Systematic Identification of Freeway Bottlenecks

Article *in* Transportation Research Record Journal of the Transportation Research Board · January 2004

DOI: 10.3141/1867-06

Pravin Varaiya
University of California, Berkeley

SYSTEMATIC IDENTIFICATION OF FREEWAY BOTTLENECKS

Chao Chen*
EECS Department
University of California, Berkeley, 94720
Tel: (510)643-5894; Fax: (510)643-2356
chaos@eecs.berkeley.edu

Alexander Skabardonis
Institute of Transportation Studies
University of California, Berkeley, 94720-1720
Tel: (510)642-9166; Fax: (410)642-1246
skabardonis@ce.berkeley.edu

Pravin Varaiya
EECS Department
University of California, Berkeley, 94720
Tel:(510)642-5270; Fax:(510)643-2356
varaiya@eecs.berkeley.edu

Transportation Research Board
83rd Annual Meeting
January 2004
Washington, DC

July 30, 2003

2748 words
3 tables
3 Figures
total 4248 words

ABSTRACT

We present an algorithm that identifies bottleneck locations, the times for which each bottleneck is active and the delay it causes. The bottlenecks are ranked in terms of their frequency of recurrence and the magnitude of their delay impact. The algorithm works with five-minute loop detector data. It uses speed difference as an indicator of bottleneck activation.

The algorithm is applied to three months of data from 270 miles of seven freeways in San Diego. It identifies 160 locations whose bottlenecks cause 64 percent of the total delay on these freeways. The 'top' ten account for 61 percent of the delay from all bottlenecks. The method can be used in any area where large amounts of data are available. Transportation authorities may use it to identify bottlenecks and to track their impact over time.

INTRODUCTION

Certain freeway locations experience congestion at nearly the same time almost every day. They are often called (recurring) bottlenecks. Bottlenecks may be caused by merging and diverging traffic, lane drops, and grade changes, for example. The Highway Management Handbook views bottleneck analysis and removal a key element of congestion relief (*1*).

Bottlenecks like the Bay Bridge in San Francisco and the I-5/I-805 split in San Diego are familiar to area motorists. However, without measurements we cannot tell where all the bottlenecks are in a region nor the severity of the delay they cause.

A transportation engineer asked to find the bottlenecks on an unfamiliar freeway might drive along the freeway for several days and at different times, noting locations downstream of which traffic is free flowing, but upstream of which traffic is significantly slower. He might further investigate his list of bottleneck locations and their activation times in order to figure out the most severe bottlenecks. This is a time-consuming way to identify bottlenecks and assess their impact.

We present an automatic method that mimics our traffic engineer. The method relies on the availability of traffic data (flows and speeds) at freeway locations over many days. The method systematically processes all the data, locates bottlenecks by speed differentials, determines how frequently each one is activated and calculates the average delay it causes.

To illustrate the efficacy of the method, we apply it to three months of data from 263 loop detector stations in 270 miles of seven San Diego freeways. The method identifies 160 bottleneck locations, accounting for 64 percent of total delay on the seven freeways; the ten worst bottlenecks account for 61 percent of the delay caused by them all.

The virtue of the method is that it can be used with no familiarity with the freeway. Its limitation is that it does not diagnose the cause of the bottleneck. As explained in a later section, information about the freeway geometry at the identified bottleneck can help to determine its cause and suggest corrective action.

Identification of bottlenecks and estimation of their activation times and delay impact can aid the transportation agency in focusing and prioritizing relief efforts. Applying the method routinely over time allows the identification of new bottlenecks and monitoring of existing ones to discern congestion trends.

BACKGROUND

The notable Twin Cities, Minnesota study by Zhang and Levinson (*2*) follows a strategy similar to that adopted here. However, their objectives are different. They identify bottlenecks locations and their activation times on the basis of occupancy differentials. Based on frequency of recurrence, and some other criteria, they select 27 bottlenecks for further investiation. For each of these, they use flow data to determine average pre-queue and queue flow rates, and the percentage flow drop. They propose a capacity measure based on these flow rates. Although they illustrate the freeway geometry at each of the 27 bottlenecks, this plays no role in the analysis.

This Minnesota study follows several earlier smaller scale studies, also intent on finding indicators of bottleneck activation. For example, Banks measured 30-second speeds upstream of the bottleneck locations and used a drop in speed as an indicator of congestion onset at four San Diego locations (*3*). Hall and Agyemang-Duah used a threshold on the ratio of 30-second occupancy to flow because speed data were not available (*4*). Bertini suggests other signals that may be used to detect bottleneck activation, such as the variance of 30-second counts (*5*).

These studies and others conclude that flow rates drop after a bottleneck is activiated (*6*, *7*, *8*). They suggest that reducing bottleneck activation by, say, ramp metering, may increase flow. Our method evaluates the impact of a bottleneck in terms of the delay it causes, rather than by the reduction in flow.

BOTTLENECK DETECTION ALGORITHM FROM LOOP DETECTOR DATA

Given a pair of upstream-downstream detector station locations, we want to determine the times (if any) when there is an active bottleneck between them. Bottlenecks are clearly visible features in two-dimensional speed contour maps. Figure 1 is an example.

It shows constant-speed contours on a time vs. distance plane. The time is between 5:50 and 9:30 AM on 1 May, 2003. The distance is along I-15 SB; the direction of travel is in order of decreasing postmiles. A speed gradient persists near postmile 26 between 5:50 and 9:30 AM. Speed is low upstream of the bottleneck. The region extending five miles upstream remains congested while the bottleneck is active.

Figure 2 shows the speed contour plots on I-15 SB over four weeks in May, 2003. Lower speeds are represented by darker colors. The bottleneck at postmile 26 is visible on 15 out of 20 days. There is an equally frequent bottleneck at postmile 15, but it appears to be causing less delay.

The algorithm is inspired by this visual analysis of speed contours. It requires 5-minute averaged speed and flow (count) measurements by freeway loop detectors. For the application, we use the data from the California freeway performance measurement system, PeMS. PeMS receives real time 30-sec loop detector flow and occupancy measurements from throughout the state. PeMS processes these data and computes 5-minute speed averages, and makes available online both the real time and historical data (*9*).

The algorithm uses the presence of a sustained speed gradient between a pair of upstream-downstream detectors to identify bottlenecks. Other signals of bottleneck activation have been used, including occupancy differentials (*2*); drop in upstream speed (*3*); ratio of occupancy to flow as speed surrogate (*4*); and variance in 30-sec counts (*5*). Because speed changes much more sharply than occupancy when a bottleneck is activated, speed differentials provide a more sensitive indicator than occupancy-based signals.

We consider a freeway with n detectors indexed by $i = 1, \cdots, n$, each of which provides speed and flow measurements, averaged over 5-minute intervals indexed by $t = 1, 2, \cdots$. Detector i is located at postmile x_i; $v_i(t) = v(x_i, t)$ is its measured speed (miles per hour, mph) and $q_i(t) = q(x_i, t)$ is its measured flow (vehicles per hour, vph) at time t. If $x_i < x_j$, it is understood that x_i is upstream of x_j.

The algorithm has three steps. First, the algorithm declares an active bottleneck at certain locations and times if the data meet criteria (1)—(4). Second, it includes additional time periods as part of bottleneck activation, provided nearby time intervals are selected in the first step. The criterion for this is (5). Lastly, it calculates the delay caused by a bottleneck, using (9).

The algorithm declares that there is an active bottleneck between two locations x_i, x_j, with $x_i < x_j$, during period t if the following four inequalities hold:

$$
\begin{aligned}
x_j - x_i &< 2 \text{ miles} && (1)\\
v(x_k, t) - v(x_l, t) &> 0 \text{ if } x_i \le x_k < x_l < x_j && (2)\\
v(x_j, t) - v(x_i, t) &> 20 \text{ mph} && (3)\\
v(x_i, t) &< 40 \text{ mph} && (4)
\end{aligned}
$$

Location x_i is upstream of x_j, but there may be other detectors at x_k, x_l between these locations. The thresholds in (1)—(4) are selected to best match visual evaluations of contour plots. Extensive analysis of data in Los Angeles shows that free flow speed is close to 60 mph and, when a bottleneck is activated, speed drops rapidly to below 40 mph (*10*). Hence the 20 mph minimum speed differential (1) and 40 mph congestion speed threshold (4). (However, for non-California freeways, different threshold speeds may be more appropriate.) The maximum separation of 2 miles in (1) is designed to include locations where speed continues to rise as we go downstream, but the difference between each neighboring pair is small. The constraint (2) that speed should drop continuously is of course the distinctive characteristic of an active bottleneck.

Recurring bottlenecks are sustained over periods longer than five minutes. Let $A_i(t) = 1$ if there is an active bottleneck at location i and time period t. We declare a bottleneck to be *sustained* between times t_1 and t_2 if

$$\sum_{\tau=t}^{t+N-1} A_i(\tau) \geq qN, \; \forall\, t_1 \leq t \leq t_2 - N + 1, \tag{5}$$

where $N = 7$ and $q = \frac{5}{7}$. That is, a sustained bottleneck has at least five active bottleneckperiods (or 25 min) within every seven consecutive periods (or 35 min). The definition is somewhat arbitrary. It is formulated in response to situations shown in Figure 1, in which at postmile 26 the bottleneck is continuously sustained between 7:00 and 8:00 except for several five-minute periods. The notion of sustained bottleneck allows us to treat this as a single bottleneck rather than two or three bottlenecks. It is customary to define the most downstream location of a sustained bottleneck as the location of an active bottleneck.

Figure 1 shows the result of applying the algorithm to data from I-15 SB on 9/18/2002, between 5 and 11 AM. The locations and times of detected bottlenecks are the squares superimposed on the speed contours. The contours visually suggest one bottleneck between 5:45 and 9:45 at postmile 26, and another between 6:45 and 8:30 at postmile 15, and indeed both bottlenecks are identified by the algorithm.

Calculating Delay

We now describe how the algorithm estimates the delay caused by a bottleneck. The speed contour in Figure 1 shows regions of congestion upstream of each bottleneck location. We assign the delay in vehicle-hours represented by these regions to their associated bottlenecks and compute its value as follows.

The n detectors divide the freeway into n segments. In the PeMS database, a segment is typically one half to one mile long. We say that a segment is congested at time t if its speed is below 40 mph. We define the congested region associated with a bottleneck as the contiguous group of congested segments immediately upstream of the bottleneck location. For an active bottleneck just downstream of segment j at time t, the congested region is the set of segments $B_j(t)$,

$$B_j(t) \stackrel{def}{=} \{i : v_k(t) < 40 \text{ mph}, \text{ for all } i \leq k \leq j\}. \tag{6}$$

The delay $D_j(t)$ associated with the bottleneck during this period is the sum of the delays in $B_j(t)$:

$$D_j(t) = \sum_{i \in B_j(t)} d_i(t) \tag{7}$$

where $d_i(t)$ is the delay in segment i at time t. Segment delay is defined as the difference between the vehicle-hours traveled and the minimum required if there is no congestion. Taking 60 mph as the free flow speed when there is no congestion, the delay per segment is

$$d_i(t) \stackrel{def}{=} l_i \times q_i(t) \times \left(\frac{1}{v_i(t)} - \frac{1}{v_{\text{ref}}}\right) \; ; \; v_{\text{ref}} = 60 \text{ mph} \tag{8}$$

where $l_i, q_i(t)$, and $v_i(t)$ are the segment length, volume, and average speed on the segment at t. (The 60 mph free flow speed is justified in (*10*).) The total delay attributed to a bottleneck at segment j that is active between times t_1 and t_2 is

$$D_j(t_1, t_2) = \sum_{t=t_1}^{t_2} D_j(t). \tag{9}$$

Using this definition, we can calculate the delay caused by each active bottleneck.

Remarks

Three observations about the algorithm are relevant. First, as already noted, instead of speed differentials one may use other signals to indicate bottleneck activation. Second, it may happen that a bottleneck may on some days be deactivated because of a "spillback" from an active downstream bottleneck. In this case, the algorithm will not identify the deactivated bottleneck, and it will attribute all of the delay to the downstream bottleneck, which is appropriate. Third, the algorithm is limited by the nature of the available data, in particular, by the detector spacing. For instance, if the detectors are widely spaced, one cannot tell that speed is decreasing as one moves downstream towards the bottleneck location, as condition (2) suggests.

RESULTS

Data

The algorithm described above was applied to loop detector data from San Diego County, Caltrans District 11, in the PeMS database (9). There are seven instrumented freeways, providing 270 directional miles of detector coverage. The data consist of five-minute lane-by-lane speed and volume. They are aggregated to represent the average speed and total volume at each location. Data between 5 AM and 10 PM on workdays from 4/1/2003 through 6/30/2003 were used; there are 64 days total.

Detection Output

The bottleneck detection algorithm found 1733 sustained bottlenecks distributed over 160 distinct locations on 64 workdays. These bottlenecks represent all locations and times that satisfy equations (1) - (5). Their causes are unknown, and may include incident or recurring conditions. We hypothesize that certain locations exhibit recurrent bottlenecks and contribute disproportionately to overall delay. This is indeed the case.

The delay associated with each detected bottleneck is computed using (9). The total delay associated with bottlenecks during the test period is 1.2 million vehicle-hours, which is 64% of the total delay measured on these freeways during this period. Of the delay caused by bottlenecks, 61% is attributed to the top ten locations alone.

Figure 3 shows histograms of delay and recurrence of the detected bottleneck locations. The upper plot shows the histogram of delay. Most locations have delays under 200 v-h per day, but a few locations have daily delays over 1000 v-h. The lower histogram shows a similarly large tail in the distribution of the number of days during which a bottleneck is active at each of the locations. These outliers represent the most delay-causing and most persistent of the locations.

Table 1 lists the locations that rank in the top ten in both recurrence and delay. That there are nine such locations implies that the most delay-causing bottleneck locations are also the most persistent. These locations exhibit sustained bottlenecks on at least 78% of the days studied. They also contribute significantly to the total delay on these freeways. This list pinpoints critical areas of the freeway where congestion originates regularly.

Verification

The bottlenecks found by the detection algorithm do indeed represent the most congested bottlenecks in San Diego. Since no definitive list is available, we use various sources of information including the Caltrans website, reports from the San Diego Association of Governments (SANDAG), area news reports, and Banks' study.

A search in Caltrans construction projects and San Diego area news articles verifies that the list of bottlenecks in Table 1 include the most congested locations in the region. The interstates I-5 and I-15 are the two major north-south routes; they contain several well-known bottlenecks. The I-5/I-805 merge is notorious with backups that last up to two hours (*11*). In response to the congestion here, Caltrans is conducting a five-year, $176 million project to widen the road between the freeway split (postmile 30) and Del Mar Heights Road (postmile 34) (*12*). Results of bottleneck detection show a bottleneck southbound in the morning and northbound in the evening, both near postmile 37. Even though this location is seven miles upstream of the actual I-5/I-805 split, it has been suggested that this southbound bottleneck is the result of vehicles changing lanes downstream (*11*). The northbound bottleneck may have other causes since it is downstream of the split. It may be unrelated to the merge and may not benefit from the project (*13*).

Three bottlenecks are found on I-15, at postmiles 26 and 14 southbound in the AM and postmile 26 northbound in the PM. I-15 is also known to be a busy route. Caltrans is currently conducting a project on I-15 to build managed lanes for Bus Rapid Transit; the location is between SR-163 and Centre City Parkway. This section contains all three bottlenecks found by the algorithm.

Our list of most severe bottleneck locations also agrees with a SANDAG study that prioritizes planned highway projects in San Diego's 20-year Regional Transportation Plan (RTP) (*14*). This study ranks 46 projects, including new freeways, widening of existing freeeways, and addition of main line and HOV lanes. The projects are ranked based on potential travel time improvement, cost to benefit ratio, and congestion relief. The location of the top eight ranked projects coincide with six out of the top nine bottleneck locations detected by the algorithm. See Table 2. This shows that the list of most severe bottlenecks in Table 1 are located in freeway segments that are considered most in need of improvement.

Banks analyzed four bottlenecks in two papers (*3*, *15*) in 1989 and 1990. One of the four locations, I-8 EB@College in PM, is also in our 'top ten' list, and appears to be the most s erious bottleneck on this freeway. The other three locations studied by Banks exhibit bottleneck behavior but are not among the most congested. Table 3 shows that the delay associated with these locations is much smaller than those attributed to the top ranked locations. Their ranking in terms of delay and frequency are presented as columns in this table. Note, however, that Banks' studies were performed more than ten years ago—traffic patterns and freeway configurations may have changed.

CONCLUSION

We described an algorithm that uses loop detector data to locate active bottlenecks and estimates their delay impact. While there are well-known bottlenecks in a region, our algorithm systematically locates all bottlenecks that satisfy a set of criteria. The algorithm locates an active bottleneck on the basis of speed differentials between upstream and downstream detector locations.

The algorithm was applied to data from San Diego which include 64 days of workday data from 263 detector stations on seven freeways. We found 160 distinct freeway locations that gave rise to 1733 sustained bottlenecks. Several of these locations exhibit persistent bottlenecks and cause large delays. We presented a list of the most delay-causing and persistent bottlenecks; independent sources confirm that they indeed are the most serious. By identifying bottleneck locations and quantifying their impact on delay, our method pinpoints the locations where bottleneck remediation is likely to provide the greatest benefit.

ACKNOWLEDGEMENTS

This study is part of the PeMS project, which is supported by grants from Caltrans to the California PATH Program. We are very grateful Fred Rooney of Caltrans and Professor Jim Banks for their encouragement.

The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views of or policy of the California Department of Transportation. This paper does not constitute a standard, specification or regulation.

REFERENCES

[1] James D. Carvel Jr., Kevin Balke, Jerry Ullman, Katerine Fitzpatrick, Lewis Nowlin, and Christopher Brehmer. Freeway management handbook. Technical Report FHWA-SA-97-064, Department of Transportation, Federal Highway Administration, 1997.

[2] L. Zhang and D. Levinson. Some properties of flows at freeway bottlnecks. In *Proceedings of 83rd Transportation Research Board Annual Meeting*, Washington, DC., January 11-15 2004.

[3] James H. Banks. Two-capacity phenomenon at freeway bottlenecks: a basis for ramp metering? In *Transportation Research Record* 1320, pages 83–90, Washington, DC, 1991. TRB, National Research Council.

[4] Fred L. Hall and Kwaku Agyemang-Duah. Freeway capacity drop and the definition of capacity. In *Transportation Research Record* 1320, pages 91–98, Washington, DC, 1991. TRB, National Research Council.

[5] Robert Bertini. Toward the systematic diagnosis of freeway bottleneck activation. In *IEEE 6th International Conference on Intelligent Transportation Systems*, 2003. In press.

[6] M. Cassidy and R. Bertini. Some traffic features at freeway bottlenecks. *Transportation Research B*, 33B:25–42, 1999.

[7] C. Chen, Z. Jia, and P. Varaiya. Causes and cures of highway congestion. *Control Systems Magazine*, 21(4):26–33, December 2001.

[8] Boris Kerner. Theory of congested traffic flow: self-organization without bottlenecks. In *Fourteenth International Symposium on Transportation and Traffic Theory*, pages 147–171, Jerusalum, Israel, 1999.

[9] Chao Chen, Karl Petty, Alex Skabardonis, and Pravin Varaiya. Freeway performance measurement: Mining loop detector data. In *Transportation Research Record* 1748, pages 96–102, Washington DC, 2001. TRB, National Research Council.

[10] Z. Jia, P. Varaiya, C. Chen, K. Petty, and A. Skabardonis. Maximum throughput in la freeways occurs at 60 mph v. 4. online at pems.eecs.berkeley.edu, January 2001.

[11] Jeff Ristine. Operation: bypass. *The San Diego Union Tribune*, June 29 2003. `http://signonsandiego.printthis.clickability.com/pt/cpt?action=cpt\&expire=\&urlID=6761833\&fb=Y\&partnerID=621`, accessed 7/31/2003.

[12] I-5/I-805 widening project. Caltrans website, January 2003. `http://www.dot.ca.gov/dist11/projectinfo/index.htm`, accessed 7/20/2003.

[13] Denis Devine. Widening of I-5/805 interchange no panacea. *North County Times*, March 31 2002. `http://www.nctimes.net/news/2002/20020331/52902.html`, accessed 7/31/2003.

[14] Kenneth E. Sulzer. Regional highway project evaluation - final results. Technical Report 00-9-15 (Agenda Report), San Diego Association of Governments, September 2000. `http://www.sandag.org/index.asp?projectid=31\&fuseaction=projects.detail`, accessed 7/15/2003.

[15] James H. Banks. Flow processes at a freeway bottleneck. In *Transportation Research Record* 1287, pages 20–28, Washington, DC, 1990. TRB, National Research Council.

LIST OF TABLES

LIST OF FIGURES

Location (postmile)	AM/PM	Recurrence	Average duration (hr)	Average daily delay (v-h)	Percent of freeway total
I-5 NB@Santa Fe Dr. (37.51)	PM	97%	2.6	1617	38%
I-5 SB@Santa Fe Dr. (37.26)	AM	94%	2.9	1678	32%
I-8 EB@College Ave. (8.34)	PM	80%	1.5	447	30%
I-15 NB@Pomerado Rd. (26.05)	PM	92%	2.0	2692	46%
I-15 SB@Miramar Rd. (14.4)	AM	78%	2.1	542	10%
I-15 SB@Pomerado Rd. (25.91)	AM	97%	2.5	2137	38%
I-805 NB@SR-52 (24.60)	AM	78%	1.5	696	28%
I-805 SB@SR-252 (11.5)	PM	88%	1.6	834	28%
I-805 SB@SR-52 (24.33)	PM	94%	1.8	464	15%

TABLE 1 Recurring bottlenecks in San Diego County.

Rank	Project	Number of bottlenecks
1	new freeway SR-905	–
2	HOV lanes on I-805 from SR-52 to I-5	2
3	new freeway SR-52 from SR-125 to SR-67	–
4	ML lanes on I-15 from SR-56 to Centre City Pkwy.	1
5	ML lanes on I-15 from SR-56 to SR-163	1
6	Widen + HOV lanes on I-5 from Birmingham Dr. to La Costa Ave.	0
7	HOV lanes on I-15 from SR-94 to SR-163	0
8	Widen and HOV lanes on I-5 from Del Mar Heights Rd. to Birmingham Dr.	2
	Total	6

TABLE 2 SANDAG 2000 highway projects priority list and corresponding bottlenecks.

Location	Time	Delay	Frequency	Delay Rank	Frequency Rank
I-8 WB@College	AM	6.0	41%	38	23
I-805 NB@El Cajon	AM	9.5	8%	25	75
I-8 EB@College	PM	28.6	80%	10	8
SR-163 SB@Washington	AM	2.9	17%	58	44

TABLE 3 Evaluation of bottleneck locations in Banks' study (*3*, *15*).

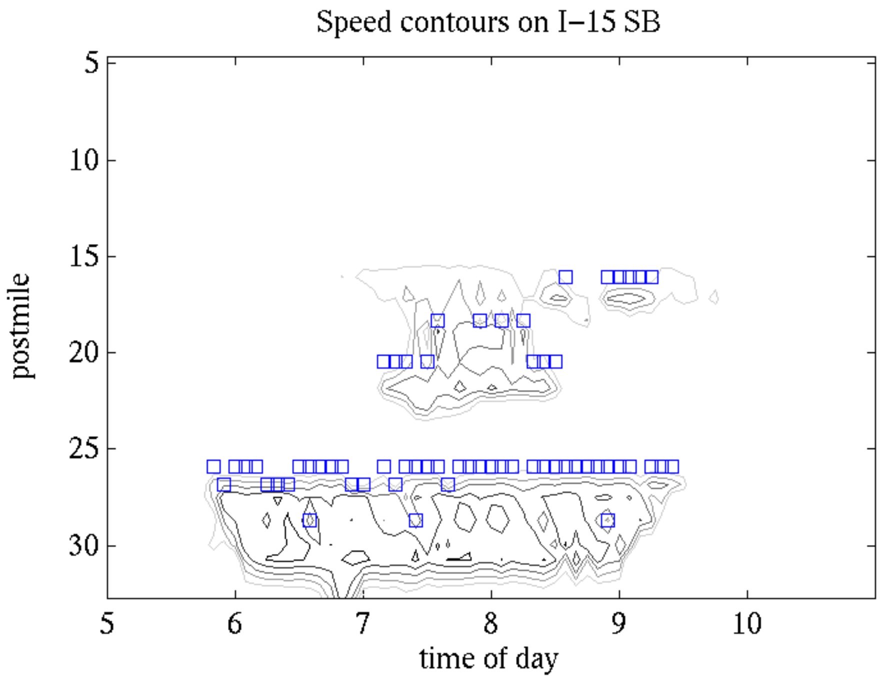

FIGURE 1 Bottleneck and detection demonstration, I-15 SB. Traffic flows in order of decreasing postmile. Data recorded 5/1/2003.

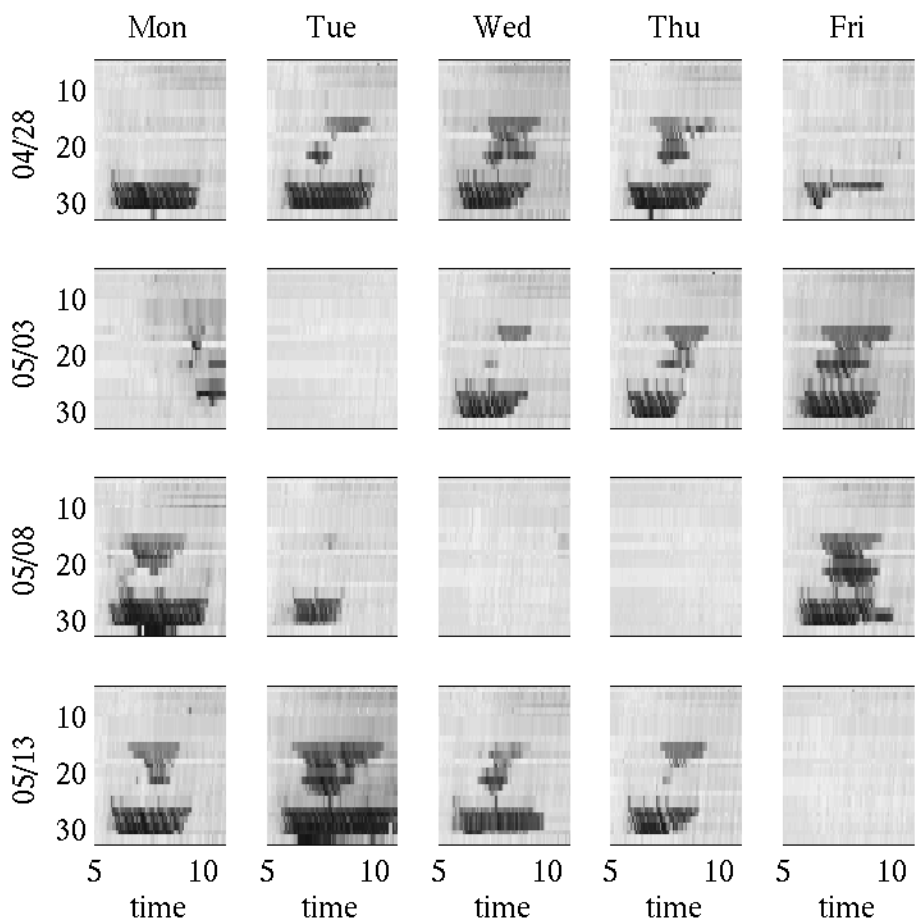

FIGURE 2 I-15 SB, daily speed contours, May, 2003. Bottlenecks at postmile 26 and 15 are consistently observed.

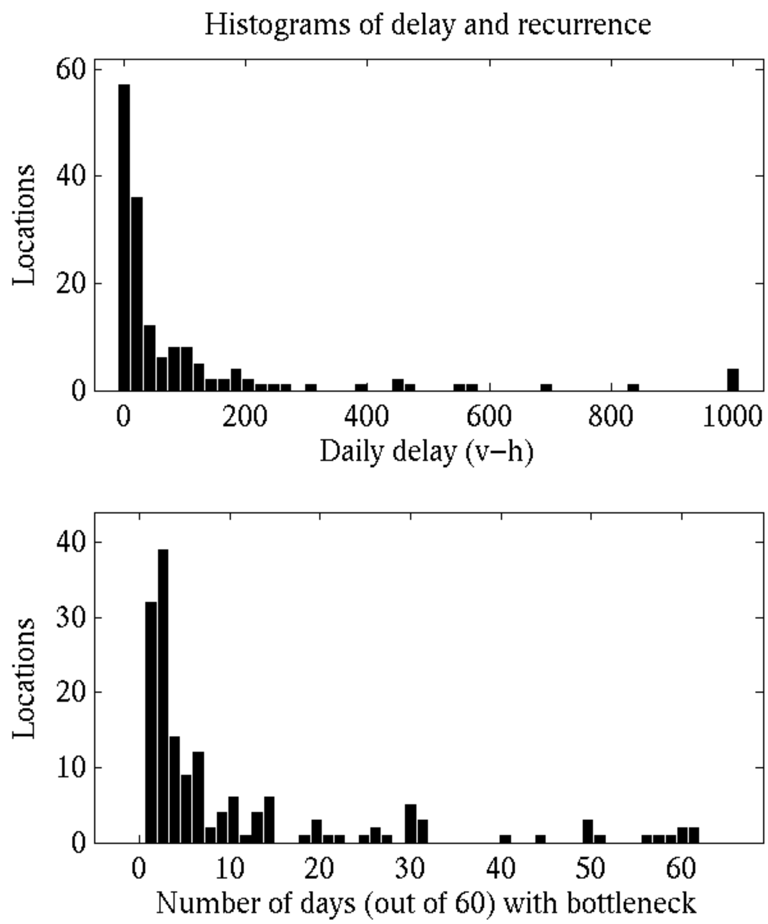

FIGURE 3 Recurrence and delay statistics of San Diego bottlenecks.