Monitoring Freeway Congestion Using Single-Loop Measurements

Yinhai Wang
Research Associate
Department of Civil Engineering
University of Washington
Box 352700
Seattle, WA98195-2700
Tel: (206) 616-2696
Fax: (206) 543-5965
Email: yinhai@u.washington.edu

Xiaoping Zhang
Research Assistant
Department of Civil Engineering
University of Washington
Box 352700
Seattle, WA98195-2700
Tel: (206) 543-5481
Fax: (206) 543-5965
Email: xpzhang@u.washington.edu

Nancy L. Nihan
Professor
Department of Civil Engineering
University of Washington
Box 352700
Seattle, WA98195-2700
Tel: (206) 543-9639
Fax: (206) 543-5965
Email: nihan@u.washington.edu
Abstract

Traffic congestion is getting worse in large cities around the world. Two of the most effective approaches to the solution of congestion problems, Advanced Traffic Management Systems (ATMS) and Advanced Traveler Information Systems (ATIS) have been developing rapidly since the 1980s. Single-loop detectors are major data sources for such systems.

In this study, we propose an approach that takes in nine consecutive 20-second interval measurements from a single-loop detector, processes these measurements, estimates speeds for 3-minute periods, and produces congestion information based on the estimated speeds. The entire procedure includes three steps: loop data preprocessing, traffic speed estimation, and congestion detection. To automate the procedure, a Freeway Congestion Monitoring System (FCMS) has been developed in C++. The FCMS can produce both speed and congestion severity curves in real time. Site data from a loop station on Southbound I-5 were applied to demonstrate the FCMS system. The system performed consistently well under both congested and un-congested conditions. After further tests, the FCMS system may be incorporated into current ATMS and ATIS systems to provide more reliable and more intuitive traffic information to transportation professionals and road users.

Key words: single loop, filter, congestion severity index, and speed.
INTRODUCTION

With increasing economic development and urbanization, road traffic is becoming more and more congested in U.S. cities across the nation. From 1980 to 1998, yearly roadway vehicle-kilometers of travel in the U.S. increased dramatically from 2,457 billion to 4,229 billion, or 72%, while, at the same time, highway kilometers increased only 2% from 3.86 billion to 3.95 billion (US DOT, 2000). The huge gap between travel demand and infrastructure supply is the main reason for traffic congestion. According to estimates by the Texas Transportation Institute, congestion costs (in 1997 dollars) have risen from $21 billion in 1982 to $72 billion in 1997, and the increasing trend continues to be strong (USDOT, 2001). Reduction of traffic congestion has become an urgent issue for policy-makers. Some of the most effective approaches to the solution of congestion problems have come from research in Intelligent Transportation Systems (ITS) such as Advanced Traffic Management Systems (ATMS) and Advanced Traveler Information Systems (ATIS), which have been developing rapidly since the 1980s.

A common objective for both ATMS and ATIS is to detect traffic problems (such as incidents) quickly and solve these problems through appropriate responses by traffic managers and/or travelers. Real-time traffic data are, therefore, essential for ATMS and ATIS success. Inductance single-loop detectors are the major data sources for such systems (ITE, 1998). A single-loop detector, however, measures only volume (the number of vehicles passing per unit time) and lane occupancy (the fraction of some total time interval that a loop is occupied by vehicles). More desirable information, such as traffic speed, travel time, etc. must be estimated from these measured volumes and lane occupancies. Using single-loop measurements to detect congestion has been a common interest for transportation researchers and practitioners since the 1970s.

A number of congestion-detection algorithms have been developed over the past three decades, and some of them have been deployed in urban freeway systems in major cities. In general, these congestion/incident detection algorithms can be classified into four categories: comparative, time-series, combined, and artificial neural network based. Comparative algorithms establish rough incident patterns, and attempt to recognize these patterns in traffic measurements by comparing detection variables to pre-selected thresholds (Stephanedes and Chassiakos, 1993). Examples of such algorithms are described in Payne et al (1976), Levin and Krause (1978), Collins et al (1979), Persaud et al (1990), and Masters et al (1991). Time-series algorithms utilize the serial correlation of disturbances across measurement intervals to predict the trend of traffic flow or occupancy rate. When abrupt changes in the trend are detected, the occurrence of a congestion/incident is signaled. Examples of time-series algorithms are given in Dudek and Messer (1974), Cook and Cleveland (1974), and Ahmed and Cook (1982).

Stephanedes and Chassiakos (1993) combined certain features of the comparative and time-series algorithms, and developed a new algorithm commonly regarded as the Minnesota algorithm. Stephanedes et al (1992) compared the performance of the Minnesota algorithm by with several other algorithms, including comparison and time-series algorithms, and concluded that the Minnesota algorithm surpassed the others in congestion detection. Ritchie and Cheu (1995) developed an incident detection algorithm based on artificial neural networks (ANNs). Their ANNs algorithm is capable of performing a non-linear mapping between a set of distributed information processing structures that mimic the simplified operation of a human
brain. Such ANNs algorithms offer an advantage over conventional incident detection algorithms in that no mathematical model of traffic operation or the incident detection process is required, thus eliminating imperfections in model formulation.

In a more recent study, Nihan (2000) developed a simple method for identifying and/or predicting freeway congestions using single-loop detection systems. She suggested using the flow-to-occupancy ratio as a congestion indicator because, when using a constant speed estimation parameter \((g)\), the flow-to-occupancy ratio is proportional to the traffic speed and, therefore, a surrogate speed variable. This current study is an extension of Nihan’s work aimed at making the estimation results more intuitive and more suitable for ATMS-and ATIS-related applications.

In this study, we propose an approach that collects nine consecutive 20-second interval measurements from a single-loop detector at one time, processes these measurements to estimate the average speed for this 3-minute period, and produces congestion information based on the speeds estimated for each period. (In this paper, the terms “period” and “interval” are used with significant distinction. An interval is 20 seconds in length determined by the hardware of the loop detection system in Washington State. Single-loop detectors provide volume and lane-occupancy measurements for each interval. A period represents a 3-minute time duration that contains 9 intervals. The proposed algorithm produces traffic speed and congestion information for each period based on the interval measurements.) In the proposed approach, a Freeway Congestion Monitoring System (FCMS) is implemented in C++ to automate the entire procedure. The FCMS system was tested with data from a Seattle freeway loop detector station as described in the next section.

DATA

The data used for this study were collected from the loop detector station ES-163R on Southbound I-5 in Seattle. Station ES-163R is located under the 130th Street over-bridge of southbound I-5. It is equipped with dual-loop detectors (formed by two single-loop detectors several meters apart) for all five lanes of the study freeway section. As shown in Figure 1, one of the five lanes is an HOV lane and the other four are general-purpose (GP) lanes. The third general-purpose lane from the right was chosen for this study. The two single-loop detectors that form the dual loop for this lane are ES-163R: MMS___3 and ES-163R:

![Figure 1 Snapshot of southbound I-5 at 130th street.](image)
Measurements from the ES-163R: MMS__3 loop were used as inputs for the speed and congestion severity calculations. Dual-loop (ES-163R: MMS__T3) measurements were used to verify the speed estimates produced by the proposed approach.

METHODOLOGY

Since congestion detection is based on estimated speeds, reliable speed estimation is critical for successful detection. For this study, the equation adopted for period speed estimation was:

\[ \bar{s}(j) = \frac{N(j)}{T \cdot O(j) \cdot g} \]

Where:
- \( j \) = index of time period,
- \( \bar{s} \) = space-mean speed for the period;
- \( N \) = volume (vehicles per period),
- \( O \) = percentage of time loop is occupied by vehicles per interval (lane occupancy),
- \( T \) = time length per period (180 seconds or 3 minutes for this study), and
- \( g \) = the speed estimation parameter, which is equivalent to the reciprocal of the mean effective vehicle length (MEVL).

Equation (1) was derived from the fundamental flow, speed and density equation based on the assumption that the mean effective-vehicle length differences from interval to interval are negligible (see Athol, 1965, or Hall and Persaud, 1989). Normally, \( g \) is treated as a constant when using Equation (1), and this makes it quite straightforward to calculate speeds with single-loop measurements. However, since traffic composition usually changes over time, \( g \) is actually not a constant and needs to be updated periodically to adapt to the oscillation of MEVLs (Wang and Nihan, 2000). To calculate the value of \( g \), the traffic composition information for each period is needed. Unfortunately, such composition data are not directly available from single-loop detectors, and determining the \( g \) values typically involves very complicated calculations.

An alternative to updating \( g \) periodically is to filter single-loop data to maintain a relatively stable MEVL through all the periods. If the MEVLs vary narrowly, \( g \) may be treated as a constant without inducing significant errors. Compared to methods that frequently calculate \( g \) values, this method, which uses a constant \( g \) and filtered data, is both intuitive and practical. Therefore, the proposed approach filters single-loop data and then uses a constant \( g \) in Equation (1) for speed estimation with the filtered data.

The entire procedure involves three steps: loop data preprocessing, traffic speed estimation, and congestion detection. In the first step, single-loop measurements are preprocessed for traffic speed estimation. A simple yet effective filtering algorithm is applied to screen out interval measurements that may significantly drift the MEVL for the period. After preprocessing, the filtered data should contain only measurements for vehicles with nearly identical lengths. These filtered data are the inputs for speed estimation using Equation (1). The estimated period speeds are then used for congestion detection and congestion magnitude calculation in the final step. The flow chart for the algorithm is shown in Figure 2.
Figure 2 Flow chart of the proposed algorithm
Single-Loop Data Preprocessing

Features of vehicle length distribution

Dual-loop detector data were used to examine the characteristics of typical vehicle length distributions on the freeway system. We extracted single-vehicle length information from the dual-loop detector by using selected intervals for the length measurements, i.e., intervals that had only one vehicle present over the entire interval duration. From May 3 to May 16 in 1999, 4915 valid single vehicle lengths (those with nonzero error flags were excluded) were observed by the dual-loop detector (ES-163R:MMS_T3) on Southbound I-5. The frequency distribution of the observed vehicle lengths is shown in Figure 3. Two peaks are obvious in the plot: one at about 5m, representing the concentration of short vehicle (SV) lengths, and the other at about 23m, representing that for long vehicles (LVs). The fact that the first peak is much higher than the second peak indicates a good length concentration of SVs. If we choose 11.89m as the partition value for the two vehicle categories, the SV and LV categories in this study correspond to the Bin1&Bin2 and Bin3&Bin4 classes in the WSDOT (Washington State Department of Transportation) dual loop vehicle classification system, respectively (see Wang and Nihan, 2001 for details). The descriptive statistics for SVs and LVs are provided in Table 1.

Figure 3 Length Distribution of Vehicles on Southbound I-5
TABLE 1. Descriptive Statistics of SV and LV lengths

<table>
<thead>
<tr>
<th>Categories</th>
<th>Number of Cases</th>
<th>Mean</th>
<th>Std Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV (Bin1 + Bin2)</td>
<td>4443</td>
<td>5.48m</td>
<td>0.87m</td>
<td>1.83m</td>
<td>11.89m</td>
</tr>
<tr>
<td>LV (Bin3 + Bin4)</td>
<td>472</td>
<td>22.50m</td>
<td>3.59m</td>
<td>12.19m</td>
<td>30.18m</td>
</tr>
<tr>
<td>Both SV and LV</td>
<td>4915</td>
<td>7.12m</td>
<td>5.20m</td>
<td>1.83m</td>
<td>30.18m</td>
</tr>
</tbody>
</table>

The standard deviation for SV lengths is only 0.87m, or about 17% of that for both SV and LV lengths. This indicates that when intervals contain only SVs, the MEVL for each interval should be very stable considering that when \( n(i) \) vehicles present in an interval \( i \), the variance for the MEVL of the interval decreases to only one \( n(i) \)th of the 0.87m. Therefore, we can use a constant value to approximate the speed estimation parameter \( (g) \) when no LVs are present in the periods.

Filtering Algorithm

Based on the above analysis, if the traffic stream is composed of only SVs, speed estimation should be straightforward. However, traffic streams typically contain both SVs and LVs; therefore we need a filter to screen out the intervals with LVs in order to apply our speed estimation strategy. We propose a simple filtering algorithm to separate intervals with LVs from those without. The algorithm is based on the fact that lane occupancy for an interval will be proportional to the sum of vehicle lengths for vehicles detected in that interval, provided that the speed variation in the 3-minute period containing the interval is trivial. Another requirement for the calculation is that the length difference between LVs and SVs is significant enough that, in most cases, the existence of LVs in an interval can be recognized by a simple comparison of the interval occupancy rates. (Occupancy rate is defined as the measured lane occupancy divided by the interval volume, or 0 if the interval volume is zero.)

The proposed algorithm utilizes the relative relationships between the nine occupancy rates (one for each interval) to screen out intervals containing LVs. Initially, all the nine occupancy rates in any period \( j \) are sorted in ascending order as follows:

\[
0 \leq \frac{o_1(j)}{n_1(j)} \leq \frac{o_2(j)}{n_2(j)} \leq \cdots \leq \frac{o_9(j)}{n_9(j)} \tag{2}
\]

Where \( o_i(j) \) and \( n_i(j) \) denotes occupancy and volume for interval \( i \) of period \( j \), respectively.

Since intervals containing zero vehicles do not contribute anything to our analysis, we drop all such intervals and Equation (2) becomes

\[
0 < \frac{o_p(j)}{n_p(j)} \leq \frac{o_{p+1}(j)}{n_{p+1}(j)} \leq \cdots \leq \frac{o_9(j)}{n_9(j)} \quad \text{for} \quad 1 \leq p \leq 9 \tag{3}
\]

Where \( p \) is the starting index for intervals with detected vehicles (including both SVs and LVs).

Then, calculate the difference between two consecutive intervals

\[
d(i, i+1) = \frac{o_{i+1}(j)}{n_{i+1}(j)} - \frac{o_i(j)}{n_i(j)} \quad \text{for} \quad p \leq i \leq 8 \tag{4}
\]
and compare the interval with a threshold denoted by $D_i(j)$. If

$$d(i,i+1) \geq D_i(j) \quad \text{for } p \leq i \leq 8 \quad (5)$$

interval $i+1$ is identified to contain LVs, all the remaining intervals for the period (from interval $i+1$ to 9) must be discarded. Otherwise, the interval measurements are kept and the next interval is tested for LVs, and so on until all remaining intervals have been tested (or discarded).

Choice of a proper threshold value is critical for this filter. Since an LV’s contribution to the occupancy rate depends on interval volume and the volume changes from interval to interval, a constant threshold value is obviously inappropriate. Equation (6) was adopted for calculating the threshold for each interval.

$$D_i(j) = \beta \cdot \frac{\bar{l}_lv - \bar{l}_sv}{S \cdot T \cdot n_{sv}(j)} \quad (6)$$

Where
- $\bar{l}_sv$ = observed mean vehicle length for SVs (5.48m for this study);
- $\bar{l}_lv$ = observed mean vehicle length for LVs (22.50m for this study);
- $T$ = interval length (20 second for this study);
- $S$ = free flow speed;
- $\beta$ = adjustment factor.

By choosing different $\beta$ values, the algorithm’s threshold can be adjusted, to control the tightness of the filtering. Due to differences in loop detector sensitivities, the same vehicle passing over different loop detectors at the same speed may be measured differently. Adjustments to the $\beta$ value can also control or eliminate the effects of such detector sensitivities.

If no interval contains detected LVs, all nine-interval measurements are kept for the period’s speed estimation.

### Speed Estimation

We use a constant $g$ value in Equation (1) for speed estimation. The specific value of $g$ can be calculated using Equation (7).

$$g = \frac{1}{\bar{l}_sv + l_{loop}} \quad (7)$$

Where $l_{loop}$ is the length of the single-loop detector. In our case, the length of a single-loop detector is 1.83m, and the mean SV length is 5.48m. This corresponds a $g$ value of 0.137m$^{-1}$.

If interval $q$ is identified to have LVs, then only the measurements for the $r$ (where $r = q-p$) intervals, from interval $p$ to interval $q-1$, is used for speed estimation. These $r$ sets of measurements are used to calculate the total applicable volume and corresponding occupancy as follows,

$$N_{sv}(j) = \sum_{i=p}^{q-1} n(i, j) \quad (8)$$

$$O_{sv}(j) = \sum_{i=p}^{q-1} o(i, j) \quad (9)$$

Since only the measurements corresponding to SVs are applied, Equation (1) can be rewritten as
Since all variables on the right side of Equation (10) are known, space-mean speed for each period can be easily calculated.

**Congestion Detection**

Traffic congestion is a direct reflection of insufficient roadway capacity. Three key aspects of congestion are its severity, extent, and duration (USDOT, 2001). The severity of congestion refers to the magnitude of the problem as measured by the average overall travel speed, travel time delay, or the length of queues behind bottlenecks. The extent of congestion reflects the portion of the population or portion of total travel affected. The duration of congestion is the length of time that the traffic stream is congested. Our proposed algorithm can provide information on two of these three key aspects: severity and duration.

Typically, congestion severity is expressed by a volume-to-capacity ratio. The higher the ratio, the more congested the facility. This volume-to-capacity ratio works well for roadway sections with recurrent congestion. For incident-induced congestion, however, the reduced capacity is not immediately known and the ratio may not be calculable.

To avoid this problem, this paper uses an index to reflect the congestion severity. The congestion severity index, \( C(j) \), is defined as

\[
C(j) = \frac{S - \bar{s}(j)}{S} \quad \text{for } \bar{s}(j) \leq S
\]

Our congestion detection algorithm performs three tasks: detection of the onset and dissipation of congestion, estimation of congestion severity and calculation of congestion duration time. The algorithm begins by checking a boolean variable called the congestion flag. This variable records whether the monitored freeway section was under congestion during the last period. If it was not congested during the last period, the algorithm detects whether congestion is beginning during the current period. If the answer is yes, the congestion flag is set to “true”, the congestion onset time \( (t_o) \) is recorded, and the congestion severity index is calculated. If the onset of congestion is not detected during the current period, the value of the congestion severity index is set to 0. If the congestion flag indicates congestion during the last period, the algorithm checks whether congestion dissipates during the current period. If not, it calculates the congestion severity index. Otherwise, it resets the congestion status flag to “false”, records the congestion dissipation time \( (t_d) \) and calculates the congestion duration time \( (t_d - t_o) \).

Correctly detecting the onset of congestion is very important to the algorithm’s success. Considering that, when congestion starts, traffic speed drops significantly in at least two consecutive periods, we use the following criteria for congestion onset detection:

1. Traffic speed drops in two consecutive periods, i.e. \( \bar{s}(j) < \bar{s}(j-1) < \bar{s}(j-2) \);
2. Arithmetic mean speed of three consecutive periods is less than 90% of the free flow speed, i.e. \( \frac{1}{3}(\bar{s}(j)+\bar{s}(j-1)+\bar{s}(j-2)) < 0.9 \cdot S \).
After congestion starts, traffic speed drops to a much lower level than before. Even if speed oscillates significantly, its peak values are still well below the free flow speed in most cases. Therefore, we simply use the free flow speed as a threshold value to detect congestion dissipation. If

\[ \bar{s}(j) \geq S \]  

then congestion has dissipated.

**APPLICATION**

The proposed algorithm has been implemented in C++ in the FCMS system developed by the authors. If real-time single-loop data are provided, the FCMS system can provide real-time speed estimation and congestion information.

On weekdays, Station ES-163R suffers recurrent congestion during the morning peak hours. For the current study, we used data collected by ES-163R: MMS__3 on two consecutive weekdays, May 13 (Thursday) and May 14 (Friday), 1999, to demonstrate the FCMS system. Our test data contained 8640 interval-measurement sets (corresponding to 48 hours).

To setup the system, one must identify the free flow speed \( S \) and the adjustment factor \( \beta \) for each specific station. The mean length of SVs (5.48m) and mean length of LVs (22.50m) can be considered as constants for all I-5 stations in the Seattle area. Also, the length of every single-loop detector on the WSDOT loop detection system is 6 ft (1.83m). The speed-limit posted at Station ES-163R was 60 mph (96.54 km/h), but 63 mph (101.37 km/h) was chosen as the free flow speed for this study station since people actually drive slightly faster than the speed limit. The adjustment factor was chosen to be 0.38 by trial and error.

With the selected free flow speed and adjustment factor, speeds for each period and congestion related information such as the severity index and duration time was calculated. The estimated speeds versus the speeds observed by the dual-loops are plotted in Figure 4. We can see that the estimated speeds and observed speeds synchronize reasonably well. The correlation coefficient between the estimated speeds and the observed speeds is 0.80. To further evaluate the speed estimation result, we define a statistical variable, estimation error, as follows

\[ \varepsilon(j) = \text{estimated speed for period } j - \text{observed speed for period } j \]  

The estimation errors range from –30.15 to 29.86 km/h. The standard deviation for \( \varepsilon(j) \) is 7.06 km/h. The expected value for \( \varepsilon(j) \) is 0.51 km/h, indicating the estimated speed is slightly higher than the observed speed. Considering the dynamic feature of traffic movement and the possible noises in single loop measurements, this result is quite favorable.
Using the estimated speeds, congestion related information is calculated. For the two-day study period, congestion occurred three times, once on the first day and twice during the second day. A comparison between the estimated and dual-loop observed congestion data is shown in Table 2. For congestion period 1 and congestion period 3, the proposed approach provided very good estimates of congestion onset, dissipation, and duration times. However, the estimated duration of congestion period 2 was much shorter than that observed by the dual loops. This was largely due to the light severity of congestion for that period. When congestion is not serious, the oscillation of the measured occupancy rates may cause the estimated speeds to reach the dissipation threshold \( (D_{i,j}) \) and therefore generate false alarms. This problem becomes trivial as congestion severity increases.

**TABLE 2. Comparison between Observed and Detected Congestion Data**

<table>
<thead>
<tr>
<th>Congestion No.</th>
<th>Onset time</th>
<th>Dissipation time</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Observed</td>
<td>6:39</td>
<td>9:00</td>
</tr>
<tr>
<td></td>
<td>Estimated</td>
<td>6:42</td>
<td>9:00</td>
</tr>
<tr>
<td></td>
<td>Estimated - Observed</td>
<td>0:03</td>
<td>9:00</td>
</tr>
<tr>
<td>2</td>
<td>Observed</td>
<td>6:21</td>
<td>7:42</td>
</tr>
<tr>
<td></td>
<td>Estimated</td>
<td>6:30</td>
<td>7:36</td>
</tr>
<tr>
<td></td>
<td>Estimated - Observed</td>
<td>0:09</td>
<td>0:06</td>
</tr>
<tr>
<td>3</td>
<td>Observed</td>
<td>16:06</td>
<td>16:51</td>
</tr>
<tr>
<td></td>
<td>Estimated</td>
<td>16:06</td>
<td>16:51</td>
</tr>
<tr>
<td></td>
<td>Estimated - Observed</td>
<td>0:00</td>
<td>0:00</td>
</tr>
</tbody>
</table>
The calculated values for the congestion severity index are plotted in Figure 5. The goodness of fit should be the same as that for the speed estimates, since this is simply another expression of the estimated speeds. This type of congestion severity curve intuitively illustrates how congestion has changed over time.

![FIGURE 5 Congestion severity curve](image)

**SUMMARY**

This paper describes a new traffic congestion monitoring approach using single-loop data. Instead of using lane occupancy rates for congestion detection, thereby producing only a few abstract statistics, the proposed approach determines whether a freeway section is congested based on estimated speeds obtained from single-loop measurements. The proposed algorithm continually calculates a congestion severity index for the identified freeway section. The proposed approach involves three steps: loop data preprocessing, traffic speed estimation, and congestion detection. To automate the procedure, a Freeway Congestion Monitoring System (FCMS) was developed in C++. The FCMS produces both speed and congestion severity curves in real time, thus making the congestion monitoring process more intuitive. The FCMS system was tested with loop data from station ES-163R on Southbound I-5. The system performed consistently well under both congested and un-congested conditions.

More tests are needed to evaluate the transferability and improve the performance of the proposed approach. After further tests, the FCMS system may be incorporated into current ATMS and ATIS systems to provide more reliable and more intuitive traffic congestion information to transportation professionals and road users.

**REFERENCES**


