Some observed queue discharge features at a freeway bottleneck downstream of a merge

Robert L. Bertini [a]* *, Michael J. Cassidy [b]

[a] *Department of Civil Engineering, Portland State University, P.O. Box 751, Portland, OR 97207-0751, USA*
[b] *Department of Civil and Environmental Engineering and Institute of Transportation Studies, University of California, 109 McLaughlin Hall, Berkeley, CA 94720-1720, USA*

Abstract

Details of traffic evolution were studied upstream and downstream of a freeway bottleneck located near a busy on-ramp. It is shown that on certain days the bottleneck became active upon dissipation of a queue emanating from somewhere further downstream. On such occasions, the bottleneck occurred at a fixed location, approximately one kilometer downstream of the merge. Notably, even after the dissipation of a downstream queue, the discharge flows in the active bottleneck were nearly constant, since the cumulative counts never deviated much from a linear trend. The average bottleneck discharge flows were also re-producible from day to day. The diagnostic tools used in this study were curves of cumulative vehicle arrival number versus time and cumulative occupancy versus time constructed from data measured at neighboring freeway loop detectors. Once suitably transformed, these cumulative curves provided the measurement resolution necessary to observe the transitions between freely flowing and queued conditions and to identify some important traffic features. [1]

1. Introduction

This study reports on observations from a freeway bottleneck (in the vicinity of a merge) when it became active upon dissipation of a queue emanating from somewhere further downstream. The active bottleneck's discharge flow was higher than the flow that was governed by the capacity of the downstream obstruction. This study is distinct from previous work that reported on obser-vations when an unobstructed bottleneck (at the same site on different days and at one additional site) exhibited higher flows prior to queue discharge at a lower rate.

In earlier studies, the authors examined traffic conditions upstream and downstream of a freeway bottleneck located near a busy on-ramp (Cassidy and Bertini, 1999a,b; Bertini, 1999) and certain reproducible features were observed. For example, it was shown that the bottleneck was located more than one kilometer downstream of the busy merge, farther than had been found previously. Also, on certain days, high bottleneck flows of nearly 7000 vehicles per hour (vph) were measured across three lanes for up to 40 min before queueing eventually occurred imme-diately upstream of the bottleneck, giving rise to lower average discharge rates. In particular, periods of rather low discharge flow accompanied the onset of upstream queueing. While the bottleneck was active, [1] the average discharge flows were observed to be nearly constant and did not vary considerably from day to day.

To promote the visual identification of time-dependent features of the traffic stream, these previous studies used transformed curves of cumulative vehicle count and curves of cumulative occupancy constructed from data measured at neighboring freeway loop detectors. These cu-mulative curves provided the measurement resolution necessary to observe the transitions from freely flowing to queued conditions and to identify a number of notable, time-dependent traffic features in and around the bottleneck.

Cumulative curves were also used in this study. The present study adds to the previous findings by reporting on observations taken during an afternoon rush on different days when a queue emanating from somewhere further downstream spilled over into the merge area prior to the activation of its bottleneck. Upon dissipation of the downstream queue, the bottleneck consis-tently arose at the same location, more than one kilometer downstream of the on-ramp. Further, vehicles discharged from the subject bottleneck without exhibiting a particularly high flow or an especially low discharge flow. Rather, it is shown that the bottleneck's average discharge rate during its active period was nearly constant and exhibited only small deviations from one day to the next.

Accordingly, observations on three additional days when the bottleneck's activation was pre-ceded by the spillover of a queue from further downstream were used to demonstrate the re-producibility of these findings. The following section contains a brief description of the freeway site and the loop detector data used in this study. Section 3 describes the bottleneck's location and discharge features that were uncovered using cumulative curves. Section 4 presents features that were found to be reproducible on three additional days and finally, Section 5 presents some brief comments on the study's findings.

2. Data

The observations that follow were taken during an afternoon peak (March 12, 1997) from the segment of westbound Gardiner Expressway near the Spadina Avenue on-ramp, as shown in Fig. 1. The site is located in Toronto, Canada. Inductive loop detectors recorded vehicle counts, [1]

---

[1] An *active* bottleneck arose when vehicles discharged from an upstream queue (to guarantee that the bottleneck served vehicles at a maximum rate) and vehicles were unimpeded by traffic conditions emanating from further downstream (Daganzo, 1997).
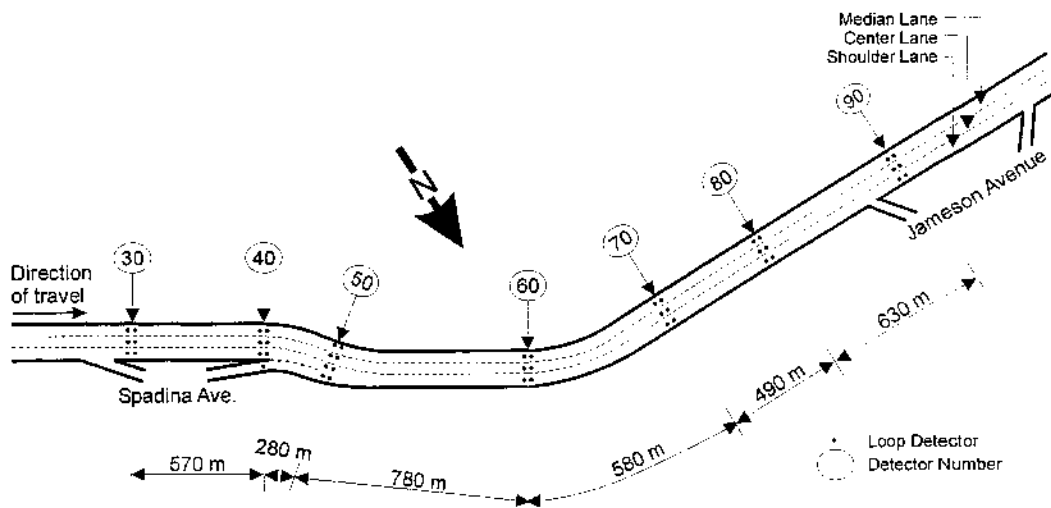
Fig. 1. Gardiner Expressway Site, Toronto, Canada.

occupancies [2] and time mean speeds in each lane at 20-second intervals. The detectors are labeled according to the numbering strategy used by the City of Toronto. Ramp metering is not installed on this facility, although the Jameson Avenue on-ramp is closed between 15:00 and 18:00 each afternoon. The lanes at the study site are referred to, from left to right, as the median lane, center lane and shoulder lane.

3. Observations

Fig. 2 presents transformed curves of cumulative vehicle arrival number constructed from counts measured across all lanes at detectors 40-80 and collected during an 80-minute period surrounding the activation of the bottleneck between detectors 60 and 70. The curves were constructed by taking linear interpolations through the 20-second counts so that a curve's slope at time t would be the flow past location x during the time interval containing the t. The counts for each curve in Fig. 2 were started (N = 0) relative to the passage of a hypothetical reference vehicle (curve 40 includes counts from the Spadina Avenue on-ramp so that all curves describe the same collection of vehicles). Hence, the horizontal and vertical separations between curves would have been the trip times and vehicle accumulations between detectors, respectively (Newell, 1982, 1993).

However, each curve (along with its time axis) was shifted horizontally to the right by the average free-flow trip time between the respective detector and downstream detector 80. Following such shifts, vertical displacements between curves are the excess vehicle accumulation between detectors due to vehicular delays. Furthermore, the curves' features have been magnified

---

[2] Measured occupancy is the percentage of time a detector is covered by a vehicle in the measurement interval.
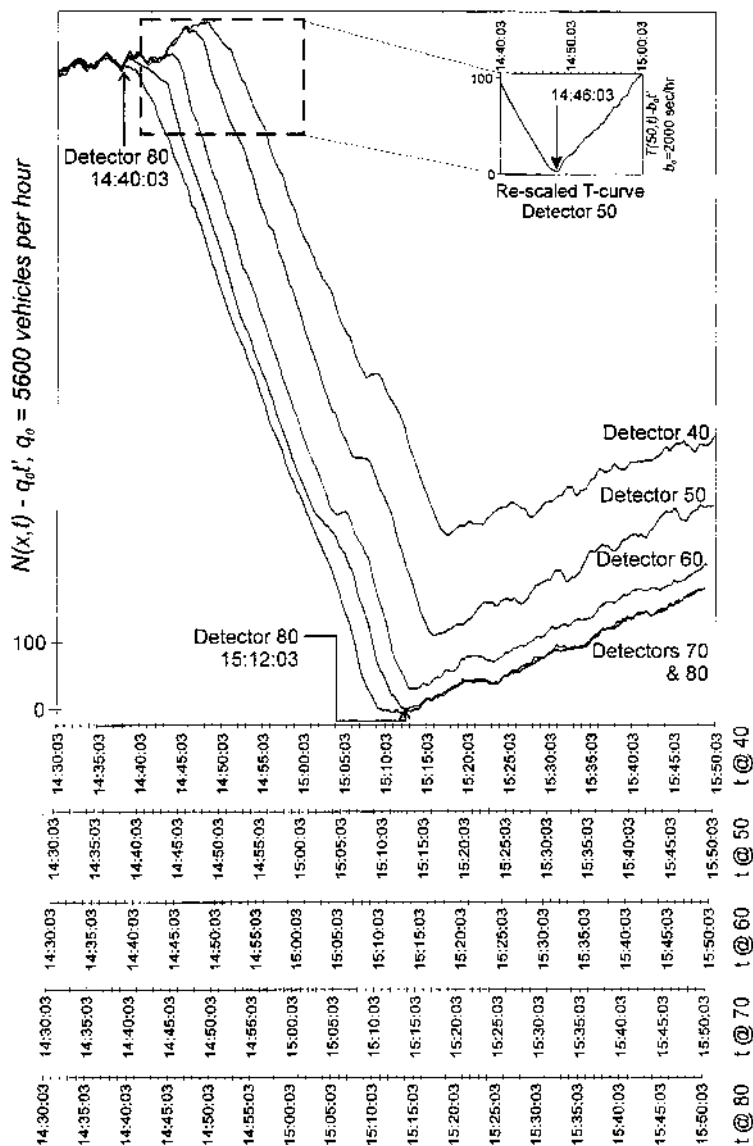
Fig. 2. Transformed *N*-curves, detectors 40–80, 3/12/97.

by plotting only the difference between the cumulative count and a line $N = q_0 \blacksquare t'$, where q_0 is a re-scaling rate and t' is the elapsed time from the beginning of each curve. This re-scaling helps visually diagnose important traffic details, and does not affect the vertical separations (Cassidy and Windover, 1995).

The nearly superimposed portions of the transformed curves in Fig. 2 reveal that traffic was flowing freely between all detectors from 14:30:03 until 14:40:03. At 14:40:03, curve 80 diverged from curve 70, marking the arrival (at detector 80) of a backward-moving queue from some-

where further downstream. Subsequently, the progress of this queue is mapped on the figure as curves 60 and 70 diverged (at 14:41:03), followed by the divergences of curves 50 and 60 (at 14:44:03) and curves 40 and 50 (at 14:46:03). Therefore, this queue emanating from a restriction located downstream of detector 80 had arrived at detector 50 by 14:46:03. To confirm this, the small window in Fig. 2 shows a re-scaled curve of cumulative occupancy versus time, or a T-curve (for detector 50), where cumulative occupancy is the total vehicle trip time measured over the detectors by time t. Again for the purpose of magnifying details, the T-curve shown is the difference between the cumulative occupancy actually measured at detector 50 (across all lanes) and a line $T = b_0 \blacksquare t,$ where b_0 is a re-scaling rate and t is the elapsed time from the beginning of the curve. A reduction in flow at detector 50 (at 14:46:03) was accompanied by an increase in the occupancy rate; these are features used to verify the arrival of the backward-moving queue.

The continued vertical separation between all pairs of curves from 14:40:03 until 15:12:03 reveals that a queue occupied the entire section between detectors 50 and 80 during this 33-minute period. The average flow measured during this interval was only 3770 vph and this value was governed by the capacity of the downstream restriction. At 15:12:03 the curves at detectors 70 and 80 again became superimposed, revealing that the downstream queue had dissipated, leaving traffic to flow freely between detectors 70 and 80. [3] The continued excess accumulation upstream of detector 70 (indicated by the vertical separations which remained after 15:12:03) reveals that the bottleneck between detectors 60 and 70 was activated immediately thereafter. Thus, the transformed N-curves in Fig. 2 were necessary for identifying the time (15:12:03) at which the bottleneck between detectors 60 and 70 became active. [4]

Further analysis of the N-curves showed that the bottleneck between detectors 60 and 70 remained active until another spillover from some downstream restriction arrived at detector 80 at approximately 18:04:03. To demonstrate this, Fig. 3 displays transformed N-curves for detectors 50 and 80 that span a much longer period. The queue's enduring presence between detectors 50 and 80 is visible in Fig. 3 by virtue of the continued vertical displacement between the two curves. To trace the spillover, one can see from Fig. 3 that the slope of the curve for detector 80 dropped at 18:04:03 (as shown by the vertical arrow) and that a slope reduction was also displayed by the curve at detector 50 shortly thereafter (at approximately 18:11:23). Even after these flow reductions, the queue between detectors 50 and 80 persisted.

Fig. 4 confirms that the backward-moving queue arrived at detector 80 at approximately 18:04:03. To this end, the figure contains a re-scaled N-curve for detector 80 along with a re-scaled T-curve. The two curves reveal that a reduction in the N accompanied a rise in the T at approximately the same time. This confirms that the downstream queue arrived at detector 80 at

---

[3] It is plausible that the dissipation of the downstream queue was caused by the closing of the Jameson Avenue on-ramp at approximately 15:00:00, assuming that there was an active bottleneck in the vicinity of this ramp prior to this time.

[4] There were no uphill or downhill gradients, construction activities, lane-changing restrictions or speed limit changes at this freeway site. As discussed in Cassidy and Bertini (1999b), it is possible that the freeway's horizontal curve (see Fig. 1) is the homogeneity creating the bottleneck. However, it was shown in Cassidy and Bertini (1999a) for another freeway location that a bottleneck also formed more than a kilometer downstream of an on-ramp absent of any obvious geometric inhomogeneity.
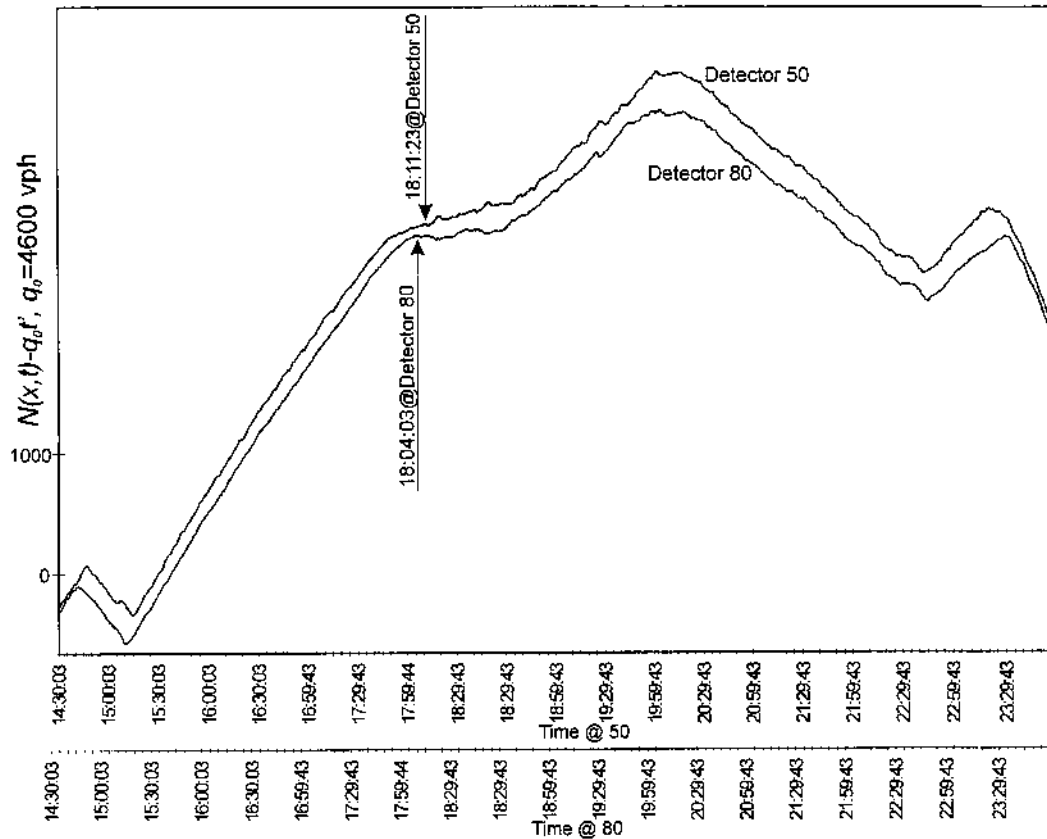
Fig. 3. Transformed N-curves, detectors 50 and 80, 3/12/97.

about 18:04:03, deactivating the bottleneck at that time by restricting its discharge flow. A re-scaled T-curve at detector 50 (not shown) confirmed that the downstream queue arrived at this detector at approximately 18:11:23.

To study discharge flows while the bottleneck was active, Fig. 5 shows re-scaled N- and T-curves for all lanes at detector 80, spanning about 3 h. This interval includes a few minutes prior to bottleneck activation where a lower flow (governed by the capacity of the downstream re-striction) was observed; the period during which the bottleneck was active (beginning at 15:12:03); and several minutes after the downstream spillover at 18:04:03. N- and T-curves (not shown) measured in individual lanes were examined to determine whether the bottleneck continued to serve vehicles at a maximum rate between 15:12:03 and 18:04:03. This examination revealed that a flow reduction (caused by an incident upstream of detector 60) arrived at detector 80 at 16:58:23. Therefore, bottleneck queue discharge flows will only be measured from the period between 15:13:03 and 16:58:23.

Fig. 5 shows that after the queue from the downstream restriction dissipated at 15:12:03, a sequence of nearly stationary flows prevailed. These sequences (and the times delineating them)
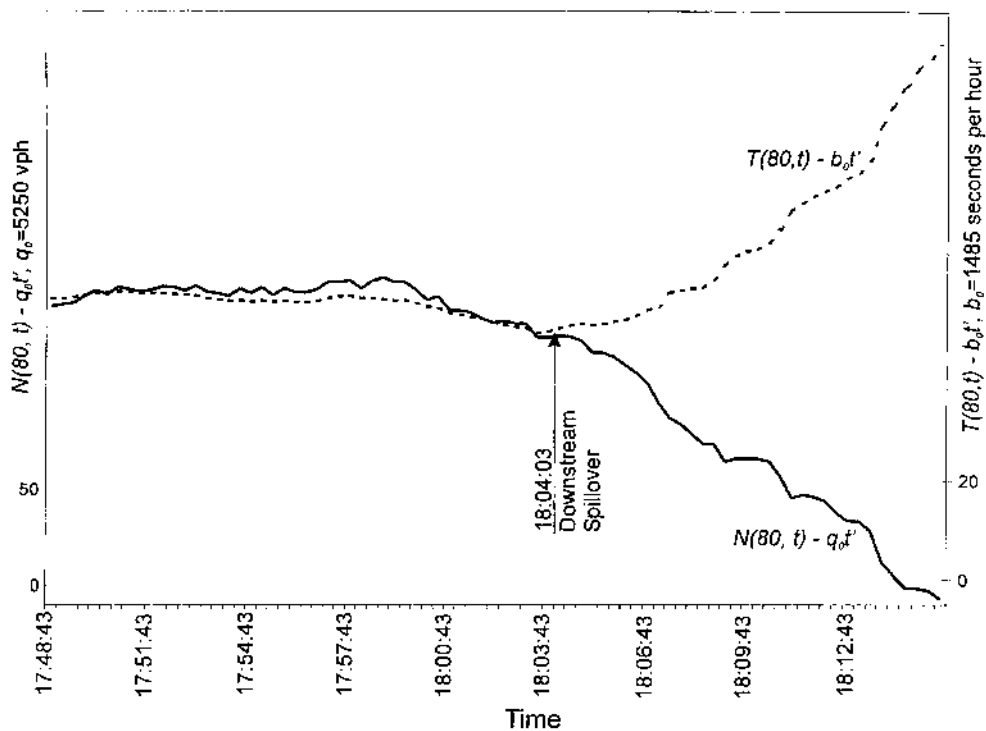
Fig. 4. Re-scaled N- and T-curves, detector 80, 3/12/97.

are superimposed on the N to highlight periods of nearly constant flow (corresponding rates are shown in units of vph, with average counts per minute in parentheses). The onset of queue discharge immediately carried a flow of 5910 vph for approximately 45 min, followed by rates that happened to diminish slightly over time on this particular day only. That the discharge rate decreased over time was not a reproducible feature on three additional days studied, as will be described in Section 4. Fig. 5 also shows that vehicles discharged through the active bottleneck at an average rate of 5830 vph (shown by the dashed line). Using the vertical scale on the left edge of Fig. 5, one can determine that the N-curve never deviated from the dashed line (denoting the average discharge rate) by more than 77 vehicles. It thus seems reasonable to designate the average discharge flow as ''nearly constant''.

The origins of the changes observed in the discharge flows shown in Fig. 5 are of interest. As a preliminary indication that the flow changes may have emanated from the downstream end of the queue, Fig. 6 shows re-scaled N-curves constructed from counts measured at detectors 40-80 during the bottleneck's active period. The curves have been vertically displaced by arbitrary distances to enhance their clarity. The sequence of nearly stationary flows measured by detector 80 (and shown in Fig. 5) has been superimposed on the N in Fig. 6. Similarly the periods of nearly stationary flow (a straightedge may be used to verify these delineations) have been depicted on the curves corresponding to detectors 40-70. After connecting the points marking the flow changes, as shown by the arrows in the figure, it appears that the flow changes emanated from somewhere
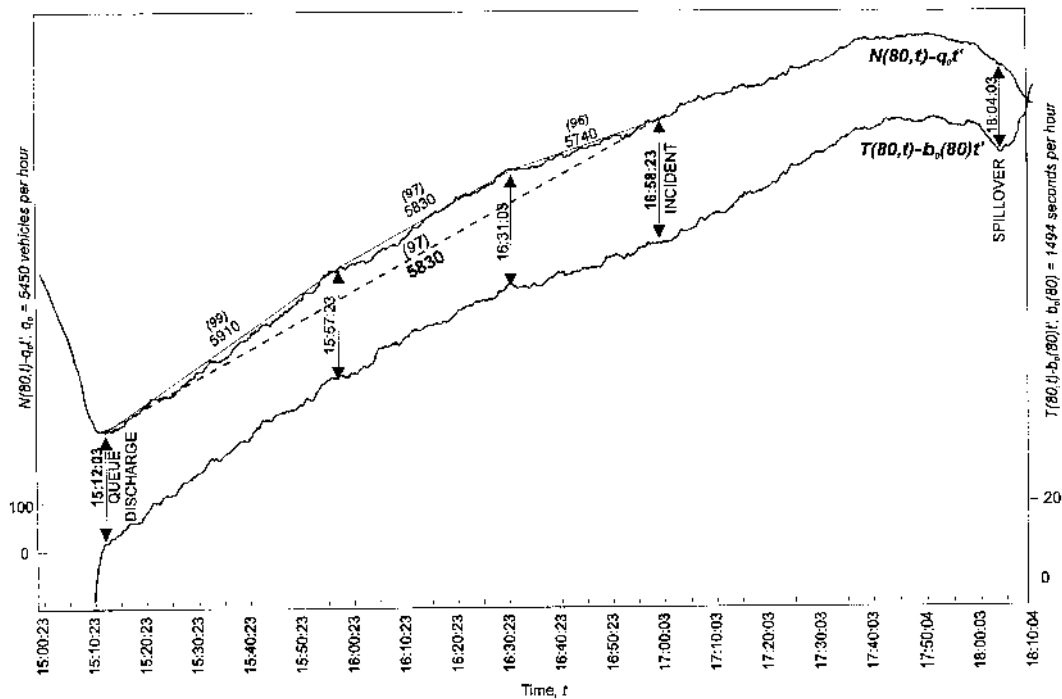
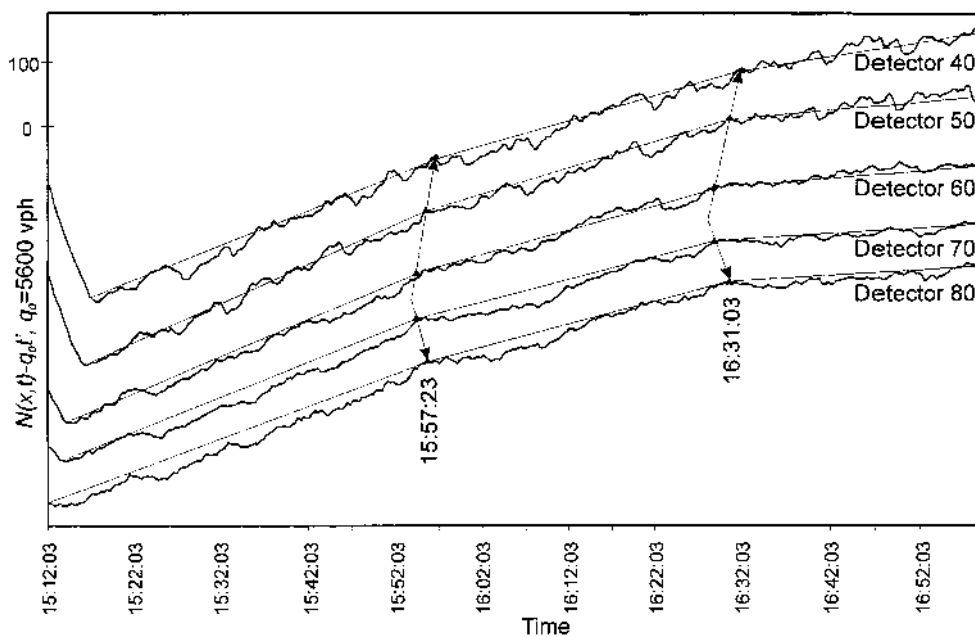Fig. 5. Re-scaled *N*- and T-curves, detector 80, 3/12/97.

Fig. 6. Re-scaled N-curves, detectors 40-80, 3/12/97.

between detectors 60 and 70. The driver behavior that leads to this apparent observation is the subject of ongoing research.

4. Reproducing the observations

As mentioned above, the queue discharge rate measured at detector 80 on March 12, 1997 decreased slightly over time between 15:12:03 and 16:58:23, whereas on other days studied, the discharge flow exhibited an alternating sequence of higher and lower rates. As an indication of this, Figs. 7-10 illustrate the activation of the subject bottleneck subsequent to the dissipation of a downstream queue on October 8, 1997. These figures are similar to Figs. 2-5, so their descriptions will be abbreviated.
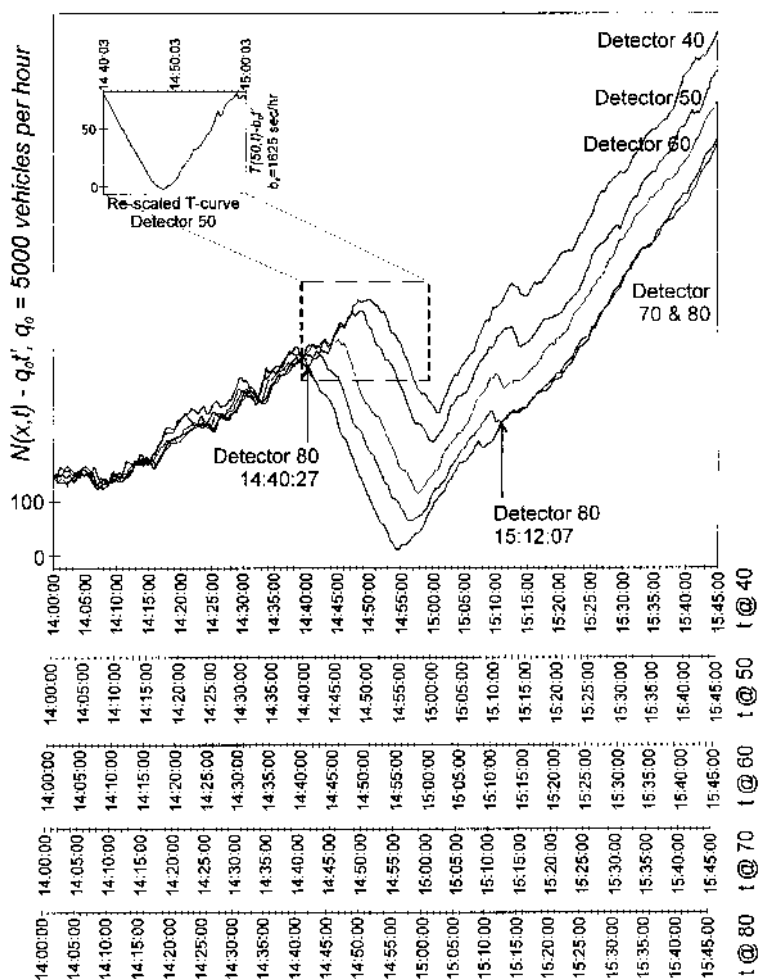

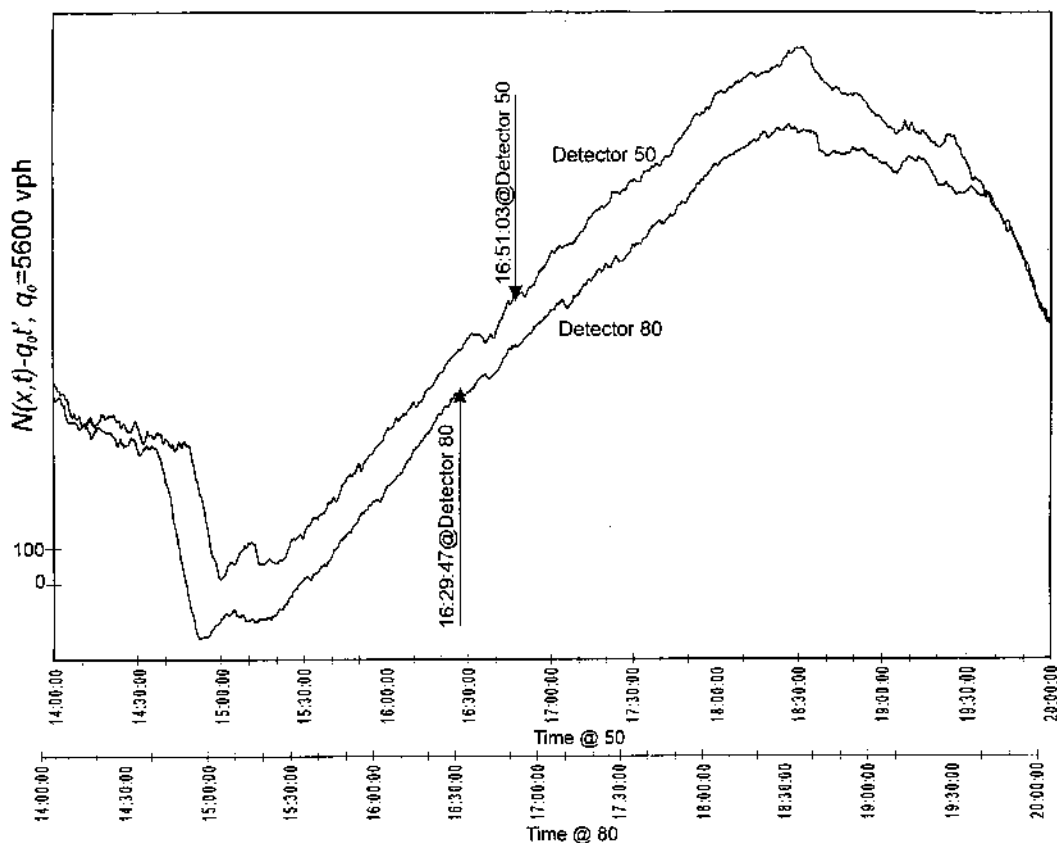Fig. 7. Transformed N-curves, detectors 40-80, 10/8/97.

Fig. 8. Transformed *N*-curves, detectors 50 and 80, 10/8/97.

The transformed N-curves [5] in Fig. 7 (spanning 105 min) reveal that traffic was flowing freely between all detectors until about 14:40:27. At about this time, curve 80 diverged from curve 70 marking the arrival of a queue that spilled over from further downstream. The queue later arrived at detector 50 (at 14:48:27), as corroborated by the re-scaled T-curve in the small window in the figure. The curves also reveal that curves 70 and 80 became superimposed at about 15:12:07, indicating that the subject bottleneck (between detectors 60 and 70) became active at that time. Fig. 8 confirms that a queue was present between detectors 50 and 80 for a much longer period. The figure also reveals that a spillover arrived at detector 80 when the slope of its N-curve dropped at about 16:29:47 (and later at detector 50), thus deactivating the bottleneck between detectors 60 and 70. The downstream queue's arrival at detector 80 is verified by Fig. 9, which reveals that the slopes of detector 80's re-scaled *N*- and T-curves dropped and rose, respectively, at about 16:29:47.

---

[5] Several 20-second periods contained detector errors on October 8, 1997; linear interpolation was used to repair the reported data.
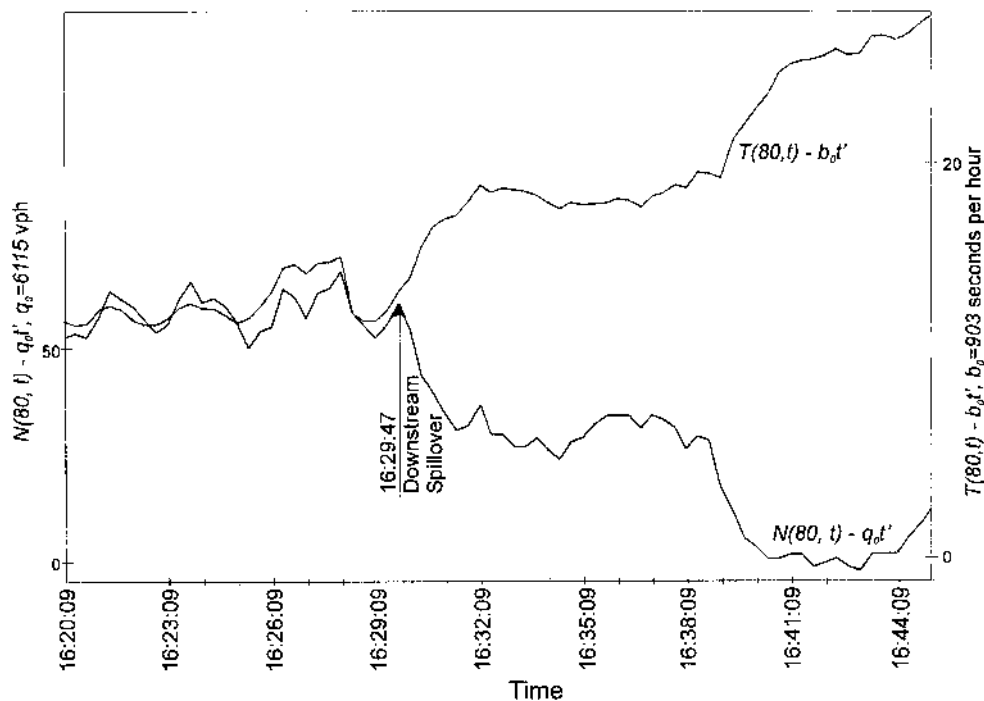
Fig. 9. Re-scaled N- and T-curves, detector 80, 10/8/97.

Fig. 10 shows re-scaled N- and T-curves for detector 80 spanning 2 h on October 8, 1997. As described, the bottleneck's active period began when a downstream queue dissipated at about 15:12:07 and ended when another spillover arrived at detector 80 at about 16:29:47. As shown in the figure, the discharge rate did not decrease with time. Rather, an alternating sequence of higher and lower rates prevailed, and a nearly constant average discharge rate (6070 vph) was measured.

Finally, Fig. 11 displays re-scaled N- and T-curves for detector 80 spanning more than three hours on November 26, 1998. On this day, the bottleneck became active upon dissipation of a downstream queue at 15:10:03 and was deactivated by a spillover at 18:08:03. Again, an alternating sequence of higher and lower rates was observed, where the average discharge rate (5950 vph) was again nearly constant.

Table 1 contains the summary of observations from four days when the bottleneck between detectors 60 and 70 was activated subsequent to the dissipation of a downstream queue, including the three days described above. Columns 2 and 3 of the table display the mean and index of dispersion of the 20-second counts (measured across all lanes) calculated from the periods during which the bottleneck remained active. The mean and index of dispersion do not vary considerably from day to day, providing further evidence that the average discharge flow is nearly constant. In addition, the vertical deviation between the *N*-curve and the dashed line was measured at each 20-second observation, and column 4 reports the maximum of this difference. In all cases, these maxima are less than about 80 vehicles.

On all four days, the flow observed prior to bottleneck activation was governed by the (lower) capacity of some downstream restriction. As shown in column 5 of Table 1, once the bottleneck
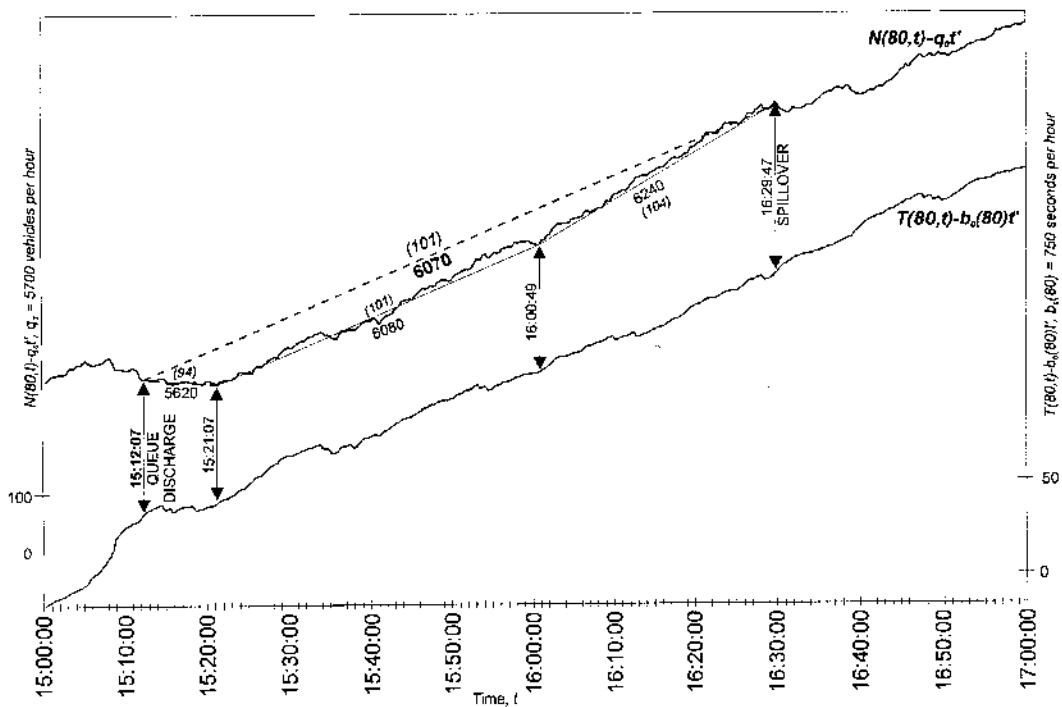
Fig. 10. Re-scaled N- and T-curves, detector 80, 10/8/97.

became active between detectors 60 and 70, the average discharge rates measured across all lanes exhibited small day to day deviations and were sustained for long periods (column 9).

The flows through the bottleneck were also studied in the individual lanes during the time that the bottleneck remained active (see Bertini, 1999). Columns 6, 7, and 8 of Table 1 contain the average discharge rates measured in the individual lanes on the four days studied. As shown, despite the absence of a nearby downstream diverge, the discharge flows were not balanced across all lanes. In fact (as further discussed in Cassidy and Bertini, 1999b), it was observed that large numbers of vehicles gradually moved into the median lane as they approached and passed through the bottleneck. These data indicate that, while the bottleneck was active, the average discharge flow varied across lanes, but the discharge rates in each lane exhibited only small variations across days.

Previous studies (Cassidy and Bertini, 1999a,b; Bertini, 1999) considered four different days at the same site (March 5, February 20, July 21 and February 11, 1997). On those days, there was no queueing in the study section just prior to the activation of the bottleneck, and a high flow was measured through the unobstructed bottleneck prior to queue discharge at a lower average flow (further explanation is provided in (Cassidy and Bertini, 1999a,b; Bertini, 1999). Despite this difference, several features were reproducible across the four days analyzed in previous studies and the four days presented here. On all eight days, the bottleneck on the Gardiner Express-way downstream of the Spadina Avenue on-ramp was activated between detectors 60 and 70 (see Fig. 1). Most importantly, once the Spadina Avenue bottleneck was activated, the average
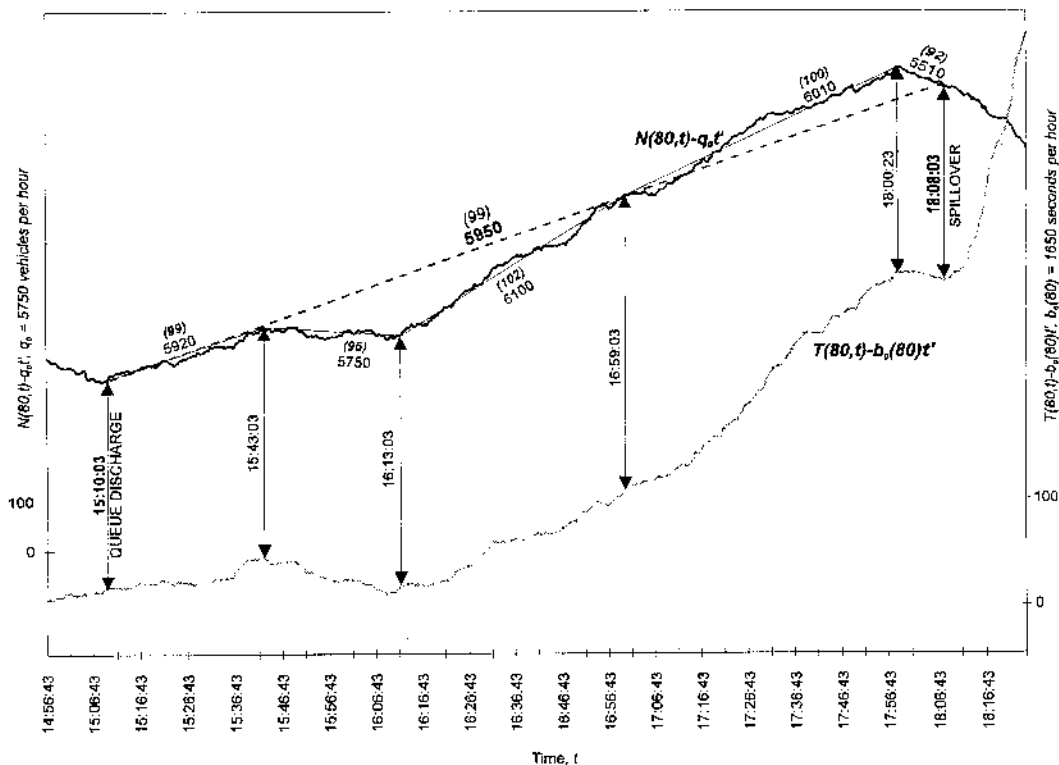
Fig. 11. Re-scaled N- and T-curves, detector 80, 11/26/97.

Table 1
Summary of queue discharge rates and features

| Date mm/dd/yy | 20-Second count | | Maximum vertical difference from trend | Average discharge rate | | | | Duration h:mm |
	Mean (vph)	Index of dispersion		Total (vph)	Median lane (vph)	Center lane (vph)	Shoulder lane (vph)	
03/12/97	32.4	0.35	77	5830	2320	1900	1610	1:46
10/08/97	33.9	0.41	71	6070	2350	1990	1730	1:17
11/26/98	33.1	0.32	72	5950	2290	1920	1730	2:58
01/17/97	32.2	0.31	20	5800	2390	1910	1560	0:41

queue discharge rates were consistent when measured across all lanes (and in the individual lanes), with a mean discharge flow of 5910 vph. These queue discharge rates always persisted for long periods.

Fig. 12 displays points representing the average discharge flows measured across all lanes (at detector 80) on the eight days studied. The magnitudes of the flows are represented by the vertical axis and the date of each observation is labeled. The mean discharge flow (also labeled on the figure) is shown as a horizontal line. In order to show where each observation falls with respect to
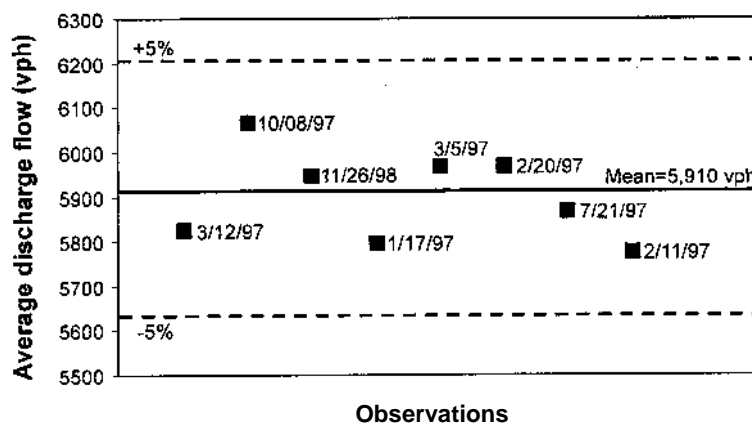
Fig. 12. Average discharge flow, detector 80.

a range of plus or minus five percent of the mean flow, dashed horizontal lines span the range from 95% to 105% of the mean discharge rate. The average discharge flow measured across all lanes only varied between -2.2% and +2.7% of its mean rate.

5. Final comments

This study's findings have been made possible by the use of transformed cumulative curves to pinpoint the bottleneck location, to guarantee that the bottleneck was active and to diagnose some of the traffic features that prevailed. The analysis of data collected from the Gardiner Expressway bottleneck has shown that on four particular days, the subject bottleneck was activated only after a queue emanating from somewhere further downstream had dissipated. On such days, vehicles immediately began discharging through the bottleneck at a nearly constant rate without exhibiting an especially high flow or a particularly low discharge flow upon queue discharge.

As a result of studying data collected from the Gardiner Expressway bottleneck on a total of eight days, it was shown that the bottleneck always occurred at a fixed location approximately one kilometer downstream of the merge. Most importantly, whether or not the bottleneck was initially impacted by the presence of queues from further downstream, the discharge flow in the active bottleneck was nearly constant, since the cumulative counts never deviated much from a linear trend. Moreover, its long-run discharge flow was reproducible from day to day in each lane and in total. This further suggests that the long-run queue discharge flow should be viewed as the bottleneck capacity given that the nearly constant rates were sustained for prolonged periods and that they were replicated (approximately) each day. Fortunately, many freeways are instrumented with loop detectors, making it possible for a jurisdiction to estimate capacity values for a particular bottleneck empirically as an alternative to using capacity values prescribed by nationwide handbooks such as the Highway Capacity Manual.

Further analyses of traffic data are required to address some unresolved issues and to confirm the reproducibility on other sites of the observed traffic features described herein. This study neither analyzed the propagation of bottleneck queues beyond the upstream limits of the study

site nor did it attempt to analyze the details of queues arriving from further downstream. It is important to study the details of queue propagation, and Cassidy and Mauch (2001) present a study of one such queue. Their research to understand the properties of the instabilities arising in queued traffic is still ongoing.

As shown in this study, queue discharge flows consisted of sequences of nearly uniform flow that gradually changed over time. From the loop detector data it was not clear why these changes existed in the traffic stream. Thus, cumulative curves constructed from loop detector data could be augmented with vehicle trajectories extracted from video surveillance, perhaps using machine vision technology (Coifman, 1997). If possible, the video should be obtained from a vantage point that is high enough such that a large section of freeway can be observed. Comparing transformed cumulative curves with vehicle trajectories extracted from video for limited time periods may help explain why the discharge rates vary with time. It may also be possible to explain the detailed mechanism of queue formation at an individual vehicle level. Such explanations would contribute toward the development of future theories of traffic flow and strategies for managing freeway traffic.

Acknowledgements

References

Bertini, R.L., 1999. Time-dependent traffic flow features at a freeway bottleneck downstream of a merge. Ph.D. Thesis, University of California, Berkeley, CA.

Cassidy, M.J., Bertini, R.L., 1999a. Some traffic features at freeway bottlenecks. Transportation Research B 33, 25-42.

Cassidy, M.J., Bertini, R.L., 1999b. Observations at a freeway bottleneck. In: Proceedings of the Fourteenth International Symposium on Transportation and Traffic Theory, Jerusalem, Israel, pp. 107-124.

Cassidy, M.J., Mauch, M., 2001. An observed feature of long freeway traffic queues. Transportation Research A 35, 149-162.

Cassidy, M.J., Windover, J.R., 1995. Methodology for assessing dynamics of freeway traffic flow. Transportation Research Record 1484, 73-79.

Coifman, B., 1997. Time space diagrams for thirteen shock waves. California PATH Working Paper UCB-ITS-PWP-97-1, Institute of Transportation Studies, University of California, Berkeley, CA.

Daganzo, C.F., 1997. Fundamentals of Transportation and Traffic Operations. Elsevier, New York.

Newell, G.F., 1982. Applications of Queueing Theory. Chapman Hall, London.

Newell, G.F., 1993. A simplified theory of kinematic waves in highway traffic I: General theory. II: Queueing at freeway bottlenecks. III: Multi-destination flows. Transportation Research B 27, 281-313.