Predicting the Duration of Congestion on Bay Area Freeways

Steven B. Colman, AICP (F)

Abstract. This paper reports on a preliminary investigation of variables influencing the duration (weekday hours) of recurring congestion on freeways. It is based on floating car data collected in 2002 on San Francisco Bay Area freeways. Congestion was defined (Caltrans, 2003a) as operation at speeds averaging less than 35 mph during weekday periods. The author assembled and explored variables that might help predict the duration of congestion, including the ADT, number of basic lanes, percent trucks, terrain, the percent of trips that are for commute purposes, interchange spacing, and the presence of HOV lanes. The result was identification of significant variables in predicting the duration of congestion on freeways, and equations that can be used in planning studies provided a few variables, such as future ADT, are known.

INTRODUCTION

The San Francisco Bay Area encompasses the nine counties that touch San Francisco Bay, including approximately seven million residents and 3.5 million jobs (ABAG, 2005). There is an extensive freeway system including 621 route miles (MTC, 2005). In 2002, 369 miles were defined as “congested” by Caltrans (Caltrans, 2003a). The “Bay Area,” as it is widely known, also has the dubious distinction of having (depending on how and when measured) the second to fourth worst traffic congestion of any metropolitan area in the United States (TTI, 2005). According to TTI’s estimates, the average Bay Area resident spends 72 hours a year—the equivalent of almost two work weeks—stuck in traffic congestion.

This exploration of the prediction of congestion duration was motivated by the author’s observation that this is an important indicator of system performance. Other studies have suggested this (TTI, 1997; Dey, 1998), yet the topic has received relatively little attention in the profession, perhaps because of the scarcity of directly observed congestion duration data. Many highway studies continue to focus on level of service (LOS) as the key performance measure. Traditional highway planning focused on achieving a LOS standard (typically C or D) during peak hours, yet in many cases it is not possible to provide adequate capacity to do so, especially in built-up urban areas. Furthermore, opponents of highway improvements have used the LOS standard to oppose projects. For example, in one Bay Area corridor study, all five improvement options (including No Build) resulted in future peak LOS of F during the peak hour, which opponents seized on as ‘proving’ that any increased capacity in the corridor was futile, and therefore no improvements should be made to the freeway. Clearly, this is untrue, since additional capacity would have the benefit of shortening the duration of congestion, even if the peak hour LOS did not improve.

The duration of congestion—the number of hours that a facility is operating below a benchmark density or speed—has been widely noted as a desirable measure of system performance, but no simple, effective technique for its prediction in a planning study was found in reviewing the traffic congestion literature. Although descriptive information on the duration of congestion is often available, usually from floating car data or embedded sensors—there appear to have been relatively few attempts to develop a predictive model of the duration of congestion.

Congestion duration is an important indicator of overall system performance because it:

- Can be easily understood by the public and decision-makers
- Measures the degree to which motorists (and their passengers) must engage in time-shifting behaviors to avoid peak congestion
- May be a better metric than LOS to effectively measure the impact of improvements to congested urban freeways

Source of Data on Congestion and Definition

Annual monitoring data from Caltrans Highway Congestion Monitoring ("HICOMP") report for the year 2002 were used as the dependent variable in this study. Although more recent data have been collected, after 2002 data collection responsibilities were transferred to the Metropolitan Transpiration Commission (MTC), and changes were made in the program.

Data were collected using the floating car (actually, floating truck) technique, with data collected in the spring and fall to represent average weekday conditions, in so far as possible. Caltrans defines congestion as, "a condition where the average speed drops below 35 mph for 15 minutes or more on a typical weekday" (Caltrans, 2003a). Duration is the length of time a freeway segment remains congested. Caltrans monitors all congested freeway segments. The morning and afternoon total hours of congestion were summed together to get the total weekday congestion.

Scope of Study

In order to develop a simple planning tool, some observations were excluded from the analysis, because of the desire to develop a technique applicable to freeway mainlines:

- Toll Plazas: The Bay Area has eight major bridges, seven of which have significant delay related to toll plazas. Because of delays related to toll collection, and the lane drops typically associated with bridges, these were excluded from the data analyzed.

- Tunnels and One-Lane Freeways: Tunnels were ignored because of their lower capacity, and in one case because the number of lanes in the Caldecott Tunnels is varied through the day. Also, there is a freeway (Route 37) that has only one lane in each direction; it was deemed a special case, and therefore ignored.

- Congestion influenced by freeway-to-freeway connections: Because these often involve lane drops, and because traffic can back upstream into areas that would otherwise not be congested, these congestion locations were excluded.

- San Francisco County: Because of the small land area of San Francisco, freeway design standards dating from the 1950s, and the frequency of interchanges, observations from San Francisco were excluded.

- Variable Lane Configurations: The sections included in this study have a constant number of basic lanes, or at most, one change in the number of lanes. Where the basic number of lanes changed several times over the congested segment, the location was excluded. This was because one of the predictive variables is daily traffic per lane, and it is difficult to make a meaningful calculation of this variable when the number of basic lanes changes.

This left 46 congestion locations that were analyzed as part of the study. The HICOMP data collection specifically is intended to reflect average weekday (recurring) congestion, and not special events or incidents.

Organization

After this introductory section, the model development section describes the choice of variables for analysis, and provides an estimation of the model using various statistical techniques. Some applications, limitations, and sources of error are then covered.

DEVELOPMENT OF THE MODEL

Underlying Theory and Variables Included

At its most basic level, traffic congestion occurs when demand exceeds capacity and thus a queue or backlog of traffic is created. The queue travels at speeds that are much less than free-flow speeds. The queue dissipates when demand has dropped sufficiently that the ‘reservoir’ of queued vehicles can be accommodated through the congestion point. In the case of significant, recurring congestion points, this can take several hours. Figure 1 below illustrates a simple case of congestion at a bottleneck, with the solid line representing the rapidly increasing demand in the morning peak, and congestion beginning at 6:30 AM. Even though the demand drops to less than capacity within an hour (7:30), congestion endures until the queued vehicles dissipate (approximately 8:45). The downstream flow is relatively constant is relatively constant during the congestion period, and may actually be less than capacity under forced flow conditions.

This simple illustration is a useful departure point for discussing congestion duration, although it ignores the fact that demand is not fixed in a time period, because some individuals will have the ability and willingness to shift the starting time of their trip (temporal shifting), as well as change the route chosen (spatial shifting). This paper does not deal with those issues, as they are better addressed through travel forecasting models.

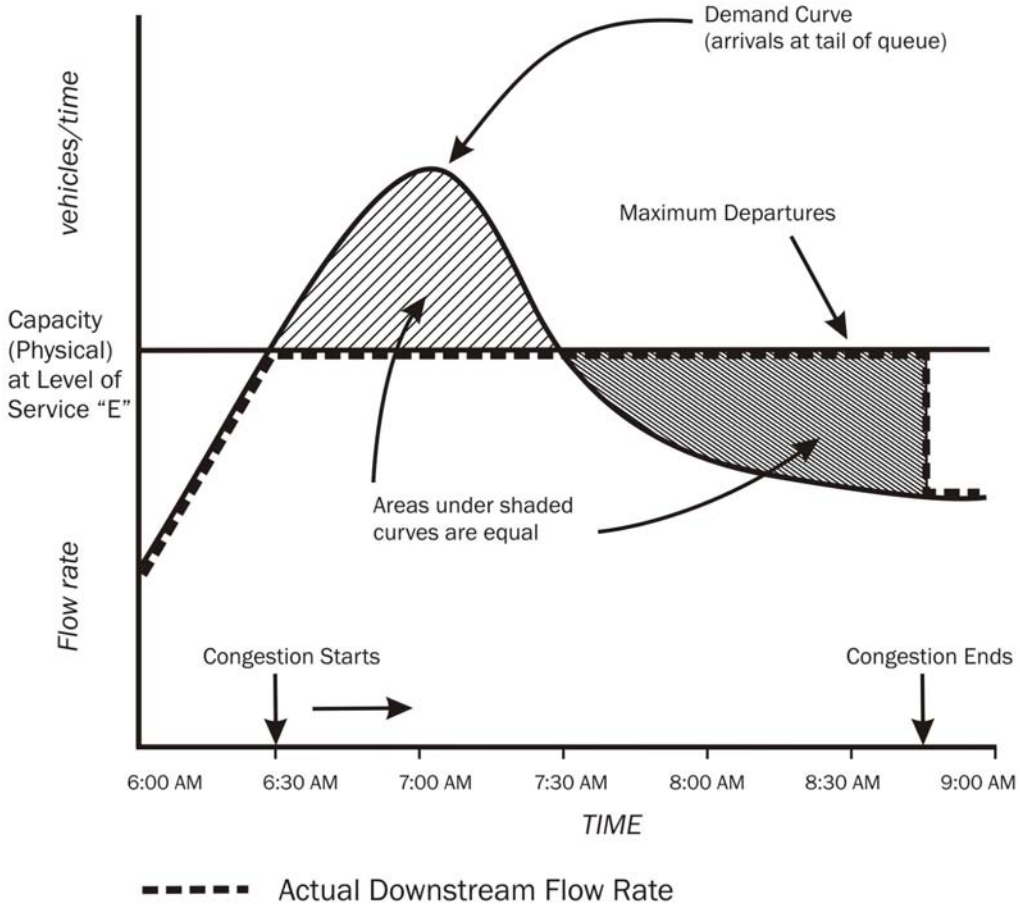

Figure 1—A Simple View of Congestion Duration at a Bottleneck

The figure also makes it clear that an important factor underlying the duration of congestion is the shape of the demand curve, which is usually not a known quantity (or at least is not easily obtainable for planning purposes). A flatter demand curve will result in less congestion, whereas a sharper, more "peaked" demand curve will result in congestion lasting longer.

In this paper, variables included in the model were chosen because of *a priori* indications that they could be significant in predicting the duration of congestion, and because they would be readily available for both the base year and for future horizon years. These variables are shown in Table 1 below, along with there associated minimum, maximum, and mean values:

Variable Name	Variable (units)	Minimum	Maximum	Mean
CONDUR	Congestion Duration (total hrs/weekday)	0.67	10.08	4.26
ADTLN	AADT per lane (in thousands of vehicles/weekday)	12.750	34.000	24.300

ICSPAC	Interchange spacing in miles (segment length/(interchanges-1))	0.40	2.36	1.03
PCTTRK	Percent trucks (0-100 percent) during eight hours of the day	0.55	12.2	5.1
GRADE	Grade (0=flat, 1=hilly or mountainous)	0	1	0.28
PCTHBW	Percent of weekday trips that are home based work purpose on the congested link (0-100 percent)	26	58	43
HOVDUM	High occupancy vehicle lane dummy (0=no HOV lane; 1=HOV lane)	0	1	0.37

Table 1– Variables Used in the Study for 2002 Conditions

Several of the variables had to be entered as dummies: for example, because some congested areas include a mix of differing grades, a dummy was used rather than an exact value. Of the 47 congested segments studied, 28% involved hilly or mountainous terrain, and 37% included an HOV lane. It should be noted that in some cases the HOV lane(s) spanned some, but not all, of the congested highway segment, and that in the lane count used in the ADTLN variable, HOV lanes were included in the lane count. The ADT on congested segments varied from 51,000 to 282,000, and the *total* number of lanes (in both directions) from 4 to 12 (with the average number of lanes in study segments at 6.7). These variables were readily available from highway inventories, and in the case of PCTHBW, from the regional travel forecasting model.

Congestion duration was the sum of both AM and PM peak period congestion. Bay Area freeways tend to experience slightly more PM than AM congestion (averaging 2.06 morning hours vs. 2.26 PM hours). Also, it should be noted that Caltrans' monitoring did not stop until average speeds rose above 35 mph; thus, one segment (I-880 in Santa Clara County) was observed to have more than 10 hours a day of congested operation in 2002.

Approach to Analysis

First, a simple linear regression model using all six variables (and a constant) was used to look at the relative significance of different variables. Then, two regression approaches were applied to the data: a stepwise linear regression and a log-log model. Estimations were done using Statistical Package for the Social Sciences, version 15 (SPSS, 2006). The model specification used the unadjusted congestion duration as a function of the other six variables. In the simple linear regression, four of the six variables were statistically significant at the 95-percentile level. The two that were not were interchange spacing and the percent of trips that are home based work. The HBW percentage had been used in the regression on the assumption that the greater the commuter usage of the freeway, the "sharper" the demand peaking would be, which would in turn result in congestion lasting longer. Although the coefficient on this variable had the expected sign (positive), the coefficient value was small, and as noted, not statistically significant.

A stepwise regression was also employed, in which predictive variables are entered and removed one at a time, until the F statistics do not indicate that any variables in the equation should be removed nor added. The ‘best’ model appeared to be one using just three variables—ADTLN, PCTTRK, and the dummy variable GRADE. The linear equation is:

DURATION OF CONGESTION (in hours/weekday)=
.343 (PCTTRK) + .193(ADTLN)+ 1.52(GRADE) - 2.63

All independent variable signs are as expected (i.e., positive). The interpretation of the results, provided the model is correctly specified, are that each additional percentage of truck traffic on the freeway adds 21 minutes of congestion; each additional 1,000 daily vehicles/lane adds just under 13 minutes of congestion, and hilly/mountainous terrain will increase congestion duration by 1.5 hours. Some versions of the model suggested HOV lanes are associated with increased congestion duration. Clearly, HOV lanes do not cause congestion. However, HOV lanes typically do not carry as many vehicles per hour as mixed flow lanes, compounded by the fact that Caltrans has a deliberate policy of adding HOV lanes in the most congested areas, not uncongested areas (the ‘correlation does not imply causation’ argument). Additional details are provided in Table 2 below:

Variable Name	Variable (units)	Coefficient	Std. Error	*t*-statistic	Signif-icance
Constant	Constant term (hours)	-2.628	1.882	-1.396	.170
ADTLN	AADT per lane (in thousands of vehicles/weekday)	0.193	0.063	3.059	.004
PCTTRK	Percent trucks (0-100 percent)	0.343	0.123	2.778	.008
GRADE	Grade (0=flat, 1=hilly or mountainous)	1.524	.647	2.357	.023

Table 2– Final Stepwise Linear Regression Results on 2002 Congestion Data

The negative constant term may seem odd, but actually makes sense, particularly for a linear regression. The data indicate that most Bay Area freeways do not experience congestion until their ADT/lane reaches a threshold value per lane. Were the constant positive, it would imply that even a freeway with no traffic (ADTLN= 0) would experience some congestion. The constant should be thought of as the intercept of the Y-axis; the implication is that congestion only occurs when ADT per lane > 13,600 (rounded), which makes sense.

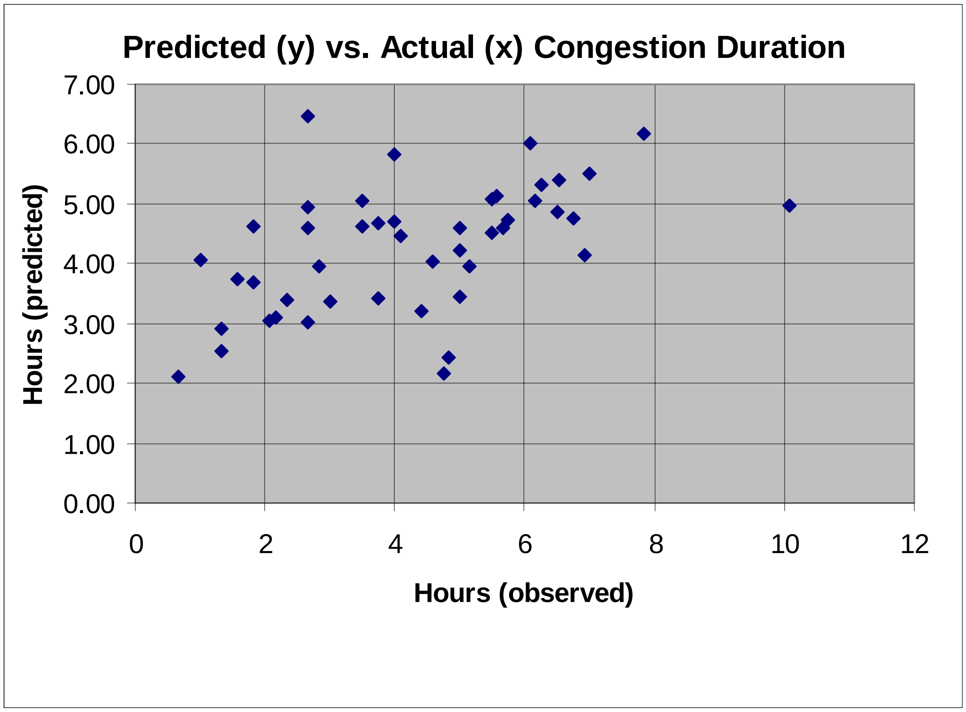

Figure 2—Plot of Residuals from the Stepwise Linear Model

Figure 2 shows an examination of the residuals (the difference between the predicted and actual congestion duration values) for the stepwise model. A 45 degree line from the origin indicates a perfect model fit; in reality, the model is accurate in most cases only to about +/- 1 to 1.5 hours of congestion. An example application might be considering the impact of widening an existing 4-lane freeway to 6 lanes. The freeway is flat and has five percent trucks. The travel forecasting model indicates that the no project alternative (4 lanes) has a forecast traffic volume of 80,000 ADT (20,000 per lane); and with the widening (6 lanes) the volume would be 90,000 ADT (15,000 per lane), due to additional capacity and spatial diversion from parallel routes. The predicted congestion duration is thus (.193)(20)+(.343)(5)-2.628 = 2.95 hours of congestion without the project. With the project, the congestion duration is estimated at (.193)(15)+(.343)(5)-2.628= 1.98 hours, or approximately one hour less per day.

Model with Logarithmic Transformation of Congestion Duration

There are several indications that the relationship between congestion duration and the predictive variables might not be linear. For example, most speed/flow curves are non-linear. Also, Figure 2 shows that the distribution of residuals was correlated somewhat with the actual congestion observed. This often indicates a non-linear relationship, and it's sometimes helpful to take logs of the dependent variable (congestion duration) to improve the model's fit. In developing several exploratory models, it appeared that the best t-statistics were achieved when a model using both the log of the total number of lanes and the total ADT was used, rather than ADT per lane. This seemed to improve the mode's performance.

Note in this model both the HOV dummy and GRADE have an extremely powerful influence on the duration of congestion. In this model, the dependent variable is natural log (ln) of the congestion duration; statistics from this version of the model are:

Variable Name	Variable (units)	Coefficient	Std. Error	*t*-statistic	Signif-icance
Constant	Constant term (hours)	-2.686	1.274	-2.109	.041
LnADT	Log of the *total* AADT (in thousands of vehicles/weekday)	1.319	0.351	3.760	<.001
LnLanes	Log of the total number of lanes, in both directions	-1.953	0.421	4.640	<.001
PCTTRK	Percent trucks (0-100 percent)	0.1233	0.030	4.047	<.001
GRADE	Grade (0=flat, 1=hilly or mountainous)	0.501	0.159	3.154	.003
HOVdum	HOV lane dummy (0=no HOV lanes, 1 if present)	0.590	0.148	3.995	<.001

Table 3– Log-Log Regression Results on 2002 Congestion Data

Use this log-log (or perhaps "partial-log log is more accurate) yields a somewhat lower estimate of the congestion using the example values in the linear model (80 ADT 4 lane freeway, 5% trucks, flat, no HOV). It predicts such a freeway would experience 2.72 hours of daily congestion, compared with 2.95 hours using the stepwise linear model discussed earlier. However, when comparing it with a 6 lane, 90,000 ADT freeway, the difference is fairly similar (the log-log model predicts a 1.28 hour reduction in congestion duration, to 1.44 hours daily).

In this model, the presence of HOV lanes turned out to be significant, and in general, the t-statistics much better. However, the size of the coefficient on the HOV dummy seems larger than reasonable, and this may reflect the previously noted issue that most HOV lanes in the Bay Area are provided on the most congested freeways. More than half the observed congestion points were predicted within +/- 64 minutes of their actual (observed) daily congestion duration.

This log-log model can be simplified (re-written) to a multiplicative form, which is somewhat easier computationally, where:

CONGESTION DURATION (in hours)=

$0.068 * \text{ADT}^{1.319} * \text{Lanes}^{-1.953} * \exp[(.1233)(\text{PctTrk})] * \text{HOVdum} * \text{GRADE}$

where:

ADT= Total AADT (in thousands)—note this is *not* per lane
Lanes= Total number of basic lanes in both directions
PctTrk= Percent trucks (0 to 100)
HOVdum= 1 if no HOV lanes; 1.79 if HOV present
GRADE= 1 if flat; 1.65 if hilly or mountainous
Exp= Exponential operator

Writing the equation this way makes it readily apparent that this formulation will always predict congestion >0 (even if a small value), because of the logarithmic nature of the model. Use of this model should probably be restricted to cases where there is at least 15,000 ADT per lane.

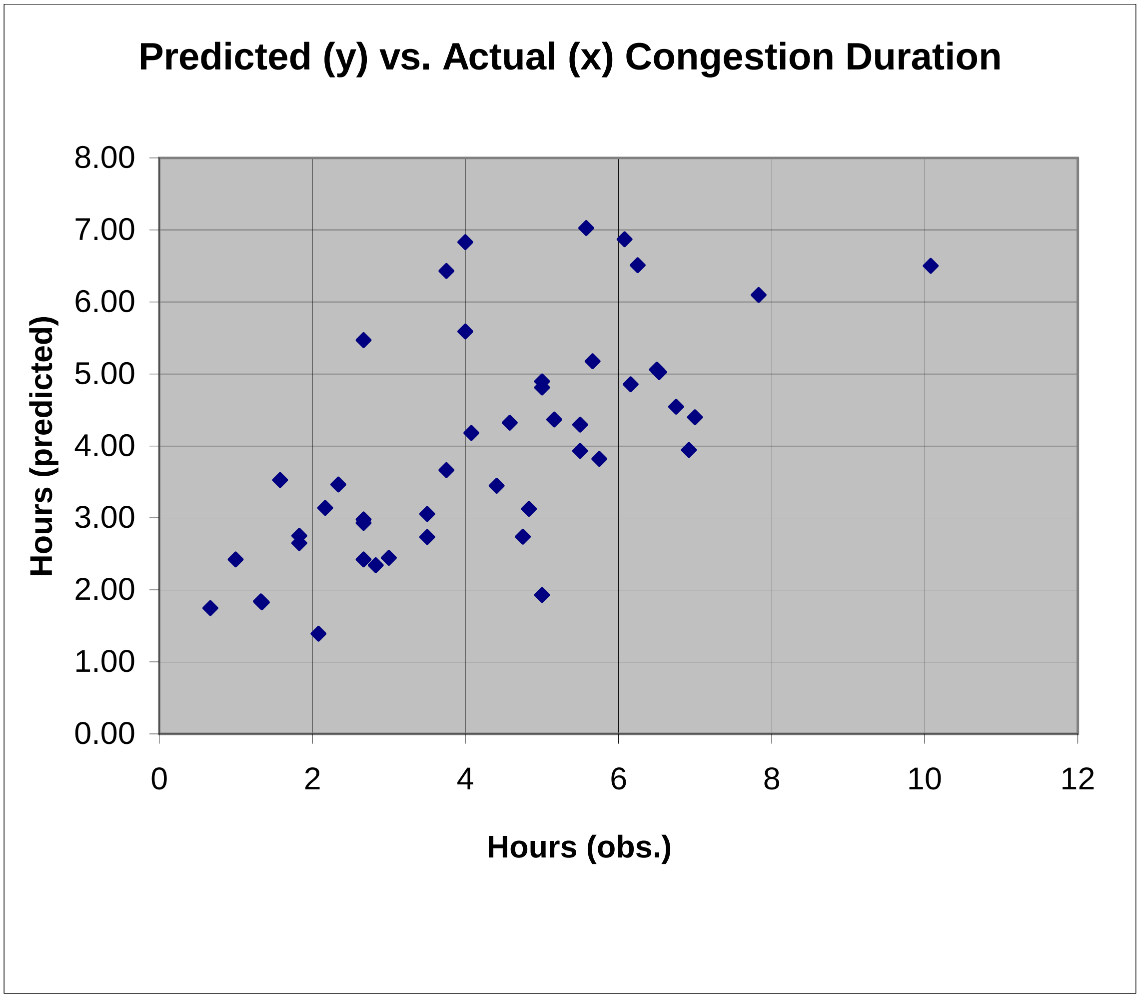

Figure 3—Plot of Residuals from the Log-Log Model

CONCLUSIONS

This paper presented an exploration in search of a simple, aggregate method for estimating existing and future congestion duration on freeways, using a relatively small number of variables typically available from planning studies. Although the results were not entirely satisfactory, work continues on improving the estimation models. The approach shows some promise in providing order-of-magnitude estimates of congestion duration that can be easily applied with a limited amount of data, and no fieldwork. This is particularly true in the area of predicting the differences between two or more project scenarios (the differences may be more important for decisionmaking than the actual values).

Some of the observations the author considers noteworthy are:

- Interchange spacing does not appear to be an important factor in determining congestion duration. This was unexpected and is somewhat contrary than the result found in other congestion studies.

- The percentage of home-based work trips in the congestion area did not appear to be important. The initial expectation was that this might represent the sharpness of the peaks mentioned earlier. It may be that there is not enough variation in the percent of HBW trips to capture the importance of this variable, or the regional travel forecasts may not be accurate enough in this regard.

- Both the daily percentage of trucks and the terrain appear to be important factors in predicting congestion duration.

- The role of HOV lanes in predicting traffic congestion is difficult to disentangle, particularly in the Bay Area. All Bay Area HOV lanes are restricted only part of the day, and are open to all traffic the remainder of the day. There is a wide variation in the actual operating details; restricted hours may range from a few to eight hours a day, and in some locations lanes operate only in the peak direction of traffic, but are open to all traffic in the reverse peak direction. This makes it extremely difficult to capture the impact of HOV lanes with a simple dummy variable.

Limitations and Sources of Errors

As with any study, there are several possible sources of error. They include:

- Measurement of traffic volumes—measurement of traffic volumes is subject to at least two types of errors: first, the volumes themselves are based on estimates typically made based on counts for a few weeks of the year, not the entire year. “Permanent count” stations in the Bay Area are very limited, and so “profiles” are used to develop estimates of volumes between permanent stations. Second, published data is for *annual average* daily traffic (counted over seven days), which in urban settings, is likely to be somewhat less than average weekday traffic. This is a refinement to the approach that will be made in the future.

- Measurement of congestion duration—although the estimates have been directly estimated using floating trucks, there may have been as few as two days of data used to represent an annual average. Because of normal fluctuations in traffic volumes, special events, etc. this could lead to erroneous predictions. The author notes that one outlier—where the models predicted substantially lower congestion duration than the observation—that data collected in years before and after indicated substantially shorter congestion duration (with no changes to the physical characteristics of the freeway). This segment, #26, had 10.08 hours of congestion in the 2002 data set, but in surveys taken before and after (2000 and 2005) using the identical data collection methodology, the congestion was significantly shorter. This problem may be remedied by the use of embedded loop detectors, which can provide a much larger number of days over which to average the congestion duration.
- Presence of auxiliary lanes—the presence of auxiliary lanes is ignored; only basic lanes are used in determining the number of lanes, and thus ADT per lane.
- Directionality—the method proposed in this paper uses the total, bi-directional ADT; no allowance is made for the peak hour directionality (sometimes called the 'D factor'), which typically is between 0.55 and 0.70 in urban areas.
- Selecting a Traffic Volume—traffic volumes vary through almost all the segments studied; choosing the correct volume is important, generally for predictive purposes. In most cases, the highest volume in the congested segment was used, on the assumption that it is the most likely to be the cause of the bottleneck.
- Estimation of Truck Percentages—These are manual counts, so available for an eight-hour period of the day, which may be different than the congested period. Also, since truck counts are not done every year, some truck data is somewhat older than 2002.
- Definition of Congestion—Other areas may use definitions of congestion different than the one employed in this paper.
- Applicability to Other Urban Areas—the applicability is unknown, as congestion duration may be a function of the localized peaking of demand, which may or may not be transferable to other areas. Local peaking characteristics of traffic in other areas are also likely to affect the duration of congestion.

ACKNOWLEDGMENT/ DISCLAIMER

The author would like to thank the California Department of Transportation (Caltrans) District 4 for the primary data upon which this paper was based; and Christopher Ferrell and David Reinke for help with the statistical analysis. However, the author is solely responsible for the correct interpretation of the data and any opinions expressed herein. The data used to estimate the models herein will be available after March 25, 2007 at the author's website, www.dowlinginc.com by clicking on "Publications."

REFERENCES

ABAG-- Association of Bay Area Governments (2005). *Projections 2005 – Forecasts for the San Francisco Bay Area to the Year 2030.* Oakland: Association of Bay Area Governments.

Caltrans District 4, Office of Highway Operations (2003a). *Information Memorandum: Year 2002 Bay Area Freeway Congestion Data*. Oakland: California Department of Transportation.

ibid. (2003b), *2002 Travel Times & Speed Profiles—Caltrans District 4 Highway Congestion Monitoring Program.* November.

ibid. (1996), *Congestion Monitoring Procedures & Guidelines.* September.

Dey, Soumya S. (1998). "'Hours of Congestion' as a Transportation Measure of Effectiveness Under Capacity Constrained Conditions." In the *ITE Journal*, December. Washington, DC: Institute of Transportation Engineers.

MTC Metropolitan Transportation Commission of the San Francisco Bay Area (2005). *Bay Area Transportation- State of the System 2005*, page 8. December.

SPSS, Inc. (2006) *SPSS 15.0 Brief Guide.* Chicago: SPSS, Inc.

TTI, The Texas Transportation Institute. *The 2005 Urban Mobility Report*, by David Schrank and Tim Lomax, Table 1, page 12. College Station, TX: Texas A&M University.

ibid. (1997) *Quantifying Congestion*, by Tim Lomax, Shawn Turner, *et. al.* College Station, TX: Texas A&M University, March.

Author Information

Steven B. Colman, AICP
Principal
Dowling Associates, Inc.
180 Grand Avenue, Suite 250
Oakland, CA. 94612
Phone 510/ 839-1742 ext. 121
Fax 510/ 839-0871
Email: scolman@dowlinginc.com