**Title**
Empirical Comparison of Travel Time Estimation Methods



**Authors**
Zhang, Xiaoyan
Rice, John
Bickel, Peter

**Publication Date**
1999-12-01

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

# Empirical Comparison of Travel Time Estimation Methods

**Xiaoyan Zhang, John Rice, Peter Bickel**
*University of California, Berkeley*

**California PATH Research Report**
**UCB-ITS-PRR-99-43**

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.



Report for MOU 353

December 1999

ISSN 1055-1425

**This paper uses Postscript Type 3 fonts. Although reading in on the screen is difficult it will print out just fine.**

# Empirical comparison of travel time estimation methods

Xiaoyan Zhang, John Rice and Peter Bickel
Department of Statistics, University of California at Berkeley

September 3, 1999

**Abstract**

In this paper, we conduct an empirical comparison of travel time estimation methods based on single-loop detector data. The methods of concern are the regression method based on an intuitive stochastic model as proposed by Petty et al. in [7], and the conventional method of using an identity relating speed, flow and occupancy with the assumption of a common vehicle length. The analysis is tailored to fit in the limitations imposed by available field data sets. We also introduce several variations of the regression method and give examples which suggest directions for future work to further improve the regression method.

The comparison is composed of three interrelated parts, each with a different focus: local comparison (concerning a single link of freeway), comparison of estimated section travel times over a prolonged stretch of freeway with multiple links and a visualized approach which enables investigation of performance patterns in time and space of the estimation methods.

## 1 Introduction

The travel time is an important variable in transportation engineering. It is a good operational measure of effectiveness of transportation systems and can be used to detect incidents and quantify congestion. Accurate travel time information is crucial in many advanced traveler information system (ATIS) functions. In recent years, many methods have been proposed in order to acquire reliable travel time information.

Some of the methods strive to measure travel times directly using vehicle reidentification technology ([2, 6, 1]). These methods generally require video cameras or other special-purpose equipments. In [1], Coifman proposed a methodology which can match a portion of vehicles using vehicle length measurements, but the method still requires double-loop speed measurements.

The single-loop detector is widely present in existing freeway infrastructure due to its relatively low cost and maturity of technology used. It is thus desirable to understand methods which estimate travel times using single-loop data only. In this paper we conduct a thorough empirical comparison of two such methods: the regression method proposed by Petty et al. in [7] and the conventional method of estimating speed, thus travel time, using a classical identity relating flow, occupancy and speed with the assumption of a common vehicle length. The analysis is based on two field data sets. One is the I-880 data set (see [8]) which is a comprehensive database containing double-loop data over a prolonged freeway section on many days. The other data set used is the SR-24 data set (see [9]) which has highly selected short periods of loop data accompanied by ground truth from simultaneous vehicle data. The I-880 data also has loop-independent travel time information from probe vehicles. We devise comparison strategies which endeavor to reveal characteristic performance patterns of the estimation methods when applied under many different conditions, within the limits imposed by available data. We also incorporate some extensions to the regression method in our comparison.

Although we present competing estimation methods here, it is not our intention to advocate the usage of one method over the other. We hope, by understanding the strengths and weaknesses of each method, to shed light upon the directions to be followed to improve the methods.

The rest of the paper is organized as follows: we introduce the data sets used in the analysis in §2. Each of sections 3 through 5 emphasizes a different aspect of the comparison. Specifically, we introduce the extensions to the regression method and compare the results concerning a single pair of loops in §3. In §4 we consider the situation of applying the methods over a multi-link freeway section and compare section travel times derived from each method on many days. We then take a visualized approach to investigate characteristic patterns of the methods in time and space in §5. We conclude with some discussions in §6.

# 2 Data

We used two field data sets in our analysis: the I-880 data collected for the Freeway Service Patrol Evaluation project ([11]); and the SR-24 data for testing a vehicle matching algorithm from loop signatures ([9]). Details about the data sets are introduced in §2.1 and §2.2 respectively.

The major part of our analysis is based on the I-880 data set. It is a comprehensive database on freeway operating conditions and incidents. In addition to high resolution double-trap loop detector data, it also includes probe vehicle data and incident information. This dataset turns out to be quite suitable for this analysis for a number of reasons. First, it is sizeable in both time and space with many interesting traffic patterns arising. This lends latitude to our analysis and allows us to investigate the behavior of the regression method under a wide variety of road conditions. Moreover, the relatively accurate double-loop speed estimate and the probe vehicle travel times can be used as benchmarks in our comparisons. The probe vehicle data is important to our analysis since it provides additional loop-independent traffic information and is the only source of actual observation of individual vehicle travel times as opposed to point measurements from loop detector. Last, the accompanying incident database contains information on road conditions enabling us to better interpret results.

The SR-24 data is much smaller in scale. It contains two batches of data on a single study link. Each batch has less than 20 minutes of data. The first batch is composed of free flow traffic while the other is of heavy congestion. This data set, although limited, is accompanied by simultaneously collected video data from which vehicle matches between upstream and downstream observations can be derived. This provides us with the ground truth needed to investigate the goodness of estimated travel time distribution by the regression method, a task which we had not been able to do in depth with the I-880 data.

## 2.1 I-880 Data

The I-880 data were collected on a heavily-used stretch on interstate freeway I-880 near Hayward, California. The test site was approximately 6 miles long. In this region, the freeway has four to five lanes with a high-occupancy vehicle (HOV) lane at left. The northbound direction has 18 detector stations; there are 17 stations in the southbound direction. See Figure 1 for a map of the test site. In the data collection period, there were also probe vehicles driven around the study section during peak hours. We mainly used data collected on 20 consecutive weekdays from Feb. 22 to March. 19, 1993 to conduct our analysis. [8] describes the data collection effort in detail.

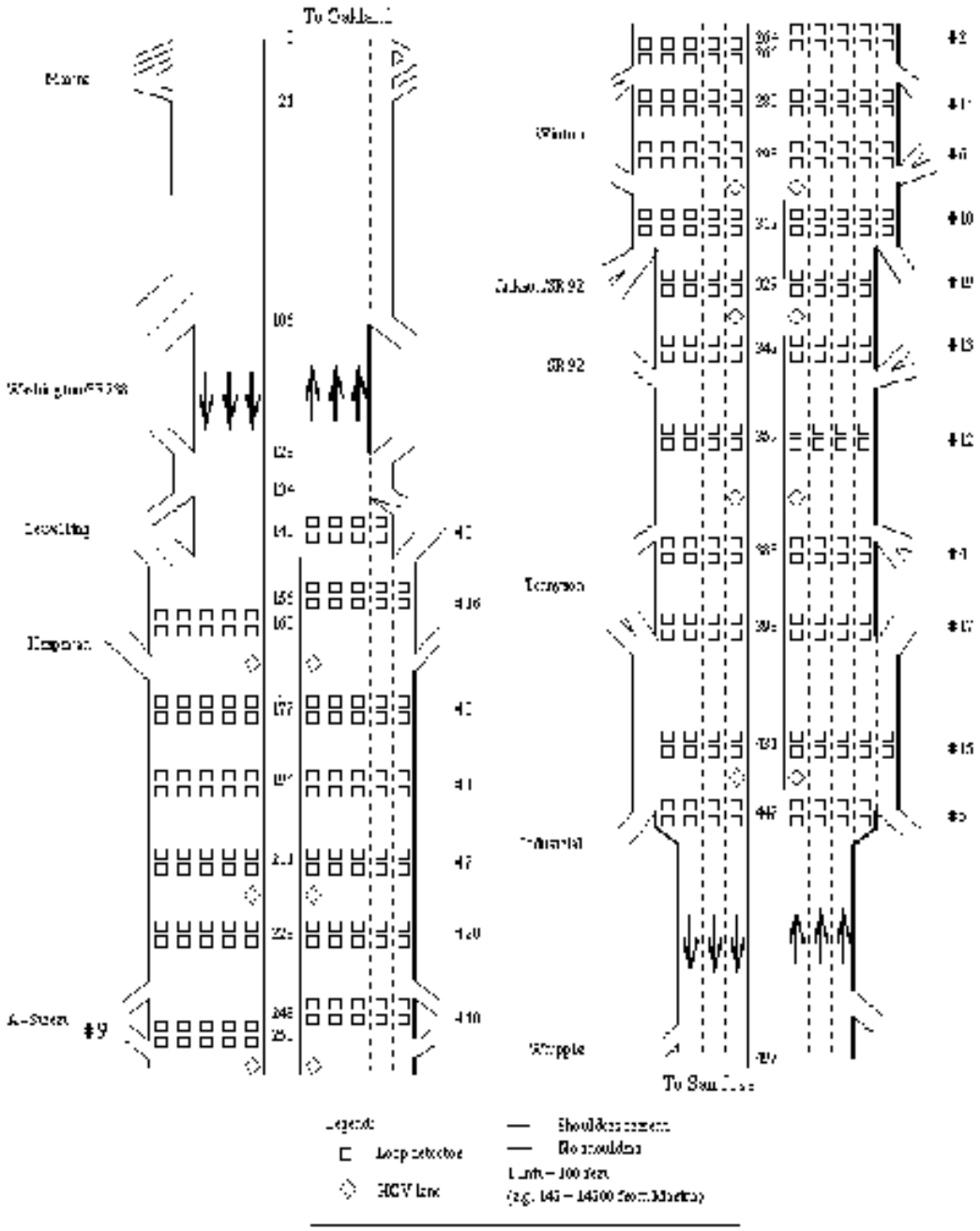


Figure 1: Schematic map of the I-880 test site. The diamond sign denotes the HOV lane.

### Loop data

Loop detectors were placed approximately 1/3 miles apart on each lane. Flows, occupancies and speeds at one second resolution were collected from 5am to 10am and then again from 2pm to 7pm on each weekday. The compressed data for each day is approximately 8 Mbytes.

As observed elsewhere, data quality was an issue. We occasionally missed data for one or two loops for an entire day. Other than that, the most perplexing problems we encountered were the unsynchronized clocks at all detectors and wrong lane labels for a few detectors. We devised data-driven methods to deal with these problems (see [3]). For no obvious reasons, we were unable to find the proper time offsets needed to synchronize detector clocks at some loops for four consecutive weekdays from February 16 to February 19, 1993. We only used data from the remaining 20 days from February 22 to March 19, 1993 whenever unsynchronized clock may be an issue.

### Probe data

Throughout the experiment period, up to four probe vehicles were instructed to drive around the study section at approximately 7 minute headway during peak hours (roughly 6:30am to 9:30am and 3:30pm to 6:30pm). The probe vehicles were equipped with instruments which recorded their trajectories. One can derive travel times over an arbitrary stretch of freeway from these trajectories. The probe vehicles provide the only chance of actually observing travel times. Travel times from loop detector data are based on point measurements of macroscopic traffic conditions. However, the design of the data collection process determines that the resulting data is subject to influences of individual behaviors on the part of probe vehicle drivers and restrictions such as limited availability of one or more probe vehicles on some days.

We observe large fluctuations of the number of probe runs on each half day as shown by the histograms in Figure 2.1. Note that the variability is larger for trips in afternoons to both directions. In the mornings, there were more probe vehicle runs during the last 14 days from March 2nd to 19th, 1993. Within any half day, the actual headways still vary considerably. The number of runs for each half day ranged from merely 4 runs to more than 20. We have a total of 855 runs in mornings and 740 runs in afternoons for both directions.

The actual vehicle headways within a study period of three peak hours still vary considerably, even on a "good" day with relatively more runs. (See Figure 3.) Moreover, occasional erratic driver behavior can mar the data, as illustrated by Figure 4. There the driver actually pulled over to the shoulder for some time before continuing on the trip. The implications of these issues will be discussed where appropriate.

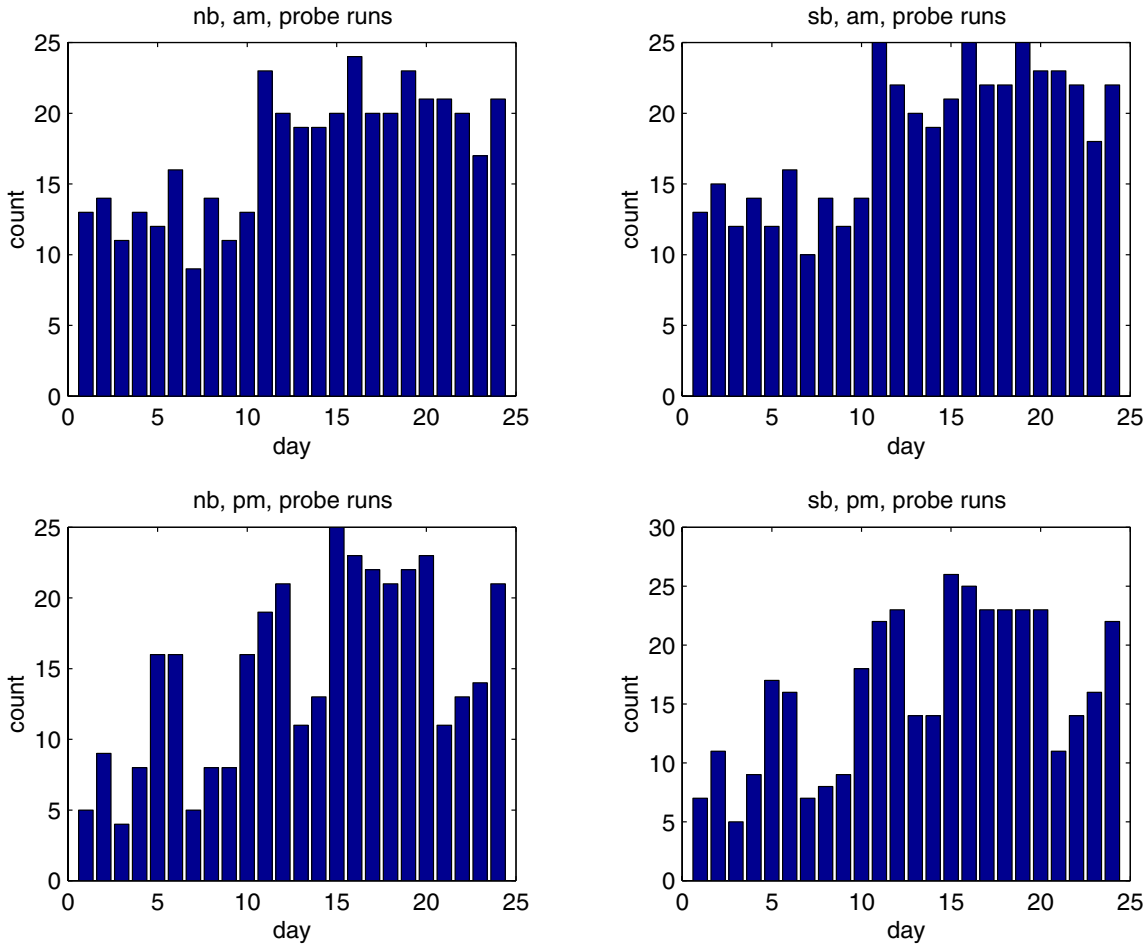


Figure 2: Histograms of the number of probe runs on each of the 24 weekdays from February 16 to March 19, 1993. Left panels are for north bound mornings and afternoons. Right panels for south bound trips. Note that the bottom right panel has a different vertical scale.

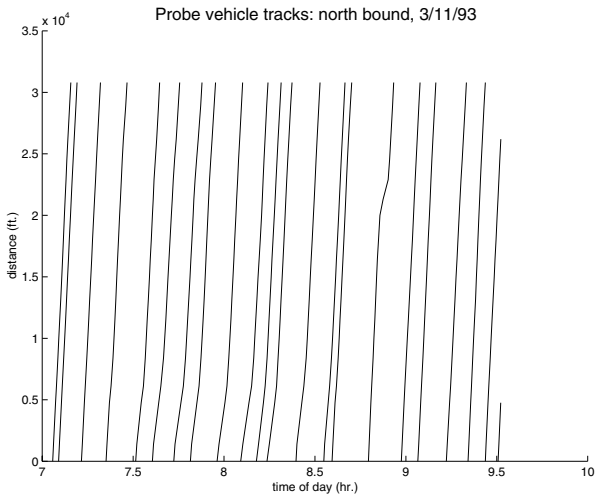


Figure 3: Probe vehicle tracks on a "good" day with 20 runs. Y-axis is location. Traffic travels upward.

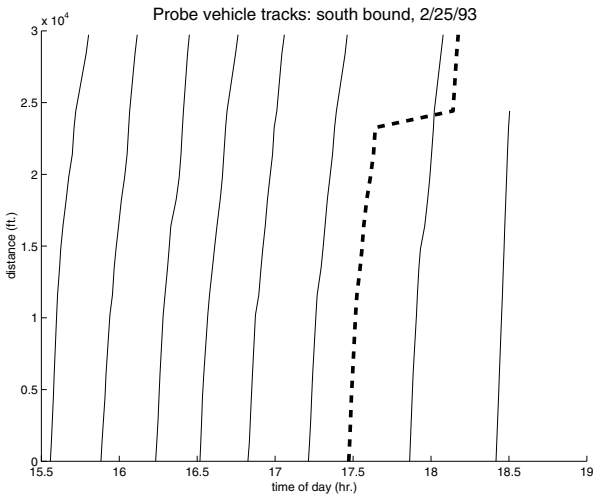


Figure 4: An erratic probe vehicle track is highlighted by the thick dash line.

### Incident data

The incident database was constructed on the basis of reports submitted by probe vehicle drivers on incidents they encountered. For obvious reasons, there may be unreported incidents and reported incident duration could be inaccurate. There were typically many incidents recorded each day. Many of them, such as ticketing events, had no noticeable impacts on traffic. We mainly used the database to confirm presence of incidents when suspected.

## 2.2 SR-24 Data

The SR-24 data were collected to test a vehicle matching algorithm based on signatures from loop detectors. The data consists of signature waveforms of circuit readings and simultaneous video data. We were attracted to this data set mainly because ground truth empirical distribution of link travel times can be derived from the accompanied video data, and because the waveforms can be reduced to *flows* and *occupancies* comparable to outputs of usual single loop detectors. [9] described the data collection effort in detail.

The test data was obtained from a field site on the westbound SR-24 freeway in Lafayatte, California in December, 1996. Two stations 1.2 miles apart were instrumented with both enhanced double-trap loop detectors and video cameras. Loop detectors produced one circuit reading every 0.013 seconds. The reduced loop detector data contains vehicle waveform at the upstream and downstream detectors together with arrival time, lane position, speed and estimated effective length (vehicle length plus detector size). Synchronized video data were collected simultaneously. Vehicles in the video were matched manually to get the ground truth vehicle match. Altogether we had 15 minutes of data with 830 vehicles under free flow

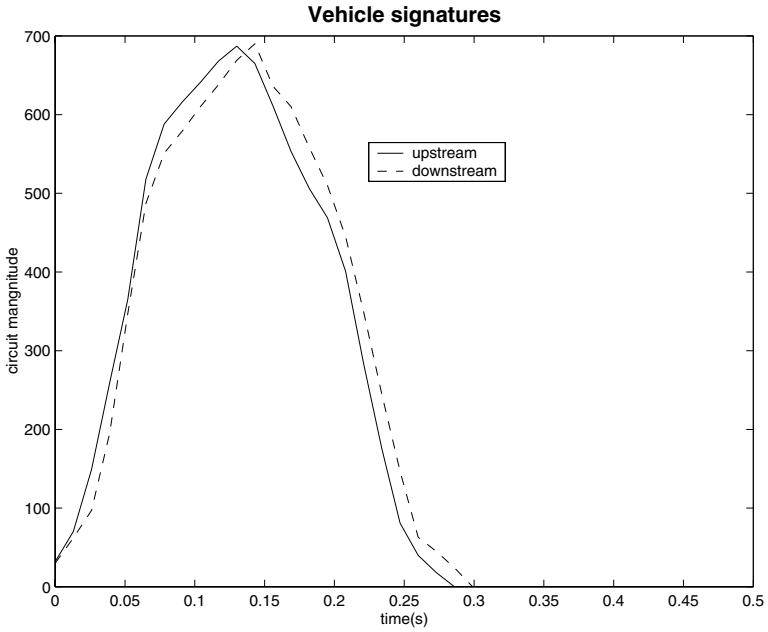


Figure 5: Overlaid loop signatures at upstream and downstream detectors of the same vehicle. Solid line is the upstream signature; dash line the downstream.

conditions on December 6, 1996 and less than 10 minutes of data from a heavy-congested period with around 920 vehicles on December 12, 1996.

Figure 5 shows overlaid waveforms of the same vehicle at the upstream and downstream stations. These raw waveforms are not suitable for our analysis. We derived time-stamped vectors of *flows* and *occupancies* for each lane at one second resolution in order to be consistent with the I-880 loop data set. The waveforms are reduced as follows: the flow(veh/second) at the second in which a waveform starts (i.e. the second at which a vehicle arrives at the detector) is imputed to be one; the occupancy(% of time for which the detector is "on") is computed by dividing the number of positive readings within the second by 77, which is the approximate number of outputs when sampling at 0.013 second interval.

# 3 Local travel time estimation

This section is concerned with comparison at a local scale. In §3.1 we briefly introduce the single-loop based estimation methods involved in the comparison. To avoid reproducing conclusions in [7], the comparison is largely between the method introduced there and some of its variants. The comparison is carried out in two aspects: as travel time estimator and as travel time distribution estimator. The section is concluded with discussions of the reasons of the methods' difficulties in dealing with congestion.

## 3.1 Basic Methods

In this section, we introduce the travel time estimation methods involved in the comparison. As mentioned before, this article is devoted to an in-depth performance evaluation of the link travel travel time estimation method proposed by Petty et al. In doing so we compare it with a frequently used estimator which utilizes a theoretical relationship involving speeds, flows and occupancies, assuming a common vehicle length constant, as well as with benchmarks such as double-trap speeds or ground truth measurements from video data (in the SR-24 dataset) when available. For convenience of narration, we will use the regression method and the CVL (common-vehicle-length) method to refer to the two estimation methods later.

Since the CVL method is incorporated in the regression method, the section is organized as follows. We start by introducing the CVL method. We then reiterate the methodology of the regression method. We go on to present some possible extensions of the regression method in efforts to improve accuracy.

### 3.1.1 The common-vehicle-length(CVL) method

In the absence of double-trap detectors, speed is often calculated on the basis of flow and occupancy:

$$\text{speed} = \text{flow}/(\text{occupancy} * g) \tag{1}$$

where $g$ is assumed to be a constant to convert occupancy to density and is related to the average effective vehicle length (vehicle length plus detector size). The method is sensitive to the value of $g$. In practice, $g$ is usually calibrated during light traffic conditions by imposing a value for free-flow speed. The reliability of the method lies in the validity of assuming $g$ to be constant over major portions of operating condition. Hall and Persaud ([5]) investigated the assumption and presented results suggesting that $g$ might be prone to a systematic bias with respect to occupancy.

### 3.1.2 The regression method

Petty et al. ([7]) proposed an accurate estimation method based on a simple stochastic model. Crudely, the model assumes that for a given brief interval of time, vehicles arriving

at an upstream detector "choose" their travel times to the downstream detector at random from a common probability distribution. This distribution is then estimated by comparing the cumulative downstream arrival process, in a single lane for now, to its conditional expectation given the upstream arrival process. In doing so the CVL method is used to derive the *fit window*, which is the assumed support of the distribution to be estimated.

More rigorously, we consider arrival processes at a pair of immediately adjacent detector stations. We assume that the arrivals are exchangeable (thus not distinguishing passenger cars and trucks, for instance)[1]. Let $X(t)$ be the cumulative arrivals and $x_t$ be the flow count at the upstream detector; similarly, $Y(t)$ and $y_t$ for the downstream one. For simplicity we take that $x_t$ and $y_t$ be measured every one second[2]. We have

$$dX(t) = x_t dt = \sum_i \delta(t - \sigma_i) dt \tag{2}$$

$$dY(t) = y_t dt = \sum_j \delta(t - \sigma_j - \tau_j) dt \tag{3}$$

where $\sigma_i$'s and $\tau_j$'s are upstream arrival times and travel times respectively. The model postulates that conditioning on all upstream events, $\tau_j$'s are exchangeable and have a common distribution independent of $j$ and $\sigma_j$'s, for $T_B \leq \sigma_j \leq T_F$. The assumption excludes such situations as a change of traffic regimes during $[T_b, T_F]$.

Let $f(\cdot)$ denote the common marginal travel time density under the conditions. It can be argued that, under realistic conditions

$$\begin{aligned} E[dY(t)|X] &= \int_{-\infty}^{t} \sum_j \delta(t - \sigma_j - \tau_j) f(\tau_j) d\tau_j dt \\ &= \sum_j f(t - \sigma_j) dt \\ &= < f, x >_t \, dt \end{aligned} \tag{4}$$

where $< f, x >$ is the discrete convolution between $f_s = f(s)$ and $x_t$. Relate the above equation to (3). We see that $f(\cdot)$ can be estimated by the choice which minimizes the discrepancy between $< f, x >$ and $y_t$, according to some criterion like least squared error. Assuming that $f(\cdot)$ has support $[a, b]$ (the *fit window*), we have

$$\hat{f}_s = min^{-1} \sum_{t=T_B+b}^{T_F+a} \left( y_t - \sum_{s=a}^{b} x_{t-s} f_s \right)^2 \tag{5}$$

[1]This assumption is necessary for loop detector datasets with no vehicle signature information, which basically include all datasets with aggregation level in the order of 1 second or more.

[2]c.f. [7] for applying the method to data with measurement interval other than one second.

Equation (5) reminds one of linear regression, which is why it has been called the regression method. Actually, the method has been implemented using the nonnegative linear regression package in MATLAB to conform to $\hat{f}_s \geq 0$.

As established in [7], we use a fix-width adaptive *fit window* $[a, b]$ which centers at the travel time estimate based on (1) with a suitable $g$. The width of the fit window is taken to be 30 seconds in the examples unless otherwise specified. We generally take the result based on a 300 second data window $[T_B, T_F]$ centered at $t$ as estimated travel time for upstream arrival time $t$.

**Mean, median or mode?** The methodology above describes a procedure for estimating the travel time distribution in the assumed model. Having determined such an estimate, the question remains as to which summary statistic to use to estimate the quantity of practical interest — the link travel time. The common choices are: the mean, the median and the mode. The authors of [7] chose the mode over the mean, arguing that the former is less sensitive to the choice of the fit window than the latter. However, the mode is quite sensitive to the inherent variability associated with simultaneous estimation of large number of parameters. We decided to settle on using the median as the travel time estimator. Our experiences with the I-880 dataset suggest that the median is, at least to say, not worse than the mode in this scenario.

### 3.1.3 Some extensions of the regression method

The regression method aims to estimate the travel time distribution under the hypothetical model formulated earlier. The extensions to the method described here endeavor to improve the estimated distribution without modification of the model. We delay the discussion about validity of this hypothetical model until later.

**B-splines** The number of parameters to be estimated in the regression method is determined by the fit window. One would want to use a sufficiently large fit window to reflect a wide range of driver behaviors. However, the large number of parameters could undermine the reliability of the resulted density estimate. We can bring down the number of parameters by imposing some parametric or semi-parametric forms on the travel time distribution. In particular, we choose to use second order B-spline (piecewise linear) approximation (see [4]).

Specifically, we dictate that the estimated travel time distribution $f(\cdot)$ be of the following form:

$$f(s) = \sum_{i=0}^{N_{sp}} a_i \cdot B_i(s) \tag{6}$$

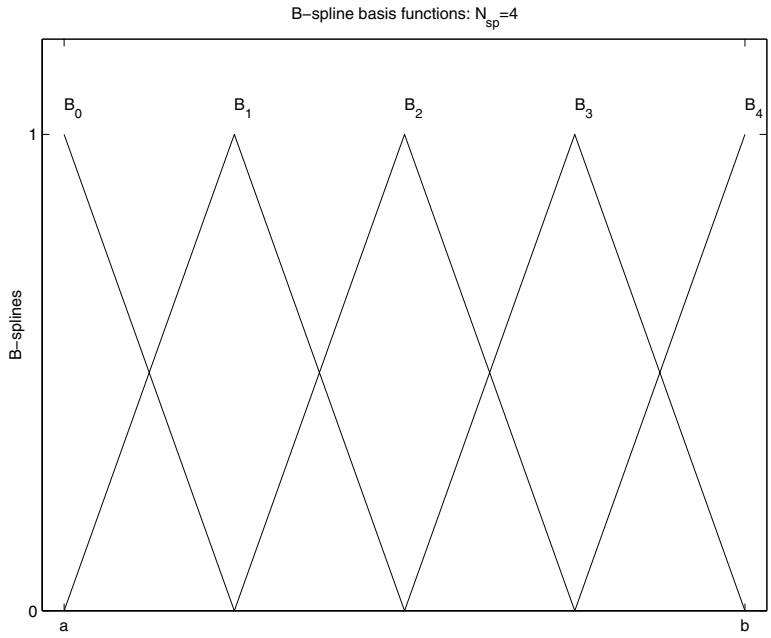


Figure 6: B-spline basis functions for $N_{sp} = 4$.

where $B_i(\cdot)$'s are second order B-spline basis functions as shown in Figure 6. $f_s$ is then estimated by $\hat{f}_s = \sum_{i=0}^{N_{sp}} \hat{a}_i \cdot B_i(s)$ where

$$\hat{\boldsymbol{a}} = min^{-1}_{\{\boldsymbol{a}: a_i \geq 0\}} \sum_{t=T_B+b}^{T_F+a} \left(y_t - \sum_{s=a}^{b} x_{t-s} \sum_{i=0} N_{sp} a_i B_i(s)\right)^2$$

The resulted $\hat{f}_s$ is then normalized to have area equal to one.

The B-spline approximation is indeed a non-full-rank linear mapping of the original parameter space. When $N_{sp} = b - a$ where $[a, b]$ is the fit window, the estimate is exactly the same as that from the original regression method. The more interesting situation is when $N_{sp}$ is considerably smaller than the number of parameters under the original regression method. The B-spline approximation amounts to a variation diminishing approximation of the original estimate. We expect the estimated distribution to be increasingly smooth with decreasing $N_{SP}$. Since it is but only an approximation, it is not expected to succeed where the original method fails. As will be demonstrated in examples later, the regression method enhanced with B-spline approximation does not lose much, if any, to the original method in terms of accuracy in estimated travel times when a suitable $N_{sp}$ is used.

**Exponential smoothing** The idea of applying exponential smoothing to the regression method is motivated by the fact that estimation at a certain time point does not take

advantage of prior estimated travel time distribution even though it is reasonable to assume travel time distribution evolves gradually rather than abruptly. Let $\hat{\boldsymbol{f}}_i$ be the estimated travel time distribution at time point $t_i$, where $t_0, t_1, t_2, \ldots$ are assumed to be evenly spaced for simplicity. The distribution estimates $\hat{\boldsymbol{f}}_i^E$'s from the exponential smoothing version of the method are:

$$\begin{aligned} \hat{\boldsymbol{f}}_0^E &= \hat{\boldsymbol{f}}_0 \\ \hat{\boldsymbol{f}}_1^E &= \lambda \hat{\boldsymbol{f}}_1 + (1-\lambda)\hat{\boldsymbol{f}}_0^E \\ &\ldots \\ \hat{\boldsymbol{f}}_i^E &= \lambda \hat{\boldsymbol{f}}_i + (1-\lambda)\hat{\boldsymbol{f}}_{i-1}^E \end{aligned} \tag{7}$$

The parameter $\lambda$ decides the importance of the estimate at current time point. $\lambda = 1$ corresponds to the original regression method. Since congested conditions are more volatile, it is sensible to use larger $\lambda$ values under such conditions.

The idea is also inspired in part by the poor performance of the estimated travel time distribution from the regression method during congested times, as we found when dealing with the I-880 dataset. We suspect that the difficulty may be due to lack of homogeneity in the batch of data used for estimation. We hope that we can improve the estimates by combining exponential smoothing with a smaller data window.

## 3.2 Examples

Since Petty et al. already conducted careful comparison of the regression method and the CVL method at a local scale, the examples here concentrate on comparing the original regression method with its variations. The examples are organized as follows: §3.2.1 compares the original regression method with its modifications in terms of accuracy of travel time estimates based on the I-880 data, using travel times from the double-loop speed as benchmark. In §3.2.2 we discuss the effect of $N_{sp}$ on estimated travel time distribution and present results which may be the basis of a rule-of-thumb for selecting the number of B-splines to be used in B-spline approximation. We use the SR-24 data set in this part of the comparison. Finally we present examples illustrating the difficulty of the regression method in dealing with heavy congestion condition.

### 3.2.1 Link travel time estimation

**B-spline approximation** In §3.1.3 we noted that the B-spline approximation is able to cut down the number of parameters in the original regression method dramatically. The extension would not be an improvement if the accuracy of the travel time estimates was inferior.

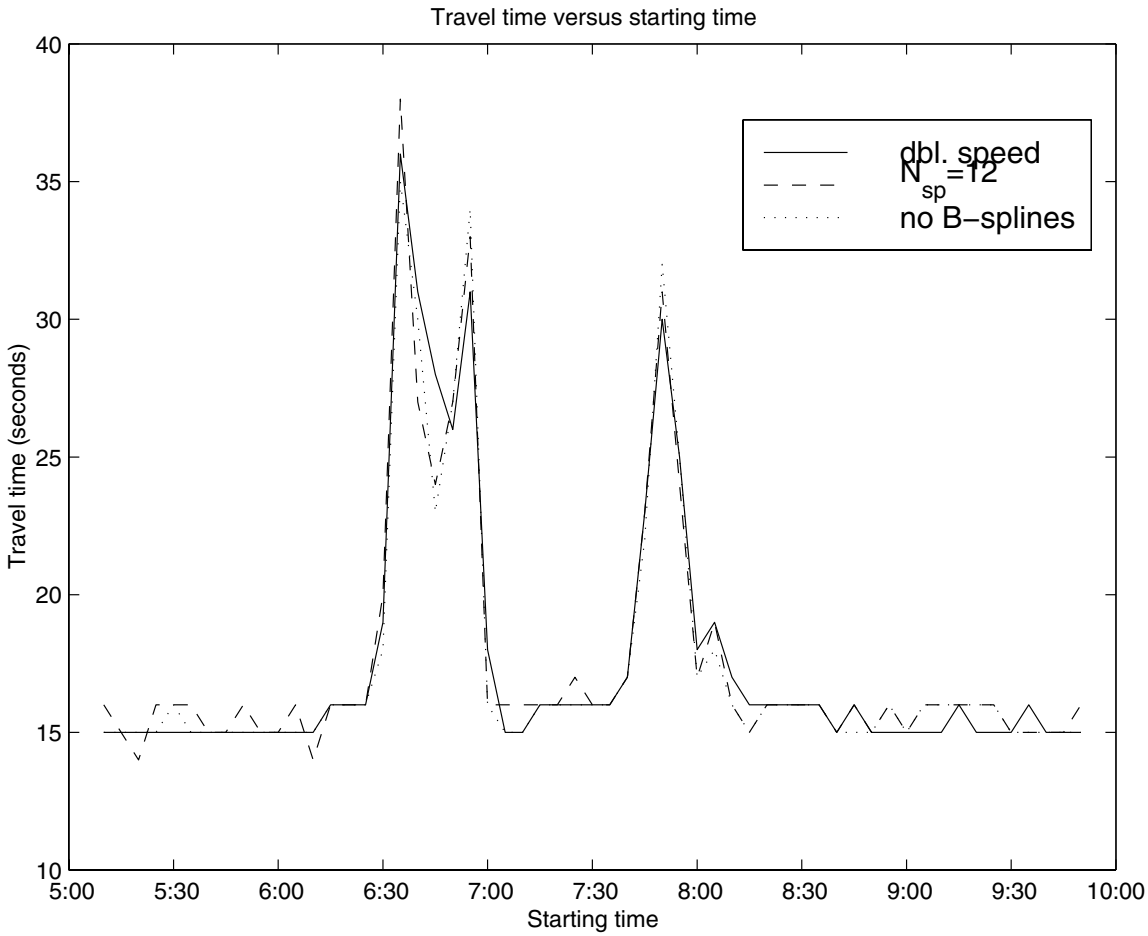


Figure 7: Link travel time estimated by B-spline extension. Solid line shows double-loop estimates, dash line is the estimate using $N_{sp} = 12$, the dotted line is the original regression method estimate.

In Figure 7 we compare the estimates from both the original regression method and the B-spline approximation with $N_{sp} = 12$. In this example we focus on two loop detectors 1360 feet apart in the I-880 dataset. It takes 15 seconds to traverse the link at 60 mph. We use a data window of $|T_B - T_F| = 300$ seconds. The fit window has a width of 30 seconds and is centered at the travel time estimate given by the CVL method with $1/g = 22.6$ ft. The double-loop travel time estimate is computed by taking the average of speeds at upstream and downstream detector to be the navigation speed through the entire link. In this figure, the dashed line and the dotted line overlaps over a large portion of time, showing that using $N_{sp} = 12$ does not deteriorate estimation accuracy. Both of the estimated travel times correspond to the double-loop travel times fairly well even during periods of considerable congestion.

**Exponential smoothing** Figure 8 contrasts the link travel time estimates using exponential smoothing with $\lambda = 0.4$ and the original method ($\lambda = 1$). Comparatively, the result with exponential smoothing shows more variation under free flow conditions. Given the smaller data window used by this method, the result is expectable. Since the variation in estimated travel times is small in scale (mostly within 1 second), this does not suffice as strong negative evidence. However, note how the dashes line slight lags behind the other two lines, most obvious when the travel time decreases rapidly. This suggests that the smoothing coefficient selected is already too small. We do not have evidence that exponential smoothing consistently improves the original regression method. We do not pursue further along this line.

### 3.2.2 Estimation of travel time distribution

**Effect of the number of B-splines** Here we concentrate on comparing the original regression method with B-spline enhanced versions with respect to estimated travel time distribution. The B-spline extension to the regression method poses a question about selection of $N_{sp}$, the number of B-splines to be used. The situation is similar in spirit to the bandwidth selection in kernel density estimation. By using fewer B-splines, one conceptually gains in terms of variance reduction but incurs larger bias.

In the I-880 dataset, we only have travel times induced from double-loop speeds as proxy for ground truth. Under light traffic, the empirical distributions of travel times derived from the upstream detector and the downstream one correspond to each other fairly well, but this is not the case under even moderate congestion. We adhere to $N_{sp} = 12$ for the I-880 dataset more or less arbitrarily.

We do have ground truth travel times from video data in the SR-24 dataset. It gives us the opportunity to observe the effect of $N_{sp}$ in terms of goodness of estimated distribution. Figure 9 shows estimated distributions with different values for $N_{sp}$. We note that the original regression travel time distribution estimate is quite bumpy with many spurious peaks. B-spline approximation effectively restricts the number of peaks and hence smoothes the resulted distribution estimate. The plots show that the smoothed estimates track the majority of mass fairly well.

The top left panel in Figure 9 shows L1 distances between the estimated distribution and the ground truth. The dashed horizontal line is the L1-distance between the regression distribution estimate and the empirical ground truth. Curiously, there is a minimum around $N_{sp} = 4$. We looked at a few other examples. The minimums all occurred in the region $N_{sp} = 3$ through $N_{sp} = 6$. The existence of a consistent region of minimums suggests intrinsic degree of freedom for the underlying travel time distribution. It raises serious hopes for parsimonious parametric models or mixture models. The B-spline estimate is itself a mixture.

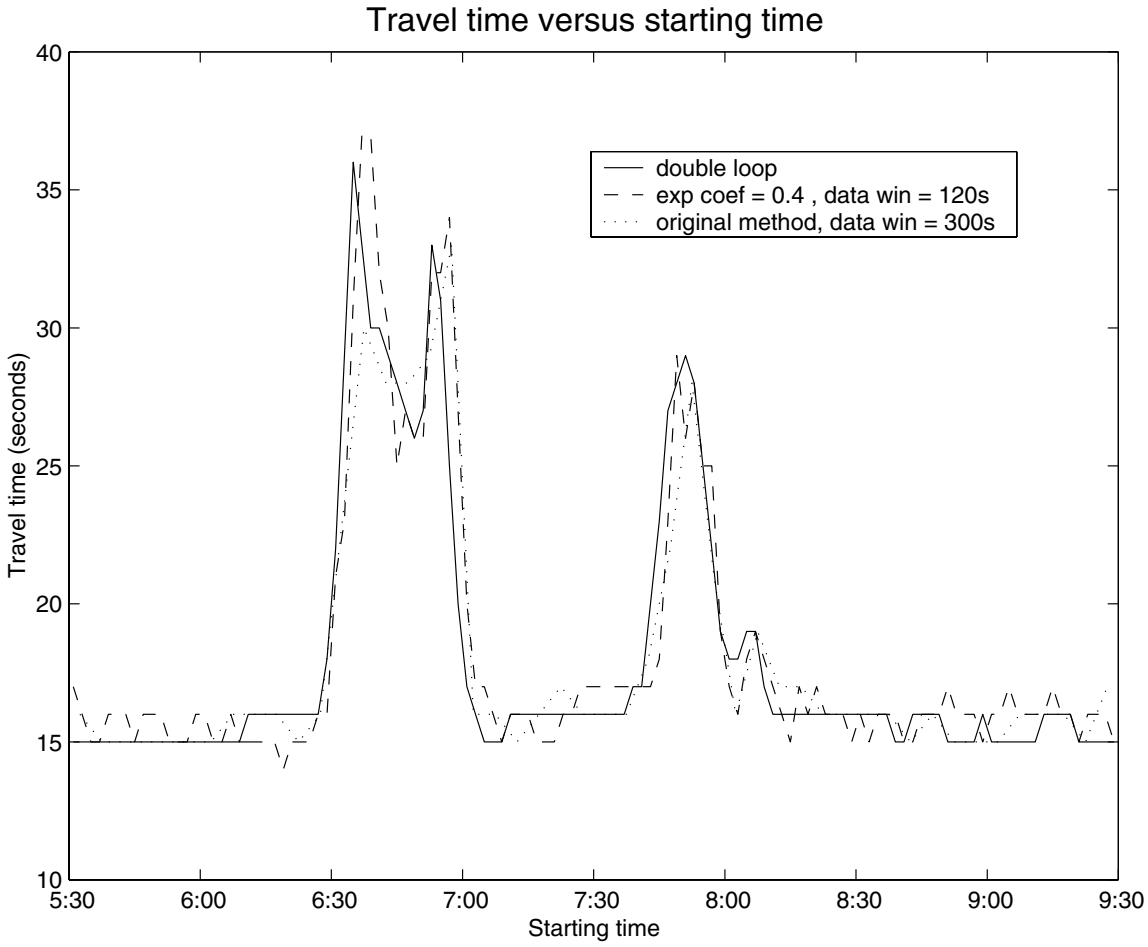


Figure 8: Link travel time estimated by B-spline extension. Solid line shows double-loop estimates, dash line is the estimate with exponential smoothing coefficient $\lambda = 0.4$ using 120 second data window; the dotted line is the original regression method with estimate with 300 second data window. Both of estimated travel times use B-spline approximation with $N_{sp} = 12$.

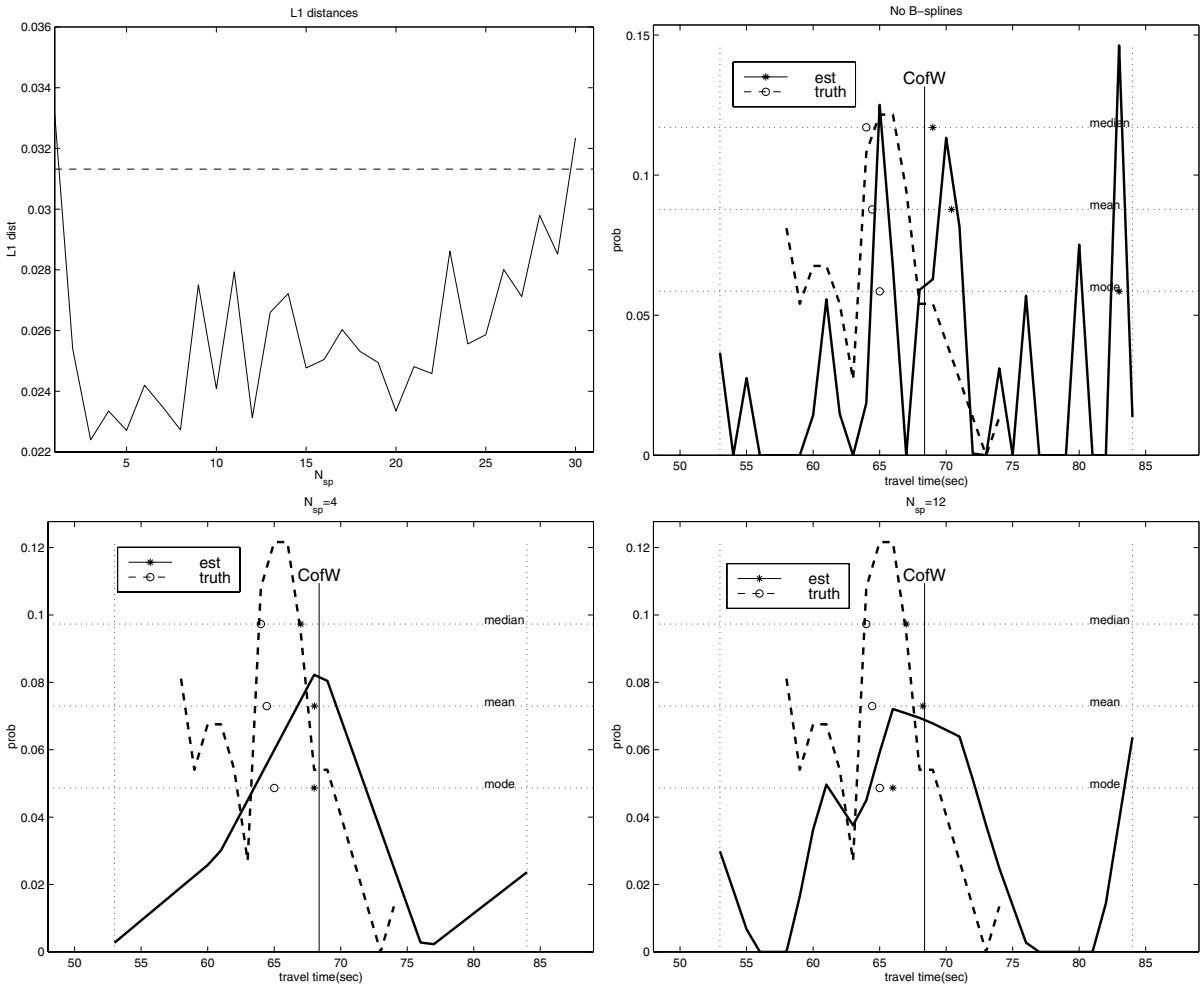


Figure 9: Effect of $N_s p$. Top left panel: L1 distances between $\hat{f}_s$ and the ground truth $f_s$ versus number of splines used; top right: solid line is the estimated distribution using the original regression method, dash line is the ground truth from video data; bottom left and right are for $N_{sp} = 4$ and $N_{sp} = 12$ respectively. The data used is the SR-24 data. The time period of the data is around noon and there was no congestion. Recall that the link is 1.2 mile long.

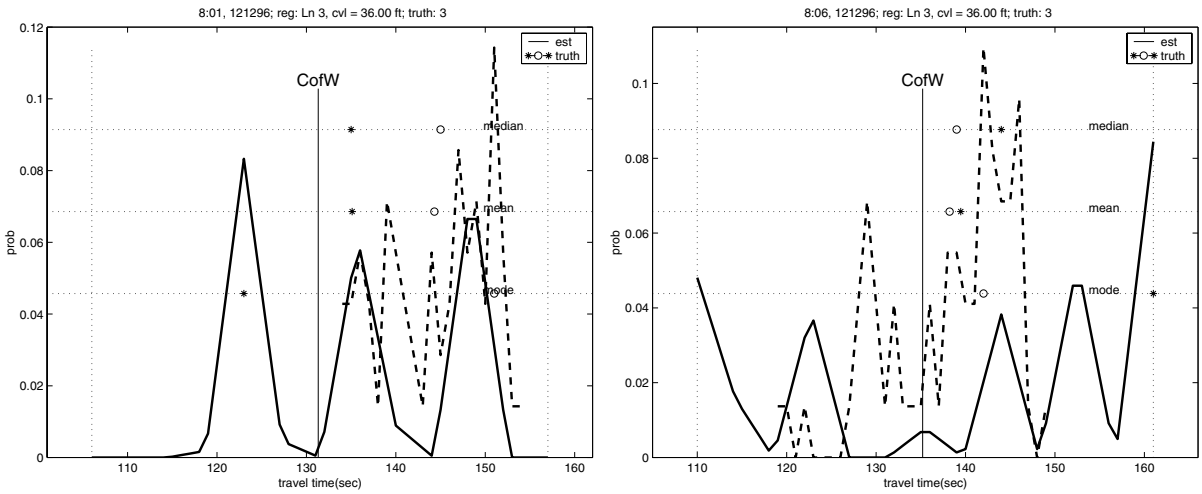


Figure 10: Estimated travel time distribution during congestion. The method used is the regression method with $N_{sp} = 12$. Note that it takes 72 seconds to traverse the 1.2 mile link at 60 mph.

### 3.2.3 Difficulty with the method during congestion

Figure 9 shows that the estimated travel time distribution during free flow condition after appropriate B-spline approximation is reasonable. We cannot say the same for estimates under heavy congestion. Figure 10 shows estimated travel time distribution using B-spline enhanced regression method at two time points approximately 5 minutes apart during a period of heavy congestion. Note that the fit window has to be enlarged to 50 seconds from 30 seconds used in previous examples so that the support of the ground truth empirical distribution is covered for both of the estimation points. The estimated travel time distribution assigns significant portion of mass in regions where none exists. The resulted travel time estimator is also far from satisfactory accordingly. The regression method fails in this case.

## 3.3 Discussion

Via the comparison in this section, we find that the B-spline approximation is a favorable addition to the original regression method which basically behaves as well as the regression method in terms of estimated travel times, and effectively improves the estimated distribution by ruling out some false peaks. (We will take the regression method to be implicitly using B-spline approximation from now on.)

However, the B-spline approximation is not a remedy for the difficulty during congestion (see Figure 10). More fundamental changes the method may be necessary to improve its behavior during congestion.

### 3.3.1 Validity for the underlying model of the regression method

The regression method and its extensions introduced so far are all based on a crude model which clearly can not be assumed to hold even approximately under heavy congestion. The difficulties we encountered with these methods prompted us to think that the model might be too primitive to accommodate the heterogeneity of traffic in some cases. Figure 11 plots travel times versus upstream arrival times for periods of free flow and congestion. It is clear to see that although the assumption could be taken to hold approximately under free flow, it clearly fails during heavy congestion in that one cannot find a practical data window in which it can be assumed that travel times of vehicles arrived within the window can taken to be exchangeable. It is also suggested in [1] that during congestion vehicles travel more in platoons as a result of mutual influence imposed by increased volume and limited road capacity. Changes to the model to accommodate such platoon behavior may be needed to improve the regression method.

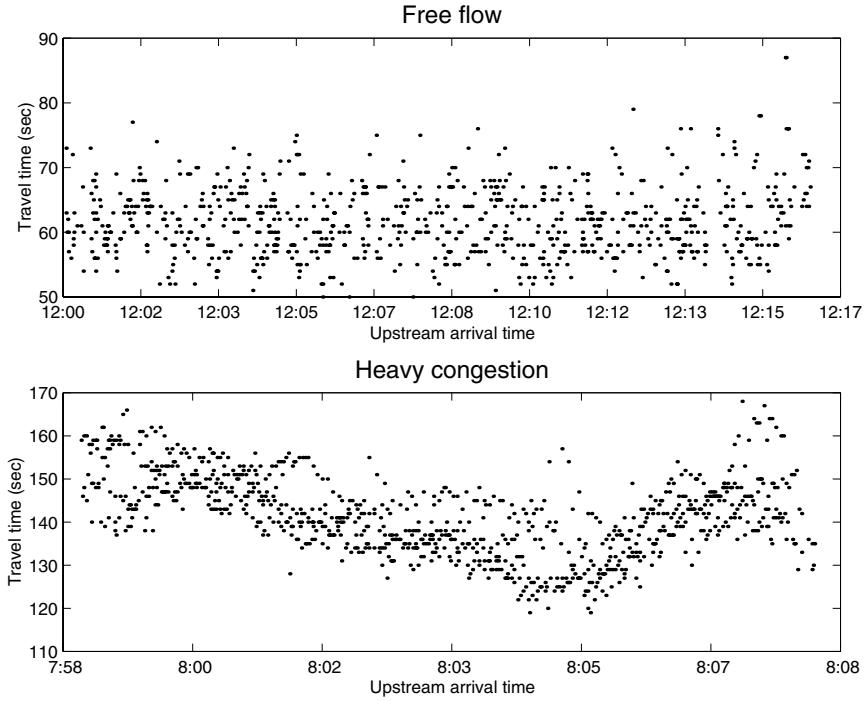


Figure 11: Travel times versus arrival times based on the SR-24 data. Top panel is for free flow situation; bottom panel for congested period. Note that Y-axis limits are different.

# 4 Comparison of estimated section travel times

In this section, we address the issue of estimating travel times over a prolonged multi-link freeway stretch, i.e. section travel times. Information on section travel times over alternative routes is essential in route guidance systems. The section travel time is also a readily understandable gauge of traffic system performance. However, direct measurement of section travel times is generally expensive to get and often requires special-purpose hardware. The analysis described later compares the section travel time estimates based on two single-loop based methods — the regression method and the CVL method. These empirical results may benefit those who need to estimate section travel times using single-loop detector data only.

Our analysis is based on the I-880 data. In the particular context, the section travel time is considered to be the amount of time needed to traverse the entire stretch of the test site in either north or south bound direction [3]. As already mentioned, this field data set contains both double-loop speeds and probe vehicle trajectories. Both of them act as proxies for benchmarks in our comparison in the absence of ground truth.

The section is organized as follows: We explain how to obtain section travel times from each of the four sources: probe vehicles, double-loop speeds, the regression method and the CVL method in §4.1. We then discuss the imperfections of the first two as source of ground truth and the comparison strategy we devised in light of the situation. Empirical results are then presented.

## 4.1 Obtaining section travel times

For the I-880 data, there are four sources to get section travel time information — probe vehicles and three loop-dependent means: double-loop speeds, the regression method and the CVL method. It is straightforward to compute section travel times from the probe vehicle trajectories by simply taking the difference between time stamps at the beginning and end of the trip. For the latter three, section travel times are derived in a similar fashion from hypothetical trajectories constructed with link travel time estimates. §4.1.1 describes the process of constructing a hypothetical trajectory from link travel times. We cover obtaining link travel time estimates in §4.1.2.

### 4.1.1 Hypothetical trajectory

We consider a $K$-link freeway stretch with detectors located at $x_0, x_1, \ldots, x_K$. A trajectory is defined by $\{(x_i, t_i), i = 0, \ldots, K\}$ where $t_i$ is the time at which a (hypothetical) vehicle passes location $x_i$. The section travel time departing at $t_0$ is thus $t_K - t_0$. Denote the

[3] Due to missing loop data, we were only able to get travel times covering a large part of the test site. See §4.3.1 for detail.

estimated link travel times for the $k$-th link at time $t$ to be $T_k(t)$ for $k = 1, \ldots, K$. The hypothetical trajectory based on collection of these estimates is constructed as follows:

$$t_1 = t_0 + T_1(t_0)$$
$$t_k = t_{k-1} + T_k(t_{k-1}), \text{ for } k = 2, \ldots K.$$

### 4.1.2 Link travel time estimation

Here we describe how to construct link travel time estimate at time $t$ using double-loop speeds, the regression method and the CVL method, respectively. For all three methods, the estimate is based on batch of data measured within a time window. We refer to this window as the *data window*. We use a 300-second data window centering at $t$ throughout the analysis in this section.

There are slightly different ways to construct link travel time estimates for each of the three methods. For example, for the double-loop speeds and the CVL method, one can choose to compute the travel time using the average of speed measurements from both the upstream and downstream detectors, instead of using only the upstream speeds (our procedure). One can also choose to use different values for parameters (such as the number of B-splines $N_{sp}$) involved in the regression method. We chose the particular form of implementation largely arbitrarily. We believe that the results would not change in essence with different implementations or parameter settings.

For now, we only used data from a single lane to estimate travel times for all three methods. The lane we used is lane 3 which is the third lane from the left and is the outer middle lane for major part of the I-880 test site. Loop data from middle lanes are usually easier to work with [4].

**Double-loop speeds:** For the double-loop speed, the link travel time is computed by assuming that the vehicle traverses the link at the average upstream speed, where the average is taken over speed measurements within the data window at the upstream detector.

**The regression method:** §3.1.2 describes the methodology of the regression method. We use the B-spline extension with $N_{SP} = 12$ for this part of the analysis. The center of the fit window is estimated using (1) with data from both the upstream and the downstream detector.

[4] The innermost and outermost lanes are usually avoided in studying traffic patterns because: the innermost lane is usually occupied by faster vehicles, the outermost lane is connected with the on/off ramps and experiences frequent changes in traffic volume caused by merging/leaving vehicles.

**The CVL method:** For the CVL method, equation (1) is used to get a speed estimate. The *flow* and *occupancy* in the equation are taken to be the corresponding average within the data window at both the upstream and the downstream detectors. The link travel time is computed accordingly. The *common vehicle length* constant $1/g$ is taken to be 22.6 feet. The same constant is also used in the regression method in computing the fit window. The resulted link travel time estimate is generally different from the center of the fit window in the regression method, since the latter uses data from both detectors.

## 4.2 Benchmarks

The object of our analysis is to compare the performances of the regression method and the CVL method as section travel time estimators. We have reasons to believe that the performances of the two estimators depend on traffic condition. Therefore, an ideal benchmark should provide ground truth section travel times under all traffic conditions in the scope of the comparison.

Before further discussion of benchmarks to be used in the comparison later, it is necessary to clarify our notion of ground truth section travel times. The section travel time of a given vehicle is unambiguously the time elapsed between the vehicle entry and exit of the section studied. Travel times from probe vehicle data are direct measurements of section travel times for the few probe vehicles. In practice, one is generally interested in some measure of "average" travel times such as the mean or median travel time for all vehicles passing through the section within a suitable time period. Hence the ground truth can only be acquired by tracing all vehicles throughout the study section for a prolonged time period, which is generally too difficult to accomplish in the real world. A more realistic alternative is to trace a sufficiently large (random) sample of vehicles via vehicle reidentification mechanisms, such as video tracking or using vehicles equipped with transponders.

For the I-880 data set, the only choices we have are section travel times derived from probe vehicle data and double-loop speeds.

### 4.2.1 Probe vehicle travel times

The probe vehicle data provides a sparse sample of direct measurement of vehicle travel times for our purpose. It is also arguable that this sample could be biased in the sense that probe vehicle drivers were consistently more or less aggressive than typical drivers. The latter assertion sounds more reflective of the actual situation given the fact that probe vehicle drivers had the additional responsibility to relay incident information they encountered on the way. To check this, we examined the scatter plot of probe vehicle speed when passing a loop detector station versus the speed measured by the corresponding double-loop detector. The plot is shown in Figure 12. The solid line is the 45 degree line. Note that the majority of the points is below the line. This is evidence supporting that probe vehicles tend to be

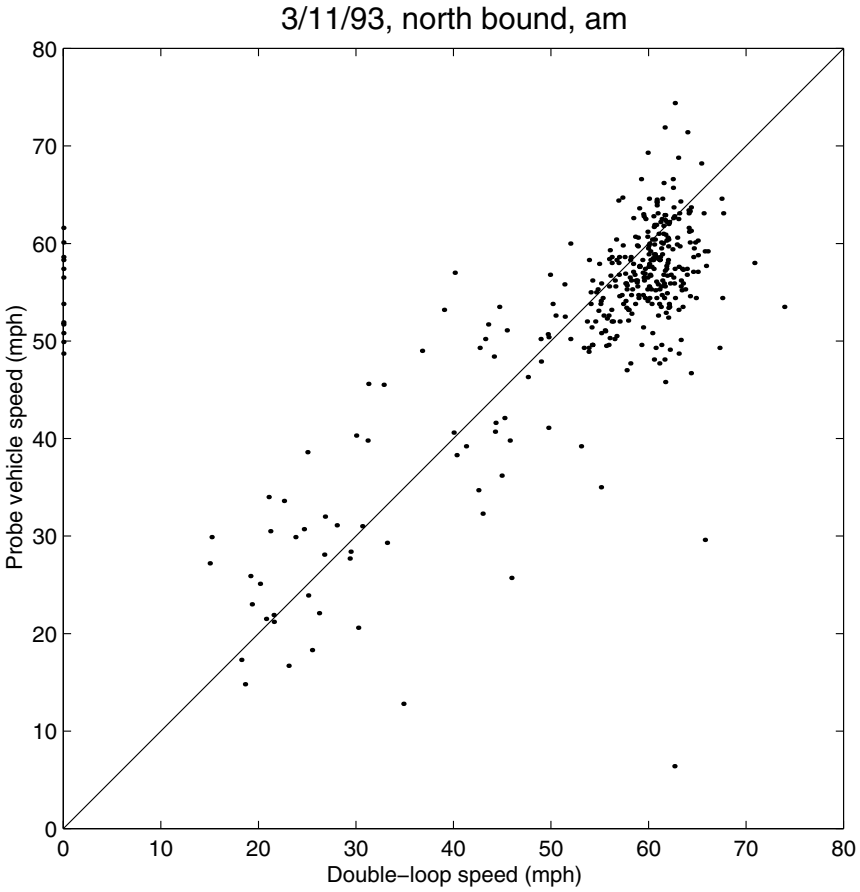


Figure 12: Probe vehicle speeds versus the double-loop speeds. The X-axis represents the double-loop speed. The plot is based on 357 points of the north bound runs in the morning of March 11, 1993. The double-loop speed plotted is the average speed of measurements within 30 seconds of the time of passing of the probe vehicle. The diagonal line is the 45-degree line.

slower than the general traffic[5].

Another source of sampling bias for the probe vehicle section travel times is caused by the limited number of available vehicles. As mentioned earlier, there were only a fixed number of probe vehicles available for driving around the test site loop as any time. The number of runs these vehicles could make is smaller when it took longer to go around the loop as in congested periods. This poses a dilemma since traffic conditions in congestion are more volatile than those in free flow, and one needs more points to characterize traffic in congestion

[5]Careful reader may notice that there are a few points on the vertical axis in Figure 12. This is normal since we only used data from a single lane in computing the double-loop speed in the plot. If there is no vehicle passing in that lane within 30 seconds of the moment of passing for a probe vehicle, the corresponding double-loop speed will be zero.

rather than less.

### 4.2.2 Section travel times from double-loop speeds

Double-loop speeds are point measurements. One implicitly assumes that vehicle speed remains constant in computing travel times from them. This works fine in free flow situation, but is more questionable in congestion when the traffic is highly volatile and actual speeds change quickly. We do not have means to rigorously quantity the bias induced as such for the I-880 data. It is reasonable to assume that the bias is small in magnitude relative to the inherent variability of a single observation as provided by probe vehicle data.

## 4.3 Comparison

As already mentioned, the goal of the comparison is to discover performance characteristics of two single-loop based travel time estimation methods — the regression method and the CVL method — in estimating section travel times. The comparison is based on the I-880 data set. In this particular data set, the only two choices of benchmarks for section travel time comparison are both flawed. The probe vehicle travel times are sparse, subject to sampling biases and occasional vagaries of the drivers. On the other hand, we do not have much information about the accuracy of the other benchmark — travel times from double-loop speeds, except that its accuracy varies with traffic conditions too. Keeping the above points in mind, we propose a point-by-point comparison in which travel times associated with each probe vehicle run are computed using both of the single-loop estimation methods and the other benchmark — the double-loop speeds. We then focus on studying the *estimation errors* relative to both of the benchmarks (differences between estimated and benchmark travel times) in relation to the trip inception time, an explanatory variable which we take to be indicative of historical traffic conditions. The comparison remains exploratory in nature as a combined result of restrictions imposed by available data and limited knowledge of the complex traffic dynamics.

We introduce notation for section travel times involved in the comparison: $INCT$ denotes the trip inception time; $TT$ with suitable subscript represents the section travel time from the respective source; namely, $TT_{probe}$ for probe vehicle travel times, $TT_{dbl}$ for double loop travel times, $TT_{reg}$ for regression method estimated travel times and $TT_{CVL}$ for those using the *common vehicle length* method.

Preparation of travel times data and some preliminary analysis of the two benchmarks are described in §4.3.1. After that, we present descriptive statistics of estimation errors in §4.3.2. The last part of the comparison is a test which compares the percentage of coverage for estimation intervals based on each of the estimation method. This is discussed in §4.3.3.

### 4.3.1 Setup and preliminary analysis of benchmarks

The task of compiling travel times is complicated by data quality problems. For now, we focus on data collected in mornings only. Frequently detailed results are only presented for the north-bound traffic. All loop-based results are computed using data for lane 3.

We describe procedures to construct hypothetical vehicle trajectories in resemblance to the observed probe vehicle trajectories in §4.1. Corresponding hypothetical trajectories are computed for each probe vehicle covering the full stretch of the test site using all three loop-based methods (double-loop speeds, the regression method and the CVL method, namely.). Occasionally, we missed data of one or more types (flow/occupancy/speed) from some of the loops. For this reason, we were not able to estimate the hypothetical tracks for the entire trip for a few days. For consistency, we restricted the comparison to travel times for the freeway stretch for which trajectories from all four methods are computable. For the north bound/AM data, the stretch is about 5 mile long from loop 5 to loop 3. For the south bound/AM, it is around 5.4 miles from loop 3 to loop 5[6]. The section travel times we discuss later are for these freeway sections unless otherwise specified.

Figure 13 and Figure 14 plot probe vehicle travel times for the north/south bound trips for all 20 days. Note how the plots for north bound trips on 3/10/1993 and south bound on 3/3/1993 stand out. There were severe incidents on those two days and the traffic patterns were vastly different in nature from the other days. For this reason, we excluded those two days from the analysis later. One can also pick up occasional stray probe vehicles from these plots. For example, the first observation on north bound 3/17/1993 and the largest one on south bound 3/11/93 look more like results of individual decisions than prevailing traffic conditions. The effect of these stray observations is twofold. First it somewhat compromises the role of probe vehicle data as source of benchmark travel time information. On the other hand, this is not entirely negative since it mirrors real-world situations. The probe data can be used as test data to check performances of estimation intervals (see §4.3.3) exactly for this reason.

Figure 13 and Figure 14 also show that probe vehicle travel times on a single day generally forms a rather smooth line. This indicates that there is strong correlation between consecutive probe vehicle runs.

We are interested in how the section travel time is related to when the trip is started, i.e. *the trip inception time* over many days. We devise a type of plot to depict the relationship. Some examples are shown in Figure 15. This type of plot is constructed in the following way. For a certain collection of travel times on many days (for example, the top left panel plots $TT_{probe}$ for northbound trips on mornings.), we first divide the range of trip inception times into small non-overlapping intervals. The travel times data are then binned according

[6] cf. Figure 1. The loop detectors are not numbered sequentially.

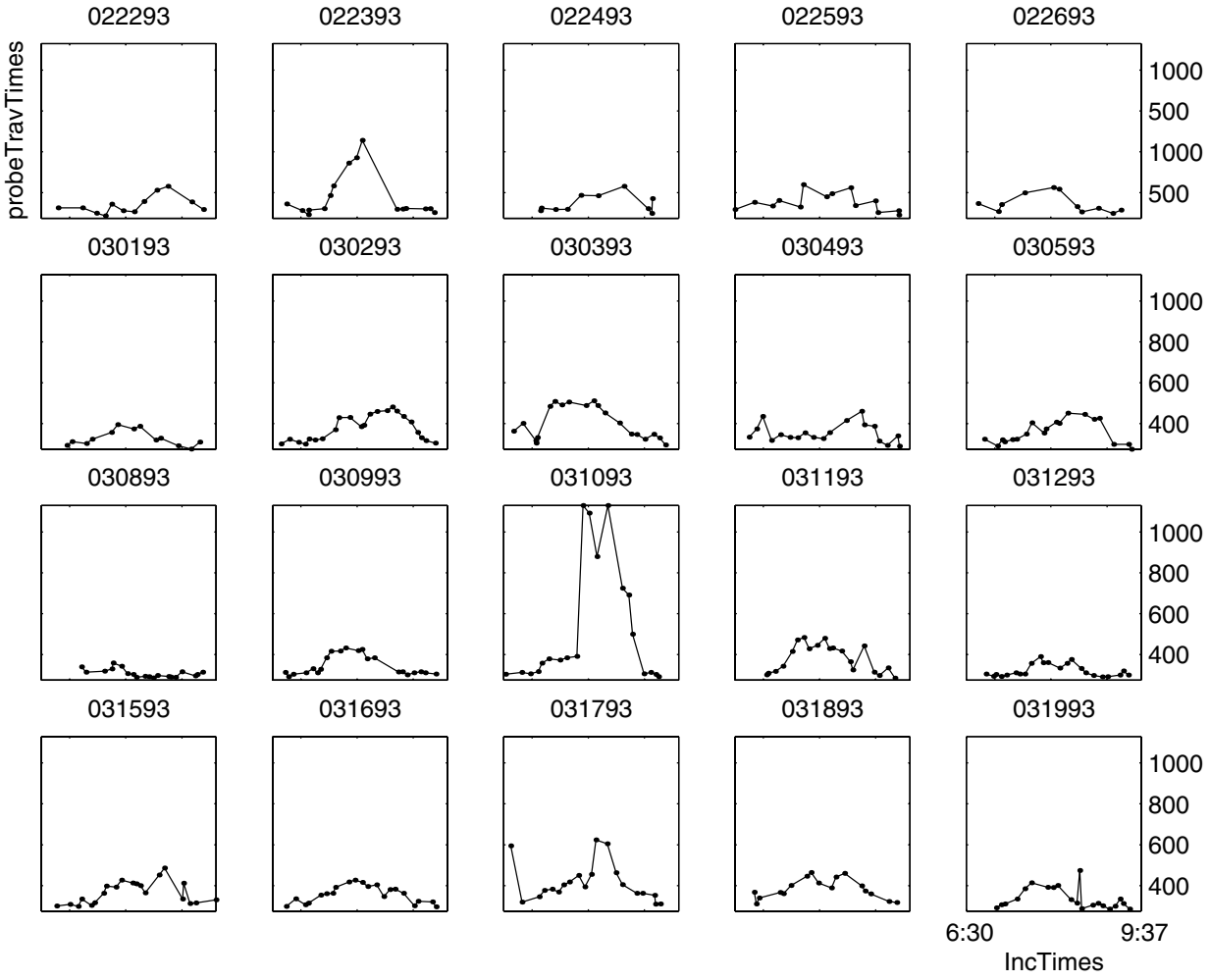


Figure 13: North bound probe vehicle runs on mornings of 20 week days.

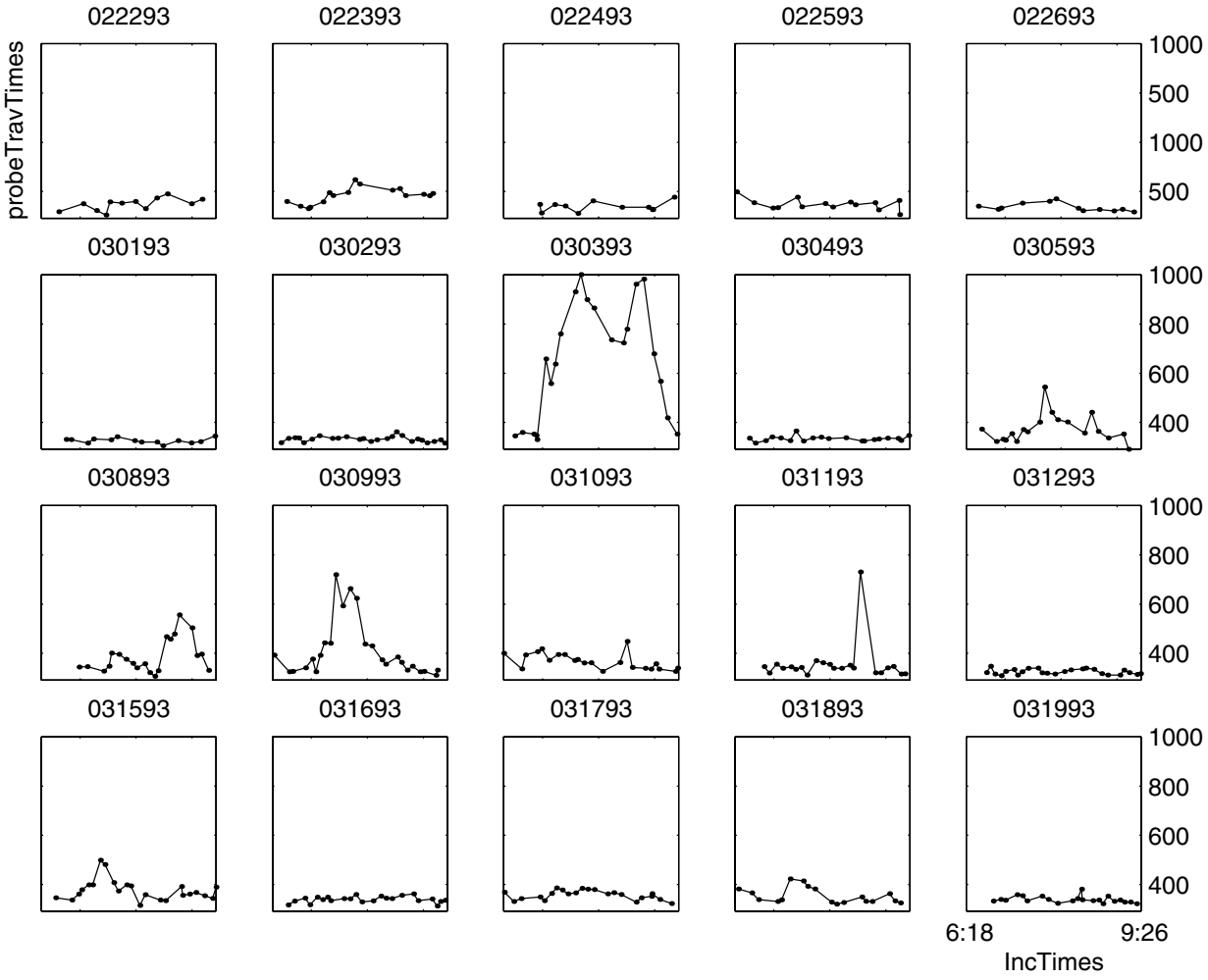


Figure 14: South bound probe vehicle runs on mornings of 20 week days.

to their trip inception times. We then plot summary statistics of data in each bin against the center of the bin. Specifically, we plot the median and the first and third quantiles. We can get a sense of how the center and the median of each bin change with $INCT$ from this type of plot. Similar plots can also be constructed for differences between travel times from different sources as in the two bottom panels in Figure 15.

In Figure 15, the thick line is the median, the two thinner ragged lines are the first and third quantiles respectively. The ends of the vertical lines depict the minimum and maximum travel times. The number at the lower end of the lines are the number of observations in each bin. Note that these numbers directly affect the accuracy of the summary statistics of the corresponding bin. When this number is relatively small, a few extreme observed/estimated travel times can dominate the summary statistics. One should keep this in mind when observing this type of plot.

Figure 15 plots benchmark travel times and their differences for both directions. The two panels in the top row are for $TT_{probe}$. We can observe distinct traffic patterns in both directions from them. The north bound/AM trip is far more congested than the opposite direction with serious congestion in the period from 7AM to 9AM. For both directions, we are not surprised to find that variability in travel times is larger during heavy traffic. Panels in the middle row show plots for $TT_{dbl}$. They reconfirm the traffic patterns observed in the probe vehicle travel times plots. The transition of $TT_{dbl}$ with time appears to be somewhat smoother than $TT_{probe}$. This is easier to observe in the south bound plots and in late morning portion of the north bound ones. Variability of $TT_{dbl}$ also appears to be smaller than that of $TT_{probe}$ for the south bound traffic. The distinction between the benchmarks is more clearly observed in the bottom panels of Figure 15 which plot $TT_{probe} - TT_{dbl}$. For both directions, the median probe vehicle travel time is almost consistently larger than that of the double loop travel times. This concurs with our previous conclusion indicating that probe drivers tend to be conservative. Looking at the two bottom panels individually, variability in the difference does not appear to change obviously with the trip inception time. However, the south bound plot displays slightly lesser variability in $TT_{probe} - TT_{dbl}$, indicating that variability of the difference is also loosely related to traffic conditions. Summary statistics for $TT_{probe} - TT_{dbl}$ is tabulated in Table 1. We notice that the difference between two benchmarks is more significant for the south bound trips. This is understandable considering that conservativeness of probe drivers is more distinguishable under lighter traffic which has more presence in the south bound traffic.

Summarizing the previous discussion, we draw the following conclusions:

1. The probe vehicle travel times is conservative. The bias is larger under free flow conditions.

2. The individual probe vehicle runs are not mutually independent. Not only consecutive runs are strongly correlated, but also runs departing around roughly the same time on

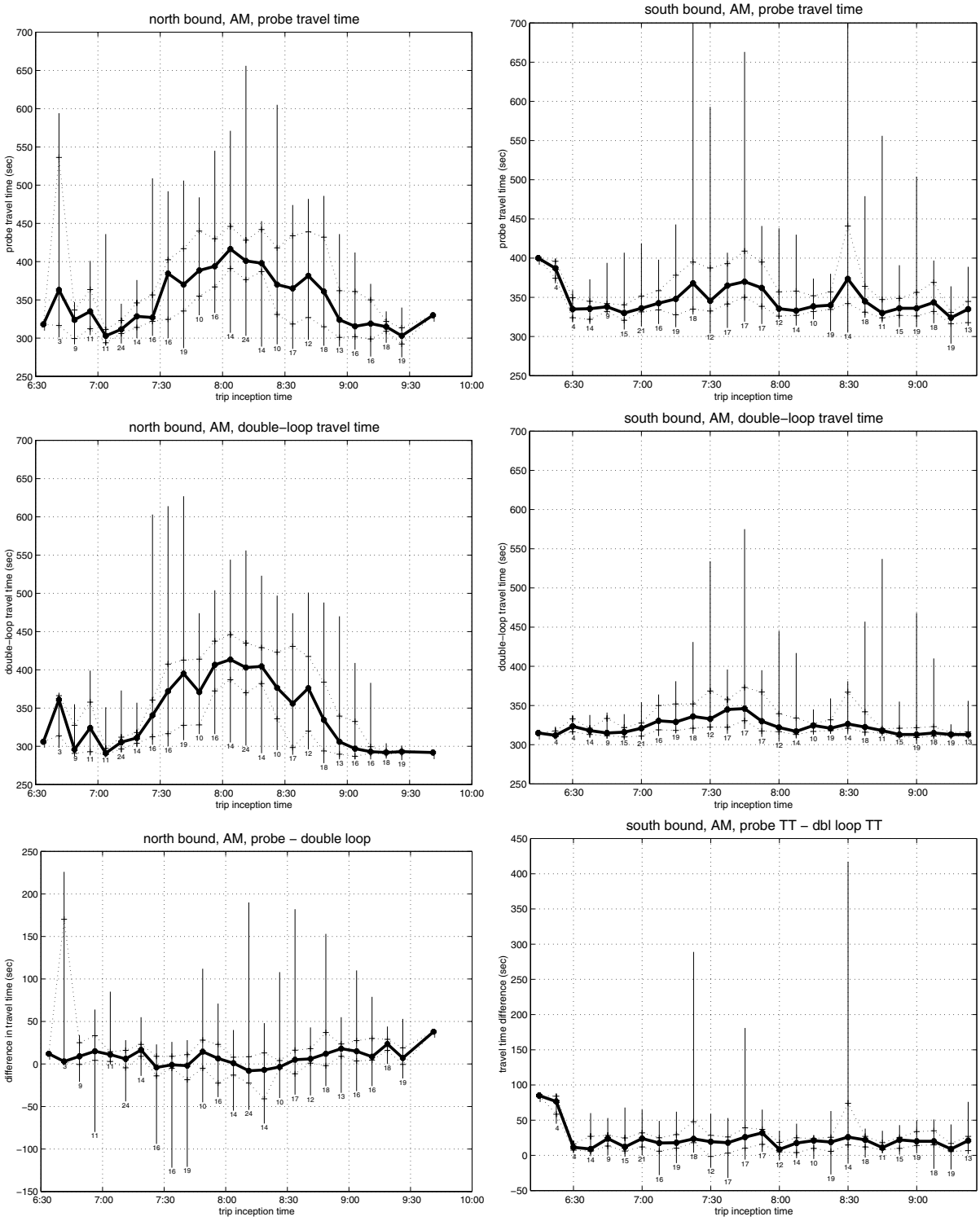


Figure 15: Benchmark travel times. Left panels are travel times from probe vehicles, double loop speeds and their differences for north bound runs. Right panels for south bound. The thick line is the median. Two thinner ragged lines are the first and third quantiles respectively. Vertical lines depict maximums and minimums. The number at the lower end of the vertical lines are the number of observations used to compute the summary statistics.

| | mean | SD | median | interquartile range |
|---|---|---|---|---|
| North/AM | 8.5 | 34.2 | 7 | 27 |
| South/AM | 23.5 | 33.1 | 19 | 21 |

Table 1: Summary statistics for $TT_{probe} - TT_{dbl}$. The unit for all entries is second. The second column is the standard deviation. The last column is the interquartile range which is the difference between the third and first quartiles.

different days are also related.

3. For the I-880 data set, the south bound traffic in the morning displays less activity than the opposite direction.

4. Variability in benchmarks are influenced by traffic conditions. In particular, $TT_{dbl}$ demonstrate smaller variation during free flow conditions compared to $TT_{probe}$. It also implies that the comparison results later can not be taken out of the context determined by the data used.

#### 4.3.2 Descriptive statistics of estimation errors

We now compare the estimation errors of the regression method and the CVL method relative to the double loop travel time. The motivation for using the double loop travel time instead of the probe vehicle travel time as the basis for this part of the comparison is that intuitively $TT_{dbl}$ is less prone to unpredictable randomness introduced by active involvement in the data collection process. The observations in §4.3.1 also suggests that $TT_{dbl}$ is more stable than $TT_{probe}$. We also use $TT_{probe} - TT_{dbl}$ as reference sometimes.

Figure 16 plots $TT_{reg} - TT_{dbl}$ and $TT_{CVL} - TT_{dbl}$ with respect to the trip inception time $INCT$ in the same way that was used in Figure 15. Panels in the middle row of Figure 15 are reproduced in the top row here to ease cross-referencing traffic conditions. Again plots for the northbound trips are in the left column, those for the south bound in the right.

We notice that the plots for the estimation error of the regression method display similar features for both of the directions. Namely, the median error holds roughly constant throughout the morning; the wider range of variation in errors (indicated by longer vertical lines in the plots) are associated with wider range of variation in $TT_{dbl}$. It is interesting to see that the regression method appears to successfully correct the bias associated with the CVL method, taking account of the close relationship between the two methods.

We can not say the same thing for the error of the CVL method displayed in the two bottom panels. The plots reveal significant and different trends with $INCT$ for the two directions. In both directions, the overall trend for the CVL error is to increase as $INCT$ approaches late morning. For the north bound, the median CVL error jumps from negligible to 40 seconds around 8AM which is roughly when the traffic reaches peak of congestion historically.

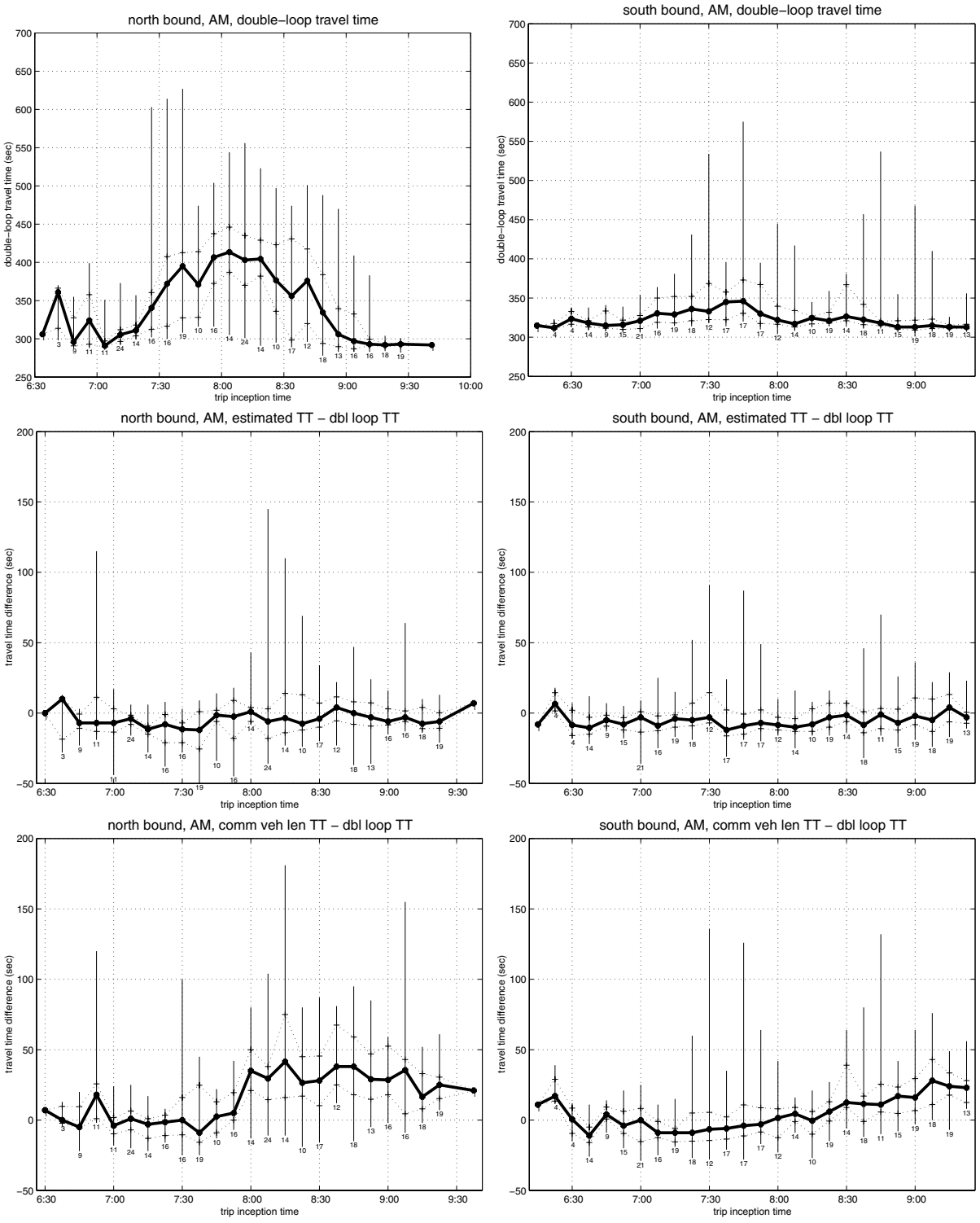


Figure 16: Estimation errors relative to the double loop travel times. Left panels are for the north bound trips, right panel south bound. In each column of plots, the first one is for $TT_{dbl}$ replicating the corresponding middle panel in Figure 15; the second and third are for estimation errors of $TT_{reg}$ and $TT_{CVL}$ relative to $TT_{dbl}$. The thick line is the median. Two thinner ragged lines are the first and third quantiles respectively. Vertical lines depict maximums and minimums. The number at the lower end of the vertical lines are the number of observations used to compute the summary statistics.

For the south bound traffic, the median error starts to increase shortly after 7AM and the trend is sustained until the end of the available data. By then, the median error reaches 25 seconds. Note that the south bound traffic heads toward minor congestion also around 7AM, as illustrated in the top right panel.

Table 2 presents summary statistics for all estimation errors relative to $TT_{dbl}$. The errors in the south bound direction demonstrate lesser variability than their north bound counterpart. This can also be inferred from the plots in Figure 16.

| | | mean | SD | median | interquartile range |
|---|---|---|---|---|---|
| North/AM | $TT_{reg} - TT_{dbl}$ | -3.4 | 18.9 | -6 | 16 |
| | $TT_{CVL} - TT_{dbl}$ | 19.5 | 28.5 | 15 | 33 |
| | $TT_{probe} - TT_{dbl}$ | 8.5 | 34.2 | 7 | 27 |
| South/AM | $TT_{reg} - TT_{dbl}$ | -2.5 | 15.2 | -5 | 13 |
| | $TT_{CVL} - TT_{dbl}$ | 7.5 | 23.1 | 4 | 24 |
| | $TT_{probe} - TT_{dbl}$ | 23.5 | 33.1 | 19 | 21 |

Table 2: Summary statistics for estimation errors of the regression method and the CVL method relative to the double loop travel time. The unit for all entries is second. The second column is the standard deviation. The last column is the interquartile range which is the difference between the third and first quartiles.

The descriptive statistics of estimation errors for the two single-loop based methods manifests that the performance of both estimators are tied in with traffic conditions. For the regression method, larger variability in estimation error can be traced back to heavier traffic conditions. The effect of traffic conditions are the CVL method is harder to summarize. We attempt to establish the missing link in §4.4. We find that the regression method travel time corresponds to the double loop benchmark very well, with error margin well under 10% most of the time. The CVL method also works well under free flow and light congestion.

#### 4.3.3 A test

For further investigation, we propose a mock test which attempts to assess the loss of applying either of the single-loop based estimation method to get section travel times in a real-world situation. The test is designed as follows: For each value of trip inception time $INCT$, we estimate the section travel time to be $\widehat{TT}$. We construct an estimation interval based on $\widehat{TT}$ with 10% margin of error attached, i.e. the interval is $[0.9\widehat{TT}, 1.1\widehat{TT}]$. The index of goodness $p$ is the percentage that the interval covers the observed probe vehicle travel time $TT_{probe}$ associated with the corresponding $INCT$.

The test originates from the idea that a reasonable measure of usability of a section travel time estimation method in many practical application is the probability that the travel time

for a random selected vehicle is in close neighborhood of the estimator. The design of the test implies that the probe vehicle is treated as randomly selected from the vehicle population departing the trip start at $INCT$, which is certainly not the case. However, since the probe vehicle runs were indeed conducted independent of loop detector data, the test is still applicable.

As already mentioned, we use a very crude estimation interval which centers at the section travel time estimate with 10% margin of error. The only justification for using this interval is that accuracy within 10% is satisfactory for many applications (including dynamic route planning, automatic incident detection etc.) requiring real-time section travel time information, as suggested in [12]. Alternative intervals could be used as well.

We apply the test to all three loop-based methods. The result for the double loop travel time is used as reference. Figure 17 shows the percentage of coverage $p$ against $INCT$ for the north and south bound data. Overall $p$ for $TT_{dbl}$ stays near or above 0.8 in both directions and is better than either $TT_{reg}$ or $TT_{CVL}$ most of the time. For the north bound, $TT_{reg}$ works slightly better during the time frame of historical congestion. In the right panel of Figure 17, the percentage of coverage for $TT_{CVL}$ embarks on a increasing trend right after 7AM until it dominates that of $TT_{reg}$ and even $TT_{dbl}$. The time frame again reminds us of historical congestion time period for the south bound. Summarily, the plots deliver conflicting messages about the two single-loop based methods, with $TT_{reg}$ preferable for the north bound traffic and $TT_{CVL}$ for the south. Comparing the mean percentage of coverage in either direction (presented in Table 3) suggests the same conclusion.

It should be noted that the accuracy of the percentage of coverage plotted in Figure 17 is tied in with the number of the observations used to compute them, which is in turn determined by how many probe vehicle runs originated in the corresponding time slots. In particular, the first point in both of the panels in Figure 17 are based on 2 and 5 points respectively. Those values for $p$ are unreliable since a few erratic probe runs had huge influence on the resulted percentage of coverage. We presented the number of observations for each time slot as the small typeface numbers along the ragged lines in Figure 17. It is worthwhile to pay attention to them when examining the figure.

The conclusion is contrary to our expectation, since the descriptive statistics of errors seem to suggest that $TT_{reg}$ is no worse than $TT_{CVL}$ even in the south bound direction. A plausible explanation may be that the upward trend of estimation error for $TT_{CVL}$ (see the bottom left panel in Figure 16) works in favor of the CVL method since $TT_{probe}$ tends to bias in the same direction relative to $TT_{dbl}$ (See the bottom panels in Figure 15 which shows that $TT_{probe}$ is generally larger than $TT_{dbl}$. The bias is more clearly observed in the bottom right panel showing the south bound traffic.). It will be interesting to look at results using adjusted probe vehicle travel times as test data where the adjustment is made to diminish the inherent bias of $TT_{probe}$. We delay further discussion of the results to §4.4.

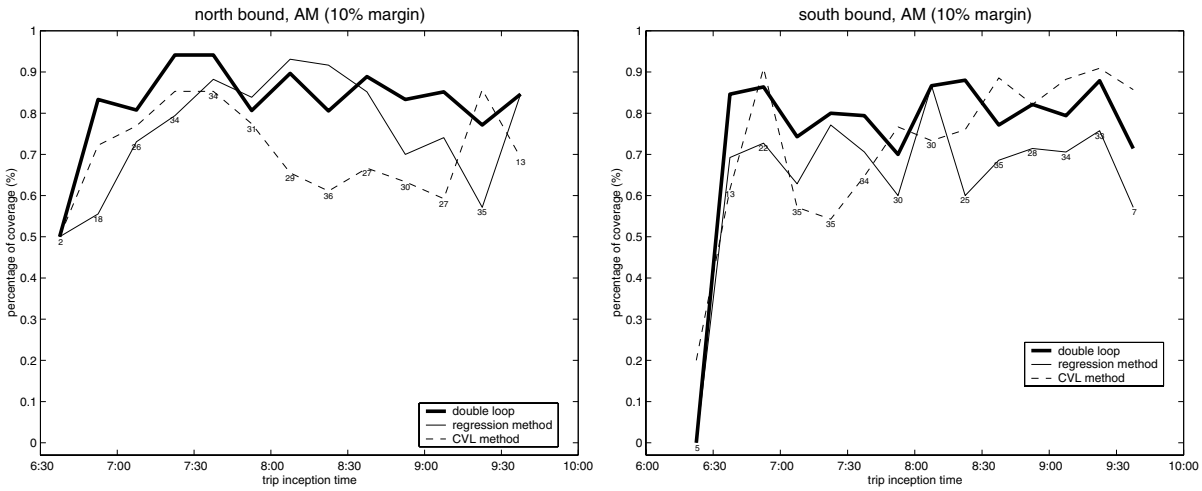


Figure 17: $p$: the percentage of coverage versus the trip inception time. Left panel is for the north bound. Right the south. The numbers in the plots are the number of observations used to compute $p$ for results at the corresponding trip inception time.

| $p(\%)$ | $TT_{dbl}$ | $TT_{reg}$ | $TT_{CVL}$ |
|---|---|---|---|
| North | 85 | 78 | 73 |
| South | 80 | 69 | 75 |

Table 3: The mean percentage of coverage over all data.

## 4.4 Discussion

From the various analyses in this section, it is our conclusion that the regression method work satisfactorily as section travel time estimator in the particular setting of the I-880 data. Its performance during congestion is less reliable than that under lighter traffic. Its error relative to the double loop travel time is constant and small on average. The results for the CVL method are less predictable with complex relation with traffic conditions and maybe other factors. We have more confidence to extend the regression method to other roadway network than the CVL method.

The comparison analysis above produced some intriguing results. We attempt to suggest possible explanation for some of them in §4.4.1. In §4.4.2, we look at how estimation errors are distributed over links. The results presented later are usually only based on data on a single day. The assertions made upon them can only be taken as heuristics rather than rigorous arguments.

### 4.4.1 The constant $1/g$?

We observe some peculiarities of the CVL estimator in the comparison earlier; particularly, the distinct trends displayed by its error relative to $TT_{dbl}$ (see §4.3.2) and the percentage of coverage of probe vehicle travel times by estimation intervals in §4.3.3. Somehow, these results can be loosely tied to the prevailing historical traffic conditions. It is plausible that the traffic condition induces drifts in the *common vehicle length constant* $1/g$ and thus contribute to the results aforementioned. An alternative (or complementary) explanation of drifting in $1/g$ is that vehicle population may change with time.

We compute the "true" *common vehicle length constant* by plugging the measured double loop speeds into equation (1) and solve for $1/g$. Denote the vehicle length constant computed this way by $L_{dbl}$. That is,

$$L_{dbl} = \text{double-loop speed} \times \frac{\text{occupancy}}{\text{flow}} \tag{8}$$

Let $L - 0 = 1/g_0 = 22.611$ ft. be the constant we used in the analysis. Note that the estimated CVL speed is proportional to the constant $L_0$, from (1). Thus, $TT_{cvl}$ underestimates $TT_{dbl}$ when $L_0$ is greater than the "true" constant $L_{dbl}$. The results are plotted in Figure 18 together with some reproduction of related results for the CVL estimator. The horizontal lines in the plots indicate the $L_0 = 22.611$ ft.. Note that the thin lines (which plot $L_{dbl}$ were computed using one half day of data only, while the thick lines are based on data from many days.

In the left panel of Figure 18, we compare the trend of $L_{dbl}$ with that of the median CVL error relative to $TT_{dbl}$ for the north bound data. The CVL estimator starts to overestimate considerably when $L_{dbl}$ goes above the value used in the estimation.

For the south bound data displayed in the right panel, we plot $L_{dbl}$ with $p$ (the percentage of coverage of probe travel times) for the CVL based interval. We observe that the two ragged lines do not start to increase at the same time: $L_{dbl}$ starts to increase between 7:30 and 8:00, while $p$ begins the increasing trend about half an hour earlier. In interpreting the results, recall the following: first, $TT_{probe}$ tends to bias upward relative to $TT_{dbl}$; second, the bias is greater during free flow. The result for the south bound data is then easier to understand. When the traffic heads into moderate congestion at around 7:00 (see the top right panel of Figure 16), $p$ starts to increase since the test data $TT_{probe}$ is less biased relative to $TT_{dbl}$ during congestion, and thus increasingly more likely to be in the neighborhood of the CVL estimate.

Also note that the range of $L_{dbl}$ is roughly within 20% for either direction. This suggests that with careful calibration of $1/g$, the CVL method is capable of delivering reasonable travel time estimates for many practical applications. Calibrating $1/g$ separately for each direction and lane may be necessary.

The above results give tentative explanation of the peculiar results we saw before. In particular, the drift in $1/g$ is partly accountable. However, the drift in $1/g$ itself is the synergy effect of traffic conditions, vehicle population and many other factors. Further investigation is needed to understand the complex interplay between all related factors.

### 4.4.2 Distribution of estimation errors over links

As have been mentioned, the section travel time induced from each source are essentially sum of link travel times. It is interesting to investigate how each constituent link travel time estimate contributes to the total section travel time error; in particular, whether the error in the resulted section travel time estimate is dispersed over all links or there exist one or more ill-conditioned links which consistently dominate the resulted error.

The question can be answered by plotting errors in estimated link travel times versus link numbers as shown in Figure 19. We observe that errors in link travel times are localized for both the regression method and the CVL method. For both of the methods, major part of the estimation error in the resulted section travel times come from link 2 (loop 15 to 17) through link 4 (loop 4 to loop 13)[7]. The correspondence between the two methods is not surprising considering the close relationship between them.

We offer some heuristics as to why both of the methods tend to behave badly around the area encompassing link 2 and link 4. The first notable fact is that link 2 and link 4 more or less border the region where recurrent congestion occurs in the test site (see examples in [13]). It is also worth of mentioning that those two links happen to be the two longest links at 3360 feet and 3900 feet respectively, more than doubling the average length of the remaining links. Consulting the map (Figure 1), we also observe that the freeway goes from

[7] Data for loop 12 is missing for this day.

5 lanes to 4 lanes in link 2, and that the exit to west bound SR-92 is right before the end of link 4[8]. The above points all suggest comparatively more complex nature of the traffic flow around the problematic region, which could be the reason that both of the single-loop based methods fail.

[8]Westbound SR-92 leads to the San Mateo bridge which is an important transbay channel.

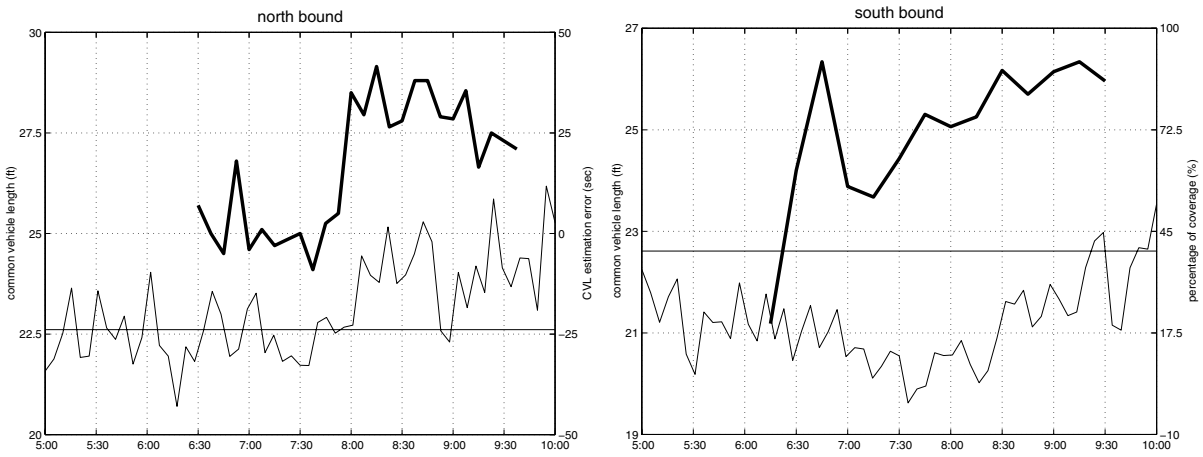


Figure 18: Variability in the *common vehicle length* constant $1/g$. Left panel is for north bound traffic, right panel for the south bound. In both panels, the thin ragged lines plot $L_{dbl}$ in (8), the horizontal line is the default value $L_0 = 22.611\ ft.$ used in the analysis. $L_{dbl}$ were computed using data collected in the morning of 3/11/1993. The thick line in each panel is relative to the top-right coordinate system. In the left panel, it is the median CVL estimation error relative to $TT_{dbl}$ (replication of the thick line in the bottom left panel in Figure 16). In the right panel it plots the percentage of coverage for the CVL estimator (replication of the dash line in the right panel of Figure 17).

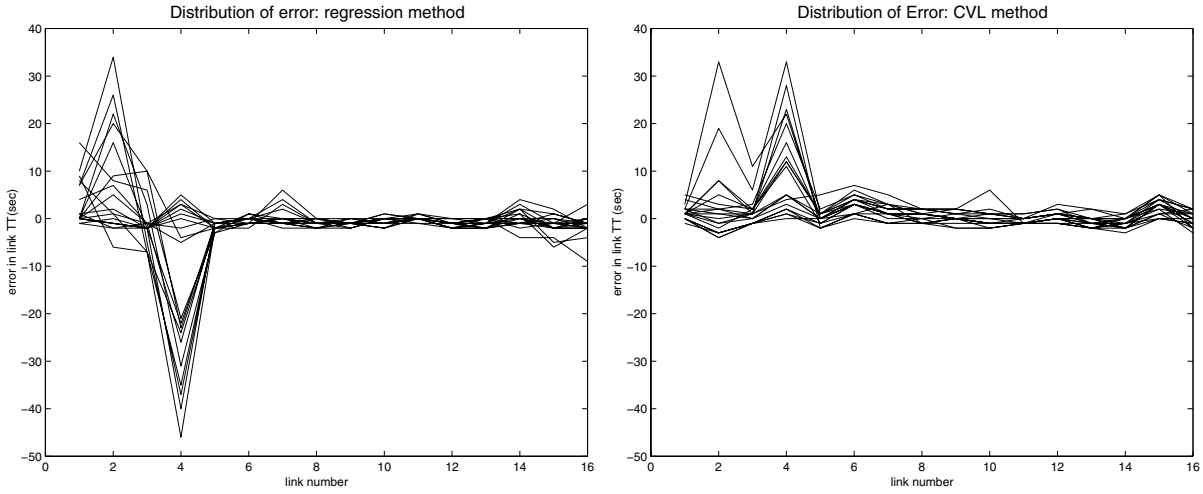


Figure 19: Estimation error in link travel times versus link number. Left panel is for the regression method, right the CVL method. The X-axis the link number, the Y-axis is the corresponding estimation error for the corresponding link relative to the double loop speeds induced travel time. Link travel times are computed in the procedure outlined in §4.1.2. The plots are based on north bound tracks for the morning of 3/11/1993. Each of the line in the plots corresponds to one hypothetical track associated with a particular probe run.

# 5 Visualization

A thorough comparison of travel time estimation methods necessarily involves examining results under a variety of traffic conditions collectively to discover patterns in the behavior of the estimators. The ability to easily convey relevant information is essential. Here we propose a visualization approach to address the issue.

In [13] we presented a visualization technique for loop detector data. We pointed out that the technique is not restricted to measured data only. Here we present examples where the technique is applied to estimated speeds and functions of their errors relative to the double loop measured ones. We illustrate, via one set of examples based on one half day, how the technique can be used in the comparison scenario. As one will observe later, the visualization highlights previously unclear or unobservable patterns in the behavior of the estimation methods.

## 5.1 The technique

In a nutshell, the technique is applicable to data defined on an uneven grid in the space-time plane. Let $U(x_i, t_j)$ represent the data value associated with time $t_j$ and location $x_i$. $U(x_i, t_j)$ may be unavailable for some $(x_i, t_j)$ pairs. The technique visualizes measurement values by mapping $U(x_i, t_j)$ to a colored pixel at $(x_i, t_j)$, where the pixel color encodes data value. A color image is generated after interpolation and/or smoothing. Color pattern of the image is the visual cue for changes in $U(x_i, t_j)$.

The technique is readily applicable to speeds measured by the double-loop detectors or estimated by the CVL method, which are point quantities. For the regression method, we impute the speed derived from the link travel time estimate to be located at the mid-point of the corresponding link.

In the examples to be presented, we produced speed estimates using each of the two single-loop based methods every two minutes for the duration of available data. The double-loop speed measurements were aggregated to the same level. The (aggregated) data were then smoothed along the time axis and interpolated along the space axis before visualization. The interpolation step handles missing data problems and the technicality that the regression method estimated speeds were defined on different locations than the other two sources of speeds.

In addition to applying the technique directly to estimated speeds, we display visualizations of functions of estimation errors which are the differences between estimated speeds and the double-loop speeds. Visualizations of selected functions of estimation errors further reveal hidden patterns characteristic of the corresponding estimator. We find that visualizations of

absolute logarithm of relative errors and the sign of the errors are particularly informative and suitable to gray scale presentations.

## 5.2 An example

Here we illustrate the application of the visualization technique in comparing travel time estimation methods via a set of examples. These examples are based on the north bound data from lane 3 collected on the morning of 3/11/1993 in the I-880 data set. There was significant non-accident induced congestion during the peak traffic hours on that day. We think the set of examples can be taken as typical of a busy urban freeway to some extent. Many of the findings also apply to other examples we have examined.

The top left panel in Figure 20 shows the visualization for the double-loop speeds. Congestion registers as dark blobs on the image. The most conspicuous congestion happened near the southern end of the test site around 8AM. Note that the most significant congestion is shown as a roughly tip-down triangular shape in the image, manifesting that the congestion propagated against traffic direction and started to dissipate from the upstream. As mentioned in [13], the triangular shape is common for many congestions for the I-880 data set.

The two bottom panels in Figure 20 exhibit visualizations for estimated speeds from the regression method and the CVL method. In both of the panels, we find dark blobs in rough correspondence with those in the double-loop visualization, signaling that both of the estimation procedure captures prominent changes in traffic conditions. Closer examination reveals that in the image for the regression method, the contour of the dark region in the bottom takes a more rectangular shape compared with the other images. The dark region also appears to be somewhat larger. These observations indicates that the regression method tends to underestimate speeds (hence overestimate travel times) when the prevailing traffic regime changes from free-flow congestion or vice versa. We can also observe obscure horizontal stripes of various shades in the bottom panels, implying that there are system biases related to locations for both of the methods.

For a closer look, we visualize functions of the estimation errors of each of the two methods, in particular the absolute values of the difference in logarithm between estimator and benchmark (absolute logarithm relative error) and the sign of the estimation error. We visualize these two functions of estimation errors instead of the errors themselves mainly because of the limitation imposed by gray scale images which render conspicuous features in the visualizations for the estimation errors obscure.

Figure 21 displays visualizations for the absolute logarithm of the relative estimation errors. In the two images, lighter shades signal smaller estimation errors relative to the double-loop benchmark and the darkest shade approximately corresponds to error margin over 10%. We notice that the striping effect observable in visualizations of the estimated speeds are even

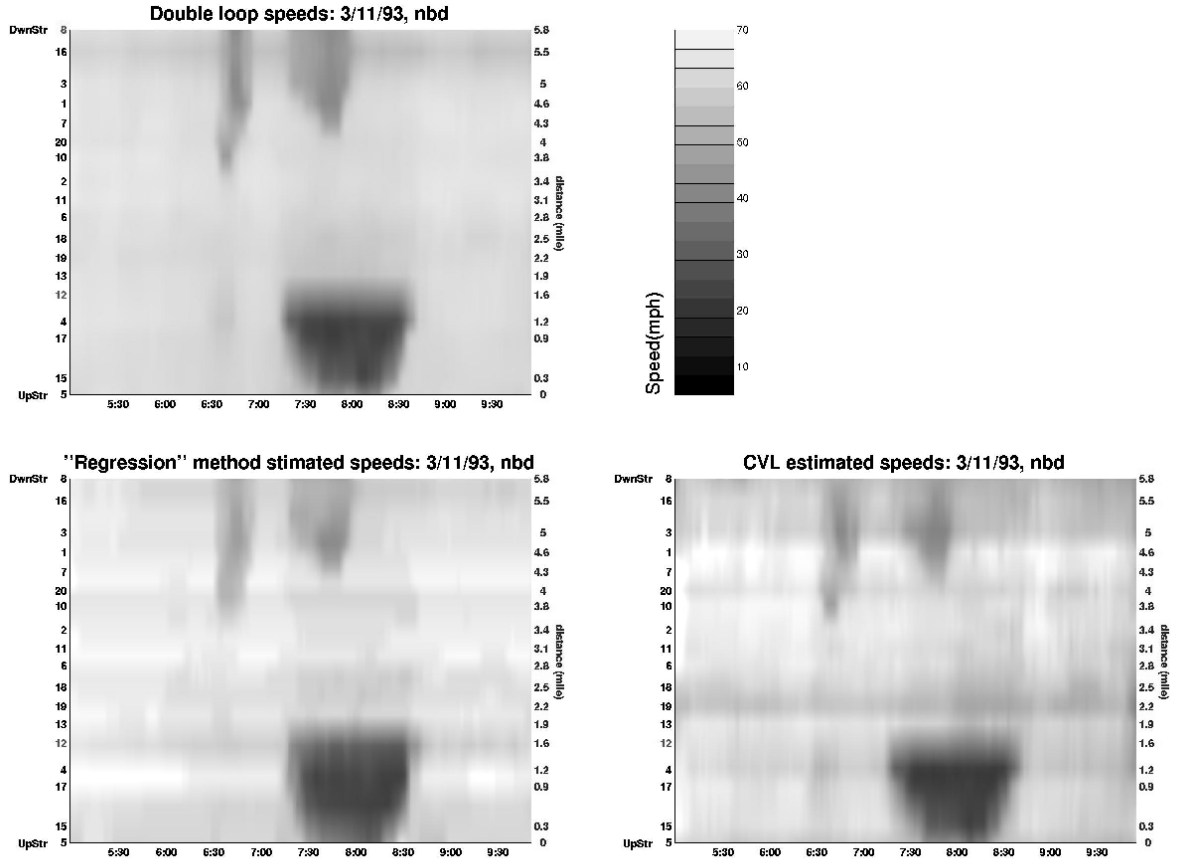


Figure 20: Visualization of estimated speeds. Top left panel: double-loop speeds; bottom left: the regression method estimated speeds; bottom right: CVL estimated speeds. Traffic goes upward. Time period is 5AM to 10AM. The left hand side Y-axis labels show the locations of loop detector. The right hand side ones are the distances in miles relative to the start of the trip.

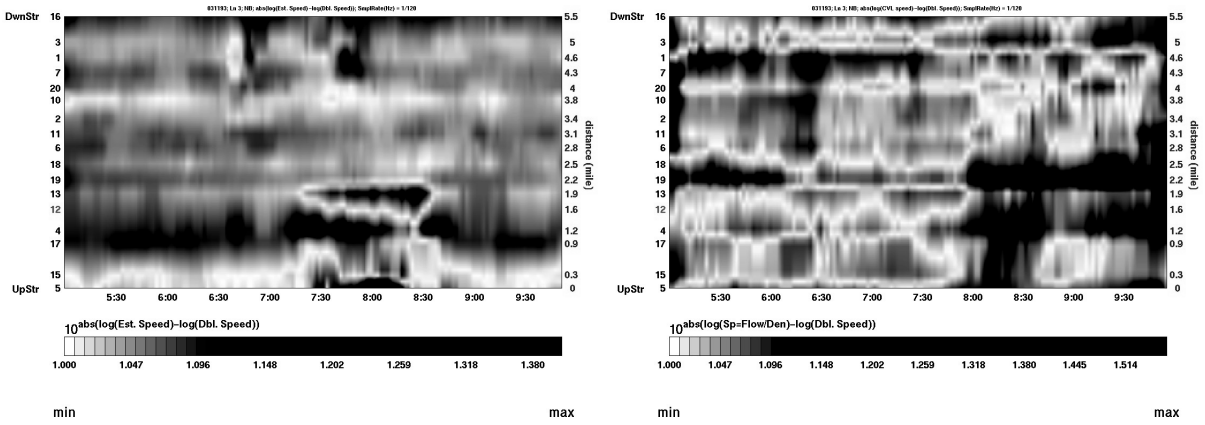


Figure 21: Absolute logarithm relative errors. Left panel is for the regression method, right the CVL method.

more prominent here. The locations of the horizontal stripes in the two panels do not line up exactly, which could be the result that the regression method estimated speeds using a different grid. In the image for the regression method, the shape of the signature of the longest lasting congestion is still quite distinguishable as the area enclosed by clear-cut dark regions. We observe higher volatility in the image for the CVL method demonstrated by mingled shades. We also observe that the CVL method grows more prone to larger errors towards late morning at a few locations.

To observe the direction of biases associated with the estimation methods, we visualize the signs of estimation errors in Figure 22. The left panel emphasizes that the regression method is subject to significant underestimation of speeds (overestimation of travel times) around the region of considerable congestion. We also strengthen our previous observation about the CVL method by concluding that it has a tendency to overestimate speeds as time goes towards late morning. There are also a few identifiable locations where the CVL method consistently overestimates speeds.

## 5.3 Discussion

To summarize our observations with the visualization approach, we find in terms of travel time estimation: for the regression method, proneness to overestimate around the region of congestion; for the CVL method, a temporal trend to overestimate and relatively higher level of volatility compared to the regression method. For both of the methods, we observe location-specific biases.

Keeping the scope of the visualization comparison in mind, we relate our findings with observations previously made. The findings in this section reaffirms previous discussion that estimation errors at a few locations can account for most of the error in travel times over a

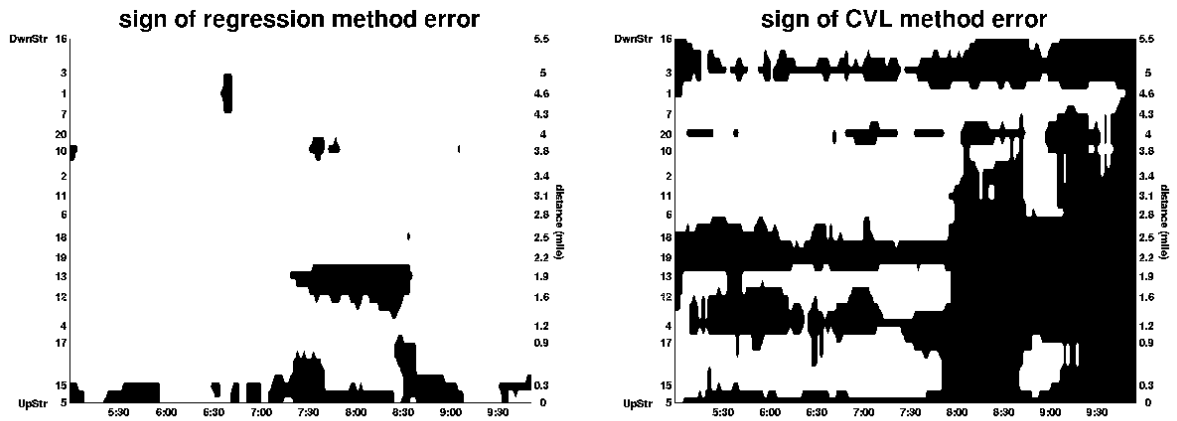


Figure 22: Signs of estimation errors. Left panel: the regression method; right:the CVL method. Black is associated with positive values while white signals negativity.

prolonged section of freeway. The localized temporal trend associated with the CVL method are consistent with previous conclusions on similar trend with the errors in section travel time estimates over many days, indicating that the observation made here on the basis of one day's data might not be isolated.

We observe location-specific biases for both of the methods throughout the duration of available data. It is a bit far-fetched to attribute this observation entirely to consistent variations in traffic conditions per se. It is tempting to conjecture instead, that it is the result of some location-specific features (such as pavement or loop configuration and calibration) which affects the *common vehicle length constant* $1/g$ and hence both of the estimation methods. For the regression method, inexact clock synchronization might also contribute to the striping effect.

For the regression method, it is possible that the pattern may be partially explained by the construction procedure of the regression method speeds estimates, aside from the inherent difficulty in handling congestion. Note that the speed estimated by the regression method uses information from both the upstream and downstream detector while the other two sources of speeds only use the upstream data. As have been observed, the duration of congestion is shorter towards the upstream direction. It is likely that speeds at upstream detectors are actually faster than downstream. Hence the regression method estimated speeds tends to bias upward around the region of congestion.

# 6 Conclusions

In this paper, we employed various techniques to conduct a thorough empirical comparison of two single-loop based travel time estimation methods — the regression method and the *common vehicle length* (CVL) method. We have demonstrated, via examples, that the regression method is more stable in performance and does not rely heavily on the choice of the common vehicle length constant. It is very accurate under free flow conditions. For the CVL method, its accuracy is completely tied in with the choice of the common vehicle length constant. Its performance varies considerably with the traffic condition, location and other identifiable factors.

As noted elsewhere ([5]), we have observed drift in the common vehicle length constant. Examples based on limited data suggest that changes in the prevailing traffic condition may be one of the contributing factors. Variations in this assumed constant are responsible for the intriguing patterns demonstrated by the CVL method. Although the regression method is also affected, the influence is much less and is negligible during free flow.

For many practical purposes, travel times estimated using either of the methods are good enough. In this case, the decision on which method to use is largely driven by consideration of infrastructure requirements. For example, the regression method is generally only applicable to high resolution (1 second resolution is preferable) loop data. The CVL method does not have this constraint. On the other hand, applying the CVL method with a single imputed value for the common vehicle length constant could be problematic with system biases concentrated in some or all lanes for one direction. Detailed calibration is called upon to avoid this. The calibration procedure for the regression method is much more simpler.

Another advantage of the regression method is that it actually estimates the travel time distribution instead of a point estimate of the travel time. The travel time distribution itself might be of interest in some settings. More importantly, understanding the travel time distribution is directly linked to building models for travel time prediction.

We have also proposed some extensions to the regression method based on the same stochastic model. In particular, the B-spline approximation significantly reduces the number of parameters without the penalty of inferior performance. Examples have also shown that difficulties of the regression method in dealing with congestion may be caused by a overly simplified model. A new approach to improve the regression method could involve significant modification of the underlying model. This is also the direction of future research.

# References

[1] B. Coifman. Vehicle reidentification and travel time measurement in real-time on freeways using the existing loop detector infrastructure. *Transportation Research Record*, (1643):181–191, 1998.

[2] B. Coifman, D. Beymer, P. MeLaughlan, and J. Malik. A real-time computer vision for vehicle tracking and traffic surveillance. *Trasnportation Research: Part C*, 6(4):271–288, 1998.

[3] B. Coifman, X. Zhang, D. Liddy, and A. Skarbardonis. The Berkeley Highway Laboratory, building on the i-880 field experiment. submitted for publication, 1999.

[4] Carl de Boor. *A Practical Guide to Splines*. Springer-Verlag, 1978.

[5] F.L. Hall and B. N. Persaud. Evaluation of speed estimates made with single-detector data from freeway traffic management systems. *Transportation Research Record*, (1232):9–16, 1989.

[6] R. Kühne and S. Immes. Freeway control systems for using section-related traffic variable detection. In *Pacific Rim TransTech Conference Proceedings*, volume 1, pages 56–62. ASCE, 1993.

[7] K. Petty, P. Bickel, J. Jiang, M. Ostland, J. Rice, Y. Ritov, and F. Schoenberg. Accurate estimation of travel times from single-loop detectors. *Transportation Research, Part B: Methodological*, 32(1):1–17, 1998.

[8] K. Petty, 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: Emerging Technologies*, 4(3), 1996.

[9] S. Ritchie and C. Sun. Section related measures of traffic system performance: final report. Technical Report PATH Research Report UCB-ITS-PRR-98-33, Institute of Transportation Studies, University of California at Berkeley, 1998.

[10] S. Seki. Travel-time measurement and provision system using avi units. In *Proceedings of 2nd World Congress on Intelligent Transport Systems*, pages 50–55. VERTIS, 1995.

[11] A. Skarbardonis, H. Noeimi, K. Petty, D.Rydzewski, P.P. Varaiya, and H. Al-Deek. Freeway service patrols evaluation. Technical Report PATH Research Report UCB-ITS-PRR-95-5, Institute of Transportation Studies, University of California, Berkeley, 1994.

[12] M. Westerman and L.H. Immers. A method for determining real-time travel time on motorways. In *25th International Symposium on Automotive Technology and Automation*, pages 221–228, 1992.

[13] Xiaoyan Zhang and John Rice. Visualizing loop detector data.