

FINAL CONTRACT REPORT
ITS DATA QUALITY:
ASSESSMENT PROCEDURE FOR FREEWAY POINT DETECTORS

Brian L. Smith, Ph.D.
Assistant Professor
Department of Civil Engineering
University of Virginia

Ramkumar Venkatanarayana
Transportation Systems Engineer

Clinton D. Smith
Research Assistant

Project Managers

Catherine C. McGhee, Virginia Transportation Research Council
Michael A. Perfater, Virginia Transportation Research Council

Contract Research Sponsored by
the Virginia Transportation Research Council

Virginia Transportation Research Council
(A Cooperative Organization Sponsored Jointly by the
Virginia Department of Transportation and
the University of Virginia)

Charlottesville, Virginia

December 2003
VTRC 04-CR10

NOTICE

The project that is the subject of this report was done under contract for the Virginia Department of Transportation, Virginia Transportation Research Council. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Virginia Department of Transportation, the Commonwealth Transportation Board, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

Each contract report is peer reviewed and accepted for publication by Research Council staff with expertise in related technical areas. The contractor performs final editing and proofreading of the report.

Copyright 2003 by the Commonwealth of Virginia.

ABSTRACT

The Virginia Department of Transportation (VDOT) has made significant investments in the traffic-monitoring infrastructure that supports intelligent transportation systems (ITS). The purpose of this infrastructure is to provide accurate, real-time information on the status of the transportation system; thus it is critical that the monitoring infrastructure provide accurate data. Although detectors are usually tested immediately after installation, it is well known that they operate in a very harsh environment and thus are susceptible to degradation in accuracy and/or complete failure. Consequently, a long-term commitment to data quality assurance is required through maintenance, data quality assessment testing, and repair/replacement.

The quality of data from ITS applications is becoming increasingly important as the data are more widely used. Not only is the data used in real-time operations, but also in myriad other, often more traditional, transportation applications. The purpose of this research project was to develop a procedure that VDOT could use to assess its ITS data quality. The report includes a data quality assessment procedure that is based on theory, practice, and empirical investigation. The procedure has the following key features:

- Benchmark data collection using temporary installation of non-intrusive detectors. This data collection technique provided the best approach to collect large quantities of validated data without disrupting traffic flow.
- Data quality assessed at the lane level to pinpoint problem detectors.
- Data quality assessed at the 1-minute (or minimum practical measurement interval) interval level to provide sufficient quantities of data in reasonable periods of time.
- Analysis techniques including both measures and plots, which provide quantitative and visual indications of data quality.

The report recommends that VDOT begin to use the procedure on both an ad-hoc basis and in a statewide program as a means of protecting its significant investment in ITS data collection.

FINAL CONTRACT REPORT

ITS DATA QUALITY: ASSESSMENT PROCEDURE FOR FREEWAY POINT DETECTORS

Brian L. Smith, Ph.D.
Assistant Professor
Department of Civil Engineering
University of Virginia

Ramkumar Venkatanarayana
Transportation Systems Engineer

Clinton D. Smith
Research Assistant

INTRODUCTION

Nearly all transportation management systems (TMSs) (referred to as Smart Traffic Centers, or STCs, by VDOT) employ the use of electronic surveillance equipment, such as inductive loop detectors, to monitor traffic. These surveillance subsystems play a critical role by sensing the “state” of the system, allowing personnel in the STCs to make informed decisions when managing the system. Furthermore, data from STCs—and other intelligent transportation systems (ITSs)—are now being used in a myriad of other, non-ITS applications. For example, the Virginia archived data management system (ADMS) is currently providing data from the Hampton Roads STC (HRSTC) to individuals applying it to applications such as environmental modeling, incident management, performance measurement, transit operations, and regional planning.

While surveillance equipment is generally rigorously tested when initially installed, experience has shown that the equipment frequently malfunctions or fails altogether. This can be attributed to the very harsh conditions that the equipment is subject to in the highway environment. Considerable research is focused on identifying detector-reported data that are clearly erroneous (hence indicating a detector malfunction). This research, for example, Turochy and Smith (2000) and Cleghorn et al. (1991), has developed a variety of data screening methods currently used by VDOT and other agencies. However, degradation of data quality from a surveillance subsystem often is not of such poor quality that a screening method can identify it as such. Therefore, given that VDOT and other agencies depend on the quality of this data, it is imperative that state transportation agencies establish sound data quality assessment procedures to ensure that the data collected by surveillance subsystems is not only reasonable, but accurate as well.

While the phrase “data quality” sounds simple on the surface, establishing a data assurance procedure is a complex undertaking. First, and foremost, is the need to collect an accurate, dependable benchmark data set that can be used to compare with data collected by the surveillance subsystem. Second, many details concerning the procedure must be addressed, such as the amount of benchmark data needed to support an adequate assessment and the types of data analysis techniques that should be used to support final conclusions. For each of the items listed above, there are an endless number of alternatives at the disposal of a traffic engineer.

Therefore, it should be made clear that the development of an indisputable, ideal, or optimal, data quality assurance procedure is simply not feasible. However, there is a need to carefully consider the details of such a procedure in order to develop one that effectively meets the needs of VDOT.

PURPOSE AND SCOPE

The purpose of the project was to develop an ITS data quality assessment procedure for use by VDOT. The scope of the project was limited to point detector-based surveillance subsystems used by ITS deployments on freeways (primarily STCs). In addition, the scope was limited to the two most widely used measures collected by point detectors—volume and speed.

METHODOLOGY

The following tasks were conducted in order to meet the objectives of this research effort. Please note that in the methodology, and the remainder of the report, data collected by the ITS detector under assessment is referred to as “field data.”

Review of Literature

A literature review was conducted to serve as a foundation for this project. The literature review focused on research and practice related to ITS data quality assessment.

Investigate Data Quality Assessment Procedure Alternatives

As stated in the introduction, transportation engineers have a wide variety of alternatives to consider when designing an ITS data quality assessment procedure. The purpose of this task was to explore alternatives in critical aspects of the procedure. These aspects are described below.

Benchmark Data

This subtask explored alternative methodologies available to collect a sound, validated set of benchmark data to use in a data quality assessment procedure. This was accomplished by identifying functional alternatives, and then assessing their effectiveness based on a core set of attributes required for benchmark data collection. These requirements include:

- *Ease of installation/set-up*—it is desirable to set up the benchmark data collection alternative at the site of the ITS detector in a timely fashion with minimal impact on traffic.
- *Ability to collect a large sample*—given that the traffic parameters volume and speed are random variables, comparisons between two sets of data must be based on relatively large data sets to minimize the impact of the natural variability of the data.
- *Ease of data validation*—benchmark data must be carefully validated to provide “true” conditions for comparison with field detector data. It is desirable to use an alternative that is relatively easy to validate.

Temporal Components

ITS data generally are collected at relatively short measurement intervals (usually somewhere in the range of 10 seconds to 2 minutes, depending on the system). A key aspect of an ITS data quality assessment procedure is whether or not to assess data quality at such short measurement intervals, or to aggregate measures to intervals as long as 15 minutes before comparisons between field and benchmark data. This consideration was explored by collecting sample benchmark and field data from the HRSTC and then examining comparisons using measurement intervals of 1, 5, and 10 minutes.

Spatial Components

Point detectors generally are installed in each travel lane at a particular directional location on a freeway. While the detectors collect data specific for each lane, it is common practice to aggregate the data across all lanes to report volume and average speed over all lanes (usually referred to as link or station data). When assessing data quality, a traffic engineer, therefore, can either compare data at the individual lane level, or at the station level. This subtask explores the impact of this decision by examining sample benchmark and field data from the HRSTC.

Data Analysis

In order to draw consistent, well-founded conclusions regarding ITS data quality, it is necessary to define clear data analysis procedures to support the comparison of field and benchmark data. This task explores data analysis techniques, both graphical and statistical, for

inclusion in the ITS data quality assessment procedure. As with previous subtasks, this question was explored using sample benchmark and field data collected at the HRSTC.

Document & Demonstrate ITS Data Quality Assessment Procedure

Based on the results of the earlier tasks, the purpose of this task is to define and document a complete ITS data quality assessment procedure for use by VDOT. In order to clearly demonstrate the procedure, a sample application of the procedure is illustrated using benchmark data collected by the Smart Travel Van and ITS field data collected on I-66 by the Northern Virginia STC.

RESULTS

Review of Literature

Despite the large investments that transportation agencies have made in real-time traffic surveillance systems, relatively little literature dealing with the topic of data quality assessment procedures exists. While there has been significant research in data screening techniques to identify infeasible detector values, these techniques provide only a “minimal examination of credibility” (Cleghorn et al., 1991). Generally, these approaches simply use lower or upper limits on reasonable values for different traffic parameters (Turner et al., 2000). Besides the papers and reports describing data screening alternatives, two key publications were identified that directly address data quality assessment. These are described below.

Texas Transportation Institute (TTI) researchers published a paper in 2000 detailing an accuracy assessment of ITS field data in San Antonio (Turner et al., 2000). In this case, they only considered traffic count, or volume data. The actual measure used was hourly volumes. Benchmark data were collected by manually counting vehicles from video recordings of the field site. Then, this data were compared with ITS field data using percentage error measures, and plots of field data on the x-axis and benchmark data on the y-axis. Finally, R^2 values for the linear relationship were used to quantify the level of “agreement” between the two data sets. While this study provides a good example of a data quality assessment procedure, the authors do not justify the decisions they made in terms of temporal or spatial aggregation, nor data analysis.

In the mid-1990s, the Federal Highway Administration, along with the Minnesota Department of Transportation and SRF Consulting Group, Inc. conducted an extensive test of non-intrusive traffic detection technologies (US DOT, 1997). To do so, the research team essentially was faced with the challenge of measuring data quality provided by the alternative technologies. As a result, the procedures developed for the test provide background for data quality assessment procedures explored in this study. In this study, inductive loop detectors were used to provide the benchmark data for comparison with the alternative detector technologies. The researchers noted that they placed considerable effort into validating the loop detector data given that it would serve as “ground truth” when evaluating the effectiveness of the alternative

detector technologies. Data from the loops were validated using a manual count procedure (volume) and by comparing speeds with those captured by a radar gun and probe vehicle (speed).

Given available benchmark data, the research team identified a number of metrics and tools to use in assessing data quality. These can be categorized as follows:

- Correlation coefficient
- Volume & speed scatter plots
- Root mean square error
- Percent differences

It is important to note that the research team in this case found that the high level of complexity involved in data quality assessment made it inappropriate to directly rank the alternative technologies considered. Thus, this illustrates the difficulty in deeming detectors as either “good” or “bad.”

Investigate Data Quality Assurance Procedure Alternatives

Benchmark Data

Based on the literature review and a critical assessment of technical and nontechnical options for benchmark data collection, four functional categories of alternatives were identified. The emphasis in this task was to assess functional alternatives rather than specific technical implementations. Clearly, nearly an infinite number of combinations of specific detectors and portable platforms exist and change on a frequent basis. By focusing on functional alternatives, the results of the research will be applicable for years to come.

Four functional alternatives were identified. They are described below:

Manual

The manual alternative consists of personnel being deployed to the same site as the field detectors and collecting benchmark data in a “manual” fashion. This technique has been used in the past primarily to collect count data for signal system timing optimization (i.e., turning movement counts, etc.). Speed data collection using the manual approach requires the use of a radar or lidar device (such as those used for speed enforcement). These devices are generally factory calibrated to ensure very accurate speed measurement of individual vehicles, on the order of +/- 1 mile/hour error (Kustom Signals, 1997).

Video Analysis

The video analysis alternative consists of videotaping traffic at the site of the field detector and then postprocessing the video to derive volume and speed data. The video can be collected either by existing STC closed-circuit television (CCTV) cameras or using a camera mounted temporarily at the site (e.g., on a nearby overpass). Volume data are derived manually simply by visually counting vehicles. Speed data derivation requires accurate time-based video

editing equipment to allow the collection of exact times when vehicles pass subsequent points of known distance. Thus, some level of field calibration (i.e., measuring distance between the reference points) is necessary.

Temporary Installation of Intrusive Detectors

In order to avoid the large manual component of data collection found in the previous two alternatives, one may choose to temporarily install detectors in the travel lanes at the same locations as the permanent field detectors being analyzed. These are referred to as “intrusive” detectors given that they are installed directly in the travel lanes. Examples of temporary intrusive detectors are temporary inductive loops and pneumatic tubes. Note that once installed, these detectors require a short-term manual calibration/validation process.

Temporary Installation of Non-intrusive Detectors

This alternative is the same as the previous approach, with the exception that it uses detectors *not* installed in the travel lanes (i.e., non-intrusive detection). Examples of non-intrusive detectors include video image vehicle detection systems, acoustic detectors, and radar detectors. Non-intrusive detectors must be mounted above or beside travel lanes. Thus, a key consideration for this alternative is the availability of a portable detector-mounting platform.

Given these four functional alternatives, the final challenge is to identify the alternative best suited to support an ITS data quality assessment procedure. As stated in the methodology, three key criteria are to be considered: ease of installation/set-up, ability to collect large samples, and ease of data validation. Table 1 rates each alternative on each criterion as poor, moderate, or good.

TABLE 1. Comparison of Benchmark Data Collection Functional Alternatives

Alternative	Criterion 1 Ease of Installation/ Set-Up	Criterion 2 Ability to Collect Large Sample	Criterion 3 Ease of Data Validation
Manual	Good	Poor	Good
Video Analysis	Moderate	Moderate	Moderate
Temp Intrusive	Poor	Good	Moderate
Temp Non-intrusive	Moderate	Good	Moderate

Upon examining Table 1, no functional alternative means of collecting benchmark data is clearly superior. Given that the temporary non-intrusive detectors and video analysis received no poor ratings, it is concluded that either of these alternatives be used. If technically feasible, the temporary non-intrusive detector alternative is preferred because of the ability to collect large

samples. However, the ability to secure an appropriate temporary platform for the sensor is a key issue.

The researchers found that an excellent platform for temporary non-intrusive detectors is a remote television van equipped with a telescoping mast and a portable power generation and distribution system. This type of van serves as the foundation for the University of Virginia and Virginia Transportation Research Council's Smart Travel Van, which uses a video image vehicle detection system as the non-intrusive detector. This system, shown in Figure 1, was used as the prototype benchmark data collection system for this research effort.



Figure 1. Smart Travel Van

Determination of Program Temporal Components

In order to address the temporal issue, benchmark and field data were collected at a number of locations monitored by the HRSTC. With this data, time series comparison plots were created at 1-minute, 5-minute, and 10-minute aggregation levels. One-minute aggregation served as the minimum level since it is generally the shortest level of aggregation used at STCs, while

5-, and 10-minute levels were considered since they are frequently used at STC's. Examples of these are provided in Figures 2, 3, and 4, respectively.

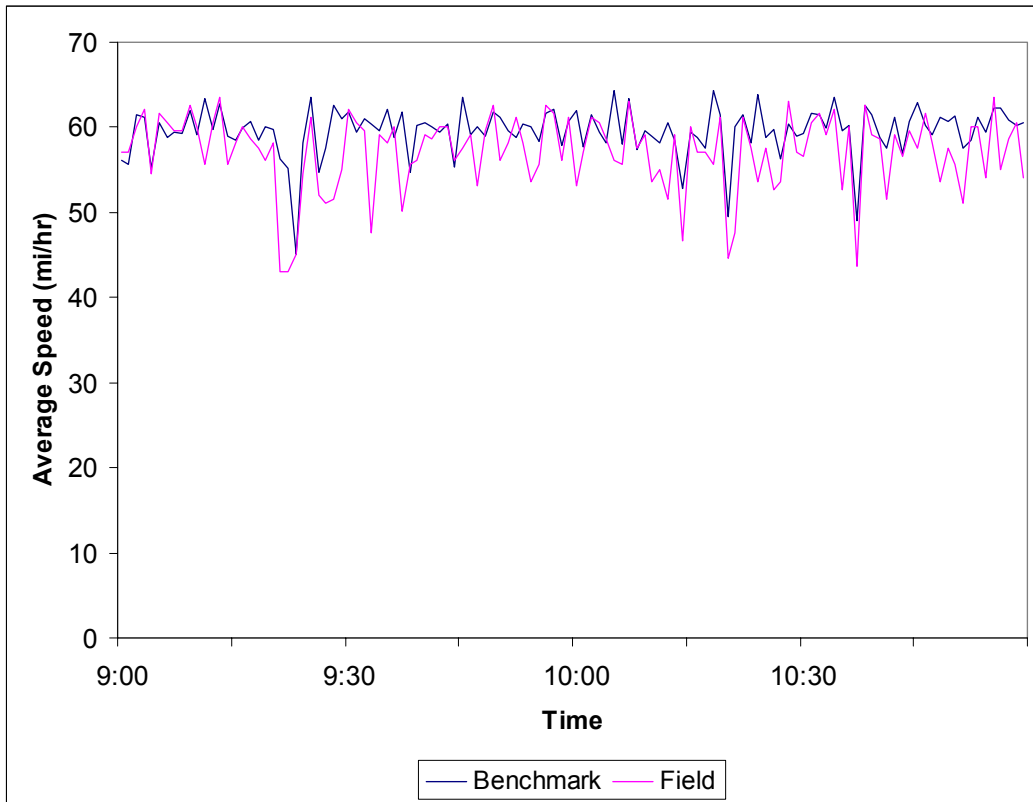


Figure 2. 1-minute Aggregated Average Speeds

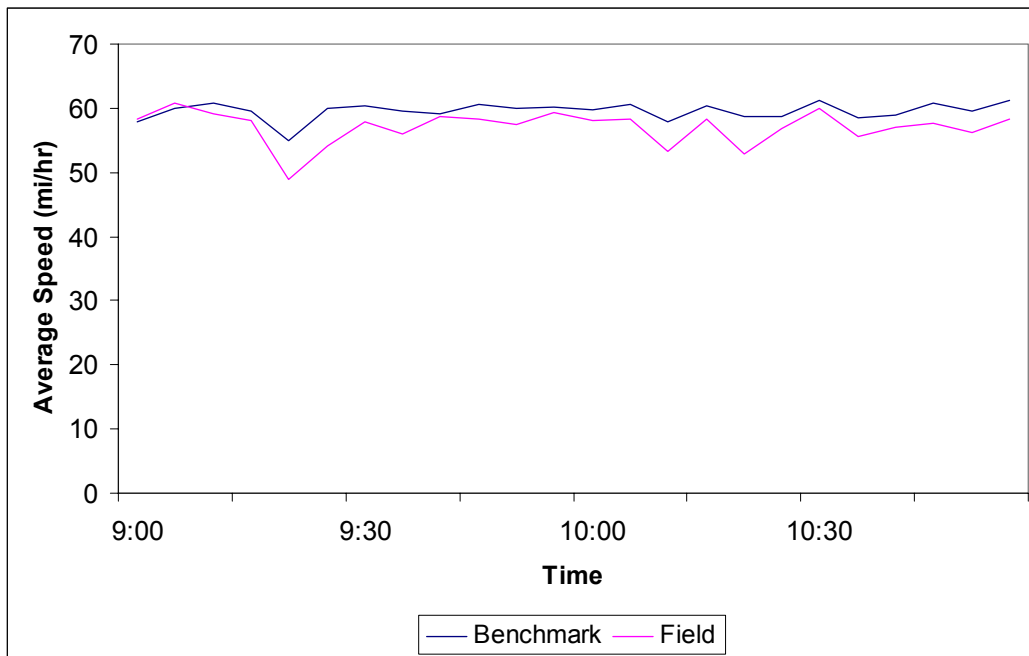


Figure 3. 5-minute Aggregated Average Speeds

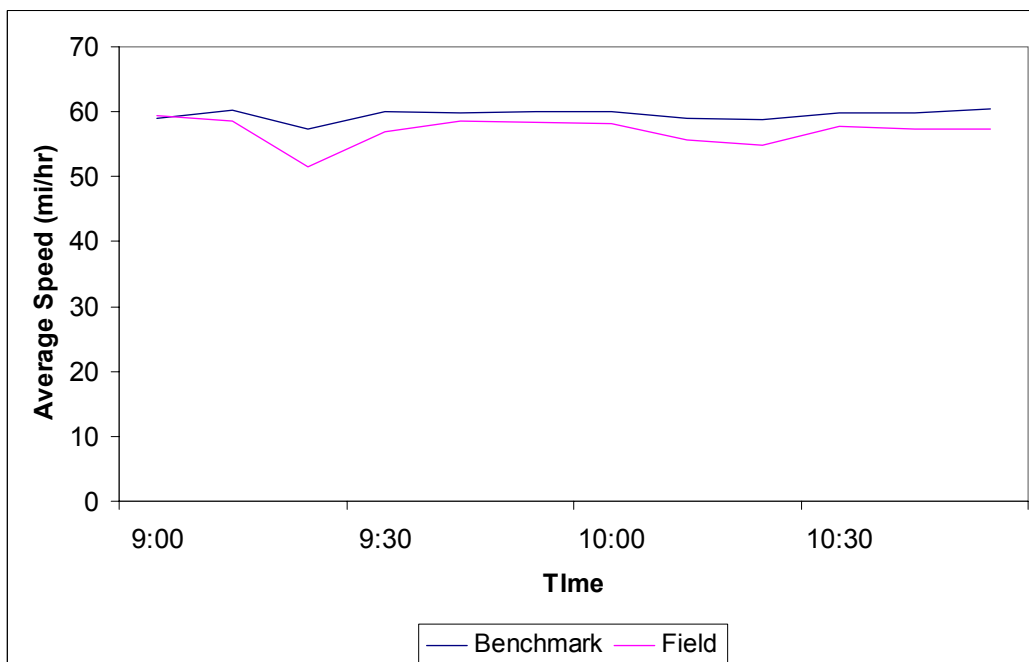


Figure 4. 10-minute Aggregated Average Speeds

Figures 2–4 are consistent with similar plots created for other HRSTC detector locations. When examining these figures, it is evident that as the aggregation interval becomes larger, overall biases in field detector data become more readily apparent. For example, Figure 4 indicates that at the 10-minute level, the field detectors (HRSTC detectors) are measuring average speed consistently 2-3 miles/hour below the actual average speed (benchmark data). However, when considering the 1-minute aggregation interval in Figure 2, it is clear that the bias is not entirely consistent at this finer level of temporal resolution.

Finally, a key consideration for the data quality assessment program is collecting sufficient samples to adequately account for the natural variability of the data when making decisions. By virtue of the central limit theorem, average random variables (such as the mean speed) can be considered to be normally distributed when 30 or more samples were available to compute the averages. Therefore, it is desirable to collect at least 30 independent samples of average speeds and volumes for both the benchmark and field data when assessing data quality. At the 1-minute level, one will collect 60 independent samples in 1-hour, whereas one will only collect 6 samples in an hour at the 10-minute level. Given this practical consideration, and the fact that the 1-minute level does not introduce a large amount of noise (as seen in Figures 2-4), it was decided that the data quality assessment procedure should aggregate data at a 1-minute level.

Determination of Program Spatial Components

In order to explore the level of spatial aggregation (i.e., lane vs. station), data from the HRSTC were examined using a common data screening approach. The commonly used Greenshield's model of traffic flow theory provides a basic relationship of macroscopic traffic flow measures (i.e., average speed, volume (or flow rate), and density) (May, 1990). These relationships are shown graphically in Figure 5. The concept behind this data screening approach is that data collected, both at the lane and station level, should indicate similar relationships as witnessed in the Greenshield's model. Finally, note that the measure "occupancy" collected by most ITS systems is an indication of the percentage of time that a sensor is "occupied" by a vehicle. It can be shown that occupancy is linearly related to the measure of density given the average length of vehicles at a location during the measurement period (May, 1990).

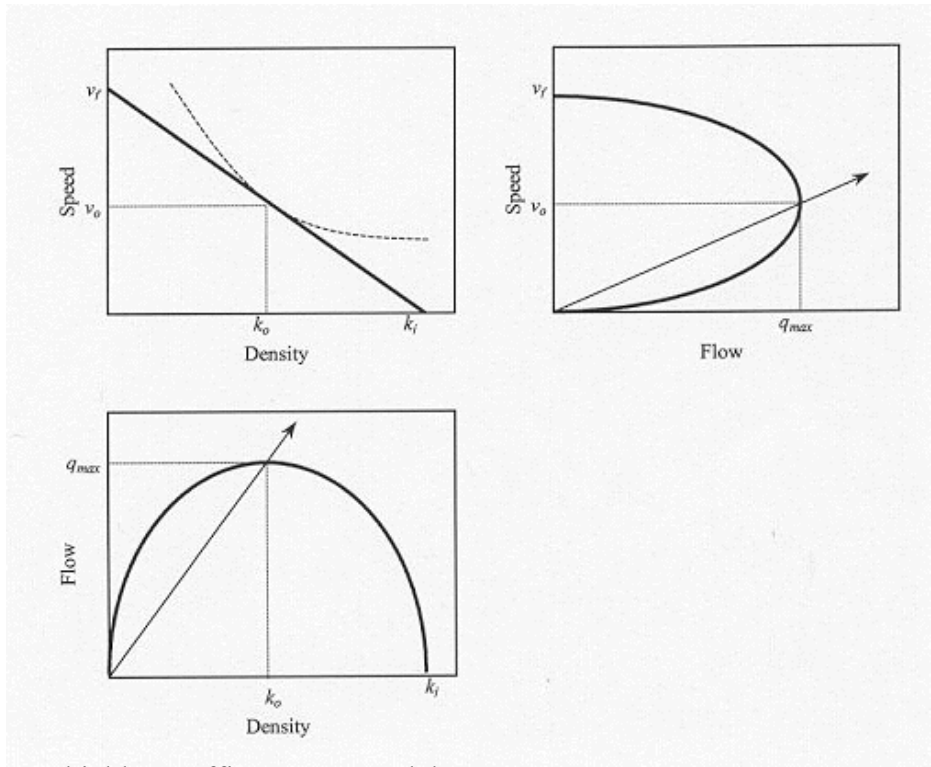


Figure 5. Greenshield's Traffic Stream Model

Figure 6 presents station level data for station 24, a four-lane section on I-264 monitored by the HRSTC. The two plots illustrate the volume-occupancy and speed-volume relationships. Carefully examining the speed-volume plot, one will note the parabolic shape expected from traffic flow theory. However, it appears that the link reaches capacity (a value of 2,500 vehicles/hour/lane—as value in agreement with theory) at speeds on the order of 25 miles/hour. This is in the realm of plausibility, yet is much lower than expected.

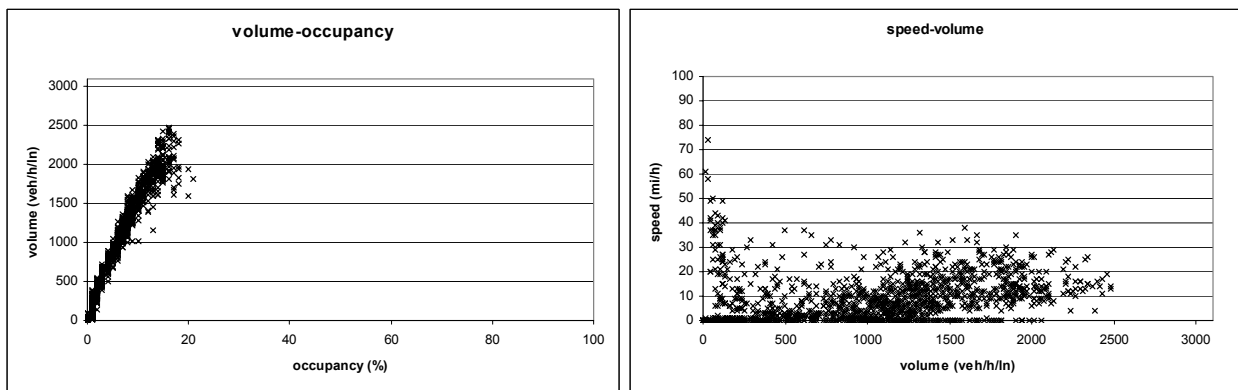


Figure 6. Station Level Data – Station #24, HRSTC

Figure 7 provides the same set of plots as Figure 6, but only for lane 2. One will note in this figure that the plots correspond nicely to what one would expect from the Greenshield's model. However, in this case, it is evident that speeds at capacity are on the order of 40–50 miles/hour. Thus, it appears that lane 2 data are at least reasonable. Furthermore, when one examines lanes 1 and 3, the plots are also similar to Figure 7 .

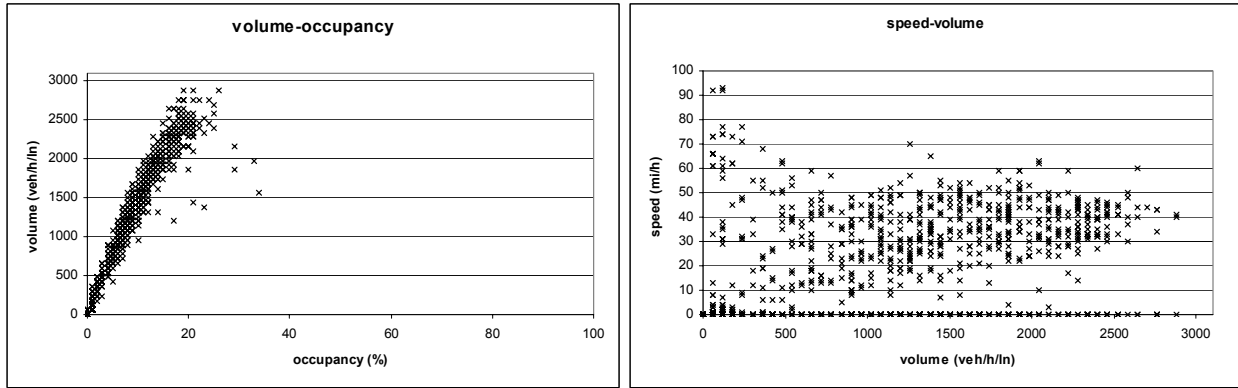


Figure 7. Lane 2 Data – Station #24, HRSTC

Finally, Figure 8 presents the set of plots for lane 4. It is clear from this figure that the detector in this lane is not functioning properly. Thus, from this example it is clear that errors in a single lane may be somewhat masked when considering data at the station level. Based on this assessment, it is concluded that the ITS data quality assessment procedure should be conducted on the individual lane level.

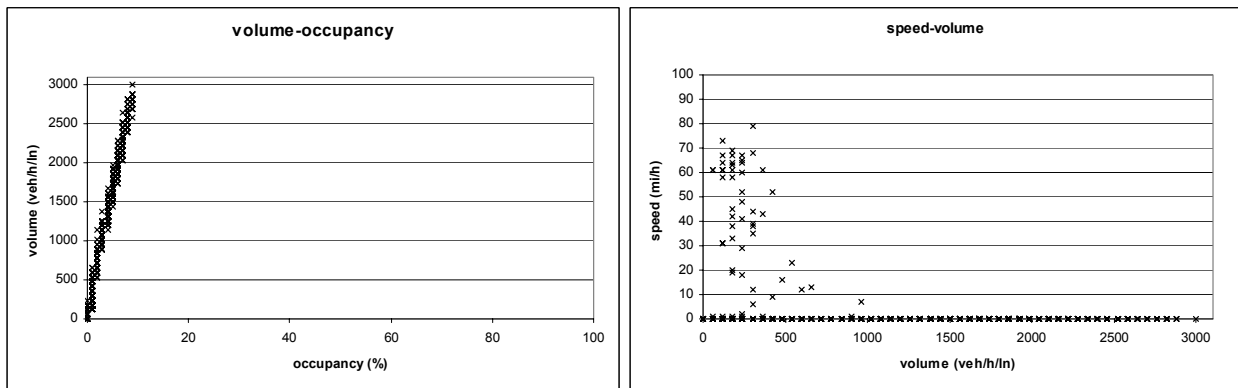


Figure 8. Lane 4 Data – Station #24, HRSTC

Data Analysis

Two basic approaches are available to compare the benchmark data to the field data in order to draw data quality conclusions. The first approach is to conduct rigorous statistical hypothesis tests to determine if the means of the two samples (benchmark and field) are statistically equivalent, representing “good” data quality. Note that this approach also may be applied to directly testing the difference in the means. The second approach is to compare the data on a more qualitative level using plots and less rigorous statistics. This section details the comparison of these approaches.

Hypothesis Testing

The advantage of hypothesis testing is that it provides a rigorous approach that directly accounts for the randomness in traffic data. However, hypothesis tests are very sensitive to small differences in means between two samples. Because of this, it is very likely (and was evident when considering data collected in Hampton Roads), that these tests will indicate poor data quality when, from a practical application perspective, the data are of acceptable quality. Furthermore, the proper interpretation of hypothesis test results requires a solid understanding of probability and statistics. It is likely that a large proportion of field personnel will not have this background. For these reasons, it was concluded that hypothesis testing was poorly suited for use in the data quality assessment procedure.

Qualitative Testing

The term qualitative testing is used in this research to refer to a combination of visual and numerical approaches intended to illustrate data quality (or lack thereof). The emphasis in qualitative testing is on practical illustration of quality, without focusing on rigorous statistical testing. Because of the practical, visual orientation of qualitative testing, this approach was selected by the research team for use in the ITS data quality assessment procedure. There are a myriad of combinations of graphics and measures that may be used in qualitative testing. This section describes those chosen for the procedure and provides the reasoning behind their selection.

First, two types of plots were selected for visual representation. The first, the time series plot, simply plots the measures of the benchmark data and field data on the y-axis and time on the x-axis. Examples of time series plots can be seen in Figures 2-4 earlier in the document. As seen in these figures, the time series plot allows one to quickly determine rough levels of “agreement” between the two data series and also is very effective in demonstrating bias in the field data. The second type of plot is referred to as the scatter plot. In this case, benchmark and field data are paired according to the time period of measurement. Then, the pairs are plotted with the benchmark data on the x-axis and the field data on the y-axis, as seen in Figure 9. Interpretation of this type of plot is very simple. Perfect accuracy of the field sensor will be evident with a straight line plotted at a 45-degree angle. The magnitude of deviation from this angle indicates the magnitude of inaccuracy. Furthermore, the Pearson correlation coefficient

corresponding to this plot also may be computed with a range of 0–1, with a value of 1 indicating perfect agreement.

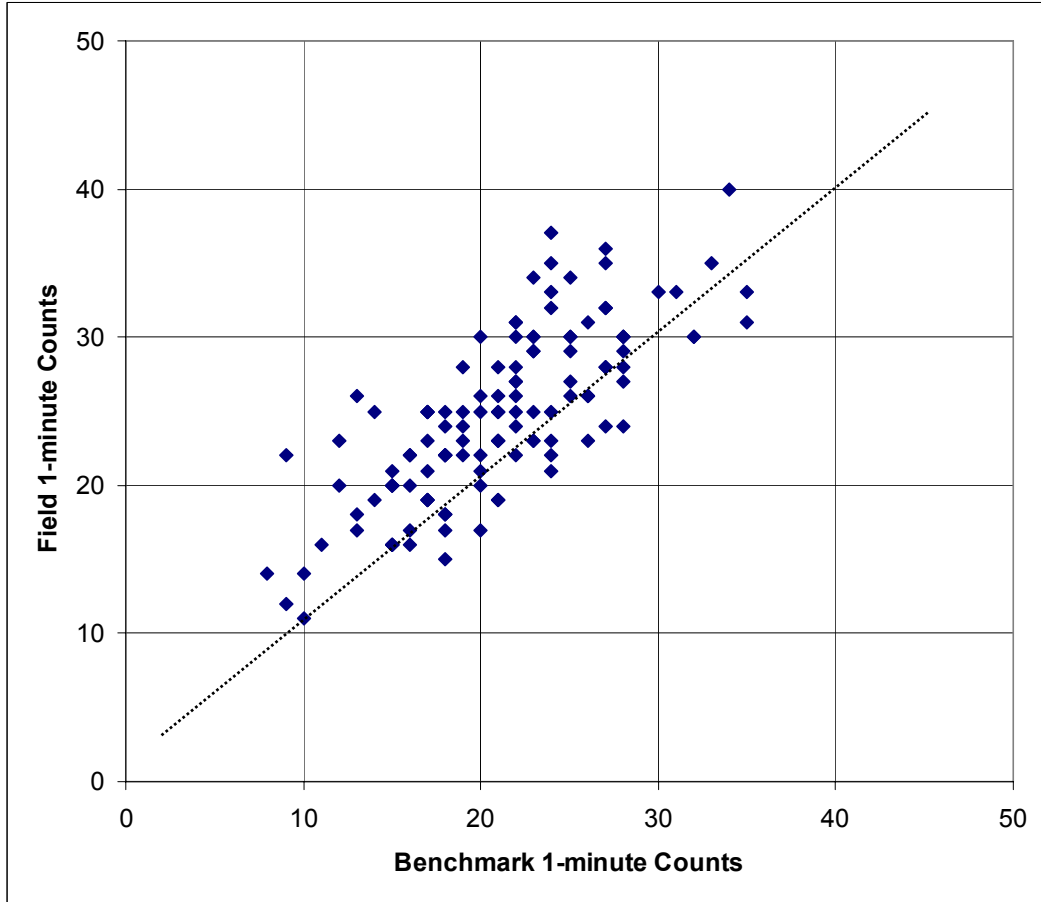


Figure 9. Scatter Plot - Volume

Finally, two different measures have been selected for use in the qualitative approach for analyzing volume and speed accuracy, respectively. In terms of volume, the simple absolute error measure works very well. It provides the engineer with the magnitude of error on a percentage basis, accounting for differing volumes at different locations (i.e., levels of traffic and numbers of lanes). For speed, however, small errors result in large percentage errors, which may be misleading (e.g., a 5-mile/hour error at a 30-mile/hour average speed location corresponds to a 17 percent error). To account for this, a truth table approach was used. In this case, each 1-minute field speed estimate is compared to the benchmark data to determine if the estimate falls within some threshold (i.e., error tolerance). This tolerance will be dictated by the uses of the data. For general ITS applications, it is recommended that a ± 2.5 -mile/hour tolerance be used. Then, one simply determines the percentage of field data measurements that fall within the tolerance. Again, a high percentage corresponds to a highly accurate sensor.

Document & Demonstrate ITS Data Quality Assessment Procedure

Based on the exploration of quality assessment procedure alternatives described above, an ITS data quality assessment procedure was developed. This procedure is documented below. Note that examples of the plots called for in the procedure are provided in the succeeding section in which the procedure is demonstrated using data from the Northern Virginia STC.

ITS Data Quality Assessment Procedure

1. Data Collection

- a. Install temporary non-intrusive detection device as close as practically possible to the field detectors under consideration.
- b. Make sure that the clock used by the benchmark data collection system (the temporary non-intrusive detection system) is synchronized with the clock of the ITS system polling and recording data for the field detectors.
- c. Validate the benchmark data manually using a radar/lidar speed detector and visual volume counts.
- d. Collect benchmark data by lane for a period of 1 hour, minimum, and aggregate data to 1-minute records (i.e., mean speeds per minute and volumes per minute for the entire data collection period).
- e. Query ITS system database to retrieve field detector data (by lane) for the same time interval. Aggregate this data to 1-minute records if necessary.

Note: If field detectors are polled at longer intervals than 1-minute (e.g., 5- or 15-minutes), simply aggregate *benchmark* data to the shortest possible interval corresponding to the field system. When this occurs, however, it is desirable to increase the length of benchmark data collection to provide a minimum of 30 independent volume and speed measurements for comparison.

2. Data Analysis

- a. Speed Analysis
 - i. Copy all of the benchmark and field detector speeds, from each lane, along with their appropriate times, into a spreadsheet for analysis.
 - ii. For both the benchmark and field data, average the lane speeds to acquire a station speed average for each minute.
 - iii. Compare field data with benchmark data
 1. *Time Series Graphs*: For all lanes and for the station being examined, plot simple comparative line graphs for the recorded speeds. Time lies on the x-axis, while average speed is plotted on the y-axis. The graph should consist of two series: benchmark data

and field data. One graph should be created for each lane, plus one for the entire station .

2. *Scatter Diagrams*

- a. Create simple scatter plots to directly compare average speeds collected at corresponding time intervals. Pair corresponding speed readings according to time (e.g., pair the 9:04 benchmark speed average and the 9:04 field speed average) into two separate columns. Plot the column with the benchmark data on the x-axis and the field data column on the y-axis. Note that for “perfect” agreement, this plot should appear as a straight line at a 45-degree angle.
- b. Using the spreadsheet function “PEARSON,” calculate a Pearson correlation coefficient between the two sets of data for each lane and the entire station.

3. *Truth Table*

- a. In the spreadsheet, create six columns: “Time,” “Benchmark,” “Field,” “Benchmark + 2.5,” “Benchmark – 2.5,” and “Pass.”
 - i. Under “Time,” list each minute of the data collection period.
 - ii. Under “Benchmark,” list each corresponding speed average according to the adjacent time.
 - iii. “Field” is the same as the “Benchmark” column, but with field detector data.
 - iv. “Benchmark + 2.5” and “Benchmark – 2.5” are calculations of the benchmark average speed, +/- 2.5 miles per hour.
 - v. “Pass” is a simple true/false statement, written with an (=IF(C1>E1,D1>C1)) statement. In this case, the field data “passes” if it is within +/- 2.5 miles/hour of the benchmark data.
- b. Once the true and false statements are determined, calculate “% True” with the statement
(=(COUNTIF(D:D,TRUE)/n)), where “D” is the “Pass” column, and “n” is the number of readings.

b. *Volume Analysis*

- i. Copy all of the benchmark and field detector counts from each lane, along with their appropriate times, into a spreadsheet for analysis. Note that the data should be in the units vehicles/minute. If desired, this can be converted to an hourly volume by simply multiplying the minute counts by 60.
- ii. For both the benchmark and field data, total the lane counts to acquire a station volume for each minute.
- iii. Compare field data with benchmark data

1. *Time Series Graphs*: For all lanes and for the station being examined, plot simple comparative line graphs for volume. Time lies on the x-axis, volume is plotted on the y-axis. The graph should consist of two series: benchmark data and field data. One graph should be created for each lane, plus one for the entire station.
2. *Scatter Diagrams*
 - a. Create simple scatter plots to directly compare volumes collected at corresponding time intervals. Pair corresponding volume readings according to time (e.g., pair the 9:04 benchmark volume and the 9:04 field detector volume) into two separate columns. Plot the column with the benchmark data on the x-axis and the field detector column on the y-axis. Note that for “perfect” agreement, this plot should appear as a straight line at a 45-degree angle.
 - b. Using the spreadsheet function “PEARSON,” calculate a Pearson correlation coefficient between the two sets of data for each lane and the entire station.
3. *Compute Absolute Percentage Error*
 - a. For each lane, and the entire station, compute the absolute percentage error for volume for the entire data collection period.

3. Drawing Data Quality Conclusions

Once the data has been collected and analyzed, it is necessary to reach final conclusions regarding field detector accuracy. The challenging aspect of this step of the procedure is that “acceptable” accuracy is not easily defined. In fact, it is dependent on the uses of the data. In some ITS applications, such as dynamic traffic assignment, very little volume error may be tolerated. While, in other general traveler information applications, significant speed errors are relatively inconsequential. The purpose of this section of the procedure is to propose some *general guidelines* that can be used to draw conclusions regarding data quality. These are appropriate when ITS data is used for general traffic monitoring and deriving system-level performance measures. Finally, note that it is possible that field detectors may be providing sufficiently accurate volume data, while also producing poor speed data.

- a. Speed analysis.
 - i. *Visually analyze the time series graphs.*
 1. Do the two series overlap?
 2. If not, do the two series follow the same trend, separated by some bias (i.e., – benchmark data is consistently 5 miles/hour lower than field data)? In the case of a bias, the field data may still be usable by removing the bias systematically.

3. If the two series do not overlap and do not follow the same general trend, the detector is producing erroneous data.
 - ii. *Consider Pearson Correlation Coefficients.* If the correlation coefficient is 0.70 or higher for the lanes and entire station, then the field detectors are sufficiently accurate.
 - iii. *Consider Truth Table.* If the percentage of “True” statements relating the field detector to an interval of +/- 2.5 mph of the benchmark reading is 70 percent or better, then the field detector is sufficiently accurate.
- b. Volume analysis.
- i. *Visually analyze the time series graphs.*
 1. Do the two series overlap?
 2. If not, do the two series follow the same trend, separated by some bias (i.e., – benchmark data is consistently 5 vehicles/minute lower than field data)? In the case of a bias, the field data may still be usable by removing the bias systematically.
 3. If the two series do not overlap and do not follow the same general trend, the detector is producing erroneous data.
 - ii. *Consider Pearson Correlation Coefficients.* If the correlation coefficient is 0.70 or higher for the lanes and entire station, then the field detectors are sufficiently accurate.
 - iii. *Consider Average Percentage Error.* If the average percentage error is 10 percent or less, the volume data produced by the field detectors are sufficiently accurate.

Demonstrate ITS Data Quality Assessment Procedure

In order to demonstrate the application of the data quality assessment procedure described above, benchmark and field data were collected on August 19, 2003, on Interstate 66 Eastbound in Northern Virginia, in the vicinity of the Route 28 interchange. At this location, I-66 is a 4-lane facility that experiences frequent congestion. The field sensors at this location are inductive loop detectors (ILDs) and the benchmark data were collected using the Smart Travel Van (STV) for a 90-minute period following manual benchmark data validation.

Speed Analysis

First, time series graphs were created. As seen in Figure 10, the graph for the station-level data indicates very good correspondence between the field data and benchmark data. Furthermore, there is no bias apparent in the field data.

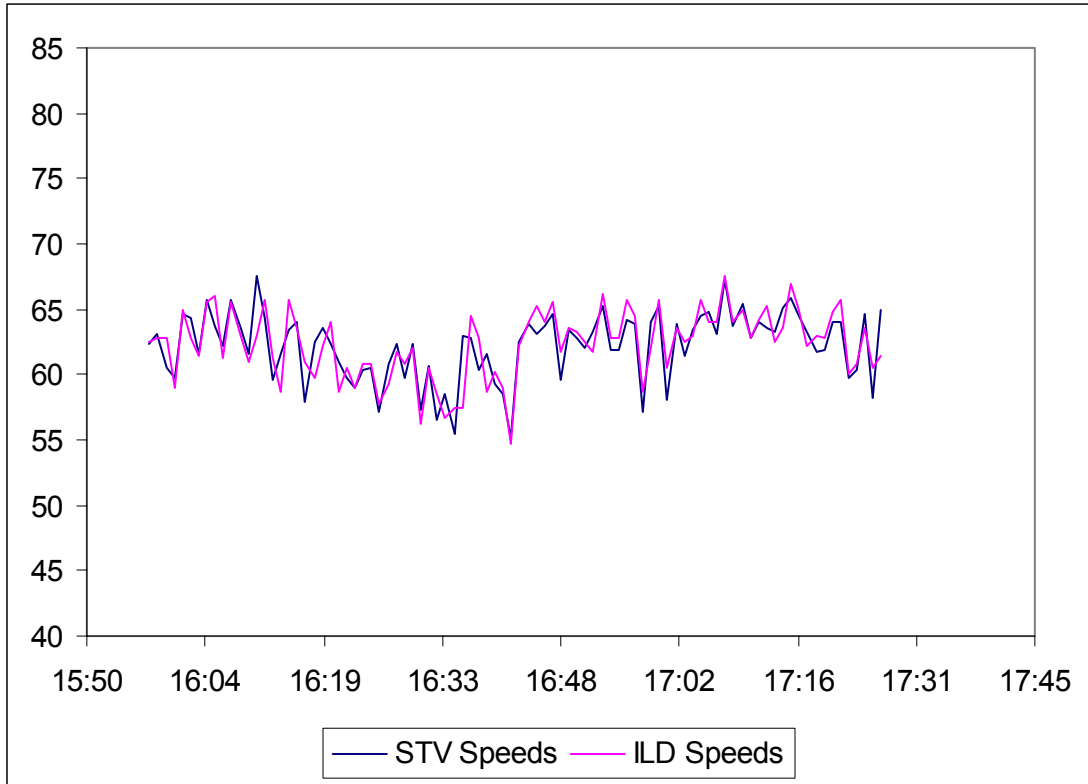
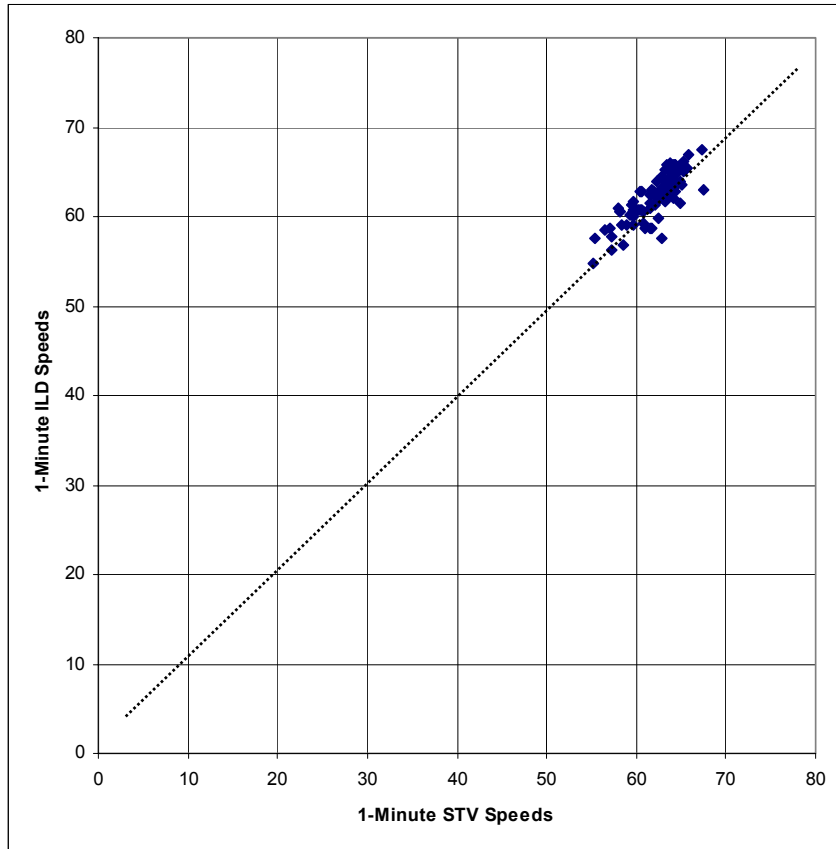


Figure 10. Time series graph of station average speed

Next, scatter diagrams were created for each lane to provide a visual representation of the correlation between the two data sets. Along with the diagram, the Pearson coefficient of correlation was calculated for each lane to measure the individual detectors' consistency with the benchmark data. The scatter diagram and Pearson coefficient for the station-level data are shown below in Figure 11.



Pearson Coefficient of Correlation: 0.8251

Figure 11. Scatter diagram of overall speeds, with Pearson correlation

The final step in the speed analysis—the calculation of the truth table—was applied. Table 2 below illustrates the table for each individual lane detector. One will note that this level of spatial resolution is certainly more descriptive than the more aggregate plots and measures illustrated in Figures 10 and 11.

Table 2 – Speed Truth Table

Detector, Lane	% Truth
427, Lane 1	70.0%
425, Lane 2	77.8%
423, Lane 3	85.6%
421, Lane 4	66.7%

Volume Analysis

As described in the data quality assessment procedure description section, the volumes were treated in a similar fashion as the speeds, but the overall volume (station-wide) was totaled from each lane, rather than averaged. Time series graphs were created for each lane, as were scatter diagrams and their associated Pearson correlation coefficients. The station level time series graph is shown in Figure 12.

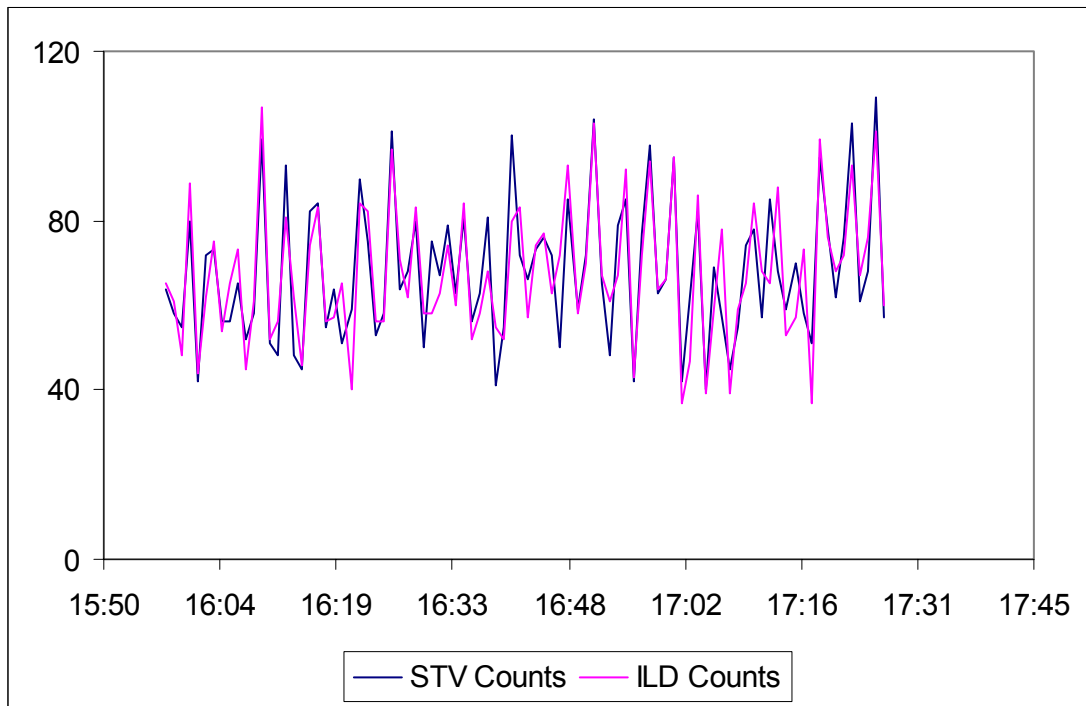
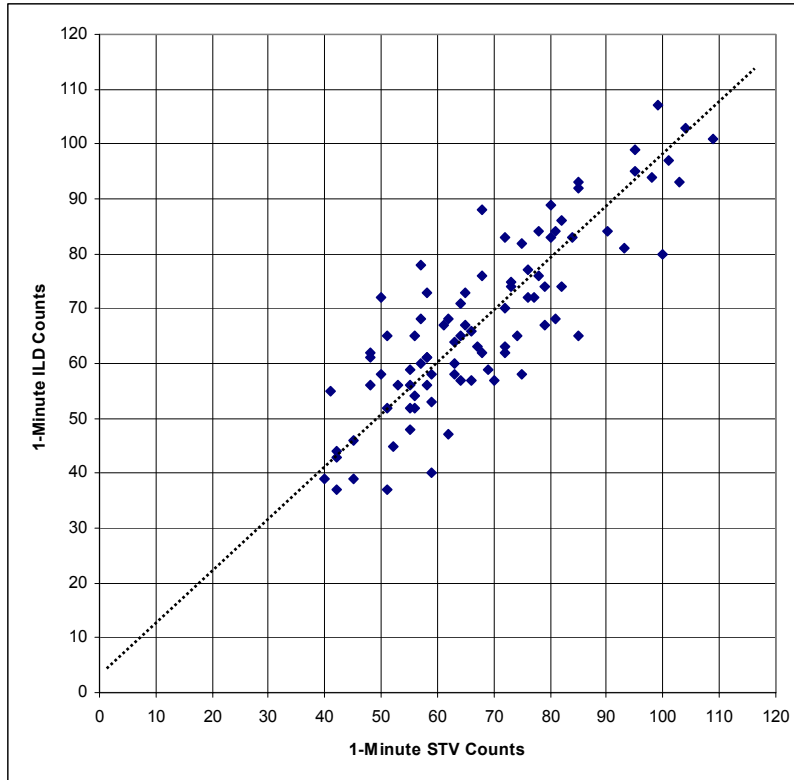


Figure 12. Time series graph of station volume

The station scatter diagram and Pearson coefficient are shown in Figure 13.



Pearson Coefficient of Correlation: 0.8492

Figure 13. Scatter diagram of volume, with Pearson correlation

Finally the station-level average percentage error was calculated at a level of 0.44 percent.

Quality Assessment

Using the criteria proposed in the previous section, one could conclude that the field detectors at this location are working as intended and that they may be deemed accurate. The one exception to this assessment is speed data from lane 4 (detector 421). As seen in Table 2, slightly less than 70 percent of average speeds from this detector are within 2.5 miles/hour of the benchmark data. This indicates that it would be beneficial to better tune/calibrate this single detector.

CONCLUSIONS

As evident in the results presented in this report, rigorously assessing the quality of data collected by ITS surveillance subsystems is a complex undertaking. First, a rich set of validated benchmark data must be collected. This benchmark data must correspond directly both temporally and spatially with the ITS sensor data being considered. Lastly, a sound set of analysis tools must be available to support conclusions regarding data quality. While this is complex, and consequently requires resources to implement, one must consider the alternative. VDOT has invested heavily in ITS surveillance subsystems since the 1980s. Without assurance of the quality of the data, this investment will yield no return for VDOT. Thus, the logical conclusion is that VDOT must routinely assess the quality of data collected by surveillance subsystems in order to protect its investment.

Beyond the general conclusions presented above, more specific conclusions may be drawn based on the results of this research:

- The lack of papers and reports addressing ITS data quality assessment indicates that this topic has not received sufficient attention in the transportation operations community.
- Collection of significant quantities of readily validated benchmark data is very difficult. As demonstrated by the results of this research, temporary installation of non-intrusive detectors provides the most advantageous approach to benchmark data collection.
- Field data should be compared to benchmark data at the minimal data collection interval possible, ideally 1-minute, in order to provide sufficiently large data sets within reasonable periods of time.
- The quality of point sensor data should be assessed at the lane, rather than the station, level.
- A combination of graphical plots and measures should be used as tools to compare benchmark data with field data. This will provide both qualitative visual indications of quality as well as quantitative measures.

Based on these conclusions, a data quality assessment procedure has been developed and demonstrated in this report that provides VDOT with a tool to use in maintaining its investment in ITS technology.

RECOMMENDATIONS

1. *VDOT systems that use point detectors on freeways (primarily STCs) should use the ITS data quality assessment procedure described and demonstrated in this report. STCs can immediately begin using this procedure to assess the quality of detectors that are either (a) high priority to regional traffic monitoring or (b) providing suspect data that may indicate a quality problem. (It is proposed that this recommendation be implemented immediately to serve as an interim measure while VDOT considers establishment of a routine data quality assessment program as described in recommendation #2).*

2. *VDOT's Smart Travel Program within the Mobility Management Division should investigate initiating a statewide ITS data quality assessment program.* The purpose of this program would be to routinely monitor data quality at all VDOT ITS deployments. While recommendation #1 is intended for as-needed quality checks, the purpose of this recommendation is to provide a program to support “preventive maintenance”, that is, identifying data quality concerns before they become a problem in applications. Note that this investigation should include careful consideration of the resources required for such a program. These resources would consist largely of labor and equipment needed for benchmark data collection.

3. *VDOT's Mobility Management Division should procure a mobile non-intrusive detection system (similar to the Smart Travel Van) to support ad-hoc data quality assessment as described in recommendation #1, and to use for the statewide data quality assessment program proposed in recommendation #2.* The research team found the VTRC/UVA Smart Travel Van to be an ideal resource to support data quality assessment. However, given that this van is fully committed to research, and taking into account the high level of use that the benchmark data collection system will experience, the procurement of a mobile non-intrusive detection system to assess data quality statewide is warranted.

REFERENCES

- Cleghorn, D., F. L. Hall, and D. Garbuio. Improved Data Screening Techniques for Freeway Traffic Management Systems. *Transportation Research Record 1320*, TRB, National Research Council, Washington, D.C., 1991.
- Kustom Signals, *ProLaser II- Infrared LIDAR system: User Manual*. Kustom Signals, Inc., Lenexa, Kansas, 1997.
- May, A. D. *Traffic Flow Fundamentals*. Prentice Hall, New York, 1990.
- Turner, S., Albert, L., Gajewski, B., and Eisele, W. Archived Intelligent Transportation System Data Quality. *Transportation Research Record 1719*, TRB, National Research Council, Washington, D. C. 2000.
- Turochy, R. E., Smith, B. L., New Procedure for Detector Data Screening in Traffic Management Systems. *Transportation Research Record 1727*. TRB, National Research Council, Washington, D.C., 2000.
- U.S. Department of Transportation. *Field Test of Monitoring of Urban Vehicle Operations Using Non-Intrusive Technologies*. FHWA-PL-97-018. Washington, D.C., 1997.