A Probabilistic Approach to Defining Freeway Capacity and Breakdown

MATT LORENZ
LILY ELEFTERIADOU
The Pennsylvania Transportation Institute
The Pennsylvania State University
201 Transportation Research Building
University Park, PA 16802 USA

ABSTRACT

This paper addresses the need for an enhanced freeway capacity definition that incorporates the probabilistic nature of the freeway breakdown process. It consists of an extensive analysis of speed and volume data collected at two freeway bottleneck sites in Toronto, Canada. At each site, the freeway breakdown process was examined in detail for over 40 congestion events occurring during the course of nearly 20 days. The paper develops preliminary models for each site describing the probability of breakdown versus observed flow rate and examines the implications that this probabilistic approach to breakdown has on the current definition of freeway capacity. A revised, probabilistic freeway capacity definition is proposed for use in future editions of the *Highway Capacity Manual*.

1. INTRODUCTION

From a transportation perspective, the term “capacity” is essentially a description of the limit of the vehicle-carrying ability of a roadway. The concept of capacity for freeways (and its associated numerical value) plays an important role within the transportation profession, since it is applied in the planning, design, and operation states of virtually every freeway facility in the United States. However, determining just what capacity is—and quantifying it—has occupied transportation researchers for decades.

The definition of freeway capacity and the numerical value associated with it have evolved over time. The current published version of the *Highway Capacity Manual* (1997) defines freeway capacity as “the maximum sustained 15-min rate of flow, expressed in passenger cars per hour per lane (pcphpl), that can be accommodated by a uniform freeway segment under prevailing traffic and roadway conditions in a specified direction.”

Implied in the current definition and understanding of freeway capacity is the notion that the facility will “break down” (transition from an uncongested state to a congested state) when demand exceeds a specified capacity value. Thus, freeway capacity is currently viewed as a deterministic phenomenon. However, several recent capacity-related research efforts have studied the breakdown phenomenon for freeways more closely—suggesting specific causes, examining its characteristics, and speculating on its relationship to capacity. It hasbecome evident that breakdown does not necessarily occur at maximum flow and breakdown can occur at flows lower or higher than those traditionally accepted as capacity. As a result, recent studies have suggested that breakdown is a stochastic event, whereby probabilities of breakdown are associated with specific vehicular flows. However, analyses supporting this claim are limited.

If freeway capacity continues to be defined with respect to breakdown conditions, then the definition of capacity should be modified in light of this probabilistic view of freeway breakdown. This paper addresses the need for a revised freeway capacity definition that incorporates the probabilistic nature of the breakdown process. Through an extensive analysis of speed and flow rate data collected at two freeway bottleneck locations in Toronto, Canada, the paper investigates the definitions of freeway capacity and breakdown and their relationship while taking into account the stochastic nature of traffic flow. In addition, this paper examines the implications that a probabilistic approach to breakdown has on the current definition of freeway capacity and develops a suggested definition for use in future editions of the *HCM*.

2. DESCRIPTION OF STUDY FACILITY AND SITES

Two capacity-constrained sites along the Highway 401 freeway system in Toronto, Canada, were selected for in-depth analysis. For the purposes of this report, the two sites are denoted as Site “A” and Site “B.” Highway 401 was selected as the study facility since detailed speed and volume data from its numerous detector stations are archived and made available through the Ministry of Transportation of Ontario and additional data for each site was available from McMaster University. The sites were selected for study since they are both bottleneck locations and instrumented with detector stations within the bottleneck capable of recording speeds and volumes.

2.1 Description of Study Facility

Highway 401, or the “401 freeway” as it is commonly referred to, is a primary freeway facility located north of the downtown Toronto city center, and serves as a major east-west corridor for local, regional, and inter-province travel. Lanes along many portions of the 401 freeway in the vicinity of the two study sites are divided into groupings for “collector” and “express” travel. Three major interchanges (Highway 427, Highway 400-Black Creek Drive, and Allen Road) are located in the vicinity of the two study sites. Roadway cross-sections vary from between two and five lanes for both “collector” and “express” lane groups.

The posted speed limit on the 401 freeway is 100 km/h (approximately 62 mph). Free flow speeds during off-peak time periods were found to range between 100 km/h and 120 km/h (approximately 62 mph and 75 mph). Traffic volumes on the 401 freeway system are generally heavy during weekday morning and afternoon peak periods, and breakdowns in traffic flow are typical during these peak times. It is not uncommon for congested conditions to persist for several hours.

2.2 Description of Study Sites

Figure 1 illustrates the lane geometry at Site “A,” located in the vicinity of the Highway 427/Highway 401 interchange. At this site, traffic traveling northbound-to-eastbound in three ramp lanes on 427 merges with express traffic traveling in three eastbound mainline lanes on 401. The left ramp lane from 427 continues as an additional (fourth) through lane on the 401 mainline. The middle and right ramp lanes merge to form a fifth mainline lane on Highway 401. A detector station (denoted in this report as "Station 1") is located approximately 550 meters downstream of the ramp gore point.

Figure 2 illustrates the lane geometry at Site "B," located in the vicinity of the Highway 400-Black Creek Drive/Highway 401 interchange. Here, two "collector" ramp lanes merge with three "express" lanes on the westbound mainline. Both ramp lanes are dropped in succession, resulting in a freeway cross-section transitioning from five to three lanes. One detector station (denoted in this report as "Station 2") is located approximately 1,250 meters downstream of the ramp gore point.

At both stations, paired detectors are located in the each of the travel lanes and are instrumented to provide vehicle counts and speed estimates continuously at 20-second intervals. The driver population in the vicinity typically consists of urban commuters and other drivers familiar with the area's transportation system. Breakdowns in traffic flow at both Site "A" and Site "B" are common during weekday peak periods.

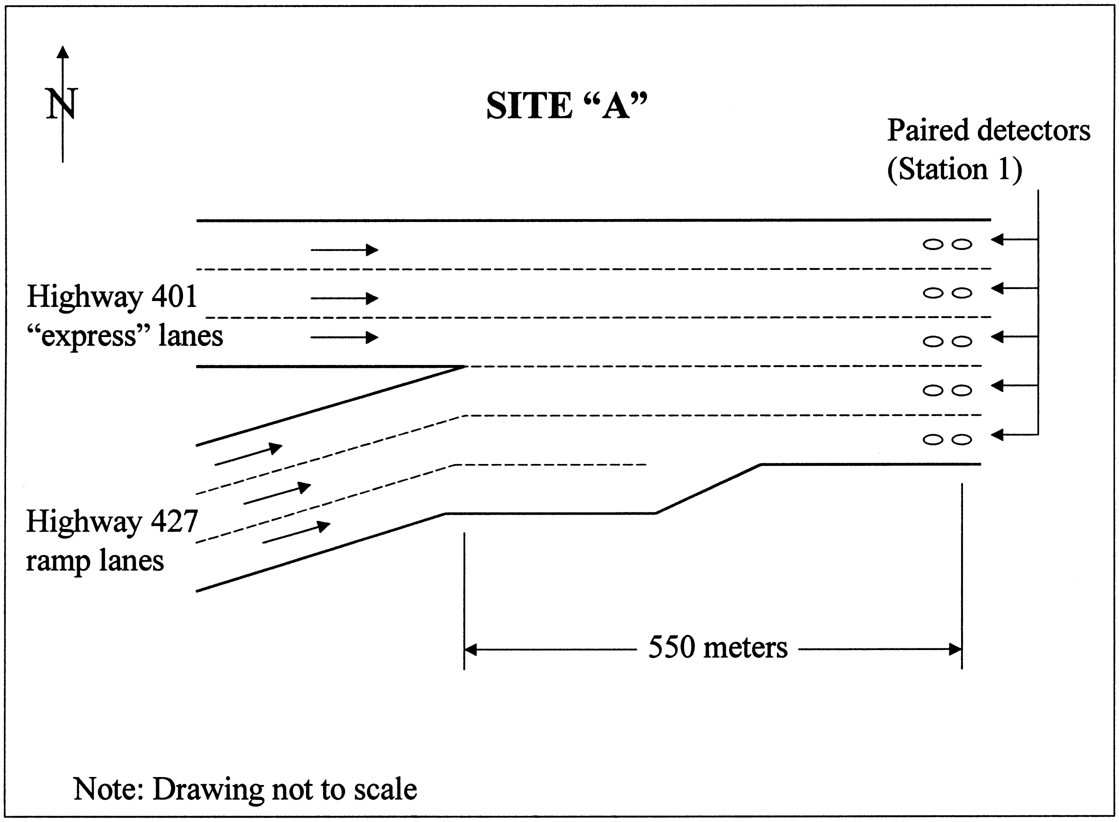

FIGURE 1 Schematic of Site "A": existing lane configuration.

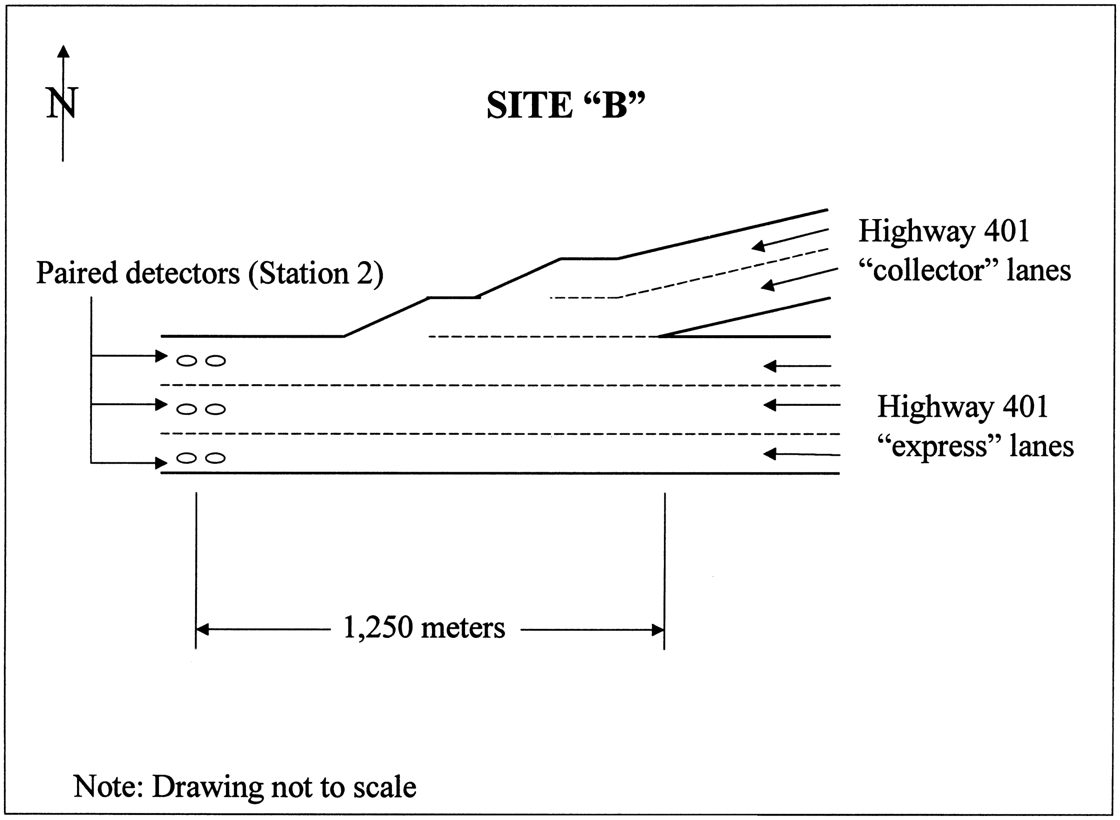

FIGURE 2 Schematic of Site "B": existing lane configuration.

3. DATA COLLECTION AND ANALYSIS

3.1 Introduction

The analysis consisted of a detailed examination of data collected during breakdown processes at the two freeway bottleneck sites described in the previous section. Each site was analyzed independently. At each site, speeds and flow rates for over 40 breakdown events were analyzed and used to generate models for probability of breakdown versus flow rate. For any given flow rate observed at one of the sites, a comparison was made between the frequency of breakdown at that flow rate and the frequency of the same rates that did not result in breakdown.

3.2 Data Collection and Summary

Historical freeway detector data for Sites "A" and "B" (detector Stations "1" and "2," respectively) were obtained from McMaster University and the Ministry of Transportation of Ontario. Each detector station recorded average vehicle speed (in km/h) and vehicle counts at both the upstream and downstream detector in the pair. These data were recorded for each lane at 20-second intervals over the duration of the data sampling period. The sampling periods ranged from 8 to 24 hours per day. Seventeen days of data were obtained at Site "A" and 20 days of data were obtained at Site "B."

The data were reviewed in detail for erroneous detector readings and summarized using a spreadsheet. In addition, the speed and flow rate data were cross-checked with data at adjacent detector stations to ensure that all subsequent analyses reflected site-imposed capacity constraints, as opposed to merely queue spillback resulting from downstream congestion.

3.3 Analysis

3.3.1 Time-series speed plots

Each of the study sites was analyzed independently. At each site, the speed and vehicle count data were summarized in 20-second intervals for each individual travel lane and over all lanes. The vehicle count data were expressed as equivalent hourly flow rates and the average speed across all lanes was determined using the volume-weighted average speed of all vehicles crossing the detector station. The speed and flow rate data were then plotted in time-series over each sample period. In these plots, time was displayed on the "x"-axis and speed on the "y"-axis. Figure 3 illustrates a representative time-series speed plot for a data sample collected on March 17, 1998 at Site "B." The points shown in Figure 3 represent the volume-weighted average of speeds in all lanes.

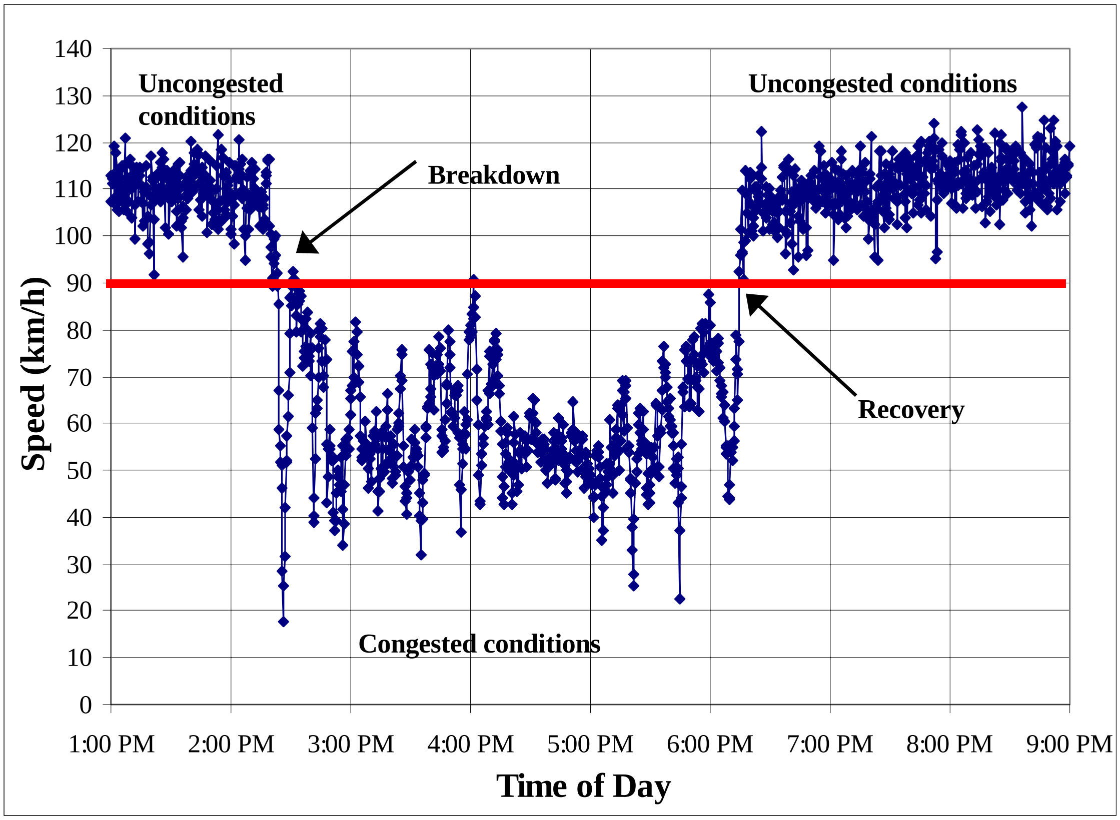

FIGURE 3 Time-series speed plot: Site "B" — March 17, 1998

The period of breakdown is easily identifiable in Figure 3. Prior to about 2:15 p.m., the average travel speed across all lanes is relatively high, fluctuating between approximately

100 km/h and 120 km/h. However, at approximately 2:15 p.m., the average travel speed across all lanes drops sharply to below 90 km/h (56 mph) and generally remains well below the 90 km/h threshold until about 6:15 p.m. At this time, speeds rise back to their pre-congestion levels.

Examining the time-series speed plot shown in Figure 3, it is evident that a speed “boundary” or “threshold” of approximately 90 km/h exists between the congested and uncongested regions. When the freeway operates in an uncongested state, average speeds across all lanes generally remain above the 90 km/h threshold at all times. Conversely, during congested conditions, average speeds rarely exceeded 90 km/h, and even then were not maintained for any substantial length of time. This 90 km/h threshold was observed to occur at both study sites and in all of the daily data samples evaluated as part of this research. It should also be noted that this threshold is a close approximation of the 53 mph (85 km/h) speed threshold for level-of-service “F” denoted in Chapter 3 of the 1997 *Highway Capacity Manual.*

3.3.2 Definition of “breakdown” and “recovery”

Similar time-series plots of speeds and flow rates were examined for each of the daily data samples. Small disturbances, similar to the initial one in Lane 1 described above, were observed to occur at various times, but did not always result in a breakdown in the traffic stream. Often, these disturbances were simply absorbed by the stream. Since the traffic stream was observed to recover from small disturbances in most cases, only those disturbances that caused the average speed over all lanes to drop below 90 km/h for a period of five minutes or more (15 consecutive 20-second intervals) were considered a true breakdown. The same criterion was used for “recovery periods”—those periods when average speeds in all lanes returned to over 90 km/h. A period of increased speeds was not considered a true recovery period unless speeds over 90 km/h were maintained for more than five minutes.

In the majority of the daily data samples, this five-minute criterion was not needed to distinguish congestion periods from recovery periods. For the most part, speed drops in the traffic stream were observed to last either a very short time (less than one or two minutes, with no residual effects) or a very long time (complete breakdowns of the freeway facility lasting more than an hour). Similarly, intervals during congested periods where speeds increased to over 90 km/h were rare and usually did not last very long (less than one or two minutes). In a few borderline cases, the five-minute criterion was needed and applied. For example, brief periods of congestion (five or six minutes in length) were sometimes followed by a brief period of recovery (another five or six minutes in length), and then by a second, sustained period of congestion. Applying the five-minute criterion, this pattern of congestion and recovery was identified as two separate breakdown events, although it could be argued that this pattern constitutes a single congestion event. Nonetheless, for this analysis the five-minute criterion remains.

3.3.3 Definition of “breakdown flow rate”

The “breakdown flow rate” is defined here as the flow rate (expressed as a per-lane, equivalent hourly rate) observed immediately prior to breakdown. However, a breakdown flow rate value determined for a single breakdown event at a particular site merely represents one of many possible breakdown flow rate values for that site. Additional flow rate values associated with other breakdown events at that site (on the same day and on other days) need to be determined as well. Furthermore, the breakdown flow rate values need to be compared to equivalent flow rates observed at times when breakdown did *not* occur. Therefore, for each site, the model describing the probability of breakdown is determined by comparing:

1. the frequency of breakdown at a given flow rate, with
2. the number of times breakdown did not occur at that given flow rate,

and plotting the results over the range of flow rates observed at the site. Assuming a typical freeway bottleneck breaks down only once or twice on a given weekday (usually during the morning or evening peak period, or both), it is apparent that a large quantity of data—spanning many breakdown periods—is required to estimate the probability of breakdown for any bottleneck site. For this paper, over 40 breakdown events were investigated at each site.

3.3.4 Analysis intervals for breakdown flow rates

Breakdown flow rates summarized at 20-second, one-minute, two-minute, five-minute and 15-minute intervals were considered for this analysis. The selection of any particular interval necessitates a trade-off between smoothing the data (when using a "long" interval) and capturing short-term fluctuations (when using a "short" interval). The one-minute, five-minute, and 15-minute intervals were all selected for further analysis here since they provide interesting comparisons over the range, and also provide balance between the smallest interval (20-seconds) and the largest interval typically investigated in capacity analyses (15-minutes).

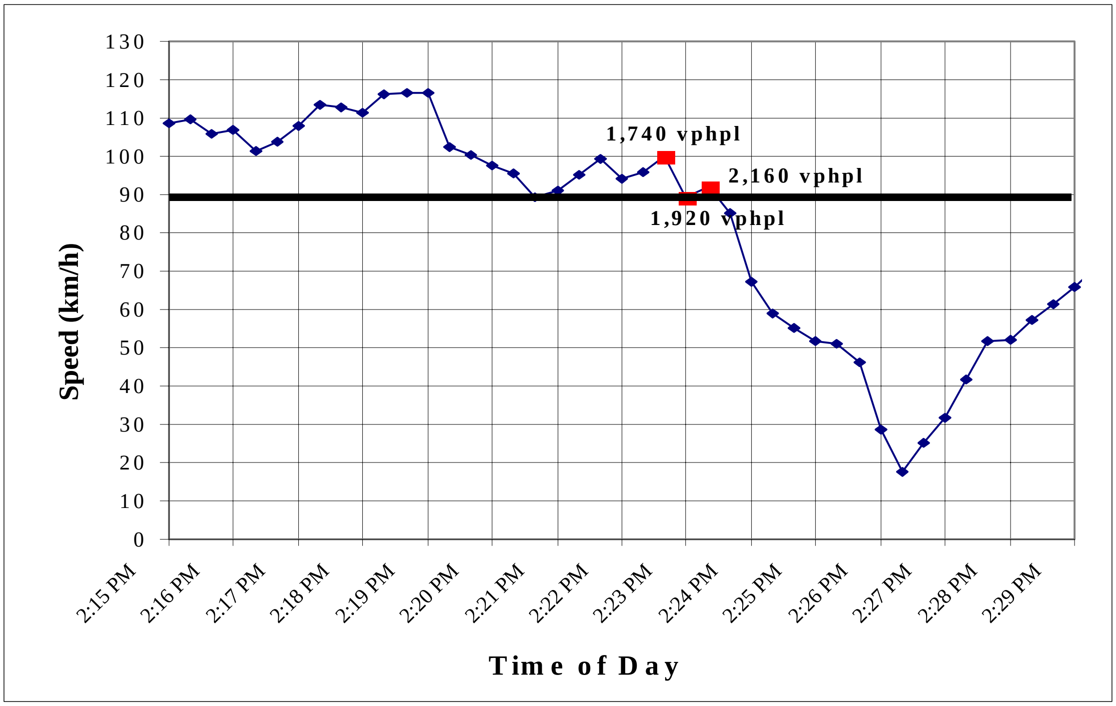

FIGURE 4 Time-series speed plot: breakdown event, Site "B" — March 17, 1998.

Figure 4 illustrates in greater detail the initial breakdown event, based on the time-series speed data plotted in Figure 3. Figure 4 highlights the three data points associated with the breakdown flow rate—the three 20-second intervals occurring immediately prior to the sustained speed drop below 90 km/h. The observed equivalent hourly flow rate (averaged across the three lanes) corresponding to each of these three data points is noted in Figure 4 as well. For illustrative purposes, only the one-minute aggregation interval is highlighted in Figure 4. Aggregations at five-minute intervals and 15-minute intervals were also conducted as part of the analysis. These aggregation intervals correspond to conditions occurring during the five-minute and 15-minute periods immediately preceding breakdown. All breakdown flow rate values were rounded to the nearest 100 vehicles per hour per lane (vphpl) for subsequent analyses. The breakdown flow rate was defined here as the per-lane, equivalent hourly flow rate that occurs during a selected interval (one-minute, five-minutes, or 15-minutes) immediately prior to the breakdown of operations on the freeway. This is not to suggest that realization of the breakdown flow rate at any other time necessarily causes the breakdown of the facility. Rather, it is expected that higher flow rates will merely lead to a higher probability of breakdown.

3.3.5 Breakdown frequency

After determining the one-minute, five-minute, and 15-minute breakdown flow rates for each daily sample, all of the flow rates at each study site were reviewed. This review revealed that, in many cases, flow rates equal to (or exceeding) the breakdown flow rate were observed to occur prior to breakdown on the freeway, yet *not* result in a breakdown—despite the fact that they occurred on the same freeway facility, and even on the same day. In addition, flow rates significantly lower than traditional capacity flow rates (1,000 vphpl, for example) were observed just prior to breakdown.

Breakdown flow rates were determined for each of the daily data samples using the methodology described in the previous section. Since each study site has unique geometric and operational characteristics, the methodology described above was applied independently for each site. However, at each site, the comparison of flow rates was extended to include rates occurring over the course of all days in the sample at that particular site. For example, the analysis compared an observed five-minute breakdown flow rate of 1,900 vphpl at Site "A" to five-minute flow rates of 1,900 vphpl occurring on all other days at Site "A."

3.3.6 Preliminary models for probability of breakdown

Preliminary models were then prepared illustrating the probability of breakdown as a function of flow rate were prepared for Sites "A" and "B." These models are illustrated in Figures 5 and 6, respectively. As shown in Figures 5 and 6, separate models have been prepared for the one-, five- and 15-minute aggregation intervals at each site. Examining Figures 5 and 6, one trend is apparent: the probability of breakdown at each site increases with increasing flow rate.

A comparison among the one-, five- and 15-minute time intervals is worthy of some discussion. Examining the data points for a one-minute aggregation interval at either site, it is apparent that the breakdown probabilities associated with the various flow rates are rather low (less than 10 percent), even for flow rates exceeding 2,000 vphpl.

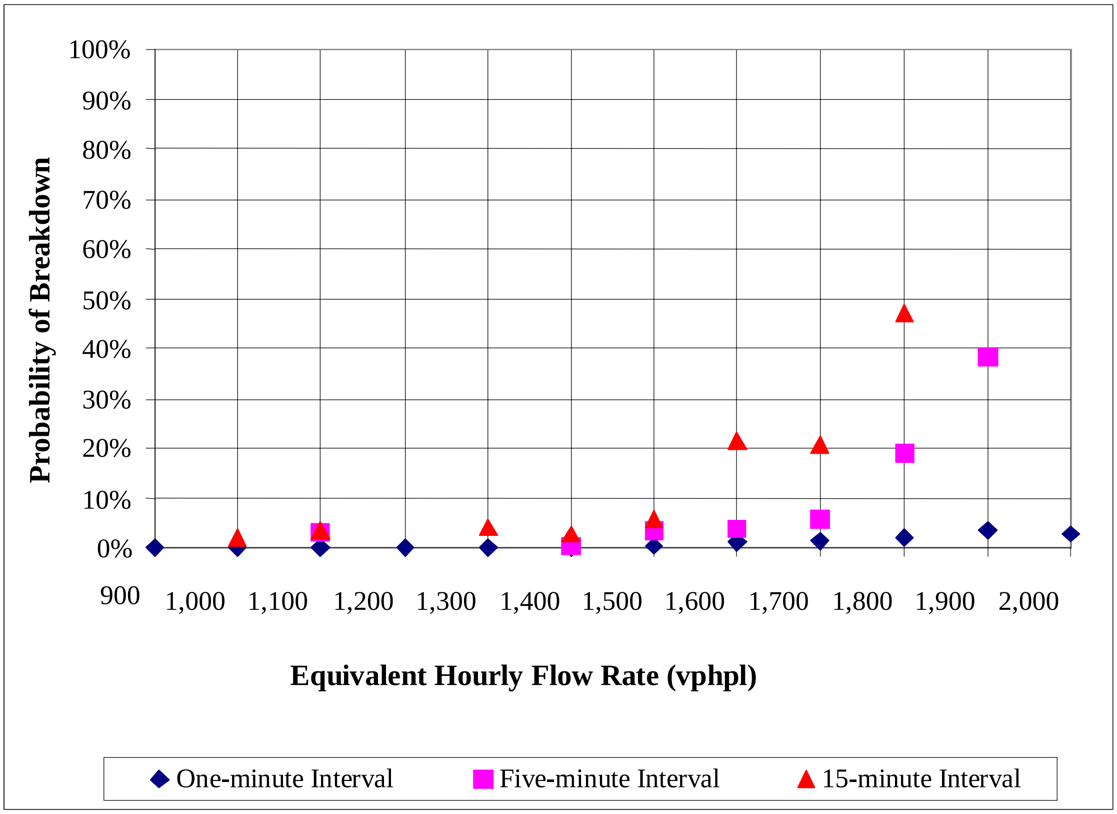

FIGURE 5 Probability of breakdown versus observed flow rate — Site "A."

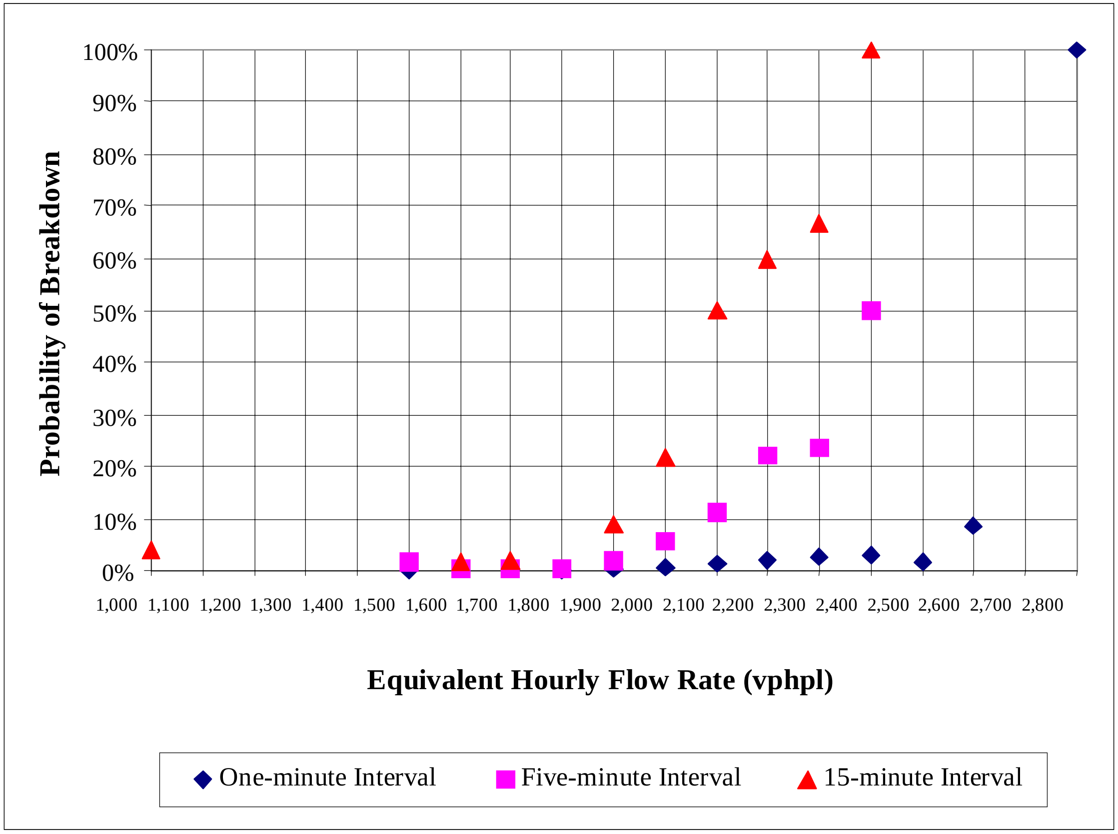

FIGURE 6 Probability of breakdown versus observed flow rate — Site "B."

Essentially, the traffic stream is capable of absorbing brief fluctuations in the flow rate—even those above 2,000 vphpl—without resulting in a high risk of breakdown. This is because the 2,000-vphpl rate is only sustained over a short period (one minute). On the other hand, if the aggregation interval is increased to five minutes, the breakdown probability for 2,000 vphpl is substantially higher because the rate is sustained over a much longer time period. Similarly, the 15-minute interval has the highest probability of breakdown for a given flow rate.

4. INTERPRETATION OF RESULTS

4.1 Comparison of Results to the Existing *HCM* Capacity Definition

Two fundamental implications within the *HCM* definition are refuted by the results described in this paper. The first is the notion that the magnitude of the capacity for freeways can be defined by a single, unchanging numerical value—specifically, a rate of flow. The second implication in the *HCM* is that any flows exceeding this maximum flow rate for a given time period are incapable of being sustained—breakdown must follow. The term "capacity" is defined by the *HCM* with respect to the breakdown event through specification of a maximum sustainable flow rate. According to the *HCM* definition, breakdown is expected to occur on the freeway once the flow rate corresponding to capacity is met or exceeded for the specified time period.

However, the results of the analysis described in this paper do not support such a predictable relationship between breakdown and capacity. Nor do they support the notion that a single flow rate value is an adequate representation of capacity for freeways. On the contrary, breakdown was observed to occur over a wide range of flow rates—rates both lower, and higher, than those traditionally associated with capacity—and were sustained for varying lengths of time. Therefore, all of these rates should be considered as contributing to the description of "capacity" for a given freeway.

4.2 A Proposed Probabilistic Capacity Definition

The results of the analysis described in this report suggest that breakdown is not a deterministic event—it was not found to occur at any particular predictable flow rate value. In addition, the results showed that the probability of breakdown increased with increasing flow rate. Given these findings, it is worthwhile to propose an alternative definition of capacity in light of the probabilistic relationship observed between breakdown and flow rates.

To begin revising the definition of capacity, it is appropriate to first refer back to the concept of capacity and what it should represent: a description of the limit of a freeway's vehicle-carrying ability. Based on the findings described in this paper, this limit is defined not only by a numerical flow rate value, but also by the likelihood that the freeway will break down at that value. Therefore, it seems appropriate to add a component to the capacity definition—one that describes the expected probability that the freeway will break down at a given flow rate.

Further, if a "probability of breakdown" component is to be included in the definition of capacity, the value of the probability component should correspond to the maximum breakdown risk deemed acceptable for a particular time period (for example, during the weekday p.m. peak hour). A target value for "probability of breakdown" or "acceptable breakdown risk" for a freeway might initially be selected by the facility's design team, and later revised by the operating agency or jurisdiction based on the facility's actual operating characteristics.

Lastly, since the probability of breakdown at a particular flow rate varies based on the time interval under consideration, the time interval (one-minute, 15-minutes, etc.) must be specified as well. In summary, a proposed new freeway capacity definition is:

> ...the rate of flow (expressed in pcphpl and specified for a particular time interval) along a uniform freeway segment corresponding to the expected probability of breakdown deemed acceptable under prevailing traffic and roadway conditions in a specified direction.

For example, the capacity of a particular freeway section could be expressed as follows:

> The capacity for Highway 401 at Site "A" is a five-minute flow rate of 2,100 pcphpl, with an allowable probability of breakdown of 36-percent.

As defined above, the capacity of a given freeway could take on a range of values, depending upon an agency's operational objectives during the time period under consideration. In a sense, the capacity of the freeway can be thought of as a "practical" capacity, varying based on the agency's considerations of what is a tolerable risk level at any given time. In addition, if freeway entrance ramps can be metered, the operating agency can exercise control over the entering traffic stream and ensure that breakdown risk thresholds are not exceeded.

It would be useful to prepare plots like those shown in Figures 5 and 6 to describe capacity at other freeway bottleneck locations. Examining curves from other existing sites with similar operational and geometric characteristics might reveal underlying relationships that could be used to assemble curves for more general applications.

4.3 Implications of the Probabilistic Capacity Definition

It is appropriate to review the benefits achieved by reducing the probability of freeway breakdown. As noted by Persaud et al. (1998), preserving uncongested freeway operations results in both safety and travel time benefits to the traveling public. On the other hand, when breakdown occurs and the traffic stream transitions to congested operation travel times increase as queues form, and accidents are more frequent. The freeway capacity definition proposed here continues to recognize the benefits provided by uncongested freeway flow, and was defined with the ultimate objective of permitting a design team or operating agency to select an appropriate breakdown probability that preserves the likelihood of uncongested operations. In other words, as proposed here, the capacity of a particular freeway would be defined solely by the operating agency's willingness to balance the risk of breakdown (and its associated negative consequences), with the desire to accommodate higher traffic volumes (and the associated positive consequences).

5. CONCLUSIONS AND FUTURE RESEARCH

The research described in this report resulted in the following conclusions:

- Breakdown is a probabilistic event and can occur over a range of flow rates, including rates both lower and higher than those traditionally accepted as capacity.
- The probability of breakdown increases with increasing flow rate and with the time interval considered.
- The current *HCM* freeway capacity definition does not accurately reflect the relationship between breakdown and flow rate.
- Freeway capacity may be more adequately described by incorporating a probability of breakdown component in the definition. A suggested definition reads:

  > ...the rate of flow (expressed in pcphpl and specified for a particular time interval) along a uniform freeway segment corresponding to the expected probability of breakdown deemed acceptable under prevailing traffic and roadway conditions in a specified direction.
- The value of the probability component should correspond to the maximum breakdown risk deemed acceptable for a particular time period. A target value for the acceptable probability of breakdown (or "acceptable breakdown risk") for a freeway might initially be selected by the facility's design team, and later revised by the operating agency or jurisdiction based on actual operating characteristics.

In addition, the following research needs were identified as part of this study:

- Plots similar to those illustrated in Figures 5 and 6 should be prepared for other existing freeway bottleneck sites with similar operational and geometric characteristics. Doing so may reveal underlying relationships that may be used to assemble probability of breakdown curves for more general applications.

REFERENCES

Banks, J.H. The Two-Capacity Phenomenon: Some Theoretical Issues, *Transportation Research Record 1320,* 1991, pp. 234–241.

Elefteriadou, L., R.P. Roess, and W.R. McShane. Probabilistic Nature of Breakdown at Freeway Merge Junctions, *Transportation Research Record 1484,* 1995, pp. 80–89.

Persaud, B.N., S. Yagar, and R. Brownlee. *Exploration of the Breakdown Phenomenon in Freeway Traffic,* Preprints of the 77th Annual Meeting of the Transportation Research Board, 1998.

Transportation Research Board, National Research Council. *Special Report 209: Highway Capacity Manual*, 1997 Edition.