Further Analysis of the Flow-Concentration Relationship

Fred L. Hall and Margot A. Gunter

On the basis of inspection of time-traced plots of daily flow and concentration data for a freeway, additional support is provided for some tentative new ways of looking at the relationship between these variables. Occupancy data are used to directly measure concentration, rather than converting to density. The overlaid, daily time traces help to make clear the nature of operations and of transitions between congested and uncongested regimes. Four principal conclusions are supported by these plots. First, the underlying inverted V-shape for the flow-occupancy relationship found earlier at a single lane and station on the same freeway has been confirmed at several locations for both of the nonshoulder lanes. Second, for the shoulder lane, it is not possible to determine from these data whether the inverted V-shape correctly describes the relationship. Third, definite differences exist in the parameters of the flow-occupancy relationship that appear to be attributable to lane and location. Fourth, there is additional support for the finding reported in an earlier paper of the authors (Transportation Research, Vol. 20A, 1986) that discontinuous relationships are not necessary to describe the data obtained from freeway operations. Better sense may be made by assuming continuous relationships and trying to explain sparseness of data by the nature of operations on the facility.

Despite some 50 years of research on the relationships between speed, flow, and vehicular concentration, disagreement still exists about what exactly occurs on freeways. An extensive test (using nearly 50 different data sets) of single- and two-regime models, as derived from first principles, has indicated that the match between theory and data is not very good (1). Therefore, a different approach is taken in this paper: the data are inspected closely to determine the form of relationship that is suggested.

In a previous paper, the authors examined flow-occupancy data for one lane at one location on an expressway (2); on the basis of that examination, three conclusions were reached about the relationship. The purpose of this paper is to determine if those same conclusions are applicable to other lanes and locations. That analysis was directed at the conceptual problem caused by gaps in observed data patterns, and suggested that it is not necessary to construct discontinuous functions to account for those gaps. The three main conclusions of that paper are summarized briefly here because they provided the starting point for the current analysis.

First, there are advantages to looking closely at daily data to discern operating relationships, rather than relying uncritically on scatter diagrams of all available data. In particular, it was found helpful to utilize time-connected traces of the daily record of operations. Second, inspection of daily time-traced plots showed a variety of types of transitions from uncongested to congested flow occurring, but the combined result of these types of transitions led to an appearance of sparsity, or even gaps in the data. It was therefore hypothesized that (2, p.210)

The nature of the data that are collected at any particular freeway location depends as much on the specifics of the location as on underlying relationships. In particular, there will be an absence of data for particular parts of the relationship if a queue backs into the location while flow is lower than capacity.

Third, because of this explanation for areas of sparse data, arguments for a discontinuous flow-occupancy (or flow-density) curve do not appear to be convincing. An inverted V-shape for a continuous curve appears to be the most representative shape, given the data examined.

The problem with the latter two conclusions is that they are based on data from only one lane (the median, or left-most, lane), at one station on the roadway (4 km upstream from a bottleneck). In this paper, flow-occupancy data for other lanes and locations along the roadway are examined to determine the extent to which those two conclusions are affected by location along the roadway—particularly with reference to entrance ramps—and by lane.

In particular, three questions are addressed in this paper:

1. Is the inverted V-shape observed at other stations and for other lanes?
2. If it is, does it appear to have similar or different parameters at different locations on the road?
3. Are the patterns seen at the several lanes and stations consistent with the idea of a continuous curve, or does a discontinuous curve appear to be appropriate?

It should be recognized from the start that any conclusions must still be tentative because all of the data come from the same freeway control system. It will remain to be seen if the relationship is different elsewhere.

The flow-occupancy relationship was selected for consideration over the speed-flow or speed-occupancy relationships because, in the authors' initial analysis of one lane and station, it provided the clearest distinction between congested and uncongested regimes (Figure 1). The flow-occupancy relationship therefore offered the most promise for being able to clearly specify the nature of the relationship, and subsequently identify the nature of differences in it between stations and lanes on a roadway. Occupancy is used rather than density for two reasons. First, it is the variable directly measured by the freeway management system and is a point or very short section measurement, which corresponds well with the
The idea that the relationships for different lanes and locations are not the same is not new. Others who have demonstrated or mentioned these differences include the 1965 Highway Capacity Manual (4), Ceder and May (5), Mahabir (6), and Hurdle and Datta (7). However, it is logical to expect that more precise specification of the traffic flow relationships for different locations will be necessary as freeway management systems become increasingly important and widespread.

The first section of this paper contains a description of the data set used in the study. The second section provides an account of the computerized procedures used to reduce the set to a form suitable for analysis. The results of the analysis are presented in the third section. Discussion of the implications and possible interpretation of the results are given in the next section. The final section contains the conclusions and suggestions for future work.

DATA

The data used in this study come from the freeway control system on Ontario’s Queen Elizabeth Way between the cities of Oakville and Toronto. A 5-km section of this roadway experiences congestion in the eastbound direction each weekday.
morning and has therefore been equipped with a freeway management system that collects traffic flow data at nine locations or stations [described in more detail by Case and Williams (8)]. At each lane at each station (lane-station), the following were compiled by the freeway management system for 5-min time intervals: occupancies at the upstream and downstream loops, number of vehicles longer than 7.6 m [a passenger car equivalency of 2 was used for these long vehicles to convert the total volume to passenger car units (pcu’s)], total number of vehicles, and average speed. In addition, a log was kept by the system operator, noting poor weather or incidents such as accidents or stalled vehicles.

For this analysis, 8 months of data were available, collected in 1978 and 1979. The 5-km section of the freeway system is shown in Figure 2, along with the locations of some of the data collection stations. The primary bottleneck is just downstream of the entrance ramp merge at Highway 10. (Unfortunately, data were not collected there.) This is a secondary constriction at the entrance ramp from Mississauga Road, which also has heavy entrance-ramp volumes.

The choice of stations for analysis was important because the effect of location is a major component of the study. The primary criterion was that the traffic conditions be as different as possible. The second criterion was that geometric conditions should be as close to ideal as possible. These two criteria led to the choice of three stations: Stations 9, 7, and 4.

Station 9 is located immediately upstream of the bottleneck and the Highway 10 entrance ramp and is on a slight vertical curve. It experiences congestion for the longest duration of all locations along the section.

Station 7 is situated about 1.6 km upstream of Station 9, with one exit ramp intervening between them. It has almost ideal geometric conditions, but is immediately downstream of a bridge. Just before the bridge is the end of the merging lane for the entrance ramp from Mississauga Road.

Station 4 has good geometrics. It is about 3.9 km from Station 9, and there are two entrance ramps and one exit ramp between it and Station 7.

**DATA REDUCTION**

After the locations and lanes were chosen, the raw data had to be reduced to a set that was appropriate for analysis, and displayed in a form that would be useful for subsequent interpretation. Specifically, the goal for the data-reduction stage was to produce for each lane-station a figure containing overlaid time-traced plots of the nontransitional data for those days that represented ideal conditions.

Because this analysis was intended to compare ideal relationships, days were rejected if weather had been poor or if incidents (stalled vehicles, accidents) had occurred. This reduced the data to 72 days of operation.

The next step was to distinguish between the two regimes of operations for the data for each lane. A critical occupancy value, at which flow was at a maximum, was identified. The choice of this value was somewhat subjective, but could be made with reasonable confidence based on examination of the overlaid daily plots and of plots of average values, following procedures described by Hall, Allen, and Gunter (2).

This procedure worked well with all the middle and median lane plots, but was not suitable for the shoulder lane. In the shoulder lane a peak flow could not easily be distinguished, and the averaging procedures did not help. Therefore, a subjective, and admittedly somewhat arbitrary, value was selected for the critical occupancy to indicate the value at which congested operation began based on the appearance of the plot.

After a critical occupancy had been established, it was possible to identify and reject inappropriate points. For the authors’ previous paper (2) a manual technique was used, which involved examining each day’s time-traced data separately. To analyze all nine lane-stations efficiently, a computerized procedure was developed that used the same criteria.

These criteria defined four categories of rejection for individual data points:

1. Equipment malfunction,
2. Serious doubt as to the data’s numerical validity,
3. Inconsistency with the equilibrium relationship, and
4. Operation in transition between the two branches of the curve.

Each of these categories of rejection will be discussed further.

Most equipment malfunctions had been identified automatically by the system. Two other types were found and rejected: zero recorded for flow rate, speed, or occupancy; and more than a 40 percent difference between upstream and downstream occupancy (which usually differed by less than 20 percent in these data).

The second criterion dealt with serious doubt about particular observations. This was characterized by a lack of reliable

---

**FIGURE 2** Schematic map of freeway system (not to scale, only eastbound ramps shown).
The data-reduction procedure thus provided nine plots: one for each lane at each of the three locations. (See Figure 3; note the following in this figure: top row, Station 9; middle row, Station 7; bottom row, Station 4; left column, median lane; center column, middle lane; right column, shoulder lane.) These plots are intended to represent nontransitional (or equilibrium) operation. The analysis consists primarily of visual inspection and comparison of these plots in order to answer the three questions raised at the beginning of this paper:

1. Is the same general shape (an inverted V-shape) observed at all lanes and stations?
2. If the shape is the same, are the parameters (particularly maximum flow rates and critical occupancies) similar?
3. Are areas of sparse data (or discontinuities) present at all locations, and, if so, are they related in a sensible way?

The answer to the first question is that there is an overall consistency in the general shape of the relationships for two lanes, but not for the shoulder lane. The six plots for the middle and median lanes all show relationships similar to that at Station 4, median lane (Figure 1a): a well-defined uncongested regime, a sharp peak in flow, and a congested regime with a large amount of scatter in the data. In the three shoulder lane plots, the maximum observed flow rates are attained in the congested regime, and no distinct peaks are observed.

Qualitative comparisons were made of the parameters for these relationships. Six comparisons were made: one for each of the median, middle, and shoulder lanes comparing the relationships at the three stations along the roadway to find the effect of location; and one for each of the three stations, comparing across the three lanes. A summary figure was drawn for each combination to be compared, which indicates the outline of the general location of the data points for the relevant lane stations.

For the median lane, distinct differences exist between the flow-occupancy relationship at Station 7 on the one hand and Stations 4 and 9 on the other (Figure 4). One possible explanation for this behavior is that Station 7 is located in a secondary bottleneck (secondary in the sense that there is another one farther downstream that causes a queue as far back as Station 7 or farther.) Hurdle and Datta (7) have suggested that speed-flow relationships may differ in a bottleneck; the flow-occupancy curve may also be affected in some way. The maximum flows observed at Station 7 are 200 pcu/hr higher than those at the other locations. The occupancies corresponding to these maximum observed flows are also 2 to 4 percent higher, which necessarily follows from the higher flows and the similarity of the nearly linear relationship in the uncongested regime over the three locations. This similarity does not occur in the congested regime, where two separate curves operate, one for Stations 4 and 9 and one for Station 7. Both curves have a similar shape, but it would appear that after the road is congested, the mean flow rates for the secondary bottleneck are consistently 400 pcu/hr higher at any given occupancy.

The middle lane data exhibit a different pattern in that the three stations appear to have consistent speed-occupancy patterns (Figure 5). For both the uncongested and congested regimes, the data for all three stations lie almost directly on top of each other, clearly suggesting that a single relationship can represent them all well. No obvious reasons for this difference from the median lane occur to the authors.

Three observations can be made from consideration of the plot for the shoulder lanes (Figure 6). First, the uncongested regimes for Stations 4 and 7 coincide, but Station 9 has generally lower flows for any given occupancy. (The Station 9 shoulder lane is heavily affected by an entrance ramp merging immediately downstream.) Second, in the congested regime the data for the three stations are again separated, somewhat as in the median-lane comparison. For any given occupancy, Station 7 has the highest flows, Station 9 the lowest, and Station 4 is in between. Third, all shoulder lane plots exhibit higher maximum flows in the congested regime than in the uncongested regime, which is contrary to their definitions. (This may be a consequence of increased flows on the metered ramps, leading to increased main-line flows as the system becomes more congested.)

Two plausible interpretations of the flow-occupancy relationship are consistent with the second and third of these observations. Either there is an underlying relationship in the shoulder lane similar to that in the middle and median lanes,
with a distinct peak in the relationship even though it has not been observed in the data, or the relationship in this lane is fundamentally different. In the first case, the gap in the data, or the unobserved portion of the curve, includes all of the operation around capacity. The alternate explanation implies that what the authors have termed transitions in fact represent operations at the top of those curves, near the capacity for shoulder lanes.

The resolution of this problem may be aided by a comparison across lanes at each station (Figure 7). The first observation is that in the uncongested regime the slope of the relationship becomes increasingly steep from the shoulder to
The net result is to leave the question about the shape of the shoulder-lane plot unresolved. The inverted V-shape found in nonshoulder lanes may not apply to the shoulder lane, even ignoring Station 9 as being too close to an entrance ramp to include in any generalization. Whether it is the same curve depends ultimately on the missing data, or the gaps in the data, which can only be resolved by looking at data from other freeways and locations.

The third question addressed by the analysis dealt with the question of discontinuities in the functions. The authors have suggested that the discontinuities others have hypothesized for relationships describing freeway data are unnecessary for understanding the nature of operations. Further, the wide variety of different functional forms that have been calculated by proponents of discontinuous two-regime models [e.g., Payne (9), Ceder (10), and Easa (11)] may well have obscured the systematic variation that is to be expected from the queueing process that takes place on freeways.

One should expect to find different apparent discontinuities in the data, depending on the location relative to high-volume entrance ramps and on the flow rates on the main line when a queue reached that location. The results support this argument. For example, it is clear that the concentration of data in the congested regime for Station 4 occurs at flows lower than for the other stations (Figure 3). This is reasonable because the metered entrance ramps at Mississauga Road add 1,000 or so vehicles per hour to the flow. Station 4, when congested, must move fewer vehicles than does Station 7 or 9. The consequence of this for the data is that Station 4 will have sparse data in the upper portion of the congested regime in a flow-occupancy curve. In a speed-concentration curve such as other researchers have focused on, the sparse data will occur somewhere in the middle of the curve, which may lead one to infer the need for a discontinuous function. However, that sparsity of data need not
DISCUSSION OF RESULTS

Two important results arise from the analyses, in addition to the support provided for the authors' earlier arguments against the necessity for discontinuous functions. First, the inverted V-shape for the flow-occupancy relationship is found for both of the nonshoulder lanes at all stations examined. Whether the shoulder lane conforms to this same pattern is still an open question. Second, the parameters of the flow-occupancy relationship for a freeway differ across stations along the roadway. In particular, maximum flow rates were higher at Stations 7 than at 4 and 9 (median and shoulder lanes), and this appears to be reflected in a slightly shifted curve in the congested regime. The analysis was begun with the hope of demonstrating that the results obtained earlier for one lane at one station could be generalized. The current results suggest that simple generalizations will be inadequate; therefore, in this section the simplest way the authors can think of to deal with the differences observed is considered.

One relatively simple explanation was rejected, namely, that the differences arise from differences in the geometric characteristics of the highway. Lane width across all four stations is almost identical. The shoulders are wider at Station 4 than at Station 7 or 9, but that would suggest that Station 4 should have the highest maximum flows. Likewise, Station 4 is on a straight level section of highway, whereas Station 7 is on a small grade and at the end of a horizontal curve. Again, this would suggest that Station 4 should have the higher flows. Station 9 is on a vertical curve, so it is surprising that in several respects it is similar to Station 4. Geometric characteristics do not appear to provide the explanation for the differences across stations.

FIGURE 7 Comparison across lanes at each of three stations: Station 4, Station 7, and Station 9.
With the obvious solution discounted, what is left is the idea that the high flow rates and shifted congested regime at Station 7 are due to the station's location in a secondary bottleneck. This idea about bottleneck flows deserves a careful discussion. The authors are not convinced that it is the explanation for the observed differences, but any other possibilities have been ruled out.

The origins of the hypothesis come from a paper by Hurdle and Datta, in which they observed some surprising results in speed-flow data (7). They hypothesized that very high flows (above 1,850 pcu/hr/lane) are associated with a slight drop in speed, and that these speeds and flows occur in a bottleneck when the vehicles are “being discharged from an upstream queue.”

However, in the Queen Elizabeth Way system careful examination of daily data from the secondary bottleneck and from Station 6 upstream reveals that Station 7 becomes congested because of a queue from the primary bottleneck before a queue can form from the secondary bottleneck. It is therefore not possible to be certain that the traffic flow relationships are different in the congested regime because they are in a secondary bottleneck location that has been fed by a queue. Nevertheless, at Station 7 there is a situation in which it is suspected, from the geometry of the situation, that a secondary bottleneck exists. The extra queue upstream of Station 7 is caused by heavy mainline flow merging with two heavily-used entrance ramps from Mississauga Road (which are metered at rates of approximately 12 vehicles per minute during this period), thereby making Station 7 act in part as a bottleneck.

Two possible explanations are offered for the observed differences in the flow-occupancy relationship in this situation. The first is the presence of metered ramps. The logic for ramp-metering systems is to control entering traffic to maintain a relatively smooth traffic flow downstream. It is entirely possible that these results are an eloquent demonstration of just how much ramp-metering systems have accomplished. The data in this paper, and in Hurdle's, come from a functioning freeway management system. It may well be that operating characteristics have changed, particularly around capacity and in the congested regime, in which ramp-metering systems are most often used. In other words, there is the possibility of a different relationship in the congested regime because of intervention by traffic engineers.

Another possible explanation is that any time the congested flow is fed by an upstream queue, the flows will be higher than if it is not. In other words, the shift in the flow-occupancy function occurs because the vehicles are coming from an upstream queue. This extends Hurdle and Datta's reasoning to a different relationship than they had proposed, but appears to be in agreement with their suggestions. The underlying mechanisms behind this behavior are certainly not obvious. Perhaps, after experiencing stop-and-go conditions combined with merging vehicles, drivers will take advantage of an uninterrupted stretch of roadway. It may therefore be a measure of the drivers' increasing frustration that higher flow rates and occupancies are found at Station 7 than at Station 4 (which has better geometrics, but has not been preceded by a merging section).

Whichever explanation is correct, the generalization of these results is shown in Figure 8. In the uncongested regime, there is a well-defined, nearly linear function, but with lower slopes and lower maximum flows as one moves from the median to the shoulder lane. In the congested regime, there is a broad band of possible data points, which should not be represented by a single line. Despite the scatter, in a bottleneck the curve is shifted up and to the right. The authors are unable to draw the shoulder lane plot near its capacity.

**CONCLUSIONS**

In this paper, additional support has been provided for some tentative new ways of looking at relationships between variables describing freeway operations. The main conclusions are as follows:

1. An underlying inverted V-shape for the flow-occupancy relationship has been confirmed at various locations along a freeway for both of the nonshoulder lanes.

2. For the shoulder lane, it is not clear whether the inverted V-shape holds or whether there is a different, inverted U-shape relationship, such as that shown in Transportation Research Circular 281 (12, Figure 3-3).

3. Definite differences have been found in the parameters of the flow-occupancy relationship that appear to be due to lane and location. In particular, the following trends were noted.

   a. Locations in a secondary bottleneck exhibit higher maximum uncongested flows and occupancies and higher flows at all occupancies in the congested regime.

   b. Within a single lane, the uncongested regime is well defined, but the flow-occupancy slope becomes increasingly steep moving across lanes toward the median lane.

   c. Proximity to an entrance ramp decreases flows at all occupancies in the shoulder lane.
4. Discontinuous relationships are not necessary to describe the data obtained from freeway operations. Better sense may be made by assuming continuous relationships, and trying to explain sparseness of data by the nature of operations on the facility.

These conclusions are admittedly tentative. To verify or discredit them, data need to be obtained from other freeway systems. In particular, data are needed from various locations upstream from a bottleneck, locations that experience different flow rates at the time they switch to congested flow. In addition, data on the shoulder lane in the bottleneck would help to establish the nature of the flow-occupancy relationship for that lane.

These findings, if confirmed, have important implications in many areas of theoretical and practical traffic engineering. The most important of these is that one needs to pay attention to the location of the data acquisition in order to make sense of the results. As freeway management systems become more common, this requirement becomes increasingly important.

ACKNOWLEDGMENTS

The data used in this analysis were provided by the Ontario Ministry of Transportation and Communications, and were made workable on the McMaster University computer facilities by Geddes Mahabir several years ago. Funding for this project was through the Natural Sciences and Engineering Research Council's Undergraduate Summer Research Award programme. The authors should also acknowledge the helpful and thought-provoking comments of two reviewers of a previous paper.

REFERENCES


Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.