# A Robust Bottleneck Identification Method using Noisy and Inconsistent Fixed-Point Detector Data

Jing Jin
Department of Civil and Environmental Engineering,
University of Wisconsin-Madison,
Madison, WI 53706
Phone: 1-608-262-2524
E-mail: jjin2@wisc.edu

Wenjiao Yu
Department of Civil and Environmental Engineering,
University of Wisconsin-Madison,
Madison, WI 53706
Phone: 1-608-262-2524
E-mail: wyu7@wisc.edu

Jie Fang
Department of Civil and Environmental Engineering,
University of Wisconsin-Madison,
Madison, WI 53706
Phone: 1-608-262-2524
E-mail: jfang3@wisc.edu

Bin Ran
Department of Civil and Environmental Engineering,
University of Wisconsin-Madison,
Madison, WI 53706
Phone: 1-608-262-0052
E-mail: bran@wisc.edu

Corresponding Author: Jing Jin

Submitted for Presentation and Publication
to the 89th Transportation Research Board Meeting
Submission Date: July 29th, 2009

**5522 Words + 3 Tables + 4 Figures = 7272 Words**

## ABSTRACT

Bottleneck identification locates problematic segments on a freeway corridor and meanwhile provides information about the cause and characteristics of the congestion. It is a critical step in mitigating the urban congestion problem. Due to the wide availability of traffic surveillance data, researchers have been designing bottleneck identification algorithms based on archived traffic flow data. Those algorithms include rule-based, contour-map-based and simulation-based methods. However, these existing methods require traffic data with high accuracy and consistency, which may not always be the case in reality. In this paper, a new bottleneck identification method based on coordinate transformation on fundamental diagram is proposed. The algorithm is designed for fix-location detector data and can tolerate noise and inconsistency. Three loop detector datasets were collected at the city of Madison and the city of Milwaukee, WI, USA. The three datasets have different levels of data quality so that the effectiveness and robustness of the proposed algorithm can be tested. Meanwhile, a novel evaluation strategy for bottleneck identification in the absence of ground truth data was first introduced in this paper. Using this strategy, the proposed algorithm is compared with Chen's method. The evaluation results indicate superior effectiveness and robustness of the proposed algorithm comparing to earlier methods.

# INTRODUCTION

## Freeway Bottlenecks and Bottleneck Identification

Congestion caused by bottlenecks contributes about 40% of the total urban congestion (*1*). As a result, the understanding, detecting and managing highway bottlenecks has long been a primary focus of freeway operations. By definition, a bottleneck is a short segment of highway with insufficient capacity (*2*). Based on the cause of bottlenecks, they can be classified into two categories, recurrent and non-recurrent bottlenecks. Recurrent bottlenecks are caused by periodic traffic demand changes. When traffic demand exceeds the capacity of a roadway segment, bottleneck appears. Non-recurrent bottlenecks are caused by temporary capacity-reduction events such as incidents, slow moving vehicles etc. When bottleneck is causing queue accumulation and congestion, the bottleneck is considered activated. Otherwise, the bottleneck is inactive.

The scope of this paper focuses on freeway bottlenecks, whose characteristics have been studied for decades. There are three major phases for an active freeway bottleneck, pre-activation, bottleneck activation, and bottleneck continuation. Each phase has its own characteristics. At pre-activation period, the transition from saturated traffic to congested traffic is the primary focus. There are two major features, the duration of the transition period (*3*) and the probability of such transition with respect to flow and other factors (*4*). During the breakdown phase, the formulation of queues and flow breakdown are the primary characteristics (*4*). At the bottleneck continuation phase, as pointed out by several researchers (*3*, *5*, *6*, *7*, *8*), the most characteristic feature is the queue discharge flow (QDF). As concluded in Cassidy and Bertini's (*5*) research, QDF, during an active bottleneck period, can exhibit nearly stationary patterns that alternating between high and low flow level and gradually diminish over time. At the same bottleneck location, the QDF can be significantly lower than normal flow before breakdown for as much as 10% or more (*3*). Considering the normal duration of bottleneck congestion is about half an hour (*3*), the resulting large vehicle delay can seriously reduce the Level of Service (LOS) at a freeway segment.

In order to cope with the bottleneck congestion problem, different strategies can be used according to congestion severity, budget or resource limitations. With enough budgets, space and needs, large construction projects, for instance adding lanes, building interchanges, can be conducted to increase capacity for road sections with severe congestion problem. In other cases, when large construction is not an option, many ITS (Intelligent Transportation System) technologies can be used, including ramp metering, traveler information based re-routing, and more recently the "Active Traffic Management" concept introduced by European highway engineers (*9*), which includes a series of ITS operations to relieve the congestion, e.g. temporary should lane opening, variable speed limit for speed harmonization, dynamic message sign for driver notification and detour advising. And many of the above methodologies are found to be able to improve the service performance at bottlenecks. However, a critical step before taking such operations is to identify problematic road sections, that is, bottleneck identification. Identifying the bottleneck locations from a large urban freeway network is of great importance for further analysis and the search for alleviation alternatives. To this sense, we shall focus on recurrent bottleneck identification in this paper. Traditionally, bottleneck identification relies on floating car method. Floating vehicles are dispatched at scheduled peak periods, several times a year to investigate a freeway corridor. And crews on board inspect any congestion problems and take notes. Such method is labor-intensive and has very little temporal or spatial coverage. Now, due to the wide deployment of loop detecting systems on major freeways, more efficient detection methods are found. The archived traffic flow data of these systems allow engineers to inspect bottlenecks in a large roadway network by investigating their performance with abundant and complete measurements.

Over the past decade, several research work has been done on bottleneck identification, including the original rule-based method proposed by Chen etal.(*10*), speed contour map based method proposed by Ban etal. (*11*), and a fuzzy logic based algorithm introduced in 2009 (*12*). Another direction is to investigate the possibility of using micro-simulation models to identify bottleneck (*11*, *13*). However, due to the time-consuming model building and calibration process, simulation is usually considered as a detailed bottleneck analysis method after the identification of critical bottleneck sections or corridors.

Chen's algorithm has been tested against California and later Virginia loop detector data (*10*, *14*). Ban's algorithm is implemented also based on California data. The dataset used for testing by these two algorithms are both based on well-calibrated and maintained detectors, however, such good data quality is not a common case among other existing loop detecting systems (*15*). Loop detecting systems typically experiences two major types of errors: measurement errors and data inconsistencies. Measurement errors include missing or repeating values, exceeding valid range etc. These errors can cause complete failure of bottleneck identification algorithms if happened on a large scale or for a long period of time and cannot be fixed. However, if such errors occur only for a short period of time at a few locations, they will not reduce the algorithm performance too much and there are techniques such as zero-filling (*10*) and interpolation to fix them. Data inconsistencies include the follow cases (*15*):

- Rapid fluctuations in values across successive time periods;
- Reported values that are significantly different from the location's history for similar days of the calendar.
- Detectors in adjacent lanes at the same location reporting significantly different values or trends;
- Detectors in adjacent upstream or downstream locations reporting significantly different values or trends;
- Detectors from multiple locations reporting the same values (indicative of a system problem);

The above inconsistencies, no matter temporally (the first two) or spatially (the latter three), can result in serious problems for the existing algorithms because they are based on the assumption that detectors are behaving consistently between upstream and downstream, between the previous and the current time intervals. A typical example of such inconsistency is the difference of measured "free-flow speed" between adjacent detectors. Since the speed measurement is distorted. It will be quite difficult to conduct any type of bottleneck identification using speed. However, the dataset itself still contains very useful information about congestion, after proper normalization, these data can still be used. Fei etal.(*12*) try to solve the data quality problem with fuzzy logic. However, their fuzzification process, which determines traffic condition levels for fuzzy logic, requires human interpretation of loop measurement. Obtaining and validation of such knowledge is difficult, especially such knowledge can vary from station to station and from time to time. An experienced traffic operator may solve the knowledge issue. But when implementing such algorithm on a large scale, the processing load for the operators may be too high. Enlightened by HCM (Highway Capacity Manual) method for determining the Level of Service (LOS) for highway segments, this paper introduces a new bottleneck identification algorithm that can reduce the impact of the above noises and inconsistencies issues.

## HCM Method of Evaluating Highway Traffic Condition

In traffic operations, fundamental diagrams of traffic flow (FDs) have long been used to evaluate the performance of highway facilities. For example, in the Chapter 23 of Highway Capacity Manual 2000 (*17*), the speed-flow diagram is used to determine the level of service (LOS) for basic freeway segment. Several lines are drawn to divide the entire speed-flow diagram into six regions and the LOS for each region goes from A to F. The underlying assumption for this approach is that there exists correlation between the "intensity" of traffic condition and the relative location of traffic states on FDs. This idea can be used in bottleneck identification because the key for bottleneck identification is the detection of traffic congestion, which is a severe change of traffic condition. Furthermore, a major benefit of this method is that it automatically eliminates the impacts of data inconsistency because the determination of traffic condition "intensity" is entirely based on the local traffic flow features (the shape of FDs at a detector station) and the output is a traffic condition evaluation (LOS) which is comparable and uniform among different sites.

## The Coordinate Transformation on Fundamental Diagrams of Traffic Flow

As proposed in Jin (*18*)'s paper, coordinate transformation can be used to convert the original flow-occupancy diagram into a more descriptive coordinate system of traffic flow, the URS (Uncongested Regime Shift)-CRS (Congested Regime Shift) system. The axes of the new coordinate system align with two distinct regimes usually found in the flow-occupancy diagram, the free-flow regime and the congested regime. The parameters of these two axes can be found using simple linear regression. And the transformed coordinate system can track traffic condition changes more sensibly and descriptively than the original flow, occupancy and speed readings. However, a major drawback of the URS-CRS projection is that the CRS axis requires large amount of congested flow data to calibrate, which may not always be available. And also, traffic states are sparsely distributed at congested regime so it is difficult to justify the existence of such congested regime line statistically. Using similar coordinate transformation techniques, in this paper, we shall improve such projection so that the projection can still be accurately conducted when there are not enough congested traffic measurements.

## METHODOLOGY

### *The URS-PUS System*

To overcome the lack of congested traffic states for CRS line calibration, we propose another coordinate system similar to the URS-CRS system. The two axes are "uncongested regime shift" (URS) and "perpendicular to uncongested regime shift" (PUS). URS is the same as in the URS-CRS system but PUS represents an axis that is perpendicular to the uncongested regime. This coordinate system only needs uncongested traffic states to calibrate. Transformation to this coordinate system consists of two steps. The first step is the translation of the origin from (0, 0) to ($o_0$, $v_0$). The second step is to rotate the new coordinate system by (90-θ) degree clockwisely, where θ is the angle between the free flow regime and the occupancy axis. In traffic flow, θ is between (0, π/2). Then the transformation matrix formula from a coordinate P(o,v) in the flow-occupancy coordinate to its new coordinate P(p, u) (p is new PUS coordinate and u is new URS coordinate), is as follows:

$$\begin{bmatrix} p \\ u \end{bmatrix} = \text{diag}\left(\frac{1}{d_{O'Q}}, \frac{1}{d_{O'O}}\right)\begin{bmatrix} \sin\theta & -\cos\theta \\ \cos\theta & \sin\theta \end{bmatrix}\begin{bmatrix} o - o_0 \\ v - v_0 \end{bmatrix}$$

Where

$$\begin{cases} d_{O'O} = \sqrt{o_0^{\,2} + v_0^{\,2}} \\ d_{O'Q} = v_0 / \cos\theta \end{cases}$$

The new URS-PUS system has similar characteristics as URS-CRS system but is easier to calibrate and update since it only corresponds needs free flow data. Also divisor vector using $d_{O'O}$ and $d_{O'Q}$ unifies the transformed coordinates. The unification allows all flow-occupancy diagrams to be mapped onto a single template diagram. In this way, any two URS-PUS coordinates are comparable even though they may come from different detectors.

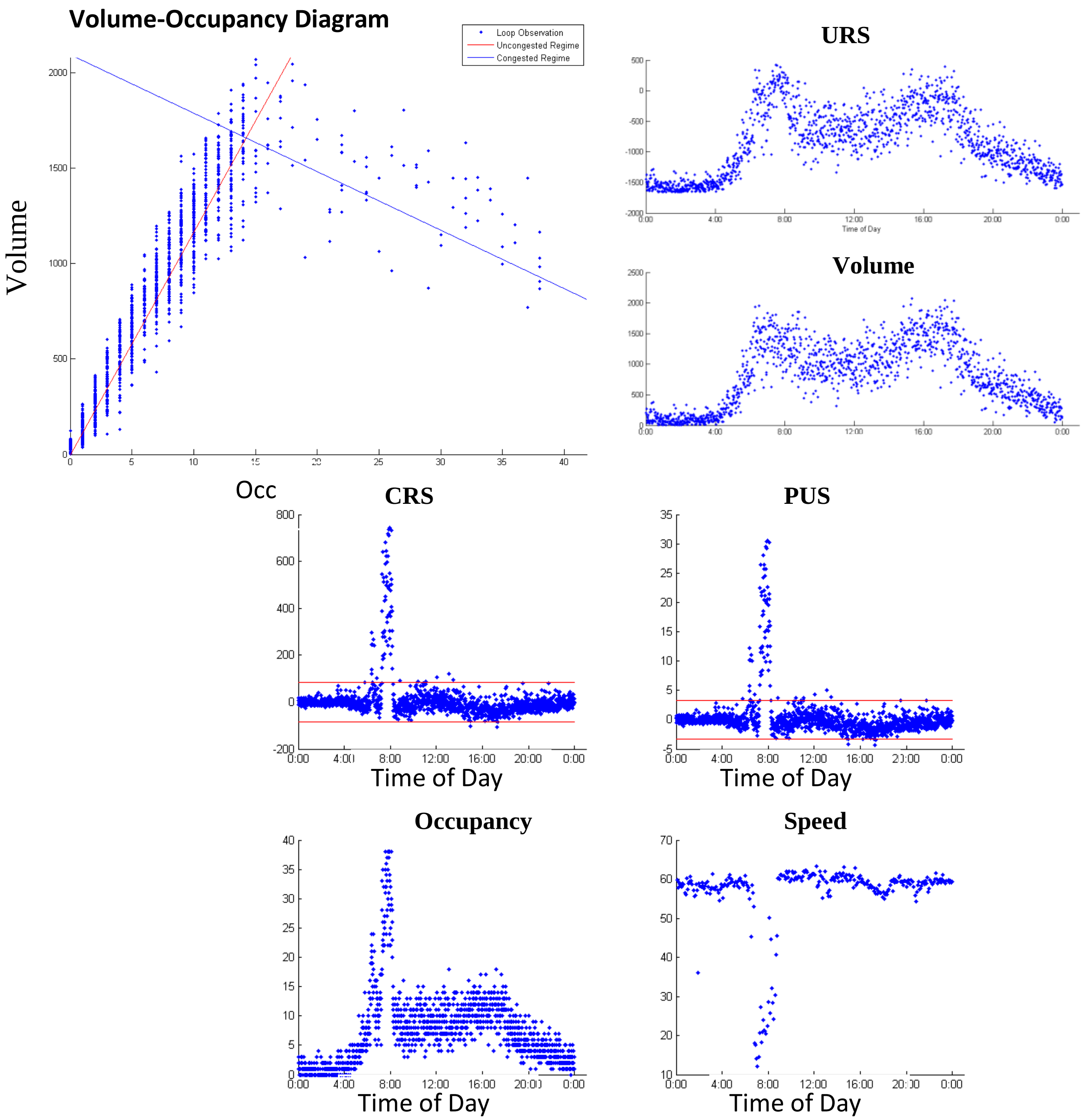



**FIGURE 1 Characteristics of URS, CRS and PUS based on field data collected at link 4017 (Feb. 4th, 2008), on I894 freeway, Milwaukee, WI, USA.**

In Figure 1, URS, CRS and PUS are calculated for each 1-min time interval using one-day data collected at link 4017 of I-894 freeway in Milwaukee, WI, USA. Because the 1-min speed data available in WisDOT website are truncated at speed limit to prevent promoting speeding, 5-min data are used to show the temporal pattern of speed. The results show good characteristics of URS, CRS and PUS. The pattern of URS is quite similar to volume which is an good indicator of demand. CRS and PUS are as sensitive as speed in detecting congestion, but it can provide more details about traffic condition changes during congested period rather than sudden jumps observed in speed measurements.

The URS-PUS system still has its limitations when data quality and completeness cannot be ensured. For example, if a road segment always experiences low traffic volume, then determining the critical point $(o_0, v_0)$ becomes difficult. The transformation still works but it will be difficult to conduct unification and the resulting URS-PUS values may not be comparable with those from another site. Large variation and data noise can still pollute the URS and PUS results though they have better resistance to such impact comparing with the original flow and occupancy.

*The Proposed Bottleneck Identification Algorithm*

The proposed algorithm includes two major steps: congestion map creation and frequency analysis. Details of each step are shown in Figure 2.

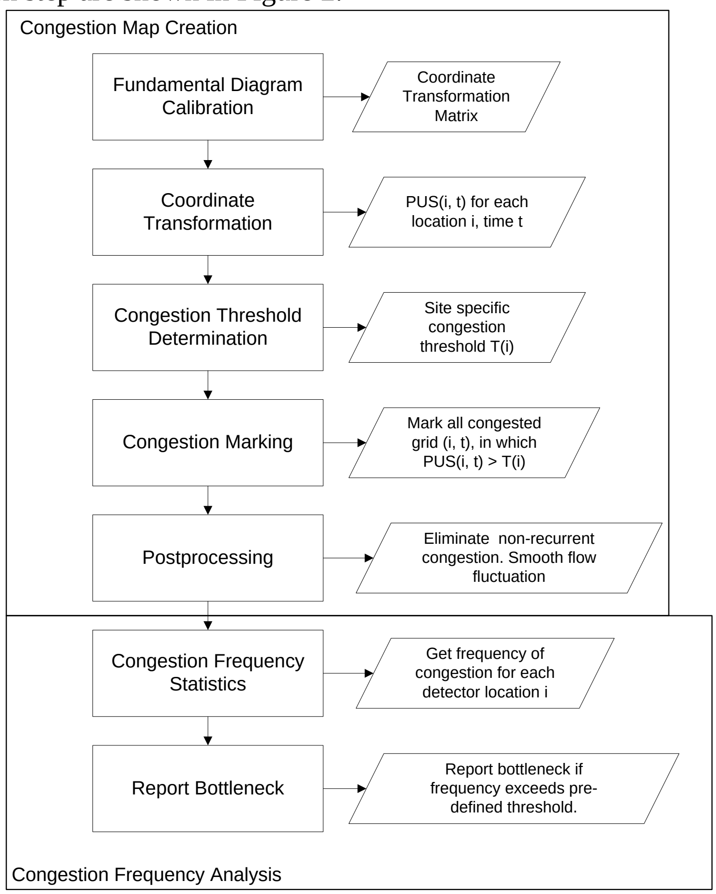



**FIGURE 2 Flow chart for the proposed algorithm**

**Fundamental Diagram Calibration** The calibration is a simply linear regression based on traffic measurements within uncongested flow. However, a critical problem at this step is to classify traffic states into uncongested and congested. The proposed classification method is based on the relative locations of traffic state with respect to the critical point ($o_C$, $v_C$), which represents maximum flow rate. If a traffic state has an occupancy value less than $o_C$, then it is considered to be uncongested. Admittedly, traffic states near capacity point cannot be explicitly classified into congested or uncongested. But their impact on the accuracy of the calibrated coefficients is small. This is because 1) those traffic states are not dominant traffic states at uncongested regime, 2) they are quite close to the regression trend line and the resulting deviation is small. Another problem is the linear formulation of the URS regime. Since physically when flow is zero, the occupancy should be zero (no vehicles are on the detector), the trend line of URS is designed to start from the origin. As a result, the linear regression method with no intercept is used. Assume the coordinate of each observed uncongested traffic state is ($o_i$, $v_i$), where i=1,2,… is the index of all traffic states in uncongested flow. Then the slope of URS regime becomes:

$$k_{URS} = \sum_i o_i v_i \Big/ \sum_i o_i^2$$

**Coordinate Transformation** Coefficients in the transformation matrix include the coordinate ($o_0$, $v_0$) of the new origin and the angle θ between URS and occupancy axis in the flow-occupancy diagram.

Let the new origin for URS-PUS system be ($o_0$, $v_0$), where $o_0 = o_C$, the critical occupancy and v0 = $k_{URS} \cdot o_0$. Note that here $v_0$ is not the maximum flow $v_C$. It is an estimated maximum flow based on the regression model. And these coefficients can be calculated as the follow:

$$\begin{cases} o_0 = o_C \\ v_0 = k_{URS}\, o_C \\ \theta = \operatorname{atan}\left(k_{URS}\right) \end{cases}$$

The above coefficients are calculated for each day of week within each month. Using these coefficients, the transformation matrice can be established. And all measurements are converted to the new URS-PUS coordinates based on their corresponding transformation matrices. In the proposed bottleneck identification method, only PUS value is used as the congestion indicator.

**Congestion Threshold and Congestion Marking** The congestion threshold is determined by both statistical analysis over the uncongested data and visual inspection of the PUS contour maps. Statistical analysis provides some candidate values for thresholds and visual inspection helps to determine the actual threshold. First, the mean, minimum, maximum and standard deviation are calculated for uncongested PUS values for each month and for all links. Then candidate thresholds are selected based on the above statistics. The congestion marking step goes through every detector location i at every time interval t and check if PUS(i, t) exceeds the threshold. If so, congestion flag is set for location i and time interval t. Otherwise, no congestion flag is marked.

**Post-Processing** Post-processing includes two parts: the elimination of non-recurrent congestion and filling holes within a congestion period caused by flow fluctuation. Incident logs maintained at traffic operation centers can be used to eliminate congestion caused by incidents. If such operator logs are not available, this step can be skipped. The reason is that as long as such non-recurrent congestion does not happen repeatedly at the same segment, they will not yield high frequency, hence not be considered as recurrent congestion. Another part of post processing is to smooth the flow fluctuation. Ban (*11*) introduced an effective zero filling technique in his bottleneck identification algorithm. In the proposed algorithm, similar techniques are used.

**Congestion Frequency Statistics and Bottleneck Report** Spatially, congestion frequency is analyzed for links between each pair of detectors. And temporally, it is estimated for each 15-minute period in a day. For each 15 minutes in a day, the algorithm finds the starting point of congestion by checking the start of congestion flag within each 15 minutes. Then the starting location and its time period is recorded for a candidate bottleneck. And a bottleneck is identified and reported if the frequency of a candidate bottleneck exceeds the pre-defined frequency threshold (e.g. 60%, 70%, 80% etc.).

## EXPERIMENTAL DESIGN

### Data Source

Data sources for our study are dual loop detector measurements at two freeway corridors in Wisconsin (See Figure 3). One corridor is located at the I-894 freeway (between W Greenfield Avenue and S 27th Street) at Milwaukee, WI, USA. The total length is 8.5 mile (about 13.7 km). A total of 27 detector stations (19 at west-to-north direction, 18 at south-to-east direction) are within the testing corridor. Detector stations are located near or at the interchanges. The average spacing between detectors is about half a mile (805m). A supplementary incident log obtained from Milwaukee State Truck Operations Center (STOC) is used to eliminate non-recurrent congestion. The other corridor is on the USH 12/18 at Madison, WI, USA. The length is about 13.1 mile (about 21.1 km). And it is covered by 28 detector stations (14 at both directions), with an average spacing about 1 mile (about 1.5 km). No incident data log is available at this point, however, the incident log is not necessary since incidents are rare events comparing to a recurrent bottleneck.

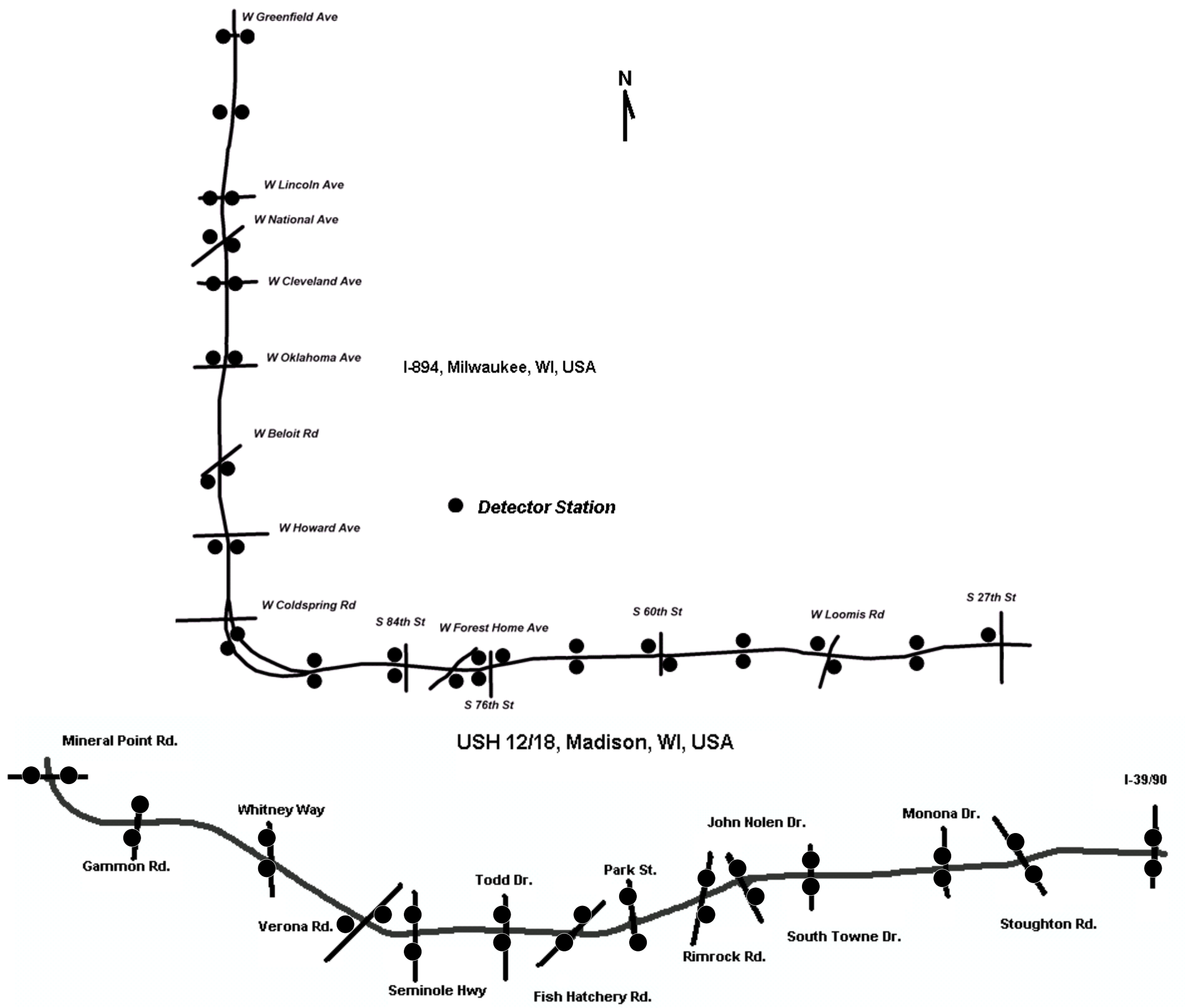



**FIGURE 3 Detector map for two test sites.**

All detectors are dual loop detectors with spot speed readings. However, data qualities are quite different between the two sites. I-894 detectors are well-maintained and well-calibrated, while USH 12/18 data has serious data consistency and data noise issues. For I-894 data, two different frequencies are available. One is 1-min data archived from the traveler information website of Wisconsin Department of transportation (WisDOT) (*19*). The other one is the archived 5-min data from the detector data archiving database. Due to archiving system issues, 5-min data does not have same data quality as 1-min data. Since three datasets (1-min I894 data, 5-min I894 data and 5-min USH12/18 data) represent three different levels of data quality, they are suitable for testing both effectiveness and robustness of the proposed algorithm. The time range for all three datasets is from January to May, 2008.

## Model Validation and Evaluation

As mentioned before, so far there has not been a comprehensive, widely-accepted benchmark or evaluation framework available for comparing bottleneck identification algorithms. And in reality, the evaluation of identified bottlenecks relies on the judgment of traffic operators. And there are several techniques to assist such human evaluation e.g. contour maps, temporal profile of measurements. But the best supplementary technique is to correlate actual surveillance videos with detected bottlenecks so that one can see the actual queue formulation in the congestion. In our study, such video data is not available. Then we have to use common knowledge about the characteristics of freeway bottlenecks to design our evaluation criteria. The criteria used in our evaluation are as the following:

- *Activation Time*: the activation time found for a bottleneck is usually within morning or afternoon peak hours. Any bottleneck activated at other periods is considered to be invalid. This criterion subjects to operator experiences if the dataset comes from large metropolitan areas. For example, for large urban area such as Chicago, Los Angeles, the range of activation time should be set to much larger, e.g. 6 am to 9 pm.
- *Activation Period*: the total length of a bottleneck should be reasonable, which is interpreted as about 15 minutes to 2 hours. Any bottleneck activated for more than 2 hours is considered invalid unless local experience is available about such bottlenecks. This criterion is designed for the tested dataset in this research. Changes should be made to allow longer activation period for large urban area.
- *Propagation Speed*: At the boundaries of an active bottleneck on spatial-temporal diagram, one can calculate the speed of congestion propagation. The propagating speed should be reasonable. This can eliminates global detector failures and global events such as severe weather conditions, which will generate horizontal or vertical boundaries that are not reasonable.

Note that these criteria can only eliminate some false alarms. However, under the absence of “true” data, accurate estimation of detection rate is impossible. One possible way to evaluate is to allow two candidate algorithms to produce the same amount of detection over the same dataset under “fair” condition and compare the number of false alarms found within their detection. And if algorithm A is better than B, at the same detection rate, A should generate fewer false alarms than B. The key point is how to establish “fair” condition. This is doable for bottleneck identification. Since the output of existing bottleneck identification methods, including the proposed algorithm in our study, all reports frequency or percentile of bottleneck activation as an accompanying output. And there is natural ordering for the output based on their frequencies or percentiles. We can use the ranking of frequency to control the algorithms to “fairly” produce the same number of detected bottlenecks. Then we compare the number of false alarms. The more the false alarms, the worse the algorithm performance. The only defect of this strategy is that it is possible that our criteria to identify false alarms may be incomplete and some missing false alarms may benefit certain algorithms. Nevertheless, this is the best we can reach to compare two bottleneck identification algorithms where there are no true bottleneck data available. In this study, the proposed algorithm and the reference algorithm are both applied to three datasets: 1-min I-894 (dataset A), 5-min I-894 (dataset B) and 5-min USH12/18 (dataset C). And the three datasets serve as the major three scenarios. And each algorithm will produce the top five or ten bottlenecks and each bottleneck will be checked to see if it is a false alarm.

## Reference Algorithm

The reference algorithm in this study is Chen’s method (*10*). Chen’s method uses the raw detector data. The algorithm includes two steps: congestion identification and congestion frequency test. For any two locations, $x_i$ and $x_j$, with $x_i<x_j$ (i is to the upstream of j), congestion is detected at $x_j$ if the following four inequalities hold:

- $x_j - x_i < 2$ mile (Spacing constraint)
- $v'(x_k, t) - v'(x_j, t) > 0$, if $x_i \le x_k < x_l < x_j$ (speed decrease from $x_i$ to $x_j$)
- $v'(x_j,t) - v'(x_i,t) > 20$mph (significant changes)
- $v'(x_i, t) < 40$mph (location i is at congestion)

The second inequality indicates that although location $x_i$ is upstream of $x_j$, but there may be other detectors at $x_k$, $x_l$ between these locations. In order to determine whether or not there is a bottleneck, a symbol $A_i(t)$ is used and $A_i(t) = 1$ if there is an active bottleneck at location i and time period t.

The frequency test is based on the following inequality for each time period $[t_1, t_2]$ and each location *i*. If the inequality holds, a sustained bottleneck location is found.

$$\sum_{\tau=t}^{t+N-1} A_i(\tau) \geq qN \quad \text{, for all } t_1 \leq t \leq t_2 - N + 1,$$

Where N = 7 and q = 5/7. That is, a sustained bottleneck has at least five active bottleneck periods (or 25 minutes) within every seven consecutive periods (or 35 minutes).

## RESULTS ANALYSIS

### Model Evaluation Results

The performance of the proposed algorithm is compared with the Chen's algorithm following the criteria introduced in previous section. The number of top bottlenecks generated is five for dataset A and ten for both dataset B and C. The number is small for A because dataset A is a relatively clean dataset and very few false alarms are generated.

In Table 1, since the top 5 bottlenecks detected by both algorithms are the same, those detected by Chen's method are all reasonable. However, for the proposed algorithm, one unreasonable bottleneck is generated. And we can see that top bottlenecks detected by both algorithms are quite consistent with only a different of three bottlenecks. The false alarm is mainly caused by unusual high speed measurement (up to 117mph) for several days causing in-accurate calibration of transformation matrix (See Figure 5.3a to 5.3c). Table 2 provides the comparison for noisier dataset B. Chen's algorithm output four false alarms while the proposed algorithm reports none. And most of the top incidents detect by Chen's method for 1-min data are among the top bottlenecks detected by the proposed algorithm. Again, in Table 3, under the noisiest dataset C, four false alarms are still found for Chen's algorithm and the algorithms almost failed at the Rimrock Road detectors. But the proposed algorithm still generates reasonable bottlenecks. And the bottlenecks identified are quite consistent with the authors' driving experiences on that corridor. Based on the three tables, we can clearly see the effectiveness and robustness of the proposed algorithm.

**TABLE 1 Evaluation Results for Dataset A (1-min I-894 Data)**

| Chen's Algorithm | | | | | |
|---|---|---|---|---|---|
| Month | Corridor | Cross Street | Location(mile) | Duration (24 hr) | Frequency |
| **MAY** | **I894 WB-NB** | **Cleveland Ave.** | **6** | **07:30-08:00** | **72.40%** |
| **APR** | **I894 SB-EB** | **Howard Ave.** | **2.7** | **07:30-07:45** | **73.30%** |
| **APR** | **I894 WB-NB** | **Howard Ave.** | **4.4** | **07:30-07:45** | **73.30%** |
| **FEB** | **I894 WB-NB** | **Beloit Rd.** | **5** | **07:30-08:00** | **72.40%** |
| **JAN** | **I894 SB-EB** | **Howard Ave.** | **2.7** | **07:30-07:45** | **72.00%** |
| JAN | I894 WB-NB | Howard Ave. | 4.4 | 07:30-07:45 | 72.0% |
| MAR | I894 WB-NB | Lincoln Ave. | 6.6 | 08:00-08:15 | 71.4% |
| The Proposed Algorithm | | | | | |
| Month | Corridor | Cross Street | Location(mile) | Duration (24 hr) | Frequency |
| ***JAN** | **I894 SB-EB** | **National Ave.** | **1.2** | **06:00-21:45** | **70.70%** |
| **FEB** | **I894 WB-NB** | **National Ave.** | **6.5** | **07:15-08:30** | **69.00%** |
| **FEB** | **I894 WB-NB** | **Beloit Rd.** | **5** | **07:15-08:00** | **67.80%** |
| **APR** | **I894 WB-NB** | **Howard Ave.** | **4.4** | **07:30-07:45** | **66.70%** |
| **JAN** | **I894 WB-NB** | **Howard Ave.** | **4.4** | **07:30-08:00** | **66.00%** |
| JAN | I894 SB-EB | Howard Ave. | 2.7 | 07:45-08:00 | 64.00% |
| MAR | I894 WB-NB | Cleveland Ave. | 6 | 07:30-08:00 | 61.40% |

* Unreasonable bottleneck.

**TABLE 2 Evaluation Results for Dataset B (5-min I-894 Data)**

| Chen's Algorithm | | | | | |
|---|---|---|---|---|---|
| Month | Corridor | Cross Street | Location(mile) | Duration (24 hr) | Frequency |
| *MAY | I894 SB-EB | 68th St. | 5 | 18:45-00:00 | 64.5% |
| *MAR | I894 SB-EB | 68th St. | 5 | 02:45-05:00 | 64.5% |
| FEB | I894 WB-NB | 35th St. | 0.5 | 16:45-17:30 | 62.1% |
| *MAY | I894 SB-EB | 68th St. | 5 | 00:00-07:15 | 61.3% |
| *JAN | I894 SB-EB | 68th St. | 5 | 01:45-05:00 | 61.3% |
| JAN | I894 WB-NB | Coldspring Rd. | 3.7 | 07:30-08:00 | 61.3% |
| JAN | I894 WB-NB | Lincoln Ave. | 6.6 | 06:30-07:00 | 61.3% |
| FEB | I894 WB-NB | 27th St. | 0 | 17:00-17:30 | 60.4% |
| APR | I894 WB-NB | Oklahoma Ave. | 5.4 | 07:00-08:00 | 59.2% |
| APR | I894 WB-NB | Howard Ave. | 4.4 | 07:15-08:00 | 58.9% |
| The Proposed Algorithm | | | | | |
| Month | Corridor | Cross Street | Location(mile) | Duration (24 hr) | Frequency |
| MAY | I894 WB-NB | Beloit Rd. | 5 | 07:00-08:15 | 62.8% |
| JAN | I894 WB-NB | Beloit Rd. | 5 | 06:30-08:15 | 62.5% |
| MAR | I894 WB-NB | Cleveland Ave. | 6 | 06:30-08:00 | 59.5% |
| MAY | I894 SB-EB | 35th St. | 6.8 | 07:30-08:00 | 58.7% |
| FEB | I894 WB-NB | Howard Ave. | 4.4 | 07:15-08:00 | 58.3% |
| MAY | I894 WB-NB | Cleveland Ave. | 6 | 06:30-08:30 | 57.9% |
| APR | I894 WB-NB | Coldspring Rd. | 3.7 | 07:30-07:45 | 57.7% |
| MAY | I894 WB-NB | 84th St. | 3 | 07:30-07:45 | 56.5% |
| MAR | I894 WB-NB | Howard Ave. | 4.4 | 07:30-08:00 | 56.3% |
| MAY | I894 SB-EB | Greenfield Ave. (Belton OP) | 0.4 | 16:00-16:15 | 56.0% |

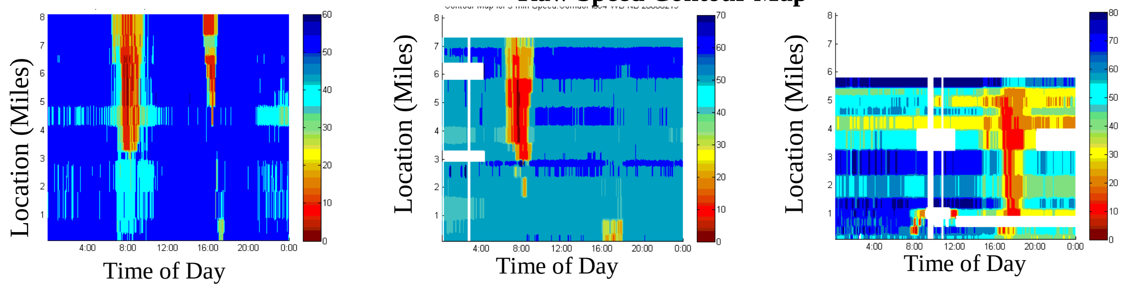
* Unreasonable bottleneck.

**TABLE 3 Evaluation Results for Dataset C (5-min USH 12/18 Data)**

| Chen's Algorithm | | | | | |
|---|---|---|---|---|---|
| Month | Corridor | Cross Street | Location(mile) | Duration (24 hr) | Frequency |
| *JAN | US-12/14 EB | Rimrock Rd. | 4.2 | 03:00-04:45 | 54.8% |
| *JAN | US-12/14 EB | Rimrock Rd. | 4.2 | 01:00-02:00 | 51.6% |
| *JAN | US-12/14 EB | Rimrock Rd. | 4.2 | 10:00-11:00 | 50.8% |
| JAN | US-12/14 EB | Rimrock Rd. | 4.2 | 17:00-18:00 | 50.8% |
| JAN | US-12/14 EB | Rimrock Rd. | 4.2 | 08:00-09:00 | 50.0% |
| MAY | US-12/14 WB | Monona Dr. | 1.3 | 07:30-08:00 | 50.0% |
| MAY | US-12/14 WB | Rimrock Rd. | 2.3 | 16:00-17:00 | 50.0% |
| JAN | US-12/14 EB | Rimrock Rd. | 4.2 | 05:00-07:00 | 49.2% |
| JAN | US-12/14 WB | Rimrock Rd. | 2.3 | 16:00-17:00 | 49.2% |
| *JAN | US-12/14 EB | Rimrock Rd. | 4.2 | 00:15-00:45 | 48.4% |
| The Proposed Algorithm | | | | | |
| Month | Corridor | Cross Street | Location(mile) | Duration (24 hr) | Frequency |
| FEB | US-12/14 WB | Stoughton Rd. | 0.8 | 07:15-08:15 | 86.3% |
| FEB | US-12/14 WB | Monona Dr. | 1.3 | 07:15-08:15 | 75.4% |
| FEB | US-12/14 WB | Fish Hatchery Rd. | 2.9 | 17:00-17:30 | 68.8% |
| FEB | US-12/14 WB | Todd Dr. | 3.7 | 07:45-08:00 | 64.3% |
| FEB | US-12/14 EB | Todd Dr. | 1.7 | 16:30-17:00 | 62.6% |
| FEB | US-12/14 EB | Todd Dr. | 1.7 | 17:00-17:30 | 62.5% |
| FEB | US-12/14 WB | Fish Hatchery Rd. | 2.9 | 08:00-08:15 | 61.5% |
| FEB | US-12/14 EB | John Nolen Dr. | 4.5 | 16:00-17:30 | 60.1% |
| APR | US-12/14 EB | South Towne Dr. | 4.8 | 16:45-17:30 | 59.4% |
| APR | US-12/14 EB | Park St. | 3.6 | 16:45-17:30 | 58.8% |

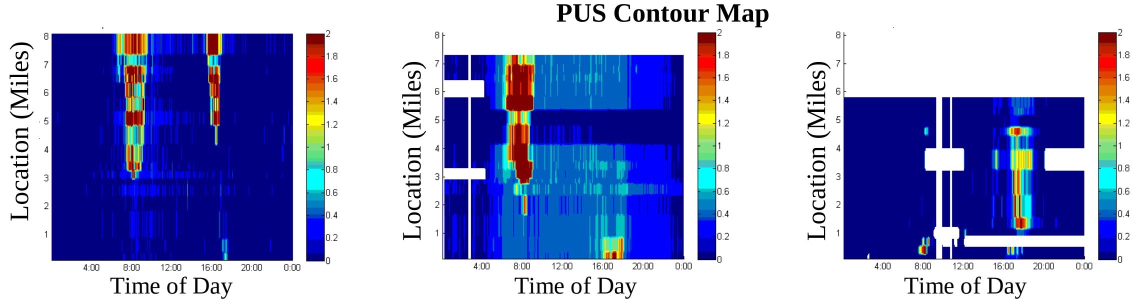
* Unreasonable bottleneck.

**Raw Speed Contour Map**



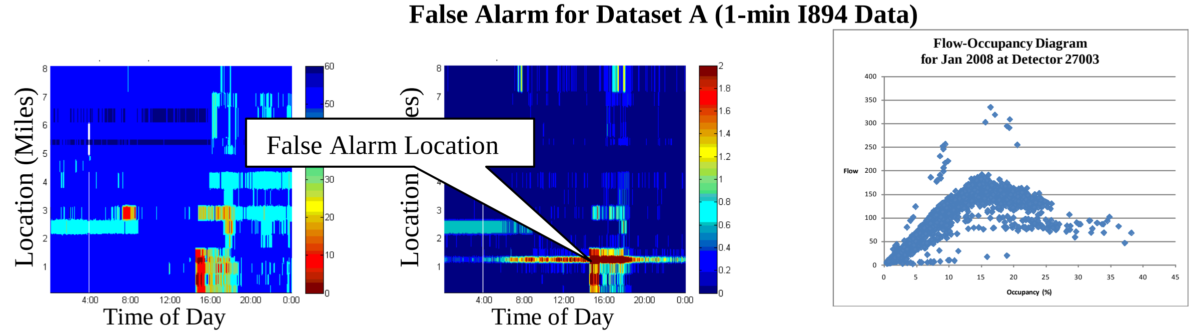

**(5.1a) Dataset A** **(5.1b) Dataset B** **(5.1c) Dataset C**

**PUS Contour Map**




**(5.2a) Dataset A** **(5.2b)Dataset B** **(5.2c)Dataset C**

**False Alarm for Dataset A (1-min I894 Data)**




**(5.3a) Raw Speed** **(5.3b) PUS Value** **(5.3 c) Flow-Occ Diagram**

**FIGURE 4 Sample diagrams for evaluation test.**

## CONCLUSION AND FUTURE STUDY

### Conclusion

The paper proposed a new bottleneck identification algorithm, which can be used against noisy, inconsistent fixed-location traffic data. In this study, Wisconsin loop detector data are used as a case study for the algorithm. The coordinate transformation technique used in the algorithm automatically converts and unifies the flow and occupancy data to the URS-PUS coordinate system. And the resulting PUS value is a good replacement for the "unreliable" speed variable under noisy condition. The algorithm is compared with a reference algorithm, Chen's algorithm, by running them against three data sets with different data qualities. For the dataset with best data quality, Chen's method is slightly better. However, for noisier dataset B and C, the proposed algorithm keeps performing much better than Chen's method. The comparison results proved the effectiveness and robustness of the proposed algorithm. Moreover, the algorithm is quite easy to implement that it can be deployed within an Oracle 10g database. Except for the above evaluation results, there are two other highlights in this study that worth being mentioned.

*Evaluation of Bottleneck Identification Algorithms*

The evaluation strategy implemented is a novel approach. First, the detection rate is fixed for candidate algorithms under "fair" condition based on frequency. Then, false alarms are identified in the detection results. The fewer false alarms identified, the better the algorithm. In this way, the comparison between two bottleneck identification algorithms in the absence of ground truth data becomes possible. However, the strategy is still not statistically sound. For more accurate evaluation, one still needs to obtain enough true bottleneck data.

*Data Quality Requirements for Bottleneck Identification*

Although partially solved the problem, data quality is still a serious issue for bottleneck identification. Different data quality issues can cause different problems. Data inconsistency can cause failure of some bottleneck identification algorithms. Even one- or two-day of unreasonable speeds can serious reduce the performance of the proposed algorithm. It is highly recommended that data cleaning and de-noising should be a crucial first step before conducting bottleneck identification using archived fixed-location data. And all bottleneck identification results should be further evaluated based on video data, driving test or operator experiences to ensure the validity of identification results.

### Future Study

There are several topics that can be further explored. First, sensitivity of the two thresholds used in the algorithm, the congestion test threshold and the frequency threshold, should be further tested. Second, we need to further test if other regression shapes of flow-occupancy diagram, for example, the bell shape can be more efficient than the proposed one. Third, the validity and effectiveness of the evaluation criteria should be further tested and investigated if true bottleneck data can be obtained. And last but not least, so far, there has not been a good evaluation framework and benchmark for bottleneck identification. This really impedes future research on this topic. This paper made some contribution towards a comprehensive evaluation framework by exploring proper evaluation strategies for bottleneck identification algorithm without ground truth data. However, more work needs to be done in order to further improve the bottleneck identification research.

**REFERENCES**


(1) Texas Transportation Institute (TTI), Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation, Prepared for FHWA, Sep. 1st, 2005. http://www.ops.fhwa.dot.gov/congestion_report/. Accessed July 2009

(2) Freeway Management and Operations Handbook, FHWA Report No. FHWA-OP-04-003, September 2003 http://ops.fhwa.dot.gov/freewaymgmt/publications/frwy_mgmt_handbook/ . Accessed May 2009

(3) Zhang, Lei; Levinson, David; Some Properties of Flows at Freeway Bottlenecks, In *Transportation Research Record: Journal of the Transportation Research Board, No. 1883,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 122–131.

(4) Kerner, Boris S.; *Chapter 7 Congested Patterns at Freeway Bottlenecks in Three-Phase Traffic Theory*, *The Physics of Traffic: Empirical Freeway Pattern Features, Engineering Applications and Theory*, Springer Berlin Heidelberg, New York, 2004

(5) Cassidy, Michael J.; Bertini, Robert L.; Some Traffic Features at Freeway Bottlenecks, Transportation Research Part B, Vol. 33, 1999, pp. 25-42

(6) Banks, James H.; Flow Processes at a Freeway Bottleneck, In *Transportation Research Record: Journal of the Transportation Research Board, No. 1287,* Transportation Research Board of the National Academies, Washington, D.C., 1990, pp. 20–28.

(7) Banks, James H.; Two-Capacity Phenomenon at Freeway Bottlenecks: A Basis for Ramp Metering, In *Transportation Research Record: Journal of the Transportation Research Board, No. 1320,* Transportation Research Board of the National Academies, Washington, D.C., 1991, pp. 83–90.

(8) Ringert, John; Urbanik, Thomas II; Study of Freeway Bottlenecks in Texas, In *Transportation Research Record: Journal of the Transportation Research Board, No. 1398,* Transportation Research Board of the National Academies, Washington, D.C., 1993, pp. 31–41.

(9) Mirshahi, M, Obenberger, J, Fuhs, C.A., Howard, C.E., Krammes, R.A., Kuhn, B.T., Mayhew, R.M, Moore, M.A., Sahebjam, K., Stone, C.J, and Yung, J.L, Active traffic Management: The Next Step in Congestion Management, Report No. FHWA-PL-07-012, Federal Highway Administrative, Department of Transportation. http://international.fhwa.dot.gov/pubs/pl07012/atm_eu07.pdf. Accessed July 2009.

(10) Chen, C; Skabardonis, A; Varaiya, P; Systematic Identification of Freeway Bottlenecks, In *Transportation Research Record: Journal of the Transportation Research Board, No. 1867*, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 46–52.

(11) Ban, Xuegang; Chu, Lianyu ; Benouar, Hamed; Bottleneck Identification and Calibration for Corridor Management Planning, In *Transportation Research Record: Journal of the Transportation Research Board, No. 1999,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 40–53.

(12) Fei, X.; Vu, Nhan N.; Liu, Ke; Li, Ling, A Fuzzy-Based System for Freeway Bottleneck Severity Diagnosis in a Sensor Network. *In Transportation Research Board 88th Annual Meeting*. CD-ROM, Transportation Research Board of the National Academies, Washington, D.C., 2009

(13) Halkias, B; Kopelias, P; Papandreou, K; Politou, A; Prevedouros, P; and Skabardonis, A; Freeway Bottleneck Simulation, Implementation, and Evaluation., In *Transportation Research Record: Journal of the Transportation Research Board, No. 2012,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 84–93.

(14) Wieczorek, Jerzy; Li, Huan; Fernández-Moctezuma, Rafael; Bertini, Robert; Integrating an Automated Bottleneck Detection Tool Into an Online Freeway Data Archive. *In Transportation Research Board 88th Annual Meeting*. CD-ROM, Transportation Research Board of the National Academies, Washington, D.C., 2009

(15) Turner, Shawn. *Quality Control Procedures for Archived Operations Traffic Data: Synthesis of Practice and Recommendations*. http://www.fhwa.dot.gov/policy/ohpi/reports/qcproc/index.htm. Accessed May 2009.

(16) Transportation Research Board (TRB), *Highway Capacity Manual*, Fourth Edition, Washington, DC.

(17) Daganzo, C.F., Fundamentals of Transportation and Traffic Operations, Elsevier Science Inc, New York. p. 259

(18) Jin, Jing; Ran, Bin; Automatic Incident Detection based on Fundamental Diagrams of Traffic Flow. *In Transportation Research Board 88th Annual Meeting*. CD-ROM, Transportation Research Board of the National Academies, Washington, D.C., 2009

(19) *Diving Conditions – Wisconsin Department of Transportation*, http://www.dot.wisconsin.gov/travel/driving-cond.htm, Accessed July, 2009.