Synthesis of Recent Work on the Nature of Speed-Flow and Flow-Occupancy (or Density) Relationships on Freeways

Fred L. Hall, V. F. Hurdle, and James H. Banks

Research published during the past 5 years has provided a revised picture of the relationships among the key traffic variables of speed, flow, and concentration. A review of the data from earlier studies with this revised picture in mind shows that many of these data are also compatible with the new picture.

The purpose of this paper is to pull together some of the ideas about speed-flow-concentration relationships that have appeared in the past decade. Because the results are not consistent with the commonly accepted depiction of these relationships, it is useful to trace the relationships back to their original development, more than 50 years ago, and to look at some of the old data from a new perspective based on the more recent data. Our conclusion is that little has changed except for the interpretation: freeway traffic may move a little faster and at somewhat smaller headways than in the past, but the old data are remarkably consistent with both the new data and the new interpretations.

The consequence is that a good case can be made for how we ought now to be interpreting the nature of these fundamental relationships, though some qualitative issues remain unresolved. It should also be made clear at the outset that our paper is empirical rather than mathematical. We believe that it is important to have a correct picture of the relationships before attempting to construct detailed mathematical models and that the classical picture found in most textbooks and in the Highway Capacity Manual (HCM) (1,2) is seriously defective. In this paper, we attempt to draw and defend a new picture that is in better agreement with recent empirical research.

The starting point for this discussion is the 1985 HCM (1). That publication raised several unanswered questions about the speed-flow relationship, in particular "the difficulty in firmly fixing the shape and location of the curve" beyond about 1,500 vehicles per hour (1, pp. 2-24). However, in Chapter 3, Basic Freeway Segments, definitive statements had to be and were made about the shape of these curves. In the 1965 HCM (2), the speed-flow curves were parabolic, with speeds decreasing with each increase in flow, even for very low flows; the 1985 HCM reduced the rate of the speed drop at low flows, but the basic shape remained the same: the slope of the curve is always negative and the right end of the curve is vertical.

There have been a number of papers in the past few years dealing with speed-flow-occupancy relationships on freeways. Our synthesis of that work is given added timeliness by the recently approved revised version of Chapter 7 of the HCM (3), which contains speed-flow relationships for multilane rural highways radically different from those that appear in Chapter 3. Not only do these new curves keep speeds constant until 75 percent of capacity, but their slope is quite modest even as the flow approaches capacity, with a total speed decrease of only 5 mph over the entire range of flows.

The first section presents our conclusions about the shapes of the speed-flow and flow-occupancy graphs. These conclusions are supported by citations from the literature of the past half dozen years. The second section is a review of older material, which on the face of it might appear to be in conflict with the recent results. One of the primary purposes of this section is to see whether some of the older data might also be consistent with our interpretation of the newer data, or whether freeway traffic behavior has clearly changed over the intervening 30 years.

DEPICTION OF THE FUNDAMENTAL CURVES

This section presents graphically our current understanding of the key bivariate relationships: speed-flow and flow-occupancy. The speed-occupancy relationship follows as a consequence of the other two. We do not discuss it here because of a lack of recent data on it and a lack of space. The two figures are drawn in general terms, without numerical values on the axes, but were constructed on the basis of published studies.

Each figure consists of three segments, represented by series of asterisks, squares, and triangles. The intention is that each of these points be taken as representing the average of some large number of observations. Thus stochastic variation has been suppressed. Our discussion of these figures is based on a hypothetical freeway like the one shown in Figure 1. The entire section has the same capacity, but two entrance ramps are followed by an exit. Clearly the flows will be greatest in Section CD, between the downstream entrance ramp and the exit. Section CD, therefore, is the bottleneck, and if a queue forms anywhere, it will be in Section BC, then spread into AB. It seems obvious that the speed within the queue will be slow and the density or occupancy high. Within the bottleneck...
Section CD, however, one would suppose that the traffic might accelerate, such that there will be higher speeds at D than at C, and measurements confirm this commonsense conclusion (4).

**Speed-Flow Relationships**

Figure 2 shows the most likely interpretation we have found of the speed and flow data acquired from numerous freeway systems over the past 30 years, as applied to the hypothetical freeway in Figure 1. The nearly level upper line (indicated by * in the figure) represents traffic behavior in the absence of any queue. On the freeway in Figure 1, we would expect the conditions represented by this curve to occur everywhere until a queue formed in Sections BC and AB, but after that only in the part of AB upstream from the end of the queue and in DE, downstream from the bottleneck. The 1985 HCM (1) reduced the rate of the speed drop from that shown in the 1965 HCM, and the 1990 revisions to Chapter 7 (3) enhance that tendency, keeping speeds constant until 75 percent of capacity is reached. Those results for multilane rural highways are consistent with comparable observations available for freeways. The only disagreement between the recent freeway studies and the Chapter 7 curve might be in the magnitude and shape of the speed decrease for higher flow rates. The revised Chapter 7 shows drops of only 5 mph, or roughly 10 percent of the free-flow speed. Hurdle and Datta (5), Persaud and Hurdle (4), and Hall and Hall (6), as reinterpreted by Hall and Agyemang-Duah (7), suggest about a 20 to 25 percent drop. Wemple et al. (8) found a 10 to 15 percent drop. Banks (9) found a drop of less than 10 percent.

An alternative way of looking at these data is not to focus on the speed drop but to look at the speed at the endpoint of the curve. Several of the freeway studies cited find that point to be about 80 km/hr, or 50 mph. If speed at the right end of the curve is approximately the same for some large class of freeways, the magnitude of the speed drop is simply a function of the free-flow speed. Such a conclusion would go a long way toward explaining the differences between data from North American urban freeways and the high-speed German Autobahnen. Unfortunately, compatibility with Japanese data seems less likely.

Several studies, however, report different results. Persaud and Hurdle (4) cite a lower value, but the relevant figures in their paper suggest considerable scatter, with a range that includes 80 km/hr. Wemple et al. (8), on the other hand, report speeds of more than 60 mph at their I-680 site when the flow is more than 2,000 vehicles per hour per lane and show data for a site on I-880 that could be interpreted as indicating no decrease, or even an increase, in speed at very high flows. Hurdle and Datta (5), Persaud (10), and Hall and Hall (6) include similar evidence of circumstances in which speed seems to be completely independent of flow, but it will be presumed in this paper that this is not ordinary behavior.

For constructing Figure 2, a speed drop of 20 percent of free-flow speeds has been used, which would be consistent with a 100-km/hr or 60-mph free-flow speed and an endpoint speed of 80 km/hr or 50 mph. Despite the possible discrepancy between this value and the Chapter 7 curve, it is important to note that flows near capacity, when they occur before queue formation, occur at much higher speeds than are currently shown in the HCM and that none of the recent data indicate that this portion of the curve becomes vertical. The shape shown for the right end of the curve is arbitrary. The new Chapter 7 curves change slope in a more or less quadratic manner, but several data sources seem to imply a straight line.

The second segment to consider is the vertical band represented in Figure 2 by the squares. This is the behavior to be expected in bottleneck Section CD when there is a queue upstream. Everywhere within the bottleneck section, the mean flow rate will be the same, but the mean speed will be a function of the location of the observation point since drivers are accelerating from the slow speeds within the queue to their desired speeds for that section of roadway (4). Thus, the farther downstream from the front of the queue one measures, the higher the average speed will be. The vertical segment of the curve shown in Figure 2 is not really a speed-flow relationship at all, but a speed-location relationship plotted on a graph that has no location axis.
Although any one location is depicted in Figure 2 as a single point, both flow and speed are random variables, so actual data will contain a good deal of scatter in both directions. For example, Hurdle and Datta (5) reported standard deviations of 205 passenger car units per hour (for flows based on 2-min counts) and 11.0 km/hr for average queue discharge flow and speed of 1,984 passenger car units per hour per lane and 79.5 km/hr, at a point 2 km downstream from the head of a queue. Furthermore, the flow and speed random variables are not independent, so the queue discharge data at one location may exhibit a distinct slope. Data from several locations suggest that at a given location lower queue discharge flow will be accompanied by lower speeds. Hence, data from this part of the curve may look as if they are on the third part (the triangles). Figure 3 provides an example of both the scatter and the apparent slope. The data are from an Ontario freeway from two locations downstream of an entrance ramp, one 0.2 km (the filled squares) the other 2.5 km (the open squares) downstream. The data from 0.2 km downstream are clearly centered at an average speed well below that of the data from 2.5 km downstream, which are themselves still below the uncongested data. There is a trend in the data from 0.2 km downstream that might be thought to look like part of the lower branch of a two-regime speed-flow curve despite the fact that these data come from queue discharge flow. Koshi (11) has reported finding a decrease in queue discharge flows on Japanese freeways as queuing delays increased from 0 to about 10 min, but we have seen no evidence of such a trend in North American freeway data.

The final segment of the curve, the triangles, represents behavior within the queue. This segment of Figure 2 has been drawn on the basis of logical considerations rather than from data. There is very little modern data for this segment and hardly any below flows of 50 percent of capacity. Some even question the existence of a speed-flow relationship for these conditions. It is reasonable to suppose that there is a relationship of averages taken over periods long enough that the effect of stop-and-go conditions within the queue are smoothed out. The average time a vehicle spends in the queue (and thus its average speed) is determined by the number of vehicles to be “served” by the bottleneck before this vehicle, and by the “service rate” of the bottleneck. Thus, if more vehicles enter downstream of the point observed, both the flow and the average speed will necessarily decrease. In Figure 1, for example, the flow in Section AB will be lower than that in BC by the amount of the entrance ramp traffic at B, and average speeds will therefore be lower. This simple argument is persuasive, but it is not sufficient to define the curve. Furthermore, it is clear from the data that the scatter of data about this curve is far greater than for the other two portions of the relationship.

The logic of the curve begins with the conventional idea that the left end must be at the origin since zero speed and zero flow obviously occur together. The right end has been drawn more or less joining the bottom of the portion representing queue discharge flow because data we have seen tend to look like this, because flows in the queue cannot exceed those in the bottleneck, and because it seems illogical to suppose that vehicles being served moved slower than those waiting in line behind them. The last statement, however, is an oversimplification. Flows within Section BC of the Figure 1 freeway can approach those within the bottleneck only if the flow on the entrance ramp at C approaches zero, so the speed just upstream from the bottleneck will usually be less than that indicated by the right end of the triangles curve, which is an indication of what would happen if the ramp flow did approach zero. It follows, since the speed at the upstream end of Section CD cannot be much different than at the downstream end of BC, that the line of squares really should extend below the triangles curve. Like the right end of the plus-sign curve, the bottom end of the line of squares is not well defined. The lowest average speed one will observe in a bottleneck like CD is a function of the entrance ramp flow and the roadway geometry; the figure is only an indication of the possibilities. Furthermore, the right end of the triangles portion is unlikely to be observed in practice, so real data are likely to exhibit a gap, rather than a crossing of curves.

Flow-Occupancy Relationships

Figure 4 shows the flow-occupancy relationship. Although density was almost always the measure of traffic concentration used in the early days, and is still used extensively in theo-
retical work, most present empirical studies have used occupancy because it is commonly measured by freeway management systems. The difference is not important for purposes of this discussion. Athol (12) stated that there is a linear relationship between occupancy and density, but more recent analyses [Koshi et al. (13) and Hall and Persaud (14)] indicate that the linear relationship holds only a portion of the range of the variables, with the nonlinear aspect of the relationship between occupancy and density depending on the covariance between vehicle lengths and vehicle speeds. Given the small magnitude of the nonlinearity and the level of generality at which we will be discussing the relationships, what we say about occupancy can be applied to density as well.

The flow-occupancy graph in Figure 4 also has three segments. The location of the first segment, the asterisks, is well documented; the logic used to explain the squares in the previous section leaves little doubt about where they must lie in this representation; but we are unsure about the third segment, so have shown two versions. Looking at the three portions one at a time, we see that the constant speed portion of the asterisk curve in Figure 2, out to perhaps 75 percent of capacity, is reflected in a linear flow-occupancy relationship up to the same volume; beyond that point, the relationship curves off slightly to the right, reflecting the increase in occupancy that occurs for a given flow rate as speeds decrease.

The second segment of the diagram (the squares) represents queue discharge flows. As in the speed-flow diagram, the average rate of flow everywhere within the bottleneck must be the same and slightly less than the maximum flows observed during prequeue operations. The mean occupancy, on the other hand, will vary with location over a range that is not entirely well defined. At the left end, it seems obvious that the line of squares should meet the line of asterisks, since queue discharge speeds continue to increase until they regain normal uncongested speeds. On the right end, the obvious limitation is that the occupancy should not exceed that at the head of the queue, but just as with the speed-flow curves, this is not a clearly defined limitation since the occupancy in the queue clearly depends on the ramp flows. Thus, in a generalized diagram like Figure 4, the point at which the line of squares ends on the right is necessarily arbitrary. In a diagram showing the results of a specific experiment, there would, of course, be a rightmost point, but this would not mean that points further to the right could not occur if the traffic pattern changed.

The two possibilities for the segment representing the congested regime within the queue (the triangles and the circles) are based as much on logic, and even conjecture, as on data, though we did look at a good deal of data before drawing them. Just as was the case for the speed-flow relationship, the twin difficulties are that all of the studies we have seen show more scatter in these data than in the uncongested case and that empirical information about very low flows is scarce.

Banks (15) has suggested that this portion of the relationship is linear. The very extensive data set presented by Koshi et al. (13) shows slight but definite convexity, but that uses density rather than occupancy. The linear possibility being the simpler one, we have drawn straight lines, but without any intention of implying that this is necessarily correct.

The remaining question is where to place the upper end of this line. Three possibilities have been suggested. May and others working at the Chicago Freeway Surveillance and Control Center proposed that the congested segment start at the right-hand end of the queue discharge flow as indicated by the circles [May et al. (16), Figure 7, p. 53], and it would appear that Chicago continues to accept that depiction [McDermott (17), Figure 4, p. 338]. Koshi et al. (13) proposed a reverse lambda shape, which seems to imply that the line joins the plus-sign curve below the queue discharge curve. Both Hall et al. (18) and Banks (15) have suggested an inverted V, implying that the line should start near the peak of the prequeue flow in somewhat the manner of the line of triangles. The choice among these proposals should be based in part on logic and in part on the available data, but our reaction to the data we have seen is that it leaves as many questions as answers.

If one accepts the triangle line, the queue discharge curve (the line of squares) will stick out from the inverted V like a windblown flag—a situation that seems counterintuitive. However, as discussed in connection with Figure 2, these relationships in Figures 2 and 4 represent a wide variety of situations, not all of which are likely to happen at any one location. The “flag” represented by the squares is a natural consequence for the flow-occupancy graph of the vertical line that is the queue discharge operation, as shown in the speed-flow curve of Figure 2. As was discussed for Figure 2, the queue discharge data from any one location will show a range of flows and occupancies. Since these variables are not independent, a pattern will often be observable within them. A number of examples we have looked at tend to show a “drooping flag,” which might easily be confused with the congested portion of this curve. Figure 5 shows an example of such a pattern from San Diego data downstream of a bottleneck. The downward trend in the queue discharge data is apparent in these data, but the figure also supports the constant volume segment shown in Figure 4 and the interpretation of it as queue discharge flow.

Summary
As was pointed out as long ago as 1958 (16,19), the nature of the data acquired from a freeway depends on where the
measurements are taken with respect to bottlenecks and queues. May (20, p. 288) provides a helpful discussion and set of diagrams explaining the general nature of this phenomenon. In brief, upstream of a limiting bottleneck (e.g. a lane drop or closure), there will be congested flow, but capacity operations will probably not occur. Within the bottleneck section, capacity flow will occur, but there will be no congested operations. If capacity increases downstream of the bottleneck section (for example, the dropped lane has been restored or reopened), at such a location there will be neither congested operation nor capacity operation. One important consequence is that it is impossible to obtain data covering the full range of possible operations at any one station. Indeed, the data needed to create even the one segment for queue discharge operations must necessarily come from a series of locations within the bottleneck.

The consequences of this dependence on location for the specific case of flow-occupancy data are shown in Figure 6. The freeway segment shown in Figure 1 forms the basis of the diagram, with conditions as they would be observed during a peak traffic period. The triangle at the foreground of the picture represents the flow-occupancy inverted V proposed by Banks (15) and Hall et al. (18). There is a queue upstream of Ramp C, extending back beyond Ramp B, but not as far as Location A. Thus for some distance downstream of A, traffic is still uncongested. At some intermediate location between A and B, drivers encounter the back end of the queue and experience an abrupt change of conditions, to heavily congested. When they pass Ramp B, conditions remain congested, but average flow rates increase (and occupancies decrease, because speeds also increase). Upon arrival at Ramp C, drivers are able to accelerate through the bottleneck. Hence the early part of the data in Section CD is the “flag” of Figure 4, and the later part is back on the uncongested surface (if Section CD is sufficiently long). Past Ramp D, volume decreases, due to vehicles exiting. In the absence of incidents somewhere along the road, one should not expect to see Section CD behavior anywhere upstream of C or BC behavior downstream of C.

**REVIEW OF EARLIER WORK**

In reviewing the historical roots of the existing HCM depictions of speed-flow-occupancy relationships, we consider the possibility that much of the earlier data might be consistent with the representations given in the preceding section and that the conventional interpretations arise because of particular historical events. In particular, we suggest that Greenshields’s seminal work in 1935 (21) has had an unduly dominant influence on all subsequent interpretations of such data.

Consider the details of Greenshields’s paper. The data were collected on one lane representing one direction of a two-lane two-way rural road. The seminal graphs of speed versus density and speed versus flow are based on seven data points, six of which are at densities below 60 veh/mi and come from one highway; the seventh is at a density of 150 veh/mi and is taken from a different highway. The straight-line relationship for speed versus density assumes away a lot of missing data between 60 and 150 veh/mi and in turn determines the parabolic shape of the speed-flow curve. Despite the presence of such heroic assumptions, it is possible that these visual representations, with their accompanying elegant mathematics, have had a determining impact on the perception of relationships within freeway data.

Besides the Greenshields model, another possible influence inclining people to U-shaped flow-concentration models may well have been the fact that that was the shape derived by Gazis et al. (22) when they originally linked car-following models and these macroscopic flow models. This additional influence would have been reinforced by the simplicity of single-regime models and the flexibility of the nonlinear models as developed by Gazis et al. (23).

The generalized flow-occupancy relationship presented in Figure 4 (using the circles for the third segment) is similar to one put forth by May et al. (16). It is important to note the differences in interpretation of their flow-occupancy diagram. Their interpretation was that speeds remain constant up to capacity. Ours, on the other hand, has a slight decrease in speeds approaching maximum flow. We agree completely with them that there is a zone of constant volume. However, we have identified this region as queue discharge flow, whereas May et al. say that it “represents impending poor operations” (16, p. 52). Another similarity is the use of a linear relationship within congested operations, although they change the slope of that line at a density of 100 veh/mi.

Despite efforts such as that by May et al. to produce a newer picture of these relationships, the curves appearing in the 1965 manual are still clearly based on Greenshields’s parabola. Two out of the three empirical studies summarized in Figure 3.37 of the 1965 HCM gave clear evidence of nearly constant speeds out to volumes of 1,000 or 1,200 vph/lane, yet the very next page of the 1965 HCM puts forth the parabolic shape as “typical” for freeways and expressways.

Soon after the appearance of the 1965 HCM, Drake et al. (24) published a thorough empirical testing of a number of

**FIGURE 6** Expected approximate location of flow-occupancy data in the vicinity of a bottleneck during congested conditions.
mathematical expressions for these three relationships. One of their two subjective choices of best models was a three-regime linear model—and our proposed diagrams also reflect three distinct “regimes” of behavior. May and his students have continued to work on the question of the best type of mathematical model to fit empirical data [Ceder (25), Ceder and May (26), and Easa and May (27)]. The models still do not fit the data well in the vicinity of capacity, and each data set requires different parameters for the mathematical model, the extreme instance of this occurring for 2 days of data from the same location.

Important insights into the task of estimating the relationships among these data were presented by Duncan (28,29). In his 1976 paper, he dealt with studies in which speed and flow data were generated and density data calculated from the relationship volume = speed \times density. He noted that the consequence of taking a relatively good-fitting speed-density function and transforming it (via the same equation) to a speed-flow function was a function that did not fit well with the original speed-flow data. In his 1979 paper, Duncan used real data, rather than randomly generated data, to enlarge on the difficulties: minor changes in the nature of the speed-density function resulted in major changes in the speed-flow function. The data in his paper would also be consistent with Figure 2, although that is not the nature of the function that Duncan shows.

Overall, then, it can be seen that previous studies do not provide strong support for the parabolic speed-flow curve, which nevertheless continues to influence the shape of the HCM (and other) representations of this relationship [see, for example, May (20, p. 288) or McShane and Roess (30, p. 286)]. On the other hand, it is possible that many of the data in the earlier studies are consistent with Figures 2 and 4. Indeed, as early as 1961, the Chicago group identified many of the same features of these relationships. Hence, it may be that driver behavior has not changed in any fundamental way during the past 30 years. This, however, is a stronger conclusion than is warranted from the available data. Suffice to say that the earlier studies do not contradict our proposed model, and provide some support for it.

**SUMMARY AND CONCLUSIONS**

In all likelihood, no one location will be able to provide data for the full range of operations; hence it will be impossible to identify the shape of the speed-flow-occupancy relationships on the basis of data from one location. Consequently, the curves in Figures 2 and 4 are composite curves, drawing together information from numerous locations. It would in fact be a mistake to attempt to construct separate curves for specific locations, since any particular location can be expected to be missing important parts of the overall relationship.

The location of the congested branch of the curve is still a problem. Two possibilities were offered in Figure 4. Perhaps more important, however, the location of the other two segments of the curve seems to be clear. Uncongested operations on freeways are consistent with the speed-flow figure for multiline rural highways approved last year for a revised Chapter 7 of the HCM. The remaining questions are the flow at which speeds begin to decline, the exact shape and magnitude of the decline, and the value of capacity. One feasible interpretation of the data is that the speed at the right-hand end of this segment of the curve is constant, at least for similar types of freeways and similar driver populations. (Japanese data might show results different from these North American studies, for example). Should this interpretation be correct, the speed drop depends on free-flow speeds, which are probably affected by such things as posted speed limits and level of enforcement. The presence of queue discharge speeds which have been clearly demonstrated, confirming (with a different interpretation) the zone of constant flow identified 30 years ago in the Chicago studies.

**REFERENCES**


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