Techniques for Validating an Automatic Bottleneck Detection Tool Using Archived Freeway Sensor Data

Jerzy Wieczorek, Rafael J. Fernández-Moctezuma, and Robert L. Bertini

Bottlenecks are key features of freeway systems. Their effects in performance and emissions are of increasing importance as congestion worsens in urban areas. In the United States, FHWA has been working to identify and monitor key bottlenecks in each state. In Oregon, a freeway data archive known as the Portland Oregon Regional Transportation Archive Listing archives measured count, density, and speed data from more than 600 locations at 20-s intervals. This archive has enabled development of online freeway performance and reliability analysis tools. This paper describes the rigorous evaluation and refinement of an automated tool for identifying recurrent freeway bottlenecks using historical data within the framework of the data archive. Efforts have focused on identification and display of active bottleneck features by using graphical tools and the selection of optimal variables that enabled careful identification of active bottlenecks. This research aims to detect bottleneck activation historically and, through future work, in real time as well. Ultimately, the results of this research will enhance the prioritization of improvements and implementation of operational strategies on the freeway network.

Understanding traffic dynamics at freeway bottlenecks is a foundation for understanding how the freeway system operates. A bottleneck is defined as a point upstream of which one finds a queue and downstream of which one finds freely flowing traffic. Bottlenecks can be static (e.g., a tunnel entrance) or dynamic (e.g., an incident or a slow-moving vehicle). Bottlenecks are said to be active when they impose a restriction on the flow. They can be activated or deactivated due to a decrease in demand or spillover from a downstream bottleneck (*1*).

Great costs imposed by congestion on the movement of people and freight motivate the development of new freeway performance measures and reporting systems. To improve transportation planning, management, and operation, the Portland Oregon Regional Transportation Archive Listing (PORTAL) has been established to collect count, occupancy, and speed data from more than 600 locations at 20-s intervals. Performance measurement tools are available for generating reports on travel time, delay, vehicle miles traveled, and vehicle hours traveled. Also available are statistical tools that describe travel reliability across days, weeks, months, and years. With the availability of this rich source of archived data, new analytical tools can be developed for use with historical data and in a real-time environment informed by past performance. The objective of this paper is to describe the techniques used for rigorous validation and refinement of an automated system to identify freeway bottlenecks within the PORTAL environment. Using ground truth knowledge of when and where bottlenecks occurred during a substantial sample period, previous work (*2*) developed and tested a working prototype with the intent to accurately identify, track, and display active bottleneck features using graphical tools. This paper presents an extended analysis of that prototype, and contains new results and applications:

- Performance analysis of the automated bottleneck detection tool, based on more extensive ground truth, confirms earlier results with greater certainty.
- The use of lane-by-lane analysis, instead of pooling multiple lanes, is evaluated with negative results.
- Fundamental diagrams are discussed as a means for developing improvements to the tool.
- The automated bottleneck detection technique is applied toward detecting shockwave speeds and visualizing historical patterns in the archive data.

BACKGROUND

Past research has sought greater understanding of where freeway bottlenecks formed and how and when they were activated. One method uses oblique curves of cumulative vehicle arrival number versus time and cumulative occupancy (or speed) versus time constructed from data measured at neighboring freeway loop detectors (*3–7*). Using this method, it is possible to observe transitions between freely flowing and queued conditions and to identify important features over time and space. This method, which requires visual inspection, has been applied here as ground truth to systematically define bottleneck activations and deactivations.

Previous research has also developed techniques for automated bottleneck activation and deactivation identification by comparing measured traffic parameters and applying thresholds for those values. Notably, Chen et al. (*8*) developed an algorithm to identify bottleneck locations and their activation and deactivation times using freeway loop detector data from San Diego, California, focusing on speed differences between consecutive detectors. Zhang and Levinson (*9*) implemented a similar system to identify bottlenecks

J. Wieczorek, Department of Mathematics and Statistics; R. J. Fernández-Moctezuma, Department of Computer Science; and R. L. Bertini, Department of Civil and Environmental Engineering, Portland State University, P.O. Box 751, Portland, OR 97207-0751. Corresponding author: J. Wieczorek, jerzy@pdx.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2160, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 87–95.
DOI: 10.3141/2160-10

using occupancy differentials. Banks (*10*) used 30-s speed drop thresholds to identify bottleneck activation in San Diego, and Hall and Agyemang-Duah (*11*) developed a threshold using the ratio of 30-s occupancy divided by flow on a Canadian freeway. Bertini (*12*) tested other signals including the variance of 30-s count as characterizing bottleneck activation at several sites. Building on past research, the objective of this paper is to evaluate the Chen et al. method, using a sensitivity analysis approach, with data from a freeway in Portland, Oregon. The performance of the automated method is judged by comparing its output to the ground truth found by the oblique-curve method (*3–7*).

The research mentioned above assumes a two-phase traffic theory, where traffic states are either free flow or congested. Several other approaches to identifying bottlenecks and congestion are based on Kerner's three-phase traffic theory, notably the ASDA–FOTO model (*13*), which separates traffic into free flow, synchronized flow, and wide-moving jams. Our proposed automated detection tool and our ground truth approximation method are both based on two-phase traffic theory. Although the different assumptions between two- and three-phase traffic models mean that these approaches are not directly compatible, approaches in Palmer et al. (*14*) have been compared and contrasted.

EXTENDING PORTAL'S FUNCTIONALITY

In addition to providing loop detector data at different aggregation levels, the PORTAL archive includes automatic data analysis tools such as oblique plots, travel time analysis, congestion reports, and information layers in time–space surface plots with incident and variable message sign information (*15*). In response to an FHWA initiative of identifying bottleneck locations, a new PORTAL module is being developed to provide bottleneck identification and analysis capabilities. One of PORTAL's main features is the use of surface plots to represent the time–space evolution of a particular measurement. One goal for the bottleneck identification module is to visually represent the bottleneck activation and deactivation times, as shown hypothetically in Figure 1. Bottleneck activations are marked and the average propagation speed of the shockwave is indicated as well. This tool will provide the most relevant information to the user: activation time, duration, deactivation time, location, and shock propagation speed.

A further objective of this research is to incorporate historical data into live data displays. Figure 1 also illustrates how previously known bottleneck locations can be displayed on the time–space diagram in a real-time environment. While bottlenecks A, B, and C were detected from the current day's data and mapped on the plot, the bounding boxes illustrate durations and locations that occurred 90% of the time over a previous period of time. These boundaries can be learned by analyzing previously archived data. Bottleneck detection thresholds may lead to false positives or false negatives. In this case, displaying the additional layer of historical bottleneck locations supplements the current analysis output with historical knowledge. Such historical knowledge of when and where to expect bottlenecks may also improve forecasting of propagation speeds as new data arrives, although the current paper's method has not yet been applied to forecasting. Bottleneck information could also be displayed in combination with incident information, providing a rich tool for users who want to distinguish recurrent congestion from incident-related congestion.

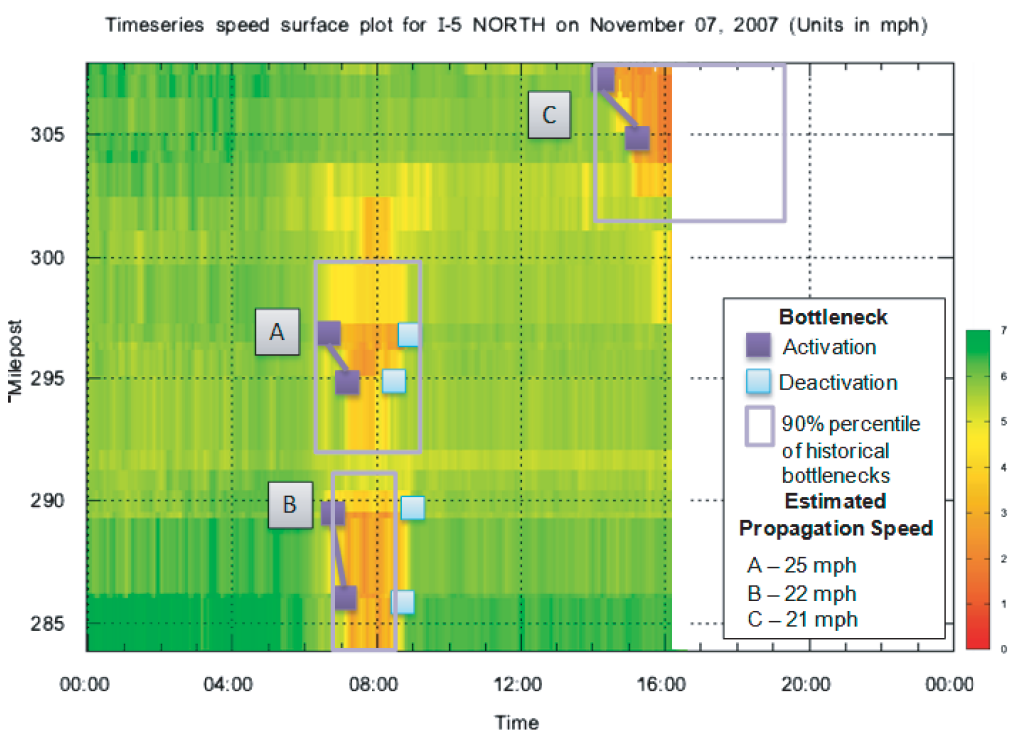


FIGURE 1 Mock-up of desired output of incoming data analysis in combination with historically learned parameters, such as expected bottleneck locations and activations. [Source: Oregon Department of Transportation.]

ANALYTICAL FRAMEWORK

Traffic theory describes six types of shock waves, as shown in Figure 2 (*16*), and bottlenecks can form and evolve in many ways. Many Portland freeway corridors contain recurrent bottlenecks that form with front-stationary shockwaves coupled with backward-forming queues. Therefore, this analysis begins with identifying front-stationary bottleneck locations and activation and deactivation times, followed by tracking backward-forming queue propagation and mapping the congested conditions in time and space.

To identify a bottleneck, Chen et al.'s automated algorithm (*8*), described in further detail below, focuses on consecutive detector pairs along a freeway corridor and searches for bottlenecks (i.e., places with queuing upstream and free-flow downstream). This automated method uses speed data from loop detectors and involves three parameters: the data aggregation time interval and two speed thresholds. Chen et al. arbitrarily selected parameter values that appear to work well on San Diego freeway speed data. However, the optimal values of these parameters may be different in other cities with different freeway network configurations, geometry, weather, pavement, traffic conditions, and the nature of the available data. For example, the Portland freeway network contains several locations where one freeway "tees" into another, resulting in recurrent congestion. In this paper, the speed thresholds and data aggregation levels comprising the Chen method were tested for bottleneck identification in Portland using PORTAL data. Ground truth locations and activation or deactivation times for 91 bottlenecks over 24 days, identified using the method described in Cassidy and Bertini (*4, 5*), were used as baseline reference points for evaluating the outcomes of the Chen method using a range of parameters. Thus, the current paper's contributions include a rigorous evaluation of Chen's approach, parameter selection guidelines for other cities (as well as specific parameter values optimized for Portland), assessment of potential refinements to the automated algorithm (lane-by-lane analysis and/or incorporating additional inputs such as flow or occupancy), and description of the algorithm's applications (shockwave speed measurement and historical bottleneck visualization).

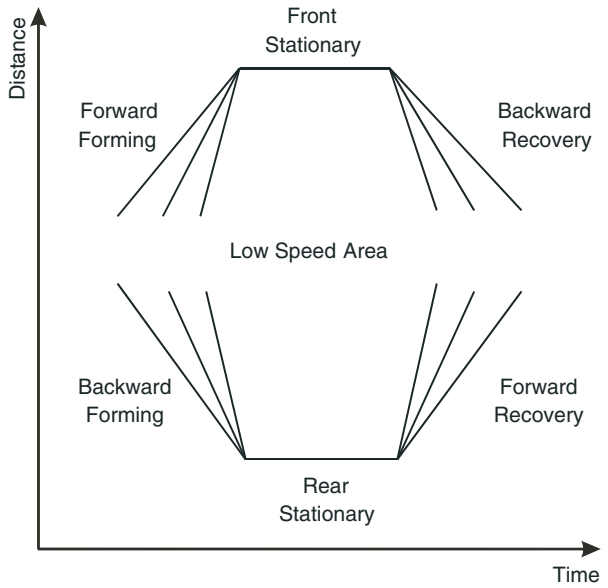


FIGURE 2 Shock wave types.

SITE DESCRIPTION AND DATA PREPARATION

The implementation of the Chen method was tested with archived data from the northbound Interstate 5 (I-5) corridor in the Portland, Oregon, metropolitan region, shown in Figure 3. There are 22 detector stations along this 22-mi (35-km) corridor. However, two of them (mileposts 293.74 and 296.6) consistently suffer from poor data quality, so they were ignored for the purposes of this research. The remaining 20 detectors have an average spacing of approximately 1.15 mi (1.85 km). Detectors are located just upstream of metered freeway on-ramps.

In previous work on this corridor (*2*), only 5 days of ground truth testing were chosen for an initial algorithm evaluation and parameter optimization. Furthermore, only algorithm performance at detecting the front-stationary bottleneck itself was evaluated, rather than the entire queue propagating upstream of the bottleneck.

With an updated algorithm capable of tracking the boundaries of the upstream queue as well, the current analysis encompasses a larger sample of ground truth days, leading to greater confidence in the results. To evaluate the method against ground truth on Portland data, a sample was chosen of 24 representative, high-data-quality, midweek nonholiday days between February and December 2008.

Data were extracted from PORTAL at the lowest available resolution of 20 s between 5:00 a.m. and 10:00 p.m. for each day, including count, occupancy, and time mean speed in each lane. Roll-forward imputation was used to impute missing data and abnormally low

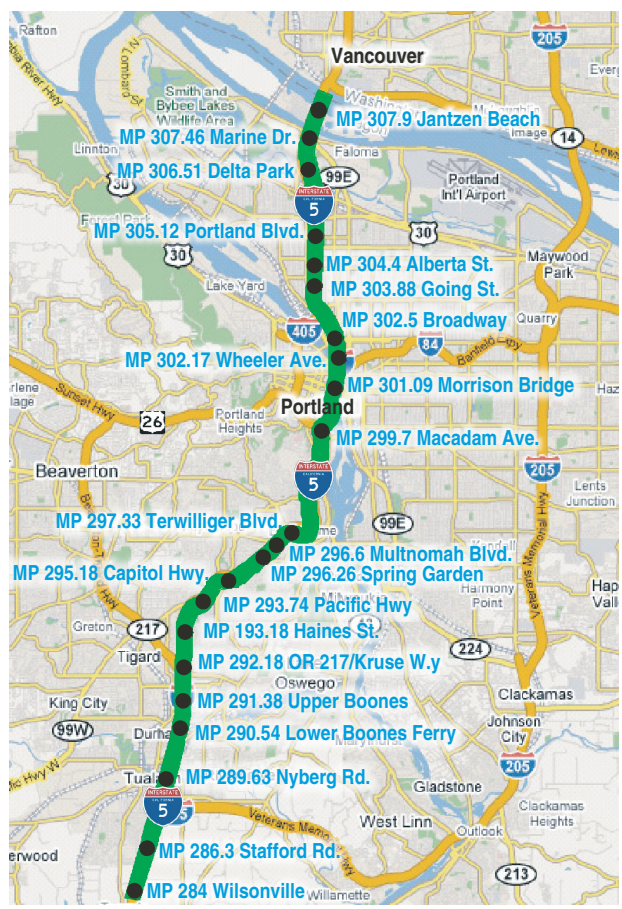


FIGURE 3 Northbound Interstate 5 corridor in Portland region.

speeds. This facilitated aggregation into the 1-, 3-, 5- and 15-min data sets. For each station, a volume weighted mean speed for 1-min aggregation was compared with arithmetic averaging; the results showed that over 95% of the differences between the two averages were smaller than 5 mph (8 km/h) and thus of no practical significance. Thus, for each station, measurements from multiple lanes were aggregated into a single quantity using simple arithmetic averaging, and mean speeds by station were calculated at further levels of aggregation (1, 3, 5, and 15 min) also by using arithmetic averages. The obtained value was extrapolated across a region of influence of a sensor station (i.e., the segment of the corridor between the current sensor station location and the next downstream sensor location).

EXPERIMENTAL DESIGN

Baseline Bottleneck Analysis

To establish a baseline to act as ground truth for this experiment, the activation and deactivation times of each candidate bottleneck were carefully diagnosed and verified using oblique curves of cumulative vehicle arrival number versus time and cumulative occupancy (or speed) versus time constructed from data measured at neighboring freeway loop detectors. This manual procedure was used to carefully confirm queue formation and propagation, such that active bottlenecks separated freely flowing traffic from queued traffic. Previous research describes the procedures in detail, particularly those described in Horowitz and Bertini (*7*) for the detector configuration in Portland (only upstream of on-ramps) and other sources (*3–6, 12*), as well as for the current project in Bertini et al. (*17*). With this visual tool, the prevailing flows and speeds measured at the particular station are seen as slopes of the oblique cumulative plot. This method filters out noise without altering the times at which flow and speed changes occurred.

The above procedure was performed for all 91 candidate bottleneck activations and the associated bottleneck-related congestion. This heuristic method provides a solid baseline for the ground truth, but requires visual inspection by the user, hence the search for an automated algorithm close to this method in accuracy. The performance of our automated bottleneck detection system (described below) was judged by how well its results corresponded to the visual method's ground truth.

Automated Detection Algorithm

To test the Chen et al. (*8*) method on the Portland data set from PORTAL, the method for identifying the locations of bottlenecks was implemented in the MATLAB programming environment. The Chen method compares each pair of longitudinally adjacent detectors at each 5-min time point and declares that there is an active bottleneck between them when

- Speed at the upstream detector is below the maximum upstream speed threshold and
- Difference in the speeds at the upstream and downstream detectors is above the minimum speed differential threshold.

Given that a bottleneck separates free-flow (downstream) from congested (upstream) traffic states, Chen et al. chose reasonable values of MaxUpstreamSpeed = 40 mph (64 km/h), MinSpeedDifferential = 20 mph (32 km/h), with data aggregated at 5-min intervals. Thus, if the upstream speed is less than 40 mph and the downstream speed is more than 20 mph greater than the upstream speed, the Chen method identifies a bottleneck between those detectors during that interval.

However, Chen et al. recognized that each of the three parameters (maximum upstream speed, minimum speed differential, and aggregation level) may need to be adjusted for the algorithm to work optimally in new situations. An analysis for the extended algorithm has been conducted that tracks not only bottlenecks but also the resultant congestion found directly upstream of bottlenecks. Five values each of MaxUpstreamSpeed, MinSpeedDifferential, and aggregation level were tested, in a total of 125 distinct combinations. Analyzing each of the 24 sample days, MaxUpstreamSpeed was allowed to range from 30 mph (48 km/h) to 50 mph (80 km/h) in 5 mph (8 km/h) increments; MinSpeedDifferential was varied from 10 mph (16 km/h) to 30 mph in 5 mph increments; and data were aggregated at 20-sec, 1-min, 3-min, 5-min, and 15-min aggregation levels.

Also following Chen et al., a "sustained bottlenecks" filter was added that smoothes the results of the base algorithm. Even after data cleaning and aggregation, there remains noise in the data that can lead to false positives (identifying a bottleneck where none exists) or false negatives (failing to find a true bottleneck). This filter is intended to discard transient false positives that are isolated in the time dimension from other bottleneck detection instances at the same detector. Similarly, it "fills in" small gaps in bottleneck detection at a given detector, under the assumption that they are false negatives. For instance, if several consecutive points in time at a given detector are classified as bottlenecks, but one point in the middle of these others fails to register as a bottleneck, the sustained bottleneck filter fills in this gap, assuming that it is unlikely for a bottleneck to deactivate only for a few minutes during heavy congestion.

After the original bottleneck detections are performed, this filter runs by scanning each detector one at a time and checking to see whether each detection is part of a sustained, active bottleneck. If, within the span of seven consecutive time periods, fewer than five of those periods include bottleneck detections, the individual detections in that period are assumed to be false positives and are deleted. However, if five or more of the seven consecutive time periods registered as active bottlenecks, then any gaps in between them are also reregistered as bottleneck detections. This filter smoothes the data as described, improving overall accuracy. Technically, Chen et al.'s default values of 5 and 7 are additional parameters and could undergo optimization as well. However, the authors felt that this would unnecessarily complicate the multivariate analysis that was performed.

The analysis so far has focused on automatically identifying the location and activation times of front-stationary bottlenecks. In the next step after the "sustained bottlenecks" filter, the algorithm examines the duration and location of the backward-forming congestion that builds up upstream of the bottlenecks. To locate this congested area on the speed contour plot, our algorithm steps to the next detector upstream of each active bottleneck and determines whether its measured speed is also below the maximum upstream speed threshold. If this is true, the algorithm continues stepwise further upstream until a detector is reached with a speed greater than the threshold. At this point in time and space, the detector is considered to be free of congestion, but all the detectors downstream toward the bottleneck are labeled as part of the congested regime for that time period. It is assumed that such a technique will only find congestion that is directly caused by an active bottleneck. Thus, the

total congested region on the time–space plane for the corridor under scrutiny can be combined with measured flow information to estimate the total delay caused by a given bottleneck. During historical analysis of a recurrent bottleneck, its consequent delay in vehicle hours could be translated into the financial cost of wasted time and fuel, environmental cost of pollutant and carbon emissions, etc., caused by this bottleneck, which could help transportation planners to prioritize their congestion reduction efforts.

RESULTS

Assessment Indices

Our tests compared the ground truth bottleneck detections against the bottleneck instances captured by the automated method under each combination of aggregation level and thresholds for the speed data. With this sensitivity analysis on the automated method's parameters, the authors hoped to achieve a high similarity between the ground truth bottlenecks and the automated bottleneck detections, which would indicate that the automated method is successfully detecting most of the real bottlenecks and few false alarms. Our assessments are measured in terms of several common quality indices and "summary scores."

Previous research has used similar indices for evaluating the accuracy of a congestion reporting system as compared with ground truth data (*17, 18*). In particular, the detection rate and false alarm rate are considered as functions of the input parameter values (and of the given day's data set). These indices were chosen to be compatible with Bogenberger et al. (*18*), but similar indices are commonly used in many other settings such as information retrieval or medical diagnostic test evaluation. In such settings, the detection rate may be called recall or sensitivity. "Positive predictive value" or "precision" is defined as 1 minus the false alarm rate.

For each point in the time–space plane, our automated algorithm decides whether or not there was active bottleneck-related congestion (at that detector, at that time). Also, for each point in this plane, the authors know according to ground truth whether the point was truly congested or not. In other words, all the evaluated points can be partitioned into true positives (correctly labeled "congested"), false positives (labeled "congested" but are actually free flow), true negatives (correctly labeled "free flow"), and false negatives (labeled "free flow" but are actually congested). Then the detection rate is the proportion of truly congested points successfully captured by the automated method. The false alarm rate is the proportion of points "captured" by the automated method that are not real bottlenecks. TruePos and FalseNeg are used as the true positive and false negative quantities, respectively.

$$\text{DetectionRate} = \frac{\text{TruePos}}{(\text{TruePos} + \text{FalseNeg})}$$

$$\text{FalseAlarmRate} = \frac{\text{FalsePos}}{(\text{TruePos} + \text{FalsePos})}$$

Unfortunately, the parameter values that produce a high DetectionRate also tend to raise the FalseAlarmRate. For instance, increasing the maximum upstream speed and decreasing the minimum speed differential tends to "loosen" the requirements on labeling a point as congested, so it increases the DetectionRate since the algorithm is less likely to miss true bottleneck instances. However, it also means the algorithm is more likely to detect false alarms and thus the FalseAlarmRate also increases, which is undesirable. Meanwhile, at low or no data aggregation (20-s or 1-min data) the noise in the data is too high and leads to high FalseAlarmRates, whereas at a high aggregation (15-min data) the data are so smoothed that the algorithm fails to find the real bottlenecks and consequently has a low DetectionRate. Because optimizing DetectionRate and FalseAlarmRate separately results in incompatible recommendations, additional summary scores are necessary for evaluating the effects of parameter values.

The denominators of the DetectionRate and FalseAlarmRate are different, so one should not naively add or subtract them to obtain a single score. However, if the denominators are expected to be of similar size (as they should be for good values of the parameters: the total number of detections ought to be close to the total number of truly congested points), and if the "cost" of a missed detection is comparable to the "cost" of a false alarm, then the sum of DetectionRate + (1 – FalseAlarmRate) can arguably be used as a reasonable summary score for evaluating algorithm effectiveness. When these assumptions are not valid, it can still be reasonable to combine the two indices into a summary score by multiplying them. A final simple summary score is Accuracy: out of all the points evaluated, how many were correctly labeled?

$$\text{SumScore} = \text{DetectionRate} + (1 - \text{FalseAlarmRate})$$

$$\text{ProductScore} = \text{DetectionRate} * (1 - \text{FalseAlarmRate})$$

$$\text{accuracy} = \frac{(\text{TruePos} + \text{TrueNeg})}{(\text{TruePos} + \text{FalsePos} + \text{TrueNeg} + \text{FalseNeg})}$$

High SumScore, ProductScore, and accuracy values would indicate that our automated algorithm is performing well, with high DetectionRate and low FalseAlarmRate. Using different summary scores helps to judge the results: if all three agree, then a single result is best, but otherwise there are clearly tradeoffs to be made.

Statistical Analysis of Results

A statistical analysis of variance was performed independently for each of the three score functions (SumScore, ProductScore, and Accuracy), treating our results as the output of a three-factor, five-level full factorial experiment with 24 replications (one for each sample day). The terms of the analysis of variance (ANOVA) model included all three main effects, all three two-way interactions, and the three-way interaction. For example, the ANOVA table for the values of the SumScore score function is reproduced in Table 1.

The variation in the scores was partitioned differently for each score function, but in each case the ANOVA results showed that each of the main effects and at least one of the two-way interactions were highly significant. In other words, changing two factors at once does not necessarily have the same effect as changing each factor separately one at a time. Thus, it is misleading to select the factor values that independently have the most desirable main effects; instead, it is necessary to compare all possible sets of parameter settings. SumScore had the simplest analysis: since only one two-way interaction was significant (aggregation × minimum speed differential), the parameter that is not involved in the interaction (maximum upstream speed) could be optimized at 35 mph and then the remaining two-way interaction can be viewed on a simple graph, shown

TABLE 1 Analysis of Variance for SumScore

Source	DF	SS	MS	*F*	*P*
Day	23	62.45	2.72	385.18	0.000
Aggregation	4	1.23	0.31	43.75	0.000
Min. speed difference	4	2.67	0.67	94.84	0.000
Max. upstream speed	4	7.05	1.76	250.14	0.000
Agg × MinDiff	16	0.34	0.02	3.04	0.000
Agg × MaxUpstream	16	0.10	0.01	0.91	0.562
MinDiff × MaxUpstream	16	0.08	0.01	0.74	0.756
Agg × MinDiff × MaxUpstream	64	0.24	0.00	0.54	0.999
Error	2,852	20.10	0.01		
Total	2,999	94.28			

NOTE: $S = 0.08396$; $R^2 = 78.7\%$; $R^2(\text{adj}) = 77.6\%$.

in Figure 4*a*. This figure makes it clear that the highest average SumScore at 35 mph (56 km/h) MaxUpstreamSpeed occurs for 3-min Aggregation with essentially a tie between 10 mph (16 km/h) and 15 mph (24 km/h) MinSpeedDifferential. Since the other two summary scores involved several significant two-way interactions, their analysis requires the full interaction plot matrices, shown for ProductScore in Figure 4*b* and for accuracy in Figure 4*c*. (In those subfigures, the top-left-hand interaction plot is comparable to Figure 4*a*.) These subfigures seem to show that 35 mph MaxUpstreamSpeed, 15 mph MinSpeedDifferential, and 3-min Aggregation is also close to optimal, but it is not as clear.

As it turned out, each score function was optimized at a slightly different set of the three parameters, but in each case the second-best setting was the same as in our earlier work (*2*): 3-min data aggregation level, 15 mph minimum speed differential, and 35 mph maximum upstream speed. Statistical pairwise-comparison in each of the three ANOVAs showed that the best and second-best settings were never significantly different: for the sake of agreement among the three score functions and for continuity with a previous paper, the authors saw no significant reason to change the reported optimal settings. For each index, the differences between the optimal and second-to-optimal score were no greater than 0.0005.

Hence, the original Chen et al. settings used for the San Diego data (20 mph minimum speed differential, 40 mph maximum upstream speed, and 5-min aggregation) are close to, but not the same as, the optimal settings for this Portland freeway [15 mph (24 km/h) differential, 35 mph (56 km/h) maximum upstream speed, and 3-min aggregation]. This kind of discrepancy between cities is unsurprising when using such parametric methods. Researchers and transportation operations analysts in other cities wishing to implement a system using the Chen et al. algorithm should perform a similar analysis to configure the optimal parameter settings for their own network.

Lane-By-Lane Analysis

A further extension of the bottleneck detection algorithm was attempted, and used the data from individual lanes in the freeway corridor. As described above, measurements from multiple lanes had so far been aggregated using simple arithmetic averaging (henceforth referred to as the "all-lanes-pooled" algorithm). This can obscure the true traffic pattern when the lanes are not all equally congested. Thus, the authors tested the process of analyzing each lane separately and then having the lanes "vote" on whether or not bottleneck-related congestion is active. It was thought that lane-by-lane analysis may improve the algorithm's sensitivity and lead to greater agreement with the ground truth.

The authors decided to test four voting schemes that seemed likely to improve performance. At each moment at each detector, the roadway is labeled as congested if lane-by-lane analysis detects congestion in:

1. At least one of the lanes,
2. At least two of the three lanes (or at least one lane if there are only two lanes),
3. At least two of the three lanes (or both lanes if there are only two lanes), and
4. All three lanes.

For each of the 24 ground truth days, the authors analyzed the day's data using each lane-by-lane voting method. As expected, voting Method A tended to label more points as congested, so it tended to increase the DetectionRate but also the FalseAlarmRate. Method D had the opposite effect. Methods B and C are two variants of the same rule, differing only for the detectors where the freeway is only two lanes wide. Neither of these two voting methods tended to affect the DetectionRate or FalseAlarmRate much.

As it turned out, none of the methods led to meaningful improvements. Only Method A had a consistently positive change in DetectionRate, but the maximum increase for a given day was only 4 percentage points, and the corresponding rise in FalseAlarmRate outweighed the benefits of the increased DetectionRate. The maximum reduction for a given day in FalseAlarmRate was only 2 percentage points, under Method D, and again this was outweighed by the corresponding drop in DetectionRate for Method D. Finally, under Methods B and C, improvements were too small and inconsistent to be worth the doubled or tripled processing time needed to analyze each lane separately. Thus, lane-by-lane analysis is not a practical extension to this congestion tracking tool.

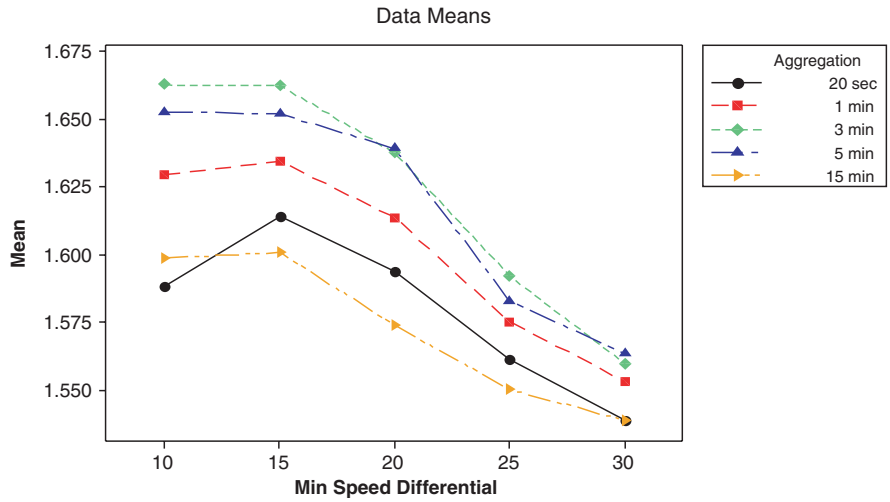
Evaluation of Fundamental Diagrams

To generate more ideas on how to improve the algorithm, fundamental diagrams were constructed for various detectors on several analyzed days, such as in Figure 5. The points on these diagrams were coded to show where errors occurred. In this figure, there is a clear cluster of false negatives (squares), which were instances of congestion according to the ground truth but which were not detected by our speed-based algorithm. Because the false negatives cluster together in the flow–occupancy plane, incorporating rules about flow or occupancy might improve the algorithm's ability to classify such observations correctly. For example, perhaps the optimal speed thresholds are different above a certain occupancy threshold. This potentially promising improvement will require more detailed analysis.

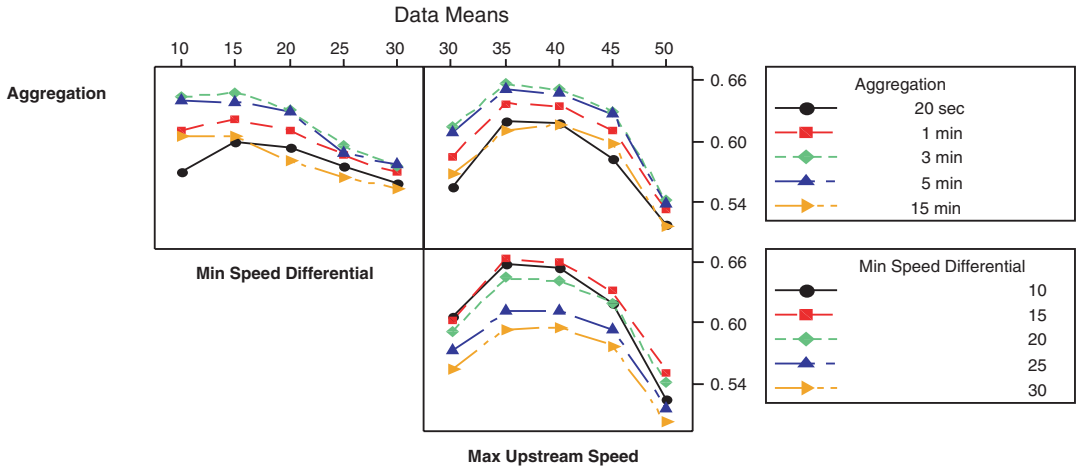
ADDITIONAL APPLICATIONS

Shockwave Speed Estimation

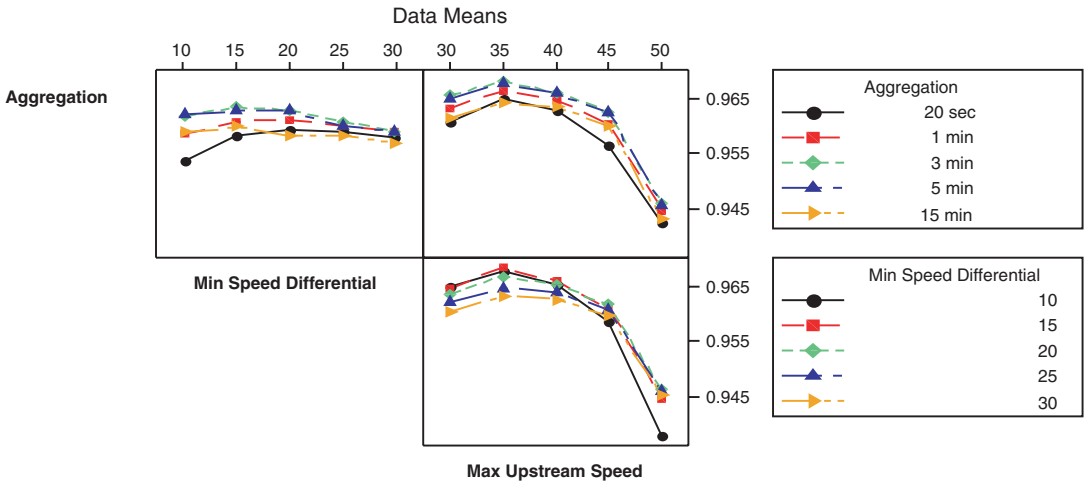
Our previous paper (*2*) illustrated how, after detection of a bottleneck and its associated upstream congested regime, one can compute the slope of the backward-forming shockwave edge (as shown in Figure 2), which is the speed of the queue propagation. In that paper algorithmically derived shockwave speeds were compared



(a)



(b)



(c)

FIGURE 4 (*a*) Interaction plot for SumScore after MaxUpstreamSpeed is optimized, (*b*) full interaction plot matrix for ProductScore, and (*c*) full interaction plot matrix for Accuracy.

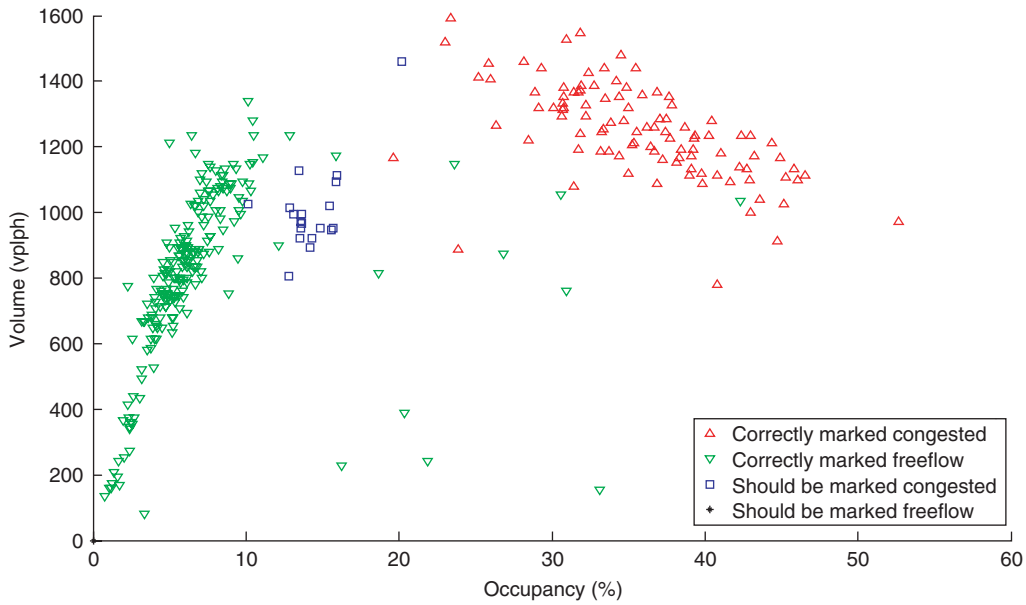


FIGURE 5 Fundamental diagram of occupancy versus volume for milepost 307.45 on 1 day's worth of analyzed data, coded to show where algorithm's results were incorrect according to the ground truth.

against those found using the ground truth data. The algorithm's results tended to be close to the ground truth but occasionally overestimated the propagation speed. With additional work to improve accuracy, the bottleneck detection tool could measure real-time queue propagation speed to estimate when the bottleneck-related congestion will affect traffic at a given point upstream.

Visualizing Years of Data

One immediate benefit of automatic detection of bottlenecks is the ability to process large amounts of historical data. The authors have analyzed 1,000 days of the I-5N corridor to detect the most frequent bottleneck activation sites. First each day is iterated independently to find activation sites and the extent of a bottleneck. The collection is then analyzed to see which bottleneck sites were active with frequency above a certain threshold. In Figure 6, the bottleneck sites that are active at least 25% of the time, at least 40% of the time, and at least 75% of the time are shown. By varying the threshold, a visualization of a particular corridor's most active bottleneck sites is also produced. The historical percentile methodology (perhaps coupled with incident information) also makes it possible to distinguish between recurrent and nonrecurrent congestion. In this case, there are several recurrent bottlenecks but the northernmost afternoon bottleneck near Jantzen Beach is active most often. Clearly, this is a useful tool for performance measurement and has potential to be used for predicting conditions in real time.

CONCLUSIONS

This paper performed a detailed graphical and statistical analysis to test the best combination of parameters toward implementing an automated bottleneck detection procedure in Portland, Oregon. The paper also presented a useful implementation that detects bottleneck activations and deactivations; considers their persistence over time; traces and maps resulting congestion upstream; analyzes historical congestion patterns; and measures shock propagation velocities. Based on a rigorous and detailed comparison with ground truth data, it appears that the procedures can be extended to the remaining freeway corridors in Portland. In terms of traffic information, appropriate quality assessment criteria should characterize the timeliness of traffic messages and their consistency with the traffic situation that would be experienced by the driver on a given route. This paper presents information that will be useful in the planning and operations environment but also to travelers, by mapping recurrent congestion in time and space. It is clear that each user's optimal choice of thresholds and data aggregation level will depend on variable factors in terms of geography, traffic pattern, and driver behaviors in a certain region.

FUTURE WORK

The promising results described here are leading to additional research toward automating the process of identifying bottlenecks on Portland freeways. One next step consists of comparing the percentiles of congestion area by day of the week and by weather or seasonal conditions, using additional historical data. A further step is to incorporate volume data from PORTAL for quantifying total delay caused by such recurrent bottlenecks that can be translated into the cost of externalities such as time wasted, fuel consumption, and emissions unnecessarily produced by these bottlenecks. The automated method in this paper should also be evaluated with regard to real-time bottleneck detection and future bottleneck forecasting.

When this bottleneck detection tool is fully implemented in PORTAL, users will be able to make such comparisons and explore what might be causing this congestion. This will facilitate some simple forecasting that can be shown on the time–space speed plots as the loop detector live feed proceeds. Finally, the reliability of travel time predictions on a given corridor may be more important than the travel times themselves for travelers, shippers, and transportation managers. Thus, in addition to identifying recurrent bottlenecks using measures of delay, the authors plan to provide measures of

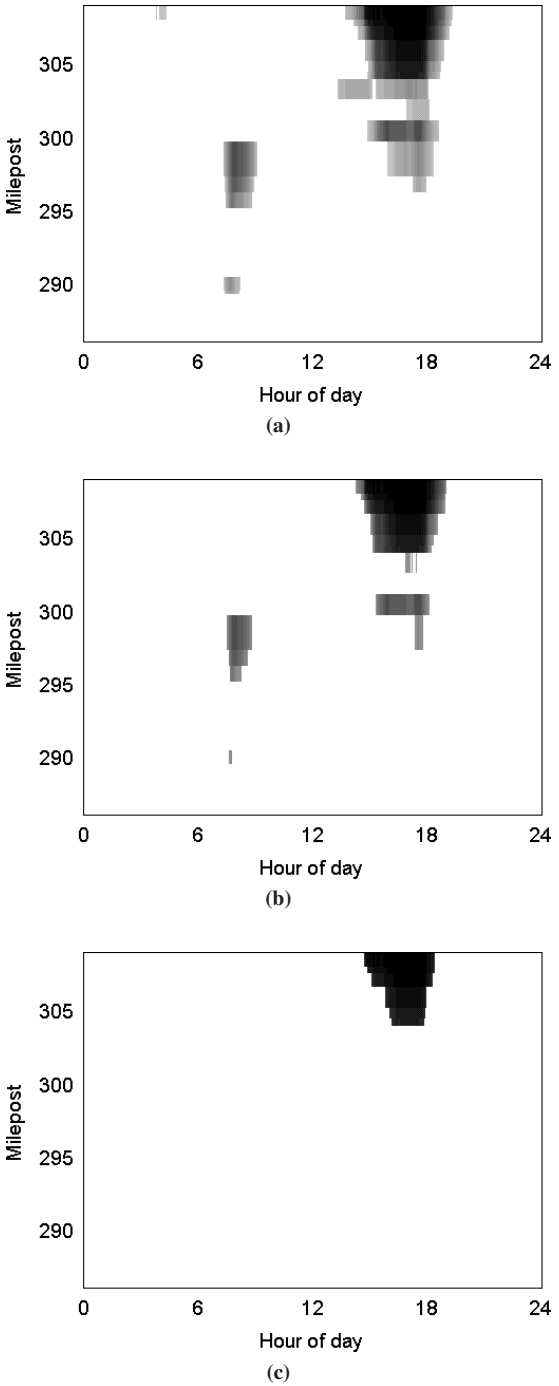


FIGURE 6 Bottleneck activation sites (with darkness proportional to frequency of recurrence): (*a*) recurrence >25%, (*b*) recurrence >40%, and (*c*) recurrence >75%.

reliability such as 95th percentile travel time, buffer index, planning time index, travel time index, and congestion frequency.

ACKNOWLEDGMENTS

The authors acknowledge the support of the Oregon Department of Transportation, FHWA, TriMet, and the City of Portland for their ongoing support and guidance in the development of the Portland Archived Data User Service. This work is supported by the National Science Foundation and the Mexican Consejo Nacional de Ciencia y Tecnología. The Portland metropolitan area's TransPort ITS committee has also encouraged development of this project. Huan Li and other students in the Portland State University ITS lab assisted with this analysis.

REFERENCES

1. Daganzo, C. F. *Fundamentals of Transportation and Traffic Operations.* Elsevier Science, New York, 1997.
2. Wieczorek, J., L. Huan, R. J. Fernández-Moctezuma, and R. L. Bertini. Integrating an Automated Bottleneck Detection Tool into an Online Freeway Data Archive. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
3. Cassidy, M. J., and J. R. Windover. Methodology for Assessing Dynamics of Freeway Traffic Flow. In *Transportation Research Record 1484,* TRB, National Research Council, Washington, D.C., 1995, pp. 73–79.
4. Cassidy, M. J., and R. L. Bertini. Some Traffic Features at Freeway Bottlenecks. *Transportation Research Part B,* Vol. 33, No. 1, 1999, pp. 25–42.
5. Cassidy, M. J., and R. L. Bertini. Observations at a Freeway Bottleneck. *Proc., 14th International Symposium on Transportation and Traffic Theory,* Jerusalem, Israel, 1999, pp. 107–124.
6. Bertini, R. L., and A. M. Myton. Use of Performance Measurement System Data to Diagnose Freeway Bottleneck Locations Empirically in Orange County, California. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1925,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 48–57.
7. Horowitz, Z., and R. L. Bertini. Using Portland Archival Listing Data to Empirically Diagnose Freeway Bottlenecks Located on Oregon Highway 217. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
8. Chen, C., A. Skabardonis, and P. Varaiya. 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.
9. Zhang, L., and D. Levinson. Ramp Metering and Freeway Bottleneck Capacity. *Transportation Research Part A,* Vol. 44, No. 4, 2010, pp. 218–235.
10. Banks, J. H. Two-Capacity Phenomenon at Freeway Bottlenecks: A Basis for Ramp Metering? In *Transportation Research Record 1320,* TRB, National Research Council, Washington, D.C., 1991, pp. 83–90.
11. Hall, F. L., and K. Agyemang-Duah. Freeway Capacity Drop and the Definition of Capacity. In *Transportation Research Record 1320,* TRB, National Research Council, Washington, D.C., 1991, pp. 91–98.
12. Bertini, R. Toward the Systematic Diagnosis of Freeway Bottleneck Activation. *Proc., IEEE 6th International Conference on Intelligent Transportation Systems,* Shanghai, China, 2003.
13. Kerner, B., H. Rehborn, M. Aleksic, and A. Haug. Recognition and Tracking of Spatial–Temporal Congested Traffic Patterns on Freeways. In *Transportation Research Part C,* Vol. 12, No. 5, 2004, pp. 369–400.
14. Palmer, J., R. L. Bertini, H. Rehborn, J. Wieczorek, and R. J. Fernández-Moctezuma. Comparing a Bottleneck Identification Tool with the Congested Traffic Pattern Recognition System ASDA/FOTO Using Archived Freeway Data from Portland, Oregon. *Proc., 16th World Congress on ITS,* Stockholm, Sweden, 2009.
15. Bertini, R. L., S. Hansen, A. Byrd, and T. Yin. Experience Implementing a User Service for Archived Intelligent Transportation Systems Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1917,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 90–99.
16. May, A. D. *Traffic Flow Fundamentals.* Prentice Hall, New York, 1989.
17. Bertini, R. L., H. Li, J. Wieczorek, and R. Fernández-Moctezuma. Using Archived Data to Systematically Identify and Prioritize Freeway Bottlenecks. Presented at 10th International Conference on Application of Advanced Technologies in Transportation, Athens, Greece, 2008.
18. Bogenberger, K., N. Henkel, E. Göbel, and R. Kates. How to Keep Your Traffic Information Customers Satisfied. Presented at 10th World Congress on Intelligent Transport Systems, Madrid, Spain, 2003.

The Highway Traffic Monitoring Committee peer-reviewed this paper.