Title
Time-Dependent Traffic Flow Features at a Freeway Bottleneck Downstream of a Merge

Author
Bertini, Robert Lawrence

Publication Date
1999

Time-Dependent Traffic Flow Features at a Freeway Bottleneck Downstream of a Merge

by

Robert Lawrence Bertini

B.S. (California Polytechnic State University, San Luis Obispo) 1988
M.S. (San Jose State University) 1991

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering-Civil and Environmental Engineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Michael J. Cassidy, Chair
Professor Gordon F. Newell
Professor John A. Rice

Fall 1999

Time-Dependent Traffic Flow Features at a Freeway Bottleneck Downstream of a Merge

by

Robert Lawrence Bertini

ACKNOWLEDGMENTS

Sincere appreciation is extended to Professor Michael Cassidy for supervising this dissertation; his time, energy and support were invaluable. Very special thanks to Professor Gordon Newell for his enthusiasm, guidance and for truly caring about his students. Many thanks to: Dr. Alex Skabardonis for support on earlier projects and for providing excellent advice; Professor Carlos Daganzo for his kind tutelage and guidance; Professor John Rice for serving on my committee; and Professor Jan Botha for his mentorship.

My graduate studies would not have been possible without the generous financial support of the Eisenhower Graduate Fellowship Program, the University of California Transportation Center, the Institute of Transportation Studies, California PATH and Caltrans. David Nesbitt, City of Toronto, graciously supplied the data. Thanks to the people who make Berkeley special, particularly Professors Martin Wachs and William Garrison, Catherine Cortelyou, Mari Cook and Briggs Nisbet.

I am thankful for all the Berkeley students with whom I have been privileged to study. Special thanks to C. Atienza, A. Erera, R. Garcia, T. Lawson, F. Long, D. Lovell, M. Mauch, L. Melendy, J.C. Muñoz, K. Petty, K. Smilowitz, P. Thananjeyan, C. Venter, J. Windover and S. Young for their contributions to me as a person and as a scholar.

It is with great joy that I offer my profound gratitude to the amazing Joëlle De Jonge for the precious gifts of her love, motivation and inspiration. Finally, none of this would have been possible without the extraordinary support and encouragement of my parents Tullio and Jeanne, my brother Bil and my grandmother Nida Canida Daly. It is to them that this dissertation is dedicated with much love and gratitude.

Contents

1 Introduction **1**

1.1 Background 3

1.2 Motivation and research scope 7

1.3 Organization of the dissertation 7

2 Data **9**

3 Methodology **11**

3.1 Curves of cumulative vehicle count versus time 11

3.2 Curves of cumulative occupancy versus time 18

4 Analysis **20**

4.1 Preliminary data analysis 21

4.2 Day 1: Bottleneck activation preceded by freely-flowing traffic 22

4.2.1 Day 1: Diagnosis of the active bottleneck 22

4.2.2 Day 1: Lane changing and high median lane flow 31

4.2.3 Day 1: Bottleneck input flows 35

4.2.4 Day 1: Queue discharge features 40

4.2.5 Day 1: Reproducing the observations 47

4.3 Day 2: Bottleneck activation preceded by a queue 48

4.3.1 Day 2: Diagnosis of the active bottleneck 49

4.3.2 Day 2: Queue discharge features 53

4.3.3 Day 2: Reproducing the observations 55

4.4 Reproducing the observations: All days 56

5 Conclusions **61**

5.1 Summary of the study's findings 61

5.2 Areas of further research 63

References 65

Appendix A - Methods for processing traffic data 68

Appendix B - Analysis of the Queen Elizabeth Way 71

Appendix C - Weather data for Gardiner Expressway 79

Appendix D - Comparison of loop detector data with video data 80

Appendix E - Gardiner Expressway observed wave speeds 82

List of Figures

1 Time series plot of flow. 5

2 Definition of an active bottleneck. 5

3 Gardiner Expressway, Toronto, Canada. 9

4 Hypothetical cumulative vehicle arrival curve and piecewise linear approximation. 11

5 Hypothetical cumulative vehicle arrival curves at two neighboring detector locations. 12

6 Hypothetical queueing diagram. 13

7 Day 1 Shifted N-curves, detectors 60 and 80. 14

8 Day 1 Transformed N-curves, detectors 60 and 80. 16

9 Day 1 Re-scaled N- and T-curves, detectors 60 and 80. 17

10 Day 1 Transformed N-curves, detectors 40 through 80. 23

11 Day 1 Transformed T-curves, detectors 40 through 80. 25

12 Day 1 Re-scaled N- and T-curves, detector 40 26

13 Day 1 Transformed N-curves, detectors 60 and 80 28

14 Day 1 Re-scaled N- and T-curves, detector 80. 29

15 Day 1 Re-scaled N- and T-curves, detector 40, shoulder lane. 30

16 Day 1 Re-scaled N- and T-curves, detector 60, shoulder lane. 31

17 Day 1 Re-scaled N-curves, median lane, stations 30–90, Gardiner Expressway. 32

18 Day 1 Re-scaled N-curves, center lane, stations 60–80, Gardiner Expressway. 34

19 Day 1 Re-scaled N-curves, shoulder lane, stations 60–80, Gardiner Expressway. 36

20 Day 1 Re-scaled N-curve, Jameson Avenue off-ramp. 37

21 Day 1 Re-scaled N-curves, all lanes, stations 60–80, Gardiner Expressway. 38

22 Day 1 Re-scaled N-curves, station 60 of the Gardiner Expressway. . . 39

23 Day 1 Re-scaled N- and T-curves, station 80 of the Gardiner Expressway. 41

24 Day 1 Re-scaled N-curves for individual lanes, Station 80. 44
25 Day 1 Re-scaled N-curve, Spadina Avenue on-ramp. 45
26 Day 2 Transformed N-curves, detectors 40 through 80. 50
27 Day 2 Re-scaled N- and T-curves, detector 50, median lane. 51
28 Day 2 Re-scaled N- and T-curves, detector 60, median lane. 52
29 Day 2 Re-scaled N- and T-curves, station 80 of the Gardiner Expressway. 53
30 Gardiner Expressway bottleneck location 56
31 Reproducibility of queue discharge flows 58
32 Time series plot of count. 69
33 Bivariate plot of occupancy (in percent) versus flow. 69
34 Occupancy (in percent) versus time and space. 70
35 Queen Elizabeth Way, Toronto, Canada. 71
36 Re-scaled N-curves, Queen Elizabeth Way. 72
37 Re-scaled T-curves, Queen Elizabeth Way. 74
38 Upstream and downstream N-curves, Queen Elizabeth Way. 75
39 Re-scaled N- and T-curves, station 25 of the Queen Elizabeth Way. . 77
40 Comparison of vehicle counts, median lane, April 23, 1997. 81
41 Re-scaled N-curves, video versus loop, April 23, 1997. 81

List of Tables

1	Gardiner Expressway days analyzed.	21
2	Day 1 Gardiner Expressway queue discharge features.	43
3	Summary of high flows in the median lane on the Gardiner Expressway.	47
4	Summary of measured traffic features on the Gardiner Expressway on days when a flow drop is observed.	48
5	Day 2 Gardiner Expressway queue discharge features.	54
6	Summary of Gardiner Expressway queue discharge flows as seen across all lanes.	55
7	Summary of Gardiner Expressway queue discharge flows as seen in individual lanes.	57
8	Variance-to-mean ratios of queue discharge flows as seen in individual lanes.	60
9	Summary of measured traffic features on the Queen Elizabeth Way.	77
10	Toronto weather.	79
11	Day 1: Gardiner Expressway wave speeds as seen in individual lanes.	82
12	Day 1: Gardiner Expressway flow drops as seen in individual lanes.	82

1 Introduction

This dissertation describes an empirical study of the spatial and temporal evolution of traffic conditions upstream and downstream of a freeway bottleneck located near a busy on-ramp. To this end, curves of vehicle arrival number and curves of measured occupancy were plotted cumulatively versus time using data from neighboring freeway loop detectors. Once transformed in ways described herein, these cumulative curves provided the measurement resolution necessary to observe the transitions from freely-flowing to queued conditions and to identify some notable, time-dependent traffic features in and around the bottleneck.

It is shown that the bottleneck was located more than 1,000 meters (more than 3,500 feet) downstream of the busy merge. On certain days, queues formed at this bottleneck without prior interference from any other queues emanating from further downstream. On such days, bottleneck flows over 6,600 vehicles per hour (vph) were measured across all three freeway lanes. These high flows were sustained each day for up to 40 minutes before a queue eventually formed immediately upstream of the bottleneck, giving rise to lower average discharge rates.

Notably, the initiations of the bottleneck's queues were always signaled by subtle, yet reproducible traffic features. For example, by viewing flows in individual lanes, it is shown that large numbers of vehicles steadily moved into the median (left) lane as drivers approached and passed through the bottleneck. As a consequence, median lane flows as high as 2,600 vph persisted for up to 40 minutes at locations well downstream of the on-ramp. Furthermore, just prior to queueing, flows were approximately evenly distributed across the center and shoulder lanes and the total flows measured across all three lanes surged to approximately 7,000 vph; these surges were sustained for only about two minutes before the queues arose.

Upon queue formation, the flow dropped such that the the average queue discharge rates measured over the entire rush were as much as ten percent lower than the high flows observed (for sustained periods) prior to the onset of the upstream queue. In addition, for the first 20 minutes of queue discharge, a more sizable flow drop was typically observed which was followed by a higher *recovery* discharge flow. After this initial sequence, the discharge flow usually settled into a nearly-constant rate.

It is also shown that vehicles entered the freeway from the upstream on-ramp at very high rates, even after the bottleneck's queue propagated beyond the ramp and obstructed this flow. Thus, the on-ramp vehicles did not share the available capacity with vehicles in the adjacent freeway lane in a strictly alternating or one-to-one basis. Rather, motorists from the on-ramp forced themselves into the queue in such a way that the gaps between the freeway vehicles were filled by multiple on-ramp vehicles.

The flow drop accompanying queue formation may suggest that there is some potential to increase bottleneck capacity by deploying traffic control strategies, such as on-ramp metering. Further, the presence of tell-tale signals of queueing suggests that such strategies might be readily operated in an on-line fashion. The initial periods marked by especially low discharge rates further underscore the potential benefits of these.

On other days, however, the subject bottleneck was activated only after a queue emanating from somewhere further downstream had dissipated. In these instances, queued vehicles discharged from the subject bottleneck without ever exhibiting a higher flow and without exhibiting especially low discharge rates at the onsets of the bottleneck's queue discharge.

But on all days examined, the bottleneck always arose at the same location (more than 1,000 meters, or 3,500 feet, downstream of the on-ramp). That this bottleneck

occurred so far from the ramp is noteworthy since most previous studies of bottleneck flows have relied only upon data taken in the immediate vicinity of merge activity. Although the queue discharge rates did exhibit certain time-dependencies, such as the initial low flows that sometimes accompanied queue formations, the overall discharge rates exhibited nearly stationary patterns that slowly alternated about a fixed rate. Thus, these flows might be described as *nearly constant*. Reassuringly, the average discharge rate measured during each rush exhibited only small deviations from one day to the next.

Given these predictable features of the bottleneck, it seems reasonable to postulate about how queues evolve upstream. The reproducible features further suggest that the long-run (daily average) queue discharge rate should be viewed as the bottleneck capacity; its near-constant rate was sustained for prolonged periods and was reproduced each day. In contrast, for the days examined, the higher flows sometimes measured immediately prior to the bottleneck queue's formation exhibited daily variations in their magnitudes and persisted for relatively short periods with durations that likewise varied each day.

1.1 Background

Though bottlenecks are the critical elements of a freeway system, over the past 50 years little progress has been made toward understanding their behavior. There have been a number of past studies of freeway bottlenecks. For example, some studies have reported that freeway capacity diminishes by an average of 3 or 4 percent following the formation of an upstream queue (Agyemang-Duah and Hall, 1991; Banks, 1990, 1991) while other studies report no such reductions (Hall and Hall, 1990; Persaud, 1986; Persaud and Hurdle, 1991; Newman, 1961).[1] Still other studies

[1] In fact, two studies of the same bottleneck have claimed both extremes (Agyemang-Duah and Hall, 1991; Hall and Hall, 1990).

have reported that queue discharge rates decrease with time (Koshi, *et al.*, 1992). However, throrough examination of these past studies indicates that these apparent contradictions may be explained by the limitations of the studies themselves.

Specifically, several studies of bottleneck flows have been limited by the absence of automated data collection equipment; the data were measured for rather short durations and over limited roadway lengths. Palmer (1959), for example, examined how changing conditions propagated in highway traffic by viewing cumulative vehicle arrival curves measured at neighboring points upstream and downstream of a bottleneck. Likewise, Persaud (1986) constructed a small number of vehicle trajectories from a series of aerial photos taken in the vicinity of a freeway bottleneck.

Additionally, other studies have been plagued by the limitations of the methods used for processing measured traffic data.[2] As one example, many previous studies (Hall and Hall, 1990; Koshi, *et al.*, 1992; Roess and Ulerio, 1994; Kerner and Rehborn, 1997; Persaud, *et al.*, 1998) relied upon plots of $q(t)$ versus t, flow versus time, from raw loop detector data measured over short intervals (such as 20 seconds). A sample of this type of data treatment is shown in Figure 1, where the flow (in vph) measured during each 20-second interval was plotted versus time as the solid line. This measured variable exhibits statistical fluctuations along with possible time-dependent changes in the average; however, this type of data treatment does not always reveal the distinctions between the two. In this case, note that the flows in Figure 1 have a range of more than 3,000 vph during the 30-minute period shown. If one attempted to superimpose linear approximations of average flow on this graph, the wide fluctuations would make it difficult to delineate the time-dependent changes in flow.

[2]See Appendix A for brief illustrations of some typical methods.

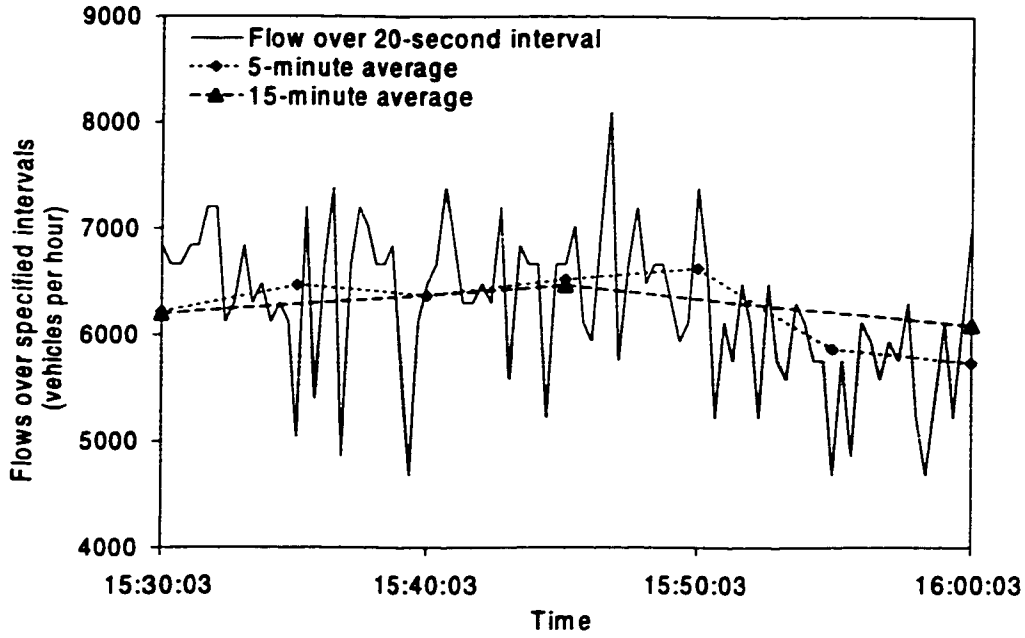

Figure 1: Time series plot of flow.

In some previous work, a common technique for attempting to reduce statistical fluctuations in plots of $q(t)$ versus t has been to examine the data over longer, pre-specified intervals. Figure 1 includes example curves of 5-minute (short dashed line) and 15-minute (long dashed line) averages of the raw detector data. Unfortunately, these treatments can average out the features of interest, reveal little about changing flow states and often include data from two distinct flow states as one data point. These characteristics can also apply when using moving averages.

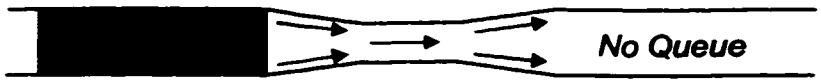

Figure 2: Definition of an active bottleneck.

In particular, one finds in the literature that such limitations in the methods used to process loop detector data have hampered efforts to examine traffic variables measured downstream of an *active* bottleneck (as defined by Daganzo, 1997), whereby vehicles discharge from an upstream queue (to guarantee that the bottleneck serves

vehicles at a maximum rate) and vehicles are unimpeded by traffic conditions emanating from further downstream (see Figure 2). For example, Hall and Hall (1990) studied a bottleneck on the Queen Elizabeth Way (QEW) in Toronto, Canada in the vicinity of the Cawthra Road on-ramp. Appendix B illustrates that the counts used to study discharge flows were taken from a point (detector number 22) that is actually located upstream of the bottleneck.

Other studies may have found the bottleneck's location (Agyemang-Duah and Hall, 1991; Koshi *et al.*, 1992), but have not confirmed whether or not the bottleneck was deactivated by queues that may have eventually spilled over from further downstream. For example, Agyemang-Duah and Hall (1991) also studied the bottleneck on the QEW in the vicinity of the Cawthra Road on-ramp. That study did examine vehicle counts measured downstream of the active bottleneck, but did not confirm whether the observed changes in flow arrived from upstream or if they were instead caused by downstream queues.

In sum, these limitations suggest that past findings have been inconclusive and in particular, conclusions that bottleneck capacities can be increased by metering on-ramps (Persaud, *et al.*, 1998) may have been premature.

To overcome the obstacles of past research, it is the purpose of this dissertation to achieve a greater understanding of bottleneck flow features via the detailed examination of one bottleneck where loop detector data are readily available. Toward this end, this study visually examined transformed curves of cumulative vehicle arrival number versus time and cumulative occupancy versus time measured at neighboring detectors.

1.2 Motivation and research scope

Given that little is known about how traffic flows through bottlenecks, a greater understanding is needed to formulate, to enhance or to verify mathematical models of vehicular traffic, so that they are consistent with the actual traffic features that are found to be reproducible. This understanding is also required before one can conclude whether or not bottleneck flows can be increased by eliminating or postponing freeway queues with control measures such as ramp metering.

Toward this end, this study observed the spatial and temporal evolution of traffic conditions for eight afternoon peak periods upstream and downstream of a freeway bottleneck located downstream of a merge in Toronto, Canada. Because the study used data that were automatically collected by the City of Toronto's freeway surveillance system (identical to ubiquitous data that are collected by numerous traffic management centers worldwide), a special data collection effort was not required and data from numerous days were readily available for assessing the reproducibility of the observations. The surveillance system provided loop detector and video data both upstream and downstream of the bottleneck, making it possible to study the time-dependent bottleneck flow processes. This study complements the work of Windover (1998) who observed in detail the dynamic features of freeway traffic upstream of a bottleneck. Time-dependent bottleneck flow processes were not examined in that earlier study because data were not available in the immediate vicinity of the bottleneck, nor downstream of the bottleneck.

1.3 Organization of the dissertation

The remainder of the dissertation is organized as follows. Section 2 describes the study site and the data collected there. Section 3 describes the use of cumulative curves for identifying certain details of freeway traffic evolution. Section 4 presents

the findings, including the bottleneck's location, the times that it remained active, certain reproducible features that were observed in uncongested traffic prior to queue formation, cues that signalled the impending onset of queueing and the reproducible features of queue discharge. In addition to recounting the findings of this study, Section 5 provides an outline of some areas of further research.

2 Data

The data were collected from the 3.3-kilometer (2.1-mile) segment of westbound Gardiner Expressway (in metropolitan Toronto, Canada) shown in Figure 3, in the vicinity of the Spadina Avenue on-ramp. Loop detectors record vehicle counts, occupancies[3] and time mean speeds in each lane over 20-second intervals. The City of Toronto synchronizes the clocks among detector controllers from their Traffic Management Center (TMC), thus no special clock synchronization strategy was necessary. The detector stations (and the video cameras) are labeled here according to the numbering strategy adopted by the City of Toronto's Road Emergency Services Communication Unit (RESCU). Tapes of concurrent video surveillance for several weekdays were obtained from the cameras shown in Figure 3.

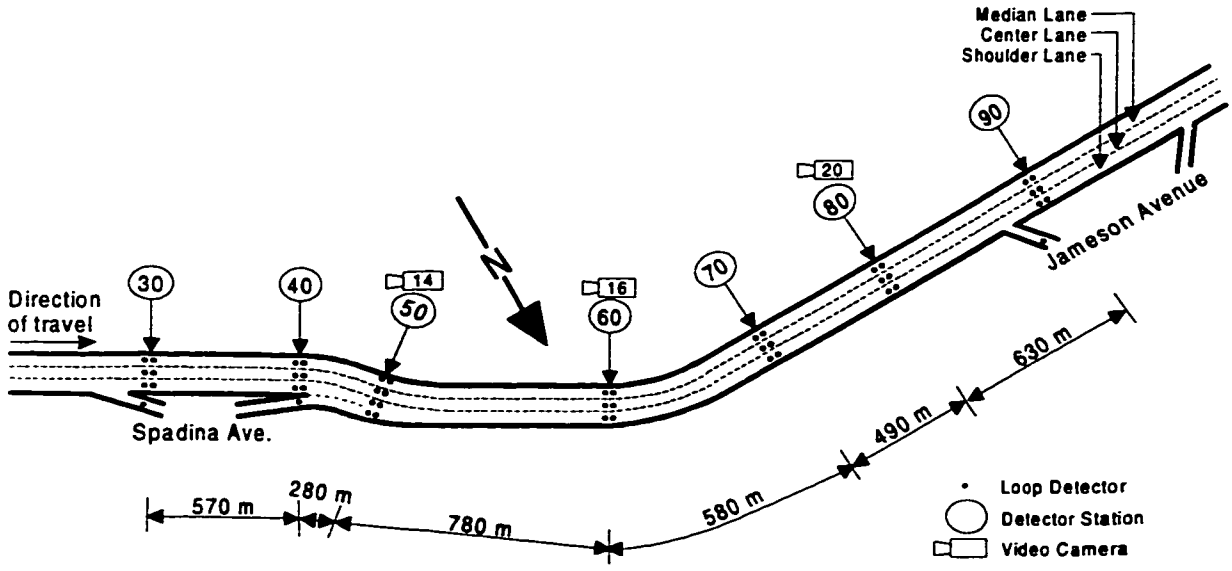

Figure 3: Gardiner Expressway, Toronto, Canada.

Ramp metering is not installed at this site, although the Jameson Avenue on-ramp is closed between 15:00 and 18:00 each afternoon. The lanes in the study site

[3]Measured occupancy is the percentage of time a detector is covered by a vehicle in the twenty-second measurement interval.

are referred to, from left to right, as the median lane, the center lane and the shoulder lane. The freeway is situated on an elevated structure, it has (essentially) no shoulders, and it is signed for a speed limit of 90 kilometers per hour (56 miles per hour).[4] Others have estimated that the traffic stream consists of six percent trucks (Persaud, *et al.*, 1998) and by statute, trucks are not permitted in the median lane. Buses are visible in the traffic stream (from video surveillance) during peak periods, even in the median lane. Appendix C contains weather (temperature and precipitation) data for Toronto on the days studied. As an aside, the site has been used in several earlier studies of freeway capacity (Persaud, 1986; Persaud and Hurdle, 1991; Persaud, *et al.*, 1998).

[4]The horizontal curve between stations 60 and 70 has a radius of 800 meters (2,600 feet) which is deemed to be comfortable for vehicular travel at 105 kilometers per hour (65 miles per hour) under ideal conditions, according to California Department of Transportation (1995).

3 Methodology

To promote the visual identification of time-dependent features of the traffic stream, this study used transformed curves of cumulative vehicle count and curves of cumulative occupancy constructed from data measured at neighboring freeway loop detector stations (Cassidy and Windover, 1995). These curves are described in the following two subsections.

3.1 Curves of cumulative vehicle count versus time

Figure 4 shows a hypothetical curve of $N(x,t)$, the cumulative number of vehicles to arrive at detector location x by time t. By constructing piecewise linear interpolations to this stepwise curve, the time derivative (slope) is the flow past station x during the 20-second measurement interval containing the t. It would actually be preferable to obtain the arrival time of each vehicle (Windover, 1998); but most available detector data are aggregated over 20 or 30 seconds and this was sufficient resolution for this study.

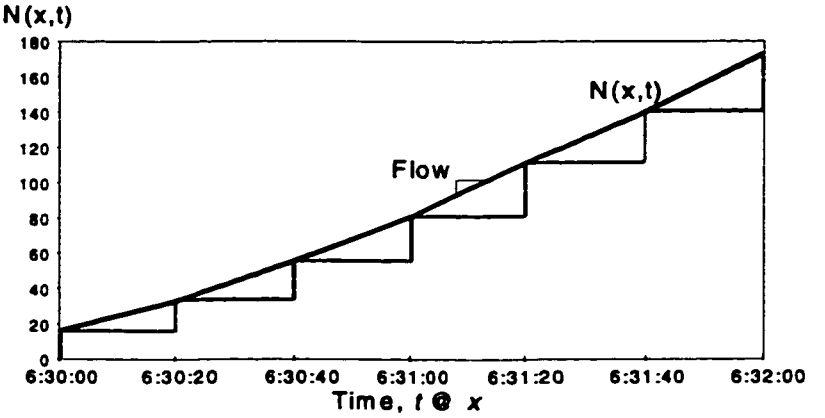

Figure 4: Hypothetical cumulative vehicle arrival curve and piecewise linear approximation.

Cumulative curves constructed at two neighboring detector locations can be used to see how trip times and vehicle accumulations changed with time over the intervening roadway segment. Figure 5 illustrates two hypothetical N-curves constructed in series, where the counts for each curve began (N=0) with the passage of some reference vehicle and vehicles were conserved between the two detectors (Newell, 1982).[5] To visualize how the N were constructed from the same collection of vehicles, consider that an upstream observer at location x_1 recorded and cumulatively plotted the arrival times of vehicles as they passed x_1. The times at which these same vehicles passed downstream x_2 were also recorded (and plotted). In the absence of vehicle overtaking maneuvers (so that a first in, first out discipline was maintained) and if vehicles were conserved between measurement locations, the horizontal distance between curves at $N = j$ is the jth vehicle's trip time from x_1 to x_2 and the vertical distance between curves at time t_1 is the vehicle accumulation between x_1 and x_2 at that time.

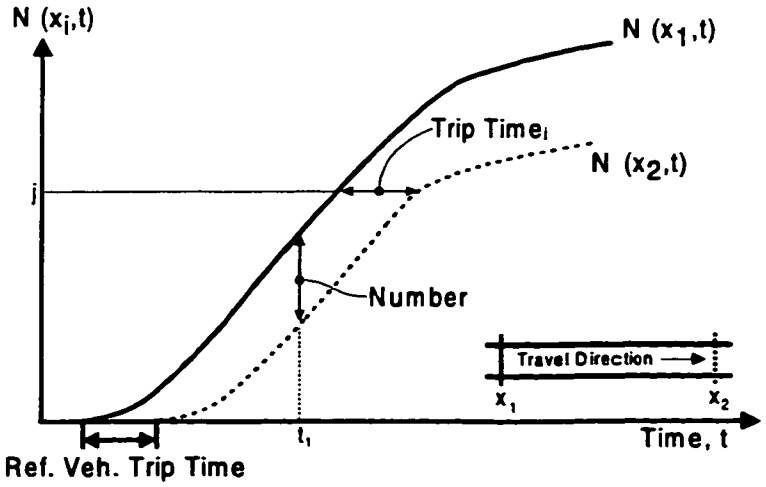

Figure 5: Hypothetical cumulative vehicle arrival curves at two neighboring detector locations.

[5]Cassidy and Windover (1995) provide discussion on how a reference vehicle can be identified (approximately) from data measured by neighboring loop detectors. Henceforth, when two or more N-curves are shown on the same graph, and vehicles are conserved between their measurement locations and their counts began with the passage of some reference vehicle unless otherwise noted.

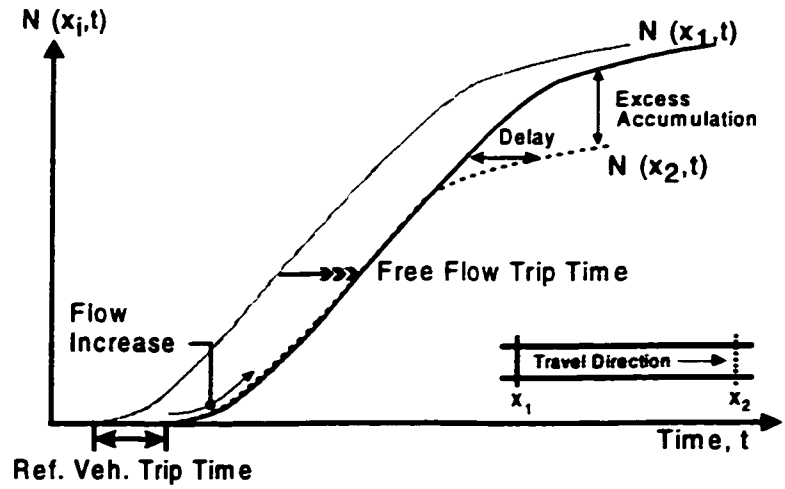

Figure 6: Hypothetical queueing diagram.

As shown in Figure 6, a queueing diagram can be constructed by shifting the upstream curve horizontally to the right by the presumed average free-flow trip time between x_1 and x_2 (Newell, 1993). In the absence of delay, the two curves are superimposed and any flow features (such as the initial flow increase shown in Figure 6) are passed downstream with the vehicles. Any remaining horizontal displacements between the shifted curves are the vehicular delays and the vertical distances are the excess vehicle accumulations between x_1 and x_2 due to delays. This horizontal shift is advantageous because the delay and excess vehicle accumulation can be seen directly from the figure.

Figures 5 and 6 are idealized because N-curves that are constructed from actual freeway counts collected over an extended period (such as 20 minutes or more) and plotted on standard-sized paper seldom reveal changing traffic conditions in a pronounced manner. For example, Figure 7 presents shifted N-curves constructed from approximately 30 minutes of counts at detector stations 60 and 80 on the Gardiner Expressway; these counts were measured across all three lanes during the afternoon

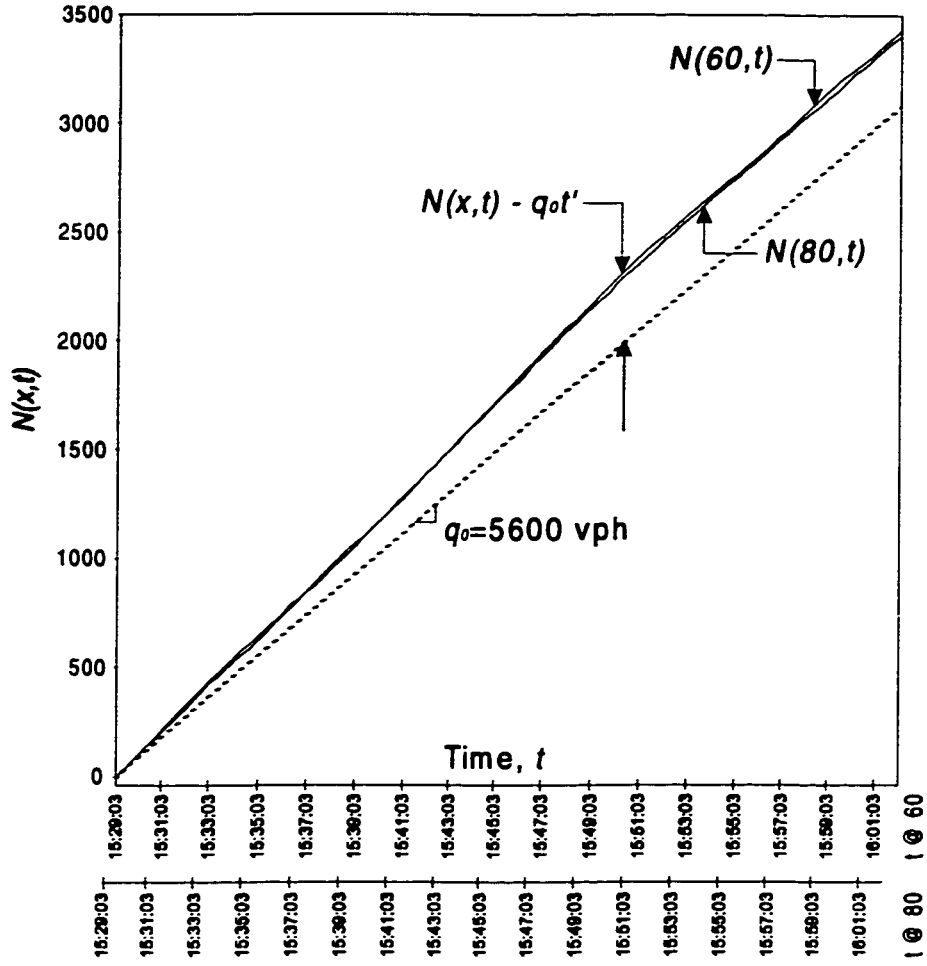

Figure 7: Day 1 Shifted N-curves, detectors 60 and 80.

of March 5, 1997. Shifted along with the upstream curve was its corresponding time axis and separate axes for each x are shown in Figure 7.

The scale of Figure 7 makes it difficult to observe changing flows and vehicle accumulations. In fact, the two N-curves appear to be essentially straight lines that are nearly superimposed. To promote the visual identification of the time-dependent features of the traffic stream, the N-curves were then re-scaled by subtracting a background cumulative count, $q_o \cdot t'$, at all time t, where q_o is defined as the back-

ground flow and t' is the elapsed time from the start of the upstream curve.[6] With a suitable choice of q_o, this re-scaling will magnify details without changing the excess accumulations (Cassidy and Windover, 1995; Cassidy, 1998). Note that a re-scaled cumulative curve will display the vertical displacements between the original curve and a chord (such as the line $N = q_o \cdot t'$). Thus, some diagnoses made with the re-scaled curves are independent of the magnitude of the background rate reduction (slope of the chord), including diagnoses based on the presence or absence of displacements between neighboring N-curves; the convexity or concavity of a curve; and the vertical deviations between a re-scaled curve and some chord. The re-scaling can, however, alter certain attributes of a curve, such as the magnitudes of its local minima and maxima and the times at which they occur. Generally speaking, the times marking visible flow changes on a re-scaled N-curve are independent of the magnitude of the background rate reduction as long as this rate is within approximately 20 percent of the prevailing average flow (Windover, 1998). Also of note, a re-scaled N-curve may exhibit negative slopes denoting prevailing flows less than the specified background flow q_o.

Figure 8 shows the transformed N-curves for the same period as in Figure 7, with the new vertical scale indicated on the y-axis. A background flow of 5,600 vph was chosen and as such, the minor deviations in N and the changes in excess vehicle accumulation can be readily seen from the curves. In Figure 8, the times at which long-run flow changes occurred are marked on the $N(60, t)$ with vertical arrows. Note that the start and end times for each period of near-constant flow were selected *by eye*. The estimated flow changes are also marked on the $N(60, t)$ by piecewise linear approximations and the flows measured during each interval are labeled in vph.[7]

[6] Henceforth, the term *rescaled N-curve* will describe a single N-curve (not necessarily constructed in series with other N-curves) from which a background flow has been subtracted. *Transformed curves* refer to those constructed in series that have been both shifted and rescaled.

[7] Henceforth, N-curves will often be annotated with piecewise linear approximations which are labeled with the corresponding flows in vph, unless otherwise noted.

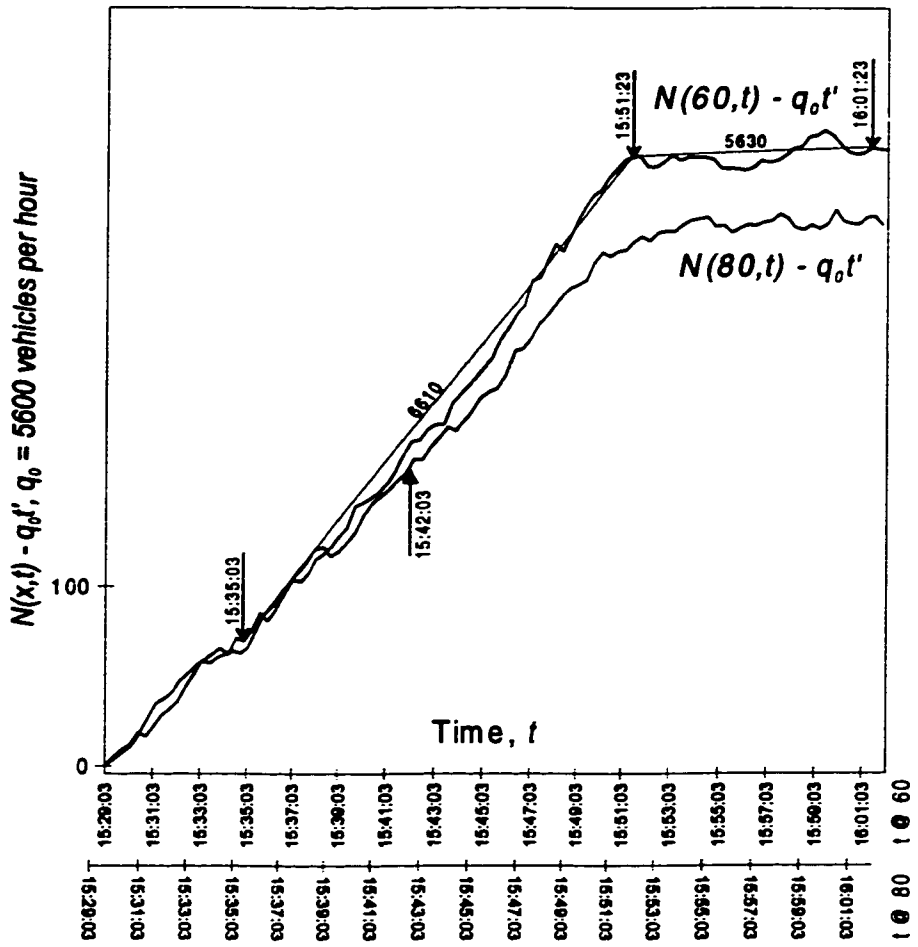

Figure 8: Day 1 Transformed N-curves, detectors 60 and 80.

In Figure 8 it is clear that traffic was initially freely flowing between stations 60 and 80 because the initial portions of the two curves are approximately superimposed, between 15:29 and 15:42. The vertical displacements between these transformed N-curves after about 15:42 clearly reveal the presence of vehicular delay, and hence the presence of a queue between stations 60 and 80 after that time.

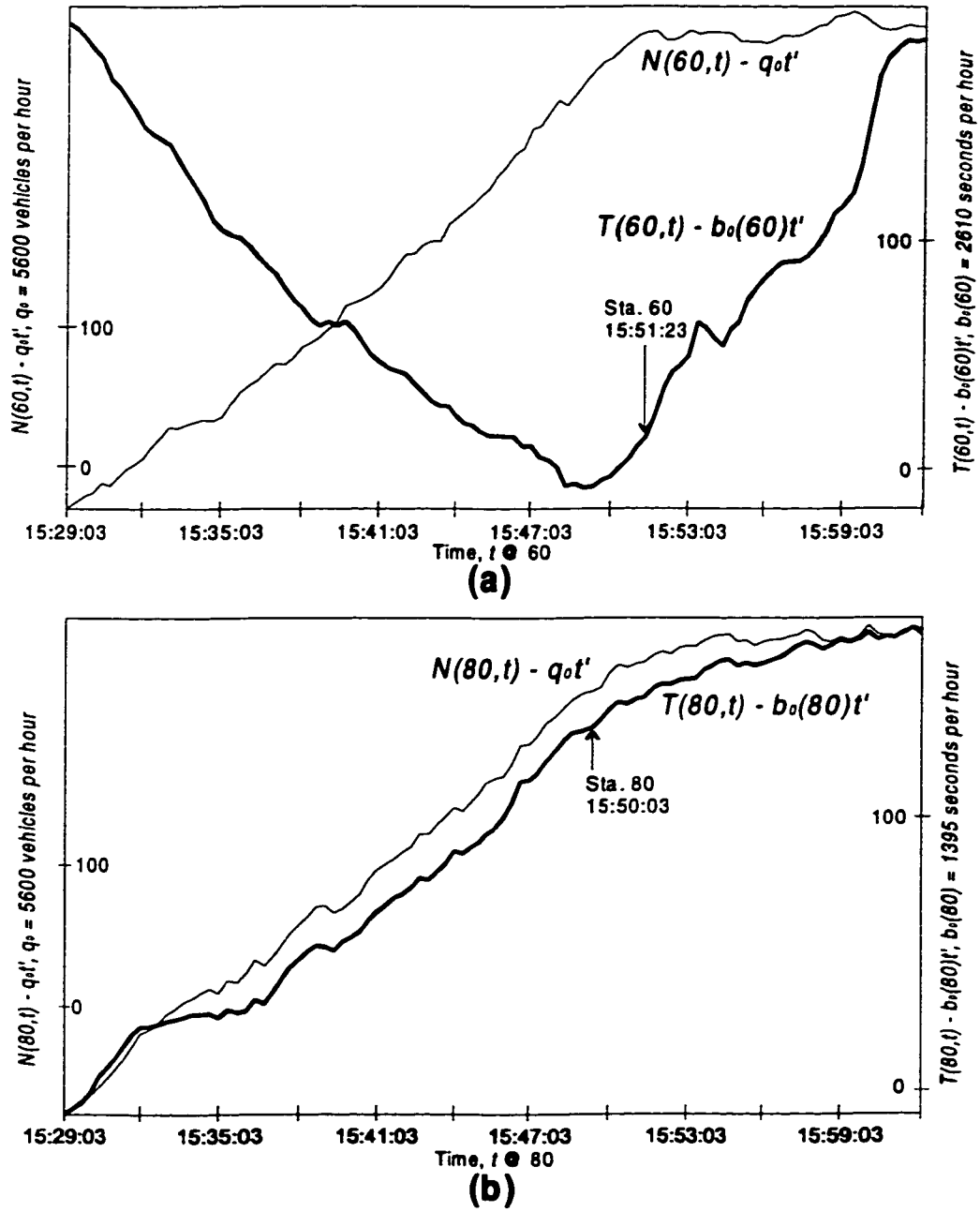

Figure 9: Day 1 Re-scaled N- and T-curves, detectors 60 and 80.

3.2 Curves of cumulative occupancy versus time

Curves of cumulative occupancy versus time were also used in this study to corroborate features observed from re-scaled N-curves and to identify periods of near-stationarity. Figures 9(a) and (b) again present the re-scaled N-curves for station 60 and 80, respectively, that were shown in Figure 8. These are shown in Figure 9 as light solid lines. Each station's re-scaled N-curve was plotted on a separate graph and was paired with its corresponding curve of $T(x,t) - b_o(x) \cdot t'$ (shown as dark lines), where:

1. $T(x,t)$ = the cumulative occupancy at station x to time t, measured in units of total vehicle trip time spent over the loop detectors (across all three lanes) by time t.[8]; and

2. $b_o(x)$ = the background rate used for re-scaling the T-curve (at a given x) to promote the visual identification of its time-dependent features.

The t' at each x is the elapsed time from the arbitrary starting point used for the curve; thus, a T-curve's starting time may also be viewed as the passage time of some reference vehicle. As with the N-curves, piece-wise linear approximations to the $T(x,t)$ were constructed so that the slope of the curve at some t is the occupancy rate at detector station x during the measurement interval containing that t.

Figure 9(a) reveals that at station 60, a reduction in flow was accompanied by increased occupany rates. These features are consistent with what would be expected to accompany the arrival of a backward-moving queue. Figure 9(b) reveals that at station 80, the reduction in flow was accompanied by a reduction in occupancy

[8]Re-scaled T-curves have also been used to identify time periods of nearly stationary freeway traffic (Cassidy, 1998). In addition, Lin and Daganzo (1997) discuss how $T(x,t)$ curves can be used in incident detection.

rates. In addition, this latter set of curves does not display any abrupt rises in T accompanied by reductions in N. To the contrary, the curves display very similar features in that their fluctuations appear to be correlated. These features confirm the passage of a forward-moving expansion wave of lower flow and lower occupancy.

In the next section, the above data treatment methods are used to observe traffic features at the study site. The notable findings from this are described.

4 Analysis

Cumulative curves of vehicle arrival number versus time and cumulative occupancy versus time were used to guarantee that the bottleneck was active and to identify time-dependent bottleneck features. This section contains the analysis of the traffic data measured during the spatial and temporal evolution of traffic conditions upstream and downstream of the bottleneck.

A total of eight days were analyzed and Table 1 introduces the notation used to identify them. Section 4.1 briefly describes the preliminary data analysis procedures used to check the overall integrity of the data on all days. In Section 4.2, data collected from the site on Wednesday, March 5, 1997 (henceforth referred to as Day 1) were used as an example whereby bottleneck activation was preceded by freely flowing traffic through the merge section. It is shown that on Day 1, an especially high flow was observed prior to bottleneck activation and that high on-ramp flows were sustained for an extended period. Days a, b, and c were also days when the bottleneck's activation was preceded by freely flowing traffic through the merge section. These latter observations were used to demonstrate the reproducibility of Day 1 findings.

In Section 4.3, data collected from the site on Wednesday March 12, 1997 (henceforth referred to as Day 2) showed that a queue from a bottleneck somewhere further downstream spilled over into the study site prior to the activation of its bottleneck. Reassuringly, on Day 2, no high flows were observed to arrive from upstream preceding this activation. Days d, e, and f were also days when the bottleneck's activation was preceded by the spillover of a queue from further downstream. These observations were used to demonstrate the reproducibility of the Day 2 findings. Finally, Section 4.4 describes the observations that were reproducible across all eight days.

Table 1: Gardiner Expressway days analyzed.

Day	Date	Bottleneck activation	
		Preceded by free flow	Preceded by a queue from downstream
1	March 5, 1997	x	
a	February 20, 1997	x	
b	July 21, 1997	x	
c	February 11, 1997	x	
2	March 12, 1997		x
d	January 17, 1997		x
e	October 8, 1997		x
f	November 26, 1998		x

4.1 Preliminary data analysis

The overall integrity of the data was first examined by inspecting time-series plots of count and occupancy and re-scaled N- and T-curves constructed for each travel lane at each detector station.[9] This visual scan quickly revealed any systematic loop detector or controller errors since negative values would typically be listed for count, occupancy and velocity whenever the detectors or controllers were not functioning. In these cases, the data were discarded.

As another assessment of the data reliability, vehicle counts manually extracted from video were compared with the counts obtained from a nearby loop detector. From this comparison, which is shown in Appendix D, it was concluded that the loop detector data reasonably matched the sample of vehicle counts obtained from the video.

[9]Daganzo (1997) and Cassidy and Coifman (1996) provide details of how loop detector data are measured.

4.2 Day 1: Bottleneck activation preceded by freely-flowing traffic

This section contains the analysis of data collected on the afternoon of Day 1. Section 4.2.1 demonstrates how the bottleneck location and the duration of its active period were determined and Section 4.2.2 presents the notable lane-changing features and high median lane flows observed prior to this activation. Section 4.2.3 describes the observations of bottleneck input flows at detector station 60 just prior to queue formation and Section 4.2.4 contains the analysis of queue discharge features, as observed across all lanes and in individual lanes; an analysis of on-ramp flows is also included in this section. Finally, Section 4.2.5 presents features that were found to be reproducible on the three days similar to Day 1 (Days a, b and c).

4.2.1 Day 1: Diagnosis of the active bottleneck

This section demonstrates how the location of the active bottleneck was identified from data collected on Day 1, and how the period during which it remained active was ascertained. Toward this end, Figure 10 presents transformed N-curves constructed from counts measured across all lanes at stations 40, 50, 60, 70 and 80 and collected during a 30-minute period that spanned the onset of queueing. The curve for station 40 includes the counts at the Spadina Avenue on-ramp so that each curve was constructed from the same collection of vehicles. Note that a curve for detector 30 was not included in Figure 10 since the vehicles measured at this location were not identical to those measured at the detector locations further downstream.

These transformed curves reveal that the bottleneck was activated between stations 60 and 70 (see Figure 3) for the reasons now explained. The five curves in Figure 10 are initially (nearly) superimposed, indicating that traffic was flowing freely between all stations from about 15:29:03 until 15:42:03. The N at stations 70

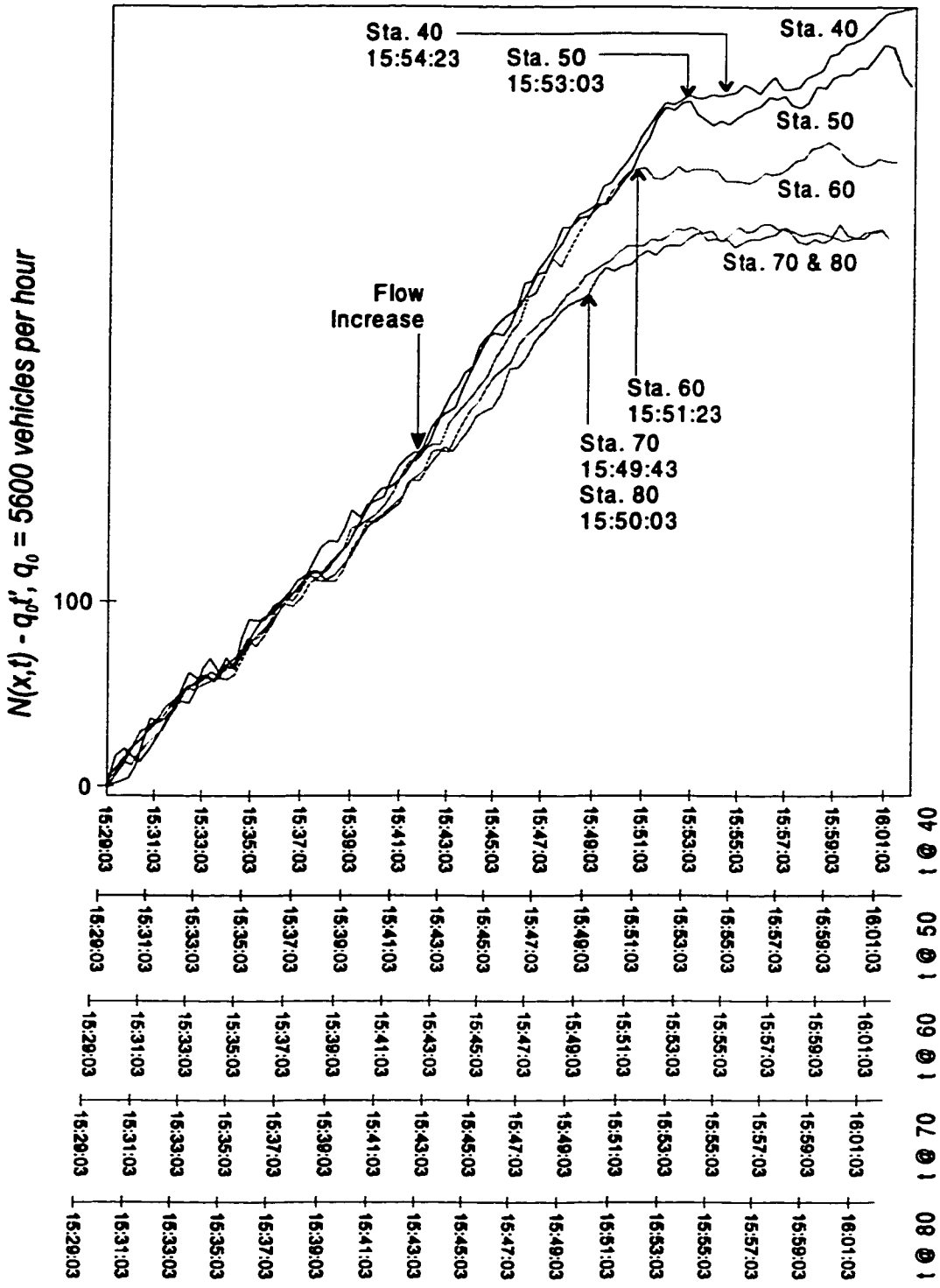

Figure 10: Day 1 Transformed N-curves, detectors 40 through 80.

and 80 remain superimposed for the entire 30-minute period shown in the figure, indicating that traffic continued flowing freely between those two downstream stations. Conversely, upstream curves 40, 50 and 60 first begin to diverge from downstream curves 70 and 80 as early as 15:42. In fact, a straightedge may be used to verify that flows at upstream stations 40, 50 and 60 (slopes of the N-curves) did begin to increase at about this time. This divergence reveals the presence of a queue between stations 60 and 70. Of note, flow reductions at downstream stations 70 and 80 are visible at approximately 15:49:43 and 15:50:03, respectively. Several minutes later, at 15:51:23, curves 60 and 50 diverge, and this is accompanied by a pronounced flow reduction at station 60. These attributes mark the arrival of a backward-moving queue at station 60.[10] The upstream curves 50 and 40 then diverge a few minutes later, at 15:53:03. This divergence along with the flow reduction at station 50 indicates the time by which the backward-moving queue eventually propagated to this station. From these transformed N-curves, excess vehicle accumulations are visible upstream of station 70 and freely-flowing traffic is visible immediately downstream; this verifies that the bottleneck was activated somewhere between stations 60 and 70. Examination of the N-curves also reveals the presence of a flow drop; a higher flow was followed by a lower flow once vehicles were discharging from the queue.

The evolution described above is corroborated by Figure 11, which presents rescaled T-curves constructed from data collected in all travel lanes at the same five detector stations; separate time axes are shown for each station. Referring to Figure 11(a), the T-curves at stations 70 and 80 exhibit concave shapes. The periods marked by relatively low occupancy rates coincide closely with the flow reductions (at stations 70 and 80) that accompanied the onset of the upstream queue. The starting times for these downstream flow reductions (previously identified in Figure 10) are

[10] In previous work, temporary flow reductions such as this were observed when the rear of a queue passes a measurement location (Cassidy and Windover, 1995).

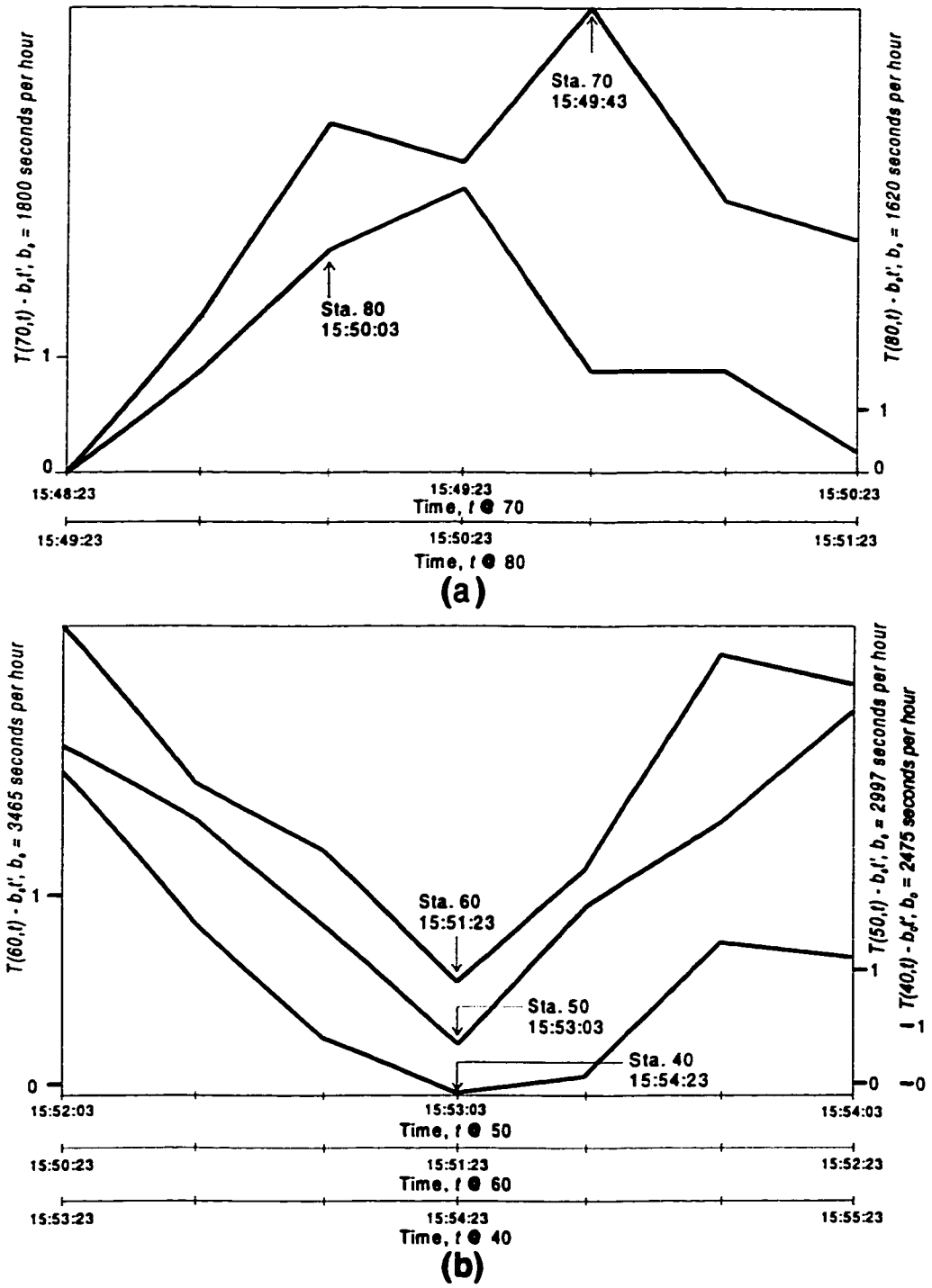

Figure 11: Day 1 Transformed T-curves, detectors 40 through 80.

also marked in Figure 11(a). The sequence of the changes shown in Figure 10, and Figure 11(a), describes the passage of a forward-moving expansion wave of lower flow and lower occupancy. One would expect such a wave to emanate from a sudden flow restriction upstream.

In Figure 11(b), the T-curves at stations 40, 50 and 60 display convex shapes. At each of these upstream stations, the times marked by an increase in the occupancy rates coincide closely with the queue's arrival times as identified from Figure 10; these latter times are shown for the appropriate T-curve in Figure 11(b). These reductions in flow accompanied by increased occupancy rates verify the arrival times of the backward-moving queue.

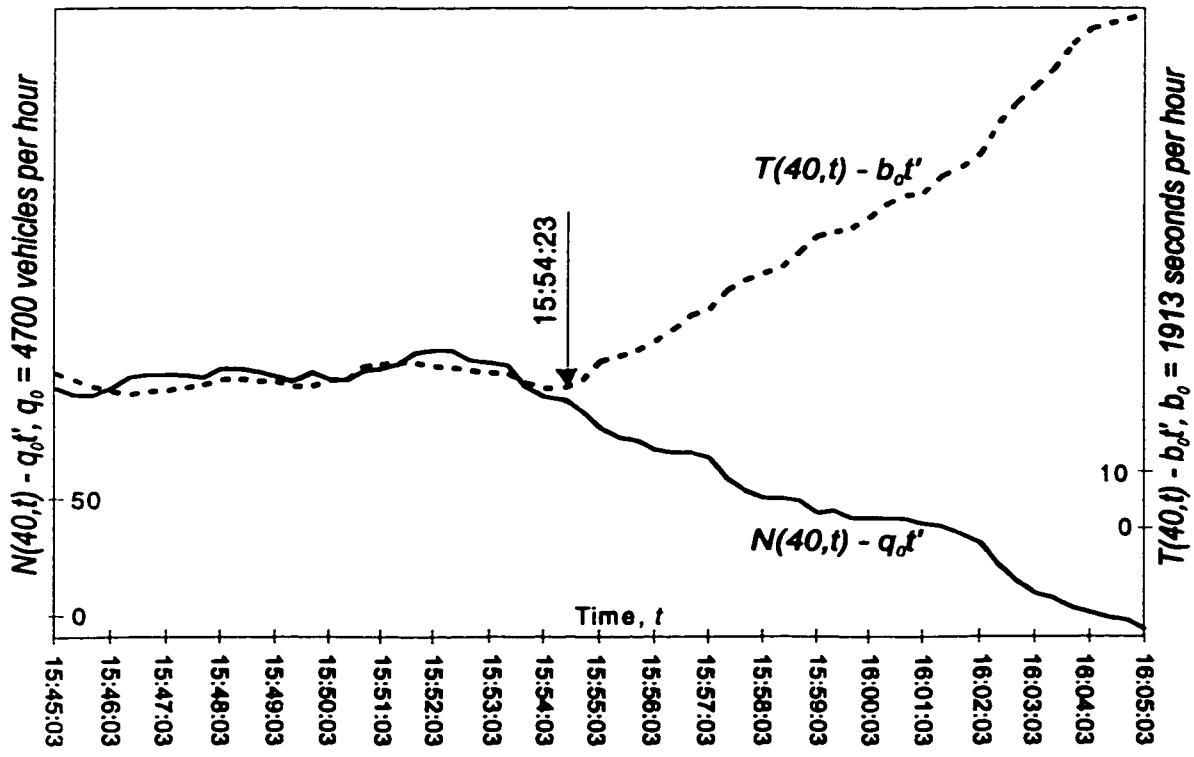

Figure 12: Day 1 Re-scaled N- and T-curves, detector 40

Toward determining the approximate time by which the freeway queue had propagated past the Spadina Avenue on-ramp, Figure 12 presents a re-scaled N-curve (solid line), and a re-scaled T-curve (dashed line), for all freeway lanes at station 40. The two curves show that a sharp reduction in flow was followed closely by an increase in the occupancy rate at about 15:54:23; this marks the arrival of the queue at station 40. That this arrival occurred at approximately 15:54 will be an important part of later discussion (in Section 4.2.4) regarding the observed time-dependencies in the on-ramp flows from Spadina Avenue.

The transformed N and T have shown that a forward-moving expansion wave, along with a backward-moving queue, emanated from between stations 60 and 70 sometime around 15:50.[11] That the bottleneck thus arose between these stations is the only reasonable interpretation.

Further analysis of N- and T-curves showed that this bottleneck remained active until a spillover from some other downstream bottleneck arrived at 18:12:03. To demonstrate how this was determined, Figure 13 displays transformed N-curves for station 60 and station 80 that span a period of more than four hours. The queue's enduring presence between stations 60 and 80 is visible in Figure 13 by virtue of the continued vertical displacement between the two curves.

To trace the spillover, one can see from Figure 13 that the slope of the curve at station 80 drops dramatically at 18:12:03 (as shown by the vertical arrow) and that a similar slope reduction is displayed by the curve at station 60 shortly thereafter (at 18:14:23).[12] Even after these flow reductions, the queue between stations 60 and 80 persisted until after 19:30. Figure 14 shows that the sustained drop in flow

[11]The speeds at which these waves were observed to travel are recorded in Table 11 in Appendix E.

[12]The short-lived drop in flow past station 60 at 18:02:43, and the flow increase shortly thereafter, emanated from upstream (as revealed by N- and T- curves shown next in Figures 15 and 16).

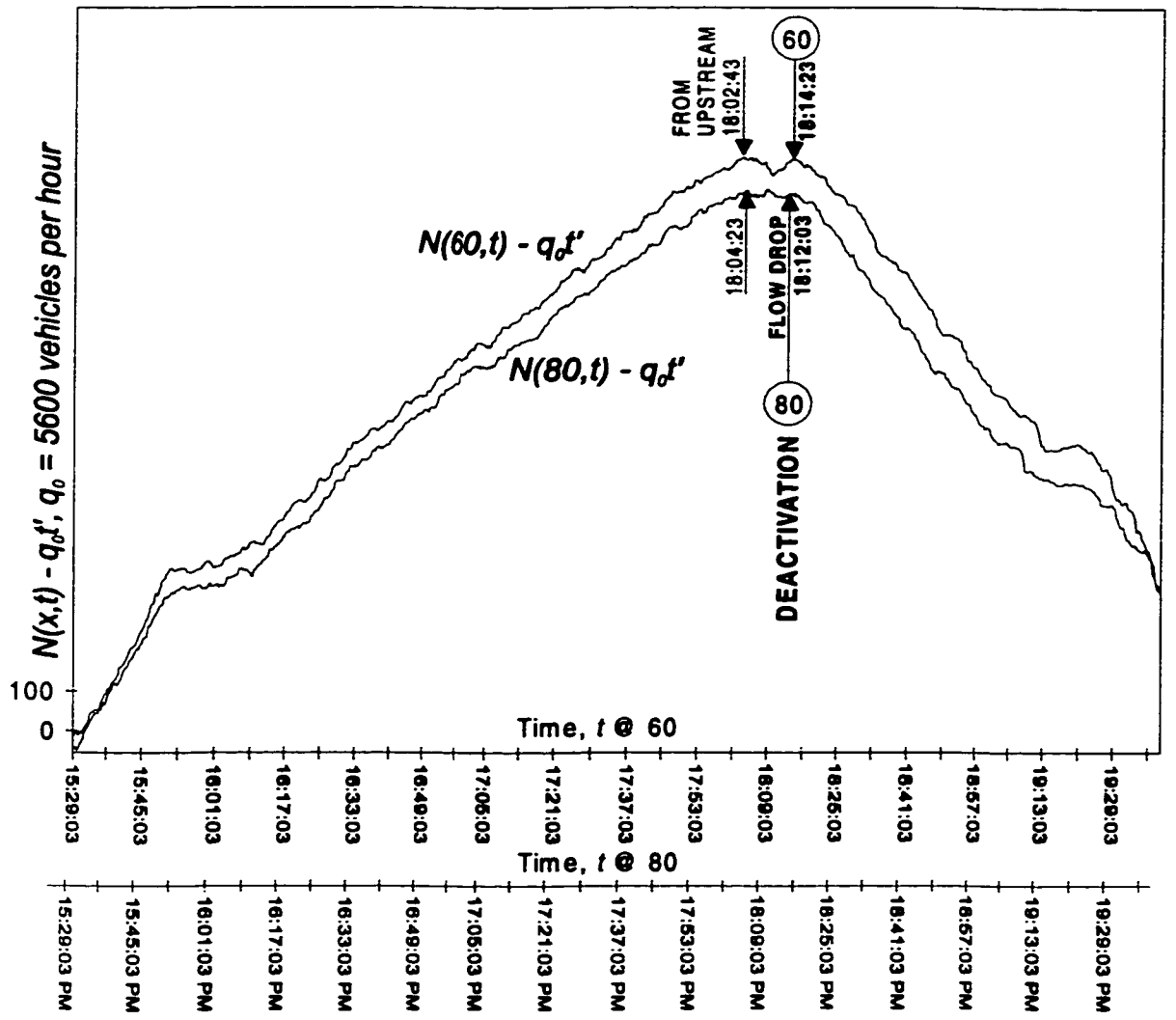

Figure 13: Day 1 Transformed N-curves, detectors 60 and 80

measured at station 80 at 18:12:03 was accompanied by an increase in the occupancy at the same time. This divergence in the re-scaled curves of N and T for station 80 indicates that a downstream queue arrived at station 80 at about 18:12:03 (note that this is shortly after the re-opening of the Jameson Avenue on-ramp) and thereby deactivated the bottleneck at that time by restricting its flow. A re-scaled T-curve at station 60 (not shown) confirmed that the queue arrived at that station at 18:14:23.

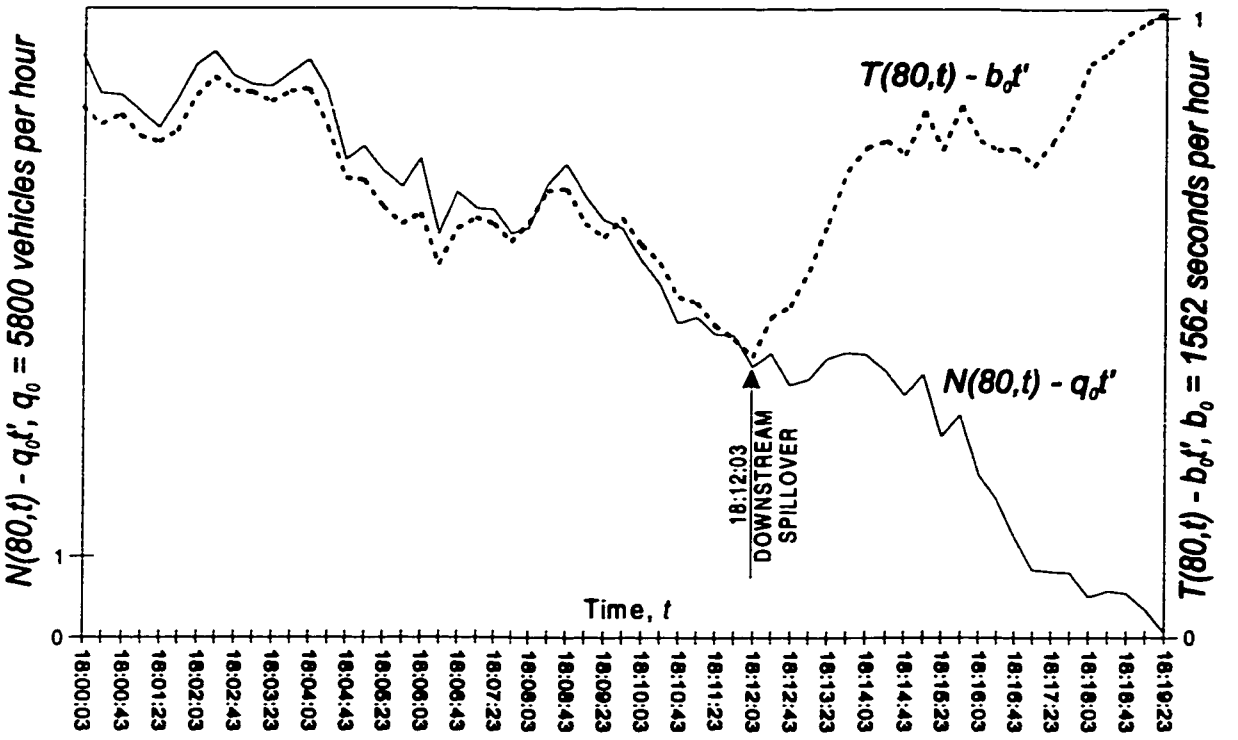

Figure 14: Day 1 Re-scaled N- and T-curves, detector 80.

As labeled in Figure 13, an earlier flow reduction was measured first at station 60 (at 18:02:43) and then at station 80 (at 18:04:23); both reductions occurred several minutes before the bottleneck was eventually deactivated at 18:12:03. Visual inspection of re-scaled N- and T-curves measured in individual lanes revealed that these earlier flow drops were caused by an incident (perhaps a vehicle stall or a small collision) that occurred in the shoulder lane near detector 50. Notably, detector

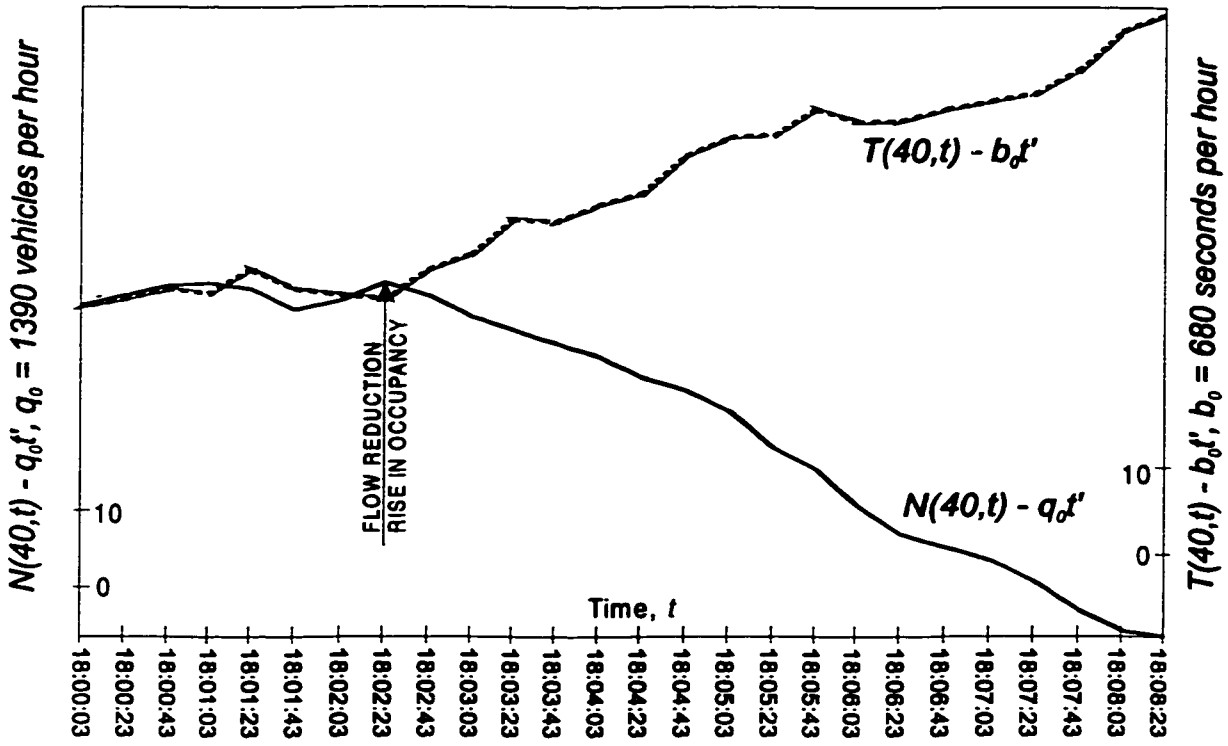

Figure 15: Day 1 Re-scaled N- and T-curves, detector 40, shoulder lane.

50 measured near-zero counts and occupancies in the shoulder lane from 18:02:23 to 18:08:23. Figure 15 shows that during this period, the shoulder lane traffic at upstream station 40 exhibited a sharp reduction in flow coupled with a rise in occupancy, features which marked the passage of a backward-moving queue. Conversely, Figure 16 shows that, within this same period, shoulder lane traffic at downstream station 60 exhibited reductions in both the flow and the occupancy. These latter features would be expected to occur downstream of a sudden restriction (for example, as traffic passed the obstruction and moved into the shoulder lane). In later sections of this dissertation, the discharge flows that accompanied this incident near detector 50 were excluded from measurements of flows through the active bottleneck since the occurance of this incident obviously created a *new* bottleneck, one of lower capacity.

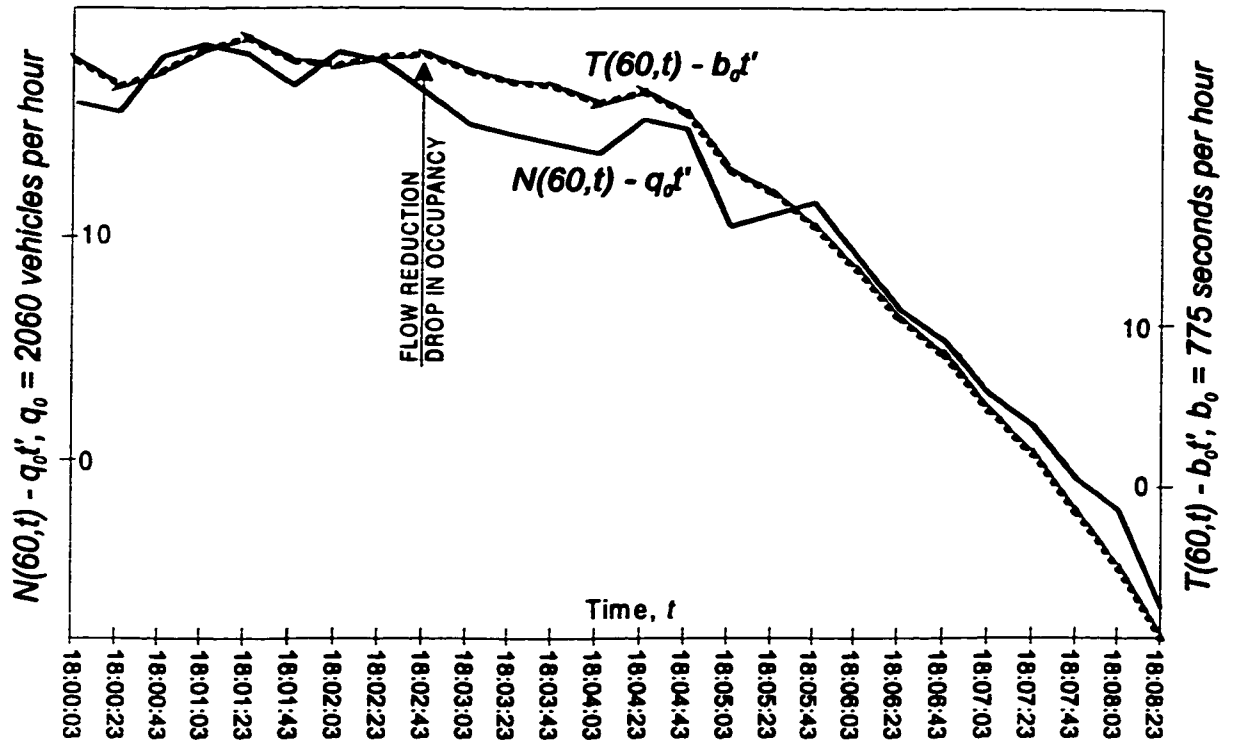

Figure 16: Day 1 Re-scaled N- and T-curves, detector 60, shoulder lane.

4.2.2 Day 1: Lane changing and high median lane flow

The bottleneck's location and the period during which it remained active have been determined using re-scaled N- and T-curves. In the attempt to understand some of the mechanisms that precipitated queue formation on Day 1, traffic patterns were studied in the individual lanes both upstream and downstream of the bottleneck, before and after queueing. Vehicle lane-changing prevented the use of queueing diagrams from individual lanes.[13] Nonetheless, single lane N-curves from neighboring stations were compared to examine how changing flows propagated through a single traffic stream and to examine the net lane changing that occurred between two locations.

[13]If a set of N-curves do not describe node conservation, their vertical displacements are not the excess accumulations (Newell, 1982; Newell, 1993; Cassidy and Windover, 1995).

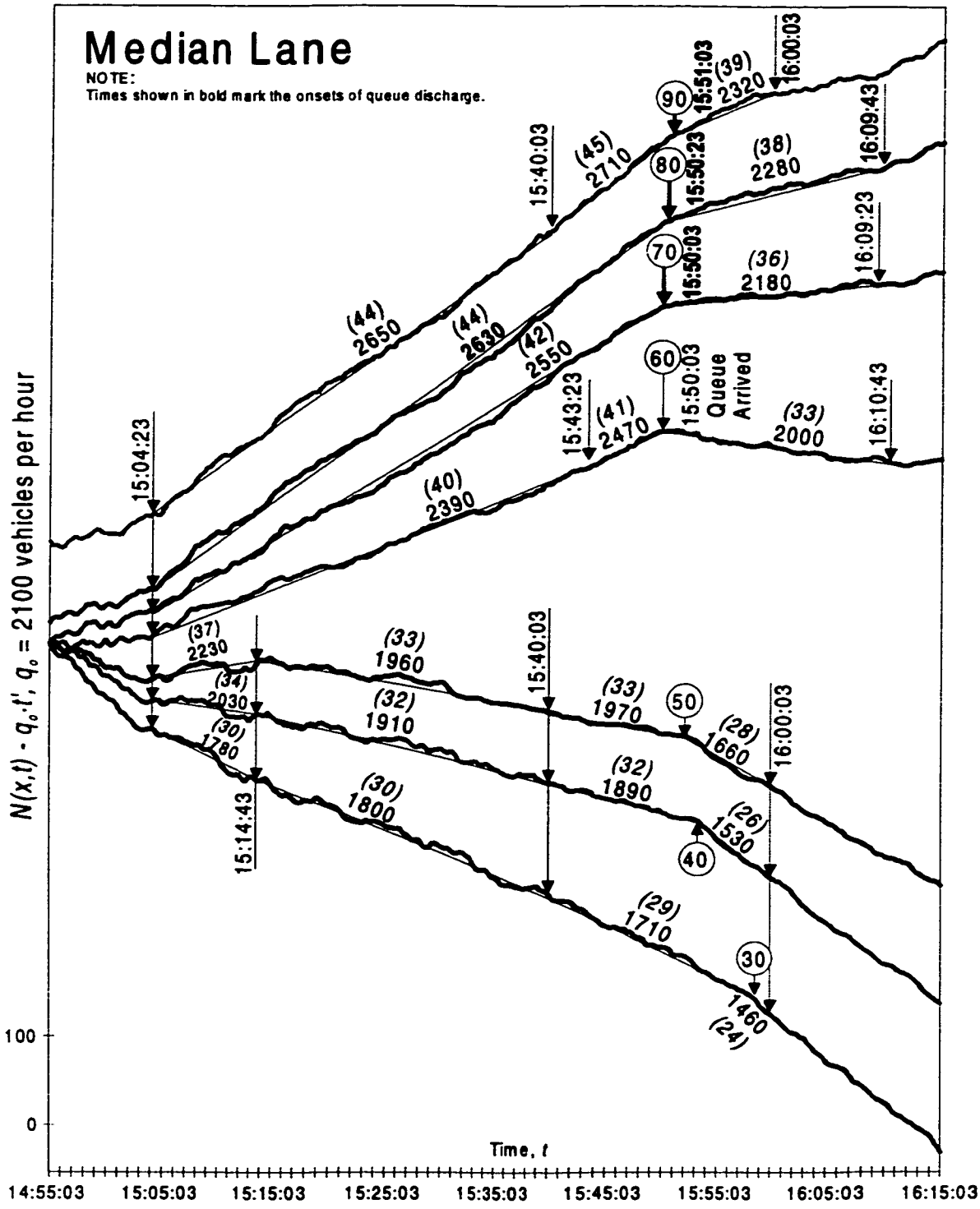

Figure 17: Day 1 Re-scaled N-curves, median lane, stations 30–90, Gardiner Expressway.

Figure 17 shows a series of N-curves (each re-scaled using a consistent value of q_o) constructed from counts in the median lane (only) at stations 30 through 90; they are presented in sequence as vehicles proceed down the freeway. Note that the curves for stations 80 and 90 have been vertically displaced by small arbitrary distances so that their details are visible in the figure. Each curve in Figure 17 was initiated with the same value of N and the average flows are labeled in vph. The labels in parentheses (accompanying the average flows in vph) are the corresponding average counts per minute. The small vertical arrows accompanying the encircled station labels mark the beginnings of discharge flow at downstream stations 70, 80 and 90, and the arrivals of the queue at upstream stations 30, 40, 50 and 60. During the period preceding the bottleneck's activation, it is clear from the increasing slopes of these curves that large numbers of vehicles moved into the median lane as they approached and passed through the bottleneck. Figure 17 shows that the median lane flow measured at station 80 was very high; a rate of 2,630 vph persisted for more than 40 minutes (from 15:04:23 until 15:50:23) before the queue formed. The figure also shows that the median lane flow even exceeded 2,700 vph at station 90, indicating that this lane-changing behavior continued more than 2.7 kilometers (1.7 miles) downstream of the Spadina Avenue on-ramp.

Knowing that the bottleneck resided between stations 60 and 70, particular focus was placed on measuring the net lane-changing between stations 60 and 80, which span the bottleneck. Figure 17 shows that some vehicles entered the median lane through this section via lane changing since the flow measured at station 80 beginning at 15:04:23 was higher than the flow measured at station 60 during the same period. This is also visible directly from the curves in the figure since the slope of the curve for station 80 is higher than the slope of the curve for station 60.

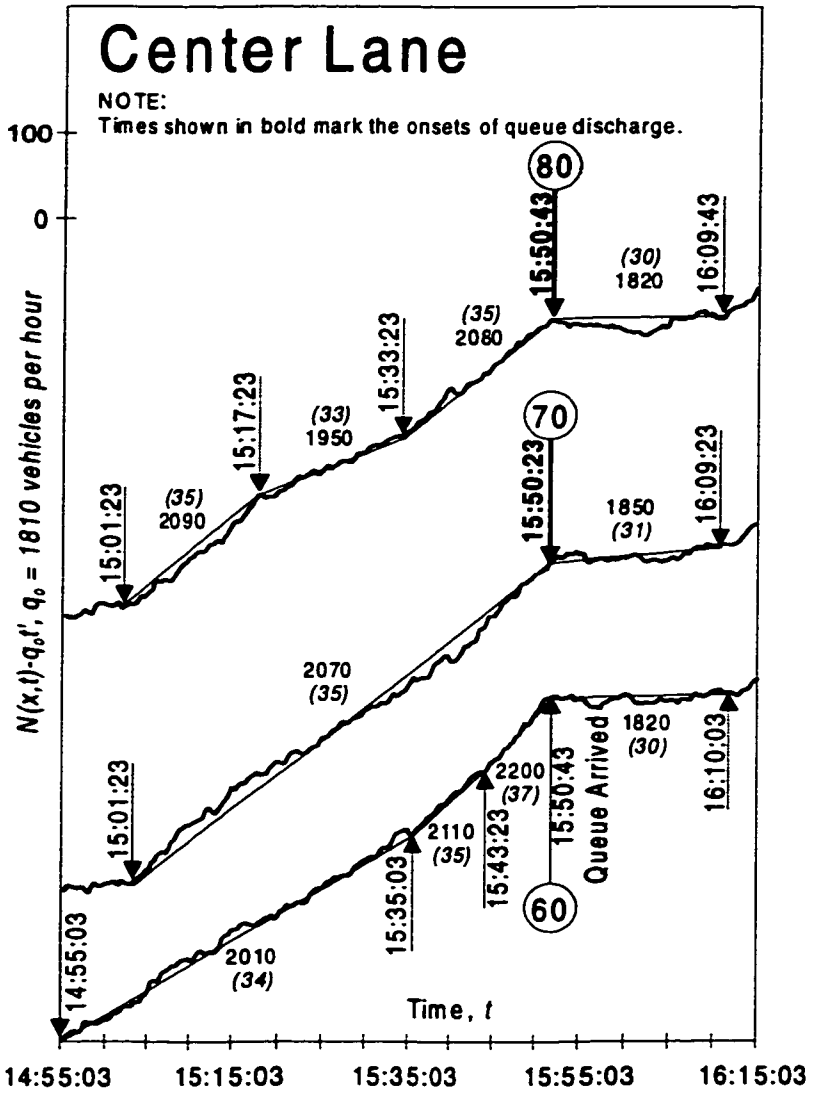

Figure 18: Day 1 Re-scaled N-curves, center lane, stations 60–80, Gardiner Expressway.

Single lane re-scaled N-curves constructed from counts measured in the center lane at stations 60, 70 and 80 are shown in Figure 18 (and are annotated in the same way as Figure 17). The curves for stations 70 and 80 have been shifted vertically by an arbitrary amount. The annotations on the N-curves in Figure 18 reveal that between detectors 60 and 80 the net flows in the center lane remained nearly unchanged as traffic moved through the bottleneck. Thus, between these detectors, the number of vehicles that moved into the center lane (e.g., from the shoulder lane) nearly equaled the number that moved out (e.g., into the median lane.

Re-scaled N-curves constructed from counts in the shoulder lane (Figure 19, annotated as the two previous figures) show that curve 60 remains well above curve 80 for all time t, indicating that vehicles continued to exit the shoulder lane at locations well downstream of the Spadina Avenue on-ramp.

As a final note, it appears that flow on the Jameson Avenue off-ramp had little impact on lane-changing through the bottleneck region. Figure 20 shows a re-scaled N-curve for the Jameson Avenue off-ramp, the labels of which indicate that the off-ramp flow was very low throughout the entire peak period (recall that the Jameson Avenue on-ramp was closed from 15:00 to 18:00).

4.2.3 Day 1: Bottleneck input flows

After the bottleneck's location and active period were determined, the period exhibiting higher flows prior to queue formation was examined in greater detail. Toward this end, Figure 21 shows re-scaled N-curves (each was re-scaled using the same value of q_o) constructed from data measured across all lanes at stations 60, 70 and 80 for an 80-minute period spanning the onset of queueing. The curves for stations 60 and 70 were shifted arbitrarily upward so that the curves' details are visible. The figure indicates that when flow was measured across all lanes at station 80, a high rate of

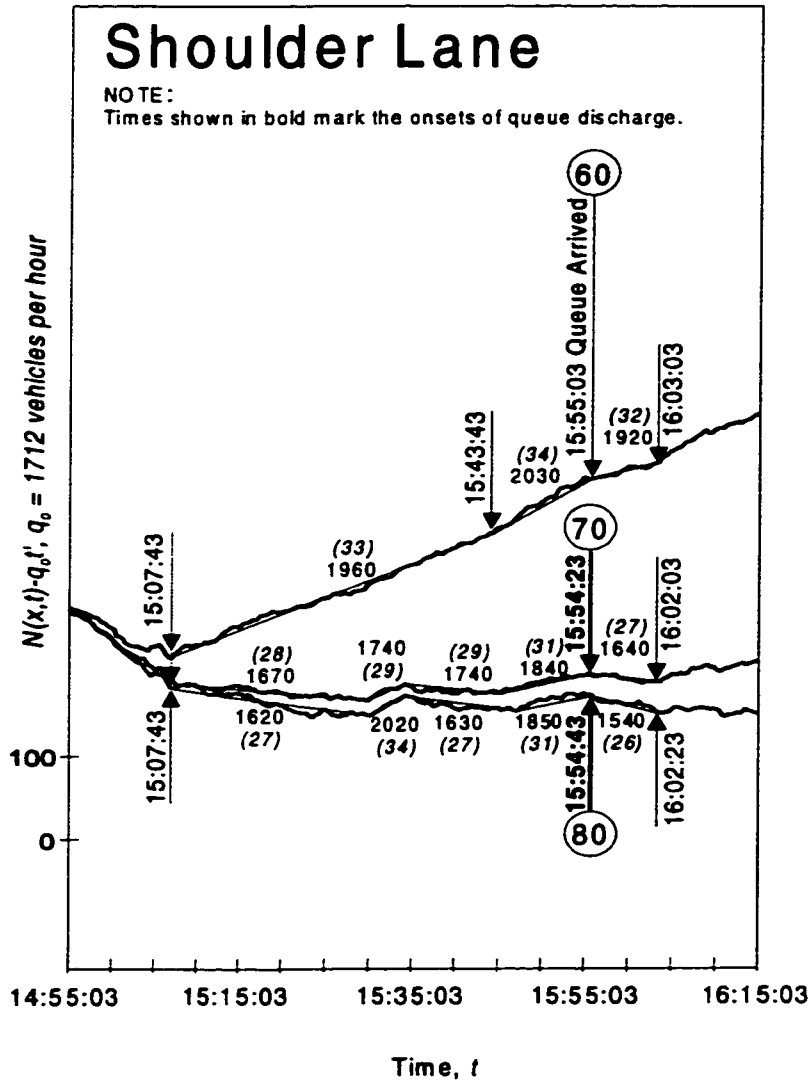

Figure 19: Day 1 Re-scaled N-curves, shoulder lane, stations 60–80, Gardiner Expressway.

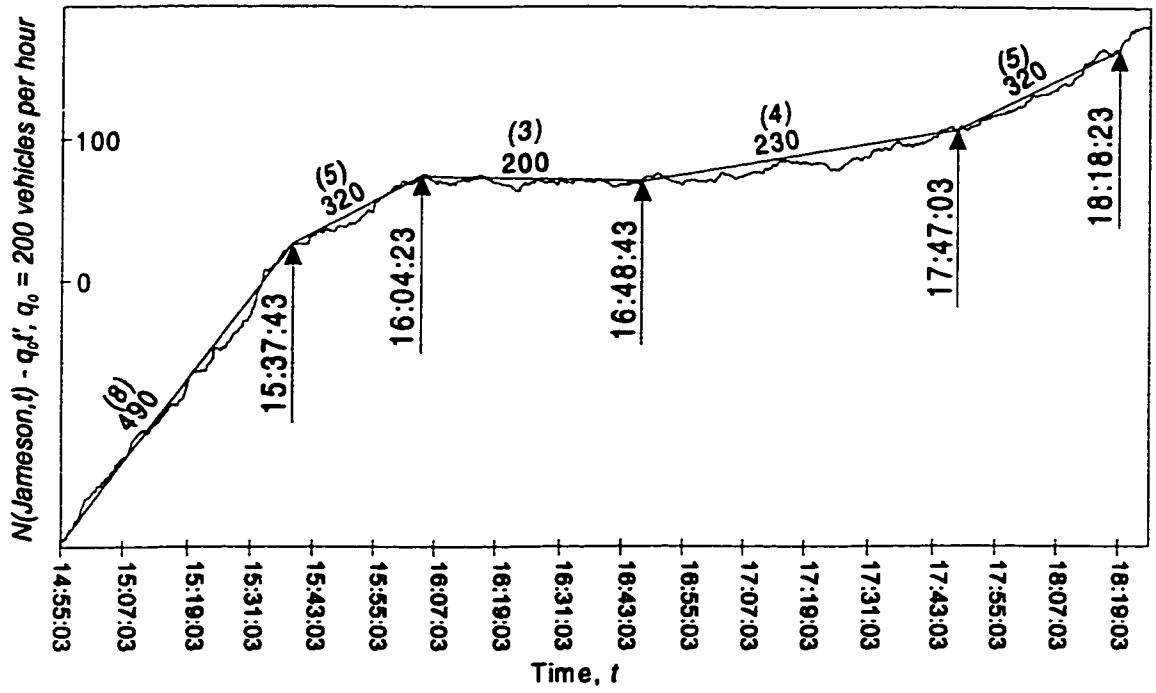

Figure 20: Day 1 Re-scaled N-curve, Jameson Avenue off-ramp.

6,490 vph prevailed for 21 minutes (from 15:29:03 until 15:50:03) prior to the arrival of the forward expansion wave.

As shown in Figure 21, a high flow was measured across all lanes prior to queue formation on Day 1. Toward a better understanding of how this high bottleneck input flow influenced the formation of the queue, traffic features were studied in the individual lanes at station 60, the measurement location immediately upstream of the bottleneck.

Figure 22 contains the single lane re-scaled N-curves constructed from data measured at station 60 for a 40-minute period spanning the onset of queueing. The same value of q_o was used for the median lane curve (upper curve in the figure), center lane curve and shoulder lane curve. Note that the small vertical arrows accompanying the lane labels mark the arrival of the queue in each lane and that these were determined from Figures 17, 18 and 19. The dotted line superimposed on the upper curve shows

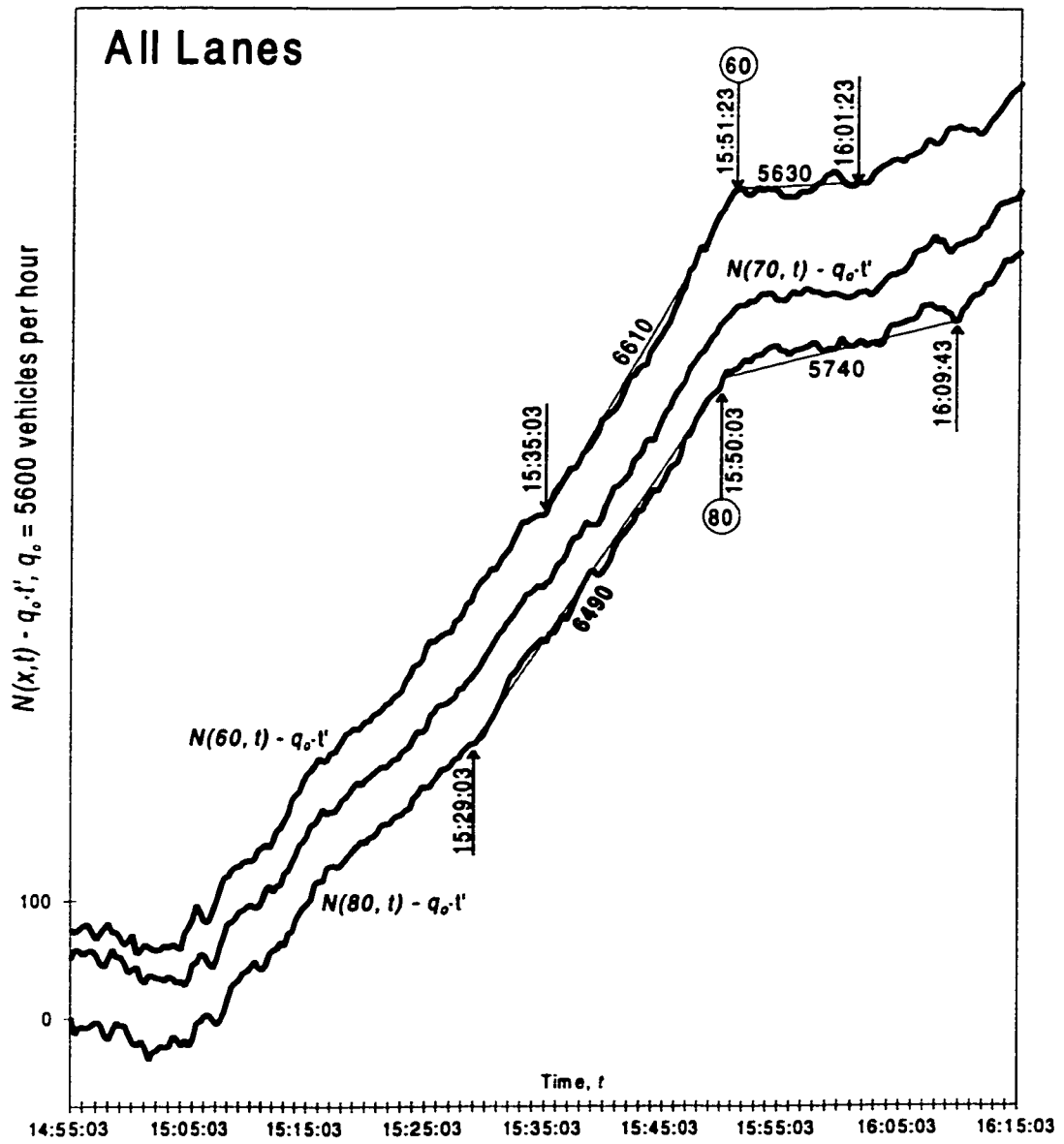

Figure 21: Day 1 Re-scaled N-curves, all lanes, stations 60–80, Gardiner Expressway.

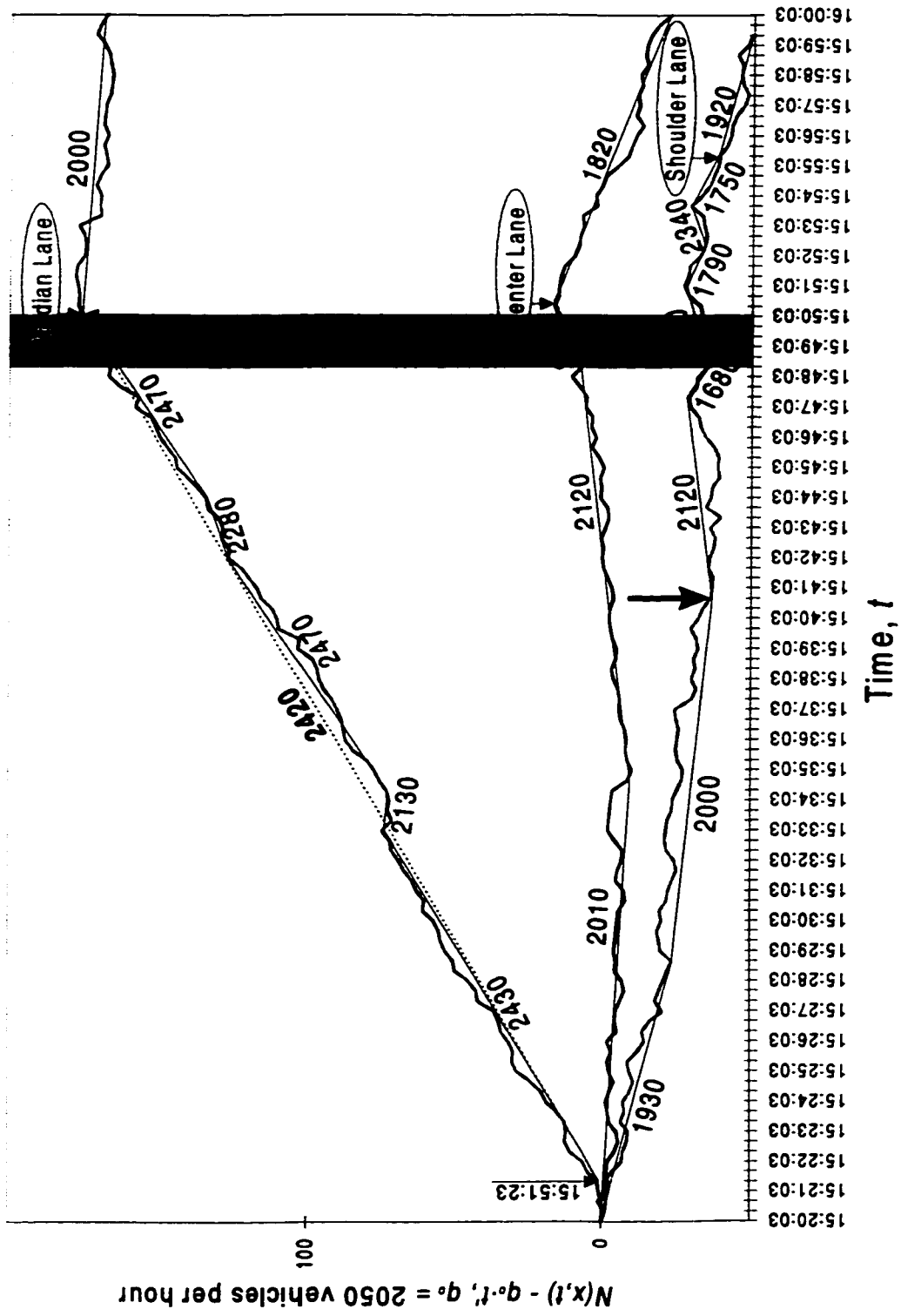

Figure 22: Day 1 Re-scaled N-curves, station 60 of the Gardiner Expressway.

that the average flow in the median lane between 15:21:23 and 15:50:03 was 2,420 vph. This high flow apparently constrained lane changing into the median lane, such that the flows in the center and shoulder lanes became progressively higher. The general upward bending of the center and shoulder lane N-curves indicates that the flows in these two lanes were increasing. Then, beginning at 15:40:43 (marked by the large vertical arrow above the shoulder lane curve), about ten minutes prior to the arrival of the queue, flow in the center and shoulder lanes became approximately equalized. Finally, for the two-minute period just before the queue formed (highlighted by the shaded vertical strip), a total flow of nearly 7,000 vph was achieved. The achievement of this high flow immediately preceding queue formation marked the first and only time that this high flow measurement was sustained for as long as a two-minute period. Total flows nearing 7,000 vph were sometimes sustained for periods of 80 seconds or less prior to the interval delineated by the shaded vertical strip in Figure 22.[14] Knowledge of this apparently critical flow signaling the impending queue formation would be important when considering any on-line techniques for attempting to sustain higher flows.

4.2.4 Day 1: Queue discharge features

To study the discharge flows through the bottleneck, Figure 23 shows re-scaled cumulative count and re-scaled cumulative occupancy curves constructed from data measured at station 80 (given the node conservation law, detector station 70 would have been an equally suitable location for measuring bottleneck flows). These curves span $3\frac{1}{2}$ hours, and include the active period which began with the arrival of the forward expansion wave at 15:50:03 and ended when the queue arrived from down-

[14] Earlier in the peak period (between 15:05 and 15:13) total flow measurements exceeding 7,000 vph were observed for periods of 40, 60 and 80 seconds, but never exceeding 80 seconds.

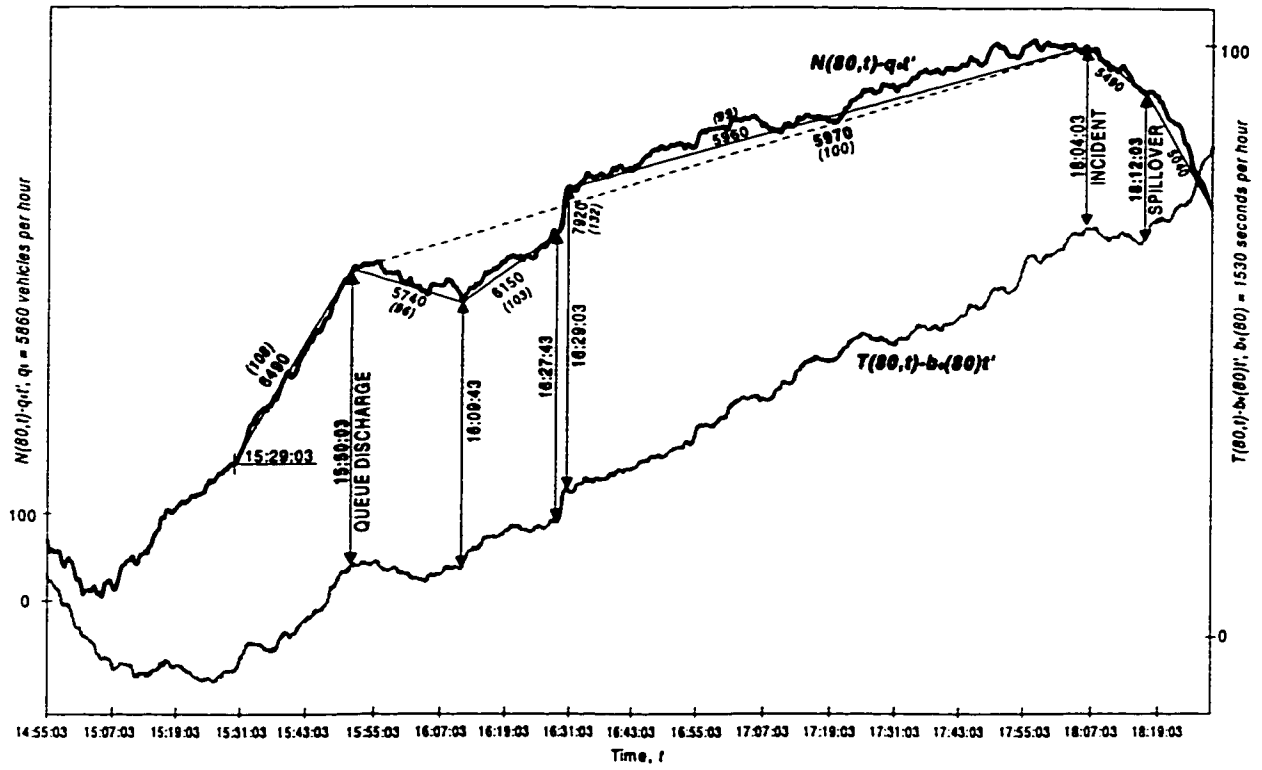

Figure 23: Day 1 Re-scaled N- and T-curves, station 80 of the Gardiner Expressway.

stream at 18:12:03. As described earlier, queue discharge flows will only be measured until 18:04:03 when the incident occurred.

During the active period, the Figure 23 curves do not display any abrupt rises in the T accompanied by reductions in the N, features that would mark the arrival of a queue from downstream. To the contrary, because both curves display remarkably similar features during the period of queue discharge, it is confirmed that all flow changes came from upstream during this active period. In fact, using curves not shown here, it was confirmed that these flow changes emanated from the downstream end of the queue. The correlations in the fluctuations in the re-scaled N and T in the figure, and their quasi-linear trends (which are delineated by the vertical arrows in Figure 23), reveal that the discharging vehicles exhibited sequences of nearly stationary traffic patterns. The intervals delineated by vertical arrows in

Figure 23 mark periods during which the re-scaled N and T exhibited only small deviations from chords. Thus, each of these intervals was characterized by some nearly constant flow and a nearly uniform vehicle speed; the rationale behind this diagnostic is explained in Cassidy (1998).

The flow pattern in Figure 23 began with the expansion wave which carried a discharge flow of 5,740 vph, the lowest rate observed during the peak. This rate persisted for 20 minutes. Following this initial lower flow, the next 18-minute period was marked by a higher discharge rate of 6,150 vph, followed by a short-lived surge of 7,920 vph; these flows are referred to as *recovery* discharge flows. After the surge, the discharge flow settled into a nearly constant rate of 5,960 vph that prevailed until about 18:04:03 (when the incident occurred upstream).

The variation in the queue discharge flow occurred about a constant rate and this average discharge rate of 5,970 vph is shown by the dashed line in Figure 23. From the vertical scale on the left edge of the figure, one can observe that while the bottleneck was active and free from any nearby incidents, the N never differed by more than 75 vehicles from the dashed line. The observation of this *nearly constant* discharge rate is important since it means that the bottleneck capacity is time-invariant and that the minor changes in flow likely had little influence on the evolution of the back of the queue.

Table 2 confirms that it is reasonable to consider the average discharge rate as nearly constant. First considering the 20-second counts, the table shows that the variance-to-mean ratio of the counts measured across all lanes at detector 80 is 0.39. In addition, 97.0 percent of the count observations fall within plus or minus two standard deviations of the mean (33.2 vehicles/20-second interval, or 5970 vph) of the counts. Furthermore, the vertical difference between the N-curve and the dashed line was measured at each 20-second observation, and the mean and variance

Table 2: Day 1 Gardiner Expressway queue discharge features.

Statistic	20-second count	Vertical difference from trend line
Mean	33.2 veh	6.36 veh
Variance	13.1 veh^2	898.4 veh^2
γ^a	0.39 veh	141.2 veh
Percent within $\pm 2\sigma^b$	97.0%	95.5%
Maximum^c	-	75 veh

[a] Variance-to-mean ratio.
[b] Percent of observations within plus or minus two standard deviations of the mean.
[c] Absolute value.

of this difference were calculated. Table 2 shows that these vertical differences were distributed rather compactly, such that 95.5 percent of the observed differences were within plus or minus two standard deviations of the mean.

The average queue discharge rate measured across all lanes was compared to the maximum flow absent the queue (this measured flow of 6490 vph was reported in Section 4.2.2 and was shown in Figure 21); the average discharge rate was about eight percent lower than the maximum flow measured prior to queue formation. The magnitude of this difference in flows is more severe than those reported in the previous studies of freeway bottlenecks (among those reporting any difference whatsoever).

The flows through the bottleneck were also studied in the individual lanes during the time that the bottleneck remained active (Figures 17, 18 and 19 only showed the high flows in the individual lanes prior to the bottleneck queue formation). Toward this end, Figure 24 presents re-scaled N-curves constructed from counts measured in the median, center and shoulder lanes at station 80, only for the active period. Each curve was constructed using a different value of q_o as labeled on the y-axis. This figure shows that the active period began at nearly the same time in the median

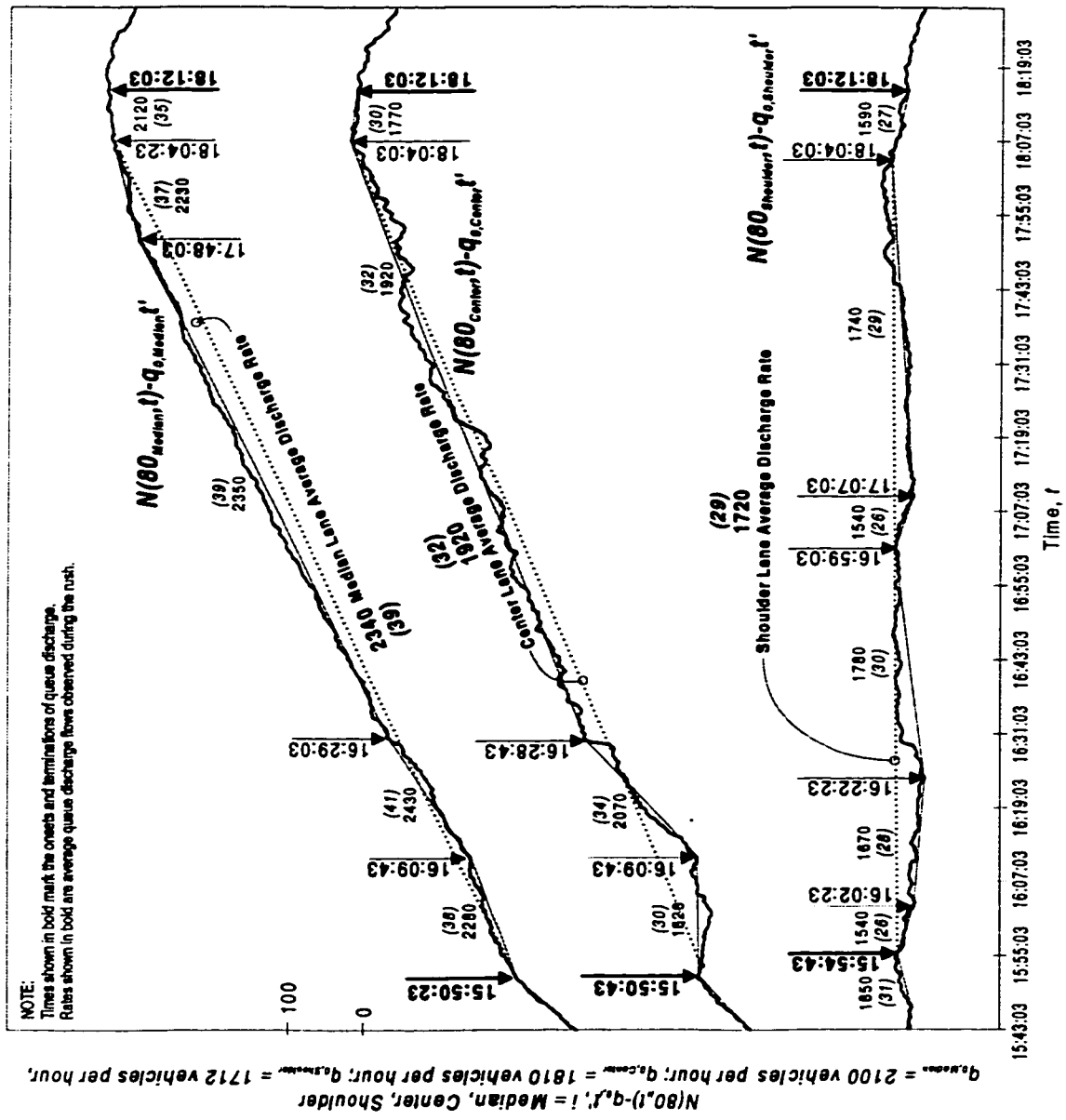

Figure 24: Day 1 Re-scaled N-curves for individual lanes, Station 80.

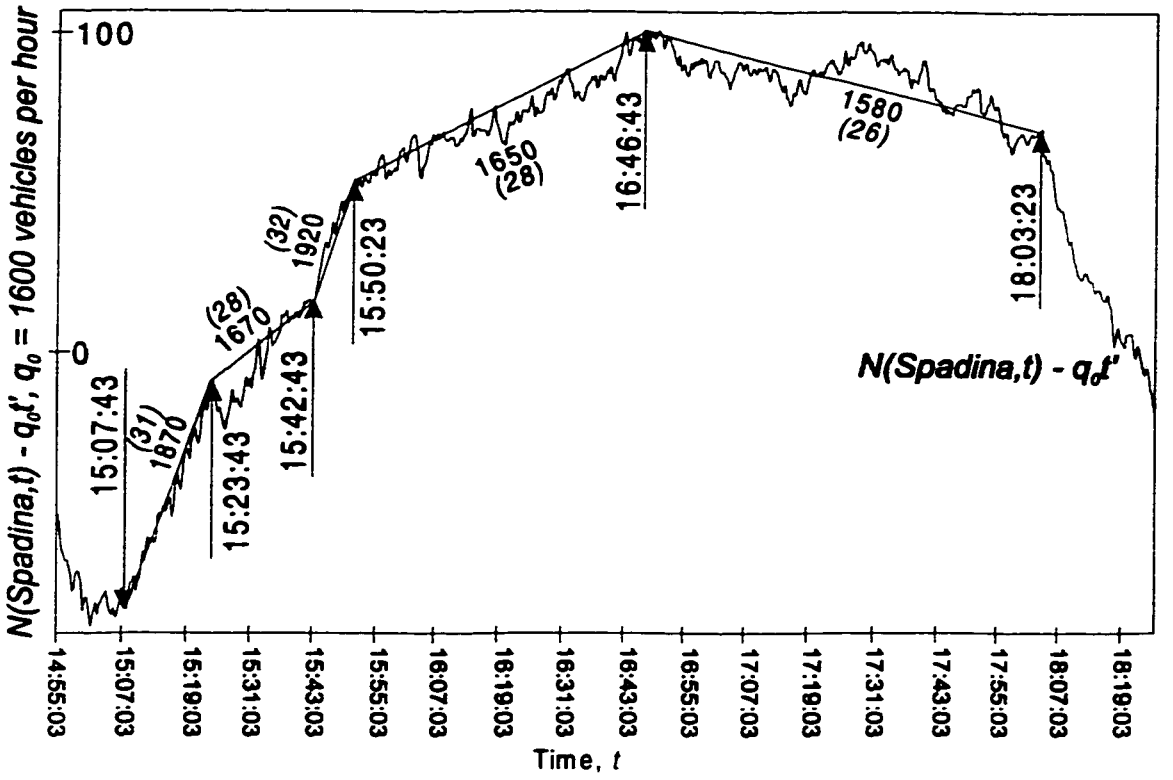

Figure 25: Day 1 Re-scaled N-curve, Spadina Avenue on-ramp.

and center lanes and that the queue appears to have formed in the shoulder lane several minutes after it formed in the adjacent lanes. The times marking the onset and termination of queue discharge flows are noted in boldface type.

The onset of the queue was also accompanied by rather dramatic flow reductions in the individual lanes. Figure 24 shows that the periods immediately following the bottleneck's activation were marked by some of the lowest discharge rates observed during the rush, but that these relatively low flows were short-lived relative to the duration of the entire rush. In the center and median lanes, these flow collapses prevailed for just under 20 minutes before higher discharge flows prevailed. In the shoulder lane, the collapses persisted for less than 10 minutes.

Figure 24 also reveals that the queue discharge rates measured in the individual lanes at station 80 never deviated much from the linear trends marked by the dotted lines. Thus, for each lane, the discharge rates can be described as being nearly constant over the rush. By comparing the average discharge rates with the flows that prevailed prior to the bottleneck's activation, it is clear that queueing was accompanied by long-run flow reductions, especially in the median and center lanes.[15] It is also evident from this figure that the average queue discharge rates varied across lanes, with the highest rate in the median lane and the lowest rate in the shoulder lane.

Finally, Figure 25 shows a re-scaled N-curve constructed solely from the counts measured on the Spadina Avenue on-ramp (adjacent to detector station 40). The sustained surge in the ramp flow evident at 15:42:43 corresponds to the measured increase in upstream flow previously revealed in Figure 10. Likewise, the sharp reduction in this ramp flow at 15:50:23 corresponds closely with the arrival of the queue from the active bottleneck; recall that this queue was shown (in Figure 12) to have arrived to the freeway detectors at station 40 at approximately 15:54. Although this queue apparently suppressed the on-ramp flow, vehicles continued to enter the freeway via this ramp at a high rate; note that an average ramp flow of 1,650 vph persisted for nearly an hour. This flow dropped at about 16:46:43, perhaps due to a reduction in the on-ramp demand (although additional ramp detectors do not exist to confirm this). In any event, these high ramp flows mean that, just downstream of the merge, more than half of the vehicles traveling in the shoulder lane originated from the on-ramp. Thus, the merging process did not exhibit the *zipper effect* (Newman, 1986) whereby freeway and ramp vehicles would have shared the shoulder lane in a strictly alternating fashion.

[15]Table 12 in Appendix E shows the flow drop percentage accompanying the onset of the queue, in each lane at each station.

Table 3: Summary of high flows in the median lane on the Gardiner Expressway.

Day	Flow prior to queue				
	Long run				Short run
	Median lane		All lanes		All lanes
	Rate *vph*	Duration *mm:ss*	Rate *vph*	Duration *mm:ss*	Rate *vph*
1	2630	43:40	6490	21:00	6970
a	2630	05:40	6620	02:40	6870
b	2400	19:40	6120	21:40	7250
c	2590	24:20	6270	16:40	7120

4.2.5 Day 1: Reproducing the observations

This section summarizes traffic features on days similar to Day 1 (when there were no queue spillovers from downstream bottlenecks early in the peak period). On Days a, b and c, certain observed traffic patterns were similar to those found on Day 1.

As shown in the second and third columns of Table 3, very high flows were observed in the median lane each day prior to queueing; these high rates were always sustained for periods of more than 5 minutes, and in three instances, for much longer. The fourth and fifth columns of Table 3 show the high flow measured across all lanes (and its duration) prior to queue formation. Notably, the sixth column of the table shows that high flows prevailed for two minutes immediately prior to queue formation. These values approach 7,000 vph.[16] That the high flow observed for two minutes prior to queue formation appears to be approximately the same each day implies that any kind of on-line control strategy might seek to maintain flow below this rate.

[16]Total flows nearing 7,000 vph were sometimes observed for periods of 80 seconds or less prior to the observation of the rates shown in column six of Table 3. The latter were always sustained for approximately two minutes.

Table 4: Summary of measured traffic features on the Gardiner Expressway on days when a flow drop is observed.

Day	Flow prior to queue (α) Rate *vph*	Disch. rate immediately follow. queue		Recovery discharge rate		Average discharge rate (β)		Percent diff. ($\alpha - \beta$)
		Rate *vph*	Dur. *mm:ss*	Rate *vph*	Duration *mm:ss*	Rate *vph*	Dur. *h:mm*	*%*
1	6490	5740	19:40	6450, 7920	28:00, 1:20	5970	2:22	8%
a	6620	5650	09:20	5940, 6360	22:00, 6:20	5970	2:08	10%
b	6120	5700	02:00	6180	06:40	5870	0:59	4%
c	6270	5640	57:20	6010	12:40	5790	2:42	8%

Columns three and four of Table 4 show that the presence of a particularly low discharge flow accompanying the onset of queueing was also a reproducible feature on each of these four days. That the onset of queueing was accompanied by a low discharge flow followed by a higher recovery rate (or a sequence of two high rates) was also a reproducible feature on each of these days (columns five and six of Table 4). On these four days, the average discharge rate (β) was substantially lower than the sustained flow immediately prior to queue formation (α). On three of the four days, the percent difference between these rates ($\alpha - \beta$) was at least 8 percent. The day with a difference of only 4 percent (Day b) was the only observation taken during the summer. To what extent *seasonal effects* influence the features of bottleneck capacity will be described as part of future research possibilities in Section 5.

4.3 Day 2: Bottleneck activation preceded by a queue

This section contains the analysis of data collected during a portion of the afternoon of Day 2. Section 4.3.1 describes the determination of the bottleneck location and

the duration of its active period. Section 4.3.2 describes the bottleneck's queue discharge features, and finally, Section 4.3.3 presents features that were found to be reproducible on Days d, e and f.

4.3.1 Day 2: Diagnosis of the active bottleneck

On Day 1, a high flow passed through the bottleneck prior to queue discharge at a lower average flow. There was no queueing in the study section prior to the activation of the bottleneck. Early in the peak period on Day 2, however, a queue spilled over into the subject freeway section from some restriction further downstream and propagated through the Spadina Avenue merge area. When this queue dissipated, the bottleneck between stations 60 and 70 became active immediately and vehicles discharged through the bottleneck without ever achieving a higher flow and without producing the particularly low discharge flow associated with the initial period of the bottleneck's activation on Day 1.

To demonstrate this, Figure 26 shows a set of transformed N-curves for stations 40, 50, 60, 70 and 80 on Day 2. These curves were constructed using counts measured across all lanes (as described in Section 4.2.1), for an 80 minute period surrounding the activation of the bottleneck between stations 60 and 70. As shown in the figure, the five curves are initially superimposed, revealing that traffic was flowing freely between all stations from 14:30:03 until 14:40:03. At 14:40:03, the divergence in the curve at station 80 from the one at station 70 marks the arrival (at station 80) of a backward-moving queue from further downstream. Subsequently, curves 60 and 70 diverge (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, by 14:46:03, this queue resulting from some downstream bottleneck had arrived at Station 50.

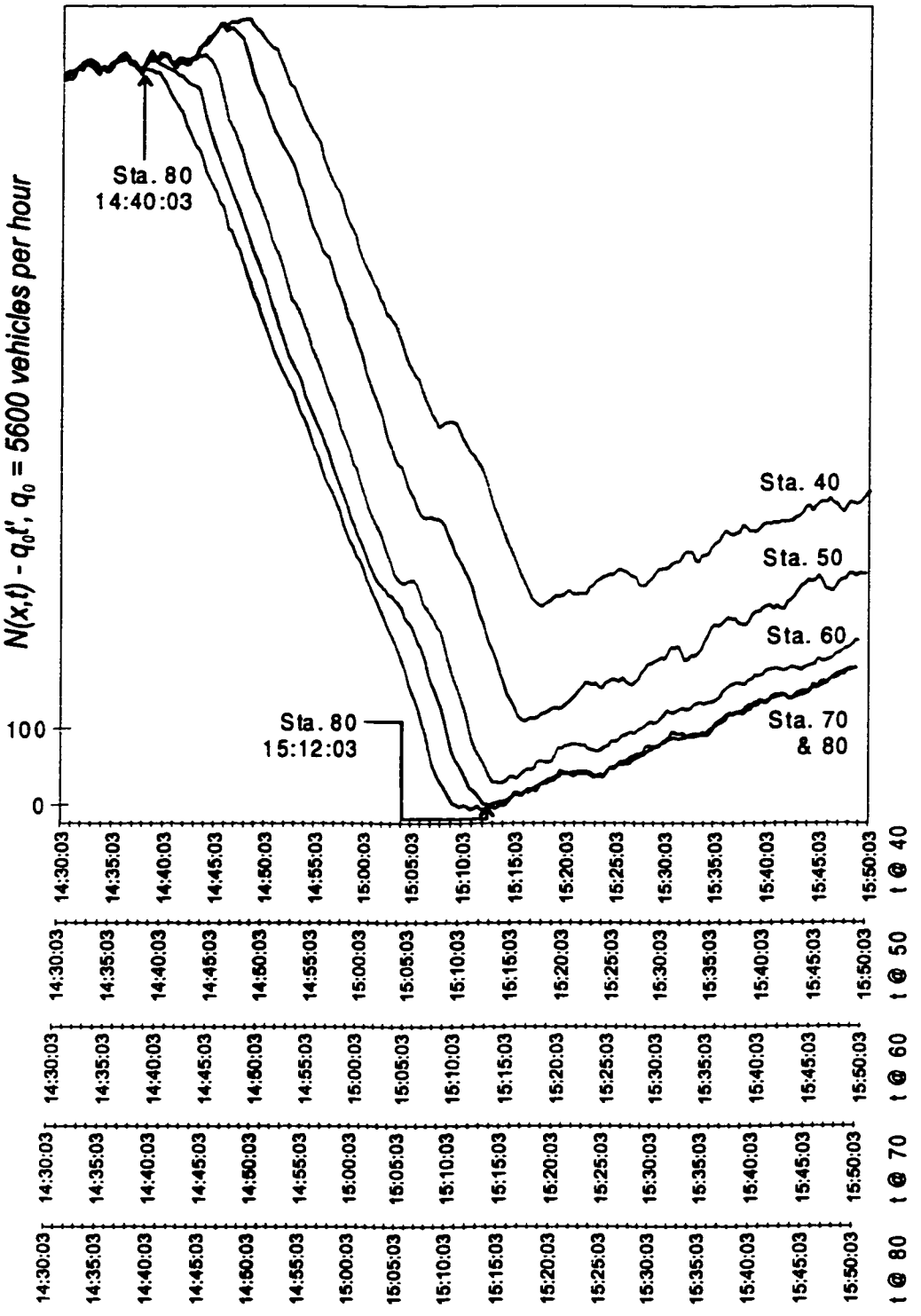

Figure 26: Day 2 Transformed N-curves, detectors 40 through 80.

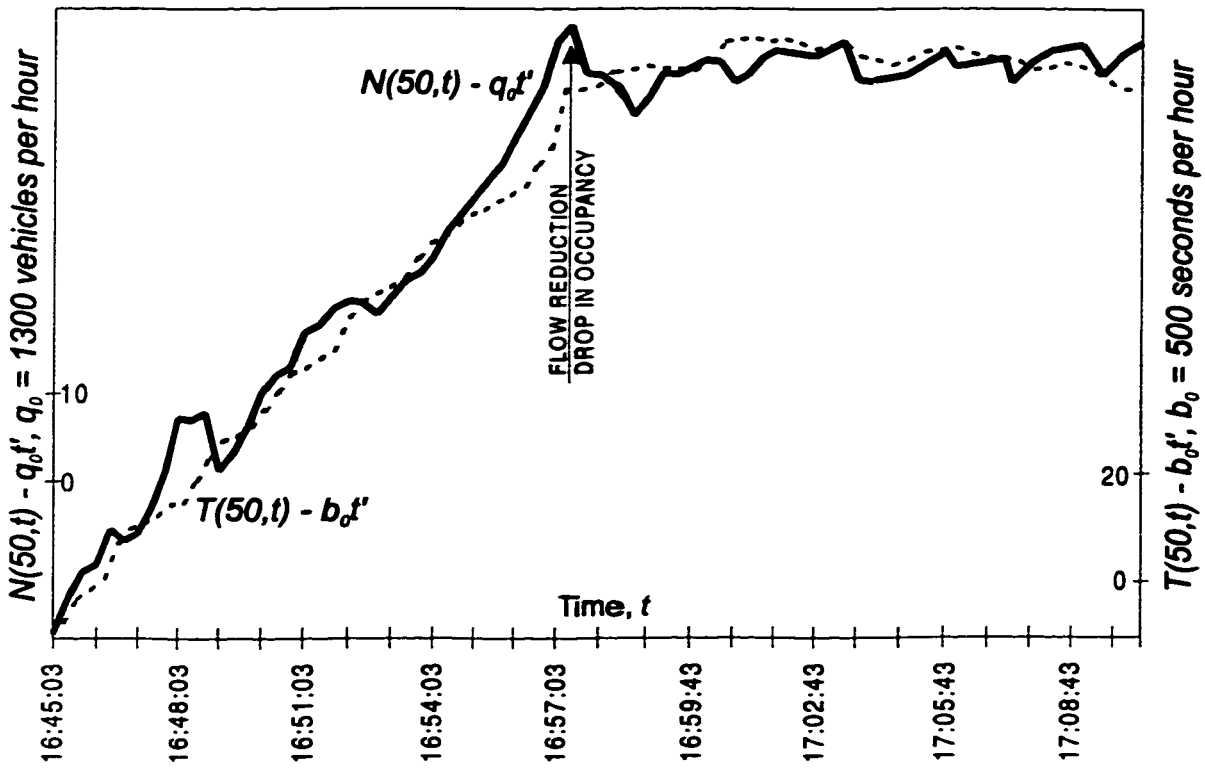

Figure 27: Day 2 Re-scaled N- and T-curves, detector 50, median lane.

The continued vertical separation between all pairs of curves for the 33-minute period from 14:40:03 until 15:12:03 reveals that a queue occupied the entire merge area throughout this period. The average flow measured during this period was only 3,770 vph and this flow was governed by the capacity of the downstream restriction.

At 15:12:03 the curves at stations 70 and 80 again became superimposed, revealing that the downstream queue had dissipated. The continued excess accumulation upstream of station 70 (shown by the vertical separations which remain after 15:12:03) indicates that the bottleneck between stations 60 and 70 was activated immediately thereafter. The transformed N-curves in Figure 26 were necessary for identifying the time at which the bottleneck between stations 60 and 70 became active. Further analysis of the N-curves in Figure 26 (and other N- and T-curves not

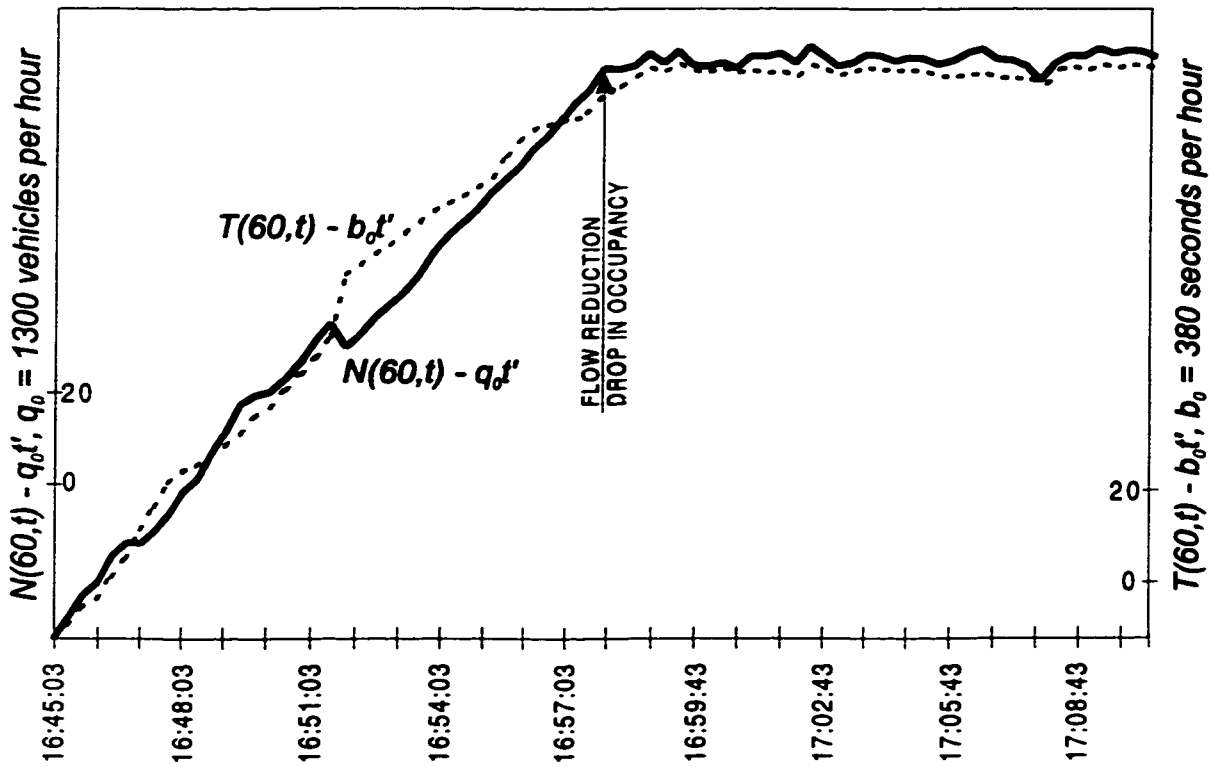

Figure 28: Day 2 Re-scaled N- and T-curves, detector 60, median lane.

shown here) indicated that this bottleneck remained active until a queue spilled over from downstream and arrived at station 80 at approximately 18:04:03.

To confirm that the bottleneck continued to serve vehicles at a maximum rate during its active period, N- and T-curves measured in individual lanes (at all stations in the study section) were inspected. This examination revealed that a sizable flow reduction in the median lane arrived from somewhere upstream of station 50 at about 16:57:23. The flow reduction (perhaps caused by an incident) arrived at station 60 shortly afterward (at 16:57:43). To verify this, Figure 27 shows re-scaled N- and T-curves constructed from data measured in the median lane at station 50. These curves reveal a sharp reduction in flow along with a sharp reduction in the occupancy at 16:57:23, marking the arrival of a flow reduction from further upstream. Figure 28 displays re-scaled N- and T-curves constructed from data measured in the median

lane at station 60. This figure shows that the effects of the flow reduction (incident) were measured at station 60 shortly following its appearance at station 50. Therefore, queue discharge flows will only be measured from 15:13:03 until 16:58:23 when the apparent incident occurred. The analysis of discharge flows measured through the active bottleneck during this period is contained in the next section.

4.3.2 Day 2: Queue discharge features

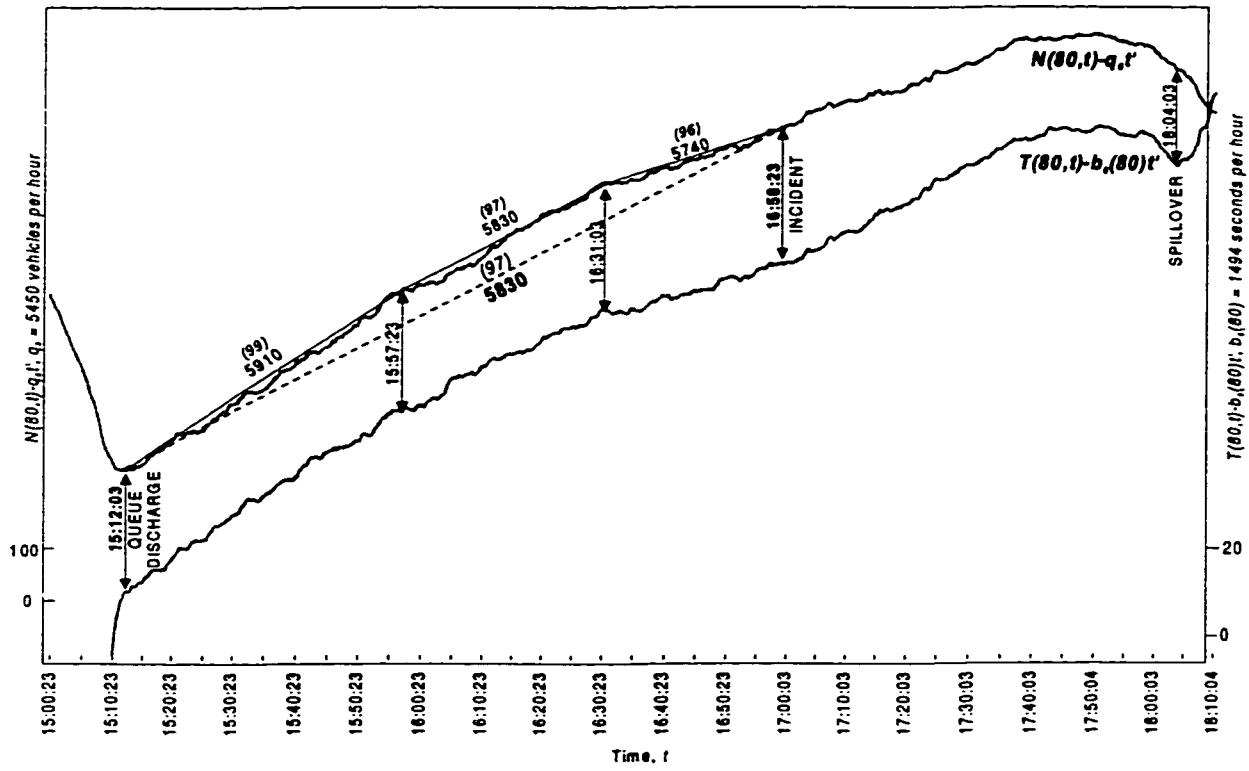

Figure 29: Day 2 Re-scaled N- and T-curves, station 80 of the Gardiner Expressway.

To study the bottleneck queue discharge flows on Day 2, Figure 29 shows re-scaled N- and T-curves for station 80 which span about 3 hours. This includes a few minutes prior to bottleneck activation where a low flow (governed by the capacity of the downstream restriction) was observed, the period during which the bottleneck was active and several minutes after its deactivation. In the figure, one can observe

that the bottleneck's deactivation was marked by the drop in the N accompanied by the sharp rise in the T at approximately 18:04:03. This marked the arrival of a queue that had spilled over from further downstream.

Figure 29 shows that during the bottleneck's active period (beginning when the queue from the downstream restriction dissipated at 15:12:03), a sequence of nearly stationary flows prevailed, without an initial, substantial drop in flow as observed on Day 1. The onset of queue discharge immediately carried a flow of 5,910 vph for approximately 45 minutes, followed by rates that slightly diminished over time. That the discharge rate decreased over time was not a reproducible feature on Days d, e and f. Further discussion of the consistency of the average discharge rates across multiple days is included in Section 4.4.

Table 5: Day 2 Gardiner Expressway queue discharge features.

Statistic	20-second count	Vertical difference from trend line
Mean	32.4 veh	34.4 veh
Variance	11.3 veh^2	673.9 veh^2
γ^a	0.35 veh	19.6 veh
Percent within $\pm 2\sigma^b$	96.3%	100.0%
Maximum^c	–	77 veh

[a] Variance-to-mean ratio.
[b] Percent of observations within plus or minus two standard deviations of the mean.
[c] Absolute value.

Figure 29 also shows that vehicles discharged through the active bottleneck at an average rate of 5,830 vph (shown by the dashed line), which was consistent with the average discharge rate observed on Day 1. From the vertical scale on the left edge of Figure 29, it is obvious that the N-curve deviated from the dashed line (denoting the average discharge rate) by no more than 77 vehicles. That the average discharge

may also be described as *nearly constant* on Day 2 can be confirmed by the data contained in Table 5. When observing the 20-second counts, the table shows that the variance-to-mean ratio of the counts measured across all lanes is 0.35, which is similar to the variance-to-mean ration for Day 1 included in Table 2. Also, 96.3 percent of the count observations fall within plus or minus two standard deviations of the mean (32.4 vehicles/20-second interval, or 5830 vph) of the counts. Furthermore, Table 5 shows that the observed vertical differences between the N-curve and the dashed average discharge trend line are distributed such that 100.0 percent of the observations are within plus or minus two standard deviations of the mean difference.

4.3.3 Day 2: Reproducing the observations

Table 6: Summary of Gardiner Expressway queue discharge flows as seen across all lanes.

Day	Date	Average discharge rate	
		Rate	Duration
	mm/dd/yy	*vph*	*h:mm*
2	03/12/97	5830	1:46
d	01/17/97	5800	0:41
e	10/08/97	6070	1:17
f	11/26/98	5950	2:58

Similar to Day 2, on Days d, e and f, the bottleneck between stations 60 and 70 was also activated subsequent to the dissipation of a downstream queue. On all four days, no high flow was observed prior to bottleneck activation since the flow was governed by the (lower) capacity of some downstream restriction. As shown in Table 6, once the bottleneck was activated between stations 60 and 70, the average discharge rates measured across all lanes exhibited small day to day deviations and were sustained

for long periods. As stated in Section 4.3.2, the discharge flow was not observed to decrease over time on Days d, e and f. In Section 4.4 the discharge flows observed in individual lanes and as measured across all lanes will be compared for all eight days considered in this study.

4.4 Reproducing the observations: All days

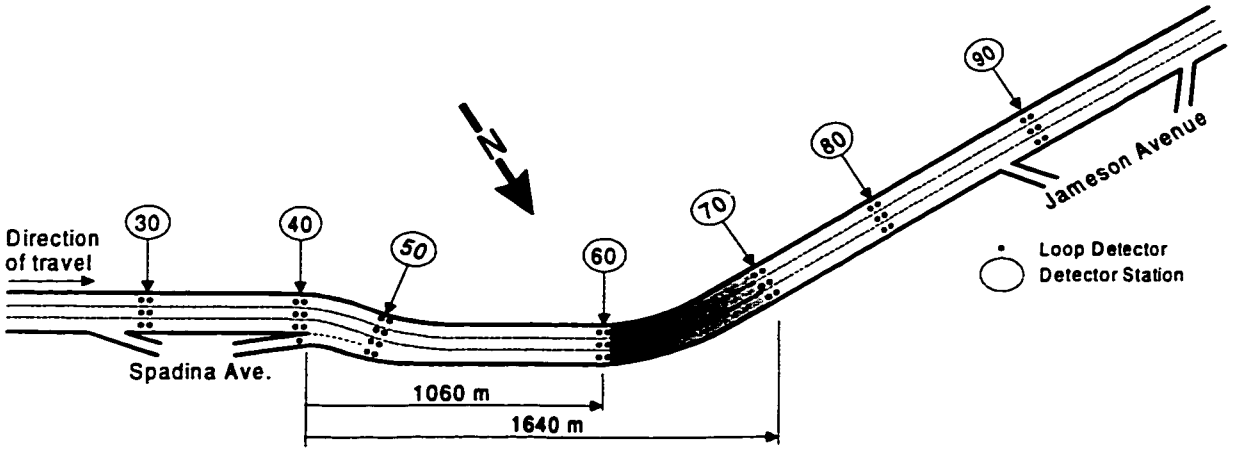

Figure 30: Gardiner Expressway bottleneck location

Several features were reproducible across the eight days that were analyzed. On all days, the bottleneck on the Gardiner Expressway near the Spadina Avenue on-ramp was activated between stations 60 and 70 (see Figure 30). Given the resolution provided by the loop detectors, this bottleneck formed between 1.06 and 1.64 kilometers (between 0.66 and 1.02 miles) downstream of the Spadina Avenue on-ramp. Because this region is located within a horizontal curve, another freeway site without a horizontal curve or other obvious inhomogeneity was analyzed (see Appendix B). At this second site, the bottleneck also consistently formed more than 1,000 meters (more than 3,500 feet) downstream of a busy on-ramp. Given the evidence revealed by this study, it is possible that the bottleneck occured well downstream of the ramp because drivers (mainly commuters very familiar with the route) were temporarily

willing to accept shorter headways as they approached and passed through the merge area. This willingness was likely influenced by drivers' desire to prevent other vehicles from moving into their lane. Once the drivers passed the merge area, they apparently relaxed, gradually increasing their headways. Section 5 will describe possible future experiments for confirmation of this conjecture.

Table 7: Summary of Gardiner Expressway queue discharge flows as seen in individual lanes.

Day	Date	Average discharge rate				Duration
		Median Lane	Center Lane	Shoulder Lane	Total	
	mm/dd/yy	*vph*	*vph*	*vph*	*vph*	*h:mm*
1	03/05/97	2340	1920	1720	5970	2:22
a	02/20/97	2290	1910	1690	5970	2:08
b	07/21/97	2330	1950	1690	5870	0:59
c	02/11/97	2290	1890	1620	5780	2:42
2	03/12/97	2320	1900	1610	5830	1:46
d	01/17/97	2390	1910	1560	5800	0:41
e	10/08/97	2350	1990	1730	6070	1:17
f	11/26/98	2290	1920	1730	5950	2:58
	Mean	2310	1920	1680	5910	$T = 14{:}51$

Table 7 summarizes the observations obtained from data collected on all eight days. Most importantly, on all days, once the bottleneck between stations 60 and 70 was activated, the average queue discharge rates were consistent when measured across all lanes (see the sixth column of the table). The mean discharge flow was observed to be 5,910 vph.[17] As shown in the table, the queue discharge rates always persisted for long periods, in contrast to the short-lived periods of high flow (observed on Days 1, a, b and c) shown in Table 4.

[17]The total number of vehicles observed on all days was divided by the total duration, T.

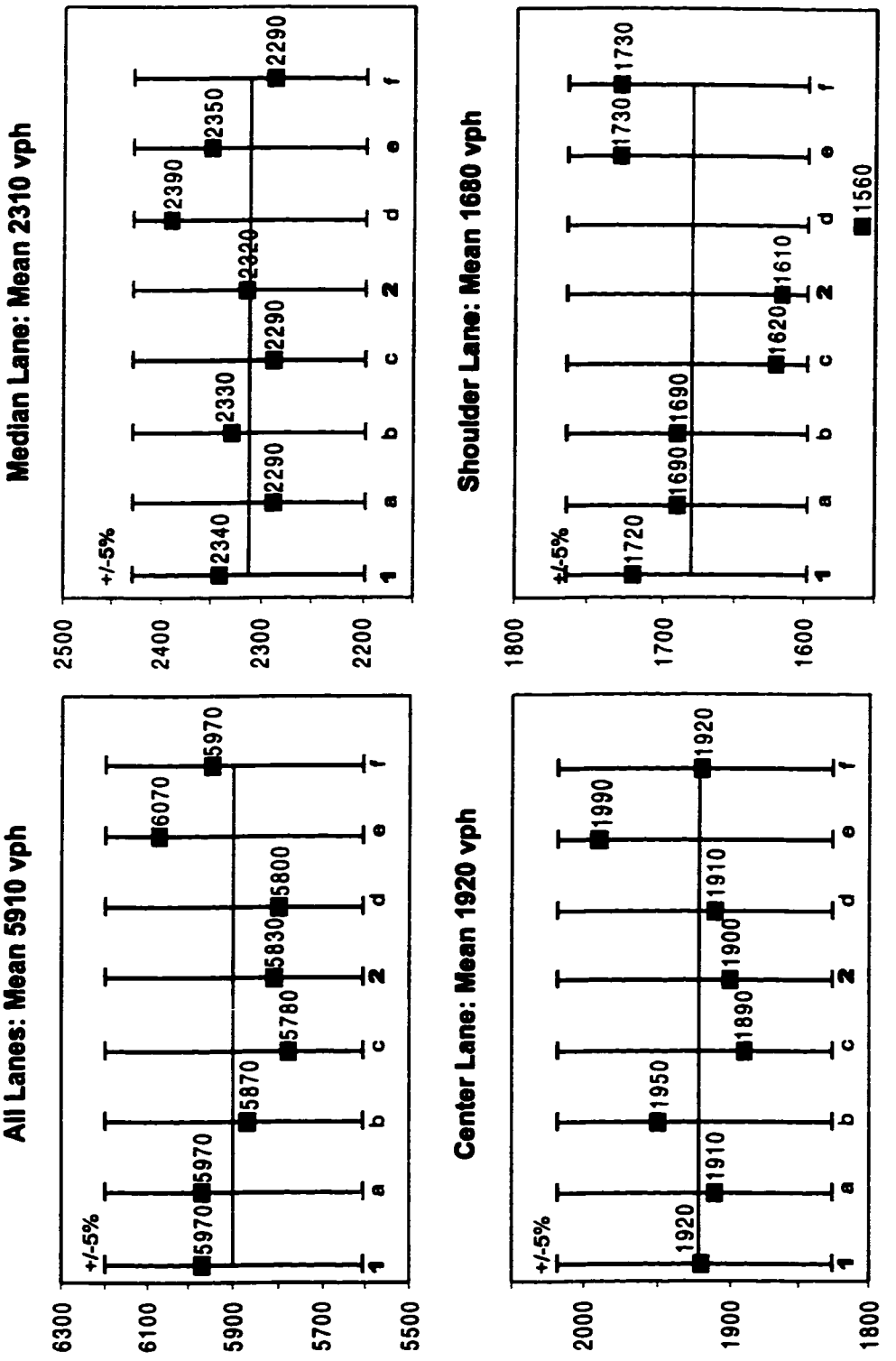

Figure 31: Reproducibility of queue discharge flows

Also of note, Table 7 indicates that, while the bottleneck was active, the average discharge flows varied across lanes, while exhibiting only small variations across days. As shown, the mean discharge rates for the median, center and shoulder lanes were 2,310 vph, 1,920 vph and 1,680 vph respectively.

To gain a visual sense of the variability of the discharge flows, Figure 31 displays points representing the average discharge rates measured across all lanes and in the individual lanes. The magnitudes of the flows are represented by the vertical axes and the eight observations are labeled across the horizontal axes. In the four quadrants of the figure, each 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 a range of plus or minus five percent of the mean flow, vertical bars in the figure span a range from 95 percent to 105 percent of the mean discharge rate. One can see from the figure that the discharge flow in the shoulder lane exhibits the highest degree of variability, while the flows measured in the median and center lanes have somewhat less variability. The only observation to fall outside of the band of plus or minus five percent is for Day d in the shoulder lane. The average discharge flow measured across all lanes only varies between -2.2% and +2.7% of its mean rate.

Finally, Table 8 displays the variance-to-mean ratios (γ) of the queue discharge counts observed during the bottleneck's active period on all eight days. To calculate γ, the mean and variance of the 20-second counts were calculated for each lane and for the total 20-second count measured across all lanes. These statistics were calculated only using counts from the periods during which the bottleneck remained active each day. It is clear from the table that the variance-to-mean ratios are generally consistent from day to day, with possibly one exception (Day e). This provides further confirmation that the queue discharge flow (in each lane and in total) should be viewed as nearly constant.

Table 8: Variance-to-mean ratios of queue discharge flows as seen in individual lanes.

Day	Date	Variance-to-mean ratio				Duration
		Median Lane	Center Lane	Shoulder Lane	Total	
	mm/dd/yy	*vph*	*vph*	*vph*	*vph*	*h:mm*
1	03/05/97	0.21	0.25	0.40	0.35	2:22
a	02/20/97	0.22	0.27	0.38	0.39	2:08
b	07/21/97	0.24	0.28	0.38	0.40	0:59
c	02/11/97	0.21	0.29	0.40	0.32	2:42
2	03/12/97	0.20	0.26	0.44	0.31	1:46
d	01/17/97	0.20	0.26	0.40	0.33	0:41
e	10/08/97	0.26	0.34	0.58	0.50	1:17
f	11/26/98	0.23	0.34	0.43	0.44	2:58

5 Conclusions

This final section contains a summary of the study and an outline of some areas for further research.

5.1 Summary of the study's findings

This dissertation described an empirical study of the spatial and temporal evolution of traffic conditions upstream and downstream of a freeway bottleneck located near a busy merge. To this end, curves of cumulative vehicle arrival number versus time and cumulative occupancy versus time were constructed from data measured by neighboring freeway loop detectors. The 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.

As a result of studying data collected from the Gardiner Expressway bottleneck on eight days, it was shown that the bottleneck always occured at a fixed location approximately 1,000 meters (more than 3,500 feet) downstream of the merge. Most importantly, no matter what happened prior to bottleneck activation, the discharge flow in the active bottleneck was nearly constant, since the cumulative counts never deviated much from a linear trend. This long-run bottleneck discharge flow was reproducible from day to day in each lane and in total. It therefore appears that the long-run queue discharge flow should be viewed as the bottleneck capacity given that the near-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.

On the four days when the bottleneck became active without interference from any downstream effects, it was shown that flow dropped substantially following the formation of an upstream queue. On these days, drivers adopted apparently aggressive behavior that resulted in high bottleneck flows prior to queue formation. These flows were followed by substantially lower queue discharge flows. On such days, reproducible signals of impending queue formation were observed for the final few minutes prior to queue formation, including the achievement of a very high flow of approximately 7,000 vph. On these days it appeared that the higher flows observed prior to queue formation were unstable, as they varied in duration and in magnitude. At this time it is uncertain whether these higher flows can be prolonged by a control strategy such as ramp metering.

It was also shown that on four other days, the subject bottleneck was activated only after a queue emanating from somewhere further downstream had dissipated. On these days, vehicles immediately began discharging through the bottleneck at a nearly constant rate without ever exhibiting a higher flow.

Finally, the City of Toronto might be able to use the results of this research as part of their current freeway management strategy. The city currently closes the Jameson Avneue on-ramp between exactly 15:00 and 18:00 every day, apparently without consideration of traffic conditions on the freeway. If the freeway managers could trace the propagation of the queues through the subject site, they might consider closing the Jameson Avenue on-ramp slightly before 15:00 if they could prevent or delay a queue from backing up into the study section. Similarly, the managers might want to re-open the ramp slightly after 18:00 to prevent the subject bottleneck from

being deactivated by queues emanating from downstream. Of course this would be subject to other important institutional and technical considerations.

5.2 Areas of further research

This research has sought a greater understanding of where a freeway bottleneck formed and how and when it was activated. This study has also examined the features of the bottleneck's discharge flow in detail. 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. To this end, some possible areas of further research are suggested in this section.

This study did not analyze 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 (1999) presented 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 periods of nearly uniform flow. From the loop detector data it was not clear why these periods existed in the traffic stream. Thus, cumulative curves constructed from loop detector data could be augmented with vehicle trajectories extracted (over short freeway segments) from video, perhaps using machine vision technology (Coifman, 1997). If possible, the video should be obtained from vantage points that are 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.

From this study it appears that that as freeway and on-ramp flows increased, traffic transitioned from free flow to queued conditions at the bottleneck due to reproducible, exogenous reasons. However, to what extent freeway bottleneck features in Toronto, Canada are similar to those in other parts of the world is not known. Thus, freeway merges should be examined using data from the United States, Europe and Asia. Similarly, merge bottlenecks should be studied under different seasonal, weather and geometric (such as the number of lanes, grades, shoulder widths, ramp design) conditions to determine if there are any notable differences.

Empirical evaluations should also be conducted to determine to what extent, if at all, ramp metering can increase bottleneck capacity. Such a study should be designed to determine to what extent eliminating, postponing or minimizing freeway queues via ramp metering can increase bottleneck capacity. If capacity gains are found to be possible, then alternative ramp metering strategies that might be used to achieve these gains should be explored and tested.

Since freeway bottlenecks come in many forms, more empirical analysis is also required at other types of inhomogeneities such as diverges, weaving areas, crests, sags, curves, tunnel entrances and lane reductions. Since the traffic patterns at each type of bottleneck may exhibit their own peculiarities, further study of bottlenecks in each of their forms seems warranted. Cumulative curves like those described in this study might be used to conduct these studies since they provide a robust way of diagnosing the details of bottleneck traffic. A greater understanding of where and how freeway bottlenecks of all types arise and how traffic features such as vehicle lane-changing affect their capacities would contribute toward the development of future theories of traffic flow and strategies for managing freeway traffic.

References

Agyemang-Duah, K. and Hall, F.L. (1991) Some issues regarding the numerical value of freeway capacity. *Proceedings of the International Symposium on Highway Capacity,* Karlsruhe, Germany, pp. 1-15.

Banks, J.H. (1991) Two-capacity phenomenon at freeway bottlenecks: a basis for ramp metering? *Transportation Research Record,* 1320, 83-90.

Banks, J.H. (1990) Flow processes at a freeway bottleneck. *Transportation Research Record,* 1287, 20-28.

California Department of Transportation (1995) *Highway Design Manual.*

Cassidy, M.J. (1998) Bivariate relations in nearly stationary highway traffic. *Transportation Research,* 32B(1), 49-59.

Cassidy, M.J. and Bertini, R.L. (1999a) Observations at a freeway bottleneck. *Transportation and Traffic Theory: Proceedings of the Fourteenth International Symposium on Transportation and Traffic Theory,* Jerusalem, Israel (in press).

Cassidy, M.J. and Bertini, R.L. (1999b) Some traffic features at freeway bottlenecks. *Transportation Research,* 33B(1), 25-42.

Cassidy, M.J. and Coifman, B. (1996) Relation among average speed, flow, and density and analogous relation between density and occupancy. *Transportation Research Record,* 1591, 1-6.

Cassidy, M.J. and Mauch, M. (1999) An observed traffic pattern in long freeway queues. Institute of Transportation Studies, University of California at Berkeley.

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

Cassidy, M.J. and Windover, J.R. (1998) Driver memory: motorist selection and retention of individualized headways in highway traffic. *Transportation Research,* 32A(2), 129-137.

Ceder, A. (1975) Investigation of two-regime traffic flow models at the micro- and macroscopic levels. Ph.D. Dissertation, University of California at Berkeley, USA.

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 at Berkeley.

Daganzo, C.F. (1997) *Fundamentals of transportation and traffic operations.* Elsevier Science Inc., New York.

Hall, F.L. and Hall, L.M. (1990) Capacity and speed-flow analysis of the Queen Elizabeth Way in Ontario. *Transportation Research Record,* 1287, 108-118.

Kerner, B.S., Konhäuser, P. and Schilke, M. (1995) Deterministic spontaneous appearance of traffic jams in slightly inhomogeneous traffic flow. *Physical Review E,* 51(6), 6343-6246.

Kerner, B.S. and Rehborn, H. (1996) Experimental features and characteristics of traffic jams. *Physical Review E,* 53(2), R1297-R1300.

Kerner, B.S. and Rehborn, H. (1997) Experimental properties of phase transitions in traffic flow. *Physical Review Letters,* 79, 4030-4033.

Koshi, M., Kuwahara, M, and Akahane, H. (1992) Capacity of sags and tunnels on Japanese motorways. *ITE Journal,* May 1992, 17-22.

Lighthill, M.J. and Whitham, G.B. (1955) On kinematic waves; I: Flow movement in long rivers, II: A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society,* 229, 281-345.

Lin, W.H. and Daganzo, C.F. (1997) A simple detection scheme for delay-inducing freeway incidents. *Transportation Research,* 31A, 141-155.

May, A.D. (1962) *The California Freeway Operations Study.* Thompson Ramo Wooldridge, Chatsworth, California, USA.

Newell, G.F. (1982) *Applications of queueing theory.* Chapman and Hall, New York.

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,* 27B(4), 281-313.

Newman, L. (1961) *Study of traffic capacity and delay at the merge of the North Sacramento and Elvas freeways.* California Division of Highways, USA.

Newman, L. (1986) *Freeway operations analysis course notes.* Institute of Transportation Studies, University of California at Berkeley, USA.

Palmer, R. (1959) *The development of traffic congestion.* M.S. Thesis, Yale University, USA.

Persaud, B.N. (1986) *Study of a freeway bottleneck to explore some unresolved traffic flow issues.* Ph.D. Dissertation, University of Toronto, Canada.

Persaud, B.N. and Hurdle, V.F. (1991) Freeway capacity: definition and measurement issues. In: *Proceedings of the international symposium on highway capacity,* Karlsruhe, Germany, pp. 289-307.

Persaud, B.N., Yagar, S., and Brownlee, R. (1998) Exploration of the breakdown phenomenon in freeway traffic. *Transportation Research Record,* 1634, 64-69.

Richards, P.I. (1956) Shock waves on the highway. *Operations Research*, 4, 42-51.

Roess, R.P. and Ulerio, J.M. (1994) New capacity analysis procedures for ramp-freeway terminals. In: Akcelik, R. *Proceedings of the second international symposium on highway capacity*, Sydney, Australia, pp. 503-522.

Windover, J.R. (1998) *Empirical studies of the dynamic features of freeway traffic.* Ph.D. Dissertation, University of California at Berkeley, USA.

Appendix A - Methods for processing traffic data

Experiments in highway traffic present a number of challenges, particularly since freeway systems are large and it is usually not possible design experiments by placing detectors at optimal locations. The data usually available come from fixed, predetermined points along the freeway, typically spaced at large intervals. Given these limitations, many previous studies have analyzed data measured at single points, using raw time series plots of flow, occupancy (the dimensionless measure of density) and average speed (Figure 32 is a time series plot of flow). In other past work, the relative features of flow-density scatterplots were sometimes compared for multiple locations (Cedar, 1975; Persaud, 1986). Figure 33 is a plot of occupancy versus flow. Other studies have attempted to analyze intensity plots of measured variables such as density or occupancy (May, 1962; Kerner, *et al.*, 1995; Kerner and Rehborn, 1996) in the attempt to understand the evolution of conditions over time and space. Figure 34 is an occupancy intensity plot for the Gardiner Expressway on March 5, 1997. While in this example it is possible to observe the general evolution of traffic from uncongested to congested conditions, as described in Cassidy and Mauch (1999), occupancy contour diagrams do not clearly show how vehicle storage may change in the queue over space.

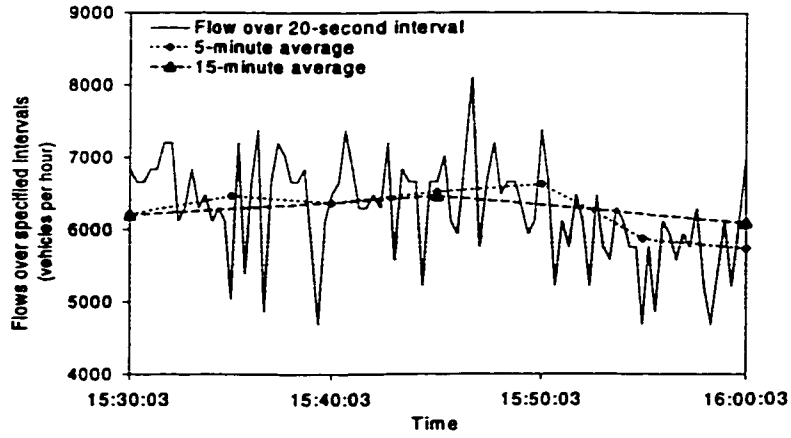

Figure 32: Time series plot of count.

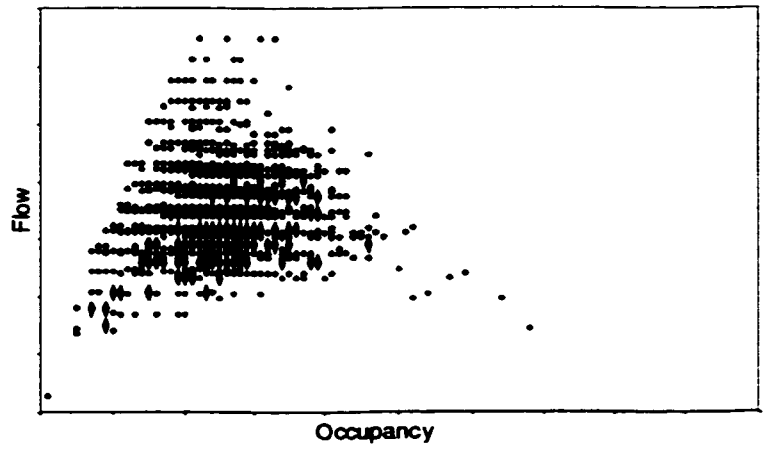

Figure 33: Bivariate plot of occupancy (in percent) versus flow.

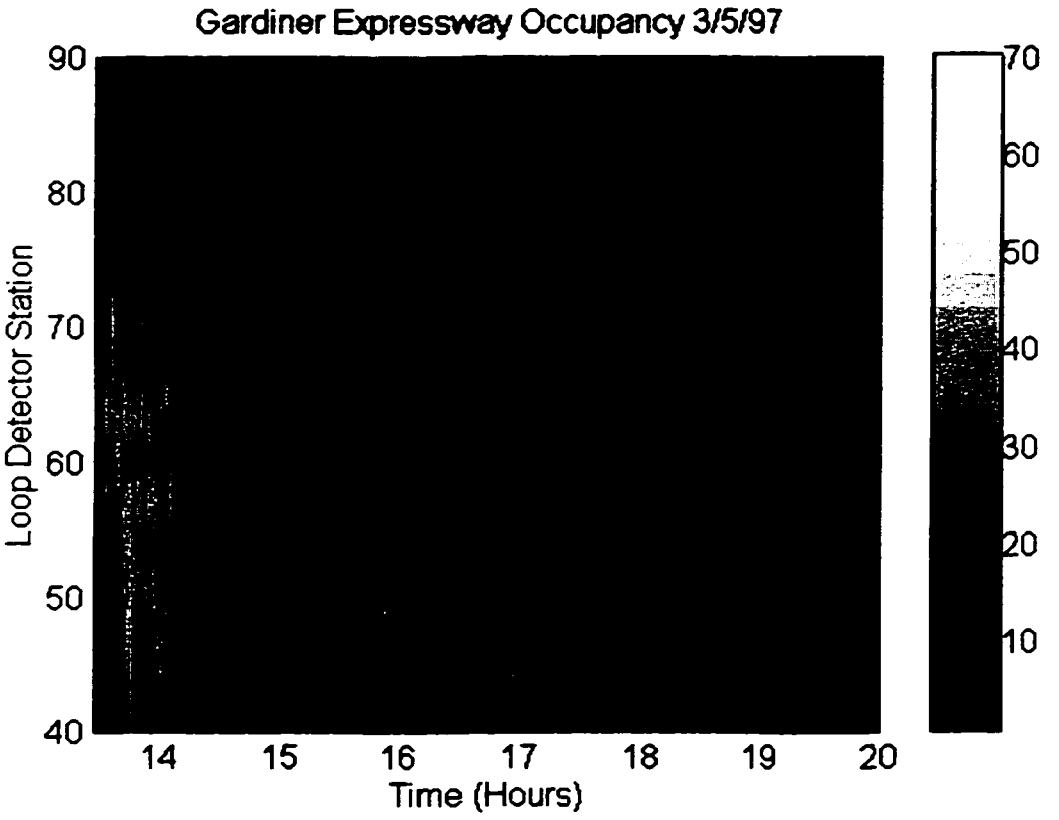

Figure 34: Occupancy (in percent) versus time and space.

Appendix B - Analysis of the Queen Elizabeth Way

Because of a concern about the influence of the horizontal curve on the location of the bottleneck on the Gardiner Expressway, another site without a horizontal curve was studied. It is shown that the bottleneck was also found about 1,000 meters (3,500 feet) downstream of the merge, again on multiple days. The Queen Elizabeth Way (QEW), illustrated in Figure 35, is also located in metropolitan Toronto, Canada and has been featured in previous studies of capacity (Agyemang-Duah and Hall, 1991; Hall and Hall, 1990). The loop detector stations for measuring traffic data are labeled in Figure 35 as per the numbering strategy adopted by the Ontario Ministry of Transport. The QEW detectors record counts, occupancies and (time) mean speeds in each lane over 30-second intervals and meters are deployed at the on-ramps upstream of the bottleneck.[18]

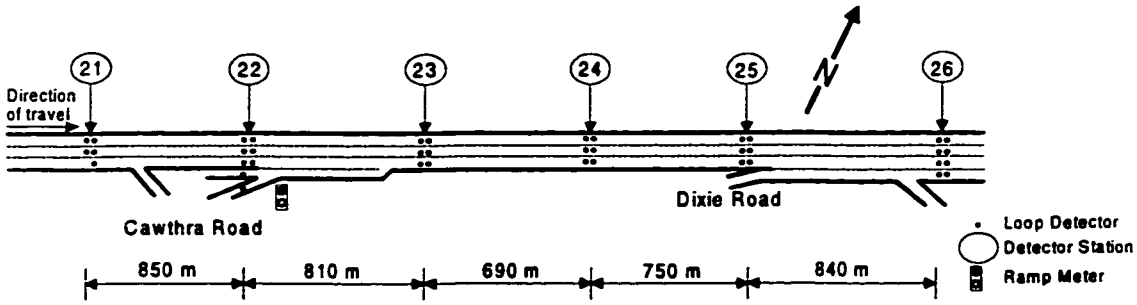

Figure 35: Queen Elizabeth Way, Toronto, Canada.

Figure 36 presents the transformed $N(x,t)$ curves for this segment of the QEW during a 24-minute portion of the morning peak period. These curves were constructed as previously described in Section xx and here x refers to detector stations 22 through 25 (as illustrated in Figure 35). The curve for station 22 includes the counts at the Cawthra Road on-ramps (to maintain vehicle conservation between

[18]David Tsui and Mark Fox, Ontario Ministry of Transportation, provided the data used in this Appendix.

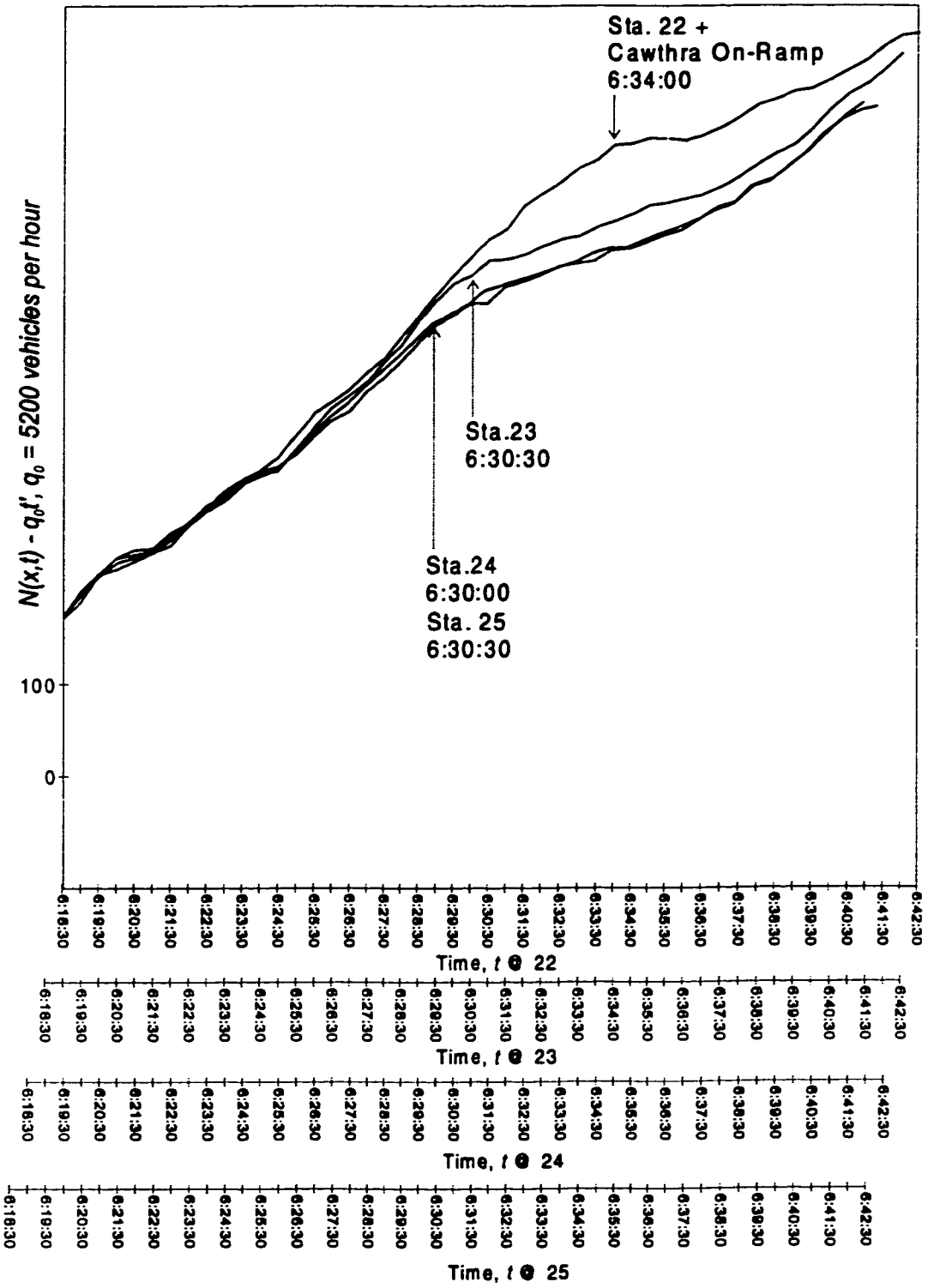

Figure 36: Re-scaled N-curves, Queen Elizabeth Way.

curves). These curves reveal that the queue formed between stations 23 and 24. Similar to the Gardiner Expressway curves in Figure 10, the Figure 36 curves begin to diverge when flows increase at upstream stations 22 and 23. Shortly thereafter, flow reductions occurred at downstream station 24 (at 6:30:00) and station 25 (at 6:30:30); these reductions were temporary as shown later in Figure 39. The queue's arrival at station 23 (at 6:30:30) is signaled by a (temporary) flow reduction at this station and this causes a divergence in curves 22 and 23.

The queue's evolution is corroborated via the re-scaled T-curves shown in Figure 37; the times annotated on this figure correspond to the starting times of the flow reductions (as observed from Figure 36). By examining the lower portion of Figure 37, the T-curves for downstream stations 24 and 25, together with the corresponding N-curves in Figure 36, it is evident that a forward-moving expansion wave propagated past these two stations in sequence. Viewing the upper portion of Figure 37 along with Figure 36 reveals that the queue emanated from downstream of station 23 and propagated in the upstream direction.

Having established that a bottleneck was between stations 23 and 24, it is shown that this bottleneck was de-activated when a queue from downstream spilled-over some time later. As a consequence, station 25 is a suitable location for measuring bottleneck flows for only a portion of the peak period.

To verify the upstream queue's continued presence, Figure 38 displays N-curves for station 22 (plus the Cawthra Road on-ramps) and for station 25. The Figure 38 curves were constructed in the same way as those in Figure 36, although the former span a longer duration. The displacements between the curves in Figure 38 show that the queue (upstream of station 25) persisted until 7:54; it dissipated following the abrupt reduction in upstream flow measured at 7:49.

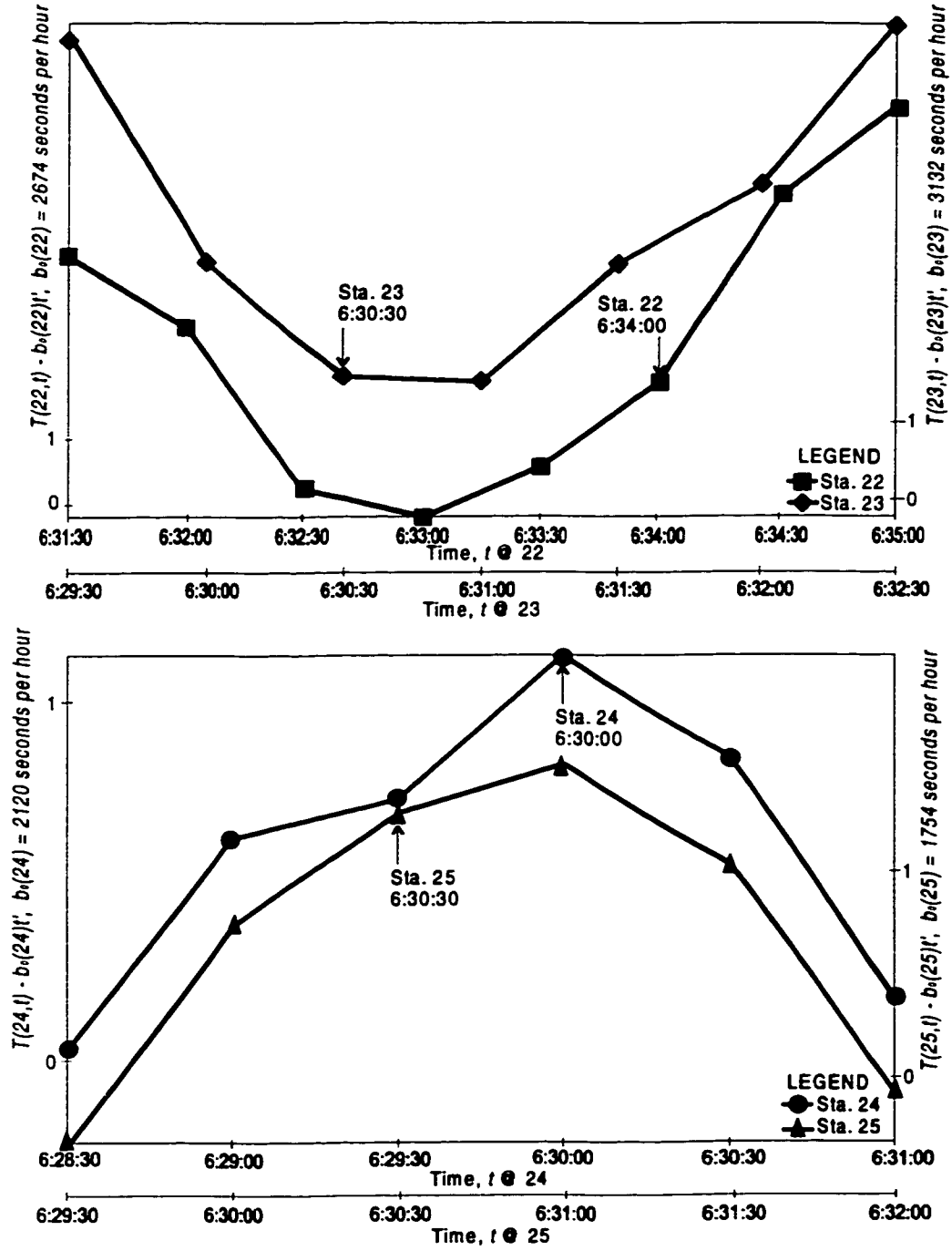

Figure 37: Re-scaled T-curves, Queen Elizabeth Way.

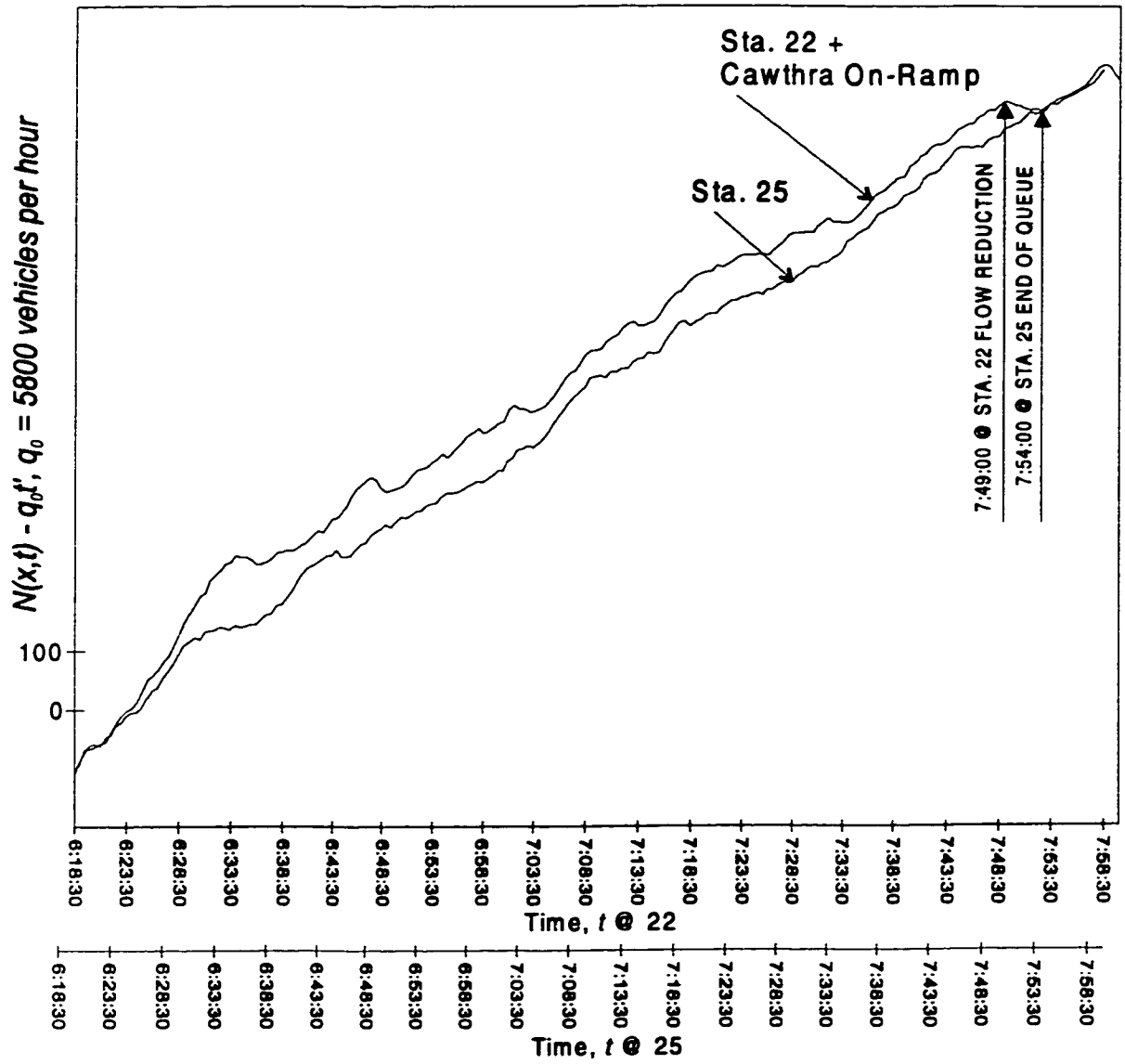

Figure 38: Upstream and downstream N-curves, Queen Elizabeth Way.

Finally, Figure 39 shows that station 25 measured discharge flows from an active bottleneck beginning with the arrival of the forward expansion wave at 6:30:30 until the queue dissipated at 7:54:00. During this period, the Figure 39 curves do not display any rises in the T accompanied by reductions in the N. To the contrary, the period is marked by near-stationary traffic with an alternating pattern of higher and lower discharge rates. As on the Gardiner Expressway, a dramatic flow reduction occurred at the onset of the upstream queue; in this instance, the low discharge rate of 6,090 vph persisted for seven minutes. Following this period, a recovery discharge rate of 6,890 vph is observed. Figure 39 also shows that, on occasion, discharging vehicles exhibited flows that approached, and even exceeded, 7,000 vph, the rate observed prior to upstream queueing. While the bottleneck was active, the N never deviated by more than about 40 vehicles from the trend line; thus, the variation in the queue discharge flows occurred about a constant rate. This average discharge rate of 6,420 vph is shown with the dashed line in Figure 39. Also shown is the flow of 7,000 vph that prevailed for 12 minutes before the queue formed upstream. Thus, the average queue discharge rate was about eight percent lower than the maximum flow absent the queue.

The analyses described were repeated using data taken during two additional mornings on the QEW. The general traffic features reported above were reproduced for the additional days studied, although some variations were observed. Table 9 summarizes the observations. The observations presented above are shown as *5/3/95* in the table. Although not shown in Table 9, the bottleneck location was consistent each day, at least within the resolution provided by the loop detectors. The QEW bottleneck always occurred between stations 23 and 24, about 1,000 meters (3,500 feet) or more downstream of the on-ramp (as shown in Figure 36), even in the absence of any horizontal curve.

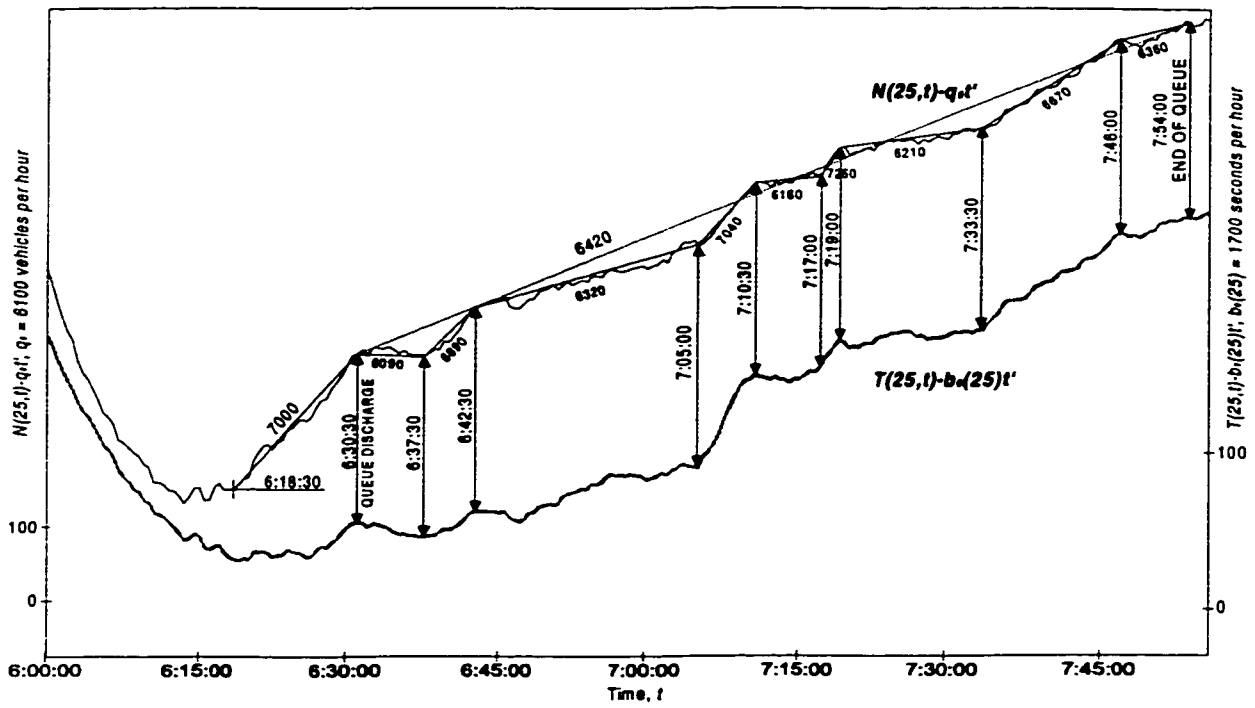

Figure 39: Re-scaled N- and T-curves, station 25 of the Queen Elizabeth Way.

Table 9: Summary of measured traffic features on the Queen Elizabeth Way.

Date	Flow immediately prior to queue		Discharge rate immediately following queue		Recovery discharge rate		Average discharge rate		Percent diff.
m/dd/yy	(a) Rate *vph*	Duration *mm:ss*	Rate *vph*	Duration *mm:ss*	Rate *vph*	Duration *mm:ss*	(b) Rate *vph*	Duration *h:mm*	a-b *%*
5/03/95	7000	12:00	6090	7:00	6890	5:00	6420	1:23	8%
5/01/95	6890	25:00	4980	5:30	6480	5:00	6420	1:23	8%
5/12/95	7120	10:30	6410	26:00	7030	4:30	6500	2:02	9%

Table 9 shows that the observed durations of the high flows prior to queueing varied considerably. Likewise, the magnitudes and the durations of the discharge flows that followed the onset of queueing (including the recovery discharge rates) exhibited daily variations.

Table 9 also shows that some bottleneck features were reproducible from day to day. For example, that the onset of queueing was accompanied by a low discharge flow followed by a higher recovery rate was a reproducible feature on each day observed. Likewise, Table 9 shows that the average queue discharge rate was consistent in that the day to day differences never exceeded 2 percent. This discharge rate always persisted for long periods.

Also consistent was the daily observation that the flow immediately prior to queue formation was substantially larger than the average discharge rate. On all three days, the difference between these rates was at least 8 percent.

Appendix C - Weather data for Gardiner Expressway

Table 10 shows relevant weather data (including high and low temperatures and precipitation quantity and type) for the city of Toronto for the days analyzed. The weather data apply to the Toronto Pearson Airport which is within a few kilometers of the Gardiner Expressway site. The data cover a 24-hour period, so there is not any particular indication of whether the small amounts of precipitation recorded on a few of the days was occurring during the afternoon peak period. The sources of these data were the Toronto *Globe and Mail* newspaper and personal contact with the Ontario Climate Center at the Canadian Ministry of the Environment.

Table 10: Toronto weather.

Day	Date	Day	Temperature				Precipitation		
			Low		High				
			°C	*°F*	*°C*	*°F*	*mm*	*in*	*type*
1	03/05/97	Wed	2.8	37.0	0.5	32.9			none
a	02/20/97	Thu	1.8	35.2	-4.5	23.9			none
b	07/21/97	Mon	14.2	19.7	57.6	67.5	4.0	0.1	rain
c	02/11/97	Tue	-1.0	30.2	-7.7	18.1	8.0	0.3	rain
2	03/12/97	Wed	-1.4	29.5	-6.2	20.8			none
d	01/17/97	Fri	-14.1	6.6	-17.6	0.3	2.0	0.1	snow
e	10/08/97	Wed	22.8	73.0	11.4	52.5			none
f	11/26/98	Thu	3.3	37.9	8.6	47.5	4.0	0.1	rain

Appendix D - Comparison of loop detector data with video data

Loop detector data were tested against ground truth traffic observations from video tape for a portion of the afternoon peak period on Wednesday, April 23, 1997. Time-stamped video surveillance from camera number 20 (shown in Figure 3) and simultaneous loop detector data from the median lane at station 80 were obtained for a 20-minute period (15:10:12 to 15:29:12) during uncongested conditions.

The arrival times of individual vehicles at an observation point just upstream of station 80 were recorded manually from the video. Using a simple computer program to append time values with keystrokes, each vehicle's arrival time was recorded. The arrival times were plotted versus vehicle number in Figure 40. Cumulative loop detector counts were also plotted, where each point reflects a 20-second observation. It is clear that the two N-curves are approximately superimposed, whereby a total of 795 vehicles were counted using both methods during the selected 20-minute period. This indicates that the loop data are well-tuned and compares favorably to ground truth information.

As confirmation of this assessment, the N-curves were then re-scaled, as shown in Figure 41, with a background flow reduction of 2,325 vph. As shown, the re-scaled N-curve obtained from video (solid line) is approximately superimposed with the re-scaled N-curve from the 20-second loop detector data. This confirms that the loop data are well-tuned, and that the 20-second aggregation does not affect the times at which flow changes occur. In fact, even with re-scaling the details are remarkably similar.

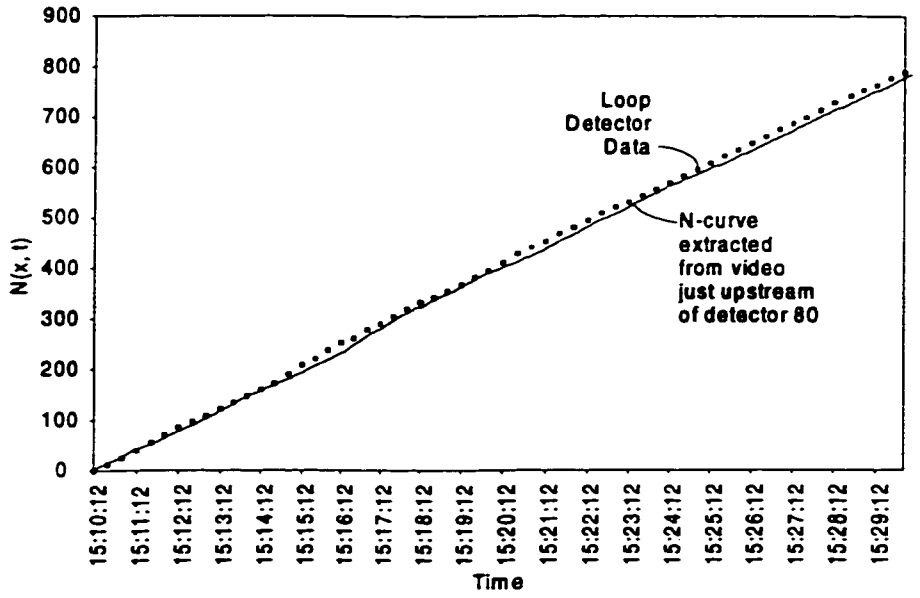

Figure 40: Comparison of vehicle counts, median lane, April 23, 1997.

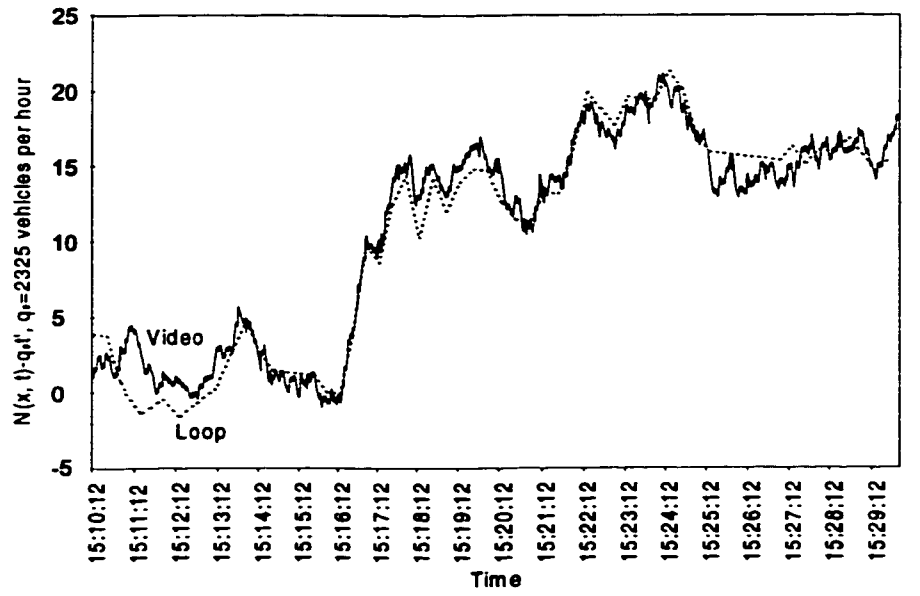

Figure 41: Re-scaled N-curves, video versus loop, April 23, 1997.

Appendix E - Gardiner Expressway observed wave speeds

Table 11: Day 1: Gardiner Expressway wave speeds as seen in individual lanes.

Detector pair	Median Lane Wave speed		Center Lane Wave speed		Shoulder Lane Wave speed		All Lanes Wave speed	
	km/h	*mi/h*	*km/h*	*mi/h*	*km/h*	*mi/h*	*km/h*	*mi/h*
40-30	-6.8	-4.3	-6.8	-4.3	–	–	-8.6	-5.4
50-40	-12.6	-7.8	-12.6	-7.8	-16.8	-10.4	-10.1	-6.3
60-50	-20.1	-12.5	-35.1	-21.9	-46.8	-29.1	-28.1	-17.4
70-80	+88.2	+54.8	+88.2	+54.8	+88.2	+54.8	+88.2	+54.8
80-90	+113.4	+70.5	+113.4	+70.5	+113.4	+70.5	–	–

Table 12: Day 1: Gardiner Expressway flow drops as seen in individual lanes.

Detector	Median Lane Flow drop		Center Lane Flow drop		Shoulder Lane Flow drop	
	Initial	Long run	Initial	Long run	Initial	Long run
30		16%		17%		
40		19%		16%		
50		17%		16%		12%
60		18%		17%		3%
70	15%	14%	12%	9%	11%	1%
80	13%	11%	12%	7%	17%	8%
90	12%	11%	9%	5%	19%	0%