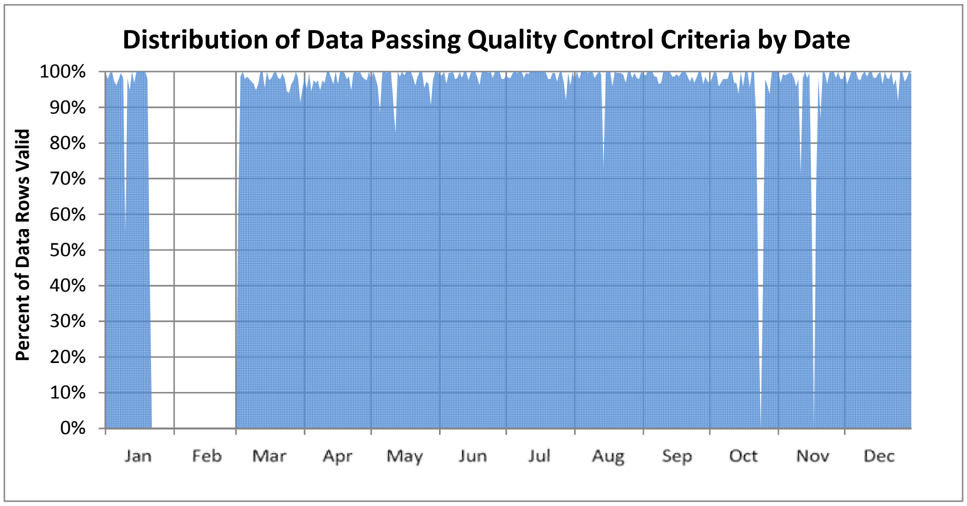


Chart 5, detector 411, 2009

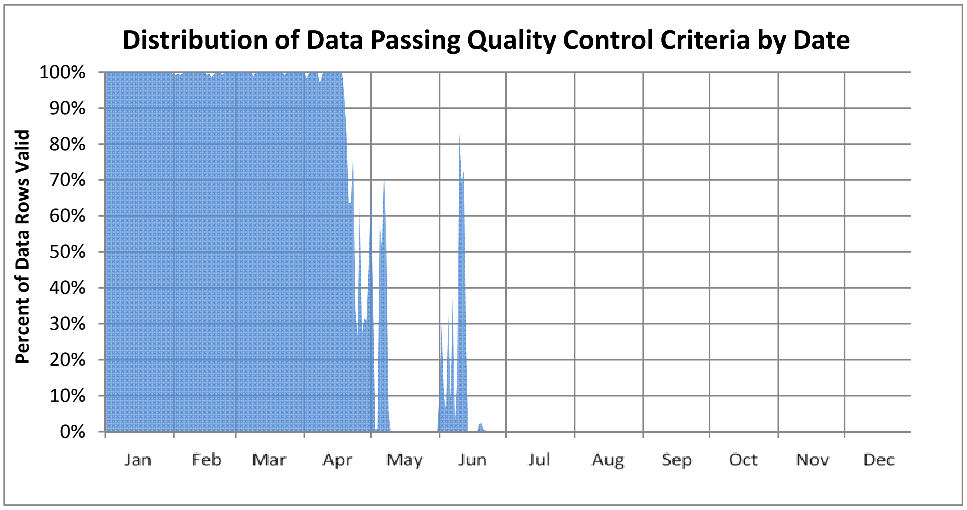


Chart 5, detector 230, 2009

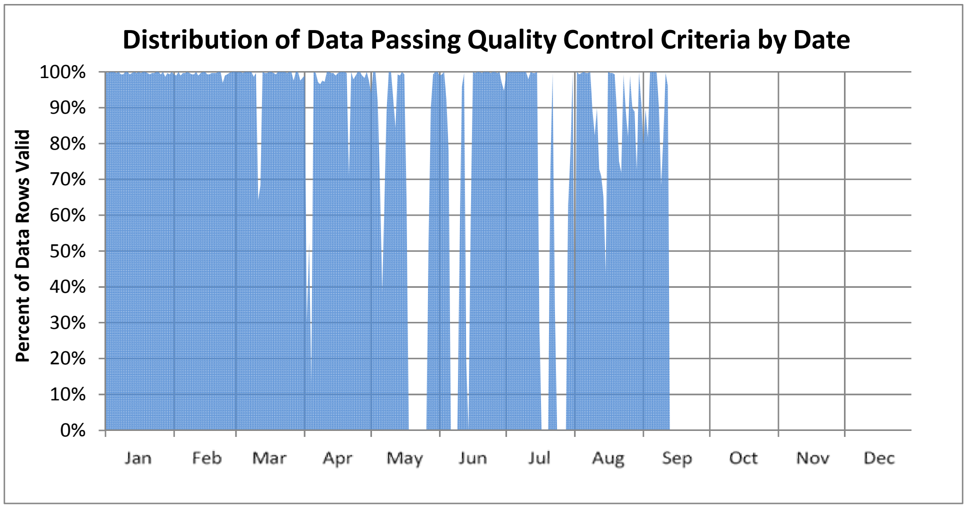


Chart 5, detector 241, 2009

Detector 2 shows a single location with two parts of the year showing very different conditions. During April, May and June the pattern suggests that the majority of the data is invalid and is being systematically filtered out. From August to December data is actually missing. Note that it is easier to have confidence in detectors where large amounts of the data are missing than it is for detectors where available data are filtered out as invalid.

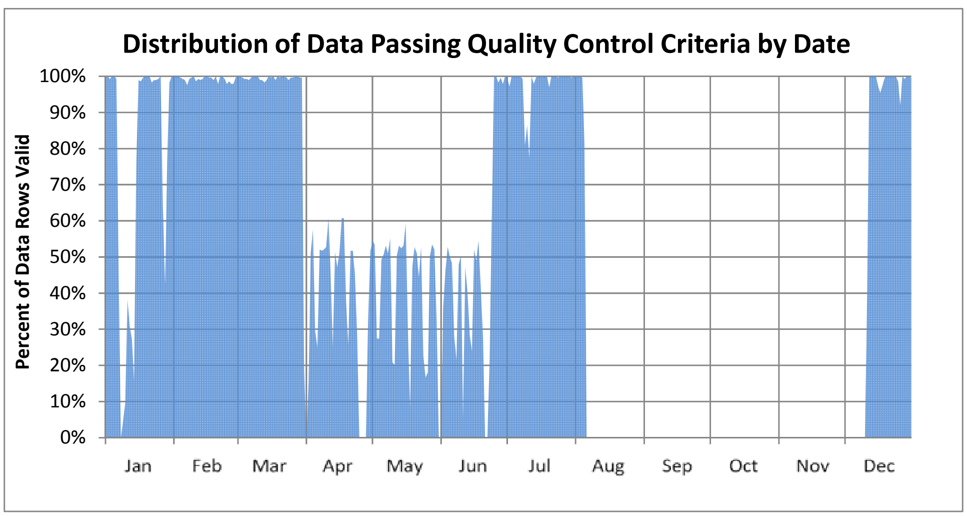


Chart 5, detector 2, 2008

Informatino from Combination of Charts:

Examples are shown here of information povided by a combination of more than one visualization tool.

At detectors 46, and 89, charts 5 show the rejection of nearly all data, and chart 4 shows the specific reason for the rejection of the data, as well as a potential solution.

At detector 46 data is missing in lane b, that is, the lane b volume column contains mostly zeros making the annual average volume with zero values included much lower than it should be. The ananlysis could almost certainly be improved by calculating detector results as a five lane location, rather than six, with lane b removed.

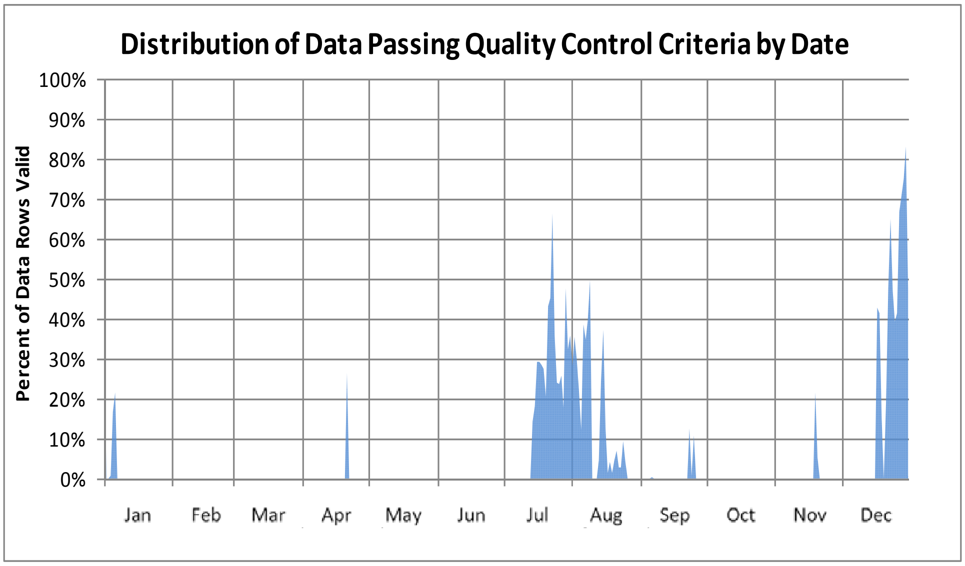


Chart 5, detector 46, 2009

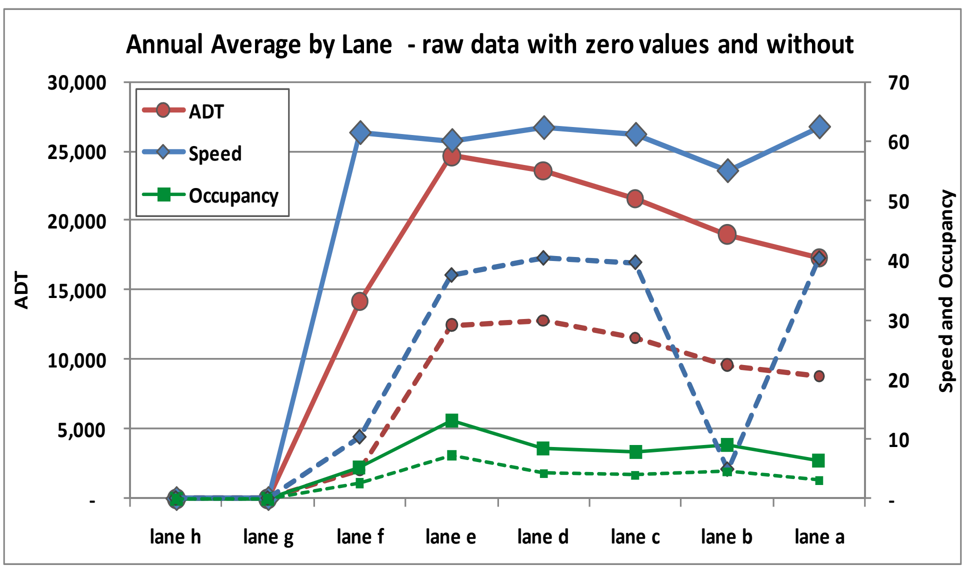


Chart 4, detector 46, 2009

At detector 89 the same situation is seen for lane c rather than lane b where data is missing. The ananlysis could almost certainly be improved by calculating detector results as a four lane location, rather than five, with lane c removed.

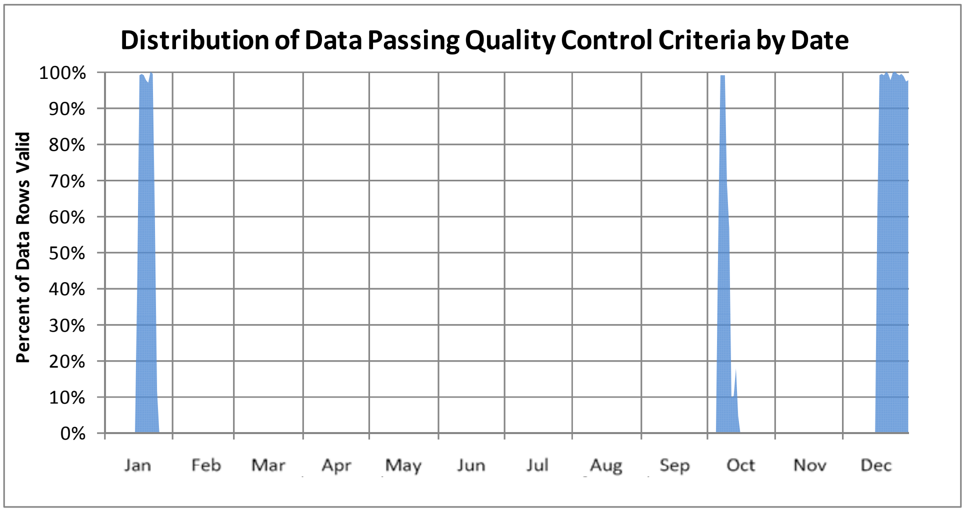


Chart 5: detector 89, 2009

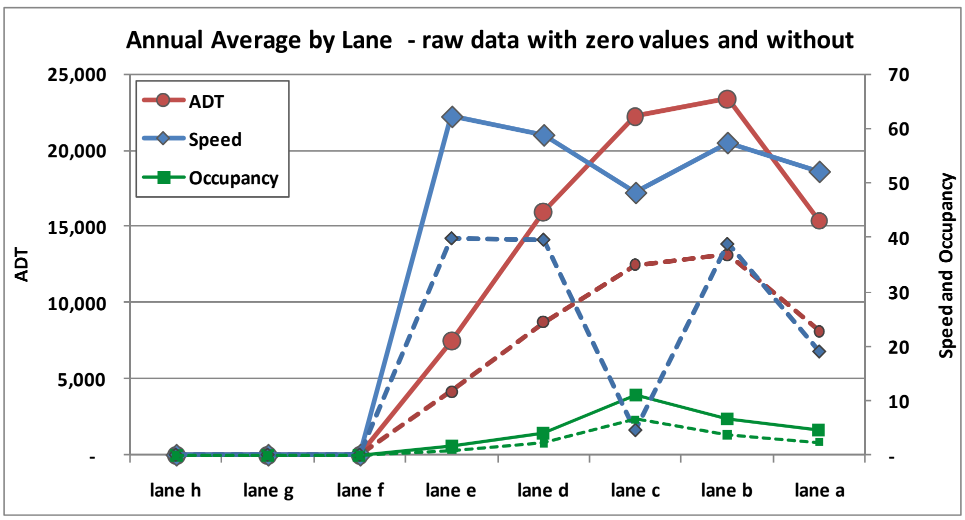


Chart 4: detector 89, 2009

Detector 500 shows a large amount of invalid data. Note that the valididty threshold for the occupancy error flag is 80%. It's clear in looking at charts four and seven below that more than half the data rows (1,700 or so of the 3,000 total hourly rows) contained occupancy values greater than 80% and the remaining less than half were either zeros or nearly 80%.

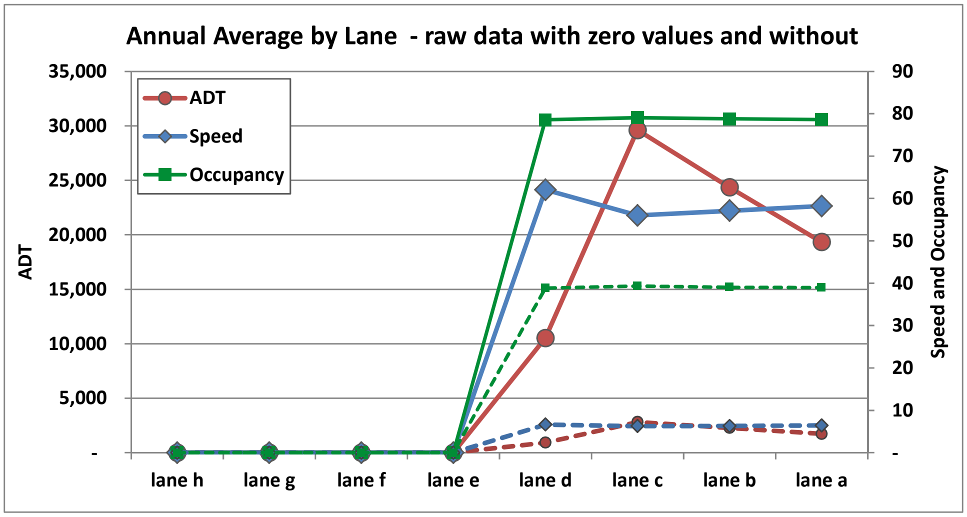


Chart 4, detector 500, 2009

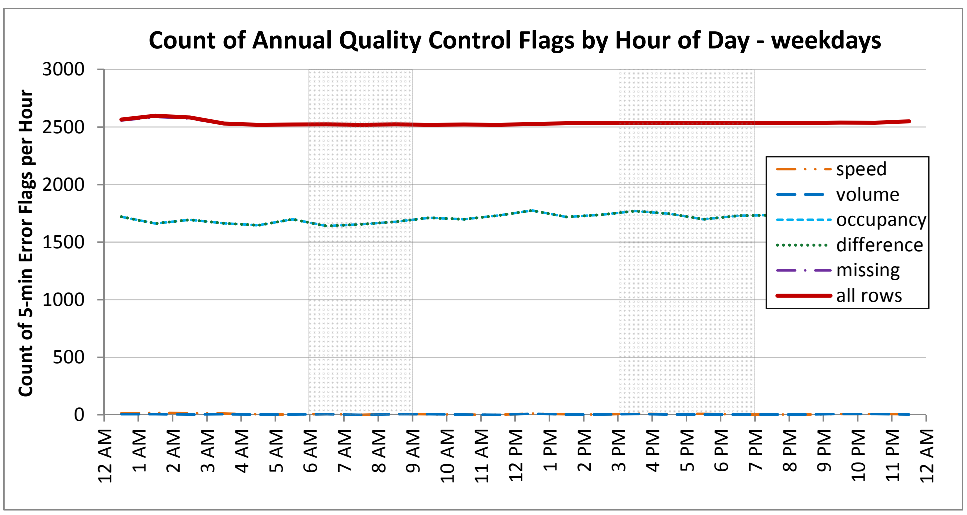


Chart 7, detector 500, 2009

Detector 239 is another location with a high amount of occupancy error. The occupancy error here appears to exist in one lane only. In this situation the hov lane (lane d) might be rejected and the analysis performed with the first three lanes only.

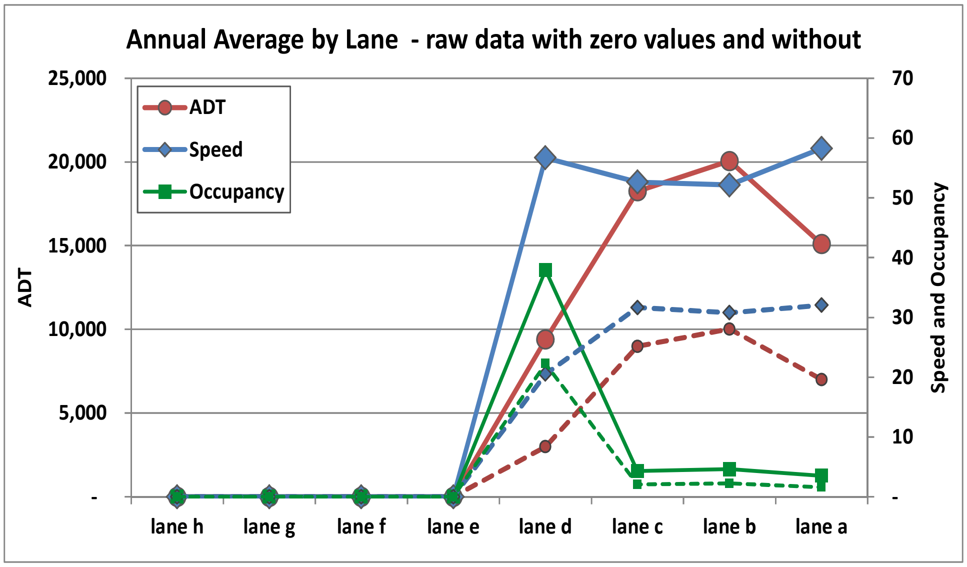


Chart 4, detector 239, 2009

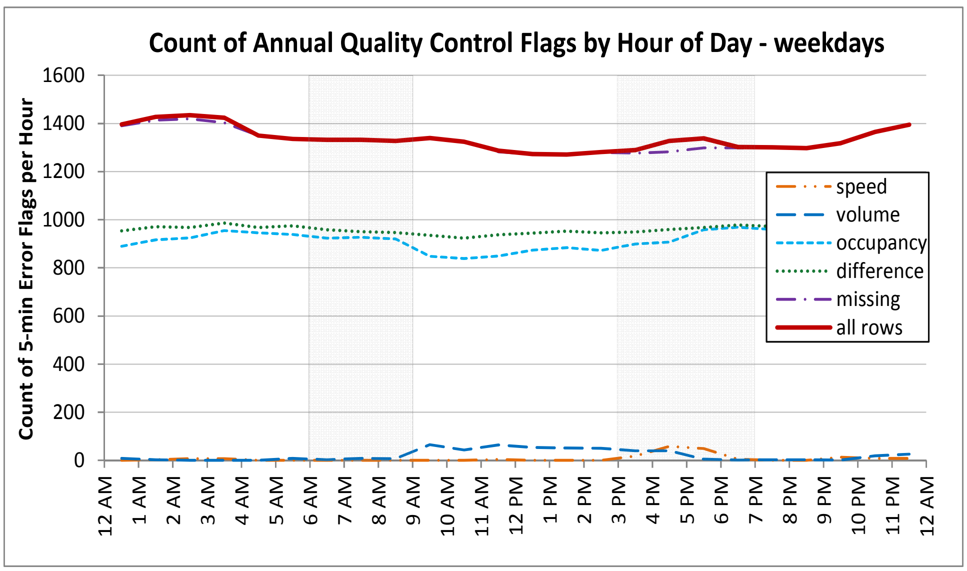


Chart 7, detector 239, 2009

APPENDIX B

SPATIAL COMPARISON OF ADJACENT DETECTOR CHARTS

In the following example, the adjacent detectors 358, 364, and 370 along the I-17 Southbound corridor are compared and detector 364 (middle detector) is rejected from the analysis. In this instance charts 1, 2, and 4 are used.

Speed Profile:

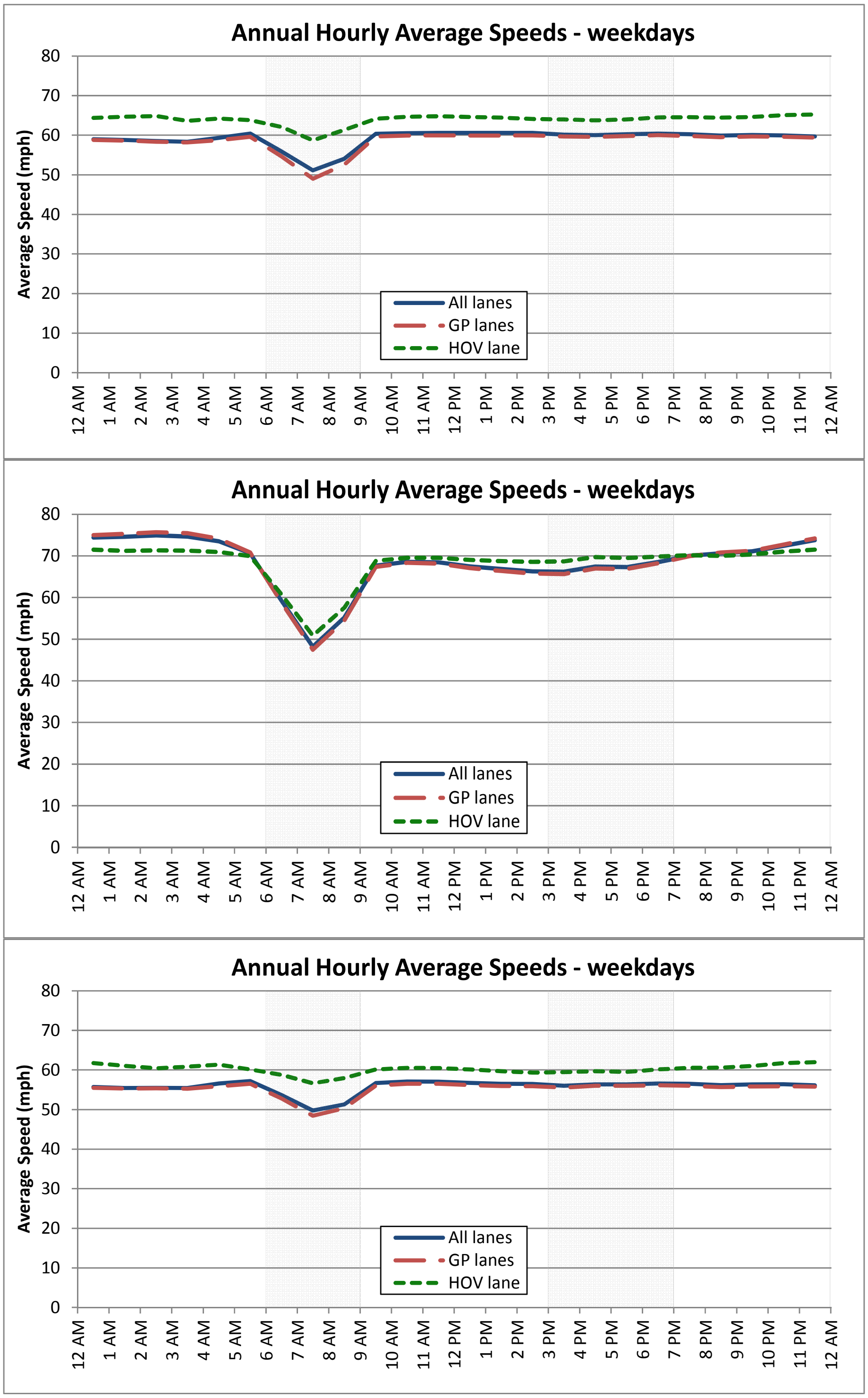
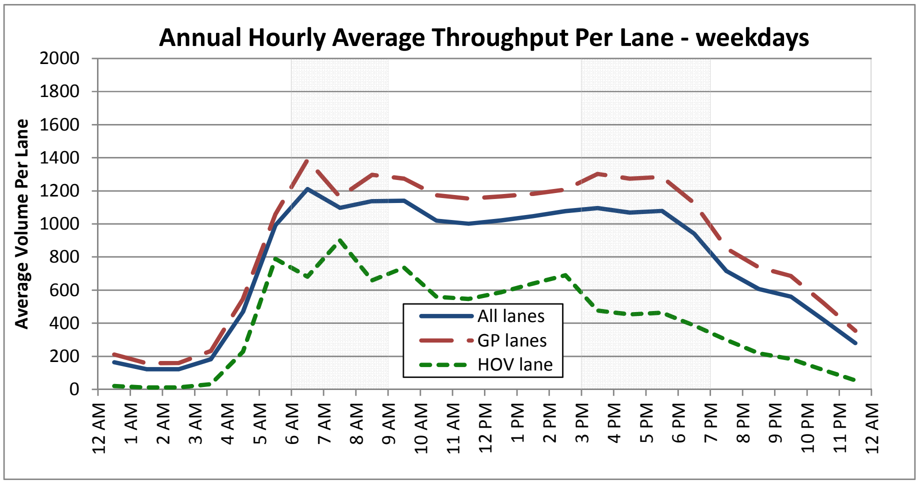
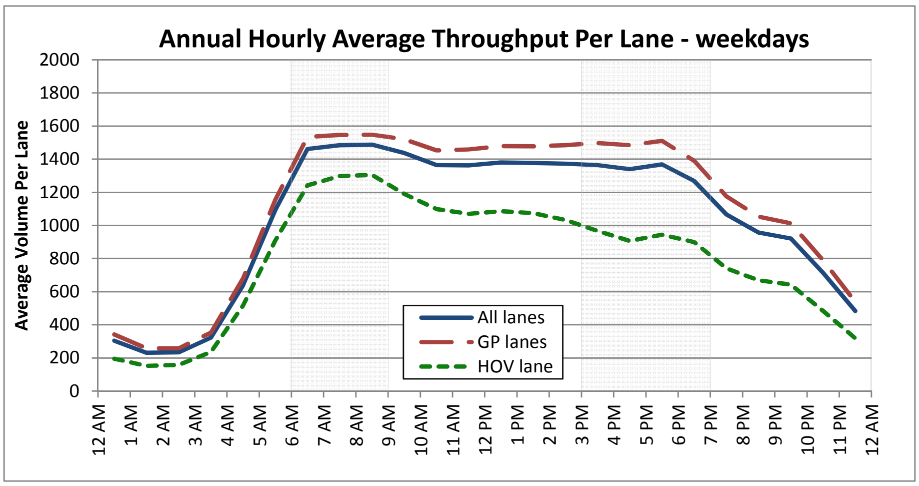


Chart 1, detectors 358, 364, and 370 respectively, 2009

Volume profile:





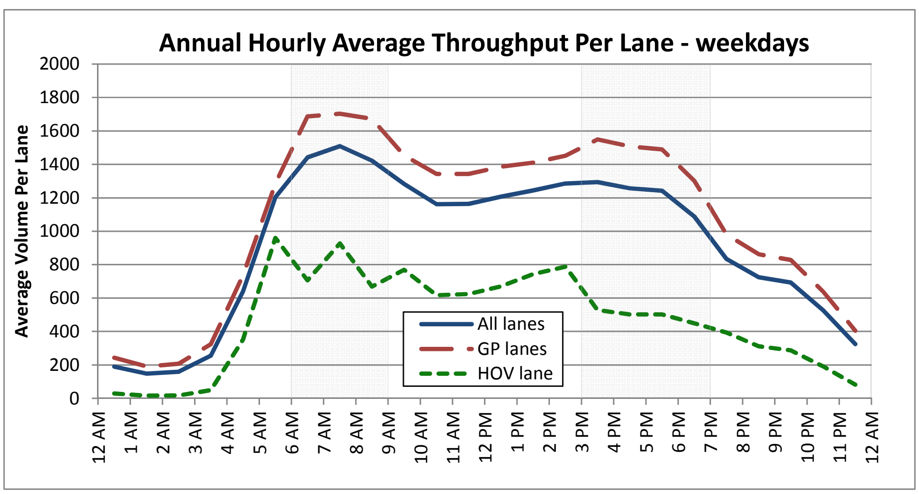
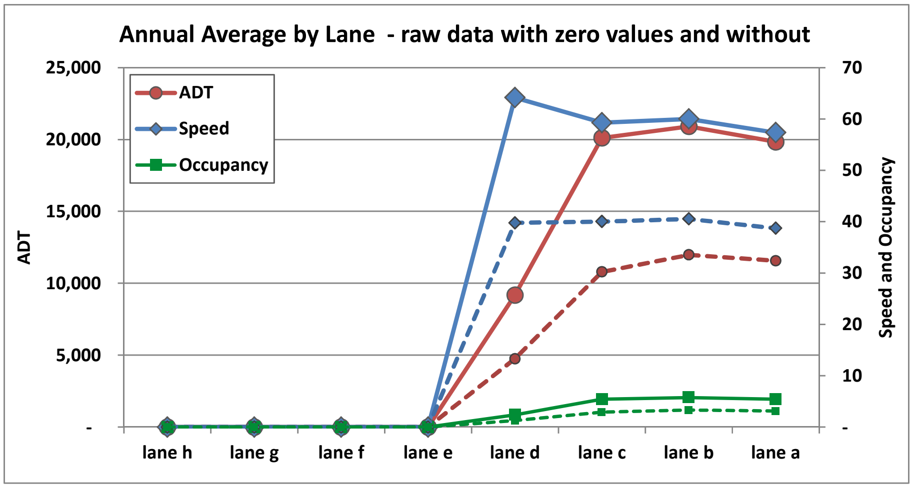
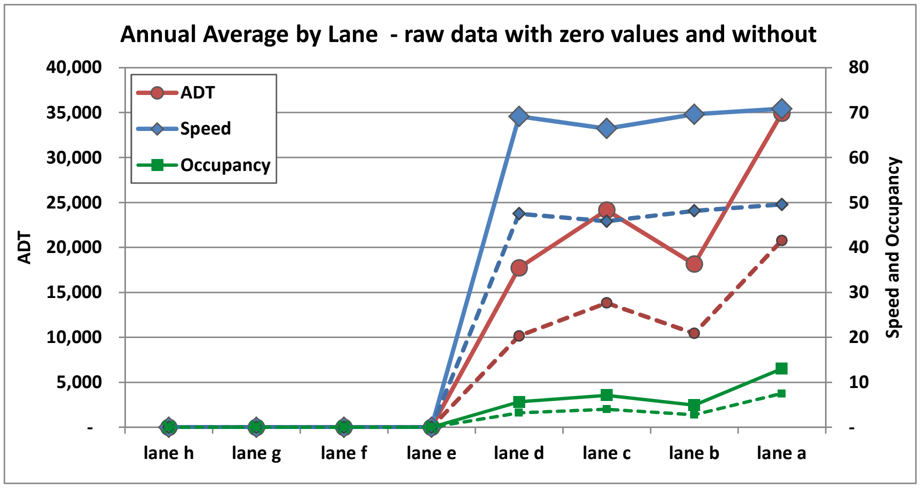


Chart 2, detectors 358, 364, and 370 respectively, 2009

By-lane Profile of Speed, Volume, and Occupancy:





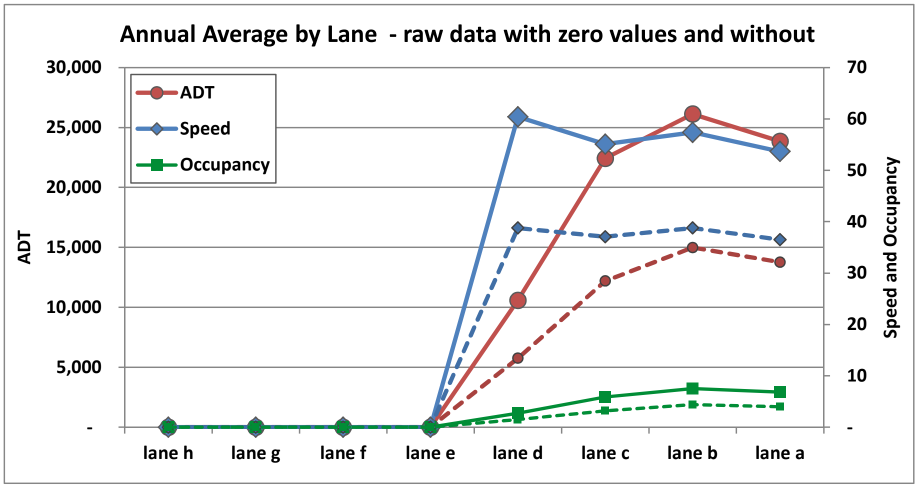


Chart 4, detectors 358, 364, and 370 respectively, 2009