Transpn. Res.-A Vol. 20A, No. 3, pp. 197-210, 1986
Printed in Great Britain.

0191-2607/86 $3.00+.00

EMPIRICAL ANALYSIS OF FREEWAY FLOW-DENSITY RELATIONSHIPS

FRED L. HALL, BRIAN L. ALLEN and MARGOT A. GUNTER
Department of Civil Engineering, McMaster University, Hamilton, Ontario, Canada

(*Received* 2 *April* 1985)

Abstract—Many researchers have reported on the occurrence of gaps in freeway speed-density and flow-density data and have suggested that discontinuous functions are necessary to properly describe "observed" traffic behavior. This paper investigates the flow-occupancy (spot-density) relationship using an extensive data set collected on the Queen Elizabeth Way in Ontario. Daily time-traced plots of 5-minute average flow rates versus occupancy were analyzed. Results indicate that there is another interpretation of gaps in data, which does not imply a discontinuous function, but rather, an inverted V shape (continuous, but not continuously differentiable). Three conclusions were reached: (a) it is essential to provide details of data collection locations if one is to know whether a particular pattern in resulting data represents a "true" relationship, or just the specifics of a particular place; (b) there are clear advantages to examining daily time traces of traffic behavior, rather than relying on scatter diagrams of numerous days of accumulated data; and, (c) previously documented arguments for a discontinuous flow-occupancy relationship do not seem convincing, because knowledge of daily operations at a particular location could easily explain the occurrence of gaps in the data.

INTRODUCTION

Over the past several years, authors such as Ceder (1976), Koshi, Iwasaki and Ohkura (1983), and Easa (1983) have commented on the occurrence of gaps in freeway flow-density and speed-density data. It has been suggested on the basis of such gaps that discontinuous functions are required to properly describe traffic behaviour. Specifically, those gaps usually occur in high-flow ranges, at speeds normally associated with near-capacity operations, and are typically located in the congested regime. A good example of this observation is presented in Fig. 1.

It is our contention that there is another interpretation which is also consistent with such data, but which does not imply a discontinuous function. Further, conventional wisdom about the nature of traffic operations on a freeway should lead one to *expect* a gap in data collected at any one location, thus removing the existence of such gaps as sufficient justification for choosing functions that are discontinuous. We are *not*, however, arguing against the existence of two regimes of operation: clearly congested flow behaves differently from uncongested flow—but that does not necessarily imply discontinuity of the underlying functions, even though they may not be continuously differentiable.

Our argument in this paper is not a theoretical one. Instead, it is solely empirical, based upon analyses of an extensive set of data obtained from the QEW Freeway Surveillance and Control Project in Ontario (Case and Williams, 1978). The omission of theoretical rationales (hydrodynamic, car-following, etc.) from the paper is not meant to imply that they are unimportant. Rather, it is our feeling that much better data on traffic flow are now available than were used in the early development of those theories. It is advisable therefore to look at what the data actually show in order to establish macroscopic freeway traffic flow relationships. This approach is somewhat in contrast to one that *a priori* assumes that relationships can be identified, for example, by integrating even good approximations of microscopic relationships. Also, it is entirely possible that the macroscopic relationships have changed over the past 30 years as a result of changing driver experience, highway design, vehicle mix, and vehicle weight-to-horsepower ratios.

In the process of developing our own arguments, and tracing the basis for others' ideas, the need for considerably more detail about researchers' data became obvious. For example, where on the roadway were data collected: in a bottleneck or if not, how far upstream of one? How were density data obtained: calculated from speed and flow data, or actually measured? If measured, then over what distance? The paper by Koshi *et al.* (1983) is a good example of thorough documentation of data. Others, regrettably, are not usually so complete.

The need for this kind of detail will become obvious as the discussion develops. The next section contains a discussion of some relevant work to set the background for the current analysis. The nature of the data utilized is briefly described, and then a fairly detailed discussion of the analytical procedures is included. The final section presents the conclusions.

BACKGROUND

Koshi *et al.* (1983) discuss flow-density scatter diagrams for locations 0.4 and 5.7 km upstream of a bottleneck, and assert that a discontinuity between the uncongested and congested flow regions "becomes obvious some distance upstream of the bottleneck, while the boundary between the regions is not clear immediately upstream of the bottleneck" (p. 406). They further assert that "the two regions of flow form not a single downward

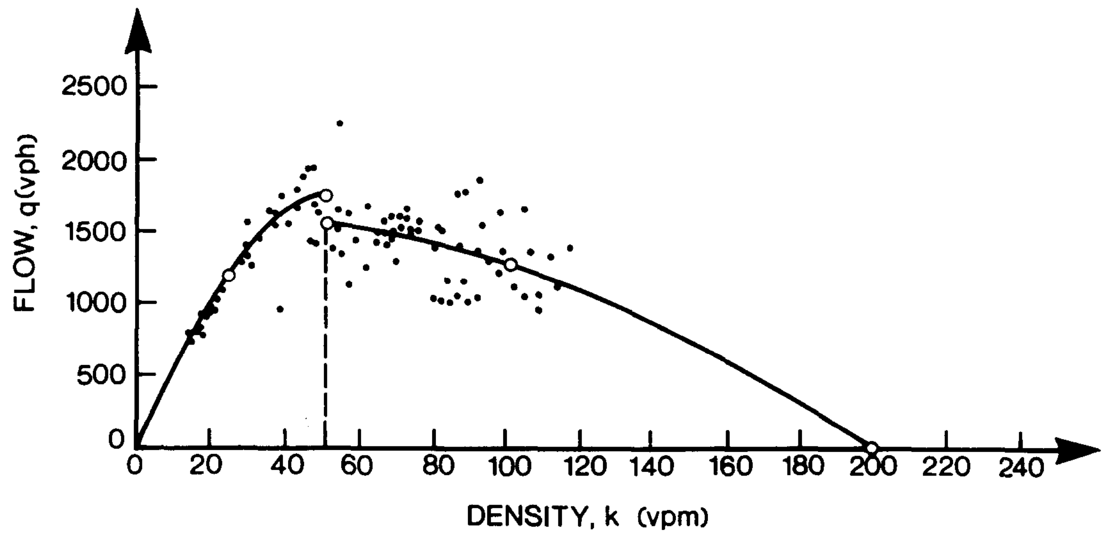

Fig. 1. Typical flow-density data with gap and assumed discontinuous function, from Easa (1983). Redrawn with permission of the author.

concave curve. . . but a shape like a mirror image of the Greek letter lamda [sic] (λ)" (Fig. 2). Both of the assertions are indeed consistent with the data they present, although perhaps a bit exaggerated. There are *fewer* data points in the area they note, but not a *complete absence* of data. The upper arm of the lambda is quite short, and is really visible only for the median lane. Nonetheless, they have described a reasonable way of viewing the scatter of data. In addition, they have documented the data well: 1-minute data acquisition; density calculated from speed and flow data (or from occupancy using a regression result based on speed-flow derived density); and 168 hours of data. Further, they have noted and investigated the implications of this interpretation for car-following models and for wave propagation.

Easa's paper is primarily concerned with a procedure for estimating two-regime models, rather than with specific data, but he does provide an example application of the procedures. His comments about the data were important to our thinking, and thus provide some of the background for this work. In addition, tracing the history of ideas from his references led to some interesting findings. Easa presents both speed-density and flow-density diagrams, containing about 100 data points each. Other than stating that the data are from "the Eisenhower Freeway at Harlem" and that the freeway was built in the 1950s, he tells us nothing about the data acquisition. He notes that "the data exhibit some discontinuity" (p. 33), especially in the flow-density graph, and cites a Gazis, Herman and Rothery paper (1961) as providing "an excellent discussion of the possible reason for the existence of discontinuity." Inspection of his flow-density graph (Fig. 1) suggests that one can as easily see Koshi's reverse lambda shape as the curves Easa fits to the data. Thus, already (i.e. before introducing our interpretation of such data) there are two quite different views of the type of curve to fit, both of which are reasonably compatible with observed data.

Easa's reference to the Gazis *et al.* discussion led us back through that paper to several early ones on this general topic, out of which arose a possibly very important point. We offer it here for consideration by others who have longer experience with this topic then we have, and as an illustration of the central importance of specifying the source of data precisely.

The Gazis *et al.* paper reproduces (in their Figure 8) a data set used by Edie (1961; his Figure 6), in order to compare Gazis' equation with Edie's for the same data. Edie in turn cites Greenberg's (1959) paper as the source of his data. Greenberg, finally, provides a reasonably complete description of the source of the data. Individual vehicle speeds across a short section and headways between vehicles were obtained in the Lincoln Tunnel in New York City. The data were then separated into speed classes, with a 1 mph level of resolution, and headways were averaged within each class. Presumably, although this is not stated in the paper, flow rates and densities were calculated from the headways. These last assumed steps are not nearly as important as the grouping of data on the basis of individual vehicle speeds. What this procedure means is that there is no guarantee that the vehicles that have been grouped together were in fact travelling together. Indeed, they may have been several minutes (or hours) apart on the roadway.

The implication of this procedure is that the flow-density data which resulted, and that all three papers were trying to explain, represent conditions which probably never occurred on the roadway, because it is unlikely that the specific densities could be observed at the same time as the flow rates they were paired with.

There are two important points to be drawn from this retracing of developments: first, there is a good chance that three important papers in this field were attempting to explain traffic operations that never actually occurred, and second, it was possible to identify that fact only because Greenberg documented the data acquisition and reduction sufficiently.

Interestingly, Greenberg's paper also discusses a sec-

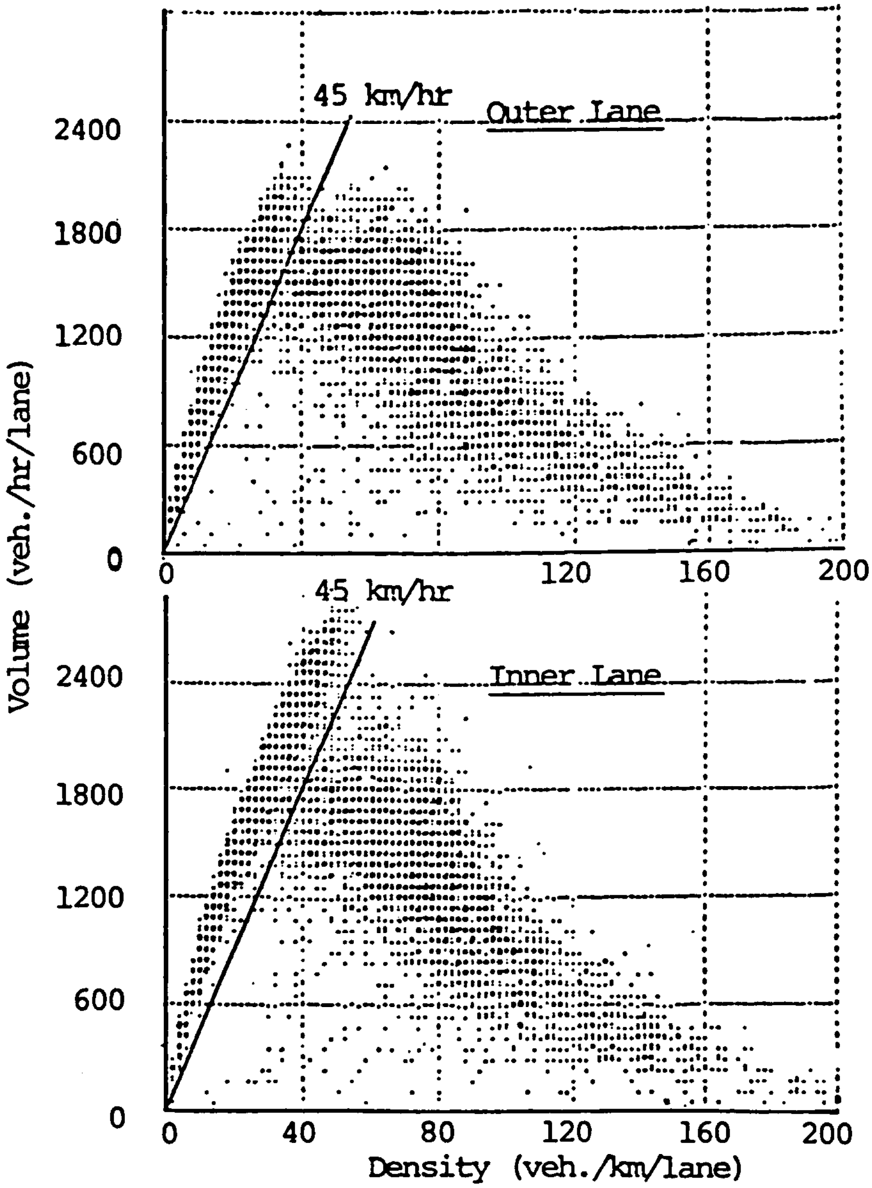

Fig. 2. Typical flow-density data (5.7 km upstream of bottleneck), from Koshi *et al.* (1983). Reproduced by permission of the author. (The 45 km/hr line appears in the original, but is not related to the current discussion.)

ond data set, which subsequent researchers appear to have ignored. These data, from the Merritt Parkway, were collected and averaged over 5-minute intervals, and show two distinct clusters, with a sizeable gap between them. This data pattern seems to us to be exactly the type one should expect, given the conventional understanding of how congested flow conditions can arise: upstream of a bottleneck a queue forms, the trailing edge of which moves upstream at a rate dependent on demand/capacity conditions. When this queue tail arrives at any upstream location, operation on the freeway "jumps" from the uncongested branch (e.g. of a speed-flow curve) to the congested branch, maintaining approximately the same flow rate. This procedure has been sketched and discussed many times; we include a schematic of the process in Fig. 3 in order to focus on the data one can therefore expect to obtain. The curves have been drawn intentionally to emphasize the potential for missing data. Obviously, if the underlying functions are different, the location of the gap in the data will be different. The important point is that there will be an absence of some data unless the facility is operating at or near capacity when the transition to congested flow occurs. The size of the gap will depend therefore on the pattern of demand over time on the mainline and on any ramps between the bottleneck and the data collection location. Consequently, gap size could depend on how far upstream of the bottleneck the data were collected. It is for this reason that it is very important to specify when and where on the roadway the data are obtained. This explanation pro-

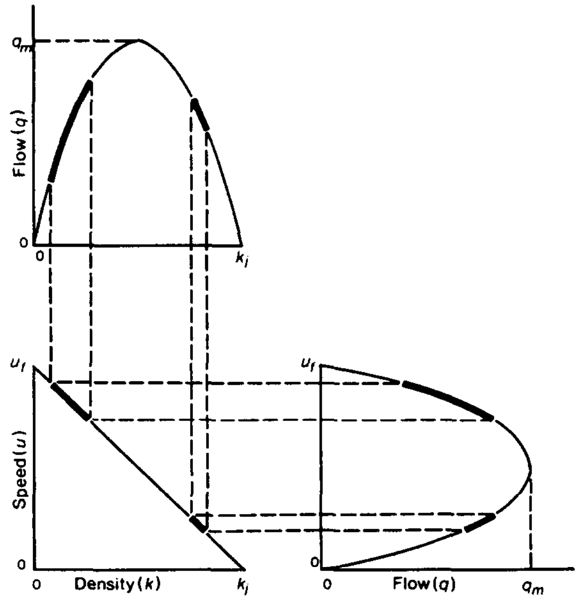

Fig. 3. Representative functions for speed-flow-density, showing expected data acquisition at one station upstream of a bottleneck (heavy solid lines), and gaps in that data (light line).

vides a plausible reason for the Koshi *et al.* result that the gap in their data is more visible 5.7 km upstream than 0.4 km from the bottleneck.

If a gap in the data is to be expected, then it follows that techniques that use the location and size of such a gap to assist in fitting curves are at best misleading, and at worst wrong. A good example of such a technique is Ceder's (1976) extensive effort fitting two-regime models to 45 different data sets. Each data set required a different function, and in many cases the resulting functions were quite distinctly different. No underlying theoretical reasons were given to suggest why they should be so different. That result contravenes the principal of parsimony in scientific theory selection. Even though the same functional form is used, the many widely varying empirical outcomes cannot be explained. A simpler outcome would certainly be more appealing.

The object of this paper, then, is to seek such a simpler outcome for speed-density relationships. The resulting model should be able to explain Koshi's reverse-lambda data pattern as well as the gaps observed in so many data sets. In addition, it should subsume the single-regime models that currently exist. This is a lot to expect from an investigation of one data set, but an awareness of all of these theories may help to focus that investigation properly. In the final analysis, however, the best we will be able to do is put forward a proposal that is not internally contradictory. The inevitable question as to whether one chooses to accept it, or chooses to remain an advocate of discontinuous two-regime models will not be settled by reference to data alone, because the data will support either reading.

DATA

The data used in this study come from the Ontario Ministry of Transportation and Communications (MTC) Freeway Surveillance and Control System (FSCS) as described by Case and Williams (1978). The data were collected in 1979–1980, and at that time the system operated on a 5 kilometer section of the Queen Elizabeth Way between Oakville and Toronto, where morning commuter traffic created congested flow conditions on the three eastbound lanes (Fig. 4). The system comprised nine mainline detector stations with induction loop pairs in each of the three directional lanes, ramp metering on five entrance ramps, and closed-circuit television surveillance cameras operated from a control centre. The limiting capacity restriction (bottleneck) was immediately downstream of Station 9, where heavy Highway 10 entrance ramp traffic merged with the three through lanes.

The data obtained from that system span 8 months from July 1979 to February 1980, and consist of approximately $2\frac{1}{2}$ hours of collection during weekday morning peak periods. From each pair of induction loops, occupancy at the downstream loop, occupancy at the upstream loop, volume of vehicles longer than 7.6 m, total volume of vehicles and average speed were obtained for each lane in 5-minute intervals. The data were stored on magnetic tapes, and a complete log of daily weather conditions and incidents (accidents, breakdowns, etc.) was available.

The data used in this analysis were carefully selected from that rather extensive set, in order to maximize the opportunity to identify the "true" underlying nature of traffic flow. Particular attention was given to daily operating conditions and to station location with respect to the bottleneck. First, to eliminate undesirable operating effects, all data were deleted for days during which inclement weather or traffic incidents were reported. The remaining 71 days of data then represented traffic flow under "ideal" conditions. Second, sample data from the 71 days were reviewed for each of the nine station locations to determine the range of data available, that is, low to high flow rates, speeds and occupancies. It was clear that the most complete range of values was present at Station 4, located approximately 4 kilometers upstream of the bottleneck. At that location, flow rates and speeds were quite high immediately preceding the onset of congestion, moved through lower speeds and high occupancies, and generally returned to high speeds at lower occupancies by the end of data collection. All other stations did not exhibit such a wide range of operating conditions, and thus were not considered as suitable for identification of the full relationship between the variables.

In our investigation of speed-flow-density relationships, no attempt was made to convert raw occupancy data to the more commonly used density parameter. Occupancy is a point measure of the percentage of time a detector loop is occupied by a vehicle, and more properly corresponds to the point measures of speed and volume that were available. Density, on the other hand, requires measurement over a section of roadway in order to calculate the number of vehicles per kilometer. An additional reason for not using density is that the often used practice of converting measured occupancy to calculated density inherently creates an unnecessary degree of uncertainty as can be seen by the large scatter of data about

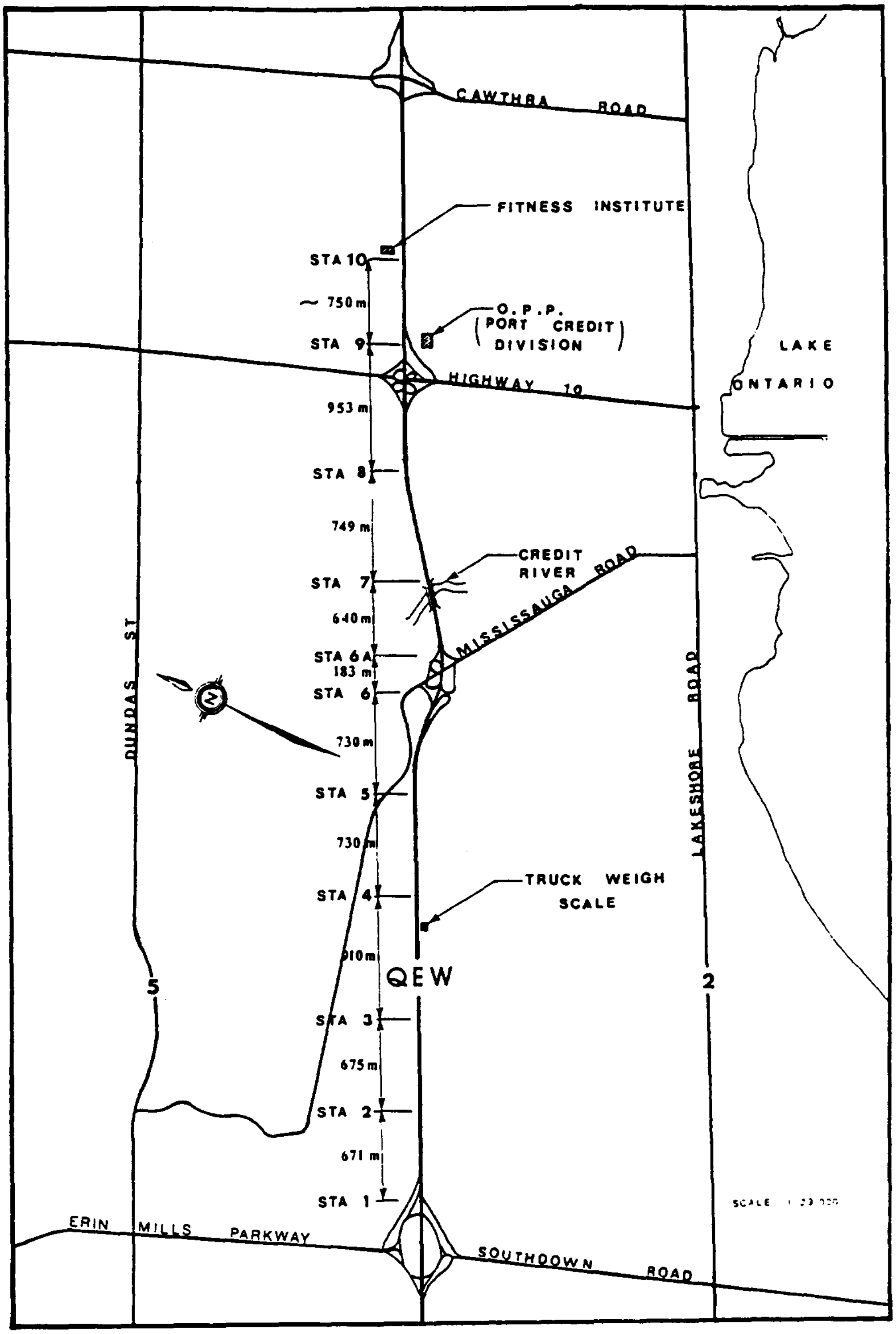

Fig. 4. Q.E.W. freeway surveillance and control site, from Mahabir (1981). Reproduced by permission of the author.

the Koshi *et al.* conversion equation shown in Fig. 5. Therefore, for pragmatic as well as fundamental reasons, actual measured occupancies were used so that the analyses that follow would not *a priori* be cast in doubt.

Finally, because previous work suggested considerable operational differences between lanes (Mahabir, 1981), that distinction was maintained. To allow comparisons between lanes, which had varying percentages of trucks on generally level terrain, all volumes were converted to passenger car units (pcu) using an equivalency of two pcu's per truck as suggested in TRB Circular 212 by Roess, Linzen, McShane and Pignataro (1980).

Data reduction

The final data set contained 71 ideal days, for which 5-minute average speeds, average occupancies (%) and total volume (passenger car units) were available, for each lane at Station 4. This section describes the evaluation and manipulation of that data carried out prior to the actual analysis.

Our analytical approach is extremely simple, but with sound reasons. There are two key assumptions that underlie the main procedures. First, we assume that there is an identifiable pattern in the flow-occupancy data. Without this assumption, there is no reason to conduct any analysis. Although all researchers begin with this assumption, we go further: the underlying pattern may well be identifiable by inspection, at least in so far as one can identify what type of operation (congested, uncongested or transitional) is occurring for a given data point on a particular day. This leads to the second key assumption: scatter-plots of data are not adequate to show the nature of changes in operations, as they are reflected in flow-occupancy data. Instead, it is necessary to look at each day separately, and to connect the 5-minute data chronologically, to see how flow and occupancy values jointly change over time. These daily time-traced data are fundamental to the analysis, and require inspection 1 day at a time. The important difference between using a normal scatter plot and a time-traced plot is that in using a time-traced plot to visualize patterns, as much weight has been given to the lines as to the data points they connect.

Taken together, these two assumptions imply that a thorough screening of the data should take place prior to quantitative analysis. There are, however, some clear disadvantages to a subjective screening (or selection) of data in this fashion. The most obvious of these is the potential for bias in the resulting data. That objection is best rebutted by the outcome, so discussion of it will occur within the context of the actual selection. The second disadvantage is a practical one: it is tedious and time-consuming to inspect the time-trace for every day for every lane. In recognition of this problem, the decision was made to limit the analysis to one lane only (the median lane—i.e. the left-most lane, closest to the median divider) at Station 4, in order to see what could be developed from the use of such a screening procedure. (The median lane was selected because it contains the highest flow rates of the three lanes.)

The primary screening of the daily speed-occupancy data for Station 4, median lane, proceeded as follows.

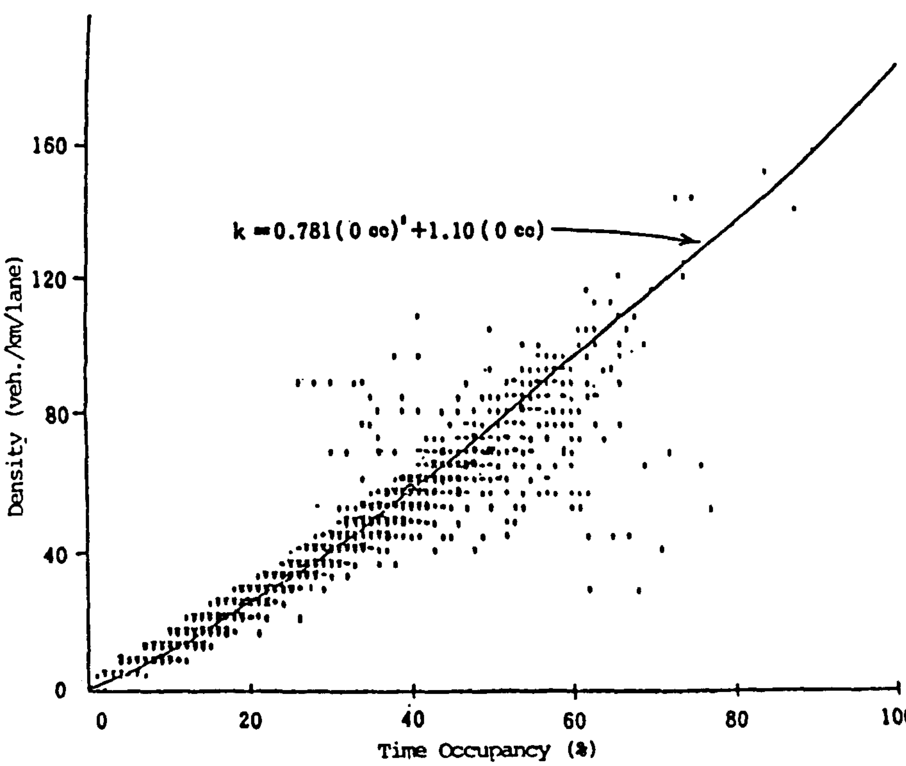

Fig. 5. Density-occupancy data, from Koshi *et al.* (1983). Reproduced by permission of the author.

First, time-connected scatter-plots were produced for each day. When these were superimposed on each other, a very clear overall pattern emerged. The next step was to compare each day to this overall pattern, and at the same time to consider the raw data that led to the plot. Two principal criteria were then applied to each day's data before these were accepted for further analysis. First, was the data acquisition equipment functioning properly for more than one-half hour? Twenty days were rejected because the data covered less than one-half hour. Second, was the day's pattern consistent with the overall pattern observed? This did not mean the day matched the full pattern, only that the data fell on the overall pattern rather than off it. Here is where the risk of biased selection clearly enters. In fact, only six of the scatter-plots were identified as not meeting this criterion, and for five of the six, inspection of the raw data showed an unusual malfunctioning of the data acquisition system: the downstream detector reported only 30% of the occupancy of the upstream detector, where normally they would differ by less than 10%. For the sixth day, the two detectors were consistent, but both reported occupancies very much lower than normal for the time of day. Given the evidence of other detector malfunctions on 5 days, and the discrepancy in the pattern of the data, we decided to delete this day also. Thus, in practice, a subjective decision was made for only 1 day. For the rest, the approach helped to identify days to inspect more closely. In all but one of the questionable days, there turned out to be a clear data collection anomoly that warranted deletion of the day.

This primary screening resulted in the original 71 days of "ideal" data being reduced to 45 days. During the screening procedures, it became apparent that there were only a small number of daily patterns. Examples of the main ones are shown in Fig. 6, where the arrow heads denote the path of increasing time. The most informative pattern (Fig. 6a) is one where the system was turned on early enough to observe increasing uncongested flow (very briefly) to rates around 2300 veh/hr, followed by a steep

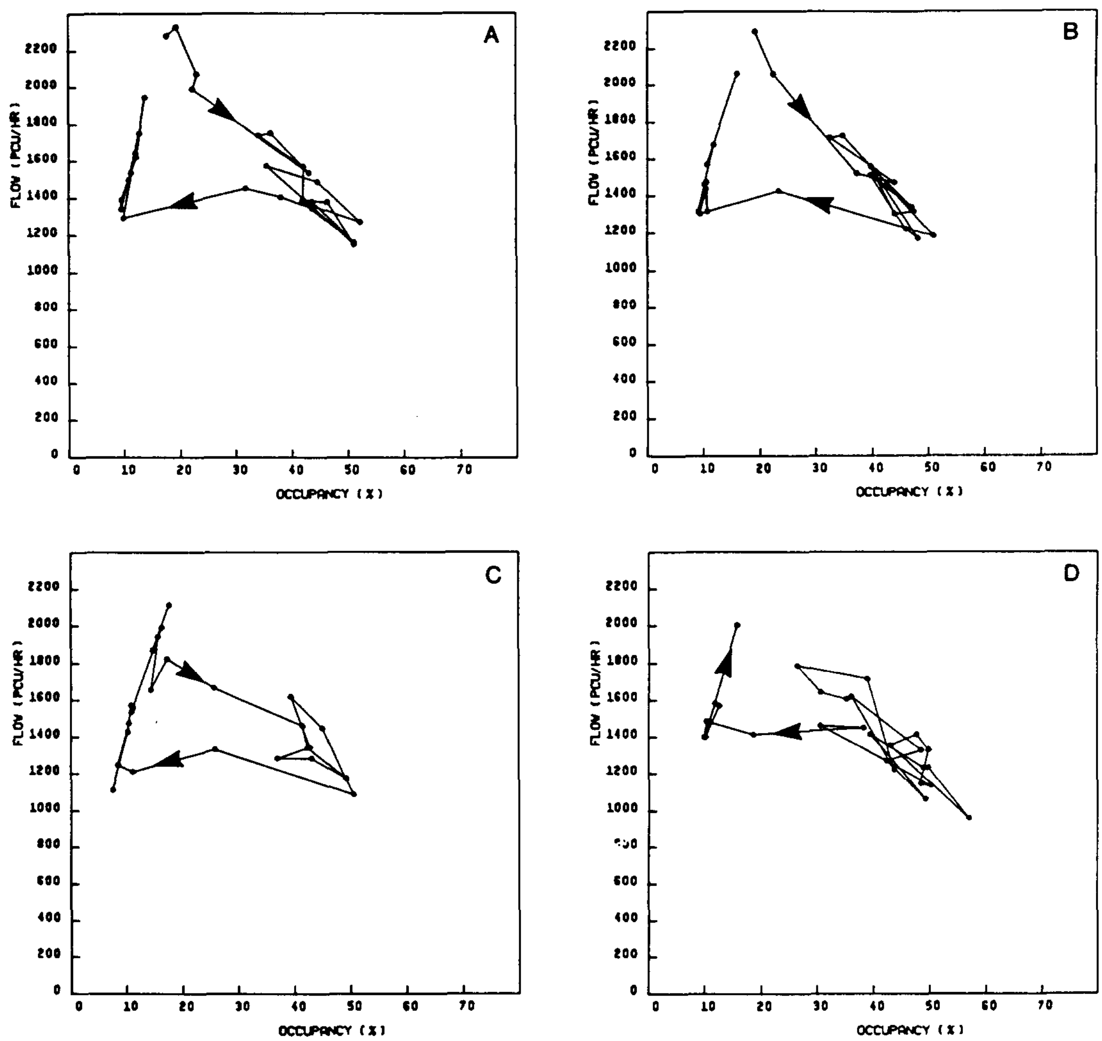

Fig. 6. Typical daily time-traced plots of flow-occupancy data: (a) high and increasing flow, leading to an uncongested–congested–uncongested pattern; (b) high flow, congested–uncongested pattern; (c) lower flow, uncongested–congested–uncongested pattern; (d) congested–uncongested pattern.

drop down the congested flow branch of the function, then later an obvious jump across to the uncongested branch, and a subsequent increase in flow prior to termination of data collection. Several days, such as that shown in Fig. 6b, did not show the early building of flow, but did have the first data point at very high flow rates, followed by the same general pattern shown in Fig. 6a. Only a few days showed the pattern that conventional wisdom leads one to expect (Fig. 6c), with a jump from the uncongested to the congested branch at roughly constant flow rates, and the remaining pattern as shown in Fig. 6a. Many more days began with operations already on the congested branch, and then later jumped back to uncongested operation as shown in Fig. 6d. There were also some days for which the data remained on only one branch or the other, and a few days with more than two transitions between branches.

One result of looking at these daily plots closely is to reconfirm the need for detail about data acquisition. For example, a scatter-plot from days with data like those shown in Fig. 6d would clearly lead to the interpretation of a gap in the data, or a reverse lambda. Yet in the case of the QEW data it is clear that such a pattern arises only because data acquisition began after the transition to congested flow had already occurred; the upper portion of the apparent lambda shape comes from later operation.

When all of the data points from the acceptable days are superimposed on one scatter-plot, the resulting pattern (Fig. 7) is very similar to that shown by Koshi or Easa. There is a very clearly defined uncongested branch, a small amount of widely scattered data for occupancies in the 20–30% range, and a reasonably well-defined congested-flow branch above 30% occupancy. Further, a best-fit line drawn through the congested flow branch would intersect the uncongested branch, resulting in Koshi's reverse lambda shape.

However, because of the detailed inspection of the daily time-traces, two important insights have been gained that argue against such a discontinuous function. First, it is known that the lack of data at occupancies of 20–30% is a result of the timing of data acquisition and the types of transitions from uncongested to congested flow occurring at the site. This implies that at different locations and perhaps for different data collection schedules, the location and size of the gap in the scatter-plot may well be different. Second, and perhaps more important, many, if not all of the data points that do occur in that area (20–30% occupancy) are produced by days in which operations move from very high flows into the congested regime (such as Figs. 6a and 6b). Sometimes the operations do not remain in that region for very long (e.g. in Fig. 6b, operations are there for only 5 minutes). On other days, operations remain at high flows, with what appear to be congested occupancies for longer periods (e.g. in Fig. 6a, flows are at 2000–2100 pcu/hr for occupancies clearly above 20% for at least 10 minutes).

The general tendencies of overall daily operations can

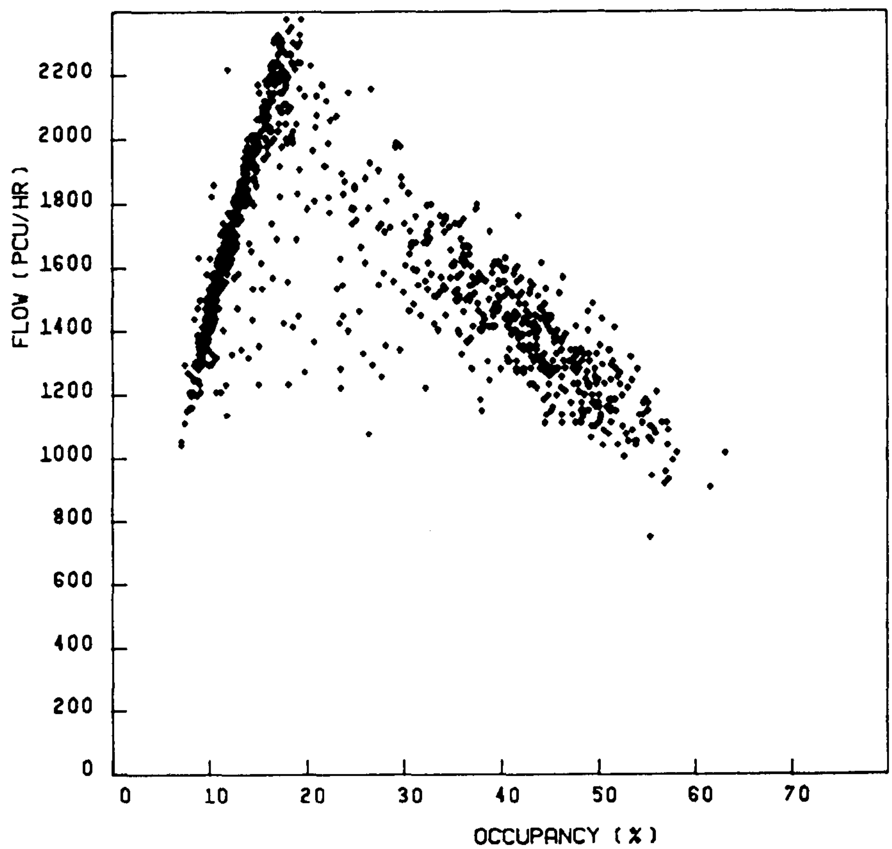

Fig. 7. Flow-occupancy scatter plot of 45 acceptable days, Q.E.W. data.

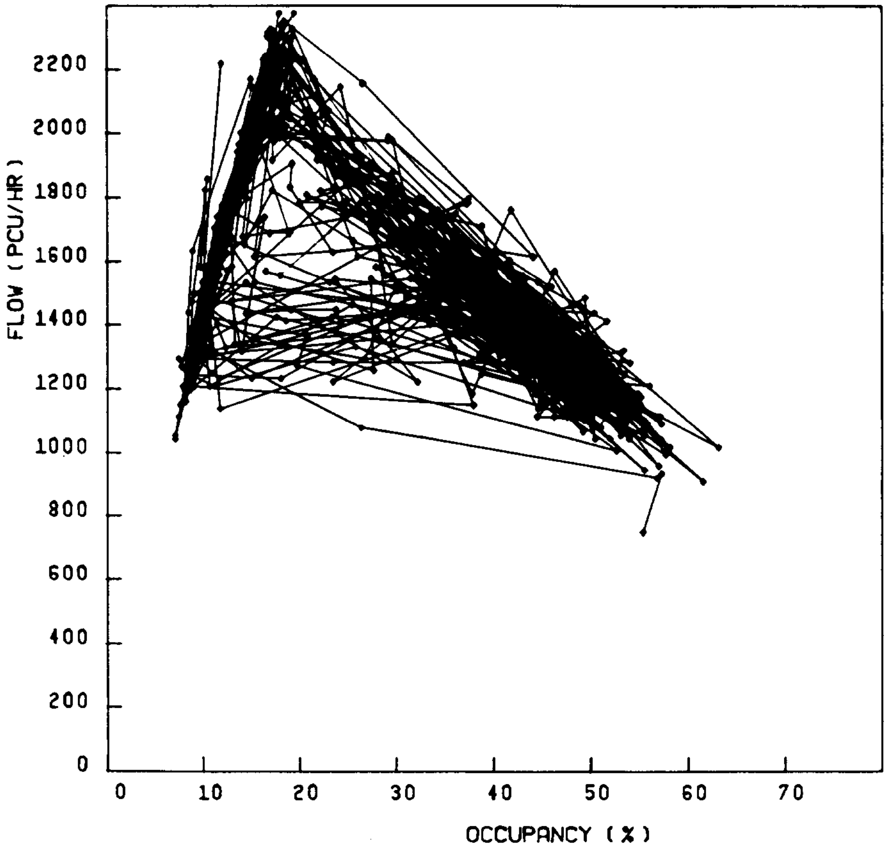

Fig. 8. Time-traced plots of flow-occupancy data for 45 days, Q.E.W.

perhaps best be seen in Fig. 8, which overlays the time-connected traces for all 45 of the days. Clearly, there is considerable movement of operations *through* the area that would normally be considered the gap in the data, as well as a number of points showing operation *within* that area. The fact that 20 of the 45 days show operations in this area leads us to attempt to construct a continuous function through these data, rather than to assume a discontinuous function is necessary just because the data are sparse. The following section contains a description of the analysis used to investigate potentially acceptable relationships.

Analysis

The analytical procedures for constructing a function or functional shape require some deliberation. Standard procedures, such as those adopted by Easa and many others, have *not* been used in this work for two fundamental reasons. First, as already stated, we are not convinced that a gap in the data necessarily implies a discontinuous function, and the conventional approach assumes potentially discontinuous functions. Equally important, but more practical, is the fact that the parameters for which values need to be specified by inspection are precisely the ones that we are unable to identify from the data: for example, jam density, free-flow speed, optimal (i.e. critical) density, and maximum flow.

Regression analysis (linear or nonlinear) also seems inappropriate, in that it has never been clear which variable is dependent and which is independent. Clearly, different curves would result depending on the choice made. Further, for a piece-wise regression (e.g. two-regimes), it would be necessary to specify the breakpoint between regimes (critical density), which is unknown.

Having rejected the sensible sophisticated procedures for placing a line through the data, we resorted to the simplest of procedures: averaging one variable for given values of the other. Because flow might be a single-valued function of occupancy, whereas clearly occupancy takes on two values for many values of flow (if not for the whole range), it made sense to start by averaging flows for each value of occupancy. Optimistically, it was even possible that this procedure would help to identify the critical occupancy value that distinguishes the two regimes.

The result, shown in Fig. 9, contains several useful findings. First, the mean pattern shown by the connected points is quite well defined for all of the uncongested regime, and for the congested regime above 30% occupancy. Second, the variation around the mean (standard deviation) is much greater for congested flow than for uncongested flow. Third, the area of greatest variation is the troublesome region comprising occupancies from the high teens to around 30%. This approach has not resolved the question of critical occupancy, but it has helped to narrow the likely range of values to 17–20%, where the variation in flow rates suddenly gets quite large.

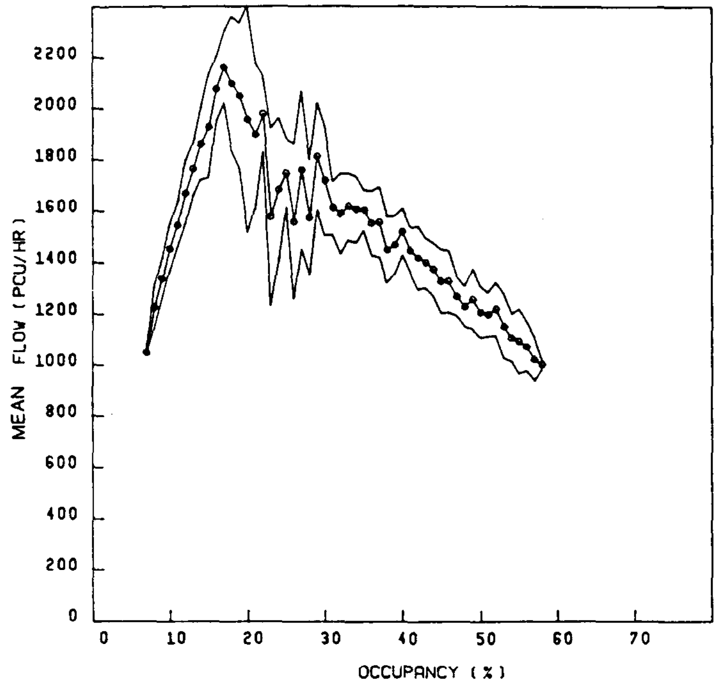

Fig. 9. Mean flow versus occupancy for 45 days, Q.E.W. data.

Before turning to averaging occupancies for given values of flow rates, however, one further improvement to the data set can be made, stemming from knowledge of the daily patterns. In all likelihood, much of the data scatter in the 15–30% occupancy range is due to inclusion of data points that represent part of the transition back from congested flow to uncongested flow. If these data were removed, the standard deviations around the mean, and even the abrupt variations in the mean, might well be improved. Here again, data screening might be thought to bias results. However, the rationale is clear: there is no justification for including a data point from a transition in an average meant to represent operation on a branch. For example, Fig. 6b shows one point at an occupancy of 23% and a flow rate of 1400 pcu/hr that is, in Fig. 9, being averaged with congested branch flows of 1800–2000 pcu/hr. Clearly that is inappropriate and therefore data must be carefully screened to remove such points.

In the 45-day data set, there were many such cases; there were also cases that were less obvious. When in doubt, we chose to leave a point in the data set rather than unduly reduce variation. In total, 17 days were deemed to show clear transitional points, with data lying between the congested and uncongested branches; a total of 30 points were deleted. The resulting occupancy-based mean flows are shown in Fig. 10. Variation is marginally reduced, and more importantly, mean flows at occupancies of 17–20% are no longer decreasing so rapidly, which is probably a better representation of the uncertainty in this area.

The first averaging approach has provided a reasonable shape for the flow-occupancy curve, but has not defined the critical occupancy precisely. Unfortunately, a value for the critical occupancy is necessary in order to calculate mean occupancy for a given flow rate. Clearly there are two values of occupancy for each flow rate, representing congested and uncongested regimes, and it is unreasonable to average uncongested occupancies with those in the congested regime. To avoid this, a critical occupancy must be identified such that operations at lower occupancies represent only uncongested flow, and the reverse. Because the precise value of critical occupancy was not apparent from the previous plots, integer values within the feasible range were selected, namely 17, 18, 19 and 20%. For each value of flow, occupancies less than the selected critical value were averaged for the uncongested regime, and occupancies greater than the selected critical value were averaged for the congested regime. Flow ranges of 50 pcu/hr (e.g. 2000 ± 25 pcu/hr) were used. The resultant plots are presented in Fig. 11.

Although the four plots are similar in overall appearance, there is one very important distinction. For the 17 and 18% plots in Figs. 11a and 11b, maximum flow rates in the congested regime are higher than those on the uncongested branch. Because such traffic behaviour is not intuitively acceptable and seems contrary to definitions of congestion, it is apparent that the critical occupancy must be greater than 18%. Maximum flow rates for the two regimes exhibit the expected trend for 19 and 20%, as shown in Figs. 11c and 11d, with slightly lower maximum rates in the congested branch. Consequently, one could reasonably conclude that the critical occupancy was approximately 19 or 20% and that these plots could be used for further interpretation.

There are several aspects observed in both of these plots worthy of mention. First, the uncongested branch is virtually linear with a very narrow band of data scatter,

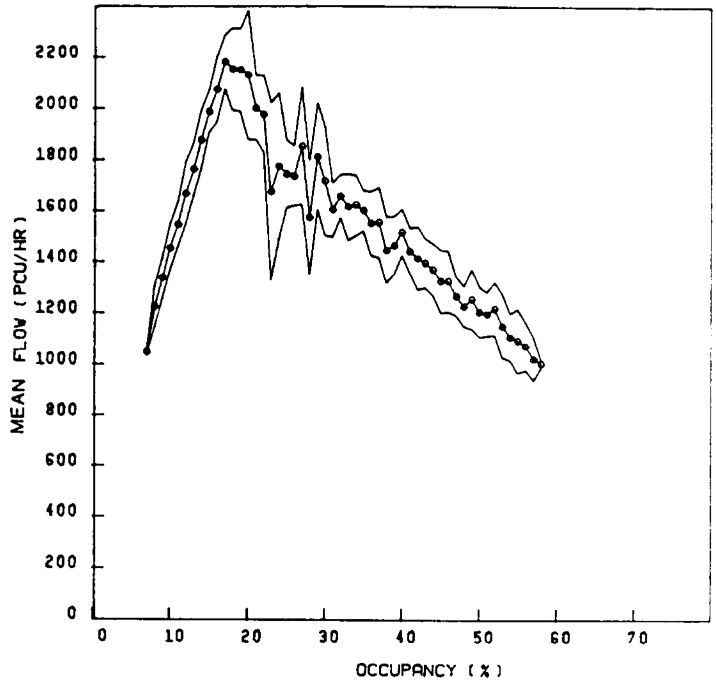

Fig. 10. Mean flow versus occupancy for 45 days, Q.E.W. data, transition points removed.

and shows maximum flow rates of 2350–2400 pcu/hr at the critical occupancy. The relationship is obviously very well defined through an occupancy range of 0–20%. The congested regime exhibits somewhat more data scatter, but a marginally convex relationship is observed in the 30–60% occupancy range. Despite the larger scatter, the relationship in this higher occupancy range is relatively smooth, consistent and well defined. Such is not the case for the intermediate range of 20–30% occupancy. Both the 19 and 20% plots contain considerable variations, particularly with widely varying occupancy trends as flow rates decrease. This lack of consistency might be expected, however, due to the relatively small amount of data available in that occupancy range. Finally, the 20% plot appears very similar to the Koshi reverse lambda, with the tendency of the congested branch to intersect the uncongested branch at flow rates less than the maximum.

From these observations, three conclusions can be drawn. First, the choice of a threshold occupancy is indeed critical to the averaging procedure and to the pattern that results; that critical occupancy seems to be about 19–20%. Second, the relationship is well defined in low (<20%) and high (>30%) occupancy ranges. There is little doubt that the uncongested regime is nearly linear in nature, and that the congested regime at occupancies greater than 30% follows a linear or slightly convex pattern. Third, although the relationship in the intermediate (20–30%) occupancy range is not as well defined, there is evidence that it is continuous.

DISCUSSION OF RESULTS

Before stating firm conclusions concerning the possibility of a continuous function, it is worthwhile to investigate other possible interpretations. There are at least four feasible configurations that could visually and/or conceptually fit the data and define the underlying relationship:

1. Koshi's reverse lambda shape, as shown in Fig. 12a (and in Fig. 1);
2. the typical discontinuous function shown in Fig. 12b (and in Fig. 2);
3. continuous *and* continuously differentiable functions such as those shown by the solid and dashed line in Fig. 12c; or
4. the continuous, but *not* continuously differentiable function shown in Fig. 12d.

In the first case, the reverse lambda shape in Fig. 12a is quite credible, at least in a cursory visual sense. Intuitively and pragmatically, however, the concept is not plausible. For example, why would one expect that commencement of the congested branch occurs at much lower flow rates than are possible in the uncongested regime? Furthermore, as was discussed previously, the daily data from traffic operations clearly showed that there were days when the lambda shape would not desscribe the true nature of the flow-occupancy interaction.

Similarly, the conventional discontinuous shape shown in Fig. 12b does not present a particularly satisfying possibility. To obtain the discontinuity, while maintaining equal slopes for the two regime shapes at the discontinuity, a considerable degree of inventiveness must be exercised when interpreting the data. It is not at all clear from viewing the data that there is a discontinuity. It is equally unclear that there exists a curvilinear shape at the upper reaches of the congested or uncongested branches, although the notion is slightly more acceptable

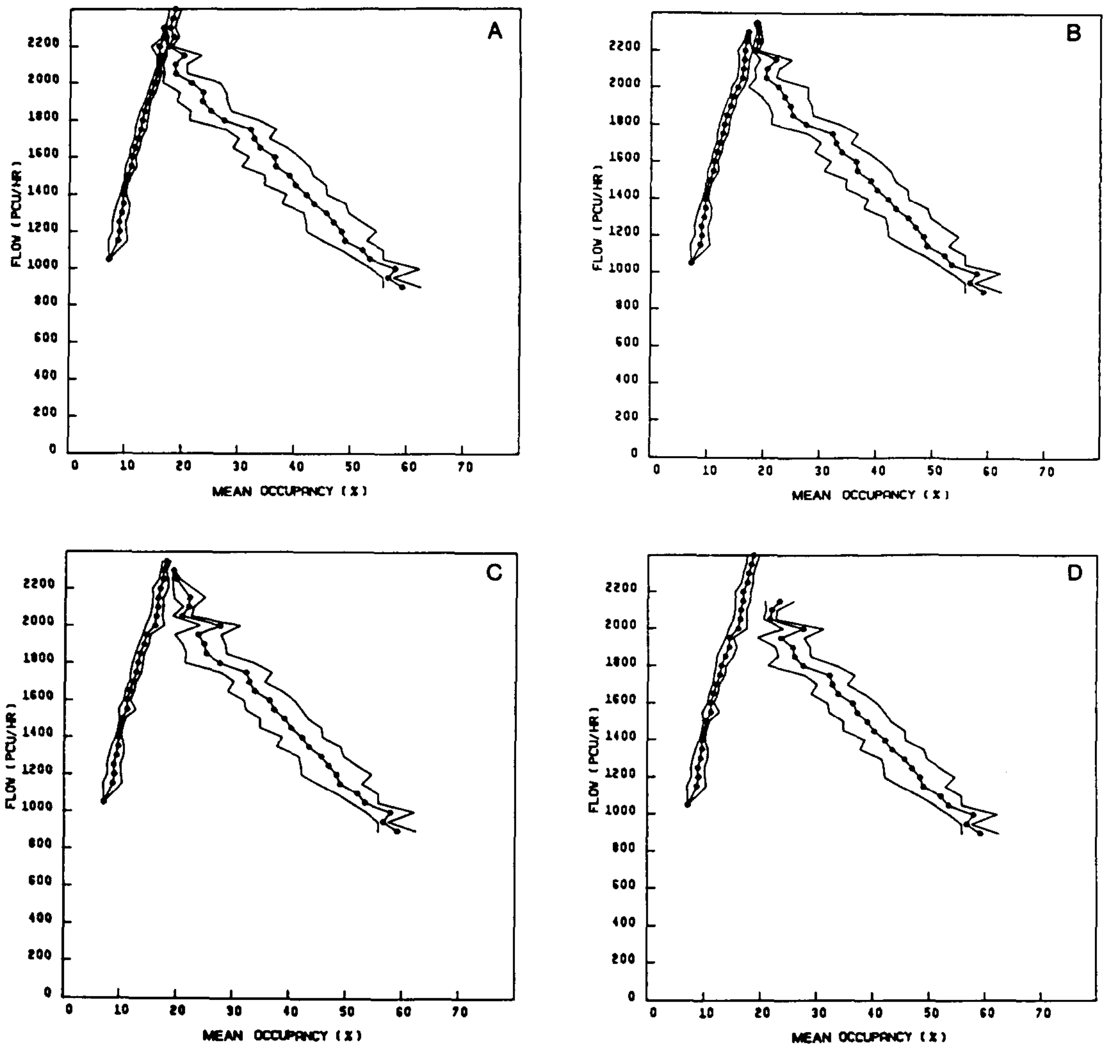

Fig. 11. Flow versus mean occupancy for 45 days, Q.E.W. data, transition points removed, for four values of critical occupancy: (a) 17%; (b) 18%; (c) 19%; (d) 20%.

in the case of 20% critical occupancy shown in Fig. 11d. Further, it is not obvious why one should assume that the slopes of the two curves are equal and zero at the point of discontinuity. Equally important for the rejection of this idea, however, is the fact that plausible operational reasons can be identified that would necessarily give a gap in the resulting data (i.e. Fig. 3), even for an underlying continuous curve.

For the continuously differentiable case in Fig. 12c, it is apparent that the curves do not fit the data well in the troublesome intermediate occupancy range. However, there are several logical reasons as to why this might occur. First, it is conceivable that the data points shown are actually in transition between the uncongested and congested regimes. The data collection location is a few kilometers upstream of the bottleneck, and one would expect that queue intrusion would create a series of jumps from one branch to the other at high, but randomly varying flow rates. Because of the 5-minute averaging in the original data, this could lead to points lying between the two regimes. It is reasonable to expect that jumps in occupancy at high flow rates would occur very rapidly when the queue shock wave entered the section. Five minute volumes and associated averaged occupancies may not provide enough definition to distinguish the complete transition in this unstable zone.

Second, if operations in the 20–30% occupancy range are inherently unstable, this could easily result in data points falling between the two regimes; very short-term oscillations in occupancy at relatively high flow rates and speeds could occur, particularly in the median lane. Such operational instabilities could result as the wave front reached the location, with drivers adjusting and readjusting their headways in response to increased density. In addition, instability in the actual location of the wave front could result in what appears to be transition data. In other words, the front could conceivably move into and out of the location at the onset of congestion, prior to reaching a more steady-state condition. A further consideration, as we argued earlier, is that the location of

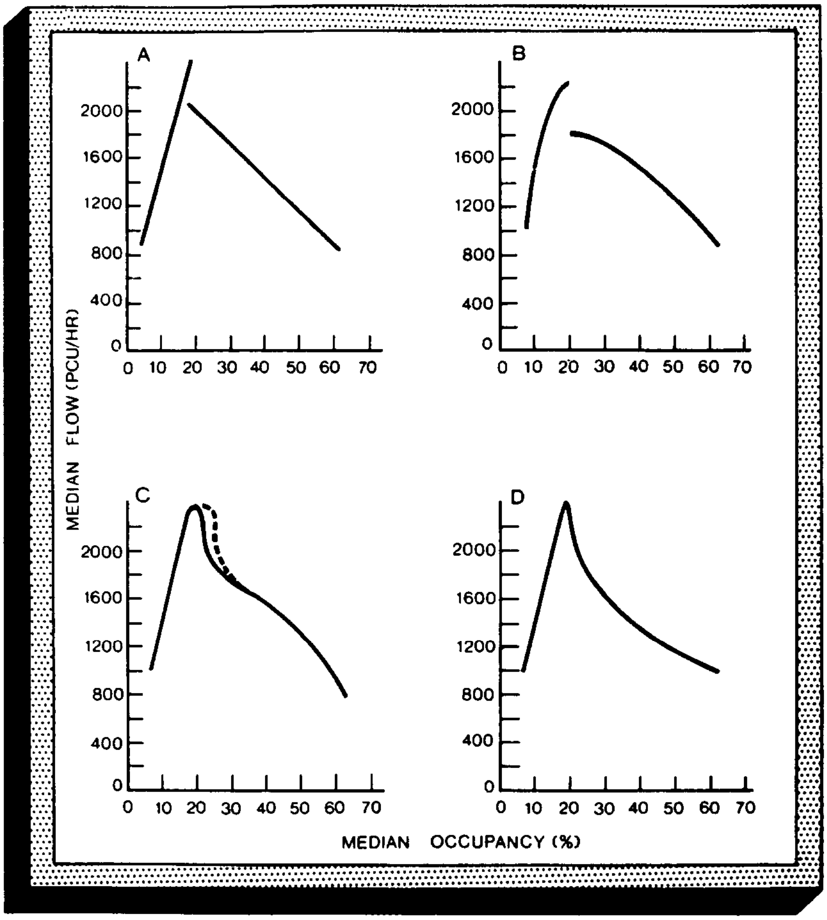

Fig. 12. Feasible flow-occupancy relationships: (a) reverse lambda; (b) discontinuous; (c) continuous (and continuously differentiable); (d) continuous (but not continuously differentiable).

the data collection (with respect to the bottleneck) will affect the nature of the data, so that one particular data set is not in itself sufficient cause to reject the smooth curve. Considering all possibilities, then, there is not sufficient conceptual evidence to indicate that a continuously differentiable shape could not exist. However, given the facts that the data fit is not particularly good, *and* that the daily plots do not provide evidence of such smooth curves, the shapes shown in Fig. 12c are not strongly appealing. One would expect a hypothesized flow-occupancy curve to bear a closer relationship to observed data. This curve, in the critical region, does not come very close.

On the other hand, the alternative continuous shape shown in Fig. 12d does fit the data rather well. The precise location and configuration of the congested branch is unimportant compared to the concept of an inverted V shape in the intermediate occupancy range. Although previously published work certainly does not show such sharply peaked relationships, it is entirely conceivable that they exist. In constrast to the preceeding argument for transition points, there is at least one plausible explanation for the very narrow range of occupancies at higher flow rates. For example, consider two extreme situations at a location experiencing high uncongested flow rates: one where the shock wave is moving slowly with a very small rate of increasing queue size, and the other where the shock wave has a high velocity and high rate of increasing queue size. In the first instance, when the slow shock wave front "hits" the location, the occupancy will increase only slightly, but operations will be on the congested branch at approximately the same flow rate (i.e. near the point of the V). In the second case, when a fast shock wave hits, the occupancy will necessarily increase very sharply, say to values greater than 30%. The high flow rates can no longer be sustained, decreasing almost instantaneously down the congested branch (i.e. to a much wider portion of the V). Consequently, with high uncongested flow rates, one may never witness a jump between the branches as could occur for the shapes in Fig. 12c. The upper portion of the congested regime (occupancies 20–30%) would therefore be defined by combinations of mainline demand and shock wave velocities varying between the two extremes presented. Because the number of combinations is potentially very large, particularly at locations far removed from the bottleneck, one would not expect a very well defined trend in the data. In fact, even for large data sets (i.e. 45 days), some lack of, and instability with, data is to be expected. Therefore, other than the fact that the inverted V shape in Fig. 12d is not consistent with conventional interpretations, very little can be found to negate the concept.

CONCLUSIONS

Despite the very simple analyses used for this paper, three important conclusions seem warranted. The first should have been obvious, but appears to have been ignored in earlier work on traffic flow relationships: the nature of the data that are collected at any particular freeway location depend 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. Consequently, it is essential to provide details of data collection locations, especially with respect to downstream bottlenecks and intervening ramps. Without that information, it is not possible to know whether a particular pattern in resulting data represents the full relationship or only the specifics of a particular place, subject to unique demand/capacity/queueing conditions.

The second conclusion also relates to the specifics of the data. There appear to be clear advantages to looking at daily traces of traffic behaviour, rather than relying on scatter diagrams of numerous days of accumulated data. The first advantage is that from the daily plots one can confidently obtain some idea of actual behaviour of the variables, which the scatter diagram cannot provide. The second advantage is that inspection of the daily plots permits one to identify points that represent transition between congested and uncongested flow and that therefore should not be included in estimation of mean values for either branch of the curve. Reliance on a scatter diagram leaves all such transitional points in the analysis, and can result in misleading analyses and consequent interpretation.

The third, and most important conclusion, is that the arguments for a discontinuous flow-occupancy (or flow-density) curve do not seem convincing. Our data clearly suggest a continuous curve is called for. More important, however, is the fact that a continuous curve and knowledge of operations at a particular location can together easily explain the occurrence of gaps in the data. It is not necessary to postulate discontinuous functions to account for this result. Therefore, we have proposed an inverted V shape for the flow-occupancy curve as the most logical form in the face of data we and others have reported.

This last conclusion requires further work to support it. The result is based on one lane at only one location, which is admittedly limited. Such a limited approach was necessary to develop the first two conclusions. Those in turn point the way toward the type of further data collection and analysis that need to be done to verify the third conclusion.

REFERENCES

Case E. R. and Williams K. M. (1978) Queen Elizabeth Way freeway surveillance and control system demonstration project. *Transpn. Res. Rec.* **682**, 84–93.

Ceder A. (1976) A deterministic traffic flow model for the two-regime approach. *Transpn. Res. Rec.* **567**, 16–30.

Easa S. M. (1983) Selecting two-regime traffic-flow models. *Transpn. Res. Rec.* **869**, 25–36.

Edie L. C. (1961) Car-following and steady-state theory for noncongested traffic. *Oper. Res.* **9**, 66–76.

Gazis D. C., Herman R. and Rothery R. W. (1961) Nonlinear follow-the-leader models of traffic flow. *Oper. Res.* **9**, 545–567.

Greenberg H. (1959) An analysis of traffic flow. *Oper. Res.* **7**, 79–85.

Koshi M., Iwasaki M. and Ohkura I. (1983) Some Findings and an Overview on Vehicular Flow Characteristics. In Proceedings of the 8th International Symposium on Transportation and Traffic Flow Theory, 1981 (Edited by Hurdle V. F., Hauer E. and Stewart G. N.) University of Toronto Press, Toronto, Ontario, pp. 403–426.

Mahabir G. P. (1981) Speed, Flow and Capacity Relations on Multilane Highways; M.Eng. Thesis, Dept. of Civil Engineering, McMaster University, Hamilton, Ontario.

Roess R. P., Linzen E., McShane W. R. and Pignataro L. J. (1980) Freeway Capacity Procedures. In *Transpn. Res. Circular 212, Interim Materials on Highway Capacity*, Transportation Research Board, Washington, DC, 151–266.