Title
Causes of Freeway Productivity Decline and the Opportunities for Gain: A Quantitative Study

Author
Varaiya, Pravin

Publication Date
2008-11-01

CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

Causes of Freeway Productivity Decline and the Opportunities for Gain: A Quantitative Study

Pravin Varaiya

California PATH Research Report
UCB-ITS-PRR-2008-31

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation, and the United States Department of Transportation, Federal Highway Administration.

Final Report for Task Order 5306

November 2008

ISSN 1055-1425

Causes of freeway productivity decline and the opportunities for gain: A quantitative study

Final Report for PATH Task Order 5306

Pravin Varaiya
University of California, Berkeley, CA 94720-1770
Tel: (510) 642-5270, Fax: (510) 642-7815
varaiya@eecs.berkeley.edu

Abstract

Work done under TO 5306 led to three accomplishments. First, a measure of freeway productivity was proposed. Second, the causes of productivity decline led to the notion of "congestion pie." Both productivity loss and congestion pie are available as PeMS applications. Third, the study entitled "An Empirical Assessment Of Traffic Operations" [1] provides a detailed empirical account of congestion.

Keywords: freeway productivity; productivity loss; congestion pie

EXECUTIVE SUMMARY

When a section of a freeway gets congested, both speed and flow are reduced. We propose to measure this reduction as *lost productivity*. This is the number of lane-mile-hours that are lost due to the freeway operating under congested conditions. When the freeway section is congested—the speed drops below a certain, user-defined threshold, e.g. 35 or 60mph—one finds the ratio r between the measured flow and the capacity for this location. The productivity loss is the product of $(1 - r)$, the length of the segment, and the congestion duration, expressed as the number of equivalent lane-miles-hours of freeway. (If the freeway is uncongested, the productivity loss is zero.) The calculation can be carried out at any scale: freeway segment, district, state. Figure 1 displays the productivity loss for District 4 during September 9, 2008-October 5, 2008.

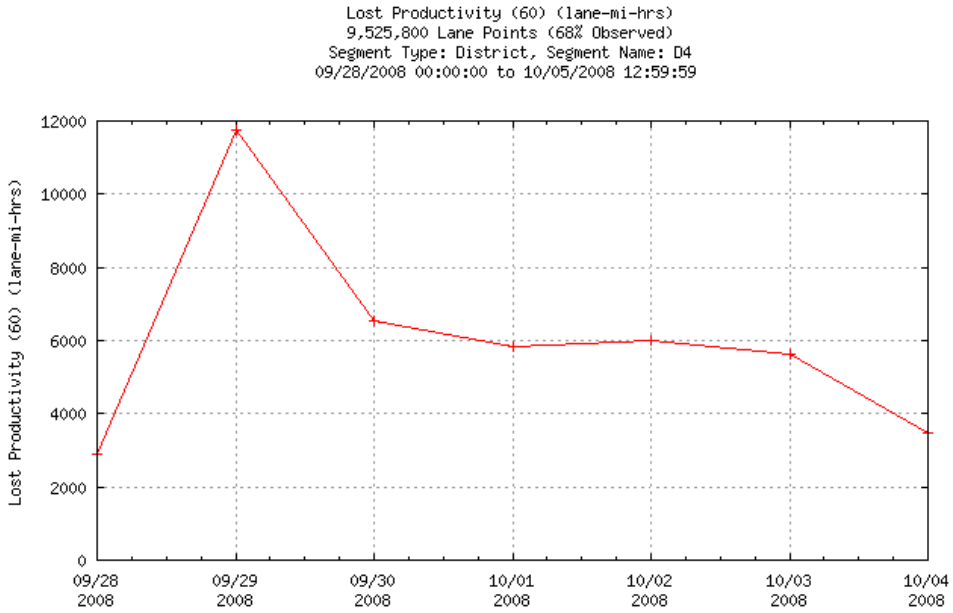


Figure 1: Productivity Loss for District 4. Source: PeMS

Congestion (hence productivity loss) has many causes whose impact can be statistically estimated: there is recurrent and non-recurrent congestion that can potentially be reduced by ideal ramp metering; there is excess demand that cannot be mitigated even under ideal ramp metering; accidents; and, lastly, the residual congestion. These estimates can be displayed in the form of a *congestion pie* as illustrated by Figure 2. A detailed study [2] examines the causes in more detail for I-880 as illustrated in the congestion pie of Figure 3

Data from PeMS provide an unparalleled opportunity to assess freeway performance and suggest ways to improve freeway management. The study [1] takes up this opportunity using six studies of freeway congestion, ranging from bottleneck identification to HOV lane effectiveness.

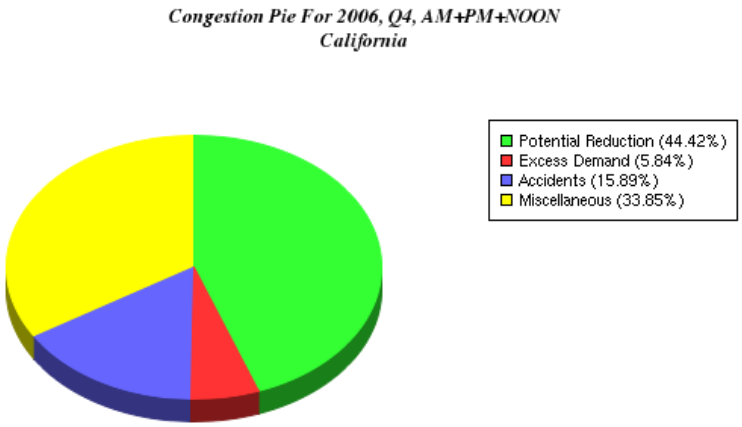


Figure 2: Congestion Pie for California. Source: PeMS

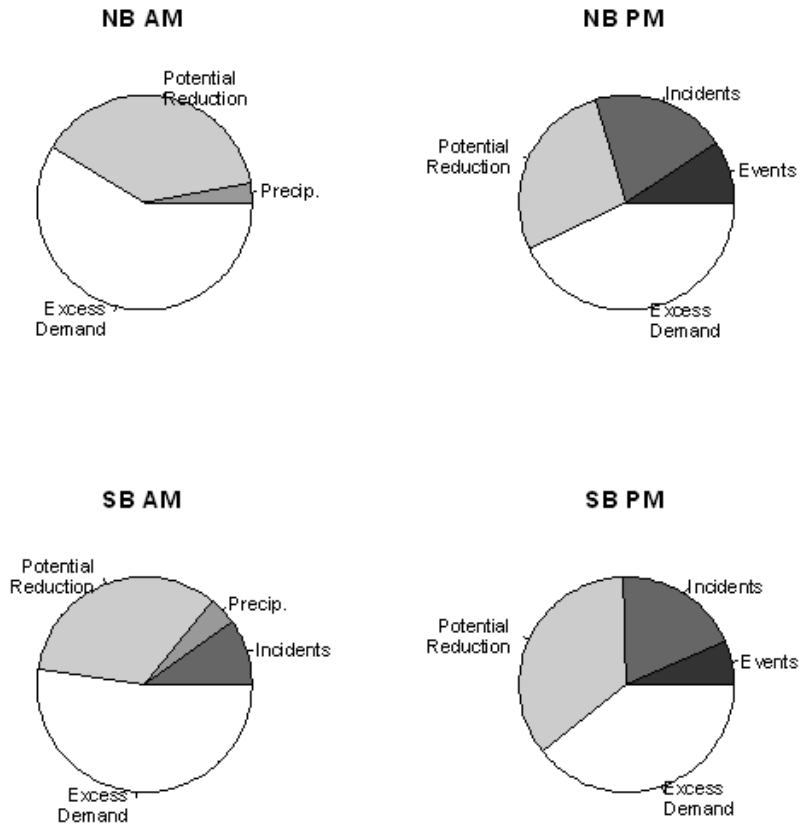


Figure 3: Congestion pie chart for four scenarios on I-880. Source: [2]

ACKNOWLEDGMENTS

The work summarized in this report was carried out jointly with Jaimyoung Kwon, Michael Mauch and Chao Chen. It was supported by the California Department of Transportation through the California PATH Program. The contents of this report reflect the views of the author who is 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 report does not constitute a standard, specification or regulation.

REFERENCES

[1] Chen, C., J. Kwon, and P. Varaiya. An empirical assessment of traffic operations. In H.S. Mahmassani, editor, *Proceedings of the 16th International Symposium on Transportation and Traffic Theory*, pages 105–124. Elsevier, 2005.

[2] J. Kwon, M. Mauch, and P. Varaiya. Components of congestion: Delay from incidents, special events, lane closures, weather, potential ramp metering gain, and excess demand. *Transportation Research Record*, 1959:84–91, 2006.

APPENDIX

The appendix reproduces [1] and [2].

1

AN EMPIRICAL ASSESSMENT OF TRAFFIC OPERATIONS

Chao Chen and Pravin Varaiya,
University of California, Berkeley 94720-1770
Jaimyoung Kwon,
Statistics Department, California State University, Hayward, CA 94542

ABSTRACT

The California Freeway Performance Measurement System stores real-time data from 26,000 loop detectors. PeMS is accessed via an internet browser (http://pems.eecs.berkeley.edu/). It currently has 3 TB of data, growing at 2 GB/day. PeMS extracts useful information from these data and displays it in graphical or tabular form. These data provide an unparalleled opportunity to assess freeway performance and suggest ways to improve freeway management. The paper illustrates this opportunity using six studies of freeway congestion, ranging from bottleneck identification to HOV lane effectiveness. The paper is not a contribution to theory, but it may encourage theoreticians to use a rich data set to formulate and address practical questions.

INTRODUCTION

Operational since 2001, PeMS receives real time data from 26,000 loops grouped into 8,040 Vehicle Detector Stations (VDS) covering 3,000 directional miles of freeways in major California urban areas. PeMS also collects incident data from the Traffic Accident Surveillance and Analysis System (TASAS) and the California Highway Patrol.

The principal aim of this paper is to examine congestion as a performance measure and demonstrate that data can be processed to reliably estimate the causes of congestion, and the gains from better ramp metering, incident management, and traveler information. Each of the following six sections addresses a different aspect of congestion. Some sections report previous research by the PeMS Development Group.

The section BPR CURVE suggests replacing the standard BPR curve by two curves: one for the free flow regime, the other for the congestion regime. For Los Angeles the two regimes separate at 50 mph. Drivers in Los Angeles spend 30% of their time in the congestion regime, so congestion delay can be reduced if this regime can be avoided. IDEAL METERING presents an empirical procedure to rapidly obtain a rough estimate of this reduction by preventing the onset of the congested regime at recurrent bottlenecks. For Los Angeles the procedure estimates an annual saving of 50 million vehicle-hours.

Not all bottlenecks cause significant congestion. BOTTLENECKS summarizes an automated procedure to identify all bottlenecks and rank them by frequency of occurrence and severity of impact. For San Diego County the procedure locates 160 bottlenecks, the ten most severe of which account for 61 percent of the delay from all bottlenecks.

To estimate the delay from a collision, its effect must be separated from congestion caused by bottlenecks. CONGESTION PIE reviews a technique that predicts what the congestion would have been had the collision not occurred. Collisions and bottlenecks cause congestion, and delay from bottlenecks can be reduced by ramp metering. Putting these considerations together yields three congestion pie slices corresponding to collisions; congestion that can be eliminated by ramp metering; and 'residual' congestion due to all other causes, the largest being 'excess' demand.

Congestion delay measures system performance. Travelers experience congestion as large variations in travel time. Because the travel time stochastic process exhibits a large temporal autocorrelation, real time data can be processed to reliably predict travel time, as shown in PREDICTING TRAVEL TIME. Travel time prediction increases welfare: It can suggest a shorter alternative route if one is available; and it can reduce the uncertainty in travel time, even when that time itself cannot be reduced.

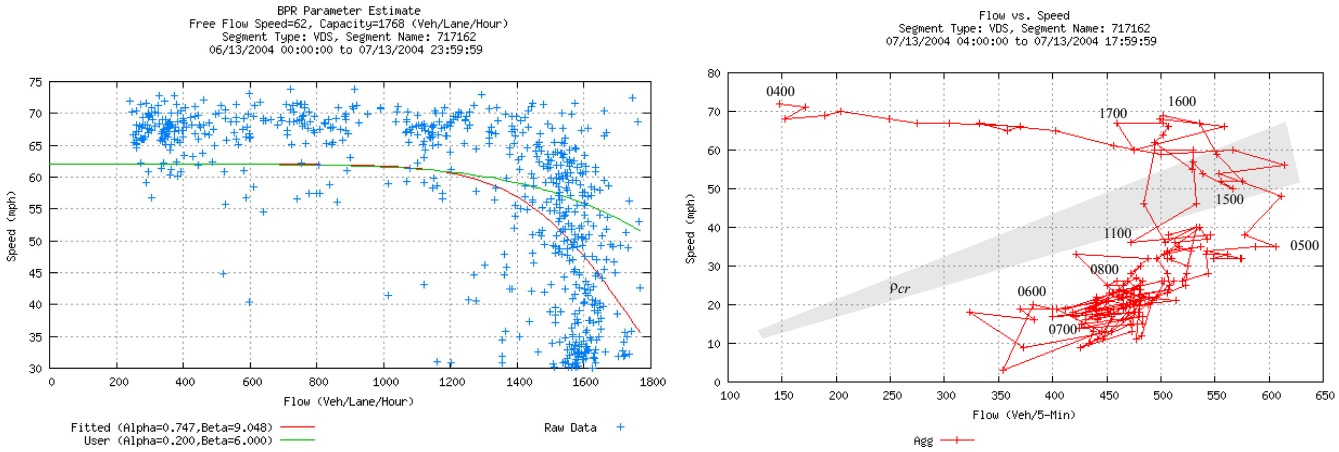
The Bay Area provides a unique opportunity to study the impact of HOV lanes on non-HOV traffic because the HOV lanes are time-actuated. HOV LANE EFFECTIVENESS presents limited evidence suggesting that HOV actuation *increases* overall congestion, by imposing a congestion penalty on non-HOV traffic (which loses one lane) and a capacity penalty on the HOV lane (which acts as a one-lane highway with much lower speed).

BPR CURVE

Figure 1(a) is a scatter plot of speed vs. flow across all four lanes of I-10W in Los Angeles at vehicle detector station (VDS) 717162. Each point represents a one-hour average for the 30-day period 13 June-13 July, 2004. Also displayed are two curves fitted to the BPR (Bureau of Public Roads) equation

$$v = \frac{v_f}{1 + \alpha(q/C)^\beta},$$

in which v is speed, v_f is free flow speed, q is flow, and C is capacity. Data with average hourly speed below 30 mph are discarded. The capacity C is estimated to be the maximum hourly flow observed during the 30-day period, and the free flow speed is the median speed when occupancy is below 10%. The parameters α, β are either user-specified or obtained using a nonlinear least-squares, Marquardt-Levenberg algorithm (Martin, W., 1998).



(a) BPR

(b) Speed vs. flow

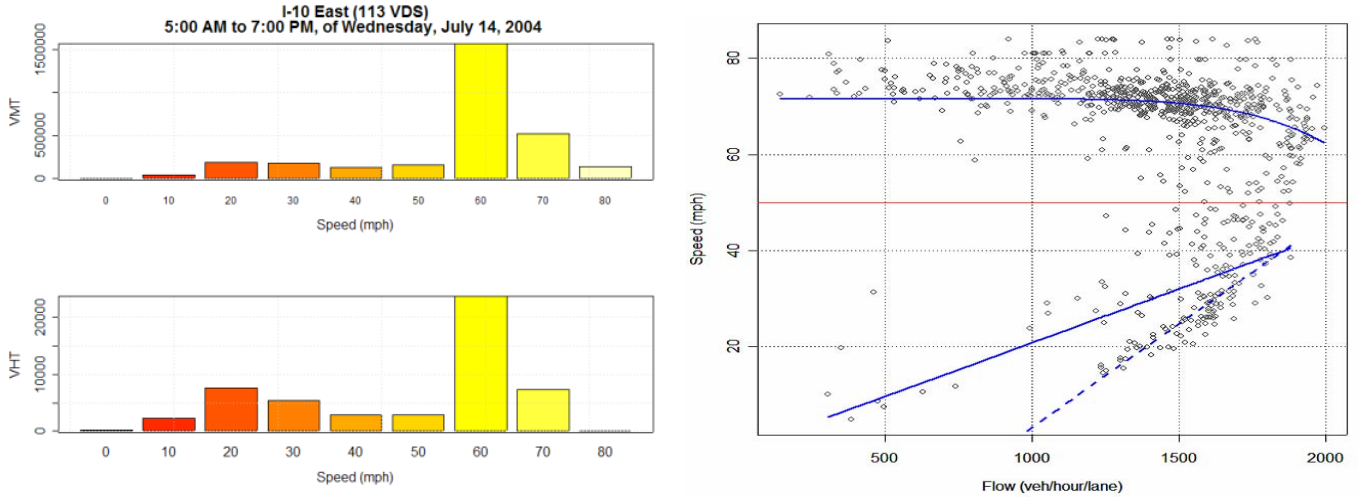
FIGURE 1 (a) Scatter plot of 1-hour average speed vs. flow and (b) trajectory of 5-min average speed vs. flow. Two BPR curves are fitted to the scatter plot in (a). The shaded region, ρ_{cr}, in (b) is the critical density separating free flow from congestion.

Figure 1(b) plots the temporal evolution of speed vs. flow at the same location as in figure 1 during 0400-1700 (4:00 AM-5:00 PM) on 13 July, 2004. Each point now represents a five-minute average. The figure suggests a modified BPR procedure that divides traveler experience into two distinct 'metastable' regimes: the *free flow regime* implicit in the BPR curve, and a low-speed *congestion regime*, separated by a 'critical density' band, ρ_{cr}. The likelihood of the two regimes can be empirically computed for any location, freeway, or an entire region.

Figure 2(a) gives the frequency distribution of VMT (veh-miles traveled) and VHT (veh-hours traveled) on I-10E during 0500-1900, 14 July, 2004. Drivers spent 35% of their time at an average speed of 30 mph and 65% at an average speed of 60 mph, suggesting the two-regime BPR model of Figure 2(b), separated at 50 mph. The free-flow BPR curve is as before. Two linear regressions are fitted to data in the congestion regime,

$$\frac{v}{v_f} = \delta + \epsilon \frac{q}{C}.$$

The solid line is obtained by least-squares; the dotted line is the least quantile regression, which is less sensitive to outliers. The likelihood of each regime, determined by frequency counting, is $P(v > 50 \text{ mph}) = 0.79$, $P(v < 50 \text{ mph}) = 0.21$.



(a) VMT, VHT on I-10E

(b) Two-regime BPR model

FIGURE 2 (a) Distributions of VMT and VHT vs. speed for I-10E and (b) A two-regime BPR model.

IDEAL METERING

Figure 1(b) suggests that holding back volume surges by metering on-ramps may prevent the occurrence of the congestion regime at some bottlenecks, and figure 2(a) implies a large reduction in delay if this can be done. Designing a ramp metering algorithm for a specific freeway section is arduous. Many set points and feedback gains must be selected (Papageorgiou, M., 1983; Papageorgiou, M. et al., 1991), based on a calibrated simulation model. But there is

a simple procedure to roughly estimate the benefits from ramp metering without detailed simulations, based on the hypothesis that the congestion regime can be avoided by controlling flow according to the *Ideal Metering Principle* (IMP) (Jia, Z. et al., 2000):

> If volume surges at on-ramps are held back by a metering policy that always keeps flow below its capacity in every link, freeway speed will be maintained at 60 mph and congestion will not appear. As a consequence of metering vehicles may be stopped at the ramps for some time.

The IMP hypothesis has two parts. One part is that if flow is always maintained below capacity, or equivalently, if density is always less than critical (ρ_{cr}), traffic will be kept in the free flow regime. Data like in figure 1(b) provide indirect support: If the traffic density is never allowed to enter the critical region, traffic will always stay in the free flow regime. The definition of 'capacity' is

empirical: It is taken to be (say) 95% of the maximum sustained observed flow. The second part of the hypothesis is that maximum flow occurs at free flow speeds, nominally 60 mph, as in Los Angeles (Jia, Z. et al., 2001) and Orange County (Chen, C. and P. Varaiya, 2001).

Of course not all congestion is due to volume surges at on-ramps and practical considerations, such as ramps of insufficient length, may prevent implementation of a proper metering policy. The planner should ask: "What will be the impact of implementing IMP-conforming ramp metering if the IMP hypothesis is true?"

A procedure to answer this question is illustrated in (Jia, Z. et al., 2000), using data for a 7-mile section (postmiles 0-7) of I-405N in Orange County, during 0500-1000 for 10 weekdays in June 1998. The section is divided into 13 links, each corresponding to one VDS; eight links have one on- and off-ramp each. A virtual on-ramp is created at the beginning of the most upstream link in order to account for metering of on-ramps upstream of the study section.

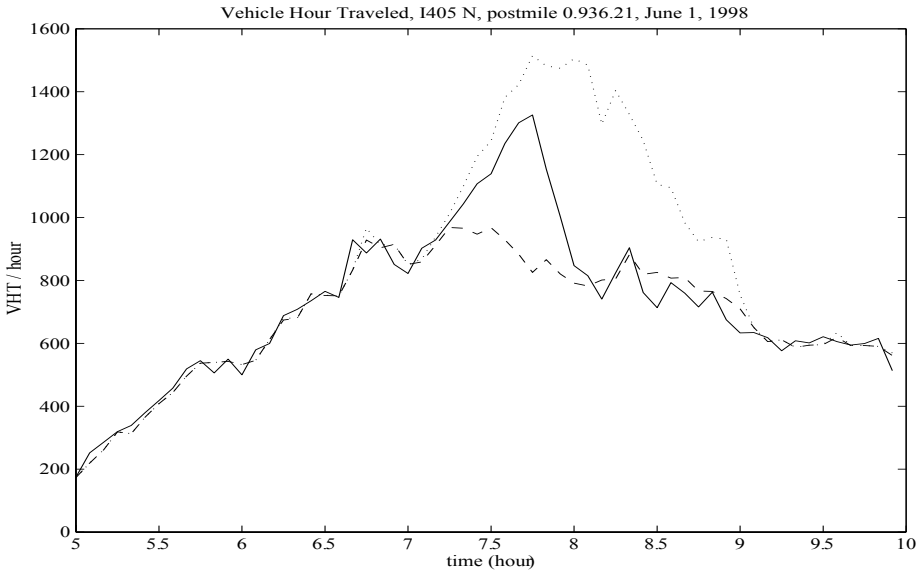


FIGURE 3 The top graph is the time in VHT actually spent on the freeway section, every 5 minutes. Units are normalized to VHT per hour, so the total VHT on this section, between 0500 and 1000 is the area under the top graph. The middle graph is the VHT per hour under IMP metering, including time in ramp queues. The bottom graph excludes time spent on the ramps, so it is the VHT per hour that would be spent traveling at 60 mph. The area between the top and middle graphs is the time saved by metering. The area between the middle and bottom graphs is the time spent at the ramps. Source: (Jia, Z. et al., 2000).

The capacity of each link is calculated as the maximum sustainable aggregate flow. Inflows at on-ramps and exit flows at off-ramps are assumed to remain unchanged despite the metering, whose impact is estimated as follows:

1. At each on-ramp, inflow is metered so that the link flow remains 5% below the link capacity;

2. Traffic on each link after metering is assumed to move at 60 mph;
3. The queue at each on-ramp is calculated by accumulating the net inflow.
On average, two-thirds of the total delay (defined as additional vehicle-hours traveled (VHT) driving below 60 mph) is eliminated by ramp metering. More insight is gained from figure 3, which shows how metering holds back large surges in demand.

This procedure was repeated in (Chen, C. et al., 2001) for five freeways (I-5, I-10, US 101, I-110 and I-405) in Los Angeles during 0000-1200, 3-9 October, 2000. That exercise found that IMP-metering reduces delay by 70%. PeMS calculates the total congestion delay (from driving below 60 mph) for Los Angeles for 2003 to be 83 million vehicle-hours. The procedure suggests that ramp-metering may eliminate 57 million vehicle-hours of delay, which at $20/veh-hr is in excess of 1 billion dollars. Even if only one-half of the delay savings from IMP-metering can be practically realized, this represents an enormous productivity gain that good management can achieve.

BOTTLENECKS

Bottlenecks can cause congestion, which may be reduced by ramp metering. A bottleneck may be associated with physical features such as ramps, lane drops, grade changes, curvature, lane closures, and accidents; but traffic jams and congestion may 'spontaneously' arise in locations with none of these features. In the absence of a guide to locating bottlenecks and estimating their severity, we need an algorithm to automatically (1) identify all bottlenecks, and (2) calculate the delay each one causes.

Such an algorithm is reported in (Chen, C. et al., 2004b), and applied using flow and speed data from 263 VDSs on 270 miles of seven freeways in San Diego. The algorithm uses a sustained speed gradient between a pair of upstream-downstream detectors to identify bottlenecks. We describe the algorithm. Consider a freeway with n detectors indexed $i = 1, \cdots, n$, each giving speed and flow measurements, averaged over 5-minute intervals indexed $t = 1, 2, \cdots$. Detector i is located at postmile x_i; $v_i(t) = v(x_i, t)$ is its speed (miles per hour, mph) and $q_i(t) = q(x_i, t)$ is its 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 four steps. First, it declares an active bottleneck at certain locations and times if the data meet criteria (1)–(4) below. Second, it includes additional time periods as part of the same bottleneck activation, provided nearby time intervals are selected in the first step. The criterion for this is (5). Third, it calculates the delay caused by a bottleneck, using (9). Lastly, identified bottlenecks are ranked in terms of frequency of occurrence and severity to isolate recurrent from transitory bottlenecks and to help prioritize mitigation efforts.

Step 1 Declare an active bottleneck between locations $x_i < x_j$ during t if all four inequalities hold:

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

Thresholds in (1)–(4) are selected on the basis of experience. In Los Angeles free flow speed is 60 mph and, when a bottleneck is activated, speed drops rapidly to below 40 mph (e.g. figure 1(b)). Hence the 20 mph minimum speed differential (3) and 40 mph congestion speed (4) thresholds. The maximum separation of 2 miles in (1) is designed to include locations where speed continues to drop as we go downstream, but the difference between each neighboring pair is small. Location x_i is upstream of x_j, but there may be other detectors at x_k, x_l between these locations. The constraint (2) that speed should drop continuously is the algorithm's characterization of an active bottleneck.

Step 2 Sustained bottlenecks last longer than five minutes. Let $A_i(t) = 1$ if there is an active bottleneck at location i and time period t; otherwise $A_i(t) = 0$. A bottleneck is *sustained* between times t_1 and t_2 if

$$\sum_{\tau=t}^{t+N-1} A_i(\tau) \ge qN, \ \forall\, t_1 \le t \le t_2 - N + 1, \qquad (5)$$

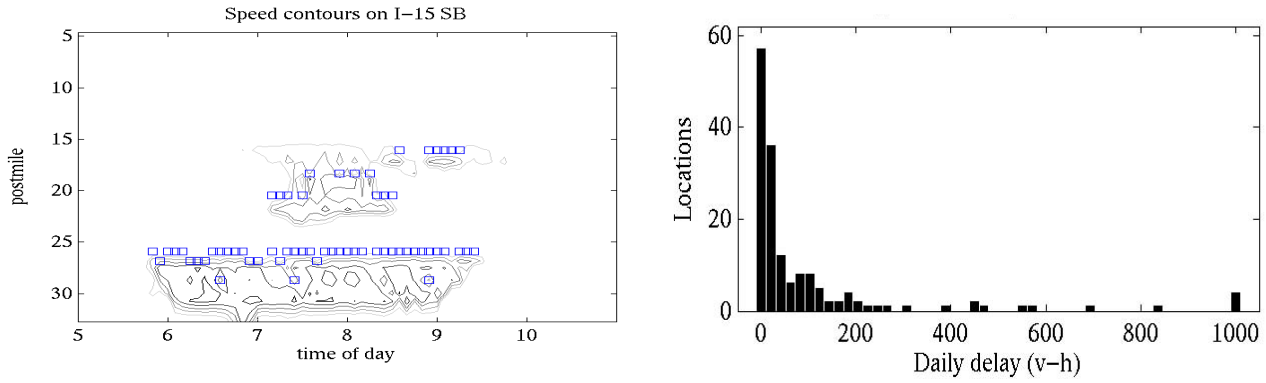
with $N = 7$ and $q = 5/7$. That is, a sustained bottleneck has at least five active bottleneck periods (25 min) within every seven consecutive periods (35 min). This *ad hoc* definition accounts for situations like in figure 4(a), in which at postmile 26 the bottleneck is continuously sustained between 0700 and 0800 except for several five-minute periods. The notion of sustained bottleneck allows treating this as a single bottleneck rather than two or three bottlenecks. The most downstream location of a sustained bottleneck is the location of an active bottleneck.

Figure 4(a) shows the result of applying the algorithm to data from I-15S. The locations and times of detected bottlenecks are the squares superimposed on the speed contours. The contours visually suggest one sustained bottleneck between 0545 and 0945 at postmile 26, and another between 0645 and 0830 at postmile 15, and indeed both bottlenecks are identified by the algorithm.

Step 3 To calculate the delay, the algorithm first delineates the space-time congested region of each bottleneck and then the delay in vehicle-hours associated with the region. As an example, the speed contour in figure 4(a) shows regions of congestion upstream of two bottleneck locations.

The n detectors divide the freeway into n segments. A segment is declared *congested* at time t if its speed is below 40 mph. The congested region associated with a bottleneck is 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 $B_j(t)$,

$$B_j(t) = \{i : v_k(t) < 40 \text{ mph, for all } i \le k \le j\}. \qquad (6)$$



(a) Bottlenck location

(b) Delay distribution

FIGURE 4 (a) Bottleneck detection on I-15 SB on 5/1/2003. Traffic flows in order of decreasing postmile. (b) Distribution of daily delay among 160 bottleneck locations; the 10 'outliers' account for 61% of the delay. Source: (Chen, C. et al., 2004b).

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

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

in which $d_i(t)$ is the delay in segment i at time t. Segment delay is the additional vehicle-hours traveled driving below the free flow speed, 60 mph,

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

Here $l_i, q_i(t)$, and $v_i(t)$ are the segment length, volume, and average speed on the segment at t. The total delay attributed to a bottleneck at segment j that is active between times t_1 and t_2 is

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

Step 4 The steps above 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 incidents or recurring conditions. 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. These are the outliers in figure 4(b).

THE PIE OF CONGESTION

Bottlenecks cause congestion, some of which can potentially be removed by ramp metering. Collisions also cause congestion. These considerations lead to the 'congestion account' (10)-(13), for a contiguous section of freeway with n detectors indexed $i = 1, \cdots, n$. Using the earlier notation,

$$d_i(t) = l_i \times q_i(t) \times \left(\frac{1}{v_i(t)} - \frac{1}{v_f} \right) \text{ vehicle-hours}, \tag{10}$$

$$D_{tot} = \sum_{i=1}^{n} \sum_{t=1}^{T} d_i(t), \tag{11}$$

$$D_{rec} = D_{tot} - D_{col}, \tag{12}$$

$$D_{tot} = D_{col} + D_{pot} + D_{rem}\,. \tag{13}$$

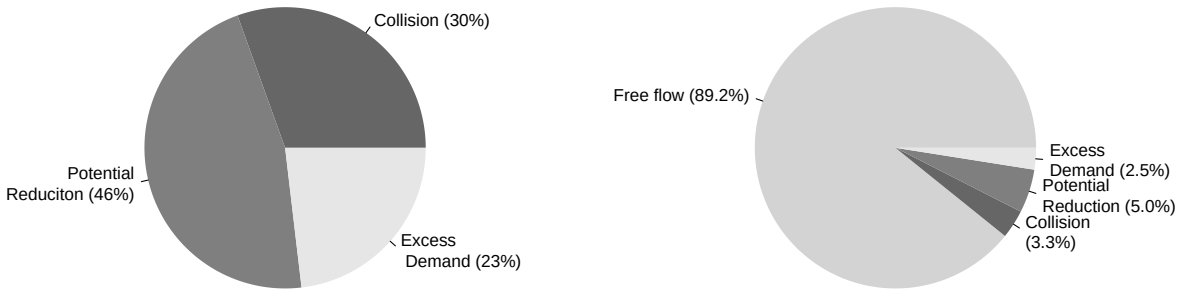
Here v_f is the reference speed, 60 mph. So $d_i(t)$ is the delay in segment i in interval t, and D_{tot} is the total delay in the section. Both $d_i(t)$ and $D_{tot}(t)$ are directly obtained from PeMS.

D_{col} is the delay caused by collisions, which has to be estimated. D_{rec}, as defined in (12), is often called the 'recurrent' congestion, much of which occurs at bottlenecks. A significant amount of D_{rec} can be potentially eliminated by ramp metering. We call this amount D_{pot}, which also needs to be estimated. Putting all these definitions together gives the summary (13) in which D_{rem} is the 'residual' congestion. D_{rem} is largely due to 'excess demand' whose impact cannot be eliminated by ramp metering, and shows up as delay at ramps. D_{rem} also includes the contribution to congestion of all other causes, such as adverse weather and special events.

The study in (Kwon, J. and P. Varaiya, 2005) proposes an automated procedure to estimate all three components in (13), using PeMS loop data and collision data from Traffic Accident Surveillance and Analysis System (TASAS) maintained by Caltrans. The procedure is applied to a 22.5 mile (postmile 4.5 to 27) section of I-15N in San Diego County. The time period is from 0500 to 2200, for 44 weekdays (2 September-31 October, 2002).

Figure 5 summarizes the study's conclusions. The total average daily congestion pie is divided into three slices. If D_{pot} and D_{rem} are reported together as D_{rec}, recurrent congestion would amount to 70%. As Hallenbach et al. (Hallenbach, M.E. et al., 2003, p. 11) observe, this large 'recurrent' congestion may in part be caused by "unusual volume surges at ramps ... that are not being effectively handled by the ramp metering program." Figure 5(a) indicates that 30% of the total congestion (or 60% of recurrent congestion) can be removed by IMP-metering that effectively handles these volume surges. Figure 5(b) summarizes traveler exposure to congestion in the study section. Travelers spend 89% of their time in the free-flow regime, and 11% in the congestion regime. The pies in figure 5 are for the study section. We are in the process of constructing congestion portraits for all freeways in California for which PeMS and TASAS data are available.

We now discuss the four-step procedure in (Kwon, J. and P. Varaiya, 2005). The first step delineates



(a) Congestion pie

(b) Exposure pie

FIGURE 5 The congestion pie (a) and the exposure pie (b) automatically constructed for the I-15N study section. Source: (Kwon, J. and P. Varaiya, 2005).

for each collision its time-space region of impact. The second step predicts how much delay this region would have experienced had that collision *not* occurred; this is the recurrent congestion. The third step calculates how much of this recurrent congestion can be eliminated by IMP-metering. The fourth step puts the estimates together in the congestion pie.

Step 1 Following a collision, congestion propagates upstream up to some maximum spatial *extent*. The congestion lasts a certain amount of time, called its *duration*. Empirically, freeway segment i is declared congested during a 5-minute interval t if the speed $v_i(t) < 50$ mph. (This is slightly different from the 40 mph criterion in (3).) Formula (10) is then used to calculate the total delay in this duration-extent 'rectangle'. The precise algorithm is similar to that of (6)-(7). The step leads to the estimate $D_{tot,a}(t)$ of the total delay at time t in the impact region of each collision a. (The same procedure can be used to delineate the impact of non-collision incidents.)

Step 2 This step predicts $D_{rec,a}(t)$, the recurrent congestion at time t that would have occurred in the absence of collision a. This is the K-nearest neighbor prediction of the recurrent delay, based on historical data of the delay $D_a(t,d)$ during the same time t and over the same spatial extent, for several other days $d = 1, \cdots, T$. More precisely, the estimate is the median value

$$D_{rec,a}(t) = \text{median}\{D_a(t, d'_k) \mid k = 1, \cdots, K\},$$

in which $d'_k, k = 1, \cdots, K$ are K days with smallest values of $|D_a(t_a, d) - D_{tot,a}(t_a)|$ for $d = 1, \cdots, D$. (In the empirical study of figure 5, $K = 3$.) Here t_a is the time just before collision a occurred. The recurrent congestion that would have occurred in the absence of collision a is

predicted to be

$$D_{rec,a} = \sum_t D_{rec,a}(t),$$

in which the sum is over the duration of impact. Finally,

$$D_{col,a} = \sum_t \max(D_{tot,a}(t) - D_{rec,a}(t), 0), \tag{14}$$

is the contribution to congestion of collision a.

Step 3 This step estimates the potential reduction in delay at recurrent bottlenecks by IMP-metering discussed in IDEAL METERING. The procedure first identifies all bottlenecks following the algorithm in BOTTLENECKS, and restricts attention to those that occur for more than 20% of the days. Next, if the bottleneck time-space region overlaps with the impact region of a collision, that day is excluded. An estimate of the reduction in delay is then computed using the procedure in IDEAL METERING. This gives an estimate of D_{pot} in (13). Details are in (Kwon, J. and P. Varaiya, 2005).

Step 4 The three delays estimated above, together with overall VHT from PeMS are displayed in the pies of figure 5.

An important side-effect of the procedure is an estimate of the delay caused by each collision, $D_{col,a}$ (14). Of 74 collisions during the study period, two-thirds cause no additional delay. These occur either when recurrent congestion is very low or very high. Eight 'outliers' (10% of collisions) account for 90% of total collision delay. Incident management could be made more effective if the high delay-causing accidents could be quickly diagnosed once they occur.

TASAS provides crash information including type of collision, number of vehicles involved, weather. From this information, we find the strongest predictor of high delay-causing accidents is the number of vehicles involved; adverse weather is a moderately strong predictor; all others, including injury and trucks, are weak predictors. Note, however, that the data set in the study contains only 74 collisions.

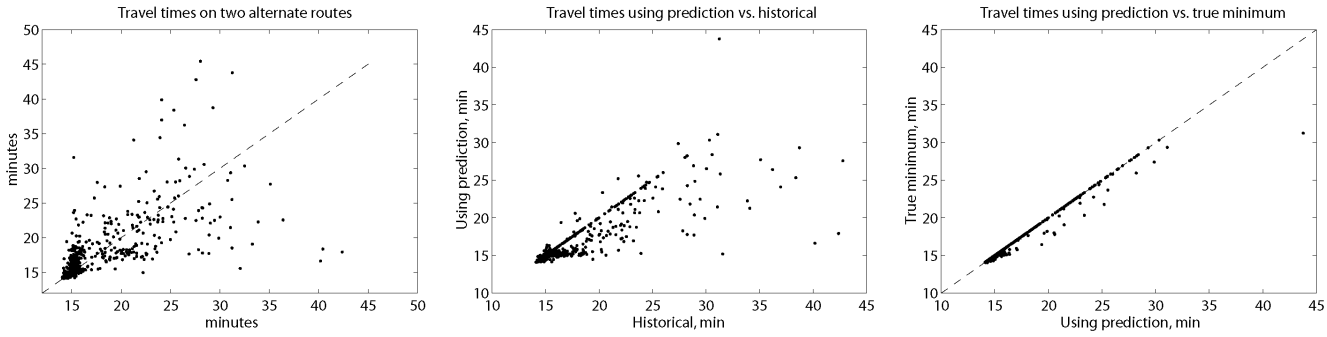
PREDICTING TRAVEL TIME

The delay on a freeway on the same day of week varies much more than the total demand. Travelers experience this variation as large uncertainty in their travel time. Let $T(t)$ be the travel time of a trip over a fixed route starting at time t. $T(t)$ is a stochastic process with trends that can be calculated from historical data, and a large variance due to congestion. Let $\sigma^2(t)$ be the (unconditional) variance of $T(t)$, and let $\sigma^2(t,s)$ be the variance of the predictor $\hat{T}(t,s)$ of this travel time conditioned on knowledge of traffic conditions up to time $s \leq t$. Because the travel time process

has a large autocorrelation, $\sigma^2(t,s)$ is much smaller than $\sigma^2(t)$. We summarize a study (Chen, C. et al., 2004a) which estimates the benefits of prediction.

The study compares travel time along two alternate routes between the I-5/I-805 interchange and the I-5/I-163 interchange in San Diego. Route 1 is entirely along I-5S, Route 2 has its first segment on I-805 and the second segment on I-163S. Travel times $T_1(t)$ and $T_2(t)$ along the two routes are computed for departure times t between 0500 and 2200 during the 22 weekdays between 1 and 31 August, 2002. There are 1320 departure times over the study period at every 17 minutes.

Each point in the scatter plot of figure 6(a) represents $(T_1(t), T_2(t))$ with the same departure time t. There are 1320 points. Two features of the scatter plot are clear. First, the travel time distributions on the two routes are similar. Second, there is a large uncertainty: 90% of the distribution lies between 12 and 35 minutes, with a median below 20 min.



(a) Travel times (b) Prediction vs. historical (c) Prediction vs. optimum

FIGURE 6 Scatter plot of travel times along the two routes (a). Comparison of minimum predicted travel time vs. historical (b) and vs. true minimum travel time (c). Source: (Chen, C. et al., 2004a).

A PeMS application predicts travel time $\hat{T}(t,t)$ for a trip starting at any time t, based on historical data and real time data available up to time t (van Zwet, E. and J. Rice, 2001). We now estimate the travel time savings using the PeMS prediction. Figure 6(b) compares the travel time that would be experienced by a traveler who selects the route with the shorter predicted travel time $\min_i \hat{T}_i(t,t)$, with that of a traveler who selects the route with the shorter expected travel time $\min_i ET_i(t)$, which can be estimated from historical data alone. Most of the points lie on or below the 45 degree line, indicating that reliance on PeMS prediction is much better than historical experience. The travel time saving is the horizontal distance to the 45 degree line.

Figure 6(c) compares the travel time based on PeMS prediction with that of a clairvoyant traveler who unerringly chooses the route with the shorter travel time, $\min_i T_i(t)$. Naturally, all points lie below the 45 degree line, but the significant feature is how frequently the points lie on the 45

degree line, indicating that prediction correctly selects the *ex post* shorter route.

When there are alternative routes as is the case here, accurate travel time prediction reduces both the average travel time and the uncertainty. Even when alternative routes are not available, the reduction in uncertainty increases traveler welfare. Estimates in (Chen, C. et al., 2004a) suggest that the benefits are significant for the example presented here.

As a final remark we note that the travel time estimate in (van Zwet, E. and J. Rice, 2001) involves *predicting* the traffic conditions that the traveler will encounter along the route. Such a predictor performs much better than the commonly used predictor which simply adds up the most recently reported travel times on the segments along the route.

EVALUATION OF HOV LANE EFFECTIVENESS

Several studies reach the obvious conclusion that HOV travelers benefit from lower travel times, see e.g. (DKS Associates, 2003; The PB Study Team, 2002). But these studies do *not* evaluate the impact of HOV lanes on overall congestion, including the congestion on mixed-flow lanes. San Francisco Bay Area data are especially helpful in evaluating this impact, because its HOV lanes are time-actuated. To facilitate comparison, the evidence below is for freeways with heavier PM peak traffic. In all cases, lane 1 (the fast lane) is HOV actuated on weekdays between 0500-0900 (5:00-9:00 AM) in the morning and 1500-1900 (3:00-7:00 PM) in the evening; at all other times HOV is deactuated. We argue that in the Bay Area, HOV lanes *increase* overall congestion.

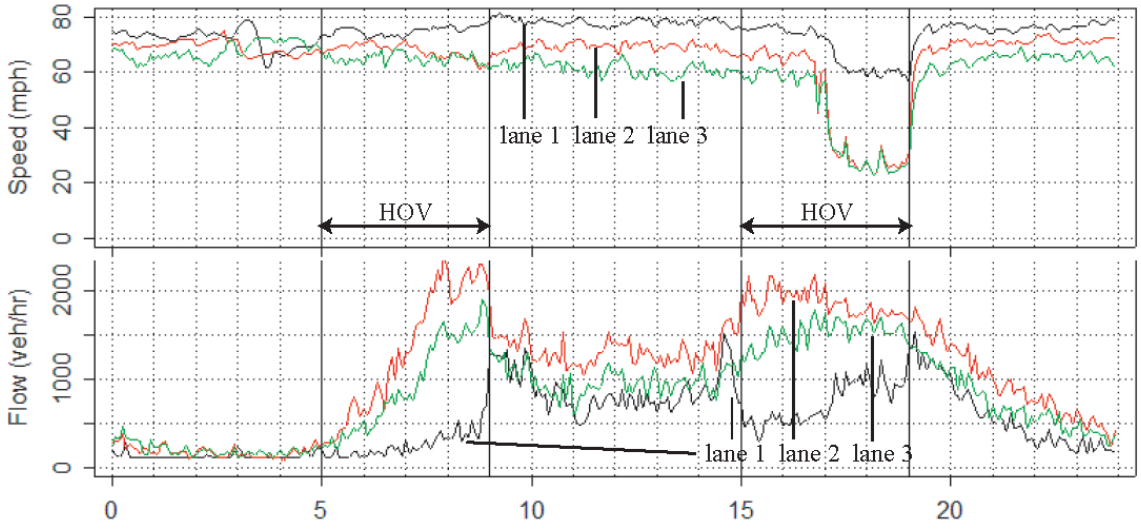


FIGURE 7 Speed and flow in lanes 1(HOV), 2 and 3 on 18 August 2004 at VDS 400104 on SR-237E.

Figure 7 shows speed and flow on all three lanes, 1(HOV), 2 and 3, of SR-237E at a particular

location on 18 August 2004. During the 0500-0900 HOV actuation period, the HOV lane is under-utilized, but since overall traffic is low, all lanes are in the free flow regime. (Most HOV lanes in the off-peak direction are underused (DKS Associates, 2003, Table 3, p.7).)

Immediately after deactuation at 0900, speed and flow are (nearly) equalized on all lanes, and they remain in the free flow regime until HOV re-actuation at 1500. At 1500 HOV flow drops dramatically, compensated by increased flows in lanes 2 and 3. But until 1700, all three lanes remain in the free flow regime, and flows in lanes 2 and 3 reach a maximum. From 1700 until 1900, HOV flow increases and speed decreases, and the HOV lane remains in free flow. However, lanes 2 and 3 enter the congestion regime. They suffer a large reduction in *both* speed and flow. The decline in flow is severe enough to reach the level of the HOV lane at 1900.

The impact of HOV actuation on overall congestion can be seen by comparing the behavior before and after HOV deactuation at 1900 in figure 7. HOV activation during 1700-1900 reduces capacity for non-HOV traffic (which loses one lane), pushes non-HOV lanes into the congestion regime, and reduces total non-HOV flow. Thus traffic suffers a *non-HOV congestion penalty*. Shortly after deactuation at 1900, all lanes enter the free flow regime, and total flow reaches a maximum over the entire day. More surprisingly, even HOV lane performance improves after deactuation: both speed and flow increase. Put inversely, both speed and flow in the HOV lane decline during HOV actuation, even though it is in free flow. We call this the *HOV capacity penalty*. In summary: HOV actuation imposes a congestion penalty on non-HOV lanes and a capacity penalty on the HOV lane.

The HOV capacity penalty—increased HOV speed and flow *after* deactuation—is seen in the six freeway locations we examined. Figure 8 shows speeds in six different freeways during 1400-2000, starting one hour before the afternoon HOV actuation at 1500 and ending one hour after HOV deactuation at 1900. (Flows are not shown as they have the expected behavior, similar to that in figure 7.) In all cases, speeds in all lanes, including lane 1(HOV), increase after deactuation; moreover, flow in lane 1(HOV) increases, and flows in the other lanes decrease.

The lane 1(HOV) capacity penalty is explained as follows. The flow increases after deactuation because drivers in lane 2 move into the lower density lane 1. The speed decreases during HOV actuation because the HOV lane becomes a one-lane highway whose speed is governed by the low speed vehicles–the 'snails'. As the non-HOV congested lanes are even slower, a faster HOV driver cannot pass the slower snail in front of it. However, as soon as HOV is deactuated, slower drivers move to the outer lanes and the fastest drivers move to (what was) the HOV lane. Speed in all lanes increase—usually dramatically as in figure 8.

The hypothesis that during HOV actuation speed is controlled by snails is confirmed in the scatter plots of figure 9. Each point is a 5-minute average of flow and speed. Plot (a), during HOV actuation, shows a sharp decrease in speed as flow (and hence the number of snails) increases, even though the lane is in free flow. Plot (b) shows no decrease in speed, as only the fast drivers

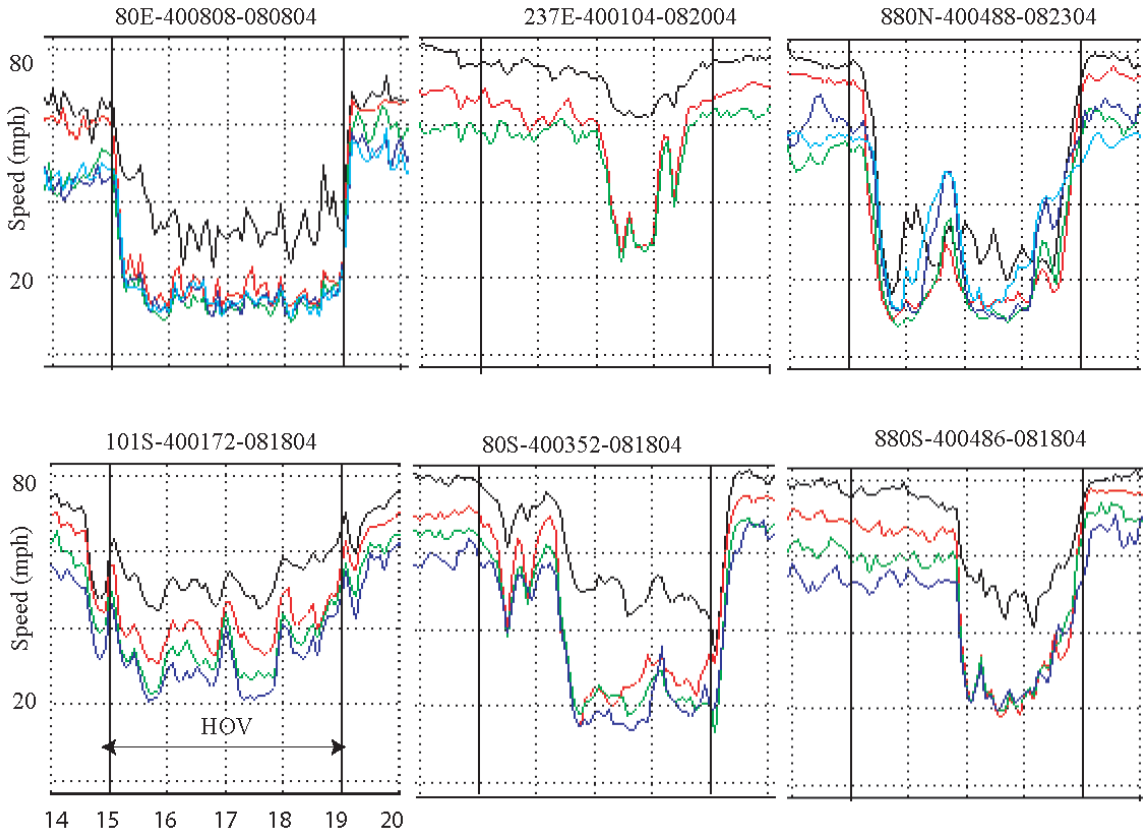
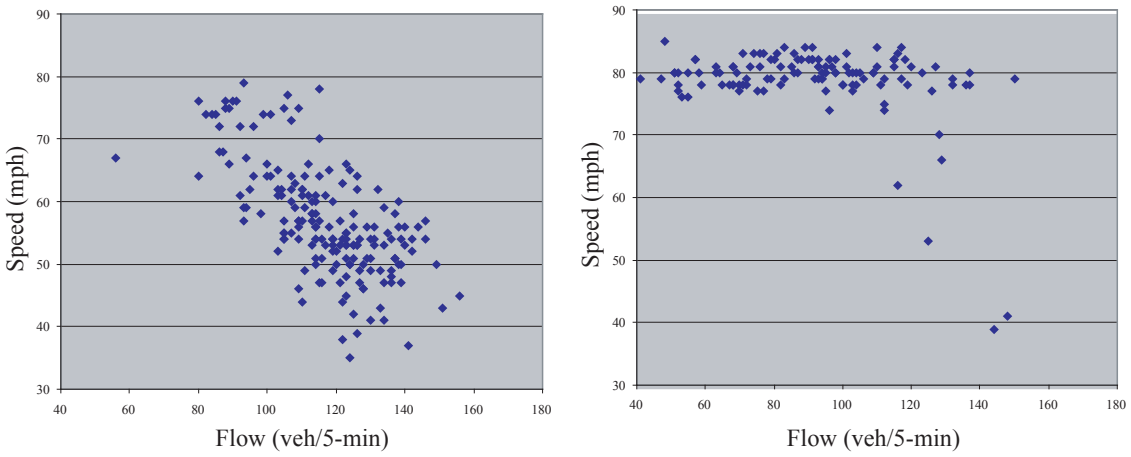


FIGURE 8 Speeds in all lanes at locations on six different Bay Area freeways, 1400-2000, beginning one hour before HOV activation (at 1500) and ending one hour after HOV activation (at 1900). In all cases, speed is highest in lane 1 (HOV), followed by lane 2, lane 3, etc. The notation 80E-400808-080804 means VDS 400808 on I-80E on August 8, 2004.

are in lane 1. The difference between plots (a) and (b) is typical of a one-lane vs. a multi-lane highway in free flow.

Three different (non-exclusive) causes may account for snails. A proportion of HOV drivers may be intrinsically slow, so their number grows as HOV flow increases. Second, the slowdown may be caused by lane changes by HOV drivers (and SOV violators) from the slower lane 2 into the HOV lane. The lane changes increase in proportion with HOV lane flow, further reducing HOV lane speed. Third, as the speed differential between the HOV and the adjacent non-HOV lane increases, drivers in the HOV lane may slow down due to the increased perceived risk of a collision should someone from the non-HOV lane merge into the HOV lane. In the last two cases, an HOV lane that is physically separated from lane 2 would not exhibit the slowdown seen in figure 9 (a). In either case, the slowdown would not be seen in freeways with two HOV lanes.

We finally arrive at the interesting question: “Will the overall congestion in the six cases in figure



(a) 1600-1900, HOV actuated

(b) 1900-2100, HOV de-actuated

FIGURE 9 Speed vs. flow (5-min averages) in lane 1, (a) 1600-1900, HOV actuated, and (b) 1900-2100, HOV de-actuated, for five weekdays in August, 2004, at VDS 400352 on I-880S.

8 be reduced by eliminating the HOV lane?" The answer would be unreservedly 'yes', but for two qualifications: one having to do with freeway management, the other with mode choice. It is obvious that a management strategy with no HOV lane and no metering will lead to greater congestion than a strategy with one HOV lane and no metering, because HOV actuation serves as a (one-lane) metering mechanism. So to fairly compare an HOV vs. a non-HOV regime, we must assume that proper ramp metering is in place to guarantee *vehicle* flow in non-HOV lanes that is close to maximum observed vehicle flow.

The second qualification is more interesting. It is based on either of two claims: (1) HOV lanes move significantly more people overall (even if they don't move more vehicles), (2) HOV lanes induce enough drivers to switch from SOV to HOV to compensate for both the congestion penalty imposed on non-HOV lanes and the capacity penalty imposed on the HOV lane by HOV actuation. We cannot address the second claim because there are no empirical estimates of the SOV-HOV mode shift for the Bay Area. We evaluate the first claim that HOV actuation increases flow of persons/hour.

We calculate flow of persons per hour (PPH) by multiplying vehicle flow (from PeMS) and AVO (average vehicle occupancy). Since the accuracy of vehicle counts exceeds 90-95%, the single most important empirical quantity in any study of HOV effectiveness is the AVO. Unfortunately, AVO estimates are very unreliable for many reasons (Levine, N. and M. Wachs, 1994), so we will use a range of estimates.

According to (California Department of Transportation, District 4, Office of Highway Operations, 2002, p. 66) on the section of I-880S that includes VDS 400486 in figure 8, during the afternoon peak hour, the HOV lane AVO is 2.1, and the AVO on the three non-HOV lanes is 1.1. We use

these estimates for the HOV actuation period. (The HOV AVO rate should be reduced by a highly variable HOV violation rate measured at 5.8% on 5 July, 2002.)

AVO estimates during HOV deactuation are not available, and we have several alternatives: the State Household Travel Survey gives an AVO of 1.5 for all trips and 1.1 for home to work trips; the Metropolitan Transportation Commission for the Bay Area gives an AVO of 1.4 for all trips and 1.1 for home to work trips; lastly, the California Life-Cycle Benefit/Cost Analysis Model uses a default of 1.38 for peak period AVO. We will use 1.25, 1.3 and 1.4 for AVO during HOV deactuation. Figure 10 (a) plots the flow in persons per 5-minutes, aggregated over all lanes, with HOV AVO = 2.1 and non-HOV AVO = 1.1 during HOV actuation, and AVO = 1.25, 1.3 or 1.4 during HOV deactuation.

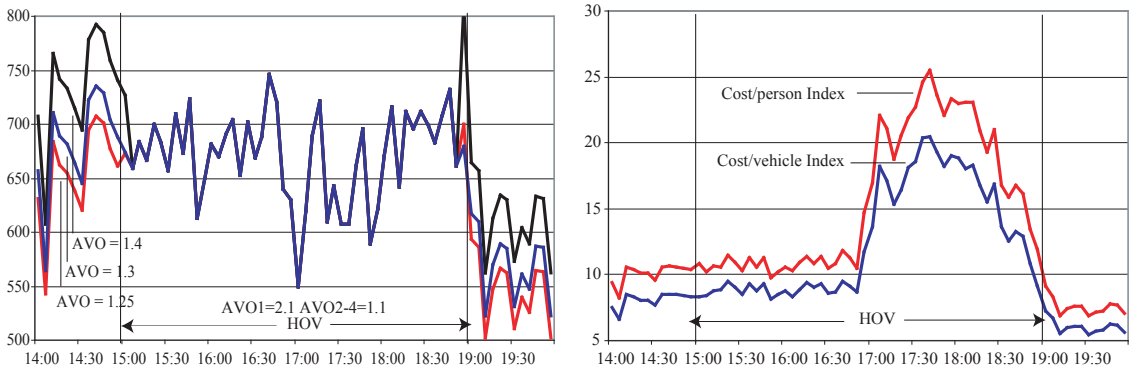


FIGURE 10 (a) Flow in persons per 5-min using the indicated AVO values, and (b) cost index per person-mile and per vehicle-mile, 1400-2000, 18 August 2004, at VDS 400486, I-880S.

With the two higher AVO estimates, HOV actuation causes a *reduction* in the flow of persons per hour compared with the period 1400-1500 before actuation. With the lowest AVO estimate, HOV actuation causes a small *increase* in PPH compared with the period 1400-1500. So the data do not support the claim that HOV actuation significantly increases (say by 10%) the flow in persons per hour.

In comparing the HOV vs. non-HOV regime, we should not ignore the travel time cost imposed by HOV actuation. Knowing the speed and the flow in persons/5-min and vehicles/5-min in each lane, we can calculate the amount of time that each person and vehicle takes to travel a fixed distance. This gives us a 'cost index', which will vary over time, as the flow and speed vary. Figure 10 (b) displays the two cost indices (AVO = 1.25 is used for these plots). Evidently, the average person (on *all* lanes) pays a travel time cost during HOV actuation (1700-1900) that is two-and-a-half times higher. Of course, a significant part of this higher cost is due to inadequate ramp metering.

If we think of the freeway as a 'people-mover' and the cost of its operation to be travel time, we must conclude that the cost is increased during HOV actuation. This is a much better indicator of productivity *loss* than the productivity *gain* measured as the ratio between HOV AVO and non-

HOV AVO in (DKS Associates, 2003, p.6,8). The latter producitivity gain merely reflects the fact that HOV actuation causes carpools to move into the HOV lane.

We close this section with some remarks. First, the analysis above leads to conclusions that run counter to those reached by most studies of HOV effectiveness. Because the evidence presented here is fragmentary, the analysis must be repeated with a more complete data set before the conclusions can be trusted.

Second, it is possible from Bay Area data to estimate the SOV-HOV mode shift, based on the hypothesis that the shift will be more pronounced the larger is the travel time differential between HOV and non-HOV lanes. Also, people may find the SOV-HOV shift to be less inconvenient on some routes than on others.

Third, when a 2+ (i.e. two or more persons) HOV lane becomes congested, it is sometimes converted to a 3+ lane. The SOV-3+HOV shift will certainly be lower than the SOV-2+HOV shift. So the conversion from 2+ to 3+HOV lane may increase overall congestion.

Lastly, because HOV lanes in the Bay Area are time-actuated, it is straightforward to estimate both the non-HOV congestion penalty and the HOV capacity penalty. This distinction is less obvious in a 24-hour HOV facility, although it, too, imposes both penalties.

There is interest in increasing the utilization of underused HOV lanes by converting them into HOT (HOV/Toll) lanes. The snail phenomenon implies, however, that even modest increases in volume following conversion will bring down HOT speed to that of non-HOT lanes (which, moreover, will have higher speed because they carry less traffic). That is, the HOV capacity penalty does not leave much room for additional traffic, so that even the cautious estimates for revenue enhancement in the Bay Area may be overly optimistic (DKS Associates, 2003, p. 22). A recent proposal to permit hybrid vehicles into HOV lanes will certainly increase congestion.

From a purely technical viewpoint, this discussion suggests that a better way to manage freeways is to eliminate HOV lanes, institute ramp metering, and permit HOV/HOT bypass at ramps. This will eliminate the HOV penalties, while encouraging mode shift from SOV to HOV. On the other hand, by having weak or no ramp metering, the HOV regime can always be made to look better. For policy considerations this technical viewpoint has to be weighed with many other factors.

CONCLUSIONS

In its draft *Transportation Management Systems (TMS) Master Plan* (System Metrics Group, Inc., 2003), Caltrans proposes an action plan to improve incident management, traffic control, and traveler information. Central to the plan is its reliance on specific performance indicators to serve “as monitoring and evaluation tools, and establish an accountability framework for the implementation

of planned TMS improvements." Caltrans has invested significant resources to develop a performance measurement system (PeMS). This paper illustrates why PeMS became a major source of 'performance indicators' and suggestions for performance targets.

The paper uses PeMS data to study freeway congestion from six different perspectives, ranging from identification of bottlenecks to evaluating the benefits of ramp metering and the effectiveness of HOV lanes. In each study, the aim is to measure the severity of congestion and reveal the opportunity for improvement. The approach is to argue on the basis of statistical models that the data are used to estimate. Qualitatively of course the models are inspired by prior theory, but the emphasis is always on quantitative conclusions.

Partly motivated by the success of PeMS, some universities and DoTs are developing small-scale protoypes of PeMS-like systems. These efforts will have a small impact until state DoTs invest in data collection infrastructure. The availability of these data will shift DoT focus from construction to operations improvements. Academic research, too, will change as it exploits opportunities opened up by access to large-scale data sets and pays more attention to questions that address the opportunities for operations improvements and conducting experiments that demonstrate improvements.

ACKNOWLEDGEMENT

We are grateful for comments and criticsm from Professor Martin Wachs of U.C. Berkeley; Robert Copp, Fred Dial and David Seriani of Caltrans; and Tarek Hatata of the System Metrics Group. This study is partly based on research that was supported by grants from Caltrans to the California PATH Program.

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

California Department of Transportation, District 4, Office of Highway Operations. HOV lanes in the Bay Area, 2002.

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

Chen, C. and P. Varaiya. Max flow in D12 occurs at 60 mph. `http://pems.eecs.berkeley.edu`, October 2001.

Chen, C., A. Skabardonis, and P. Varaiya. A system for displaying travel times on changeable message signs. In *Proceedings of 83rd Transportation Research Board Annual Meeting*, Washington, D.C., January 2004a.

Chen, C., A. Skabardonis, and P. Varaiya. Systematic identification of freeway bottlenecks. In *Proceedings of 83rd Transportation Research Board Annual Meeting*, Washington, D.C., January 2004b.

DKS Associates. 2002 High Occupancy Vehicle (HOV) Lane Master Plan Update. Prepared for Metropolitan Transportation Commision, Caltrans District 4 and the California Highway Patrol Golden Gate Division, March 2003.

Hallenbach, M.E., J.M. Ishimaru, and J. Nee. Measurement of recurring versus non-recurring congestion. Washington State Transportation Center (TRAC), October 2003.

Jia, Z., P. Varaiya, C. Chen, K. Petty, and A. Skabardonis. Congestion, excess demand and effective capacity in California freeways. `http://pems.eecs.berkeley.edu`, December 2000.

Jia, Z., P. Varaiya, C. Chen, K. Petty, and A. Skabardonis. Maximum throughput in LA occurs at 60 mph. `http://pems.eecs.berkeley.edu`, January 2001.

Kwon, J. and P. Varaiya. The congestion pie: delay from collisions, potential ramp metering gain, and excess demand. In *Proceedings of 84th Transportation Research Board Annual Meeting*, Washington, D.C., January 2005.

Levine, N. and M. Wachs. Methodology for Vehicle Occupancy Measurement. Report submitted to the California Air Resources Board and the California Department of Transportation (Office of Traffic Improvement)., 1994.

Martin, W. Travel estimation techniques for urban planning. Technical report, NCHRP Report 365, Transportation Research Board, Washington, D.C., 1998.

Papageorgiou, M. *Applications of Automatic Control Concepts to Traffic Flow Modeling and Control*. Lecture Notes in Control and Information Sciences, 50. Springer, 1983.

Papageorgiou, M., H. Hadj-Salem, and J. Blosseville. ALINEA: a local feedback control law for on-ramp metering. *Transportation Research Record*, 1320, 1991.

System Metrics Group, Inc. Transportation Management Systems Plan, 2003.

The PB Study Team. HOV Performance Program Evaluation Report. Los Angeles County Metropolitan Transportation Authority, 2002.

van Zwet, E. and J. Rice. A simple and effective method for predicting travel times on freeways. In *2001 IEEE Intelligent Transportation Systems Proceedings*, pages 227–232, Oakland, CA, 2001.

The Components of Congestion: Delay from Incidents, Special Events, Lane Closures, Weather, Potential Ramp Metering Gain, and Excess Demand

Jaimyoung Kwon*
Department of Statistics
California State University, East Bay
Hayward, CA 94542
Tel: (510) 885-3447, Fax: (510) 885-4714
jaimyoung.kwon@csueastbay.edu

Michael Mauch
DKS Associates
8950 Cal Center Drive, Suite 340
Sacramento, CA 95826-3225
Tel: (916) 368-2000, Fax: (916) 368-1020
mvm@dksassociates.com

Pravin Varaiya
Department of Electrical Engineering and Computer Science
University of California, Berkeley CA 94720
Tel: (510) 642-5270, Fax: (510) 642-7815
varaiya@eecs.berkeley.edu

For Presentation and Publication
85th Annual Meeting
Transportation Research Board
January 2006
Washington, D.C.

July 30, 2005

Words: 4,279 (excluding Figure and Table captions)
Plus 2 Table and 3 Figures (1,250)
TOTAL: 5,529

*Corresponding Author

ABSTRACT

A method is presented to divide the total congestion delay in a freeway section into six components: the delay caused by incidents, special events, lane closures, and adverse weather; the potential reduction in delay at bottlenecks that ideal ramp metering can achieve; and the remaining delay, due mainly to excess demand. The fully automated method involves two steps. First, the components of non-recurrent congestion are estimated by statistical regression. Second, the method locates all bottlenecks and estimates the potential reduction in delay that ideal ramp metering can achieve. The method can be applied to any site with minimum calibration. It requires data about traffic volume and speed; the time and location of incidents, special events and lane closures; and adverse weather. Applied to a 45-mile section of I-880 in the San Francisco Bay Area, the method reveals that incidents, special events, rain, potential reduction by ideal ramp metering, and excess demand respectively account for 13.3%, 4.5%, 1.6% 33.2% and 47.4% of the total daily delay. The delay distribution of the various components is different between the AM and PM peak periods and between the two freeway directions. Quantifying the components of congestion at individual freeway sites is essential in developing effective congestion mitigation strategies.

Keywords: freeway congestion; incidents; weather; ramp metering; loop detectors

1. INTRODUCTION

Congestion is caused by incidents, special events, lane closures, weather, inefficient operations, and excess demand. Their impact can be summarized in the division of the congestion 'pie' into its component as in Figure 1. Knowledge of the congestion pie is essential to the selection of effective congestion mitigation strategies (*1*).

The paper presents a method to divide the total congestion D_{total} into six components: (1) D_{col}, the congestion caused by incidents, which could be reduced by quicker response; (2) D_{event}, the congestion caused by special events, which could be reduced by public information and coordination with transit; (3) D_{lane}, the congestion caused by lane closures, which could be reduced by better scheduling of lane closures; (4) $D_{weather}$, the congestion caused by adverse weather, which could be reduced by demand management and a better weather response system; (5) D_{pot}, the congestion that can be eliminated by ideal ramp metering; and (6) the residual delay, D_{excess}, largely caused by demand that exceeds the maximum sustainable flow. The method is applied to a 45-mile section of I-880 in the San Francisco Bay Area, using data for January-June, 2004.

The method refines previous studies (*2*,*3*,*4*) that group D_{pot} and D_{excess} together as 'recurrent' congestion. It also refines our recent work (*15*), which considers only three components (D_{col}, D_{pot} and D_{excess}). Transportation agencies measure recurrent congestion in various ways, and find it accounts for 40%-70% of total congestion (*5*). The availability of more comprehensive data has prompted attempts to separately estimate the contribution of different causes of congestion. There are studies that divide total congestion into 'recurrent' and 'non-recurrent' congestion; and studies that divide the non-recurrent congestion into accident-induced congestion and other incident-induced congestion. There also are estimates of the congestion caused by adverse weather. These studies are reviewed in the next section.

These studies leave a large fraction (between 40 and 70 percent) of the total congestion unexplained. This unexplained residual is often called 'recurrent' congestion. As Hallenbeck et al. observe, "Many large delays still occur for which incidents are not responsible, and for which no 'cause' is present in the [data]." They suggest that one cause of these delays may be "unusual volume surges at ramps ... that are not being effectively handled by the ramp metering program" (*2*, p.11). The proposed method estimates this potential reduction in delay, D_{pot}.

The paper is organized as follows. Previous studies are reviewed in Section 2. The proposed method is described in Section 3. The congestion components of I-880 are determined in Section 4. Section 5 concludes the paper.

2. PREVIOUS STUDIES

Transportation agencies until recently only reported recurrent congestion. (For an example see (*7*); for an extensive survey of the practice see (*5*).) The availability of more comprehensive data has inspired studies to quantify the relative impact of different causes of congestion.

Several studies estimate the impact of incidents. The earliest studies relied on correlating specially-collected incident data using 'floating cars' with loop-detector data (*8*). These data provide a great deal of information about the nature of incidents, but the data collection efforts are too expensive to replicate on a large scale or on a continuing basis.

Date from California Highway Patrol computer aided dispatch (CAD) and Freeway Service Patrol (FSP) logs were used to evaluate FSP effectiveness in Los Angeles freeways (*9*) and in Oregon (*10*). These studies need much human effort, data analysis skill, and subjective judgment in determining the spatial and temporal region of the congestion impact of an incident. Our previous work (*15*) developed an automated method to delineate an incident's impact region. But that approach requires accurate time and location of incidents, which may not be available.

Determining every individual incident's impact region can be avoided if one is willing to average out the impact of individual incidents as in (*2*, *3*). Both studies separate 'non-recurrent' and 'recurrent' congestion, but they differ in definition and method.

Skabardonis et al. (*3*) consider a freeway section during a peak period. The total congestion on each of several days is calculated as the additional vehicle-hours spent driving below 60 mph (see equation (1) below). Each day is classified as 'incident-free' or 'incident-present'. The average congestion in 'incident-free' days is defined to be the recurrent delay. Total congestion in incident-present days is considered to be the sum of recurrent and incident-induced congestion. Subtracting average recurrent congestion from this gives an estimate of the average non-recurrent or incident-induced congestion. On the other hand, Hallenbeck et al. (*2*) take the median traffic conditions on days when a freeway section does *not* experience lane-blocking incidents as the "expected, recurring condition."

A less data-intensive approach is taken by Bremmer et al. (*4*). In the absence of incident data, they simply assume that an incident has occurred if a trip "takes twice as long as a free-flow trip for that route." The aim of this study is to forecast travel times, measure travel time reliability, and conduct cost-benefit analysis of operational improvements, rather than to measure the congestion contribution of different causes.

Lastly, the impact of inclement weather on freeway congestion is studied in (*11*, Chapter 22) and (*12*), which find that light rain or snow, heavy rain, and heavy snow reduces traffic speed by 10, 16, and 40 percent, respectively.

3. PROPOSED METHOD

The method applies to a contiguous section of freeway with n detectors indexed $i = 1,\ldots, n$, whose flow (volume) and speed measurements are averaged over 5-minute intervals indexed $t = 1,\ldots, T$. Days in the study period are denoted by $d = 1,2,\ldots,N$. Detector i is located at postmile x_i; $v_i(d,t) = v(x_i, d, t)$ is its speed (miles per hour, mph) and $q_i(d,t) = q(x_i, d, t)$ is its flow (vehicles per hour, vph) at time t of day d.

The n detectors divide the freeway into n segments. Each segment's (congestion) delay is defined as the additional vehicle-hours traveled driving below free flow speed v_{ref}, taken to be 60 mph. So the delay in segment i in time t is

$$D_i(d,t) = l_i \times q_i(d,t) \times \max\{1/v_i(d,t) - 1/v_{ref}, 0\} \text{ vehicle-hours}, \quad (1)$$

in which l_i is the segment length in miles. The total delay in the freeway section on day d is the delay over all segments and times,

$$D_{total}(d) = \sum_{i=1}^{n} \sum_{t=1}^{T} D_i(d,t). \quad (2)$$

The average daily total delay is simply

$$D_{total} = \frac{1}{N} \sum_{d=1}^{N} D_{total}(d). \quad (3)$$

In the application below we separately consider the daily delay over two peak periods, 5-10 AM for the morning peak and 3-8 PM for the afternoon peak.

Incidents are indexed a = 1, 2, … . The time τ_a when incident a occurs and its location σ_a are approximately known. The incident clearance time and the spatial and temporal region of the incident's impact are not known.

Decomposition of Delay

The method divides the average daily total delay (3) into six components,

$$D_{total} = D_{col} + D_{event} + D_{lane} + D_{weather} + D_{pot} + D_{excess}. \quad (4)$$

It will be useful to define

$$D_{non-rec} = D_{col} + D_{event} + D_{lane} + D_{weather}, \quad (5)$$
$$D_{rec} = D_{tot} - D_{non-rec} = D_{pot} + D_{excess}. \quad (6)$$

Above,

D_{col} is the daily delay caused by incidents,
D_{event} is the daily delay caused by special events,
D_{lane} is the daily delay caused by lane closure,
$D_{weather}$ is the daily delay caused by adverse weather condition,
D_{pot} is the potential reduction of D_{rec} by ramp metering,
D_{excess} is the residual delay, attributed mostly to excess demand,
D_{rec} is the daily 'recurrent' delay, and
$D_{non\text{-}rec}$ is the daily 'recurrent' delay.

D_{total}, calculated from flow and speed data, is the average daily total delay. D_{col}, D_{event}, D_{lane} and $D_{weather}$ are components of so-called 'non-recurrent' congestion. The difference between their

sum and D_{total} is the 'recurrent' congestion (*2*, *3*). A portion of recurrent congestion due to frequently occurring bottlenecks could, in principle, be reduced by ramp metering. That potential reduction is estimated as D_{pot}. The remaining delay, D_{excess}, is due to all other causes, most of which is likely due to demand in excess of the maximum sustainable flow. The delay due to excess demand can only be reduced by changing trip patterns. We now describe how each component of (4) is estimated.

Non-Recurrent Delays

The components of non-recurrent delay are identified using the following model,

$$D_{total}(d) = \beta_0 + \beta_{col} X_{col}(d) + \beta_{event} X_{event}(d) + \beta_{lane} X_{lane}(d) + \beta_{weather} X_{weather}(d) + \varepsilon(d), \quad (7)$$

Where

$\varepsilon(d)$ is the error term with mean zero,
$X_{col}(d)$ is the number of incidents on day d,
$X_{event}(d)$ is the number of congestion-inducing special events such as sport games on day d,
$X_{lane}(d)$ is the number of lane-closures on day d, and
$X_{weather}(d)$ is the 0-1 indicator of adverse weather condition on day d.

The explanatory variables listed above are used in our application, but the list could be augmented if additional data are available. For example, $X_{event}(d)$ could be the attendance at special events instead of the number of special events; $X_{lane}(d)$ could be the duration instead of the number of lane closures; and $X_{weather}(d)$ could be the precipitation (as in our application).

The model assumes that each incident, special event, lane-closure, and adverse weather condition contributes linearly to the delay. Figure 2 illustrates that such model is reasonable for our study site. More complicated causality between explanatory variables, such as between the bad weather and the number of accidents, is not considered to keep the number of parameters in the model small. But if one has enough data and the interaction is strong enough, such interaction terms could be included. (For the San Francisco Bay Area data considered below, the correlation coefficient between precipitation and number of accidents is only 0.032.)

Fitting the model to the data via linear least squares gives the parameter estimates, again denoted β_0, β_{col}, β_{event} , β_{lane} and $\beta_{weather}$. The components of the total delay then are

$$D_{col} = \beta_{col} \times \text{avg}\{X_{col}(d)\}, \quad (8)$$
$$D_{event} = \beta_{event} \times \text{avg}\{X_{event}(d)\}, \quad (9)$$
$$D_{lane} = \beta_{lane} \times \text{avg}\{X_{lane}(d)\}, \text{ and} \quad (10)$$
$$D_{weather} = \beta_{weather} \times \text{avg}\{X_{weather}(d)\}, \quad (11)$$

in which the average is taken over days, $d = 1,\ldots,N$.

The intercept β_0 in (7) is the delay when there are no incidents, special events, lane-closures, or adverse weather. Thus, consistent with convention, it may be identified with recurrent congestion, since it equals total delay minus the non-recurrent delay $D_{non\text{-}rec}$ defined above,

$$\beta_0 = D_{rec} = D_{total} - D_{non\text{-}rec}. \quad (12)$$

Recurrent Delay Algorithm: Separating Recurrent and Non-recurrent Congestion

The next step is to divide the recurrent delay into the delay that can be eliminated by ramp metering and the delay due to excess demand. For this, the method identifies recurrent bottlenecks on the freeway section using the automatic bottleneck identification algorithm proposed in (*13*). Then the ideal ramp metering (IRM) is run on those recurrent bottlenecks that are activated on more than 20% of the weekdays considered (*14*, *15*).

Here is a brief description of the IRM algorithm. For a specific recurrent bottleneck, let segment *i* and *j* be the upstream and downstream boundaries of the bottleneck, respectively. For the upstream boundary *j*, we use the median queue length of the bottleneck. Then we compute the total peak period volume at the two locations. The difference between the two would be the difference between the total number of cars incoming or exiting the freeway between the two segments. We assume that all those cars contributing to the difference are arriving (or leaving) at a virtual on-ramp (off-ramp) at the upstream segment *i*. Also, the time-series profile of that extra traffic is assumed identical to the average of those at segment *i* and *j*. That enables us to compute the modified total input volume profile at the segment *i*. The capacity of the whole section is the maximum sustainable (over 15-minute) throughput at location *j* and we compute this from the empirical data. We meter the virtual input volume at segment *i* at 90% of C_j to prevent the breakdown of the system, assuming:

(1) The metered traffic will be free flow (60 mph) throughout the freeway section, and
(2) The upstream meter has infinite capacity.

Thus, under IRM, the delay occurs only at the meters. The potential savings from IRM at these bottlenecks for each day *d* is then computed as,

$$D_{pot}(d) = D_{BN,\ before\ IRM}(d) - D_{BN,\ after\ IRM}(d). \quad (13)$$

Here $D_{BN,\ before\ IRM}(d)$ and $D_{BN,\ after\ IRM}(d)$ is the delay at the bottlenecks before and after IRM is run. The average daily potential saving is

$$D_{pot} = \min\ \{\text{median}(D_{pot}(d),\ d = 1,\ \ldots),\ D_{\text{rec}}\}. \quad (14)$$

In (14) the median instead of the mean is used to ensure that the influence of incidents and special events etc. is minimized in the computation. Also, the potential saving can't be larger than the total recurrent delay D_{rec}.

Due to the 'ideal' nature of IRM, D_{pot} need to be interpreted with caution. Especially, the assumption of a very large, though not infinite, capacity at the meter is not realistic for many

urban freeways and metering at certain locations can lead to breakdown of arterial traffics nearby. Thus, it is recommended that D_{pot} be viewed as the maximum possible saving in the recurrent delay by metering.

Congestion Pie

The method described above divides the average daily total delay D_{total} into six components, summarized in easily understood pie charts like those in Figure 1.

4. CASE STUDY

The method is applied to a 45.33 mile (postmile .39 to 45.72) section of southbound (SB) and northbound (NB) I-880 in the San Francisco Bay Area. Two time periods are considered: AM peak, 5-10 AM; and PM peak, 3-8 PM. Data cover 110 weekdays during January 5–June 30, 2004. There are four scenarios, distinguished by peak period and freeway direction: SB AM, SB PM, NB AM and NB PM.

Data Sources

Traffic Speed and Volume Data

The 90 (NB) and 94 (SB) loop detector stations in the section provide 5-minute lane-aggregated volume and speed data, available at the PeMS website (*16*).

Freeway Service Patrol (FSP) Incidents

Incident data are for Freeway Service Patrol (FSP) assisted incidents. On an average non-holiday weekday the FSP assists upwards of 80 motorists on I-880 during 6:00-10:00 AM and 3:00-7:00 PM. FSP peak hours are an hour shorter than peak hours used for computing total delay (5-10 AM and 3-8 PM) but we don't expect the effect would be substantial. On weekends and holidays, FSP assistance is not provided. FSP drivers record the date and time, duration, freeway name and direction, incident description (e.g. traffic accident, flat tire, out-of-gas), and location (e.g. on- or off-ramp, left shoulder, right shoulder, in-lane). We only consider in-lane incidents (as opposed to those on the left or right shoulder or on a ramp) during peak hours. There were 829 such incidents during the study period.

Special Events

On 45 out of 110 weekdays, there were special events in the Oakland Coliseum, near postmile 36 of I-880, including baseball (the Oakland A's) and basketball (the Golden State Warriors) games and show performances, mostly starting at 7 PM. Data were provided by Networks Associates Coliseum & The Arena in Oakland.

Weather

Weather data were collected from California Department of Water Resource (DWR) for "Oakland north" (station ID "ONO") station (*17*). The station reports daily precipitation, temperature, wind speed and direction, etc; only precipitation was considered in the analysis.

Lane closure

Lane closure data were obtained from the Lane Closure System (LCS) managed by California Department of Transportation (*18*). LCS records include, for each lane closure:

> Location: freeway, direction, county, and postmile,
> Begin/End date and time,
> Facility/Lanes: on/off-ramp, # lanes, which lanes, and
> Type of work: sweeping, construction, etc.

For the first half of 2004, for NB I-880, there were 224 lane closures, 126 of them in the traffic lanes. It turns out that all day time closures were 'sweeping' or 'call box remove/repair', which involve a moving closure of at most one lane and have negligible impact on congestion. All congestion-inducing lane closures (repair, striping, and paving) occurred at night (after 10 PM and before 5 AM) or on weekends outside the AM and PM peaks. This was also the case for SB 880. Thus we assign $D_{lane} = 0$ for all scenarios.

Results

Table 1 summarizes the regression results for non-recurrent congestion. The last column shows the multiple R-squared values for each scenario, which is the ratio of the sum of squares of the delay explained by the regression model and the total sum of squares around the mean. The F-statistic for testing whether the fit of the model is valid is significant with practically zero P-value for all four scenarios, suggesting the linear regression model successfully explains the delay variation. We also observe:

1. β_{event} is statistically significant (P-value < .10) only for PM shifts. This is to be expected since most special events occur in the afternoon or evening. Each special event, on the average, contributes a delay of 1,084 and 705.5 veh-hrs for NB and SB respectively.
2. β_{col} is statistically significant (P-value < .001) only for PM shifts. This suggests that congestion in the morning peak hours is more recurrent in nature than in the afternoon/evening. In PM shifts, each incident contributes a delay of 486.13 (NB) and 383.75 (SB) veh-hrs on the average.
3. $\beta_{weather}$ is statistically significant (P-value < .001) only during AM shifts. On average, one inch of rain adds 1305.7 (NB) and 2125.6 (SB) veh-hrs of delay. Note that it rained on 29 out of 110 weekdays; the median precipitation was .13 inches, and the maximum was 2.44 inches.

Figure 2 shows the relationship between D_{total} and some of the explanatory variables illustrating the correlation between the total delay and those variables.

Next, formulas (8)-(11) are used to compute the delay components shown in Table 2. Before applying the formula, we set to zero those regression coefficients that are not statistically significant at significance level 0.1.

The automatic bottleneck detection algorithm is applied to speed data of the kind whose contour plot is shown in Figure 3. Clearly visible in the figure are an AM bottleneck near postmile 10 and a larger PM bottleneck near postmile 27. D_{pot} and D_{excess} are computed from the IRM algorithm and shown in the right columns of Table 2. About 44% of recurrent delay can potentially be eliminated by ideal ramp metering: (D_{pot} and D_{excess} are extrapolated from district wide quantities; freeway-specific computation is underway in PeMS v. 6.0.)

From the charts in Figure 1 one can conclude:

1. One-third of the congestion delay occurs at recurrent bottlenecks and can be potentially eliminated by ideal ramp metering.
2. One-half of the delay is due to excess demand in both directions, and can be reduced only by changing trip patterns.
3. Incidents and special events contribute 18% of the delay. The former can be reduced by more rapid detection and response; impact of special events may be reduced by information on changeable message signs.

The 486.13 (NB) and 383.75 (SB) vehicle-hours of delay per incident for the PM shift is in rough agreement with other estimates. A regression of total daily delay vs. number of accidents for all of Los Angeles yields a slope of 560 vehicle-hours per accident (*6*, p.20). For southbound I-5 in Seattle, Hallenbeck et al. find that a lane-blocking incident causes between 318 (conservative estimate) and 591 (liberal estimate) vehicle-hours of delay (*2*, p.15).

The average daily delay caused by incidents, D_{col}, is 986 and 837 vehicle-hours, which is 20.3% and 18.8% of total PM delay for NB and SB, respectively. By way of comparison, Hallenbeck et al. find that "for the urban freeways examined [in the Central Puget Sound region of Washington State] lane-blocking incidents are responsible for between 2 and 20 percent of total daily delay" (*2*, p.8). These average numbers must be used with caution because the delay impact of incidents varies considerably from freeway to freeway and over different times of day. For example, in our study, during the AM peak (5-10 AM), the average incident-induced delay is 0 (because β_{col} is not significantly different from 0) for NB and 9.9% of the total peak hour delay for SB.

Aggregating over both peaks and both directions, the delay components are 13.3%, 4.5%, 1.6%, 33.2%, and 47.4% for incidents, special events, rain, potential reduction and excess demand.

5. CONCLUSION

Between 1980 and 1999, highway route-miles increased 1.5 percent while vehicle miles of travel increased 76 percent (*1*). In 2000, the 75 largest metropolitan areas experienced 3.6 billion hours of delay, resulting in $67.5 billion in lost productivity, according to the Texas Transportation Institute. Mitigating congestion through more efficient operations is a priority of transportation agencies. The first step in designing an effective mitigation strategy is to know how much each cause contributes to congestion. One can then design a set of action plans, each aimed at

reducing the contribution of a particular cause. The more detailed the set of causes that are considered, the more effective the strategy that can be devised.

The paper proposes a fully automated method that calculates six components of congestion: delay attributed to incidents, special events, lane closures, and weather; delay that can be eliminated by ramp metering; and the remaining delay, mostly due to excess demand.

The method is applied to a 45-mile section of I-880 in the San Francisco Bay Area for AM and PM peaks and for both directions. Incidents and special events together account for 17.8% of total delay. Lane closures caused no delay because delay-causing closures were not scheduled during peak hours. Rain caused 1.6% of total delay. A surprisingly large 33% of all delay could be eliminated by ideal ramp metering. Lastly, 47% of the delay is due to excess demand. Certainly, as discussed in the text, the 33% potential reduction due to metering needs to be interpreted with caution, as the maximum possible reduction. Even with such precaution, if these estimates are supported in more detailed studies, it is likely that most congestion mitigation strategies would harvest large potential gains from ramp metering.

ACKNOWLEDGEMENT

We are grateful for comments and criticism from John Wolf, Fred Dial, Jose D. Perez and Lisa Davies of Caltrans; Tarek Hatata of the System Metrics Group; and Alex Skabardonis and Karl Petty of the PeMS Development Group. This study is supported by grants from Caltrans to the California PATH Program.

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] FHWA. FHWA Congestion Mitigation website. http://www.fhwa.dot.gov/congestion/congest2.htm, Last Accessed July 1, 2005.

[2] Hallenbeck, M.E., J.M. Ishimaru, and J. Nee. Measurement of recurring versus non-recurring congestion. Washington State Transportation Center (TRAC), October 2003.

[3] Skabardonis, A., K. Petty, and P. Varaiya. Measuring recurrent and non-recurrent traffic congestion. In *Proceedings of 82nd Transportation Research Board Annual Meeting*, Washington, D.C., January 2003.

[4] Bremmer, D., K.C. Cotton, D. Cotey, C.E. Prestrud, and G. Westby. Measuring congestion: Learning from operational data. In *Proceedings of 83rd Transportation Research Board Annual Meeting*, Washington, D.C., January 2004.

[5] Dowling Associates, Berkeley Transportation Systems and System Metrics Group. Measuring Non-Recurrent Traffic Congestion: Final Report. Prepared for California Department of Transportation, June 2002.

[6] System Metrics Group. Freeway performance report. Prepared for California Department of Transportation, 2003.

[7] California Department of Transportation. 2002 HICOMP Report. State Highway Congestion Monitoring Program, November 2003.
[8] Petty, K., H. Noeimi, K. Sanwal, D. Rydzewski, A. Skabardonis, P. Varaiya, and H. Al-Deek. The freeway service patrol evaluation project: Database support programs, and accessibility. *Transportation Research, Part C*, 4 (2):71–85, April 1996.
[9] Skabardonis, A., K. Petty, P. Varaiya, and R. Bertini. Evaluation of the Freeway Service Patrol (FSP) in Los Angeles. Research Report UCB-ITS-PRR-98-31, California PATH, University of California, Berkeley, CA 94720, 1998.
[10] Bertini, R., S. Tantiyanugulchai, E. Anderson, R. Lindgren, and M. Leal. Evaluation of Region 2 Incident Response Program using archived data. Transportation Research Group, Portland State University, July 2001.
[11] Transportation Research Board. *Highway Capacity Manual 2000*, December 2000.
[12] Chin, S.M., O. Franzese, D.L Greene, H.L. Hwang, and R.C. Gibson. Temporary losses of highway capacity and impacts on performance. Technical Report ORNL/TM-2002/3, Oak Ridge National Laboratory, National Transportation Research Center, Knowville, TN 37932, May 2002.
[13] Chen, C., A. Skabardonis, and P. Varaiya. Systematic identification of freeway bottlenecks. In *Proceedings of 83rd Transportation Research Board Annual Meeting*, Washington, D.C., January 2004.
[14] Jia, Z., P. Varaiya, C. Chen, K. Petty, and A. Skabardonis. Congestion, excess demand and effective capacity in California freeways. Online at pems.eecs.berkeley.edu, December 2000.
[15] Kwon, J. and P. Varaiya. The congestion pie: delay from collisions; potential ramp metering gain, and excess demand. *Proceedings of 84th Transportation Research Board Annual Meeting*, Washington, D.C., January 2005.
[16] PeMS. PeMS website. http://pems.eecs.berkeley.edu.
[17] California Department of Water Resource Website, http://cdec.water.ca.gov/intro.html, Last Accessed November 11, 2004.
[18] California Department of Transportation, Lane Closure System (LCS) Website http://www.lcs.dot.ca.gov/, Last Accessed November 3, 2004.

LIST OF TABLES

LIST OF FIGURES

TABLE 1 Regression Result for Non-Recurrent Delay

Scenario	Factor	Estimate	Std. Error	t value	Pr(>|t|)[1]	Multiple R-squared
NB AM	(Intercept)	3,301.1	191.1	17.28	0.000 ***	0.12
	Event	-221.5	216.2	-1.03	0.308	
	Incident	115.8	74.2	1.56	0.122	
	Weather	1,305.7	384.4	3.40	0.001 ***	
NB PM	(Intercept)	3,419.7	408.1	8.38	0.000 ***	0.14
	Event	1,084.6	416.0	2.61	0.010 *	
	Incident	486.1	133.9	3.63	0.000 ***	
	Weather	75.4	732.7	0.10	0.918	
SB AM	(Intercept)	3,402.6	339.6	10.02	0.000 ***	0.17
	Event	-482.0	342.2	-1.41	0.162	
	Incident	221.1	127.6	1.73	0.086 .	
	Weather	2,125.6	598.5	3.55	0.001 ***	
SB PM	(Intercept)	3,311.1	374.8	8.83	0.000 ***	0.12
	Event	705.5	419.9	1.68	0.096 .	
	Incident	383.8	116.9	3.28	0.001 **	
	Weather	28.7	751.3	0.04	0.970	

1. Significance codes “***”, “**”, “*” and “.” mean the P-value is between 0 and .001, between .001 and .01, between .01 and .05, and between .05 and .1, respectively.

TABLE 2 Delay Contributions from Each Cause and Congestion Pie[1]

Scenario	Factor	β	Mean Weakday Occurrences	Delay Contributions (veh-hrs)	Factor, after Bottleneck Analysis	Delay Contributions (veh-hrs)	Percent of Total Delay
NB AM	Recurrent	3,301	NA	3,301	Pot	1,307	38.4%
		NA	NA	NA	Excess	1,994	58.6%
	Event	0	0.42	0	Event	0	0.0%
	Incident	0	1.55	0	Incident	0	0.0%
	Weather	1,306	0.08	102	Weather	102	3.0%
NB PM	Recurrent	3,420	NA	3,420	Pot	1,336	27.5%
		NA	NA	NA	Excess	2,084	42.9%
	Event	1,085	0.42	454	Event	454	9.3%
	Incident	486	2.03	986	Incident	986	20.3%
	Weather	0	0.08	0	Weather	0	0.0%
SB AM	Recurrent	3,403	NA	3,403	Pot	1,327	33.5%
		NA	NA	NA	Excess	2,076	52.4%
	Event	0	0.42	0	Event	0	0.0%
	Incident	221	1.78	394	Incident	394	9.9%
	Weather	2,126	0.08	166	Weather	166	4.2%
SB PM	Recurrent	3,311	NA	3,311	Pot	1,565	35.2%
		NA	NA	NA	Excess	1,746	39.3%
	Event	705	0.42	295	Event	295	6.6%
	Incident	384	2.18	837	Incident	837	18.8%
	Weather	0	0.08	0	Weather	0	0.0%

1. NA means the number is not needed.

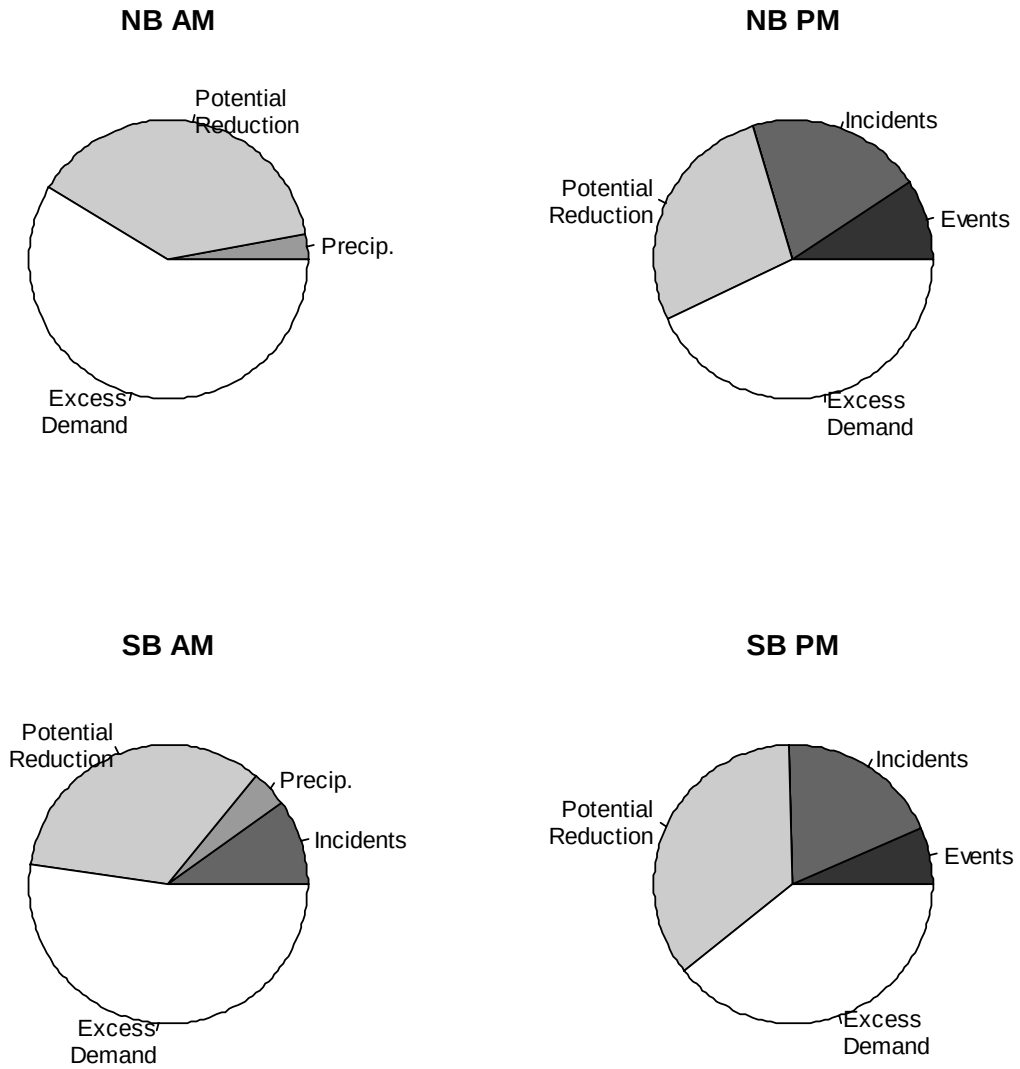


FIGURE 1 Congestion pie chart for four scenarios on I-880.

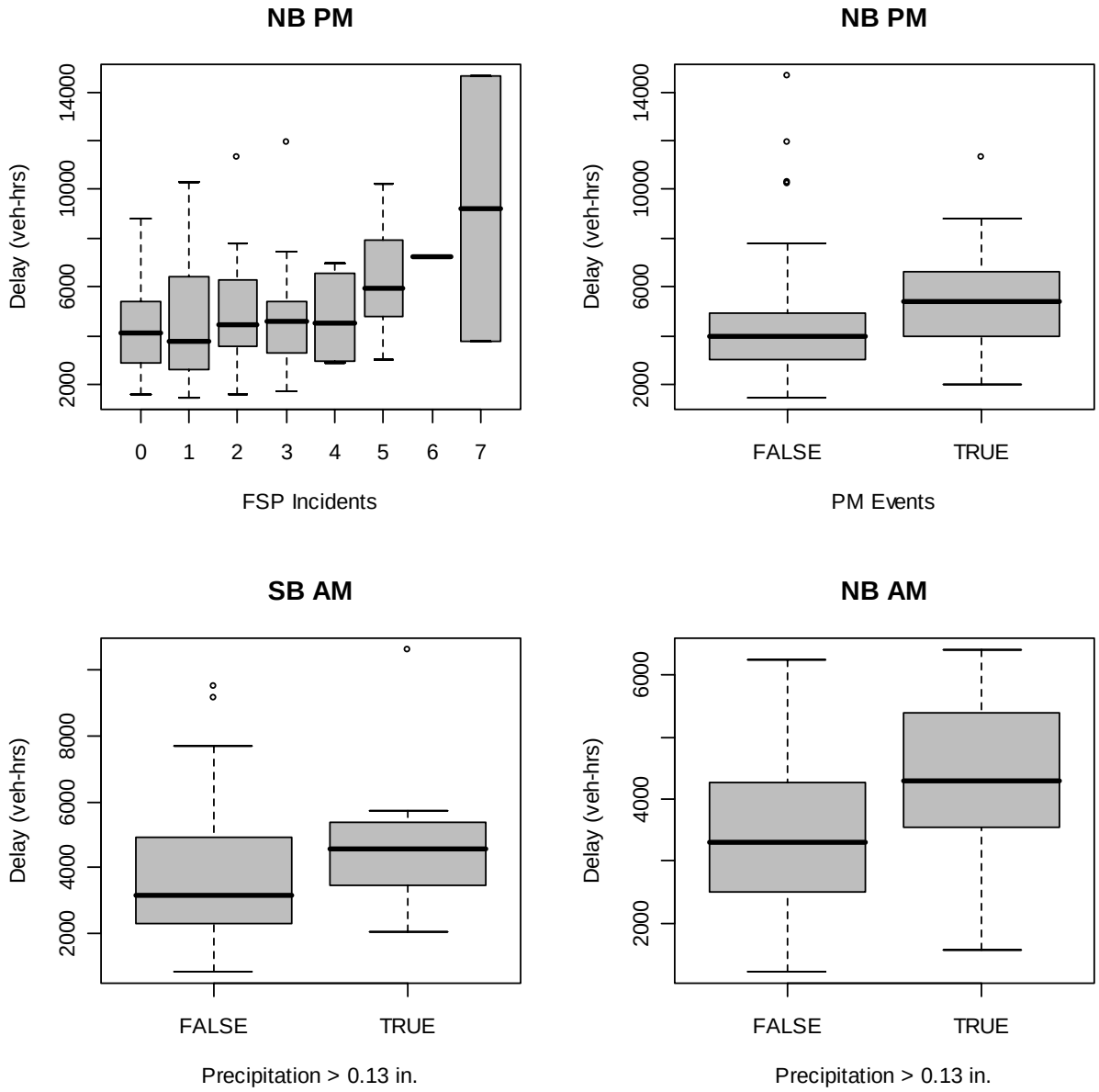


FIGURE 2 Relationship between delay and selected factors. The distribution of the average daily total delay $D_{total}(d)$, summarized as the box-and-whisker plot, is shown for each level of the number of incidents (upper left), special event occurrence (upper right), or adverse weather condition (bottom plots).

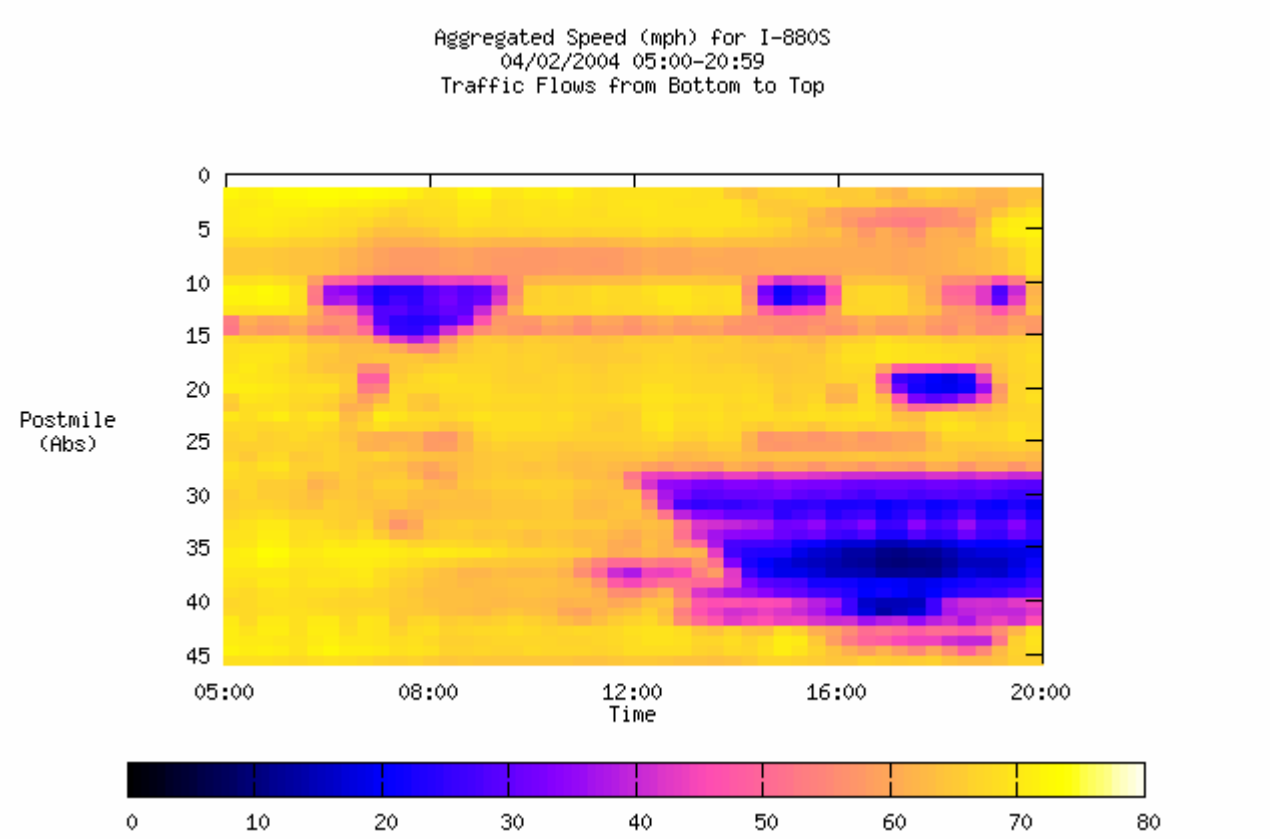


FIGURE 3 Lane-aggregated speed by postmile and time of day for I-880 S on April 2, 2004.