Draft Research Report
Agreement T2695, Task 64
Arterial FLOW

ARTERIAL PERFORMANCE MONITORING USING STOP BAR SENSOR DATA

by

Mark E. Hallenbeck
Director

John M. Ishimaru
Senior Research Engineer

Katherine D. Davis
Graduate Research Assistant

Jaime M. Kang
Research Engineer

Washington State Transportation Center (TRAC)
University of Washington, Box 354802
1107 NE 45th Street, Suite 535
Seattle, Washington 98105-4631

Washington State Department of Transportation Technical Monitor
Morgan Balogh
Regional Operations Traffic Engineer, Northwest Region

Sponsored by

Washington State Transportation Commission
Department of Transportation
and in cooperation with
U.S. Department of Transportation
Federal Highway Administration

July 2007

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO. WA-RD 681.1	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
4. TITLE AND SUBTITLE ARTERIAL PERFORMANCE MONITORING USING STOP BAR SENSOR DATA		5. REPORT DATE July 2007
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Mark E. Hallenbeck, John M. Ishimaru Katherine D. Davis, Jaime M. Kang		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Washington State Transportation Center (TRAC) University of Washington, Box 354802 University District Building; 1107 NE 45th Street, Suite 535 Seattle, Washington 98105-4631		10. WORK UNIT NO.
		11. CONTRACT GRANT NO. Agreement T2695, Task 64
12. SPONSORING AGENCY NAME AND ADDRESS Research Office Washington State Department of Transportation Transportation Building, MS 47372 Olympia, Washington 98504-7372 14 Doug Brodin, Project Manager, 360-705-7972		13. TYPE OF REPORT AND PERIOD COVERED Draft Research Report
		14. SPONSORING AGENCY CODE
15. SUPPLEMENTARY NOTES This study was conducted in cooperation with the University of Washington and the U.S. Department of Transportation		

16. ABSTRACT

The primary objectives of this project were to analyze the potential of a hypothesis regarding the ability of traffic data from basic loop sensors to represent approximate arterial traffic conditions (congestion); to develop a prototype analytical method to implement that relationship, and to determine requirements and other issues associated with future application of the method.

The hypothesis being tested was that there is a general relationship between a) the data from an arterial loop sensor, and b) the overall level of congestion on the arterial segment near the loop for that direction of travel. Specifically, lane occupancy percentage values from a loop sensor located just upstream from a stop bar for an arterial traffic signal were hypothesized to be a surrogate value for nearby arterial performance (specifically, congestion or traffic delay), and this relationship was proposed to be used to develop a basic arterial performance monitoring method to aid in planning and operational analyses, as well as potential real-time monitoring applications such as online traffic maps.

This research provided additional understanding of the feasibility of using basic sensor data to monitor performance on arterials, as well as outstanding issues. Given the results thus far, the use of green-plus-amber occupancy values from a stop bar loop shows the most promise among the options analyzed; furthermore, the hardware and analytical requirements are not restrictive. There is a need for additional testing of the robustness of this method before it can be implemented.

17. KEY WORDS Arterial, performance measurement, monitoring, loop detectors		18. DISTRIBUTION STATEMENT	
19. SECURITY CLASSIF. (OF THIS REPORT) None	20. SECURITY CLASSIF. (OF THIS PAGE) None	21. NO. OF PAGES	22. PRICE

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 Washington State Transportation Commission, Department of Transportation, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

CONTENTS

FIGURES

TABLE

EXECUTIVE SUMMARY

INTRODUCTION

The Washington State Department of Transportation (WSDOT) has a long history of developing freeway data archives and freeway performance monitoring capabilities. As a result of the WSDOT's research efforts, a performance monitoring methodology and tool set have been successfully developed and deployed to monitor instrumented freeways by using the WSDOT's FLOW system of detectors and data archiving processes. This project focuses on developing an analogous method to monitor arterial performance that would be useful from a traveler information perspective and a roadway operations and planning perspective.

The following were the objectives of this research project:

1) Explore potential methods to monitor arterial performance, using basic sensors and signal data.
2) Develop a prototype method for arterial monitoring and test it by using simulated data calibrated for a real-world corridor.
3) Evaluate the feasibility for deployment, and the potential functional specifications of signal systems and sensor systems required to make the method workable in practice.

This research effort to develop arterial performance monitoring techniques is an outgrowth of a longstanding WSDOT research effort to archive traffic data and use that archive to develop useful performance monitoring capabilities for Washington state's freeways. WSDOT has supported an ongoing loop-based freeway data collection and archiving effort in the Puget Sound region since the early 1980s. This sustained effort has produced one of the nation's most comprehensive freeway traffic data archives. In 1995, WSDOT initiated an applied research and development program with the Washington State Transportation Center at the University of Washington (TRAC-UW) to develop an analytical methodology and tool set that more fully exploits the potential of the data archive to support WSDOT freeway performance monitoring and management activities. The resulting freeway performance monitoring methodology, named TRACFLOW, was developed by TRAC-UW to analyze mainline general-purpose and HOV freeway lanes in instrumented regions of the state, using site-specific, corridor-wide, and trip-based performance metrics derived from the WSDOT FLOW archives. The TRACFLOW methodology has been successfully used to perform technical analyses for WSDOT as well as other public and private sector groups.

The TRACFLOW system was developed by following these guidelines:

- Use existing data whenever possible.
- Avoid reliance on potentially time-consuming and costly supplementary data collection.
- Use data based on readily available sensor technology.
- Produce performance metrics that are useful for a variety of users, including engineers, planners, and the traveling public.

The strategies used to develop that freeway monitoring methodology were also used as guidelines for this arterial performance monitoring research.

BACKGROUND

The approach used for this arterial monitoring research effort was designed to complement the objectives and methods used for the TRACFLOW freeway monitoring system; namely, to develop a versatile method by taking advantage of frequently used sensing hardware and data collection processes, and to produce metrics of performance that would be useful to engineers, planners, and travelers.

There are some aspects of arterial performance monitoring that differ from freeway monitoring, however. **First**, while urban freeways such as those in the Seattle area are often equipped with a dense network of sensors (loops, typically), arterials have a variety of instrumentation, ranging from sophisticated sensors and signal systems to virtually no sensors at all. Arterial performance on the significant number of lane-miles of arterials with less extensive sensors is more difficult to evaluate.

Second, signalized intersections introduce inherent variability in monitored arterial road performance, as vehicles stop for lights, then resume travel when the light changes; this characteristic must be considered when using typical sensors for performance monitoring. This project made use of loop sensors such as those commonly found on freeway and arterial systems. Their use was based on the plausible assumption that as congestion varies on a roadway, corresponding sensor values (particularly the occupancy percentage) will generally vary in response. This is not a new idea; this basic concept is currently used for freeway monitoring by TRACFLOW and others. The idea has also been used by others for arterials as well.

The difficulty with the loop occupancy concept is the interrupted nature of arterial flow, caused by the signal operations themselves and associated queuing and queue dissipation. Unlike freeway flows that commonly exhibit fluid state changes, at least under recurring congestion conditions, arterial flow, by its very nature, displays frequent uneven fluctuations in performance over time, even during uncongested conditions. The question is how simple sensor data can be best used to track the central tendency of performance over time, and also recognize the localized transient changes that might be associated with signal state changes rather than congestion.

Third, the relationship between interrupted flow roadway performance and loop detector occupancy is further complicated by the location of the detector relative to the traffic signals that are interrupting that flow. The basic problem is that the relationship between loop occupancy percentage and roadway performance is a function of where the vehicle queues form while waiting for red lights to turn green, relative to the location of the loop being used as a detector. For stop bar loops, when the signal indication is red, a single vehicle stopped at that light creates an occupancy value at the stop bar loop detector of 100 percent until the light changes to green and the vehicle departs. However, if the loop is placed 100 feet back from the stop bar, and only that one vehicle is present at the red light, loop occupancy is zero for that same roadway performance scenario. Thus, the same roadway performance can generate completely different loop occupancy values, depending on loop location.

TEST APPROACH

The approach taken in this research was to identify a common arterial scenario and arterial data collection device (inductance loop), then analyze the potential for using data that are available in that scenario to estimate arterial performance. The location and density of loops can vary among arterials; however, arterial signal loops are often installed at or near the stop bar to enable the signal system to detect the presence of vehicles at the signal. The stop bar loop thus represents a basic sensor that is relatively common because it provides the most basic detection—"a vehicle is waiting"—required to operate actuated traffic signals. From a roadway performance/traffic monitoring perspective, stop bar loops are not at the "best" location for monitoring volume or detecting the size of traffic queues, so use of stop bar loop detection as a performance monitoring tool represented a "most difficult case" situation to be tested as part of this research. Successful use of a commonly used sensor such as the stop bar loop would enable the resulting method to be used in a broader variety of scenarios.

Because the research involved analyses of various hypotheses regarding the effects of changes in sensor configuration and location on arterial monitoring capabilities, it was impractical to use an actual fixed loop installation in the field. Instead, a microscopic traffic simulation model of a real arterial segment was developed to determine whether a more complex relationship between volume, occupancy, and signal timing might exist, and whether that relationship could be detected by using a more sophisticated analysis technique.

TEST HYPOTHESIS

The hypothesis being tested was that there is a useful relationship between a) the data from an arterial loop sensor, and b) the overall level of congestion on the arterial segment near the loop for that direction of travel. Specifically, lane occupancy percentage values from a loop sensor located just upstream from a stop bar for an arterial traffic signal were hypothesized to be a surrogate value for nearby arterial performance (specifically, congestion or traffic delay).

TEST SCENARIO

The specific location modeled was a 1-mile segment of State Route 522 near Bothell, Washington, featuring three intersections: 61st Avenue NE, 68th Avenue NE, and 73rd Avenue NE (from west to east). Most of the analyses focused on the eastbound segment from 61st Avenue NE to 68th Avenue NE. The arterial scenario being modeled for this project used a common field sensor configuration, namely, inductance loops at each signal's stop bar.

The methods explored in this research assumed that a) there would be access to all data that <u>could be</u> detectable by sensors in that scenario, and b) there would be a mechanism available to collect and store those data. In many real-world cases, though, data from arterial loops are not stored after use, and even if they can be stored, the proposed research would require previously unused mechanisms of data storage and reporting, requiring new data archiving capabilities (software) within the traffic signal control software to support the monitoring process.

PRELIMINARY TASKS

Before the potential relationship between stop bar loop data and congestion levels was explored, a simulation model was constructed to represent a segment of State Route 522 near Bothell. The model was developed by using the VISSIM microscopic simulation environment and was calibrated by using the most recent available complete data for turning movements, average vehicle volumes, and signal timing plans.

The resulting simulation model was set to generate several values associated with delay (congestion). The per-vehicle intersection delay value is based on a comparison between modeled travel times of vehicles on a defined segment in the simulation network and the theoretical minimum travel time for that segment if no vehicles or signals are slowing traffic. The model also generates average and maximum queue lengths on specified arterial segments. These values were used to quantify the associated magnitude and variability of congestion. The delay value and queue lengths would also allow test cases to be more objectively and quantitatively compared to one another.

ANALYSIS TASKS AND RESULTS

After the SR 522 model was developed, three phases of analyses were performed. The first two phases consisted primarily of exploratory tests and model development tasks, while the third phase involved applying the results from the previous phases to perform a broader analysis of the test hypothesis.

Phases 1 and 2: Experimentation and Model Enhancement

Phase 1 focused on analyzing a single approach to a signalized intersection along the primary peak direction of traffic, specifically the eastbound approach to 68th Avenue NE during one hour of the PM peak period, from 4:00 PM to 5:00 PM. These model experiments were exploratory, designed to determine whether the test hypothesis warranted further research and how to best analyze and present the results.

The following data were collected for each test:

- loop volume and occupancy percentage from a loop located just upstream from the stop bar, at one second intervals.

- signal event (changes of signal status) data for the eastbound signal at the same intersection
- upstream segment average speed and delay, for a segment leading to the stop bar.

One key concern was that the stop bar occupancy percentage might be biased upward by the normal presence of vehicles that were stationary at the stop bar during a red light, making it difficult to distinguish between high occupancy values caused by stationary vehicles without congestion, and high occupancy values caused by vehicles moving slowly past the stop bar under congested conditions and on through the intersection during a green light. To minimize the ambiguity produced by that situation, the occupancy data during the signal's red phase were removed from the analysis process, and only a "30-second green time" stop bar occupancy, based on the occupancy during the first 30 seconds of green signal status in each signal cycle, was computed.

Phase 2 of the project analyses focused on model enhancements. Areas of enhancement were a) more accurate roadway geometry and lane configurations, b) updated signal timing plans, and c) more realistic vehicle movement logic. After the model enhancements were made, the Phase 1 experiments were re-run.

Phases 1 and 2: Observations

A review of the analyses in Phase 1 suggested the potential utility of the stop bar loop data to monitor performance. As hypothesized, there was a relationship between higher occupancy values and heavier congestion (as congestion grew, 30-second green time occupancy values grew). For lower to moderate congestion levels, the trend appeared promising. However, it was more difficult to determine the nature of that association at higher congestion levels because of a difficulty in achieving very heavy congestion scenarios in the model.

The Phase 2 experiments showed stop bar loop data patterns similar to those from Phase 1, with a similar limitation at the higher congestion levels. Sensitivity analyses to move the sensor location upstream from the stop bar showed that there were, as expected, distinct differences in the occupancy patterns as the sensor moved upstream. While use of sensor data at an upstream location might be promising for detecting queue lengths, the sensitivity analyses did not fully clarify where in a particular arterial segment the sensor should be, relative to the typical queuing patterns, block lengths, and intermediate traffic sources and sinks. To simplify subsequent experiments, the researchers determined that the focus of additional research in Phase 3 should remain on the stop bar loop.

Phase 3: Experiment Modifications

The results from the first two phases suggested that a relationship between loop data and congestion could be developed, although the results were still inconclusive. Phase 3 focused on a revision to the hypothesis, modeling and testing approach, followed by a broader series of tests. The most important change was to the hypothesis. The original test hypothesis focused on data from the first 30 seconds of green time of each cycle. In Phase 3, two variations to that original approach were explored: 1) The green time occupancy hypothesis was modified by using data from the entire green and amber time of each signal cycle, rather than just the first 30 seconds; and 2) for comparison, the tests were also performed using all data points in a cycle regardless of signal phase.

Another significant change addressed one of the limiting aspects of the previous analyses, namely the difficulty of determining the extent of the association between loop data and congestion during heavily congested conditions because of the difficulty in generating heavy congestion delay in the model. Specific causes were identified for this, such as transient upstream bottlenecks that developed over time and prevented traffic from arriving at the test segment. Several methods were developed to overcome this limitation and produce heavy congestion. These methods increased the upstream arriving volume to induce more congestion, removed upstream bottlenecks, and/or reduced downstream capacity to induce queuing backups.

After these features were implemented, a new matrix of cases was tested. The experiments continued to focus on the same eastbound peak direction approach to a signalized intersection (68th Avenue NE) that was used in phases 1 and 2. The matrix of cases represented various combinations of

- different arriving (upstream) volumes (to attempt to generate a range of traffic congestion conditions)
- different scenarios of arterial conditions (coordinated signals, uncoordinated signals, heavy turning movements, or a blocking incident/construction)
- different stop bar loop occupancy computation methods, i.e., use all data values, or use first 30 seconds of green values, or use all values during the green and amber phases

Similar tests were also performed at two other intersections on SR 522.

Phase 3 Observations

a) Overall Loop Data Patterns

The revised modeling approach produced results that were consistent with those in the previous phases, i.e., occupancy percentage grew as congestion grew. In addition, data from new heavy congestion test scenarios suggested the continuation of the pattern at heavier congestion levels, with volume throughput leveling off and beginning to decline as congestion grew. The various test cases showed similar patterns to one another, although some were less distinct than others.

b) Exploring Patterns through Data Aggregation

The results described above suggested a relationship between occupancy and congestion. Those results were based on data aggregated at the individual cycle level, i.e., each data point represented the average occupancy at the stop bar loop during the green and amber time of a single cycle. For a given test (a specific congestion level), the aggregated cycle data points were generally clustered in the volume-occupancy space; however, the clusters were not always compact and well defined. This variability made the determination of a clearer relationship between the occupancy and congestion more difficult. The researchers then hypothesized that if the data were aggregated at a higher level, any transient cycle-level variability of the loop data would have a less direct effect on the analysis and perhaps enable a clearer picture of the overall nature of the relationship between loop data and congestion to emerge. So, for each simulation test run, an aggregate average occupancy percentage for the entire 1-hour test period (after the initial start-up time of the run) was computed, rather than cycle-by-cycle values. As for the corresponding congestion indicator, the aggregate value used was the average per-vehicle speed (or alternatively, delay) associated with the arterial segment upstream from the stop bar.

When those two aggregated variables (occupancy and speed) were tracked for each test case, and the various test outputs were combined, the results still show some variability, particularly at the higher congestion levels (slower speeds), but the overall pattern showed an upward trend in occupancy percentage as a function of congestion (represented by average speed), as proposed in the original hypothesis.

c) Using Occupancy to Estimate Congestion

The relationship between occupancy and speed indicated above suggested that average occupancy could be used as a general indicator of associated congestion levels. For example, one approach would be to subdivide the occupancy range, with each subrange corresponding to a different congestion level (based on speed). For example, a Web-based display of arterial conditions might show congestion in three categories: light, moderate, and heavy. In that case, the occupancy range would also be split into three subsets, one for each congestion category. As traffic conditions varied over time, the occupancy would be tracked, and the congestion level corresponding to that occupancy value would be displayed.

The use of such a relationship could involve the following steps:

1) Produce occupancy vs. speed data, using a particular occupancy computation method (e.g., using data during green and amber signals). Verify that the data show an upward trend in occupancy together with a downward trend in speed.
2) If the occupancy data have some variability, consider smoothing the data to produce a central trend of the occupancy data (vs. congestion) that is less influenced by fluctuations. (See note below.)
3) Define congestion categories on the basis of speed on the arterial segment being analyzed. For example, the light/moderate/heavy website described above might use categories based on Highway Capacity Manual guidelines for Level of Service standards on that type of arterial. Or, a more direct approach might be to simply specify speed ranges that are either consistent with other existing applications, or coincide with existing local performance standards. Simple speed categories also have the benefit of being more easily interpretable by travelers.
4) Determine the range of occupancy values that correspond to each congestion category defined in step 3. Do this by looking at the distribution of congestion delay values (speeds) across all the test cases and dividing them into groups by congestion level (i.e., each specified speed range), then looking at the corresponding occupancy range. The result is a functional relationship between occupancy ranges and speed/congestion ranges (e.g., LOS A corresponds to occupancy values of between 0 and N percent). Once this function has been established, it can then be applied to a performance monitoring application.

Note that in this process, the development of a function based on subdivisions of the occupancy range requires the existence of a one-to-one relationship between occupancy and congestion, i.e., occupancy should grow monotonically as congestion grows (or as speeds slow). In reality, though, some variability of the field data might occur. That is why some type of smoothing operation may be desirable, one that removes transient spikes but still displays the overall character and trend of the data. Although this smoothing approach helps facilitate the specification of the threshold values for each congestion category, the precision of the category definitions will be reduced, something that should be considered when defining the categories.

d) Effects of Occupancy Computation Methods

The alternatives tested for aggregating occupancy data for each signal cycle were as follows:

1) average all occupancy data during each cycle (one-second data)
2) average all occupancy data during the first 30 seconds of the relevant (e.g., eastbound thru-traffic) green phase of each cycle
3) average all occupancy data during the entire green and amber phases of each cycle.

A comparison of the three methods showed that the "all occupancy" method produced data that were clustered around a combination of high occupancy values and low volume values, as one would expect given the method's inclusion of the red phase data, when vehicles are stopped at the stop bar and no vehicles are moving across the stop bar. The 30-second green method appeared to noticeably clarify the pattern of occupancy versus congestion. The green+amber method produced the clearest association between occupancy and congestion, providing more tightly clustered data for a given test case, and clearer results at the heavy congestion levels. (Unless otherwise noted, all results described throughout this report are based on the green+amber method.)

e) Results at Other Locations

Similar tests were also performed at two other intersections, 61st Avenue NE and 73rd Avenue NE, both in the eastbound direction. The results at the two locations were consistent with those from the original test location, showing data points that either followed the pattern of data from the primary test location or were a logical extension of the pattern to higher congestion levels.

f) Variability of Occupancy during the Reporting Process

Given the stop-and-go nature of traffic on signalized arterials, it is not surprising that the analyzed green and amber data showed fluctuations in occupancy values over time. At low to moderate congestion levels, average occupancy values per signal cycle tended to be more clustered, varying more smoothly over time, as one would expect. At heavy congestion levels, associated occupancy values appeared to vary more noticeably over time.

This variability becomes an issue when one applies a proposed occupancy-based monitoring method. Namely, how should the method include this variability when arterial performance is reported? Should short-duration oscillations be considered useful indicators of performance, or should they be considered transient values that distract from a more important goal of showing the central tendency of the traffic conditions? There are several options to address this:

- Aggregate the occupancy data over time to remove the influence of short-term oscillations that are cyclical, e.g., associated with the signal cycle.
- Use broader congestion categories that cover larger subsets of the occupancy range, so that variability of occupancy from cycle to cycle is less likely to cause frequent changes in reported performance. However, the result would also be less specific.
- Reclassify the congestion categories to include the variability, e.g, develop new categories that represent transition values between existing categories.

- Do not change the original data.

IMPLEMENTATION REQUIREMENTS

The method described above would require the following:

Supporting Data:

- Stop bar sensors capable of producing occupancy percentage values at the desired level of frequency
- A data storage capability (or data transfer capabilities to a central facility)

Method:

- A specified congestion categorization approach (e.g., speed ranges)
- Threshold occupancy values for each congestion category

Processing:

- Software to smooth data as required.

The supporting data are producible by a basic loop sensor at the stop bar. The implementation software and associated parameters would have to be developed or specified by the user. Given the relatively straightforward algorithms employed in this method, the processing software should be relatively inexpensive to develop.

A more likely upgrade requirement might be an archiving capability. The proposed method requires some storage and processing, performed centrally or on-site. For arterial networks that either did not store their data locally at all, or did store the data but did not centrally archive the values, some mechanism to transfer data would also be required. Either way, the data would need to be archived to support the desired performance monitoring activities.

The data types required would likely involve a change of traffic signal controller software to record volume/occupancy by using variable time frames, and to store along with those timeframes the length of the actual green+amber condition. (That is, our recommended approach to performance monitoring is to use stop bar statistics of volume and occupancy collected only when that phase is green and amber. To do this would require the traffic signal system to no longer use a "fixed time" reporting framework but, instead, one that varied with signal phase lengths.) This would be particularly important when some type of adaptive traffic control was used (including actuated and semi-actuated traffic signals and signal timing plans) that did not have fixed phase lengths.

The good news is that this new capability to support arterial monitoring would allow traffic signal engineers to not only examine the level of congestion present but also determine how flexible signal timing algorithms were actually being used in the field and thus might be modified to improve congestion. The result would be a more robust capability for measuring not only how congested each approach had become but how arterial signal timing plans were actually being used. This would provide engineers with detailed information that could be used to routinely tune timing plans to decrease delay without having to pay for new, short duration traffic data collection.

IMPLEMENTATION ISSUES

While the tests conducted thus far suggest that this method is potentially useful, there are other methodology development issues that should be considered prior to implementation.

Method Calibration and Robustness

The results described in this report were based on simulation of a typical 1-mile arterial section. While the analyses suggest that the results were consistent for different intersections and approach directions in this single model, it would be desirable to perform additional tests to further validate and calibrate the proposed monitoring methodology.

Arterial Segment Definition

The first step in the monitoring process described previously requires development of an occupancy versus speed data set. However, while the occupancy value refers to a specific point (the stop bar loop), the speed value could be associated with any segment. In the initial tests, we chose to use the intersection spacing (i.e., a segment from the upstream intersection to the intersection in question). However, because those spacings can vary, the resulting segment length being represented by the single stop bar loop will vary also.

One approach would be to determine whether there was a well defined and predictable family of occupancy versus speed curves based on intersection spacing, i.e., "If the spacing is X feet, the corresponding occupancy versus speed curve is…". The threshold values would then be based on that particular curve. Another alternative would be to identify significant discrete drops in the speed curve of the occupancy versus speed graph (something that was in fact observed in the data) and use the corresponding occupancy values. This would essentially allow the speed curve to dictate the congestion categories, and take advantage of the fact that regardless of segment length, the relative speed curve stays the same, i.e., as length changes, the overall speed curve shifts up and down, but relative patterns do not change.

Stop Bar Limitations

It may be that there are limitations in the ability of the stop bar loop alone to estimate overall arterial performance as opposed to only "intersection approach performance." The approach described in this report could be a good "arterial performance" metric as long as delays are only intersection based, and not from mid-block occurrences. In addition, because we only have measurements at the intersection, we can say that the signal has failed, but we cannot say that one is likely to sit through one, two, three or more light cycles to get through that signal. One alternative that might provide more overall monitoring capabilities might be to look at the volume associated with each green phase, the green phase length, and the number of signal cycles in a row that the cycle has shown “failed” performance levels.

CONCLUSIONS

The primary objectives of this project were to analyze the potential of a hypothesis regarding the ability of traffic data from basic loop sensors to represent approximate arterial traffic conditions (congestion); to develop a prototype analytical method to implement that relationship, and to determine requirements and other issues associated with future application of the method. This research provided additional understanding of the feasibility of using basic sensor data to monitor performance on arterials, as well as outstanding issues. Given the results thus far, the use of green+amber occupancy values from a stop bar loop shows the most promise among the options analyzed; furthermore, the hardware and analytical requirements are not restrictive. There is a need for additional testing of the utility and robustness of this method before it can be implemented.

INTRODUCTION

The Washington State Department of Transportation (WSDOT) has a long history of developing freeway data archives and freeway performance monitoring activities. As a result of the WSDOT's research efforts, a performance monitoring methodology and tool set have been successfully developed and deployed to monitor instrumented freeways by using the WSDOT's FLOW system of detectors and data archiving processes.

Analogous monitoring capabilities are important and useful for the arterial component of the state road network. Such capabilities are also beneficial to freeway planners and engineers, given the inherent interaction between the freeway and arterial networks. For a traveler considering a trip that uses both freeways and arterials, the distinction between the two is an artificial one; ultimately, performance of an overall trip is the prime concern, not simply performance of the freeway portion of a trip. Therefore, a method of monitoring arterial performance that complements the existing freeway monitoring techniques would be desirable, not only from a traveler information perspective but also from a roadway operations and planning perspective.

The following were the objectives of this research project:

4) Explore potential methods to monitor arterial performance, using basic sensors and signal data.
5) Develop a prototype method for arterial monitoring and test it by using simulated data calibrated for a real-world corridor.
6) Evaluate the feasibility for deployment, and the potential functional specifications, of signal systems and sensor systems required to make the method workable in practice.

This research effort to develop arterial performance monitoring techniques is an outgrowth of a longstanding WSDOT research effort to archive traffic data and use that archive to develop useful performance monitoring capabilities for Washington state's freeways[1]. WSDOT has supported an ongoing loop-based freeway data collection and archiving effort in the Puget Sound region since the early 1980s. This sustained effort has produced one of the nation's most comprehensive freeway traffic data archives. In 1995, WSDOT initiated an applied research and development program with the Washington State Transportation Center at the University of Washington (TRAC-UW) to develop an analytical methodology and tool set that more fully exploits the potential of the data archive to support WSDOT freeway performance monitoring and management activities. The resulting freeway performance monitoring methodology, named TRACFLOW, was developed by TRAC-UW to analyze mainline general-purpose and HOV freeway lanes in instrumented regions of the state, especially urban areas, by using site-specific, corridor-wide, and trip-based performance metrics derived from the WSDOT FLOW archives. TRAC-UW continues to enhance, test, and support this methodology and tool set.

The TRACFLOW methodology has been successfully used by TRAC-UW to perform technical analyses for WSDOT as well as other regional and state agencies. Examples include the yearly WSDOT Puget Sound freeway performance monitoring analysis, the yearly WSDOT Puget Sound HOV lane network evaluation, and travel time and HOV performance analyses in the yearly congestion component of the WSDOT Gray Notebook. These tools are also used by analysts at the WSDOT, as well as by other public and private sector groups.

The TRACFLOW system was developed by following these guidelines:

- Use existing data whenever possible.

[1] The freeway monitoring discussion is an abridged version of a background discussion in "FLOW Analysis and Advanced Technology Support", a TRAC proposal to WSDOT for FY 2008-2009, May 2007.

- Avoid reliance on potentially time-consuming and costly supplementary data collection that would reduce the cost-effectiveness of freeway monitoring and affect the ability to produce results in a timely manner.
- Use data based on readily available sensor technology. In this case, basic inductance loops were used to produce the WSDOT data archive.
- Produce performance metrics that are useful for a variety of users, including engineers, planners, and the traveling public.

These strategies used to develop the freeway monitoring methodology were also used as guidelines for the arterial performance monitoring research.

METHODOLOGY

BACKGROUND

The approach used for this arterial monitoring research effort was designed to complement the objectives and methods used for the freeway monitoring system. In particular, the focus was to take advantage of frequently used sensing hardware and data collection processes, in order to expand the versatility of the resulting arterial monitoring method, balanced with the desire to produce metrics of performance that would be useful to engineers, planners, and travelers.

There are some aspects of arterial performance monitoring that differ from freeway monitoring, however. **First**, while urban freeways such as those in the Seattle area are often equipped with a dense network of sensors (loops, typically), arterials have a variety of instrumentation, ranging from sophisticated sensors and signal systems to virtually no sensors at all. While well-instrumented arterials can collect a broad array of data, making them easier to monitor, arterial performance on the significant number of lane-miles of arterials with less extensive sensors is more difficult to evaluate.

Second, signalized intersections introduce inherent variability in monitored arterial road performance, as vehicles stop for lights, then resume travel when the light changes. Nevertheless, this project made use of sensors such as those commonly found on freeway systems. Their use was based on the plausible assumption that as congestion varies on a roadway, corresponding sensor values (e.g., the occupancy percentage) will generally vary in response. This is not a new idea. This basic concept is currently used for freeways. Its use for arterials has also been discussed by other researchers. Perrin, et. al. used occupancy data from stop bar detectors, combined with upstream system detector data, to estimate vehicle/capacity ratio and arterial level of service[2]. In Bellevue, Washington, occupancy values from loops located approximately 100 feet upstream from signal stop bars are used to estimate arterial segment performance for an online traffic map[3].

The difficulty with the concept is the interrupted nature of arterial flow, caused by the signal operations themselves and associated queuing and queue dissipation. Unlike freeway flows that commonly exhibit fluid state changes, at least under recurring congestion conditions, arterial flow, by its very nature, displays frequent uneven fluctuations in performance over time, even during uncongested conditions.

Third, the relationship between interrupted flow roadway performance and loop detector occupancy is further complicated by the location of the detector relative to the traffic signals that are interrupting that flow. The basic problem is that the relationship between loop occupancy percentage and roadway performance is a function of where the vehicle queues form while waiting for red lights to turn green relative to the location of the loop being used as a detector. For stop bars, when the signal indication is red, a single vehicle stopped at that light creates an occupancy value at the stop bar loop detector of 100 percent until the light changes to green and the vehicle departs. If the loop is placed 100 feet back from the stop bar, and only that one vehicle is present, loop occupancy is zero for that same roadway performance scenario. Thus, depending on loop location, the same roadway performance can generate two completely different loop occupancy values.

As an aside, the interrupted nature of arterial flow introduces some interesting issues associated with the display of that variability in a way that is useful for traveler information purposes. For example, is it important to track performance at a small update

---

[2]Joseph Perrin, Peter T. Martin, Brad Coleman, "Monitoring Commuter Congestion on Surface Streets in Real Time", Transportation Researcb Record, Volume 1811, 2002

[3] http://trafficmap.cityofbellevue.net/FAQ.htm

increment (e.g., every 30 seconds) to show the variability, or is it more useful to track performance in the longer term (e.g., every signal cycle or every 5 minutes) to show the trend? How does the answer to this question change if the data are used for planning or operations purposes?

TEST APPROACH

The approach taken in this research was to identify a common arterial scenario and arterial data collection device, then analyze the potential for using data that are available in that scenario to estimate arterial performance. Ideally, an arterial performance monitoring method would enable basic performance monitoring for arterials without requiring sophisticated instrumentation; that method would be usable with different levels of sensor instrumentation, i.e., it would not be dependent on sophisticated data collection installations. With that in mind, this research explored the feasibility of developing an arterial monitoring method that only requires data that are commonly available and widely used, i.e., loop data.

While the location and density of loops can vary among arterials, arterial signal loops are often installed at or near the stop bar to enable the signal system to detect the presence of vehicles at the signal. The stop bar loop represents a basic sensor that is relatively common because it provides the most basic detection—"a vehicle is waiting"—required to operate actuated traffic signals. From a roadway performance/traffic monitoring perspective, stop bar loop locations are not the "best" location for monitoring volume or detecting the size of traffic queues, so use of stop bar detection as a performance monitoring tool allowed a "most difficult case" situation to be tested as part of the research. Successful use of the stop bar loop would also enable the resulting method to be used in a broader variety of scenarios.

(At the same time, it is not an absolute requirement that a monitoring method operate at the lowest common denominator. If incremental enhancements to data collection and archiving capability in the field could produce a significant enhancement in monitoring capabilities, this might be a cost-effective approach for many communities. Enhanced monitoring capabilities could be achieved through upgrades of existing systems or new sensor installations. This research primarily explored methods that can use the more common forms of data available for arterials, but it is hoped that if a basic method is developed, it could then be used as the basis for defining software and hardware specifications for more enhanced future traffic sensor installations.)

Because the research involved analyses of various hypotheses regarding the effects of changes in sensor configuration and location on arterial monitoring capabilities, it was impractical to use an actual fixed loop installation in the field. Instead, a microscopic traffic simulation model of a real arterial segment was developed to test various hypotheses. The experiments performed in this project were designed to determine whether a more complex relationship between volume, occupancy, and signal timing might exist, and whether that relationship could be detected by using a more sophisticated analysis technique. The experiments involved varying the level of congestion on a modeled arterial, then observing corresponding modeled stop bar occupancy values to analyze the nature of the relationship between the two attributes.

TEST HYPOTHESIS

The hypothesis being tested was that there is a general relationship between a) the data from an arterial loop sensor, and b) the overall level of congestion on the arterial segment near the loop for that direction of travel. Specifically, lane occupancy percentage values from a loop sensor located just upstream from a stop bar for an arterial traffic signal were hypothesized to be a surrogate value for nearby arterial performance (specifically, congestion or traffic delay), and this relationship was proposed to be used to develop a basic arterial performance monitoring method to aid in planning and operational analyses, as well as potential real-time monitoring applications such as online traffic maps.

TEST SCENARIO

The specific location modeled was a 1-mile segment of State Route 522 near Bothell, Washington (Figure 1a), featuring three intersections: 61st Avenue NE, 68th Avenue NE, and 73rd Avenue NE (from west to east) (Figure 1b). Most of the analyses focused on the eastbound segment from 61st Avenue NE to 68th Avenue NE. The arterial scenario being modeled for this project used a common field sensor configuration, namely, inductance loops at each signal's stop bar. With this common scenario in mind, research was conducted to determine the potential of the stop bar loop as a performance indicator for the arterial segment leading to the signal.

Note that while the loop configuration assumed for the purposes of this project was relatively common and simple, the methods explored in this research also assumed that a) there would be access to all data that could be detectable in that scenario, and b) there would be a mechanism available to collect and store those data. In many cases, data from arterial loops are not stored after use, and even if they can be stored, proposed research often requires previously unused mechanisms of data storage and reporting, requiring new data archiving capabilities (software) within the traffic signal control software to support the monitoring process.

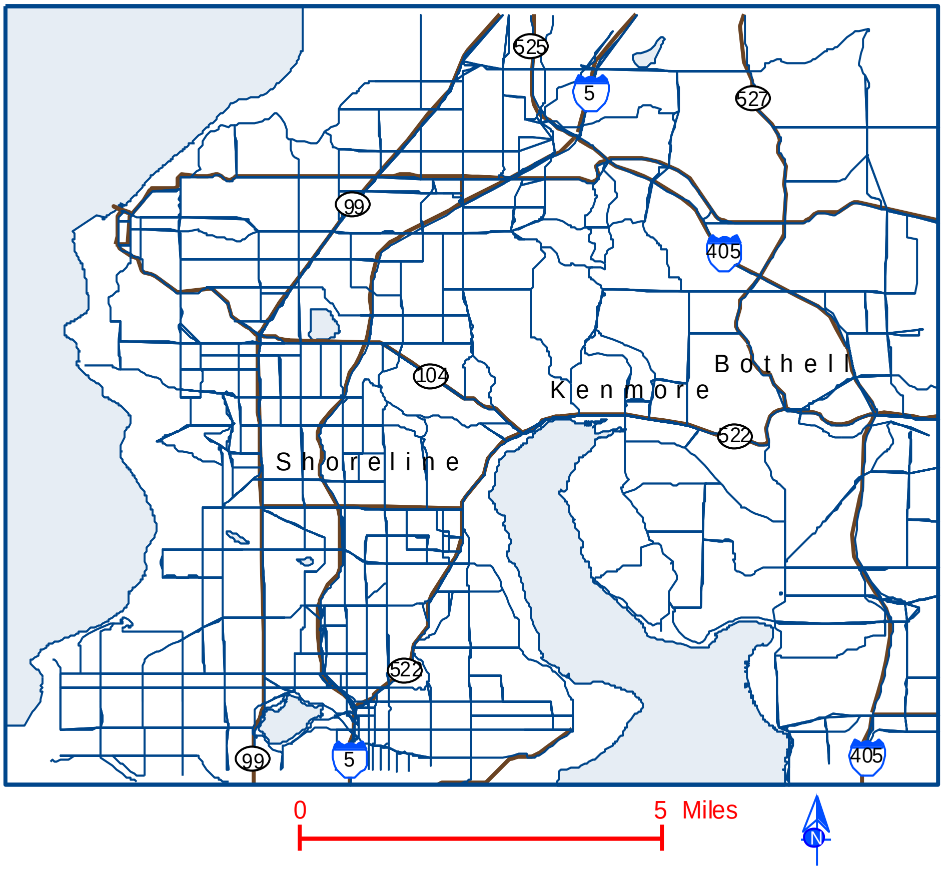

Figure 1a. Location of test scenario

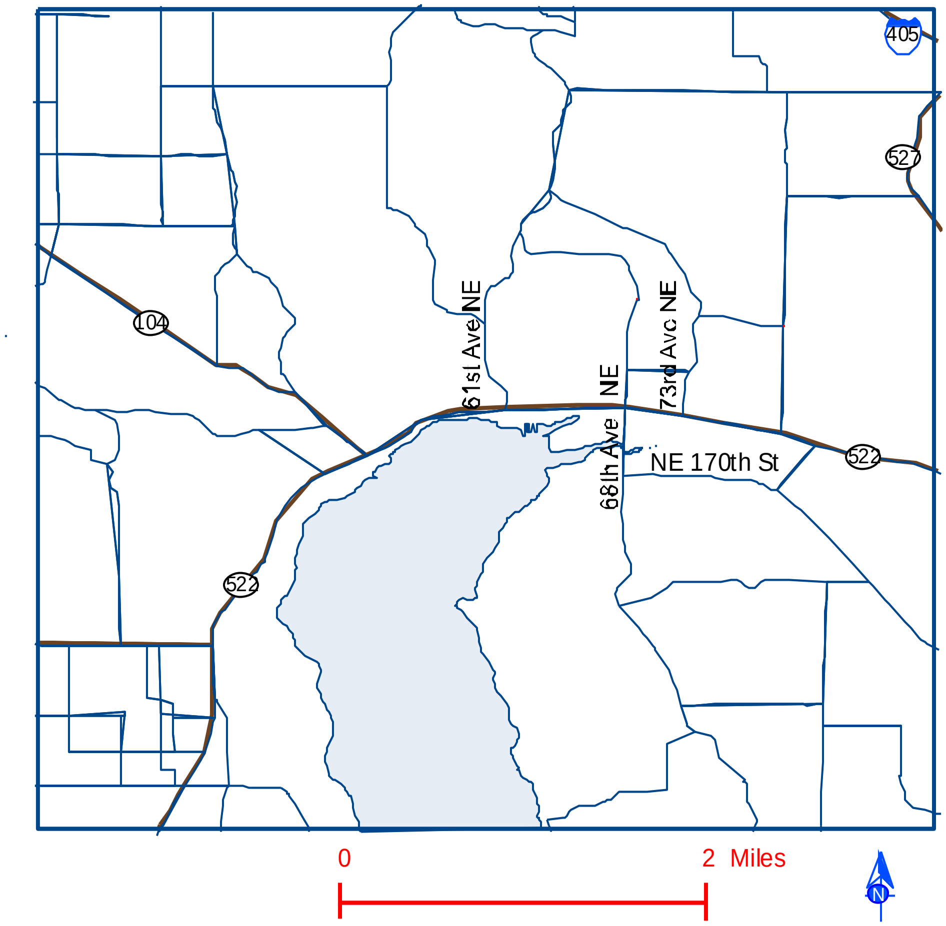

Figure 1b. Location of modeled segment on SR 522 in Kenmore, Washington (from 61st Avenue NE to 73rd Avenue NE)

PRELIMINARY TASKS

Before the potential relationship between stop bar loop data and congestion levels was explored, several preliminary tasks were performed to prepare for those analyses.

First, a simulation model was constructed to represent a segment of State Route 522 in Bothell. The model was developed by using the VISSIM microscopic simulation environment and was calibrated by using the most recent available complete data for turning movements, average vehicle volumes, and signal timing plans.

Second, the resulting simulation model was used to establish a method to associate a given test scenario with resulting congestion levels. Note that the test hypothesis was based on the assumption that one can observe both a) the stop bar loop occupancy values, and also b) the associated congestion conditions, so that one can determine the extent to which those two values are associated with one another. While the simulation model could directly generate the former (occupancy values at a stop bar loop) as an output, it was also necessary to determine what model output should be used for the latter (congestion conditions).

Several alternative values that could represent congestion were tested. Initially, traffic conditions during a model run were determined by visual inspection of the associated animation. While very high and very low congestion could be visually detected this way, the method was difficult to apply unambiguously for traffic conditions other than those at the extremes.

Another value that was considered was the input volume for the model. This volume is specified by the user and represents the arriving traffic volume along the primary arterial segment being analyzed. A series of test cases could be generated, each using a different input volume. This range of test cases could then represent a variety of congestion conditions, if the input volumes were associated with specified levels of congestion (for example, the higher the input volume, the higher the congestion). In the case of the eastbound approach to the 68^{th} NE intersection, different user-specified upstream input volumes from the intersection could be used to represent various congestion conditions approaching the intersection. However, a higher volume by itself would not necessarily guarantee congestion, since higher volumes can operate without congestion; a higher input volume, while potentially associated with congestion, does not automatically produce congestion. Furthermore, even if higher volumes did appear to produce more congestion (based on the animation, for example), there would still be a desire to more objectively quantify the congestion level.

Fortunately, the model does generate several values associated with delay. The per-vehicle intersection delay value is based on a comparison between modeled travel times of vehicles on a defined segment in the simulation network and the theoretical minimum travel time for that segment if no vehicles or signals are slowing traffic. The delay value could therefore be used to quantify the associated level of congestion. The model also generates average and maximum queue lengths on specified arterial segments. These values also provide valuable information about the magnitude and variability of congestion. By using this approach, a series of test cases at varying levels of congestion could be developed, by varying the input volume and then looking at the corresponding delay value and queuing attributes. The delay value and queue lengths would also allow test cases to be more objectively and quantitatively compared to one another.

The method eventually chosen for defining the sequence of experiments in this project was to vary the primary arriving vehicle volume at regular intervals, then observe the delay and queue values to determine whether congestion was growing and to what extent. In addition, the animation proved to be very useful for reviewing results and determining why non-intuitive results might be occurring.

ANALYSIS TASKS AND RESULTS

After the SR 522 model was developed, three phases of analyses were performed. The first two phases consisted primarily of exploratory tests and model development tasks, while the third phase involved applying the results from the previous phases to perform a broader analysis of the test hypothesis. Below are described the tasks performed in each phase and the resulting observations.

PHASE 1: INITIAL EXPERIMENTATION

Experiment Description

The initial experiments focused on analyzing a single approach to a signalized intersection along the primary peak direction of traffic, specifically the eastbound approach to 68th Avenue NE during one hour of the PM peak period, from 4:00 PM to 5:00 PM. These experiments were exploratory, designed to determine whether the test hypothesis warranted further research and how to best analyze and present the results.

The model was configured to produce output files of both signal status and loop data, at 1-second intervals, throughout the modeled 1-hour period, as well as produce overall system delay estimates. The signal state data were collected for the eastbound signal at the 68th Avenue NE intersection, while the lane occupancy data were collected at a simulated stop bar loop at the eastbound approach to the 68th Avenue NE intersection. These data were produced by using a range of arriving traffic conditions representing light, moderate, and heavy congestion, as defined by VISSIM's estimate of system delay and queues and confirmed by visual inspection of the associated animations. The stop bar occupancy was then compared to the traffic conditions to determine the extent to which there appeared to be some association between specific ranges of stop bar occupancy and varying levels of congestion, as well as whether this association suggested that stop bar occupancy could be used as a surrogate indicator of congestion.

One concern of this approach was that the stop bar occupancy percentage might be biased upward by the normal presence of vehicles that were stationary at the stop bar during a red light, making it difficult to distinguish between high occupancy values caused by stationary vehicles without congestion and high occupancy values caused by vehicles moving slowly past the stop bar under congested conditions and on through the intersection during a green light. To minimize the ambiguity produced by that situation, the occupancy data during the red phase were removed from the analysis process, and only a "30-second green time" stop bar occupancy, based on the occupancy during the first 30 seconds of green signal status in each signal cycle, was computed for the 68th Avenue NE signal in the eastbound direction. This required comparing the stop bar occupancy data from the model's loop detector output file to the signal event data from the signal output file, and then filtering only the occupancy values during the first 30 seconds of green signal status per signal cycle. This was performed with the assistance of various software utilities and macros built for that purpose.

Phase 1 Observations

A review of the analyses in Phase 1 suggested the potential use of the stop bar loop data to monitor performance. As hypothesized, there was a relationship between higher occupancy values and heavier congestion. For lower to moderate congestion levels, the results appeared promising. However, it was more difficult to determine the nature of that association at higher congestion levels because of a difficulty in achieving very heavy congestion scenarios in the model. Figure 2 shows typical results. Each circle on the graph represents the average occupancy and volume for the first 30 seconds of each signal cycle of a test run. Note that as the volume and delay grow, data points move to the right, representing higher occupancy values. At the heaviest congestion test cases, however, the data points tend to overlap more, and the trend becomes less clear.

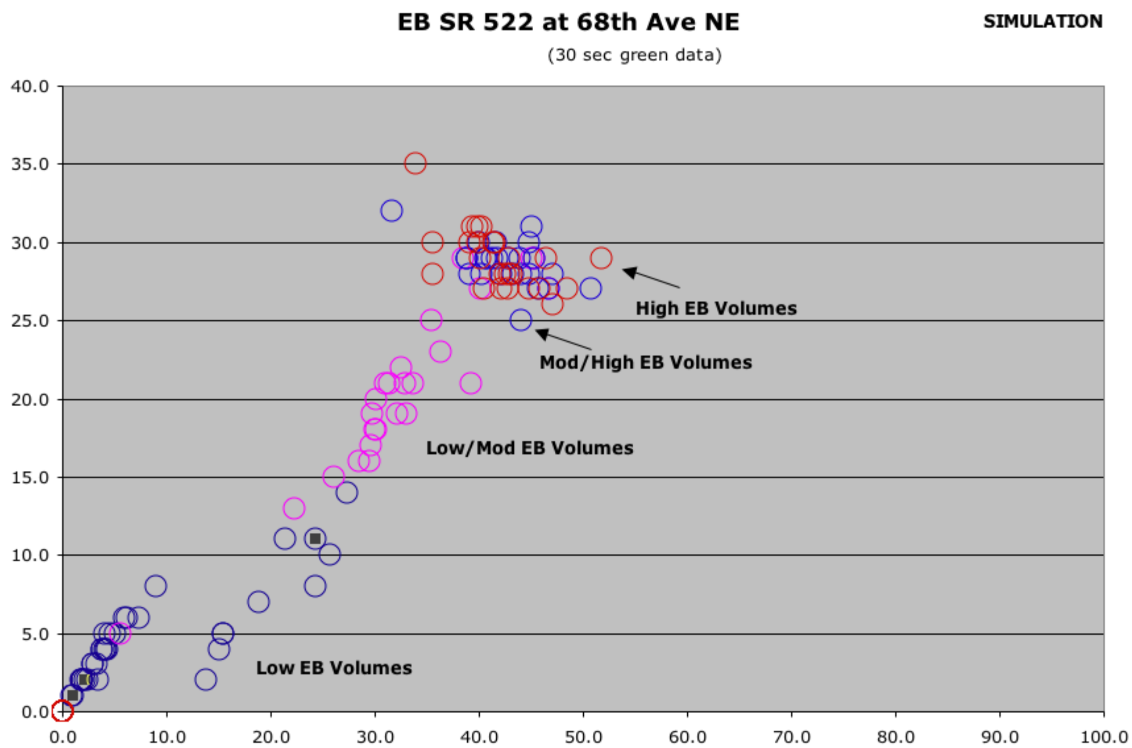

Figure 2. A comparison of occupancy percentage values (x axis) and corresponding arterial congestion (volume is on y-axis).

Following these experiments, the researchers also determined that the model required some enhancements to its network modeling and signal timing plans to improve its realism.

PHASE 2: UPDATED MODELING AND SENSITIVITY ANALYSES

Model Enhancements

Phase 2 of the project analyses focused first on model enhancements. Areas of enhancement were a) more accurate roadway geometry and lane configurations, b) updated signal timing plans, and c) more realistic vehicle movement logic.

After the model enhancements were made, the Phase 1 experiments were re-run, with a continued focus on the same eastbound approach to 68th Avenue NE during the PM peak period, to see whether the enhanced tools and model knowledge would produce results that were more conclusive and that suggested a potentially useful arterial performance monitoring methodology. This was followed by sensitivity analyses to evaluate the potential effect of moving the sensor location upstream from the stop bar.

Phase 2 Observations

The Phase 2 experiments showed stop bar loop data patterns similar to those from Phase 1, with a similar limitation at the higher congestion levels. The sensitivity analyses to move the sensor location upstream from the stop bar showed that there were, indeed, distinct differences in the occupancy patterns as the sensor moved upstream. In particular, at upstream locations where queues would be expected throughout the first 30 seconds of green, there was a distinct separation between lower and higher congestion, with the latter moving to higher values. The data from the higher congestion scenario were often near 100 percent and zero volume, corresponding to stopped vehicles. This pattern changed as the loop moved farther upstream, becoming less predictable.

While use of sensor data at a midblock location might result in a promising method for detecting queue lengths from congestion, the sensitivity analyses did not clarify where in the block that sensor location should be, nor how it would be dependent on the typical queuing patterns, block lengths, and intermediate traffic sources and sinks of each arterial segment. To simplify subsequent experiments, the researchers determined that the focus should remain on the stop bar loop.

PHASE 3: MODIFIED HYPOTHESIS AND TEST MATRIX

Experiment Modifications

The results from the first two phases suggested that a relationship between loop data and congestion could be developed, although the results were still inconclusive. Phase 3 focused on a revision to the hypothesis and a broader series of tests. Before the additional experiments, the following adjustments to the modeling and testing approach were made:

- The original test hypothesis focused on data from the first 30 seconds of green time of each cycle. In Phase 3, two variations to that original stop bar occupancy approach from Phase 1 were explored: 1) The green time occupancy hypothesis was modified by using data from the entire green and amber time of each signal cycle, rather than just the first 30 seconds; and 2) tests were also performed by using all data points in a cycle regardless of signal phase.
- Previous work in Phase 1 and Phase 2 used the overall system delay value from VISSIM as the indicator of congestion. However, that value processes vehicles on all approaches of all intersections, not just the one intersection and eastbound approach of interest. For Phase 3, a per-vehicle delay of vehicles specifically approaching the intersection in question (68th Avenue NE) from 61st Avenue NE was used, in keeping with the research focus on the segment from 61st Avenue NE to 68th Avenue NE.
- The detector loop position at the 68th Avenue NE stop bar was adjusted to more reliably detect vehicles that stopped slightly behind the stop bar.

- Side street volumes approaching a model intersection were kept constant across all test cases, rather than incrementally increasing them as the eastbound volumes increased.
- Observations of previous analyses suggested that the simulation should run longer to reach equilibrium. Therefore, Phase 3 model runs used longer run times to compensate for the longer initial transition period.

In addition to those changes, efforts were made to address one of the limiting aspects of the previous analyses, namely the difficulty of determining the extent of the association between loop data and congestion during heavily congested conditions because of the difficulty in generating heavy congestion delay. Very heavy congestion was not occurring in the base model, even when very high input volumes were chosen; in fact, at times the delay actually leveled off or decreased somewhat even as arriving volumes grew. This was contrary to the expectation that congestion would likely grow as more vehicles traveled on the arterial.

Several methods were tested to overcome this limitation and produce heavy congestion approaching 68th Avenue NE. These methods either increased the upstream arriving volume to induce congestion, or reduced downstream capacity to induce queuing backups. For the first alternative method, analyses determined that high input volumes were triggering occasional conflicts approaching the upstream intersection, 61st Avenue NE, in the form of a high number of vehicles turning left there that was greater than the storage capacity of the left turn lane. This produced a queue of turning vehicles backing up to the SR 522 through-lanes that caused conflicts with vehicles attempting to continue eastbound through the intersection. (Note that the vehicles arrive upstream from 61st Avenue NE and then flow eastbound through each intersection.) The result was effectively a limit on the number of vehicles arriving at the next intersection (68th Avenue NE, i.e., the focus of the test), thus preventing heavy congestion from building there. To reduce conflicts and allow more vehicles through the intersection, the length of the 61st

Avenue NE left turn lane was increased to provide more storage. (As an aside, this was a good illustration of a modeling problem that was solved much more easily by viewing the model animation than by analyzing the data files.)

The second method for inducing heavier congestion involved reducing throughput at a location downstream from the 68th Avenue NE intersection to encourage congestion to back up to that intersection. This was done by reducing the eastbound green time at the downstream intersection (73rd Avenue NE) to produce heavier queuing there and inhibit through-traffic. A variation of this method reduced through-traffic by directing a higher percentage of vehicles to a left turn lane at 73rd Avenue NE with limited green time. The resulting queue caused backups that affected through-traffic.

The third method reduced throughput by reducing road capacity. This was done by removing a section of one of the two eastbound lanes downstream from 73rd Avenue NE, thus causing vehicles to merge into a single lane and producing queues.

After these features were implemented, a matrix of new cases were tested. The experiments continued to focus on the same eastbound peak direction approach to a signalized intersection (68th Avenue NE) that was used in phases 1 and 2. The matrix of cases represented various combinations of

- different arriving (upstream) volumes (to roughly represent a range of traffic congestion conditions)
- different scenarios of arterial conditions (coordinated signals, uncoordinated signals, heavy turning movements, or a blocking incident/construction)
- different stop bar loop occupancy computation methods: use all data values, use first 30 seconds of green values, use all values during the green and amber phases

The resulting matrix of scenarios was as follows:

Table 1. Matrix of cases tested

Scenario	Description	ID No.
Base Case	Original calibrated model	3
Mod A (heavy congestion test 1)	Base Case + shortened green time eastbound at 73rd Avenue NE	5
Mod B (heavy congestion test 2)	Base Case + longer left turn lane to northbound at 61st Avenue NE	9
Mod C (heavy congestion test 3)	Base Case + Mod A + Mod B	8
Mod D ("non-recurring incident")	Mod C + lane blockage downstream from (east of) 73rd Avenue NE	10
Mod E (heavy congestion test 4)	Mod C + more vehicles turning northbound at 73rd Avenue NE	12
Mod F (site 73)	Mod C + additional sensors at eastbound approach to 73rd Avenue NE	9b
Mod G (site 61, heavy congestion)	Mod C + additional sensors at eastbound approach to 61st Avenue NE	11
Mod H (site 61, light to mod cong.)	Mod C + additional sensors at eastbound approach to 61st Avenue NE	11b

Mod A and B were attempts to induce heavy congestion scenarios in the model at 68th Avenue NE by reducing downstream capacity or increasing upstream volume, respectively (Mod C combined the two). Mod D also produced heavy congestion, this time by using a downstream capacity restriction like one would encounter in a non-recurring event such as a blocking incident or construction. Mod E utilized heavy turning movements and limited turning green time at the downstream intersection (73rd Avenue NE) to induce heavy congestion upstream. Mod F, Mod G, and Mod H produced results from two additional locations for comparison. For each scenario, a range of input volumes was tested, from 500 eastbound vehicles per hour to 3000 eastbound vehicles per hour, generally in 250 or 500 vehicle per hour increments. Each test was analyzed by using all three occupancy computation alternatives to determine stop bar loop occupancy for each signal cycle (combining all data values throughout the signal cycle, combining data from only the first 30 seconds of data during the green phase of each cycle, or combining all data during the green and amber phase of each cycle).

The following data were collected for each test:

- loop volume and occupancy percentage from a loop located just upstream from the stop bar approaching the SR 522/ 68th Avenue NE interchange (traveling eastbound), from :00-5:00 PM, at one second intervals.
- signal event (signal status change) data for the eastbound signal at the same intersection from 4-5 PM; this data output shows the time of each change in signal status (green, amber, red)
- upstream segment average speed and delay, for a 2900-foot (intersection spacing) segment leading to the stop bar.

The same data types noted above were also collected for the other test sites.

Phase 3 Observations

Overall Loop Data Patterns

The revised modeling approach produced results that were consistent with those in the previous phases, i.e., occupancy grew as congestion grew, as shown by the pattern of the clusters of occupancy values corresponding to increasing levels of congestion. In addition, new data from heavy congestion scenarios suggested the continuation of the pattern at heavier congestion levels. Figure 3 illustrates this pattern for the revised base case (Mod C), showing how the occupancy values (based on the green+amber occupancy method) tended to be larger as congestion grew. The figure shows the combined results of six simulation runs representing different levels of congestion. Each run is a different color; each circle represents the average occupancy for the green+amber time of one signal cycle during that run. When traffic was light and delay was minimal (e.g., green circles), volume and occupancy were both relatively low. As traffic grew, both volume and occupancy also grew (e.g., yellow, orange, red circles). At approximately 30 percent occupancy, congestion built significantly, and while occupancy continued to grow, volume throughput leveled off and began to decline. The values for the heavy congestion tests (blue and black circles) continued the pattern observed at the low to moderate congestion levels, with occupancy values continuing to grow.

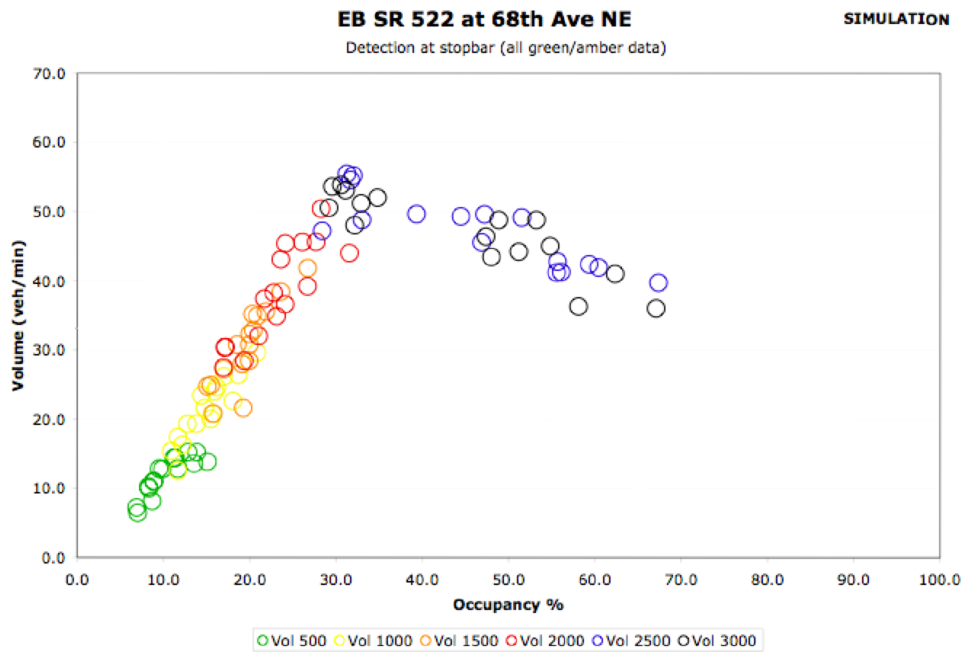

Figure 3. A comparison of occupancy percentage values (x axis) and corresponding arterial congestion (volume is on y-axis) for Mod C base case.

Other test cases showed similar patterns, although some were less distinct than others

<u>Exploring Patterns through Data Aggregation</u>

The results described above suggested a relationship between occupancy and congestion. Those graphs were based on data aggregated at the individual cycle level, i.e., each data point or circle represented the average occupancy at the stop bar loop during the green and amber time of a single cycle. For any given test, the associated data points were generally clustered; however, the clusters were not always condensed and well defined. This variability made the determination of a clear relationship between the occupancy and congestion more difficult. The researchers then hypothesized that if the data were aggregated at a higher level, the transient cycle-level variability of the loop data would have a less direct effect on the analysis and perhaps enable a clearer picture of the overall nature of the relationship between loop data and congestion to emerge. So, for each simulation test run, an aggregate average occupancy percentage for the entire 1-hour test period (after the initial start-up time of the run) was computed, rather than cycle-by-cycle values. As for the corresponding congestion indicator, the aggregate value used was the average per-vehicle speed (or alternatively, delay) associated with the arterial segment upstream from the stop bar. In the case of the primary test location, 68th Avenue NE, this segment started just downstream from the 61st Avenue NE intersection and continued to the eastbound stop bar at the 68th Avenue NE intersection.

When those two aggregated variables were tracked for each test case, and the outputs were combined, the results were as shown in Figure 4. While there was still some variability, particularly at the higher congestion levels (slower speeds), the overall pattern showed an upward trend in occupancy percentage as a function of congestion (represented by average speed), as proposed in the original hypothesis.

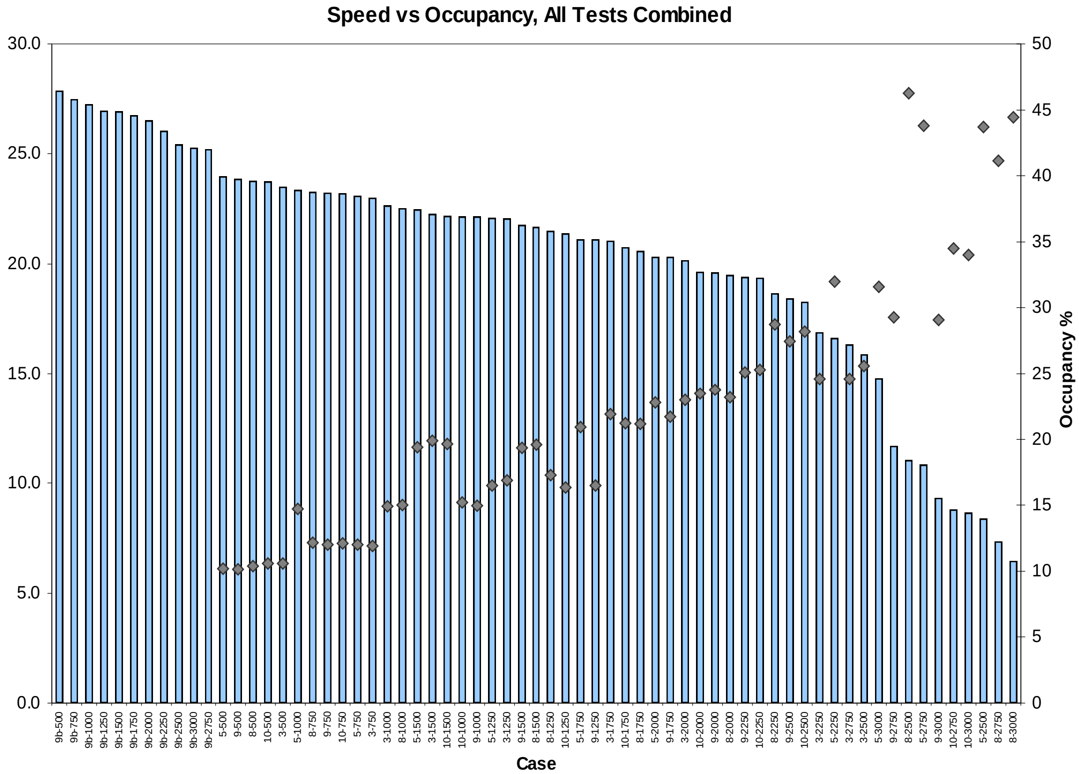

Figure 4. A comparison of occupancy percentage values (right axis) and corresponding arterial speed (left axis), for each test case.

(See "Variability of occupancy," below, for a discussion of the variability of loop data from a performance monitoring and reporting standpoint.)

<u>Using Occupancy to Estimate Congestion</u>

The relationship between occupancy and speed indicated in Figure 4 suggested that average occupancy could be used as a general indicator of associated congestion levels. For example, one approach would be to subdivide the occupancy range, with each subrange corresponding to a different congestion level (based on speed). For example, a web-based display of arterial conditions might show congestion in three categories: light, moderate, and heavy. In that case, the occupancy range would also be split into three subsets, one for each congestion category. As traffic conditions varied over time, the occupancy would be tracked, and the congestion level corresponding to that occupancy value would be displayed.

The use of such a relationship as suggested by Figure 4 would involve the following steps:

4) Produce occupancy vs. speed data, such as that shown in Figure 4, by using a particular occupancy computation method. Verify that the data show an upward trend in occupancy together with a downward trend in speed.
5) If the occupancy data have some variability, such as that shown in Figure 4, consider smoothing the data to produce a central trend of the occupancy data (vs. congestion) that is less influenced by fluctuations.
6) Define congestion levels on the basis of speed on the arterial segment being analyzed. For example, the light/moderate/heavy website described above might use categories based on Highway Capacity Manual guidelines for Level of Service standards on that type of arterial. Or, a more direct approach might be to simply specify speed ranges that are either consistent with other existing applications, or coincide with existing local performance standards. Simple speed categories are also more easily interpretable.
7) Determine the range of occupancy values that correspond to each congestion category defined in step 3. Do this by looking at the distribution of congestion delay values (speeds) across all the test cases and dividing them into groups by congestion level (i.e., a specified speed range). For each congestion level, the corresponding occupancy value range is determined. The result is a functional relationship between occupancy ranges and speed/congestion ranges:

$$Occ = f(m)$$

where

f(m) = a one-to-one relationship between a given congestion category m and a specified range of occupancy values OCC (e.g., LOS A corresponds to occupancy values of between 0 and N percent).

8) Once f(x) has been established, it can then be applied to a performance monitoring application.

Note that in the process above, the development of f(m) with subdivisions of the occupancy range requires the existence of a one-to-one relationship between occupancy and congestion, i.e., each value of occupancy corresponds to exactly one congestion category. This means that occupancy should grow monotonically as congestion grows (or as speeds slow). In reality, though, some variability is likely. That is why some type of smoothing operation may be desirable, one that removes transient spikes but still displays the overall character and trend of the data. This smoothing approach helps facilitate the specification of the threshold values for each congestion category, although it will reduce the precision of the category definitions.

An Application of the Occupancy versus Congestion Relationship

The following is an example application of the approach described above.

1) Produce occupancy vs. speed data. Figure 4 shows an example of such data. The graph also suggests the presence of an occupancy trend that grows with congestion.
2) Because the occupancy data have some variability, the data are smoothed to reduce the fluctuations by applying a six-period moving average to the data.
3) Congestion levels are defined on the basis of speed by using the Highway Capacity Manual LOS guidelines[4]. For example, for a Type II arterial, the LOS values are as follows:

[4] Highway Capacity Manual, Transportation Researcb Board, 1997 Section II-4.

LOS A	>30 mph
LOS B	24 to 30 mph
LOS C	18 to 24 mph
LOS D	14 to 18 mph
LOS E	10 to 14 mph
LOS F	<10 mph

Alternatively, simply divide the speed range (up to the speed limit) into 5-mph or 10-mph increments.

4) For each congestion category above, the corresponding occupancy value range is determined by using the data on Figure 4. This process is shown in graphical form in Figure 5a, where each speed threshold value is matched to the closest corresponding occupancy value. In this example, the smoothed value associated with that occupancy value is used instead. (Note that LOS A, B, and C have been combined in this example.)

5) The resulting function is

LOS A ,B,C	0 to 26 percent
LOS D	26 to 29 percent
LOS E	29 to 34 percent
LOS F	>34 percent

6) Each incoming occupancy data point can then be assigned to one of the LOS categories.

Note that even with the smoothing process, occupancies do not always monotonically grow with congestion, which means that the threshold values may be ambiguous. For this example, that was not the case. If another smoothing option is chosen, such as a 3rd order polynomial, the results may be as shown in figures 5b and 5c. This smoothed curve produces monotonically growing values that simplify the process of determining threshold occupancies. The resulting thresholds in Figure 5b are approximately the same as those from the moving average, but with some differences of a few percentage points.

This illustrates the limits to precision of the congestion category thresholds with this method.

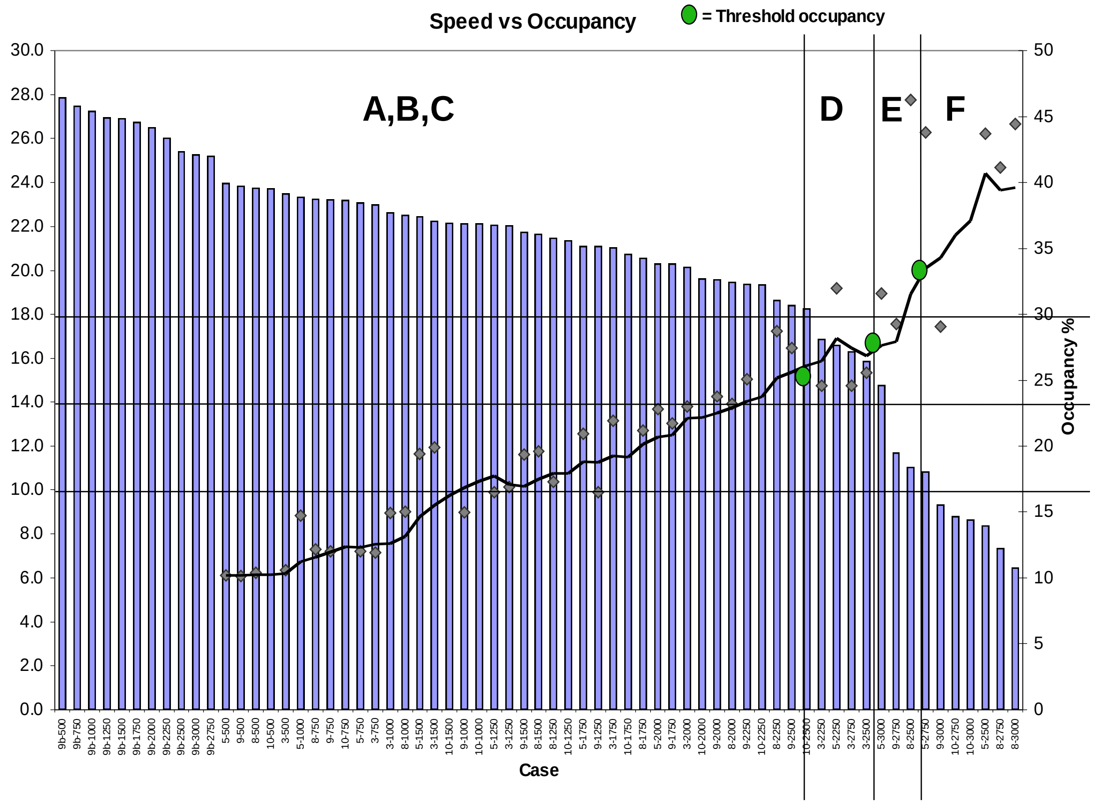

Figure 5a. Determining congestion category thresholds from a comparison of occupancy percentage values (right axis) and corresponding arterial speed (left axis) from each test case, using a moving average smoothing fit.

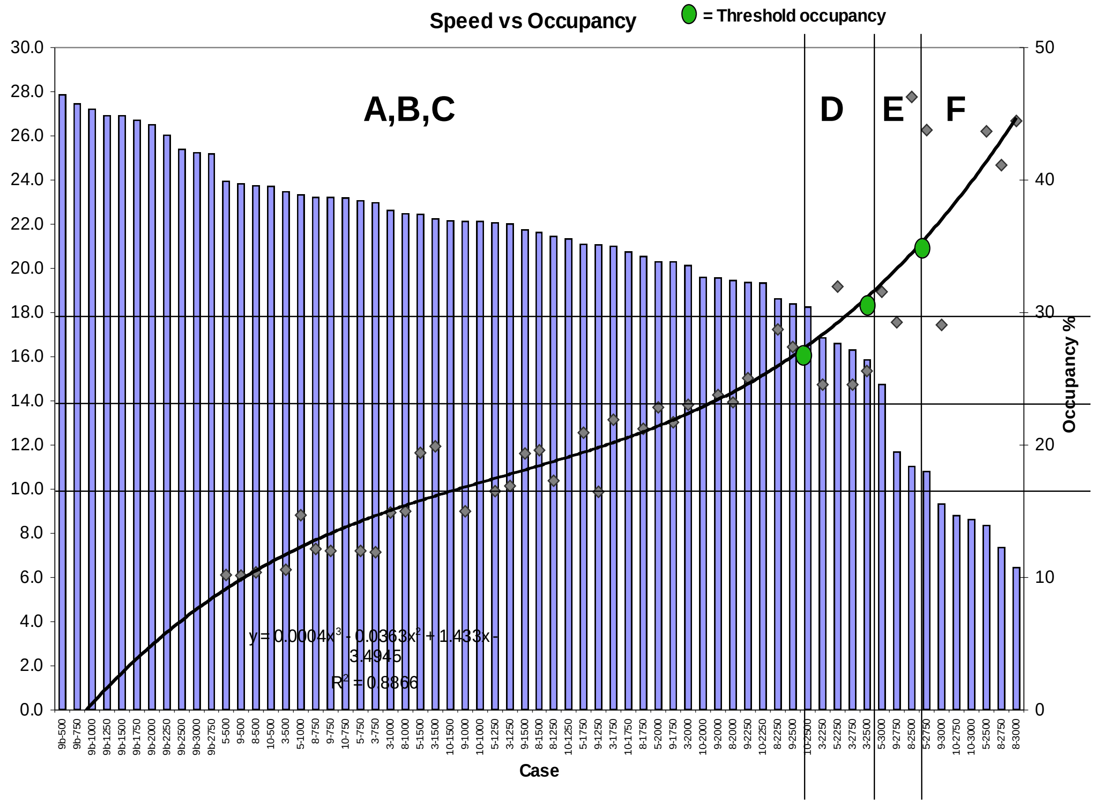

Figure 5b. Determining congestion category thresholds from a comparison of occupancy percentage values (right axis) and corresponding arterial speed (left axis) from each test case, using a polynomial smoothing fit.

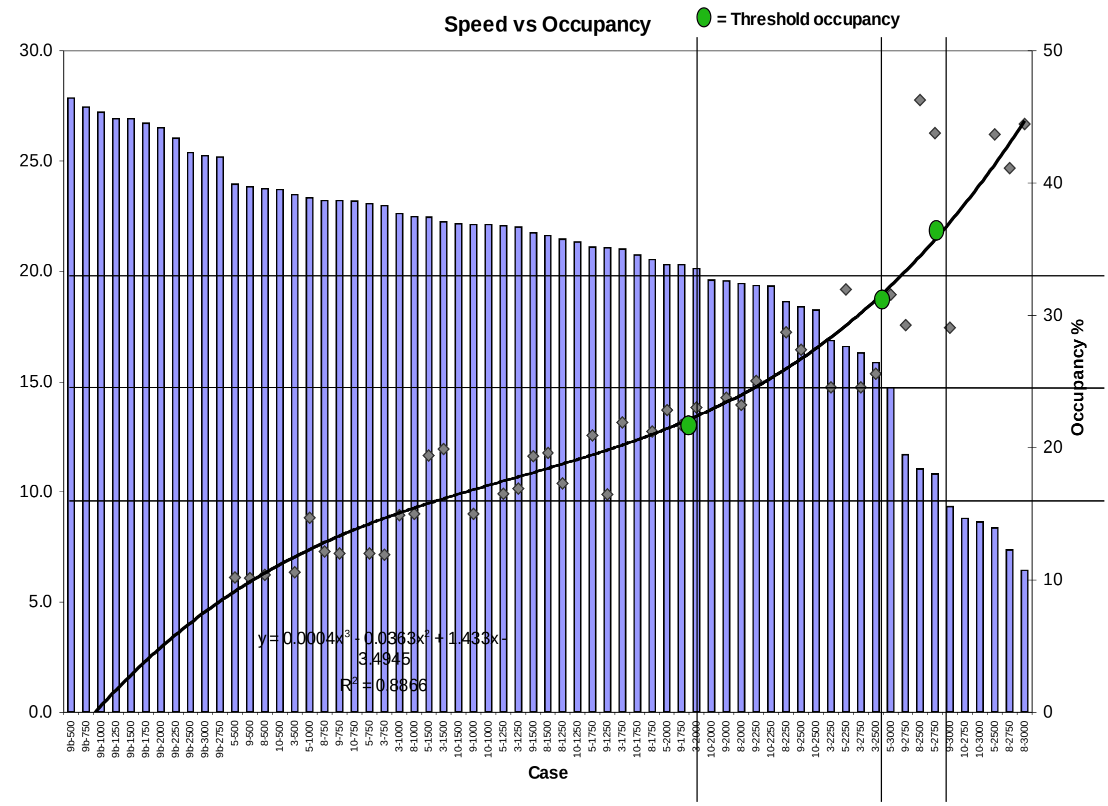

Figure 5c. Determining congestion category thresholds from a comparison of occupancy percentage values (right axis) and corresponding arterial speed (left axis) from each test case, using a polynomial smoothing fit and a simple 5 mph speed range.

Effects of Different Occupancy Computation Methods

The alternatives tested for computing occupancy for each signal cycle were as follows:

4) average all occupancy data during each cycle (one-second data)
5) average all occupancy data during the first 30 seconds of the relevant (e.g., eastbound thru-traffic) green phase of each cycle
6) average all occupancy data during the entire green and amber phases of each cycle.

A comparison of the three methods showed that the "all occupancy" method produced data that were clustered around a combination of high occupancy values and low volume values, which one would expect given the method's inclusion of the red phase, when vehicles are stopped at the stop bar and no vehicles are moving across the stop bar. The 30-second green method appeared to clarify the pattern noticeably. The green+amber method produced the clearest association between occupancy and congestion, providing more tightly clustered data for a given test case, and somewhat clearer results at the heavy congestion levels. See figures 6a, 6b, 6c for a comparison of the results from the Mod C test. (Unless otherwise noted, all results described throughout the report are based on the green+amber method.)

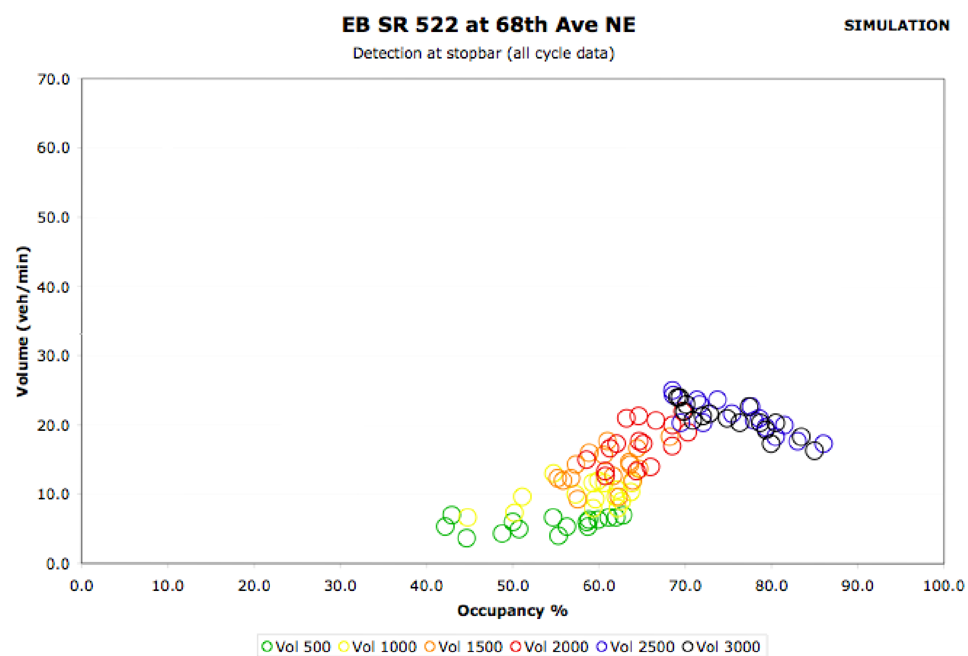

Figure 6a. A comparison of occupancy percentage values (x axis) and corresponding arterial congestion (volume is on y-axis) for Mod C base case, using all values in each signal cycle.

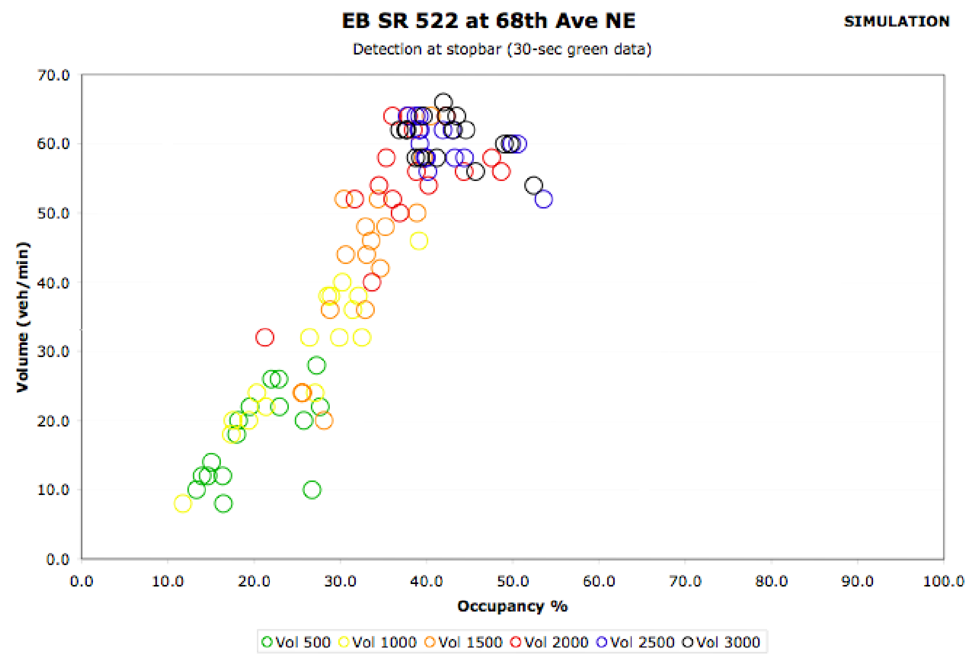

Figure 6b. A comparison of occupancy percentage values (x axis) and corresponding arterial congestion (volume is on y-axis) for Mod C base case, using data from the first 30 seconds of green time per signal cycle.

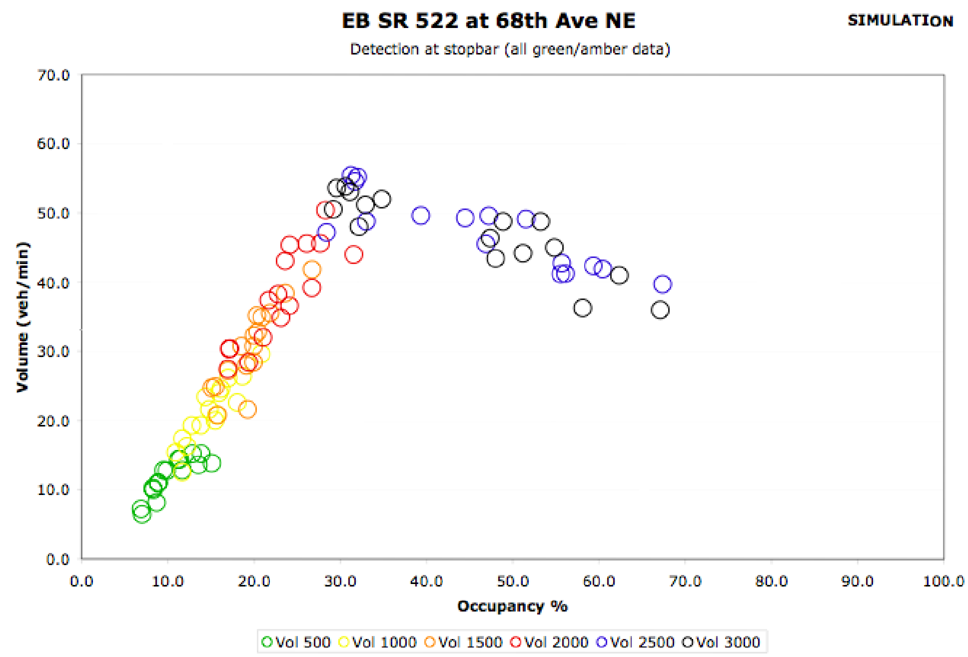

Figure 6c. A comparison of occupancy percentage values (x axis) and corresponding arterial congestion (volume is on y-axis) for Mod C base case, using data from the all green and amber time per signal cycle.

Effects at Other Locations

Similar tests were also performed at two other intersections, 61st Avenue NE and 73rd Avenue NE, both in the eastbound direction. Figure 7 shows the data from the two locations (shown as solid markers) superimposed on the data from the primary test location (open markers). The 61st Avenue NE samples (11b1500, 11b3000, and 113000 shown as yellow, blue, and gray solid squares, respectively) assumed 1500, 3000, and 3000 vehicles per hour respectively, while at 73rd Avenue NE, case 9b (solid triangles) assumed 1500 vehicles per hour. As the figure suggests, the results at the additional two locations were consistent with those from the original test location, showing data points that either followed the pattern of data from the primary test location or were a logical extension of the pattern to higher congestion levels.

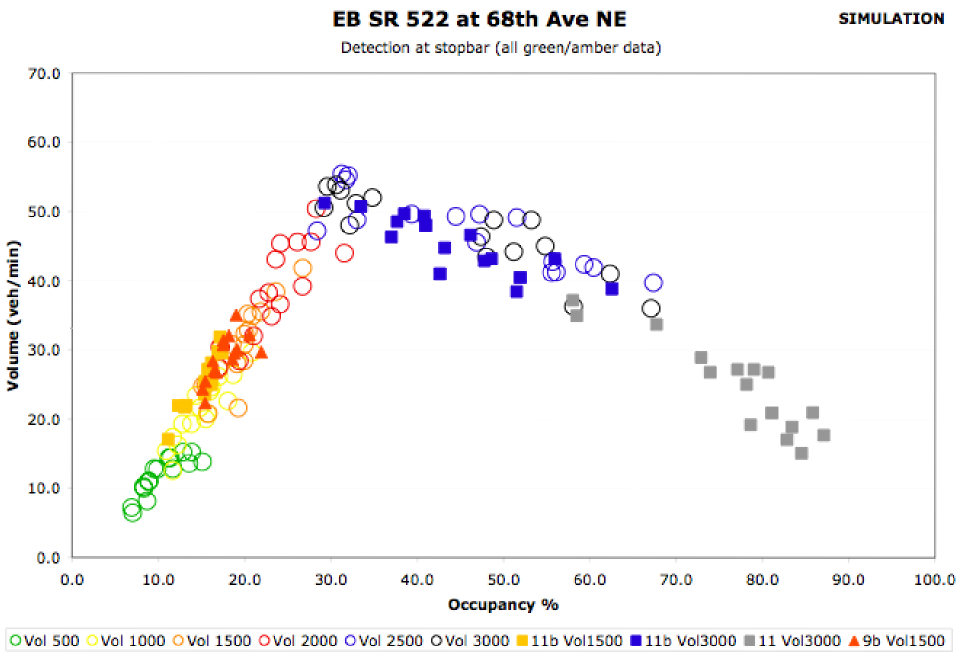

Figure 7. A comparison of occupancy percentage values (x axis) and corresponding arterial congestion (volume is on y-axis) for Mod C base case, plus two new locations.

Variability of Occupancy during Reporting Process

Given the stop-and-go nature of signalized arterials, it is not surprising that the analyzed data showed fluctuations in occupancy values over time. At low to moderate congestion levels, average occupancy values per signal cycle tended to be more clustered, varying more smoothly over time. At heavy congestion levels, associated occupancy values appeared to vary more noticeably over time. See Figure 8 for examples of the time-varying patterns for successive signal cycles using the Mod C case.

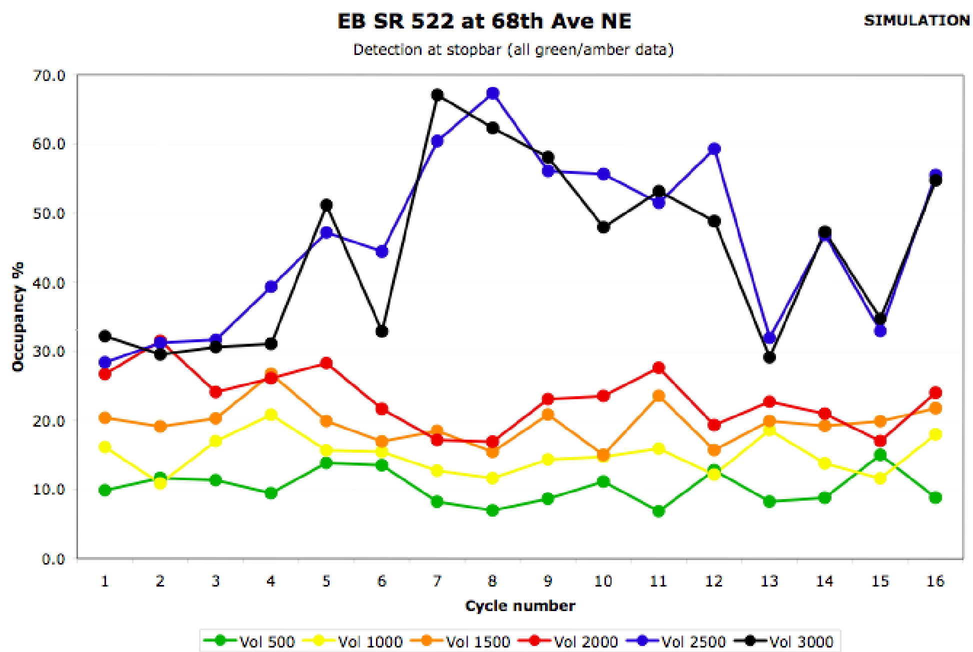

Figure 8. Variability of cycle occupancy over time, for Mod C base case and green+amber values.

In the previous discussion about developing a monitoring algorithm, occupancy variability was addressed through smoothing techniques to better define the central trend of the data; that was specifically for the purposes of establishing the threshold occupancy values used by that method. However, occupancy variability is also an issue when one applies the resulting method. Namely, how should the method include this variability when arterial performance is reported? Should short-duration oscillations be considered useful indicators of performance, or should they be considered transient values that distract from a more important goal of showing the central tendency of the traffic conditions? Given the previous discussion of precision limitations for the threshold values, the significance of using those congestion levels for these transient changes is not clear. It might be the case that the application dictates the reporting increment, to some extent; if the time increment for monitoring performance that is required for the application is not small, then it might not be necessary to report the transient values. For a near-real-time traveler information application, for example, if a Web display is updated every signal cycle, oscillating conditions from cycle to cycle might produce "flickering" traffic condition indicators that would perhaps not be useful to the traveler. In that case, it might be more helpful to report the results in another way. There are several options to address this:

- Aggregate the occupancy data over time. If oscillations that occur are cyclical, e.g., associated with the signal cycle, then it can be argued that they are perhaps not the most useful indicators of performance. It might be more useful to smooth the data over time, by using a method such as moving averages or by using a median value to reduce the effect of outliers.
- Use broad congestion categories to compensate for the variability. If the congestion categories covered large subsets of the occupancy range, variability of occupancy from cycle to cycle would be less likely to cause flickering performance conditions. However, the result would also be less specific.
- Reclassify the congestion categories to take into account the variability. The congestion categories could be refined to include the oscillations. For example, a new category of congestion could be introduced, between LOS B and LOS C, that represented a transition period win which conditions were a blend of B and C.
- Do not change the original data. If the application would benefit from reporting of the transient conditions, no processing would be required.

IMPLEMENTATION AND OTHER ISSUES

IMPLEMENTATION REQUIREMENTS

The method described above would require the following:

Supporting Data:

- Stop bar sensors capable of producing occupancy percentage values at the desired level of frequency
- A data storage capability (or data transfer capabilities to a central facility)

Method:

- A specified congestion categorization approach (e.g., speed ranges)
- Threshold occupancy values for each congestion category

Processing:

- Software to smooth data as required.

The supporting data are producible by a basic loop sensor at the stop bar. Arterials often have sensors at the stop bar to detect arriving vehicles and help control the associated signal. The implementation software and associated parameters would have to be developed or specified by the user. Given the relatively straightforward algorithms employed in this method, the processing software should be relatively inexpensive to develop.

While the basic hardware and software requirements do not appear to be onerous, a more likely upgrade requirement might be an archiving capability. The method requires some storage and processing. This processing could be performed centrally or on-site. Either way, the data would need to be archived (or transmitted to an archive facility) to support the desired performance monitoring activities, whether planning tasks, operational tasks, or real-time information tasks.

For arterial networks that either did not store their data locally at all, or did store the data but did not centrally archive the values, some mechanism to transfer data would be required. This would likely involve a change of traffic signal controller software to record volume/occupancy by using variable time frames, and to store along with those timeframes the length of the actual green+amber condition. (That is, our recommended approach to performance monitoring is to use stop bar statistics of volume and occupancy collected only when that phase is green and amber. To do this would require the traffic signal system to no longer use a "fixed time" reporting framework but, instead, one that varied with signal phase lengths.) This would be particularly important when some type of adaptive traffic control was used (including actuated and semi-actuated traffic signals and signal timing plans) that did not have fixed phase lengths.

The good news is that this new capability to support arterial monitoring would allow traffic signal engineers to not only examine the level of congestion present but also determine how flexible signal timing algorithms were actually being used in the field and thus might be modified to improve congestion. For example, if we have a permissive phase extension of 20 seconds, how often are all 20 seconds being used? How often is none of that possible extension being used, and why was it not used? Was it the result of a pedestrian button on a perpendicular approach that forced off the signal, or a lack of traffic volume on that approach while a conflicting approach had a waiting vehicle? These types of questions could be addressed with a more powerful data collection capability. The result would a more robust capability for measuring not only how congested each approach had become but how arterial signal timing plans were actually being used. This would provide engineers with detailed information that could be used to routinely tune timing plans to decrease delay without having to pay for new short duration traffic data collection.

IMPLEMENTATION ISSUES

While the tests conducted thus far suggest that this method is potentially useful, there are other methodology development issues that should be considered prior to implementation.

Method Calibration and Robustness

The results described in this report were based on simulation of a typical 1-mile arterial section. While the analyses suggest that the results were consistent for different intersections and approach directions in this single model, it would be desirable to perform additional work to further validate and calibrate the proposed monitoring methodology. This would include analyzing additional locations and geometric configurations, other road types (e.g., Type I, II, III categories from the HCM), and other types of traffic flow characteristics. The result could be a matrix of threshold values, based on different roadway attributes, rather than a single universal threshold standard.

Arterial Segment Definition

The first step in the monitoring process described previously requires development of an occupancy versus speed data set. However, while the occupancy value refers to a specific point (the stop bar loop), the speed value could be associated with any segment. In the initial tests, we chose to use the intersection spacing (i.e., a segment from the upstream intersection to the intersection in question). However, because those spacings can vary, the resulting segment length being represented by the single stop bar loop will vary also. This introduces several questions:

- What segment is the stop bar loop attempting to represent?
- What is the effect of varying segment lengths on the methodology?

For urban arterials with short intersection spacings, it might be reasonable to say that a stop bar detector can represent the segment immediately upstream (to the next signal). For other arterials with larger spacings, it might be necessary to use intermediate loops to monitor performance in areas where there were no intersection and stop bar nearby, limit the segment length represented by each stop bar loop, or not provide information on those segments.

The occupancy to speed relationship in Figure 5 is based primarily on 68th Avenue NE data and the use of a fixed segment length for computing speed (about 2900 feet). When data were analyzed at 61st Avenue NE, a similar length was used. However, for 73rd Avenue NE this was not possible because the spacing was less than half that value. In general, intersection spacings vary, and speeds are at least partly a function of those segment lengths. This means that for a given scenario, changing the segment length could change the speeds and, therefore, the threshold occupancy values.

One potential approach would be to determine whether there was a family of occupancy versus speed curves based on intersection spacing, i.e., "If the spacing is X feet, the corresponding occupancy versus speed curve is…". The threshold values would then be based on that particular curve. However, this assumes that the family of curves has been shown to be reasonably well defined and predictable.

Another alternative would be to identify significant relative drops in the speed curve and use the corresponding occupancy values. This would take advantage of the fact that regardless of segment length, the relative speed curve stays the same, i.e., as length changes, the overall speed curve shifts up and down, but relative patterns do not change.

Stop Bar Limitations

It may be that there are limitations in the ability of the stop bar loop to estimate arterial performance as opposed to "intersection approach performance." The approach described in this report could be a good "arterial performance" metric as long as delays are only intersection based, and not from mid-block occurrences. In addition, because we only have measurements at the intersection, we can say that the signal has failed, but we cannot say that one is likely to sit through one, two, three or more light cycles to get through that signal. One alternative might be to look at the volume associated with each green phase, the green phase length, and the number of light cycles in a row that the cycle has shown "failed" performance levels. (An alternative is that, with fast data transmission, it might also be possible to use upstream and downstream volumes to determine a crude measure of the number of vehicles stored on a segment. This might in turn indicate the number of signal cycles required to get through the intersection, which in turn would provide travel times and average speeds for the segment. However, this would require an additional level of complexity beyond the approach described here.)

CONCLUSIONS

The primary objectives of this project were to analyze the potential of a hypothesis regarding the ability of traffic data from basic loop sensors to represent approximate arterial traffic conditions (congestion); to develop a prototype analytical method to implement that relationship, and to determine requirements and other issues associated with future application of the method. This research provided additional understanding of the feasibility of using basic sensor data to monitor performance on arterials, as well as outstanding issues. Given the results thus far, the use of green+amber occupancy values from a stop bar loop shows the most promise among the options analyzed; furthermore, the hardware and analytical requirements are not restrictive. There is a need for additional testing of the robustness of this method before it can be implemented.

ACKNOWLEDGMENTS

The authors wish to thank Thomas Tumola, who developed the original VISSIM model and the initial exploratory analyses while a graduate student at the University of Washington. We also appreciate the work of Adam Sanderson, Software Engineer at TRAC, for his assistance in developing a database of arterial data to support the modeling work. Thanks also go to the Washington State Department of Transportation for its assistance with data and other technical support during the development of the VISSIM model. The authors also wish to thank the team of researchers at the University of Washington who developed and tested Phase 2 model enhancements as part of a graduate-level special projects class on Traffic Simulation Applications. Those researchers were

Matthew Beaulieu
Kathy Davis
David Kieninger
Kevin Mizuta
Euneka Robinson-McCutchen

We appreciate the contributions of all these individuals and organizations.