Detecting Signals of Bottleneck Activation for Freeway Operations and Control

Robert L. Bertini  
Department of Civil & Environmental Engineering  
Portland State University  
P.O. Box 751  
Portland, OR 97207-0751  
Phone: 503-725-4249  
Fax: 503-725-5950  
Email: bertini@pdx.edu

August 2004
Abstract

Any application of intelligent transportation systems (ITS) to freeway operations and control relies on understanding the behavior of freeway bottlenecks. By improving our understanding of the formation and dissipation of queues at active freeway bottlenecks (points characterized by the presence of queued traffic immediately upstream and unqueued traffic immediately downstream) we can improve the techniques used to manage freeway traffic—one of the hallmarks of ITS. It is shown that freeway bottlenecks’ activations can be diagnosed definitively using archived inductive loop detector data (measured at 20- or 30-sec intervals) from two sites in Toronto, Canada. Once the bottlenecks’ locations and times of activation were determined, potential signals of their activations were explored. It is shown that certain potential signals are evident immediately before upstream queue formation, and that these are reproducible from day to day (over three or four days examined) at the two sites studied. As a result, it is suggested that these signals be used for systematic investigation of bottleneck behavior in either real time or in retrospect, on a more widespread basis as part of a real-time ITS control system.

Introduction

Any application of intelligent transportation systems (ITS) to freeway operations and control relies on an understanding of the behavior of freeway bottlenecks. These critical freeway system components are the building blocks for improving system operations, the foundation of ITS. By improving our understanding of the formation and dissipation of queues at active freeway bottlenecks we can improve the techniques used to manage freeway traffic—one of the hallmarks of ITS. In this paper we consider an “active” bottleneck to be a point upstream of which there is a queue and downstream of which there is freely-flowing traffic. Our use of the term “active bottleneck” is consistent with Daganzo (1997), in that the bottleneck would be
serving vehicles at a maximum rate and that the discharge flows measured downstream of a queue were not affected by traffic conditions further downstream. Due to their temporal and spatial variability, bottlenecks can be difficult to detect. Detection can be further hampered by the typically large spacing between freeway surveillance detectors.

The study of freeway bottlenecks near on-ramps is not a new subject, nor has it been a topic without debate. Previous studies have examined the concept of breakdown from a probabilistic standpoint, including work by Elefteriadou, Roess and McShane (1995) and Persaud, Yagar and Brownlee (1998). In addition to the large body of past research (some of which is mentioned below), in earlier studies, the author systematically has examined traffic conditions upstream and downstream of freeway bottlenecks located near busy on-ramps. In these particular studies, several reproducible features were observed (Cassidy and Bertini, 1999; Cassidy and Bertini, 1999a; Bertini, 1999; Bertini and Cassidy, 2002). For example, it was shown that these bottlenecks were located more than 1 km (0.62 mi) downstream of the on-ramps, farther than had been found previously in some studies (Persaud, 1986; Persaud and Hurdle, 1991), but consistent with what was proposed theoretically by Buckley and Yagar (1974). Also, on certain days, sustained high bottleneck flows of more than 7,000 vehicles per hour (vph) were measured (across three lanes) for up to 40 minutes before queueing arose upstream of the bottleneck. This gave rise to measurably (on the order of 10%) lower discharge rate that prevailed for much longer periods. This is consistent with other studies that have reported that freeway capacity diminishes by an average of 3 or 4% following the formation of an upstream queue (Agyemang-Duah and Hall, 1991; Banks, 1990, 1991). However, this was in contrast with other studies that reported no such flow reductions upon bottleneck activation (Persaud, 1986; Persaud and Hurdle, 1991; Hall and Hall, 1990; Newman, 1961; Roess and
Ulerio, 1994). It should also be noted that two studies of the same bottleneck have claimed both extremes (Hall and Hall, 1990; Agyemang-Duah and Hall, 1991).

To promote the identification of time-dependent traffic features, these studies used cumulative curves of vehicle count and occupancy constructed from data measured at neighboring freeway detectors (Cassidy and Windover, 1995). Transformations of these curves provided the enhanced measurement resolution necessary to observe transitions between unqueued and queued conditions and to identify notable, time-dependent features. The same technique was also used in this study, which adds to previous findings by explicitly diagnosing the activation of two bottlenecks located near freeway merges and examining potentially detectable signals of their activation. On a total of seven days examined at two different sites, particularly high flows were measured just upstream of the bottleneck location beginning about two minutes prior to bottleneck activation. The next section contains a brief description of the freeway sites and the loop detector data used in this study.

**Detector Data**

The observations that follow were taken from the 3.3 km (2.1 mi) segment of the Gardiner Expressway and the 3.9 km (2.5-mile) segment of Queen Elizabeth Way (QEW) in Toronto, Canada, illustrated in Figure 1. Both sites have been featured in previous studies of freeway bottleneck capacity features (Bertini, 1999; Bertini and Cassidy, 2002; Cassidy and Bertini, 1999; Cassidy and Bertini, 1999a; Agyemang-Duah and Hall, 1991; Hall and Hall, 1990; Persaud, 1986; Persaud and Hurdle, 1991).

Inductive loop detectors recorded vehicle count, occupancy (the percent time a detector was covered by a vehicle) and time mean speed in each lane over 20-second intervals on the Gardiner Expressway and over 30-second intervals on the QEW. All measurements were made
during days when the local weather bureau reported clear skies and no measurable precipitation. The Gardiner Expressway has no ramp metering, although the Jameson Ave. on-ramp is closed between 15:00 and 18:00 each afternoon. The horizontal curve between detectors 60 and 70 has a radius of 800 meters (2,600 ft) which is deemed to be comfortable for vehicular travel at 105 km/hr (65 mi/hr) under ideal conditions. On the QEW, ramp meters are in place at the on-ramps.

The archived detector data from the Gardiner Expressway were validated using video observations taken during a portion of the afternoon peak period on Wednesday, April 23, 1997. Time-stamped video surveillance from the closed-circuit television camera shown in Figure 1 and simultaneous loop detector data from the median lane at detector 80 was obtained for a 20-minute period (15:10:12 to 15:29:12) during unqueued conditions.

The upper portion of Fig 3 presents curves of $N(x,t)$ plotted from both video and loop detector data. Here, $N(x,t)$ is defined as the cumulative number of vehicles to pass detector 80 by time $t$. The arrival time for each vehicle was recorded manually from archived video recorded just upstream of detector 80. Each passing vehicle was numbered sequentially. Next, a set of points was plotted using the passage time as the abscissa and the vehicle number as the ordinate, resulting in a step function with each step height equal to one vehicle. Next, linear interpolations were taken through the outer edge of each stair step. The resulting curve (thin line on the figure) has a slope equal to the flow past $x$ at some time $t$.

Similarly, the loop detector from the median lane at detector data reported vehicle counts in 20-second intervals. The $N(x,t)$ from the loop detector data was also plotted as a step function, but this time the width of each step was 20 seconds. A piece-wise linear approximation was also used, such that the slope of the $N(x,t)$ is the flow measured by the loop detector past station 80 at any time $t$. As shown in the figure, a total of 795 vehicles were counted by the two independent
data sources during the selected period. This indicates that the detector data compare favorably to ground truth information obtained from the video.

In order to magnify the curves’ features in the lower portion of Figure 2, each \( N(x,t) \) was re-scaled by subtracting a background cumulative count, \( q_0(t') \), at all time \( t \), where \( q_0 \) is defined as the oblique rescaling rate and \( t' \) is the elapsed time from the beginning of the curve. A suitable choice of \( q_0 \) promotes the visual identification of changing flows directly from the oblique curve, as shown in Cassidy and Windover (1995). Note that the oblique \( N(x,t) \) is now shown on an amplified vertical scale and that changes in flow can be observed clearly from the figure. This procedure is also described in several references (Cassidy and Windover, 1995; Cassidy and Bertini, 1999). As shown, the oblique \( N(x,t) \) obtained from video (solid line) is approximately superimposed with the oblique \( N(x,t) \) from the detector. This confirms that the detector data were well tuned, and that the 20-second aggregation did not affect the times at which flow changes occurred. In fact, even with oblique plotting the details are remarkably similar.

The next section describes the bottleneck’s definitive diagnosis, including its location, time of activation and deactivation and some discharge features that were uncovered on the Gardiner Expressway. Subsequent sections present findings related to bottleneck input flows and other signals of bottleneck activation on the Gardiner Expressway. The next section describes the bottleneck features observed on the QEW followed by a section describing the QEW’S reproducible bottleneck input flows. The observations were reproduced on additional days at both sites, as described in appropriate sections. Lastly, some final comments and suggestions for future research are provided.

**Diagnosing the Bottleneck in Detail on the Gardiner Expressway**

In order to pinpoint bottleneck activation and deactivation, Gardiner Expressway traffic features
were analyzed using data from Feb. 11, 1997. Figure 3 presents oblique $N(x,t)$, cumulative vehicle arrival number, constructed from counts measured across all lanes at detectors 40 through 80, over a 50-minute period during the afternoon. As in Figure 2, these $N(x,t)$ were constructed by taking linear interpolations through the 20-second counts so that a curve’s slope at time $t$ would be the flow past location $x$ at that time. The counts for each curve in Figure 3 began ($N=0$) so that all curves describe the same collection of vehicles (curve 40 includes counts from the Spadina Ave. on-ramp). In this and later figures a relative vertical scale is shown for convenience. As in a standard queueing diagram, any horizontal and vertical separations between unaltered curves would have been the trip times and vehicle accumulations between detectors, respectively (Newell, 1982; Newell, 1993).

However, these curves were altered by shifting each curve in Figure 3 horizontally to the right by the average free-flow trip time between the respective detector and downstream detector 80. Thus, any resulting vertical displacements between curves are the *excess* vehicle accumulation between detectors and resulting horizontal displacements are the *excess* trip time (delay) between detectors. The same value of $q_0$ was used for all curves and therefore did not affect the vertical separations (Cassidy and Windover, 1995). The use of an oblique coordinate system magnifies changes in vehicle arrival rate and the times at which these changes occurred.

All five curves in Figure 3 appear superimposed between 15:00 and 15:18, indicating the presence of freely flowing traffic throughout this freeway section during this period. The curves for detectors 70 and 80 remained nearly superimposed for the entire period shown, indicating that traffic continued to flow freely between these detectors. Beginning about 15:18, excess accumulation was visible between detectors 60 and 70 followed by flow reductions at detectors 70 and 80 at 15:21:23 and 15:22:03 respectively. The divergence of the detector 50 curve from
the one at detector 60 (at 15:19:03) marked the arrival of a backward-moving queue at detector 60. A pronounced flow reduction at detector 60 accompanied this divergence. The continued presence of freely flowing traffic between detectors 70 and 80 accompanied by excess vehicle accumulation upstream of detector 70 revealed that the bottleneck was located somewhere between detectors 60 and 70. The exact location is beyond the spatial resolution of the detectors.

Figure 3 also maps the propagation of the queue further upstream. A flow reduction at detector 50 was observed at 15:27:23, when the curve at detector 50 deviated from that at detector 40. To verify the arrival of the backward-moving queue at each detector, curves of cumulative occupancy, $T(x,t)$, were also constructed. Cumulative occupancy is proportional to the cumulative travel time measured across the detectors, which by definition is inversely related to the vehicles’ speed (see Cassidy and Coifman (1997) for a discussion of the relation between occupancy and velocity). Oblique coordinates were also used, where $T(x,t)$ was reduced by $b_0(t')$, $b_0$ was an oblique scaling rate and $t'$ was the elapsed time from the beginning of each curve. The inset in Figure 3 contains an oblique curve of cumulative occupancy, $T(x,t)$, versus time, measured at detector 50. A sharp increase in cumulative occupancy occurred at 15:27:23 signaling an abrupt increase in vehicles’ travel times across the detectors. This verifies the arrival of the queue at that time and location. Thus, Figure 3 has made it possible to definitively diagnose the bottleneck’s location (between detectors 60 and 70), the time it became active (15:19:03) and to map the queue’s upstream propagation. Figure 3 also shows the average measured velocities (across all lanes) both before and after queue discharge/arrival.

The bottleneck between detectors 60 and 70 remained active until a spillover from a downstream restriction arrived at detector 80 at 18:12:03. To demonstrate this, Figure 4 displays oblique $N(x,t)$ from detectors 60 and 80 for a longer period. The queue’s presence between
detectors 60 and 80 is confirmed by the continued vertical displacement between the two curves.

The inset in Figure 4 demonstrates that the backward-moving queue arrived at detector 80 at approximately 18:12:03. The inset contains oblique $N(x,t)$ and oblique $T(x,t)$ curves for detector 80. The two curves reveal that a reduction in the $N(x,t)$ accompanied a rise in the $T(x,t)$ at approximately the same time, deactivating the bottleneck by restricting its flow. To trace the spillover, one can see from Figure 4 that the flow reduction was also measured at detector 60 a short time later (at 18:19:23). An oblique $T(x,t)$ at detector 60 (not shown) confirmed that the downstream queue arrived at that detector at approximately 18:19:23. Even after these flow reductions, the queue between detectors 60 and 80 persisted until 19:00.

Figures 3 and 4 have verified this bottleneck’s location on the Gardiner Expressway and its activation and deactivation times. The bottleneck’s queue discharge features were examined during its active period using data recorded from detector 80 (downstream of the bottleneck). Figure 5 shows oblique $N(80,t)$ and $T(80,t)$ which reveal no abrupt reductions in the $N(x,t)$ accompanied by rises in $T(x,t)$ between 15:22:03 and 18:12:03. Thus it is apparent that there was no disruption of active bottleneck discharge caused by a queue from further downstream.

Figure 5 also shows that a flow of 6240 vph was measured between 15:05:23 and 15:22:03 (the beginning of queue discharge). Upon queue formation, a lower flow of 5650 vph prevailed for about 57 minutes. This was followed by sequences of nearly constant flow until the downstream spillover deactivated this bottleneck. From the perspective of modeling queue evolution, this sequence of flows did not deviate much from the mean discharge flow of 5790 vph (dashed line), which was eight percent lower than the flow prior to bottleneck activation. This mean bottleneck discharge flow prevailed for 3 hours 40 minutes. Figure 5 confirms that queue discharge was accompanied by a reduction in flow, apparently consistent with Agyemang-
Duah and Hall (1991) and Banks (1990, 1991) and contradicting Persaud (1986), Persaud and Hurdle (1991), Hall and Hall (1990) and Newman (1961). As noted in Cassidy and Bertini (1999) and Bertini (1999), this may be due to methods used for processing data or other factors.

Figure 6 shows oblique $N(x,t)$ constructed from counts measured on the un-metered Spadina Ave. on-ramp. The increase in ramp flow observed at 15:06:43 corresponded with the measured increase in upstream flow previously revealed in Figures 3 and 5. Similarly, the timing of the reduction in ramp flow at 15:16:03 matched closely with the arrival of the queue from the active bottleneck (the queue arrived at detector 60 at 15:19). Although this queue may have suppressed the on-ramp flow, vehicles continued to enter the freeway at a high rate; an average ramp flow of 1,740 vph persisted for more than 25 minutes. The ramp flow dropped at about 15:48:03, perhaps due to a reduction in on-ramp demand. The relatively high ramp flows indicate that, just downstream of the merge, more than half of the vehicles traveling in the shoulder lane originated from the on-ramp. Figure 7 shows an oblique curve illustrating the ratio of the number of vehicles passing the Spadina on-ramp detector to the number of vehicles traveling in the shoulder lane at detector 40. As shown by the slope of this curve, this ratio was always greater than 1. 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.

**Bottleneck Input Flows on the Gardiner Expressway**

Toward a better understanding of how this high bottleneck input flow influenced and possibly signaled queue formation, traffic features were studied in the individual lanes at detector 60, immediately upstream of the bottleneck. Figure 8 contains oblique $N(x,t)$ constructed from data measured in the individual lanes at detector 60 for a 40-minute period spanning the onset of queueing. The same value of $q_0$ was used for the median lane curve (upper curve in the figure),
center lane curve and shoulder lane curve. The times at which the queue arrived in each lane are also shown. The dashed line superimposed on the upper curve shows that the mean flow in the median lane between 15:00:03 and 15:18:03 was 2,540 vph. This high flow apparently constrained lane changing into the median lane, leading to successively higher flows in the center and shoulder lanes. At 15:10, ten minutes prior to queue arrival, flows in the center and shoulder lanes became approximately equal. For the final two-minute period before queue formation (highlighted by the shaded vertical strip), a flow of more than 7,100 vph was measured across three lanes. This high flow immediately preceding queue formation marked the first and only time that a flow of this magnitude was sustained for as long as a two-minute period.

This is also visible in Figure 9, which shows the total bottleneck input flow measured across all three lanes at detector 60. In this figure, the 20-second counts were averaged over two-minute periods. As shown, the high flow of 7,100 vph was measured immediately preceding bottleneck activation, but was not measured at any other time. To explore the concept of a “maximum” flow as detected over several measurement periods, Figure 10 shows step functions recording the maximum flow measured at detector 60 (total across all lanes) just before and immediately after queue formation. The flow is averaged across measurement periods between 40 seconds and 2 minutes. As shown, these functions reach their maxima either just before or upon bottleneck activation. The heavy line used for the two-minute average clearly reveals a flow increase about two minutes before queue formation—thus from a traffic management perspective a two-minute average may help signal impending queue formation.

Table 1 shows that a maximum two-minute flow was also achieved on the Gardiner Expressway on a total of four days that were analyzed. Knowledge of this apparently critical flow (mean value of 7,050 vph) signaling the impending queue formation would be important
when considering any on-line techniques for attempting to sustain these higher flows, such as through ramp metering or variable speed control.

If one wanted to develop an automated procedure for searching a large database containing freeway loop detector data for signs of bottleneck formation, it would be possible to compare upstream and downstream sensor locations while searching for abrupt signals of excess vehicle accumulation. Figure 3 revealed 15:18 as the time at which excess vehicle accumulation was visible between detectors 60 and 70 by showing that the oblique $N(x,t)$ measured at the two detectors began to diverge at that time. Figure 11 shows an alternative method of plotting the same data. The scatter plot $(N(60,t)-N(70,t))$ shows the difference between cumulative counts measured across all lanes at detectors 60 and 70 at 20-sec intervals. The oblique cumulative curve of these data superimposed on the figure shows a sharp upward surge at 15:17:03, two minutes before the queue was shown to arrive at detector 60 (in Figure 3). This provides another way to diagnose potential excess vehicle accumulation upstream of a bottleneck’s location.

In Figure 8, it was shown that heavy net lane inflow rates resulted in very high flows in the median lane just prior to bottleneck activation (see also Cassidy and Bertini, 1999a). In order to explore whether notable changes in net lane inflow were visible in the immediate vicinity of the bottleneck, Figure 12 shows curves of net lane inflow for each lane between detectors 60 and 70. As shown, several minutes before bottleneck activation, there was a reduction in center lane net inflow, followed by an increase in net shoulder lane inflow, followed immediately by a reduction in the median lane net inflow.

As another potential signal of bottleneck activation, Figure 13 shows a bar chart of the variance of the counts recorded at detector 60, measured over rolling 2-minute intervals. As shown, the count variance dropped visibly upon queue formation. This is not surprising, since
vehicles were discharging from a queue. To amplify this feature, the variance is also plotted cumulatively in the figure, using an oblique axis to magnify the details of the curve. This shows that the variance dropped at the beginning of the period of sustained high flow prior to bottleneck activation (15:05:23 as shown in Figure 5) and then dropped again at 15:16:03, several minutes before the queue arrived at detector 60. Observing the cumulative variance during the bottleneck’s active period indicates that the count variance remained constant while vehicles were discharging from the queue. The variance then increased at approximately the time at which the bottleneck was deactivated (18:12:03) due to the downstream spillover. This may be a promising signal for further analysis.

**Other Signals of Queue Formation on the Gardiner Expressway**

As further indicators of changing traffic conditions in individual lanes just upstream of the bottleneck, Figure 14 shows oblique $V(x,t)$ for detector 60. The $V(x,t)$ were the cumulative time mean speeds measured at station $x$ by time $t$, where the slopes of the $V$ were “speed rates” measured at location $x$ by time $t$. Oblique coordinates were also used to identify periods of nearly constant mean speed and times marking notable changes in measured speed. As shown, speed reductions were observed in each lane at the precise time that the queue arrived at detector 60.

The next sections illustrate that the findings reported for the Gardiner Expressway are reproducible at another freeway site in the Toronto metropolitan area, with different geometric and control characteristics.

**Diagnosing the Bottleneck on the QEW**

It is shown that a bottleneck became active about 1,000 meters (3,500 feet) downstream of a merge at the Queen Elizabeth Way (QEW) site, illustrated in Figure 1. Figure 15 presents transformed oblique $N(x,t)$ for this segment of the QEW during a 38-minute portion of the
morning peak period on May 12, 1995. The curve for detector 22 includes the counts at the
Cawthra Road on-ramps to maintain vehicle conservation. This figure reveals that the detector
23 and 24 curves began to diverge after flows increased at upstream detectors 22 and 23. Shortly
thereafter, flow reductions occurred at downstream detector 24 (at 6:43:00) and detector 25 (at
6:43:30). The queue’s arrival at detector 23 (at 6:41:00) was signaled by a flow reduction at this
station and this caused a divergence in curves 22 and 23. This confirms that the bottleneck
became active between detectors 23 and 24 at 6:41:00.

The queue’s arrival time at detector 23 is verified by the inset in Figure 15 showing
oblique $T(x,t)$ measured at detector 23. The increase in cumulative occupancy and flow
reduction measured at detector 23 at 6:41:00 confirm the queue’s arrival time at that location. To
verify the upstream queue’s continued presence, Figure 16 shows oblique $N(x,t)$ for detectors 23
and 25 for a longer period. As indicated by the continued vertical displacement between the two
curves, the queue between detectors 23 and 25 persisted until around 8:43:00 when the $N(x,t)$
again became superimposed.

Oblique curves from detector 25 (downstream of the bottleneck) were used to examine
the bottleneck’s discharge features. Figure 17 consists of oblique $N(25,t)$ and $T(25,t)$ measured
from the active bottleneck, beginning with the arrival of the forward expansion wave at 6:43:00
until the queue dissipated at 8:43:00. During this period, the Figure 17 curves do not display any
reductions in the $N(x,t)$ accompanied by rises in the $T(25,t)$. 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 Exspressway, a high flow (7,120 vph) prevailed here for almost 16 minutes
and was followed by a substantial flow reduction at the onset of the upstream queue; in this
instance, the low discharge rate of 6,410 vph persisted for 26 minutes. Following this period, a
recovery discharge rate of 7,030 vph was observed. The mean discharge rate of 6,490 vph (shown with a dashed line in Figure 17) prevailed for two hours. On this day the mean queue discharge rate was about eight percent lower than the maximum flow absent the queue.

Figure 18 shows an oblique $N(x,t)$ constructed from the counts measured on the metered Cawthra on-ramps (adjacent to detector 22). The sustained surge in ramp flow evident at 7:08:30 corresponds with the increased discharge flow previously revealed in Figure 17. Further, Figure 19 shows an oblique curve of the ratio of number of vehicles passing the Cawthra on-ramp detectors to the number of vehicles traveling in the shoulder lane at detector 22. As shown, as a result of the metering strategy, this ratio was always less than 1.

**Bottleneck Input Flows on the QEW**

After the QEW bottleneck’s location and active period were determined, the period exhibiting higher flows prior to queue formation was examined. Traffic features were studied in the individual lanes at detector 23, the measurement location immediately upstream of the bottleneck.

Figure 20 contains single lane oblique $N(x,t)$ for detector 23 for a 60-minute period spanning the onset of queueing. The same value of $q_o$ was used for each curve. The small vertical arrows accompanying the lane labels mark the arrival of the queue in each lane and that these were determined from figures not shown here. The thin solid line superimposed on the upper curve shows that the mean flow in the median lane between 6:18:30 and 6:41:30 was 2,660 vph. This high flow may have constrained lane changing into the median lane, leading to increased flows in the center and shoulder lanes. The general upward bending of the center and shoulder lane $N(x,t)$ indicates that the flows in these two lanes were increasing. Then, beginning at 6:29, 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 more than 7,700 vph was recorded.

Figure 21 shows the total bottleneck input flow measured across all three lanes at detector 23. In this figure, the 30-second counts have been averaged over two-minute periods. As shown, the high flow of more than 7,700 vph was measured just before bottleneck activation, but was not measured at any other time before or after that time. As on the Gardiner Expressway, the achievement of this high flow immediately preceding queue formation on the QEW marked the first and only time that this high flow measurement was sustained for as long as a two-minute period. Table 2 shows that a maximum two-minute flow was achieved on three days that were analyzed. This indicates that the similar finding using data from the Gardiner Expressway is reproducible at one other Toronto metropolitan area freeway site with different characteristics.

Final Comments
As a major linchpin of ITS, substantial investments have been made in freeway surveillance systems that make freeway management, operations and information systems possible. These systems are aimed at reducing both recurrent and nonrecurrent congestion, in order to make our transportation network safer and more efficient. It is thus reasonable to attempt to use data generated by these surveillance systems to understand the details of recurrent bottleneck formation as one ingredient of a more expansive ITS-oriented freeway management strategy. By using the data reported by the surveillance system to improve our understanding of these details, we can improve the current strategies that we apply, such as ramp metering and traveler information, and expand our horizons toward other future strategies.

Toward this end, in this paper, the locations, activation and deactivation times and other features have been diagnosed for two freeway bottlenecks in Toronto, Canada, using archived loop detector data. Further, as the main point of the paper, it is shown that an apparently reproducible
signal of high bottleneck input flow was measured two minutes before bottleneck activation at both sites. This signal would be helpful toward developing any system for prolonging the high flow that prevailed prior to queue formation. For example, since many bottlenecks are recurrent, an improved understanding of their operation could inform the design of a corridor-wide ramp metering system. Prior to bottleneck activation, metering rates could be temporarily reduced. The research described in this paper did not include attempt to design such a system. In addition, since the research was only able to use archived loop detector data, it was not possible to conduct real-time experiments involving changes to the metering strategy at the QEW site. However, this should be the subject of future research and field experimentation. Further, other potential signals of impending bottleneck activation are presented, which may be helpful tools for systematic approaches for detecting bottlenecks on freeway systems, as described in Skabardonis, Choe and Varaiya (2002).

These signals are not presented as substitutes for the methodical approach described in this paper, but are proposed as possible tools for narrowing down the number of candidate locations on a larger system. Further, any system for pinpointing recurrent freeway bottlenecks must be capable of distinguishing them from queues caused by incidents (Lin and Daganzo, 1997). For example, an incident resulting in a blocked lane would reveal a much larger flow reduction (on the order of 25% for a three-lane section) upon queue formation than would a recurrent bottleneck (on the order of 10%). This is the subject of ongoing research.

Acknowledgements

Lisa Maasland and David Nesbitt, City of Toronto, and David Tsui and Mark Fox, Ontario Ministry of Transportation generously provided the data used in this study. Prof. Michael Cassidy and the late Prof. G.F. Newell also provided helpful guidance and encouragement.
Discussions with Prof. Alexander Skabardonis and Prof. Pravin Varaiya also prompted some aspects of this paper. The author appreciates the excellent suggestions provided by two anonymous reviewers that resulted in vast improvements to this paper.

References


Newman, L. (1986). Freeway operations analysis course notes, Institute of Transportation Studies, University Extension, University of California, Berkeley, U.S.A.


**LIST OF FIGURES**

Figure 1 Freeway sites, Toronto Canada
Figure 2 Loop detector validation, Gardiner Expressway, detector 80
Figure 3 Oblique N(x,t), Gardiner Expressway, Feb. 11, 1997
Figure 4  Oblique $N(x,t)$, detectors 60 and 80, Gardiner Expressway, Feb. 11, 1997.
Figure 5  Oblique $N(x,t)$ and $T(x,t)$, detector 80, Gardiner Expressway, Feb. 11, 1997.
Figure 6  Oblique $N(x,t)$, Spadina Ave. on-ramp, Gardiner Expressway, Feb. 11, 1997.
Figure 7  Oblique merge ratio curve, Gardiner Expressway, Feb. 11, 1997.
Figure 8  Oblique $N(x,t)$, detector 60, Gardiner Expressway, Feb. 11, 1997.
Figure 9  Average flow, detector 60, Gardiner Expressway, Feb. 11, 1997.
Figure 10  Maximum flow step function, detector 60, Gardiner Expressway, Feb. 11, 1997.
Figure 11  Oblique cumulative count difference between detectors 60 and 70, Gardiner Expressway, Feb. 11, 1997.
Figure 12  Net lane inflow for each lane between detectors 60 and 70, Gardiner Expressway, Feb. 11, 1997.
Figure 13  Count variance (detectors 60) and cumulative variance, Gardiner Expressway, Feb. 11, 1997.
Figure 14  Oblique $V(x,t)$, detector 60, Gardiner Expressway, Feb. 11, 1997.
Figure 15  Oblique $N(x,t)$, QE W, 5/12/95
Figure 16  Oblique $N(x,t)$, QE W, detectors 23 and 25, 5/12/95
Figure 17  Oblique $N(x,t)$ and $T(x,t)$, QE W, detector 25, 5/12/95
Figure 18  Oblique $N(x,t)$, QE W, Cawthra on-ramp, 5/12/95
Figure 19  Oblique merge ratio curve, QE W, 5/12/95
Figure 20  Oblique $N(x,t)$, QE W, detector 23, 5/12/95
Figure 21  QE W detector 23 total flow (30 second average), 5/12/95

LIST OF TABLES
Table 1  Gardiner Expressway Bottleneck Input Flows
Table 2  QE W Bottleneck Input Flows

Table 1 Gardiner Expressway Bottleneck Input Flows

<table>
<thead>
<tr>
<th>Date</th>
<th>Maximum Input Flow (vph)</th>
</tr>
</thead>
<tbody>
<tr>
<td>February 11, 1997</td>
<td>7,120</td>
</tr>
<tr>
<td>March 5, 1997</td>
<td>6,970</td>
</tr>
<tr>
<td>February 20, 1997</td>
<td>6,870</td>
</tr>
<tr>
<td>July 21, 1997</td>
<td>7,250</td>
</tr>
<tr>
<td>Mean</td>
<td>7,050</td>
</tr>
</tbody>
</table>

Table 2 QE W Bottleneck Input Flows

<table>
<thead>
<tr>
<th>Date</th>
<th>Maximum Input Flow (vph)</th>
</tr>
</thead>
<tbody>
<tr>
<td>May 12, 1995</td>
<td>7,710</td>
</tr>
<tr>
<td>May 3, 1995</td>
<td>7,818</td>
</tr>
<tr>
<td>May 1, 1995</td>
<td>7,324</td>
</tr>
<tr>
<td>Mean</td>
<td>7,620</td>
</tr>
</tbody>
</table>
Figure 1. Freeway sites, Toronto, Canada.

Figure 2. Loop detector validation, Gardiner Expressway, detector 80.
Figure 3. Oblique $N(x,t)$, Gardiner Expressway, Feb. 11, 1997.

Figure 4. Oblique $N(x,t)$, detectors 60 and 80, Gardiner Expressway, Feb. 11, 1997.
Figure 5. Oblique $N(x,t)$ and $T(x,t)$, detector 80, Gardiner Expressway, Feb. 11, 1997.

Figure 6. Oblique $N(x,t)$, Spadina Ave. on-ramp, Gardiner Expressway, Feb. 11, 1997.

Figure 7. Oblique merge ratio curve, Gardiner Expressway, Feb. 11, 1997.

Figure 8. Oblique $N(x,t)$, detector 60, Gardiner Expressway, Feb. 11, 1997.
Figure 9. Average flow, detector 60, Gardiner Expressway, Feb. 11, 1997.

Figure 10. Maximum flow step function, detector 60, Gardiner Expressway, Feb. 11, 1997.

Fig. 11. Oblique cumulative count difference between detectors 60 and 70, Gardiner Expressway, Feb. 11, 1997.
Figure 12. Net lane inflow for each lane between detectors 60 and 70, Gardiner Expressway, Feb. 11, 1997.

Figure 13. Count variance (detector 60) and cumulative variance, Gardiner Expressway, Feb. 11, 1997.
Figure 14. Oblique $V(x,t)$, detector 60, Gardiner Expressway, Feb. 11, 1997.

Figure 15. Oblique $N(x,t)$, QEW, 5/12/95
Figure 16. Oblique N(x,t), QEW, detectors 23 and 25, 5/12/95

Figure 17. Oblique N(x,t) and T(x,t), QEW, detector 25, 5/12/95

Figure 18. Oblique N(x,t), QEW, Cawthra on-ramp, 5/12/95
Figure 19. Oblique merge ratio curve, QEW, 5/12/95

Figure 20. Oblique N(x,t), QEW, detector 23, 5/12/95

Figure 21. QEW detector 23 total flow (30 second average), 5/12/95