Some Observations on Speed-Flow and Flow-Occupancy Relationships Under Congested Conditions

FRED L. HALL, ANNA PUSHKAR, AND YONG SHI

There is a need for a better understanding of freeway operations under congested conditions. Not only are major freeways in many cities congested several hours a day, but intelligent vehicle-highway system programs to be able to function effectively under all conditions, aggregate traffic behavior under all conditions needs to be accurately modeled. Three models for congested flow-occupancy data have been put forward in the past. Using data from the Highway 401 freeway management system in Toronto, evidence in support of any of those is sought, as it is for implied models of speed-flow relationships. The data suggest that all three models can be supported. Each may be appropriate under different conditions.

This paper attempts to begin to provide insight into the nature of congested operations on freeways, with particular reference to the flow-occupancy curve and the speed-flow curve. This may seem like a pointless task. Transportation agencies would prefer to avoid operating their facilities under such conditions. Why then should the 1985 *Highway Capacity Manual* (HCM) (*1*) or theoretical discussions pay attention to those conditions?

It is important for three reasons to know how traffic behaves within congestion. The first is the fact that in many jurisdictions, the major urban freeways are congested several hours a day, a situation that appears likely to get worse before it gets better. Consequently, transportation professionals need a better understanding of those conditions. The second is that the Committee on Highway Capacity and Quality of Service is currently attempting to gain that understanding and is in the process of revising the chapters of the HCM that describe these fundamental relationships. This is therefore a good time to contribute to that discussion. The third reason looks to the future, and to intelligent vehicle-highway systems (IVHS). For the promise provided by IVHS to come to fruition, there must be a solid understanding of aggregate traffic behavior under congested conditions as well as under free-flow conditions. For IVHS to provide realistic advice to motorists (or to make the vehicle route decisions for them, as has been suggested in more comprehensive IVHS schemes), the system will need to be able to predict travel conditions in the near future. An understanding of behavior within congestion is critical to that task.

The first section of this paper provides the background for the investigation that follows, by looking at what is known about congested operations and what has previously been suggested about how best to study them. The second section describes the freeway system from which the data for the paper have been drawn and the methods used to select and reduce the data. The third and fourth sections contain the substance of the investigation, in the form of discussion of a number of figures displaying relevant data and different models derived from them, first for flow-occupancy curves, then for speed-flow curves. The final section provides the conclusions that can be drawn from the study.

Department of Civil Engineering, McMaster University, Hamilton, Ontario L8S 4L7, Canada.

BACKGROUND

To judge by the 1985 HCM, which is the most recent representation of the consensus view of North American highway professionals, there is not currently a sound understanding of freeway operations within congestion. The figures in the 1985 HCM showing speed-flow relationships provide only a dotted representation for the congested part of the curve, suggesting some uncertainty about its location (e.g., HCM Figures 3-4 or 3-6). The density-flow graphs provide an even stronger indication of uncertainty, showing only the first quarter or so of the congested portion of the curve beyond capacity (HCM Figures 3-3 and 3-6). This hesitation about defining the congested portion of the curve is continued in the revised Chapter 7 of the HCM, recently published (*2*), which shows only the uncongested portion of the speed-flow relationship and makes no effort to describe the congested portion (for multilane roadways).

With regard to the freeway material, the reluctance to provide a representation for congested operations is entirely understandable in light of the paucity of information available in the literature. Going back as far as Greenshields' (*3*) seminal work (which admittedly was not of freeway data), there is only one data point within the congested part of the curve. The best data set for dealing with the congested part of the curve is probably that used by Drake et al. (*4*) to compare a number of models of traffic relationships. For their analysis, they had between 50 and 60 observations within the congested area (depending on how that is defined), in which each observation represented 1 min of data for the middle lane of a 3-lane expressway. The other study with a considerable amount of congested traffic data was that conducted by Ceder (*5,6*). In both of these analyses, however, parameters related to the fitting of relationshps to the congested data were determined in part by the uncongested data. The most recent analyses, by Hall et al. (*7*), suggest that the uncongested data are misrepresented by those models; hence it is very likely that Ceder's models distort the congested part of the curves. The 1992 paper by Hall et al., although making a start at defining the shape of the two curves, could define closely only the uncongested and queue discharge portions of the curves. Because of a lack of data, the authors had to rely on abstract logic rather than data to specify the congested part of the speed-flow curve; they could not decide between two placements of the congested part of the flow-occupancy curve on either logical or empirical grounds.

An additional reason for the lack of willingness to represent the congested portion of the curves appears in one of the earlier studies. Greenberg, 1959, noted (*8*, p.84) that in the queue upstream of a bottleneck section, flow is controlled by the bottleneck. [May (*9*, p.288) illustrates this effect.] If flow in the queue is controlled by somewhere else, one might decide that there is little sense in studying congested operations with reference only to what happens within the queue. Certainly the effect of the downstream bottleneck needs to be considered in any effort to understand congested operations. Because it may not be clear where that bottleneck is, a recursive approach may help, in which operations at one section are identified as a function of conditions there and at the next downstream section (which in turn may require looking farther downstream). Nevertheless, the presence of this effect should only serve as a warning for analyses of congested flow; it should not preclude such study entirely.

Wattleworth (*10*) discussed the nature of the flow-density curve that can arise in a bottleneck when its flow is governed by an upstream queue and ramp that together do not supply enough traffic to the bottleneck. He concluded that "it may not be possible to obtain empirically a true volume-density curve for a bottleneck since part of the observed curve may merely be reflecting the influence of conditions upstream of the bottleneck." (*10*,p.20) This observation complements that just noted by May, in the sense that it is likely that no single location can provide the full range of conditions necessary to define the curves and that any location's data can be affected by upstream or downstream conditions. Hence it is particularly important to be aware of surrounding conditions during the time data are being acquired.

To assist the discussion of the congested section of the curves, it is helpful to have a model for the other parts of them. The curves identified by Hall et al. (*7*) will be accepted for subsequent discussion: a linear uncongested section that falls off slightly at higher flows, and a queue discharge segment at constant volume. In the following discussion, the three parts of the curves will be referred to as uncongested, queue discharge, and congested operations. Although queue discharge may be regarded as a form of congestion, in that drivers are traveling at less than their desired speed, for convenience only operations within a queue will be called congested.

For the flow-occupancy curve, Banks has suggested that within-queue flows can be represented by a straight line (*11*). If this is accepted, the question that remains is how this section relates to the other two—in other words, how and where the within-queue section of the curve is to be placed. Three suggestions appear in the previous literature, two of which were included in the flow-occupancy figure by Hall et al. All three have been included in Figure 1. For convenience in later

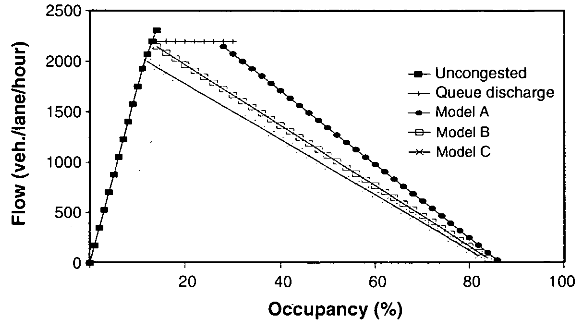

FIGURE 1 Proposed models for congested flow-occupancy.

discussion, they have been termed Models A through C. Model A arises from work in Chicago and joins the congested segment to the right-hand end of the queue-discharge flow section (*12,13*). Model Bel arises from both Ontario and California data and has the congested segment at least aiming toward; if not joining, the uncongested section at the peak of pre-queue flows (*11,14*). Model C arises from Japanese data and has been called the reverse lambda model (*15*).

There is a necessary connection between the decision made as to the placement of the uncongested segment on a flow-occupancy graph and that on the speed-flow graph. Not only must they reflect identical flow conditions within the queue, but the speeds and occupancies at any given flow are inversely related: the lower the speed, the higher the occupancy. If occupancy were a constant multiple of density, as Athol (*12*) suggested, then one could specify the nature of the inverse relationship precisely, starting from the fundamental identity,

$$\text{flow} = \text{speed} * \text{density} \tag{1}$$

to calculate that

$$\text{flow} = \text{speed} * \text{occupancy} * k \tag{2}$$

where k is a constant of proportionality. However, Hall and Persaud (*16*) showed that the linear relationship between occupancy and density holds only under restrictive assumptions, namely, either uniform vehicle lengths or uniform vehicle speeds. Under congested conditions, vehicle speeds are not uniform across vehicles. Vehicle lengths are not uniform even within the lane closest to the median of the roadway, although use of data from only that lane would minimize the variation. However, for data from the full roadway, vehicle lengths vary widely. Hence the preceding equation cannot be used to determine one of the speed-flow and flow-occupancy relationships from the other. Nevertheless, as a starting point for this analysis, Figure 2 shows the three possibilities for the congested portion of the speed-flow curve that correspond with the flow-occupancy curves in Figure 1 based on Equation 2.

Given this background to the problem, the task can now be more clearly specified: what, if any, are the conditions under which each model is correct? That they might all be correct appears likely when one considers the fact that the queue discharge portion of the curve is due to upstream conditions and the congested portion is due to downstream conditions. Each station on the freeway is related differently to

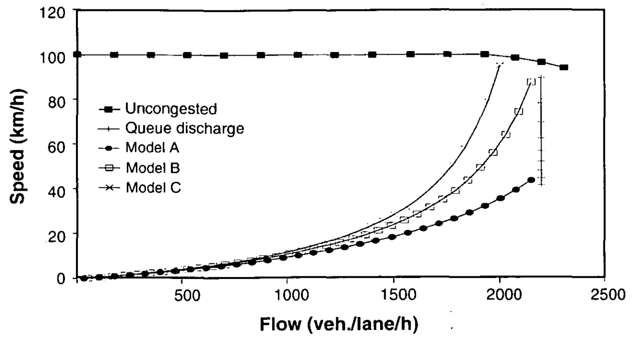

FIGURE 2 Implied models for congested speed-flow relationships.

the upstream and downstream bottlenecks that create these traffic patterns. Hence the queue discharge and congested portions of the curve may not be related in exactly the same way for different stations. The data patterns may then depend on how congestion is arrived at, in the sense that the congested portion of a curve will look different for a location that has been experiencing queue discharge flow, as compared with one that moves into congestion from normal uncongested operations. The investigation needs to take into account two considerations. The first, and potentially more important, is the governing downstream bottleneck. The second is the fact that it is unlikely that the full range of the congested parts of the curves can be specified with data from a single location. Hence it may be helpful to combine data from several locations.

DESCRIPTION OF AVAILABLE DATA

Although it will be necessary to investigate this problem with data from a number of locations, such an investigation must start somewhere. Relevant data were available from a freeway traffic management system (FTMS) that was put into operation January 30, 1991, on Highway 401 in Toronto. Although there is more to the system than inductive loop detectors, those are the only aspects of the system used for this analysis, so they are the only parts described.

Spacing of detector stations on the 401 system averages roughly 500 m, with some as close as 380 m and others as much as 770 m apart. There are double-loop detectors at every third station, where speeds are calculated for each vehicle over the 4.5 m distance between the loops, and average speeds are reported every 20 sec, along with the volume counts and occupancies. For single-loop stations, speeds are estimated on the basis of vehicle lengths calculated at a nearby station, based on the speeds there. These speed estimates are not accurate enough for our purposes. Hence there are potentially three times as many stations available for the flow-occupancy discussion as for the speed-flow graphs.

The part of the 401 highway on which the FTMS has been installed is an express-collector system. Within the 15.7 km of the section on which the FTMS is operating, there are five transfers from the express lanes to the collectors, and three transfer opportunities from collector to express, in each direction. Major bottlenecks occur at some of these, and at a few particular entrance ramps to the collector system. Because the focus of this study was congested traffic, locations upstream of these bottlenecks were of primary interest. Despite the size of the 401 system, and the perception that there is considerable congestion in the system, in fact only a limited number of stations could be used for the analysis. The congestion is concentrated at a relatively small number of stations, and not all of those provided speed data. This study focuses on the section of the 401 east of Weston Road and west of Yonge Street (Figure 3), as it contains several potential bottleneck areas, and most of the detectors were operating properly in the area.

On the 401, the data are available for 20-sec intervals. Although this level of detail is useful for identifying traffic incidents quickly, plots of such data contain considerable random variation that sometimes obscures the underlying relationship. Figure 4 shows an example of the 20-sec data obtained from one of the westbound stations. Although a pattern is apparent, considerable variation is clear within the congested data. To see which presented the clearest picture of the relationship, intervals of 1, 2, 5, 10, and 15 min were examined. [The treatment of missing data within the averaging program was to assume it had the same value as the rest of the interval if up to one-third of the intervals (e.g., five of fifteen 20-sec periods in a 5-min interval) were missing or invalid, and to treat the whole interval as missing if more than one-third of the data was missing.] The 5-min interval was chosen as the best one for the present purpose as it minimized the random variation yet retained some indication of the data patterns. Figure 5 shows the 5-min data for the same station and time period as Figure 4. The transitions to and from congestion occur at approximately the same flow rates in each figure, and the congested data center on 40

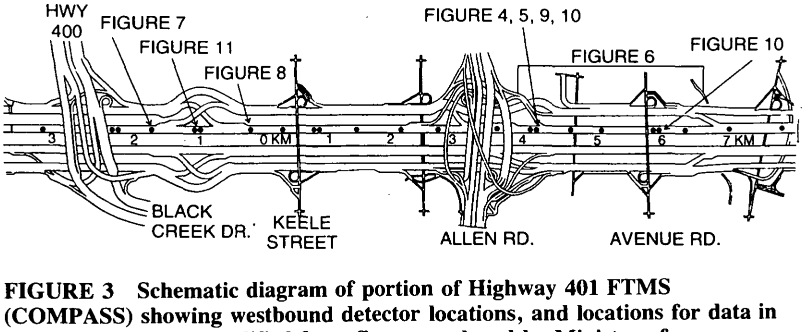

FIGURE 3 Schematic diagram of portion of Highway 401 FTMS (COMPASS) showing westbound detector locations, and locations for data in subsequent figures (modified from figure produced by Ministry of Transportation, Ontario).

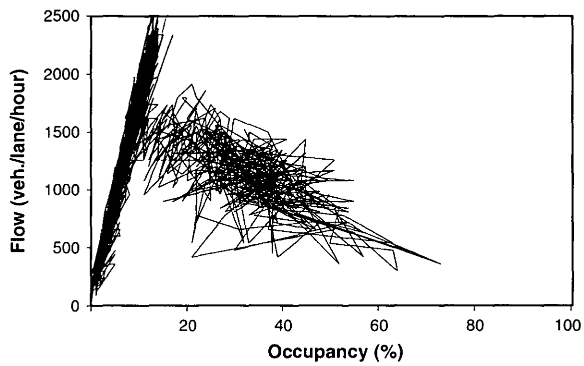

FIGURE 4 Twenty-second data for two peak periods (0500–0900 and 1600–1900).

percent occupancy and a flow of 1,000 vehicles per hour (vph) in both figures. However, the range and scatter of the congested data have been greatly reduced by the averaging.

For the 401 system, the station data are stored in separate files from the loops data, which would provide information on the individual lanes. Although it might be of interest to study the behavior in each separate lane across a station, the analyses here have focused on the station data. Within congestion it appears that all lanes operate similarly, and the primary question for applications is how the roadway as a whole operates.

FLOW-OCCUPANCY RELATIONSHIPS

Before turning to consideration of the shape and position of the speed-flow curve, it is useful to examine the flow-occupancy curves, primarily because more flow-occupancy data are available from the Highway 401 system. Any conclusions from the study of these graphs will help to determine the speed-flow congested curve, on the basis of the approximate relationship specified in Equation 2 and reflected in the deviation of Figure 2 from Figure 1.

The 401 system has some lengthy congested sections in which queues can back up for several miles. The first approach was to take data from a number of stations within one queue, to learn if they provided a similar picture for the congested part of the curve. This was done for five consecutive stations on the westbound express lanes, with data for March 6, 1991. The five are the seventh through the eleventh east of the control center. One of the five is the station shown in Figures 4 and 5. At this station, the data went from normal uncongested operations to congestion at flows of about 1,500 vph, over perhaps 15 min, then stayed within congestion for more than an hour before returning, at lower flows, to uncongested operations. This pattern is fairly typical of the five stations, in both the flows at which congestion set in and the duration of congestion.

The combined data for the five stations appear in Figure 6. The difference between movement into and out of congestion is lost in the general scatter of the combined data. An attempt was made to fit a third-order polynomial on occupancy to the congested data, using a stepwise regression analysis, but only one of the three variables could be made to enter. The first-order correlation coefficients are very similar for all three variables in the polynomial: the square of occupancy produces an R^2 of .748, occupancy alone an R^2 of .737, and the cube of occupancy an R^2 of .722. However, on the basis of the jam occupancy that results from putting each of these into a separate equation, the linear function is most sensible, as it is the only one to produce an x-intercept of more than 70 percent. It might be expected that the jam occupancy should be 100 percent, but data such as shown in Figure 4 suggest that the x-intercept lies between 80 and 90 percent. Of course, few if any of these data come from completely stopped traffic, but even when traffic is at a standstill, space remains between vehicles. The equation selected to represent these data is

$$\text{flow} = 1919 - 21.8 * \text{occupancy} \qquad (3)$$

This equation, as is obvious from Figure 6, is not consistent with either Models A or B of Figure 1. The line would join the uncongested line in the vicinity of a flow of 1,700, as in Model C, the reverse lambda suggested by Koshi (*15*).

However, these data are not the only pattern observed. Figure 7 shows data from a station that was experiencing what appears to be queue discharge flow, at a flow rate in the vicinity of 1,750 vph and occupancies of 30 percent, when a queue from a downstream incident reached the location. The data seem to follow Model A, the Chicago model, quite well.

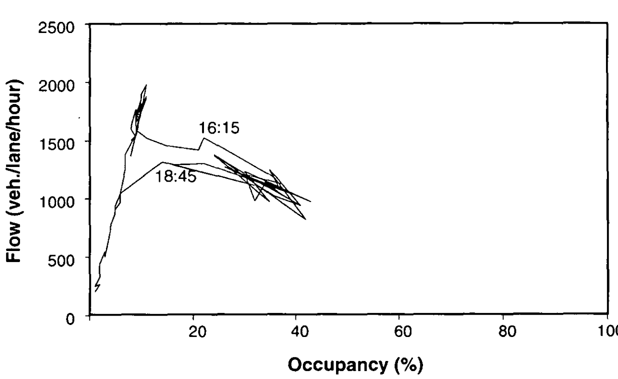

FIGURE 5 Five-minute aggregate data for two peak periods.

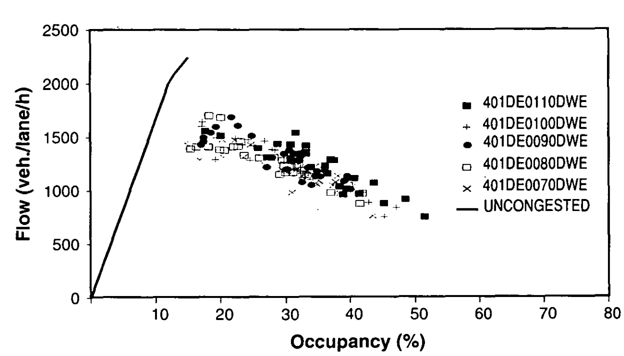

FIGURE 6 Combined congested data for five westbound express stations.

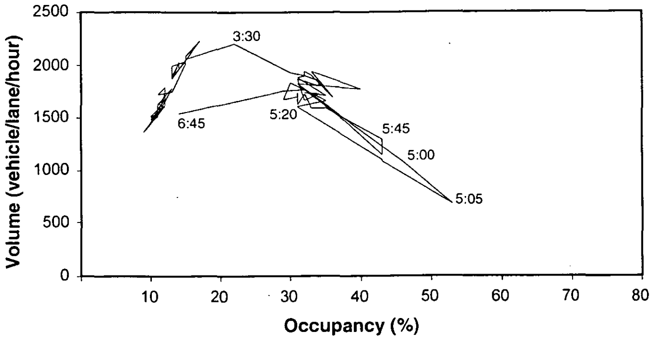

FIGURE 7 Flow-occupancy pattern where queue discharge precedes operations within a queue.

In addition, if the congested segment of the curve is extrapolated back to the uncongested segment, it would appear to be consistent with Model B. Regardless of whether Model A or B (or both) is supported by Figure 7, it is clear that these data occur at higher flow rates than the data in Figure 6, for the same occupancy. Hence at least two different patterns of congested flow-occupancy behavior can occur, depending on the circumstances.

The picture of flow-occupancy behavior can be even more complicated. Figure 8 shows the data pattern for a location that was experiencing normal queue behavior (as represented by the large cluster of points at flows of 600 to 800, and occupancies of 40 percent or so) at the time that a jackknifed tractor-trailer downstream of this location caused a severe capacity reduction. The series of points from 12:10 to 13:15 do not fall on any of the curves that have been identified. The relationship between flow and occupancy within congestion may be much more complex than one might hope.

SPEED-FLOW RELATIONSHIPS

Because of the absence of speed traps at two-thirds of the stations on the 401, only two of the five stations used in Figure 6 provide good speed data. Consequently, two days from each station have been used to provide a comparable estimate of the shape of the congested portion of the speed-flow curve. As an example of the full context for such data, Figure 9 shows the 5-min averages of the speed-flow data corresponding to the flow-occupancy data of Figure 5.

The four sets of congested data are combined in Figure 10. As with the flow-occupancy data, an attempt was made to fit a third-order polynomial, in this case forced through the origin, to the data. For the speed-flow data, two terms were needed, the linear term and either the quadratic or cubic term. With the former, the R^2 was .936; with the latter, .937. The resulting equations give, respectively, speeds of 72 and 81 km/hr at a flow of 2,000 vph, and speeds of 87 and 103 km/hr at flows of 2,200 vph. Since the 103 km/hr speed is certainly too high, the quadratic equation will be accepted for the moment. The equation is therefore

$$\text{speed} = 0.003 * \text{flow} + 0.0000166 * \text{flow}^2 \tag{4}$$

This model would seem to be most nearly consistent with Model B of Figure 2.

It is possible that the top four points in the graph are really transitional ones and not truly congested data. Both equations changed when these points were omitted from the analysis, with the coefficient on the linear term becoming larger and that on the quadratic or cubic term being reduced by about a third. The R^2 for both equations became .939. Speeds at 2,000 vph were only 62 and 68 km/hr respectively. Such a model might seem to be consistent with Model A of Figure

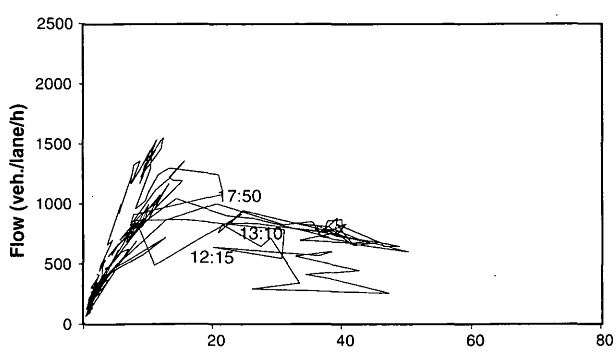

FIGURE 8 Flow-occupancy pattern with queueing resulting from a severe capacity restriction.

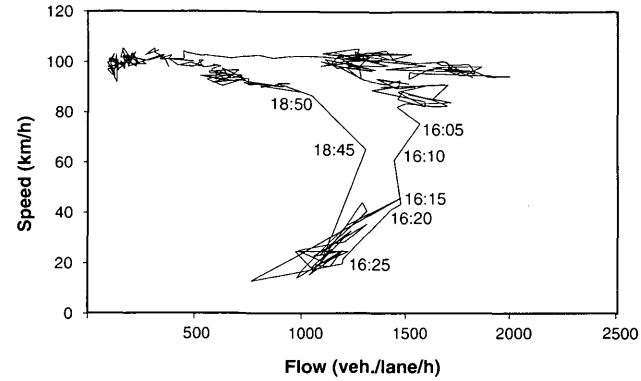

FIGURE 9 Speed-flow data corresponding to flow-occupancy data of Figure 5.

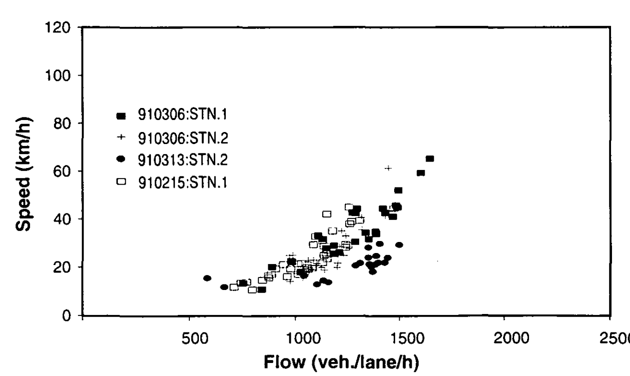

FIGURE 10 Combined congested data from two stations and 3 days.

2, but other data suggest that those speeds are not low enough to reflect Model A.

Figure 11 shows the pattern for data immediately upstream of the major transfer from collector lanes to express lanes. When congestion sets in at this station, speeds drop abruptly, to the vicinity of 30 km/hr, for flows ranging from 1,500 to 2,000 vph/lane. These data are clearly not consistent with the data in Figure 10, yet the pattern in Figure 11 is reproduced regularly. These data are consistent with Model A, suggesting that exclusion of the top four points in Figure 10 would not be appropriate, and that the best representation of the combined speed-flow data in Figure 10 is as in Equation 4.

CONCLUSIONS

Each of the models identified in Figures 1 and 2 seems to have data to support it, based on the examination of some data from Highway 401 in Toronto. It seems most likely that each could apply at different locations along the freeway or under different conditions. For example, Model A in the flow-occupancy case seemed to apply when a queue extended upstream to a station that had previously been operating in queue discharge mode. The speed-flow version of Model A was found in the first station upstream of a major transfer lane. Flow-occupancy Model C was found within an extended queue; the speed-flow data from within that queue seemed most likely to match Model B.

One conclusion of the analysis, then, is that the simple approximation provided by Equation 2 (relying on occupancy being a constant multiple of density) is not a close enough approximation to be of help in specifying the nature of relationships within congested operations. If it were, then the same model would apply to both the flow-occupancy and the speed-flow data from a particular location. A theoretical derivation explaining the problems with that approximation appeared in Hall and Persaud (*16*), as mentioned earlier. These results emphasize the problems of continuing to use the fundamental equation within congested operations.

Some data were also found, in the queue upstream of a major accident, that did not match any of the models. Contrary to the simple smooth curves that might have been expected, such as those that provide reasonable representations of uncongested operations, the situation within congested operations seems quite complex. Further investigation of congested operations is warranted. It may also help to unravel the complexity if some analyses are conducted lane by lane.

One important question that should be addressed for further analyses is what the appropriate averaging interval should be. This paper has relied primarily on 5-min data, in part for reasons explained above, but also in part for consistency with the types of analyses that underlie the current HCM. Certainly much information is lost by moving from the detail available at shorter intervals (such as in Figure 4) to a 5-min average. If regression analyses are to be used to try to develop equations from the data, then it may be better to retain the data for the shortest available interval. If the results of analyses such as these are to be used in IVHS installations, shorter intervals would also make sense. On the other hand, few facilities collect data at the 20-sec intervals used on the 401. There is an argument for some standardization of analyses.

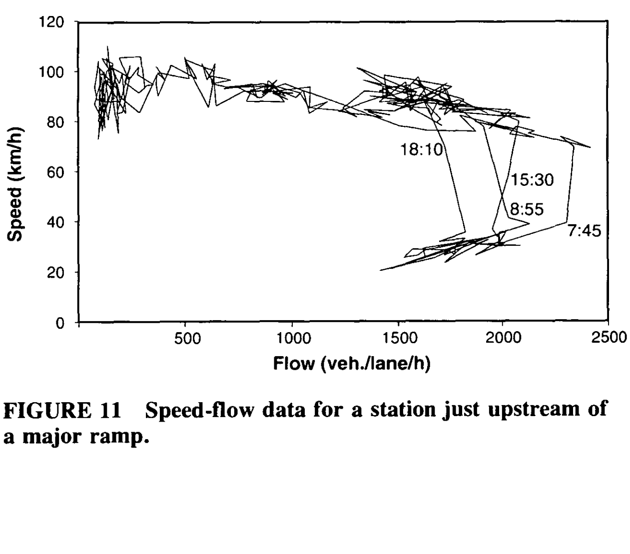

FIGURE 11 Speed-flow data for a station just upstream of a major ramp.

ACKNOWLEDGMENTS

The work described here has been supported by the Natural Sciences and Engineering Research Council of Canada, through both a research grant to Fred Hall and a summer research assistantship to Anna Pushkar. The assistance of the Ministry of Transportation of Ontario, through the Freeway Traffic Management Section, in making available the data is greatly appreciated.

REFERENCES

1. *Special Report 209: Highway Capacity Manual*. TRB, National Research Council, Washington, D.C., 1985.
2. *Special Report 209S: Supplement to Highway Capacity Manual: Revised Chapter 3, Chapter 7, Table of Contents, and Index*. TRB, National Research Council, Washington, D.C., 1992.
3. B. D. Greenshields. A Study of Traffic Capacity. *HRB Proc.*, Vol. 14, 1935, pp. 448–477.
4. J. S. Drake, J. L. Schofer, and A. D. May Jr. A Statistical Analysis of Speed Density Hypotheses. In *Highway Research Record 154*, HRB, National Research Council, Washington, D.C., 1967, 53–87.
5. A. Ceder. *Investigation of Two-Regime Traffic Flow Models at the Micro- and Macroscopic Levels*. Ph.D. thesis. University of California, Berkeley, 1975.
6. A. Ceder and A. D. May. Further Evaluation of Single- and Two-Regime Traffic Flow Models. In *Transportation Research Record 567*, TRB, National Research Council, Washington, D.C., 1976, pp. 1–15.
7. F. L. Hall, V. F. Hurdle, and J. M. Banks. Synthesis of Recent Work on the Nature of Speed-Flow and Flow-Occupancy (or Density) Relationships on Freeways. In *Transportation Research Record 1365*, TRB, National Research Council, Washington, D.C., 1992.
8. H. Greenberg. An Analysis of Traffic Flow. *Operations Research*, Vol. 7, 1959, pp. 79–85.
9. A. D. May. *Traffic Flow Fundamentals*. Prentice-Hall, Englewood Cliffs, N.J., 1990.
10. J. A. Wattleworth. Some Aspects of Macroscopic Freeway Traffic Flow Theory. *Traffic Engineering*, Nov. 1963, pp. 315–320.
11. J. H. Banks. Freeway Speed-Flow-Concentration Relationships: More Evidence and Interpretations. In *Transportation Research Record 1225*, TRB, National Research Council, Washington, D.C., 1989, pp. 53–60.
12. P. Athol. Interdependence of Certain Operational Characteris-

tics Within a Moving Traffic Stream. In *Highway Research Record 72*, HRB, National Research Council, Washington, D.C., 1965, pp. 58–87.
13. A. D. May, Jr., P. Athol, W. Parker, and J. B. Rudden. Development and Evaluation of Congress Street Expressway Pilot Detection System. In *Highway Research Record 21*, HRB, National Research Council, Washington, D.C., 1963, pp. 48–70.
14. F. L. Hall, B. L. Allen, and M. A. Gunter. Empirical Analysis of Freeway Flow-Density Relationships. *Transportation Research A*, Vol. 20A, 1986, pp. 197–210.
15. M. Koshi, M. Iwasaki, and I. Ohkura. Some Findings and an Overview on Vehicular Flow Characteristics. *Proc., 8th International Symposium on Transportation and Traffic Theory*, 1983, pp. 403–426.
16. F. L. Hall and B. N. Persaud. An Evaluation of Speed Estimates Made with Single-Detector Data from Freeway Traffic Management Systems. In *Transportation Research Record 1232*, TRB, National Research Council, Washington, D.C., 1989, pp. 9–16.

The ideas expressed here represent the opinions of the authors only.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.