Title
Finding and Analyzing True Effect of Non-recurrent Congestion on Mobility and Safety

Permalink
https://escholarship.org/uc/item/3mh7d161

Author
Varaiya, Pravin

Publication Date
2007-08-01

Finding and Analyzing True Effect of Non-recurrent Congestion on Mobility and Safety

Pravin Varaiya
University of California, Berkeley

CALIFORNIA PARTNERS FOR ADVANCED TRANSIT AND HIGHWAYS

Finding and Analyzing True Effect of Non-recurrent Congestion on Mobility and Safety

Final Report for PATH TO 5321

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

April 29, 2007

Abstract

This report summarizes empirical research about the causes and impact of non-recurrent congestion. 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 method can be applied to any site with minimum calibration, but it requires data about traffic volume and speed; the time and location of incidents, special events and lane closures; and adverse weather.   The method is illustrated by applying it to a 45-mile section of I-880 in the San Francisco Bay Area.

Data limitations preclude applying the method statewide. A simpler method, which depends only on routine data collected in the PeMS system, has been implemented. A PeMS application now provides a 'congestion pie' for any district or freeway segment. The pie divides the total congestion delay into three categories: potential reduction, excess demand, and accidents; and an unexplained, 'miscellaneous', category.

Keywords: freeway congestion; non-recurrent congestion; congestion pie; weather; ramp metering; loop detectors

Executive Summary

Traditionally, congestion delay has been divided into recurrent and non-recurrent delay. The availability of more detailed and massive amounts of data, and the need to design congestion mitigation strategies whose benefits are more predictable, has spurred research into the causes and impact of congestion. This report summarizes empirical research about the causes and impact of non-recurrent congestion. The study has two major components.

The first component develops a method to divide the total congestion 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. 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.

The method is fully automated and 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. Because such detailed data are not routinely available, the method cannot be implemented today statewide.

The second component of the research develops a method that only uses data routinely available in PeMS. That method is now implemented as part of the PeMS 'congestion pie' application. PeMS now provides a congestion pie for any district or freeway segment, dividing the total congestion delay into three categories: potential reduction, excess demand, and accidents; and an unexplained, 'miscellaneous', category.

1. **Introduction**

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

The research summarized here proposes and tests 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 six components in our method correspond closely with the six slices of the FHWA pie of Figure 1, with three differences. First, the cause 'poor signal timing' is replaced by 'poor ramp metering' or, in our terminology, 'potential reduction in delay'. Also, what is regarded in the FHWA pie as the delay due to 'bottlenecks' is in our terminology 'excess demand'. Lastly, the 'work zone' slice in the FHWA pie is called 'lane closures' in our method.

Our 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 report 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 summarizes a simpler method that relies on routine data collected by PeMS. Section 6 concludes the paper. Some remarks about data collection are provided in the Appendix.

---

[1] Congestion (delay) is measured as the extra vehicle-hours spent driving below 60 mph on a freeway.

2. Previous Studies

Transportation agencies until recently only reported recurrent congestion. (For an example see (7)[2]; 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.[3]

Data 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 routinely 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).[4] Both studies separate 'non-recurrent' and 'recurrent' congestion but differ in definition and method. We pause to briefly discuss how this 'separation' poses conceptual and measurement questions.

From knowledge of the location and time of an incident one can estimate the total delay in a space-time region that includes the impact of the incident (15). However, one needs to separate this total delay into *recurrent* delay—that is the delay that would have occurred in the *absence* of the incident and the *additional* delay that was caused by the incident. The conceptual challenge is how to estimate the recurrent delay component, since it cannot be directly measured. The measurement question is how to estimate the space-time region of impact of an incident when the exact time and location of the incident are not known with the required accuracy.

We also note more subtle questions of causation that our study does not address. For example, how should one divide the total congestion caused by two nearby incidents? How should one account for the fact that (recurrent) congestion may itself increase the likelihood of accidents? How should one estimate the impact of an accident that causes congestion to spread into a different freeway (via a freeway-freeway connector) or into arterials as drivers respond to traffic information by diverting?

While we have discussed 'separation' in terms of estimating the impact of incidents, the same questions arise in estimating the impact of other causes: special events, adverse weather, lane

---

[2] The HICOMP report publishes *recurrent* congestion on urban area freeways for typical weekday commute periods. The report defines recurrent congestion as a "condition lasting for 15 minutes or longer where travel demand exceeds freeway design capacity and vehicular speeds are 35 miles per hour (mph) or less during peak commute periods on a typical incident-free weekday."

[3] The rich information revealed in (8) inspired the PeMS project, which automates the collection and processing of freeway loop detector data.

[4] This limits the use of the results predicted by our model. The prediction is reasonable in terms of the *average* contribution of incidents to delay, but not in terms of contribution of particular incidents.

closures, etc. For each cause the total delay is (relatively) easy to measure; the difficulty is to separate this delay into a recurrent component and a component that should be attributed to the cause being analyzed. Naturally, the difficulty is compounded when there are multiple causes. For example, adverse weather increases incidents as well as recurrent delay, and it is not clear how to separate the different impacts.

We return to review how previous studies have addressed separation. 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 $i = 1,\dots, n$, whose flow (volume) and speed measurements are averaged over 5-minute intervals indexed $t = 1,\dots, T$.[5] Days in the study period are denoted by $d = 1,2,\dots,N$. 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

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

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,

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

---

[5] The reader may recognize that this formulation of the available data conforms to the data available in PeMS.

The average daily total delay is simply

$$D_{total} = \frac{1}{N} \sum_{d=1}^{N} D_{total}(d) .$$ (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 $a = 1, 2, \ldots$. 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,

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

It will be useful to define

$$D_{non-rec} = D_{col} + D_{event} + D_{lane} + D_{weather},$$ (5)

$$D_{rec} = D_{tot} - D_{non-rec} = D_{pot} + D_{excess}.$$ (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
$D_{non-rec}$ 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,

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

in which

$\varepsilon(d)$ 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 or *additively* to the delay. Figure 2 illustrates that such model is reasonable for our study site.

More complicated causality relations between explanatory variables, such as between the bad weather and the number of accidents, are 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 example, interaction between adverse weather and collisions could be modeled by adding a multiplicative term such as $\beta X_{col}(d) X_{weather}(d)$. For the San Francisco Bay Area data considered below, however, the correlation coefficient between precipitation and number of accidents is only 0.032 so such a multiplicative term would contribute nothing to explaining the total delay.

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*).[6] 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 in free flow (60 mph) throughout the freeway section, and
(2) The upstream metered ramp has infinite capacity.

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

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

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

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

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}, hence the 'min' in formula (14).

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.[7]

---

[6] The algorithm in (*13*) has been implemented as part of PeMS' bottleneck application, which provides a list of bottlenecks, the delay and extent (or queue length) of the congestion each bottleneck causes, and its frequency of occurrence.

[7] The emphasis in this research is on developing a method to estimate the congestion pie statewide, using readily available data. For this reason, we resort to a rough, but easily automated, method. A more refined estimate would

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 4, which may be compared with the FHWA pie of 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 to 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.

---

require a simulation of how much delay reduction is possible using realistic ramp metering strategies. Such an approach is being pursued under California PATH Task Order 6611, Tools for Operations Planning (TOPL).

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 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 a larger portion of the congestion in the morning peak hours is 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. Note that D_{pot} and D_{excess} were extrapolated from district wide quantities; freeway-specific computation is now available in PeMS.

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. Statewide implementation

A modified (simpler) version of the method described above has been implemented in PeMS. The differences stem from data sources and method.

PeMS uses 1) the Caltrans TASAS incident database, 2) its bottleneck identification algorithm, and 3) the Ideal Ramp Metering algorithm described above. PeMS divides

D_{total}, the total daily delay on the freeway, into

D_{col}, the delay that is assigned to collisions;

D_{pot}, the potential delay that can be saved by running an ideal ramp metering algorithm at the major bottlenecks on the freeway;

D_{excess}, the delay caused by excess demand that no ramp metering algorithm could reduce; and

D_{misc}, the delay that cannot be assigned to collisions or bottlenecks.

There is a slight difference in method having to do with the division of the recurrent delay D_{rec} into D_{pot} and D_{excess}. In the method described here, by definition, $D_{excess} = D_{rec} - D_{pot}$ comprises *all* of the recurrent delay that cannot be reduced by the Ideal Ramp Metering. By contrast, in PeMS, D_{excess} (PeMS) $= D_{bn} - D_{pot}$, in which D_{bn} is the total delay that occurs only at bottlenecks; whereas $D_{misc} = D_{total} - D_{col} - D_{bn}$. Thus D_{excess} (PeMS) is smaller than the estimate of D_{excess} provided here.

Figure 5 displays the congestion pie from PeMS for 2004 Q1, which roughly coincides with the study period.

6. 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 research summarized here 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.

A slight modification of the method proposed here has already been implemented in PeMS' 'congestion pie' application and can be applied to any freeway or district.

Acknowledgement

Department of Transportation. This report 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.

[19] FHWA. Trends and Advanced Strategies for Congestion Mitigation, 2005. http://www.ops.fhwa.dot.gov/congestion_report/, Last Accessed May 3, 2007.

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%

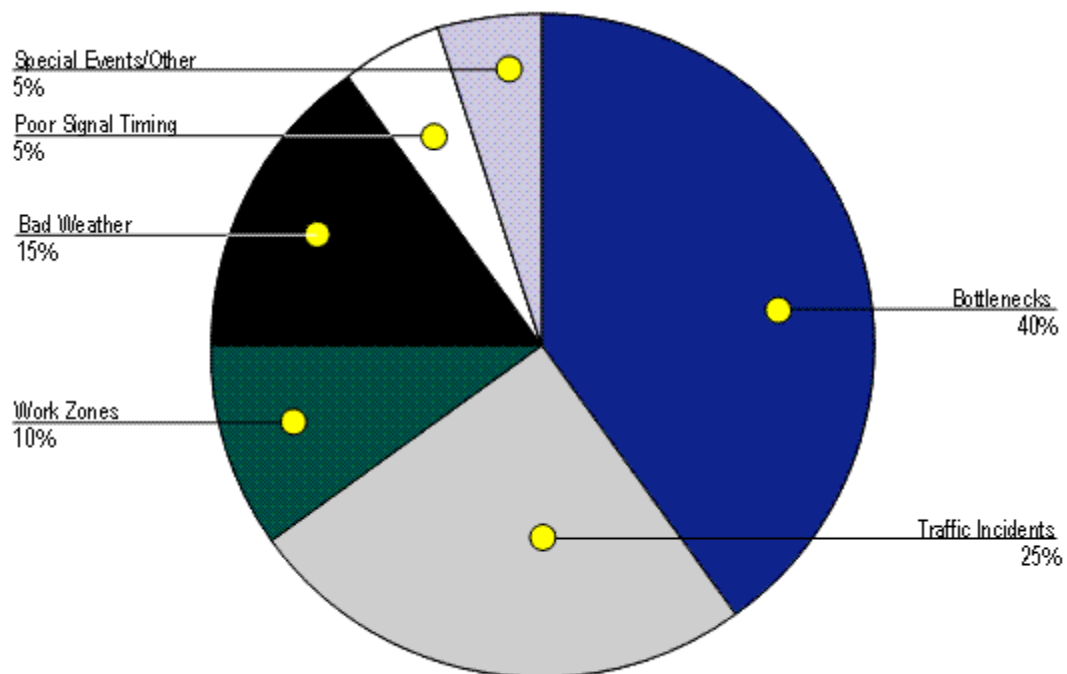1. NA means the number is not needed.

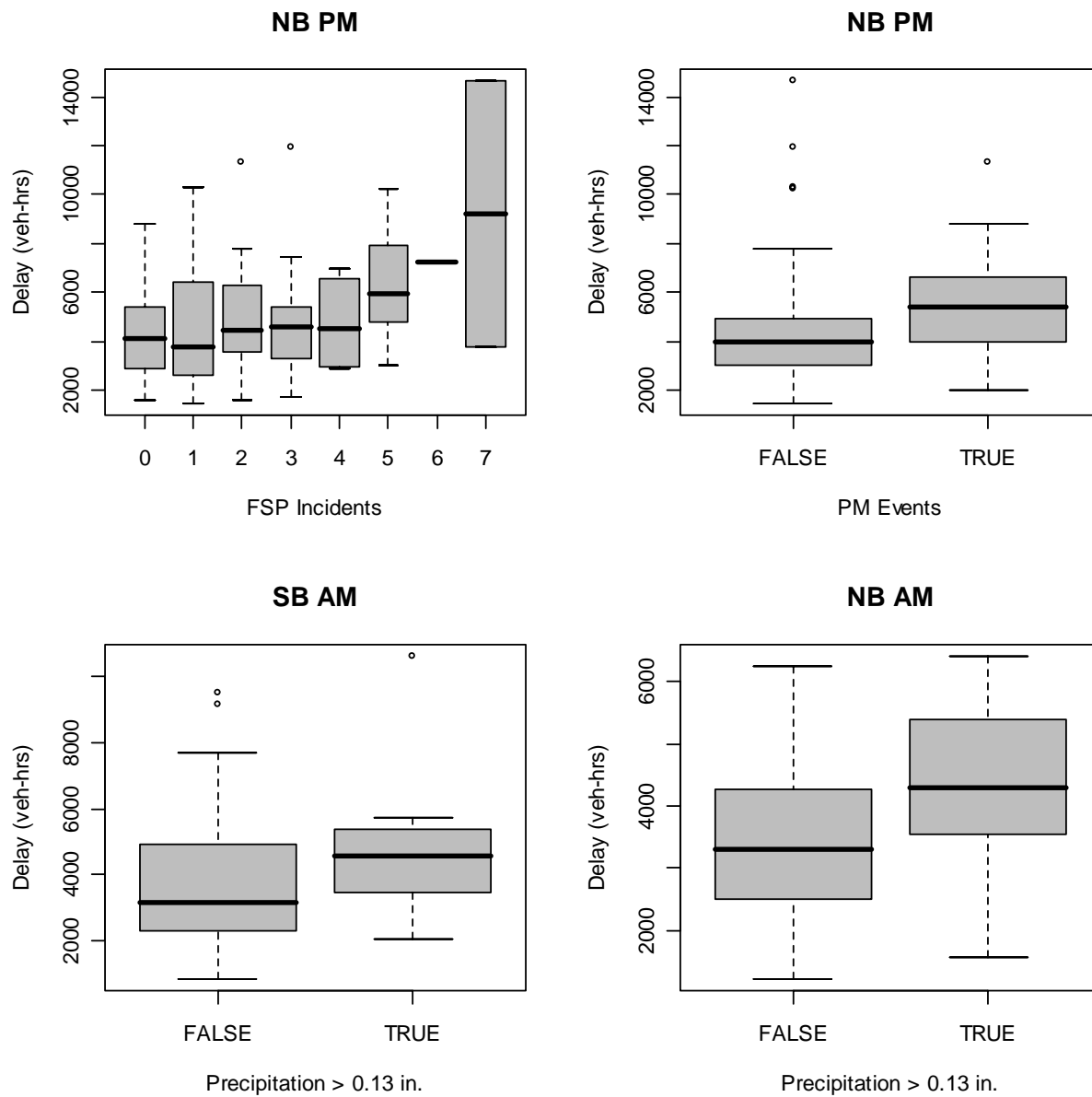**Figure 1 The FHWA congestion pie. Source** *(19)***.**

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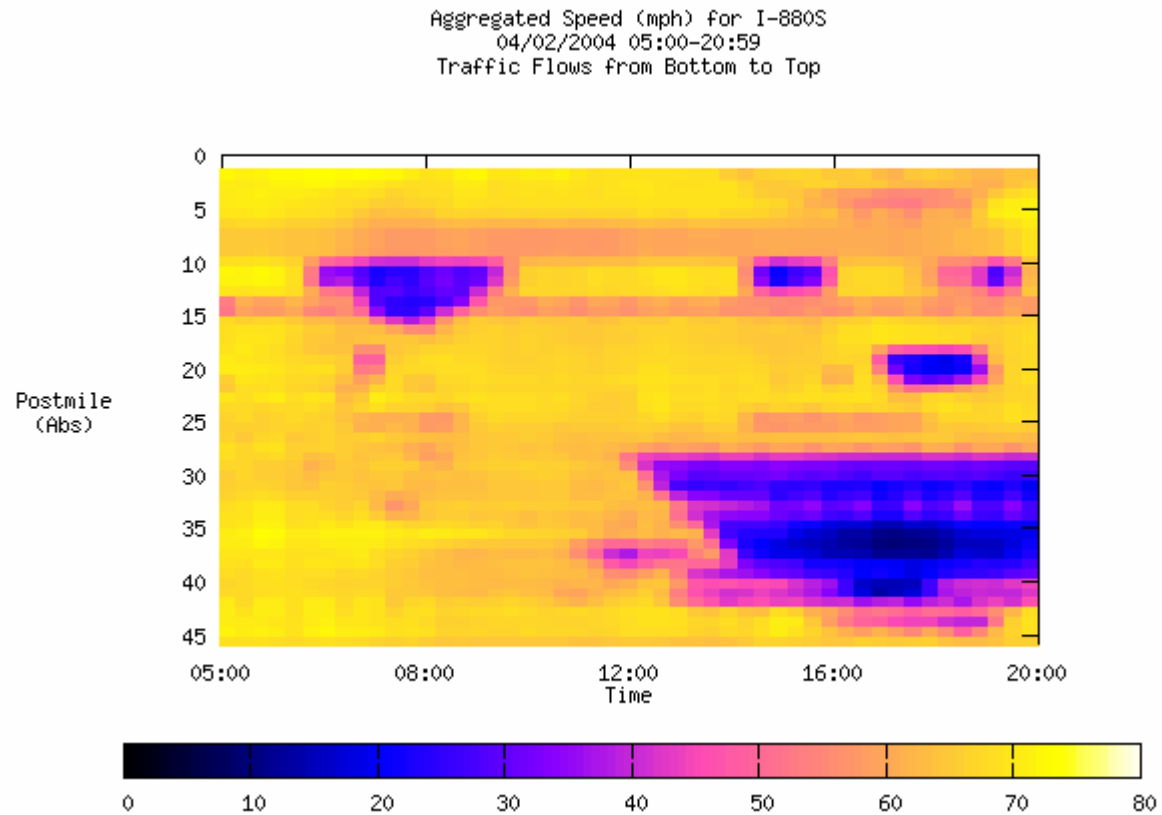
Figure 3 Lane-aggregated speed by postmile and time of day for I-880 S on April 2, 2004.
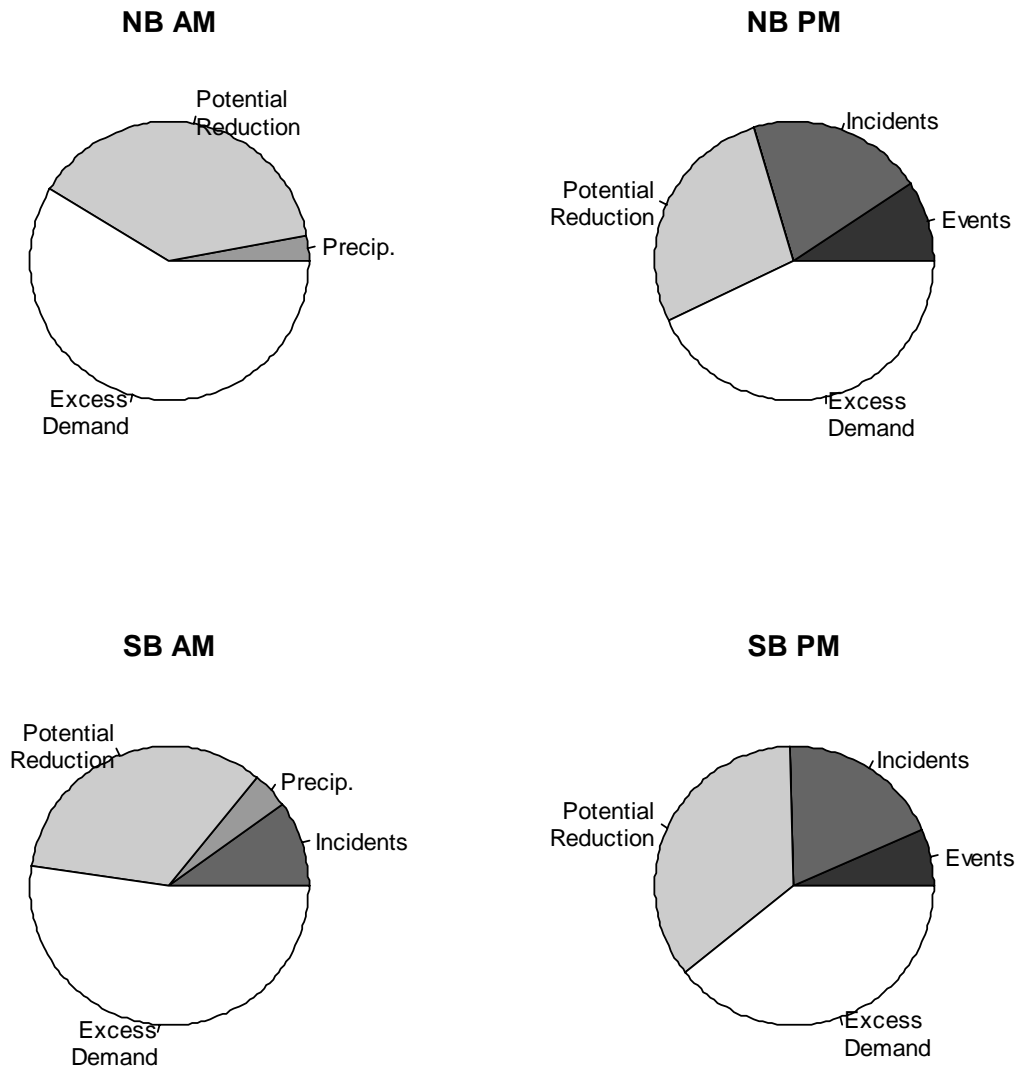
NB AM

NB PM

SB AM

SB PM

Figure 4 Congestion pie chart for four scenarios on I-880.

Congestion Breakdown
☐ Potential Reduction (40.7%)
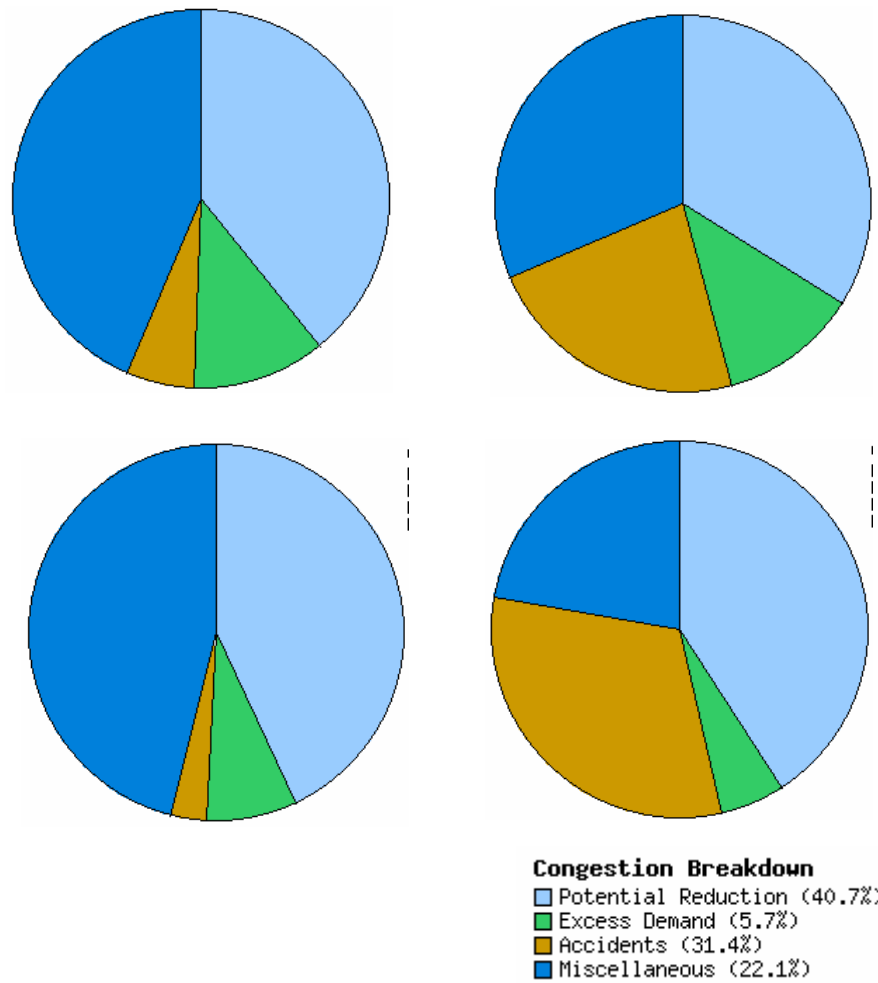☐ Excess Demand (5.7%)
☐ Accidents (31.4%)
☐ Miscellaneous (22.1%)

Figure 5 Congestion breakdown for I-880 for 2004Q1. NB: AM (top left), PM (top right); SB: AM (bottom left), PM (bottom right). Source: PeMS *(16)***.**

Appendix: Data Issues

This appendix discusses issues relating to data.

Routine data

Routine data for volume (flow), speed, freeway configuration (length) etc. are obtained from PeMS, which provides 5-min aggregated values and freeway configuration.

Incidents

There are many choices, each of which has problems. Ideally, one would like to know the location of an incident (post mile, lane), its nature (collision, breakdown), and the response (clearance time). Unfortunately, these are not available.

PeMS uses TASAS data, which do provide location and nature of incidents, but no clearance time. TASAS data do not contain all incidents, e.g. breakdowns are not recorded. The major difficult is that there is a significant time lag (up to six months) in the availability of TASAS data.

FSP (Freeway Service Patrol) data provide a rough location and nature of the incidents. However, FSP data are only available for the time periods when FSP is in operation.

CHP (California Highway Patrol) data records are available in PeMS, which gets them from the CHP website. PeMS parses the CHP logs and for some of the incidents it is able to determine the nature of the incident.

BAIRS (Bay Area Incident Response System) identifies and directs local personnel and. equipment to traffic obstructions. It has proved successful in reducing the response time by quickly locating and dispatching crews and equipment.

Lane closures

This study relied on the Lane Closure System database. The major difficulty in using the database is that it records lane closure requests but not whether the lane closure was actually implemented. This makes the database not useful.

Weather

The Department of Water Resources (DWR) (http://cdec.water.ca.gov/intro.html) provides historical and hourly (in some cases) measurements of (1) precipitation, (2) temperature, and (3) wind. Unfortunately, there are no measurements of fog, which is sometimes significant in the Bay Area.

There is one station near North Oakland, measurements from which are used in this study. However, the micro-climate variations in the Bay Area means that we use county-wide measurements for I-880 conditions, which is not accurate.