# Toward the Systematic Diagnosis of Freeway Bottleneck Activation

Robert L. Bertini

*Abstract*—**An active freeway bottleneck is a point characterized by the presence of queued traffic immediately upstream and unqueued (freely-flowing) traffic immediately downstream. It is shown that a freeway bottleneck's activation can be diagnosed definitively using archived inductive loop detector data. Once the bottleneck's location and time of activation were determined, potential signals of its activation were explored. It is shown that certain potential signals are evident immediately before upstream queue formation. 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.**

***Index Terms*—intelligent systems, networks, time series, transportation.**

## I. Introduction

BOTTLENECKS are critical components of freeway systems and understanding their behavior is important for improving system operations. Fig. 1 shows that an "active" bottleneck arises when vehicles discharge from an upstream queue (guaranteeing that the bottleneck serves vehicles at a maximum rate) and vehicles are unimpeded by traffic conditions emanating from further downstream [1]. Bottlenecks can be difficult to detect since they are temporally and spatially variable. Detection can be further hampered by large spacing of freeway surveillance detectors (see Fig. 1).

In earlier studies, the author examined traffic conditions upstream and downstream of freeway bottlenecks located near busy on-ramps and several reproducible features were observed [2], [3], [4], [5]. For example, it was shown that the bottleneck was located more than 1 km (0.62 mi) downstream of the on-ramp, farther than had been found previously [6],

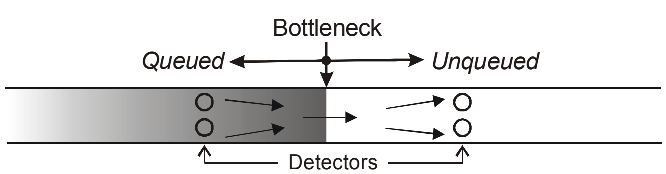


Fig. 1. Active bottleneck.

Manuscript received March 1, 2003. This work was supported in part by the U.S. Department of Transportation under the Eisenhower Graduate Fellowship Program and the University of California Transportation Center.

Robert L. Bertini is with the Department of Civil & Environmental Engineering, Portland State University, P.O. Box 751, Portland, OR 97207-0751 USA (phone: 503-725-4249; fax: 503-725-5950; e-mail: bertini@pdx.edu).

[7]. Also, on certain days, a high bottleneck flow of 7,000 vehicles per hour (vph) was measured (three lanes) for up to 40 minutes before a queue formed upstream of the bottleneck. This gave rise to a measurably (on the order of 10%) lower discharge rate that prevailed for a much longer period. This was in contrast with other studies that reported no such flow reductions upon bottleneck activation [6], [7], [8], [9].

On other days, it was shown that a queue emanating from somewhere further downstream spilled over into the merge area prior to the activation of its bottleneck. Upon dissipation of the downstream queue, the bottleneck consistently arose at the same location, more than 1 km (0.62 mi) downstream of the on-ramp. Further, vehicles discharged from the subject bottleneck without exhibiting a particularly high flow. On all days studied, while the bottleneck was active, the average discharge flows were observed to be nearly constant and exhibited only small deviations from one day to the next. This is in contrast with [10] that found noticeably different bottleneck discharge flow rates from day to day.

To promote the visual identification of time-dependent traffic features, these studies used cumulative curves of vehicle count and occupancy constructed from data measured at neighboring freeway detectors [11]. 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.

Cumulative curves were also used in the study presented here, which adds to previous findings by explicitly diagnosing the activation of a bottleneck located near a busy on-ramp and examining potential signals of its activation. Most notably, on the four days examined, beginning about two minutes prior to bottleneck activation, a particularly high flow was measured just upstream of the bottleneck location. The next section contains a brief description of the freeway site and the loop detector data used in this study.

## II. Data

The observations that follow were taken from the 3.3 km (2.1 mi) segment of the Gardiner Expressway in Toronto, Canada, illustrated in Fig. 2. 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. This freeway has no ramp metering, although the Jameson Ave. on-ramp is closed between 15:00

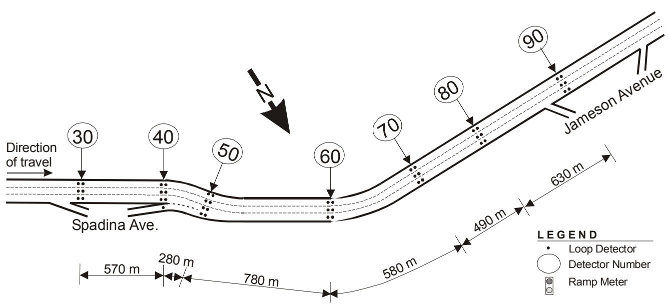


Fig. 2. Gardiner Expressway site, Toronto, Canada.

and 18:00 each afternoon. The next section describes the bottleneck's definitive diagnosis, including its location, time of activation and deactivation and some discharge features that were uncovered (for a single day). These observations were reproduced on three additional days.

## III. Diagnosing the Bottleneck in Detail

In order to pinpoint bottleneck activation and deactivation, Gardiner Expressway traffic features were analyzed using data from Feb. 11, 1997. Fig. 3 presents oblique curves of *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 shown in Fig. 4(a), the *N* 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 Fig. 3 began (*N*=0) relative to the passage of a hypothetical reference vehicle (curve 40 includes counts from the Spadina Ave. on-ramp) so that all curves describe the same collection of vehicles. Any horizontal and vertical separations between curves would have been the trip times and vehicle accumulations between detectors, respectively [12], [13].

Each curve in Fig. 3 was shifted horizontally to the right by the average free-flow trip time between the respective detector and downstream detector 80. Resulting vertical displacements between curves are the *excess* vehicle accumulation between detectors due to vehicular delays. In order to magnify the curves' features, as shown in Fig. 4(b), an oblique coordinate system was used where *N(x,t)* was reduced by $q_0(t')$ where $q_0$ was an oblique scaling rate and $t'$ was the elapsed time from the beginning of each curve. The same value of $q_0$ was used for all curves and therefore did not affect the vertical separations [11]. The use of this oblique coordinate system magnifies changes in vehicle arrival rate and the times at which these changes occurred. This procedure is also described in several references [11], [2].

All five curves in Fig. 3 were superimposed between 15:00 and 15:18, indicating the presence of freely flowing traffic throughout this freeway section. 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

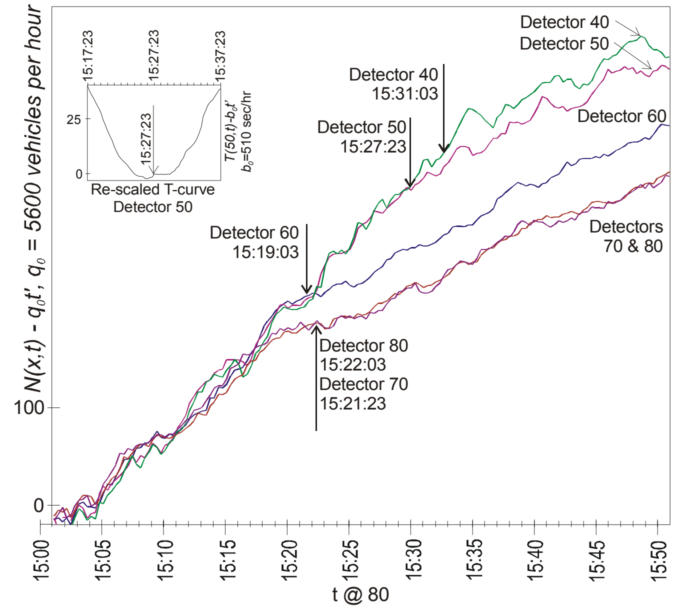


Fig. 3. Oblique N(x,t), Gardiner Expressway, Feb. 11, 1997.

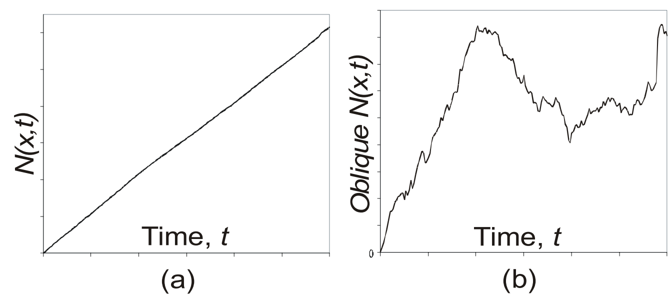


Fig. 4. Sample construction of cumulative curve of vehicle counts (a) and oblique cumulative curve (b).

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.

Fig. 3 also maps the propagation of the queue further upstream. As shown, a flow reduction at detector 50 was observed at 15:27:23, when the curve at detector 50 deviated from the detector 40 curve. 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 [14] for a discussion of the relation between occupancy and velocity). Oblique coordinates was 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. Fig. 3's inset 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, verifying the arrival of the queue at that time and location. Thus, Fig. 3 has made it possible to diagnose the bottleneck's location (between

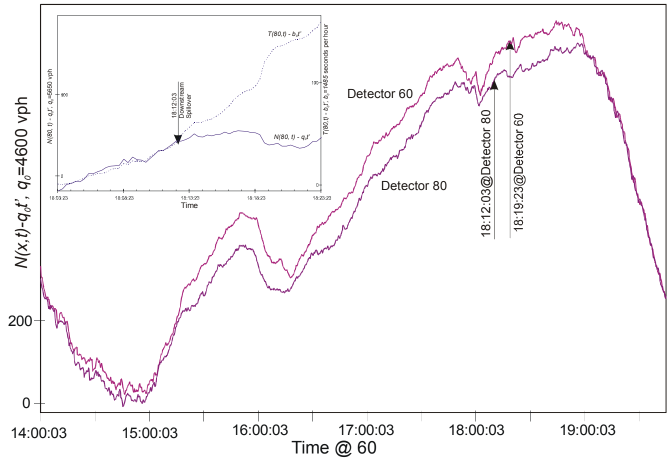


Fig. 5. Oblique *N(x,t)*, detectors 60 and 80, Gardiner Expressway, Feb. 11, 1997.

detectors 60 and 70) as well as the time it became active (15:19:03).

Further analysis showed that 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, Fig. 5 displays oblique *N(x,t)* for detectors 60 and 80 for a longer period. The queue's presence between detectors 60 and 80 is visible in Fig. 5 as the continued vertical displacement between the two curves. To trace the spillover, one can see from Fig. 5 that the flow reduction was also displayed by the curve at detector 60 shortly thereafter (at 18:19:23). Even after these flow reductions, the queue between detectors 60 and 80 persisted until 19:00.

The inset in Fig. 5 confirms that the backward-moving queue arrived at detector 80 at approximately 18:12:03. The figure 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. This confirms that the downstream queue arrived at detector 80 at about 18:12:03, deactivating the bottleneck at that time by restricting its flow. An oblique *T(x,t)* at detector 60 (not shown) confirmed that the downstream queue arrived at that detector at approximately 18:19:23.

Figs. 3 and 5 have verified the bottleneck's location, the time it became active, and the time that it was deactivated. Fig. 3 also clearly maps the passage of the backward-moving queue. Now it is possible to examine the active bottleneck's queue discharge features in detail. Cumulative curves from detector 80 (downstream of the bottleneck) were used to examine the bottleneck's discharge features. Fig. 6 shows oblique *N(80,t)* and *T(80,t)* which reveal no abrupt reductions s in the *N(x,t)* accompanied by reductions 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 anywhere further downstream.

Fig. 6 also shows that between 15:05:23 and 15:22:03 (the beginning of queue discharge) a flow of 6240 vph was measured. Upon queue discharge, 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 average discharge flow of 5790 vph (dashed line), which was 8% lower than the flow prior to bottleneck activation. The bottleneck discharge flow prevailed over a period of 3 hours 40 minutes. This figure confirms that queue discharge was accompanied by a drop in flow, apparently contradicting [6], [7], [8], and [9].

Fig. 7 shows an oblique *N(x,t)* constructed from counts measured on the Spadina on-ramp. The surge in ramp flow at 15:06:43 corresponds with the measured increase in upstream flow previously revealed in Figs. 3 and 6. 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 the on-ramp demand. These 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. Fig. 8 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

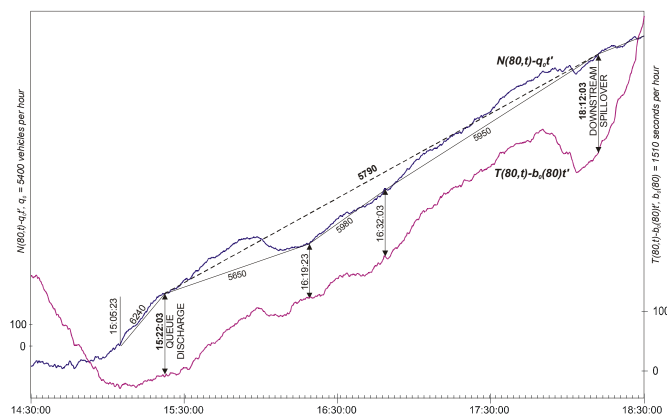


Fig. 6. Oblique *N(x,t)* and T(x,t), detector 80, Gardiner Expressway, Feb. 11, 1997.

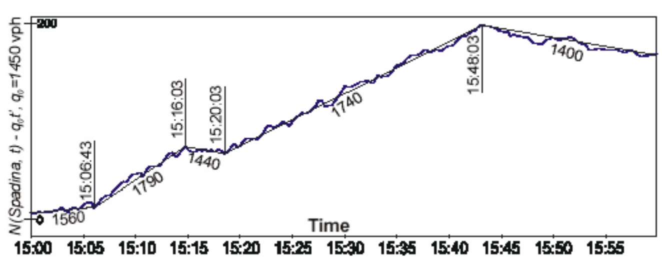


Fig. 7. Oblique *N(x,t)*, Spadina Ave. on-ramp, Gardiner Expressway, Feb. 11, 1997.

not exhibit the zipper effect [15] whereby freeway and ramp vehicles would have shared the shoulder lane in a strictly alternating fashion.

## IV. Bottleneck Input Flows on the Gardiner Expressway

After the bottleneck's location and active period were determined, the period exhibiting higher flows prior to queue formation were examined in detail. Toward a better understanding of how this high bottleneck input flow influenced and possibly signaled the queue formation, traffic features were studied in the individual lanes at detector 60, immediately upstream of the bottleneck.

Fig. 9 contains the single lane oblique *N(x,t)* constructed from data measured 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. Note that the times at which the queue arrived in each lane are also shown in the figure. The dashed line superimposed on the upper curve shows that the average 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 higher flows in the center and shoulder lanes. Then, at 15:10, about ten minutes prior to the arrival of the queue, flows in the center and shoulder lanes became approximately equalized. For the final two-minute period just before the queue formed (highlighted by the shaded vertical strip), a total flow of more than 7,100 vph was achieved. The achievement of 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 indicated in Fig. 10, which shows the total bottleneck input flow measured across all three lanes at detector 60 as a scatter plot. In this figure, the 20-second counts have been 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 before or after that time. Fig. 11 explores the concept of a "maximum" flow as measured over several measurement periods. The figure shows step functions recording the maximum flow measured at detector 60 (total across all lanes) throughout the period 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 large jump about two minutes before queue formation—thus a two-minute average may help signal impending queue formation.

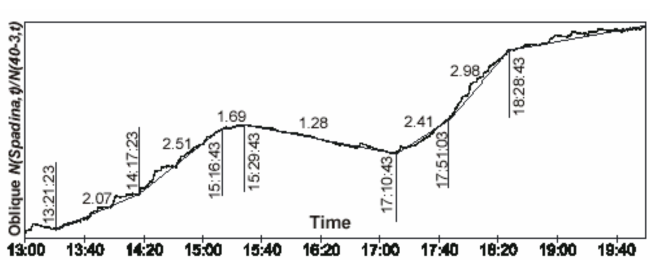


Fig. 8. Oblique merge ratio curve, Gardiner Expressway, Feb. 11, 1997.

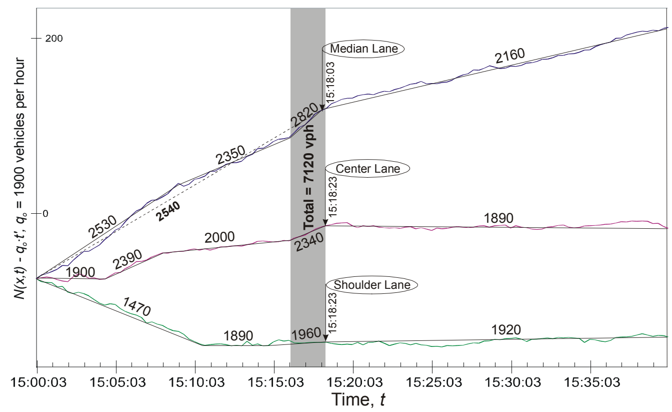


Fig. 9. Oblique *N(x,t)*, detector 60, Gardiner Expressway, Feb. 11, 1997.

TABLE I
Gardiner Expressway Bottleneck Input Flows

| Date | Maximum Input Flow (vph) |
|---|---|
| February 11, 1997 | 7,120 |
| March 5, 1997 | 6,970 |
| February 20, 1997 | 6,870 |
| July 21, 1997 | 7,250 |
| Mean | 7,050 |

Table 1 shows that a maximum two-minute flow was also achieved on three other 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.

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 searching for abrupt signals of excess vehicle accumulation. Fig. 3 revealed 15:18 as the time at which excess vehicle accumulation was visible between detectors 60 and 70 by showing that the cumulative count curves measured at the two detectors began to diverge at that time. Fig. 12 shows an alternative method of plotting the same data. The scatter plot (*N(70,t)-N(60,t)*) shows the difference between cumulative counts measured across all lanes at detectors 60 and 70 at 20-

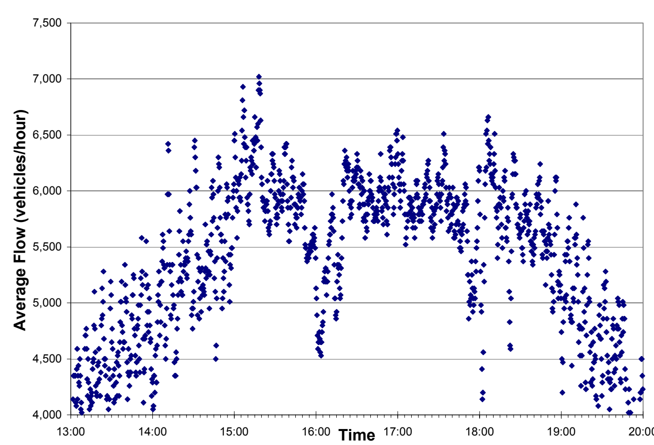


Fig. 10. Average flow, detector 60, Gardiner Expressway, Feb. 11, 1997.

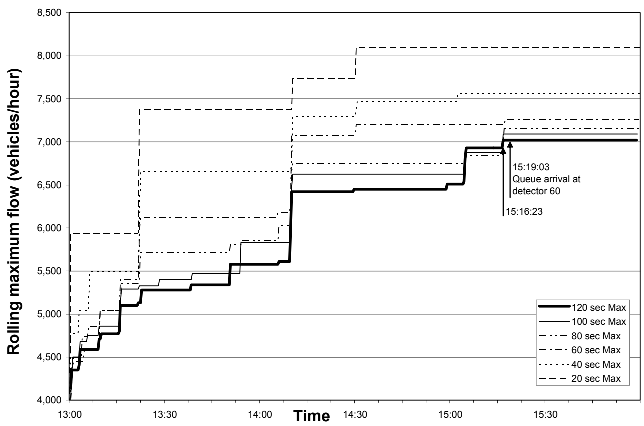


Fig. 11. Maximum flow step function, detector 60, Gardiner Expressway, Feb. 11, 1997.

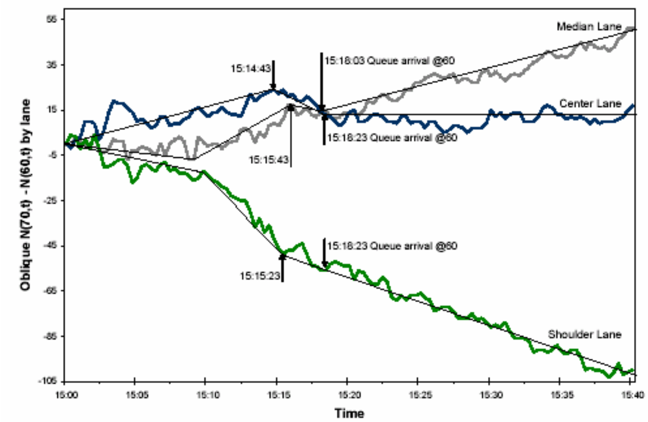


Fig. 13. Net lane inflow for each lane between detectors 60 and 70, Gardiner Expressway, Feb. 11, 1997.

sec intervals. The oblique cumulative curve of these data shows a sharp upward surge at 15:17:03, two minutes before the queue was shown to arrive at detector 60 (in Fig. 3). This would provide another way to diagnose potential excess vehicle accumulation upstream of a bottleneck's location.

In Fig. 9 as well as a previous study using data from the same site [3], it was shown that heavy net lane inflow rates resulted in very high flows in the median lane during the time just prior to bottleneck activation. In order to explore whether notable changes in net lane inflow are visible in the immediate vicinity of the bottleneck, Fig. 13 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, Fig. 14 shows a bar chart of the variance of the counts recorded at detector 60, measured over 2-minute periods. As shown, the count variance drops 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 drops at the beginning of the period of sustained high flow prior to bottleneck activation (15:05:23 as shown in Fig. 6) and then drops 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 future analysis.

## V. Other Signals of Queue Formation on the Gardiner Expressway

As further indicators of changing conditions in individual lanes just upstream of the bottleneck, Fig. 15 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 average speed and times marking speed changes. As shown, speed reductions were observed in each lane at the precise time that the queue arrived at detector 60.

As suggested by [16] in the context of incident detection, Fig. 16(a) shows oblique cumulative occupancy curves constructed from data measured across all lanes at detectors

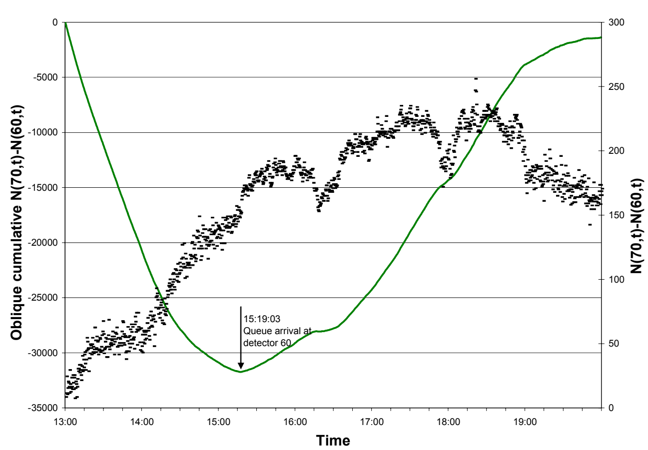


Fig. 12. Oblique cumulative count difference between detectors 60 and 70, Gardiner Expressway, Feb. 11, 1997.

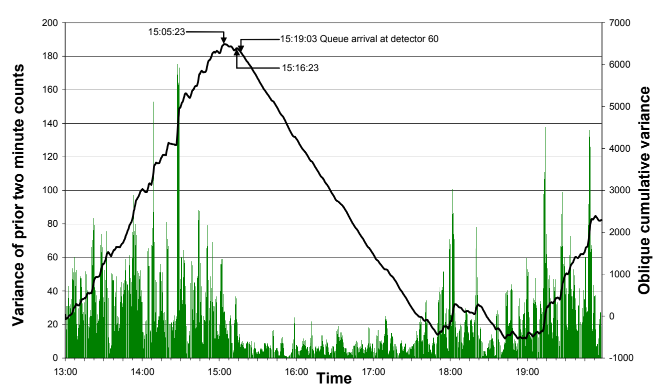


Fig. 14. Count variance (detectors 60) and cumulative variance, Gardiner Expressway, Feb. 11, 1997.

60 and 70. Upon bottleneck activation, it is clear that the curves diverge dramatically. It may be desirable to automate a search process for many locations on a freeway network. Toward this end, Fig. 16(b) shows a curve constructed from the cumulative difference between occupancy measured at detectors 60 and 70. This curve shows a sharp increase in the occupancy difference at the time (15:19:03) that the queue arrived at detector 60. It would be important to note that an automated procedure would need to be designed to differentiate between incidents (non-recurrent crashes, breakdowns and other random events) and recurrent bottleneck activation.

## VI. Final Comments

It is shown that it is possible to definitively diagnose a freeway bottleneck's location, activation and deactivation times and other features using loop detector data. Further, it is shown that an apparently reproducible signal of high bottleneck input flow was measured two minutes before bottleneck activation. It is possible that this signal would be helpful toward developing any system for attempting to prolong the high flow that prevailed prior to queue formation. 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 [17]. These signals are not presented as substitutes for the methodical approach described in Sec. III, 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 [16]. 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.

## Acknowledgment

Lisa Maasland and David Nesbitt, City of Toronto, generously provided the data used in this study. Appreciation is extended to Prof. Michael Cassidy, for his helpful guidance, and to the late Prof. G.F. Newell for his inspiration and encouragement. Discussions with Prof. Pravin Varaiya, also prompted some aspects of this paper.

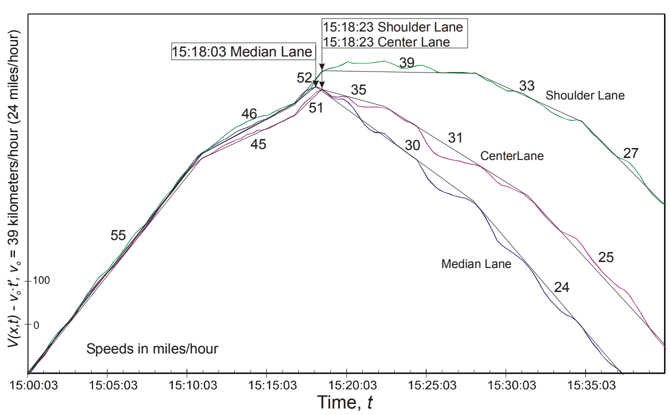


Fig. 15. Oblique $V(x,t)$, detector 60, Gardiner Expressway, Feb. 11, 1997.

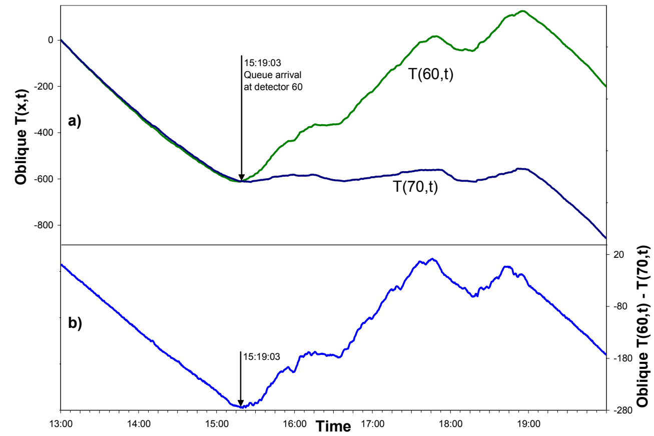


Fig. 16. Oblique cumulative occupancy from detectors 60 and 70 (a); oblique cumulative difference between T(70,t) and T(60,t), (b).

## References

[1] C.F. Daganzo, Fundamentals of transportation and traffic operations. New York: Elsevier, p. 133.

[2] M.J. Cassidy and R.L. Bertini, "Some traffic features at freeway bottlenecks," Transp. Res., vol. 33B, no. 1, pp. 25-42, Jan. 1999.

[3] M.J. Cassidy and R.L. Bertini, "Observations at a freeway bottleneck, Proc. 14th Int. Sym. Transportation and Traffic Theory, Jerusalem, Israel, 1999, pp. 107-124.

[4] R.L. Bertini, "Time-dependent traffic flow features at a freeway bottleneck downstream of a merge," Ph.D. dissertation, Dept. Civil & Envir. Eng., Univ. of California, Berkeley, CA, 1999.

[5] R.L. Bertini and M.J. Cassidy, "Some observed queue discharge features at a freeway bottleneck downstream of a merge," Transp. Res., vol. 36A, no. 8, pp. 683-697, Oct. 2002.

[6] B.N. Persaud, "Study of a freeway bottleneck to explore some unresolved traffic flow issues," Ph.D. dissertation, Dept. Civil Eng., University of Toronto, 1986.

[7] B.N. Persaud and V.F. Hurdle, "Freeway capacity: definition and measurement issues. Proc. Int. Sym. Highway Capacity, Karlsruhe, Germany, 1991, pp. 289-307.

[8] F.L. Hall and L.M. Hall, "Capacity & speed-flow analysis of the Queen Elizabeth Way in Ontario," Transp. Res. Rec. 1287, pp. 108-118, 1990.

[9] L. Newman, "Study of traffic capacity and delay at the merge of the No. Sacramento and Elvas freeways," California Div. of Highways, 1961.

[10] B.S. Kerner, "Theory of congested highway traffic: empirical features and methods of tracing and prediction," Proc. 15th Int. Sym. Transp. and Traffic Theory, Adelaide, Australia, 2002, pp. 417-439.

[11] M.J. Cassidy and J.R. Windover, "Methodology for assessing dynamics of freeway traffic flow," Transp. Res. Rec. 1484, pp. 73-79, 1995.

[12] G.F. Newell, Applications of queueing theory. London: Chapman Hall, 1982.

[13] G.F. Newell, "A simplified theory of kinematic waves in highway traffic I: General theory. II: Queueing at freeway bottlenecks. III: Multi-destination flows," Transp. Res., 27B, no. 4, pp. 281-313, Aug. 1993.

[14] M.J. Cassidy and B. Coifman, "The relation between average speed, flow and occupancy and the analogous relation between density and occupancy," Transp. Res. Rec. 1591, pp. 1-6, 1997.

[15] L. Newman, "Freeway operations analysis course notes," unpublished course notes, Univ. of California, Berkeley, CA, 1986.

[16] W.H. Lin and C.F. Daganzo, "A simple detection scheme for delay-inducing freeway incidents," Transp. Res., vol. 31A, no. 2, pp. 141-155, Mar. 1997.

[17] A. Skabardonis, T. Choe and P.P. Varaiya, "Freeway performance measurement systems (PEMS): operational analysis tool," Trans. Res. Board Preprint, 02-3679, 2002.