Title

New Aggregation Strategies to Improve Velocity Estimation from Single Loop Detectors

Authors
Coifman, Benjamin
Lee, Zu-Hsu

Publication Date
2000-08-01

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

New Aggregation Strategies to Improve Velocity Estimation From Single Loop Detectors

Benjamin Coifman
Zu-Hsu Lee

California PATH Working Paper
UCB-ITS-PWP-2000-12

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 Transportation, Federal Highway Administration.

Report for MOU 3000

August 2000

ISSN 1055-1417

Estimating Median Velocity Instead of Mean Velocity at Single Loop Detectors

Benjamin Coifman, PhD

Assistant Professor, Civil and Environmental Engineering and Geodetic Science
Assistant Professor, Electrical Engineering

Ohio State University
470 Hitchcock Hall
2070 Neil Ave
Columbus, OH 43210

Coifman.1@OSU.edu
http://www-ceg.eng.ohio-state.edu/~coifman
614 292-4282

Zu-Hsu Lee

University of California
Department of Industrial Engineering and Operations Research
Berkeley, CA 94720

leez@ieor.berkeley.edu

Submitted for publication in Transportation Research-B

7/11/00

ABSTRACT

Loop detectors are the preeminent vehicle detector for freeway traffic surveillance. Although single loops have been used for decades, debate continues on how to interpret the measurements. Many researchers have sought better estimates of velocity from single loops. The preceding work has emphasized post-processing techniques. Although rarely noted, these techniques effectively seek to reduce the bias due to long vehicles in measured occupancy and flow. This paper presents a different approach, using a new aggregation methodology to estimate velocity and reduce the impact of long vehicles in the original traffic measurements.

Keywords:
traffic surveillance, loop detectors, velocity estimation

INTRODUCTION

Loop detectors are the preeminent vehicle detector for freeway traffic surveillance. They are frequently deployed as single detectors, i.e., one loop per lane per detector station. Recent work by our group has identified a fundamental shortcoming with conventional estimates of velocity from single loop detectors. These estimates presume a "mean vehicle length"[1] that applies to all samples, but Coifman (2000) has shown that this assumption breaks down because a given sample may not be representative of the "average vehicle".

Many researchers have investigated post-processing techniques to reduce the influence of long vehicles, e.g., Mikhalkin et al (1972), Pushkar et al (1994), Dailey (1999), and Coifman (2000). All of these studies used aggregate flow (q) and occupancy (occ) to estimate mean velocity. Rather than manipulating aggregate data, this paper examines new aggregation methods to reduce the estimation errors.

The first section of this paper reviews the state of the practice and the related shortcomings of parameter estimation from single loop detectors. The next section proposes an alternative method for estimating velocity and the final section contrasts the new approach against conventional estimates.

CONVENTIONAL VELOCITY ESTIMATION

Provided that vehicle lengths and vehicle velocities are uncorrelated, it can be shown that harmonic mean velocity (mean v) and arithmetic mean vehicle length (L) for a given sample are related by the following equation:

$$mean\ v \approx \frac{q \cdot L}{occ} \tag{1}$$

However, these two parameters can not be measured independently at a single loop. Typically, an operating agency will set L to a constant value and use Equation 1 to estimate velocity from single loop measurements. But this approach fails to account for the fact that the percentage of long vehicles may change during the day or the simple fact that a sample may not include "typical" vehicle lengths. Particularly during low flow, when the number of vehicles in a sample is small, a long vehicle can skew occupancy simply because it takes more time for that vehicle to pass the detector. For example, Coifman (2000) found that approximately 85 percent of the individual

---

[1] Throughout this paper, "length" refers to the "effective vehicle length" as *seen* by the detectors.

vehicle lengths observed at one detector station were between 15 and 22 feet, but some vehicles were as long as 85 feet, or roughly four times the median length.

In accordance with the law of large numbers, the sample distribution should become more representative of the entire population as the sample size increases, which in turn, increases with flow. Figure 1 illustrates this phenomena using two common sampling periods (T). In part A, T=30 sec and the maximum number of vehicles per sample is so small that the observations fall into distinct columns, i.e., the first column contains observations with only one vehicle, the second column contains observations with only two vehicles, and so on. Notice that for both values of T, the range of L is inversely proportional to q.

ALTERNATIVE PARAMETERS

For this study, we examine 24 hours of detector actuations, sampled at 60 Hz, from a dual loop directional-detector[2] station, where a directional-detector consists of two closely spaced loop detectors in the same lane. In this configuration it is possible to measure true vehicle velocities by the quotient of the loop separation and the difference in arrival times at each loop. While an individual vehicle's length is simply the product of its velocity and *on time*[3]. The detector station includes five lanes in each direction. In an attempt to capture the temporal changes in vehicle lengths, the data were arbitrarily subdivided into three hour long segments by lane. The two distributions shown in Figure 2 represent the "best" distribution (lowest vehicle length variance) and "worst" distribution (highest vehicle length variance) observed across the 80 subsets. Removing the temporal component, consider a sample of N vehicles drawn at random from the "worst" distribution. Intuitively, the sample mean vehicle length is likely to be biased towards long vehicles because of the extended tail. The median vehicle length, however, should be less sensitive to the outliers. This hypothesis was verified using Monte Carlo simulation. The simulation consisted of 10,000 samples of N vehicles from each distribution, where N was set to 10, 50, 100, 500 and 1000 vehicles, and the sample mean and median were calculated. Table 1 summarizes the 99 percent confidence intervals for the mean and median lengths. The mean length confidence interval was significantly worse than the median length under all conditions. In fact, the range of the median length confidence interval for N vehicles was proportional to that of that of the mean for $10 \cdot N$. If we continue to assume that individual vehicle length and velocity are uncorrelated, then the simulation results lead to the following postulate:

---

[2] Commonly referred to as *speed trap detectors.*

$$median\ v \approx \frac{L}{median\ on\ time} \tag{2}$$

ESTIMATING VELOCITY

For both Equations 1 and 2, with a fixed L, one can consider the function on the right hand side as an estimate of the parameter on the left hand side. Using the entire day's worth of data from each lane and setting the sample size to N consecutive vehicles in a given lane, Figures 3 and 4 show scatter plots of the estimates versus the corresponding measured parameters for two different values of N. In each figure, the top and bottom plots come from the same samples and L is assumed to be 20 feet in all plots. Any error in L would skew the estimates proportional to the true values. In both figures, the mean velocity estimate is much noisier than the median velocity estimate. Notice that the median velocity and its estimate tend to fall into discrete columns and rows, respectively, due to the resolution of the 60 Hz measurements. With N=10, the mean estimate is subject to errors from long vehicles throughout all traffic conditions, as highlighted with the circles in Figure 3. At larger N, Figure 4, the error is only evident during free flow conditions. This bias is due to the fact that trucks represent a larger percentage of the vehicle fleet in the early morning hours, a period when there was no observed congestion. To quantify these errors, we define the Measure of Variance (MOV) and Measure of Bias (MOB) over n samples as follows:

$$MOV = \frac{\sum_{i=1}^{n}\left(x_i^* - \hat{x}_i\right)^2}{n} \tag{3}$$

$$MOB = \frac{\sum_{i=1}^{n}\left(x_i^* - \hat{x}_i\right)}{n} \tag{4}$$

where x_i^* is the true value of the parameter for the i-th sample and $\hat{x}_i$ is the corresponding estimate. The resulting MOV and MOB for the velocity estimates from five different sample sizes are shown in the first few columns of Table 2. In each case, the MOV for the median velocity estimator is approximately one third of that for the mean velocity estimator. Of course, the MOV is sensitive to the choice of L. So the analysis is repeated in the latter columns with L selected such that MOB=0 for the given sample size and parameter.

---

[3] *On time* is the amount of time that a given vehicle occupies the detector.

The estimates thus far are based on samples of a fixed number of vehicles. But the fixed number sampling is not very informative if the freeway is blocked or flow drops for some other reason. So in practice, it is better to sample over fixed time periods. Fixed time sampling has the added benefit that all samples can be synchronized at a detector station and thus, requires less computational and communications overhead. It is important, nonetheless, to ensure that the sample period is large enough to ensure a sufficient number of vehicles in a sample during any period when surveillance is desired. Looking at Figure 3 and Table 2, N=10 vehicles appears to provide satisfactory results for the median estimate, but this sample size is a little low for the mean estimate. With T=30 seconds, this criteria would require q>1200 veh/hr throughout the entire day. Applying this sampling period to the data results in very poor performance by both estimation techniques (see Table 2). On the other hand, if T=5 minutes, the criteria only requires q>120 veh/hr. Repeating the preceding analysis with T=5 minutes yields Figures 5 and 6. In Figure 5, one can see significant errors in the mean estimate and few errors in the median estimate. The corresponding statistics are reported in Table 2 and the performance appears to be on the order of fixed samples with N=10 vehicles. Figure 6 shows that the flow is quite low at this site for at least four hours in the early morning and it is above 1200 veh/hr for only about half of the day. The reader should also note that this location sees a relatively large volume of traffic, with an average of approximately 25,000 vehicles/lane/day for this data set. Most other locations will have a lower average flow, reaffirming the need for longer sample periods.

Although T=5 minutes appears to provide sufficient sample size, the long delay between measurement updates may be undesirable. Fortunately, many applications only need a single estimate of velocity for a given detector station or link. To keep N high while reducing T, one can sample across multiple lanes before estimating velocity. In particular, setting T=30 seconds and sampling individual vehicle measurements across the four outside lanes in a given direction yields Figure 7 and the final row of statistics in Table 2.[4]

Finally, note that the median estimates will degrade in the presence of high truck volumes. To address this fact, the methodology could be modified to use a lower percentile, e.g., the 25th percentile rather than the median. Additional information could also be used, such as the presence of truck restrictions in specific lanes or slightly more complicated models that exclude low flow conditions, e.g., Coifman (2000).

---

[4] The inside lane was excluded because it is a high occupancy vehicle lane in both directions.

CONCLUSIONS

Many researchers have sought better estimates of velocity from single loops. The earlier works have emphasized post-processing techniques to reduce the bias from long vehicles in mean velocity estimates. This paper has taken a different approach, it used a new aggregation methodology to estimate median velocity and it was shown that the estimate is less sensitive to the presence of long vehicles. This fact leads to the added benefit that the assumed value of L is less sensitive to site-specific characteristics of the traffic flow. Furthermore, the proposed method of estimating median velocity is simple enough that it can be deployed on existing traffic controllers.

It may seem intuitive that the median is less sensitive to outliers than the mean, but it does not appear that this fact has previously been employed for estimating velocity from single loops. Although the median is less sensitive to outliers, it is still necessary to observe several vehicles in a given sample to reduce the impact of long vehicles and it is not advisable to estimate velocity with short sample periods during low flow conditions. To this end, one method for increasing N while keeping T low was proposed. The approach combined data from multiple lanes before estimating velocity and yielded satisfactory results on the experimental data set.

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; and the United States Department of Transportation, Federal Highway Administration.

REFERENCES

Coifman, B. (2000) Improved Velocity Estimation Using Single Loop Detectors. Transportation Research: Part A, [accepted for publication]. (draft available at: http://www-ceg.eng.ohio-state.edu/~coifman/documents/SingleLoop.pdf)

Dailey, D. (1999) A Statistical Algorithm for Estimating Speed from Single Loop Volume and Occupancy Measurements. *Transportation Research-B*, Vol 33B, No 5, June 1999, pp 313-322.

Coifman, B. and Lee, Z.

Mikhalkin, B., Payne, H., Isaksen, L. (1972) Estimation of Speed from Presence Detectors. *Highway Research Record 388*, pp 73-83. Highway Research Board, Washington, DC.

Pushkar, A., Hall, F., Acha-Daza, J. (1994) Estimation of Speeds from Single-Loop Freeway Flow and Occupancy Data Using Cusp Catastrophe Theory Model. *Transportation Research Record 1457*, pp 149-157. Transportation Research Board, Washington, DC.

Figure 1, Observed average effective length versus flow for five lanes at one detector station, over one day sampled at (A) T = 30 sec, (B) T = 5 min.

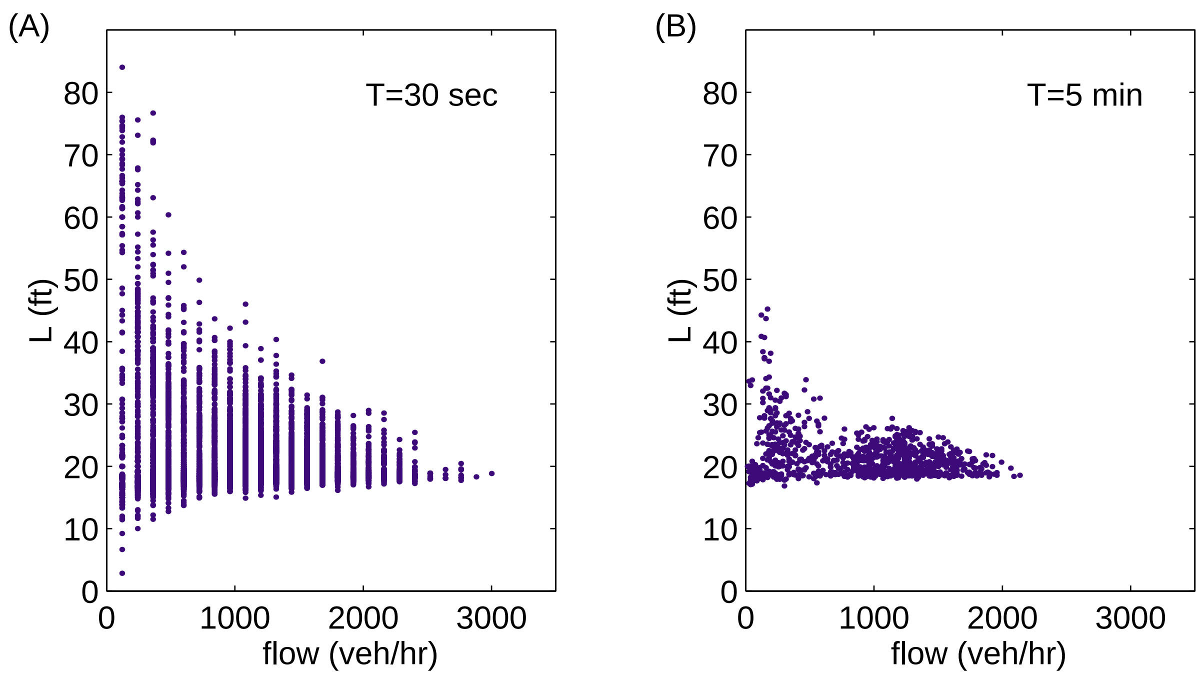
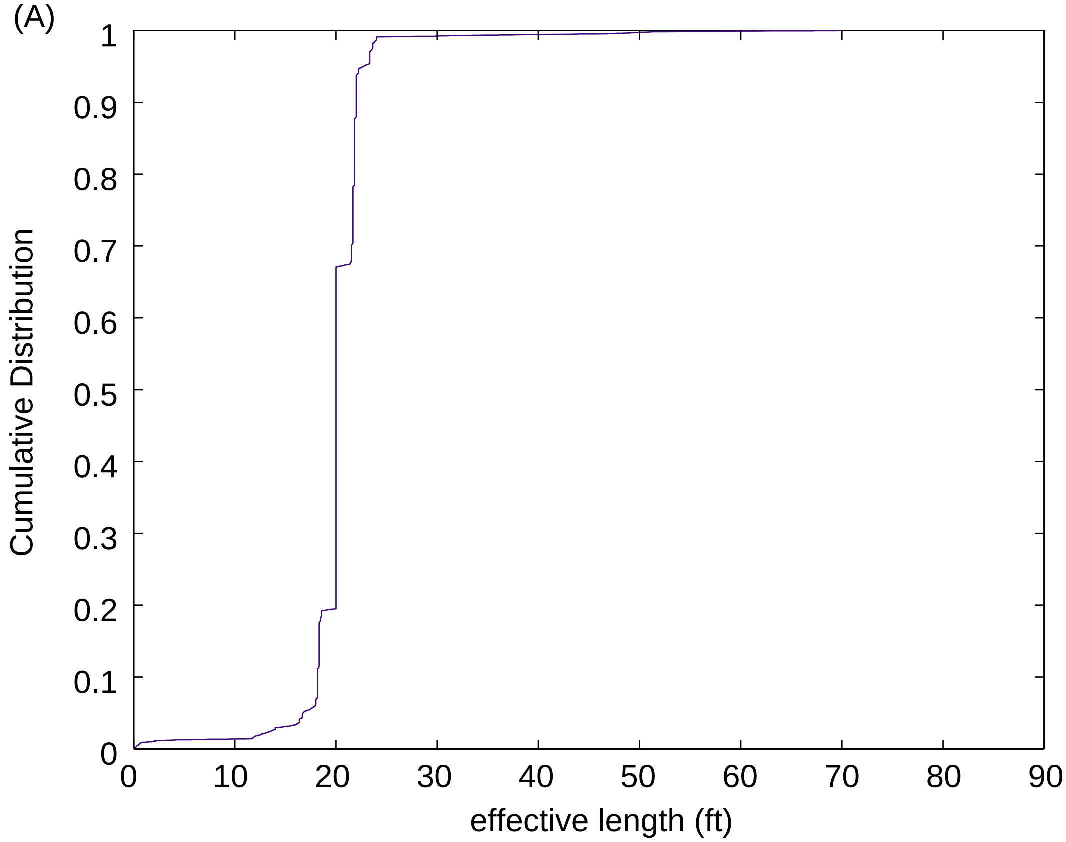


Figure 2, Observed distributions of individual effective vehicle lengths in a single lane for three hour periods. (A) lowest variance or "Best" case, (B) highest variance or "worst" case.

(A)



(B)

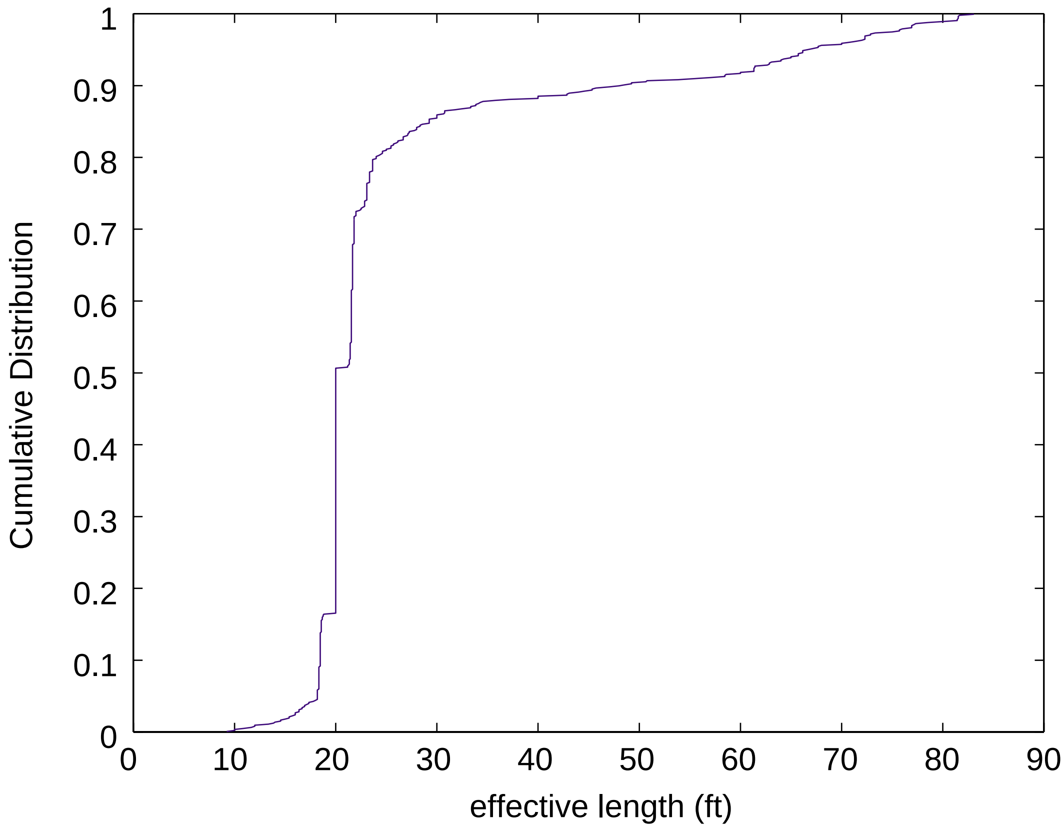


Figure 3, This figure uses real traffic data to compare estimated versus measured (A) mean velocity (B) median velocity for 24,640 samples of 10 vehicles each. Note that the circles were added to the same locations in both plots to highlight the differences.

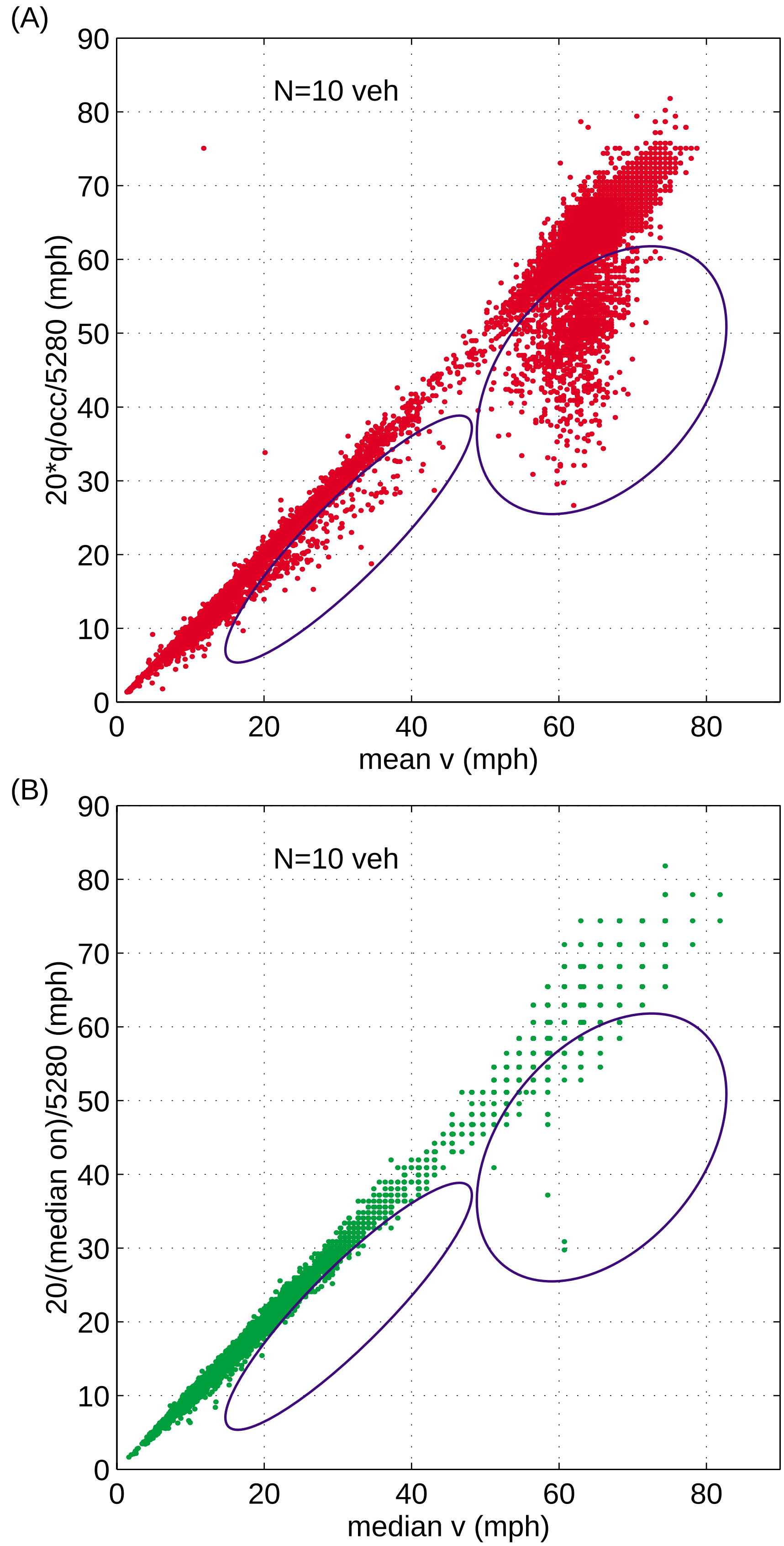


Figure 4, Now using samples of 100 vehicles each, this figure compares estimated versus measured (A) mean velocity (B) median velocity for 2,460 samples. Again, the circles were added to the same locations in both plots to highlight the differences.

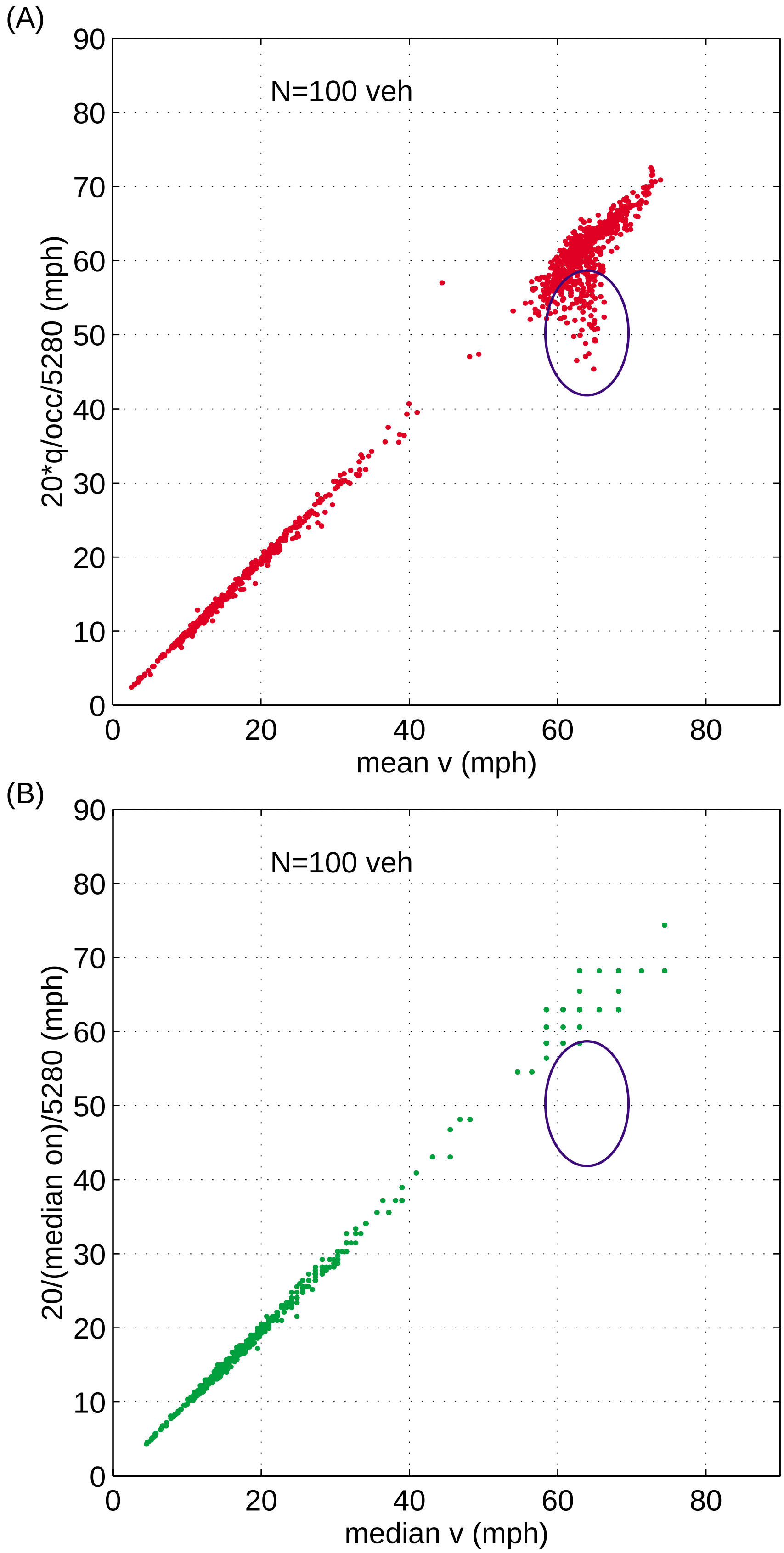


Figure 5, Moving to a fixed sample period of 5 minutes, this figure compares estimated versus measured (A) mean velocity (B) median velocity for 2,870 samples. Again, the circles were added to the same locations in each plot to highlight differences.

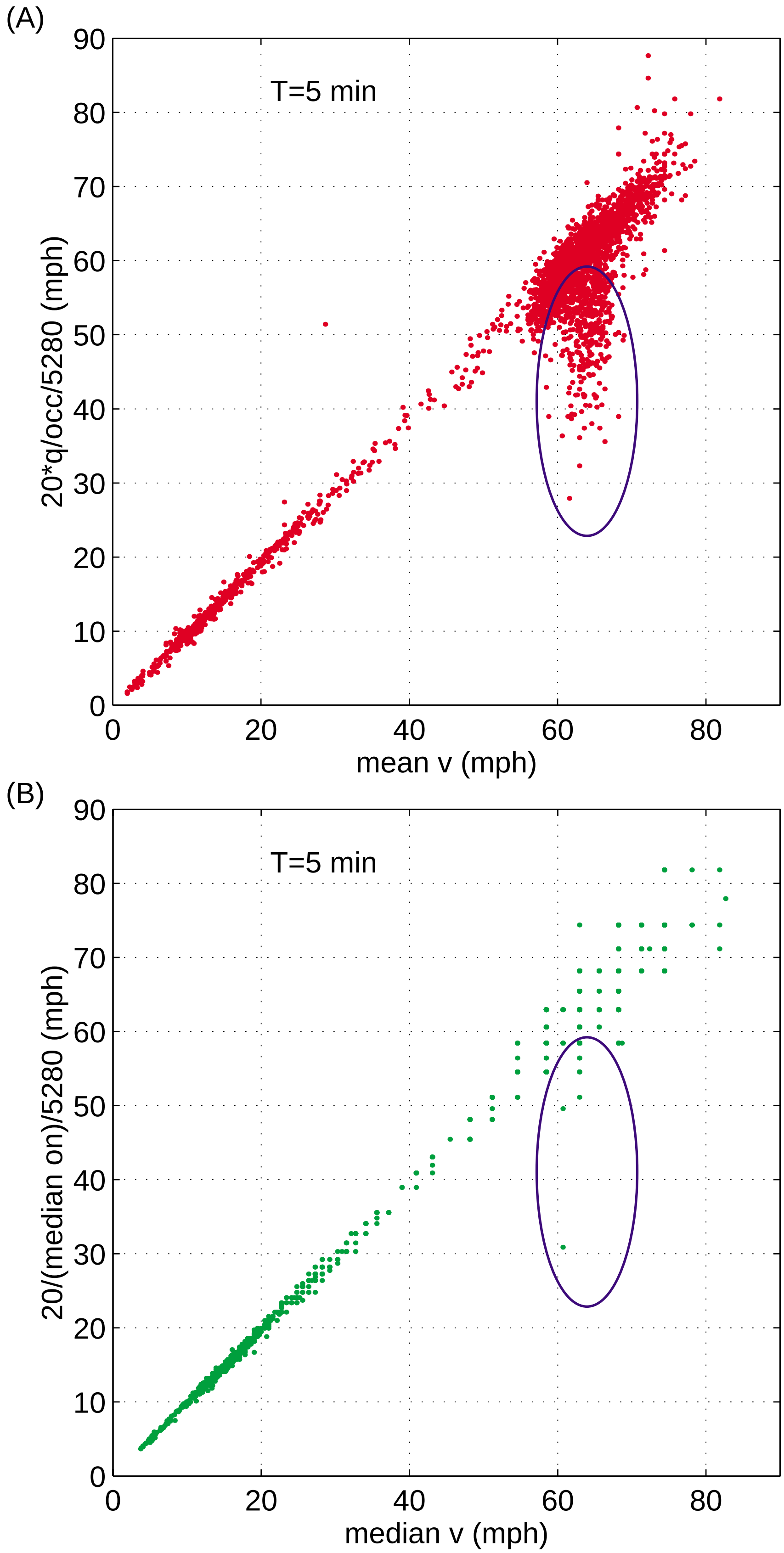
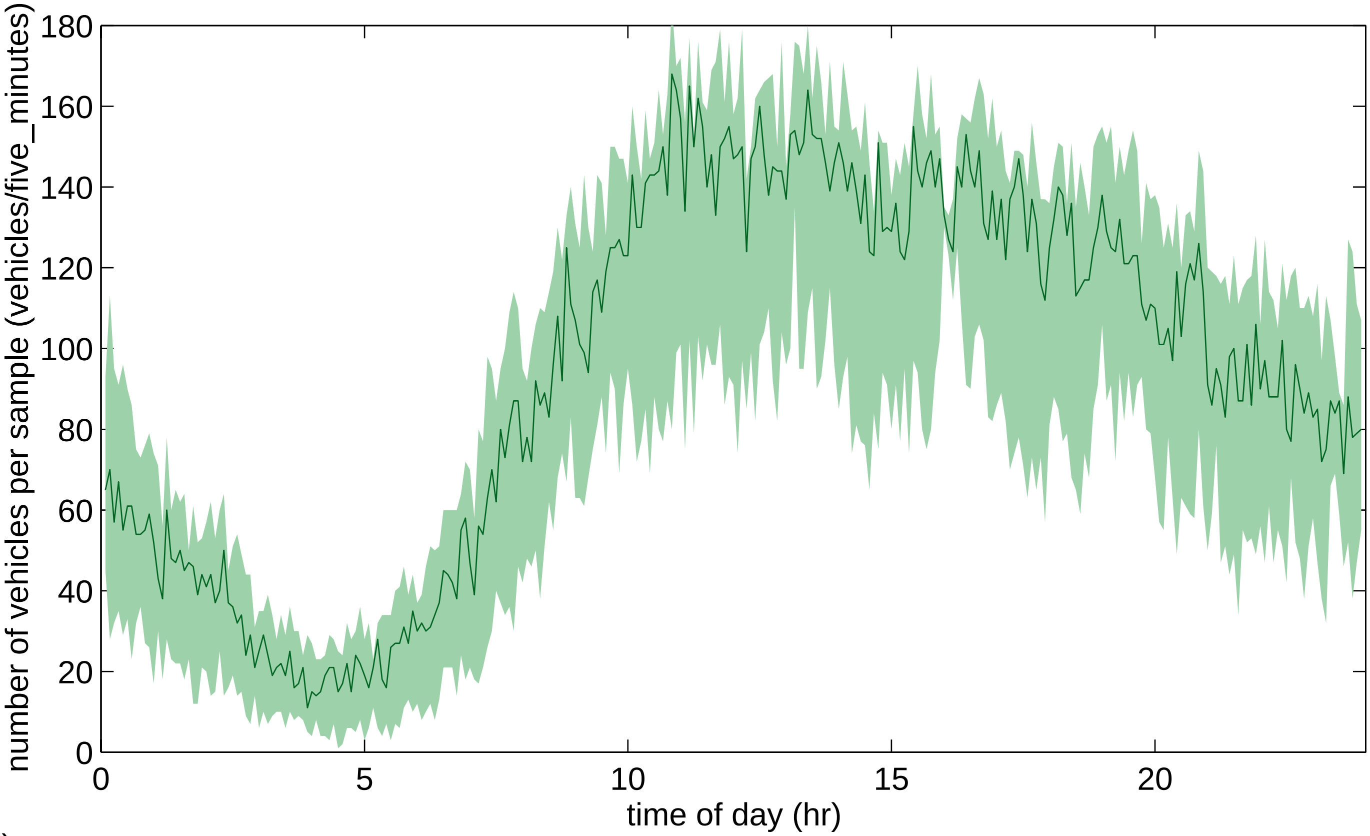


Figure 6, Range of observed sample sizes (light region) and median sample size (solid line) across five adjacent lanes for the data shown in the previous figure (A) northbound lanes, (B) southbound lanes.

(A)



(B)

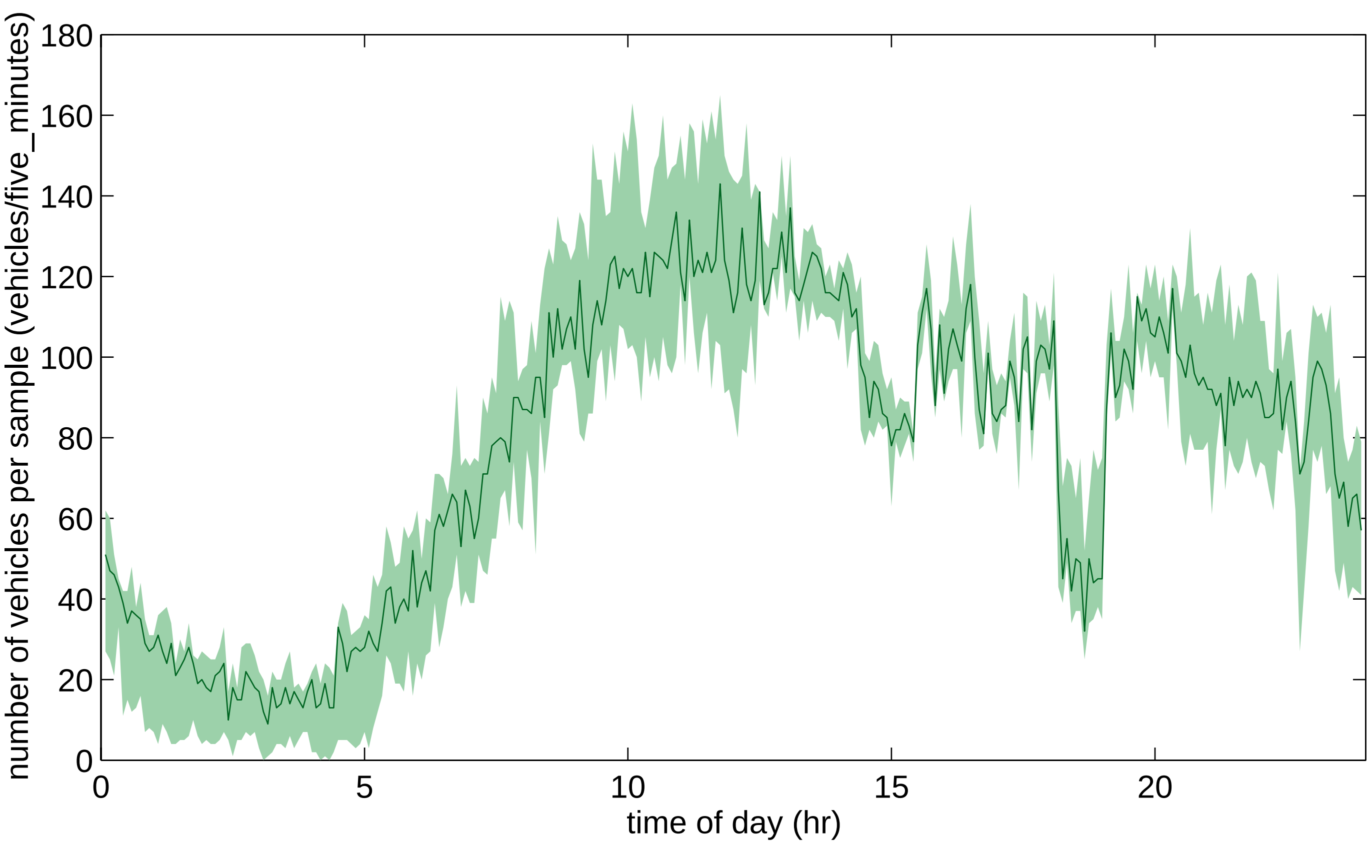


Figure 7, Finally, using a fixed sample period of 30 seconds and combining data over four lanes, this figure shows estimated versus measured (A) mean velocity (B) median velocity for 5,760 samples. Once more, the circles highlight the differences betwen the plots.

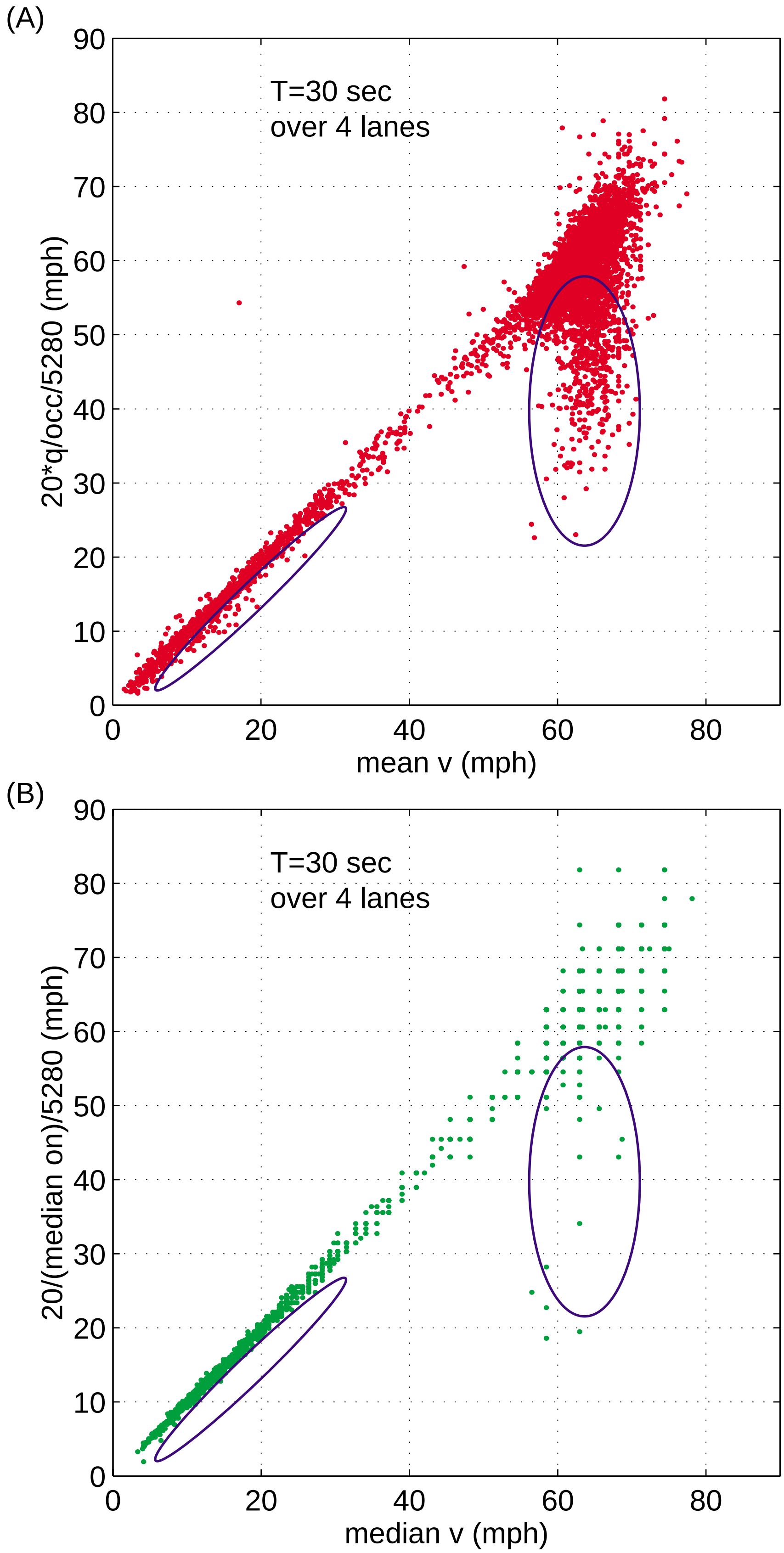


Table 1, Confidence intervals for mean and median effective vehicle length from Monte Carlo simulation for various sample sizes from the "best" and "worst" observed distributions.

		99 percent confidence intervals of vehicle length (ft)					
		for mean			for median		
		upper bound	lower bound	difference	upper bound	lower bound	difference
from "best" distribution	N=10 veh	25.07	16.65	8.42	21.82	18.45	3.37
	N=50 veh	21.78	18.75	3.03	20.77	20.00	0.77
	N=100 veh	21.27	19.16	2.11	20.00	20.00	0.00
	N=500 veh	20.59	19.71	0.88	20.00	20.00	0.00
	N=1000 veh	20.45	19.82	0.63	20.00	20.00	0.00
from "worst" distribution	N=10 veh	40.77	19.03	21.74	27.69	19.23	8.46
	N=50 veh	32.07	21.42	10.65	21.74	20.00	1.74
	N=100 veh	30.23	22.62	7.61	21.67	20.00	1.67
	N=500 veh	27.89	24.42	3.47	21.54	20.00	1.54
	N=1000 veh	27.34	24.88	2.46	21.45	20.00	1.45

Table 2, Measure of Variance (MOV) and Measure of Bias (MOB) for estimated mean and median velocity using different sampling criteria on the same set of vehicle measurements. Note that these data come from real observations rather than simulation.

		L set to 20 feet				L set to eliminate bias			
		MOV $(mph)^2$		MOB (mph)		MOV $(mph)^2$		L (ft)	
sampling criteria	number of samples	for mean v	for median v	for mean v	for median v	for mean v	for median v	for mean v	for median v
N=10 veh	24640	26.63	7.22	2.44	1.03	21.62	6.15	20.96	20.39
N=50 veh	4920	17.10	5.58	2.71	1.09	9.36	4.28	21.08	20.42
N=100 veh	2460	15.48	5.38	2.74	1.11	7.31	4.04	21.09	20.42
N=500 veh	490	13.99	4.92	2.77	1.00	5.37	3.84	21.12	20.38
N=1000 veh	235	13.54	5.09	2.78	1.05	4.77	3.86	21.13	20.40
T=30 sec	28750	63.69	29.23	2.82	1.35	60.94	28.48	21.08	20.50
T=5 min	2870	34.36	7.02	3.59	1.03	22.37	5.96	21.39	20.38
T=30 sec and combining 4 lanes	5760	47.39	8.73	4.02	0.98	33.12	7.85	21.59	20.36