OBSERVATIONS AT A FREEWAY BOTTLENECK

Michael J. Cassidy and Robert L. Bertini
Department of Civil and Environmental Engineering and Institute of Transportation Studies,
University of California at Berkeley

ABSTRACT

Traffic was studied upstream and downstream of a bottleneck on a freeway in Toronto, Canada using transformed curves of cumulative vehicle count and cumulative occupancy. The bottleneck was located more than a kilometer downstream of a busy on-ramp. After diagnosing its location and the times that it remained active each day, the study focused on the traffic patterns that arose in each travel lane. It was observed that prior to the bottleneck’s activation, the vehicles’ lane-changing trends created extraordinarily high flows in the median (i.e., left-most) lane and that these high flows were sustained for extended durations. When a queue eventually formed at the bottleneck, its discharge rates were considerably lower than those flows measured prior to queueing. Within each lane, the queue discharge rates remained nearly constant over the rush and the average rates varied only slightly across days. Finally, vehicles arriving to the bottleneck from the nearby upstream on-ramp entered the freeway at high rates, even after the bottleneck’s queue propagated beyond this ramp.

1. INTRODUCTION

In an earlier study, the authors examined traffic at two bottlenecks on freeways in and near Toronto, Canada (Cassidy and Bertini, 1999). From this study, certain reproducible patterns were observed. For example, each bottleneck formed more than a kilometer downstream of an on-ramp and their formations occurred at these same locations each day. While the bottlenecks
were active, the vehicles discharged through them at nearly constant rates, although some time dependencies were observed for short periods following the onset of queueing. Furthermore, a bottleneck’s average queue discharge rate did not vary much from one day to the next and these average rates were typically 8 to 10 percent lower than the flows measured prior to queueing upstream.

These earlier findings came by visually comparing sets of transformed cumulative curves. Each curve was constructed from either the counts or the occupancies collected at one of several neighboring loop detector stations. Of note, these curves described measurements that were taken across all travel lanes, meaning that the above findings came by grouping together the traffic streams in multiple lanes and studying them in the aggregate.

In this paper, we add to the previous findings on bottleneck flow by reporting on some observations taken from individual lanes. At each of three neighboring detector stations located upstream and downstream of a bottleneck, curves of cumulative vehicle count and cumulative occupancy were separately constructed for each travel lane. Visual comparisons of these curves revealed certain details of traffic evolution, some of which were unexpected.

It was observed, for example, that large numbers of vehicles gradually moved into the median lane as they approached and passed through the bottleneck. This lane-changing pattern even continued at locations more than 2 kilometers downstream of the neighboring on-ramp. As a consequence, flows measured in the median lane were remarkably high; e.g., at a location well downstream of the on-ramp, the median lane flows sometimes exceeded 2,600 vehicles per hour. These high rates persisted each day for durations of up to 40 minutes before queues formed upstream and lower discharge rates ensued.

The bottleneck’s queue formed at nearly, but not exactly, the same times in each lane; i.e., this formation occurred in the shoulder lane several minutes after it had occurred in the adjacent lanes. Following the queue formations, the flow reductions observed downstream were most pronounced in the median and center lanes and less so in the shoulder lane. The discharge rates remained nearly constant so long as the bottleneck was active and free of any incidents nearby. Although these average rates varied across lanes, each lane’s average was reproduced from day to day.

It was further observed 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

---

1 The term “active bottleneck” denotes that the discharge rates measured downstream of a queue were not affected by traffic conditions from further downstream (Daganzo, 1997).
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 following section provides descriptions of the bottleneck and of the loop detector data used in this study. Section 3 describes the use of cumulative curves to identify the bottleneck’s location and the times that it remained active; uncovering these details was a requisite step for studying the evolution of the bottleneck flows. Section 4 presents the bottleneck’s traffic patterns that were observed in each lane during a single rush. Some findings from repeating these analyses with data from other days, and comments regarding future research directions, are provided in the fifth and final section.

2. The Data

The segment of the Gardiner Expressway shown in Figure 1 was the site used in our study. As an aside, this site has been used in several earlier studies of freeway capacity (Persaud, 1986; Persaud and Hurdle, 1991; Persaud, et al., 1998), including the previous one by the authors (Cassidy and Bertini, 1999). It is located in Toronto, Canada and meters are not deployed on its on-ramps (although the Jameson Avenue on-ramp is closed during a portion of each afternoon rush).

Figure 1: Gardiner Expressway, Toronto, Canada

The loop detector stations for measuring traffic data are labeled in Figure 1 as per the numbering strategy adopted by the City of Toronto. These detectors record counts, occupancies and (time) mean speeds in each lane over 20-second intervals. In total,
measurements were made during three weekday afternoons when the local weather bureau reported clear skies and no measurable precipitation.

In the next section, data from one of the observation days are used to demonstrate that a bottleneck was activated between detector stations 60 and 70. This bottleneck remained active for more than two hours before a queue spilled over from further downstream and restricted its discharge. We will also show that, some minutes prior to its deactivation, the bottleneck’s flow was impeded by an incident that occurred near detector 50.

3. The Active Bottleneck

Figure 2 presents transformed curves of cumulative count, \( N \), versus time, \( t \), for detectors 40 through 80. These were constructed using the vehicle counts taken in all lanes during a 30-minute period that spanned the onset of queueing.\(^2\) Untransformed \( N \)-curves give the cumulative number of vehicles to have passed (detector) location \( x \) by \( t \). By constructing the curves as linear interpolations through the cumulative counts measured every 20 seconds, each curve’s slopes would be the flows past \( x \) during the 20-second measurement intervals. Moreover, since the counts for each curve in Figure 2 started \((N = 0)\) with the passage of a reference vehicle, the horizontal and vertical separations between the curves would have been the trip times and the vehicle accumulations between detectors, respectively (Newell, 1982; Newell, 1993).

In Figure 2, however, each curve, along with its corresponding time axis, was shifted to the right by the average free-flow trip time between the respective detector and downstream detector 80. Consequently, the vertical separations in the curves are the excess vehicle accumulations between detectors due to vehicular delays. These shifts are advantageous because two superimposed curves indicate that traffic in the intervening segment was freely-flowing; every feature of an upstream \( N \)-curve is passed to its downstream neighbor a free-flow trip time later. In addition, Figure 2 shows only the differences between each curve of cumulative count and the line \( N = q_0 \cdot t' \), where \( q_0 \) is the rate used to re-scale the curves and \( t' \) is the elapsed time from the start of each curve. This is important because reducing the curves’ cumulative count by a background flow \( q_0 \) magnifies details, such as the time dependencies in the flows, without changing the excess accumulations (Cassidy and Windover, 1995).

\(^2\) The on-ramp counts at Spadina Avenue were also used to construct the curve for detector 40 so that all of the curves in Figure 2 describe the same collection of vehicles. Conversely, a curve for detector 30 was not included in Figure 2, since the vehicles measured at this location were not identical to those measured at the detector stations further downstream.
The (nearly) superimposed curve portions in Figure 2 indicate that traffic was initially in free flow and remained in free flow between detectors 70 and 80. The reader may use a straightedge to confirm that curves 40, 50 and 60 exhibit increased slopes sometime shortly after 15:40, as delimited by the large arrow in Figure 2, and this caused these three curves to

![Figure 2: Transformed N-curves, Detectors 40 through 80.](image-url)
diverge from their (two) downstream counterparts. These curve features indicate that a bottleneck was activated between detectors 60 and 70 when increased flows arrived from upstream. The subsequent separation between curves 60 and 50 (at about 15:51:23) reveal when the queue arrived at detector 60. Likewise, the queue’s arrival at detector 50 is made evident by the divergence in curves 50 and 40 (at about 15:53:03).

In short, the transformed $N$-curves in Figure 2 conclusively diagnose the bottleneck’s location by showing that excess vehicle accumulations occurred upstream of detector 70 while free-flow conditions prevailed immediately downstream. Especially notable are the pronounced flow reductions (i.e., the reduced slopes of the $N$-curves) that followed the queue’s formation; further details on this are provided in the next section.

Figure 3 reveals the approximate time that the backward-moving queue arrived at the (freeway) detectors at station 40. Presented in this figure is a re-scaled $N$-curve for the freeway lanes at this detector station along with a re-scaled curve of cumulative occupancy versus time, or a $T$-curve, where cumulative occupancy is the total vehicular trip time over the detectors by time $t$ (Lin and Daganzo, 1997). Again for the purpose of magnifying details, the $T$-curve shown here is the difference between the cumulative occupancy actually measured in the three freeway lanes at detector 40 and the line $T = b_0 \cdot t'$, where $b_0$ is the background occupancy rate used to
re-scale the curve and \( t' \) is the elapsed time from the curve’s start. 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. These features reveal the arrival of the queue at station 40. That this arrival occurred some time around 15:54 will be an important part of later discussion regarding the observed time dependencies in the on-ramp flows from Spadina Avenue.

Finally, Figures 4a and 4b reveal the period that the bottleneck remained active. Figure 4a shows transformed \( N \)-curves for detectors 60 and 80; these were constructed in the manner previously described, but for an extended period of over 4 hours. The slope of curve 80 drops noticeably some time around 18:00 and a similar reduction is displayed by curve 60 soon thereafter. Notably, the queue between detectors 60 and 80 persisted, even after these flow reductions; this is evident from the continued displacement between the two curves. These curve features indicate that a queue from further downstream arrived at these detector stations and thereby deactivated our bottleneck. In Figure 4b, the divergence in the re-scaled curves of \( N \) and \( T \) reveal that the queue from downstream arrived at station 80 at about 18:12:03.

Before concluding this section, it is worth re-emphasizing that the curves in Figures 2 and 4a were instrumental in identifying the bottleneck’s location and the period that it remained active. These curves derived their value, in part, by displaying the excess vehicle accumulations that arose between detectors and this required that the curves be constructed from the counts taken over all travel lanes.\(^3\) Having now identified these bottleneck details (i.e., its location and the time it was active), traffic patterns could be studied in the individual lanes at locations upstream and downstream of the bottleneck. Some notable findings from this study are presented next.

4. SOME OBSERVATIONS IN INDIVIDUAL LANES

Figure 5 presents re-scaled \( N \)-curves in the shoulder lane for detectors 60, 70 and 80. Figures 6 and 7 present the re-scaled curves for these same detector stations in the center and median lanes, respectively. These three detector stations (i.e., 60, 70 and 80) were selected because they were situated immediately upstream and downstream of the bottleneck. In each figure, the curves span a period of more than 3 hours, which includes the time that the bottleneck was active.

Before presenting some of the findings obtained by jointly examining Figures 5, 6 and 7, a brief explanation of their annotations is warranted. First, piece-wise linear approximations

\(^3\) If a set of \( N \)-curves do not describe node conservation, their vertical displacements would not be the excess accumulations (Newell, 1982; Newell, 1993; Cassidy and Windover, 1995).
Figure 4a: Transformed $N$-curves, Detectors 60 and 80.

Figure 4b: Re-scaled $N$- and $T$-curves, Detector 80.
were superimposed on the $N$ to highlight periods of nearly constant flow; the $N$ usually deviated from its corresponding linear approximation by no more than about 10 vehicles. The start and end times for each period of near-constant flow were selected “by eye” and re-scaled $T$-curves (not shown here) were also used to aid in these delimitations, since (sizable) changes in flow are accompanied by changes in occupancy. The times marking these flow changes are labeled on each curve.\textsuperscript{4} Further, the times marking the onset and the termination of queue discharge flows are noted in boldface type for the curves at downstream stations 70 and 80. Also shown are the rates corresponding to each period of near-constant flow; the numbers shown without parentheses are in units of vehicles per hour (vph) and those in parentheses are the corresponding average counts per minute. Finally, dotted lines are used to highlight the average queue discharge rates measured over the rush at downstream detectors 70 and 80. Labels specifying these average discharge flows are shown on the figures in boldface type.

We now turn our attention to the traffic patterns revealed by Figures 5 through 7. From even cursory examination of these figures, one observes a trend in vehicle lane changing; namely, that large numbers of vehicles changed lanes to the left while they traveled between detectors 60 and 80. The curves in Figure 5, for example, were each started with the same value of $N$, but curve 60, $N(60, t)$, lies well above curve 70 for all time $t$. In similar fashion, curve 70 rises above curve 80, indicating that vehicles continued to exit the shoulder lane at locations well downstream of the Spadina Avenue on-ramp.

Figure 6, on the other hand, shows no such obvious trend. Rather, the (net) flows in the center lane remained nearly unchanged as traffic moved through the bottleneck, indicating that, between detectors, the number that moved into the center lane nearly equaled the number that moved out. In fact, the Figure 6 curves have been vertically displaced (by arbitrary distances) because not separating these curves would have made it difficult to view their details.

Figure 7 shows that large numbers entered the median lane between detectors 60 and 80 and that this trend gave rise to some extraordinarily high flows. For example, a flow in excess of 2,600 vph was measured in the median lane at detector 80 prior to the bottleneck’s activation. Remarkably, this very high rate was observed for over 40 minutes before the queue formed upstream and a lower discharge rate ensued. (The queue’s formation is signaled by the onset of its discharge flow at detector 70).

The onset of queueing occurred at nearly, but not precisely, the same times in each lane; i.e., the queue appears to have formed in the shoulder lane several minutes after it formed in the adjacent lanes. The curves at detector 60 (in all three lanes) exhibit sustained surges some

\textsuperscript{4} A few of these times are not labeled in Figure 5 so that the figure would not become cluttered.
Figure 5: Re-scaled N-curves, Shoulder Lane.
Figure 6: Re-scaled $N$-curves, Center Lane.
Figure 7: Re-scaled N-curves, Median Lane.
minutes prior to the queue’s formation. Conversely, the onset of this queue was accompanied by rather dramatic flow reductions. Figures 5 through 7 show 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 rush. In the center and median lanes (Figures 6 and 7), these so-called flow collapses (Cassidy and Windover, 1995) prevailed for just under 20 minutes before being replaced by higher discharge flows. In the shoulder lane (Figure 5), the collapse persisted for less than 10 minutes.

Also of note, Figures 5 through 7 show that sizable reductions in the discharge flows were measured in all lanes some minutes before the bottleneck was deactivated by the arrival of the queue from downstream (recall that this queue arrived at detector 80 at about 18:12). Visual inspection of N- and T-curves measured in individual lanes revealed that these flow drops were caused by an incident (perhaps a vehicle stall or a small collision) that occurred in the shoulder lane near detector station 50. Notably, detector 50 measured near-zero counts and occupancies in the shoulder lane from 18:02:23 to 18:08:23. Figure 8a shows that during this period, the shoulder lane traffic at upstream station 40 exhibited a sharp reduction in the flow coupled with a rise in the occupancy, features which mark the passage of a queue. Conversely, Figure 8b shows that, within this period, shoulder lane traffic at downstream station 60 exhibited sudden reductions in both the flow and the occupancy. These features would be expected to occur downstream of a sudden restriction (i.e., as traffic passed the obstruction and moved into the shoulder lane). A reduction in the discharge flow also occurred in the shoulder lane at about 16:59 (see Figure 5), but this reduction was short-lived and of no real consequence.

Despite the flow variations that occurred during the bottleneck’s active period, Figures 5, 6 and 7 reveal that the queue discharge rates never deviated much from the linear trends shown with the dotted lines (the reduced discharge flows that accompanied the incident near detector 50 were excluded from consideration here). Thus, for each lane, the discharge rates can be described as being nearly constant over the rush. By comparing the average rates that correspond to each dotted line with the flows that prevailed prior to the bottleneck’s activation, it is clear that queuing was accompanied by long-run flow reductions, especially in the median and center lanes. Also by examining these three figures collectively, it is evident that the average queue discharge rates varied across lanes.

Finally, Figure 9 shows a re-scaled N-curve constructed solely from the counts on the Spadina Avenue on-ramp (at 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
Figure 8a: Re-scaled N- and T-curves, Detector 40, Shoulder Lane.

Figure 8b: Re-scaled N- and T-curves, Detector 60, Shoulder Lane.
Likewise, the sharp reduction in this ramp flow at 15:50:23 corresponds closely to the arrival of the queue from our active bottleneck; recall that this queue was shown (in Figure 3) to have arrived to the freeway detectors at station 40 some time shortly after 15:50. Although this queue apparently suppressed the on-ramp flow, vehicles continued to enter the freeway via this ramp at a high rate; i.e., 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 so-called “zipper effect” (Newman, 1986) whereby freeway and ramp vehicles share the shoulder lane in a strictly alternating fashion.

Study of the N-curves constructed from freeway counts at detector stations 30 and 40 (not shown here) revealed these increased flows observed in Figure 2 were part of sustained surges in the freeway flows as well as in the flow from the Spadina Avenue on-ramp.
5. **Findings From Repeated Experiments and Future Research Directions**

Data from detectors 40 through 80 were extracted during two other weekday afternoons and were examined in the manner previously described. On each of these two additional days, the observed traffic patterns were similar to those presented above. As a means of exemplifying some of these day to day similarities, Table 1 presents certain observations taken each day at detector 80. Row 1 of this table shows that, in the median lane, very high flows were observed each day prior to queueing. The table’s second row reveals that these high rates were always sustained for periods of at least 5 minutes, and in two instances, for much longer. Row 3 of Table 1 shows that the flow collapse at the onset of queueing was a reproducible feature in the median lane (note that the rates shown in this row are lower than their corresponding average discharge rates measured over the rush and presented in row 5). However, these collapses persisted for durations that varied across days, as shown in row 4 of the table. The flow collapse was likewise reproduced each day in the center lane, although this information is excluded from the table.

Important features also not shown in Table 1 are that the bottleneck always formed at the same location (i.e., between detectors 60 and 70) and that it was always activated by a sustained surge in the flow from upstream. Thus, traffic transitioned from free flow to queued conditions in a predictable way; the queues formed at an inhomogeneity, the bottleneck, due to reproducible, exogenous reasons, i.e., the increased flows. In this instance, one might presume that the freeway’s horizontal curve is the inhomogeneity creating the bottleneck (see Figure 1). While this may indeed be the case, it is worth noting that the same analysis methods applied to data from another freeway location found that a bottleneck consistently formed more than a kilometer downstream of an on-ramp, even though there was no obvious inhomogeneity at this location (Cassidy and Bertini, 1999). In any event, our studies to date have revealed no evidence suggesting that traffic can break down and form queues in a spontaneous manner.

Also of note, rows 5, 6 and 7 of Table 1 indicate that, while the bottleneck was active, the average discharge flow (in a given lane) exhibited only small variation across days. On each day, these rates can be described as “near-constant” since the cumulative counts never deviated much from a linear trend. Given these predictable features of its discharge rates, it seems reasonable to postulate about how queues might evolve upstream of this bottleneck; e.g., by using a continuum model of highway traffic (Lighthill and Whitham, 1955; Richards, 1956; Newell, 1993).
Table 1
Some Observations Taken From Station 80

<table>
<thead>
<tr>
<th>Row</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3/5/97</td>
<td>2/20/97</td>
<td>7/21/97</td>
</tr>
<tr>
<td>2</td>
<td>2630</td>
<td>2630</td>
<td>2400</td>
</tr>
<tr>
<td>3</td>
<td>43:40</td>
<td>5:40</td>
<td>19:40</td>
</tr>
<tr>
<td>4</td>
<td>2280</td>
<td>2110</td>
<td>2300</td>
</tr>
<tr>
<td>5</td>
<td>19:20</td>
<td>7:00</td>
<td>5:20</td>
</tr>
<tr>
<td>6</td>
<td>2340</td>
<td>2290</td>
<td>2330</td>
</tr>
<tr>
<td>7</td>
<td>1920</td>
<td>1910</td>
<td>1950</td>
</tr>
<tr>
<td>8</td>
<td>1720</td>
<td>1690</td>
<td>1690</td>
</tr>
</tbody>
</table>

Note: Data from Day 1 were used in Figures 2 through 9.

In closing, we note that the observations reported above bring to light a number of unanswered questions. For example, the reason(s) why the observed lane-changing trends persisted well downstream of the on-ramp, and the extent to which the high flows observed in the median lane might be reproduced at other bottlenecks, are unknown. Also unknown are the causes of the flow reductions that accompanied queueing, especially the relatively large reductions at the onset of queueing. Finally, the potential for using control measures, such as ramp metering, to extend the periods marked by high flows (observed prior to queueing, for example) is uncertain.

Answers to the above will only come through additional empirical study. Since (freeway) bottlenecks come in many forms, including merges, diverges, weaves and lane reductions, and since the traffic patterns on each type of bottleneck may exhibit their own peculiarities, the study of bottlenecks in each of their forms seems warranted. Cumulative curves like those described here might be used to conduct these studies since they provide a robust way of diagnosing the details of bottleneck traffic.

ACKNOWLEDGEMENTS

The authors are indebted to Mr. David Nesbitt, City of Toronto, for providing the data used in this study, and to G.F. Newell for his helpful comments.
REFERENCES


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


