Improved Speed Estimation from Single Loop Detectors with High Truck Flow

Benjamin Coifman, PhD
Associate Professor, Department of Civil, Environmental, and Geodetic Engineering
Associate Professor, Department of Electrical and Computer Engineering
The Ohio State University
Hitchcock Hall 470
2070 Neil Ave
Columbus, OH 43210
Coifman.1@OSU.edu
http://www.ceg.osu.edu/~coifman
614 292-4282

Srikanth Neelisetty
Graduate Research Assistant
Department of Civil, Environmental, and Geodetic Engineering
The Ohio State University
Columbus, OH 43210
neelisetty.1@osu.edu
Coifman, B. and Neelisetty, S.

**ABSTRACT**

Single loop detectors are the most common sensors employed by freeway traffic management agencies. The data are used for traffic management and traveler information. Single loop detectors can only measure flow and occupancy. Although speed is often the most useful metric, it can only be estimated at conventional single loop detectors. Typically this estimate comes from the quotient of flow and occupancy multiplied by the fixed, assumed average effective vehicle length. This conventional approach is limited because the actual average effective vehicle length will vary from sample to sample. Many researchers have proposed alternatives to address this problem and although many of the methods work well under normal conditions, there has been limited research into methods that yield reliable estimates under heavy truck traffic. Heavy truck flows may arise as a function of location or time of day, e.g., proximity to a trucking facility or early mornings when the number of passenger vehicles drops, respectively. This paper presents a new methodology to estimate speed from single loop detectors in conditions where trucks comprise a large percentage of the fleet. While the focus is on single loop detectors, the work is equally applicable to side-fire microwave radar detectors that emulate single loop detectors.

**Key words:**

Single loop detectors; speed estimation; trucks; traffic surveillance;
INTRODUCTION

Loop detectors are the most widely used sensors for freeway traffic surveillance and they are most commonly deployed as single loop detectors with one detector per lane. One can only estimate vehicle speed using a conventional single loop detector. Typically this estimate comes from the quotient of flow, q, and occupancy, occ, multiplied by the fixed, assumed average effective vehicle length, L. As will be discussed shortly, these conventional estimates can yield poor performance because the true but unknown average vehicle length can vary dramatically across samples. Since operating agencies often require accurate speed information, many avoid the speed estimation problem by deploying dual loop detectors, with a pair of loops in a lane separated by a known distance. Although more expensive than a single loop detector to install, one can measure vehicle speeds directly from the quotient of spacing and the difference in arrival times at the two detectors. The fact remains that single loop detectors are more common than dual loop detectors and there is a need to estimate speed as accurately as possible from these detectors. Furthermore, many emerging non-intrusive vehicle detection technologies emulate the operation of single loop detectors, e.g., most of the commercially available side-fire microwave radar detectors. While the focus of this paper on single loop detectors, this work is equally applicable to the non-intrusive vehicle detection technologies that emulate single loop detectors.

Several researchers have proposed alternatives to address the shortcomings of the conventional single loop detector speed estimation. Mikhalkin et al. (1972), Pushkar et al. (1994), Dailey (1999), Wang and Nihan (2000) and Coifman (2001) proposed new techniques to estimate mean speed using aggregate flow and occupancy data. Others have proposed statistical and analytical methods to estimate speed, e.g., Dailey (1999), Hellinga (2002). A few researchers have used individual vehicle actuations to improve estimates under mixed traffic conditions, e.g., Coifman et al (2003), Wang and Nihan (2003). Although many of the methods work well under normal conditions, there has been limited research into methods that yield reliable speed estimates under heavy truck traffic. At the crux of the speed estimation problem is the fact that the true individual vehicle effective lengths vary from 20 ft to 80 ft and thus, the true but unknown average vehicle length for a given sample will vary as well. Typically passenger vehicles dominate, and so the constant L is usually set near their effective length, on the order of 20 ft. As shown in Coifman (2001), the wide range of vehicle lengths skews the occupancy, and hence speed estimates, from one sample to the next. On most urban freeways trucks make up a small percentage of the passing vehicles (on the order of 5-15%). This study considers a detector with 40% long vehicles over the day and thus, a given sample may contain almost all trucks or almost no trucks due to random arrivals. Clearly a single assumed value of L should not work well for such a location. In general, heavy truck flows may
Coifman, B. and Neelisetty, S.

arise as a function of location or time of day, e.g., proximity to a trucking facility or early mornings when the number of passenger vehicles drops, respectively.

This paper seeks to improve the performance of speed estimation from the deployed single loop detector infrastructure in the presence of heavy truck flows. Our approach uses the improved speed estimates from Coifman et al (2003) in conjunction with basic traffic flow theory assumptions. The remainder of this paper is broken into four sections. The first section provides the necessary background for the work. The second section focuses on identifying distinct vehicles in a single loop detector's time series data for which one can accurately estimate speed: namely when one sees a sequence of the longest feasible vehicle followed by the shortest feasible vehicle or vice versa. The next section formalizes the process into an algorithm that accommodates the fact that many minutes may pass between observing such sequences and then this section evaluates the performance over a month of data. Finally, this paper closes with a discussion and conclusions.

**Methodological Background**

This study uses 24 hours of individual vehicle actuations sampled at 240 Hz from 17 dual loop detector stations on I-70 and I-71 in Columbus, Ohio. Most loop detector stations collect raw individual vehicle actuation data as used in this study; however, conventionally the controller software quickly aggregates the data to longer sample periods (typically ranging between 20 sec and 5 min) and then discards the raw data. The dual loop detectors provide both single loop data and measured vehicle speeds to verify the single loop detector speed estimates. To mimic a single loop detector, only the upstream loop data in a given dual loop detector is used to estimate speeds in that lane using the conventional method via Equation 1. Figure 1A shows the plot of speed estimated using Equation 1 with a sampling period of T=30 seconds at a typical detector. The estimates are very noisy, as evidenced by the scatter in the plot.

\[ \hat{v} = \frac{L \times q}{occ} \]  

(1)

Naturally, the bias in Equation 1 depends on the choice of L. At an individual vehicle level the true, but unknown, effective vehicle length is simply the product of its speed and on-time (i.e., the time the detector is occupied by the vehicle). The following correction process from Neelisetty (2004) and refined in Lee and Coifman (2012a) is used to remove the bias due to L. By definition, during free flow periods all vehicles should have speeds close to the posted speed limit, and over a 24 hr period, the vast majority of vehicles passing a freeway detector station should do so under free flow conditions. We find the mode on-time and calculate the individual vehicle speed as if this on-time came from a passenger
vehicle. Finally, we find the multiplicative correction factor that would adjust this speed to the posted speed limit (or any assumed free speed via some other means). This correction multiplicative factor is then applied to the original $L$ to reduce the bias in the speed estimates from the original choice of $L$. In this case yielding a corrected $L=24$ ft for the detector used in Figure 1A. Although not shown, this result was verified by examining the true individual effective lengths measured from the dual loop detector. Various modifications can be employed using engineering judgment, e.g., a few locations may experience many hours of recurring congestion, in which case a time of day strategy can be used to exclude these periods from the sample. Though care must be taken to ensure that the late night hours are not over-represented since these periods are often characterized by a higher percentage of trucks in the flow (Coifman, 2001). As one might expect, the errors highlighted by the oval in Figure 1A occur most frequently during the early morning (12 midnight to 5 AM) because the flow of passenger vehicles decreases while the flow of trucks remains similar to the midday levels. In any event, as shown in Lee and Coifman (2012a), the correction factors generally remain stable for years, typically only changing after a technician adjusts the detector settings in the cabinet.

After removing the bias in Equation 1 due to the constant $L$, in this case the average absolute error for the estimated speeds in Figure 1A is 6.3 mph. These results are consistent with Coifman (2001), which showed that the assumption of constant vehicle length could lead to poor estimates because a sample’s vehicles can differ from the assumed $L$.

Coifman et al (2003) showed that using Equation 2 to estimate the median speed instead of the mean speed from Equation 1 can reduce the effect of long vehicles in a sample since the median operator is less sensitive to outliers in a distribution compared to the mean. This change can lead to significant improvements when long vehicles represent fewer than 15% of the passing traffic. Figure 1B shows the results from Equation 2 applied to the same raw detector actuations and using the same $L$ as Figure 1A. Now, however, speeds are estimated using a moving window of 11 successive vehicles rather than a fixed time $T$. The oval highlights the same region in both plots, showing clear improvement in Figure 1B, while the average absolute error dropped by half to 3.2 mph (again after correcting for bias due to length).

\[
\hat{v} = \frac{L}{\text{median on-time}}
\] (2)

Repeating this analysis at a detector station with heavy truck traffic (roughly 40% of the vehicles are longer than 30 ft) correcting for the bias resulted in $L=20$ ft. Figure 1C-D show the performance from Equation 1 and 2, respectively, applied to the raw detector actuations for the new detector on the same
day as was used in the top of the figure. The conventional estimate from Equation 1 has an average absolute error of 25.9 mph while Equation 2 has an average absolute error of 23.8 mph. Figure 2 shows the distribution of measured individual vehicle effective lengths for the entire day from the dual loop detector used in Figure 1C-D. Although Equation 2 improved performance at the typical detector in Figure 1B, both of the methods performed poorly at the detector with heavy truck traffic in Figures 1C-D.

**Speeds from Distinct Vehicle Sequences**

The speeds of two successive vehicles should usually be close to one another, even during congestion, e.g., Figure 3 shows a scatter plot comparing the measured speed of successive vehicles for an entire day at a dual loop detector. Since each vehicle’s on-time is proportional to its length, the ratio of successive on-times should be close to the corresponding ratio between the unknown vehicle lengths. For example, the ratio of on-times of successive passenger cars will be close to 1, while the ratio of on-times of a car followed by a truck will be far from 1, proportional to the ratio of their lengths. From Figure 2, we see that the typical passenger vehicle has a length of approximately 20 ft while the longest trucks are about 80 ft. We would expect the maximum ratio to occur at these extremes, i.e., 1:4.

**Finding Distinct Vehicle Sequences**

Assuming that there are no detector errors, one would only expect to see an on-time ratio of 4 whenever a car is immediately followed by long truck. Thus, although we cannot measure length directly, we can deduce individual vehicle lengths when observing these extreme on-time ratios. This idea is exploited to find sequences of long and short vehicles when they are moving together in traffic. Two such sequences: short vehicle followed by a long vehicle (SL) and long vehicle followed by a short vehicle (LS), are identified by comparing the ratio of on-times of successive vehicles in the given lane. Instead of using a discrete ratio, a range of 3.5-4.5 is used to accommodate for the small variations in the lengths and speeds of vehicles, e.g., during free flow periods, short vehicles often travel at slightly higher speeds than trucks. Once the range of the ratio is established with the lower and upper bounds (LB and UB respectively), the SL and LS sequences can be found in the time series of on-times from Equations 3 and 4. Where, $on_i$ is the on-time of i$^{th}$ vehicle, numbered successively, LB is set to 3.5, and UB is set to 4.5.

We then calculate the individual speeds for the pair of vehicles in the sequences using Equation 5, where $L_s$ and $L_L$ are mode lengths from a measured or estimated length distribution. These distributions are usually bimodal, with strong narrow peak for short vehicles and a broader peak for long vehicles, e.g., as evident in Figure 2. For the stations considered in this work, the mode for short on-times is 13/60 sec and for long on-times is 50/60 sec during known free flow periods. As expected, the ratio of the two modes is
Coifman, B. and Neelisetty, S.

close to 4. For ratios below 3.5 the error rate climbs rapidly, because one or both vehicles are not of the assumed extreme length.

If \( LB \leq \frac{on_i}{on_{i-1}} \leq UB \), then a SL sequence has been observed. \hspace{1cm} \text{(3)}

If \( LB \leq \frac{on_{i-1}}{on_i} \leq UB \), then a LS sequence has been observed. \hspace{1cm} \text{(4)}

\[ \hat{V}_s = \frac{L_s}{on_s} \text{ and } \hat{V}_L = \frac{L_L}{on_L} \] \hspace{1cm} \text{(5)}

**Accuracy of the Speed from Sequences**

To evaluate the performance of conventional estimates: Equation 1; median on-time estimates: Equation 2; and the new sequence method: Equation 5, we use two measures of error: average absolute percentage error, AAPE, and bias, as defined via Equations 6 and 7, respectively. Where \( v_i^* \) is the speed measured from the dual loop detector and \( \hat{v}_i \) the estimated speed from the single loop detector for vehicle \( i \). Table 1 shows the accuracy of these speed estimates for sequences as compared to the conventional and median on-time estimates over the 17 dual loop detector stations on the day of the experiment. Since each method has a different sampling criteria, care is taken to use the same set of vehicles for \( v_i^* \) and \( \hat{v}_i \) for the given method. In the case of the sequence method we use the speeds of the two individual vehicles. The conventional estimate comes from Equation 1 with \( T = 30 \) seconds. To facilitate comparison, we exclude all 30 sec samples that do not include a sequence used in the sequence method from the totals in Equations 6 and 7. The median on-time estimate comes from Equation 2 applied to the second vehicle in a given sequence (LS or SL) and the 10 vehicles preceding it, so here too we effectively exclude the samples without sequences from Equations 6 and 7.

\[ AAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{v}_i - v_i^*|}{v_i^*} \times 100 \] \hspace{1cm} \text{(6)}

\[ Bias = \frac{1}{n} \sum_{i=1}^{n} (\hat{v}_i - v_i^*) \] \hspace{1cm} \text{(7)}
Accommodating Detector Errors by Cleaning the Data and Establishing Benchmark Vehicles

The sequence method is sensitive to detector errors since it compares the ratio on-times. The number of false sequences identified increases in the presence of detector errors such as flicker or pulse break-ups. For example, an on-time that is terminated prematurely may be a quarter of the duration of a normal short vehicle, resulting in an erroneous SL or LS detection. So, it is important to eliminate as many such detector errors as possible by assessing the health of the detector, e.g., via Coifman and Dhoorjaty (2004), Lee and Coifman (2011, 2012a, 2012b). If such errors are too frequent, the detector may require field maintenance no matter what speed estimation strategy is used.

Online techniques can also be used to remove detector errors from the data. For this work, first, we remove unreasonably short on-times from the data assuming a maximum feasible speed for the shortest possible vehicle, yielding the minimum feasible on-time, anything shorter is regarded as erroneous. This approach works well during free flow but may still encounter problems in congestion due to extremely low speeds, detector errors and truly short vehicles such as a motorcycle. During heavy congestion, the sequence matching can erroneously pick up two short vehicles going significantly different speeds (e.g., 2 mph and 8 mph are off by a factor of 4). One could use an occupancy threshold (Coifman, 2001) to turn off the sequence matching when congestion is heavy. Alternatively, within the moving window, one could look for an on-time in excess of what is feasible for a free flow truck and/or the absence of any on-times that are feasible for a free flow car (we employ this latter idea below to accommodate for extremely low flow). Meanwhile, a more robust approach to transient errors than our implementation presented herein would compare the short on-time in a sequence against the neighborhood of recent vehicles, e.g., the previous 2 min, to make sure a given on-time was not unusually short either due to a detector error or a truly short vehicle such as a motorcycle. Given sufficient sample size (e.g., at least 30 vehicles) the passenger vehicles will almost always exhibit a strong narrow peak even under changing speeds (we have pursued this idea in subsequent work: Coifman and Kim, 2009). A slightly more complicated scheme could be implemented where on-times that fall below a certain percentile within the distribution from the neighborhood vehicles are not used for sequence detection. Of course during low flow periods it may take many minutes to observe the required number of vehicles, but such periods can be easily identified using an occupancy threshold (Coifman, 2001) to turn off sequence matching when flows are very low. Alternatively, one could look at the spectrum of possible speeds arising from all of the individual on-times within the sample assuming a 20 ft vehicle, and then again with an 80 ft vehicle.
Accommodating Extremely Low Flow

During periods of extremely low flow many minutes can elapse between observed sequences, leading to latency in the sequence method. Likewise, many minutes could pass before one observes 11 vehicles, leading to latency in the median on-time method. These extremely low flow conditions only arise during very heavy congestion or extremely low demand in the lane. To circumvent the latency we do not use data more than 2 min old (or the sample period, whichever is longer). Under the extremely low flow conditions this time window may yield so few vehicles that the unmodified sequence method or unmodified median on-time method do not have enough vehicles to yield an accurate speed estimate. So it is critical to distinguish between the two extremely low flow conditions using other metrics and there are many ways to do so. The simplest metric is probably the occupancy threshold from Coifman (2001). In this paper we use a slightly more generalized approach and look for telltale signs in the time series data that clearly reveal the underlying source of extremely low flow. First we check for evidence of free flow traffic and then verify there is no evidence of congested traffic, as outlined below. If both tests indicate free flow we will override an estimate of congested speeds from the sequence method or median on-time method (as will arise when there are mostly trucks in the sample and no sequences are found). Neelisetty (2004) observed the characteristics of on and off-times over a window of 11 vehicles ending on the current vehicle to establish whether traffic is congested or free flowing, independent of the speed estimates.

Free flow traffic is characterized by relatively short on-times and occasional long off-times (the time gap between two successive vehicles during which the detector is unoccupied). Neelisetty established that a vehicle must be in free flow conditions (speed > 45 mph), if it satisfies all of the tests in Equation 8 using an 11 vehicle window to calculate the measures. Equation 8a requires all off-times in the sample to be above 1.7 sec, which rarely occurs for 11 successive vehicles in congestion, while Equation 8b requires the maximum off-time to be at least 25 sec, indicative of the long gap between platoons.

\[
\begin{align*}
\text{Minimum off-time} & > T_1 = 100/60 \text{ sec} \\
\text{Maximum off-time} & > T_2 = 1500/60 \text{ sec} \\
\text{Standard deviation of off-times} & > T_3 = 400/60 \text{ sec} \\
\text{Median off-time} & > T_4 = 400/60 \text{ sec}
\end{align*}
\]

Congested traffic is characterized by the occasional on-time that is excessively long. After excluding all vehicles labeled free flow by Equation 8, Neelisetty (2004) established that a vehicle must
Coifman, B. and Neelisetty, S.

be in congested conditions (speed < 45 mph), if it satisfies all of the tests in Equation 9. Equation 9a requires all 11 on-times in the sample to be at least 0.5 sec (a free flow passenger car should have an on-time of 0.2 sec) while Equation 9b requires the maximum on-time to be at least 1.7 sec (even the longest free flow vehicle should have an on-time below 0.7 sec). In Equation 9c, \( \frac{on_i \times 100\%}{on_i + off_{i-1}} \) gives the individual vehicle occupancy and \( off_{i-1} \) is the preceding off-time. Neelisetty's values for the constants, \( T_1 \) to \( T_8 \), should be a good starting point for other detector stations, but they can easily be calibrated for a given location by observing the data from many days.

Minimum on-time > \( T_5 = 30/60 \) sec \hspace{1cm} (9a)

Maximum on-time > \( T_6 = 100/60 \) sec \hspace{1cm} (9b)

Individual vehicle occupancy > \( T_7 = 40/60 \) sec \hspace{1cm} (9c)

Standard deviation of on-times > \( T_8 = 45/60 \) sec \hspace{1cm} (9d)

**Hybrid Algorithm for Estimating Speed**

The speed estimates from Equation 2 are good under typical urban vehicle populations but they are poor when there are heavy truck flows. In these severely heterogeneous traffic conditions, the true but unknown average vehicle length for the median on-time will also exhibit a large range. Our new sequence method excels under these severely heterogeneous traffic conditions and provides good estimates, but the sequence method is dependent on the passage of sequences, and depending on the location, there could be gaps between successive sequences in excess of 15 min. So in this section we develop a hybrid, which uses speed estimates from Equation 2 and sequences from Equation 5 to produce new aggregate speed estimations for a given sample period \( T \), set to 30 sec for illustrative purposes.

**Aggregation Process**

Figure 4 shows the flow chart for the hybrid algorithm. We start with the current sample. Whenever sequences fall within that sample during aggregation, we trust their speed estimates as we have eliminated most detector errors before finding sequences (as per the previous section). If more than one sequence is observed, we take median of the speed from the vehicles in those sequences to further reduce the sensitivity to errors. When there are no sequences in the current sample, the median of speed of any sequences within the preceding two minutes is used because traffic conditions do not normally change significantly within that time frame, e.g., as implicitly accepted via the common used 5 min sampling
period via Equation 1. When there are no sequences in the current sample or within the preceding two minutes, we compare find the median on-time estimate from Equation 2 from the previous 2 min (although the input period is 2 min, the resulting speed is only attributed to the current sample). This estimate will underestimate speed if the sample has mostly trucks, as often occurs between midnight and 5 am. If the speed from Equation 2 > 45 mph, it will be kept outright, since long vehicles can only bring the estimate down. If the speed from Equation 2 < 45 mph, we then check to make sure this measurement does not come from Extremely Low Flow conditions using the tests from Equations 8 and 9. Whenever the flow is moderate, neither test will pass, and thus, the median on-time speed will be kept. If Equation 8 indicates that traffic is free flow and Equation 9 does not indicate that traffic is congested, a congested median on-time speed will be discarded since Equations 8 and 9 are more restrictive than Equation 2 and thus, the sample likely contains mostly long vehicles. We can then use a fixed assumed free flow speed at the detector as an estimate for the current sample. This process is repeated for each sample. Finally, if $T > 2$ min, then all occurrences of "2 min" in Figure 4 are changed to "$T$".

**Evaluation**

Data from 34 days were used to test the new hybrid method and the results presented here are from a typical day among the evaluation set. The new hybrid method of speed estimation is first tested at a detector with a moderate flow of trucks (the same detector that was used in Figure 1A-B). Figure 5 presents the results for a single day of input data and show the results from the hybrid method (plots C & F) contrasted against the concurrent conventional (Equation 1, in plots A & D) and median on-time (Equation 2, in plots B & E) estimates for two sampling criteria, 30 sec and 5 min over the same data set. The ovals are added for reference, and are shown in the same location on all of the plots. Though the median on-time method works better than the conventional method for $T=30$ sec, as seen from the density of points in the circled region, it is still not able to eliminate all of the estimation errors. The hybrid method performs the best in this case. When the sampling period is increased to 5 min, the contrast between the three methods decreases but is still evident. The top rows of Table 2 show the performance of the three methods for $T=30$ sec before and after bias removal relative to the dual loop detector measured speed.

The situation changes when we examine the results from the detector that has a truck flow in excess of 40% of its daily traffic flow of 10,000 vehicles/day (the same detector that was used in Figure 1C-D). This detector sees such high truck flow because it is outside the city’s ring road and is in the shoulder lane just after an exit for a major shopping center. So, most of the passenger vehicles depart this lane just upstream of the detector while the through trucks remain. This station does not see recurring
Coifman, B. and Neelisetty, S.

congestion, so we only present results for free flow. Figure 6 shows the results for T=30 sec and 5 min using the same format as the previous figure. Both the conventional and median on-time estimates perform poorly at this location, while the new hybrid method shows significantly better performance. Even after increasing the sampling period to 5 min, both of the earlier methods show poor performance. We see frequent errors for the conventional and median on-time methods because the true but unknown average vehicle length for the given sample comes from a much larger range than the discrete L assumed by the two older methods. The lower rows of Table 2 show the performance of the three methods for T=30 sec before and after bias removal relative to the dual loop detector measured speed.

Figure 7 shows the cumulative distribution function for the errors, $v_i^* - \hat{v}_i$ for T=30 sec via each of the three speed estimation methods from the entire 34 days of data at the two detectors presented in Figure 1. Specifically, plots A-B show the results before and after correcting for bias, respectively, for the detector with moderate truck flow, around 10% of the vehicle fleet. Plots C-D repeat the process for the detector with heavy truck traffic. It can be seen that the new hybrid method performs better than the conventional method and the median on-time method at both detectors, with a major improvement when heavy truck traffic is present. Consistent with Figure 1, the L used in plots A & C differs. After removing the bias, Figure 7B & D, compare the error distributions of from each of the three speed estimation methods and shows that our new Hybrid method exhibits a smaller range of errors. Furthermore, our method is the only one of the three that automatically removes the bias.

**Discussion and Conclusions**

Single loop detectors are the most widely deployed traffic sensors on freeways. Though they provide valuable information of flow and occupancy, there is no single robust method to estimate speed accurately at all locations. This topic has been the focus of research for years, but little work has been done to estimate speed from single loop data in presence of heavy truck flows. This paper has shown the limitations of earlier speed estimations in this situation and developed a hybrid methodology that addresses this problem without degrading performance under other traffic conditions. The results show that the new hybrid estimate is a significant improvement over conventional speed estimates.

Most loop detector stations collect raw individual vehicle actuation data as used in this study; however, conventionally the controller software aggregates the data to longer sample periods (typically ranging between 20 sec and 5 min) and then discards the raw data. This conventional aggregation approach made sense when data storage and communication were expensive, but it impedes detailed numerical analysis that can extract much more information about the traffic stream (e.g., Coifman 2002,
2003; Coifman and Krishnamurthy 2007) and detector health (e.g., Lee and Coifman 2011, 2012a, 2012b). In the short term the advances presented herein can be realized through new controller software, changing the algorithm used to aggregate data. In the longer term, we think operating agencies should consider revising the communications network to collect the raw actuation data rather than the aggregated data. As demonstrated by this paper and the earlier work by our group cited above, there is a wealth of information that can be extracted from the raw actuation data.

**ACKNOWLEDGMENTS**

This work was performed as part of the California PATH (Partners for Advanced Highways and Transit) Program of the University of California, in cooperation with the State of California Business, Transportation and Housing Agency, Department of Transportation. The Contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification or regulation.

**REFERENCES**


Coifman, B. and Neelisetty, S.


Figure 1. Comparison of estimated versus measured speeds for (A) conventional method with $T=30$ sec and (B) median on-time with $N=11$ vehicles, for a typical detector. Repeating the comparison for a detector with heavy truck flows (C) conventional method with $T=30$ sec and (D) median on-time with $N=11$ vehicles.
Figure 2, Distribution of effective vehicle lengths observed at a detector with heavy truck flows (A) Cumulative distribution (B) Frequency distribution.
Figure 3, Speeds of successive vehicles are typically close, as illustrated using one day of measured speed data from a dual loop detector, the outer lines show the 10 mph bounds.
Figure 4. Flow chart for the hybrid method of estimating speed. Note that if $T > 2$ min, then all occurrences of “2 min” in the flow chart are changed to “$T$.”
Figure 5. The performance of the conventional (first column), median on-time (second column) and the hybrid speed estimation method (third column) for a typical detector. A, B, C show the results for T=30 sec and D, E, F for T=5 min. All plots use the same raw detector actuations. The circles are drawn for reference over the same regions in each plot to facilitate comparison.
Figure 6, The performance of the conventional (first column), median on-time (second column) and the hybrid speed estimation method (third column) for the detector with heavy truck flows. A, B, C show the results for T=30 sec and D, E, F for T=5 min. All plots use the same raw detector actuations. The circles are drawn for reference over the same regions in each plot to facilitate comparison.
Figure 7, Distribution of errors for the conventional, median on-time and hybrid speed estimation methods for an entire month of data (A) and (B) show error distribution before and after correcting for bias at the same typical detector (C) and (D) show error distribution before correcting for bias for the detector with heavy truck traffic. Text in (A) and (C) shows mean error for each method, while text in (B) and (D) shows vehicle length that made mean error zero for each method.
TABLE 1. Comparison of AAPE and Bias for all the three methods: conventional (L*q/occ), sequences (SL and LS) and median on-time (Median) for an entire day of data.

<table>
<thead>
<tr>
<th></th>
<th>L*q/occ AAPE</th>
<th>Congestion SL AAPE</th>
<th>Median LS AAPE</th>
<th>Median LS Bias</th>
<th>Median LS AAPE</th>
<th>Median LS Bias</th>
<th>Free flow L*q/occ AAPE</th>
<th>Free flow SL AAPE</th>
<th>Free flow Median AAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>AAPE</td>
<td>Bias</td>
<td></td>
<td>Bias</td>
<td></td>
<td>AAPE</td>
<td>Bias</td>
<td>AAPE</td>
</tr>
<tr>
<td>Long</td>
<td>13.4</td>
<td>1.3</td>
<td>10.7</td>
<td>5.5</td>
<td>0.1</td>
<td>18.7</td>
<td>6.0</td>
<td>7.0</td>
<td>-3.6</td>
</tr>
<tr>
<td></td>
<td>Short</td>
<td>13.4</td>
<td>1.3</td>
<td>6.2</td>
<td>-2.5</td>
<td>18.7</td>
<td>6.0</td>
<td>5.5</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>L*q/occ AAPE</td>
<td>LS AAPE</td>
<td>Median LS AAPE</td>
<td>Median LS Bias</td>
<td>Median LS AAPE</td>
<td>Median LS Bias</td>
<td>L*q/occ AAPE</td>
<td>LS AAPE</td>
<td>Median LS AAPE</td>
</tr>
<tr>
<td>Long</td>
<td>14.0</td>
<td>1.8</td>
<td>6.8</td>
<td>6.7</td>
<td>0.3</td>
<td>19.9</td>
<td>6.8</td>
<td>4.9</td>
<td>-2.5</td>
</tr>
<tr>
<td>Short</td>
<td>14.0</td>
<td>1.8</td>
<td>5.2</td>
<td>6.8</td>
<td>0.1</td>
<td>19.9</td>
<td>6.8</td>
<td>4.3</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Comparison of AAPE and Bias for an entire day of data.
TABLE 2, AAPE comparison of the three methods for a typical detector and a detector with heavy truck flows for T=30 sec. Note at the typical detector that the large AAPE during congestion corresponding to average absolute errors less than 5 mph. We only present free flow results for the heavy truck flow detector because it does not see recurring congestion.

<table>
<thead>
<tr>
<th></th>
<th>AAPE before bias removal</th>
<th></th>
<th>AAPE after bias removal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L*q/occ</td>
<td>Median</td>
<td>Hybrid</td>
</tr>
<tr>
<td><strong>typical detector</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free Flow</td>
<td>30.3</td>
<td>13.2</td>
<td>6.7</td>
</tr>
<tr>
<td>Congestion</td>
<td>26.0</td>
<td>26.4</td>
<td>19.9</td>
</tr>
<tr>
<td><strong>Heavy truck flow detector</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free Flow</td>
<td>56.7</td>
<td>54.3</td>
<td>5.5</td>
</tr>
<tr>
<td>Congestion</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>