TRAFFIC STREAM CHARACTERISTICS

BY FRED L. HALL[4]

[4]Professor, McMaster University, Department of Civil Engineering and Department of Geography, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4L7.

CHAPTER 2 - Frequently used Symbols

k density of a traffic stream in a specified length of road

L length of vehicles of uniform length

c_k constant of proportionality between occupancy and density, under certain simplifying assumptions

k_i the (average) density of vehicles in substream *I*

q_i the average rate of flow of vehicles in substream *I*

$\bar{u}$ average speed of a set of vehicles

A $A(x,t)$ the cumulative vehicle arrival function over space and time

k_j jam density, i.e. the density when traffic is so heavy that it is at a complete standstill

u_f free-flow speed, i.e. the speed when there are no constraints placed on a driver by other vehicles on the road

2.
TRAFFIC STREAM CHARACTERISTICS

This chapter describes the various models that have been developed to describe the relationships among traffic stream characteristics. Most of the work dealing with these relationships has been concerned with uninterrupted traffic flow, primarily on freeways or expressways. Consequently, this chapter will cover traffic stream characteristics for uninterrupted flow. In discussing the models, the link between theory and measurement capability is important since often theory depends on measurement capability.

Because of the importance of measurement capability to theory development, this chapter starts with a section on historical developments in measurement procedures. That section is followed by one providing detailed descriptions and definitions of the variables of interest. Some of the relationships between the variables are simply a matter of definition. An example is the relationship between density of vehicles on the road, in vehicles per unit distance, and spacing between vehicles, in distance per vehicle. Others are more difficult to specify. The final section, on traffic stream models, focuses on relationships among speed, flow, and concentration, either in two-variable models, or in those that attempt to deal simultaneously with the three variables.

2.1 Measurement Procedures

The items of interest in traffic theory have been the following:

- rates of flow (vehicles per unit time);
- speeds (distance per unit time);
- travel time over a known length of road (or sometimes the inverse of speed, "tardity" is used);
- occupancy (percent of time a point on the road is occupied by vehicles);
- density (vehicles per unit distance);
- time headway between vehicles (time per vehicle);
- spacing, or space headway between vehicles (distance per vehicle); and
- concentration (measured by density or occupancy).

A general notion of these variables, based on the intuitive idea self-evident from their names, will suffice for the purposes of discussing their measurement. Precise definitions of these variables are given in Section 2.2.

Measurement capabilities for obtaining traffic data have changed over the nearly 60-year span of interest in traffic flow, and more so in the past 40 years during which there have been a large number of freeways. Indeed, measurement capabilities are still changing. Five measurement procedures are discussed in this section:

- measurement at a point;
- measurement over a short section (by which is meant less than about 10 meters (m);
- measurement over a length of road [usually at least 0.5 kilometers (km)];
- the use of an observer moving in the traffic stream; and
- wide-area samples obtained simultaneously from a number of vehicles, as part of Intelligent Transportation Systems (ITS).

For each method, this section contains an identification of the variables that the particular procedure measures, as contrasted with the variables that can only be estimated.

The types of measurement are illustrated with respect to a space-time diagram in Figure 2.1. The vertical axis of this diagram represents distance from some arbitrary starting point along the road, in the direction of travel. The horizontal axis represents elapsed time from some arbitrary starting time. Each line within the graph represents the 'trajectory' of an individual vehicle, as it moves down the road over time. The slope of the line is that vehicle's velocity. Where lines cross, a faster vehicle has overtaken and passed a slower one. (The two vehicles do not in fact occupy the same point at the same time.) Measurement at a point is represented by a horizontal line across the vehicular trajectories: the location is constant, but time varies. Measurement over a short section is represented by two parallel horizontal lines a very short distance apart. A vertical line represents measurement along a length of road, at one instant of time, such as in a single snapshot taken from above the road (for example an aerial photograph). The moving observer technique is represented by one of the vehicle trajectories, the heavy line in

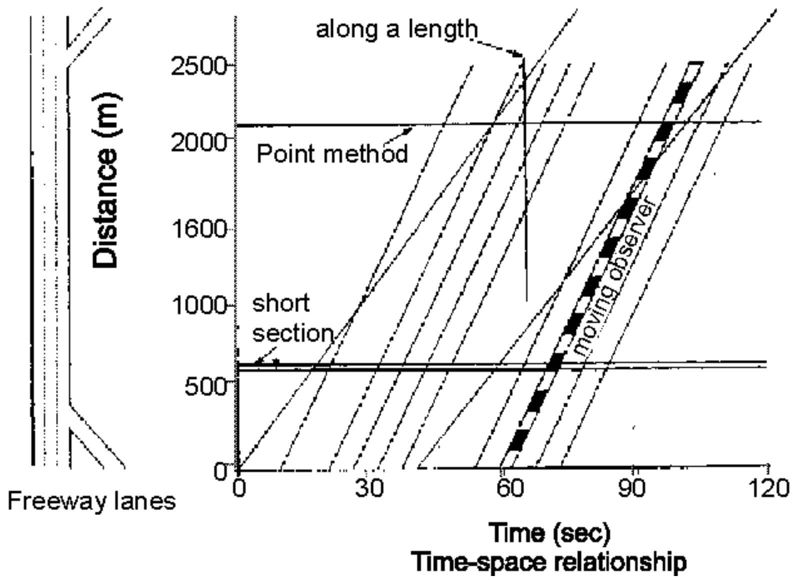

Figure 2.1
Four Methods of Obtaining Traffic Data (Modified from Drew 1968, Figure 12.9).

Figure 2.1. Details on each of these methods can be found in the ITE's Manual of Traffic Engineering Studies (Box 1976).

The wide-area samples from ITS are similar to having a number of moving observers at various points and times within the system. These new developments will undoubtedly change the way some traffic measurements are obtained in the future, but they have not been in operation long enough to have a major effect on the material to be covered in this chapter.

2.1.1 Measurement at a Point

Measurement at a point, by hand tallies or pneumatic tubes, was the first procedure used for traffic data collection. This method is easily capable of providing volume counts and therefore flow rates directly, and with care can also provide time headways. The technology for making measurements at a point on freeways changed over 30 years ago from using pneumatic tubes placed across the roadway to using point detectors (May et al. 1963; Athol 1965). The most commonly used point detectors are based on inductive loop technology, but other methods in use include microwave, radar, photocells, ultrasonics, and television cameras.

One of the more recent data collection methods draws upon video camera technology. In its earliest applications, video cameras were used to acquire the data in the field, which was then subsequently played back in a lab for analysis. In these early implementations, lines were drawn on the video monitor screen (literally, when manual data reduction was used). More recently this has been automated, and the lines are simply a part of the electronics. This procedure allows the data reduction to be conducted simultaneously with the data acquisition.

Except for the case of a stopped vehicle, speeds at a 'point' can be obtained only by radar or microwave detectors. Their frequencies of operation mean that a vehicle needs to move only about one centimeter during the speed measurement. In the absence of such instruments for a moving vehicle, a second observation location is necessary to obtain speeds, which moves the discussion to that of measurements over a short section.

Density, which is defined as vehicles per unit length, does not make sense for a point measurement, because no length is involved. Density can be calculated from point measurements when speed is available, but one would have to question the meaning of the calculation, as it would be density at a point. In the abstract, one would expect that occupancy could be measured at a point, but in reality devices for measuring occupancy generally take up a short space on the roadway. Hence volume (or flow rate), headways, and speeds are the only direct measurements at a point.

2.1.2 Measurements Over a Short Section

Early studies used a second pneumatic tube, placed very close to the first, to obtain speeds. More recent systems have used paired presence detectors, such as inductive loops spaced perhaps five to six meters apart. With video camera technology, two detector 'lines' placed close together provide the same capability for measuring speeds. Even with such short distances, one is no longer dealing strictly with a point measurement, but with measurement along a section of road, albeit a short section. All of these presence detectors continue to provide direct measurement of volume and of time headways, as well as of speed when pairs of them are used.

Most point detectors currently used, such as inductive loops or microwave beams, take up space on the road, and are therefore a short section measurement. These detectors produce a new variable, which was not available from earlier technology, namely occupancy. Occupancy is defined as the percentage of time that the detection zone of the instrument is occupied by a vehicle. This variable is available because the loop gives a continuous reading (at 50 or 60 Hz usually), which pneumatic tubes or manual counts could not do. Because occupancy depends on the size of the detection zone of the instrument, the measured occupancy may differ from site to site for identical traffic, depending on the nature and construction of the detector. It would be possible mathematically to standardize the measurement of occupancy to a zero-length detection zone, but this has not yet happened. For practical purposes, many freeway management systems rely solely on flow and occupancy information. (See for example, Payne and Tignor 1978; Collins 1983.)

As with point measurements, short-section data acquisition does not permit direct measurement of density. Where studies based on short-section measurements have used density, it has been calculated, in one of two ways to be discussed in Section 2.3.

The large quantity of data provided by modern freeway traffic management systems is noteworthy, especially when compared with the quantity of data used in the early development of traffic flow theory. As a consequence of this large amount of relatively recent data, there has been in the past decade a considerable increase in the amount of research investigating the underlying relationships among traffic stream characteristics, which is reflected in Section 2.3.

2.1.3 Measurement Along a Length of Road

Measurements along a length of road come either from aerial photography, or from cameras mounted on tall buildings or poles. It is suggested that at least 0.5 kilometers (km) of road be observed. On the basis of a single frame from such sources, only density can be measured. The single frame gives no sense of time, so neither volumes nor speed can be measured.

Once several frames are available, as from a video-camera or from time-lapse photography over short time intervals, speeds can also be measured, often over a distance approximating the entire section length over which densities have been calculated. Note however the shift in the basis of measurement. Even though both density and speed can be taken over the full length of the section, density must be measured at a single point in time, whereas measurement of speed requires variation over time, as well as distance. In general, flow refers to vehicles crossing a point or line on the roadway, for example one end of the section in question. Hence, flow and density refer to different measurement frameworks: flow over time at a point in space; density over space at a point in time.

Despite considerable improvements in technology, and the presence of closed circuit television on many freeways, there is very little use of measurements taken over a long section at the present time. The one advantage such measurements might provide would be to yield true journey times over a lengthy section of road, but that would require better computer vision algorithms (to match vehicles at both ends of the section) than are currently possible. There have been some efforts toward the objective of collecting journey time data on the basis of the details of the 'signature' of particular vehicles or platoons of vehicles across a series of loops over an extended distance (Kühne and Immes 1993), but few practical implementations as yet. Persaud and Hurdle (1988b) describe another way to make use of measurements along a length, following the method proposed by Makagami et al. (1971). By constructing cumulative arrival curves at several locations, they were able to derive both the average flow rate and the average density within a section, and consequently the average speeds through it.

2.1.4 Moving Observer Method

There are two approaches to the moving observer method. The first is a simple floating car procedure in which speeds and travel times are recorded as a function of time and location along the road. While the intention in this method is that the floating car behaves as an average vehicle within the traffic stream, the method cannot give precise average speed data. It is, however, effective for obtaining qualitative information about freeway operations without the need for elaborate equipment or procedures. One form of this approach uses a second person in the car to record speeds and travel times. A second form uses a modified recording speedometer of the type regularly used in long-distance trucks or buses. One drawback of this approach is that it means there are usually significantly fewer speed observations than volume observations. An example of this kind of problem appears in Morton and Jackson (1992).

The other approach was developed by Wardrop and Charlesworth (1954) for urban traffic measurements and is meant to obtain both speed and volume measurements simultaneously. Although the method is not practical for major urban freeways, it is included here because it may be of some value for rural expressway data collection, where there are no automatic systems. While it is not appropriate as the primary mode of data collection on a busy freeway, there are some useful points that come out of the literature that should be noted by those seeking to obtain average speeds through the moving car method. The technique has been used in the past especially on urban arterials, for example in connection with identifying progression speeds for coordinated signals.

The method developed by Wardrop and Charlesworth is based on a survey vehicle that travels in both directions on the road. The formulae allow one to estimate both speeds and flows for one direction of travel. The two formulae are

$$q = \frac{(x+y)}{(t_a+t_w)} \tag{2.1}$$

$$\bar{t} = t_w - \frac{y}{q} \tag{2.2}$$

where,

- q is the estimated flow on the road in the direction of interest,
- x is the number of vehicles traveling in the direction of interest, which are met by the survey vehicle while traveling in the opposite direction,
- y is the net number of vehicles that overtake the survey vehicle while traveling in the direction of interest (i.e. those passing minus those overtaken),
- t_a is the travel time taken for the trip against the stream,
- t_w is the travel time for the trip with the stream, and
- $\bar{t}$ is the estimate of mean travel time in the direction of interest.

Wright (1973) revisited the theory behind this method. His paper also serves as a review of the papers dealing with the method in the two decades between the original work and his own. He finds that, in general, the method gives biased results, although the degree of bias is not significant in practice, and can be overcome. Wright's proposal is that the driver should fix the journey time in advance, and keep to it. Stops along the way would not matter, so long as the total time taken is as determined prior to travel. Wright's other point is that turning traffic (exiting or entering) can upset the calculations done using this method. This fact means that the route to be used for this method needs to avoid major exits or entrances. It should be noted also that a large number of observations are required for reliable estimation of speeds and flow rates; without that, the method has very limited precision.

2.1.5 ITS Wide-Area Measurements

Some forms of Intelligent Transportation Systems involve the use of communications from specially-equipped vehicles to a central system. Although the technology of the various communications systems differ, all of them provide for transmission of information on the vehicles' speeds. In some cases, this would simply be the instantaneous speed while passing a particular reporting point. In others, the information would be simply a vehicle identifier, which would allow the system to calculate journey times between one receiving location and the next. A third type of system would not be based on fixed interrogation points, but would poll vehicles regardless of location, and would receive speeds and location information back from the vehicles.

The first system would provide comparable data to that obtained by paired loop detectors. While it would have the drawback of sampling only a small part of the vehicle fleet, it would also have several advantages. The first of these is that system maintenance and repair would not be so expensive or disruptive as is fixing broken loops. The second is that the polling stations could be set up more widely than loops currently are, providing better coverage especially away from freeways.

The second system would provide speeds over a length of road, information that cannot presently be obtained without great effort and expense. Since journey times are one of the key variables of interest for ITS route guidance, better information would be an advantage for that operational purpose. Such information will also be of use for theoretical work, especially in light of the discussion of speeds that appears in Section 2.2.2.

The third system offers the potential for true wide-area speed information, not simply information at selected reporting points. The major difficulty with implementing this approach is that of establishing locations precisely. Global positioning systems have almost achieved the capability for doing this well, but they would add considerably to the expense of this approach.

The limitation to all three systems is that they can realistically be expected to provide information only on speeds. It is notgenerally possible for one moving vehicle to be able to identify flow rates or densities in any meaningful way. Of course, with appropriate sensors, each instrumented vehicle could report on its time and space headway, but it would take a larger sample than is likely to occur in order to have any confidence in the calculation of flow as the reciprocal of reported time headways, or density as the reciprocal of reported spacings. Thus there remains the problem of finding comparable flow or density information to go along with the potentially improved speed information.

2.2 Variables of Interest

In general, traffic streams are not uniform, but vary over both space and time. Because of that, measurement of the variables of interest for traffic flow theory is in fact the sampling of a random variable. In some instances in the discussion below, that is explicit, but in must cases it is only implicit. In reality, the traffic characteristics that are labeled as flow, speed, and concentration are parameters of statistical distributions, not absolute numbers.

2.2.1 Flow Rates

Flow rates are collected directly through point measurements, and by definition require measurement over time. They cannot be estimated from a single snapshot of a length of road. Flow rates and time headways are related to each other as follows. Flow rate, q, is the number of vehicles counted, divided by the elapsed time, T:

$$q = \frac{N}{T} \tag{2.3}$$

The total elapsed study time is made up of the sum of the headways recorded for each vehicle:

$$T = \sum_{i=1}^{N} h_i \tag{2.4}$$

If the sum of the headways is substituted in Equation 2.3 for total time, T, then it can be seen that the flow rate and the average headway have a reciprocal relationship with each other:

$$q = \frac{N}{T} = \frac{N}{\sum_i h_i} = \frac{1}{\frac{1}{N}\sum_i h_i} = \frac{1}{\bar{h}} \tag{2.5}$$

Flow rates are usually expressed in terms of vehicles per hour, although the actual measurement interval can be much less. Concern has been expressed, however, about the sustainability of high volumes measured over very short intervals (such as 30 seconds or one minute) when investigating high rates of flow. The 1985 Highway Capacity Manual (HCM 1985) suggests using at least 15-minute intervals, although there are also situations in which the detail provided by five minute or one minute data is valuable. The effect of different measurement intervals on the nature of resulting data was shown by Rothrock and Keefer (1957).

2.2.2 Speeds

Measurement of the speed of an individual vehicle requires observation over both time and space. The instantaneous speed of an individual vehicle is defined as

$$u_i = \frac{dx}{dt} = limit(t_2 - t_1) \rightarrow 0 \ \frac{x_2 - x_1}{t_2 - t_1} \qquad (2.6)$$

Radar or microwave would appear to be able to provide speed measurements conforming most closely to this definition, but even these rely on the motion of the vehicle, which means they take place over a finite distance and time, however small those may be. Vehicle speeds are also measured over short sections, such as the distance between two closely-spaced (6 m) inductive loops, in which case one no longer has the instantaneous speed of the vehicle, but a close approximation to it (except during rapid acceleration or deceleration).

In the literature, the distinction has frequently been made between different ways of calculating the average speed of a set of vehicles. The kind of difference that can arise from different methods can be illustrated by the following example, which is a kind that often shows up on high school mathematics aptitude tests. If a traveler goes from A to B, a distance of 20 km, at an average speed of 80 kilometers per hour (km/h), and returns at an average speed of 40 km/h, what is the average speed for the round trip? The answer is of course not 60 km/h; that is the speed that would be found by someone standing at the roadside with a radar gun, catching this car on both directions of the journey, and averaging the two observations. The trip, however, took 1/4 of an hour one way, and 1/2 an hour for the return, for a total of 3/4 of an hour to go 40 km, resulting in an average speed of 53.3 km/h.

The first way of calculating speeds, namely taking the arithmetic mean of the observation,

$$\bar{u}_t = \frac{1}{N} \sum_{i=1}^{N} u_i \qquad (2.7)$$

is termed the time mean speed, because it is an average of observations taken over time.

The second term that is used in the literature is space mean speed, but unfortunately there are a variety of definitions for it, not all of which are equivalent. There appear to be two main types of definition. One definition is found in Lighthill and Whitham (1955), which they attribute to Wardrop (1952), and is the speed based on the average time taken to cross a given distance, or space, *D*:

$$\bar{u}_s = \frac{D}{\frac{1}{N} \sum_i t_i} \qquad (2.8)$$

where t_i is the time for vehicle *I* to cross distance *D*

$$t_i = \frac{D}{u_i} \qquad (2.9)$$

A similar definition appears in handbooks published by the Institute of Transportation Engineers (ITE 1976; ITE 1992), and in May (1990), and is repeated in words in the 1985 HCM. One question regarding Equation 2.8 is what the summation is taken over. Implicitly, it is over all of the vehicles that crossed the full section, *D*. But for other than very light traffic conditions, there will always be some vehicles within the section that have not completed the crossing. Hence the set of vehicles to include must necessarily be somewhat arbitrary.

The 1976 ITE publication also contains a related definition, where space mean speed is defined as the total travel divided by the total travel time. This definition calls for specifying an explicit rectangle on the space-time plane (Figure 2.1), and taking into account all travel that occurs within it. This definition is similar to Equation 2.8 in calling for measurement of speeds over a distance, but dissimilar in including vehicles that did not cover the full distance.

Some authors, starting as far back as Wardrop (1952), demonstrate that Equation 2.8 is equivalent to using the harmonic mean of the individual vehicle speeds, as follows.

$$\bar{u}_s = \frac{D}{\frac{1}{N}\sum_i t_i} = \frac{D}{\frac{1}{N}\sum_i \frac{D}{u_i}} = \frac{1}{\frac{1}{N}\sum_i \frac{1}{u_i}} \tag{2.10}$$

The difficulty with allowing a definition of space mean speed as the harmonic mean of vehicle speeds is that the measurement over a length of road, D, is no longer explicit. Consequently, the last right-hand side of Equation 2.10 makes it look as if space mean speed could be calculated by taking the harmonic mean of speeds measured at a point over time. Wardrop (1952), Lighthill and Whitham (1955), and Edie (1974) among other authors accepted this use of speeds at a point to calculate space mean speed. For the case where speeds do not change with location, the use of measurements at a point will not matter, but if speeds vary over the length of road there will be a difference between the harmonic mean of speeds at a point in space, and the speed based on the average travel time over the length of road. As well, Haight (1963) and Kennedy et al. (1973) note that measurements at a point will over-represent the number of fast vehicles and under-represent the slow ones, and hence give a higher average speed than the true average.

The second principal type of definition of space mean speed involves taking the average of the speeds of all of the vehicles on a section of road at one instant of time. It is most easily visualized with the example given by Haight (1963, 114): "an aerial photograph, assuming each car to have a speedometer on its top." Leutzbach (1972; 1987) uses a similar example. In Figure 2.1, this method is represented by the vertical line labeled "along a length". Kennedy et al. (1973) use a slightly more realistic illustration, of two aerial photographs taken in close succession to obtain the speeds of all of the vehicles in the first photo. Ardekani and Herman (1987) used this method as part of a study of the relationships among speed, flow, and density. Haight goes on to show mathematically that a distribution of speeds collected in this fashion will be identical to the true distribution of speeds, whereas speeds collected over time at one point on the road will not match the true distribution. In deriving this however, he assumes an "isoveloxic" model, one in which each car follows a linear trajectory in the space time diagram, and is not forced to change speed when overtaking another vehicle. This is equivalent to assuming that the speed distribution does not change with location (Hurdle 1994). A similar definition of space mean speed, without the isoveloxic assumption, appears in HRB SR 79 (Gerlough and Chappelle 1964, viii): "the arithmetic mean of the speeds of vehicles occupying a given length of lane at a given instant."

Wohl and Martin (1967, 323) are among the few authors who recognize the difference in definition. They quote the HRB SR79 definition in a footnote, but use as their definition "mean of the speeds of the vehicles traveling over a given length of road and weighted according to the time spent traveling that length". Regardless of the particular definition put forward for space mean speed, all authors agree that for computations involving mean speeds to be theoretically correct, it is necessary to ensure that one has measured space mean speed, rather than time mean speed. The reasons for this are discussed in Section 2.2.3. Under conditions of stop-and-go traffic, as along a signalized street or a badly congested freeway, it is important to distinguish between these two mean speeds. For freely flowing freeway traffic, however, there will not be any significant difference between the two, at least if Equation 2.10 can be taken to refer to speeds taken at a point in space, as discussed above. Wardrop (1952, 330) showed that the two mean speeds differ by the ratio of the variance to the mean of the space mean speed:

$$\bar{u}_t = \bar{u}_s + \frac{\sigma_s^2}{\bar{u}_s} \tag{2.11}$$

here σ_s^2 is defined as $\Sigma k_i(u_i - \bar{u}_s)^2/K$, k_i is the density of sub-stream I, and K is the density of the total stream. When there is great variability of speeds, as for example at the time of breakdown from uncongested to stop and go conditions, there will be considerable difference between the two. Wardrop (1952) provided an example of this kind (albeit along what must certainly have been a signalized roadway -- Western Avenue, Greenford, Middlesex, England), in which speeds ranged from a low of 8 km/h to a high of 100 km/h. The space mean speed was 48.6 km/h; the time mean speed 54.0 km/h. On the basis of such calculations, Wardrop (1952, 331) concluded that time mean speed is "6 to 12 percent greater than the space mean" speed. Note however, from the high school mathematics example given earlier, that even with a factor of 2 difference in speeds, there is only about a 13 percent difference between $\bar{u}_t$ and $\bar{u}_s$.

In uncongested freeway traffic, the difference between the two speeds will be quite small. Most vehicles are traveling at very similar speeds, with the result that σ_s^2 will be small, while $\bar{u}_s$ will be relatively large. Gerlough and Huber (1975) provide an example based on observation of 184 vehicles on Interstate 94 in Minnesota. Speeds ranged from a low of 35.397 km/h to a high of 45.33 km/h. The arithmetic mean of the speeds was 39.862 km/h; the harmonic mean was 39.769 km/h. As expected, the two mean speeds are not identical. However, the original measurements were accurate only to the nearest whole mile per hour. To the accuracy of the original measurements, the two means are equal. In other words, for relatively uniform flow and speeds, the two mean speeds are likely to be equivalent for practical purposes. Nevertheless, it is still appropriate to specify which type of averaging has been done, and perhaps to specify the amount of variability in the speeds (which can provide an indication of how similar the two are likely to be).

Even during congestion on freeways, the difference is not very great, as shown by the analyses of Drake et al. (1967). They calculated both space mean speed and time mean speed for the same set of data from a Chicago freeway, and then regressed one against the other. The resulting equation was

$$\bar{u}_s = 1.026\ \bar{u}_t - 1.890 \tag{2.12}$$

with speeds in miles per hour. The maximum speeds observed approached 96.6 km/h, at which value of time mean speed, space mean speed would be 96.069 km/h. The lowest observed speeds were slightly below 32.2 km/h. At a time mean speed of 32.2 km/h, the equation would yield a space mean speed of 29.99 km/h. However, it needs to be noted that the underlying relationship is in fact non-linear, even though the linear model in Equation 2.12 resulted in a high R^2. The equation may misrepresent the amount of the discrepancy, especially at low speeds. Nevertheless, the data they present, and the equation, when applied at high speeds, support the result found by Gerlough and Huber (1975): at least for freeways, the practical significance of the difference between space mean speed and time mean speed is minimal. However, it is important to note that for traffic flow theory purists, the only 'correct' way to measure average travel velocity is to calculate space-mean speed directly.

Only a few freeway traffic management systems acquire speed information directly, since to do so requires pairs of presence detectors at each of the detector stations on the roadway, and that is more expensive than using single loops. Those systems that do not measure speeds, because they have only single-loop detector stations, sometimes calculate speeds from flow and occupancy data, using a method first identified by Athol (1965). In order to describe and comment on Athol's derivation of that procedure, it is first necessary to define the measurement of occupancy, a topic which might otherwise be deferred to Section 2.3.3.

Occupancy is the fraction of time that vehicles are over the detector. For a specific time interval, T, it is the sum of the time that vehicles cover the detector, divided by T. For each individual vehicle, the time spent over the detector is determined by the vehicle's speed, u_i , and its length, L_i , plus the length of the detector itself, d. That is, the detector is affected by the vehicle from the time the front bumper crosses the start of the detection zone until the time the rear bumper clears the end of the detection zone.

$$\begin{aligned} occupancy &= \frac{\sum_i (L_i + d)/u_i}{T} \\ &= \frac{1}{T}\sum_i \frac{L_i}{u_i} + \frac{d}{T}\sum_i \frac{1}{u_i} \end{aligned} \tag{2.13}$$

Athol then multiplied the second term of this latter equation by $N\,(1/N)$, and substituted Equations 2.3 and 2.10:

$$\begin{aligned} occupancy &= \frac{1}{T}\sum_i \frac{L_i}{u_i} + d \cdot \frac{N}{T} \cdot \frac{1}{N}\sum_i \frac{1}{u_i} \\ &= \frac{1}{T}\sum_i \frac{L_i}{u_i} + d \cdot \frac{q}{\bar{u}_s} \end{aligned} \tag{2.14}$$

Assuming that the "fundamental equation" holds (which will be dealt with in detail in the next section), namely

$$q = k\bar{u}_s \tag{2.15}$$

this becomes

$$occupancy = \frac{1}{T}\sum_i \frac{L_i}{u_i} + d \cdot k \tag{2.16}$$

Noting that T is simply the sum of the individual vehicle headways, Athol made the substitution, and then multiplied top and bottom of the resulting equation by $1/N$:

$$\begin{aligned} occupancy &= \frac{\sum_i \frac{L_i}{u_i}}{T} + d \cdot k \\ &= \frac{\frac{1}{N}\sum_i \frac{L_i}{u_i}}{\frac{1}{N}\sum h_i} + d \cdot k \\ &= \frac{\frac{1}{N}\sum_i \frac{L_i}{u_i}}{\bar{h}} + d \cdot k \end{aligned} \qquad (2.17)$$

In order to proceed further, Athol assumed a uniform vehicle length, L, which allows the following simplification of the equation:

$$\begin{aligned} occupancy &= \frac{\frac{1}{N}\sum_i \frac{L}{u_i}}{\bar{h}} + d \cdot k \\ &= \frac{1}{\bar{h}} \cdot L \cdot \frac{1}{N}\sum_i \frac{1}{u_i} + d \cdot k \\ &= L \cdot \frac{q}{\bar{u}_s} + d \cdot k \\ &= (L+d)k = c_k k \end{aligned} \qquad (2.18)$$

Since at a single detector location, d is constant, this equation means that occupancy and density are constant multiples of each other (under the assumption of constant vehicle lengths). Consequently, speeds can be calculated as

$$\bar{u}_s = \frac{q \cdot c_k}{occupancy} \qquad (2.19)$$

It is useful to note that the results in Equation 2.18, and hence Equation 2.19, are still valid if vehicle lengths vary and speeds are constant, except that L would have to be interpreted as the mean of the vehicle lengths. On the other hand, if both lengths and speeds vary, then the transition from Equation 2.17 to 2.18 and 2.19 cannot be made in this simple fashion, and the relationship between speeds, flows, and occupancies will not be so clear-cut (Hall and Persaud 1989). Banks (1994) has recently demonstrated this result in a more elegant and convincing fashion.

Another method has recently been proposed for calculating speeds from flow and occupancy data (Pushkar et al. 1994). This method is based on the catastrophe theory model for traffic flow, presented in Section 2.3.6. Since explanation of the procedure for calculating speeds requires an explanation of that model, discussion will be deferred until that section.

2.2.3 Concentration

Concentration has in the past been used as a synonym for density. For example, Gerlough and Huber (1975, 10) wrote, "Although concentration (the number of vehicles per unit length) implies measurement along a distance...." In this chapter, it seems more useful to use 'concentration' as a broader term encompassing both density and occupancy. The first is a measure of concentration over space; the second measures concentration over time of the same vehicle stream.

Density can be *measured* only along a length. If only point measurements are available, density needs to be calculated, either from occupancy or from speed and flow. Gerlough and Huber wrote (in the continuation of the quote in the previous paragraph), that "...traffic engineers have traditionally *estimated* concentration from point measurements, using the relationship

$$k = q/\bar{u}_s \quad ". \qquad (2.20)$$

This is the same equation that was used above in Athol's derivation of a way to calculate speeds from single-loop detector data. The difficulty with using this equation to estimate density is that the equation is strictly correct only under some very restricted conditions, or in the limit as both the space and time measurement intervals approach zero. If neither of those situations holds, then use of the equation to calculate density can give misleading results, which would not agree with empirical measurements. These issues are important, because this equation has often been uncritically applied to situations that exceed its validity.

The equation was originally developed by Wardrop (1952). His derivation began with the assumption that the traffic stream could be considered to be a number of substreams, "in each of which all the vehicles are traveling at the same speed and form a random series" (Wardrop 1952, 327). Note that the randomness must refer to the spacing between vehicles, and that since all vehicles in the substream have constant speed, the spacing within the substream will not change (but is clearly not uniform). Wardrop's derivation then proceeded as follows (Wardrop 1952, 327-328), where his symbol for speed, v_i, has been replaced by the one used herein, u_i):

> Consider the subsidiary stream with flow, q_i and speed u_i. The average time-interval between its vehicles is evidently $1/q_i$, and the distance travelled in this time is u_i/q_i. It follows that the density of this stream in space, that is to say, the number of vehicles per unit length of road at any instant (the concentration), is given by

$$k_i = q_i/u_i \quad i=1, 2, ...c \tag{2.21}$$

The next step involved calculating the overall average speed on the basis of the fractional shares of total density, and using the above equation to deduce the results:

$$\overline{u_s} = \frac{\sum_i k_i u_i}{k} = \frac{\sum_i q_i}{k} = \frac{q}{k} \tag{2.22}$$

The equation for the substreams is therefore critical to the derivation to show that Equation 2.20 above holds when space mean speed is used.

There are two problems with this derivation, both arising from the distinction between a random series and its average. Wardrop is correct to say that the "average time-interval" between vehicles (in a substream) is $1/q_i$, but he neglects to include the word 'average' in the next sentence, about density: the <u>average</u> density of this stream in space. The issue is not simply one of wording, but of mathematics. Because the substream is random, which must mean random spacing since it has a uniform speed, the density of the substream cannot have a constant value. The density as measured over one portion of the substream may well be different from the density as measured over a different portion. The k_i calculated in Equation 2.21 is not the true density of the stream, but only an estimate. If that is the case, however, the derivation of Equation 2.20 is in jeopardy, because it calls upon Equation 2.21 subsequently in the derivation. In short, there appears to be an implicit assumption of constant spacing as well as constant speed in the derivation.

Gerlough and Huber (1975) reproduce some of Wardrop's derivation, but justify the key equation,

$$k_i = q_i/u_i \quad i=1, 2, ...c \tag{2.23}$$

on the basis of "analysis of units" (Gerlough and Huber 1975, 10). That is, the units of flow, in vehicles/hour, can be obtained by multiplying the units for density, in vehicles/km, by the units for speed, in km/hour. The fact that space mean speed is needed for the calculation, however, relies on the assumption that the key equation for substreams holds true. They have not avoided that dependence.

Although Equation 2.20 has been called the fundamental identity, or fundamental equation of traffic flow, its use has often exceeded the underlying assumptions. Wardrop's explicit assumption of substreams with constant speed is approximately true for uncongested traffic (that is at flows of between 300 and perhaps 2200 pcphpl), when all vehicles are moving together quite well. The implicit assumption of constant spacing is not true over most of this range, although it becomes more nearly accurate as volumes increase. During congested conditions, even the assumption of constant speed substreams is not met. Congested conditions are usually described as stop-and-go (although slow-and-go might be more accurate). In other words, calculation of density from speed and flow is likely to be accurate only over part of the range of operating conditions. (See also the discussion in Hall and Persaud 1989.)

Equation 2.20, and its rearranged form in Equation 2.15, explicitly deal in averages, as shown by $\overline{u_s}$ in both of them. The same underlying idea is also used in theoretical work, but there the relationship is defined at a point on the time space plane. In that context, the equation is simply presented as

$$q = uk \tag{2.24}$$

This equation "may best be regarded as an idealization, which is true at a point if all measures concerned are regarded as continuous variables" (Banks 1994, 12). Both Banks and Newell (1982, 60) demonstrated this by using the three-dimensional surface proposed by Makagami et al. (1971), on which the dimensions are time (t); distance (x); and cumulative number of vehicles (N). If one assumes that the discrete steps in N can be smoothed out to allow treatment of the surface $A(x,t)$, representing the cumulative vehicle arrival function, as a continuous function, then

$$q = \frac{\partial A}{\partial t} \tag{2.25}$$

$$k = \frac{\partial A}{\partial x} \tag{2.26}$$

and

$$u = \frac{\partial x}{\partial t} \tag{2.27}$$

Since

$$\frac{\partial A}{\partial t} = \frac{\partial x}{\partial t} \cdot \frac{\partial A}{\partial x} \tag{2.28}$$

It follows that $q=uk$ for the continuous surface, at a point. Real traffic flows, however, are not only made up of finite vehicles surrounded by real spaces, but are inherently stochastic (Newell 1982). Measured values are averages taken from samples, and are therefore themselves random variables. Measured flows are taken over an interval of time, at a particular place. Measured densities are taken over space at a particular time. Only for stationary processes (in the statistical sense) will the time and space intervals be able to represent conditions at the same point in the time-space plane. Hence it is likely that any measurements that are taken of flow and density (and space mean speed) will not be very good estimates of the expected values that would be defined at the point of interest in the time space plane -- and therefore that Equation 2.22 (or 2.15) will not be consistent with the measured data. Density is also sometimes estimated from occupancy, using the relationship identified by Athol, as derived in the previous section (Equation 2.18). That equation is also valid only under certain conditions. Hall and Persaud (1989) identified those as being either constant speeds or constant vehicle lengths. Banks has identified the conditions more precisely as requiring both the covariance of vehicle length with the inverse of vehicle speed and the covariance of vehicle spacing with the inverse of vehicle speed to be zero. Speeds within a lane are relatively constant during uncongested flow. Hence the estimation of density from occupancy measurements is probably reasonable during those traffic conditions, but not during congested conditions. Athol's (1965) comparison of occupancy with aerial measurements of density tends to confirm this generalization. In short, once congestion sets in, there is probably no good way to estimate density; it would have to be measured.

Temporal concentration (occupancy) can be measured only over a short section (shorter than the minimum vehicle length), with presence detectors, and does not make sense over a long section. Perhaps because the concept of density has been a part of traffic measurement since at least the 1930s, there has been a consensus that density was to be preferred over occupancy as the measure of vehicular concentration. For example, Gerlough and Huber (1975, 10) called occupancy an "estimate of density", and as recently as 1990, May referred to occupancy as a surrogate for density (p. 186). Almost all of the theoretical work done prior to 1985 either ignores occupancy, or else uses it only to convert to, or as a surrogate for, density. Athol's work (1965) is a notable exception to this. On the other hand, much of the freeway traffic management work during the same period (i.e. practical as opposed to theoretical work) relied on occupancy, and some recent theoretical work has used it as well. (See in particular Sections 2.3.5 and 2.3.6 below.)

It would be fair to say that the majority opinion at present remains in favor of density, but that a minority view is thatoccupancy should begin to enter theoretical work instead of density. There are two principal reasons put forward by the minority for making more use of occupancy. The first is that there should be improved correspondence between theoretical and practical work on freeways. If freeway traffic management makes extensive use of a variable that freeway theory ignores, the profession is the poorer. The second reason is that density, as vehicles per length of road, ignores the effects of vehiclength and traffic composition. Occupancy, on the other hand, is directly affected by both of these variables, and therefore gives a more reliable indicator of the amount of a road being used by vehicles. There are also good reasons put forward by the majority for the continued use of density in theoretical work. Not least is that it is theoretically useful in their work in a way that occupancy is not. Hurdle (1994) has drawn an analogy with the concept of acceleration in Newtonian physics. Acceleration was impossible to measure directly in Newton's time, and difficult to measure even indirectly, yet he built a theory of mechanics in which it is one of the most fundamental parameters. In all likelihood there will continue to be analysts who use each of occupancy and density: this is a debate that will not be resolved.

2.3 Traffic Stream Models

This section provides an overview of work to establish the relationship among the variables described in the previous section. Some of these efforts begin with mathematical models; others are primarily empirical, with little or no attempt to generalize. Both are important components for understanding the relationships. Two aspects of these efforts are emphasized here: the measurement methods used to obtain the data; and the location at which the measurements were obtained. Section 2.3.1 discusses the importance of location to the nature of the data obtained; subsequent sections then deal with the models, first for two variables at a time, then for all three variables simultaneously.

2.3.1 Importance of Location to the Nature of the Data

Almost all of the models to be discussed represent efforts to explain the behavior of traffic variables over the full range of operation. In turning the models from abstract representations into numerical models with specific parameter values, an important practical question arises: can one expect that the data collected will cover the full range that the model is intended to cover? If the answer is no, then the difficult question follows of how to do curve fitting (or parameter estimation) when there may be essential data missing.

This issue can be explained more easily with an example. At the risk of oversimplifying a relationship prior to a more detailed discussion of it, consider the simple representation of the speed-flow curve as shown in Figure 2.2, for three distinct sections of roadway. The underlying curve is assumed to be the same at all three locations. Between locations A and B, a major entrance ramp adds considerable traffic to the road. If location B reaches capacity due to this entrance ramp volume, there will be a backup of traffic on the mainstream, resulting in stop-and-go traffic at location A. These vehicles can be considered to be in a queue, waiting their turn to be served by the bottleneck section immediately downstream of the entrance ramp. The data superimposed on graph A reflect the situation whereby traffic at A had not reached capacity before the added ramp volume caused the backup. There is a good range of uncongested data (on the top part of the curve), and congested data concentrated in one area of the lower part of the curve. The volumes for that portion reflect the capacity flow at B less the entering ramp flows.

At location B, the full range of uncongested flows is observed, right out to capacity, but the location never becomes congested, in the sense of experiencing stop-and-go traffic. It does, however, experience congestion in the sense that speeds are below those observed in the absence of the upstream congestion. Drivers arrive at the front end of the queue moving very slowly, and accelerate away from that point, increasing speed as they move through the bottleneck section. This segment of the speed-flow curve has been referred to as queue discharge flow (Hall et al. 1992). The particular speed observed at B will depend on how far it is from the front end of the queue (Persaud and Hurdle 1988a). Consequently, the only data that will be observed at B are on the top portion of the curve, and at some particular speed in the queue discharge segment.

If the exit ramp between B and C removes a significant portion of the traffic that was observed at B, flows at C will not reach the levels they did at B. If there is no downstream situation similar to that between A and B, then C will not experience congested operations, and the data observable there will be as shown in Figure 2.2. None of these locations taken alone can provide the data to identify the full speed-flow curve. Location C can help to identify the uncongested portion, but cannot deal with capacity, or with congestion. Location B can provide information on the uncongested portion and on capacity.

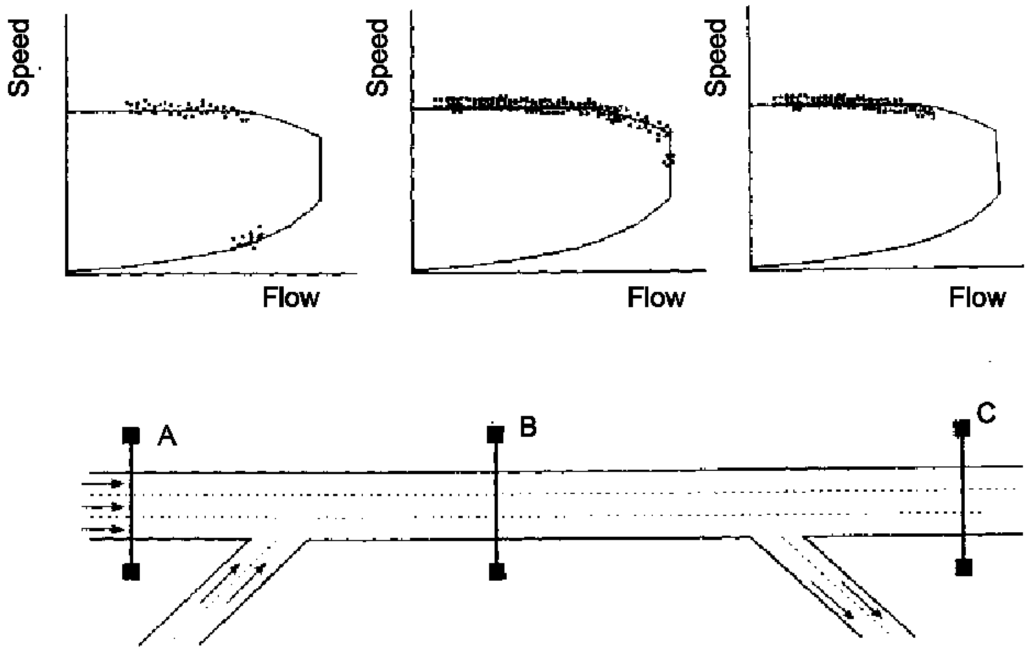

Figure 2.2
Effect of Measurement Location on Nature of Data
(Similar to figures in May 1990 and Hall et al. 1992).

This would all seem obvious enough. A similar discussion appears in Drake et al. (1967). It is also explained by May (1990). Other aspects of the effect of location on data patterns are discussed by Hsu and Banks (1993). Yet a number of important efforts to fit data to theory have ignored this key point (ie. Ceder and May 1976; Easa and May 1980).

They have recognized that location A data are needed to fit the congested portion of the curve, but have not recognized that at such a location data are missing that are needed to identify capacity. Consequently, discussion in the remaining subsections will look at the nature of the data used in each study, and at where the data were collected (with respect to bottlenecks) in order to evaluate the theoretical ideas. As will be discussed with reference to specific models, it is possible that the apparent need for several different models, or for different parameters for the same model at different locations, or even for discontinuous models instead of continuous ones, arose because of the nature (location) of the data each was using.

2.3.2 Speed-Flow Models

The speed-flow relationship is the bivariate relationship on which there has been the greatest amount of work within the past half-dozen years, with over a dozen new papers, so it is the first one to be discussed here. This sub-section is structured retrospectively, working from the present backwards in time. The reason for this structure is that the current understanding provides some useful insights for interpreting earlier work.

Prior to the writing of this chapter, the Highway Capacity and Quality of Service Committee of the Transportation Research Board approved a revised version of Chapter 3 of the Highway Capacity Manual (HCM 1994). This version contains the speed-flow curve shown in Figure 2.3. This curve has speeds remaining flat as flows increase, out to somewhere between half and two-thirds of capacity values, and a very small decrease in speeds at capacity from those values. The curves in Figure 2.3

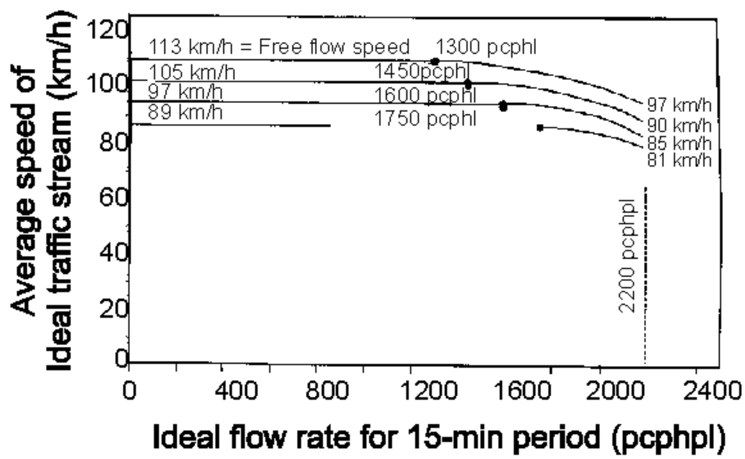

Figure 2.3
Speed-Flow Curves Accepted for 1994 HCM.

do not represent any theoretical equation, but instead represent a generalization of empirical results. In that fundamental respect, the most recent research differs considerably from the earlier work, which tended to start from hypotheses about first principles and to consult data only late in the process.

The bulk of the recent empirical work on the relationship between speed and flow (as well as the other relationships) was summarized in a paper by Hall, Hurdle, and Banks (1992). In it, they proposed the model for traffic flow shown in Figure 2.4. This figure is the basis for the background speed-flow curve in Figure 2.2, and the discussion of that figure in Section 2.3.1 is consistent with this relationship.

It is perhaps useful to summarize some of the antecedents of the understanding depicted in Figure 2.4. The initial impetus came from a paper by Persaud and Hurdle (1988a), in which they demonstrated (Figure 2.5) that the vertical line for queue discharge flow in Figure 2.4 was a reasonable result of measurements taken at various distances downstream from a queue. (This study was an outgrowth of an earlier one by Hurdle and Datta (1983) in which they raised a number of questions about the shape of the speed-flow curve near capacity.) Further impetus for change came from work done on multi-lane rural highways that led in 1992 to a revised Chapter 7 of the HCM (1992). That research, and the new Chapter 7, suggested a shape for those roads very like that in Figure 2.4, whereas the conventional wisdom for freeways, as represented in the 1985 HCM (Figure 2.6) was for lower speeds and a lower capacity. Since it seemed unlikely that speeds and capacities of a freeway could be improved by replacing grade-separated overpasses or interchanges by at-grade intersections (thereby turning a rural freeway into a multi-lane rural highway), there was good reason to reconsider the situation for a freeway.

Additional empirical work dealing with the speed-flow relationship was conducted by Banks (1989, 1990), Hall and Hall (1990), Chin and May (1991), Wemple, Morris and May (1991), Agyemang-Duah and Hall (1991) and Ringert and Urbanik (1993). All of these studies supported the idea that speeds remain nearly constant even at quite high flow rates. Another of the important issues they dealt with is one that had been around for over thirty years (Wattleworth 1963): is there a reduction in flow rates within the bottleneck at the time that the queue forms upstream? Figure 2.4 shows such a drop on the basis of two studies. Banks (1991a, 1991b) reports roughly a three percent drop from pre-queue flows, on the basis of nine days of data at one site in California. Agyemang-Duah and Hall (1991) found about a 5 percent decrease, on the basis of 52 days of data at one site in Ontario. This decrease in flow is often not observable, however, as in many locations high flow rates do not last long enough prior to the onset of congestion to yield the stable flow values that would show the drop.

The 1994 revision of the figure for the HCM (Figure 2.3) elaborates on the top part of Figure 2.4, by specifying the curve for different free-flow speeds. Two elements of these curves

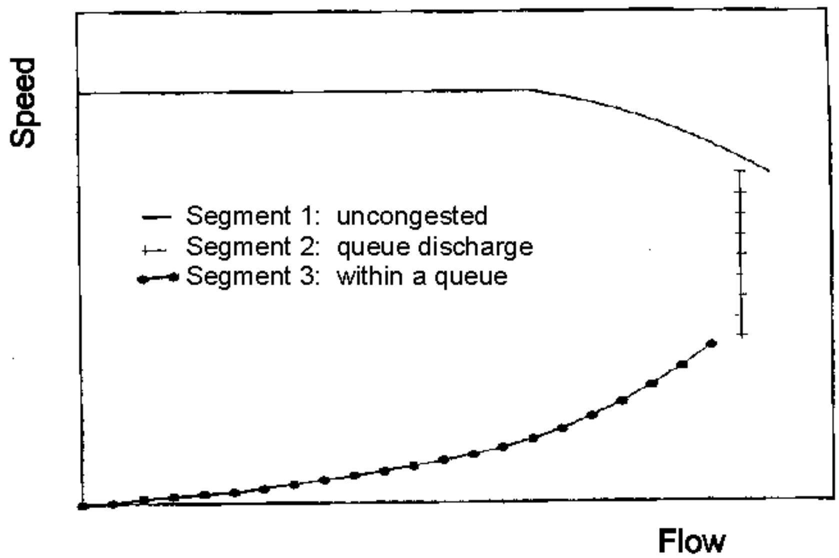

Figure 2.4
Generalized Shape of Speed-Flow Curve
Proposed by Hall, Hurdle, & Banks
(Hall et al. 1992).

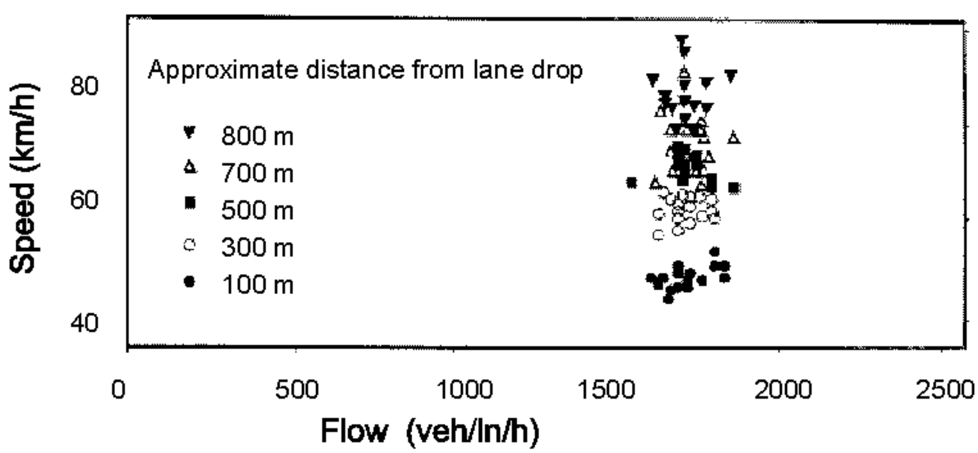

Figure 2.5
Speed-Flow Data for Queue Discharge Flow at Varied
Distances Downstream from the Head of the Queue
(Modified from Persaud and Hurdle 1988).

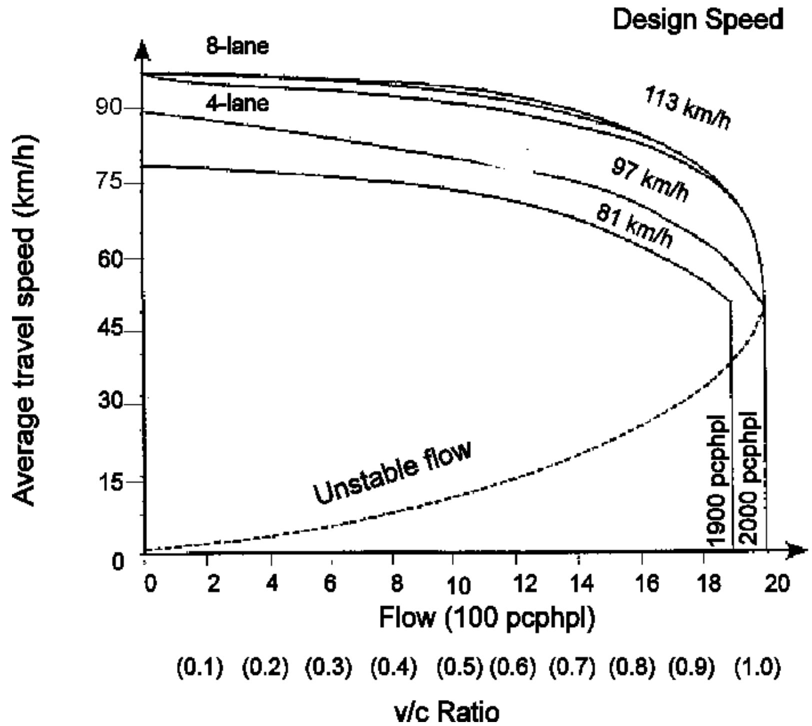

Figure 2.6
1985 HCM Speed-Flow Curve (HCM 1985).

were assumed to depend on free-flow speed: the breakpoints at which speeds started to decrease from free-flow, and the speeds at capacity. Although these aspects of the curve were only assumed at the time that the curves were proposed and adopted, they have since received some confirmation in a paper by Hall and Brilon (1994), which makes use of German Autobahn information, and another paper by Hall and Montgomery (1993) drawing on British experience.

Two empirical studies conducted recently in Germany support the general picture in Figure 2.3 quite well. Heidemann and Hotop (1990) found a piecewise-linear 'polygon' for the upper part of the curve (Figure 2.7). Unfortunately, they did not have data beyond 1700 veh/hr/lane, and had that only for two lanes per direction, so could not address what happens at capacity. Stappert and Theis (1990) conducted a major empirical study of speed-flow relationships on various kinds of roads. However, they were interested only in estimating parameters for a specific functional form,

$$u = (A - e^{BQ})\, e^{-C} - K e^{DQ} \tag{2.29}$$

where u is speed, Q is traffic volume, C and D are constant "curvature factors" taking values between 0.2 and 0.003, and A, B, and K are parameters of the model. This function tended to give the kind of result shown by the upper curve in Figure 2.8, despite the fact that the curve does not accord well with the data near capacity. The lower curve is a polygon representation, based on an assumed shape as well. In Figure 2.8, each point represents a full hour of data, and the graph represents five months of hourly data. Note that flows in excess of 2200 veh/hr/lane were sustained on several occasions, *over the full hour*.

The relevant British data and studies are those that serve as the basis for the manual for cost-benefit analyses (COBA9 1981). The speed-flow curve in that manual is shown in Figure 2.9. While it shows a decline in speed (of 6 km/h) per each additional 1000 vehicles per hour per lane) from the first vehicle on the road, detailed inspection of the data behind that conclusion (in

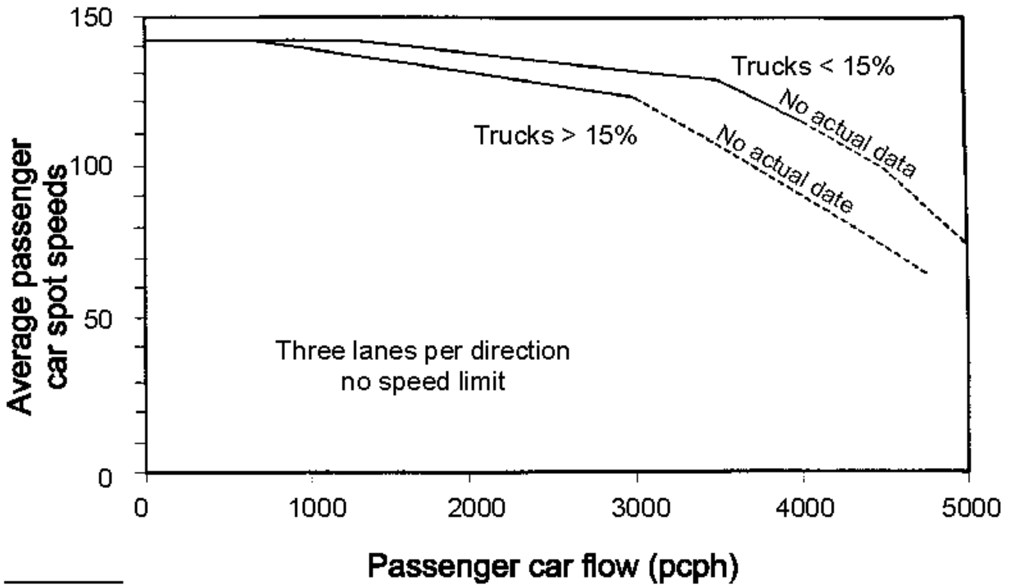

Figure 2.7
Results from Fitting Polygon Speed-Flow Curves to German Data
(Modified and translated from Heidemann and Hotop 1990).

Martin and Voorhees 1978, and in Duncan 1974) shows that the data are ambiguous, and could as easily support a slope of zero out to about the breakpoint of 1200 vphpl (Hall and Montgomery 1993). A more recent British study (Hounsell et al. 1992) also supports the notion that speeds remain high even out to capacity flows. Hence there is good international support for the type of speed-flow curve shown in Figure 2.3, and nothing to contradict the picture put forward there and in Figure 2.4.

The problem for traffic flow theory is that these curves are empirically derived. There is not really any theory that would explain these particular shapes, except perhaps for Edie et al. (1980), who propose qualitative flow regimes that relate well to these curves. The task that lies ahead for traffic flow theorists is to develop a consistent set of equations that can replicate this reality. The models that have been proposed to date, and will be discussed in subsequent sections, do not necessarily lead to the kinds of speed-flow curves that data suggest are needed.

It is instructive to review the history of depictions of speed-flow curves in light of this current understanding. Probably the seminal work on this topic was the paper by Greenshields in 1935, in which he derived the following parabolic equation for the speed-flow curve on the basis of a linear speed- density relationship together with the equation, flow = speed $*$ density:

$$q = k_j (u - \frac{u^2}{u_f}) \tag{2.30}$$

where u_f is the free-flow speed, and k_j is the jam density. Figure 2.10 contains that curve and the data it is based on, redrawn. The numbers adjacent to the data points represent the "number of 100-vehicle groups observed," on Labor Day 1934, in one direction on a two-lane two-way road (p. 464). In counting the vehicles on the road, every 10th vehicle started a new group (of 100), so there is a 90 percent overlap between two adjacent groups (p. 451). The groups are not independent observations. Equally important, the data have been grouped in flow ranges of 200 veh/h and the averages of these groups taken prior to plotting. The one congested point, representing 51 (overlapping) groups of 100 observations, came from a different roadway entirely, with different cross-section and pavement, which were collected on a different day.

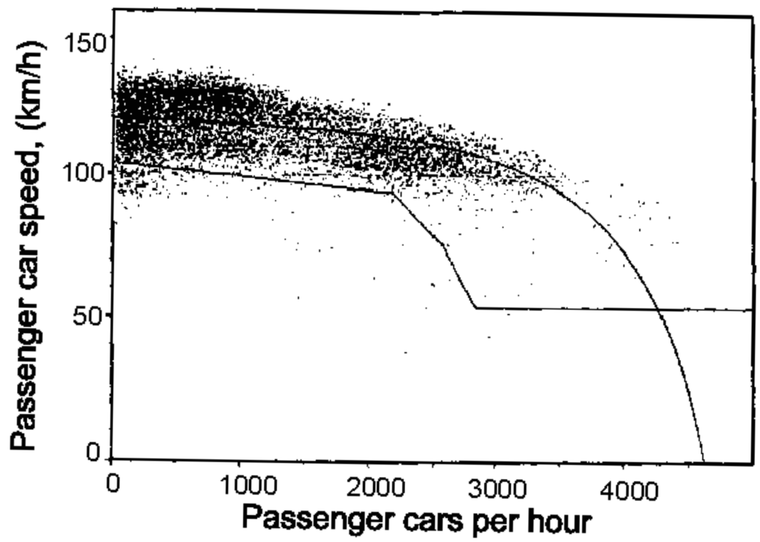

Figure 2.8
Data for Four-Lane German Autobahns (Two-Lanes per Direction), as reported by Stappert and Theis (1990).

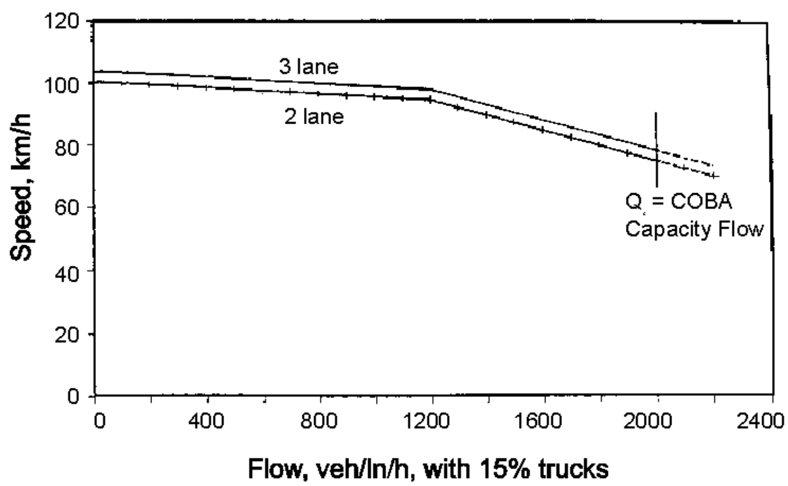

Figure 2.9
UK Speed-Flow Curve (Source: COBA9 1981).

These details are mentioned here because of the importance to traffic flow theory of Greenshields' work. The parabolic shape he derived was accepted as the proper shape of the curve for decades. In the 1965 Highway Capacity Manual, for example, the shape shown in Figure 2.10 appears exactly, despite the fact that data displayed elsewhere in the 1965 HCM showed that contemporary empirical results did not match the figure. In the 1985 HCM, the same parabolic shape was retained (Figure 2.6), although broadened considerably. It is only with the 1994 revision to the HCM that a different empirical reality has been accepted.

In short, Greenshields' model dominated the field for over 50 years, despite at least three problems. The most fundamental is that Greenshields did not work with freeway data. Yet his result for a single lane of traffic was adopted directly for freeway conditions. (This of course was not his doing.) The second problem is that by current standards of research the method of analysis of the data, with overlapping groups and averaging prior to curve-fitting, would not be acceptable. The third problem is that despite the fact that most people have used a model that was based on holiday traffic, current work focuses on regular commuters who are familiar with the road, to better ascertain what a road is capable of carrying.

There is a fourth criticism that can be addressed to Greenshields' work as well, although it is one of which a number of current researchers seem unaware. Duncan (1976; 1979) has shown that calculating density from speed and flow, fitting a line to the speed-density data, and then converting that line into a speed-flow function, gives a biased result relative to direct estimation of the speed-flow function. This is a consequence of three things discussed earlier: the non-linear transformations involved in both directions, the stochastic nature of the observations, and the inability to match the time and space measurement frames exactly.

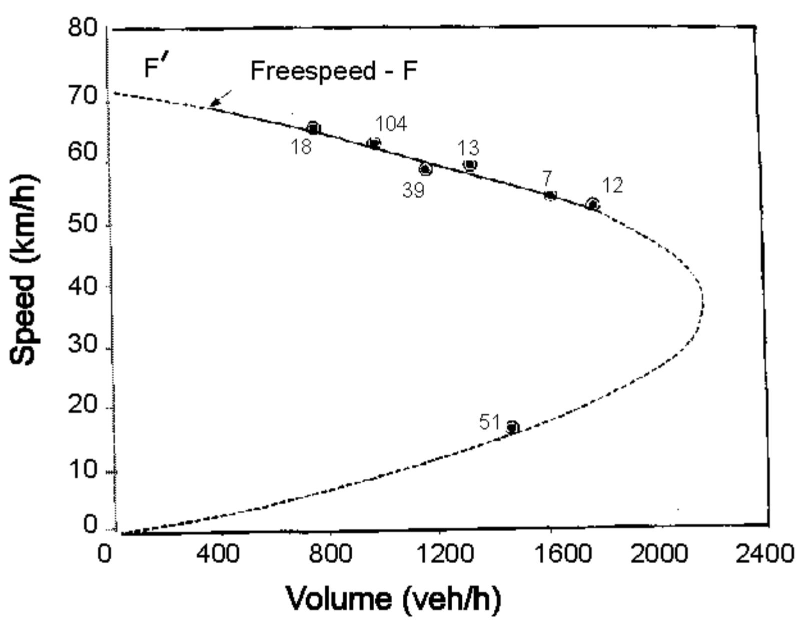

Figure 2.10
Greenshields' Speed-Flow Curve and Data
(Greenshields 1935).

It is interesting to contrast the emphasis on speed-flow models in recent years, especially for freeways, with that 20 years ago, for example as represented in TRB SR 165 (Gerlough and Huber 1975), where speed-flow models took up less than a page of text, and none of the five accompanying diagrams dealt with freeways. (Three dealt with the artificial situation of a test track.) In contrast, five pages and eleven figures were devoted to the speed-density relationship. Speed-flow models are now recognized to be important for freeway management strategies, and will be of fundamental importance for ITS implementation of alternate routing; hence there is currently considerably more work on this topic than on the remaining two bivariate topics. Twenty years ago, the other topics were of more interest. As Gerlough and Huber stated (p. 61), "once a speed-concentration model has been determined, a speed-flow model can be determined from it." That is in fact the way most earlier speed-flow work was treated (including that of Greenshields). Hence, it is sensible to turn to discussion of speed-concentration models, and to deal with any other speed-flow models as a consequence of speed-concentration work, which is the way they were developed.

.3.3 Speed-Density Models

This subsection deals with mathematical models for the

$$u = u_f (1 - k/k_j) \tag{2.31}$$

speed-density relationship, going back to as early as 1935. Greenshields' (1935) linear model of speed and density was mentioned in the previous section. It can be written as:The measured data were speeds and flows; density was calculated using Equation 2.20. The most interesting aspect of this particular model is that its empirical basis consisted of half a dozen points in one cluster near free-flow speed, and a single observation under congested conditions (Figure 2.11). The linear relationship comes from connecting the cluster with the single point. As Greenshields stated (p. 468), "since the curve is a straight line it is only necessary to determine accurately two points to fix its direction." What is surprising is not that such simple analytical methods were used in 1935, but that their results (the linear speed-density model) have continuedto be so widely accepted for so long. While there have been studies that

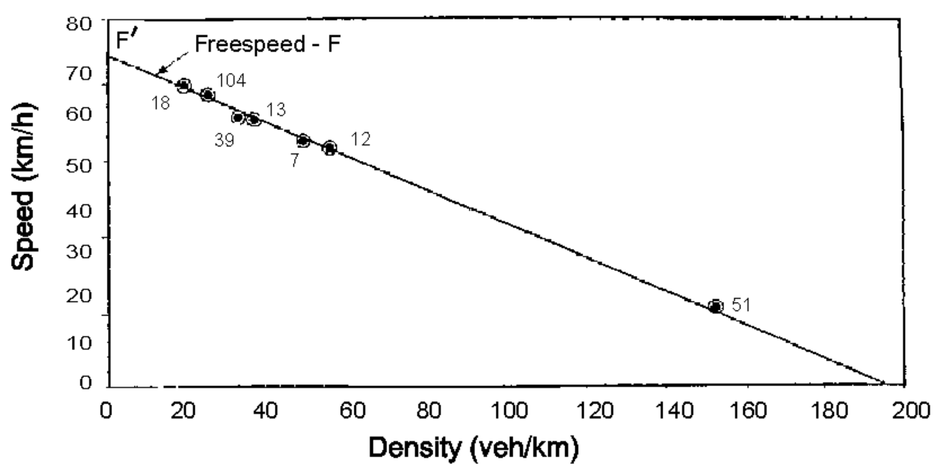

Figure 2.11
Greenshields' Speed-Density Graph and Data (Greenshields 1935).

claimed to have confirmed this model, such as that in Figure 2.12a (Huber 1957), they tended to have similarly sparse portions of the full range of data, usually omitting both the lowest flows and flow in the range near capacity. There have also been a number of studies that found contradictory evidence, most importantly that by Drake et al. (1967), which will be discussed in more detail subsequently.

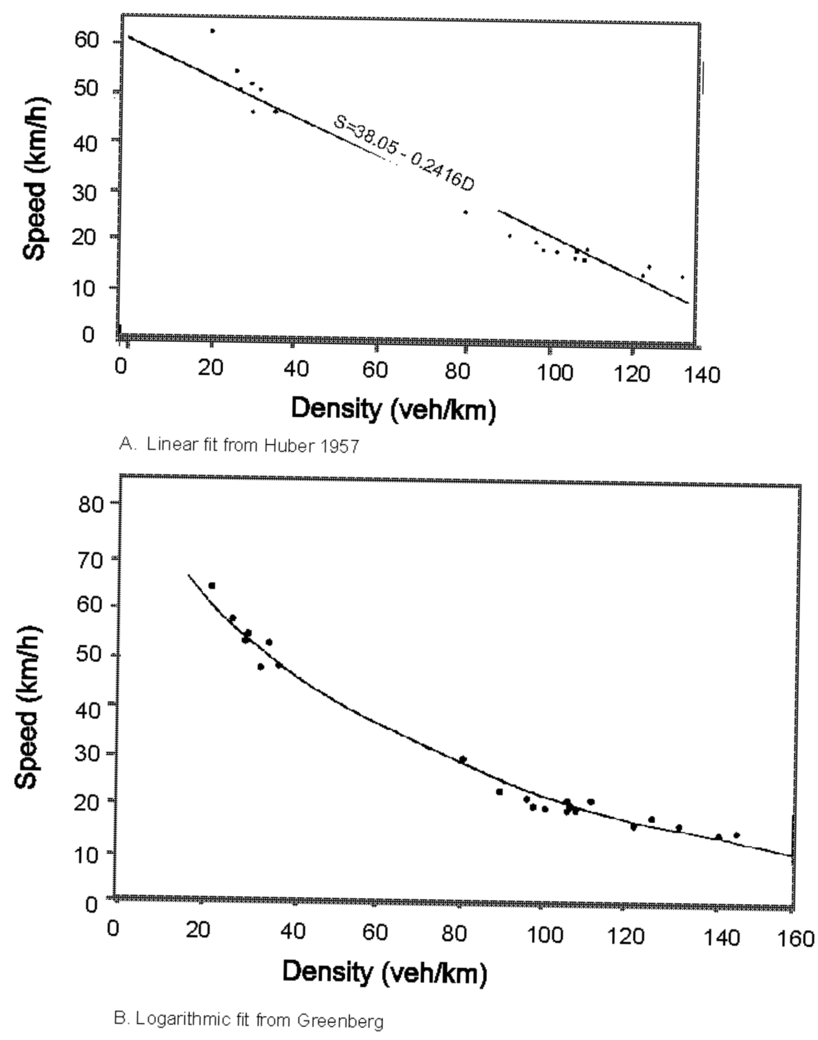

Figure 2.12
Speed-Density Data from Merritt Parkway and Fitted Curves.

A second early model was that put forward by Greenberg (1959), showing a logarithmic relationship:

$$u = c \ln(k/k_j) \tag{2.32}$$

His paper showed the fit of the model to two data sets, both of which visually looked very reasonable. However, the first data set was derived from speed and headway data on individual vehicles, which "was then separated into speed classes and the average headway was calculated for each speed class" (p. 83). In other words, the vehicles that appear in one data point (speed class) may not even have been traveling together! While a density can always be calculated as the reciprocal of average headway, when that average is taken over vehicles that may well not have been traveling together, it is not clear what that density is meant to represent. It is also the case that lane changing was not permitted in the Lincoln Tunnel (where the data were obtained), so this is really single-lane data rather than freeway data. The second data set used by Greenberg was Huber's. This is the same data that appears in Figure 2.12a; Greenberg's graph is shown in Figure 2.12b. Visually, the fit is quite good, but Huber reported an R of 0.97, which does not leave much room for improvement.

These two forms of the speed-density curve, plus five others, were investigated in an important empirical test by Drake et al. in 1967. The test used data from the middle lane of the Eisenhower Expressway in Chicago, 800 m (one-half mile) (upstream from a bottleneck whose capacity was only slightly less than the capacity of the study site. This location was chosen specifically in order to obtain data over as much of the range of operations as possible. A series of 1224 1-minute observations were initially collected. The measured data consisted of volume, time mean speed, and occupancy. Density was calculated from volume and time mean speed. A sample was then taken from among the 1224 data points in order to create a data set that was uniformly distributed along the density axis, as is assumed by regression analysis of speed on density. The intention in conducting the study was to compare the seven speed-density hypotheses statistically, and thereby to select the best one. In addition to Greenshields' linear form and Greenberg's exponential curve, the other five investigated were a two-part and a three-part piecewise linear model, Underwood's (1961) transposed exponential curve, Edie's (1961) discontinuous exponential form (which combines the Greenberg and Underwood curves), and a bell-shaped curve. Despite the intention to use "a rigorous structure of falsifiable tests" (p. 75) in this comparison, and the careful way the work was done, the statistical analyses proved inconclusive: "almost all conclusions were based on intuition alone since the statistical tests provided little decision power after all" (Drake et al., p. 76). To assert that intuition alone was the basis is no doubt a bit of an exaggeration. Twenty-one graphs help considerably in differentiating among the seven hypotheses and their consequences for both speed-volume and volume-density graphs.

Figure 2.13 provides an example of the three types of graphs used, in this case the ones based on the Edie model. Their comments about this model (p. 75) were: "The Edie formulation gave the best estimates of the fundamental parameters. While its R^2 was the second lowest, its standard error was the lowest of all hypotheses." One interesting point with respect to Figure 2.13 is that the Edie model was the only one of the seven to replicate capacity operations closely on the volume-density and speed-volume plots. The other models tended to underestimate the maximum flows, often by a considerable margin, as is illustrated in Figure 2.14, which shows the speed-volume curve resulting from Greenshields' hypothesis of a linear speed-density relationship. (It is interesting to note that the data in these two figures are quite consistent with the currently accepted speed-flow shape identified earlier in Figures 2.3 and 2.4.) The overall conclusion one might draw from the Drake et al. study is that none of the seven models they tested provide a particularly good fit to or explanation of the data, although it should be noted that they did not state their conclusion this way, but rather dealt with each model separately.

There are two additional issues that arise from the Drake et al. study that are worth noting here. The first is the methodological one identified by Duncan (1976; 1979), and discussed earlier with regard to Greenshields' work. Duncan showed that the three step procedure of (1) calculating density from speed and flow data, (2) fitting a speed-density function to that data, and then (3) transforming the speed-density function into a speed-flow function results in a curve that does not fit the original speed-flow data particularly well. This is the method used by Drake et al., and certainly most of their resulting speed-flow functions did not fit the original speed-flow data very well. Duncan's 1979 paper expanded on the difficulties to show that minor changes in the speed-density function led to major changes in the speed-flow function. This result suggests the need for further caution in using this method of double transformations to calibrate a speed-flow curve.

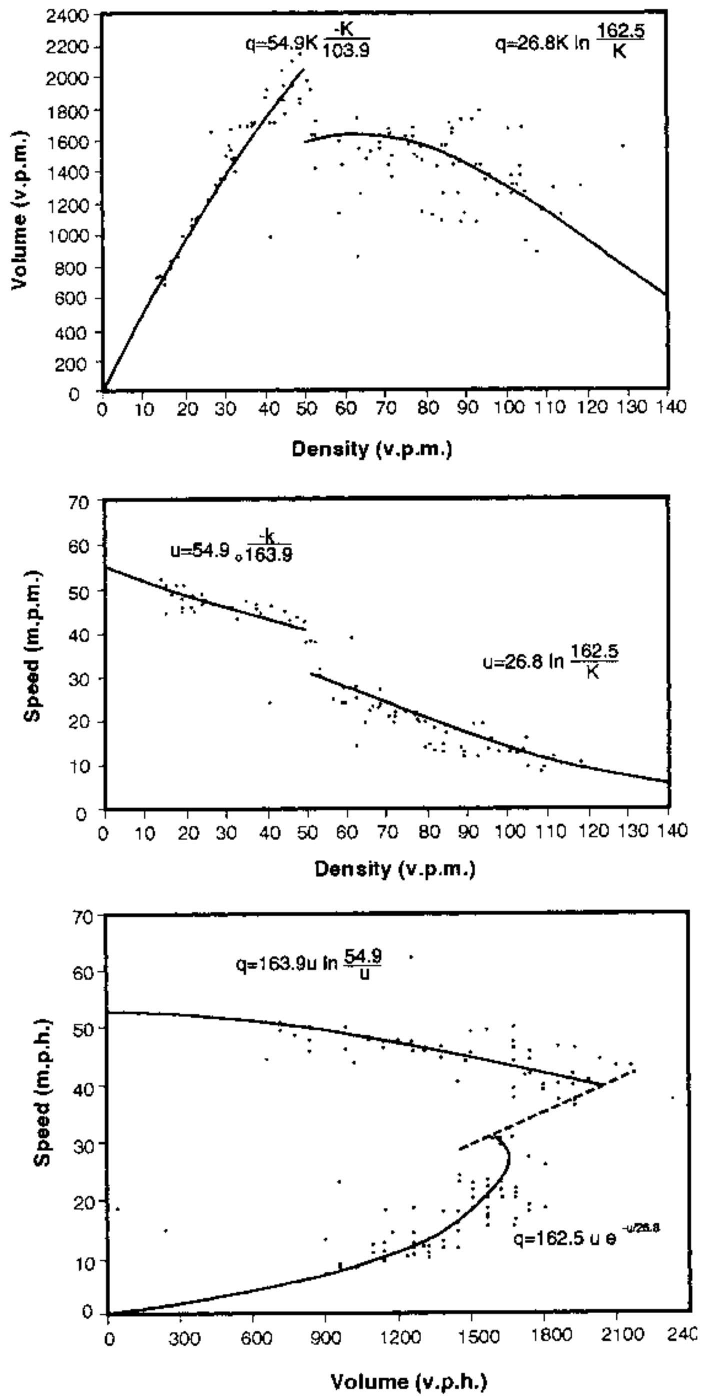

Figure 2.13
Three Parts of Edie's Hypothesis for the Speed-Density Function, Fitted to Chicago Data (Drake et al. 1967).

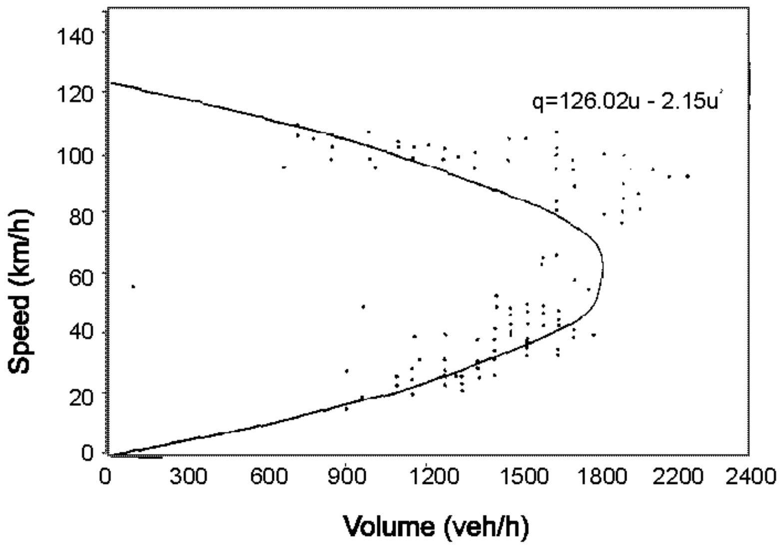

Figure 2.14
Greenshields' Speed-Flow Function Fitted to Chicago Data (Drake et al. 1967).

The second issue is the relationship between car-following models (see Chapter 4) and the models tested by Drake et al. They explicitly mention that four of the models they tested "have been shown to be directly related to specific car-following rules," and cite articles by Gazis and co-authors (1959; 1961). The interesting question to raise in the context of the overall appraisal of the Drake et al. results is whether the results raise some questions about the validity of the car-following models for freeways. The car-following models gave rise to four of the speed-density models tested by Drake et al. The results of their testing suggest that the speed-density models are not particularly good. Logic says that if the consequences of a set of premises are shown to be false, then one (at least) of the premises is not valid. It is possible, then, that the car-following models are not valid for freeways. This is not surprising, as they were not developed for this context.

2.3.4 Flow-Concentration Models

Although Gerlough and Huber did not give the topic of flow-concentration models such extensive treatment as they gave the speed-concentration models, they nonetheless thought this topic to be very important, as evidenced by their introductory paragraph for the section dealing with these models (p. 55):

> Early studies of highway capacity followed two principal approaches. Some investigators examined speed-flow relationships at low concentrations; others discussed headway phenomena at high concentrations. Lighthill and Whitham (1955) have proposed use of the flow-concentration curve as a means of unifying these two approaches. Because of this unifying feature, and because of the great usefulness of the flow-concentration curve in traffic control situations (such as metering a freeway), Haight (1960; 1963) has termed the flow-concentration curve "the basic diagram of traffic".

Nevertheless, most flow-concentration models have been derived from assumptions about the shape of the speed-concentration curve. This section deals primarily with work that has focused on the flow-concentration relationship directly. Under that heading is included work that uses either density or occupancy as the measure of concentration.

Edie was perhaps the first to point out that empirical flow-concentration data frequently have discontinuities in the vicinity of what would be maximum flow, and to suggest that therefore discontinuous curves might be needed for this relationship. (An example of his type of curve appears in Figure 2.13.) This suggestion led to a series of investigations by May and his students (Ceder 1975; 1976; Ceder and May 1976; Easa and

May 1980) to specify more tightly the nature and parameters of these "two-regime" models (and to link those parameters to the parameters of car-following models). The difficulty with their resulting models is that the models often do not fit the data well at capacity (with results similar to those shown in Figure 2.14 for Greenshields' single-regime model). In addition, there seems little consistency in parameters from one location to another. Even more troubling, when multiple days from the same site were calibrated, the different days required quite different parameters.

Koshi et al. (1983) gave an empirically-based discussion of the flow-density relationship, in which they suggested that a reverse lambda shape was the best description of the data (p.406): "the two regions of flow form not a single downward concave curve... but a shape like a mirror image of the Greek letter lamda [sic] (λ)". These authors also investigated the implications of this phenomenon for car-following models, as well as for wave propagation.

Although most of the flow-concentration work that relies on occupancy rather than density dates from the past decade, Athol suggested its use nearly 30 years earlier (in 1965). His work presages a number of the points that have come out subsequently and are discussed in more detail below: the use of volume and occupancy together to identify the onset of congestion; the transitions between uncongested and congested operations at volumes lower than capacity; and the use of time-traced plots (i.e. those in which lines connected the data points that occurred consecutively over time) to better understand the operations.

After Athol's early efforts, there seems to have been a dearth of efforts to utilize the occupancy data that was available, until the mid-1980s. One paper from that time (Hall et al. 1986) that utilized occupancy drew on the same approach Athol had used, namely the presentation of time-traced plots. Figure 2.15 shows results for four different days from the same location, 4 km upstream of a primary bottleneck. The data are for the left-most lane only (the high-speed, or passing lane), and are for 5-minute intervals. The first point in the time-connected traces is the one that occurred in the 5-minute period after the data-recording system was turned on in the morning. In Part D of the figure, it is clear that operations had already broken down prior to data being recorded. Part C is perhaps the most intriguing: operations move into higher occupancies (congestion) at flows clearly below maximum flows. Although Parts A and B may be taken to confirm the implicit assumption many traffic engineers have that operations pass through capacity prior to breakdown, Part C gives a clear indication that this does not always happen. Even more important, all four parts of Figure 2.15 show that operations do not go through capacity in returning from congested to uncongested conditions. Operations can 'jump' from one branch of the curve to the other, without staying on the curve. This same result, not surprisingly, was found for speed-flow data (Gunter and Hall 1986).

Each of the four parts of Figure 2.15 show at least one data point between the two 'branches' of the usual curve during the return to uncongested conditions. Because these were 5-minute data, the authors recognized that these points might be the result of averaging of data from the two separate branches. Subsequently, however, additional work utilizing 30-second intervals confirmed the presence of these same types of data (Persaud and Hall 1989). Hence there appears to be strong evidence that traffic operations on a freeway can move from one branch of the curve to the other without going all the way around the capacity point. This is an aspect of traffic behavior that none of the mathematical models discussed above either explain or lead one to expect. Nonetheless, the phenomenon has been at least implicitly recognized since Lighthill and Whitham's (1955) discussion of shock waves in traffic, which assumes instantaneous jumps from one branch to the other on a speed-flow or flow-occupancy curve. As well, queuing models (e.g. Newell 1982) imply that immediately upstream from the back end of a queue there must be points where the speed is changing rapidly from the uncongested branch of the speed-flow curve to that of the congested branch. It would be beneficial if flow-concentration (and speed-flow) models explicitly took this possibility into account.

One of the conclusions of the paper by Hall et al. (1986), from which Figure 2.15 is drawn, is that an inverted 'V' shape is a plausible representation of the flow-occupancy relationship. Although that conclusion was based on limited data from near Toronto, Hall and Gunter (1986) supported it with data from a larger number of stations. Banks (1989) tested their proposition using data from the San Diego area, and confirmed the suggestion of the inverted 'V'. He also offered a mathematical statement of this proposition and a behavioral interpretation of it (p. 58):

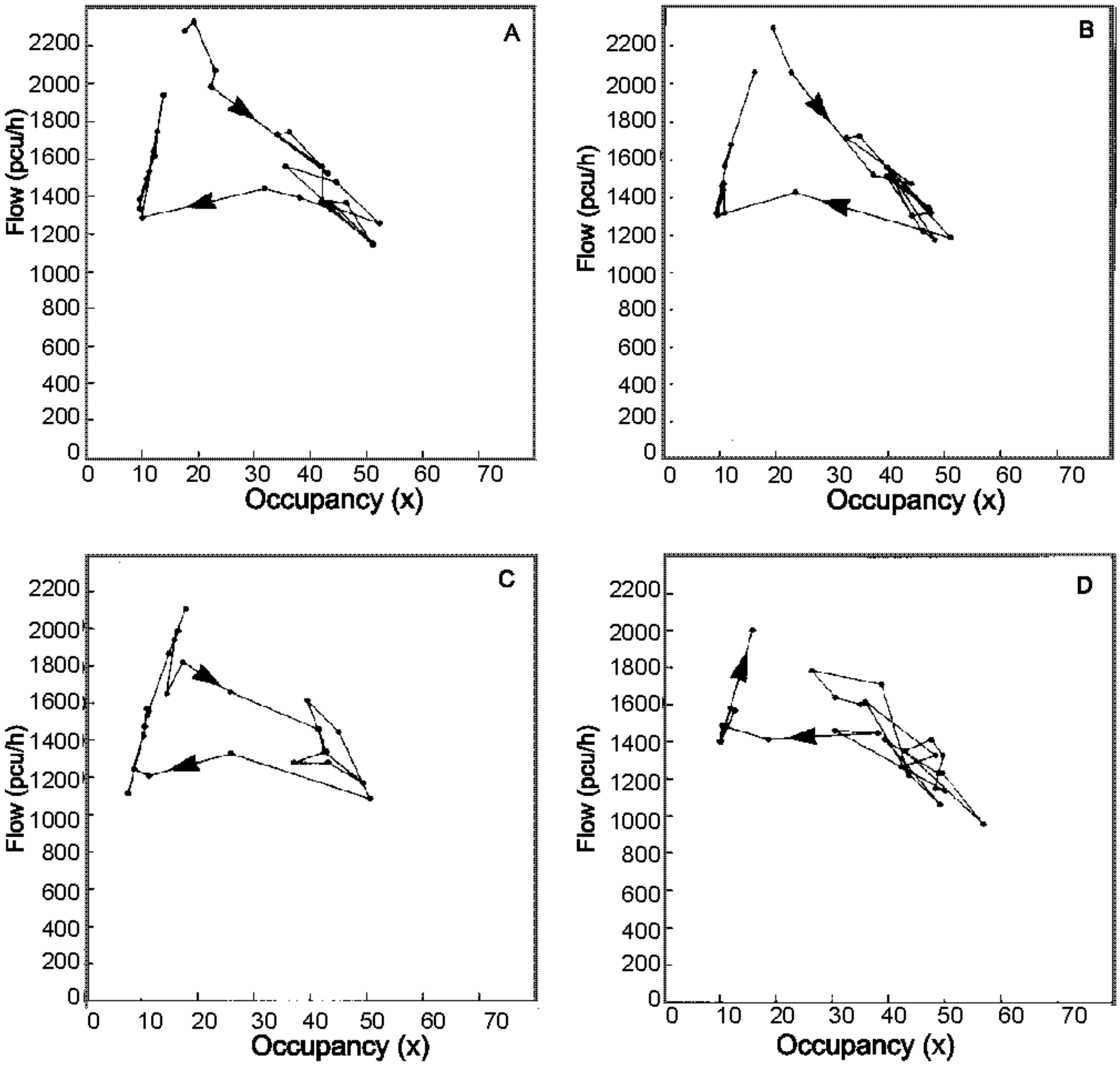

Figure 2.15
Four Days of Flow-Occupancy Data from Near Toronto (Hall et al. 1986).

The inverted-V model implies that drivers maintain a roughly constant average time gap between their front bumper and the back bumper of the vehicle in front of them, provided their speed is less than some critical value. Once their speed reaches this critical value (which is as fast as they want to go), they cease to be sensitive to vehicle spacing....

2.3.5 Three-Dimensional Models

There has not been a lot of work that attempts to treat all three traffic flow variables simultaneously. Gerlough and Huber presented one figure (reproduced as Figure 2.16) that represented all three variables, but said little about this, other than (1) "The model must be on the three-dimensional surface u = q/k," and (2) "It is usually more convenient to show the model of (Figure 2.16) as one or more of the three separate relationships in two dimensions..." (p. 49). As was noted earlier, however, empirical observations rarely accord exactly with the relationship $q=u\ k$, especially when the observations are taken during congested conditions. Hence focusing on the two-dimensional relationships will not often provide even implicitly a valid three-dimensional relationship.

these figures, and the letters on Figure 2.19, a quote from the original paper is helpful (p. 101).

> The original computer work and the photography for the original report (Gilchrist 1988) were in color, with five different colors representing different speed ranges. For this paper, black and white were alternated for the five speed ranges, which allows each of them to stand out clearly in many figures. The figure (2.19) is a good example.... Area A contains the data with speeds above 80 km/hr. Area B (light lines) covers the range 70 to 80 km/hr; area C (dark lines) the range 60 to 70 km/hr; area D (light lines) the range 50 to 60 km/hr; and area E the range below 50 km/hr.

One of the conclusions drawn by Gilchrist and Hall was that "conventional theory is insufficient to explain the data", and that the data were more nearly consistent with an alternative model based on catastrophe theory (p. 99). A different approach to three-dimensional modeling was presented by Makagami et al. (1971), as discussed in Section 2.2.3 above (and Equations 2.25 through 2.27). In that model, the dimensions were time,

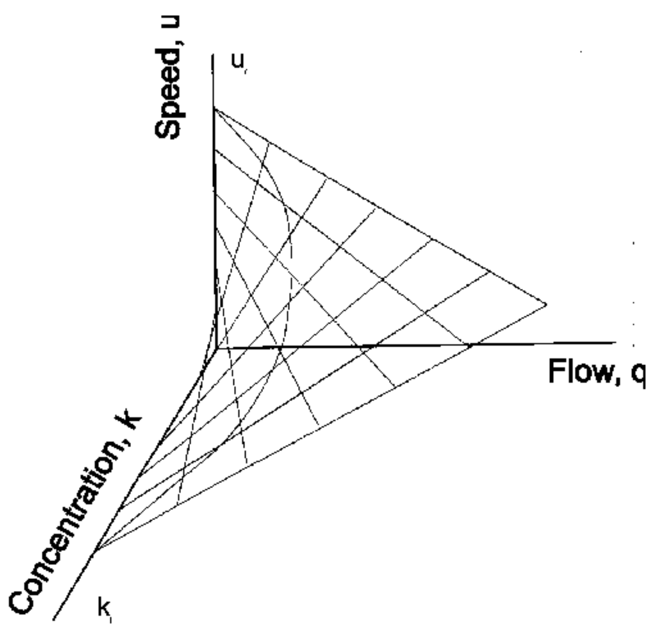

Figure 2.16
The Three-Dimensional Surface for Traffic Operations, as in Transportation Research Board Special Report 165 (Gerlough and Huber 1975).

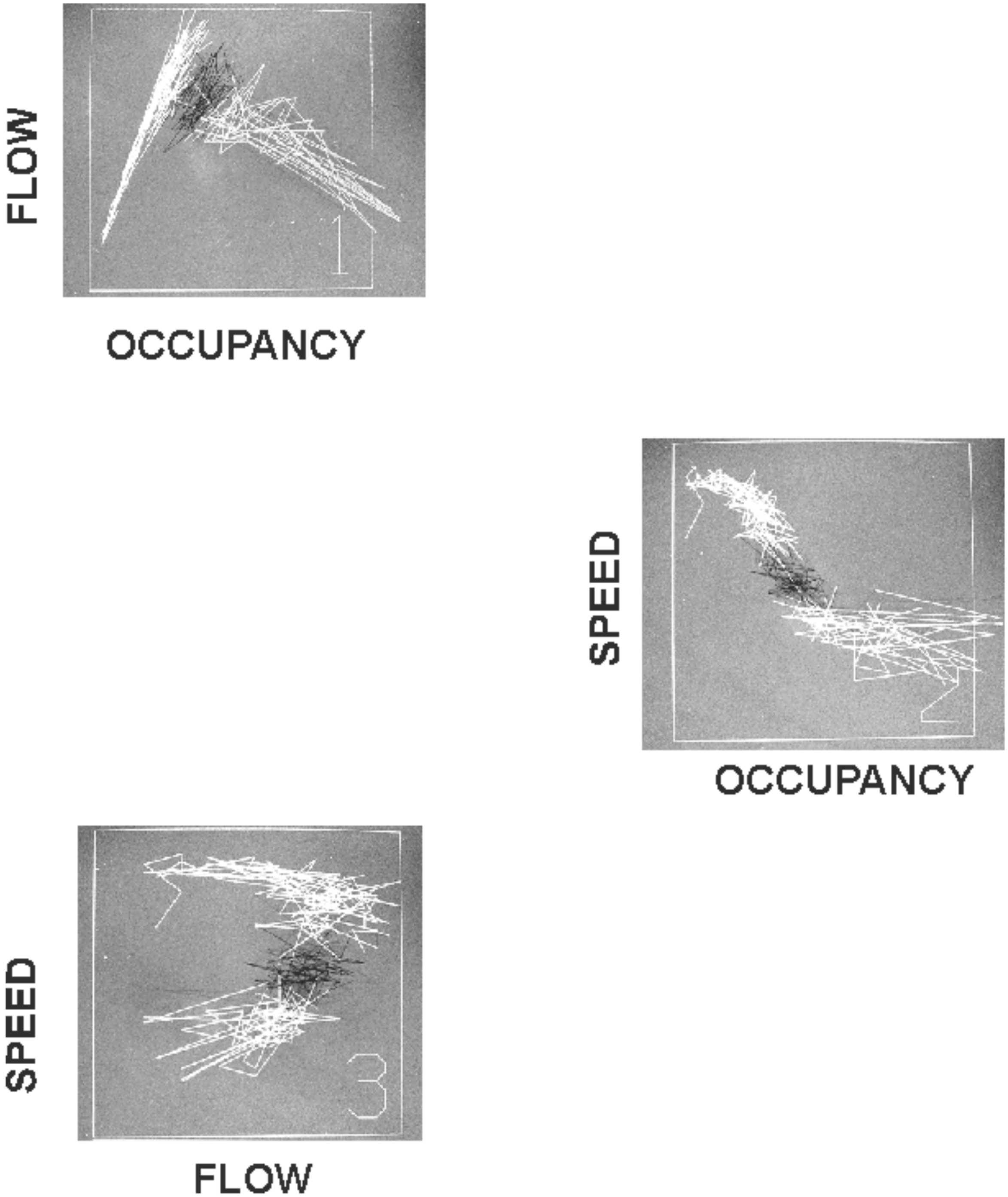

Figure 2.17
Two-Dimensional Projection of Data Used in Three-Dimensional Study
(Gilchrist and Hall 1989).

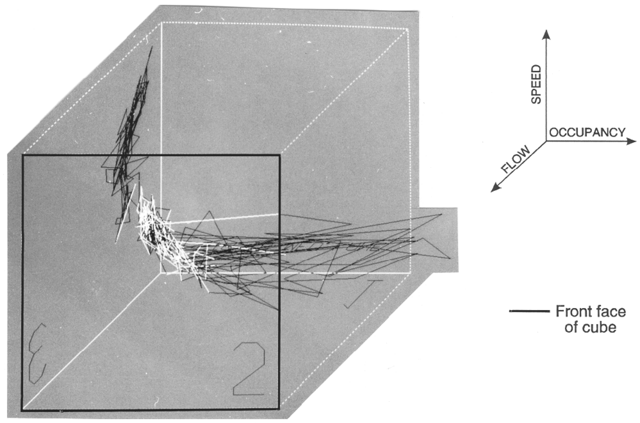

Figure 2.18
One Perspective on the Three-Dimensional Speed-Flow-Concentration Relationship (Gilchrist and Hall 1989).

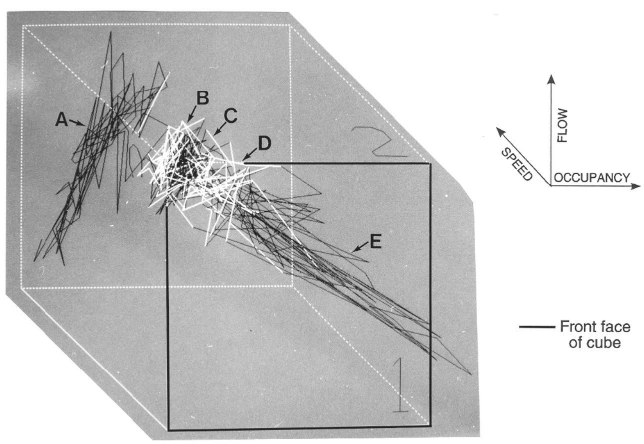

Figure 2.19
Second Perspective on the Three-Dimensional Relationship (Gilchrist and Hall 1989).

distance, and cumulative vehicle count. The derivatives of the surface representing the cumulative count are speed, flow, and density. This three-dimensional model has been applied by Newell (1993) in work on kinematic waves in traffic. In addition, Part I of his paper contains some historical notes on the use of this approach to modeling.

One recent approach to modeling the three traffic operations variables directly has been based on the mathematics of catastrophe theory. (The name comes from the fact that while most of the variables being modeled change in a continuous fashion, at least one of the variables can make sudden discontinuous changes, referred to as catastrophes by Thom (1975), who originally developed the mathematics for seven such models, ranging from two dimensions to eight.) The first effort to apply these models to traffic data was that by Dendrinos (1978), in which he suggested that the two-dimensional catastrophe model could represent the speed-flow curve. A more fruitful model was proposed by Navin (1986), who suggested that the three-dimensional 'cusp' catastrophe model was appropriate for the three traffic variables.

The feature of the cusp catastrophe surface that makes it of interest in the traffic flow context is that while two of the variables (the control variables) exhibit smooth continuous change, the third one (the state variable) can undergo a sudden 'catastrophic' jump in its value. Navin suggested that speed was the variable that underwent this catastrophic change, while flow and occupancy were the control variables. While Navin's presentation was primarily an intuitive one, without recourse to data, Hall and co-authors picked up on the idea and attempted to flesh it out both mathematically and empirically. Figure 2.20 shows the current visualization of the model.

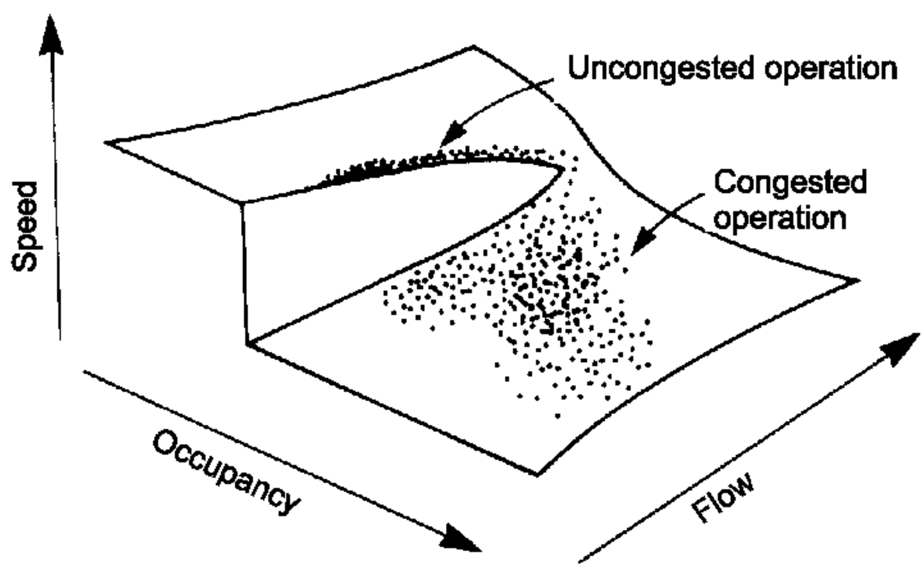

Figure 2.20
Conceptualization of Traffic Operations on a Catastrophe Theory Surface Using the Maxwell Convention (Persaud and Hall 1989).

Acha-Daza and Hall (1993) compared the effectiveness of the catastrophe theory model for estimating speeds with four of the models discussed above: Greenshields'; Greenberg's; Edie's; and the double linear regime model. The comparison was done using a data set in which all three variables had been measured, so that the speeds calculated using each model could be compared with actual measured values. Typical results for the catastrophe theory model (Figure 2.21, which yielded an R^2 of 0.92) can be compared with those for Edie's model (Figure 2.22; R^2 of 0.80), which had been found to be best by Drake et al. (1967). Although the Greenshields' and double-linear model resulted in higher R^2 values (0.87 and 0.89 respectively) than did Edie's, both models gave very clustered speed estimates, with few predictions in the 60-80 km/h range, and a similar set of points below the diagonal in the observed range of 60 to 80 km/h. It is worth noting that real data show very few observations in the range of 60 - 80 km/h also, so in that respect, both models are effective. Pushkar et al. (1994) extended this idea to develop a method that could be used for estimating speeds from single-loop detector data. The method involves estimating parameters for the model from at least one location at which speeds have been measured, and transferring those parameters to other nearby locations.

The catastrophe theory model has received some confirmation from its ability to replicate speed measurements. It has two added advantages as well on the intuitive level. First, it illustrates graphically that freeway operations do not have to stay on the (e.g. speed-flow) curve; jumps are possible from one branch to the other, and when they occur, there will be sudden changes in speeds. Second, it also illustrates graphically the fact that different locations will yield different types of data (see Figure 2.2) in that at some locations the data will go around the discontinuity in the surface, while at others the data will cross directly over the discontinuity. The catastrophe theory model provides a consistent way to explain

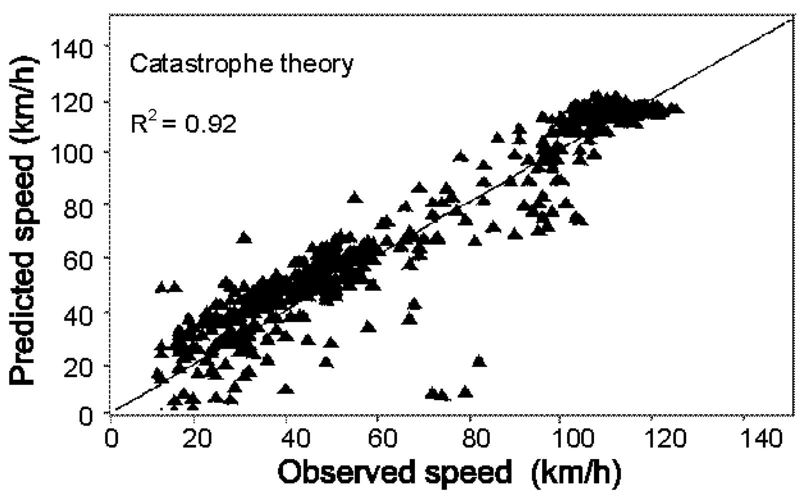

Figure 2.21
Comparison of Observed Speed with Speeds Estimated using Catastrophe Theory Model (Acha-Daza and Hall 1994).

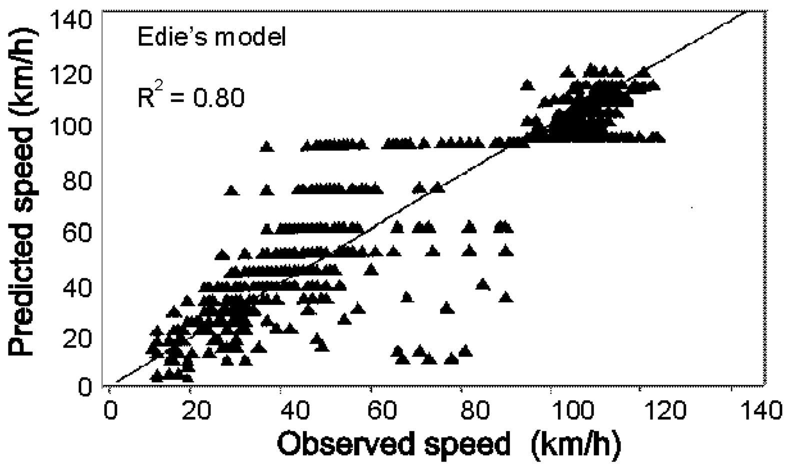

Figure 2.22
Comparison of Observed Speeds with Speeds Estimated by Edie's Model (Acha-Daza and Hall 1994).

2.3.6 Conclusions About Traffic Stream Models

The current status of mathematical models for speed-flow-concentration relationships is in a state of flux. The models that dominated the discourse for nearly 30 years are incompatible with the data currently being obtained, and with currently accepted depictions of speed-flow curves, but no replacement models have yet been developed. Part of the reason is probably that many theoreticians continue to work with density, whereas the empirical data are in terms of occupancy. The relation between those two measures of concentration is sufficiently weak that efforts to transform one into the other only muddy the picture further. The other problem was noted by Duncan (1976; 1979): transforming variables, fitting equations, and then transforming the equations back to the original variables can lead to biased results, and is very sensitive to small changes in the initial curve-fitting.

Recognition of three-dimensional relationships is also important for improved understanding. Consequently, it is important to make more use of those sets of freeway data in which all three variables have been measured and no estimation is needed, and to work with practitioners to ensure that there are more data sets for which all three variables have been measured. The models from the mid-60s (and earlier) do not measure up to those data that are available; it is not clear whether the newer models such as catastrophe theory will ultimately be any more successful.

Despite those words of caution, it is important to note that there have been significant advances in understanding traffic stream behavior since the publication in 1975 of the last TRB Special Report on Traffic Flow Theory (Gerlough and Huber). For example, the speed-flow relationship shown in Figure 2.3 is considerably different from the one in the 1965 Highway Capacity Manual, which was still accepted in 1975. The recognition that there are three distinct types of operation, as shown in Figure 2.4, will affect future analysis of traffic stream behavior.

Since the appearance of the 1985 Highway Capacity Manual, there has been a sizeable amount of research on traffic stream models, which has led to a different understanding of how traffic operates, especially on freeways. Efforts to implement ITS, with regard to both traffic management and traffic information provision, will provide challenges for applying this improvement in understanding. Equally important, ITS will likely provide the opportunity for acquiring more and better data to further advance understanding of these fundamental issues.

References

Acha-Daza , J. A. and F. L. Hall (1993). *A Graphical Comparison of the Predictions for Speed Given by Catastrophe Theory and Some Classic Models.* Transportation Research Record 1398, Transportation Research Board (TRB), National Research Council (NRC), Washington, DC, pp. 119-124.

Acha-Daza, J. A. and F. L. Hall. (1994). *The Application of Catastrophe Theory to Traffic Flow Variables.* Transportation Research Record 28B, NRC, Washington, DC, pp. 235-250.

Agyemang-Duah, K. and F. L. Hall. (1991). *Some Issues Regarding the Numerical Value of Freeway Capacity, in Highway Capacity and Level of Service.* Proceedings of the International Symposium on Highway Capacity, Karlsruhe. (U. Brannolte, ed.) Rotterdam, Balkema, pp. 1-15.

Ardekani, S. and R. Herman (1987). *Urban Network-wide Traffic Variables and their Relations*. Transportation Science 21, pp. 1-16.

Athol, P. (1965). *Interdependence of Certain Operational Characteristics Within a Moving Traffic Stream*. Highway Research Record, 72, pp. 58-87.

Banks, J. H. (1989). *Freeway Speed-Flow Concentration Relationships: More Evidence and Interpretations.* Transportation Research Record, 1225, TRB, NRC, Washington, DC, pp. 53-60.

Banks, J. H. (1990). *Flow Processes at a Freeway Bottleneck.* Transportation Research Record 1287, TRB, NRC, Washington, DC, pp. 20-28.

Banks, J. H. (1991a). *Two-Capacity Phenomenon at Freeway Bottlenecks: A Basis for Ramp Metering?* Transportation Research Record 1320, TRB, NRC, Washington, DC, pp. 83-90.

Banks, J. H. (1991b). *The Two-Capacity Phenomenon: Some Theoretical Issues*. Transportation Research Record 1320, TRB, NRC, Washington, DC, pp. 234-241.

Banks, J. H. (1994). *The Relationship of Measured to Calculated Speeds and the "Fundamental Relationship": Some Comments.* Paper submitted to the Transportation Research Board for its January 1995 meeting.

Box, P. C. and J. C. Oppenheimer (1976). Manual of Traffic Engineering Studies, Fourth Edition. Arlington, VA: Institute of Transportation Engineers. Note that this edition will soon be superseded by the 5th Edition: Manual of Transportation Engineering Studies, edited by D. Robertson, to be published in 1994 by Prentice-Hall for the Institute.

Ceder, A. (1975). *Investigation of Two-Regime Traffic Flow Models at the Micro- and Macro-Scopic Levels*. Ph.D. dissertation, University of California, Berkeley, California.

Ceder, A. (1976). *A Deterministic Traffic Flow Model for the Two-Regime Approach*. Transportation Research Record 567, TRB, NRC, Washington, DC, pp. 16-32.

Ceder, A. and A. D. May (1976). *Further Evaluation of Single- and Two-Regime Traffic Flow Models.* Transportation Research Record 567, TRB, NRC, Washington, DC, pp. 1-15.

Chin, H. C. and A. D. May (1991). *Examination of the Speed-Flow Relationship at the Caldecott Tunnel*. Transportation Research Record 1320, TRB, NRC, Washington, DC.

COBA9 (1981). *COBA9 (Cost-Benefit Analysis) Manual*. U. K. Department of Transport, Assessment Policy and Methods Division, London.

Collins, J. F. (1983). *Automatic Incident-Detection Experience with Trl Algorithm HIOCC*. Transport and Road Research Laboratory, TRRL Supplementary Report 775, Crowthorne, England.

Dendrinos, D. S. (1978). *Operating Speeds and Volume to Capacity Ratios: The Observed Relationship and the Fold Catastrophe*. Transportation Research Record 12, TRB, NRC, Washington, DC, pp. 191-194.

Drake, J. S. , J. L. Schofer, and A. D. May (1967). *A Statistical Analysis of Speed Density Hypotheses.* Highway Research Record, 154, pp. 53-87.

Drew, Donald R. (1968). *Traffic Flow Theory and Control*. New York: McGraw-Hill.

Duncan, N. C. (1974). *Rural Speed/Flow Relations*. Transport and Road Research Laboratory, TRRL LR 651, Crowthorne, Berkshire, England.

Duncan, N. C. (1976). *A Note on Speed/Flow/Concentration Relations*. Traffic Engineering and Control, pp. 34-35.

Duncan, N. C. (1979). *A Further Look at Speed/ Flow/ Concentration*. Traffic Engineering and Control, pp. 482-483.

Easa, S. M. and A. D. May (1980). *Generalized Procedures for Estimating Single- and Two-Regime Traffic Flow Models*. Transportation Research Record 772, TRB, NRC, Washington, DC, pp. 24-37.

Edie, L. C. (1961). *Car Following and Steady-State Theory for Non-Congested Traffic*. Operations Research, 9, pp. 66-76.

Edie, L. C. (1974). *Flow theories, in Traffic Science,* D. C. Gazis, Editor, Wiley, New York.

Edie, L. C., R. Herman, and T. Lam (1980). *Observed Multilane Speed Distributions and the Kinetic Theory of Vehicular Traffic*. Transportation Science, 14, pp. 55-76.

Gazis, D. C., R. Herman, and R. Potts (1959). *Car-Following Theory of Steady-State Traffic Flow.* Operations Research, 7, pp. 499-505.

Gazis, D. C., R. Herman, and R.W. Rothery (1961). *Nonlinear Follow-The-Leader Models of Traffic Flow.* Operations Research, 9, pp. 545-567.

Gerlough, D. L. and M. J. Huber (1975). *Traffic Flow Theory: a Monograph.* Special Report 165, Transportation Research Board (Washington DC: National Research Council).

Gilchrist, R. (1988). *Three-Dimensional Relationships in Traffic Flow Theory Variables.* Unpublished Master's degree report, Department of Civil Engineering, McMaster University, Hamilton, Ontario, Canada.

Gilchrist, R. S. and F. L. Hall (1989). *Three-Dimensional Relationships among Traffic Flow Theory Variables*. Transportation Research Record 1225,TRB, NRC, Washington, DC, pp. 99-108.

Greenberg, H. (1959). *An Analysis of Traffic Flow*. Operations Research, Vol 7, pp. 78-85.

Greenshields, B. D. (1935). *A Study of Traffic Capacity*. Highway Research Board Proceedings 14, pp. 448-477.

Gunter, M. A. and F. L. Hall (1986). *Transitions in the Speed-Flow Relationship*. Transportation Research Record 1091, TRB, NRC, Washington, DC, pp. 18-21.

Haight, F. A. (1960). *The Volume, Density Relation in the Theory of Road Traffic*. Operations Research 8, pp. 572-573.

Haight, F. A. (1963). *Mathematical Theories of Traffic Flow*. (New York: Academic Press)

Hall, F. L., B. L. Allen, and M. A. Gunter (1986). *Empirical Analysis of Freeway Flow-Density Relationships.* Transportation Research 20A, TRB, NRC, Washington, DC, pp. 197-210.

Hall, F. L. and M. A. Gunter (1986). *Further Analysis of the Flow-Concentration Relationship*. Transportation Research Record 1091, TRB, NRC, Washington, DC, pp. 1-9.

Hall, F. L. and B. N. Persaud (1989). *An Evaluation of Speed Estimates Made with Single-Detector Data from Freeway Traffic Management Systems.* Transportation Research Record 1232, TRB, NRC, Washington, DC, pp. 9-16.

Hall, F. L. and L. M. Hall (1990). *Capacity and Speed Flow Analysis of the Queue in Ontario*. Transportation Research Record 1287, TRB, NRC, Washington, DC, pp. 108-118.

Hall, F. L, V. F. Hurdle, and J. H. Banks (1992). *Synthesis of Recent Work on the Nature of Speed-Flow and Flow-Occupancy (Or Density) Relationships on Freeways.* Transportation Research Record 1365, TRB, National Research Council, Washington, DC, pp. 12-18.

Hall, F. L. and F. O. Montgomery (1993). *Investigation of an Alternative Interpretation of the Speed-Flow Relationship for UK Motorways*. Traffic Engineering and Control, Volume 34, No. 9, September, pp. 420-425.

Hall, F. L. and W. Brilon (1994). *Comparison of Uncongested Speed-Flow Relationships in German Autobahn and North American Freeway Data*. Transportation Research Record 1457 , TRB, NRC, Washington, DC, pp. 35-42.

Heidemann, D. and R. Hotop (1990). *Verteilung der Pkw-Geschwindigkeiten im Netz der Bundesautobahnen-Modellmodifikation und Aktualisierung (Distribution of passenger-car speeds on the Autobahn network -- model modifications and actualization).* Straße und Autobahn - Heft 2, pp. 106-113.

HCM (1965). *Special Report 87: Highway Capacity Manual*, Transportation Research Board.

HCM (1985). *Special Report 209: Highway Capacity Manual*. Transportation Research Board.

HCM (1992). *Chapter 7, Multilane Rural and Suburban Highways.* A Replacement for the Chapter That Appeared in the Third (1985) Edition of the HCM. Transportation Research Board. Washington, DC: National Academy of Sciences.

HCM (1994). *Committee A3A10, Highway Capacity and Quality of Service*. Highway Capacity Manual, Chapter 3, "Basic Freeway Segments". TRB, National Research Council, Washington, DC, 1993.

Hounsell, N. B., S. R. Barnard, M. McDonald, and D. Owens (1992). *An Investigation of Flow Breakdown and Merge Capacity on Motorways.* TRL Contractor Report 338. Transport and Road Research Laboratory, Crowthorne, Berkshire.

Hsu, P. and J. H. Banks (1993). *Effects of Location on Congested-Regime Flow-Concentration Relationships for Freeways*. Transportation Research Record 1398, TRB, NRC, Washington, DC, pp. 17-23.

Huber, M. J. (1957). *Effect of Temporary Bridge on Parkway Performance.* Highway Research Board Bulletin 167, pp. 63-74.

Hurdle, V. F. (1994). Personal communication.

Hurdle, V. F. and P. K. Datta (1983). *Speeds and Flows on an Urban Freeway: Some Measurements and a Hypothesis.* Transportation Research Record 905, TRB, NRC, Washington, DC, pp. 127-137.

ITE (1976). Institute of Transportation Engineers. *Transportation and Traffic Engineering Handbook*, J.E. Baerwald, Editor. Prentice-Hall.

ITE (1992). Institute of Transportation Engineers. *Traffic Engineering Handbook, 4th edition*, J.L. Pline, Editor. Prentice-Hall.

Kennedy, N., J. H. Kell, and W. Homburger (1973). *Fundamentals of Traffic Engineering, 8th edition*. Institute of Transportation an Traffic Engineering, Berkeley, CA.

Koshi, M., M. Iwasaki, and I. Okhura (1983). *Some Findings and an Overview on Vehicular Flow Characteristics.* Proceedings, 8th International Symposium on Transportation and Traffic Flow Theory (Edited by Hurdle, V.F., Hauer, E. and Steuart, G.F.) University of Toronto Press, Toronto, Canada, pp. 403-426.

Kühne, R. and S. Immes (1993). *Freeway Control Systems for Using Section-Related Traffic Variable Detection*. Proceedings of the Pacific Rim TransTech Conference, Seattle. New York: American Society of Civil Engineering, pp. 56-62.

Leutzbach, W. (1972). *Einffuhrung in die Theorie des Verkehrsflusses*, Springer, Berlin.

Leutzbach, W. (1987). *Introduction to the Theory of Traffic Flow*. Springer-Verlag, New York.

Lighthill, M. J. and G. B. Whitham (1955). *On Kinematic Waves: Ii. A Theory of Traffic Flow on Long Crowded Roads*. Proceedings of the Royal Society A229 No. 1178, pp. 317-145.

Makagami, Y., G. F. Newell, and R. Rothery (1971). *Three-Dimensional Representation of Traffic Flow.* Transportation Science, Volume 5, No. 3, pp. 302-313.

Martin & Voorhees Associates. (1978). *A Study of Speed / Flow Relationships on Rural Motorways and All-Purpose Dual Carriageways*. Prepared for Department of Transport: TRRL and Traffic Engineering Division.

May, A. D. (1990). *Traffic Flow Fundamentals.* Prentice-Hall.

May, A. D., P. Athol, W. Parker, and J. B. Rudden (1963). *Development and Evaluation of Congress Street Expressway Pilot Detection System.* Highway Research Record, 21, pp. 48-70.

Morton, T. W. and C. P. Jackson (1992). *Speed/Flow Geometry Relationships for Rural Dual-Carriageways and Motorways.* TRRL Contractor Report, 279. Transport and Road Research Laboratory, Crowthorne, Berkshire, England.

Navin, F. (1986). *Traffic Congestion Catastrophes.* Transportation Planning Technology, 11, pp. 19-25.

Newell, G. F. (1982). *Applications of Queueing Theory*, Second Edition. London: Chapman and Hall.

Newell, G. F. (1993). *A Simplified Theory of Kinematic Waves in Highway Traffic*. Transportation Research B, 27B: Part I, General theory, pp. 281-287; Part II, Queueing at freeway bottlenecks, pp. 289-303; Part III, Multi-destination flows, pp. 305-313.

Payne, H. J. and S. C. Tignor (1978). *Incident-Detection Algorithms Based on Decision Trees with States*. Transportation Research Record 682, TRB, NRC, Washington, DC, pp. 30-37.

Persaud, B. N. and F. L. Hall (1989). *Catastrophe Theory and Patterns in 30-Second Freeway Data -- Implications for Incident Detection.* Transportation Research 23A, TRB, NRC, Washington, DC, pp. 103-113.

Persaud, B. N. and V. F. Hurdle (1988a). *Some New Data That Challenge Some Old Ideas about Speed-Flow Relationships.* Transportation Research Record 1194, TRB, NRC, Washington, DC, pp. 191-198.

Persaud, B. N. and V. F. Hurdle (1988b). *Freeway Data from Time-Lapse Aerial Photography -- an Alternative Approach*. Transportation Science, 22, pp. 251-258.

Pushkar, A., F. L. Hall, and J. A. Acha-Daza (1994). *Estimation of Speeds from Single-Loop Freeway Flow and Occupancy Data Using the Cusp Catastrophe Theory Model.* Transportation Research Record 1457, TRB, NRC, Washington, DC, pp. 149-157.

Ringert, J. and T. Urbanik, II (1993). *Study of Freeway Bottlenecks in Texas*. Transportation Research Record 1398, TRB, NRC, Washington, DC, pp. 31-41.

Rothrock, C. A. and L. A. Keefer (1957). *Measurement of Urban Traffic Congestion*. Highway Research Board Bulletin 156, pp. 1-13.

Back to the main Publication page.

Stappert, K. H. and T. J. Theis (1990). *Aktualisierung der Verkehrsstärke- /Verkehrsgeschwindigkeitsbeziehungen des BVWP-Netzmodells (Actualization of the speed-flow relations for the federal Transportation Masterplan Network).* Research Report VU18009V89 Heusch/Boesefeldt, Aachen.

Thom, R. (1975). *Structural Stability and Morphogenesis*. English translation by D.H. Fowler of *Stabilite Structurelle et Morphogenese*. Benjamin: Reading, MA.

Underwood, R. T. (1961). *Speed, Volume, and Density Relationships: Quality and Theory of Traffic Flow.* Yale Bureau of Highway Traffic, pp. 141-188, as cited in Drake et al. 1967.

Wardrop, J.G. (1952). *Some Theoretical Aspects of Road Traffic Research.* Proceedings of the Institution of Civil Engineers, Part II, Volume I, pp. 325-362.

Wardrop, J. G. and G. Charlesworth (1954). *A Method of Estimating Speed and Flow of Traffic from a Moving Vehicle.* Proceedings of the Institution of Civil Engineers, Part II, Vol. 3, pp. 158-171.

Wattleworth, J. A. (1963). *Some Aspects of Macroscopic Freeway Traffic Flow Theory.* Traffic Engineering, 34: 2, pp. 15-20.

Wemple, E. A., A. M. Morris, and A. D. May (1991). *Freeway Capacity and Level of Service Concepts.* Highway Capacity and Level of Service, Proceedings of the International Symposium on Highway Capacity, Karlsruhe, (U. Brannolte, Ed.) Rotterdam: Balkema, pp. 439-455.

Wright, C. (1973). *A Theoretical Analysis of the Moving Observer Method.* Transportation Research 7, TRB, NRC, Washington, DC, pp. 293-311.