Study of Indicators of Recurrent Congestion on Urban Roadway Network
Based on Bus Probes

Thesis
Presented in Partial Fulfillment of the Requirements for the Degree Master of Science
in the Graduate School of The Ohio State University

By
Cheng Chen, M.S.
Graduate Program in Civil Engineering

The Ohio State University
2010

Thesis Committee:
Mark R. McCord, Co-adviser
Rabi G. Mishalani, Co-adviser
Prem K. Goel
Abstract

Congestion has long been a concern to transportation operation and planning. Various methods and technologies have been developed and applied in detecting congestion. This study focuses on using bus as probes to indicate recurrent congestion on urban roadway networks.

In this study, bus speeds are used to reflect indications of congestion. Bus speed data collected over “homogeneous days” are pooled in order to find indications of recurrent congestion. Bus speeds collected during a time period at each location are compared to bus speeds collected during other time periods (or during the whole day) at the same location. For practical purposes, the entire bus route investigated for the purpose of method development, validation, and demonstration is discretized into 10 meter spatial sections, and time of day is discretized into 30 minute time intervals. Each 10 meter long spatial section and 30 minute duration time interval constitutes a space-time cell for which existence or absence of an indication of congestion is determined.

Three different methods are used to determine the existence or absence of an indication of congestion for each space-time cell. Two of the three methods, the Mean Based
method and the Median Based method, apply traditional statistical tests. The other method, the Low Speed Threshold (LST) method, is developed in this study.

An empirical study based on Automatic Vehicle Location (AVL) data collected using The Ohio State University (OSU) Campus Transit Lab (CTL) is conducted to evaluate the three methods. The indications of congestion produced by the three methods are then compared to a priori expectation of recurrent congestions along the route.

The Mean Based method produces few indications of recurrent congestion. It misses space-time cells where recurrent congestion is expected. The Median Based method and the LST method produce similar general patterns in terms of where and when recurrent congestions are indicated. However, there exist some differences between the Median Based method and the LST method in terms of the specific locations where and specific times when recurrent congestions are indicated.

A systematic analysis and comparison based on the Median Based method and the LST method is conducted. Based on the empirical analyses and a priori expectation of recurrent congestions along the route, the LST method seems more appealing than the Median Based method (and the Mean Based method) for determining indications of recurrent congestion. Although the empirical results produced by the LST method are encouraging, more research is necessary before concluding that the LST method could be
used to detect recurrent congestion on a widespread basis. A sensitivity analysis of the parameters used in this study is suggested for future research. Other traffic related elements such as speed limits, bus trip running times, and schedule adherence might also be utilized to improve the recurrent congestion detection methods.
Dedication

Dedicated to the students at The Ohio State University
Acknowledgments

I appreciate the experience of working with my two advisors Dr. Mark R. McCord and Dr. Rabi G. Mishalani from Civil Engineering department and Dr. Prem Goel from Statistic Department. Their creativity and intense logic helped a lot in our research. They are also very generous in helping me checking all sorts of potential problems in this thesis. Although they are very busy, they still took their time in reading and commenting on every detail in the thesis. Their inspiration, encouragement, and generosity are much appreciated.

I appreciate the help of Dr. Benjamin A. Coifman in the past four years. He is nice to answer any questions I have. I learned much from his research and courses.

Thanks Dr. Mark R. McCord and Dr. Rabi G. Mishalani again for funding me as a graduate research associate. I enjoyed the weekly research meeting with them and Dr. Prem Goel. Their enthusiasm for scientific research impressed me. Working with them is a rewarding experience. Their unselfish help deserves more than thanks.

Thanks Yuxiong Ji, who has been my friend for 7 years since we were in Tongji University in Shanghai, China, for his help in many aspects of my life.

Thanks to the other fellows I have worked with in the past four years: Xi Zhang, Dawei Lu, Fanyu Zhou, Qianqian Zhang, Aaron Quan, Jinguo Gao, Dunke Zhou, and Yufang Zhang.
Thanks to my family for their unselfish support and encouragement.
Vita

October 04, 1982……………Born – Shanghai, China

2004…………………………B.S., Department of Civil Engineering, Tongji University, China

2006………………………… M.S., Department of Roadway and Railway Engineering, Tongji University, China

2006 - present……………… Graduate Research Associate, Civil and Environmental Engineering & Geodetic Science, the Ohio State University

Fields of Study

Major Field: Civil Engineering
Table of Contents

Abstract ........................................................................................................................................... ii
Dedication ........................................................................................................................................... v
Acknowledgments .............................................................................................................................. vi
Vita ................................................................................................................................................... viii
List of Figures ..................................................................................................................................... xi
List of Tables ....................................................................................................................................... xiii
Lists of Symbols ............................................................................................................................... xiv
List of Abbreviations ....................................................................................................................... xvi
Chapter 1 Introduction ..................................................................................................................... 1
Chapter 2 Methods ........................................................................................................................... 6
  2.1 General Approach ...................................................................................................................... 8
  2.2 Mean Based Method ................................................................................................................. 12
  2.3 Median Based Method .............................................................................................................. 16
  2.4 Low Speed Threshold Method ............................................................................................... 20
  2.5 Spatial Considerations and Congestion Mapping ................................................................. 27
Chapter 3 Empirical Study .............................................................................................................. 29
  3.1 Data Description ....................................................................................................................... 29
  3.2 Parameter Values and General Approach to Applying Methods ........................................... 32
  3.3 Patterns from Three Methods .................................................................................................. 36
    3.3.1 General Morning and Afternoon Peaks ............................................................................. 39
    3.3.2 Congestion Dissipation over Time .................................................................................... 39
    3.3.3 Cell by Cell Comparison ................................................................................................. 42
  3.4 Similarities and Differences between the Median Based Method and the LST Method ......... 44
    3.4.1 Examples of Space-Time Cells Where and When Both Median Based and LST Methods Indicate Congestion ........................................................... 49
List of Figures

Figure 2.1 Illustration of the Speed-Time-Distance Data Cube ......................................... 7
Figure 2.2 Illustration of Discretization of Space and Time.................................................. 9
Figure 2.3 Illustration of Comparison over Time at Fixed Location..................................... 10
Figure 2.4 General Approach for Finding Indication of Recurrent Congestion ............... 11
Figure 2.5 Flowchart of Mean Based Method for a Given Location d............................. 15
Figure 2.6 Flowchart of Median Based Method for a Given Location d........................... 17
Figure 2.7 Flowchart of Low Speed Threshold Method..................................................... 21
Figure 2.8 Illustration of Determining Low Speed Region ................................................ 23
Figure 3.1 CLS Bus Route Map ......................................................................................... 29
Figure 3.2 Strong Indication of Congestion Matrices of Three Methods ( = 0.05, C = 3) ................................................................................................................................. 38
Figure 3.3 Congested Locations (black dots) Indicated by LST Method during 8:00~8:30 ................................................................. 40
Figure 3.4 Congested Locations (black dots) Indicated by LST Method during 9:00~9:30 ................................................................. 41
Figure 3.5 Congested Locations (black dots) Indicated by LST Method during 10:00~10:30 ......................................................................................... 42
Figure 3.6 Comparison of Median Based Method and LST Method ................................ 45
Figure 3.7 Either Median Based Method or LST Method Indication of Congestion (Black) ......................................................................................................................... 47
Figure 3.8 Number of Congested Spatial Sections Indicated by Either Method .............. 47
Figure 3.9 Locations of the Six Examples ......................................................................... 49
Figure 3.10 Location of Example 1: Both Methods Have Congestion Indication............ 51
Figure 3.11 Scatter Plot of Speed against Time of Day at Location Example 1 .......... 51
Figure 3.12 Location of Example 2: Both Methods Have Congestion Indication............ 53
Figure 3.13 Scatter Plot of Speed against Time of Day at Location Example 2 ........... 54
Figure 3.14 Location of Example 3: Only Median Based Method Has Congestion Indication .......................................................................................................................... 55
Figure 3.15 Scatter Plot of Speed against Time of Day at Location Example 3 .......... 56
Figure 3.16 Location of Example 4: only Median Based Method Has Congestion Indication .......................................................................................................................... 60
Figure 3.17 Scatter Plot of Speed against Time of Day at Location Example 4 .......... 61
Figure 3.18 Location of Example 5: Only LST Method Has Congestion Indication ......... 63
Figure 3.19 Scatter Plot of Speed against Time of Day at Location Example 5 .......... 63
Figure 3.20 Location of Example 6: only LST Method Has Congestion Indication ........ 65
Figure 3.21 Scatter Plot of Speed against Time of Day at Location Example 6 .......... 66
Figure B.1 Congested Locations (black dots) Indicated by Median Based Method during 8:00~8:30 .......................................................................................................................... 80
Figure B.2 Congested Locations (black dots) Indicated by Median Based Method during 9:00~9:30 .......................................................................................................................... 81
Figure B.3 Congested Locations (black dots) Indicated by Median Based Method during 10:00~10:30 ...................................................................................................................... 82
Figure C.1 Combined Strong Indication of Congestion Matrix (Median Based vs. LST) 84
List of Tables

Table 3.1 Bus Stops for Campus Loop South Route ........................................................ 30
Table 3.2 Parameters and Parameter Values Used in the Empirical Study ................. 32
Table 3.3 Cell Level Comparison of the Median Based Method and the LST Method ... 44
Table 3.4 Number of Cells Indicated/Not Indicated as Congested by Median Based Method and LST Method................................................................. 46
Table A.1 Number of Cells Where Either Both Methods Indicate Congestion or Both Do Not Indicate Congestion (C = 3) for Each Pair of Methods ............................................. 78
Table A.2 Number of Cells with Strong Indication of Congestion by Both Methods for Each Pair of Methods........................................................................................................ 79
Table A.3 Based on Independence Assumption, Number of Expected Congestion Cell under Each Pair of Methods.................................................................................. 79
Table C.1 “Primitive” List of Cell Clusters with Congestion Indications ......................... 86
Table C.2 “Long Span” Cell Clusters with Congestion Indication .................................. 88
Table C.3 Locations for Detailed Analysis...................................................................... 91
Lists of Symbols

$\Delta d = \text{step size in space.}$

$d_{\text{smooth}} = \text{smoothing distance}$

$d_e = \text{route length.}$

$\Delta t = \text{step size in time.}$

$t_{\text{hom}} = \text{homogeneous time interval.}$

$t_s = \text{start time.}$

$t_e = \text{end time.}$

$v = \text{observed speed.}$

$N = \text{number of observations in the pooled sample (speeds obtained at location } d \text{ for the entire day).}$

$F(v) = \text{CDF of pooled sample at speed } v.$

$N' = \text{number of observations in the interval sample (speeds obtained at location } d \text{ for time interval } [t-t_{\text{hom}}/2, t+t_{\text{hom}}/2]).}$

$F'(v) = \text{CDF of interval sample at speed } v.$

$M_{d,t} = \text{Number of observations in the low speed region for space-time cell } (d, t).$

$\mu_i = \text{mean population speed in the } i^{th} \text{ time interval.}$

$r = \text{total number of time intervals.}$

$n_j = \text{number of speed observations during the } j^{th} \text{ time interval.}$
\( N_i^c \) = total number of speed observations during the complement of the \( i^{th} \) time interval.

\( \bar{v}_i \) = average speed during the \( i^{th} \) time interval.

\( \alpha \) = significance level

\( \text{Var}(\hat{C}_i) \) = estimated variance of \( C_i \).

\( \hat{\sigma}_e^2 \) = estimated pooled standard deviation of speed within each time interval.

\( \lambda_i \) = median speed of the \( i^{th} \) time interval.

\( \lambda_{ic} \) = median speed of the complement of the \( i^{th} \) time interval.

Quantile \( \alpha(s) \) = \( d^{th} \) quantile among observations in set \( s \).

\( q_{\text{freeflow}} \) = parameter specifying the value of the above \( d^{th} \) quantile.

\( plow \) = parameter specifying the ratio of the low speed threshold to the bus free flow speed.

\( p_t \) = proportion of low speed observations during \([t-t_{hom}/2, t+t_{hom}/2]\).

\( p_{\text{whole}} \) = proportion of low speed observations during the whole day.

\( Y(v) \) = the number of speed observations smaller than or equal to \( v \).
List of Abbreviations

AVL = Automatic Vehicle Location
GPS = Global Positioning System
CTL = Campus Transit Lab
OSU = the Ohio State University
CDF = Cumulative Distribution Function
LST = Low Speed Threshold
Chapter 1 Introduction

Congestion has long been a concern to transportation planners and operators. Typically, two types of congestion are considered: non-recurrent congestion and recurrent congestion. Non-recurrent congestion refers to unexpected congestion caused by unpredictable or transient events such as accidents, unusual weather, or construction (Bremmer, 2004). On the other hand, recurrent congestion refers to relatively predictable congestion caused by interaction of routine traffic demand with the fixed infrastructure. This thesis focuses on developing and analyzing indicators of recurrent congestion.

Congestion detection is of value for off-line and real-time traffic management and could potentially provide useful inputs to bus transit planners and operators. Therefore, different sensing technologies have been applied in automatic congestion detection. Typical sensors include manual traffic counters, loop detectors, and camera detectors (Coifman, 2006). More recently, probe vehicles (FHWA Travel Time Data Collection Handbook, 2009) utilizing Automatic Vehicle Location (AVL) technologies such as those based on the Global Positioning System (GPS) have been considered for detecting congestion. Assuming probe vehicles travel in a manner representative of surrounding traffic, the probe vehicle techniques are typically designed for collecting real-time traffic data. However, the collected data from probe vehicles can also be used off-line to determine recurrent congestion on urban roadway networks.
One of the advantages of using buses as probes is that they travel pre-determined routes regularly. This enables the development of a data base from repeated observations covering time-of-day or day-of-week conditions (Abdelaziz, 2009). In addition, buses have extensive space and time coverage. Furthermore, the additional cost for the purpose of detecting congestion is low when buses are already equipped with AVL systems.

Berinizon (1993) reports one of the earliest actual deployments where buses are used as probes for monitoring traffic operations. However, the cited paper only describes the efforts of deploying such system and no detailed results regarding this deployment are presented. Dailey and Cathey (2002, 2003, 2005, 2006) developed a real-time system using AVL equipped buses as probes for monitoring traffic operations for the Seattle, WA metropolitan area. The primary focus of their work appears to be freeways, with all of the independent evaluation being limited to comparisons with freeway loop detector data. Freeway loop detector data are readily available. The positive correlation between freeway travel times as measured by loop detectors and by transit AVL data has been examined and confirmed by, for example, Cathey and Dailey (2003b), El-Geneidy and Bertini (2004), and Coifman and Kim (2006). However, on arterials and urban streets, using buses as probes for inferring the surrounding traffic conditions is more challenging because transit vehicles tend to behave more differently than the general traffic on these roadways.

Bertini et al. (2003a, 2003b, 2004, 2005, 2008) studied the correlation of transit buses with surrounding vehicles on arterials. More specifically, they studied the correlation of
the AVL data from transit buses with data from floating vehicles on arterials. Bertini and Tantiyanugulchai (2003b) introduced two methods to aggregate the data from a given bus traveling through a given corridor. One method creates a hypothetical bus trajectory, which assumes buses travel as if there are no bus stops by subtracting estimated bus stop dwell times. The other method defines a pseudo bus trajectory, which assumes a bus travels at the maximum speed recorded by the transit vehicle within each link. The authors found the pseudo bus trajectory to be closest to the floating vehicle trajectories on average. Bertini et al (2005) then applied this method to evaluate traffic signal improvements and compared the results against concurrent floating vehicle data. While both the floating vehicle and transit AVL data show the signal timing yielded an improvement, the authors show that the methods produced results that differed by as much as 24%.

Chakroborty and Kikuchi (2004) studied five arterial corridors in Delaware to estimate automobile travel time as a function of the transit vehicle travel time less bus stop dwell times. The results were not encouraging, and it was concluded that more research in this area is necessary.

Hall and Vyas (2001) tried to use transit AVL data to detect the presence or absence of real-time congestion on an arterial. It was found that most of the delays in the floating vehicle trips are captured by the transit bus trips, but the reverse relationship did not necessarily hold. One of the largest factors was simply the fact that the transit vehicles would wait not only at intersections bus also at bus stops. Furthermore, the segments
chosen were so small that a single traffic light could significantly impact the accuracy of a prediction. The authors concluded that not enough data were gathered for the algorithm to be generally applicable and stated that "bus speed alone cannot be effective in detecting true incidents, because the false alarm rate would be high."

More recently, Berkow et al (2008) attempted to improve congestion monitoring and travel time estimation by combining the data from a bus AVL system and data from signal system detectors on an arterial corridor. In their study, AVL data are used with data from the signal system detectors to help the estimation of travel times between two consecutive detectors, however there is no mention of detect recurrent congestion. In this thesis, the emphasis is on using AVL data alone to detect recurrent congestion.

In summary, there is limited research on using bus probes to estimate traffic conditions on urban streets, and there is no known study that focuses on detecting recurrent congestion on urban roadways using bus AVL data.

Detecting congestion on a roadway network consisting of local streets, collectors, and arterials is challenging because of the existence of signalized intersections, stop signs, and pedestrian crossings. However, as stated earlier, the extensive space-time coverage that can be obtained when using bases as probe and relatively low additional cost of using the available data motivate addressing these challenges.
This study attempts to use historical AVL data to find indications of recurrent congestion on urban roadway networks. Three methods are developed and their results are compared using AVL data obtained from the Campus Transit Lab (CTL) at The Ohio State University (OSU). Fairly satisfactory results from two of the three methods are obtained by comparing the recurrent congestion indications to a priori expectations of the traffic patterns on the roadway network examined. One method appears to produce the most satisfactory results in terms of the consistency of indications of recurrent congestion with a priori expectations.
Chapter 2 Methods

In this study, bus speed obtained from a bus automatic vehicle location (AVL) system is used as an indicator of traffic congestion. To apply the approaches developed in this research, time-stamped bus speeds at locations along a route are either obtained as basic inputs or derived from location and time data.

Since the purpose of this study is to find indications of “recurrent” congestion, the location, time of day, and speed data should be available over multiple days. Under the assumption that the speed patterns are the same on all the weekdays, speeds obtained on different weekdays are combined. Figure 2.1 is an illustration of bus speeds together with location along the bus route and time of day after combining data across multiple week days.

Since urban roadway networks are spatially heterogeneous, it is not reasonable to search for indications of congestion by comparing speeds over different locations. Therefore, in this study, speeds at the same location are compared over time of day in order to find time periods when recurrent congestion is indicated. For a fixed location, when the speed records obtained during a certain time period are systematically lower than the speed records obtained during the other times of the day, an indication of congestion is considered to be present. The problem then is to determine an approach that provides an operational determination of “systematically lower speeds”. This study developed three
approaches for determining which time periods exhibit “systematically lower speeds” for a given location.

These three approaches use statistical tests as tools to help achieve this determination. Two of the approaches use classical statistical tests, the ANOVA test and the two sample t-test for one of the two approaches and the nonparametric K-W rank-sum test and Wilcoxon rank-sum test for the other approach. As far as the author is aware, these two tests have not been applied to detect recurrent congestion using bus probes. As for the
third approach, the procedure and the statistical test consist of an investigation and application of statistical ideas customized specifically for the purposes of this study.

2.1 General Approach

In Figure 2.1, the distance and time are continuous variables. However, in this study, location (characterized by distance along a route with respect to a reference point) is discretized into spatial sections of fixed length $\Delta d$ called step size in space, and time is discretized into fixed time intervals of fixed duration $\Delta t$ called step size in time. As a result, the combined space-time domain is discretized into space-time cells with a spatial resolution of $\Delta d$ meters and time resolution of $\Delta t$ minutes. (In this study, $\Delta d$ is set at 10 meters and $\Delta t$ is set at 30 minutes.)

In order to compare speeds for space-time cells covering different times at a fixed location $d$, enough speed data at or around location $d$ is required. By assuming speeds over a certain distance $d_{smooth}$ are homogeneous, speeds obtained within plus or minus $d_{smooth}/2$ from location $d$ are grouped for analyzing indications of congestion for the spatial section centered at location $d$. For simplicity, a spatial section centered at location $d$ will be referred to as location $d$.

Like the assumption of homogeneous speed distribution over distance $d_{smooth}$, for a fixed location $d$, speeds observed during a certain time interval $t_{hom}$ are assumed to be homogeneous. Combining the homogeneity assumptions over space and over time, speed
data obtained over space $[d-d_{\text{smooth}}/2, d+d_{\text{smooth}}/2]$ and during time $[t-t_{\text{hom}}/2, t+t_{\text{hom}}/2]$ are grouped as speeds representative for the $(d, t)$ space-time cell. (In this study, $d_{\text{smooth}}$ is set at 100 meters and $t_{\text{hom}}$ is set at 30 minutes.)

Figure 2.2 illustrates the discretization of continuous space and time into $(d, t)$ space-time cells. A bus route with a reference start location $d_s$ and end location $d_e$ is first discretized into spatial sections with $\Delta d$ increment. For each spatial section $d$, the speeds within $[d-d_{\text{smooth}}/2, d+d_{\text{smooth}}/2]$ are grouped to represent the speeds for that spatial section. The speeds at each spatial section $d$ are then further grouped into different time intervals with increment $\Delta t$, where the reference start time is denoted by $t_s$ and end time is denoted by $t_e$.

![Figure 2.2 Illustration of Discretization of Space and Time](image_url)
Figure 2.3 illustrates the comparison over time at a fixed location. As discussed in the above paragraph, speeds are first grouped into their corresponding spatial sections. Then, speeds at spatial section $d$ are further grouped into different time intervals. Some aspect of the speed distribution during a time interval is then compared to the same aspect of the speed distribution during other time intervals or during the whole day. This comparison is performed for each time interval, one after another as indicated in Figure 2.3. Indications of congestion are inferred from these comparisons.
Figure 2.4 General Approach for Finding Indication of Recurrent Congestion

Figure 2.4 shows the flow chart of the general approach for finding indications of recurrent congestion along a bus route. The process of finding indications of congestion starts from location $d_s + d_{\text{smooth}}/2$. At this location, speeds located within $(d - d_{\text{smooth}}/2, d + d_{\text{smooth}}/2)$ are selected. These selected speeds are then grouped into time intervals with duration $t_{\text{hom}}$. These time intervals with duration $t_{\text{hom}}$ are centered at time points that are

Initialization:
$$d = d_s + d_{\text{hom}}/2$$

Group speeds obtained within $(d - d_{\text{hom}}/2, d + d_{\text{hom}}/2)$ into time intervals as indicated in Figure 2.2 with duration $t_{\text{hom}}$

Use the methods described in sections 2.2~2.4 to identify time intervals that have indication of recurrent congestion

$$d = d + \Delta d$$

If $d < d_e - d_{\text{hom}}/2$, then:

Y

End

N
\( \Delta t \) apart from each other starting from \( t_s \) and ending in \( t_e \). (In this study, the starting time \( t_s \) is 7:00 and the ending time \( t_e \) is 18:30.) For each time interval, indication of recurrent congestion is checked by the methods described in sections 2.2~2.4. After checking all the time intervals, an increment of \( \Delta d \) over space is applied, and the steps are repeated, to check if there is indication of recurrent congestion at location \( d + \Delta d \). The increment of \( \Delta d \) over space continues until the distance \( d \) is larger than \( d_c - d_{smooth}/2 \).

To determine if an indication of recurrent congestion is present at a fixed location \( d \), different statistical tests are applied. Sections 2.2 to 2.4 describe three such methods. Section 2.2 describes a Mean Based method; section 2.3 describes a Median Based method; section 2.4 describes a low speed threshold method.

### 2.2 Mean Based Method

In the Mean Based method, an indication of congestion is considered to be present at location \( d \), during the time period \( [t-t_{hom}/2, t+t_{hom}/2] \), if the calculated mean speed during \( [t-t_{hom}/2, t+t_{hom}/2] \) is statistically significantly less than a weighted average of the mean speeds during the complement of \( [t-t_{hom}/2, t+t_{hom}/2] \) at that location. The Mean Based method is composed of two main steps as depicted in the flow chart of Figure 2.5.

In the first step, an \( F \)-test (Rice, 2005) is performed on the mean speeds across all the time intervals. The purpose of the \( F \)-test is to detect if there are any differences in the population mean speeds across the set of non-overlapping time intervals at each fixed
location. The *F-test* is performed once for each location \( d \). The assumption of the *F-test* is that the speeds during each time interval are normally distributed with the same variance.

The null (H\(_0\)) and alternative (H\(_a\)) hypotheses of the *F-test* are:

\[ H_0: \mu_1 = \mu_2 = \cdots = \mu_r \]
\[ H_a: \mu_i \)'s are not all equal \]

Where:

\( \mu_i \) = mean population speed in the \( i^{th} \) time interval, and
\( r \) = total number of time intervals.

If the null hypothesis is rejected under a certain significance level \( \alpha \), the second step is performed. Otherwise, it is concluded that location \( d \) has no indication of congestion, since there is no time interval having mean speed significantly different from any other interval.

In the second step of the Mean Based method, Scheffé's procedure (Rice, 2005) is performed to determine which time intervals have lower population mean speeds than a weighted average of the mean speeds during other time intervals. The null (H\(_0\)) and alternative (H\(_a\)) hypotheses are:

\[ H_0: C_i = \mu_i - \sum_{j \neq i} w_j \mu_j = 0 \]
\[ H_a: C_i < 0 \]

Where:

\[ w_j = \frac{n_j}{N_i} \]

Where:
\( n_j \) = number of speed observations during the \( j^{th} \) time interval, and
\( N_i^c \) = total number of speed observations during the complement of the \( i^{th} \) time interval.

Notice that \( \sum_{j \neq i} w_j = 1 \).

\( C_i \) can be estimated by:

\[
\hat{C}_i = \bar{v}_i - \sum_{j \neq i} w_j \bar{v}_j
\]  

(2.1)

Where:

\( \bar{v}_i \) = average speed during the \( i^{th} \) time interval.

The mean speed during the \( i^{th} \) time interval is claimed to be significantly lower than the average of the mean speeds during other time intervals if \( \hat{C} \) is less than a certain cutoff point. The cutoff point is calculated as (Rice, 2005):

\[
\beta = -\sqrt{(r - 1)F_{r-1,N-r,\alpha}} \sqrt{\text{Var}(\hat{C}_i)}
\]  

(2.2)

Where:

\( N \) = total number of speed observations,
\( \alpha \) = significance level used in the first step, and
\( \text{Var}(\hat{C}_i) \) = estimated variance of \( C_i \).

This estimated variance is calculated by (Rice, 2005):

\[
\text{Var}(\hat{C}_i) = \left( \frac{1}{n_i} + \sum_{j \neq i} \frac{w_j^2}{n_j} \right) \hat{\sigma}_e^2
\]  

(2.3)

Where:

\( \hat{\sigma}_e^2 \) = estimated pooled standard deviation of speed within each time interval.
Applying Scheffé's method produces a binary output for each time interval $i$ indicating whether the mean speed during the $i^{th}$ time interval is significantly lower than the average of the mean speeds during other time intervals.

Figure 2.5 Flowchart of Mean Based Method for a Given Location $d$
When the process is repeated for all locations $d$, a binary matrix indicating whether there is a significant decrease in the mean speed during a time interval compared to the weighted average of the mean speeds during other time intervals at each location is produced. Location $d$ is claimed to reflect an indication of congestion during $[t-t_{hom}/2, t+t_{hom}/2]$ if Scheffé's method for cell $(d, t)$ indicates a lower mean speed during $[t-t_{hom}/2, t+t_{hom}/2]$ compared to the weighted average of the mean speeds during other time intervals.

2.3 Median Based Method

The Mean Based method in section 2.2 requires the assumptions that the speed distributions are normal and have the same variance for each time interval at a given location. The normality assumption may be reasonable at some locations under some conditions. However, it is likely to be violated at some locations. An example would be intersections where the stopping of vehicles is expected to result in a bi-modal speed distribution. Therefore, a method, referred to as the Median Based method, that does not require the normality assumption is developed. The flow chart of the Median Based method for a given location $d$ is depicted in Figure 2.6.
Similar to the Mean Based method, the Median Based method also consists of two steps. However, there are two major differences between the Median and Mean Based methods.
The $K$-$W$ rank-sum test (Hollander et al, 1999) is performed in the first step of the Median Based method, rather than the $F$-test in the first step of the Mean Based method, and

The Wilcoxon rank-sum test (Hollander et al, 1999) is performed in the second step of the Median Based method.

Statistically, the $K$-$W$ rank-sum test detects whether the medians of all populations are the same. The assumption of the $K$-$W$ rank-sum test is that all the time interval population distributions have identical shape, i.e. the distributions can vary along the speed axis as long as the entire shape remains unchanged. The null and alternative hypotheses of the $K$-$W$ rank-sum test are:

$H_0: \lambda_1 = \lambda_2 = \cdots = \lambda_r$

$H_a: \lambda_i$’s are not all equal

Where:

$\lambda_i$ = median speed of the $i^{th}$ time interval.

In this study, the $K$-$W$ rank-sum test is performed to determine whether the median speeds of all time intervals are equal, again, under the assumption that all the time interval population distributions have identical shape.

Similar to the first step in the Mean Based method, if the null hypothesis of the $K$-$W$ rank-sum test is not rejected by yielding a p-value larger than a specified significance level $\alpha$, the second step in the Median Based method is not performed. On the other hand, if the $K$-$W$ rank-sum test rejects the null hypothesis by yielding a p-value less than the
specified significance level $\alpha$, then the second step of the Median Based method is performed to determine which time intervals could have indications of congestion.

In the second step of the Median Based method, the nonparametric one-sided Wilcoxon rank-sum test is performed. The test detects whether there is any difference in the medians of two populations. The assumption of the Wilcoxon rank-sum test is that the two populations have identical shape. The null and alternative hypotheses of the Wilcoxon rank-sum test are:

$$H_0: \lambda_i = \lambda_i^c$$

$$H_a: \lambda_i < \lambda_i^c.$$  

Where:

$\lambda_{tc} = \text{median speed of the complement of the } i^{th} \text{ time interval}.$

The Wilcoxon rank-sum test (Rice, 2005) is applied on two sets of populations of speeds, one on $[t-t_{hom}/2, t+t_{hom}/2]$ and the other on the complement of $[t-t_{hom}/2, t+t_{hom}/2]$. If the null hypothesis of the Wilcoxon rank-sum test is rejected under the significance level $\alpha$ used in the $K-W$ rank-sum test, then it is claimed that there is an indication of congestion for the time interval $[t-t_{hom}/2, t+t_{hom}/2]$.

When the process is repeated for all locations $d$, a binary matrix indicating whether there is a significant decrease in the median speed during a time interval compared to the median speed during other time intervals at each location is produced. Location $d$ is claimed to reflect an indication of congestion during $[t-t_{hom}/2, t+t_{hom}/2]$ if the $K-W$ rank-
sum test for cell \((d, t)\) indicates a lower median speed during \([t-t_hom/2, t+t_hom/2]\) compared to the median speed during other time intervals.

### 2.4 Low Speed Threshold Method

Like the Mean Based method, the Median Based method suffers from the assumption that the distributions of speeds during each time interval \([t-t_hom/2, t+t_hom/2]\) have the same shape at a fixed location \(d\). Like the normality assumption in the Mean Based method, the same shape assumption in the Median Based method may not hold at some locations. Therefore, another method, referred to as the low speed threshold method, is proposed.

The main idea behind this method is to consider indications of congestion associated with a relatively larger proportion of low speeds for an interval given a location. The basic premise being that congestion results in systematically lower speeds when it occurs.

More specifically, in the low speed threshold method, the proportion \(p\) of low speeds during \([t-t_hom/2, t+t_hom/2]\) is compared to the proportion \(p_{whole}\) of low speeds during the whole day at each of the fixed locations \(d\). Unlike the Mean Based method and the Median Based method, the low speed threshold method compares the speeds during \([t-t_hom/2, t+t_hom/2]\) to the speeds during the whole day rather than to the speeds during the complement of \([t-t_hom/2, t+t_hom/2]\).
Figure 2.7 Flowchart of Low Speed Threshold Method

Input: Speeds over space \([d-d_{hom}/2, d+d_{hom}/2]\)

Obtain: \(N, V_{threshold}, \) and \(F(V_{threshold})\) during whole day

\[t = t_s + t_{hom}/2\]

Obtain: \(N', F'(V_{threshold})\) during \([t-t_{hom}/2, t+t_{hom}/2]\)

Calculate p-value by equation 2.4

Store p-value in the p-value matrix at \((d, t)\)

\[t = t + t_{hom}\]

\(Y\)

Calculate p-value by equation 2.4

Store p-value in the p-value matrix at \((d, t)\)

\(N\)

Claim indications of congestion for time intervals during which the proportion test claims a increase in the low speed proportion result

End
Again, the low speed threshold method is based on the proportion of low speeds. The flowchart of the low speed threshold method for a given location \( d \) is depicted in Figure 2.7. A speed is considered a low speed if it is smaller than a specified threshold speed \( V_{threshold} \). For a fixed location \( d \), a low speed region is defined as the interval \([0, V_{threshold}]\).

Several threshold definitions could be proposed and investigated. In this study, \( V_{threshold} \) for location \( d \) is defined as a proportion of the “bus free flow speed” \( V_{freeflow} \). The bus free flow speed is defined as a percentile of the observed speeds for the entire day as follows:

\[ V_{freeflow} = \text{Quantile}_{q_{freeflow}} \text{(whole day speed observations)} \]  

(2.4)

Where:

\( \text{Quantile}_a(s) = a^{th} \) quantile among observations in set \( s \),

\( q_{freeflow} = \) parameter specifying the value of the above \( a^{th} \) quantile, and

\( \text{whole day speed observations} = \) set of empirical speeds observed during the whole day.

Given \( V_{freeflow} \), \( V_{threshold} \) is specified as:

\[ V_{threshold} = p_{low} \times V_{freeflow} \]  

(2.5)

Where:

\( p_{low} = \) parameter specifying the ratio of the low speed threshold to the bus free flow speed.

Figure 2.8 illustrates these definitions. In the illustration, the bus free flow speed is taken to be the 95\(^{th}\) quantile (i.e. \( q_{freeflow} = 95\% \)) and \( p_{low} \) is set to 0.35. The 95\(^{th}\) quantile is chosen with the intention to reflect a relatively stable value for the maximum speed. The 0.35 proportion is chosen with the intention to produce an upper bound on speeds indicative of congested conditions (in fact the resulting empirical speeds are comparable
to the minimum urban street speed standard suggested by the National Committee on Urban Transportation according to Lomax (1997)). Those speeds contained within the shaded low speed region in Figure 2.8 are therefore considered to be low speeds based on the above specification.

In the low speed threshold method, a statistical test is performed to test whether the proportion \( p_t \) of low speed observations during \( [t-t_{hom}/2, t+t_{hom}/2] \) is statistically significantly larger than the proportion \( p_{whole} \) of low speed observations during the whole day. That is, it is desirable to test whether there are a relatively larger number of low
speeds in \([t-t_{thom}/2, t+t_{thom}/2]\) than would be expected by chance given the distribution of speeds for the entire day. The null and alternative hypotheses of the test are:

\[H_0: p_t = p_{whole}\]

\[H_a: p_t > p_{whole}\]

Denote the number of observations of speeds during the whole day and during \([t-t_{thom}/2, t+t_{thom}/2]\) by \(N\) and \(N'\), respectively. And, denote the cumulative distribution functions of speed during the whole day and during \([t-t_{thom}/2, t+t_{thom}/2]\) by \(F(\cdot)\) and \(F'(\cdot)\), respectively. Also denote the number of speed observations smaller than or equal to a speed \(v\) observed within the space time cell \((d, t)\) by \(M_{d,t}(v)\). That is, \(M_{d,t}(v) = N' \times F'(v)\). Let \(Y(v)\) be a variable denoting the number of speed observations smaller than or equal to \(v\).

Define \(P_v(Y(v) \geq M_{d,t}(v))\) as the probability of obtaining an empirical low speed proportion greater than or equal to \(F'(v)\) for cell \((d, t)\) when \(N'\) samples are taken randomly from the whole day population without replacement. Theoretically, the probability \(P_v\left(Y(V_{threshold}) \geq M_{d,t}(V_{threshold})\right)\) is based on the hyper geometric distribution where:

\[
P(Y(V_{threshold}) = n|N') = \binom{M}{n} \binom{N'-M}{n'-n} \binom{N}{N'}
\]

However, if \(N'\) is small compared to \(N\), i.e., \(\frac{N-N'}{N-1}\) is close to 1, this probability under the without replacement sampling scheme can be approximated by that under the with
replacement sampling scheme. Given this approximation, this probability is given by the binomial distribution as follows:

\[ P(Y(V_{\text{threshold}}) = n|N') = \binom{N'}{n} (F(V_{\text{threshold}}))^n \times (1 - F(V_{\text{threshold}}))^{N' - n} \quad (2.7) \]

When \( P_v \left( Y(V_{\text{threshold}}) \geq M_{d,t}(V_{\text{threshold}}) \right) \) is small, then \( M_{d,t}(V_{\text{threshold}}) \) is large compared to what is expected under the random sampling hypothesis that the speeds in the given time interval were drawn at random from the collection of daily speeds at this location.

Based on equation 2.7, for \( v = V_{\text{threshold}} \) and \( F'(v) = F'(V_{\text{threshold}}) \):

\[
P_v \left( Y(V_{\text{threshold}}) \geq M_{d,t}(V_{\text{threshold}}) \right)
= \sum_{n = M_{d,t}(V_{\text{threshold}})}^{N'} P(Y(V_{\text{threshold}}) = n|N')
= \sum_{n = M_{d,t}(V_{\text{threshold}})}^{N'} \binom{N'}{n} (F(V_{\text{threshold}}))^n \times (1 - F(V_{\text{threshold}}))^{N' - n} \quad (2.8)
\]

In testing the null hypothesis that \( p_t = p_{\text{whole}} \) with the alternative hypothesis \( p_t > p_{\text{whole}} \), \( P_v \left( Y(V_{\text{threshold}}) \geq M_{d,t}(V_{\text{threshold}}) \right) \) is defined to be the approximated p-value.

Using equation 2.8, an approximated p-value is calculated for each space-time cell. As a result, a two dimensional approximated p-value matrix is produced by the low speed threshold method. For each \((d, t)\) space-time cell in the two dimensional approximated p-value matrix, a corresponding approximated p-value calculated by equation 2.8 is stored. This approximated p-value matrix enables the identification of the time intervals when
the percentage of low speeds is significantly larger than the percentage of low speeds during the whole day. Location $d$ is claimed to reflect an indication of congestion during $[t-t_{hom}/2, t+t_{hom}/2]$ if the approximated p-value at the cell $(d, t)$ in the approximated p-value matrix is less than a significance level $\alpha$.

Having introduced all the three methods, it is worthwhile to note the differences among the three methods in terms of what they consider. Firstly, they consider different aspects of a speed distribution. The Mean Based method considers the mean of the distributions, the Median Based method considers the median of the distributions, and the LST method considers the probability mass in lower tail of the distributions. Secondly, the above three methods have different type I error rates. For each space-time cell, the LST method rejects the null hypothesis depending only on the speed distributions of the corresponding space-time cell and of the corresponding spatial section across the entire day. It therefore has an on average type I error rate of 0.05 over all the space-time cells. However, it does not control the overall probability of at least one type I error across all the time intervals at each location. The Median Based method applies the $K-W$ Rank Sum test in the first step. This implies that before rejecting the null hypothesis in the second step for a certain space-time cell, the null hypothesis in the first step need to be rejected. It therefore has a more restrictive criterion for rejecting the null hypothesis for each space-time cell than the LST method. The Mean Based method applies Scheffé's procedure to determine which time intervals have significantly lower mean speeds. This results in the fact that the Mean Based method has the most restrictive criterion for rejecting the null hypothesis for each space-time cell among all the three methods. Therefore, it should be expected
that the Mean Based method yields the least number of space-time cells with indication of congestion and the LST method yields the most.

2.5 Spatial Considerations and Congestion Mapping

Given a certain significance level $\alpha$, each of the three methods described in Sections 2.2~2.4 produces an indication of congestion, according to the methods’ criteria, at location $d$ at time interval $t$. In other words, the three methods all yield a binary two dimensional matrix indicating where and when there is indication of congestion.

At this point, a “C-space rule” is applied to strengthen the conclusion about indications of congestion for the cells $(d, t)$ of the two dimensional space-time matrices by taking spatial considerations into account. This C-space rule states that the space-time cell $(d, t)$ has a strong indication of congestion if congestion indications are present at $C$ consecutive cells over space, where the space cell $d$ is in the middle of the series of $C$ cells, during the time interval $t$. The reasons for applying the C-space cell rule are the following.

1. The space time cells $(d, t)$ and $(d+\Delta d, t)$ have on average an overlap of $\left(d_{\text{smooth}} - \Delta d/d_{\text{smooth}}\right) \times 100\%$ of the speed observations. Therefore, they should be more likely to yield the same indication of congestion. If congestion is detected only at $(d, t)$ but not at the neighboring cells over space, then the indication of congestion is very likely to be a false positive, i.e., a type II error.
Congestion usually occurs over a spatial region longer than 10 meters. Therefore, it is not reasonable to conclude a strong indication of congestion at \((d, t)\) if there is no congestion indicated at the neighboring cells.

By applying the \(C\)-space rule, an updated matrix is obtained. This updated matrix is referred to as the strong indication of congestion matrix.

Finally, after applying the \(C\)-space rule, it is more meaningful to visualize the results on a corresponding map that shows the physical locations along the bus route. Such visualization allows an easier comparison of congestion indications with \textit{a priori} expectations as will become evident in chapter 3.
Chapter 3 Empirical Study

3.1 Data Description

An empirical study is conducted to investigate the performance of the three methods described in chapter 2. Bus AVL data are collected from the Campus Transit Lab (CTL) at the Ohio State University (OSU), where AVL-equipped buses run on multiple routes serving approximately 4 million passengers annually. The AVL system in use when the data were collected for this study was a “homemade” system designed to record the location of a bus every 100 meters or every 3 minutes, depending on which event occurred first (Ji 2006). The 100 meters’ criterion is generally satisfied first.

Figure 3.1 CLS Bus Route Map
The speed data used in this empirical study are computed from time-stamped location data for buses running on the Campus Loop South (CLS) route between 7:00 and 18:30 on weekdays from 03/29/2004 to 06/24/2004 (ten weeks). This period corresponds to the OSU academic spring quarter in 2004. The layout of CLS route is depicted in Figure 3.1. CLS was selected in part because of familiarity with the traffic conditions along the roadways on which the route is located. The route is 8,334 meters in length, runs in a loop, and traverses the campus bus stops indicated in Table 3.1.

<table>
<thead>
<tr>
<th>Stop #</th>
<th>Name</th>
<th>Cumulative Distance From Stop 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bevis Hall</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Carmack Corner</td>
<td>159</td>
</tr>
<tr>
<td>3</td>
<td>Carmack 4</td>
<td>312</td>
</tr>
<tr>
<td>4</td>
<td>Carmack 1</td>
<td>717</td>
</tr>
<tr>
<td>5</td>
<td>Blankenship Hall</td>
<td>1245</td>
</tr>
<tr>
<td>6</td>
<td>Ag Campus EB</td>
<td>1842</td>
</tr>
<tr>
<td>7</td>
<td>St John Arena EB</td>
<td>2371</td>
</tr>
<tr>
<td>8</td>
<td>Drake Union</td>
<td>3038</td>
</tr>
<tr>
<td>9</td>
<td>Cannon &amp; 12th (SB)</td>
<td>3478</td>
</tr>
<tr>
<td>10</td>
<td>Med Center Dr. &amp; Cannon</td>
<td>3984</td>
</tr>
<tr>
<td>11</td>
<td>Med Center Dr. &amp; 9th</td>
<td>4170</td>
</tr>
<tr>
<td>12</td>
<td>Neil Ave.&amp; 10th Ave.</td>
<td>4522</td>
</tr>
<tr>
<td>13</td>
<td>Mack Hall</td>
<td>4782</td>
</tr>
<tr>
<td>14</td>
<td>Hale Hall</td>
<td>5088</td>
</tr>
<tr>
<td>15</td>
<td>Ohio Union NB</td>
<td>5305</td>
</tr>
<tr>
<td>16</td>
<td>Arps Hall</td>
<td>5778</td>
</tr>
<tr>
<td>17</td>
<td>North Dorms</td>
<td>6075</td>
</tr>
<tr>
<td>18</td>
<td>St. John Arena (WB)</td>
<td>6803</td>
</tr>
<tr>
<td>19</td>
<td>Ag Campus (WB)</td>
<td>7301</td>
</tr>
</tbody>
</table>

Table 3.1 Bus Stops for Campus Loop South Route
The raw AVL data records include, among other things, the location of the bus in latitude and longitude, time, date, and instantaneous GPS speed when the record is triggered.

Previous efforts resulted in projecting the latitude and longitude in the raw data onto a linear reference system using cumulative distance from bus stop 1 (Bevis Hall), which served as the reference starting point for this route. That is, bus stop 1 is considered to have cumulative distance of 0. The cumulative distances of the bus stops are presented in Table 3.1.

One measure of speed, referred to as the arc speed, is considered in this study. In empirical studies not reported here, arc speeds appeared to have more distinct patterns over time and space than patterns formed from instantaneous GPS speeds for the AVL-based data used in this empirical study. The arc speed is derived from the differences in the cumulative distance and time elapsed between two consecutive AVL signals:

$$ v_{arc} = \frac{d_{i+1} - d_i}{t_{i+1} - t_i} $$

(3.1)

where:

- $d_i$ = the cumulative distance of the $i^{th}$ signal, and
- $t_i$ = time of the $i^{th}$ signal.

The midpoint $\frac{d_{i+1} + d_i}{2}$ is considered to be the associated location (cumulative distance along the route from the reference point), and the average time $\frac{t_{i+1} + t_i}{2}$ is considered to be the associated time for the corresponding arc speed. An arc speed calculated from consecutive AVL signals straddling a stop — i.e., with one signal upstream and the other signal downstream of the stop — would be affected by the dwell time at the bus stop. To
avoid the impact of dwell times on detecting recurrent congestion, the pair of AVL
signals straddling any stop is not used to calculate arc speeds in this study. As a result,
there are no speed observations for the space-time cells near bus stops, and therefore no
indication of congestion around bus stops. The signal upstream of the stop is used with
the signal previous to it to calculate an arc speed upstream of the stop, and the signal
downstream of the stop is used with the signal following it to calculate an arc speed
downstream of the stop.

3.2 Parameter Values and General Approach to Applying Methods

The methods described in chapter 2 are applied to find indications of congestion along
the CLS bus route. Table 3.2 lists the parameters used in this study.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Notation</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step size in space</td>
<td>$\Delta d$</td>
<td>10 meters</td>
</tr>
<tr>
<td>Smoothing distance</td>
<td>$d_{\text{smooth}}$</td>
<td>100 meters</td>
</tr>
<tr>
<td>Starting location</td>
<td>$d_s$</td>
<td>0 meters</td>
</tr>
<tr>
<td>Ending location</td>
<td>$d_e$</td>
<td>8334 meters</td>
</tr>
<tr>
<td>Step size in time</td>
<td>$\Delta t$</td>
<td>30 minutes</td>
</tr>
<tr>
<td>Homogeneous time interval</td>
<td>$t_{\text{hom}}$</td>
<td>30 minutes</td>
</tr>
<tr>
<td>Start time</td>
<td>$t_s$</td>
<td>7:00</td>
</tr>
<tr>
<td>End time</td>
<td>$t_e$</td>
<td>18:30</td>
</tr>
<tr>
<td>Significance Level</td>
<td>$\alpha$</td>
<td>0.05</td>
</tr>
<tr>
<td>$C$-space rule parameter</td>
<td>$C$</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 3.2 Parameters and Parameter Values Used in the Empirical Study
The step size from the beginning of one spatial section to the beginning of the next spatial section (step size in space) $\Delta d$ is chosen to be 10 meters. A 10-meter value is chosen partially because, conceptually, congestion on urban streets typically spans a length more than 10 meters over space. Thus, if there is congestion somewhere along the route, it is hoped that a step size of 10 meters over space would be short enough to capture it. On the other hand 10 meters is expected to be long enough to avoid too many consecutive spatial sections indicating the same congestion effect. As with other parameters, investigating the sensitivity of the methods to the value of $\Delta d$ is beyond the scope of this study, but such an investigation would be a valuable issue for future work.

The length of the spatial window (smoothing distance) $d_{\text{smooth}}$ is set at 100 meters, regardless of the location along the route. The 100-meter smoothing distance is chosen as a trade-off between being long enough to allow aggregation of enough empirical data to allow statistically meaningful tests in the three methods considered and being short enough to avoid covering areas with widely varied congestion patterns. In addition, as mentioned above, the empirical AVL signals tend to be collected every 100 meters. This implies that an arc speed (whose midpoint is) located within in the space region $[d-50, d+50]$ is likely to capture the effects of bus operation in the considered spatial window.

As a result of the 100-meter $d_{\text{smooth}}$, the starting location for a non-looping route would be $0+50 = 50$ meters and the ending location would be set to the length of the route (Route Length) minus 50 meters so that each spatial window would have speed observations over 100 meters. However, since CLS is a loop route, the speeds observed within [Route
Length -50, Route Length], in addition to those observed within [0, 50], can be considered as speeds for the spatial window with midpoint 0. That is, even for location \(d=0\), speeds can be considered 50 meters upstream within [Route Length -50, Route Length] and 50 meters downstream within [0, 50]. The same logic applies to location \(d=\) Route Length. Therefore, in this study, the starting location is set to \(d_s = 0\), and the ending location \(d_e\) is set to \(d_e = \) Route Length rounding down to nearest 10 meters (\(\Delta d\)).

The starting time \(t_s\) and ending time \(t_e\) are set to be 7:00 and 18:30, respectively, because the scheduled bus service frequency was the same between these times. The length of the time interval (homogeneous time duration) \(t_{hom}\) is set at 30 minutes. This 30-minute \(t_{hom}\) value was chosen as a compromise between being long enough to include a sufficient number of data points to produce statistically meaningful tests and short enough to capture the effect of temporal homogeneity. It is noted that speeds at a fixed location \(d\) are grouped into mutually exclusive 30-minute time intervals. That is, the step size \(\Delta t\) is the same as the homogeneous time duration \(t_{hom}\). The exclusivity of the intervals implies that the independence assumption, required by the \(F\)-test and the Scheffé's's procedure in the Mean Based method and the \(K-W\) rank-sum test and the Wilcoxon Rank-sum test in the Median Based method, would not be violated due to common data points. (Although multiple speed observations from a certain bus trajectory may cause the violation of the above independence assumption when they occur in the same 100 meter space window, it is believed that ignoring this violation is not likely to have any marked effect on the results of the statistical tests because of the following two reasons. Firstly, due to the 100 meter sampling rate over space, it is not very likely to obtain two speeds on the same
trajectory within the same space window. Secondly, again due to the 100 meter sampling rate over space, two consecutive speed observations should be distant enough from each other such that the correlation between the two speed observations is very small if there is any.) The significance level $\alpha$ of all the statistical tests applied in this study is set to be 0.05.

Finally, the parameter $C$ of the $C$-space rule is set to 3. Recall that $C$ specifies the minimum number of consecutive spatial sections required to have an indication of congestion during the same time interval in order to be classified as having a strong indication of congestion.

As a result, the procedure of finding congestion indications following any of the three methods can be presented in general as following:

1. Initiate $d = d_s$,
2. Select speeds from within $[d-50, d+50]$,
3. Apply method for each 30-minute time interval from the 7:00~18:30 interval with significance level $\alpha = 0.05$,
4. Set $d = d + 10$,
5. Repeat steps 2, 3, and 4 until $d = \text{Route Length} = 8330$ (8334 rounded down to nearest 10m),
6. Output binary indication of congestion matrix, and
7. Apply 3-space rule to produce strong indication of congestion matrix.
In the case of the CLS route, the procedure considers 833 spatial sections, with 23 non-overlapping (independent) time intervals for each section leading to 19,159 space-time cells. However, since arc speeds are not calculated for two consecutive signals straddling a bus stop, many space-time cells have no observation. Over the 10 week period being studied, there are a total of 99,072 empirical arc speeds. On average, there are 51.7 data points in each space-time cell. (Recall that the spatial sections overlap so that data points are used in more than one spatial section.)

3.3 Patterns from Three Methods

As described in chapter 2, each method yields a 2-dimensional binary strong indication of congestion matrix. Figure 3.2 illustrates the binary strong indication of congestion matrices produced by each of the three methods. As listed in Table 3.2, a significance level (\( \alpha \)) of 0.05 and a minimum number (C) of consecutive congestion indications across spatial sections of 3 are used in producing the strong indication of congestion matrices. A space-time cell in the plot is coded as black if there is a strong indication of congestion during the corresponding time interval at the corresponding location. Otherwise, a space-time cell is coded as white.

It can be noticed from Figure 3.2 that the Mean Based method produces much fewer congestion indications (71 space-time cells) compared to the Median Based method (1474 space-time cells) and the LST method (812 space-time cells). Additional comparisons at the aggregate level are presented in Appendix A. It is found that the Mean
Based method misses many of the expected congested cells that the LST method picks up. Therefore, in the following sections, only the Median Based method and the LST method are compared. In subsections 3.3.1 to 3.3.3, similarities in patterns resulting from the application of the Median Based methods and the LST method are illustrated and discussed.
Figure 3.2 Strong Indication of Congestion Matrices of Three Methods ($\alpha = 0.05$, $C = 3$)
3.3.1 General Morning and Afternoon Peaks

It is noticed from Figure 3.2 that over locations from $d = 0$ to $d = 4500$ and from $d = 7000$ to $d = 8330$, both the Median Based method and the LST method are more likely to produce indications of congestion during the morning peak between 8:00 and 10:00. Similarly, over locations from $d = 4500$ to $d = 7000$, both methods are more likely to produce indications of congestion during the afternoon peak between 15:00 and 18:00. These general patterns are consistent with our knowledge of the system segments. Between 0 to 4500 meters and between 7000 to 8330 meters, there is much through traffic not originated from or destined to campus. This through traffic would be particularly high during morning peak. On the other hand, the roadway along the route between 4500 to 7000 meters consists primarily of smaller roadways receiving local, campus traffic. On this portion of the route, congestion is likely to be experienced during the afternoon peak when there is a concentration of high vehicular traffic leaving campus and high pedestrian traffic at crossings interacting with vehicle flow at these locations. (Although the vehicular traffic would also peak in the morning on these sections, the volume of pedestrian traffic during the morning peak is much lower than that during the evening peak.)

3.3.2 Congestion Dissipation over Time

One can also notice from Figure 3.2 that the indications of morning congestion revealed by both the Median Based method and the LST method dissipate over time. This pattern is more clearly visible on a route map. Figure 3.3 through Figure 3.5 illustrate the
dissipation of morning peak congestion as revealed by the LST method during three 30
minute time intervals: 8:00–8:30, 9:00–9:30, and 10:00–10:30. Similar figures for the
Median Based method, which can be found in Appendix B, illustrate a similar pattern.

Figure 3.3 Congested Locations (black dots) Indicated by LST Method during 8:00–8:30
Figure 3.4 Congested Locations (black dots) Indicated by LST Method during 9:00~9:30
3.3.3 Cell by Cell Comparison

To investigate the similarities between the Median Based method and the LST method further, pair-wise comparisons based on the binary valued cells of the strong indication matrices (after applying the 3-space rule) produced by the two methods are conducted. When comparing the two matrices, each space-time cell belongs to one of the following four types of combinations: both methods indicate congestion, only the Median Based method indicates congestion, only the LST method indicates congestion, neither method indicates congestion. Table 3.3 shows the percentage of space-time cells in each of the
four categories. The large value obtained from the summation of the percentages of the upper-left cell and the lower-right cell in Table 3.3 indicates strong overall similarity between the two methods. In addition, the conditional probability of congestion indications produced by the LST method, given that there are indications of congestion from the Median Based method, is 0.2974 (0.0229 divided by 0.077). Similarly, the conditional probability of congestion indications produced by the LST method given that there is no indication of congestion from the Median Based method is 0.0211 (0.0195 divided by 0.9230). It is much more likely (probability 0.2974 compared to probability 0.0217) that the LST method indicates congestion when there is an indication of congestion from the Median Based method at the corresponding space-time cell than when there is no indication of congestion from the Median Based method at the corresponding space-time cell.

The same conclusion is drawn when investigating the probability of Median Based method producing indications of congestion conditional on the indications of congestion based on the LST method. The conditional probability of congestion indications produced by the Median Based method, given that there are indications of congestion from the LST method, is 0.5401 (0.0229 divided by 0.0424), whereas, the conditional probability of congestion indications produced by the Median Based method given that there is no indication of congestion from the LST method is 0.0565 (0.0541 divided by 0.9576). It is much more likely (probability 0.5401 compared to probability 0.0565) that the Median Based method indicates congestion when there is indication of congestion from the LST method.
method at the corresponding space-time cell than when there is no indication of congestion from the LST method at the corresponding space-time cell.

<table>
<thead>
<tr>
<th>Proportion of congestion indication combinations</th>
<th>Median Based method indicates Congestion</th>
<th>Marginal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>LST method indicates Congestion</td>
<td>2.29%</td>
<td>1.95%</td>
</tr>
<tr>
<td></td>
<td>5.41%</td>
<td>90.35%</td>
</tr>
<tr>
<td>Marginal</td>
<td>7.70%</td>
<td>92.30%</td>
</tr>
</tbody>
</table>

Table 3.3 Cell Level Comparison of the Median Based Method and the LST Method

### 3.4 Similarities and Differences between the Median Based Method and the LST Method

Further comparison between the Median Based method and the LST method is conducted. As stated in section 3.3.3, there are four combinations in terms of which method indicates congestion: both methods indicate congestion, only the Median Based method indicates congestion, only the LST method indicates congestion, neither method indicates congestion. Figure 3.6 shows where and when the four combinations occur. In Figure 3.6, a space-time cell is color-coded as: black if both methods have strong indication of congestion; blue if the Median Based method indicates congestion but not the LST
method; red if the LST method indicates congestion but not the Median Based method; and white if neither method detects congestion. As a supplement to Figure 3.6, Table 3.4 tabulates the number of space-time cells for each combination and calculates their marginal counts.

Figure 3.6 Comparison of Median Based Method and LST Method

<table>
<thead>
<tr>
<th>Color</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>Both methods indicate congestion;</td>
</tr>
<tr>
<td>Blue</td>
<td>Only the Median Based method indicates congestion;</td>
</tr>
<tr>
<td>Red</td>
<td>Only the LST Method indicates congestion;</td>
</tr>
<tr>
<td>White</td>
<td>Neither method indicates congestion.</td>
</tr>
</tbody>
</table>
### Table 3.4 Number of Cells Indicated/Not Indicated as Congested by Median Based Method and LST Method

<table>
<thead>
<tr>
<th></th>
<th>Median Based Indicates</th>
<th>Mean Based Does Not Indicate</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>LST Indicates</td>
<td>438</td>
<td>374</td>
<td>812</td>
</tr>
<tr>
<td>LST Does Not Indicate</td>
<td>1036</td>
<td>17311</td>
<td>18347</td>
</tr>
<tr>
<td>Total</td>
<td>1474</td>
<td>17685</td>
<td>19159</td>
</tr>
</tbody>
</table>

Figure 3.7 shows where and when either the Median Based method or the LST method indicates congestion with $C = 3$ for the $C$-space rule. As can be noticed from Figure 3.7, some time intervals show more indications of congestion along the route compared to other time intervals. This pattern is clearer in Figure 3.8, which shows the number of congested spatial sections detected by either method versus time. Four time intervals, 7:30~8:00, 8:00~8:30, 15:00~15:30, and 17:00~17:30, with relatively high number of congested spatial sections, are selected for detailed analysis in this section. The locations with any indications of congestion indicated by either method during these four time periods are analyzed systematically except locations around bus stops or within the west campus parking lot where bus behaviors are heavily influenced by bus operation system requirements rather than by surrounding traffic.
Figure 3.7 Either Median Based Method or LST Method Indication of Congestion (Black)

Figure 3.8 Number of Congested Spatial Sections Indicated by Either Method
The detailed comparisons consist of examining each “location” with indication of congestion produced by either the Median Based method or the LST method during the four time-intervals. The word “location” is placed in quotes because it is not defined as a 10 meter spatial section here. In general, several consecutive spatial sections having indications of congestion are combined to form one “location” if doing so makes sense given the nature of the roadway over these sections. A detailed description of how such “locations” are defined is provided in Appendix C. For each “location”, only the 10 meter spatial section in the middle of the “location” is chosen for detailed analysis. For each spatial section, there could be more than one space-time cell under investigation, since there could be multiple time intervals having indications of congestion for the same spatial section.

In subsections 3.4.1 through 3.4.3, more detailed comparisons between the Median Based method and the LST method for some space-time cells are presented and discussed to illustrate the methodology and highlight the comparative results. The “locations” and the corresponding traffic directions of the six examples illustrated in sections 3.4.1~ 3.4.3 are shown on a CLS route map in Figure 3.9.
3.4.1 Examples of Space-Time Cells Where and When Both Median Based and LST Methods Indicate Congestion

The location of the first example, where and when both the Median Based and LST method indicate congestion, is shown in Figure 3.10 with label “ex. 1”. This is an eastbound location upstream of the signalized intersection of Woody Hayes Drive and Fyffe Road. Congestion is expected at this location during the morning peak hours because a high concentration of traffic demand both on Woody Hayes Drive and on Fyffe Road coming to campus during the morning peak produces queues at the signalized intersection.
The empirical results from both the Median Based method and the LST method have indications of congestion during the two time intervals 7:30~8:00 and 8:00~8:30 in the morning peak hour. Figure 3.11 is a plot of the observed speeds at this location over time. The figure indicates that the whole day speed distribution at this location exhibits bimodality. This bimodality is likely a result of the signal control at the signalized intersection. A bus encountering a green signal can proceed through this location at a relatively higher speed (contributing to the distribution with the higher mode) than a bus encountering a red signal (contributing to the distribution with the lower mode). Thus traffic signals at intersections cause a gap in the magnitude of the speeds observed during a green phase and the speeds observed during a red phase. Figure 3.11 indicates a higher proportion of speeds in the distribution with the lower mode during 7:30~8:00 and 8:00~8:30 with respect to the proportion of speeds of similar magnitudes observed during the whole day. There are two possible transportation explanations for this pattern:

(1) More traffic is held up by the red signal due to the queue formed upstream of the signalized intersection, or

(2) An increased demand from the other approach, Fyffe Road, causes longer red signal time for the Woody Hayes Drive approach.

In either case, the pattern is a result of increased traffic during the indicated time period.
Figure 3.10 Location of Example 1: Both Methods Have Congestion Indication

Figure 3.11 Scatter Plot of Speed against Time of Day at Location Example 1
The location of the second example, where and when both the Median Based and LST methods indicate congestion, is shown in Figure 3.12 with label “ex. 2”. This is a westbound location upstream of a signalized pedestrian crossing on West Woodruff Avenue. The red signal on West Woodruff Avenue is triggered when a pedestrian presses the buttons on either side of the road. Congestion is expected at this location during afternoon peak hours because of the interaction between the high vehicular volume leaving main campus during the afternoon peak hours and the high pedestrian demand at this location during the afternoon peak hours.

The interaction between the vehicular traffic and the pedestrian traffic results from the following situation. When there is a higher concentration of vehicular demand, pedestrians are more likely to press the button requesting the red signal for vehicles because they would wait much longer if they do not do so. On the other hand, when the vehicular traffic demand is low, pedestrians usually choose to walk across West Woodruff Avenue (about 10 meters wide) in between vehicle gaps instead of pushing the button to trigger the red signal.

The empirical results from both the Median Based method and the LST method have indications of congestion at this location during the 17:30~18:00 interval, which is in the afternoon peak hour. Figure 3.13 is a plot of the observed speeds at this location over time. The figure indicates that the whole day speed distribution at this location has some low speeds. This lower tail is likely formed by buses encountering red signals that are requested by the pedestrians. Figure 3.13 indicates that a higher proportion of speeds
around the lower tail of the speed distribution is observed during 17:30~18:00 with respect to the proportion of speeds of similar magnitudes observed during the whole day. This pattern is consistent with the knowledge that higher pedestrian and vehicle demand is encountered during this time interval at this location.

Figure 3.12 Location of Example 2: Both Methods Have Congestion Indication
3.4.2 Examples of Space-Time Cells Where and When Only Median Based Method Indicates Congestion

The location of the third example, where and when only the Median Based method indicates congestion, is shown in Figure 3.14 with label “ex. 3”. This is a southbound location upstream of bus stop 8 on Canon Drive. Although congestion is indicated by the Median Based method, congestion is not expected because there is very little through traffic in the south bound direction during 7:30–8:00 on this roadway segment of Canon Drive, and there is very little traffic with destinations downstream of this location.
Figure 3.14 Location of Example 3: Only Median Based Method Has Congestion Indication

Figure 3.15 is a plot of the observed speeds at this location over time. The figure indicates that the whole day speed distribution has no clear bimodality or low speeds as seen in the first two examples. The speeds across the whole day are roughly on the same level of magnitude. Further, Figure 3.15 indicates that there is a shift in the median of the speed distribution during 7:30~8:00 compared to the median of the speed distribution during the whole day. However, it does not seem that the entire speed distribution during this time interval has such a shift because the lowest speeds observed during this time interval are relatively high compared to the lowest speeds observed during other time intervals. Analysis of many speed distributions, the details of which are beyond the scope of this study, indicates that this is a typical speed distribution formed at uninterrupted
roadway segments. Possible transportation explanations for this pattern (a slight shift in the median but not a shift in the entire distribution) are that:

(1) There is slightly more traffic during this time period than during other times. Therefore, there are few very high speeds observed during this time interval. This results in the appearance of a decrease in the median. However, the traffic demand during this time interval is still not high enough to cause an increase in the number of low speeds observed.
(2) The location is immediately upstream of bus stop 8. Bus stop 8 is a holding stop, and bus drivers may purposely slow down when approaching bus stop 8 to reduce the dwell time at the bus stop. (Continuing to hold at a bus stop after passengers have boarded can lead to greater passenger dissatisfaction.) Of course, this explanation is reasonable only when buses arrive earlier than the scheduled bus arriving time. During the time interval 7:30~8:00, there is not much traffic along the CLS route. Thus, it is likely that buses arrive at bus stop 8 earlier than the schedule.

(3) Another possible explanation is that buses run faster than usual at this uninterrupted roadway segment during other times of the day. Especially when there is congestion at other locations along the route, buses might use this uninterrupted roadway to catch up with the bus schedule. This explanation can be supported by Figure 3.15: the speeds observed during 8:00~9:00, which is part of the morning peak hours, and during 16:00~18:00, which is part of the afternoon peak hours, are relatively higher than speeds observed during other time intervals. Congestion is expected and indicated (see Figure 3.7) elsewhere in the roadway network during the 8:00~9:00 and 16:00~18:00 time intervals. As a result, when comparing the speeds during 7:30~8:00 to the speeds observed during other time intervals, which includes the fast “catch up speed”, a relative decrease in the median speed during the 7:30~8:00 interval is detected by the Median Based Method. In other words, this “relative decrease” in speed at the indicated time interval may not be caused by more traffic at this location but by the faster speeds
at other time intervals that result from congestion elsewhere in the roadway network.

The knowledge of the traffic condition at this location and the explanations given above point to this indication of congestion produced by the Median Based method as a false indication of congestion.

The location of the fourth example, another example of where and when only the Median Based method indicates congestion, is shown in Figure 3.16 with label “ex. 4”. This northbound location is at the middle of the curved roadway segment on Medical Center Drive. Congestion is indicated by the Median Based method, during the 7:30~8:00 time interval, but not by the LST method. Congestion is not expected at this location during the morning peak hours because there is very little through traffic in the northbound direction on this roadway segment and very little traffic with destinations downstream of this location in the northbound direction.

Figure 3.17 is a plot of the observed speeds at this location over time. Figure 3.17 indicates that, similar to the previous example, there is a shift in the median of the speed distribution during 7:30~8:00 compared to the median of the speed distribution during the whole day. However, it seems that the entire speed distribution during this time interval does not have such a shift because the lowest speeds observed during this time interval are relatively high compared to the lowest speeds observed during other time intervals.
The transportation explanations for this pattern (a slight shift in the mode but not a shift in the entire distribution) are that:

(1) Like the first transportation explanation for the previous example (the third example), there is slightly more traffic during this time period than during other time. Therefore, there are few high speeds observed during this time interval. However, the traffic demand during this time period is still not high enough to cause low speeds.

(2) Like the third transportation explanation for the previous example (the third example), another possible explanation is that buses run faster than usual at this uninterrupted roadway segment during other times of the day. Especially when there is congestion at other locations along the route, buses might use this uninterrupted roadway to catch up with the bus schedule. As with the previous example, this explanation can be supported by Figure 3.17: the speeds observed during 8:00~9:00, which is part of the morning peak hours, and during 16:00~18:00, which is part of the afternoon peak hours, are relatively higher than speeds observed during other time intervals. Congestion is expected and indicated (see Figure 3.7) elsewhere in the roadway network during the 8:00~9:00 and 16:00~18:00 time intervals. As a result, when comparing the speeds during the 7:30~8:00 interval to the speeds observed during other time intervals, which includes the fast “catch up speed”, a relative decrease in the median speed during the 7:30~8:00 interval is detected by the Median Based Method. As in the previous example, this “relative decrease” in speed at the indicated time interval may not be
caused by more traffic at this location but by the faster speeds at other time intervals that result from congestion elsewhere in the roadway network.

The knowledge of the traffic condition at this location and the explanations given above point to this indication of congestion produced by the Median Based method as a false indication of congestion.

Figure 3.16 Location of Example4: only Median Based Method Has Congestion Indication
3.4.3 Examples of Space-Time Cells Where and When Only LST Method Indicates Congestion

The location of the fifth example, where and when only the LST method indicates congestion, is shown in Figure 3.18 with label “ex. 5”. This is an eastbound location upstream of the intersection of Carmack Road and Kenny Road. Congestion is indicated by the LST method during the 17:00~17:30 time interval. Congestion is expected at this location during the afternoon peak hours because there is high vehicular traffic demand leaving the west campus parking lot during this time. The Kenny Road and Carmack
Road intersection downstream of this location is a major exit for this departing traffic. Therefore, it is expected that queues form at the traffic signal and extend far enough upstream to reach this location and cause congestion during the afternoon peak hour.

Figure 3.19 is a plot of the observed speeds at this location over time. The figure indicates that the speed distributions for different time intervals are quite different. Some time intervals have some low speeds while some time intervals do not. Comparing to the speed distribution in the first example of subsection 3.4.1, the distribution of these low speeds, in terms of their magnitude and the gap between them and the higher speeds, looks similar to the distribution of the low speeds in the first example in subsection 3.4.1. Possible transportation explanation for such a speed distribution is that this location usually does not have arc speeds caught in queues unless there is a heavy congestion at the downstream intersection. As a result, when there is heavy congestion at the downstream intersection, very slow arc speeds, most likely obtained from buses encountering a red signal, are observed. It can also be noticed from Figure 3.19 that there are more low speeds observed during this time period than during the whole day. Again, the transportation explanation for this phenomenon (i.e. concentration of low speeds during this time interval) is that the traffic queue formed from the downstream intersection reaches this location at the peak hours due to the higher traffic demand than the demand during other times. However, these low speeds observations during this time interval at the location are not sufficient to result in rejecting the null hypothesis of the Wilcoxon Rank-Sum test in the Median Based method due to lack of data.
Figure 3.18 Location of Example 5: Only LST Method Has Congestion Indication

Figure 3.19 Scatter Plot of Speed against Time of Day at Location Example 5
The location of the sixth example, another example where and when only the LST method indicates congestion, is shown in Figure 3.20 with label “ex. 6”. This is a westbound location upstream of the intersection of West Woodruff Avenue and Cannon Drive. Congestion is indicated by the LST method during the 16:30~17:00 time interval. Congestion is expected at this location during the afternoon peak hours because there is very high vehicular traffic demand leaving main campus during this time. A large proportion of the traffic leaving campus uses both approaches at the downstream intersection of West Woodruff Avenue and Cannon Drive.

Figure 3.21 is a plot of the observed speeds at this location over time. The whole day speed distribution shows bimodality, like the whole day distribution at location “ex. 1” in Figure 3.11. However, it appears that the bimodality is not as clear for all time intervals as it was in Figure 3.11. Rather, the variation in the bimodality of the speed distributions for each time interval has some similarity with the variation shown in Figure 3.19 (the previous, fifth example). Possible transportation explanations for such speed distribution are:

(1) This location is closer to its downstream intersection than the location “ex. 5” but farther from its downstream intersection than the location “ex. 1”. Therefore, the speed distribution is a combination of the speed distributions at location “ex. 5” and at location “ex. 1”.

(2) The traffic signal of the downstream intersection is triggered by vehicles occupying the loop detectors installed at the stop lines of the intersection. Therefore, a bus
approaching this intersection is more likely to encounter a red signal when there is more vehicular traffic on the other approach.

The above paragraph explains possible reasons for the variations in the speed distributions for each time interval. As for the 16:30~17:00 time interval, Figure 3.21 indicates that there are a higher proportion of low speeds observed during this time interval than during the whole day. Again, the transportation explanations for this phenomenon are that:

(1) The traffic queues formed from the downstream intersection are more likely to reach this location during this time interval due to the high traffic demand on this roadway segment.

(2) A bus running during this time interval is more likely to encounter a red signal due to the high traffic demand on the other approach.

Figure 3.20 Location of Example6: only LST Method Has Congestion Indication
3.4.4 Summary of Examples and Extensions

The first example illustrates that both the Median Based method and the LST method can produce indications of congestion when the speed distribution is bimodal for each time intervals. However, as example “ex.6” illustrates, there are cases where the LST method produces indications of congestion when the speed distribution is bimodal for each time interval while the Median Based method does not. The LST method is more sensitive to low speed observations than the Median Based method. From a transportation point of view, the LST method appears better able to find indications of congestion at signalized intersections than the Median Based method.
The second example illustrates that both the Median Based method and the LST method produce indications of congestion when the speed distribution is bimodal for some, but not all time intervals. However, as example “ex. 5” illustrates, there are cases where the LST method produces indications of congestion for such a speed distribution while the Median Based method does not. The LST method is more sensitive to low speed observations than the Median Based method. From a transportation point of view, the LST method appears better able to find indications of congestion at pedestrian crossings, in addition to signalized intersections.

The third and fourth examples illustrate that the Median Based method may produce false indications of congestion that might be explainable by bus driver operations. The LST method appears to be less prone to this deficiency. Based on the empirical studies conducted so far, there are quite a few location and time intervals where and when the Median Based method produces indications of congestion while the LST method does not. Out of these location and time intervals, some cases exhibits only a slight decrease in the median speed during the time interval indicated. It appears that such cases are likely to be false indications of congestion, which can possibly be explained by bus driver operations.
Chapter 4 Summary and Discussion

4.1 Summary

This study focuses on using buses as probes to indicate recurrent congestion on urban roadway networks. Since buses behave differently from the surrounding traffic, the approach compares bus speeds to other bus speeds. Specifically, for a given location bus speeds in a specified period are compared to bus speeds in other periods in order to identify periods of low bus speeds at that location. By comparing bus speeds to other bus speeds, there is no need to compare bus speeds to speeds of other traffic when determining indications of congestion. The low bus speeds could be an indicator of the effects of surrounding traffic.

In this study, bus speed data collected over “homogeneous days” are pooled in order to find indications of “recurrent” congestion. For practical purposes, the entire route is discretized into 10 meter overlapping spatial sections, and time of day is discretized into 30 minute time intervals. Each 10 meter long spatial section and 30 minute duration time interval constitutes a space-time cell for which existence or absence of indication of congestion is determined. More specifically, in this study, bus speeds collected at a fixed location (a spatial section) during a 30-minute time interval are compared to the bus speeds collected at that location during other 30-minute time intervals or during the whole day.
Three different methods are used to find indications of congestion by using this general approach. Two of the three methods, the Mean Based method and the Median Based method, apply traditional statistical tests. Therefore, they are also referred to as traditional methods in the following discussion. The other method, the Low Speed Threshold (LST) method is developed in this study. As far as the author is aware, the two traditional methods have not been applied to detect recurrent congestion using bus probes. As for the LST method, the procedure consists of an investigation and application of statistical ideas customized specifically for the purposes of this study.

The basic data structure required for applying the three methods are location (in terms of distance from a reference point), time of day, and speed. Speeds can be calculated in several ways. In this study, arc speeds are used, where the arc speed is determined as the distance traveled between two consecutive data records divided by the time between these records.

An empirical study based on AVL data collected using The Ohio State University (OSU) Campus Transit Lab (CTL) is conducted to evaluate the ability of the three methods to indicate congestion. It is found that the Mean Based method misses many of the expected congested cells that the other two methods pick up. On the other hand, both the Median Based method and the LST method show fairly similar general patterns in terms of their indications of congestion.
Both methods have major indications of congestion in the core campus during the afternoon period between 15:00 and 18:00 and in the west campus area (which is primarily devoted to remote parking) during the morning period between 8:00 and 10:00. This congestion pattern is consistent with the author’s knowledge of the traffic conditions on the OSU campus.

Both methods show dissipation of the morning period congestion as time progresses through the morning period beginning at 8:00. This dissipation is also consistent with the knowledge of the traffic conditions on the OSU campus.

Despite the above similarities, the Median Based method and the LST method also show some differences in terms of where and when strong indications of congestion are found. The space-time cells where the Median Based method gives different results than the LST method in terms of strong indications of congestion are analyzed in detail.

Based on the detailed analysis, it is hypothesized that many cases where and when the Median Based method produces a strong indication of congestion while the LST method does not are caused by bus driver behavior. Such bus driver behavior could be a result of bus drivers trying to avoid going ahead of schedule in the indicated periods (because there are less congested conditions along the route at other locations during that time) or to catch up to the schedule in the complementary periods (because there are more congested conditions along the route at other locations during that time). If bus driver behavior is the cause of these patterns in speed distribution, these strong indications of congestion are likely to be false indications of congestion. The LST method is less prone
to producing a strong indication of congestion in such cases because the type of bus
driver behavior hypothesized would be less likely to affect low speeds.

On the other hand, most of the cases where and when recurrent congestions are indicated
by the LST method but not by the Median Based method are where and when congested
conditions are expected. In such cases, the distribution of speeds where and when the
LST method has a strong indication of congestion while the Median Based method does
not are more likely to be bimodal or have some low speeds. Bimodal distributions are
likely to be formed around signalized intersections. A bus encountering a green signal
would yield a much higher arc speed at such locations than a bus encountering a red
signal. As a result, there is a clear gap between the speed distribution from buses
encountering green signals and the speed distribution from buses encountering red signals.
As a result of this clear gap in the modes, a bimodal mixed distribution is formed.

In summary, based on the detailed comparisons conducted thus far, the LST method is
preferred to be used to produce indications of recurrent congestion. However, more
research is necessary before concluding that the LST method is to be preferred to the
Median Based method.

4.2 Future Research

The Mean Based method and the Median Based method are applied in this study without
testing the validity of their assumptions. Normality of speed distributions is assumed
when justifying the *F*-test and the Scheffé's contrast test in the Mean Based method. Therefore, the normality test should be performed for each location before the Mean Based method is applied. If the normality assumption is rejected by the normality test, then the Mean Based method cannot be applied as a rigorously defensible statistical method. However, given the low rate at which the Mean Based method produced strong indications of congestion, testing for normality would not be expected to change the results much. Similarly, the same shape assumption on the shapes of speed distributions is assumed when justifying the *K-W* Rank Sum test and the Wilcoxon Rank-Sum test in the Median Based method. Therefore, the same shape test should be performed for each location before the Median Based method is applied. If the same shape test is rejected, then the Median Based method cannot be applied as a rigorously defensible statistical method. Unlike the Mean Based method, the Median Based method produced many strong indications of congestion. Testing for the same shape of the distribution is likely to eliminate strong indications of congestion.

Further investigation could be devoted to fine-tuning the parameters of the Median Based and LST methods. The general parameters used in the empirical studies are listed in Table 3.2. In addition, two other parameters for determining the low speed threshold \( V_{\text{threshold}} \) (\( p_{\text{low}} \) and \( q_{\text{freeflow}} \)) are also specified in this study. Of the list of parameters, the following are likely to be most worthwhile for further investigation.

The smoothing distance \( d_{\text{smooth}} \) is the distance span over which speeds are grouped for a fixed location. Intuitively, given a sufficient number of observations, a smaller value of
\(d_{\text{smooth}}\) should be used for finding temporal patterns of speeds at a specific location. A larger value of \(d_{\text{smooth}}\) might dilute the temporal pattern of speeds at a fixed location by including data from nonhomogenous spatial sections. Therefore, given enough data, reducing \(d_{\text{smooth}}\) may increase the likelihood of finding indications of recurrent congestion at a fixed location if there are congested conditions. On the other hand, excessively reducing \(d_{\text{smooth}}\) may cause the number of observations of the speeds during a certain time interval at a certain location to be too small to allow rejection of a null hypothesis. Therefore, it would be worthwhile to investigate the effect of \(d_{\text{smooth}}\) on the results of the statistical methods used.

The homogenous time interval length \(t_{\text{hom}}\), over which speeds are grouped, is also an important parameter. Like \(d_{\text{smooth}}\), a value of \(t_{\text{hom}}\) that is too small could reduce the number of observations to the point that a meaningful statistical test cannot be conducted, while a value of \(t_{\text{hom}}\) that is too large could dilute an interesting speed pattern by including nonhomogenous time periods.

The effect of the significance level \(\alpha\) might also be interesting to study. Significance level \(\alpha\) directly affects the number of space-time cells with indication of congestion. A large value of \(\alpha\) is likely to yield false indications of recurrent congestion, while a small value of the \(\alpha\) might eliminate some location and time combinations where and when there is recurrent congestion. It would be worthwhile to investigate the tradeoff between these different types of errors as this parameter \(\alpha\) is varied.
Finally, the low speed threshold $V_{\text{threshold}}$ in this study is determined by a fixed proportion $p_{\text{low}}$ of a percentage $q_{\text{freelow}}$ of the whole day speed distribution at each location. It might be worthwhile to examine the effect of these parameters and the resulting $V_{\text{threshold}}$ on the results in the LST methods.

Investigation of the use of GPS speed (the speed an Automatic Vehicle Location (AVL) system reports when a data point is produced) obtained from the raw AVL data would potentially be useful. Due to the large variance of the GPS speed in the data collected from the system used in this study, using GPS speed does not seem to produce results that are as “good” as those produced using arc speeds. The variability in the GPS speed data makes identification of temporal variability more difficult to identify. However, a GPS speed theoretically captures the bus running status over a much shorter time period and, thus, shorter distance than an arc speed. As discussed when considering the effect of $d_{\text{smooth}}$ above, some speed patterns at a location might be diluted because of the 100 meter span used in this study to group arc speeds.

One could also conceive of other methods for finding indications of recurrent congestion. One might investigate the indications of recurrent congestion by comparing the bus speeds to the speed limit. Using the speed limit as a reference point would be expected to control for some of the heterogeneity in speeds along the route.

This study focuses on detecting recurrent congestion. However, it is also of value to detect congestion caused by temporary changes to the infrastructure (construction on the
road, for example) or real-time congestion. Temporary congestion detection would appear to be a relatively straightforward extension of recurrent congestion detection. It would be useful to investigate the amount of data required in order to detect temporary congestions with the methods applied in this study or by some improved methods. Such an investigation would lend insight on how soon temporary congested conditions can be detected.

Despite the need for future research, it appears that the LST method, or some improved method based on the LST method, would be capable of indicating recurrent congestion from bus AVL data. Such a method would be beneficial in taking advantage of existing data to monitor traffic condition that are otherwise unmonitored.
References


Appendix A: Strong Indication of Congestion Matrices

After applying the $C$-space rule (see section 2.5), a binary strong indication of congestion matrix is produced. As described in section 2.5, a space-time cell has a strong indication of congestion if there are more than $C$ consecutive spatial sections exhibiting indications of congestion during a given time interval. In this study, the parameter $C$ is set to 3. Table A.1 compares each pair of the three methods by showing the number of space-time cells where both methods either indicate congestion or do not indicate congestion (based on the binary strong indication of congestion matrices).

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>LST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>19159</td>
<td>17756</td>
<td>18392</td>
</tr>
<tr>
<td>Median</td>
<td>17756</td>
<td>19159</td>
<td>17749</td>
</tr>
<tr>
<td>LST</td>
<td>18392</td>
<td>17749</td>
<td>19159</td>
</tr>
</tbody>
</table>

Table A.1 Number of Cells Where Either Both Methods Indicate Congestion or Both Do Not Indicate Congestion ($C = 3$) for Each Pair of Methods

Table A.2 compares each pair of the three methods by showing only the number of space-time cells where both methods have strong indication of congestion. Notice the much stronger similarity in the results produced by the Median Based and LST methods with respect to the results produced by the Mean Based method.
Table A.2 Number of Cells with Strong Indication of Congestion by Both Methods for Each Pair of Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Median</th>
<th>LST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>71</td>
<td>71</td>
<td>58</td>
</tr>
<tr>
<td>Median</td>
<td>71</td>
<td>1474</td>
<td>438</td>
</tr>
<tr>
<td>LST</td>
<td>58</td>
<td>438</td>
<td>812</td>
</tr>
</tbody>
</table>

To further examine the similarity between any two methods, the number of expected congested cells for each pair (assuming independence between the two methods), is shown in Table A.3. The probability values in Table A.3 are determined based on the total number of space-time cells, which is 19.159. By comparing Table A.3 with Table A.2, all the numbers in Table A.2 are greater than their corresponding numbers in Table A.3. While the Median Based and LST methods exhibit stronger similarities, this comparison suggests that all three methods are positively correlated in terms of where and when congested condition is indicated.

Table A.3 Based on Independence Assumption, Number of Expected Congestion Cell under Each Pair of Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Median</th>
<th>LST</th>
<th># Congested</th>
<th>Prob.(Congested)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>5.5</td>
<td>3.0</td>
<td>71</td>
<td>1474</td>
<td>0.0769</td>
</tr>
<tr>
<td>Median</td>
<td>5.5</td>
<td>62.5</td>
<td>1474</td>
<td>812</td>
<td>0.0424</td>
</tr>
<tr>
<td>LST</td>
<td>3.0</td>
<td>62.5</td>
<td>812</td>
<td>0.0037</td>
<td></td>
</tr>
<tr>
<td># Congested</td>
<td>71</td>
<td>1474</td>
<td>812</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prob.(Congested)</td>
<td>0.0037</td>
<td>0.0769</td>
<td>0.0424</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B: Dissipation of Morning Peak Congestion

Based on Median Based Method

This section illustrates the dissipation of morning peak congestion indications produced by the Median Based method. The congestion indications for three time intervals 8:00–8:30, 9:00–9:30, and 10:00–10:30 are shown separately on Figures B.1 through B.3.

Figure B.1 Congested Locations (black dots) Indicated by Median Based Method during 8:00–8:30
Figure B.2 Congested Locations (black dots) Indicated by Median Based Method during 9:00–9:30
Figure B.3 Congested Locations (black dots) Indicated by Median Based Method during 10:00~10:30
Appendix C: Systematic Selection of Space-Time Cells for Detailed Analysis

The identification of the “locations” where the Median Based and LST methods are compared in detail in section 3.4 is presented in this appendix.

There are four time intervals of the most interest: 7:00~7:30, 8:00~8:30, 15:00~15:30, and 17:00~17:30. The intervals 8:00~8:30 and 17:00~17:30 are of interest given the morning peak hour and the afternoon peak hour they represent. The interval 7:30~8:00 is of interest because there are almost as many congestion indications during 7:30~8:00 as there are during 8:00~8:30 but the locations are quite different. It is worthwhile to investigate in detail what is causing this difference. The interval 15:00~15:30 is of interest because it is not the morning peak hour or the afternoon peak hour but still exhibits quite a few congestion indications. To systematically identify the spatial sections forming the “locations” of interest for detailed analysis in section 3.4, the following steps are applied.

Step 1: Combine two strong indication of congestion matrices (one from the Median Based method and the other from the LST method) by converting the two binary matrices into a matrix with four types of values for the space-time cells. The four types of congestion indications are
type 1: only the Median Based method indicates congestion;

type 2: only the LST method indicates congestion; and

type 3: both the Median Based method and the LST method indicate congestion;

type 4: neither of the methods indicate congestion.

The corresponding matrix is shown in Figure C.1

Black: Both methods indicate congestion;
Blue: Only the Median Based method indicates congestion;
Red: Only the LST Method indicates congestion; and
White: Neither method indicates congestion.

Figure C.1 Combined Strong Indication of Congestion Matrix (Median Based vs. LST)

Step 2: For each time interval out of the four selected time intervals, combine all the
consecutive spatial sections having the same type of congestion indication (out of the
above four types) and define it as a combined location. Label each combined location by
their time interval, indication type, midpoint distance from the route reference point, and
spatial span. This produces the “primitive” list of cell clusters with congestion indications
as shown in Table C.1.
<table>
<thead>
<tr>
<th>ID</th>
<th>Type</th>
<th>MidDistance</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>95</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>435</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>485</td>
<td>60</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>570</td>
<td>110</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>635</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>650</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>915</td>
<td>180</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1345</td>
<td>140</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1530</td>
<td>50</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>1625</td>
<td>20</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>1680</td>
<td>90</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1740</td>
<td>30</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>2030</td>
<td>10</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>2220</td>
<td>150</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>2775</td>
<td>280</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>3310</td>
<td>10</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>3350</td>
<td>70</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>3400</td>
<td>30</td>
</tr>
<tr>
<td>21</td>
<td>3</td>
<td>3875</td>
<td>80</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>3930</td>
<td>30</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>4190</td>
<td>10</td>
</tr>
<tr>
<td>24</td>
<td>3</td>
<td>4215</td>
<td>40</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>4250</td>
<td>30</td>
</tr>
<tr>
<td>26</td>
<td>2</td>
<td>7245</td>
<td>60</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>7280</td>
<td>10</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>7340</td>
<td>30</td>
</tr>
<tr>
<td>29</td>
<td>3</td>
<td>7410</td>
<td>110</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>7470</td>
<td>10</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
<td>7825</td>
<td>120</td>
</tr>
<tr>
<td>32</td>
<td>3</td>
<td>7915</td>
<td>60</td>
</tr>
<tr>
<td>33</td>
<td>1</td>
<td>7960</td>
<td>30</td>
</tr>
<tr>
<td>34</td>
<td>3</td>
<td>7980</td>
<td>10</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
<td>8000</td>
<td>30</td>
</tr>
<tr>
<td>36</td>
<td>3</td>
<td>8035</td>
<td>40</td>
</tr>
<tr>
<td>37</td>
<td>1</td>
<td>8060</td>
<td>10</td>
</tr>
<tr>
<td>38</td>
<td>1</td>
<td>8205</td>
<td>220</td>
</tr>
</tbody>
</table>

Table C.1 “Primitive” List of Cell Clusters with Congestion Indications
Step 3: Select the “primitive” cell clusters with a spatial span greater than or equal to 30 meters for further consideration. Also, the locations where the cumulative distances are smaller than 800 meters (corresponding to the portion of the route in the west campus parking lot area) are eliminated from further analysis. As a result, the “long span” cell clusters with strong congestion indications are produced and are shown in Table C.2.
Step 4: Since there are only four time intervals being investigated, it is not too complicated to further combine the “long span” clusters over time if certain criteria are met resulting in the “locations” of interest for detailed analysis in section 3.4. As shown in Table C.2, a location where congestion is indicated during two or more time intervals is listed as two or more locations with partial overlap. Such locations, however, are combined as one “location” denoted by a common distance, referred to as modified distance, of the spatial spans with indications of congestion during the two or more time intervals. (In doing so, congested locations from Table C.1 are also considered). The following examples illustrate this combination process.

<table>
<thead>
<tr>
<th>Type</th>
<th>Dist</th>
<th>Type</th>
<th>Dist</th>
<th>Type</th>
<th>Dist</th>
<th>Type</th>
<th>Dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>915</td>
<td>1</td>
<td>1340</td>
<td>1</td>
<td>1530</td>
<td>3</td>
<td>1680</td>
</tr>
<tr>
<td>3</td>
<td>930</td>
<td>3</td>
<td>1130</td>
<td>1</td>
<td>1355</td>
<td>3</td>
<td>1560</td>
</tr>
<tr>
<td>1</td>
<td>2220</td>
<td>3</td>
<td>3350</td>
<td>3</td>
<td>3875</td>
<td>3</td>
<td>4215</td>
</tr>
<tr>
<td>1</td>
<td>2775</td>
<td>2</td>
<td>3355</td>
<td>3</td>
<td>3820</td>
<td>2</td>
<td>4040</td>
</tr>
<tr>
<td>1</td>
<td>2220</td>
<td>3</td>
<td>3200</td>
<td>3</td>
<td>4670</td>
<td>3</td>
<td>4915</td>
</tr>
<tr>
<td>1</td>
<td>2775</td>
<td>3</td>
<td>3200</td>
<td>2</td>
<td>4915</td>
<td>2</td>
<td>5185</td>
</tr>
<tr>
<td>3</td>
<td>3350</td>
<td>2</td>
<td>3200</td>
<td>3</td>
<td>3200</td>
<td>3</td>
<td>3625</td>
</tr>
<tr>
<td>3</td>
<td>3875</td>
<td>3</td>
<td>3355</td>
<td>3</td>
<td>3820</td>
<td>3</td>
<td>5900</td>
</tr>
<tr>
<td>3</td>
<td>4215</td>
<td>3</td>
<td>3355</td>
<td>3</td>
<td>3820</td>
<td>3</td>
<td>6205</td>
</tr>
<tr>
<td>3</td>
<td>7280</td>
<td>3</td>
<td>4215</td>
<td>3</td>
<td>7420</td>
<td>3</td>
<td>6425</td>
</tr>
<tr>
<td>3</td>
<td>7410</td>
<td>3</td>
<td>4215</td>
<td>3</td>
<td>7420</td>
<td>2</td>
<td>7550</td>
</tr>
<tr>
<td>3</td>
<td>7915</td>
<td>1</td>
<td>7695</td>
<td>3</td>
<td>7995</td>
<td>3</td>
<td>7995</td>
</tr>
<tr>
<td>1</td>
<td>8250</td>
<td>3</td>
<td>7995</td>
<td>1</td>
<td>8250</td>
<td>3</td>
<td>7995</td>
</tr>
</tbody>
</table>

Table C.2 “Long Span” Cell Clusters with Congestion Indication
Example 1: This example shows the process of the identification of “location” 5170 in Table C.3 (row 19). Location 5185 is listed under the 17:00~17:30 time interval in Table C.2 (row 6) as type 3 congestion indication. There is no location around 5185 listed in Table C.2 during other time intervals. However, in Table C.1 (row 6), there is a type 2 congestion indication at location 5165 during the 16:00~16:30 time interval spanning 20 meters (which indicates there is congestion indication at 5170). Therefore, distance 5170 is assigned as the modified distance in Table C.3 so that there are two time intervals (16:00~16:30 and 17:00~17:30) that have indication of congestion instead of just one. Doing this reduces the number of cell clusters eliminated in step 3 due to a small length span of congestion indication.

Example 2: Location 7280, during 7:30~8:00, spans only 10 meters of type 3 congestion indication as shown in Table C.1 (row 27). However, it covers a much longer span of type 2 congestion indication upstream of 7280 as shown Table C.1 (row 26). Given that there are no indications of congestion for this long span during the other time intervals, type 3 congestion indication is identified in Table C.3 at the combined location 7280 (since type 3 represents the case where both methods indicate congestion).
Step 5: After combining the desired “locations” over time, tabulate each location by its modified distance and its corresponding indication type for each time interval. The resulting list of these locations selected for detailed analysis in section 3.4 is shown in Table C.3. All 50 cell clusters forming the “locations” listed in Table C.3 are examined.
<table>
<thead>
<tr>
<th>Location</th>
<th>Indication Type</th>
<th>Expectation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time Interval</td>
<td>Time Interval</td>
</tr>
<tr>
<td></td>
<td>1  2  3  4</td>
<td>1  2  3  4</td>
</tr>
<tr>
<td>1</td>
<td>920 1 3 0 0</td>
<td>1 1</td>
</tr>
<tr>
<td>2</td>
<td>1030 0 0 2 2</td>
<td>? 1</td>
</tr>
<tr>
<td>3</td>
<td>1130 0 3 0 0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1350 1 1 0 0</td>
<td>? ?</td>
</tr>
<tr>
<td>5</td>
<td>1530 1 1 0 0</td>
<td>? ?</td>
</tr>
<tr>
<td>6</td>
<td>1560 0 3 0 0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1680 3 3 0 0</td>
<td>1 1</td>
</tr>
<tr>
<td>8</td>
<td>2040 0 2 1 0</td>
<td>? ?</td>
</tr>
<tr>
<td>9</td>
<td>2220 1 2 0 0</td>
<td>? ?</td>
</tr>
<tr>
<td>10</td>
<td>2780 1 0 0 0</td>
<td>?</td>
</tr>
<tr>
<td>11</td>
<td>3200 0 3 0 3</td>
<td>1 1</td>
</tr>
<tr>
<td>12</td>
<td>3350 3 3 0 0</td>
<td>1 1</td>
</tr>
<tr>
<td>13</td>
<td>3630 0 0 0 2</td>
<td>?</td>
</tr>
<tr>
<td>14</td>
<td>3860 3 3 0 0</td>
<td>1 1</td>
</tr>
<tr>
<td>15</td>
<td>4040 0 2 0 0</td>
<td>?</td>
</tr>
<tr>
<td>16</td>
<td>4220 3 1 0 0</td>
<td>?</td>
</tr>
<tr>
<td>17</td>
<td>4620 0 0 2 2</td>
<td>? 1</td>
</tr>
<tr>
<td>18</td>
<td>4940 0 0 2 2</td>
<td>? 1</td>
</tr>
<tr>
<td>19</td>
<td>5170 0 0 2 3</td>
<td>? 1</td>
</tr>
<tr>
<td>20</td>
<td>5560 0 0 1 0</td>
<td>?</td>
</tr>
<tr>
<td>21</td>
<td>5610 0 0 3 1</td>
<td>? ?</td>
</tr>
<tr>
<td>22</td>
<td>5660 0 0 2 3</td>
<td>? ?</td>
</tr>
<tr>
<td>23</td>
<td>5900 0 0 2 3</td>
<td>? 1</td>
</tr>
<tr>
<td>24</td>
<td>6210 0 0 3 0</td>
<td>?</td>
</tr>
<tr>
<td>25</td>
<td>6350 0 0 3 3</td>
<td>? 1</td>
</tr>
<tr>
<td>26</td>
<td>6700 0 0 0 3</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>7280 3 0 0 0</td>
<td>?</td>
</tr>
<tr>
<td>28</td>
<td>7420 3 3 0 0</td>
<td>? ?</td>
</tr>
<tr>
<td>29</td>
<td>7550 0 0 0 2</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>7700 0 1 0 0</td>
<td>?</td>
</tr>
<tr>
<td>31</td>
<td>7920 3 0 0 0</td>
<td>1</td>
</tr>
<tr>
<td>32</td>
<td>8000 3 3 0 0</td>
<td>1 1</td>
</tr>
<tr>
<td>33</td>
<td>8250 1 0 0 0</td>
<td>?</td>
</tr>
</tbody>
</table>

Indication Type: 0 Neither, 1 Median, 2 LST, 3 Both
Expectation: 1 expected, ? not sure

Table C.3 Locations for Detailed Analysis