

EXPERIMENTAL CHARACTERIZATION OF MULTI-LANE FREEWAY TRAFFIC UPSTREAM OF AN OFF-RAMP BOTTLENECK

Juan Carlos Muñoz¹ and Carlos F. Daganzo
Department of Civil and Environmental Engineering
and
Institute of Transportation Studies
University of California, Berkeley, CA 94720
(March 31, 2000)

Abstract

This report describes field observations of multi-lane freeway traffic upstream of an oversaturated off-ramp. It is based on empirical evidence from freeway I-880 (northbound) near Oakland, California. The report presents two diagnostic tools that reveal hidden features of the traffic stream and, based on these findings, proposes congestion mitigation strategies that could work for similar locations with little or no construction.

The first tool (oblique plots) was used to identify a “moving bottleneck” that occurred before the rush hour and to characterize, for the first time, the traffic stream directly upstream and downstream of it. The fine detail of these plots affords a unique glimpse into the realm of driver psychology. The data show that the first hundred drivers in the queue initially accepted relatively short average spacings, and that a little later the same drivers allowed themselves to relax into larger spacings. Later drivers also adopted the new spacings. The empirical observations are compared to the predictions of kinematic wave (KW) theory.

It is also found that once the queue caused by the off-ramp spill-back fills the study section, the system’s flows and speeds at all locations within the section stabilize in a steady state that lasts for over 30 minutes. The second diagnostic tool (cross-correlation

¹ Instructor at the Pontificia Universidad Católica de Chile, Ph.D. student at U.C. Berkeley.

of N-curves) reveals that during this time, and close to the congested off-ramp, small disturbances in the shoulder lanes move upstream while those in the median lanes move forward. This indicates that the shoulder lanes are in a congested/queued state while the median lanes are not, and that the freeway should not be modeled as a “single pipe” KW stream near the off-ramp.

Lane-specific behavior is not found beyond the 2.2-km mark, however. At these locations, it was found that: (i) the KW model with a wave speed of about 20 km/hr holds well in congested (queued) traffic, (ii) the onset of queuing is gradual and nearly simultaneous across all lanes and (iii) the transition zone where vehicles decelerate toward the end of the queue is about 1 km long, and propagates like a shock. Because the difference between the queued and unqueued flows was small, the transition zone moved slowly (at about 3.5 km/hr) and took almost 20 minutes to pass over a detector. Flow-occupancy scatter plots exhibit macroscopic patterns consistent with the above conditions.

The report also explains in a simple way some strange events that occurred during the congestion dissipation process, most notably, a large and sustained increase in the through flow that took place with no appreciable change in other traffic conditions. The report concludes with an assessment of the vehicle-hours of delay caused by the bottleneck, and various proposals for eliminating a large fraction of these.

TABLE OF CONTENTS

1. INTRODUCTION..... 5

2. DIAGNOSTIC TOOLS 6

 2.1 Oblique plots 7

 2.2 Estimation of wave speed..... 10

3. DATA ANALYSIS 12

4. THE INITIAL UNCONGESTED PERIOD (14:00 hrs -- 14:45 hrs) 19

5. THE CONGESTED STEADY STATE PERIOD (15:40 hrs -- 16:20 hrs) 26

 5.1. Traffic behavior upstream of station 22 28

 5.2. Traffic behavior downstream of station 22 31

6. THE TRANSITION PERIOD (14:45 – 15:40 hrs)..... 33

7. THE DISSIPATION PERIOD (16:20 – 18:00 hrs)..... 38

8. FINDINGS, CONGESTION REMEDIATION MEASURES AND BENEFIT ASSESSMENT 41

9. REFERENCES 48

APPENDIX A: THREE ESTIMATORS OF WAVE SPEED 50

APPENDIX B: ADDITIONAL DATA 55

LIST OF FIGURES

Figure 1.	8
Figure 2.	13
Figure 3.	15
Figure 4.	17
Figure 5.	20
Figure 6.	22
Figure 7.	26
Figure 8.	28
Figure 9.	29
Figure 10.	31
Figure 11.	33
Figure 12.	38
Figure A1.	53
Figure B1.	55
Figure B2.	56
Figure B3.	57
Figure B4.	58
Figure B5.	59
Figure B6.	60
Figure B7.	61

LIST OF TABLES

Table 1.	35
Table B1.....	62

1. INTRODUCTION

Partly motivated by the ideas in Daganzo (1997) and Daganzo et al. (1997), this report examines empirically the behavior of multi-lane freeway traffic upstream of an off-ramp bottleneck. These references propose an extension of KW theory that allows specific sets of drivers to be excluded from certain lanes; e.g., as is necessary to model freeway sections with car-pool lanes and/or surface streets with lane markings for turning movements. This extended theory should also apply to unmarked freeway sections upstream of diverges if it turns out that drivers segregate themselves by lane depending on their destination.

These generalized KW models include the possibility of a semi-congested (“2-pipe”) state where some lanes are queued and others are not. Weak experimental evidence of this phenomenon has been provided by Lawson et al. (1999), who showed that the shoulder lanes upstream of a congested diverge exhibited higher occupancies than the median lanes. These authors also noted that the higher occupancies did not remain confined to the right lanes as the queue grew further upstream. This suggests that the freeway section upstream of an off-ramp cannot be modeled with the extended KW theory as if it was homogeneous, but that an approximation where exiting vehicles are restricted to the n lanes closest to the shoulder may be reasonable if n is assumed to increase with the distance from the off-ramp. This and other related properties of the traffic stream are the focus of this report.

The six main theoretical findings of this report are: (a) showing that close to the congested off-ramp disturbances in vehicle counts propagate upstream in the shoulder lanes and downstream in the median lanes (this confirms that the Lawson et al. (1999)

observations indeed correspond to a semi-congested state); (b) showing that semi-congested states did not occur (on the regular lanes) beyond the station that was 2.2 km upstream of the off-ramp and that in this region the total freeway flow obeyed the simplest version of the KW model; (c) measuring the size and structure of the transition zone (“shock”) that marks the end of the simple KW queue at these upstream locations; (d) showing that the bottleneck flow can change for no observable reason, most likely due to changes in the O/D table, as predicted in FIFO models; (e) showing that the changes in O/D table can affect the structure of the queue; and (f) identifying a non-recurrent moving bottleneck in free-flow traffic, and offering the most detailed characterization of any such event available on record with a new visualization tool.

The report is organized as follows: Section 2 discusses the diagnostic tools. Section 3 describes the site, the data set (which encompassed the complete evening rush from the onset of congestion to its final dissipation) and what generally happened during the observation period. Section 4 examines the uncongested period before the off-ramp spills over, including the effects of a moving bottleneck. Section 5 analyzes the steady state period immediately after the queue has filled the study section. Sections 6 and 7 examine respectively, the onset of queuing and the queue dissipation process. Section 8 concludes the report by proposing general remedial measures for similar diverge bottlenecks, and then assessing the potential benefit of these measures for the site in question

2. DIAGNOSTIC TOOLS

This section describes two methodologies that will be used in this report and its sequel. Subsection 2.1 presents a visualization tool for traffic data that is related to the

rescaled cumulative plots of Cassidy and Windover (1995), and Sec. 2.2 describes three numerical approaches for comparing time-series of vehicle counts. The visualization tool, like a microscope, can help analysts determine what is going on in a traffic stream that is acting peculiarly by revealing its fine detail. The numerical methods can help analysts determine the direction and speed of traffic waves when they are so weak (due to lane changing, for example) that visualization does not work.

2.1 Oblique plots

Perhaps the most informative way of displaying traffic data is by means of the so-called “N-curves of cumulative vehicle count,” or “N-curves” for short, that depict the cumulative number of vehicles that have passed over a detector versus time, t . In this report, the curve for detector i will be denoted N_i , and the associated non-decreasing mathematical function $N_i(t)$. Figure 1a shows two hypothetical N-curves labeled by location, $i = \text{'Upstream'}$ and $i = \text{'Downstream'}$.

The defining characteristic of N-curves, as opposed to other cumulative plots, is that the vertical position of each curve on the graph is chosen so as to ensure that the vertical separation between any pair of curves at some time $t = t_0$, $|N_i(t_0) - N_j(t_0)|$, is the actual vehicular accumulation between the two corresponding detectors. This relative positioning is very useful when vehicles are conserved between detectors because then the vehicular accumulation between detectors *at any time* is still given by the vertical separation between curves. Furthermore, assuming still that vehicles are conserved, horizontal separations between N-curves are “space-mean” trip times (i.e., the trip times of “average” vehicles that are passed by as many vehicles as they pass) and the area between curves is the

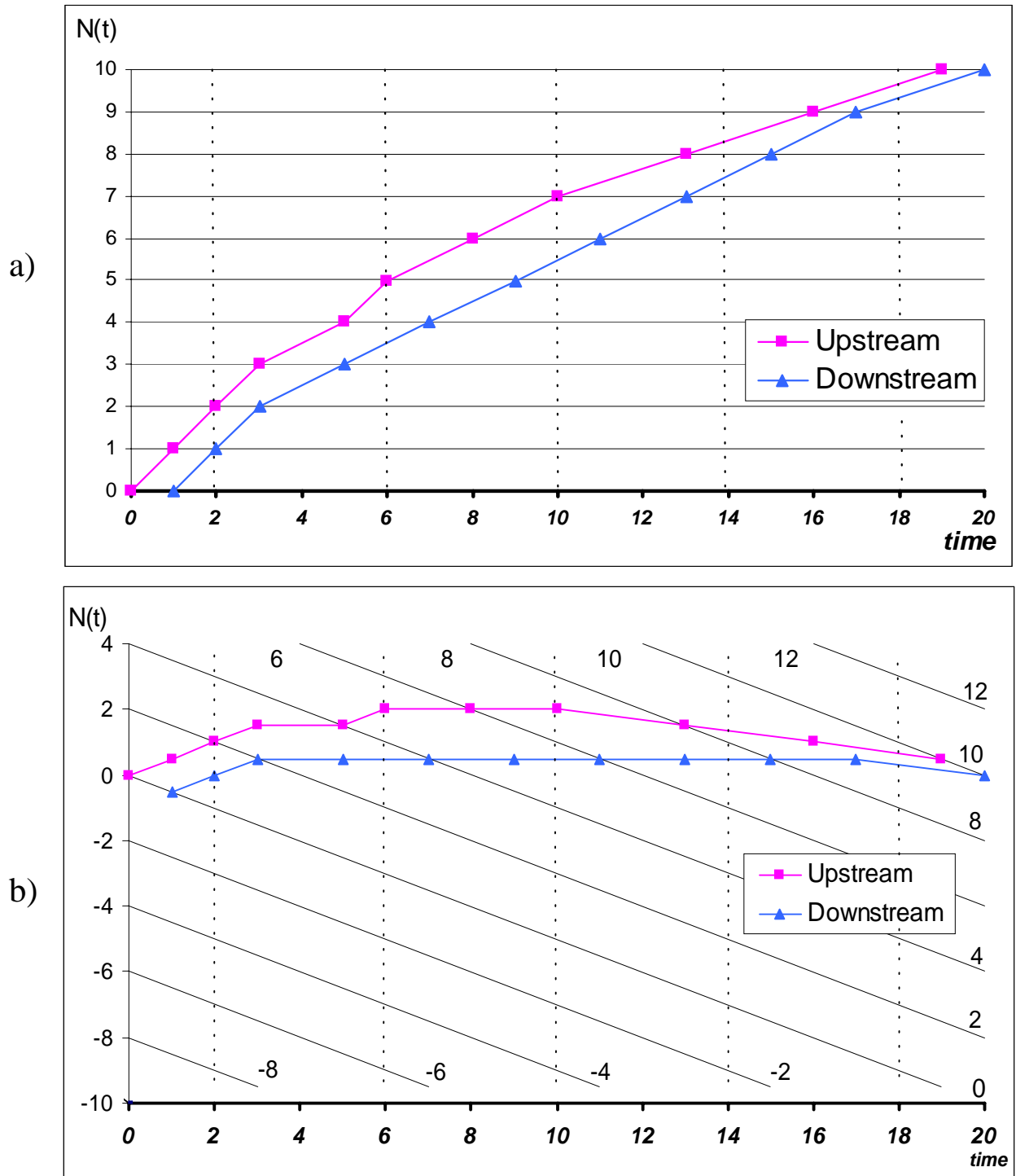


Figure 1. Arrival and departure curves: (a) orthogonal coordinate system; (b) oblique system

accumulated travel time between detectors. More information about N-curves can be found in introductory books such as Newell (1982) and Daganzo (1997a).

One disadvantage of “raw” N-curves is that when one looks at real data sets spanning long periods of time and many vehicles, the scale of the axes has to be so large that one cannot see important details. Cassidy and Windover (1995) proposed subtracting a “background flow”, q_b , from all the data and then plotting N-curves for the modified data; i.e., plotting the functions: $N_i(t) - (t-t_0)q_b$. The range of the count variable for the modified N-curves can be reduced in this way by orders of magnitude with judicious choice of q_b . This trick restores desired detail to the N-plots without changing the vertical separation between curves², which continues to represent accumulation. However, as mentioned in Cassidy and Windover (1995), horizontal distance between curves no longer represents travel time so that some information may be lost with the transformation. The rest of this subsection shows how this information can be restored.

It should be obvious that if one takes the data for two or more N-curves, e.g., the data of Fig. 1a, and plots them on an oblique coordinate system such as the one of Fig. 1b, no information is lost. [An oblique coordinate system is defined in the standard way, by two families of individually labeled, parallel lines which are either vertical or slanted in this report. The coordinates of a point are simply the labels of the two lines passing through the point. In Fig. 1b, and elsewhere in this paper, coordinate labels will be shown in boldface for vertical (time) lines and as an ordinary font for slanted (number) lines. The reader is invited to verify that the pairs of N-curves in Figs. 1a and 1b are graphs of the same two functions. Clearly, since no information is lost by a change of coordinates,

² Obviously, since: $[N_i(t) - q_b(t-t_0)] - [N_j(t) - q_b(t-t_0)] = N_i(t) - N_j(t)$.

anything whatsoever that could be found in a standard plot will also be found in an oblique plot.]

The advantage of oblique plots is that they magnify angles in the N-curves and in this way allow an analyst to see subtle things, such as changing trip times and platoon formation in unqueued traffic, without losing any information. (By changing the angle of the abscissa axis one can obtain the desired level of magnification.)

The reader is now invited to verify that if one redraws the curves of Fig. 1a after subtracting from the data a background flow equal to the negative slope of the abscissas axis of Fig. 1b (i.e., one plots the Cassidy-Windover curves) then the resulting curves are identical to those of Fig. 1b. Consideration shows that this is always true. Therefore, if one overlays a family of slanted lines with slope equal to the negative background flow on a Cassidy-Windover (CW) plot, and then labels the slanted lines (e.g., along the edges of the plot as suggested above), one has converted the picture into a full-information oblique plot.

2.2 Estimation of wave speed

Visual inspection of oblique (or CW) plots often reveals similarities in the shapes of two adjacent N-curves. This happens for example in congested freeway sections away from diverges, where visual inspection of said plots clearly reveals (Windover, 1998) that the N-curve patterns propagate in the upstream direction approximately as predicted in KW theory; see Newell (1993). [Recall that the horizontal (time) separation between similar patterns is the kinematic wave trip time.]

If the propagation of kinematic effects is masked by other phenomena such as

lane-changing and freeway inhomogeneities, numerical methods may still be able to discern weak similarities that may escape the eye. The numerical approach has been used to compare time-series of ordinary (non-cumulative) vehicle counts by various authors. For example, Mika et al. (1969) had shown, in qualitative agreement with the more recent findings of Windover (1998), that the cross-correlation of the two time-series of vehicle counts recorded at two contiguous detector stations in congested traffic reaches a maximum when one shifts one of the time series with a time lag corresponding to a wave velocity of approximately -16 mi/hr.

Since a similar cross-correlation approach can be used with cumulative curves and also with cumulative curves where a background flow has been subtracted, one is justified in asking about the relative merits of the three approaches. Appendix A presents specific recipes and compares the three approaches for uncongested traffic. It shows in particular that the two cumulative-count methods always yield virtually identical wave velocities, and that they estimate the 'free-flow wave speed' more efficiently than the ordinary count method. Therefore, cumulative-count methods will be used in this report for uncongested traffic.

[The appendix also warns that, despite their equivalence, the two cumulative-count methods will always yield a different cross-correlation for the same data because their respective pairs of time-series increase with time at different rates. Therefore, it should be stressed that cross-correlations should only be used to compare wave speeds within a scenario and not to compare scenarios with different flows.]

The statistical superiority of cumulative count methods over simple count methods does not necessarily extend to other scenarios (e.g., with congested traffic and lane-

changing). Nonetheless, methods based on cumulative counts are appealing because these curves contain traffic features that are of ultimate interest. Therefore, these methods will continue to be used for congested conditions, but will be complemented with ordinary count methods if the results are not conclusive.

3. DATA ANALYSIS

Geometry: Previously examined in Lawson et al. (1999), the study site is a section of U.S. Interstate Freeway I-880 (northbound), directly upstream of the connecting off-ramp with freeway I-238. A diagram is shown in Figure 2a. Lanes are numbered from 1 to 5 starting with the median lane, which is a high occupancy vehicle (HOV) lane. Detector station labels are shown along the bottom of the figure. They differ from those used by Caltrans and Lawson et al. (1999). The labels used in this report express the approximate distance between each station and the I-238 off-ramp in multiples of 100 meters. (For example, the first station from the left which is 2705 meters away from station 0 is labeled "27".) The HOV lane ends between stations 22 and 17. The HOV designation is in force between 15:00 hrs and 19:00 hrs.

Traffic data: The site is interesting because a troublesome queue starts to backup from the I-238 off-ramp every weekday sometime between 14:30 and 15:00. The queue eventually grows and fills all the non-HOV lanes, disrupting through flow in a big way. The resulting main-line congestion lasts until 18:00 and beyond. The *Freeway Service Patrol* data set (Skabardonis et al., 1994) was used for our study because of its fine level of detail. The data set includes 2-second counts and 2-minute occupancies at every station and lane (except for lane 2 of station 17, which had a malfunctioning detector) for a typical day in

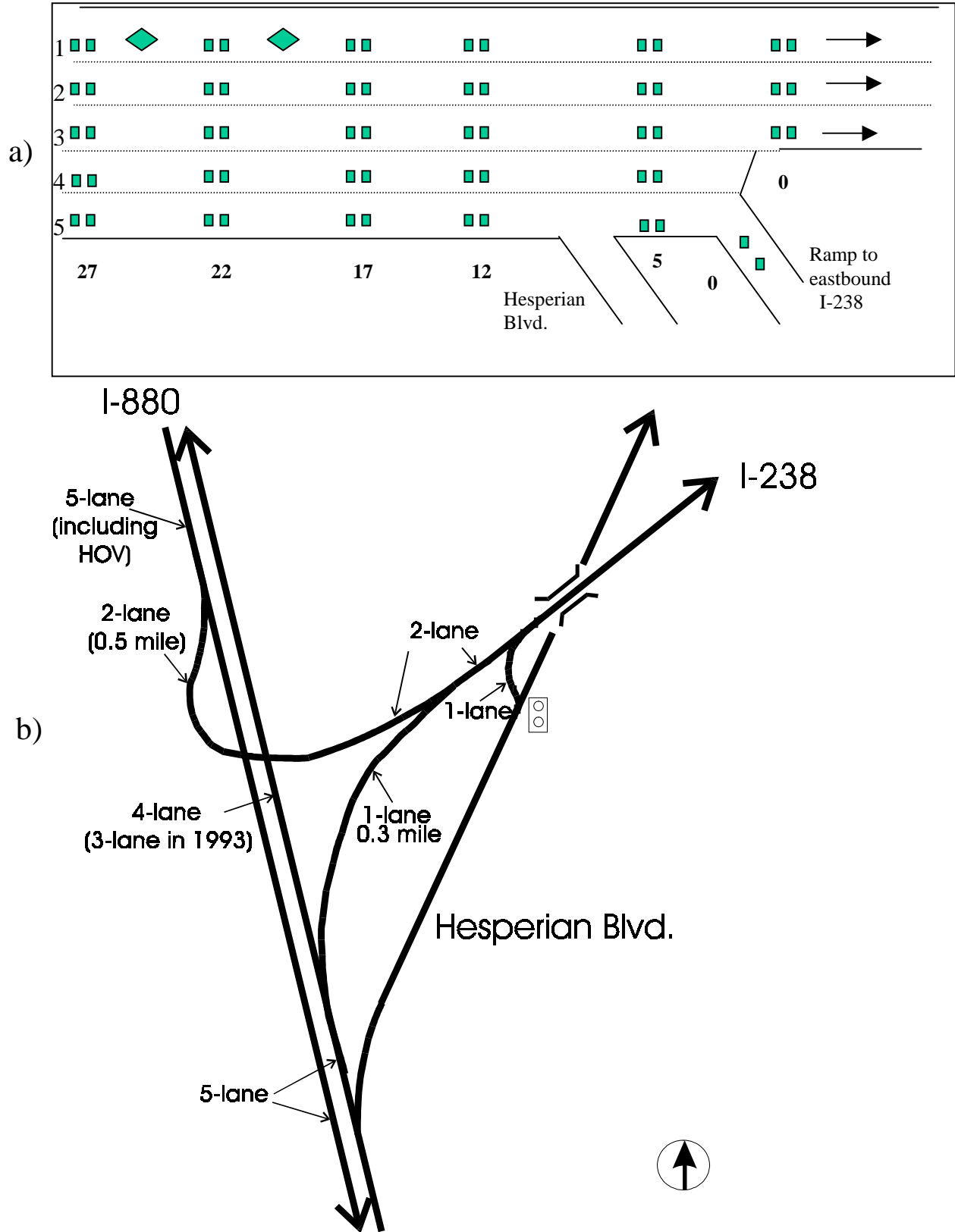


Figure 2. I-880, northbound: (a) site geometry (not to scale); (b) schematic map of the site.

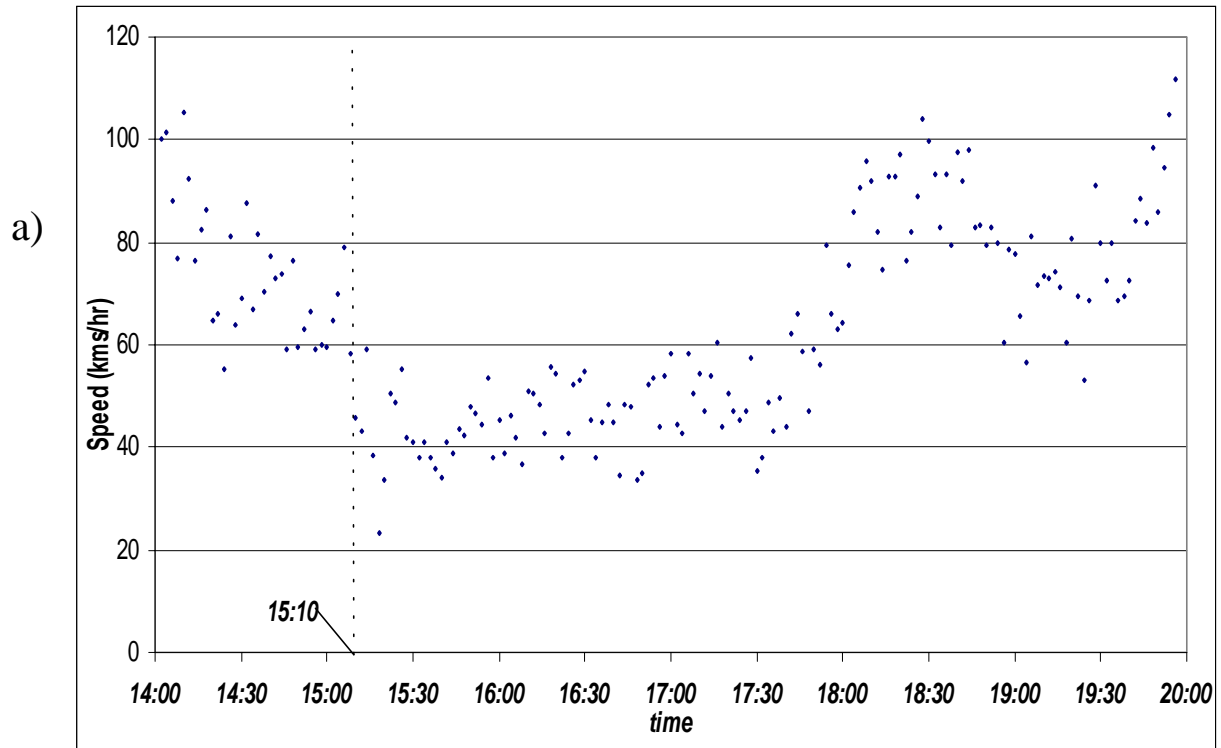
1993. The available information was treated as described below, in order to recover from it pertinent speed- and N-curve information.

Speeds: As is usual, the flow-occupancy ratio for each detector was multiplied by a suitable constant to convert it into space-mean speed. (The constants represent the space-mean vehicle length.) They were chosen so as to force a match between: (a) the predicted average speed for each detector during the uncongested interval from 14:00 hrs to 14:10 hrs, and (b) the average speed across all lanes observed for the same time-of-day interval during a recent field trip. This speed was found to be 108 km/hr.

Figure 3a shows the time-series of speed obtained from the detector on lane 3 of station 12. Note how the speeds drop gradually but substantially after 14:00 hrs and then stabilize at a lower value. This is estimated to occur at around 15:10 hrs. A similar pattern, with a gradual drop and subsequent stabilization, was observed at the remaining detectors. Part b of the figure shows the estimated times when the stabilization occurs at all the locations where data were available. The picture indicates that a queue started from within the on-ramp at 14:50 hrs, grew upstream, and about 45 minutes later reached all the lanes of station 27 simultaneously (at 15:38 hrs). From this time on, the speeds at all the locations remained approximately constant for just over 40 minutes. Appendix B contains the data.

N-curves: To construct the N-curves, the detector counts had to be corrected for systematic bias and random drift, and the detector clocks had to be synchronized.

Since vehicles should be conserved across stations, a correction for drift and bias was defined for the *station* counts. Its objective was ensuring that all detector stations counted the same number of vehicles during the study period (from 14:00 hrs to 20:00 hrs)



b)

1	15:36	15:18	15:10	15:16	15:24	
2	15:38	15:18	N/A	15:16	15:22	
3	15:38	15:34	15:16	15:10	Not clear	
4	15:38	15:30	15:10	15:02	15:00	0
5	15:38	15:22	15:10	14:58	14:52	
	27	22	17	12	5	0

14:50

Figure 3. Speed conditions on the site: (a) time series of the 2-minute average speeds recorded at station 12, lane 3; (b) times of day (hr:min) when the time-series of speeds at all locations first settle at lower level.

and then distributing the adjustments as evenly as possible. The overall correction factor for station “i” is:

$$F_i = [C_r(20)-C_r(14)]/[C_i(20)-C_i(14)] \quad \text{for } i \neq r, \quad (1)$$

where $C_i(t)$ is the recorded cumulative count of station “i” at time t, in hours, and “r” is a reference station that is not adjusted. We used $r = 12$. The same correction factor was assumed to hold during the whole study period ($14 < t < 20$), as if there was a systematic counting bias. Therefore the formula for the corrected station counts is:

$$C'_i(t) = C_i(14) + F_i [C_i(t)-C_i(14)] \quad \text{for } i \neq r. \quad (2)$$

Counts for individual detectors were corrected using the factors for their station. The factors for stations 22 and 27 were very close to 1. Since flow was not conserved between stations 5 and 12, the factors applied to stations 0 and 5 had to be about 10% greater than one.³ The use of time-independent factors assumes that the fraction of flow exiting at Hesperian was constant throughout the study period.⁴

To correct for random drift, the trip times observed from the curves were checked for consistency with the average speeds that were recorded. Surprisingly, no corrections were necessary, except a very minor one for station 22, which appeared to have drifted by

³ Such an adjustment has no effect on vehicle or wave trip times, but should be taken into account when evaluating vehicular accumulations downstream of station 12.

⁴ More precise adjustments were not made because the Hesperian Blvd off-ramp was not instrumented. The resulting imprecision does not invalidate comparisons among the N-data of stations 12 and above (and 5 and below).

about 30 vehicles by 18:00 hrs. (This is a tribute to the reliability of the special detection equipment that was installed at the site.).

Station clocks were synchronized in two steps: first, using method (b) of Appendix A to determine the *unsynchronized* time lag that best fits the free flow data from 14:00 hrs to 14:10 hrs, and then adjusting the clocks so that the adjusted time lags match the known trip times during this time (i.e., corresponding to a space-mean speed of 108 km/hr.)

Figure 4 summarizes the result of these procedures. It displays oblique N-curves for stations 0, 5, 12, 22 and 27 (lanes 1-5, combined) with a background flow of 6930 veh/hr. Note how traffic is in stationary free-flow until a little after 14:45 hrs, how the delays increase (the queue grows) after that time from station to station, how the system reaches another period of stationary (*but queued*) flow from a little before 15:40 hrs to 16:20 hrs, and how the queue dissipation process begins at the upstream detectors (around 17:30

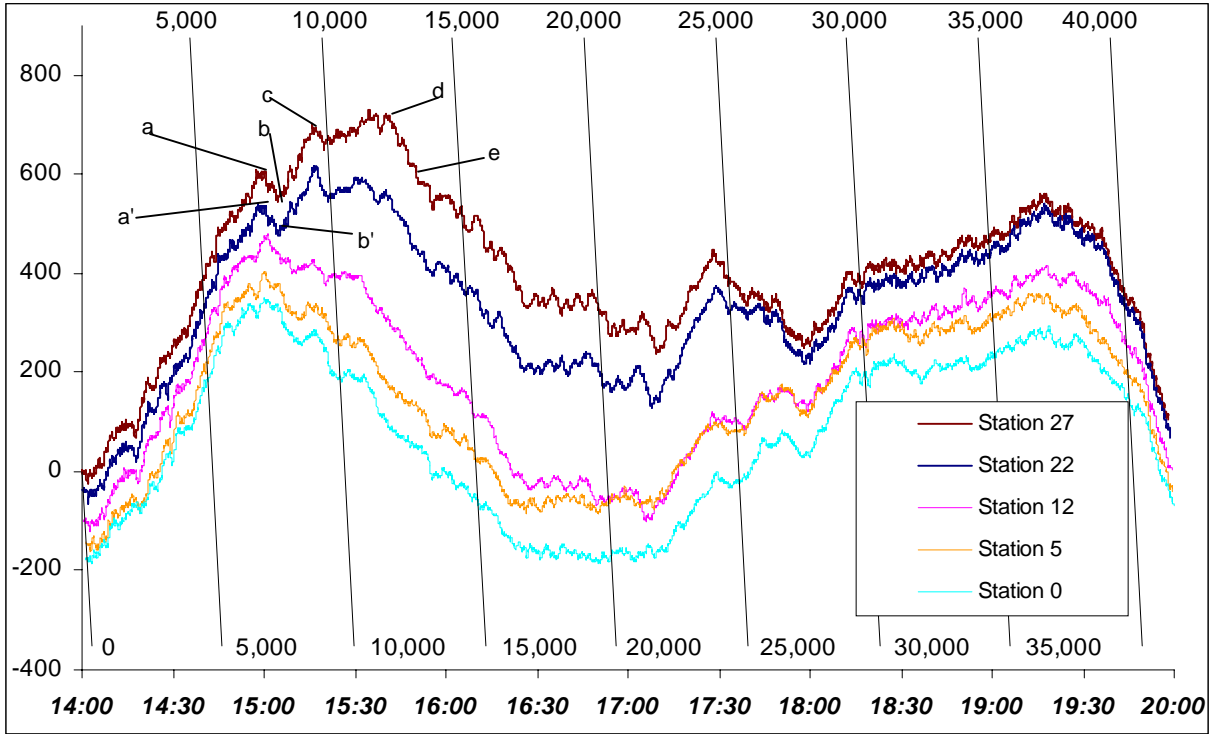


Figure 4. Oblique plot with N-curves for all lanes at all the stations for the whole study period (background flow = 6,930 vph)

hrs) because the demand subsides. Inspection of the figure also reveals two surprises: (i) that traffic flow increases considerably after 16:20 hrs although the queues continue, and (ii) that the curves of detector stations 5 and 12 intersect (and/or may be closer than they should be) between 16:20 and 19:00 hrs.

The second surprise is likely due to variations in the flow leaving on the Hesperian Blvd off-ramp between the two stations. Note that Eq. (2) would have preserved the correct trip times between stations 12 and 5 at all times if the fraction of exiting vehicles had remained constant during the study period, because in this case the pattern of exits has the exact same draining effect on the counts of station 5 as the systematic counting bias that is eliminated by Eq. (2). Because the curves are infeasibly close to each other between 16:20 hrs and 19:00 hrs (curve N_5 is too high), this suggests that the fraction of vehicles leaving on the Hesperian off-ramp during this time is smaller than at other times⁵. Surprise (i) is explained in Sec. 7.

Sections 4 and 5 of the report examine in detail the two periods of uncongested and congested stationary flow occurring from 14:00 to 14:45 hrs and from 15:40 to 16:20 hrs, respectively, including the dynamics of a moving bottleneck that occurred during the uncongested period. Sections 6 and 7 later focus on the regime transitions in and out of congestion.

4. THE INITIAL UNCONGESTED PERIOD (14:00 hrs -- 14:45 hrs)

Examination of the uncongested steady state period before congestion begins does

⁵ When the speed on the exit lanes is low, as occurs prior to 16:20 hrs, drivers can bypass the off-ramp congestion by using the Hesperian exit; see Sec. 8. The availability of this shortcut may explain the higher flows at the Hesperian exit prior to 16:20 hrs.

not reveal anything that contradicts current understanding of free-flow traffic conditions; in particular, the presence of the diverge does not seem to disrupt uncongested traffic in a significant way. Figure 5a, for example (an oblique plot for stations 12 and 27 with background flow, 6930 veh/hr) shows that disturbances in count propagate forward with the traffic stream and that trip times remain quite constant despite the fluctuations in count. The close match is shown more clearly in part b of the figure, where curve N_{27} has been translated toward N_{12} along the slanted lines by the trip time estimated with method (b) of Appendix A. The output of the estimation procedure is displayed in Fig. 5c, where one can see that $v^*_s = 100 \text{ km/hr}$ ⁶.

Although the two curves of Fig. 5b show a large discrepancy around 14:28 hrs, this is not grounds for questioning the theory since, as is shown below, oblique plots establish that the discrepancy is due to a non-recurring event, which can be characterized in detail.

The discrepancy is first magnified by means of a smaller scale plot, as shown in Fig. 6a. This figure also displays the intermediate N-curve corresponding to station 22. The new figure shows that the discrepancy corresponds to a 30-sec period of very low flow. A detector malfunction was ruled out because a similar but shorter-lived effect was found at the intermediate detector, and because the cumulative count later recovered to the “correct” level.

A sudden decrease in demand was also ruled out because demand fluctuations should propagate across detectors with little change, but this was not the case here.

⁶ This estimate is lower than the free flow speed estimate used to synchronize the clocks perhaps because the latter was obtained during the early part of the period, from 14:00 hrs to 14:10 hrs, when the flow was lower.

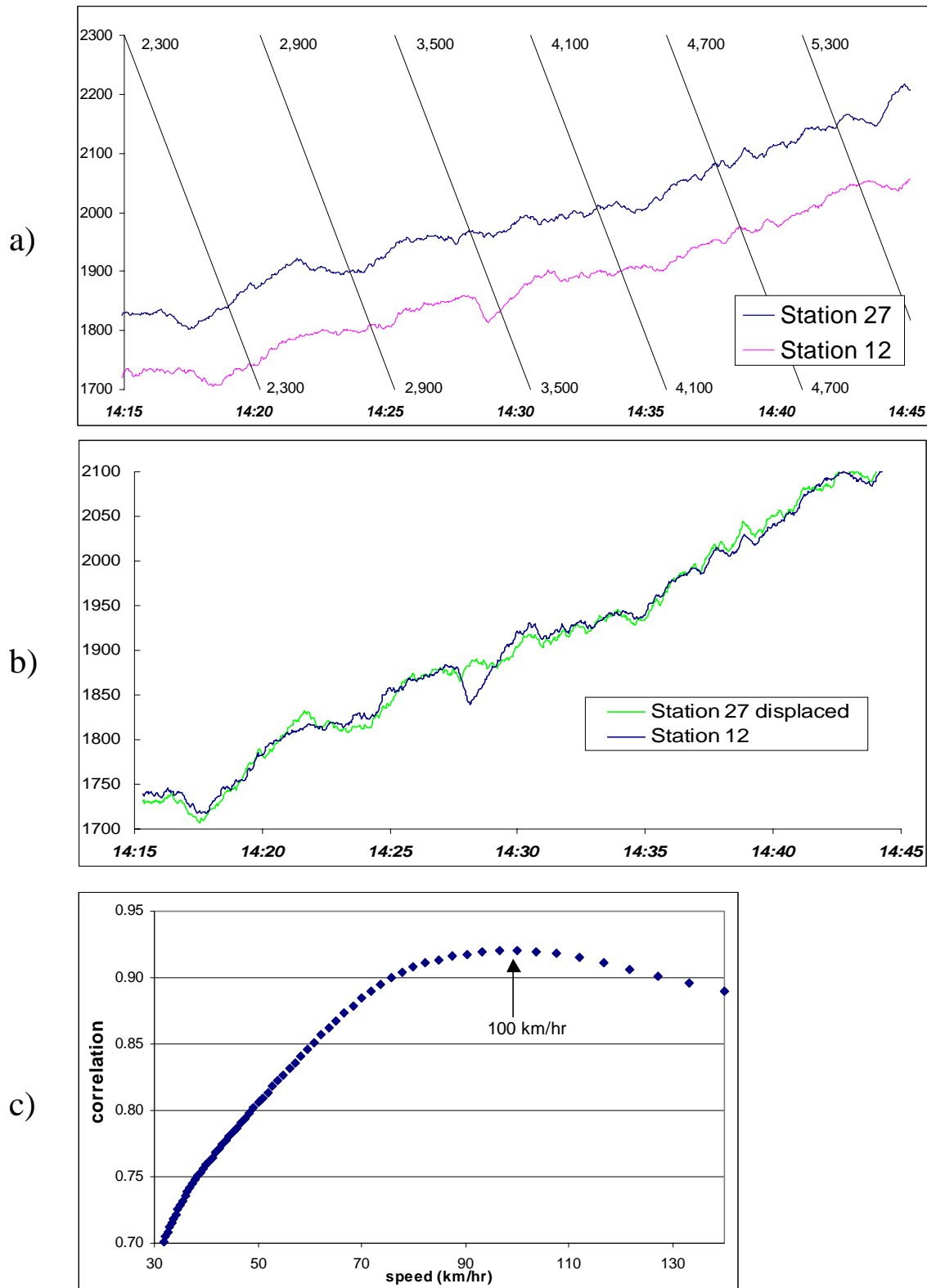


Figure 5. Traffic conditions at stations 12 and 27 between 14:15 and 14:45 hrs. (a) Oblique N-plot (background flow = 6930 vph); (b) shifted N-plot with N27 displaced along the oblique coordinates toward N12 (shift corresponds to a speed of 100 km/hr); (c) correlation between shifted curves for different speeds.

The most likely explanation for the observations is that a set of slow moving vehicles entered the road somewhere upstream of station 27 (for $N \approx 3400$) and they hampered the freeway in a way that only a small flow was allowed to pass. The position of this “moving bottleneck” in each one of the curves is the bottom of the “V”. The slope of left side of the “V” is the flow that gets past the bottleneck, and the slope of the right side the flow inside the moving queue directly upstream of the bottleneck.

Examination of the figure reveals that the bottleneck speed was 72 km/hr, and that its speed was the same between stations. We can also see (by extrapolating the length of the lull in the upstream direction) that the incident must have started approximately 3.0 km upstream of station 0.

The flow downstream of the incident is estimated by the slope of the left side of the “V”, in the oblique coordinate system. The flows are 1500 veh/hr at station 22 and 1800 veh/hr at station 12. This suggests that the incident blocked all the lanes except one.

The rate at which vehicles pass through the bottleneck (in a frame of reference attached to the bottleneck) can also be estimated from the figure since it is the rate at which the N-coordinate of the bottom of the “V” increases with time. Note that “N” increases by about 10 veh (+/- 2 veh) between stations 27 and 12, and that it does so in 75 sec. Hence, the passing rate is about $(10/75)(3600) = 480$ veh/hr (+/- 100 veh/hr)⁷.

The flow upstream of the incident is much larger: 10740 veh/hr at station 22, and 9540 veh/hr at station 12, where the queue is more fully developed. Note that the first

⁷ The passing rate could also be obtained without the figure, since vehicle conservation implies that the flow through the bottleneck (in a frame of reference moving with the bottleneck) only depends on the flow and speed prevailing on one of its sides and the speed of the bottleneck. Reassuringly, the result of this calculation is: $1800(100-72)/100 = 504$ veh/hr.

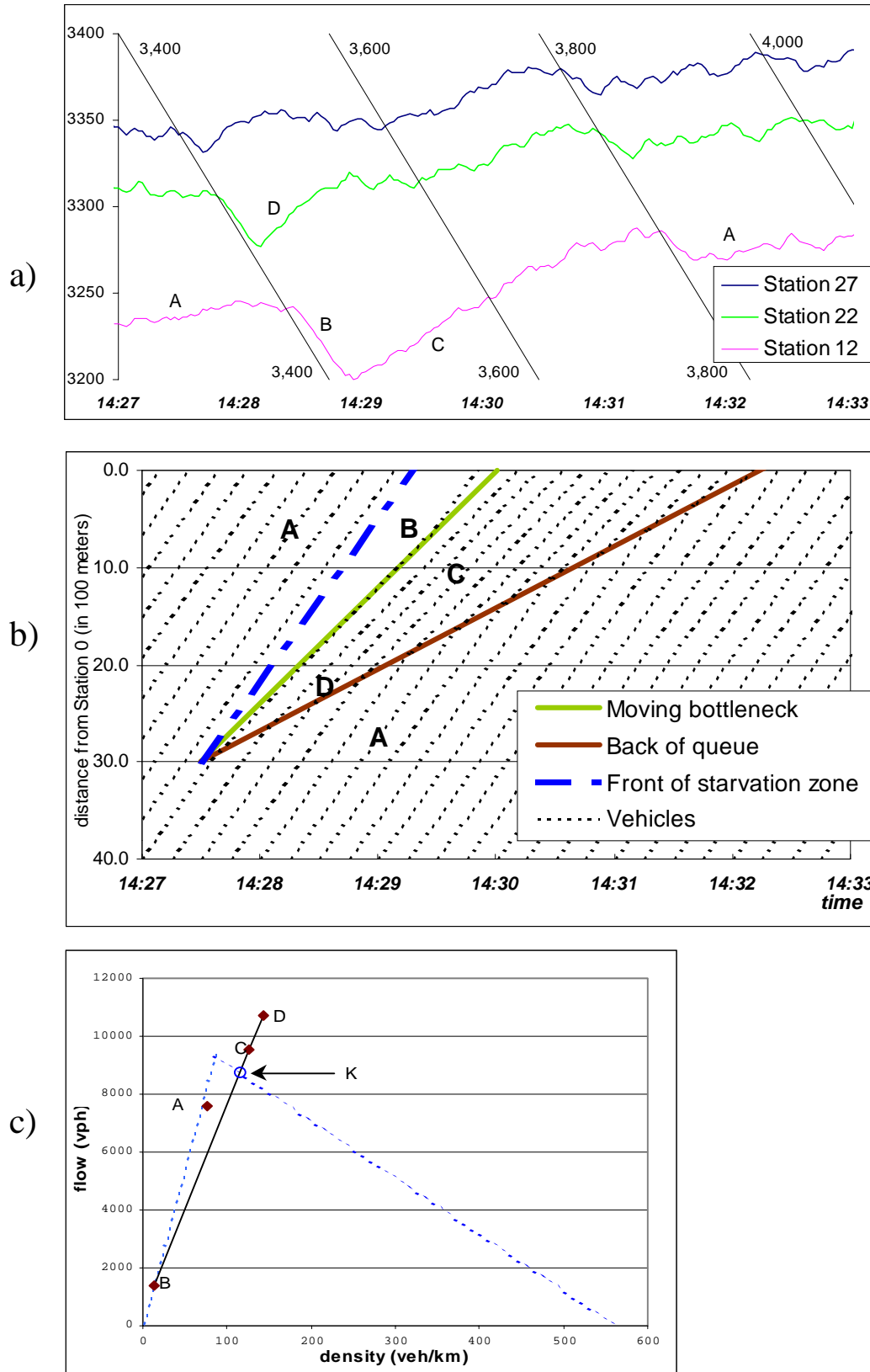


Figure 6. Traffic patterns around a moving bottleneck: (a) N-curve signature detected at stations 27, 22 and 12; (b) evolution in time and space; (c) flow and density states.

value is very high (for a 5 lane freeway) and that the two flow levels are experienced by a single set of drivers, $N \in (3430, 3500)$, which were in the queue at both locations. This suggests that *the differences in flow are not caused by driver differences, but by a change in driver behavior*. The data indicate that the first hundred drivers pack themselves closely upon joining the queue and later relax. The data also show that following drivers do not act in this way, since higher numbered drivers (up to $N= 3750$) never experience the high flows. This differential behavior is probably caused by proximity to the bottleneck when vehicles join the queue. We speculate that the first drivers may be adopting very short spacings in the hope of getting through the bottleneck quickly, and that their motivation disappears once they realize that they will be queued for a while⁸. Later drivers act differently because they may not be able to see the bottleneck and do not see a benefit from following aggressively.

Therefore, the platoon flow seen at station 12 (about 9500 veh/hr) is more likely to be sustained at succeeding locations. This flow is close to the “capacity” of a 5-lane freeway. Near-capacity flows would also be expected for an obstruction that moves with a speed close but a little smaller than the “optimum” speed⁹.

The regime transition corresponding to the back of the moving queue can also be identified in Fig. 6a by the upper right side of the “V’s”. Although the queue grows in length, it does not grow fast enough to overcome the speed of the bottleneck and

⁸ It is not unusual to see very high flows at locations where drivers can be motivated to follow closely for brief periods of time; e.g., the short weaving areas of clover-leaf interchanges, where flows in excess of 3000 veh/hr per lane have been observed (Brilon, 1998)

⁹ See Newell (1993a) and Sec. 4.4.5 in Daganzo (1997a) for more details about moving bottlenecks

propagate upstream; thus, the back of the queue propagates forward and passes all the observers. Note that it passes station 27 at about 14:28 hrs and station 12 around 14:30:30 hrs. Thus, it takes about 2.5 minutes to travel 1.5 km, yielding a forward speed of approximately 36 km/hr. Forward propagation is logical since the upstream flow is lower than the queued flow.

Figure 6 also displays the flow-density states that would be consistent with all this information, both, as part of a time-space diagram (in part b) and on an associated flow-density diagram (in part c). Figure 6b also shows the trajectories of several items: (i) the moving bottleneck, (ii) the back of the queue, (iii) the front of the starvation zone and (iv) some vehicles. This part of the figure also indicates the likely time and place where the bottleneck first emerged. The diagram is consistent with the data of Fig. 6a in the sense that the timing and magnitude of the regime transitions at the locations of the detectors coincide with the flow changes of Fig. 6a.

Figure 6c displays on the flow-density plane the states A, B and C that were observed at detector 12, and also the compressed state, D, observed at detector 22. The dotted line is the flow-density relation that fit the data best during steady state periods (see Sec. 5 and Fig. 9), and the open circle the queued state, K, that kinematic wave theory would have predicted instead of C or D.

Note how state D is only experienced at the outset and how the system quickly reaches a state C, which is close to K. Thus, except for the small difference between states C and K, this is the same diagram that would be obtained from kinematic wave theory. Therefore, despite what appear to be temporary changes in driver psychology, observations are roughly consistent with KW theory.

In summary, oblique plots have afforded us a detailed experimental view of a “moving bottleneck” and its curious properties; such plots can probably be quite useful for other applications too.

5. THE CONGESTED STEADY STATE PERIOD (15:40 hrs -- 16:20 hrs)

This section shows that during the congested steady state period between 15:40 hrs and 16:20 hrs, traffic upstream of section 22 can be modeled as a 1-pipe KW stream. It also shows that this is not the case closer to the bottleneck where waves are shown to travel in opposite directions in adjacent sets of lanes.

Figure 7 shows that conditions were indeed stationary between 15:40 and 16:20 hrs. The figure displays, both, the average speed during this period at every detector (in km/hr) and, parenthetically, the rate at which it changed (in km/hr²). These rates were estimated by the slope of linear regression lines for data such as that of Fig. 3a. Since the study period is $\frac{2}{3}$ hour long, the maximum deviation in the average speed from the value in the middle of the interval (in km/hr) is only $\frac{1}{3}$ of the parenthetical numbers. The maximum deviation is small at all the locations. The worst case, excluding HOV lane (1), was for station 12, lane 3. The data for this detector was already presented in Fig. 3a, which shows no significant trend in the relevant period (15:40 – 16:20 hrs). Therefore, we can safely assume that the traffic stream between stations 27 and 5 was in a steady-state from 15:40 hrs to 16:20 hrs.

Figure 7 also shows the distribution of average speeds over the study site. Note that the closer one is to the bottleneck the wider is the variation in speed across the non-HOV lanes. The spread is largest (76.1 km/hr) for station 5, and smallest (< 1

1	79.4 (20.6)	84.6 (8.2)	91.7 (10.4)	84.3 (7.3)	107.3 (4.8)	
2	25.4 (0.9)	32.2 (0.0)	N/A	58.4 (8.2)	92.1 (5.5)	
3	25.5 (-4.0)	31.6 (5.3)	39.7 (5.1)	45.1 (8.4)	67.4 (4.4)	
4	25.5 (-4.0)	27.8 (5.3)	31.4 (5.9)	26.4 (5.6)	39.2 (-1.0)	0
5	24.6 (-2.5)	9.8 (-1.3)	6.4 (-2.7)	6.4 (-2.5)	16.0 (-5.3)	
27	22	17	12	5	24.1 (-8.0)	0

Figure 7. Average speed in km/hr (and its rate of change in km/hr²) at different locations between 15:40 and 16:20 hrs.

km/hr) for station 27. The HOV lane remains largely unaffected by the queue. Note as well, that the speeds of lanes 2 and 3 decline in the upstream direction. This suggests that the queue does not span all the lanes close to the bottleneck, but that it does so by station 27. (Recall too from Fig. 3b that the drop in speed at this station was felt simultaneously at all lanes.) Since the distribution of speeds at station 22 is also fairly even, the above suggests that traffic may well behave as a 1-pipe KW stream between stations 22 and 27. This is checked below.

5.1. Traffic behavior upstream of station 22

A maximum cross-correlation of 0.53 was found between curves N_{22} and N_{27} with the 0-trend method of Appendix A for optimal horizontal and vertical shifts. The estimated wave velocity was -19.4 km/hr. Figure 8 shows oblique plots of the two N-curves before and after the shift. The good fit suggests that cumulative counts can be predicted using the 1-pipe KW model upstream of the 2 km mark. This distance may be different at other sites, and may also depend on the O/D mix within the queue.

Flow-density plots for station 27 (lanes 1-5, combined) are consistent with this. Figure 9a shows that despite the usual scatter, a line with a slope of -19.4 km/hr fits well through the 2-minute, dark triangles of our time period¹⁰. Recall too that the state observed upstream of the moving bottleneck (point C of Fig. 6c), which also lasted for about 2 minutes, is similarly close to the same line¹¹. Figure 9b shows that the scatter is considerably reduced when the aggregation interval is increased to 8 minutes. Similar

¹⁰ The white triangles do not fall on the line. They correspond to the period examined in Sec. 7, when traffic was not in a 1-pipe queued regime.

¹¹ Recall that the q-k diagram of Figures 9a, 9b and 6c are identical.

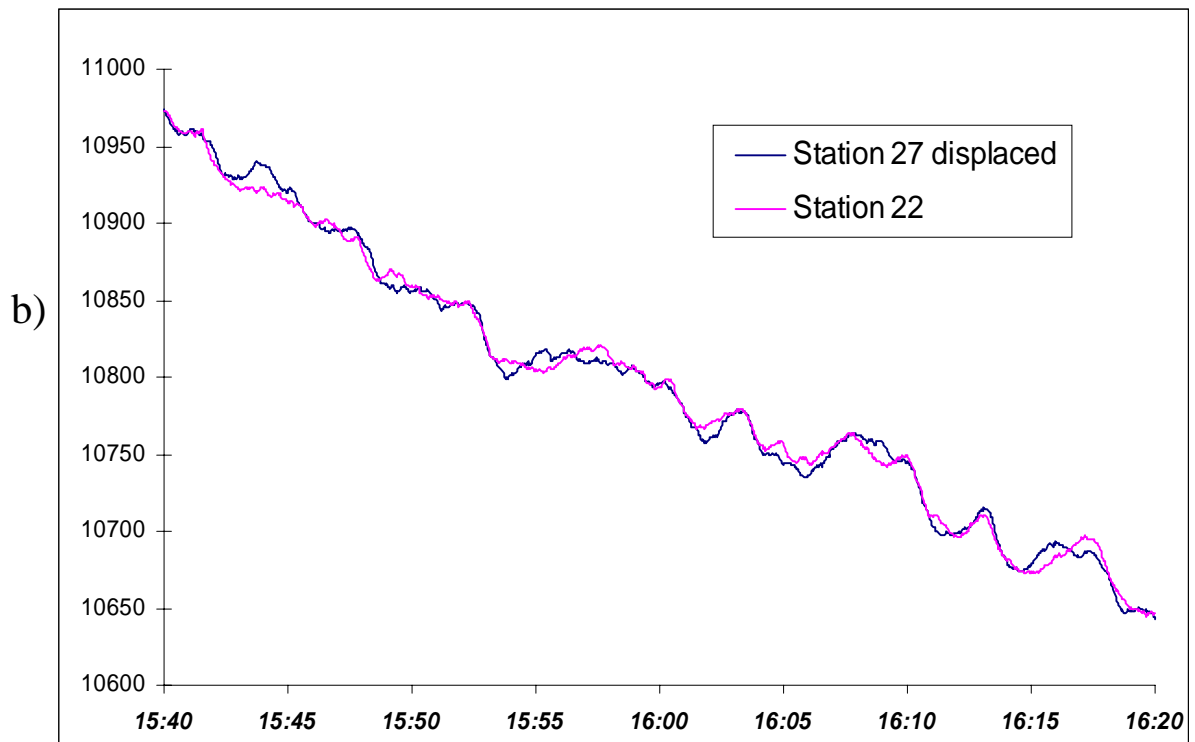
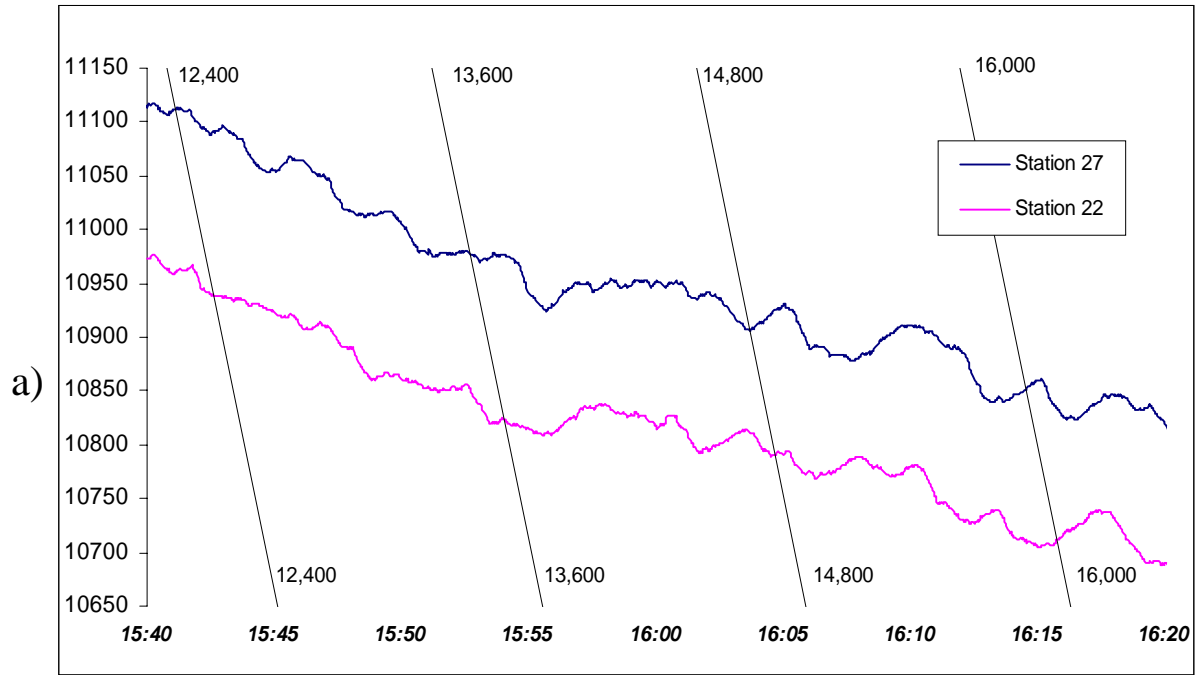


Figure 8. Oblique plots for stations 22 and 27 (without HOV) in the congested period between 15:40 and 16:20 hrs. (background flow of 6930 vph): (a) N-curves; (b) N-curves after the optimum shift.

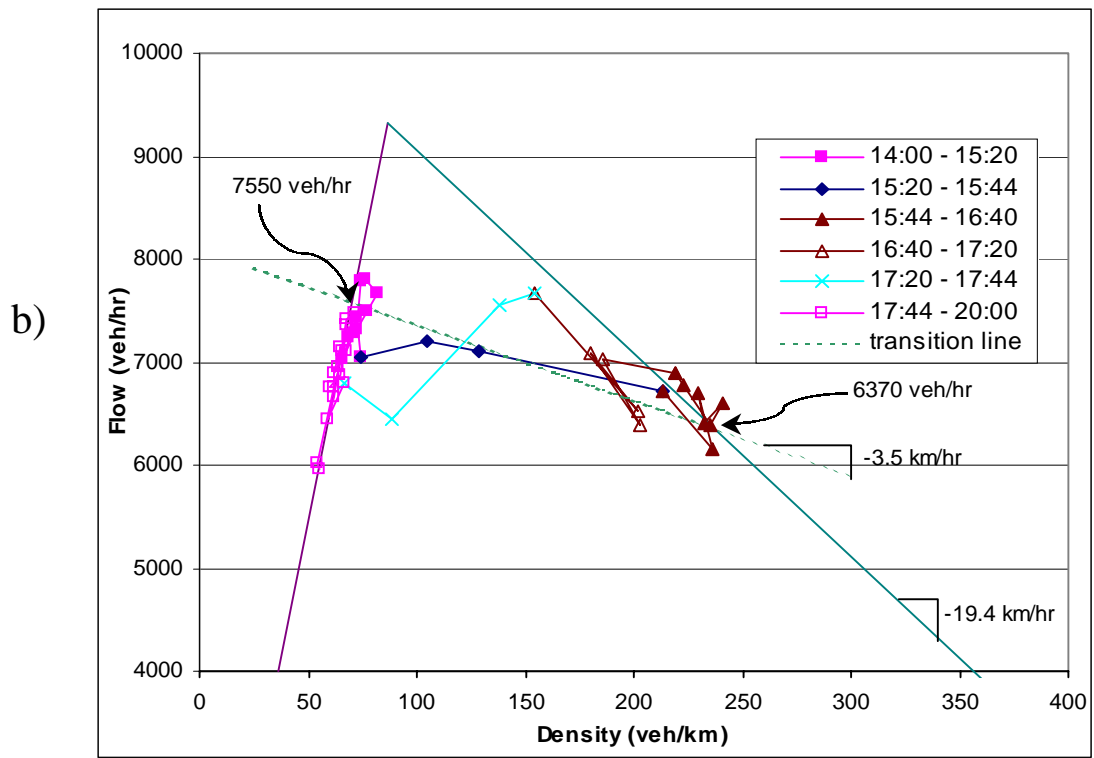
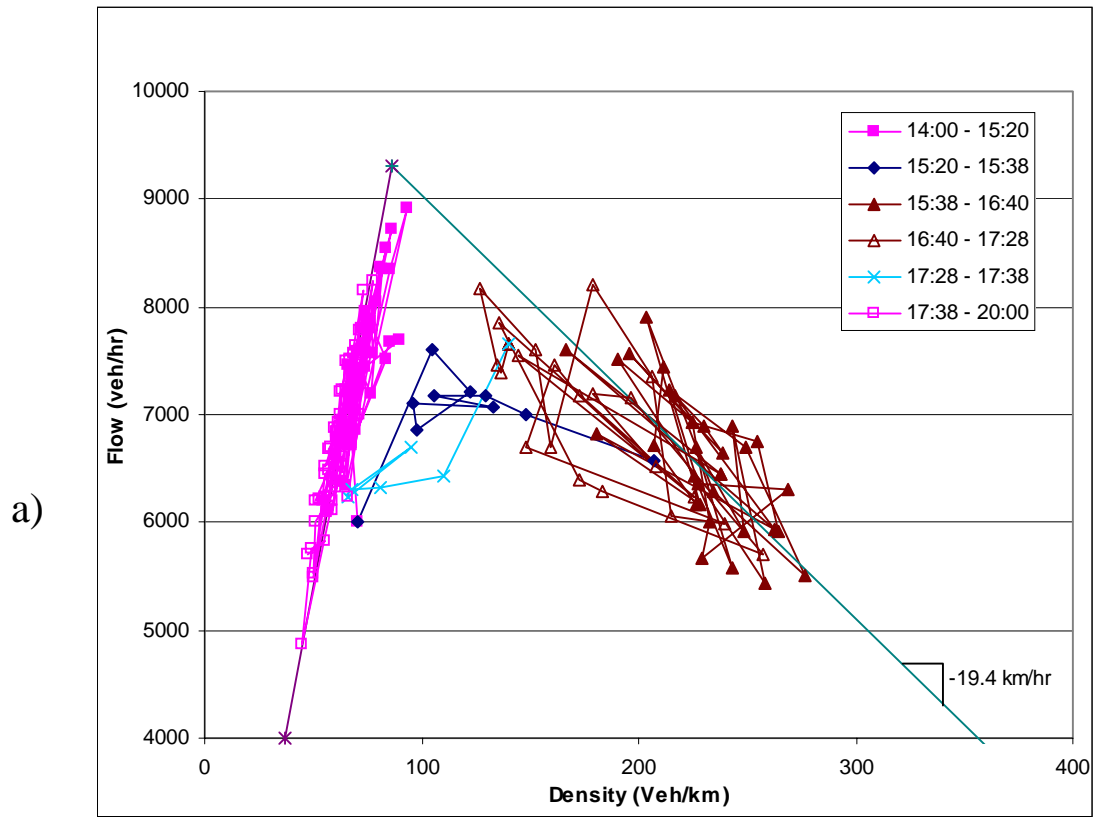


Figure 9. Scatter plots of flow v/s density at station 27 (with HOV lane) from 14:00 – 20:00 hrs: (a) 2-minute aggregation; (b) 8-min aggregation.

reductions in scatter were observed for individual lanes. Therefore, the scatter can be explained by driver differences¹².

5.2. Traffic behavior downstream of station 22

Because station 17 does not have a full complement of detectors, and an off-ramp exists between stations 5 and 12, we examine the section between stations 22 and 12. We suspect that traffic in this section is not moving as a 1-pipe stream because of the observed speed differences across lanes at station 12; see Fig. 7.

As one may expect, visual inspection of the N-curves revealed no clear way in which curve N_{12} should be displaced toward N_{22} in order to provide a best fit, and numerical that was obtained for different, negative wave velocities with the main two cross-correlation methods described in Appendix A; part (a) with cumulative counts, and part (b) with ordinary counts. The diagrams should be compared with the more peaked ones (Figs. 10c and 10d) obtained for the 1-pipe section between stations 27 and 22, that was analyzed in the previous section.

In view of this, we first looked for a semi-congested state, in which well-defined waves would be traveling in different directions on different sets of lanes between stations 22 and 12. To this end, lanes 1 and 2 were grouped together as one system, and lanes 4 and 5 as another. Lane 3 was not included in either of the systems to increase the difference in the two sets of average speeds¹³. Cross-correlation plots like those of Fig.

¹² Variations along the line are due to changes in speed.

¹³ Lane 1 was included as part of set 1 because the HOV lane ends between stations 22 and 17, and because excluding it from the set would imply too much non-conservation of flow.

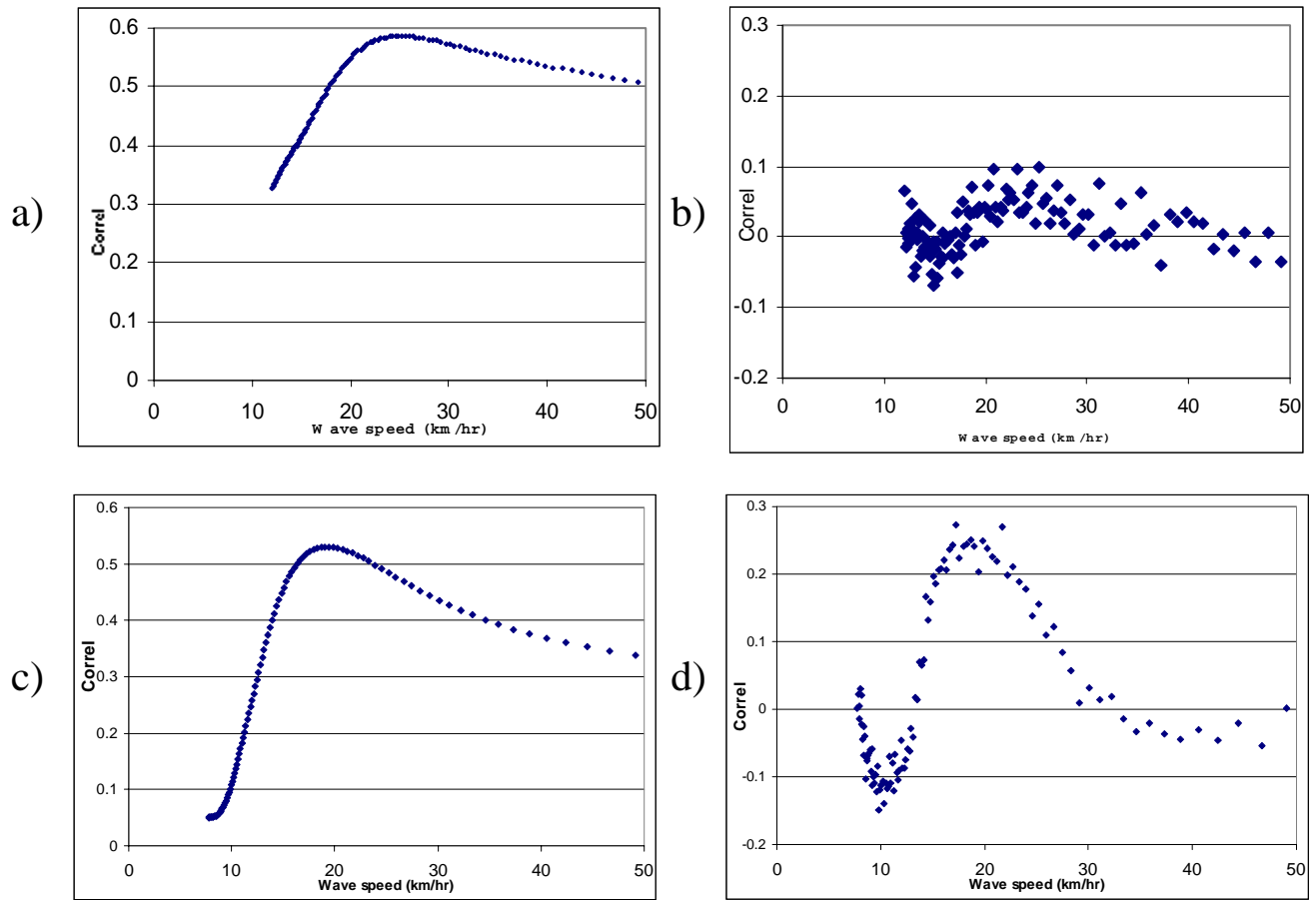


Figure 10. Cross-correlation of different time-series between 15:40 and 16:20 hrs. for different negative wave velocities (excluding the HOV lane): (a) zero-trend N-curves for stations 22 and 12; (b) counts for stations 22 and 12; (c) zero-trend N-curves for stations 27 and 22; (d) counts for stations 27 and 22.

10 for both positive and negative wave velocities were obtained for the two sets of lanes for different ten-minute intervals. We looked for different signs in the estimated wave velocities for the two sets of lanes, but no clear indication was found. We attributed this result to the low speeds in lane 2; see fig. 7. It appears that HOV vehicles are still crossing over in significant numbers between stations 22 and 12, which prevents forward wave motion by slowing down traffic on the median lanes. Thus, although speeds were lane-specific, their differences were not large enough to generate a semi-congested state.

Bi-directional wave propagation was found, however, for the time interval immediately following the end of our steady state period, when traffic speeds were higher. (Possible reasons for the speed increase are given in Sec. 7.) The best-fit wave velocities for this period (16:20 -- 16:30 hrs) were +56.6 km/hr for lanes (1,2), and -8.6 km/hr for lanes (4,5); see Fig. 11. Although the correlation curve for lanes (4,5) has another peak at -16.1 km/hr, this curve shows no evidence of forward-moving information. Thus, the two curves of Fig. 11 clearly establish the 2-pipe, semi-congested behavior of the system.

Note that information moved forward on lanes (1,2) at an average speed of 57 km/hr, which is just a bit lower than the estimated average speed of these lanes. These low speeds suggest that the 2-pipe unqueued flow on the median lanes may be modeled as suggested in Daganzo (1997) and Daganzo et al. (1997), but with a reduced speed limit.

6. THE TRANSITION PERIOD (14:45 – 15:40 hrs)

The transition between the unqueued and queued regimes was first felt around 14:45 hrs at station 5 and then with a delay at upstream stations. At stations 22 and 27 the changes took place around 15:30 PM, as highlighted for station 27 by points (a)-(e) of

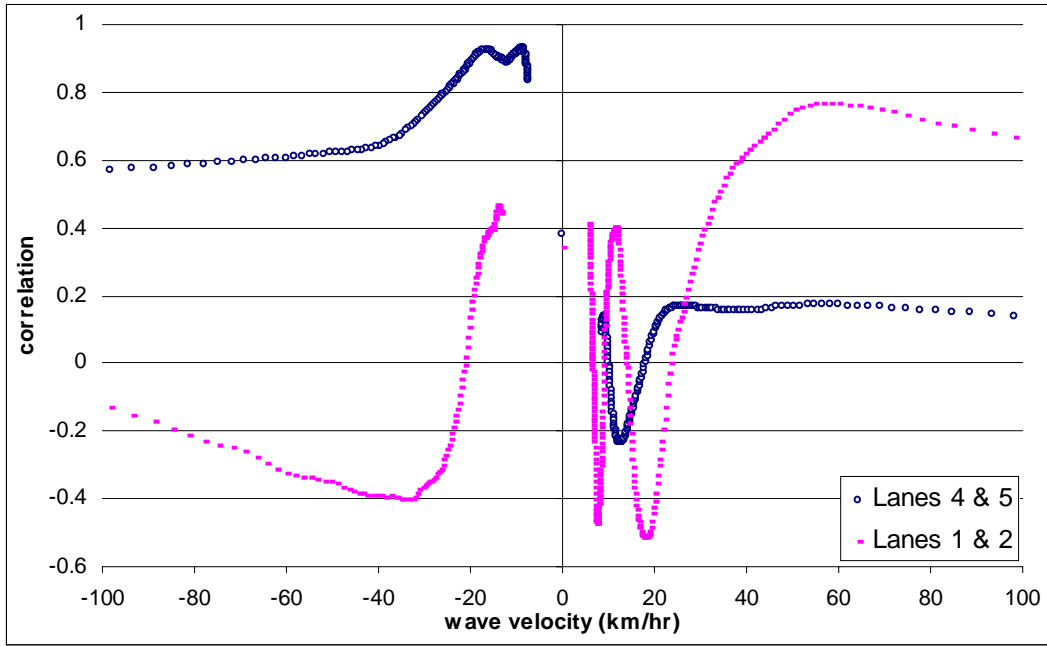


Figure 11. Cross-correlation of the zero-trend N-curves for stations 12 and 22, from 16:20 to 16:30 hrs.

Fig. 4. The regime transition reduced flow by about 1180 veh/hr and propagated from station 22 to 27 in about 8.5 min (\pm 3 min). An explanation of these assertions follows.

The drop in flow was estimated from the change in the average slope of curve N_{27} for the two 15 min time intervals immediately before and immediately after the regime transition was felt at station 27; these intervals are marked by end-points (b, c) and (d, e) in Fig. 4. The estimated flows were 7550 veh/hr and 6370 veh/hr. Arrows designate the corresponding two states in Fig. 9b. As expected, they fall near the middle of the last three black squares and the first three black triangles.

The brief drop in flow around points a and a' of Fig. 4 (from about 7300 veh/hr to 6650 veh/hr) and the ensuing recovery at points b and b' was not part of the regime transition. The drop occurred first at station 27 and propagated to station 22 with the traffic speed. No significant reduction in average speed was detected either. Therefore, the change was a drop in demand¹⁴.

The trip time of the regime transition (8.5 min) was estimated from the time series of average speeds. The time intervals when speeds declined on the three middle lanes were first obtained; see Table 1. The centers of these intervals are the most reliable estimates for the time when the apex of the transition passed through each detector. The trip times were found to be similar and fairly synchronous, although not perfectly so by any means¹⁵. The average of these times is 8.5 min. The error in this qualitative procedure is estimated to be (\pm 3 min).

¹⁴ The brief reduction in flow coincided with the time when the HOV designation begins (at 15:00 hrs). The reduction can perhaps be explained by the actions of single occupant vehicles, which may advance their arrivals to benefit from the HOV lane before it is reserved, and by the actions of HOV's, which may postpone their arrivals to make sure that the HOV lane is clear.

Table 1. Intervals of time when speed declined at stations 22 and 27.

	Lane 2	Lane 3	Lane 4
Station 22	(15:16, 15:30)	(15:12, 15:34)	(15:10, 15:18)
Station 27	(15:18, 15:38)	(15:22, 15:38)	(15:16, 15:38)

The travel time estimate indicates that the transition propagated at about -3.5 km/hr (with an error of about +/-1 km/hr). The slowness of the transition should not be surprising in view of the small difference (1180 veh/hr) between the upstream and downstream flows on both sides of the transition (15% difference). The value is consistent with the slope of the dotted line through the tips of the arrows in Fig. 9b.

Table 1 also reveals that the three middle lanes were in transition at station 27 for about 19 minutes. This information is consistent with the two “non-equilibrium” black diamonds in the flow density plot of Fig. 9b, which correspond to the time interval of the transition and confirm that the transition indeed took slightly over 16 min to pass over detector 27. Therefore, the transition zone must have a characteristic length of about 1.13 km, since it travels at 3.5 km/hr. We believe that the error in this estimate (depending on how one defines the transition zone) is about +/-0.3 km.

It is also seen from Fig. 9b that the (flow, density) black diamonds move from the left side of the diagram to the right side approximately along the dotted line. Interestingly, this is precisely what would occur if all vehicles decelerated slowly and underwent the exact same motion in the regime transition frame of reference; i.e., if every vehicle adopted

¹⁵ Lanes 1 and 5 were not used because they are special: an HOV and an exit lane.

the speed of the prior one with a fixed delay¹⁶. Obviously, this is not what happens in detail, since other forms of driving can also give rise to the same data¹⁷, but the simple characterization of the regime transition as a gradual deceleration “shock” may be useful as a macroscopic approximation.

Therefore, we can estimate the time a vehicle spends in the transition as the ratio of the transition length, 1.13 km, and the average speed of the vehicles in a frame of reference moving with the shock. If we assume that the vehicles decelerate uniformly as they travel through the transition (to the 25 km/hr speed of Fig. 7a—or Fig. 9b—from an initial speed of 100 km/hr) then their average speed while crossing the transition must be $(100+25)/2 = 62.5$ km/hr. This is 66 km/hr in the frame of reference of the transition. Thus, the time to cross the 1.13 km zone is $(1.13/66)(60) = 1.03$ min. Thus, the average deceleration rate is quite leisurely, about 0.33 m/s² or 0.033 g's¹⁸. Since these values describe the drivers' perspective, they could be disproven if they were to contradict one's driving experience. The values seem reasonable, though.

We also estimate that the number of vehicles simultaneously present in the transition zone at any given time is about 116 (29 per lane). This can be estimated using

¹⁶ The delay must depend on the speed of the shock, so as to ensure that vehicle trajectories are translationally symmetric along the path of the shock in time-space.

¹⁷ E.g., if the regime transition is expanding as it propagates.

¹⁸ This gives an upper bound on the maximum error in vehicle count that would be committed by a correct but approximate theory of traffic flow that assumes that regime transitions occur instantaneously and vehicle trajectories are piecewise linear. For the estimated deceleration rate and shock dimensions, the maximum error in vehicle position in such an idealized theory is about 160 m and the maximum error in vehicle arrival time about 10 s. For our 7000 veh/hr flow, the maximum error in count (obtained around the center of the shock) is therefore about 20 veh; i.e., the position of a vehicle on the road could be predicted to within 4 vehicular spacings.

Little's formula from the number of vehicles crossing the shock in the time it takes for their leader to cross it (1.03 min). Since the upstream vehicle flow in the non-HOV lanes is about 6550 veh/hr prior to the transition (as shown both in Figs. 4 and 9b)¹⁹ and the upstream vehicle speed is 100, the relative flow into the transition is about: $[6550(103.5/100) = 6779]$ veh/hr. Therefore the number of vehicles in the transition at any given time is the product of our two quantities: $(6779/60)(1.03) = 116$ veh, or about 29 vehicles per lane.

7. THE DISSIPATION PERIOD (16:20 – 18:00 hrs)

The dissipation period was not studied in as much detail. It included two distinct phases, which are noted in Fig. 12: (1) a period of continued congestion and increasing flow with slightly higher speeds and more significant speed differences across lanes (even at station 27); and (2) the queue dissipation period. Phase 1 lasted approximately from 16:20 to 17:30 hrs and phase 2 from 17:30 to 18:00 hrs.

The flow changes of phase 1 were felt first near the exit, and later at the upstream detectors, as indicated by the 4 dotted arrows of Fig.12. The white triangles of Fig. 9 for detector 27 were generated during this phase. During this time the outflow from the two off-ramps was constant (1350 and 500 veh/hr) and the freeway speeds at station 0 were close to 100 km/hr; i.e., traffic was flowing freely. The flow change observed in Figs.12 and 9 during this period (from 6370 veh/hr immediately before to 7600 veh/hr at the end) corresponds to an increase in the flow of through vehicles from 4520 to 5750 veh/hr. Thus, the proportion of exiting vehicles declined from 29% to 24%.

¹⁹ The HOV flow was estimated as 1000 vph during the transition period.

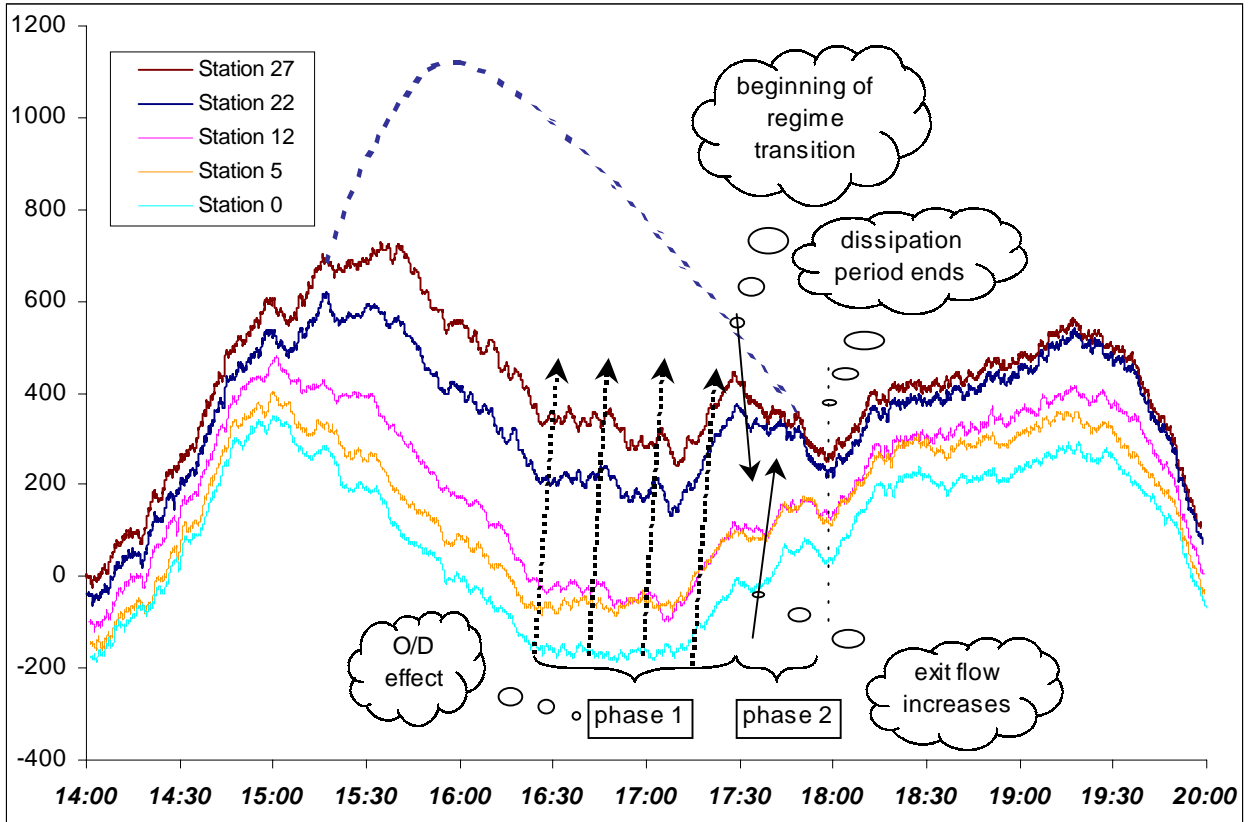


Figure 12. Oblique plot with N-curves for all lanes at all the stations for the whole study period (background flow = 6,930 vph)

To our surprise, we also observed that the speeds at station 27 did not increase equally on all lanes, and that they increased significantly more on the *exit* lanes. (The speed on lane 2 increased to approximately 32 km/hr and on lane 5 to 56 km/hr, as shown on Table B1 of Appendix B.) The most plausible explanation for this is a reduced flow of HOV vehicles cutting into the exit lanes downstream of station 27. Fewer crossovers could have increased the speeds of the middle lanes, thanks to less interference, and also the exit lane flows and speeds at station 27 because of the reduced competition downstream.

The termination of the 1-pipe regime at station 27 when the exit lanes began to move faster can be explained by the logical behavior of a queue rich in through vehicles (poor on exiting vehicles) and by the particular geometry of our site. (Through commuters probably avoided the faster-moving exit lanes since drivers would have known that these lanes were marked as 'exit only' and dropped; see Fig. 1.)

The position of the white triangles of Fig. 9, beneath the congested branch of the curve and grouped along a parallel branch, is also logical. The grouping along a lower line indicates that vehicles were allowing themselves slightly wider spacings, probably because different lanes moved at different speeds.

Phase 2 was felt from 17:20 to 17:44 hrs at detectors 27 and 22, as the back of the queue receded over these detectors. (We know that this is what happened because the speed changes were felt first at 27 and later at 22; see also the direction of the top solid arrow in Fig. 12.) The queue dissipated in a different way at the downstream locations.

Here, phase 2 started at approximately 17:30 hrs when the queue inside the off-ramp suddenly cleared. As a result, the off-ramp flow increased from 1350 veh/hr to 1875 veh/hr (see Fig. B1 of Appendix B), and this also allowed more through flow. This flow change

propagated from the off-ramp as an upstream wave, as shown by the bottom solid arrow of Fig. 12. The convergence of both solid arrows shows that the queue was gradually eaten up from both ends, finally dissipating somewhere between stations 12 and 22 at around 18:00 hrs.

Two comments regarding the queue dissipation transition at station 27 are in order. First note from the crosses of Fig. 9 that the transition went directly from the congested to the uncongested part of the diagram without visiting the “capacity state”. This often observed effect has sometimes been used as evidence against KW theory (since in that theory acceleration regime transitions should visit the capacity state). This criticism is invalid, however, because whenever a deceleration transition moves forward over a detector (as occurs in our case) detector speeds *increase* during the transition. Second, note from Fig. 9 that the transition involved a flow change similar to the one in Sec. 6 (i.e., involving a similar propagation speed) and that, as in Sec. 6, the change took place in two steps (about 15 min). Thus, it is estimated that the width of the transition zone continued to be about 1 km during the recovery.

8. FINDINGS, CONGESTION REMEDIATION MEASURES AND BENEFIT ASSESSMENT

Perhaps the main finding of this paper is evidence that a bottleneck caused by a small off-ramp on a multi-lane freeway can reduce the freeway’s free-flow downstream of the bottleneck, significantly below the capacity of the freeway. At our site, the average downstream freeway flow (at station 0) during the queued steady state of Sec. 3 was about 4520 veh/hr. This is considerably lower than the capacity of the available three lanes.

Other findings are:

- (1) Changes in the O/D table of a diverge with a congested off-ramp can alter the freeway discharge flow significantly, without any effect on the off-ramp. (In our case, a gradual decline in the exit percentage from 29% to 24% increased the discharge flow from 4520 veh/hr to 5750 veh/hr even though the off ramp flow remained constant.)
- (2) Weaving traffic from an HOV lane toward an exit can interfere with the through flow and reduce capacity.
- (3) A semi-congested traffic regime with queued and unqueued states on different lanes can exist upstream of a diverge. (In our case this was detected about 1.5 km upstream.)
- (4) A 1-pipe traffic regime with similar speeds on all lanes can also exist further upstream--2.5 km away when the queue was rich in exiting vehicles.
- (5) Drivers can change attitudes "en masse". (A group of drivers behind a moving bottleneck initially adopted very short headways and later relaxed them.)
- (6) The deceleration transition between free-flow and queuing speeds takes place in a moving zone (the back of the queue) which was about 1 km long in our case. (On that day, the zone moved slowly, first back and later forward.)
- (7) The long periods of time during which flow-density points at detector 27 are in between the branches of the "fundamental diagram" coincide with the passage of the moving zone.
- (8) Flow-density points are otherwise close to the "fundamental diagram".

Below, we estimate the total vehicle-hours of delay caused by this bottleneck on the day in question, and the fraction of delay that could have been eliminated with various remedial measures.

The status quo: The dotted line of Figure 12 represents the cumulative number of arrivals that would have passed the bottleneck at different times, if there had been no restrictions. It assumes that the desired arrival rate changed only once and that it did so relatively abruptly between the values at the beginning and the end of the rush. Since in the oblique coordinate system vertical separations between curves (projected on the vertical axis) continue to be the change in vehicle number, the area between the dotted curve and curve N_{27} , which is the total delay in passing station 27, can be evaluated as if the coordinates were rectangular. Our rough estimate is 800 veh-hrs. This yields an average delay per vehicle of 3.2 minutes since approximately 15000 vehicles were served. Therefore the queue must have extended on average 2.3 km beyond station 27—insofar as a reduction in average speed from 100 km/hr to 30 km/hr²⁰ (see Fig. 9b) for 2.3 km yields a delay of 3.2 min. The total delay between stations 27 and 0 is estimated from the separation between curves N_{27} and N_0 to be comparable and to involve 20000 vehicles. Thus, the total delay caused by this bottleneck is about 1600 veh-hrs. More specifically, about 20,000 vehicles are delayed an average of 4.8 min each.

Remedial measures: Since 1993 when our data were collected, lane 4 of I-880 was extended beyond the bottleneck. Unfortunately, this improvement has had little effect on congestion—changes in traffic conditions cannot be quantified precisely because the

²⁰ Our congested speed guess is the average at detector 27 for the whole uncongested period, obtained from Fig. 9b.

I-880 monitoring project was terminated.

In retrospect, the failure of the lane addition should not be surprising. After all, the main finding of this report is that the discharge flow was restricted by the lane changes of exiting vehicles between station 12 and station 27, and not by a lack of downstream capacity. Clearly, adding a lane was not the solution.

Three alternative improvements are proposed, which can be implemented alone or in combination: (a) modifying the off-ramp flow, (b) modifying the lane assignments, and (c) modifying the HOV lane. Their benefits are assessed.

Off-ramp actions (a): The off-ramp flow can be altered with conventional traffic engineering procedures. Our off-ramp is a one-lane road that merges with the two-lane connector from I-880 south to I-238 west; see Fig. 2b. Interstate I-238 is a short two-lane freeway. A few hundred feet after this merge there is another entrance onto I-238, which is regulated by traffic signals on Hesperian Blvd; see Fig. 2b.

If the flow on our off-ramp could be increased from 1350 veh/hr to 1900 veh/hr, then our research shows that the discharge flow on I-800 north could be expected to increase proportionately from 4500 veh/hr to 6300 veh/hr²¹. The total flow upstream of the Hesperian off-ramp (which would now carry 700 veh/hr instead of 500 veh/hr) could then be 8900 veh/hr. This exceeds comfortably the maximum arrival flows observed at station 27, as shown by Figs. 4 and 9b. Therefore, an increase of 550 veh/hr in the off-ramp flow would completely eliminate delay. Similar considerations show that a 200 veh/hr increase in the off-ramp flow would limit the maximum flow at station 27 to about 7300 veh/hr; since this value is only slightly lower than the maximum arrival flow, it would also result in

²¹ This flow should now be easily sustainable, thanks to the addition of the extra lane.

negligible delay. Finally, an increase in the off-ramp flow as small as 100 veh/hr is estimated to reduce the delay by more than 50%.

Our off-ramp would be able to carry more flow if (i) the competing flow from Hesperian Blvd onto I-238 was reduced, and/or if (ii) the merge area of our off-ramp onto I-238 were re-striped so as to give priority to the vehicles from our off-ramp. The first suggestion can be field-tested by changing the timing of the signals on Hesperian Blvd, or more drastically by closing access to I-238 from Hesperian altogether. In any case, if the Hesperian flow is reduced by “x” veh/hr, the off-ramp flow should increase (approximately) by “x/2” veh/hr. (Therefore, we estimate that if the flow allowed in by this signal is reduced by as little as 3 veh/min, then the delay on I-880 would drop by 50%.) The second suggestion is also easy to implement but one must watch for possible consequences on the southbound I-880 flow; e.g., if the re-striping creates a backup on the connector from I-880 south.

Lane assignments (b): Our research shows that upstream of station 27, traffic behaves as a single pipe; i.e., that through vehicles are entrapped in the queue of exiting vehicles. Our research also shows that when the exit lanes flow faster, through drivers do not encroach on them; i.e., for the most part drivers are disciplined enough to stay on their lanes. This suggests that “through only” lane markings could be effectively used on three through lanes to encourage exiting vehicles to stay on the exit lanes, even if these lanes are queued. These markings should be strictly enforced, and applied for a distance upstream of the off-ramp equal to the maximum extent of the off-ramp queue. The extent of this queue will depend on the number of lanes reserved for exiting vehicles (e.g., 2 lanes in our case) and on the exit vehicle flow. Simple methods (Lawson et al., 1997, and Erera

et al., 1998) can be used to estimate, a priori, this distance in general cases. In our case the distance should be slightly greater than 3 km.

[In our case, the exiting arrival rate would be 1950 veh/hr²² for about 1 hour and then decline, as suggested by the dotted curve of Fig. 12. Since the exit rate would still be 1350 veh/hr, the maximum delay for an exiting vehicle would be about 20 minutes, and the average 10 min. From the q-k diagram of Fig. 9 (for the special case of two lanes) we expect the speed in the queue to be approximately 8.5 km/hr. Assuming a 100 km/hr free-flow speed, our calculations reveal that the maximum extent of the exit-only queue would be slightly over 3 km. In actuality, this distance could be somewhat different if the ramps further upstream contribute significantly to flow]

If lane markings were used for the three km upstream of the diverge, we estimate that total delay would be reduced by more than 80%. (Approximately 3000 exiting vehicles would be delayed, 10 min on average, and the total delay would be about 500 veh-hrs.)

The strategy we have suggested is somewhat drastic because it would place severe lane restrictions at times when they are not needed. It is not difficult to imagine sites (different from ours) where static lane markings could create problems at other times of day when the traffic distribution is different. Therefore, it makes sense to use an “intelligent” control where overhead markings would be changed with the time of day, or even dynamically in response to actual traffic conditions. In situations similar to ours the overhead signs could be adjusted in response to the detection of slow speeds on the exit lanes so as to make sure that the freeway lanes are unassigned only when they are well upstream of the exit queue.

²² This is the I-238 off-ramp flow before the congested period begins.

HOV actions (c): The HOV lane ends between stations 17 and 22 at our site, but its weaving effects are felt all the way down to stations 12 and 5. Clearly, if “through only” lane assignments are used, the HOV designation should be terminated well upstream of the location where the lane assignments begin. (Otherwise HOV’s are encouraged to break the rules and generate pernicious weaving.) In our case, if a static strategy were to be used, then the HOV lane would have to be terminated about 4 km upstream of the off-ramp. The termination point, like the lane-assignments could be changed dynamically, perhaps in response to traffic conditions.

From a theoretical standpoint, this report has shown that KW theory holds quite well upstream of a diverge, provided the location in question is not too close to the diverge, even if the latter is congested. (The report has also shown that the 1-pipe KW theory holds quite well in the presence of a moving bottleneck.) Therefore, one is justified in using simple KW arguments, as we have above, for benefit assessment.

The next step is to test the above ideas at different locations, including our site. There are many congested diverges in US cities, and many could probably benefit from policies (b) and (c) above. A case in point is the I-80 approach to the Bay Bridge from the East Bay in the SF Bay Area, which is often backed up due to a diverge bottleneck. (In that case, and many others, policy (a) is likely to be expensive but policy (b) could have merit if properly implemented.) We recognize that the dynamic versions of policies (b) and (c) can be complicated, but the flexibility of a dynamic approach seems appealing. Clearly, dynamic policies require further development (e.g., to explore hardware choices, human factors, etc), and this is also an area for further work.

9. REFERENCES

Brilon, W. (1998) (private communication).

Cassidy, M. and Windover, J. (1995), "Methodology for assessing dynamics of freeway traffic flow", Trans. Res. Rec. **1484**, 73-79.

Daganzo, C. F. (1997) "A continuum theory of traffic dynamics for freeways with special lanes", Trans. Res. **31B**, (2), 83-102 (1997).

Daganzo, C. F. (1997a) *Fundamentals of Transportation and Traffic Operations*, Pergamon-Elsevier, Oxford, U.K.

Daganzo, C.F., W. H. Lin and J. M. Del Castillo (1997) "A simple physical principle for the simulation of freeways with special lanes and priority vehicles", Trans. Res. **31B** (2), 105-125 (1997).

Erera, A.L., Lawson, T.W. and Daganzo, C.F. (1998) "A simple, generalized method for analysis of a traffic queue upstream of a bottleneck" Trans. Res. Rec. 1646, pp. 132-140.

Lawson, T.W., Lovell, D.J. and Daganzo, C.F. (1997) "Using the input-output diagram to determine the spatial and temporal extents of a queue upstream of a bottleneck" Trans. Res. Rec. 1572, pp. 140-147.

Lawson, T., Lin, W. H. and Cassidy, M. (1999) "Validation of the incremental transfer model" PATH working paper UCB-ITS-PWP-99-16, Inst. Trans. Studies, Univ. of California, Berkeley, CA

Mika, H.S., Kreer, J. B., and Yuan, L.S. (1969) "Dual mode behavior of freeway traffic", High. Res. Rec. **279**, 1-13.

Newell, G. F. (1982), *Applications of queueing theory* (2nd edition), Chapman Hall, London.

Newell, G. F. (1993), "A simplified theory of kinematic waves in highway traffic, I general theory, II queuing at freeway bottlenecks, III multi-destination flows", Trans. Res. **27B**, 281-313.

Newell (1993a) "A moving bottleneck", Institute of transportation Studies, Research Report, UCB-ITS-RR-93-3, Univ. of California, Berkeley, CA.

Skabardonis, A., H. Noeimi, K. Petty, D. Rydzewski, and P. Varaiya (1994) "Freeway service patrol evaluation", Research Report UCB-ITS-PRR-94 Institute of Transportation Studies, Univ. of California, Berkeley, CA.

APPENDIX A: THREE ESTIMATORS OF WAVE SPEED

This appendix defines three statistical measures for comparing the traffic counts of neighboring detectors and compares their performance as estimators of free-flow speed.

Three statistical measures

Consider two detectors between which vehicles are conserved, and let $i = u$ and $i = d$ designate whether the detector in question is the upstream or downstream detector of our pair. The time period for the comparisons $(0, T)$ is assumed to be long compared with the maximum reasonable wave trip time, τ_{\max} . Vehicle counts are assumed to be available from both detectors in the intervals of a fine time lattice with spacing δ (e.g., $\delta = 1$ sec) from $t = -\tau_{\max}$ to $t = T + \tau_{\max}$. From this data counts can be aggregated onto a coarser time lattice with intervals of length Δ , where Δ is an integer multiple of δ . Points on the new lattice are denoted by t_j , where j can be any integer between $-\tau_{\max}/\Delta$ and $(T + \tau_{\max})/\Delta$. Summation of these data yields cumulative counts $C_i(t_j)$ for all points on the lattice, t_j , and for both detectors, $i = u, d$.²³ Note that the interval counts can be expressed as: $C_i(t_j) - C_i(t_{j-1})$.

We will look for the maximum cross-correlation (optimum) lag for different Δ using three different types of discrete-time series, $\{ S_{ij} \}$ (the first subscript refers to the detector and the second to time):

- a) Simple count curves (Mika et al., 1969), where $\{ S_{ij} \} = \{ C_i(t_j) - C_i(t_{j-1}) \}$

²³ The symbol "C" is used instead of "N" to stress that these counts can be initialized in any way. If the counts were zeroed at the times when a reference vehicle traveling with the space-mean speed would have crossed each detector station, then they would correspond to N-curves. This is not required, however.

b) Cumulative curves, where $\{ S_{ij} \} = \{ C_i(t_j) \}$, and

c) Zero-trend cumulative curves, where $\{ S_{ij} \} = \{ C_i(t_j) - C_i(T)(t_j/T) \}$

Case (c) uses CW-curves with background flows equal to the average for the whole time period. This is useful because it automatically corrects for differentially tuned detectors and at the same time reduces the artificially high cross-correlations that are obtained in case (b). These arise from the trends in the two time series, which are necessarily similar, and therefore lead to high correlations when the average flow in $(0, T)$ is high. Thus, method (b) should not be used to make comparisons across scenarios with different average flows. On the other hand, one can show that the optimum lag estimates obtained with methods (b) and (c) are virtually indistinguishable if the detectors are properly tuned.

In all three cases, the estimate of the optimum integer lag, m^* , is:

$$\operatorname{argmin}_m \{ \operatorname{corr} (S_{uj} , S_{d,j+m}) \text{ for } j \in J_m \}, \quad (\text{A1})$$

where the correlation measure is obtained over the set of j 's, J_m , for which both time series are defined; i.e., $J_m = \{ j \mid j \text{ integer ; } j \geq 0, -m ; j \leq T/\Delta, T/\Delta - m \}$.²⁴ The continuous, optimum time lag, τ^* , is estimated by $\tau^* = m^*\Delta$. If the separation between detectors is d , the propagation velocity for count information is estimated by $v^* = d/(m^*\Delta)$.

Note as an aside that if the $C_i(t_j)$ are N-curves, then the time series (b) can also yield an estimate of the rate at which vehicles overtake the count information wave. This is true

²⁴ Note that the sample size (given by the cardinality of J_m) is $\lfloor T/\Delta - |m| \rfloor$, which is a function of m . In all our applications $m^* \gg T/\Delta$, however, and therefore, the differences in sample size across m are inconsequential.

because in this case we can simultaneously look for the horizontal and vertical shifts in the N-curves that minimize the sum across j of the squared differences in their ordinates:

$$\operatorname{argmin}_{m, n} \{ \sum_j (C_u(t_j) - C_d(t_j + m \Delta) - n)^2 \} \quad (\text{A2})$$

where the summation extends over the set J_m . In this case, the first argument returned by (A2) is an estimate of the optimal lag (very similar to (A1)), and the other argument an estimate of the net number of vehicles encountered by the wave during its trip between d and u .

We expect the accuracy of method (a) to depend significantly on Δ because in the limit of $\Delta \rightarrow 0$ the $\{S_{ij}\}$ become continuous-time zero-one processes, which have zero cross-correlation for all lags. This implies that the reliability of the optimum lag estimates must deteriorate drastically as $\Delta \rightarrow 0$. The same problem does not arise with methods (b) and (c). This is illustrated in the next subsection, which compares the three approaches for the purpose of estimating the space-mean speed of a freely flowing traffic stream.

Example: estimation of the space-mean speed

It is possible to show that if $\Delta \rightarrow 0$ and cars can act independently (flow is light), then the wave propagation velocity, v^* , obtained with methods (b) and (c) is a consistent estimator for the space-mean speed, v_s . The value obtained with method (a) can also be shown to be a reasonable estimate of v_s (for non-zero Δ only), particularly when the detectors are close. Therefore, we are justified in performing a simulation test to see how well the three approaches estimate v_s .

A synthetic sample of $M = 1000$ vehicles was used to estimate v_s for a hypothetical,

infinite population of vehicles. Vehicles were observed at two detectors that were d distance units (Du) apart. They were assumed to arrive at the upstream station as a renewal process with uniformly distributed inter-arrival times ranging from 0 to 1 time units (Tu), and to travel without changing their speed. Arrival speeds were assumed to be independent uniformly distributed random variables with range [50, 80] (Du/Tu). Clearly then, the population space-mean speed was: $v_s = 63.8 \approx 65 \text{ Du/Tu}$.²⁵

The three estimation methods were then applied with lags, $\Delta = 1$ and $\Delta = 2$ (Tu's), to 200 random, M -vehicle samples. Part (a) of Fig. A1 displays the frequency distribution across samples of the estimates obtained with methods (a) and (b), for $d = 390$ (??) Du. (The distribution of method (c) was found to be virtually indistinguishable from that of (b), as expected). The figure shows that both estimation methods performed similarly with $\Delta = 2$ Tu. However, as the resolution of the method was increased, by decreasing Δ to 1 Tu, method (b) improved more than method (a). (This should not be a surprise, since we have already mentioned that for sufficiently small Δ , method (a) would fail.)

The simulation was then repeated for a longer distance between detectors (7800 Du's ??) because this increases the difficulty of the estimation problem. In this case too, methods (b) and (c) proved more reliable than (a), as shown in part (b) of the figure.

²⁵ The uniform distribution in [50, 80] is for speeds seen by a stationary observer. Therefore, space-mean speed is the harmonic (and not the arithmetic) mean of this distribution.

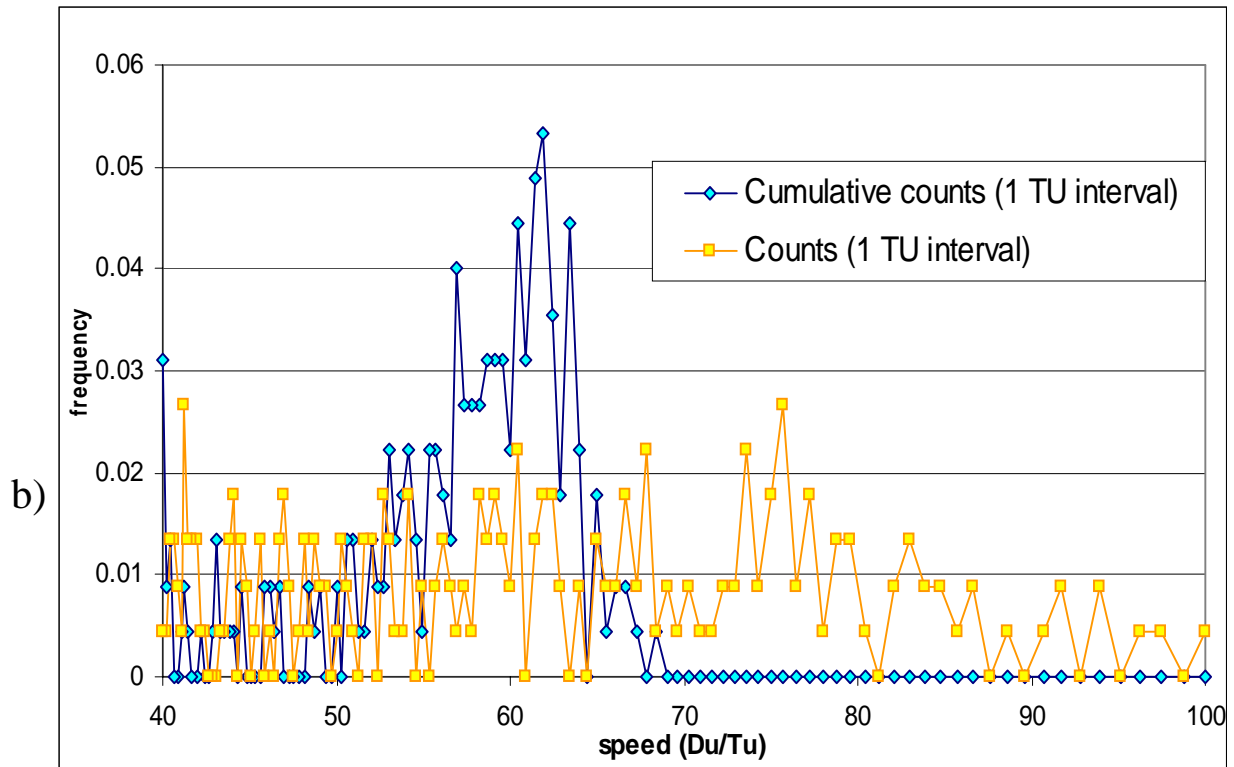
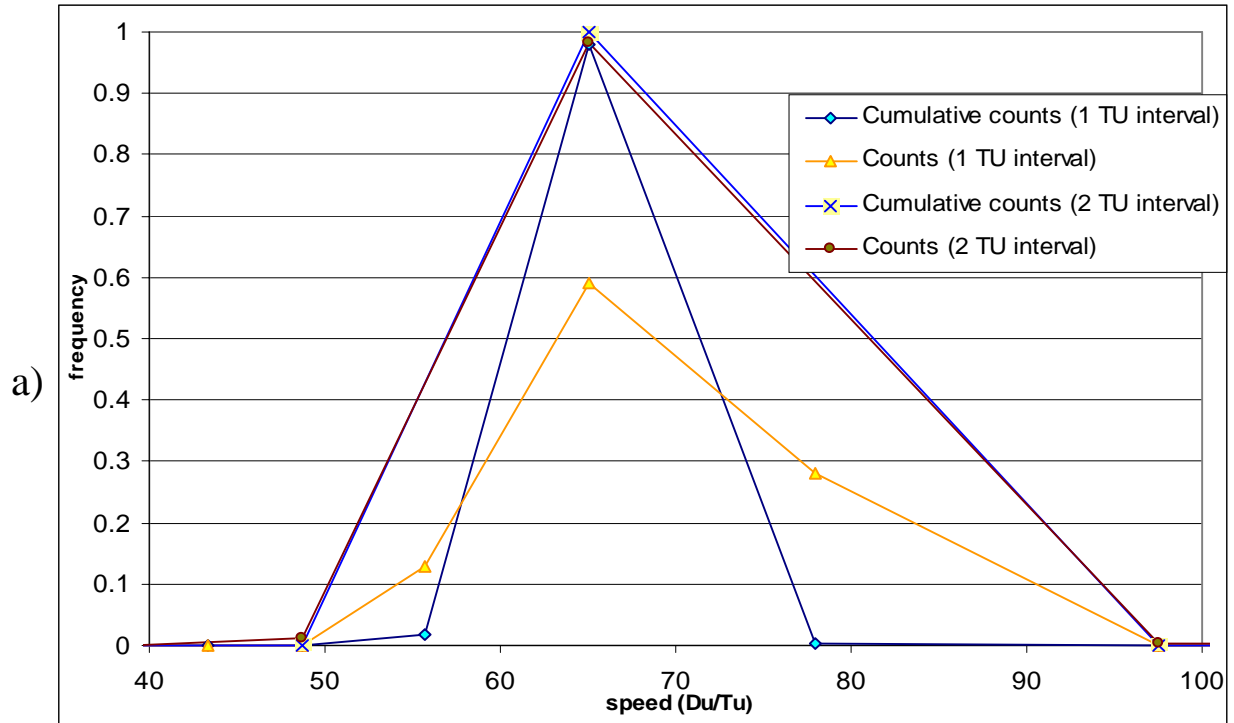


Figure A1. Distribution of the maximum correlation speed estimator for a true speed of 63.8 Du/Tu for different time series: (a) Distance of 390 Du; (b) Distance of 7800 Du.

APPENDIX B: ADDITIONAL DATA

This appendix contains supplementary data. Figure B1, is an oblique plot of the flow at I-238 off-ramp with a background flow of 1620 vph. The remaining figures are time series of average speed at all the detector locations, similar to Fig. 3a of the text. Table B1 summarizes the speed information from station 27.

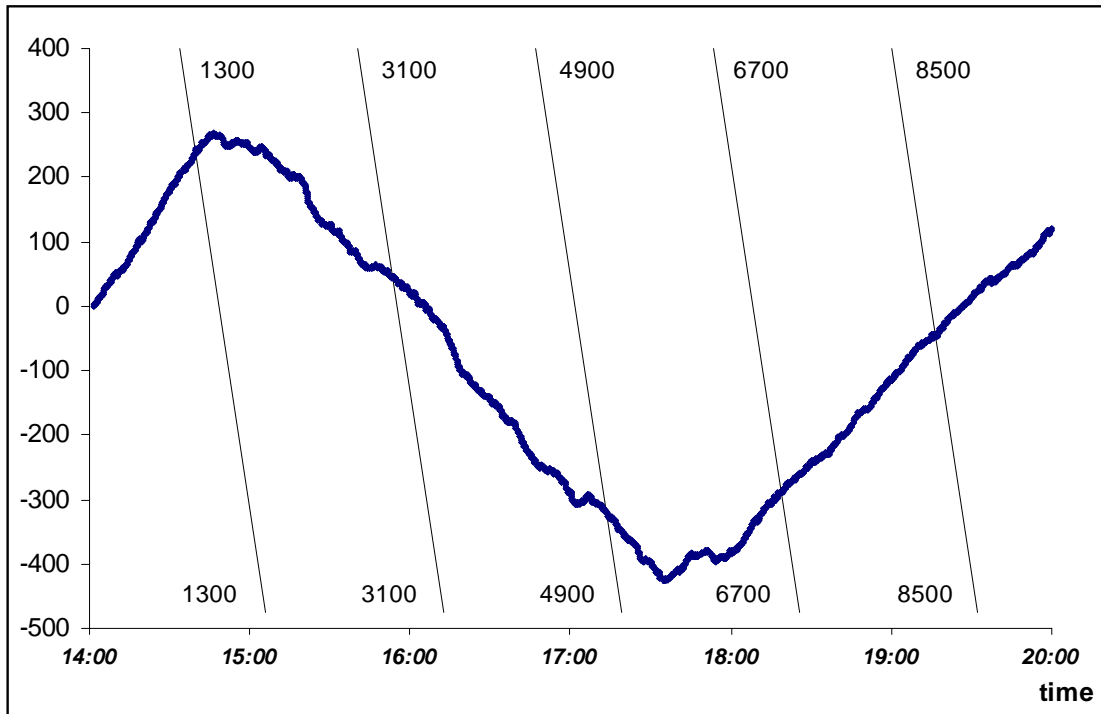


Figure B1. Oblique plot with N-curve for exit off-ramp at station 0
(background flow = 1,620 vph)

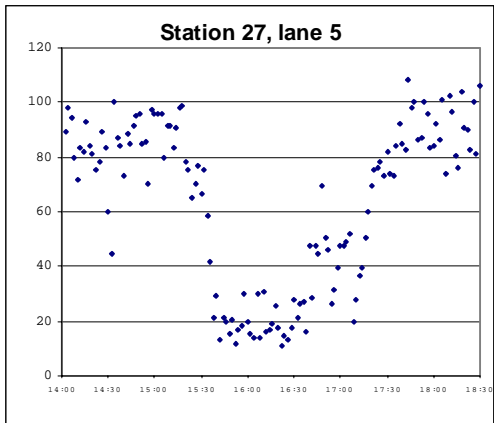
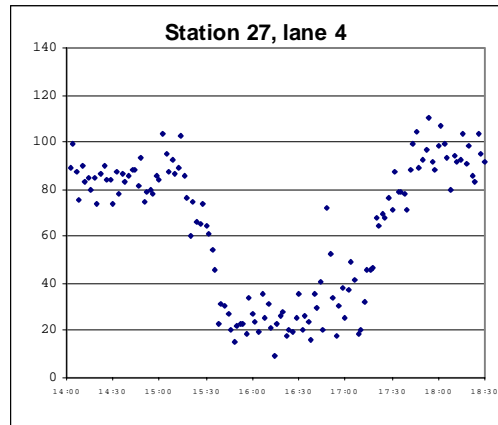
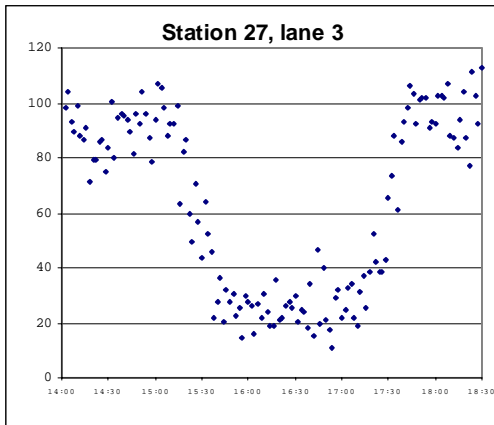
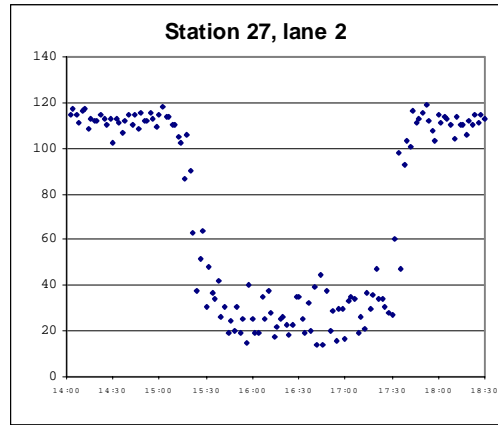
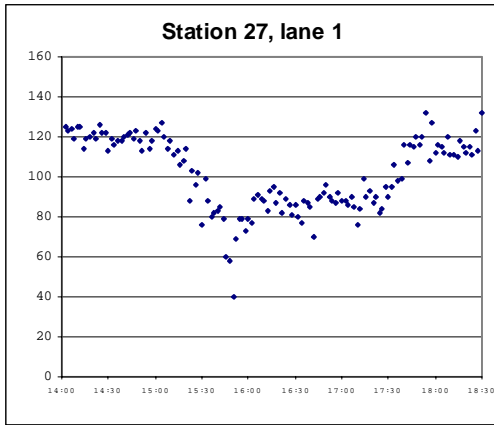


Figure B2. Speed estimation at Station 27, all lanes (in km/hr).

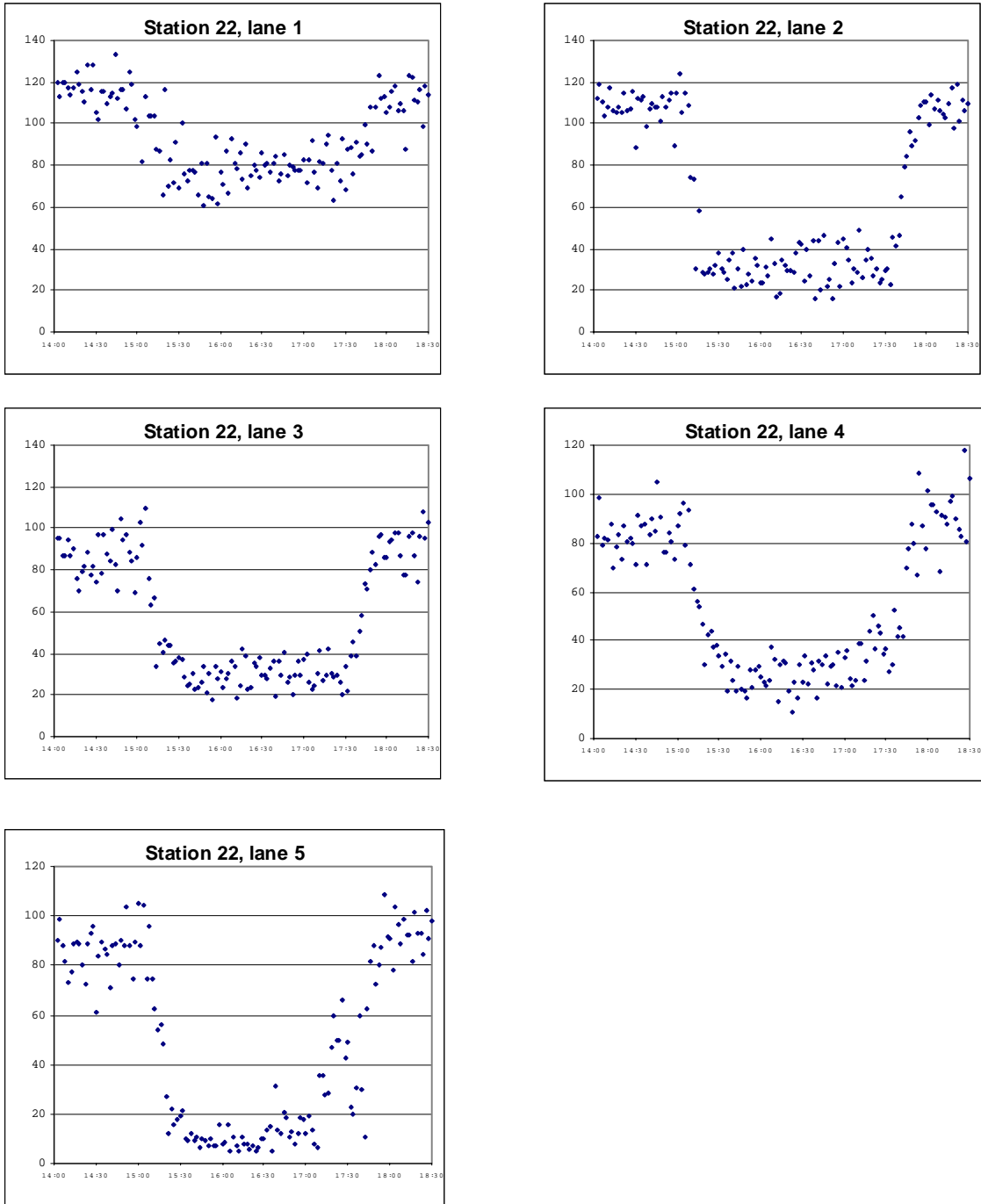


Figure B3. Speed estimation at Station 22, all lanes (in km/hr).

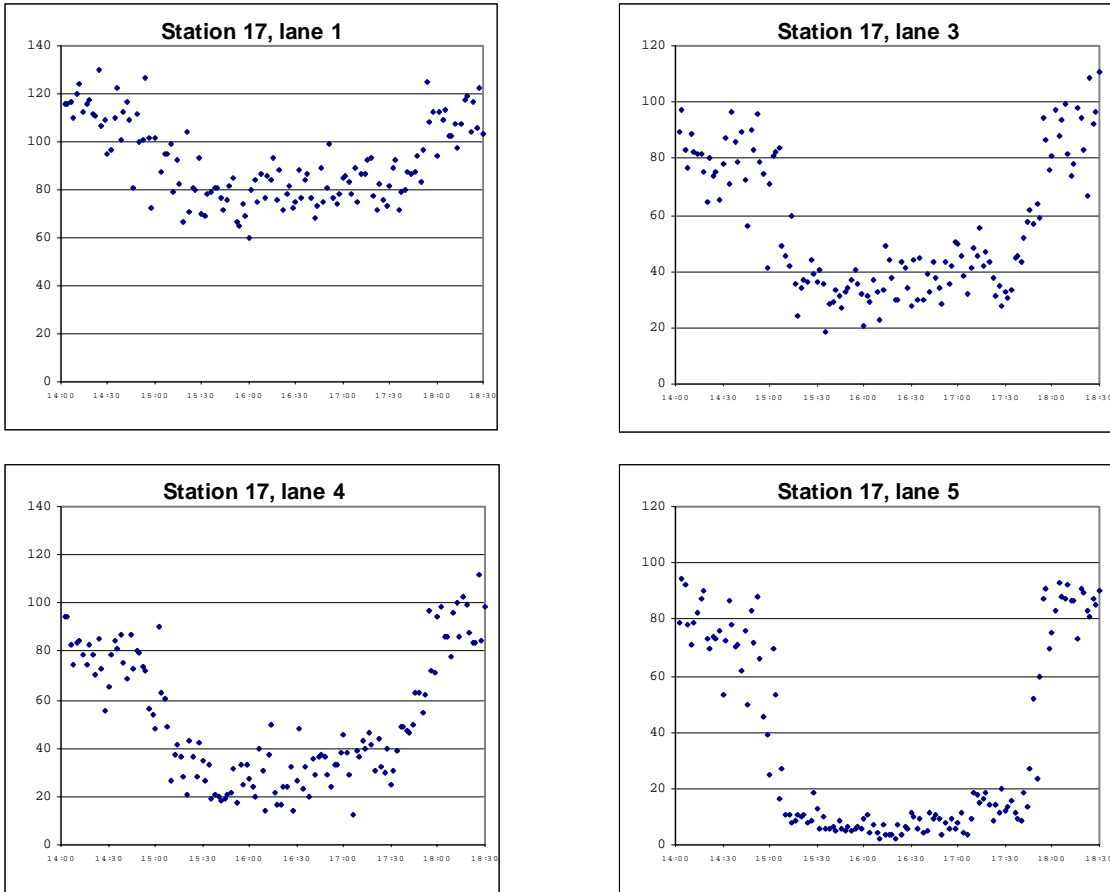


Figure B4. Speed estimation at Station 17, all lanes (in km/hr). No data was available for lane 2.

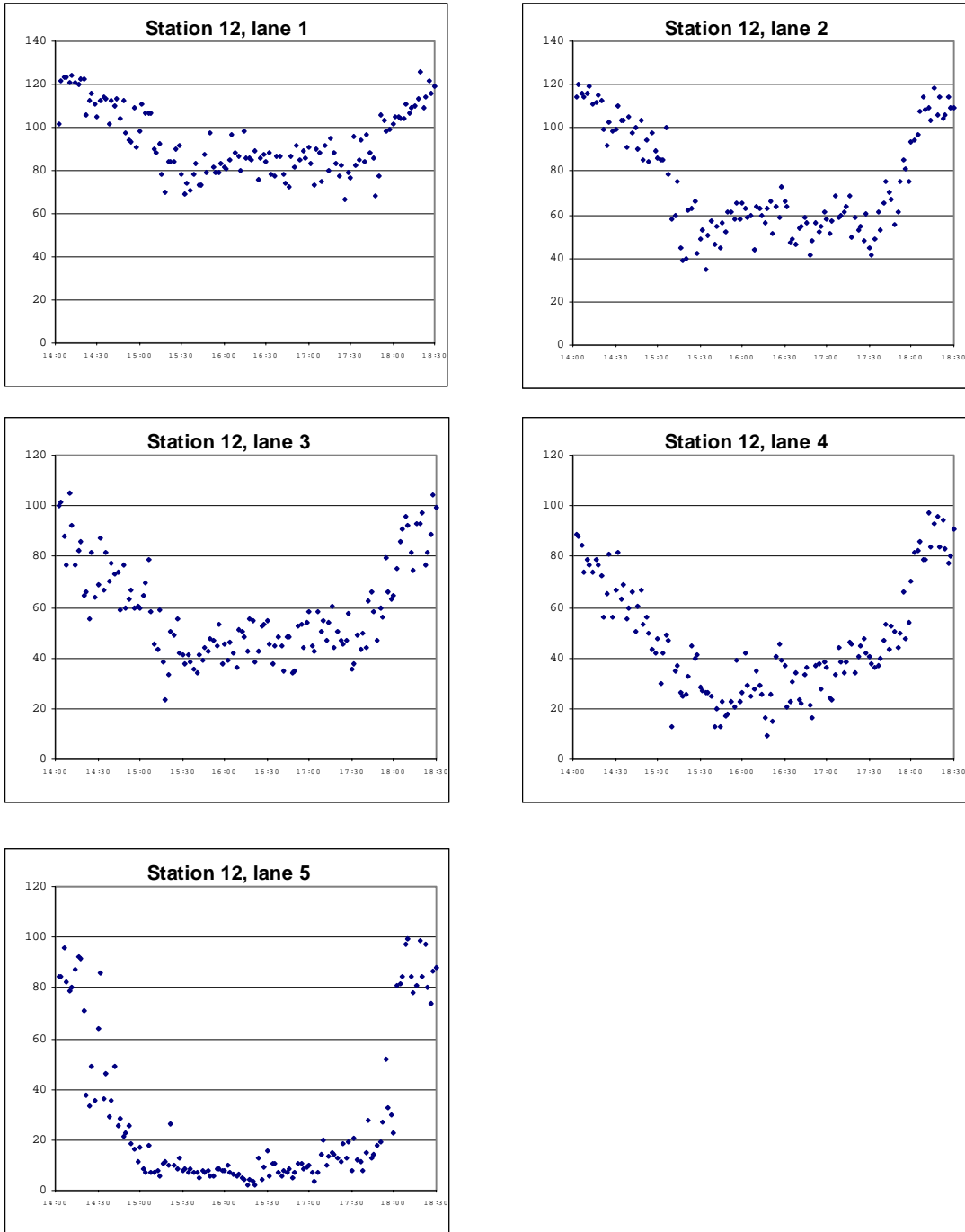


Figure B5. Speed estimation at Station 12, all lanes (in km/hr).

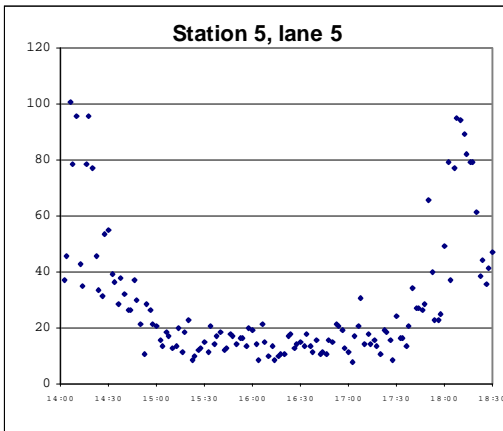
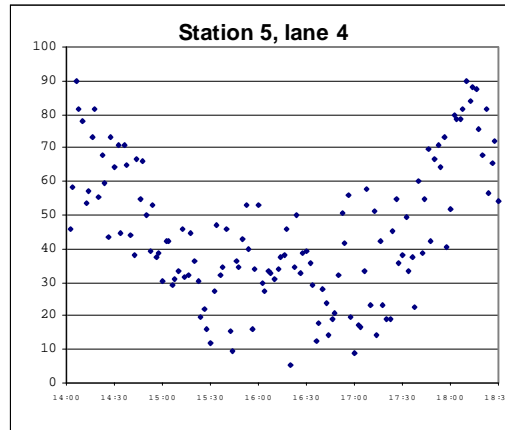
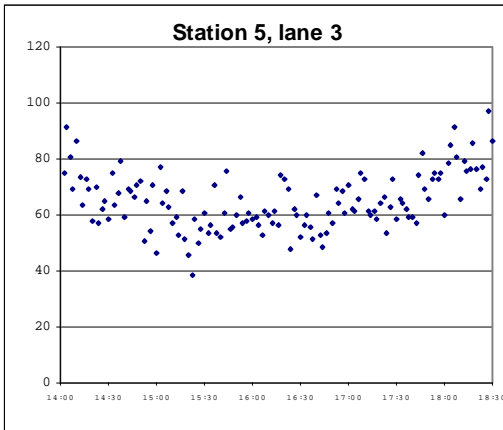
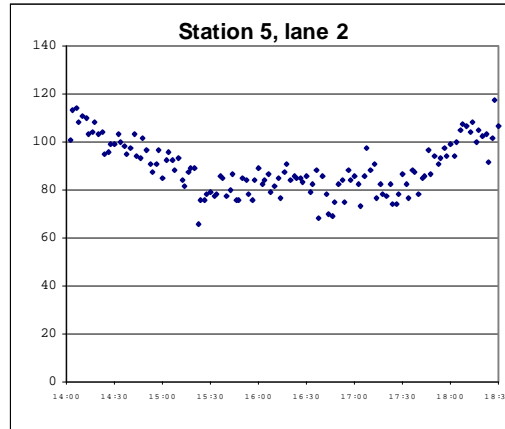
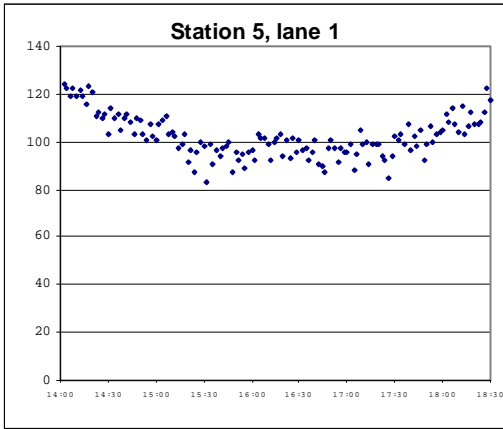


Figure B6. Speed estimation at Station 5, all lanes (in km/hr).

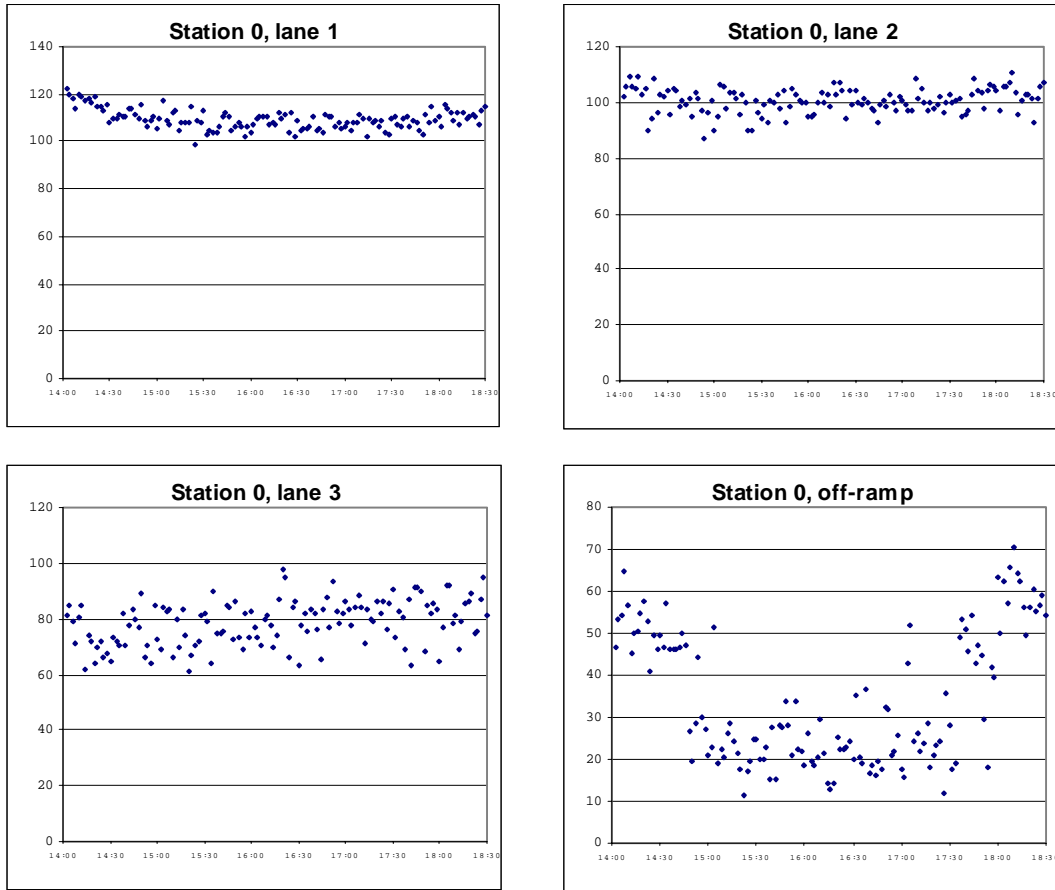


Figure B7. Speed estimation at Station 0, all lanes (in km/hr).

Table B.1 Average speed in km/hr (and its rate of change in km/hr²) for different periods of the day at station 27.

15:40-16:20

	slope	mean
Lane 1	19.23	79.38
Lane 2	0.87	25.34
Lane 3	-3.71	25.44
Lane 4	-3.71	25.44
Lane 5	-1.23	24.56

16:30-17:00

	slope	mean
Lane 1	9.53	86.22
Lane 2	-7.24	26.93
Lane 3	-1.17	25.38
Lane 4	9.17	32.10
Lane 5	22.93	36.27

17:00-17:30

	slope	mean
Lane 1	4.30	88.29
Lane 2	13.98	32.30
Lane 3	35.29	37.63
Lane 4	53.23	51.02
Lane 5	43.72	56.41

17:30-18:00

	slope	mean
Lane 1	24.21	111.31
Lane 2	56.53	97.23
Lane 3	25.52	91.28
Lane 4	25.12	90.14
Lane 5	11.53	88.80