

Theory of Traffic Flow— Statistical Models

J. C. OPPENLANDER
Assistant Professor of Civil Engineering
Purdue University

INTRODUCTION

The formulation of a science involves four distinct levels of development. Initially the technically-curious observers start with the collection and reduction of pertinent data about the phenomena of interest to them. From a mathematical point-of-view this step is analogous to the discipline of descriptive statistics. Research engineers desire to describe the central tendency, variability, and shape of the field and laboratory data that they are observing.

As the engineering scientists develop an intrinsic understanding of this growing science, they conduct simple experiments by trial-and-error methods. At this second stage of the development workable controls are devised for the important components that comprise the scientific system involved. Significant components are identified from the descriptive data collected in the exploratory investigations.

Intelligible patterns or sensible explanations are detected from the immense collection of information in the third phase of the development of a science. The techniques of statistical inference, both estimation and significance testing, are valuable aides for explaining the relationships and interactions among the many variables of the complex system. In addition, modern-day computers permit large quantities of observed data to be analyzed and reduced into simple statistical models. These models are the most important results produced at this level of scientific growth. Today the major research efforts in traffic engineering are concerned with the development of statistical models. With these models the traffic engineer has a quasi-mathematical formulation of measured conditions, and he can also make reasonable predictions of unmeasured conditions through generalizations deduced for the entire population.

The final level in the evolution of a science concerns the development of a workable theory or theories that accurately describe the scientific system or its various components. These theories must adopt the essential features of thought and terminology which have characterized the growth of this science. First, a particular problem must be resolved into

a concept that appears reasonable in comparison with real-world experiences. This conceptual analysis involves defining and delimiting the nature and scope of the problem to be investigated. After the concept has been fully described, its elements must be formulated into a mathematical model; that is, expressions are developed to describe the conceptual model in quantitative terms. This model must be verified by actual field or laboratory experiments to determine if it behaves in a way similar to the system being investigated. In general, a specific science has many theories applicable to a reasonable explanation of its growing and complex knowledge.

The purposes of this paper are to provide an introduction to the theory of traffic flow, which is the final phase in the development of the science of traffic engineering, and to illustrate several statistical models that have been developed to describe various traffic-stream characteristics.

TRAFFIC-FLOW THEORY

The subject of traffic-flow theory has been formulated only to a very limited extent. This lack of application of theoretical considerations to traffic movement is largely explained by the extreme complexity of vehicular traffic and traffic problems, by the concentration of technical efforts to upgrade quickly an inadequate highway transportation system, and by the general absence of research and engineering personnel concerned mainly with developing the theory of traffic flow. In fact, applied mathematicians are primarily responsible for the traffic-flow theories that have been developed.

Various theoretical concepts have been expressed as mathematical models to describe the complex phenomenon of traffic flow. These methods of quantitatively depicting traffic flow are classified under the following general approaches: statistical models, probabilistic models, continuous-flow analogies, car-following concept, queuing theory, traffic-network studies, mathematical experiments, and intersection situations. Although computing-machine simulation provides a valuable technique for studying uninterrupted and interrupted traffic flows, simulation is not a theory of traffic flow. Rather, complex traffic situations can be analyzed and synthesized by the tool of simulation.

The motion of vehicular traffic is not only governed externally by the physical laws of nature, but it is further complicated internally by driver behavior. Thus, theories of traffic flow must evolve from the combined application of the knowledge afforded by both human-behavioral and physical sciences to the man-machine system of highway transportation.

Because the *a priori* knowledge of the theory of traffic flow is rather limited, considerable understanding of traffic-stream characteristics can be gained through the study of statistical models.

FUNDAMENTAL CONSIDERATIONS

The theory of traffic flow is concerned with the four-dimensional movement of discrete, man-machine objects over a roadway network. The dimensions include transverse and longitudinal positions, elevation, and real time. The following general statements about the development of traffic-flow theory can be summarized from the theoretical and applied research that has been performed:

1. No single, general theory will ever be devised to describe the complete system of traffic movement. Theories are being formulated to represent specific traffic situations. At the most, computer simulation permits the synthesis of several elementary situations.
2. The discipline of traffic-flow theory is concerned only with the mathematical representation of stream and intersection characteristics. Land-use generation, growth, and assignment models are generally not included as theories of traffic flow, although the traffic estimates derived from these planning models are essential for determining the vehicular volumes on the various sections of the highway network.
3. Complete descriptions of traffic flow must be statistical to account for the variations within and among the various drivers and vehicles which comprise the discrete objects moving in the traffic stream or through an intersection. However, various components of the traffic-flow model do not need to be statistical in nature.
4. Theories of traffic flow must be limited to descriptions of central-tendency measures of traffic-stream characteristics. It would be extremely difficult to evaluate mathematical models for the movement patterns of individual vehicles because of the random and complex variations exhibited within individual drivers. At best, traffic-flow models could be developed to account reasonably well for the variability among drivers and vehicles.

Several fundamental properties are essential for the accurate representation of traffic movement. The following three characteristics have been cited by F. A. Haight as separating traffic flow from any other type of flow problems:

1. Finiteness—the discrete objects in the highway-transportation system are finite in length and travel at finite velocities which

vary from vehicle to vehicle and with time. In addition, the various sections of the roadway system are finite in length. Individual vehicles may overtake and pass other vehicles, but speeds are generally reduced as vehicular volumes increase. These two features are in counterdistinction to the situation of fluid flow, where molecules move in a fixed relation with each other and velocity is increased by a constraint.

2. Ambiguity—the control of the discrete objects is a function of the individual drivers and of the traffic system. That is, drivers can be controlled to a limited extent by traffic regulations and control devices, but they cannot be scheduled in regard to time or place. This characteristic insures that equilibrium exists in the system.
3. Time-space—the positions of vehicles are not identical in time and in space. This feature results from vehicles traveling on the same highway section at different speeds, in different lanes, and/or in different directions.

The continuity of traffic movement is represented by the expression that volume in vehicles per time is equal to the product of speed in

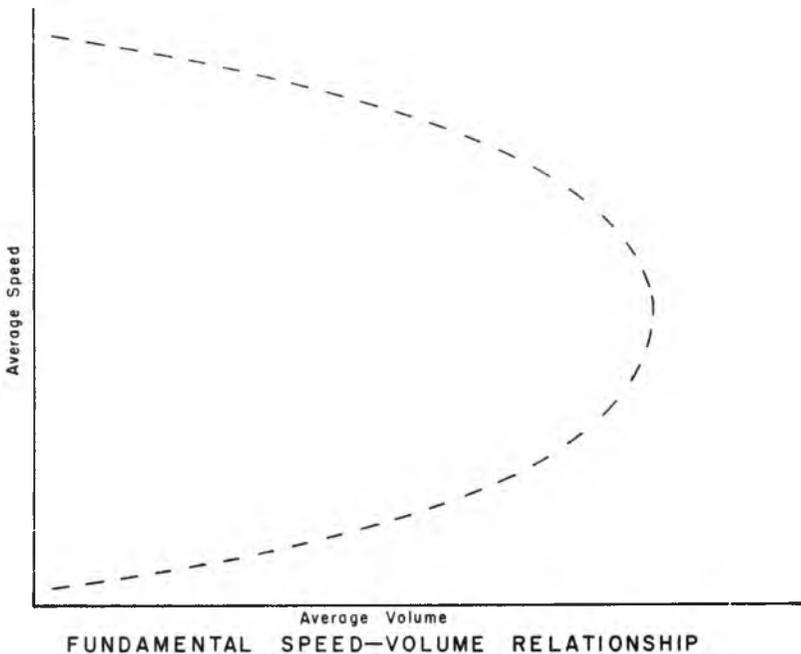


Fig. 1.

distance per time and density in vehicles per distance. This relationship is characteristic of a particular highway section with a certain population of drivers and vehicles under a given environment at a specific time. The maximum traffic volume possible is denoted as the capacity of that roadway location.

The fundamental relationships of speed and volume, speed and density, and volume and density are presented graphically in Figs. 1 to 3, respectively. The limits of these diagrams represent no flow due to the absence of any vehicles on the roadway and no flow due to the presence of a traffic jam. Highway and traffic engineering research has not completely determined these fundamental diagrams of traffic flow for the many possible combinations of driver, vehicle, roadway, traffic, and environmental variables that significantly influence the fundamental relationships. Therefore, the general shapes of these relationships are

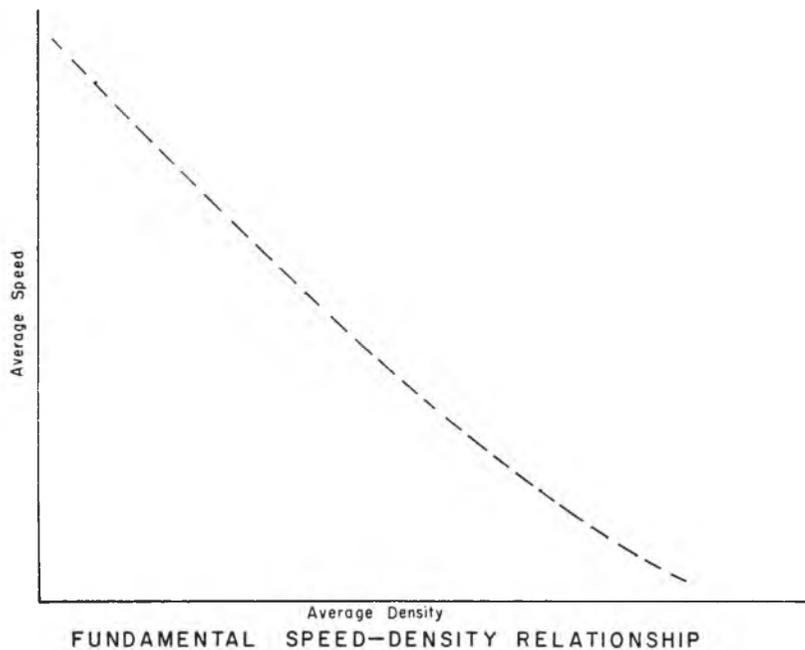


Fig. 2.

indicated by the dashed curves in the three diagrams. Research studies are urgently needed in this area of traffic-stream and intersection characteristics.

These fundamental diagrams are unique to the system of vehicular movement. Any satisfactory and realistic theory of traffic flow must be

compatible with these volume-speed-density relationships. In fact, these fundamental relations furnish the logical starting point for the development of theories of traffic flow.

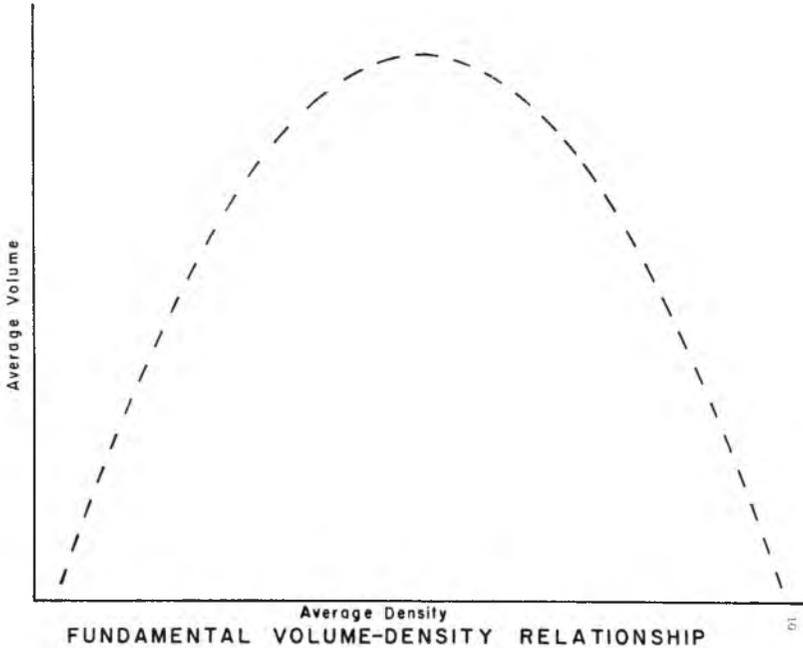


Fig. 3.

STATISTICAL MODELS

In the development of a science statistical models provide the transition from the explanation of collected data to the formulation of theories. Traffic flow is essentially a stochastic or random process. Therefore, various characteristics of traffic flow can be described by the techniques of probability and statistics.

After a statistical theory has been formulated about some traffic-stream characteristic, it must be evaluated and tested by observing samples of actual traffic flow. This process is analogous to statistical inference; that is, statistical models for the population are inferred from random and representative samples taken from this population. Statistical theories of traffic flow can be considered as estimation and hypothesis-testing models.

Because there are many topics in parametric and non-parametric statistical inference, it is possible to develop a wide variety of statistical

models in the realm of traffic flow. The evolution of these models is limited only by the imagination and ingenuity of highway and traffic engineers. Several examples of estimation and significance-testing models are presented to illustrate statistical theories of traffic-stream characteristics.

Statistical Estimation

Point or interval estimates are often desired in the evaluation of a traffic-flow parameter. In addition, statistical estimation involves the determination of a functional relationship between a dependent variable and one or more independent variables. This relationship may be linear or curvilinear, depending on the *a priori* knowledge of the subject under investigation. Thus, it is possible to predict with varying degrees of precision the value of the dependent variable when the independent variables are known or assumed.

Example No. 1—Mean Spot Speed

The following multiple linear regression equation was derived from the multivariate analysis of actual traffic flow on two-lane, rural highways:

$$S = 39.34 + 0.0267 X_1 + 0.1396 X_2 - 0.8125 X_3 - 0.1126 X_4 + 0.0007 X_5 + 0.6444 X_6 - 0.5451 X_7 - 0.0082 X_8$$

where S = mean spot speed—mph,

X_1 = out-of-state passenger cars in traffic stream—percent,

X_2 = truck combinations (tractor with one or more trailers) in traffic stream—percent,

X_3 = degree of curve—deg,

X_4 = gradient—signed percent,

X_5 = stopping sight distance—ft,

X_6 = lane width—ft,

X_7 = number of commercial roadside establishments, such as restaurants, service stations, motels, taverns, etc., per mile (counted on both sides of the roadway for one-half mile in advance of and one-half mile beyond the speed site)—no. per mile, and

X_8 = total traffic volume—vph.

The coefficient of multiple correlation was 0.788 for this investigation and was significant at the 5-percent level. The precision of this multiple estimate was measured by a standard error of estimate equal to 4.47 mph.

To illustrate an application of this statistical model, it is assumed that an advisory speed limit is to be posted on a particular horizontal curve. The values of the eight independent variables for this highway location are, respectively, 20 percent, 10 percent, 4 deg, + 2 percent,

600 ft, 12 ft, 1 per mile, and 400 vph. Solution of the multiple regression equation for these roadway and traffic conditions produces a mean spot speed of 42 mph. If one standard deviation of about 7 mph is added to the mean speed to approximate the 85th-percentile speed, then an advisory speed of 49 mph would be calculated from this analysis. Thus, the advisory speed limit on this curve would be posted as 50 mph.

Example No. 2—Accident Rate

A research study in a large metropolitan area generated the following curvilinear regression expression for the estimation of traffic accident rates:

$$A = 1.52 T^{2.25}$$

where A = accident rate—number of accidents per one million vehicle-miles of travel and

T = average travel time per mile—min. per mile.

If the average travel time on a certain arterial street was 2.0 min. per mile during the peak-hour periods, the expected accident rate would be 7.2 traffic accidents per one million vehicle-miles of travel.

Example No. 3—Intersection Capacity

A recent statistical evaluation of the approach capacities for intersections has produced the multiple linear regression equation as follows:

$$C = -4236.59 + 1731.03 X_1 - 113.87 X_2 + 373.68 X_3 - 314.49 X_4 - 171.86 X_5 + 263.14 X_6 - 65.16 X_7 + 66.18 X_8 - 13.76 X_9 + 5.20 X_{10} + 260.26 X_{11} + 245.07 X_{12} + 9.90 X_{13} + 8.84 X_{14}$$

where C = approach capacity—vph,

X_1 = ln of approach width in feet,

X_2 = coded parking density on approach,

0—none

1—few

2—medium

3—full

X_3 = ln of coded population,

1—below 50,000

2—50,000 to 99,999

3—100,000 to 249,999

4—250,000 to 499,999

5—500,000 to 999,999

6—1,000,000 and over

X_4 = ln of percent of local buses per hour per number of approach lanes,

- X_5 = ln of percent of commercial vehicles per hour per number of approach lanes,
- X_6 = ln of coded location of intersection area,
 1—central business
 2—fringe business
 3—outlying business
 4—intermediate residential
 5—outlying residential
 6—rural area
- X_7 = ln of left turns from opposite leg per hour,
- X_8 = coded adequacy of street markings,
 0—no markings or only crosswalk
 1—center line with or without crosswalk
 2—lane lines with or without other markings
- X_9 = percent of through green time,
- X_{10} = percent left-turn time per percent left turns,
- X_{11} = coded parking restriction on approach,
 0—parking restriction not in effect or for less than 100 ft from crosswalk
 1—parking restriction in effect for at least 100 ft from crosswalk
- X_{12} = ln of coded difference between approach width and exit width,
 1—exit narrower than approach by 10 ft or more
 2—exit narrower than approach by less than 10 ft
 3—exit equal to approach
 4—exit wider than approach by less than 10 ft
 5—exit wider than approach by 10 ft or more
- X_{13} = squared code of signal-cycle length, and
 1—35 to 44 sec
 2—45 to 59 sec
 3—60 to 74 sec
 4—75 to 89 sec
 5—90 to 104 sec
 6—105 to 119 sec
 7—120 to 134 sec
 8—135 sec or more
- X_{14} = percent right-turn time per percent right turns.

The complexity of the solution of this equation can be reduced with the aid of nomographs.

Hypothesis Testing

Statistical models of the hypothesis- or significance-testing type are especially useful in appraising the importance of observed statistics. These statistical tests are valuable techniques for the evaluation of before-and-after studies designed to measure the effectiveness of a traffic-engineering improvement.

In significance testing the hypothesis and alternate are first formulated. The appropriate level of significance and test statistic are then selected, and the sampling distribution of the statistic is determined. After a rejection region is chosen within the sampling distribution, the decision of accepting or rejecting the null hypothesis can be made.

Example No. 4—Travel Time

A parametric statistical test for evaluating the significance of the difference between two sample means is provided by Student's "t" test. The test statistic is obtained from the following formula:

$$t = (\bar{X}_1 - \bar{X}_2) / \sqrt{(s_1^2/N_1) + (s_2^2/N_2)}$$

where t = Student's "t" statistic,

\bar{X}_1 = sample mean,

s_1^2 = sample variance, and

N_1 = sample size.

This significance test is predicted on the assumption that the samples are drawn from normal populations. However, the condition of homogeneity of variance does not have to be assumed.

A before-and-after study was performed to measure the effectiveness of a progressive signal system in reducing travel time. Mean travel times of 10.06 and 8.31 min. and variances of 3.07 and 1.03 (min.)² were calculated for the before and after conditions, respectively. The sample size in each case was nine test runs. The computed t statistic of 2.59 is larger than the critical value of 1.75 for a 5-percent significance level and a one-tailed test. Thus, it was concluded that the progressive signal system produced a significant reduction in travel times.

Example No. 5—Speed Characteristics

Non-parametric statistics are also useful in the testing of statistical hypotheses. The sign test is a quick and easy method for comparing two selected statistics under various sets of conditions. This statistical test requires that there are pairs of observations on the two things being compared, the two observations of a given pair were made under similar conditions, and the different pairs were observed under different conditions. The formula presented below is valid for large samples of 30 or

more observations, and the "z" statistic has a normal sampling distribution:

$$z = (X - 0.5 N \pm 0.5) / 0.5 \sqrt{N}$$

where z = "z" statistic,

X = number of plus signs, and

N = sample size.

Note: use + 0.5 when $X < 0.5 N$ and - 0.5 when $X > 0.5 N$.

In applying the sign test the direction of the differences between the paired observations is noted; that is, whether the sign of the difference is plus or minus. The sampled populations in this test may have any shape and any variability.

To illustrate the use of the sign test as a statistical model in evaluating traffic-stream behavior, it is desired to determine the influence of out-of-state passenger cars on spot-speed characteristics for two-lane highways in rural areas. The hypothesis to be tested is that the mean speeds of out-of-state passenger cars are equal to those of in-state passenger cars. The mean spot speeds of these two vehicle classifications were determined at 53 study sites. The out-of-state cars had higher mean speeds than the in-state cars at 41 locations, while the reverse was evident at only 12 speed sites. The calculated z value of 3.85 is greater than the critical measure of 1.96 for a level of significance of 5 percent and a two-tailed test. Thus, the null hypothesis was rejected, and it was concluded that the speeds of in-state and out-of-state passenger cars represent different populations. Any comparisons of speed statistics among several locations on two-lane, rural highways must consider the influence of out-of-state passenger cars on spot-speed characteristics.

Example No. 6—Accident Control Charts

Quality control charts have application in appraising the safety of traffic flow. A highway section is considered to be in statistical control if all the observed variations in traffic accidents are only from chance causes. Control charts provide a graphical representation of accident patterns and indicate the extent of variation in accidents due to random and assignable causes.

Although a control chart based on the binomial distribution is probably most applicable to the study of traffic accidents, the very small values of accident rates permit the use of control charts for a Poisson distribution. The values necessary for the construction of an accident control chart can be obtained from the following expressions:

$$\begin{aligned} \bar{x} &= c \\ UCL &= c + 3\sqrt{c} \\ LCL &= c - 3\sqrt{c} \end{aligned}$$

where $\bar{\xi}$ = mean number of accidents per year,
 UCL = upper control limit,
 LCL = lower control limit,
 $c = np$ = number of accidents per year on a highway section,
 n = annual vehicle-miles of travel on a highway section,
 and
 p = accident rate—number of accidents per one million vehicle-miles of travel on a highway section.

The mean number of accidents per year is usually taken as the long-term annual average of traffic accidents on a highway section. However, the centerline of the control chart could be evaluated by a reasonable accident value for the classification of highway being studied. When the annual number of accidents exceeds the upper control limit, it is quite possible that the highway section may be out-of-control. This discrepancy indicates the need for an engineering investigation to ascertain if some assignable cause in the design or operation of the highway is responsible for the high number of traffic accidents.

A certain one-mile section of highway has an ADT of 5000 vehicles per day and a long term accident rate of 3 accidents per one million vehicle-miles of travel. A total of 10 accidents occurred on this roadway last year. The centerline, upper control limit, and lower control limit are, respectively, 5.5, 12.4, and 0.0 accidents per year. Therefore, the safety of this highway section was considered to be in-control.