FINAL REPORT

ALTERNATIVE APPROACHES TO CONDITION MONITORING IN FREEWAY MANAGEMENT SYSTEMS

Rod E. Turochy, Ph.D., P.E.
Research Scientist

Brian L. Smith, Ph.D.
Assistant Professor
Department of Civil Engineering
School of Engineering and Applied Science
University of Virginia

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

In Cooperation with the U.S. Department of Transportation
Federal Highway Administration

Charlottesville, Virginia

January 2002
VTRC 02-R8

DISCLAIMER

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.

EXECUTIVE SUMMARY

Introduction

In response to increasing traffic congestion in major metropolitan areas, transportation agencies have built traffic management systems to facilitate the safe and efficient movement of people and freight. The Virginia Department of Transportation's (VDOT) smart traffic centers (STCs) in the Northern Virginia and Hampton Roads regions monitor traffic conditions, disseminate information to travelers, and operate traffic control devices to minimize the impact of traffic congestion on highway users.

Roadway surveillance is a key function of traffic management systems, and it is performed primarily through traffic sensors embedded in either the pavement or alongside the roadway that continuously collect data on traffic conditions. Several methods, such as status maps and incident detection algorithms, have been developed to extract useful information about traffic conditions from these data. In traffic management systems across the United States, incident detection algorithms constitute the most common approach to the automation of traffic condition monitoring. Despite more than 30 years of evolutionary development, the performance of these algorithms has not met expectations; their use has been minimized or eliminated in many traffic management systems. Therefore, an investigation into alternative approaches to using the traffic data being collected for traffic condition monitoring was warranted. This research effort examined possible automated methods for monitoring traffic conditions that would maximize the value of the traffic data collected in VDOT's STCs.

An expectation of current traffic conditions can be based on relevant conditions from the past (e.g., from the same time of day and day of the week). This expectation can intuitively be viewed as normal conditions; as current conditions deviate from this expectation, conditions are increasingly likely to be considered abnormal. By comparing current traffic data to a set of data that is historically relevant, a characterization of the extent to which current conditions are abnormal can be developed.

Method Development

Approaches to condition monitoring have typically not considered the interdependent, correlated nature of the commonly measured traffic variables of mean speed, occupancy (proportion of time that a detection zone is occupied by a vehicle), and volume (traffic flow rate). Therefore, statistical techniques that exploit these characteristics of the measured traffic variables were common to all approaches considered in this study. The use of multivariate statistical quality control (MSQC) as a means to compare current conditions to a set of historical data had the desirable properties of an alternative approach. Desirable properties were identified based on the findings of a survey of U.S. traffic management systems conducted in this research, a review of traffic management literature, and discussions with several VDOT traffic managers. Among these properties are its multivariate nature (accounting for the interdependent relationships among mean speed, occupancy, and volume), its ability to characterize conditions across a continuous range (rather than simply provide a binary response), and its assessment of system state based solely on the properties of relevant historical data (without externally applied thresholds). Concentric confidence regions (e.g., 90, 95, 99, 99.9 percent) can be drawn to classify future observations. Each confidence ellipsoid is centered on the mean of the historical data set and represents the estimated probability that a future observation will fall within it.

The application development environment and archived traffic data at Virginia's Smart Travel Laboratory were used to develop an MSQC-based approach to traffic condition monitoring. A prototype computer program was written to incorporate the basic calculations of MSQC and enhancements found in the literature and developed in this research. The basic calculations result in a value for T^2, the statistical distance between the mean of a set of historical data and a new observation of interest. To calculate T^2, a set of historical data relevant for comparison to the current observation must be generated. Several strategies for sampling the archived traffic data were considered; four strategies that consider past observations based on an analysis of traffic profiles were selected for development and evaluation. These strategies were the four possible combinations from two sets of time-of-day (TOD) criteria (the same TOD as the current observation and a 10-minute window centered on the TOD of the current observation) and two sets of day-of-week (DOW) criteria (same DOW as the current observation and all weekdays or weekend-days as appropriate). The resultant value of T^2 for the current observation (T^2_k) itself tells little about the extent to which traffic conditions are abnormal. Therefore, several enhancements that yield additional information were developed.

T^2_k is related to the exact confidence region on the surface of which it falls and, therefore, can be converted to a value between 0 and 1. The distribution of T^2 follows the F distribution and is related to a specific confidence region only by a function of the number of observations in the historical data set and the number of dimensions measured (in this case, three: mean speed, volume, and occupancy). As an example, if T^2_k falls on the 92.8 percent confidence region, from a statistical perspective, this indicates that there is a 92.8 percent probability that the current observation and the historical data sets to which it is compared come from populations with different means. The current observation can, thereby, be characterized as having a 92.8 percent chance of being abnormal, or a "normality level" of 0.928. Characterizing the extent to which current conditions are abnormal on a finite scale of 0 to 1 provides more information than the abstract value of T^2_k alone and allows for an intuitive understanding of the extent to which conditions are abnormal.

Data from the Hampton Roads Smart Traffic Center (HRSTC), VDOT's traffic management system in southeastern Virginia, have been archived at the Smart Travel Laboratory since June 1998. These data were used in the development and evaluation of the MSQC-based approach; a sample subset of these data was inspected for patterns of potentially erroneous data prior to program development. A procedure that incorporates data screening tests described in the literature and tests designed to address data deficiencies found in the inspection of HRSTC data was developed. Simple threshold-value tests that apply maximum acceptable values for volume and for occupancy were adapted from documented procedures, whereas tests that apply a minimum value for collection interval and that check for positive values of mean speed when volume is positive were inspired by inspection of the HRSTC data. Tests based on traffic flow theory principles that examine the average effective vehicle length, as derived from the measured traffic variables, and the maximum feasible volume when occupancy is reported as 0, were also developed in this research and incorporated in the screening procedure.

Findings

After successfully developing a prototype computer program that extracts traffic data from the database, screens the data, and performs the calculations that yield output measures such as the normality level, evaluation of the MSQC-based traffic condition monitoring method could commence. The purposes of the evaluation were to ensure that (1) the method characterizes traffic conditions in an accurate, intuitive, and easily interpreted manner; (2) the program can operate in a real-time traffic management system environment, and (3) the method can potentially serve supplemental purposes beyond detection of abnormal conditions, such as incident detection. Three categories of evaluation measures were developed: traffic condition assessment, prototype program operating performance, and incident detection. In traffic condition assessment, the impact of the four database sampling strategies, several historical data set sizes, and an outlier exclusion rule were examined. Program operating performance was evaluated to determine program efficiency and identify potential improvements for consideration in traffic management system implementation. The MSQC-based condition monitoring method is compared with one commonly used incident detection algorithm on a limited set of accident-related traffic data to ascertain the method's potential as an incident detection technique.

The state of traffic conditions can be captured in the normality level, directly related to the mean significance level of the confidence region on which any current observation falls. Theoretically, the mean normality level would be 0.5, but since the historical data sets are not exactly multivariate normally distributed and underlying factors that influence traffic patterns change with time, the resulting values are slightly higher than 0.5. Specifically, the mean normality level ranged between 0.58 and 0.63 among the four database sampling strategies. In the six pairwise comparisons among the four sampling strategies, differences in mean normality levels were statistically significant (at the $\alpha = 0.05$ level) in four cases. Although sampling strategy 2B (all weekdays or weekend-days, 10-minute interval) yielded the lowest value (0.5864), from a statistical perspective, there is no reason to prefer sampling strategy 2B oversampling strategy 2A (all weekdays or weekend-days, same 2-minute interval). In addition, differences between sampling strategies regarding the proportion of tested conditions (observations) that fall outside the 95 percent confidence region were examined; no statistically significant differences were found.

The impact of excluding observations that fall outside the 95 percent confidence region from the historical data set was also assessed. Excluding such out-of-control observations is intended to ensure that outliers or anomalies in the historical data do not affect the ensuing analyses. Conformance of the historical data sets with the multivariate normal distribution was significantly improved when the exclusion rule was applied, whereas negligible differences in measures of traffic condition assessment were found. Therefore, the exclusion rule was retained in the prototype computer program.

The operation of the prototype computer program was evaluated to assess its ability to operate effectively in an implementation setting. Mean program operating times were analyzed, and the impact of historical data needs was ascertained. Mean operating times ranged from 4.99 seconds (using sampling strategy 1B) to 6.33 seconds (using sampling strategy 1A), for an assessment of conditions at 10 locations, among the four sampling strategies. In the six pairwise comparisons, the difference in mean operating time was statistically significant between sampling strategy 1A (same day-of-week, same 2-minute time interval) and each of the other three strategies. Although these times were measured for program executions considering only 10 locations, HRSTC has approximately 200 detector stations. Linearly extrapolating the observed times to 200 locations for sampling strategies 1B, 2A, and 2B yields estimated operating times under 2 minutes, which is the time interval at which traffic data are collected. In addition, these operating times could be substantially reduced through using distributed computing systems (rather than running the program on a single computer), generating the historical data sets external to the program so that they could be accessed by the program (rather than querying the historical database in each execution of the program), and implementing other possible program coding improvements.

Regarding the historical data requirements among the four sampling strategies, sampling strategy 1A requires a substantially deeper historical database than do other sampling strategies since only one observation per week can be considered for inclusion in the historical data set. By contrast, sampling strategies 1B (same day-of-week, 10-minute time interval) and 2A may use up to five observations from each week, and sampling strategy 2B may use up to 25 observations from 1 week of historical data. This is reflected in the fact that about 23 percent of the program executions in which sampling strategy 1A was selected reached the beginning of the database before filling the historical data set, whereas the percentage of executions with insufficient historical data ranged between 0.7 and 3.6 percent among the other three sampling strategies.

The ability of the MSQC-based condition monitoring method to detect incidents (specifically accidents) was evaluated using a set of traffic data related to 54 accidents. Modified California Algorithm 7, a commonly used incident detection algorithm, was applied to the same data. Two measures commonly used to evaluate incident detection algorithms, mean detection time and detection rate, were calculated using the 95 and 99 percent confidence regions as thresholds in the MSQC-based method and using two sets of threshold values for the parameters in the incident detection algorithm. Mean detection time was approximately 5 minutes for the two MSQC-based method thresholds and one of the threshold sets for the modified California algorithm. Detection rates were higher with both MSQC-based method thresholds than with the incident detection algorithm using either set of thresholds. It should be noted that with an incident detection algorithm, it is typically recommended (although often not done in practice) that threshold sets be established based on a study site-specific conditions. Therefore, this comparison probably does not represent the best performance that Modified California Algorithm 7 may be capable of on the traffic data set studied.

Conclusions and Recommendations

An automated method based on multivariate statistical quality control and using archived traffic data can provide an informative assessment of traffic conditions. Several output measures that describe traffic conditions have been developed in this research. The most notable is the normality level (mean value of the confidence region) that describes the extent to which current conditions are abnormal on a scale of 0 to 1. Several strategies for using the historical data were evaluated; database sampling strategies 1B (same day-of-week, 10-minute time interval) and 2B (all weekdays or weekend-days, 10-minute time interval) appear to be most promising. A prototype computer program developed in this research can operate in a real-time traffic management environment and has potential for purposes beyond condition monitoring (such as incident detection). With further refinement, the prototype program should be suitable for a pilot study and, ultimately, implementation in HRSTC and other traffic management systems with sufficient archived traffic data.

Key recommendations resulting from this research include:

- *Develop the prototype program further.* The operation of the program and the utility of its user interface can be enhanced to include features such as a map interface and alternative database query structures to maximize performance.

- *Conduct a pilot study for implementation.* An enhanced version of the prototype computer program developed in this research could be tested and evaluated in HRSTC for use as a traffic management tool.

- *Improve the data collection process.* The manner in which traffic data, specifically mean speed and occupancy, are collected and aggregated could be changed to allow for more precise and accurate traffic data.

- *Conduct further research into additional program logic and output measures.* The MSQC-based traffic condition monitoring method could be enhanced through development of an adjustment for recent abnormal observations and through an empirical approach to defining abnormal conditions.

FINAL REPORT

ALTERNATIVE APPROACHES TO CONDITION MONITORING IN FREEWAY MANAGEMENT SYSTEMS

Rod E. Turochy, Ph.D.
Research Scientist

Brian L. Smith, Ph.D.
Assistant Professor
Department of Civil Engineering
School of Engineering and Applied Science
University of Virginia

INTRODUCTION

Traffic congestion in urbanized areas has steadily worsened for decades. This phenomenon creates costly delays for motorists, increases pollution, and can jeopardize safety. Traffic management systems have been constructed by transportation agencies in an effort to manage traffic through surveillance of traffic, response to traffic incidents, operation of traffic control devices, and dissemination of traffic information. Traffic management systems, particularly those that manage freeways, typically deploy sensors at various locations (often referred to as detector stations) along the roadway to collect traffic data. These data can then be processed to derive information about the state of the transportation system. An assessment of system state allows for responsive actions to be taken by traffic managers to attempt to return the system to a state of control, or to normal conditions.

Two critical functions of traffic management systems are surveillance of traffic (i.e. data collection) and condition monitoring (i.e., data analysis). Traffic surveillance is performed using detection devices, such as inductive loop detectors and video cameras. The state of traffic conditions is then determined through analysis of the data received. Traffic condition monitoring methods attempt to estimate the state of the highway system, based on surveillance data, thereby providing information that traffic managers can use to make decisions to manage the system.

Current State of Traffic Condition Monitoring

Video cameras, status maps, and incident detection algorithms are common methods used to monitor traffic conditions. Video cameras (also referred to as closed-circuit television) provide traffic managers with an easily interpreted visual image of traffic conditions. However, as traffic management systems continue to expand in size and scope, this method of assessing system state becomes increasingly labor-intensive. A status map is essentially a graphic representation of traffic data, typically in the form of a map with highway segments color-coded according to a particular traffic variable (most commonly mean speed). Although status maps are based on collected traffic data, substantial human resources are still required to monitor and interpret them. Incident detection algorithms typically present traffic managers with a yes or no response regarding the suspected occurrence of a traffic-disturbing incident based on an interpretation of collected traffic data. Such algorithms analyze traffic data and thereby automate the interpretation of traffic conditions. However, for a variety of reasons ranging from difficulties in calibration and maintenance of these methods to erroneous data provided as inputs, performance of these methods has been below expectations. Poor performance and the availability of other detection technologies (such as cellular telephone calls from motorists) have relegated incident detection algorithms to a secondary status in many systems. In fact, several traffic management systems have discontinued use of incident detection algorithms altogether.[1]

These methods merely assist in the monitoring and analysis of traffic conditions; their outputs still require the judgment of traffic managers regarding their relevance and reliability prior to using the information to make decisions regarding the management of traffic. Further strain will be placed on the efforts of traffic managers to monitor conditions as the size and complexity of traffic management systems expand at a greater rate than available operational staff.[2] As copious amounts of data are collected and archived in many of these systems, substantial effort is required to extract information from these data. Constraints on human resources and the extensive, ongoing investment in traffic monitoring infrastructure both provide motivation to further automate condition monitoring methods and find ways to improve upon those currently in use.

Of paramount importance in accurate operation, and ultimate acceptance, of any condition monitoring method is the quality of the input traffic sensor data. From these data, information about the state of the highway system can be obtained. The performance of condition monitoring methods is directly linked to the quality of the traffic data supplied to the method. The decisions made by personnel at traffic management systems are often based on the output of such methods. As many efforts to develop condition monitoring methods have used simulated traffic data, the issue of data quality has often been overlooked. A recent effort sponsored by the Federal Highway Administration to examine the current state of development and testing of traffic condition monitoring tools notes that "data validation is [to be] addressed as an explicit processing step that is carried out prior to the application of incident detection algorithms."[3] Therefore, screening of the data prior to their use is a critical component in the automation and accurate operation of the condition monitoring function.

A New Approach

Condition monitoring methods have historically focused on attempting to detect incidents that disrupt traffic flow to an extent that consequential changes in the traffic data occur and can be observed. By using methods that automate the interpretation of traffic conditions based on the historical distribution of traffic data, a different perspective can be used to evaluate traffic conditions. Instead of producing a yes or no result regarding the occurrence of an incident, as is typically the case with incident detection algorithms, a statistical analysis of traffic data allows for the focus of traffic monitoring methods to be placed on the detection of abnormal conditions. Rather than attempting to detect a discrete event (incident) whose occurrence *may* be reflected in the surveillance data (as with incident detection algorithms), statistical methods such as multivariate statistical quality control (MSQC) can be used to evaluate the state of the transportation system over a continuous range of conditions, rather than simply classifying system state into one of two categories. Methods that can perform a comparison of current and historical data to characterize current conditions as related to a definition of normal conditions based on the historical data are considered for development. Thus, the focus of such a method is to use relationships among the data to describe the state of the system rather than to make a decision as to whether a particular discrete event has occurred in the system. In addition to evaluation of conditions across a continuous range, the desired method will not only use all measured traffic variables in this assessment but also take into account the inherent relationships among these variables, thereby making greater use of the properties of the data than typically found in traffic condition monitoring methods today.

Motivation for the use of methods that analyze traffic data and employ a fresh approach to assessing traffic conditions comes from many sources. The increasing investment in intelligent transportation systems, particularly in the surveillance elements of traffic management systems, indicates that significant benefits in terms of traffic flow, congestion reduction, and safety are expected of these systems. In addition, recent research into the current and future states of the practice of traffic condition monitoring indicates the potential for advanced statistical techniques to make positive contributions as described later.

Traffic congestion, which occurs as traffic volumes approach the physical capacity of a highway resulting in significantly reduced speeds and increased travel times, can be described as either recurrent or non-recurring. Recurrent congestion typically occurs in a predictable manner at the same times of day and locations in a highway network attributable to the relationship between traffic volumes and highway capacity. Non-recurring congestion occurs when traffic volumes approach capacity at times and locations atypical on the highway network, as may be caused by special events or reductions in highway capacity caused by incidents such as crashes or highway maintenance activities. Recurrent congestion, therefore, can be considered normal, as its presence occurs regularly and expectedly. In a recent report on the development of incident detection algorithms sponsored by the Federal Highway Administration, the chief developer of the widely used California algorithms suggested that statistical approaches to classifying traffic conditions deserve further study. Among the recommendations for further research was an approach in which "traffic zones are labeled over an extended period of time to form a database for statistical analysis that would lead to a characterization of recurrent congestion."[3] Recurrent congestion is a normal condition, as defined in this research, whereas the traffic data collected and archived in many traffic management systems today can constitute the database alluded to and used in this research.

In reference to improving upon the typical output of incident detection algorithms, Payne stated: "a probabilistic approach to classification reflects the input data characteristics more accurately than the conventional binary output and also allows for a degree of certainty to be incorporated into the results, thereby indicating the appropriate level of attention by traffic management center operators."[3] Essentially, a probabilistic, statistically based approach to monitoring traffic can provide more information than incident detection algorithms and describe system state over a continuous range rather than either as incident or non-incident. The concept of multivariate statistical quality control, as applied and enhanced in this research effort, provides such an output.

PURPOSE AND SCOPE

The purpose of this research was to develop and evaluate an effective method prototype that monitors traffic conditions and provides an assessment of system state across a range of conditions in the context of an expectation of current conditions as based on relevant historical data.

In this approach, an assessment of system state across a range of conditions (normal and varying degrees of abnormality) was to be based on the expected state of the system as defined using historical traffic data.

METHODS

The methods employed in this research included a literature review; an investigation of current traffic condition monitoring practices and future needs; identification of selection of candidate approaches; collection, inspection, preparation and screening of traffic data; application and enhancements to the selected approach; and evaluation and comparative analyses.

Task 1: Literature Review

A literature review was conducted to determine the extent of previous research in traffic condition monitoring. Areas of focus included:

- traffic management and traffic monitoring (primarily freeway management systems, traffic condition monitoring methods, and incident detection algorithms)

- statistical monitoring and classification methods (primarily on univariate and multivariate quality control, condition monitoring methods, confidence ellipses/tolerance regions)

- traffic data (collection techniques, data screening and quality assurance, traffic flow theory and relationships among traffic parameters).

Task 2: Investigation of Current Methods and Future Needs

An investigation of condition monitoring in freeway management systems was conducted to identify methods of data collection and condition monitoring methods currently in use and to assess the performance of these methods. This investigation highlighted trends in traffic monitoring practices and opportunities for improvement. The investigation consisted of discussions with officials of the Virginia Department of Transportation (VDOT) and a survey of freeway management systems in the United States. The survey provided insight into trends in staffing and resource availability, the current state of condition monitoring, and the current and future size and scope of freeway management systems and expected improvements.

Task 3: Identification and Selection of Candidate Approaches

In this task, general requirements and desirable statistical properties for traffic condition monitoring methods were identified. Findings from the literature review, investigation of current practice, and discussions with several VDOT traffic managers highlighted both desirable properties for automated traffic condition monitoring and the resulting assessment of system state based on shortcomings of existing condition monitoring methods. Chief among these properties were (1) accounting for the relationships among the measured system variables by treating them as a set rather than independent, unrelated variables, (2) providing an assessment of current conditions across a continuous range rather than a binary yes/no assessment, and (3) evaluating current conditions based on properties of the historical data themselves, rather than requiring external interpretations or the application of threshold values. Many data-driven statistical methods were considered, including cluster analysis, discriminant analysis, and MSQC. MSQC was the method that best met the described properties and, therefore, was selected for development. Using MSQC, the current multivariate observation can be compared to the mean of a set of relevant historical observations, and the statistical distance between these points can be calculated and used as a measure of deviation.

In addition to identifying MSQC as the preferred statistical method and developing its application and enhancements, the organization of the historical database and methods by which it will be sampled to obtain the relevant historical data for a current set of conditions are critical to the operation of the condition monitoring methods developed. The manner in which relevant data are selected affects the operating performance of the methods, as well as the basis for determining the extent to which current conditions are normal or abnormal. A set of strategies for retrieving a set of relevant historical data from the traffic database was developed.

Task 4: Collection and Inspection of Traffic Data

Traffic management systems typically collect data on traffic conditions in an automated, continuous manner. Data from the Hampton Roads Smart Traffic Center (HRSTC), a freeway traffic management system in southeastern Virginia, were used in this research. These data are archived in Virginia's Smart Travel Laboratory, a facility jointly supported by VDOT and the University of Virginia. Sample sets of the data were visually inspected to examine their format and determine modifications and additional calculations necessary to use data in a condition monitoring method. Information pertaining to the traffic parameters measured, the temporal and spatial properties of the data, and the organization of the traffic and incident databases were obtained in this effort.

Task 5: Preparation and Screening of Traffic Data

An inspection of the data and examination of their plausibility and consistency with traffic flow theory principles was undertaken. It quickly became evident that a procedure was necessary to screen, or filter, the data to detect and remove data suspected to be erroneous. In addition, no quality testing or screening was performed on the data prior to their receipt or archiving in the Smart Travel Laboratory. A procedure to screen the data for quality control purposes was developed as a part of this research; a review of data screening literature and a visual inspection of sample data assisted in its development.

Task 6: Application and Enhancements to the Selected Approach

Development of MSQC for use in traffic condition monitoring consists of applying the basic MSQC calculations and traffic variables to compare current data with historical data and then developing extensions beyond the basic calculations that can provide additional information on current conditions. A prototype computer application was developed to perform the necessary calculations. In addition to employing the basic MSQC calculations in traffic management, several enhancements to the basic calculations were developed to provide more information for diagnosing current traffic conditions, including adding a calculation to determine the extent to which current conditions deviate from the norm (resulting in normality level) and a characterization of system state across sets of conditions representing degrees of abnormality. In addition, a decomposition of T^2 (the performance measure in basic MSQC) developed in another field was applied to provide information on the extent to which each of the measured traffic variables contributes to T^2. Within the context of the prototype computer application that performs all of the computations noted, these extensions are identified as output measures, each of which provides information to a traffic manager beyond the value of T^2, which, as a stand-alone value without context, is of relatively little use.

Task 7: Evaluation and Comparative Analyses

After developing the MSQC-based traffic condition monitoring method, the prototype computer application was evaluated using three categories of performance measures that pertain to characterization of the observation of interest (current conditions), prototype program performance, and incident detection. Statistical analyses to determine the significance of differences in evaluation measures among the database sampling strategies were conducted. The results of these analyses were used to develop recommendations for implementation.

The manner in which the MSQC-based method characterized traffic conditions was evaluated. Specifically, the distribution of the values of the output measures from several thousand executions of the prototype program was examined and compared among the database sampling strategies. The impact of each strategy on the probability that conditions would be classified as abnormal was determined. The traffic management application of the prototype program was used to perform this evaluation.

The operating characteristics of the prototype program were evaluated. Running time and its relationship to database sampling strategy was examined. In addition, the impact of insufficient historical data coverage on each sampling strategy was assessed. The traffic management application of the prototype program was used to perform this evaluation.

The capability of the MSQC-based method to perform incident detection functions was the third evaluation technique. Only incidents that produce abnormal conditions are detected; however, such incidents will also be those that have the greatest impact on traffic and most merit the attention of traffic managers. Among the measures of effectiveness (MOEs) commonly applied to incident detection algorithms, detection rate and mean time to detect were used to evaluate each method. Since abnormal conditions are caused by non-recurring traffic congestion, regardless of whether an incident has occurred, the false alarm rate was not used in the evaluation. The typical definition of *false alarm rate* would classify all such occurrences as false alarms, thus misrepresenting the effectiveness of the candidate methods. An incident log maintained by HRSTC was searched to find the time periods in which traffic was affected by incidents at the stations used in the analysis. With this information, the MSQC-based method and a commonly used incident detection algorithm were evaluated and compared.

LITERATURE REVIEW

Historical Development and Types of Traffic Condition Monitoring Methods

Freeway management systems date to the early 1960s, when the Lodge Freeway in Detroit and the Eisenhower Expressway in Chicago were outfitted with ramp metering, inductive loop detectors, and other tools to monitor and control traffic. The data collected in previously unforeseen quantities provided a source of information for efforts to improve traffic monitoring. Early efforts at using the data to monitor traffic and detect flow disturbances were made in Houston and Los Angeles in the late 1960s and early 1970s.[4,5] On the Gulf Freeway in Houston, a method of analyzing values for mean speed and flow using control charts was employed in a "freeway control system." In the Los Angeles area "freeway surveillance and control project," a decision tree poses a series of yes-or-no questions concerning traffic data and assesses the likelihood that an incident may have occurred. These efforts in automated condition monitoring are respectively the beginnings of the statistical and decision-tree families of incident detection algorithms. The California algorithm devised in the late 1960s for the Los Angeles system, along with subsequent modifications, is generally considered to be the most widely used family of algorithms in use today.

Developments in communications technology have led to a growing range of techniques being applied to the condition monitoring problem. A wide range of tools is available to traffic managers today to assist in the monitoring of traffic conditions. Video cameras (closed-circuit television) provide traffic managers with an easily interpreted visual image of traffic conditions. A status map is essentially a graphic representation of traffic data with highway-segments color-coded according to a particular traffic variable. These tools require significant human resources to thoroughly monitor traffic and therefore offer little automation of the condition monitoring function. Incident detection algorithms analyze the data themselves and therefore provide a greater level of automation than do status maps and video cameras. Because of the computational nature of incident detection algorithms and the prevalence of their use, a significant body of literature has evolved regarding incident detection algorithms. Although the computational power was available to allow these techniques to examine data from across entire freeway management systems within a few minutes, the complexity of these techniques limited their adaptability by requiring a greater calibration effort when compared with the older yet simpler statistical and decision-tree algorithms. Despite these advances in analysis techniques, the performance of, and level of satisfaction with, incident detection algorithms still leaves much room for improvement.[2]

Statistical Techniques Used in Traffic Condition Monitoring

A few incident detection algorithms use data-based techniques in their operations; all of these techniques employ some form of statistical concepts. Four such algorithms are an unnamed method developed in Texas in the late 1960s, the standard normal deviate (SND) model, the Bayesian algorithm, and the Advance algorithm. The first two of these algorithms rely primarily on statistical techniques, whereas in the Bayesian and Advance algorithms, statistical techniques play supporting roles. Each of these algorithms employs statistical approaches to condition monitoring.

The first documented statistically based condition monitoring method was developed in a late-1960s research effort sponsored by the U.S. Bureau of Public Roads and undertaken at Texas Transportation Institute.[4] At the time, systems were operational only in Chicago, Detroit, and Houston; data from the Gulf Freeway in Houston were used in this case. In this effort, control charts were used to detect deviations in 1-minute flow rates aggregated from 15-second data collection intervals. For each detector station, a running 5-minute average of the 1-minute volumes, in vehicles per minute, was plotted, with control limits drawn at two standard deviations above and below the running average. If the most recent 1-minute flow rate (taken as the sum of four consecutive 15-second counts) fell below the lower limit for two consecutive 15-second periods, it was said to "indicate a capacity reduction of significant size."[4] This method, essentially an incident detection algorithm, was deployed on only a limited set of peak period data from one location, thus demonstrating the possibility that it could detect incidents, but it was not fully tested or evaluated. Although this algorithm was not comprehensively evaluated, it does represent the first effort to apply statistical quality control (SQC) using one variable to traffic condition monitoring.

The SND model for incident detection was developed in Texas in the early 1970s and evaluated off-line using data from the Gulf Freeway in Houston.[6] The SND was the critical parameter used to attempt to detect incidents. The SND was defined as SND = $(x_{n+1} - \bar{x})/s$, or the difference between the current value for average lane occupancy and its mean over n intervals immediately preceding the current, then divided into the standard deviation over the n intervals. Traffic data were collected in 60-second intervals. The mean and standard deviation were calculated over the previous three 1-minute intervals (n=3) and previous five 1-minute intervals (n=5). Two schemes for declaring an incident alarm were studied: one in which the first SND value greater than the critical SND value was detected, and one in which two successive SND values were greater than the critical SND value. As the critical SND value increased, both the rate of detection of incidents and the false alarm rate decreased. Essentially, this algorithm examines the rate of change in successive measurements and compares it to observations immediately prior to the current, declaring an incident when the rate of change is out of line with recent measurements. As with the method developed by Whitson et al., this method is univariate in nature and does not use historical data from previous days. However, it does make notable use of the concept of deviant observations and uses control limits in a manner analogous to using confidence intervals.

In an effort to include historical data pertaining to incidents in an algorithm, a single-parameter incident detection model was developed based on Bayesian principles of probability.[7] This method uses logic similar to the decision tree used in the California algorithms, specifically in calculating the difference between the occupancy values at two consecutive detector stations at a given time – relative spatial difference in occupancy (OCCRDF). A historical data set was used to assign a probability of incident occurrence based on the number of successive iterations yielding values for OCCRDF above a specified threshold. An advantage of the iterative process of the Bayesian algorithm is that an experienced traffic manager viewing its output on a minute-by-minute basis can estimate the likelihood that an incident signal (or series of consecutive signals) will be a false alarm or an actual incident. This algorithm appears to sacrifice detection time to reduce false alarm rate, and like those algorithms previously discussed, it uses only one traffic variable. Although this particular incident detection algorithm requires large volumes of historical data, it offers flexibility in its use and importance and at least compares well with other commonly used algorithms.

An algorithm that uses discriminant analysis to integrate types of traffic data was developed for potential use on arterial streets in the Chicago area as part of the Advance project in the early 1990s.[8,9] Information is obtained by the Advance incident detection system from three sources: loop detectors, probe vehicles, and anecdotal sources. The use of simulated traffic data in the evaluation of this algorithm avoided the issue of data quality in that no screening techniques to identify erroneous data are required; therefore, the lack of false alarms is unsurprising. However, the notion of comparing current and archived historical data first appears here.

Several automated condition monitoring methods (all in the form of incident detection algorithms) have employed statistical methods to some extent. A few aspects of this research appear among these algorithms, but all aspects do not appear in any single method. For example, the Bayesian and Advance algorithms employ historical data, but they rely on data from multiple detector stations to operate, whereas this research examines behavior of traffic on a location-specific basis, not relying on the relationship between data at multiple locations to assess system state. The two algorithms developed in Texas (control charts and SND) are similar to this research in their use of the concept of comparing current data to immediately prior observations to identify deviant trends (thus relying primarily on statistical techniques). However, they each consider only a single traffic variable (volume) and do not use historical data that can enhance the ability of a method to discern abnormal observations. All of the condition monitoring methods discussed take the approach of asking whether an incident has occurred and provide a binary yes or no answer to this question, rather than taking the approach of determining whether traffic conditions are normal, or how abnormal they are, based on current and historical data.

Data Screening in Traffic Management

Screening of the input traffic data can improve upon the performance of any traffic condition monitoring method. Since the introduction of electronic surveillance on roadways in the 1960s, procedures that examine the collected data for errors prior to their use in decision support tools such as incident detection algorithms have continued to evolve. The data being screened through the processes described in this literature review are not the raw inductance data from the loop detectors (sometimes referred to as microscopic data) but rather the traffic characteristic data that have been generated from the raw data by the system software (macroscopic data). In effect, these traffic characteristic data are surrogate values for measurements of mean speed, occupancy, and flow rate. As the use of condition monitoring systems in traffic management systems expands, and as traffic data are shared among an ever-increasing range of users, the macroscopic data are predominantly used for these purposes.

Prior to the screening of traffic characteristic data, the data are typically examined (pre-screened) to identify records that contain negative values or missing data; those records are then removed from further analysis. Tests employed for the screening of traffic characteristic data can be grouped into two categories: threshold value tests and tests based on basic traffic flow theory principles. The most commonly used tests involve the use of threshold values to determine acceptability of data. Maximum and minimum acceptable values for volume, speed, and occupancy are of this type. For example, a test that would place a maximum acceptable value of 80 mph on speed denotes records with values of speed greater than this threshold as erroneous; such records are then removed from use in traffic management applications. A threshold value test may require some simple calculations; for example, if data are collected in 60-second intervals, a maximum acceptable value for volume could be set at 50 vehicles per interval at each detector (equivalent to 3,000 vehicles per lane per hour).

More sophisticated tests make use of relationships among traffic characteristics by applying traffic flow theory. Such tests take advantage of the relationships among the commonly measured parameters in traffic management systems: mean speed, volume, and occupancy. For example, one test to screen data applies a plausible range of volume values within a given range of occupancy values.[10] All types of screening tests attempt to ensure the validity of traffic data prior to their use in traffic management applications.

Despite the dependence of incident detection algorithms and other tools on sound input data, data screening is often incomplete or altogether missing in traffic management systems. A 1984 survey found that only 53 percent of freeway management projects surveyed employed some sort of data checking procedures.[11] Only 34 percent of the systems surveyed examined the plausibility of the data obtained from traffic detectors. Of the systems that provided information on the data screening tests used, most employed upper or lower limits, or threshold values, for occupancy and/or volume. Although many traffic management systems rely on incident detection algorithms to some extent, there is widespread dissatisfaction with their performance.[2]

A substantial portion of the performance problems may be due to unchecked or unsound data being supplied to the algorithms. For example, a scenario in which positive values for speed but zero values for occupancy and volume were supplied to an incident detection algorithm on the Burlington Skyway in Canada was the cause of numerous false alarms.[12] Screening of data for possible errors is likely to improve the efficiency and accuracy of traffic management applications that rely on these data.

INVESTIGATION OF CURRENT PRACTICE AND FUTURE NEEDS

Background regarding the state of the practice of traffic condition monitoring methods used to classify system state was needed to complement the literature review and provide additional insight into appropriate direction for the research. A survey of freeway management systems in the United States was conducted as a part of this research to obtain insight into trends in their staffing and resource availability, the current and future size and scope of freeway management systems and expected improvements, and the current state of condition monitoring practices.

Survey Results

The survey (see Appendix A) was mailed to 39 freeway/traffic management systems across the United States in May and June of 1998. Of the 26 systems responding to the survey, 23 were currently operational. The other 3 planned to be fully operational within the next 2 years.

System Size and Scope

System managers were asked about the size and scope of their systems both today and after any planned expansions. Among the operational systems represented in the survey responses, the average system covers 165 miles of freeway. Four of the systems (17 percent of those responding) also covered some non-limited access highways. To manage the system during peak traffic periods, on average, 3.7 traffic operators were used. All but one of the systems that responded (96 percent) indicated plans to expand the size and scope of their operations. The horizon year for expansion completion ranged between 1999 and 2017. Twenty of these systems reported their anticipated system coverage and peak period operational staff. On a per-system basis, the average expected increase in coverage area was 204 percent, whereas the average expected increase in operational staff was only 71 percent. In fact, only 35 percent of systems responding to the survey anticipated having more than one additional operator during peak periods in the horizon year of their expansions as compared to their current operations.

Traffic Condition Monitoring Methods

Video cameras, status maps (graphical displays), and incident detection algorithms are common methods used to monitor traffic conditions and to assess the state of the transportation system. Closed-circuit television cameras are used by 87 percent of the systems responding to the survey. These cameras provide a snapshot of the system state at certain locations; however, in large systems that have dozens or hundreds of cameras, using these cameras to monitor traffic conditions require significant human resources. Graphical displays of condition information, such as status maps that color freeway segments according to mean speed, are currently used in 65 percent of freeway management systems, whereas 30 percent have such displays under construction or in design. Speed is the most common type of data used to label freeway segments on the display; 76 percent of the systems with visual displays built or under design use mean speed to code the freeway segments. About half of these systems can also label segments according to traffic volume and occupancy data. The other systems use either occupancy or the status of an incident detection algorithm as the primary means of characterizing traffic conditions on freeway segments shown in their displays. The systems that use map displays for overall traffic condition assessment rated them 4.10 on average (on a scale of 1 to 5, with 5 being the most important), whereas video cameras scored an average of 4.82 in terms of their importance as an information source. This high rating underscores the importance of having a quick diagnostic tool such as a system map or a network of video cameras to assess conditions in the system from the traffic management center. Yet, neither of these methods provides an assessment of system state in relation to conditions at other times or to a reasonable expectation of conditions; in fact, video cameras do not use data at all but rather require human memory to provide context.

Among practices used in freeway monitoring systems to monitor traffic conditions and to assess system state, incident detection algorithms were found to be most prevalent, with approximately 70 percent of freeway monitoring systems having used them. Algorithms used include various forms of the California algorithms, which incorporate changes in multiple functions of traffic flow parameters (originally developed in the mid-1970s), algorithms based on only one parameter (either speed or occupancy), all-purpose incident detection, and algorithms developed internally by particular systems. The California family of algorithms was used most widely, with 26.1 percent of survey respondents reporting its use. Respondents who specified the particular algorithm used noted either modified California algorithm 7 or modified California algorithm 8.

Incident Detection Algorithm Experiences

To learn more about the experiences of traffic managers regarding incident detection algorithms, survey respondents were asked to rate their satisfaction with them and then to rate the performance of the algorithms with regard to three commonly used MOEs associated with incident detection algorithms: mean time to detect an incident, false alarm rate, and detection rate. Satisfaction was rated on a scale of 1 (very dissatisfied) to 5 (very satisfied). The average rating was 2.92; however, the ratings were widely distributed, indicating that a wide range of opinions about these algorithms exists among system managers. The performance of the algorithms with respect to the three commonly used MOEs was rated on a similar scale, with 1 representing very poor performance and 5 representing very good performance. The algorithms' detection rate was rated most highly (3.00), followed by mean detection time (2.80); the false alarm rate fared worst (2.53). These average ratings corroborate the findings of past research and others in that satisfaction with incident detection algorithms is mixed, and that a high false alarm rate is a common reason for a system to reduce its reliance upon an incident detection algorithm or discontinue its use altogether.[1] Although a wide range of incident detection algorithms have been employed in an effort to automate condition monitoring, incident detection algorithms have proven to be fallible and in most cases are considered a secondary means of analyzing traffic conditions. As has been documented in past studies, the survey revealed that common experiences include unacceptably high false alarm rates and difficulties in calibration and adjustment to changing conditions.

Survey Summary

This survey found that although nearly all systems have plans to expand in the near future, staffing during peak traffic periods would increase at a much lower rate than system size. Additionally, although incident detection algorithms are the predominant form of condition monitoring method in use today, performance has been only fair (Smith and Turochy, 1999). In this task, it was documented that more automation will be necessary in freeway management systems, and the performance of current methods for using traffic data to monitor traffic warrant further investigation into alternative approaches to condition monitoring. The performance of automated condition monitoring methods to date has not alone justified the extensive investment in surveillance equipment to collect traffic data, indicating the need to improve upon such methods in existence today. Although the performance of incident detection algorithms has not been great, because of the greater rate of increase in system size and scope than in available staff, automated traffic condition monitoring will continue to play an important role in the future. In fact, 83 percent of survey respondents indicated that they do plan to use some form of automated condition analysis tools in the future. Additionally, they believe that an automated system for traffic condition monitoring and analysis will play a role of high importance. Therefore, a strong need exists for improvement over the automated traffic condition monitoring and analysis methods in use today, primarily incident detection algorithms. Increasing system size, smaller operational staff, improved data gathering and storage capabilities, and more reliance on automation all point to the need to develop more effective tools for the detection of abnormal conditions in freeway management systems.

IDENTIFICATION AND SELECTION OF CANDIDATE APPROACHES

From the investigation of current practice and the literature review, desirable attributes of candidate methods for traffic condition monitoring were derived. A review of statistical analysis techniques for potential use in condition monitoring and system state assessment was then conducted to select a method for further development. Among several statistical techniques considered, MSQC best met the criteria developed herein.

Criteria for and Selection of Candidate Methods

General requirements for an automated method that monitors and analyzes traffic conditions were derived through the literature review, the investigation of current practice, and discussions with several VDOT employees involved in traffic management. To improve upon existing methods, several key features of a candidate method include:

- *The method should be multivariate in nature.* Rather than examining each traffic variable independently, a method that accounts for the relationships among the variables makes better use of the data being collected and can also yield additional opportunities for data screening and for identification of abnormal conditions. Considering mean speed, volume, and occupancy as a multivariate set allows for data screening tests and condition monitoring methods to be developed that exploit the interdependence of these measured variables through traffic flow theory.

- *The method should provide an output that characterizes system state across a continuous range of conditions.* Methods that provide a simple binary assessment of conditions, such as incident detection algorithms, provide a very limited amount of information to traffic managers. More informative outputs allow traffic managers, rather than automated algorithms, to make decisions on system operations. An assessment of conditions across a continuous range may also allow for some identification and trouble shooting of system component (e.g., detector) failures.

- *The method should yield results based on properties of the data themselves.* Rather than requiring the use of externally applied thresholds or other manipulation to classify system state, the method should derive the assessment of system state through the data internally used in its calculations. For example, declaring as abnormal a value for volume that is greater than two standard deviations from its historical mean is preferable to declaration of abnormal conditions based on a value for volume that is more than some specified threshold (e.g., 200 vehicles per hour per lane) from its historical mean. Condition assessment based on classification as determined only by threshold values for data is fallible as threshold values may not be sufficiently dynamic to represent changing traffic patterns and do not necessarily indicate the presence of a particular condition.

- *The method should relate current conditions to past observations that are historically relevant.* Comparing the current system state to conditions at similar times of day and days of the week takes advantage of the cyclical nature of traffic on daily and weekly bases. The extent to which current conditions are abnormal can thereby be determined by comparing them to a set of relevant historical data that effectively serves as an expectation of current conditions.

Traffic condition monitoring methods in use today do not meet all of the preceding criteria. Incident detection algorithms, most of which operate in a rule-based decision-tree manner, typically do not consider the three commonly measured traffic variables as a set of related measures, describe conditions in a simple binary (yes/no) manner, and employ static thresholds. Additionally, the focus of these algorithms is only to detect incidents, rather than more broadly describing the state of the system in light of expected conditions (based on historical data). Several other methods to perform an assessment of current system state were subjected to the stated criteria, including discriminant analysis, cluster analysis, knowledge-based methods, and statistical quality control. However, only SQC (the multivariate form) meets all four criteria. Discriminant analysis as typically applied classifies an observation into one of two groups, rather than across a continuous range. Cluster analysis allows for more than two categories but would require an infinite number of classes to represent a continuous range of conditions. Knowledge-based methods, in addition to requiring extensive training time, also cannot provide output across a truly continuous range. Except for MSQC, all of these methods, although handling multivariate data, do not explicitly account for the correlation among the measured variables.

MSQC provides a suitable paradigm for analysis of traffic conditions, even though relatively little control can be exerted on the traffic stream when compared to industrial processes to which MSQC is traditionally applied. The concept of a range of normal conditions, within which little apparent (or important) variation in system performance exists, and the concept of several degrees, or a continuum, of abnormal conditions, is intuitive to traffic managers. In statistical quality control, an observation of current conditions is compared to the distribution of observations of relevant past conditions; traffic conditions are assessed in the same manner. Traffic managers define normal conditions based on spatial and temporal characteristics, comparing conditions in a certain location to prior conditions at the same location and at the same time of day and day of week. This concept of temporal comparability was used to develop strategies for sampling the historical database.

Introduction to Multivariate Statistical Quality Control

The candidate methods for effective monitoring of traffic conditions were developed using the multivariate form of SQC. SQC applies the properties of the normal distribution to observations that comprise historical data. In SQC, past observations are plotted and control limits can set at any desired level (usually at a specified number of standard deviations from the mean of the past data, or at a specified confidence level). These control limits can be used for the detection of abnormal observations that represent large deviations from the process mean. For example, a common decision rule is that one observation more than three standard deviations from the mean, or two successive observations more than two standard deviations from the mean, indicates that the current conditions are abnormal. In the univariate case, this chart takes the form of a Shewhart control chart. Examples of such a chart are shown in Figure 1.[13] In Figure 1, upper and lower control limits at two standard deviations greater than and less than the mean, respectively, are shown as dashed lines. An observation that falls outside these lines is indicative of an abnormal condition in SQC and condition monitoring.

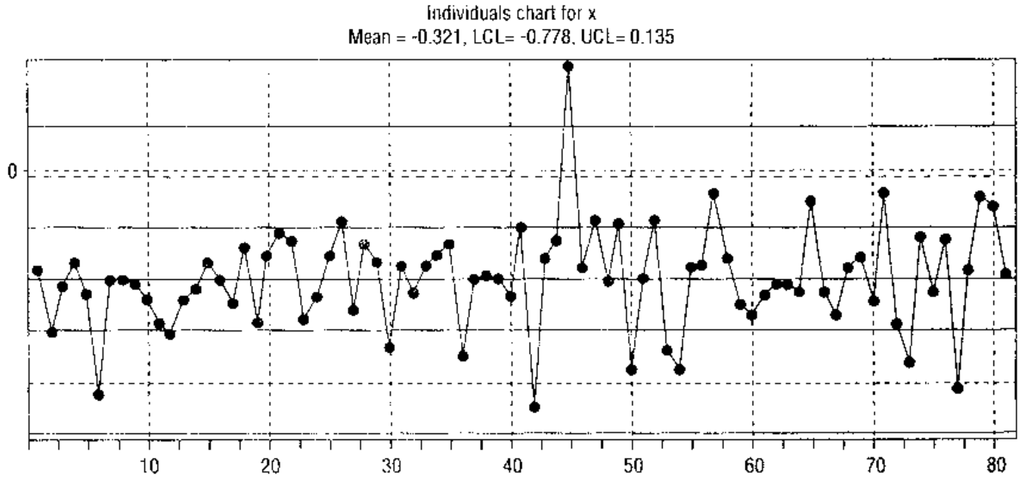

Figure 1. Example of Shewhart Control Chart (Mason and Young, 1998)
Copyright, American Society for Quality. Reprinted with permission.

In the multivariate case where the variables are jointly distributed (there is correlation among the variables), a control ellipsoid (based on some measure of variation) can be developed and new observations plotted; if a new observation falls outside the ellipsoid, it is abnormal. The control ellipsoid can be drawn using some multiple of standard deviations (as is typically done with univariate control charts) or for a certain confidence level. This control ellipsoid, drawn at a value of Hotelling's T^2 corresponding to a desired confidence level, takes advantage of the interrelationships among the measured variables and is a multivariate analog of the univariate test statistic, *t*. In the bivariate case, the control ellipsoid becomes a control ellipse that is easily represented graphically. Applying univariate control charts (therefore discounting the interrelationships among the variables) to a bivariate process results in a box shaped region, as shown in Figure 2.[13]

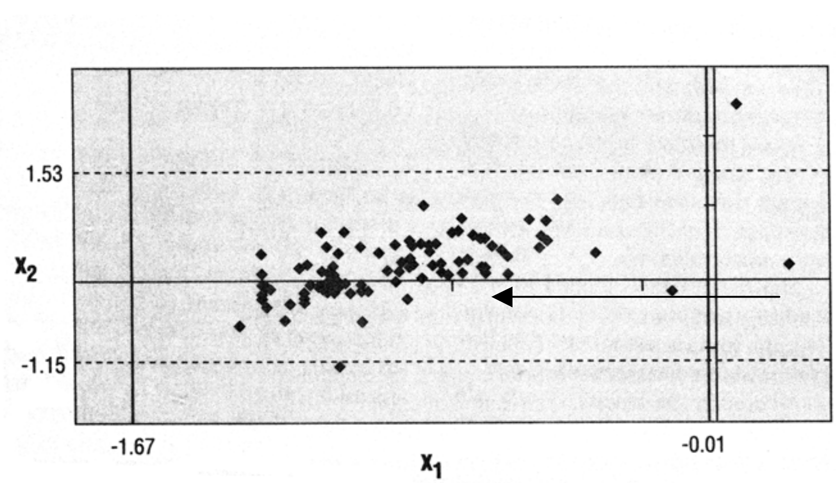

Figure 2. Example of Bivariate Shewhart Control Chart (Mason and Young, 1998)
Copyright, American Society for Quality. Reprinted with permission.

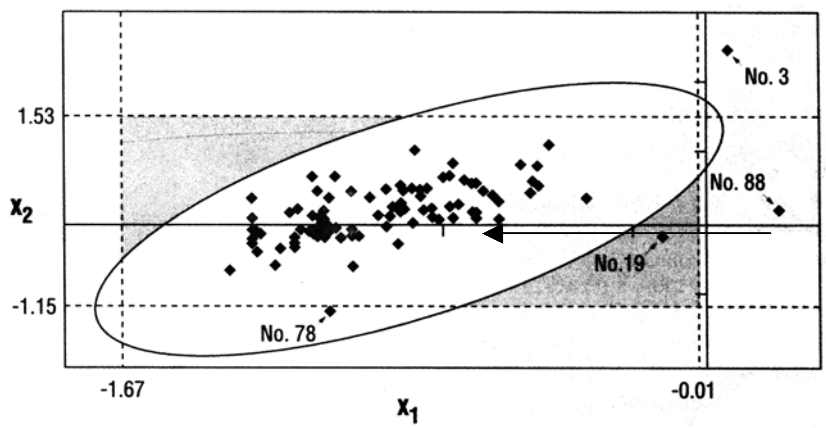

Figure 3. Example of Bivariate Control Ellipse (Mason and Young, 1998)
Copyright, American Society for Quality. Reprinted with permission.

Figure 3 superimposes the control ellipse based on some specified value of T^2 onto Figure 2, showing clearly the misinterpretation potential that would result from modeling a cross-correlated multivariate process as independent univariate processes.[13] In this example, observation 19 would be considered normal (when a specific confidence level is applied) without regard to the relationships between the variables, but when the confidence ellipse that takes correlation into account is plotted, this observation is clearly not normal. Conversely, observation 78 falls within the normal region when MSQC is applied. In this research, the trivariate traffic data (mean speed, volume, occupancy) are used in this approach. MSQC does not treat them as individual, independent variables but instead as a group of related, interdependent measures. Univariate cases are not considered in this research; in fact, such a method that uses volume data to examine observations has been used in an incident detection algorithm previously.[4]

Multivariate Statistical Quality Control Application Procedure

Hotelling's T^2 (also referred to simply as T^2) can be used to measure the distance representing the deviation of the current multivariate observation from the group mean to which it is compared. As explained later, the group to which current conditions are compared may consist of observations from the same location at similar times of the day and days of the week. The statistical distance measured by T^2, also known as Mahalanobis' distance, measures the statistical distance from an observation to the center of the data. Unlike Euclidean distance, this measure weighs each variable inversely to the size of its variance and accounts for the correlation among the variables.[13] T^2 can be viewed as a multivariate extension of the square of the univariate Student's *t*-statistic that characterizes the deviation of a multivariate observation from its group mean with a single variable, thus allowing for a simple, easily interpreted measure of deviation.

In the following description of the MSQC computational procedure, the current observation refers to the most recent observation of traffic conditions at a particular location. Each observation consists of three measured variables: mean speed, volume, and occupancy. In MSQC, the statistical distance (T^2) between a particular observation of interest (the current observation) and a reference sample of observations is measured. In this application to traffic condition monitoring, the current observation were compared to a set of observations that had similar temporal and spatial characteristics.

Conformance with the Multivariate Normal Distribution

MSQC is predicated on the assumption to the data used in the calculations are drawn from a population that approximately follows a multivariate normal distribution. Prior to execution of the basic calculations of MSQC that result in the T^2 value for a given observation (upon comparison with a set of relevant data), the data set comprising the set of relevant observations with which the observation of interest is compared was checked for conformance with the multivariate normal distribution. Each set of relevant historical data drawn from the database for comparison with a current observation was examined for normality using a computational procedure that is a surrogate for a χ^2 plot (analogous to a q-q plot for the univariate case). The data set was checked for trivariate (mean speed, volume, and occupancy) normality. Evaluation criteria for checking the normality assumption are "roughly half of the d_j^2 are less than or equal to $q_{c,p}(0.50)$, and, a plot of the ordered squared distances, d_j^2, versus $q_{c,p}((j-0.5)/n)$ is nearly a straight line having slope 1 and which passes through the origin."[14] The first of these two criteria states that approximately 50 percent of the sample observations should lie within the 50 percent confidence ellipsoid, whereas the second states that the plot of generalized distances from the mean versus χ^2 quantiles should be approximately linear and that the values in each ordered pair ($q_{c,p}((j-0.5)/n)$, d_j^2) are nearly equal. MSQC is sufficiently robust to handle minor deviations from the normal distribution.[14]

In order to strengthen the testing procedure, in addition to the 50 percent quantile, the 25 and 75 percent quantiles were also examined, as a surrogate to ensuring that the plot of the ordered squared distances d_j^2 from the mean versus χ^2 quantiles approximates a line of slope 1 that passes through the origin. Specifically, the proportions of d_j^2 values that fall within the 25, 50, and 75 percent confidence ellipsoids were compared to the values of the 25, 50, and 75 percent quantiles, respectively, of the chi-square distribution with $n-p$ degrees of freedom (where n is the number of observations and p is the number of variables observed). In this case, $p=3$ for the three measured traffic variables; d_j^2 is the statistical distance (ordered squared distance) from an observation, j, to the mean of the group of observations $j = 1,\dots,n$.

For each observation, j, in the sample, d_j^2 was calculated as follows:

$$d_j^2 = (\boldsymbol{x_j} - \bar{x})^T \boldsymbol{S}^{-1} (\boldsymbol{x_j} - \bar{x}); \text{ where} \tag{1}$$

x_j is the set of relevant past observations, a vector of dimension $p \times 1$; $j = 1,\dots,n$

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \; ; j = 1,\ldots,n \qquad (2)$$

$\bar{x}$ is the mean of the group of relevant past observations, a vector of dimension $p \times 1$

$$\boldsymbol{S} = \frac{1}{n-1}\Sigma\, (\boldsymbol{x_j} - \bar{x})\, (\boldsymbol{x_j} - \bar{x})^{\mathrm{T}} ; j = 1,\ldots,n \qquad (3)$$

$\boldsymbol{S}$ is the covariance matrix of the observations in the relevant group, a vector of dimension $p \times p$

The covariance matrix (or variance-covariance matrix), S, is computed by summing the squared differences between each past observation and the mean of the past observations and then dividing by the term $(n - 1)$; the multivariate nature of the observations requires each difference vector to be multiplied by its transpose to arrive at the squared distance for each observation. S is a function of only past observations to which the current observation is being compared, and not to the current observation itself. The covariance matrix is a symmetric matrix ($S^T = S$) in which the elements on the main diagonal are the variances for each component variable and the other elements are the covariances between two variables. Incorporating the inverse of the covariance matrix into the calculation of distance weighs each variable (or interaction between two variables) inversely proportional to its variance (or covariance between the two variables). After the proportion of d_j^2 values that fall below χ^2_p (0.50) is determined, if approximately half (between 40 and 60 percent) of the d_j^2 values are less than χ^2_p (0.50), then the underlying distribution can be considered to be multivariate normal.[14] This test can expanded upon using any χ^2_p quantiles. To strengthen the testing procedure, similar checks are performed for χ^2_p (0.25) and χ^2_p(0.75). If between 15 and 35 percent of the d_j^2 values are less than χ^2_p (0.25), and if between 65 and 85 percent of the d_j^2 values are less than χ^2_p (0.75), then the underlying distribution of the past observations to which the current observation is compared is considered to be multivariate normal.

Confidence Ellipsoid for New Observations

A confidence, or prediction, ellipsoid for quality control of new observations can be generated as a function of the expected percentage of all observations drawn from the same population (defined spatially and temporally) that fall within the ellipsoid. This concept can be used to determine whether conditions may be normal or abnormal; if the data point representing the current observation falls outside the ellipsoid defined by a desired level of confidence, then current conditions may be abnormal. T^2_k for a new observation (where k is simply an index that uniquely identifies an observation), $\boldsymbol{x_k}$, compared to a reference sample of relevant past observations (whose mean is $\bar{x}$) taken from a larger population, is given as follows:[14]

$$T^2_k = (\boldsymbol{x_k} - \bar{x})^T \boldsymbol{S}^{-1} (\boldsymbol{x_k} - \bar{x}) \text{ ; where} \qquad (4)$$

$\boldsymbol{x_k}$ is the current observation, a vector of dimension $p \times 1$

When the population mean and variance are represented by sample data, T^2 follows an F-distribution and is distributed according to $[p(n-1)(n+1)/n(n-p)]\ F_{p,n-p}$. Therefore, a $100(1-\alpha)$ *percent* p-dimensional prediction ellipsoid ($p=3$ in this application), centered on $\bar{x}$, is given by all $\boldsymbol{x}$ that satisfy

$$(\boldsymbol{x} - \bar{x})^T \boldsymbol{S}^{-1} (\boldsymbol{x} - \bar{x}) \leq \frac{p(n-1)(n+1)}{n(n-p)}\ F_{p,n-p}(\alpha) \tag{5}$$

Rearranging the preceding formula and setting it as an equality allow for calculation of F for a particular observation of interest. The F-statistic is a function of T^2, n (the number of observations in the group of relevant past observations), and p (the number of variables in each observation). The F-statistic has two (p, $n-p$) degrees of freedom. T^2_k has been shown to follow an F-distribution; $F(\alpha)$ is calculated as follows:

$$F_{p,n-p}(\alpha) = \frac{n(n-p)}{(n-1)(n+1)p}\ T^2_k \tag{6}$$

Therefore, for a known value of n, with $p=3$, the corresponding value of T^2_k can be found for any value of α. Critical values of T^2 that correspond to values of interest of α (e.g., $\alpha=0.10$, $\alpha=0.05$, $\alpha=0.01$) can be calculated, allowing direct comparison of T^2_k to any number of critical T^2 values, facilitating the determination of the normality of any observation. This simplification of calculations is expounded upon in the next section pertaining to extensions on the basic MSQC calculations.

At the specified significance level α, the null hypothesis that the new observation, $\boldsymbol{x}$, and the reference samples come from populations with equal means ($\bar{x}$) cannot be rejected if the condition of equation (5) is met.[15] In this manner, for any tolerance region (such as the 95 percent prediction ellipsoid for which $\alpha=0.05$), a determination can be made as to whether a new observation falls within that tolerance region and can therefore be considered normal.[14] Incorporation of the covariance matrix, $\boldsymbol{S}$, into the distance calculation accounts for the inherent correlation among the traffic variables and distinguishes statistical distance from straight Euclidean distance. The output of this basic condition monitoring method is then simply a notation for each location reporting data at a given time that conditions are either normal or abnormal.

Observations that do not reflect normal conditions (those that are out-of-control) are not used in the definition of tolerance regions. Upon extraction of a set of relevant historical observations from the database, the calculation in equation 1 was applied to each observation in the historical data set, prior to performing the T^2 calculation outlined in equation 4. Any observation in the historical data set for which d_j^2 is greater than the critical T^2 value for $\alpha=0.05$ (such observation would fall outside the 95 percent confidence ellipsoid or tolerance region) is rejected and replaced with a different observation from the database the meets the spatial and temporal criteria that define the set of relevant historical observations. Although this procedure helps to minimize the possibility that observations of abnormal conditions as well as erroneous data are included in the historical data set, it also introduces a bias against all observations that are highly deviant from the historical mean (e.g., outside the 95 percent confidence ellipsoid). The resultant means, variances, and covariances are also affected, with variances and covariances reported as smaller than their true values. Since normality calculations are based on these values, this practice increased the probability that an observation would be considered abnormal, as explained the subsequent evaluation of the MSQC-based condition monitoring method.

DATA COLLECTION AND MANAGEMENT

Prior to use of any data in the development and evaluation of a traffic condition monitoring method, the data must be inspected and any processing necessary for operation of the traffic condition monitoring method must be performed. The data must then be screened for potentially suspect data to minimize the possibility that erroneous data would be included in application of the method. A strategy for extracting data from the archived database must be developed to ensure that temporally relevant data are used to create the historical data sets. This section closes with an analysis of the statistical properties of the historical data sets generated from the data collected at the study sites.

Data Archiving, Inspection, and Preparation

Traffic Characteristic Data

Traffic data must be obtained in order to develop, apply, and evaluate candidate methods for traffic condition monitoring. Simulated traffic data have been used in previous efforts involving incident detection algorithms. Although the generation of artificial data requires fewer resources than collection and inspection of real data, the use of artificial data does have drawbacks. Data from the field must be collected and inspected to address data quality concerns pertinent to implementation of any condition monitoring method in practice; the use of simulated data does not ensure that data quality issues are appropriately addressed. Traffic data from HRSTC, obtained and archived at the Smart Travel Laboratory as collected, were inspected to ascertain the characteristics and quality of the data. These data are collected using inductive loop detectors in the HRSTC, owned and operated by VDOT in southeastern Virginia. Thorough inspection and preparation of the data sets to be used in the development and testing of condition monitoring methods highlight patterns of potentially erroneous data, thereby guiding development of the data screening procedure. This inspection and preparation process also guides the manner in which the data will be sampled for use in the condition monitoring methods and the organization of the database.

The Hampton Roads system currently has 203 detector stations that cover 19 miles of freeways (sections of Interstates 64, 264, and 564) and ramps at interchanges along these freeways. About 80 of these stations are on ramps; the remainder are on mainline sections of freeway. This research focused on mainline locations since abnormal conditions on the mainline tend to affect a much greater volume of traffic and create much more delay than on the ramps. An observation, or data record, consists of three traffic variables (mean speed, volume, and occupancy) measured for a given time period and detector station. In addition to reporting values for the three traffic variables, each record also reported a collection length (collection interval) and the number of lanes reporting data (operating correctly) for that interval. Unlike most traffic management systems, the interval over which traffic data are measured (often referred to as the collection interval or collection length) in Hampton Roads is variable, ranging between 110 and 210 seconds. To stabilize this variability, data are obtained from Hampton Roads by the Smart Travel Laboratory in 2-minute intervals (at the beginning of every even minute), effectively fixing the collection interval as a constant. The rare occasions in which records may be missed do not affect this research as the constant interval of 120 seconds, beginning with every even minute (e.g., 12:32:00), was used in the grouping of observations into bins. A sample of the traffic data, as obtained from the database and placed into a spreadsheet format for analysis, is shown in Table 1.

Table 1. Sample of Hampton Roads Traffic Data As Archived

Date	Time	Speed (mph)	Occupancy (%)	Volume (vehicles)	Collection length (sec)	Lanes with data (lanes)
7/17/00	16:10	53	12	180	119	3
7/17/00	16:12	37	26	250	120	3
7/17/00	16:14	31	28	223	120	3
7/17/00	16:16	30	29	217	120	3
7/17/00	16:18	36	25	235	120	3

The Hampton Roads traffic data are reported in the following units: speed is reported is miles per hour, occupancy in percent, volume in vehicles, and collection length in seconds. The data as shown in Table 1 require additional preparation. Although two of the three variables typically used to monitor traffic conditions (mean speed and occupancy) are computed directly as part of the processing software at the traffic management center, volume as a flow rate is not. Vehicles per hour per lane, or hourly equivalent volume, is a commonly used measure of traffic flow that relates traffic volumes to available highway capacity. Hourly equivalent volume is often referred to as *flow* or *volume*; this is the case within this report. This calculation is easily performed in a spreadsheet and in the procedure used to query the database by converting the volume reported in each record into a measure of average flow rate by lane and across time, as follows:

$$\textit{Flow (veh/hr/ln)} = \frac{3600\,(Volume)}{(Collection\ Length)\,Lanes\ with\ data} \qquad (7)$$

The flow rate derived in equation (7) is referred to as volume throughout the remainder of the research. Table 2 shows the data from the observations represented in Table 1 after the data have been prepared for use in the MSQC calculations.

Table 2. Sample of Hampton Roads Traffic Data As Prepared

Date	Time	Speed (mph)	Occupancy (%)	Volume (veh/hr/ln)
7/17/00	16:10	53	12	1815
7/17/00	16:12	37	26	2500
7/17/00	16:14	31	28	2230
7/17/00	16:16	30	29	2170
7/17/00	16:18	36	25	2350

The data as collected in the HRSTC and received by the Smart Travel Laboratory also have one other noteworthy peculiarity in that the maximum value of mean speed reported at each station is 65 mph; the reason for the inclusion of this feature in the firmware at the traffic management system could not be determined. However, this *cap* was not expected to affect the research adversely. A brief inspection and analysis of data from several stations in the system showed that at many locations, the mean speed never reaches 65 mph. At those stations where the mean speed was reported to be 65 mph (but could possibly be higher), this phenomenon usually occurred only during the overnight hours, when a condition monitoring system is least needed (as traffic flows are very low).

Selection of Locations for Method Development and Evaluation

For the development and evaluation of the condition monitoring methods, detector stations located on basic freeway segments were selected to ensure that traffic data used were representative of typical freeway traffic and not affected by interchanges and queues on ramps. In basic freeway segments, each detector station has either one or two loop detectors in each lane at a particular location. An installation of two loop detectors at one station (a pair of detectors, usually spaced about 10 feet apart, in each lane at the detector station, sometimes referred to as speed traps, essentially measures spot speeds. Vehicle speeds are calculated by recording the time that a vehicle begins to cross over each detector and then dividing the difference between those times into the fixed distance between the beginning points of each detection zone. The process for estimating vehicle speeds at a single loop detector is much less robust than at a paired loop detector configuration; therefore, stations with paired loop configurations were selected for use in this research.

Data from 10 stations that represent the coverage area of the freeway traffic management system in Hampton Roads were used in this study (see Figure 4); information on these locations is given in Table 3. Through the inspection of data, these 10 locations are known to be reliable (i.e., operating and reporting data a high percentage of the time) and to generate data of good quality (i.e., consistently passing the data screening procedure described in the next section). The full range of geometric conditions for mainline freeway segments that exist in the Hampton Roads region are covered in the 10 study sites. The extent of the HRSTC coverage area is also shown in Figure 4 as are the location codes for each freeway segment included in the current 19-mile coverage area.

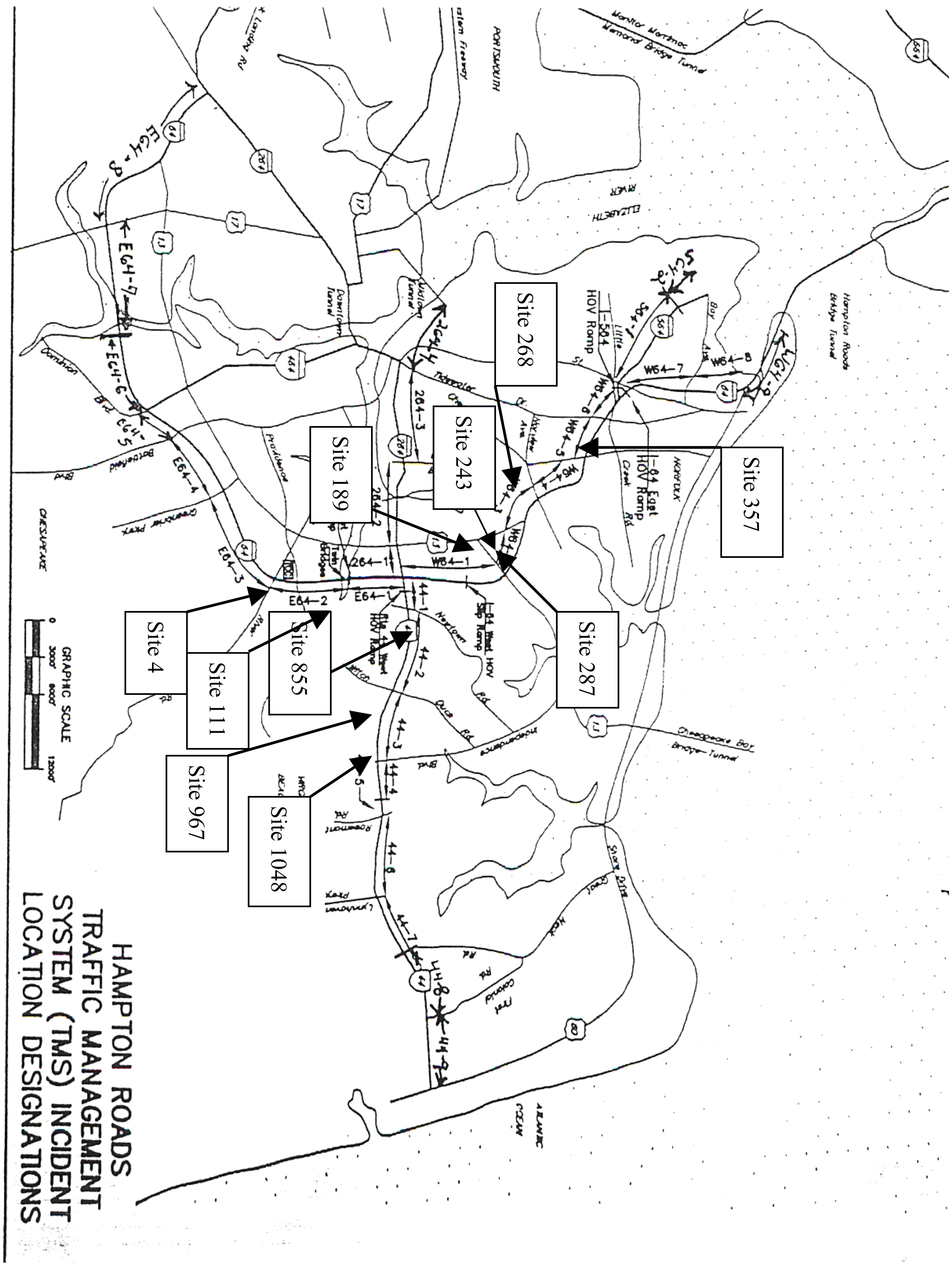

Figure 4. Map of Hampton Roads Study Area

Table 3. Selected Locations for Study Data

Route	Direction	Between Exits	Site	Station
64	Westbound	286 (Indian River Rd) and 289 (Greenbrier Pkwy)	4	15
64	Westbound	284 (I-264) and 286 (Indian River Rd)	111	4
64	Eastbound	282 (Northampton Blvd) and 284 (I-264)	189	39
64	Eastbound	281 (Military Hwy) and 282 (Northampton Blvd)	243	60
64	Eastbound	279 (Norview Ave) and 281 (Military Hwy)	268	71
64	Westbound	279 (Norview Ave) and 281 (Military Hwy)	287	69
64	Westbound	277 (Tidewater Dr) and 278 (Chesapeake Blvd)	357	92
264	Eastbound	14 (I-64) and 15 (Newtown Rd)	855	141
264	Eastbound	16 (Witchduck Rd) and 17 (Independence Blvd)	967	165
264	Eastbound	17 (Independence Blvd) and 18 (Rosemont Rd)	1048	182

Data Screening Procedure

Sample data sets were taken from the archived data to assess data quality. Upon inspection of the data, it quickly became evident that some observations were not consistent with basic traffic flow theory principles. Some other observations also had extremely short collection lengths of less than ten seconds, which led to similar inconsistencies. These examples of poor data quality gave rise to the need for development of a data screening procedure as an integral part of any condition monitoring method. A thorough search of data screening techniques used in traffic management was conducted, and new screening tests were developed to detect suspect data. These two efforts resulted in a data screening procedure that addresses the unique characteristics of the Hampton Roads data and that can also be applied to traffic data collected in other locations, thereby extending the state of knowledge in transportation operations data screening.

Traffic data screening tests can be grouped into two broad categories: threshold value (or acceptable range) and traffic flow theory-based. The focus of screening tests found in the literature is typically on data quality as determined by values of individual variables that fall outside an acceptable range. The quality of data values that are simply not possible (such as negative values) or are not considered feasible from a traffic engineering perspective (such as volumes well above estimated highway capacity) is considered suspect and therefore should not be used in any analysis of traffic conditions. In addition to applying threshold value tests to data from the Hampton Roads system, these data were also examined for consistency with basic traffic flow theory principles. The relationships among the three traffic variables measured (mean speed, volume, and occupancy), as well as the characteristics of the data collection operation itself, were examined to locate infeasible data values and to develop tests that identify them. An example of a test that employs traffic flow theory is the volume-to-occupancy ratio test.[10] In this test, minimum and maximum acceptable values for the ratio of volume to occupancy were established based on an examination of traffic data and guided by traffic flow theory principles that relate these variables to one another. A unique range of acceptable values was established for low occupancy values (i.e., 7 percent and below), medium values (i.e., 8 to

25 percent), and medium-high values (i.e., 26 to 35 percent), and very high values (i.e., 36 percent and above).

Procedure Development

A development database composed of 1 day of data from the entire Hampton Roads system (approximately 100,000 records) and 1 week of data from four detector stations in known peak-period congestion locations (approximately 21,000 records) was used to assess the effects of the tests and of different parameters within each test. Prior to application of the screening procedure, a pre-screening function was performed. Records that contained negative values or missing data were removed from the database, leaving only records with positive values for speed, volume, and occupancy. A visual inspection of detector data, application of traffic flow theory to these records, and an examination of screening tests documented in the literature led to the development of tests that together constituted a data screening procedure. The screening procedure consisted of six tests, three of the threshold value variety, one that had properties of both types of tests, and two that used traffic flow theory. The tests were independent and could be applied in any combination or order. Further details on the tests can be found in *Transportation Research Record 1727*.[16] After the screening procedure was developed and refined, it was evaluated using another day of data.[16]

Test 1 – Maximum Occupancy Threshold

This threshold value test examines occupancy values in individual records. If the occupancy value for a particular station and time period is higher than is considered feasible for traffic, that record is considered erroneous. Based on an inspection of the data, the threshold value for this test was set at 95 percent; records with reported values greater than 95 percent are rejected. This test should not be applied to data from ramps as queues may routinely occupy detectors at these locations (the sites included in this study are all on mainline freeway segments.

Test 2 – Overall Maximum Volume Threshold

A threshold value test of this type can be found in many traffic management systems; it used to detect records in which the reported volumes are infeasible based on expected highway capacity. Records of volume as high as 3,100 vehicles per hour per lane (for a collection interval of 30 seconds) have been found in the literature; visual inspection of the Hampton roads data revealed volumes as high as 2,900 vehicles per hour per lane. Therefore, if the calculated hourly equivalent volume for a record is greater than a threshold of 3,100 vehicles per hour per lane, that record is rejected.

Test 3 – Positive Volume with Zero Speed

This test has properties of both threshold value and traffic flow theory-based tests as it examines the relationship between volume and speed values and screens out records in which the reported speed is zero and the reported volume is positive (greater than 0). Such a combination is infeasible; for flow to occur (positive volume), speed must also be positive; therefore such records should be considered erroneous.

Test 4 – Maximum Volume Threshold with a Reported Occupancy of Zero

This test examines only single records in which 0 is recorded for occupancy but a volume greater than 0 is recorded. In the processing of data in the Hampton Roads system (and many other traffic management systems), the decimal component of occupancy is truncated, effectively rounding occupancy down to the nearest whole number. This scenario occurs frequently in the early morning hours when traffic is in a very low flow condition. Estimated volume, as a check against the reported volume when occupancy is reported as zero, can be derived as a function of occupancy, speed, collection interval, and the length of the detectors themselves.

Test 5 - Average Effective Vehicle Length Infeasible

This test examines a function of occupancy, volume, and speed data for individual records. The average effective vehicle length (sum of vehicle length as observed by the detectors and detector length) is calculated from the data using traffic flow theory principles. The average effective vehicle length (AEVL) has a range over which its values are feasible, based on the physical dimensions of the vehicles traveling on the freeways. The need for this test became apparent after calculation of values of AEVL, for records in the development database that did not fail other tests, yielded infeasible values in excess of 100 ft for some records and less than a few feet for others. Table 4 presents an example of several records for which the values of the individual variables (mean speed, volume, and occupancy) appear feasible but for which the derived AEVL is infeasible. These records would not be considered erroneous using screening tests that consider each variable independently, but when the relationships among the variables are considered (such as through the calculation of AEVL), these records are correctly classified as erroneous.

Table 4. Records Resulting in Infeasible Values for Average Effective Vehicle Length

Speed (mph)	Volume (veh/ln/hr)	Occupancy (%)	AEVL (ft)
2	2245	10	0.47
24	2269	1	0.56
53	1600	3	5.25
16	2045	13	5.37

There are two important traffic flow theory assumptions made in this test. First, occupancy is used as a surrogate measure for density in the conventional speed-flow-density relationship; a linear relationship between occupancy and density is assumed in the formula for AEVL. The second assumption made in the AEVL test pertains to speed as measured in the field and as used in the formula. An installation of paired loop detectors (a pair of detectors in each lane at the detector station, sometimes referred to as speed traps) essentially captures spot speeds, which are then aggregated and reported as a time mean speed for each collection interval. However, the formula relating flow and density (and thus the formula for AEVL) calls for space (harmonic) mean speed. Other recent studies using the same Hampton Roads data concluded that "the use of time mean speed instead of space mean speed in traffic flow modeling for basic freeway segments does not introduce a significant statistical error in analysis."[17]

This test should be performed only on data collected at stations that have paired loop detectors. Data from the inductive loop detector stations used in the development database, and most of the detector stations in the Hampton Roads system, are of this configuration. This type of installation allows for individual vehicle speeds to be measured directly by dividing the known distance between detectors into the time difference between activations of the detectors by a vehicle. Speeds cannot be directly calculated from single loop stations; an assumption pertaining to average vehicle length or some other parameter is required. As vehicle lengths are derived from speed data, this test is suitable only for data collected from paired loop detector stations. The basic formula that yields average effective vehicle length is the product of speed and occupancy divided by volume.

The average effective vehicle length test cannot be performed on all records. Because of the fact that this test involves multiplication by speed and by occupancy and division by volume, records in which any of mean speed, volume, or occupancy was reported as 0 were excluded from this test (as they would yield values of AEVL equal to 0 or to infinity). Fortunately, records with such values typically do not represent conditions in which diagnosis of system state is dependent only on data from the detector station whose data are in question. For example, a value of zero for volume may represent conditions downstream of an incident that completely blocks traffic; however, such a catastrophic incident would generate data at the upstream station that would signal the presence of abnormal conditions.

Test 6 – Minimum Collection Length Threshold

This test screens out records with values of 90 seconds or less for collection length. In traffic management systems, the collection length is typically fixed. This test evolved directly from inspection of the Hampton Roads data in which the collection length over which traffic data are aggregated is variable, typically falling between 110 and 210 seconds. Occasionally, a record contains a very short collection length along with volumes that are not feasible in such a short period. Therefore, unlike the other five tests included in the data screening procedure, this test likely has minimal application to data collected using loop detectors in traffic management systems other than Hampton Roads.

Additional Considerations and Limitations to the Data Screening Procedure

In this screening procedure, maximum threshold value tests were employed for volume and occupancy data. However, such a test for speed was not included due to a characteristic of the data processing software in the HRSTC. A cap on maximum reported values of speed of 65 mph is built into the software. This feature, although perhaps serving as a type of threshold value test prior to reporting of the data, can possibly impact the usefulness of the data, as true mean speeds may range above 65 mph at several locations. A screening test that would place a maximum acceptable value for mean speed could be easily implemented as a part of this screening procedure if applied to data from other traffic management systems.

An additional consideration is the manner in which occupancy data are reported. Although occupancy is a continuous variable, occupancy values are typically reported discretely as whole numbers with units of percent, as is the case with the Hampton Roads data. Occupancy as reported provides sufficient precision for many traffic flow analyses; however, at very low flows, the reported value of occupancy may differ considerably from the true value. In many traffic management systems (including Hampton Roads), occupancy is rounded down to the nearest integer when reported. Most records in which occupancy is reported as 0 actually have volumes greater than zero; however, since the occupancy is less than 1 percent, it is reported as zero. This scenario gives rise to the need for test 4 (maximum volume with 0 occupancy). Additionally, any tests and analyses that include multiplication or division by occupancy are confounded by such records, which must therefore be excluded from these types of tests, such as test 5 (average effective vehicle length).

The impact of the traffic data screening procedure described was assessed by applying it to a test database of 1 day of data pulled from the archived Hampton Roads data. After the test database was prescreened, the procedure was applied to the remaining 70,651 records; 10.32 percent of these records failed at least one test in the procedure (thereby failing the procedure itself); most of the failures pertained to test 5 (average effective vehicle length), which had a failure rate of 8.29 percent. The fact that the AEVL test detected more than 3 times as many erroneous records as did all other tests in the procedure is indicative of its robustness. Although this test represents an important contribution to the practice of traffic data screening, the impact of minimum and maximum acceptable values of AEVL is worthy of further investigation.

The traffic data screening procedure described herein was developed in the Smart Travel Laboratory and has been developed into a front-end application for accessing the Hampton Roads data that users can employ to screen their data using any combination of the six tests. This application, in Microsoft Visual Basic, presents a user interface that does not require the user to access the database directly or be familiar with Structured Query Language (SQL), the syntax used in database applications such as Oracle 8i. Since the data screening procedure was developed primarily for use with data from paired-loop detector stations (particularly the average effective vehicle length test), it should be used with caution on data from single-loop detector stations. Additionally, although the concepts behind the screening tests are applicable to traffic data collected in any traffic management system, data from systems other than Hampton Roads should be thoroughly inspected and the screening procedure tested on small samples of data prior to large-scale use. This cautionary approach should determine the suitability of the procedure's application to such data, or detect necessary changes in the values and thresholds used in the tests as described herein.

Data Sampling Strategy

A major consideration in applying traffic condition monitoring methods that employ historical data is the organization of the historical database. Specifically, an informed decision about the level of granularity at which data will be sampled to generate the historical data sets to which the most recent observation is compared must be made. The manner in which data are sampled (*sampling strategy*) affects the historical extent of data required as the sampling strategy is a function of how temporally relevant data are defined. Additionally, the ease of implementation, operating speed of the method, and applicability of the historical data sets to the current observation are also affected by the choice of sampling strategy. Although use of data at a fine level of detail (such as considering every detector station in the system at each collection interval by each day of the week) may allow for the method to be most versatile and sensitive to small changes in system state, it may also result in the most complex operational scheme and relatively high processing time. On the other hand, although simplifying the data sampling used to generate an operational database may allow for rapid startup and relatively quick response, much detailed information may be lost.

With respect to criteria for defining historical data sets, the level of temporal segmentation for time-of-day and day-of-week can be coarse, such as considering all days of the week to be the same and considering all data within a 1-hour window to be the same, thus allowing for high sample sizes for each historical data set within a relatively low-dimensioned database. Temporal segmentation can also be very fine, such as considering each day of the week separately and each collection interval (every 2 minutes for the Hampton Roads data) separately, resulting in a database with relatively high dimensions and potentially limited sample sizes for each historical data set.

For the development and evaluation of the condition monitoring methods, detector stations in paired-loop (speed trap) configurations, located on basic freeway segments, were selected, as explained previously. For the development, application, and evaluation of the methods, data from individual stations can be used without regard to the extent of the system they may represent.

Several levels of temporal segmentation were considered for application and evaluation of each of the candidate methods. Two types of temporal segmentation are used: time-of-day (TOD) and day-of-week (DOW). Initially, three levels of segmentation were proposed for both TOD and DOW, as follows:

1. DOW as each day separately, with 7 values (DOW = Monday, Tuesday, …, Sunday)

2. DOW as weekday (Monday through Friday) or weekend (Saturday or Sunday), with 2 values (DOW = Weekday, Weekend)

3. DOW as all days, with one value, effectively eliminating segmentation by DOW, and for TOD:

   (a) TOD as each collection interval separately; in the case of the Hampton Roads data this yields 720 possible values (TOD = 0:00, 0:02, …, 23:58)

   (b) TOD in a window of 10 minutes of collected data centered on the current time (5 collection intervals for data aggregated at 120-second intervals); in the case of the Hampton Roads data this yields 720 possible values (for example, if TOD = 14:00 then the time periods of interest include data collected at 13:56, 13:58, 14:00, 14:02, and 14:04)

   (c) TOD in a window of 1 hour of collected data centered on the current time (30 collection intervals for data aggregated in 120-second intervals), in the case of the Hampton Roads data, this yields 720 possible values (for example, if TOD = 14:00, then the time periods of interest include data collected at 13:30, 13:32,…, 14:28, and 14:30).

The sampling strategy naming convention defined is used throughout the remainder of this report. For example, in generating a historical data set, strategy 1B includes observations from the same day of the week as the observation of interest and within 5 minutes before or after the time of the observation of interest for possible inclusion in the historical data set.

Four of the nine sampling strategies originally proposed were selected for development. Strategies 3, 3B, and 3C were eliminated from consideration since traffic patterns on weekdays (i.e., Monday through Friday) and weekend-days (i.e., Saturday and Sunday) are substantially different. Through a visual inspection of graphs, or traffic profiles, of mean speed as a function of TOD and volume as a function of TOD, as well an examination of the deviation of mean speed and of volume between days of the week, the 5 weekdays can be considered similar enough so that they could be collapsed into one group, whereas profiles for Saturday and Sunday could also be merged into one group. Based on these findings, it would not be appropriate to treat data from a Saturday or a Sunday as from the same set of conditions as data from a weekday. Figure 5 shows typical traffic volumes for Saturdays and Sundays at the same site. Figure 6 shows typical traffic volumes for weekdays (average of Monday through Friday, with bars of two standard deviations shown) at one study site. It is evident by comparing these two figures that weekdays and weekend-days exhibit markedly different traffic patterns and thus should not be combined within a sampling strategy.

Strategies that consider data from a 1-hour period (Strategies 1C, 2C, and, again, 3C) were not studied further for two reasons. First, traffic conditions can vary widely over a 1-hour period. This changeability of traffic conditions is not a concern with strategies 1A and 2A, which do not consider more than one consecutive TOD interval (with Hampton Roads data, this means examining data for a 2-minute window by TOD), and Strategies 1B and 2B, which consider historical data over an approximately 10-minute window. Second, the observations that comprise the historical data set would be taken from a very small number of days. For example, if the method is designed to draw the most recent 100 observations from the database according

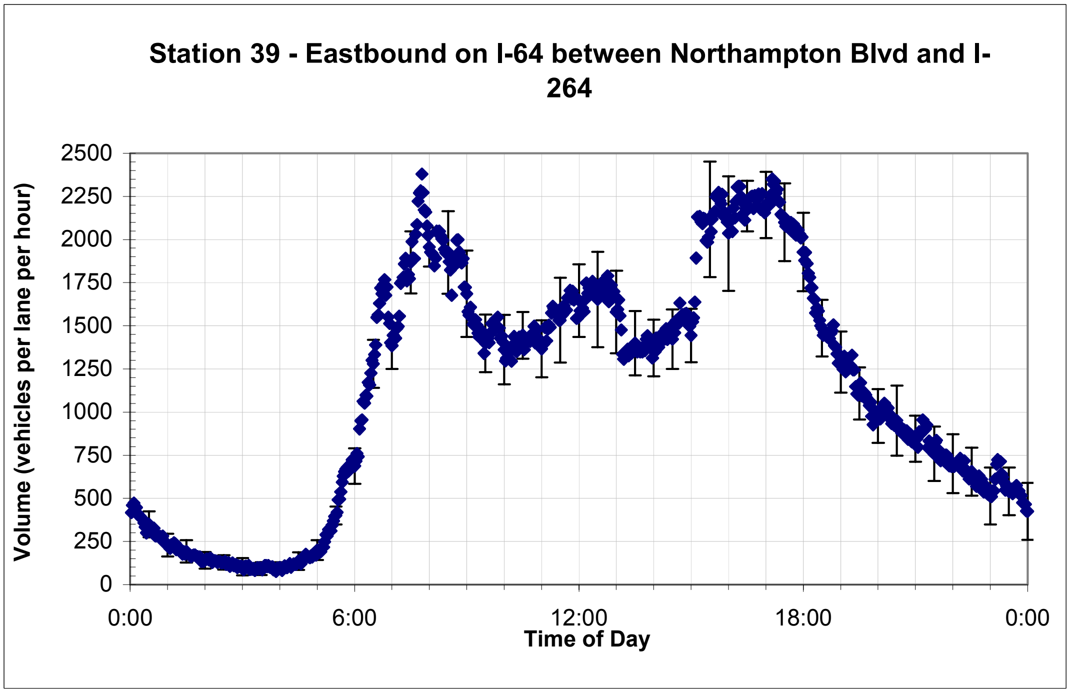

Figure 5. Typical Weekend-Day Traffic Profile

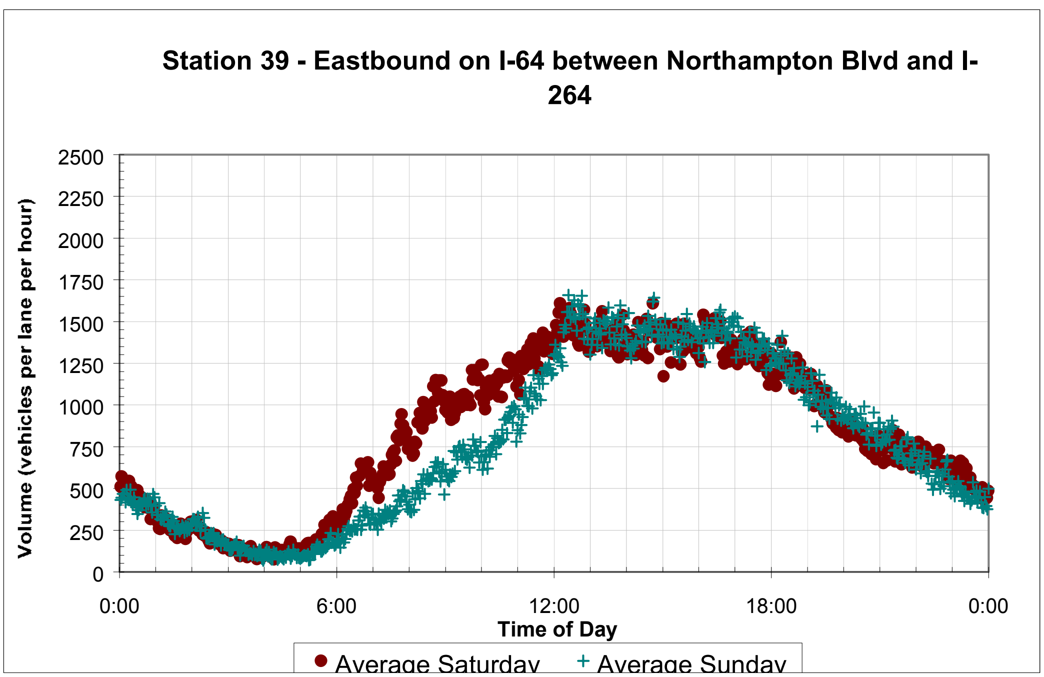

Figure 6. Typical Weekday Traffic Profile

to the sampling strategy in use and that pass the screening procedure, they would likely be taken from only four different days (at up to 30 observations per day, for data aggregated at 120-second intervals). If a non-recurring traffic disruption (such as an incident or other unusual event) were to occur on one (or more) of those days, it would have a great influence on the picture of typical or normal traffic conditions as represented by that historical data set, thereby rendering the data set inappropriate for use in a condition monitoring method. Strategies 1B, 2B, and 3B, which draw no more than five observations from any particular day, would employ data from at least 20 different days (based on 100 historical observations), thereby minimizing the impact that one day of abnormal traffic conditions would have on the resulting characterization of present traffic conditions.

Rather than limiting the condition monitoring method to one single strategy for developing historical data sets to represent normal traffic conditions, several different strategies were considered for temporally segmenting the database. After examining the potential problems of each of the strategies, four temporal strategies were retained for development and evaluation. Sampling strategies that consider observations collected over a range of 1 hour to be from the same historical data set, and strategies that consider all days of the week to be from the same set, were excluded from further analysis.

The four sampling strategies further developed and evaluated were:

1. Strategy 1A - same day-of-week, same 2-minute time interval

2. Strategy 1B - same day-of-week, 10-minute time interval

3. Strategy 2A - all weekdays or weekend-days, same 2-minute interval

4. Strategy 2B - all weekdays or weekend-days, 10-minute interval

By developing several sampling strategies, the effectiveness of MSQC in traffic condition monitoring can be evaluated over a wide range of schemes for database organization.

Statistical Properties of Historical Data Sets

After identification of sites to be used in the development and application of MSQC-based traffic condition monitoring methods, an analysis of the statistical properties of the data was performed. Of particular interest is the distribution of data in the historical data sets generated by the prototype program. The general procedure for checking conformance of a data set with the multivariate normal distribution, as given in the literature and modified in this research, is discussed previously and reviewed briefly below. Historical data sets created using any of the four database sampling strategies described in the preceding section, in general, approximately follow the multivariate normal distribution. The extent of conformance does vary somewhat by strategy, and by study site, as explained in the following discussion. Conformance was assessed using the procedure described previously.

Differences among the strategies were minor when conformance proportions are averaged across all study sites, with 75.2 percent of historical data sets generated using strategy 1B conforming to the multivariate normal distribution according to the tests previously described, 75.0 percent of historical data sets generated using strategy 2A passing the test, 74.7 percent of historical data sets generated using strategy 1A passing the test, and 73.6 percent of historical data sets generated using strategy 2B passing the test. Table 5 shows the proportion of conditions tested for which the historical data sets passed the test for conformance with the multivariate normal distribution across study sites and by site.

Table 5. Proportion of Historical Data Sets Conforming to Multivariate Normality

	Proportions			
Site, Sampling Strategy	**1A**	**1B**	**2A**	**2B**
All Study Sites	0.74660	0.75220	0.74971	0.73645
Site 4	0.67227	0.75394	0.78983	0.75776
Site 111	0.71818	0.73537	0.76862	0.78304
Site 189	0.73913	0.74627	0.69414	0.69296
Site 243	0.71282	0.73767	0.66744	0.73154
Site 268	0.77143	0.74945	0.74769	0.71269
Site 287	0.79481	0.76606	0.78795	0.77397
Site 357	0.75066	0.74882	0.79607	0.76415
Site 855	0.81132	0.81657	0.76800	0.71166
Site 967	0.76667	0.75862	0.77154	0.69825
Site 1048	0.74317	0.74897	0.75107	0.71484

Hypothesis testing was performed to determine the significance of differences in the proportions of historical data sets conforming to the multivariate normal distribution among the sampling strategies using the summary results for the ten study sites. The null hypothesis tested is that the proportion of historical data sets conforming to the multivariate normal distribution for a particular sample is the same as the proportion for another sample, that is, H_0: $p_1 = p_2$. Each of the six pairwise comparisons was performed by applying the test for comparison of proportions in independent samples (this is sometimes referred to as the 2x2 contingency table). The value of the test statistic, z, is calculated as follows:[18]

$$z = (p_1 - p_2) / \sqrt{\frac{p(1-p)}{\frac{1}{n_1} + \frac{1}{n_2}}} \text{ ; where} \qquad (8)$$

p_1 is the estimated proportion that the condition has been met for sample 1
p_2 is the estimated proportion that the condition has been met for sample 2
n_1 is the sample size of sample 1
n_2 is the sample size of sample 2
and p is the pooled proportion for the two sets:

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \quad (9)$$

In each test, the resulting probability that the underlying distributions of the two samples compared were the same is found in a normal distribution table by applying the calculated *z*-statistic. Each comparison was checked for significance at the $\alpha=0.05$ level; if so, then the difference between the two means is commonly considered to be *statistically significant*. No statistically significant differences were found in the comparisons; in each case the difference in proportions was not significant at the $\alpha=0.05$ level. The results of the six comparisons between sampling strategies for significance of differences in conformance to multivariate normality test are shown in Table 6. Without regard to site-specific considerations, the extent to which historical data sets conform to the multivariate normal distribution does not vary significantly among the four sampling strategies and therefore no preference can be given to any sampling strategy based on this criterion.

Table 6. Results of Statistical Testing on Conformance with Multivariate Normality

Strategies	Z-Statistic	Level of Significance
1A vs. 1B	-0.506	$\alpha > 0.05$
1B vs. 2A2A	0.242	$\alpha > 0.05$
2A vs. 2B	1.277	$\alpha > 0.05$
1A vs. 2A	-0.149	$\alpha > 0.05$
1B vs. 2B	1.541	$\alpha > 0.05$
1A vs. 2B	0.907	$\alpha > 0.05$

Selecting a sampling strategy that generates historical data sets with the greatest extent of conformance with the multivariate normal distribution maximizes the robustness of the MSQC-based condition monitoring approach. In general, when selecting a sampling strategy for use on a systemwide basis, the differences among the strategies are minor and insignificant. However, if a very small number of locations are to be monitored and the selection of a sampling strategy is feasible on a site-specific basis, examination of a sample of historical data set generated using each strategy, for each site, would yield the best strategy for that site as conformance proportions vary considerably across the sites in a traffic management system. In general, since MSQC can handle minor deviations from the multivariate normal distribution without losing robustness, the extent to which the historical data sets conform to the distribution as indicated by the testing procedure described is sufficient.[14] In application of the MSQC-based traffic condition monitoring method to other locations, an assessment of conformance of the historical data sets with the multivariate normal distribution should be made and a decision then made as to whether the proportion of sets that conform is sufficient to proceed with the MSQC-based approach.

DEVELOPMENT AND APPLICATION OF CONDITION MONITORING METHODS

This section explains how the concepts of basic MSQC and knowledge of the data characteristics described previously are adapted for and applied to the monitoring of traffic conditions. Enhancements to the basic application of MSQC that provide additional insight regarding the state of the highway system have been developed through a review of extensions on MSQC noted in the literature and new extensions developed in this research effort. These enhancements lead directly to output measures for the prototype computer program developed to monitor traffic conditions for real-time use in traffic management systems.

Multivariate Statistical Quality Control Applied to Traffic Data

The value of the confidence region on which the current observation falls is related to the level of confidence that can be applied to the statement that the current observation can be considered as drawn from the reference sample of observations (historical data set). The value of α (from equation 6) is the level of significance at which the hypothesis that the current observation and the historical data set come from the same population can be rejected. It is therefore convenient to state that the probability that abnormal conditions are occurring can then be represented as the value $1-\alpha$. This interpretation of MSQC allows for outputs of the condition monitoring method to convey information about the extent to which current conditions can be considered abnormal.

Basic MSQC, defined previously as the necessary operations to arrive at a value of T^2_k, results in a single number that describes the deviation of a multivariate observation from a reference point (in this context, this is the historical mean). However, basic MSQC does not provide any additional information, or context for assessing normality, about the condition of the process being monitored. Enhancements to basic MSQC, consisting of a combination of extensions found in the literature and then adapted for use in traffic condition monitoring, and of new extensions developed in this research, are described below. An extension is defined as any additional analysis beyond the basic concept of MSQC that results in T^2_k. An extension can provide additional information about $\mathrm{T}^2_{\mathrm{k}}$, such as its significance or its member components.

Enhancements to Multivariate Statistical Quality Control

Three enhancements to basic MSQC for application to traffic condition monitoring developed in this research are:

1. Classification of $\boldsymbol{x_k}$ simply as normal or abnormal can be augmented by applying several control limits simultaneously.[15] This allows for several distinct levels of abnormality to be employed.

2. Variable tolerance regions can be used to represent an exact level of abnormality (normality level). This enhancement is an original extension on MSQC that determines the exact extent to which $\boldsymbol{x_k}$ (the observation of interest) is abnormal.

3. Decomposition of T^2_k into orthogonal components representing the contribution of each measured variable to T^2_k. This enhancement is based on a decomposition of T^2 as described in the literature.[19]

The first enhancement can provide a representation of conditions across a range of conditions, the second allows for a characterization of system state on a continuous range, and the third allows for identification of each variable's contribution to T^2 for a particular observation (T^2_k).

Univariate Control Chart with Several Critical Control Limits

The MSQC approach can be represented in a manner similar to univariate Shewhart control charts, since T^2 is a single value (for an observation, comprised of several variables) that contains information about all variables measured. The T^2 Chart illustrated in Figure 7 demonstrates this concept.[13] An upper control limit for the value of T^2 corresponding to the desired confidence level can be placed on the chart. Any number of control limits for a range of confidence levels (e.g., 90, 95, 99, 99.9 percent) can be applied simultaneously, allowing for an interpretation of observed conditions over several degrees of potentially abnormal conditions, and characterization of traffic conditions across several discrete intervals, thereby approximating a continuous range.

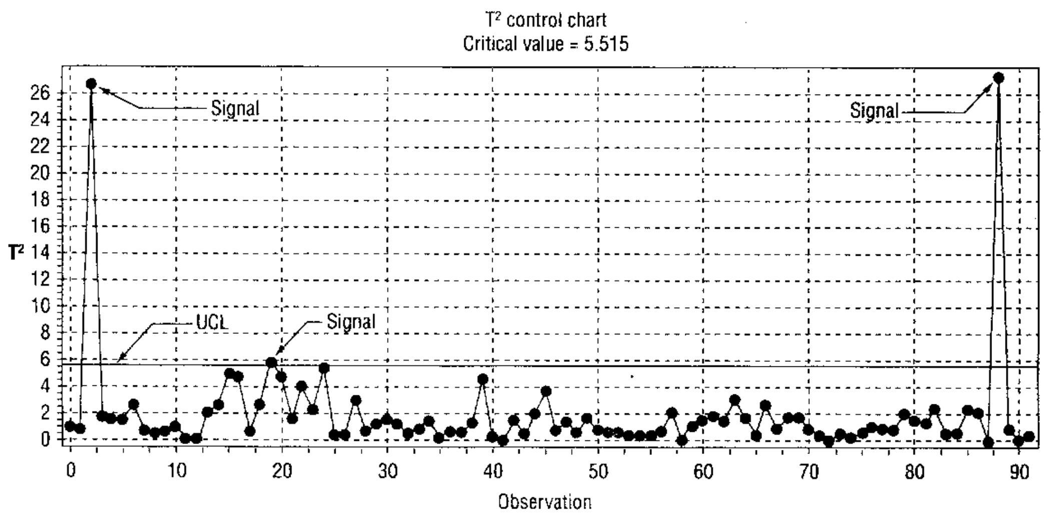

Figure 7. Example of T^2 Chart (Mason and Young, 1998)
Copyright, American Society for Quality. Reprinted with permission.

This concept of a series of control limits can be presented as a series of levels of normality or abnormality. As shown in Table 7, numerical values or words can be used to describe the ranges of normality delineated by the control limits. For example, if the value of T^2_k, when plotted, is above the value of the control limit that corresponds to the 90 percent confidence ellipsoid but below the control limit based on the 95 percent confidence ellipsoid, the observation can be classified as *abnormal level 1*, thereby eliminating the need to interpret a

Table 7. Output Measures for Degrees of Normality

T^2_K (Current Conditions):	Level of Normality:
$T^2_k <= T^2(90\%)$	Normal
$T^2(90\%)<=T^2_k<T^2(95\%)$	Abnormal 1
$T^2(95\%)<=T^2_k<T^2(99\%)$	Abnormal 2
$T^2(99\%)<=T^2_k<T^2(99.9\%)$	Abnormal 3
$T^2_k >= T^2(99.9\%)$	Abnormal 4

chart to discern the information. This classification of traffic conditions into one of several levels of abnormality is denoted as output measure 1 for the prototype program.

The values in parentheses in Table 7 refer to tolerance regions that capture the noted percentage of observations from the general population. The table essentially defines a series of concentric ellipsoidal prediction regions, centered on the mean of the historical data set.[15] An increase in the level of abnormality (representing increasing deviation of a current observation from the mean) corresponds to a larger confidence ellipsoid that envelops those regions encompassing observations that are nearer to normal conditions. The concept of concentric tolerance regions can be illustrated graphically as shown in Figure 8.

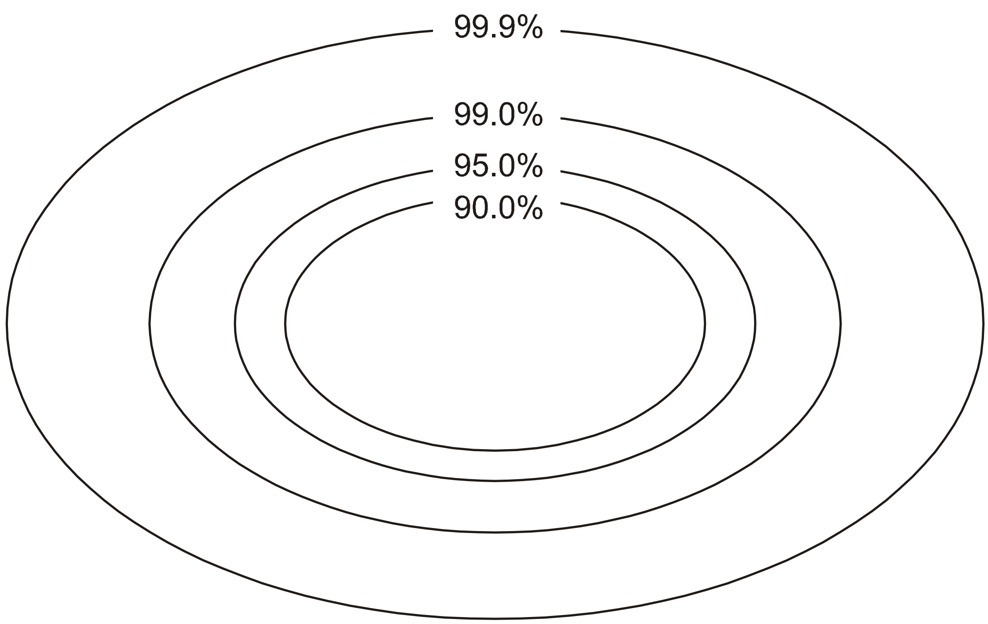

Figure 8. Graphic Representation of Concentric Tolerance Regions

The calculations needed to turn this enhancement to basic MSQC into output measure 1 for the prototype program are straightforward. It can be seen in equation (6), that for fixed values of n and p, $F(\alpha)$ and T^2_k are linearly related. Therefore, for selected values of α (such as 0.10, 0.05, 0.01, and 0.001, as applied in Table 7), and selected values of n (the number of observations that constitute the relevant historical data set) and p (the number of observed variables, fixed as 3 in this application), the values of T^2_k can be used directly in the program and checked against previously determined critical values for T^2, as shown in the preceding table. This avoids the need to calculate $F(\alpha)$ for each new observation. If flexibility is desired such that any value of n can be selected, rather than from a small set of values, then the calculation of $F(\alpha)$ can be performed in the program as described previously.

Variable Tolerance Regions

Output measure 1 of the prototype program, described previously, compares T^2_k to the critical values associated with several values of α to develop of series of levels of abnormality. This concept can be expanded upon to denote the exact level of abnormality by solving for $F(\alpha)$ in equation 6, in which $F(\alpha)$ is calculated as a function of n and p (both of which are constants), and T^2_k. Upon calculation of $F(\alpha)$, the value of α is determined using the series expansion for F, a complex function of F, n, and p. There are four series expansions, in order to accommodate whether n is even or odd and whether n-p is even or odd. The formulas for the series expansions can be found in the National Bureau of Standards publication, *Applied Mathematics Series 55, Handbook of Mathematical Functions*.[20] For brevity, only one of the four series expansions is shown here (the case for which p is odd and n-p is even):

$$\alpha(F|v_1,v_2) = 1 - [(1-y)^{\frac{v_1}{2}}]\,[1 + \frac{v_1}{2}y + \left(\frac{v_1(v_1+2)}{(2)(4)}\right)y^2 + \ldots$$

$$+ \frac{v_1(v_1+2)\ldots(v_2+v_1-4)}{(2)(4)\ldots(v_2-2)}\,y^{\frac{v_2}{2}}\,] \qquad (10)$$

where v_1 and v_2 are the two degrees of freedom associated with the F-distribution (in this case $v_1 = p$ and $v_2 = n-p$) and y in the preceding formula is defined as:

$$y = v_2 / (v_2 + v_1 F) \qquad (11)$$

In comparison with output measure 1, rather than using a series of concentric tolerance regions to assess current conditions, the surface of the tolerance region on which the current observation lies is determined, and the corresponding value (confidence level) is determined. This approach to characterizing traffic conditions is utilized in output measure 2 of the prototype program. The information can be presented by noting the abnormality level, or probability, that for each location being monitored, abnormal conditions may be occurring (statistically, this is the probability that the current observation does not come from the same population as the reference sample of historical data). Therefore, this output measure characterizes traffic conditions across a continuous range.

Decomposition into Orthogonal Components

The T^2 statistic provides a useful measure of the distance of a multivariate observation from its mean (and therefore a measure of the extent to which conditions vary from normal) that does consider the relationships among the variables. However, T^2 itself does not provide information on how far the values of the variables within an observation deviate from their respective means. The T^2 statistic can be decomposed into orthogonal components that identify each variable's contribution to the value of T^2 conditional on the other variables, according to a method developed by Mason et al.[19] This decomposition allows the condition monitoring method to identify which variables (if any) are individually abnormal based on a particular confidence level. Upon identifying the contributing component for a variable, the remainder of the decomposition is dependent upon that component. For example, one decomposition order is:

$$T^2 = T^2_1 + T^2_{2/1} + T^2_{3/1,2} \qquad (12)$$

in which T^2 is decomposed into the contribution of variable 1, plus the component for variable 2 conditioned upon (after removing the effects of) variable 1, plus the component for variable 3 conditioned upon variables 1 and 2. The general formula for the decomposed components is given as:

$$T^2_{p/1,\ldots,p-1} = [(\boldsymbol{x}_p - \bar{x}_{p/1,\ldots,p-1}) / s_{p/1,\ldots,p-1}] \qquad (13)$$

Note that due to this conditional structure, T^2 is not equal to $T^2_1 + T^2_2 + T^2_3$. Therefore, there are $p!$ possible decompositions with a total of $2p$ possible terms. In this application, with $p=3$ measured variables, there are six possible decompositions. The individual T^2 components can then individually be evaluated against a series of control limits to determine the extent to which the individual values are normal, in a manner similar to that used to evaluate the overall T^2 statistic in output measure 1. In the prototype program, when output measure 3 is executed, each of the three components from the six possible decompositions are investigated.

Although it has already been shown that T^2 values follow an F-distribution as follows:

$$T^2_k \sim \frac{p(n-1)(n+1)}{n(n-p)} F_{p,n-p}(\alpha) \qquad (14)$$

The individual components are similarly distributed as

$$T^2_{j+1/1,\ldots,j} \sim \frac{(n+1)}{n} F_{1,n-1}(\alpha) \qquad (15)$$

Therefore, the unconditional components T^2_1, T^2_2, and T^2_3 follow this distribution and can be checked for significance at any desired value of α.

This output measure can be used as an additional diagnostic tool in the event of abnormal conditions to identify the contribution of each variable (mean speed, occupancy, and volume) to the overall condition. Therefore, this output measure will be only invoked if current conditions exceed some defined threshold of abnormality (e.g., $T^2_k > T^2$ [95%]). This threshold is shown here at the 95 percent tolerance region, as exists in the prototype program, but can easily be modified to allow for user specification at any critical value. The 95 percent threshold was used in the prototype program since the 95 percent confidence level is typically used to establish statistical significance of a difference. When invoked, the prototype program will calculate each of the possible decompositions and then perform significance testing on each component to ascertain the contribution of each to the total T^2 value.

Prototype Program Operation

A test-stage prototype computer program was developed in order to evaluate the MSQC-based traffic condition monitoring method. The prototype program was developed in the Smart Travel Laboratory in order to access the database that contains historical data from the HRSTC. These programs use Microsoft Visual Basic to perform the basic MSQC calculations and enhancements that lead to output measures as defined previously.

The prototype program is intended for development and ultimate implementation in traffic management centers. One of the key functions of traffic management systems is to monitor traffic flow and detect abnormalities in the traffic stream as quickly as possible. Based on the information gathered regarding traffic conditions, decisions can then be made about how to manage traffic, including dissemination of information on traffic conditions (through media such as dynamic message signs, highway advisory radio, and the general news media), and operation of traffic control devices such as ramp metering and other traffic signals.

1. *Obtain user inputs.* The user of the program selects the locations to be examined in the operation of the program (historical traffic data from these locations must be archived in a relational database, and "current" observations must also be accessible); sampling strategy and output measures desired are also specified.

2. *Query database for observation of interest.* The remaining steps are executed for each location individually, that is, steps 2 through 10 are executed for the first location selected, then for the second location selected, through all locations selected. For each selected location, the database is queried to obtain the most recent ('current') observation (using a function of the system time). Data screening, using the procedure developed herein, is performed in the query.

3. *Query database for relevant historical data.* Historical observations are read from the database, populating an array on which the basic MSQC calculations will be performed. Data screening is performed in the query; a record will only be used to populate the array that constitutes the set of relevant historical data if it passes the screening procedure.

4. *Calculate mean and variance-covariance matrices.* The first two of the three basic MSQC calculations are then performed: the matrix of means ($\overline{X}$) for the set of relevant historical data is calculated using equation (2) and the variance-covariance matrix (***S***) is calculated and inverted using equation (3).

5. *Examine historical data set for and removal of out-of-control observations.* The historical data set is then examined to ensure that all observations fall within a

specified confidence ellipsoid (set at 95 percent) to ensure that no out-of-control observations, or outliers, are included in the historical data set. Any record discarded is replaced with another historical observation that passes the screening procedure (according to step 3), and the process returns to step 4. If all observations are within the specified control ellipsoid, then the process moves to step 6.

6. *Conformance of historical data set with multivariate normal distribution.* The historical data set is then checked for conformance with the multivariate normal distribution and the results of this check are displayed or noted in the file.

7. *Calculate T^2 for the observation of interest.* The third step in the basic MSQC calculations, the calculation of T^2 for the current observation (T^2_k), is then performed (equation 4). The T^2_k value for this record is displayed.

8. *Calculate output measures.* The output measures (enhancements to the basic MSQC calculations) are then obtained and displayed, based on user input as outlined in Step 1. Output measure 1 is the categorization of the current state into one of a set of ranges of normality as defined by a specified control ellipsoid (e.g., <90, <95, <99, <99.9 percent). Output measure 2 (variable tolerance regions) is the calculation of the control ellipsoid on whose surface the current observation falls, yielding a "normality level." Output measure 3, triggered only if the current observation falls outside a specified control ellipsoid (initially set at 95 percent, this threshold can be made user-specified), calculates the normality ranges as defined in output measure 1 for each of the traffic variables.

9. *Archive results.* The results of the current state analysis for the observation of interest are written to an output file, including date and time, location, mean speed, volume, and occupancy for the current record, the historical means of mean speed, volume, and occupancy, T^2_k, the results of the applied output measures, and the results of the multivariate normality conformance check.

10. *Repeat for next observation.* At the next time at which new observations would are received, the program returns to Step 2. In the operation of the prototype program in the Smart Travel Laboratory, this would occur at every even-numbered minute, as the data are received in the lab at such times. This time interval could be made user-specified and easily adjusted for any data collection cycle.

Figure 9 shows the graphic user interface for the prototype traffic management program. The first step for the user is to select a database sampling strategy (in this case strategy 2B was selected); the user then selects any number of desired locations for monitoring (location 111 was selected in this example). The user may then click the "verify sampling strategy and locations"

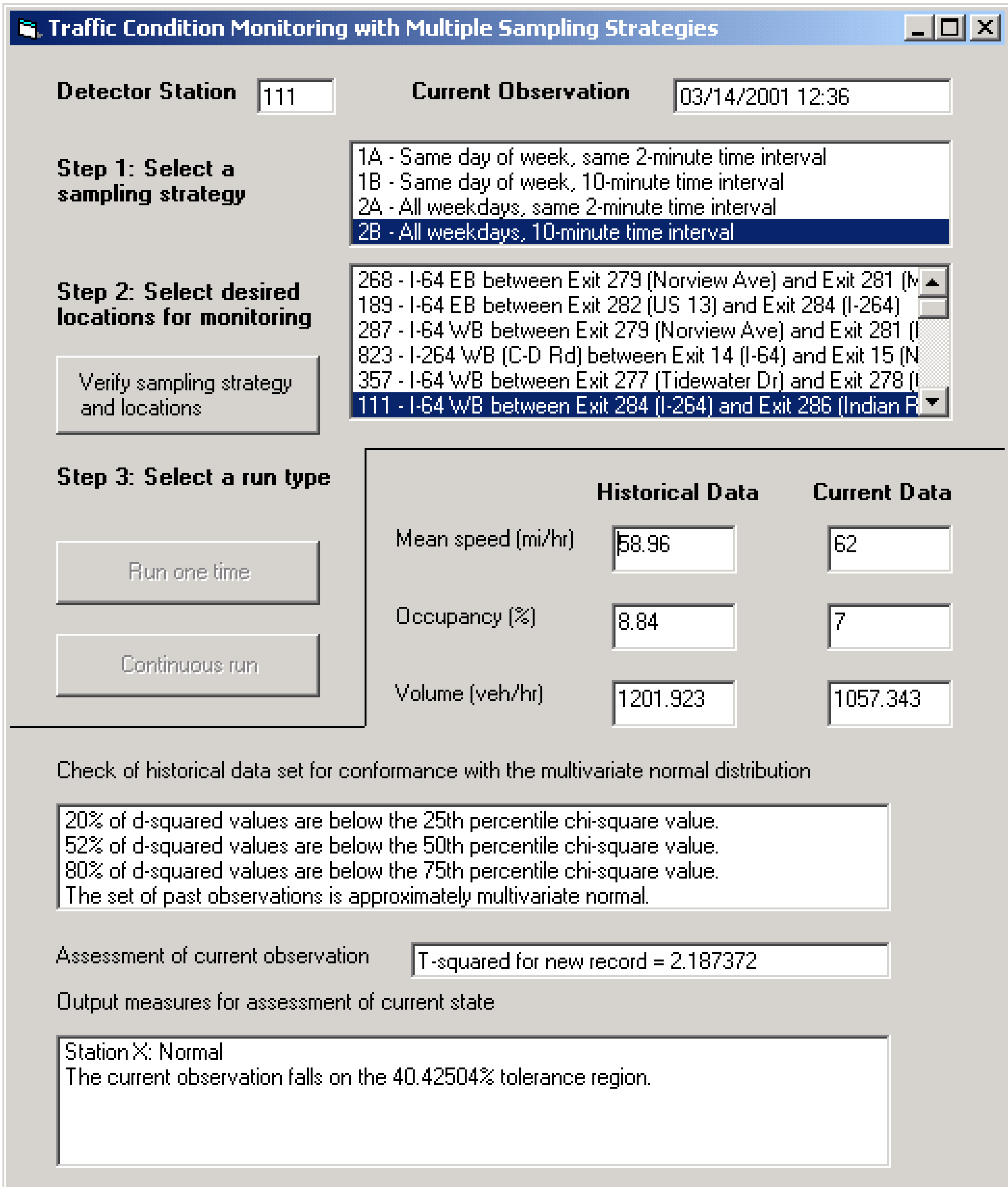

Figure 9. Prototype Traffic Management Program Interface

button to set the appropriate values for these variables. At this time the buttons under "Step 3: Select a run type" are enabled; the user can choose to have the program run one time or run indefinitely. In this case, the program ran one time and the relevant historical data and current observation data were displayed. Information pertaining to conformance with the multivariate normal distribution, T^2_k, and output measures 1 and 2 are displayed. Additionally, all of this information is written to a text file for archiving and future analysis purposes.

EVALUATION AND COMPARATIVE ANALYSES

The MSQC-based traffic condition monitoring methods developed herein have been evaluated using three categories of evaluation criteria: traffic condition assessment, program operating characteristics, and incident detection. Comparisons among evaluation measures as functions of the four sampling strategies and five sizes of historical data sets are made. The main goals of this evaluation are to show that the MSQC-based method of monitoring traffic conditions has the following characteristics:

- produces a characterization of traffic conditions that is informative and easily interpreted
- can operate in a real-time traffic management system environment
- has the potential to be developed into a technique for detecting incidents that can supplement its condition monitoring function.

Comparisons among the database sampling strategies were made to provide useful information for the selection of a sampling strategy for implementation in a traffic management system. The findings can aid decision makers in determining the amount of data generated in intelligent transportation systems that should be archived.

Measures of traffic condition assessment pertain to the probability that the MSQC-based method will characterize an observation of interest as representing abnormal conditions. Two measures were developed and applied: the average value of the confidence ellipsoid on which the observation of interest falls (output measure 2 of the prototype program) and the proportion of tested conditions that fall outside the 95 percent confidence ellipsoid. The impact of the size of the historical data set on these measures was also evaluated. Testing for statistically significant differences among the sampling strategies and historical data set sizes is also presented. Measures of program operating characteristics assess the performance of the prototype traffic management (real-time) program. Mean operating times among the sampling strategies are compared and statistically significant differences noted. Proportions of program executions that result in insufficient historical data are similarly compared; this is to demonstrate the impact of a limited historical database on each of the sampling strategies. Since measures of program operating performance are not critical to evaluation of the MSQC-based traffic condition monitoring method itself but instead to its ultimate implementation, the evaluation of program operating performance is presented in Appendix B. Measures of incident detection are presented since incident detection algorithms are the most commonly used automated traffic condition methods today, and the MSQC-based has the potential to perform this function as well. Detection rate and mean detection time are calculated for the MSQC-based method and for one of the most widely used incident detection algorithms, Modified California Algorithm 7, based on a set of accident-related traffic data.

Measures of Traffic Condition Assessment

In the MSQC-based condition monitoring method, the assessment of the state of traffic conditions can be captured in one value, the value of the confidence region, on which the current observation falls. This can also be called *normality level* and is described as output measure 2 of the prototype program. This section examines the mean values of the confidence ellipsoid on which the observations of interest fall. Results are presented by sampling strategy across all study sites, along with a statistical comparison of the accumulated results among the four database sampling strategies. The distribution of this value is also examined; the proportion of conditions tested that fall outside several confidence thresholds (90, 95, 99, and 99.9 percent) is also presented by sampling strategy across all study sites, along with a statistical analysis of the proportion of conditions tested exceeding the 95 percent confidence ellipsoid.

In order to analyze the differences among sampling strategies with regard to these two measures, the value of n (the number of observations included in the historical data set) was held at a constant value ($n=30$). However, to ascertain the impact of the size of the historical data set, the value of n was varied while the sampling strategy was held fixed. For strategy 1B, values of $n=20, 30, 40, 50,$ *and* 75 were selected in executions of the prototype program to assess the impact of this parameter on the two evaluation measures pertaining to traffic condition assessment.

Mean Values of Normality Level

Table 8 compares the mean values for output measure 2 among all runs of the prototype traffic management program by sampling strategy across all study sites and by study site. To analyze the impact of sampling strategy selection on this measure, the number of observations in the historical data set was held constant at $n=30$. Output measure 2 effectively represents the mean probability, for all runs using each sampling strategy, that the most recent observation comes from a population with a different mean than the population represented by the historical data set (for simplicity in discussion, the probability value is described as the probability that the most recent observation describes *abnormal* conditions), among sampling strategies. The value of output measure 2, for a particular condition tested, is equal to the value of $1-\alpha$ for the $100(1-\alpha)$ *percent* confidence ellipsoid, or tolerance region, on which the observation of interest falls. Among the four sampling strategies, the mean value of output measure 2 ranges between 0.58643 (for strategy 2B) and 0.62245 (for strategy 1A). Therefore, strategy 2B is least likely to characterize conditions as abnormal (with an average probability of 0.58643), whereas strategy 1A is most likely to do so (with an average probability of 0.62245). If the desired operation of the method is to be conservative, then based on this measure, strategy 2B is recommended.

Table 8. Abnormal Probabilities Across Sampling Strategies

	Mean Normality Level			
Sampling Strategy	**1A**	**1B**	**2A**	**2B**
Mean	0.62245	0.60301	0.58879	0.58643
Variance	0.10687	0.09728	0.09934	0.09768

Ideally, if the historical data were distributed exactly according to the multivariate normal distribution *and* if traffic patterns did not change over time due to population and resultant traffic growth as well as any seasonal effects, then the mean probability value would be 0.5. However, because only about 75 percent of the data sets conform to the multivariate normal distribution (achieving 100 percent conformance is ideal but asymptotic), and since the historical data sets are created by data captured over an extended period of time, mean probability values slightly higher than 0.5 are not surprising. In the test runs of the program, historical data sets consisted of $n=30$ historical observations that passed the screening procedure and fell within the 95 percent tolerance region as defined by the set itself. These historical data were typically captured over about 2 weeks for strategy 2B, 2 months for strategies 1B and 2A, and 8 months for strategy 1A; these time frames assume a small proportion of potential data fail the screening procedure or the 95 percent tolerance region check. As would be expected, the data set drawn from the shortest time span (strategy 2B) yielded the lowest mean probability value of the four strategies. Similarly, strategy 1A, which uses no more than one observation from each week thereby taking several months to yield a historical data set of $n=30$ observations, yielded the highest mean value. Most of the data used to generate historical data sets using strategy 1A are, therefore, more stale, then data sets generated using strategy 2B (with strategies 1B and 2A being in between) and, therefore, least appropriate to constitute a set of relevant historical data.

Statistically significant differences of the assessment of traffic conditions among the four sampling strategies were examined using statistical hypothesis testing. The differences in the mean value of output measure 2 (the value of α for the confidence ellipsoid on which the observation of interest falls) among the four database sampling strategies can be tested for significance by comparing the means of two populations. The test for the significance of the difference between two independent samples (t-test) in which the t-statistic is calculated and then compared to the Student's t-distribution to determine the level of significance has two cases, one in which the population variances of the two samples being compared (σ_1^2 and σ_2^2) can be considered equal, and one in which they are not considered equal. If it can be shown that, with some level of confidence, the null hypothesis that the two sample variances are equal can be rejected, then the more robust case of the t-test, in which the population variances are considered equal, can be applied.[18] To compare the values of the sample variances, the F-statistic was calculated for all pairwise comparisons among sampling strategies. In all six cases, the value of the F-statistic was not significant at the $\alpha=0.05$ level; therefore, the more robust case was applied.

The determination of significant differences between mean values of output measure 2 among the four sampling strategies results in six pairwise comparisons. In each comparison, the null hypothesis tested is H_0: $\mu_1 = \mu_2$ for the mean values of samples 1 (x) and 2 (y); the test statistic, t, is calculated as follows:

$$t = \left(\bar{x} - \bar{y} \right) / \sqrt{\frac{(n_x - 1)s_x^{\,2} + (n_y - 1)s_y^{\,2}}{n_x + n_y - 2}} \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \qquad (16)$$

In each pairwise comparison, the resulting t-statistic value was checked for significance at the $\alpha=0.05$ level, if so, then the difference between the two means is considered to be statistically significant. Four of the six pairwise comparisons yielded statistically significant differences: between strategies 1A and 1B, 1A and 2A, 1A and 2B, and 1B and 2B resulted in *t*-statistic values that were significant at the $\alpha=0.05$ level, allowing for ejection of the null hypothesis that the compared samples have the same means. Only when comparing strategies 1B and 2A, and 2A and 2B were the differences not statistically significant, as shown in Table 9.

Table 9. Results of Statistical Testing on Abnormal Probabilities

Strategies	*t*-Statistic	Level of Significance
1A vs. 1B	2.390	$\alpha < 0.05$
1B vs. 2A	1.907	$\alpha > 0.05$
2A vs. 2B	0.317	$\alpha > 0.05$
1A vs. 2A	4.065	$\alpha < 0.05$
1B vs. 2B	2.267	$\alpha < 0.05$
1A vs. 2B	4.429	$\alpha < 0.05$

This analysis of the mean values of output measure 2 (a function of the average value of the confidence ellipsoid on which the tested conditions fall) has shown that when comparing sampling strategy 1A to any of the other three strategies and when comparing strategy 1B and 2B, the differences are statistically significant. Additionally, when comparing the sample variances among the strategies, none of the differences are statistically significant. For implementation, with regard to the probability that tested conditions will be considered abnormal, strategy 1B, 2A, or 2B would be preferable to strategy 1A; however, insufficient statistical evidence exists to establish a preference among these three strategies.

Proportion of Conditions Tested Exceeding Critical Probability Values

Table 10 compares the proportion of tested conditions exceeding each member of a set of threshold values among sampling strategies. The threshold values used (90, 95, 99, 99.9 percent) are the same as those use in the method that describes the extent to which conditions are abnormal across a range of conditions (output measure 1). This table, in conjunction with Table 9, can be used both to describe the MSQC-based method's characterization of traffic patterns and to assess differences among the four sampling strategies' tendencies to describe traffic conditions in the real-time traffic condition monitoring application. The measures presented in Table 10 can be used to describe the variability (and therefore predictability) in traffic conditions for any combination of temporal characteristics. To analyze the impact of sampling strategy selection on this measure, the number of observations in the historical data set was held constant at n=30.

Table 10. Proportion of Conditions Tested By Tolerance Region

	Proportions			
Tolerance Region, Sampling Strategy	**1A**	**1B**	**2A**	**2B**
<90%	0.74018	0.75108	0.75765	0.76671
>90; <=95%	0.06936	0.06176	0.06152	0.05386
>95; <=99%	0.08367	0.07367	0.06896	0.07064
>99; <=99.9%	0.04587	0.04009	0.04378	0.03437
>99.9%	0.06092	0.07340	0.06810	0.07442

The differences among the proportions for each sampling strategy appear trivial when site-specific conditions are not considered; for example, the percentage of tested conditions that fall outside the 95 percent tolerance region/confidence ellipsoid varies from 17.9 percent (for strategy 2B) to 19.0 percent (for strategy 1A). This measure indicates that strategy 1A is most likely to classify conditions as abnormal and that strategy 2B is least likely to do so. Interestingly, the proportions exceeding each of the thresholds, for all strategies, are noticeably more than would be the case with an exact multivariate normal distribution. This tells us that the true distribution of the historical data sets used (and possibly traffic data in general and at other locations) is platykurtic, or flatter than true multivariate normal, with greater density at the tails of the distribution than true multivariate normal. As stated previously, the resultant proportions of conditions tested that fall outside the 95 percent tolerance region are greater than the 5 percent that would be expected if the data conform *exactly* to the multivariate normal distribution *and* if the data were not sampled over some extended period of time during which underlying traffic conditions may change.

The differences in proportions of tested conditions falling outside the 95 percent confidence ellipsoid, among the sampling strategies using the summary results for the ten study sites, were tested to determine if they were statistically significant. The test statistic, z, was calculated for each of the six pairwise comparisons, and the significance of the difference in proportions was determined, as described in the testing procedure for differences in proportions, employed in the multivariate normal conformance analysis. No statistically significant differences (significant at the $\alpha=0.05$ level) were found. It can be concluded that the proportion of tested conditions exceeding the 95 percent tolerance region does not vary significantly among the sampling strategies, and for implementation on a systemwide basis, this issue is not a factor in the selection of a sampling strategy.

As stated previously, the resultant proportions of conditions tested that fall outside the 95 percent tolerance region are greater than the 5 percent that would be expected if the data conform *exactly* to the multivariate normal distribution *and* if the data were not sampled over some extended period of time during which underlying traffic conditions may change.

Impact of Outlier Exclusion Using 95 Percent Confidence Ellipsoid

One step in the prototype program involves the generation of a 95 percent confidence ellipsoid based on all observations tentatively included in the historical data set and then the removal of any observations that fall outside this region. Additional observations are then drawn from the database and this procedure is then repeated until all observations included in the historical data set fall within the current set's 95 percent confidence ellipsoid. This practice was applied in the prototype program by excluding observations that fall outside the 95 percent confidence ellipsoid. It is possible, however, that application of this rule may impact the measures used to evaluate the MSQC-based traffic condition monitoring method, by excluding extreme values from the distribution that are not outliers or erroneous observations, and thus warrants investigation of its impact.

A limited analysis of the impact of the exclusion rule described on (1) conformance of historical data sets to the multivariate normal distribution and (2) the mean value of the confidence ellipsoid (output measure 2 of the prototype program) was undertaken. To isolate the impact of the exclusion rule, sampling strategy and historical data set size (*n*) were held constant. The prototype program was modified to eliminate the exclusion rule; using sampling strategy 1B and $n=30$, 3140 of the observations (across all ten study sites) for which the program was executed with the exclusion rule in place were analyzed again without the exclusion rule. The results are shown in Table 11.

Table 11. Impact of Outlier Exclusion on Selected Measures

	Proportions	
Measure	**With Exclusion Rule**	**Without Exclusion Rule**
Mean Value of Confidence Ellipsoid	0.58315	0.58638
Variance of Confidence Ellipsoid	0.09869	0.09945
Conformance with Multivariate Normal	0.79299	0.65286

It can be seen in Table 11 that exclusion of outlying observations (defined as those that fall outside the 95 percent confidence ellipsoid) has a noteworthy impact on conformance of the historical data sets to the multivariate normal distribution but a negligible effect on confidence ellipsoid mean value. The percentage of historical data sets conforming to the multivariate normal distribution was 79.3 percent when the exclusion rule was in effect but only 65.3 percent without the rule. To some extent, an improvement in conformance is expected when outlying observations are rejected. A negligible change (-0.3 percent) in the mean value of the confidence ellipsoid was observed with application of the exclusion rule. Although this analysis was undertaken on a small portion of the observations analyzed throughout this research, the results indicate that use of the exclusion rule improves conformance with the multivariate normal distribution with only negligible effects on measures of traffic condition assessment. Although the exclusion rule does not, of course, impact the underlying distribution of the observed data, it appears to improve conformance, it appears to improve conformance with the multivariate normal distribution of the data sets used in the analysis.

Incident-related Measures

The last category of evaluation measures for the MSQC-based traffic condition monitoring method consists of measures related to incident detection. This section summarizes the performance of the MSQC-based condition monitoring method and that of a commonly used incident detection algorithm, modified California algorithm 7, when applied to traffic data spanning 54 accidents. Although the detection of incidents is not the primary purpose of the MSQC-based condition monitoring method, because incident detection algorithms are the only widely used, highly automated condition monitoring methods in traffic management systems, even a limited comparison may prove informative.

The Modified California Algorithm 7 was selected for use in this comparative analysis for several reasons. It is one of the more widely studied and used incident detection algorithms; 6 of the 23 operating freeway management systems that responded to the survey of traffic management practices use either this particular algorithm or another member of the California family of incident detection algorithms. This algorithm employs a decision tree that compares functions of occupancy values captured at successive locations on the freeway network to threshold values to determine whether or not an incident may have occurred. The decision tree is shown in Figure 10.

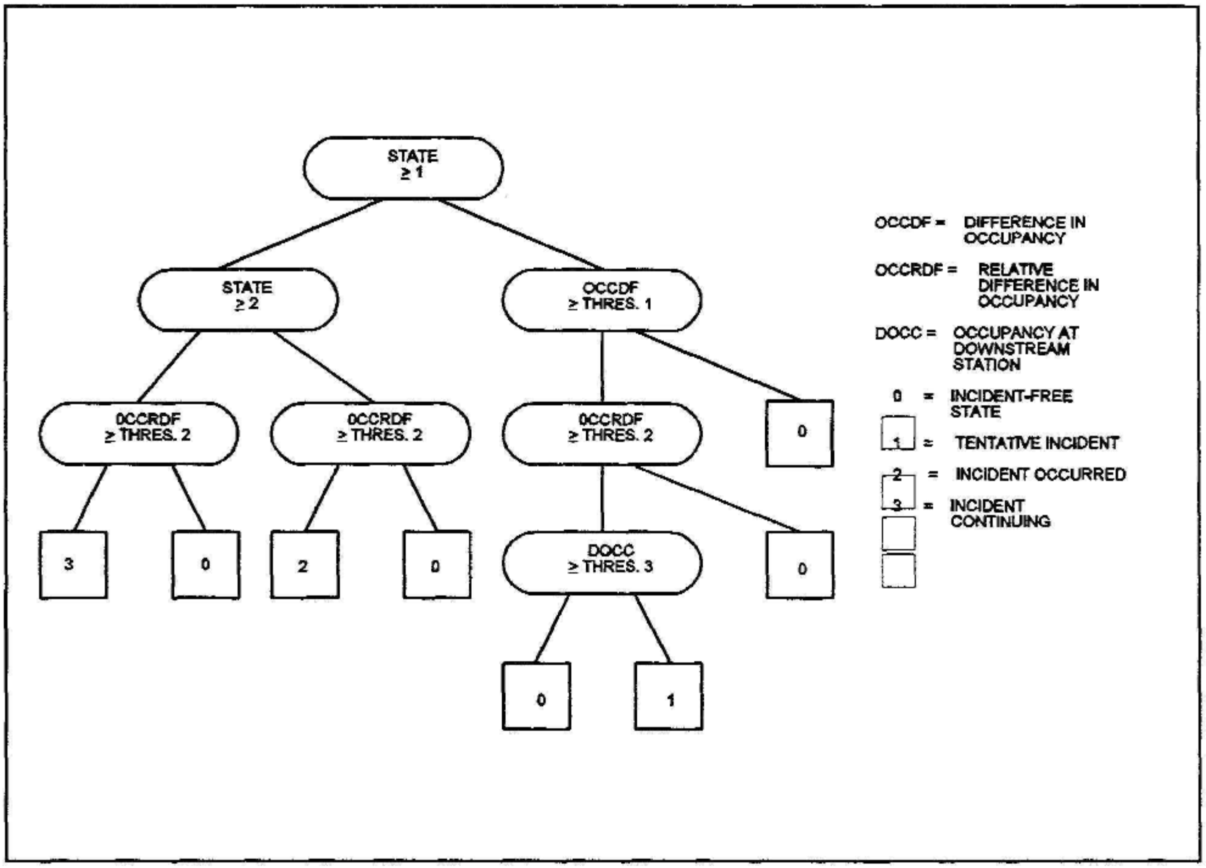

Figure 10. Decision Tree for Modified California Algorithm 7

The effectiveness of an incident detection algorithm is often measured by its ability to correctly and rapidly detect an incident based on traffic data. Three commonly used MOEs have evolved to quantitatively evaluate the performance of these algorithms: detection rate, mean time to detect (sometimes called detection time), and false alarm rate. The detection rate refers to the proportion of incidents that are detected by the algorithm over some specified period of time. Detection rate is typically expressed as the percentage of actual incidents that are detected by the algorithm. Mean time to detect is simply the average time taken for an algorithm to detect an incident after its occurrence. A variety of ways have been used to measure detection time; in some studies the time of incident occurrence is known and the detection time is the difference between that moment and the moment that an algorithm raises an incident alarm. In some cases only the time an incident was recorded by an operator is known, so the detection time as measured from that moment until detection by the algorithm is actually less than the true detection time. False alarm rate describes the extent to which false positives (alarms when no incident has actually occurred) are produced by an algorithm. The most commonly used definition for this measure is the percentage of total iterations of an algorithm's operation that result in false alarms during some specified period of time (sometimes called *off-line* false alarm rate). An iteration would typically occur every time the traffic data are aggregated and reported (for most traffic management systems, this value is 20, 30, or 60 seconds). Although these quantitative measures provide a picture of possible algorithm performance, some of the algorithms were developed and tested using simulated traffic data, thus avoiding the problem of poor data quality. A performance assessment of algorithms as implemented in the field provides a somewhat different picture, as discussed previously in the findings of the survey of traffic management systems.

As with most incident detection algorithms, decisions must be made regarding threshold values for their test parameters prior to operation. Ideally, these parameters would be selected based on site-specific studies of historical traffic data. The magnitude of this task contributes to its only fair performance in practice; similarly, threshold values were selected for this analysis based on values published in the literature. In a study during the development of the modified California algorithms in the 1970s, several sets of threshold values were tested for impacts on detection rates and false alarm rates, two of the three commonly used MOEs for incident detection algorithms. The values used for threshold 1 (difference in occupancy at successive upstream and downstream detectors at a given time, OCCDF) range from 8.1 to 26.6. For threshold 2 (the relative spatial difference between occupancies, OCCRDF), the values tested range from 0.30 to 0.36, whereas for threshold 3 (downstream occupancy, DOCC), the values tested range from 12.3 to 16.8.[5] In a more recent study, values for the three thresholds tested on data from Interstate 880 in Oakland, California ranged from 8 to 22 for threshold 1, 0.30 to 0.39 for threshold 2, and 13 to 17 for threshold 3.[21]

Based on the threshold values found in the literature, two sets of threshold values were selected for application to the Hampton Roads data used to assess the performance of Modified California Algorithm 7 on the accident data described below, and for the resulting comparison with the MSQC-based condition monitoring method. Set 1 uses values of 10, 0.30, and 10 for thresholds 1, 2, and 3 respectively; set 2 uses values of 15, 0.35, and 15. These threshold values are proximate to the midpoints of the ranges of threshold values noted in the studies cited.

Application of Incident Detection

A program written to execute the selected incident detection algorithm, using the historical database in a similar manner, was also used in this analysis. The 54 accidents for which related traffic data were analyzed were documented in the HRSTC incident database as having occurred between July 1, 2000 and December 31, 2000 on approximately 5 miles of Interstate 64 eastbound between Exit 279 (Norview Avenue) and Exit 284 (Interstate 264). Traffic data pertaining to types of incidents other than accidents (such as abandoned vehicles, roadway debris, and disabled vehicles) were not selected since accidents typically have a much greater impact on traffic conditions than do other incident types.

There are three commonly used MOEs for incident detection algorithms: detection rate, mean detection time, and false alarm rate, as described previously. Among the MOEs commonly applied to incident detection algorithms, detection rate and mean time to detect will be used to evaluate the performance of both the incident detection algorithm (modified California algorithm 7) and the MSQC-based condition monitoring method. Due to the fact that the candidate methods will likely detect many occurrences of abnormal conditions that are not related to incidents, false alarm rate is not used to evaluate the candidate methods. Many detections using the MSQC-based method that are not associated with incidents but are nonetheless occurrences of abnormal conditions would be misleadingly classified as false alarms using the typical definition of false alarm rate, thus misrepresenting the effectiveness of the MSQC-based method. It must also be noted that by excluding false alarm rate from the set of MOEs applied to an incident detection algorithm, the algorithm's performance is not fully assessed. Although the typical definition of false alarm rate, applied to the MSQC-based condition monitoring method, may generate many false alarms, the other advantages of its approach are unaffected.

The detection rate and mean detection time, as defined previously, were calculated for Modified California Algorithm 7 and for the MSQC-based condition monitoring method. A persistence check was applied to the MSQC-based methods as exists in Modified California Algorithm 7. Two cycles of this incident detection algorithm must be executed with the relevant conditions met, as illustrated in Figure 12, in order for an incident to be 'declared'; therefore, two consecutive iterations of the MSQC-based method in which the designated threshold (tolerance region) is exceeded must occur for an incident to be declared. Threshold settings of 95 and 99 percent in the MSQC-based method and threshold sets 1 and 2 as defined previously for modified California algorithm 7 were evaluated. The results in are summarized in Table 12.

Table 12. Incident Detection Performance Measures

Method	Mean Detection Time (minutes)	Detection Rate (%)
Mod Cal Alg 7, set 1	5.25	37.0
Mod Cal Alg 7, set 2	7.09	30.4
MSQC (95%)	5.48	57.4
MSQC (99%)	4.95	40.4

Findings

It can be seen from Table 12 that the MSQC-based condition monitoring method had a higher detection rate and a lower mean detection time than did the incident detection algorithm. With the detection threshold for the MSQC-based method set at the 95 percent confidence ellipsoid, the mean detection time was 5.5 minutes and the detection rate was 57.4 percent; with a 99 percent threshold the respective values are 4.9 minutes and 40.4 percent. As would be expected, as the threshold for which tested conditions are considered abnormal is increased, the detection rate decreases as a higher standard exists for declaration of an incident. With Modified California Algorithm 7, the mean detection time was 5.2 minutes and the detection rate was 37.0 percent with threshold set 1.

In evaluations of incident detection algorithms, the tradeoff between false alarm rate and detection rate is often presented. As algorithm parameters are selected to maximize the number of incidents detected, the number of false alarms typically increases; maximization of detection rate and minimization of false alarm rate are conflicting objectives. In the two studies of algorithm performance discussed, the detection rate for Modified California Algorithm 7 was between 8 and 62 percent, and between 20 and 59 percent.[5,21] A report published in 1993, in which many algorithms were evaluated, notes a detection rate of 67 percent as the highest noted in any effort.[1] The 37 percent detection rate observed in this analysis is consistent with those noted in the literature.

The values of these MOEs differ from those found in other evaluations of this algorithm due to the inconsistent spacing of detectors in the Hampton Roads system and the use of thresholds not calibrated to the specific locations considered. It must also be noted that, since false alarm rate was not computed due to reasons described previously, a complete picture of the MSQC-based method's performance simulating an incident detection algorithm is not presented. If the primary purpose of the MSQC-based condition monitoring method were to be the detection of incidents, evaluation of traffic data related to a much larger sample of incidents than considered herein, and comparisons with several incident detection algorithms would be warranted. Additionally, the issue of comparisons using false alarm rates would also need to be addressed.

CONCLUSIONS AND RECOMMENDATIONS

The automated monitoring of traffic conditions is made possible through the vast amounts of traffic data being collected and archived in traffic management systems. Automation of the monitoring function is increasingly imperative as the size and scope of traffic management systems are growing at a greater rate than that of the human resources available to operate these systems. As the archiving of traffic data has been become simple, affordable, and commonplace in recent years, these archival databases can be mined to develop a procedure for detecting erroneous data and to then allow for a comparison of current conditions with those that can be expected based on historical trends. A condition monitoring method that employs multivariate statistical quality control to perform this function, taking advantage of the relationships among commonly measured traffic variables, produces an assessment of the extent to which instrumented locations in the system are normal or abnormal. This research effort has developed and evaluated such a method; conclusions, recommendations for use in traffic management, and recommendations for further research are identified below.

Findings

The MSQC-based traffic condition monitoring method prototype was evaluated by comparing its performance among four database sampling strategies in six pairwise comparisons, using the following measures:

- mean value of the confidence ellipsoid / tolerance region (mean normality level)
- proportion of tested conditions falling outside the 95 percent tolerance region
- prototype program operating time
- proportion of historical data sets conforming to the multivariate normal distribution.

For the traffic condition assessment measure based on the mean value of the confidence ellipsoid - normality level (also output measure 2 of the prototype program), strategies 1B, 2A, and 2B exhibited significantly better performance than strategy 1A. A significant difference also existed between strategies 1A and 2A, but not between 1B and 2A, and between 2A and 2B. For the other traffic condition assessment measure, proportion of tested conditions falling outside the 95 percent tolerance region, no significant differences were found in the six pairwise comparisons. Among measures of program operating characteristics, strategies 1B and 2B yielded significantly lower mean operating times than strategy 1A; however, significant differences were not found in any of the other pairwise comparisons. When examining proportion of conditions tested with insufficient historical data (as a surrogate for required historical data), significant differences were found in all six pairwise comparisons. In order of proportion (from lowest to highest), strategy 1B fared best, followed by strategies 2B, 2A, and 1A, respectively. However, from a standpoint of practical significance, it can be argued that the differences among strategies 1B, 2A, and 2B are negligible and only that strategy 1A would not be recommended for implementation. Finally, no significant differences were found in the six pairwise comparisons for the measure selected to represent the characteristics of the historical data sets (conformance to multivariate normality).

For one particular value of historical data set size ($n=30$) and using one sampling strategy (strategy 1B, which arguably fared best in first set of comparisons described), a comparison with a commonly used incident detection algorithm (modified California Algorithm 7) was made using the following measures:

- mean detection time
- detection rate.

The performance of the MSQC-based method for these two measures was assessed using the 95 percent and the 99 percent confidence ellipsoids as the detection thresholds, and two threshold sets were applied in evaluation of the incident detection algorithm, resulting in four sets of values across the two measures of effectiveness. In this limited comparison, tests for statistically significant differences were not applied due to the relatively small sample sizes. The MSQC-based method (at both 95 and 99 percent) compared closely with one of the two incident detection algorithm threshold sets for mean detection time (approximately 5 minutes). The detection rate using the MSQC-based method at 95 percent had a notable higher detection rate than the other three method-threshold combinations. False alarm rate, commonly used in evaluations of incident detection algorithms, was not applied due to the fundamentally different nature of the MSQC-based method.

Several aspects of this research extend the state of knowledge and practice in traffic management. A new approach to traffic condition monitoring, and a prototype computer application, based on multivariate statistical quality control, have been developed. This approach is an alternative to most traditional approaches in that rather than classifying the state of the system into one of two conditions, conditions can instead be characterized over a series of discrete intervals or over a continuous range. In a related fashion, a procedure was developed to determine the significance of the deviation of an observation of interest from the mean value of a set of historically relevant data; this procedure can be applied outside the traffic management domain. Issues pertaining to archived databases, including organizational structure (sampling strategy) and data requirements (historical data sufficiency) have been studied. A wide range of database sampling strategies for the generation of relevant historical data sets was examined; based on intuitive understanding and graphical representations of typical traffic patterns, several of these strategies were selected for development and evaluation. The impact of these strategies on the necessary amount of archived traffic data was examined. Tests for the screening of traffic data that exploit the interdependence of measured traffic variables (mean speed, volume, and occupancy), based on traffic flow theory principles, have been developed. One of these tests, the average effective vehicle length test, was shown to detect several times more erroneous data records than many commonly used threshold value tests.

Recommendations for Traffic Management Practice

Implementation of the MSQC-Based Traffic Condition Monitoring Method

Implementation of the traffic condition monitoring method developed in this research can be addressed through a pilot study at a traffic management center, such as the HRSTC from which the data used in this research were collected. Some of the issues a pilot study should address include:

- the availability and quality of archived data
- the setting of program parameters
- the design of the user interface

- the potential benefits of distributed or parallel computing.

In any system implementation, access to archived data, and an assessment of the quality thereof, must be obtained. The impact of program parameter settings on output measures should be studied. Program user interface design improvements, such as the integration of a geographic information system or other visual representation, can be evaluated. Operating timesavings from the use of networked computers can also be ascertained.

In the development of a fully featured program to implement MSQC-based traffic condition monitoring in a traffic management system, sampling strategy and historical data set size will need to be determined by the user based on findings presented in this report. Regarding the extent to which the historical data sets conform to the multivariate normal distribution and the proportions of tested conditions exceeding the 95 percent confidence ellipsoid, no statistically significant differences were found among the strategies. Regarding the mean value of the confidence ellipsoid on which the tested conditions fall (output measure 2 of the prototype program), the statistical testing indicates that strategy 1A (same day-of-week, same 2-minute time interval) is not preferred because it yielded a significantly higher mean confidence ellipsoid value. Regarding program operations, no statistically significant differences exist between strategies 1B and 2B; however, in most cases, the operating times under these strategies were significantly shorter than for strategies 1A and 2A.

The impact of the depth of the historical database, as measured by the proportion of tested conditions for which insufficient historical data were not available, created statistically significant differences in all comparisons among strategies. By this measure, strategy 1B is most preferred, followed by strategy 2B and then strategy 2A. Strategy 1A resulted in significantly more occurrences of insufficient historical data than did the other three strategies studied. Site-specific considerations aside, the most preferred strategies are strategy 1B (same day-of-week, 10-minute time interval), and strategy 2B (all weekdays or weekend-days, 10-minute interval).

The ease of interpreting results of the ultimate implementation program is a factor in determining its utility and ultimate acceptance in a traffic management system. A graphic user interface that incorporates geographic information systems into the implementation-stage program could result in a more quickly and easily interpreted output. In development of an implementation-stage prototype program, the user interface could be addressed to maximize the speed of interpretation of the program's output measures.

Another hurdle to the highest level of acceptance is the reputation that other automated methods of monitoring traffic conditions have earned. Particularly, the difficulties associated with incident detection algorithms and the fact that they have been relegated to secondary status or discontinued altogether many systems may result in a resistance among some system personnel to any other automated method, even though the MSQC-based method is fundamentally different from most incident detection algorithms. To attain user acceptance, program users could be trained in a simulated "on-line" traffic management environment, such as at the Smart Travel Laboratory. During the training, the key differences between the MSQC-based method and other automated condition monitoring methods (e.g., incident detection algorithms) could be discussed.

Data Collection Improvements

Two issues regarding the implementation of data collection and management processes in traffic management systems have arisen through the data management phase of this research. The placement of a cap on reported mean speed data and the truncation of the decimal portion of occupancy values both have detrimental effects on programs that use such data for analyses. In the Hampton Roads system, the reported values of mean speed do not exceed 65 mph, although the true mean speed exceeds this value at some locations.[21] Such an artificial cap on the mean speed may skew the true distribution of the mean speed data, thereby impacting the generation of confidence ellipsoids based on the data. Additionally, a common practice in traffic management systems, including Hampton Roads, is to report occupancy data as whole numbers (in percent), with the decimal portion truncated, effectively rounding the measured values down to the next whole number. Although with higher traffic volumes the impact of this phenomenon is likely insignificant (e.g., the difference between 20.9 and 20.0 percent occupancy is very small), at lower volumes the effect is much greater (e.g., the difference between 1.9 and 1.0 percent occupancy is nearly a 2:1 ratio). Occupancy values less than 1 percent are typically reported as 0 percent, potentially skewing the distribution of the true data at this end of the range of possible values. This concern gave rise to Test 5 of the screening procedure and could be easily remedied if occupancy values were reported to the nearest tenth of a percent.

Data Archiving

In addition to developing recommendations for data collection practices, issues regarding data archiving practice have been identified. In this research, traffic data were available for a fourteen to eighteen month period prior to use of the prototype program, yet, many executions of the program did not produce sufficient data from the archived database to generate the necessary historical data sets. The selection of sampling strategy and historical data set size, of course, impact the required depth of historical data for the programs to execute successfully. For the applications developed in this research, using most combinations of sampling strategy and historical data set size (except for sampling strategy 1A, the depth of the available database was sufficient (see Table C.3) in at least 96 percent of program executions.

The procedure used in this research to assess historical data sufficiency can be applied to assess sufficiency for other traffic management applications. In the greater intelligent transportation systems arena, this issue has been raised and tradeoffs must be made between data available for archiving and data storage space (and cost) constraints. The operability of traffic management applications is, in part, a function of the depth of the databases accessed in such applications.

Recommendations for Further Research

Although this research has shown that MSQC provides an appropriate and informative framework for the automated monitoring of traffic conditions, many issues have been raised that may merit further research. Recommendations regarding potential enhancement to the MSQC-based traffic condition monitoring method include:

- *Adjustment for Recent Abnormal Observations:* Sustained abnormalities in the traffic stream can be caused by special localized events and may warrant adjustments to the manner by which the extent of normality is determined so that observations of current traffic conditions based on these events are not repeatedly classified as abnormal. Examples of such events may include traffic exiting a sports event or concert, patterns associated with holiday or tourist traffic, and in the case of the Hampton Roads study area, traffic patterns impacted by the arrival of a battle group at the Norfolk Naval Base. These events would cause traffic conditions to be classified as abnormal for several minutes or hours based on the concept applied in the prototype program, yet such conditions would quickly become normal based on a set of observations immediately prior to the current one. Such time periods should be noted in the archived database so that those observations can be excluded from subsequent historical data sets.

- *Empirical Approach to Setting Control Limits:* An examination of methods to establish levels of normality (for output measure 1 of the prototype program) based on the historical data may prove useful. The prototype program uses "round" values of 90, 95, 99, and 99.9 percent for tolerance regions as thresholds to define levels of abnormality for output measure 1. Establishing these thresholds based on the distribution of the historical data (perhaps using a clustering or nearest neighbor technique), although likely making the method more complex and increasing program execution times, would reflect any natural distinctions of abnormal extents that may exist in the data themselves.

- *Further Evaluation of the Impact of Excluding Observations Outside the 95 percent Confidence Ellipsoid:* The impact of excluding 'out-of-control' observations (i.e. outside the 95 percent confidence ellipsoid) from historical data sets on traffic condition assessment can be investigated to determine whether this practice has a significant impact on mean values of the confidence ellipsoid (output measure 2 of the prototype program) and on the size of the confidence ellipsoids. The impact of excluding 'out-of-control' observations from historical data sets on conformance of the historical data sets with the multivariate normal distribution also merits investigation as this practice may impact conformance.

- *Impact of the Multivariate Normality Assumption:* Although most of the historical data sets used in this research conform to the multivariate normal distribution, development of fully quantitative procedures for conformance testing and the impact of non-conforming historical data sets merit further investigation. The conformance test found in the literature was augmented and made more robust for use in this research; the extent to which existing procedures can be augmented impacts the validity of the multivariate normality assumption. Although it has been noted that MSQC "can handle minor deviations from the multivariate normal distribution,"[14] a quantification of this capability can assist in the evaluation of this assumption's validity.

- *Further Evaluation as an Incident Detection Technique:* The potential of the MSQC-based method to detect incidents could be further studied through a comparison of its performance with several types of incident detection algorithms using data from many different locations. Comparisons with statistics-based algorithms, such as the Standard Normal Deviate, are worthy of further study. Such a detailed and thorough analysis could produce a more complete picture of the MSQC-based method's incident detection capabilities.

- *Refinement of Parameters in the Average Effective Vehicle Length Test:* The AEVL test in the data screening procedure developed herein rejected over 8 percent of the pre-screened records in the data screening application test data set. Due to the magnitude of the test's impact on available data for analysis, it may warrant further investigation. Specifically, the minimum and maximum effective vehicle length thresholds as used in this research may not be optimal with regard to minimizing the rejection of accurate data while maximizing the rejection of erroneous data. On a site-specific basis, field studies of vehicle lengths could be undertaken to determine appropriate minimum and maximum values. The accuracy of detector measurements can also influence calculated effective vehicle lengths and could be addressed in such an effort.

- *Investigation of Causes of Abnormal Conditions:* Observations identified as abnormal conditions could then be further studied to ascertain events that may have caused such conditions and assign causes to them. Common causes of abnormal conditions, and the proportion of such conditions they cause, could be identified, possibly allowing traffic managers to anticipate abnormal conditions and plan accordingly.

ACKNOWLEDGMENTS

The authors thank the Federal Highway Administration for their support of this research, and the many VDOT employees who provided suggestions and support.

REFERENCES

1. Balke, K. N. 1993. *An Evaluation of Existing Incident Detection Algorithms.* Research Report 1232-20. Texas Transportation Institute, College Station.

2. Smith, B. L. and R. E. Turochy. 1999. Trends in Traffic Management Staffing Levels. *Proceedings of the 9th Annual Meeting, Intelligent Transportation Society of America* (CD-ROM), Washington, D.C.

3. Payne, H. J. and S. M. Thompson. 1997. *Development and Testing of Operational Incident Detection Algorithms: Technical Report.* Report No. R-009-97, Ball Aerospace and Technologies Corporation, San Diego, California.

4. Whitson, R. H., J. H. Buhr, D. R. Drew, and W. R. McCasland. 1969. Real-Time Evaluation of Freeway Quality of Service. In *Highway Research Record 289*, pp. 38-50. Highway Research Board, Washington, D.C.

5. Payne, H. J. and S. C. Tignor. 1978. Freeway Incident Detection Algorithms Based on Decision Trees with States. In *Transportation Research Record 682*, pp. 30-37. Transportation Research Board, Washington, D.C.

6. Dudek, C. L., C. J. Messer, and N. B. Nuckles. 1974. Incident Detection on Urban Freeways. In *Transportation Research Record 495,* pp. 12-24. Transportation Research Board, Washington, D.C.

7. Levin, M. and G. M. Krause. 1978. Incident Detection: A Bayesian Approach. In *Transportation Research Record 682*, pp. 52-58. Transportation Research Board, Washington, D.C.

8. Bhandari, N., F. S. Koppelman, J. L. Schofer, V. Sethi, and J. N. Ivan. 1995. Arterial Incident Detection Integrating Data from Multiple Sources. In *Transportation Research Record 1510*, pp. 60-69. Transportation Research Board, Washington, D.C.

9. Ivan, J. N. 1998. Data Fusion of Fixed Detector and Probe Vehicle Data for Incident Detection. In *Computer-Aided Civil and Infrastructure Engineering 13*, pp. 329-337. Blackwell Publishers, Malden, Mass.

10. Jacobson, L. N., N. L. Nihan, and J. D. Bender. 1990. Detecting Erroneous Loop Detector Data in a Freeway Traffic Management System. In *Transportation Research Record 1287*, pp. 151-166. Transportation Research Board, Washington, D.C.

11. Jeng, C.-Y. and A. D. May. 1984. *Monitoring Traffic Detector Information and Incident Control Strategies: A Survey.* Working Paper UCB-ITS-WP-84-2. Institute of Transportation Studies, University of California, Berkeley.

12. Cleghorn, D., F. L. Hall, and D. Garbuio. 1991. Improved Data Screening Techniques for Freeway Traffic Management Systems. In *Transportation Research Record 1320*, pp. 17-23. Transportation Research Board, Washington, D.C.

13. Mason, R. L. and J. C. Young. 1998. Why Multivariate Statistical Process Control? In *Quality Progress*, 31(12), pp. 88-93.

14. Johnson, R. A. and D. W. Wichern. 1998. *Applied Multivariate Statistical Analysis*. Prentice-Hall Inc., Upper Saddle River, N.J.

15. Fuchs, C. and R. S. Kenett. 1998. *Multivariate Quality Control: Theory and Applications.* Marcel Dekker, Inc., New York.

16. Turochy, R. E. and B. L. Smith. 2000. A New Procedure for Detector Data Screening in Traffic Management Systems. In *Transportation Research Record 1727*, pp. 127-131, Transportation Research Board, Washington, D.C.

17. Subramanyan, S. 2000. *The Feasibility of Developing Congestion Mitigation Measures That Incorporate Crash Risk*. Master's Thesis, University of Virginia, Charlottesville.

18. Snedecor, G. W. and W. G. Cochran. 1980. *Statistical Methods.* 7th edition, Iowa State University Press, Ames.

19. Mason, R. L., N. D. Tracy, and J. C. Young. 1995. Decomposition of T^2 for Multivariate Control Chart Interpretation. *Journal of Quality Technology*, 27(2), pp. 99-108.

20. National Bureau of Standards. 1972. *Applied Mathematics Series 55, Handbook of Mathematical Functions*. Washington, D.C.

21. Stephanedes, Y. J. and J. Hourdakis. 1996. Transferability of Freeway Incident Detection Algorithms. In *Transportation Research Record 1554*, pp. 184-195. Transportation Research Board, Washington, D.C.

APPENDIX A

SURVEY OF FREEWAY MANAGEMENT SYSTEMS

VIRGINIA TRANSPORTATION RESEARCH COUNCIL

SURVEY OF CONDITION MONITORING PRACTICES IN FREEWAY MANAGEMENT SYSTEMS IN NORTH AMERICA

Please note that the individual responses provided in completing this survey will be kept confidential. The information below will only be used for future contacts to clarify survey responses.

Name of system: ____________________

Name of person completing survey: ____________________

Title/position of person listed above: ____________________

Telephone number for future contact: ____________________

E-mail address: ____________________

If you have any questions about the survey, please contact:

Rod Turochy or Brian L. Smith

Phone: 804-293-1946
E-mail: turochy@virginia.edu

Phone: 804-293-1930
E-mail: briansmith@virginia.edu

Please return the completed survey in the enclosed prepaid envelope to:

Virginia Transportation Research Council
Attn: Brian L. Smith
530 Edgemont Road
Charlottesville, VA 22903

YOUR TRAFFIC MANAGEMENT SYSTEM TODAY

First, please tell us about the size and scope of your system today.

1. How many miles of freeway are included in your system? ________

2. What freeways are included in your system? ____________________________________

3. How many miles of non-limited access highways are included in your system? ________

4. How many traffic operators does your system use during a typical peak period? ________

Please tell us about the relative importance of condition information sources in your system today.

Please consider the importance in your freeway management operations of each of the following sources of traffic condition information. Please rate each item on a scale of 1 to 5, with 5 representing the highest level of importance, for your current operations. If a source listed below is not used, please indicate by circling N/A (not applicable).

	Importance:					
	Very low			Very high		
5. Detectors (inductive loops, acoustic, etc.)	1	2	3	4	5	N/A
6. State and local police or highway patrol	1	2	3	4	5	N/A
7. Freeway / safety service patrol	1	2	3	4	5	N/A
8. Telephone calls from private citizens	1	2	3	4	5	N/A
9. Broadcast media (such as TV, radio services)	1	2	3	4	5	N/A
10. Closed-circuit television cameras (CCTV)	1	2	3	4	5	N/A
11. Roadside emergency call boxes	1	2	3	4	5	N/A
12. Other (please specify) __________________	1	2	3	4	5	N/A
13. Other (please specify) __________________	1	2	3	4	5	N/A

YOUR TRAFFIC MANAGEMENT SYSTEM IN THE FUTURE

Please tell us about the planned size and scope of your system five to ten years from now.

14. Are there plans to extend the coverage area of your system during this period?

Please circle one of the following: YES (go to question **14a**)

NO (go to question **15**)

14a. How many miles of freeways will be included in the system upon completion of the planned expansion? __________

14b. How many miles of non-limited access highways will be included in the system upon completion of the planned expansion? __________

14c. How many traffic operators do you anticipate your system will utilize in a typical peak period upon completion of the expansion? __________

14d. When is the proposed expansion planned for completion? _______________

Please tell us about the relative importance of condition information sources in your system five to ten years from now.

Please consider the importance in your freeway management operations of each of the following sources of traffic condition information. Please rate each item on a scale of 1 to 5, with 5 representing the highest level of importance, for your future **operations. If a source listed below will not be used, please indicate by circling N/A (not applicable).**

	Importance: Very low				Very high	
15. Detectors (inductive loops, acoustic, etc.)	1	2	3	4	5	N/A
16. State and local police or highway patrol	1	2	3	4	5	N/A
17. Freeway / safety service patrol	1	2	3	4	5	N/A
18. Telephone calls from private citizens	1	2	3	4	5	N/A
19. Broadcast media (such as TV, radio services)	1	2	3	4	5	N/A
20. Closed-circuit television cameras (CCTV)	1	2	3	4	5	N/A
21. Roadside emergency call boxes	1	2	3	4	5	N/A
22. Other (please specify) ___________________	1	2	3	4	5	N/A
23. Other (please specify) ___________________	1	2	3	4	5	N/A

CONDITION MONITORING METHODS

Please tell us about your condition monitoring and incident detection practices.

24. Does your operations center use a visual display (such as a wall map) on which the status of freeway links or segments can be characterized using traffic data from surveillance equipment?

Please circle one of the following: YES (go to question **24a**)

IN PLANNING / DESIGN (go to question **24a**)

NO (go to question **25**)

24a. Please circle each type of data used to characterize freeway links on the display:

MEAN SPEED VOLUME OCCUPANCY OTHER_____________

24b. Please rate the role that the display plays in your overall assessment of traffic conditions on a scale of 1 to 5, with 5 representing the highest level of importance:

Importance:

Very low			Very high	
1	2	3	4	5

25. If your system has detectors, what is the time interval (in seconds) used for sampling of…

25a. Occupancy data _________________________

25b. Volume data _________________________

25c. Speed data _________________________

26. Is your system currently programmed to use incident detection algorithms to monitor traffic?

Please circle one of the following: YES (go to question **26a**)

NO (go to question **27**)

26a. Which algorithm(s) are used? ______________________________________

26b. Please rate your level of satisfaction with their performance on a scale of 1 to 5, with 5 representing the highest level of satisfaction:

Very dissatisfied				Very satisfied
1	2	3	4	5

27. If your system either currently uses incident detection algorithms or has used them in the past, please rate the performance of the incident detection algorithm(s) on a scale of 1 to 5, with 5 representing the best level of performance, with regard to the following measures of performance:

	Very poor			Very good	
Mean time to detect incidents	1	2	3	4	5
False alarm rate	1	2	3	4	5
Detection rate	1	2	3	4	5

28. In the future, do you anticipate that some form of automated or electronic incident detection will be used in the monitoring of traffic conditions?

Please circle one of the following: YES (go to question **28a**)

NO (end of survey)

28a. Please rate the role that automated condition monitoring will play in your assessment of traffic conditions on a scale of 1 to 5, with 5 representing the highest level of importance:

Importance:

Very low			Very high	
1	2	3	4	5

28b. Please tell us what type of condition monitoring methods will be employed (if incident detection algorithms are planned for use, please name them:

____

Comments and additional information:

If there is any other information that you think may help us in the development of condition monitoring methods, please tell us (and use additional sheets if needed):

____

____

If you would like a copy of the results of this survey, please place below the name and address to which you would like the results sent, or write "see page 1" if the address is the same as indicated at the beginning of the survey:

____

____

____

Thank you for your time and assistance in our research effort.

APPENDIX B

PROTOTYPE PROGRAM OPERATING PERFORMANCE

The purpose of evaluating the operating characteristics of the prototype traffic management program is to show that the program can operate in a real-time environment. Additionally, with possible refinements in program structure and availability of historical data, shorter program operating times than shown could reasonably be expected upon implementation in a traffic management system. The operating time of the program with a small number of locations is examined and comparisons made among the database sampling strategies, and the historical data requirements of each strategy will also be examined.

Prototype Program Operating Time

The mean operating times data described herein were taken from executions of the traffic management prototype program on a Dell Optiplex GX300 that has a 600MHz Pentium III processor. Table B.1 shows the trend in mean operating times for executions of the prototype traffic management program involving one location ($r = 1$) through observations at ten locations ($r = 10$) and variability of the execution times for each value of r. Mean operating times are shown in seconds. As can be seen from the table, there is a generally increasing trend in mean operating times, as would be expected, as r increases. There was almost always an increase in execution time for each increase in number of observations ($r = r + 1$); however, the high variances indicate that there is a lot of variability among executions for any particular value of r. This variability is due to the range in the number of observations that need to be examined in order to obtain n observations that can be included in the historical data set. Each observation selected for potential inclusion in the historical data set based on sampling strategy requirements is examined must also pass the data screening procedure *and* fall within the 95 percent tolerance region to be included.

Table B.1. Mean Operating Times for All Sampling Strategies

	Mean Operating Times (seconds)			
Locations, Sampling Strategy	**1A**	**1B**	**2A**	**2B**
1	1.2959	0.7613	0.8605	0.6696
2	2.0674	1.3966	1.7382	1.2748
3	2.5630	1.8352	2.4316	1.8943
4	4.0451	3.3532	3.7302	3.0715
5	3.6330	2.8815	3.5928	2.5699
6	5.1103	3.9363	4.8875	3.5778
7	5.3111	3.9183	4.8169	3.8808
8	6.3247	4.9986	5.8680	5.3159
9	5.3834	4.3386	4.9285	4.6509
10	5.7771	4.2096	4.5277	4.3494

The mean operating times for program runs of $r = 1$ *to* 10 locations were tested for statistically significant differences among the four sampling strategies. The mean operating times, for a particular value of r were compared between two sampling strategies using hypothesis testing procedures for comparing two populations to determine if statistically significant differences exist between the two means. In this case, due to the wide range in sample variances among sampling strategies for a particular value of r, the case of the "t-test" in which population variances are not the same was selected for the ensuing analysis. To test the null hypothesis, H_0: $\mu_1 = \mu_2$ for the mean values of samples 1 (x) and 2 (y); the test statistic, t, is calculated as follows:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} \tag{18}$$

Thus, for each value of r, six pairwise comparisons were made among the four sampling strategies, resulting in a total of 60 comparisons. In each comparison, the resulting t-statistic value was checked for significance at the $\alpha=0.05$ level; if so, then the difference between the two means is considered to be *statistically significant*. The results of the comparisons are shown in Table B.2. In this table are shown the values of the test statistic, t, for all comparisons; values significant at the $\alpha=0.05$ level are shown in italics. All comparisons between strategies 1A and 1B, and between strategies 1A and 2B, resulted in t-statistic values that were statistically significant at the $\alpha=0.05$ level. Seven of the ten comparisons each between strategies 1B and 2A, and strategies 2A and 2B, were statistically significant. When comparing between strategies 1A and 2A, only three of the ten comparisons produced statistically significant results, whereas between strategies 1B and 2B, no differences were statistically significant.

Table B.2. Results of Statistical Testing on Program Operating Times

	Value of Test Statistic, *t*					
Locations	**1A vs. 1B**	**1B vs. 2A**	**2A vs. 2B**	**1A vs. 2A**	**1B vs. 2B**	**1A vs. 2B**
1	*4.648*	-1.575	*3.381*	*3.840*	1.537	*5.614*
2	*3.522*	*-2.043*	*3.066*	1.614	0.920	*4.489*
3	*3.733*	*-3.353*	*2.989*	0.585	-0.419	*3.400*
4	*1.697*	-0.964	*1.975*	0.911	0.709	*2.759*
5	*2.686*	*-2.621*	*4.064*	0.136	1.331	*4.078*
6	*3.376*	*-2.673*	*4.033*	0.572	1.313	*4.852*
7	*5.130*	*-3.506*	*3.593*	*1.711*	0.155	*5.191*
8	*3.457*	*-2.590*	1.411	1.181	-0.817	*2.329*
9	*3.077*	*-1.921*	0.895	1.397	-0.962	*2.139*
10	*4.416*	-1.189	0.646	*3.267*	-0.592	*3.953*

It can be concluded that, in general, when comparing among the four sampling strategies, the operating time for strategy 1A is significantly longer than the operating times for strategies 1B and 2B. However, for most values of r, statistically significant differences in mean operating times do not exist when comparing any other combination of sampling strategies. These results make strategy 1A less favorable for implementation than any of the other strategies.

Incorporated in the mean prototype program operating times is the extraction of data from the database to create the historical data sets and resulting confidence ellipsoids / tolerance regions. Although this step was included in the program and performed for each location with each execution of the program, this step could be automated in a separate program that could operate on a regular interval external to the real-time condition monitoring program. In this scenario, files that contain the historical data sets and the resulting mean and variance-covariance matrices would be generated in a separate process from the operation of the condition monitoring program and then be accessed as needed by the condition monitoring program. Doing so would reduce the operating times shown in Table B.1 as the database connection and calculations performed on the extracted historical traffic data to generate these matrices would not be part of the condition monitoring program; reading these matrices from files previously generated would be a substantially faster operation.

Another opportunity for reducing program operating times lies in the use of distributed computing systems. In such an installation, the program can be executed on a network of several computers. Each computer would execute the program for a small subset of the total number of locations to be monitored, with all output information sent to one computer. Using several computers in this parallel manner should greatly reduce overall operating times from those observed in the prototype operation.

The operating times shown in Table B.1, for the computer equipment used in program development, represent a worst-case scenario in terms of operating times, yet still demonstrate that the prototype program operates sufficiently fast to be useful in a traffic management system to monitor traffic conditions. Even in a traffic management system responsible for monitoring 200 locations, the prototype program would be expected to operate in less than 2 minutes for any sampling strategy (and with the possible refinements described, these times would be expected to be much shorter), thereby providing an assessment of traffic conditions systemwide in a reasonable time frame.

Period of Available Historical Data

Another factor that can impact the performance of the condition monitoring program is depth of the historical database, that is, the time period for which data are included. Table B.3 compares the proportion of conditions tested, among sampling strategies, wherein insufficient data existed in the database to generate a historical data set of $n=30$ observations (the default number of observations required to complete a historical data set in the prototype program). The Smart Travel Laboratory's database is comprised of Hampton Roads traffic data beginning in July 1999, whereas the test executions of the real-time program occurred during October, November, and December of 2000. A historical data record selected according to a sampling strategy was only used in the generation of a historical data set if the data record passed the data screening procedure, *and* if the data point represented by the mean speed-volume-occupancy triple in the record fell within the 95 percent confidence ellipsoid/tolerance region as defined by the data set itself. For a specific sampling strategy, location, and time, if the beginning of the database was reached (i.e. all available data for inclusion in the historical data set have been considered) prior to the data set consisting of a user-specified number of observations, then that location is said to have insufficient historical data to create a historical data set with the required number of observations (n=30 in the prototype program).

Table B.3. Proportions of Conditions Tested with Insufficient Historical Data

Sampling Strategy	1A	1B	2A	2B
Proportions	0.22817	0.00731	0.03634	0.01245

An examination of Table B.3 shows that strategy 1A much more frequently did not find sufficient historical data (at least 30 observations to comprise the historical data sets in the case of the prototype program) than other strategies, finding insufficient historical data in 22.8 percent of cases. The other three strategies fared much better, with strategies 1B, 2A, and 2B finding insufficient historical data 0.73, 3.63, and 1.24 percent of the time, respectively. This can be due to the much greater depth in the database (further back in time) that strategy 1A requires to generate the historical data set, and that it is more likely that a candidate record for the historical data set will fail the control ellipsoid test using strategy 1A than with other strategies.

The first of these two key reasons is evident in that strategy 1A can select only one record per week for inclusion in the historical data set and, therefore, if n=30, requires at least 30 weeks of data (exactly 30 weeks if every candidate record passes the screening procedure and confidence ellipsoid test) or more. For weekdays, strategies 1B and 2A each require a minimum of only 6 weeks of historical data, and strategy 2B requires a minimum of less than 2 weeks of historical data. The second key factor in determining how well the strategies perform with regard to the ‘depth’ of the database required is each strategy’s propensity to classify conditions as abnormal, since this serves as a proxy to the likelihood that any record that is a candidate for inclusion in a particular historical data set would be classified as abnormal and therefore fail the control ellipsoid test. Table 8 indicates that an observation of interest (the current observation) is more likely to be classified as abnormal by strategy 1A than by any other strategy, whereas it has also been shown that strategy 1A classifies more observations as abnormal than does any other strategy.

The differences in proportions of tested conditions with insufficient historical data, among the sampling strategies using the summary results for the ten study sites, were tested to determine statistical significance. The test statistic, z, was calculated for each of the six pairwise comparisons, and the test for significance of difference in proportions was applied, as described in the section on multivariate normality conformance. In this application, the null hypothesis is that the proportion of tested conditions with insufficient historical data for sample 1 is equal to that of sample 2, which can be written as H_0: $p_1 = p_2$. The null hypothesis can be rejected if the test statistic, z, is significant at the $\alpha=0.05$ level. In all six pairwise comparisons, statistically significant differences (significant at the $\alpha=0.05$ level) were found. *Highly significant* differences (those at the $\alpha=0.01$ level) were found in all comparisons except that between strategies 1B and 2B, for which the difference in proportions was merely significant at only the

$\alpha \approx 0.03$ level. It can be concluded that statistically significant differences do exist among the database sampling strategies in the proportion of tested conditions for which insufficient data were available to generate the historical data sets. Based on these results, strategy 1A would *not* be recommended for implementation.

When selecting a sampling strategy for systemwide implementation, strategies 1B and 2B fare best with respect to availability of (and need for) historical data; no statistically significant difference exists between these strategies. However, this need only be a factor if the historical database covers a relatively short period (e.g., less than one year). The other chief implication for implementation considerations is that, due to its need for a much longer period of historical data than other strategies, strategy 1A will much more frequently fail to yield usable results and therefore is not preferred for implementation.