TRANSPORTATION RESEARCH
CIRCULAR
Number E-C149 June 2011

75 Years of the Fundamental Diagram for Traffic Flow Theory

Greenshields Symposium

TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

TRANSPORTATION RESEARCH CIRCULAR E-C149

75 Years of the Fundamental Diagram for Traffic Flow Theory

Greenshields Symposium

July 8–10, 2008
Woods Hole, Massachusetts

Traffic Flow Theory and Characteristics Committee

June 2011

Transportation Research Board
500 Fifth Street, NW
Washington, DC 20001
www.TRB.org

TRANSPORTATION RESEARCH CIRCULAR E-C149
ISSN 0097-8515

The **Transportation Research Board** is one of six major divisions of the National Research Council, which serves as an independent adviser to the federal government and others on scientific and technical questions of national importance. The National Research Council is jointly administered by the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal.

The **Transportation Research Board** is distributing this circular to make the information contained herein available for use by individual practitioners in state and local transportation agencies, researchers in academic institutions, and other members of the transportation research community. The information in this circular was taken directly from the submission of the authors. This document is not a report of the National Research Council or of the National Academy of Sciences.

Transportation Research Board
500 Fifth Street, NW
Washington, DC 20001
www.TRB.org

Glenda J. Beal, Production Editor; A. Regina Reid, Proofreader and Layout

Preface

The TRB Traffic Flow Theory and Characteristics Committee held a midyear meeting and symposium on the Fundamental Diagram, 75 Years (Greenshields 75 Symposium), July 8–10, 2008, at the J. Erik Jonsson Conference Center of the National Academy of Sciences in Woods Hole, Massachusetts.

The theme of the symposium was "Celebrating 75 Years of the Fundamental Diagram." The main topics included historical appreciation of traffic flow theory founders; recent developments and their influence on current practice; the impact of technological developments such as real-time measurements and remote sensing; the current status of traffic flow theory; and the identification of gaps in knowledge and development of research topics to address those areas that need improvement (or lack sufficient experience). Forty-five individuals participated in the symposium and meeting.

The idea of the Greenshields 75 Symposium was originated by Reihnart Kühne of the German Aerospace Center, who championed the importance of commemorating the seminal contributions of the founders of traffic flow theory. In particular, it was an occasion to commemorate Bruce Greenshields' studies on traffic flow during the early 1930s, which culminated in the first publication on speed-flow curves, *A Study of Traffic Capacity,* presented at the 14th Annual Meeting of the Highway Research Board in 1935 and published in its proceedings. The symposium also honored the contributions of Joseph Treiterer.

A subcommittee headed by Nathan Gartner, University of Massachusetts Lowell (UMass), managed the planning and organization of the symposium and meeting, assisted by Chronis Stamatiadis and Rahul Deshpande, also of UMass. A call for contributions was distributed in late 2007, with a December 15 deadline for abstracts. Subcommittee members assembled and screened the abstracts and organized a program consisting of an opening session and four technical sessions. The opening session included four invited presentations under the title: Foundations of Traffic Flow Theory—The Fundamental Diagram. Kühne and Gartner discussed Greenshields' contributions in highway traffic and urban streets, Benekohal discussed Treiterer's legacy, and Kerner presented modern approaches to traffic flow modeling. The four technical sessions were The Fundamental Diagram: From Theory to Practice (seven papers); Measurements and Characteristics of Traffic Flow (four papers); Empirical Observations of Traffic Flow Characteristics (four papers); and Simulation and Calibration of Traffic Flow Models (seven papers). Finally, an invited panel, chaired by Hani Mahmassani of Northwestern University, titled their discussion Perspectives on Traffic Flow Theory: Beyond Greenshields.

Rich Cunard and the TRB staff are acknowledged for their support in the logistical and organizational aspects of the symposium and midyear meeting.

—Reinhart Kühne, German Aerospace Center
Symposium Chair
—Nathan H. Gartner, University of Massachusetts Lowell
Symposium Cochair

September 2009

Contents

Foundations of Traffic Flow Theory
The Fundamental Diagram

Greenshields' Legacy
Highway Traffic

REINHART D. KÜHNE
German Aerospace Center, Transportation Studies, Berlin

HISTORY

The first beginnings for traffic flow descriptions on a highway are derived from observations by Greenshields, firstly shown to the public exactly 75 years ago (*Proc., 13th Annual Meeting of the Highway Research Board, Dec. 1933*). He carried out tests to measure traffic flow, traffic density and speed using photographic measurement methods for the first time. A short look on his CV shows that Greenshields started his career as a traffic engineering scientist with this publication which leads to a PhD thesis at the University of Michigan in 1934.

How Greenshields performed the measurements is shown in Figure 1. From his original paper we read: "The field method of securing data was quite simple. A 16-mm simplex movie camera was used to take pictures. An electric motor driven by an automobile storage battery operated the camera with a constant time interval between exposures." Figure 2 shows the camera with the motor attachment. Varying the voltage by changing the battery terminals controlled the time interval, which might be varied from ½ to 2 seconds. This method was found better than rheostat control. The time interval was carefully measured with a stopwatch over a period of 40 to 100 exposures and checked by the sweep hand of a photographic timer included in the pictures. In order that moving cars might appear in at least two consecutive pictures a field of twice the space traveled per time interval was required.

Bruce D. Greenshields

- Born in Winfield, Kansas; grew up in Blackwell, Oklahoma;
- Graduate of University of Oklahoma; earned master's degree in civil engineering at the University of Michigan;
- 1934 doctorate in civil engineering from the University of Michigan;
- Taught at different universities;
- Wrote numerous articles on traffic behavior and highway safety;
- Pioneer in the use of photography relating to traffic matters and in applying mathematics to traffic flow;
- Invented the Drivometer;
- 1956 joined the University of Michigan faculty and was acting director of the Transportation Institute there;
- 1966 retirement;
- Returned to Washington, D.C., and was a traffic consultant to various federal agencies; and
- Dr. Greenshields received the Matson Memorial Award in 1976.

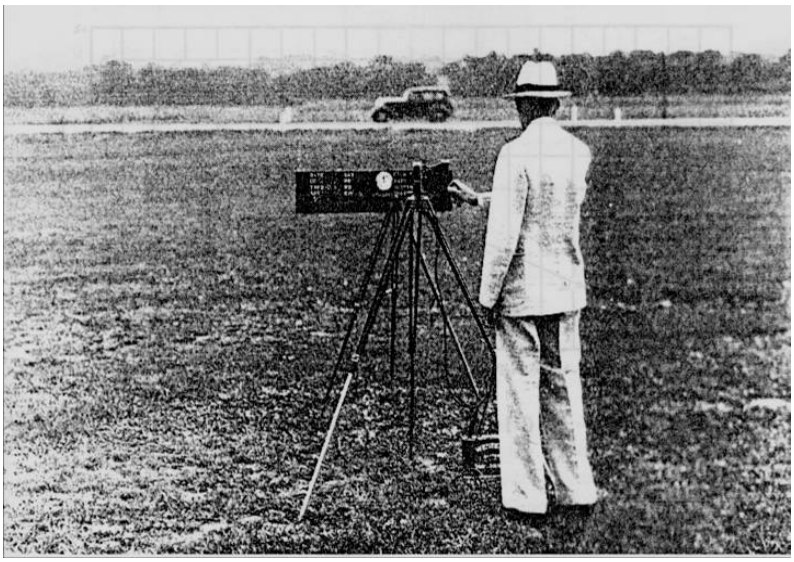


FIGURE 1 Greenshields measurement set up in the 1930s.

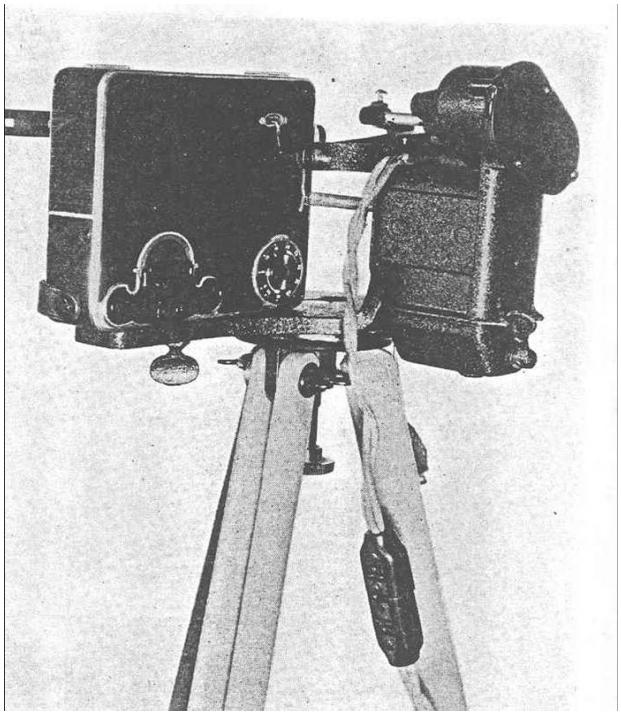


FIGURE 2 Camera with Motor Attachment used by Greenshields.

To avoid photographic blur due to motion, a moving car had to be at least 300 feet from the camera. In this case the length of road included in a picture was about 125 feet. The blur might have been lessened by using a faster shutter.

At the beginning of each film, and hourly during a run, there was included a photograph of a bulletin board giving the location date, hour, time interval, shutter opening and other pertinent information. The white cloth stretched along the opposite side of the road was used to keep the vehicles from fading into the dark background. Figure 3 shows three frames of pictures taken with the movie camera at this station.

The vertical lines are added to show how the pictures look when projected upon a screen with lines drawn upon it for scaling distance. The measured distance from the camera to the road together with the camera characteristics suffices to give the scale of dimensions which is more accurately determined if the camera is set at right angles to the road. As a check, however, a complete plan of the section of the roadway studied is recorded giving the distances from the camera and between objects in the pictures such as fence posts or poles. Where no identification exists a 100-foot tape is laid along the pavement and at every 10-foot interval a marker is held over the point and photographed. There is thus obtained a definite scale for the picture.

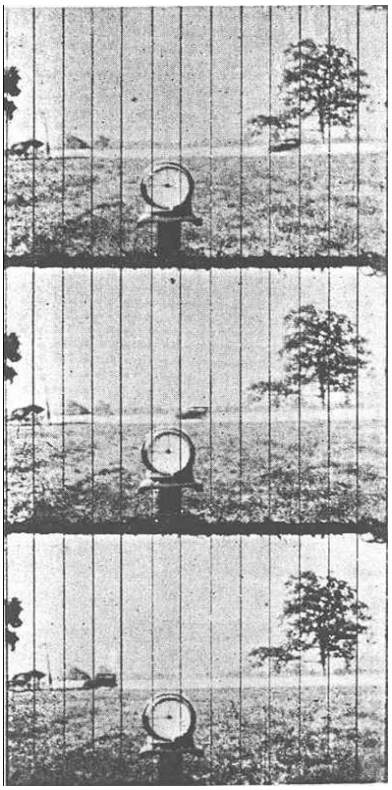


FIGURE 3 Three frames pictures taken with the movie camera.

THE LINEAR SPEED–DENSITY RELATION

Greenshields postulated a linear relationship between speed and traffic density, as shown in Figure 4. When using the relation

Flow = density * speed

the linear speed–density relation converts into a parabolic relation between speed and traffic flow (Figure 5). Increasingly even the term "flow" was not known 75 years ago and Greenshields called that term "density-vehicles per hour" or density of the second kind.

In this model some traffic flow characteristics are expressed well. It shows a maximal traffic flow with the related optimal traffic density. In the q-v-diagram exists two regimes, meaning it's possible to have two speeds at the same traffic flow. By this the traffic flow is classified in a stable and an unstable regime. Greenshields linear relation would be called an univariate model, because both regimes are calculated with the same formula.

Early studies at traffic capacity of motorways had two different approaches. On the one side speed–traffic density relations were analyzed. Here a constant (free) speed was implied

$q = v_f * k$

On the other side distance phenomenon at high traffic density were analyzed and as easiest approach a constant reaction time t_r was implied, which brings you to the gross headway

$l = l_0 + v * t_r \quad \text{with } k = 1/l$

$$q = -\frac{l_0}{t_r}(k - k_{max}),\ k_{max} = \frac{1}{l_0}$$

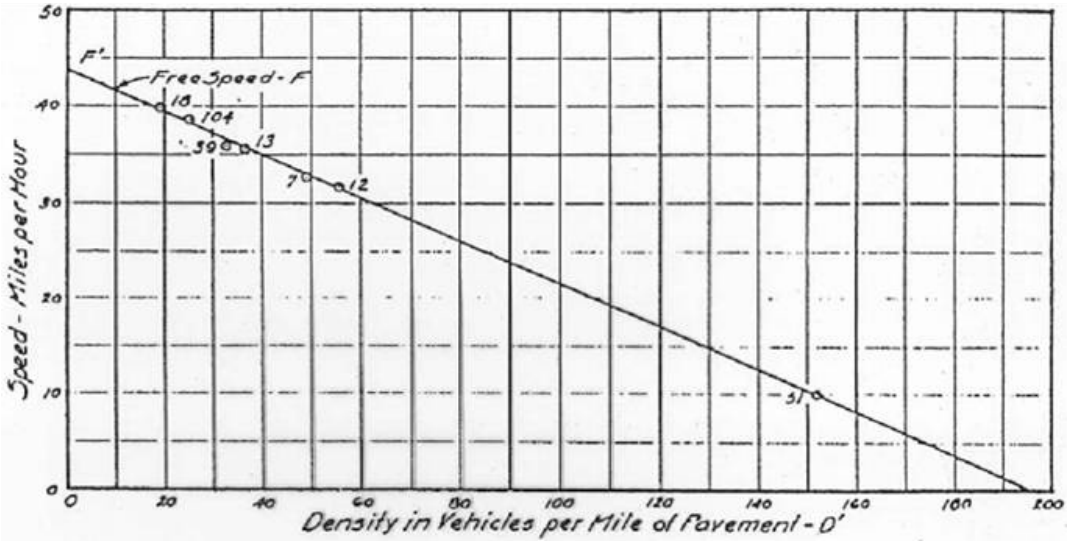


FIGURE 4 Speed–density relation V (Greenshields 1934).

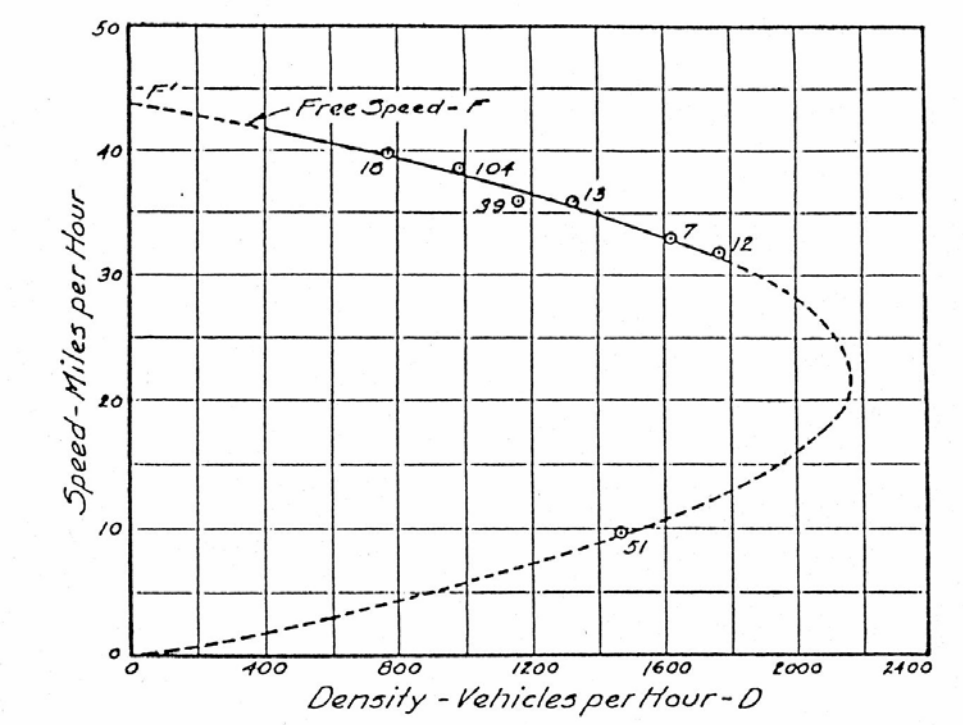


FIGURE 5 The first Fundamental Diagram as q-v diagram.

Also a linear relationship between q and k—but with a negative congestion broadening speed—l_0/t_r as proportionality constant. Summarizing both regimes creates a triangle function as traffic flow–traffic density relation. Lighthill and Whitham as well as Richards preached this triangle function as flow–density–curve and the use of the cinematic wave theory on road traffic as instrument to combine both fields and to explain the dispersion of shock waves as revertive going congestion front (LWR theory).

Also the q-k-relation established by Lighthill and Whitham has a parabolic curve progression and it's a one field model, too. The maximum stands for the expected road capacity of a motorway section. The insights of Greenshields inspired the development of two- and multiple-regime models in the aftermath.

TWO-VARIATE MODELS

Edie showed as one of the first that at the empiric q-k data often in the area of the maximum traffic flow and he suggested to describe the q-k-relation with discontinuous curvature. In 1961 he shows the first two-variate model approach for the Fundamental diagram. Here he discriminates the regime of the free traffic and the jammed traffic. His suggestions caused a series of analysis specially made by May, which aim was to specify strenghtly the characteristics and parameters of this two-field model. May and Keller developed a two-field traffic flow model, which based on the vehicle-sequence model of Gazis. In the process emerged that the traffic flow in the field of instable traffic is better shown by a hyperbolic function that by a parabolic one. Parallel to these developments Prigogine and others established a traffic flow

analogy for the kinetic gas theory. They showed a dependency of the velocity distribution of the traffic density and indirect of the overhauling probability. Model tests by means of measurement data showed that the curves q(k) only brings a realistic description at low traffic and burst off before the capacity limit was achieved. The model was also critically judged concerning the description of the effects of a speed impact. After that Prigogine and Herman tested an analogy of the traffic flow phases compared to the phase changes on the condition of aggregation of water (gas-fluid).

Two-field models were edited for the practical use of alternative route control with changing directing signs. Assuming that there are time and route sector homogeneous conditions the traffic situation could be modelized by density waves and shown for the optimization of controlling the necessary effects of traffic flow in one end function.

FOLLOW-UPS OF AERIAL PHOTOGRAMENTRY TECHNIQUES

Treiterer and Myers made first tests about hysteresis in traffic flow, where the connection of two variables depends on the previous history, if one variable grows or falls in relation to the other. Thereby a convoy of vehicles airborne observed in an interval of 4 minutes on a route of 5.3 km. Seventy vehicles were analyzed, which due to an upstream interference passed through a backwards running shock wave. The measurements (see Figure 6) show an asymmetry in driving behavior at delay and speed up as resulted in tests of Treiterer and Myers. After they run through the shock wave the vehicle convoy speeded up from 40 to 60 km/h and the traffic flow increased from 1,800 vehicles per hour to nearly 3,000 vehicles per hour. The traffic density didn't change significantly.

INFLUENCE OF GREENSHIELDS ON EUROPEAN AND JAPANESE RESEARCH WORK

Greenshields influenced traffic engineers and researchers around the world. Cremer calculated the influence of dynamic traffic flow effects on the result of the fundamental diagram on a basis of a macroscopic traffic model. Through analysis of traffic data on highways in Japan, Koshi postulates that the curve for the field of the free and dammed traffic is not constant concave in the q-k diagram, but rather similar to a mirrored lambda. A classification of vehicles for free traffic in two groups "only following vehicles" and "leading and following vehicles" followed.

Kühne and Kerner introduced a phase transition model about the traffic flow in the field of capacity and the appropriate parameter for identification of interferences in time. Banks studied the traffic flow on the Interstate 8 east of San Diego in view of a capacity drop. Using a combination of video recordings and traffic data of measurement loops he analyzed the traffic flow over 9 days. The measurement loops provided traffic volume and assignment data in 30-second intervals. With the video recording the congestion start could be specified chronologically accurate and periods of 12 minutes each prequeue as well as queue discharge could be chosen. Via these data the existence of a two-capacity phenomenon for this route

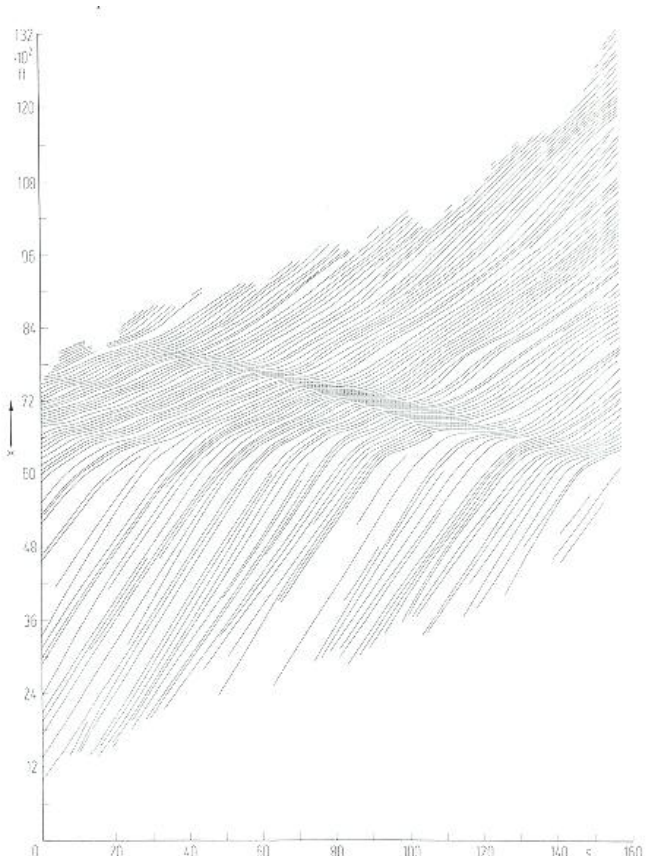


FIGURE 6 Description of the states of traffic.

section could be proved. Hall and Agyemang-Duah studied the traffic flow on the Queen Elisabeth Way west of Toronto over several days. The traffic data existed for 30-second intervals. The calculatedcapacity drop was 5% to 6%. Inquiries of Brilon and Ponzlet on German motorways showed data between 4% and 12%.

RESOURCES

Banks, J.H. Review of Empirical Research on Congested Freeway Flow. In *Transportation Research Record, Journal of the Transportation Research Board, No. 1802*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp, 225–232.

Beckmann, H., F. Jacobs, K.-H. Lenz, R. Wiedemann, and H. Zackor. Das Fundamentaldiagramm-eine Zusammenstellung bisheriger Erkenntnisse, Forschungsarbeiten aus dem Straßenwesen Heft 89, Forschungsgesellschaft für das Straßenwesen (FGSV) e.V., Köln 1973.

Daganzo, C. F. Cassidy, M. J.; Bestini, R. L. Possible Explanations of Phase Transitions in Highway Traffic. In *Transportation Research Part A*, Vol. 33, 1999, pp. 365–379.

Gazis, D. C., R. Herman, E. W. Montroll, and R. W. Rothery. Nonlinear Follow-the-Leader Models of Traffic Flow. In *Operations Research*, Vol. 9, 1961, pp. 545–560.

Greenshields, B. D., J. T. Thompson, H. C. Dickinson, and R. S. Swinton. The Photographic Method of Studying Traffic Behavior. *Highway Research Board Proceedings*, Vol. 13, 1933, pp. 382–399.

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

Hall, F. L., and Agyemang-Duah, K. Freeway Capacity Drop and the Definition of Capacity. In *Transportation Research Record 1320*, TRB, National Research Council, Washington, D.C., 1991, pp. 91–98.

Special Report 209: Highway Capacity Manual, 3rd ed. (1997 update). TRB, National Reserach Council, Washington, D.C., 1998.

Keller, H., and T. Sachse. Einfluss des Bezugsintervalls in Fundamentaldiagrammen auf die zutreffende Beschreibung der Leistungsfähigkeit von Straßenabschnitten, Forschung Straßenbau und Straßenverkehrstechnik, Heft 614, Bundesministerium für Verkehr, Bonn, 1992.

Koshi, M., M. Iwasaki, and I. Ohkura. Some Findings and an Overview on Vehicular Flow Characteristics. *Proc., 8th International Symposium on Transportation and Traffic Theory*, Toronto, 1983.

Kühne, R., and P. Michalopoulos. Continuum Flow Models in Traffic Flow Theory: A State-of-the-Art-Report. Transportation Research Board, 1992.

Litgthill, M. J., and J. B. Whitham. On Kinematic Waves, I: Flow Movement in Long Rivers, II: A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society,* Series A, Volume 229, London,1955, pp. 281–345.

May, A. D., and H. M. Keller. Evaluation of Single and Two Regime Traffic Flow Models. *Proc., 3rd International Symposium on Transportation and Traffic Theory*, Karlsruhe, 1968; in Kim, Y; Keller, H.: Zur Dynamik zwischen Verkehrszuständen im Fundamentaldiagramm, Straßenverkehrstechnik, Sept. 2001, S. 433–442.

Prigogine, I., R. Herman, and R. Anderson. On Individual and Collective Flow. Académie royale de Belgique, Bulletin de la Classe des Sciences, 5ème série, tome XLVIII, 1962, pp. 792–804.

Ponzlet, M. Auswirkungen von systematischen und umfeldbedingten Schwankungen des Geschwindigkeitsverhaltens und deren Beschreibung in Verkehrsflussmodellen, Heft 16 der Schriftenreihe des Lehrstuhls für Verkehrswesen der Ruhr-Universität Bochum, 1996.

Treiterer, J., and J. A. Myers. The Hysteresis Phenomena in Traffic Flow. *Proc., 6th International Symposium on Transportation and Traffic Flow Theory*, 1974, pp. 13–38.

Wardrop, J. G., and G. Charlesworth. A method of estimating speed and flow traffic from a moving vehicle. *Proceedings of the Institution of Civil Engineers, Part II*, February 1954.

Zackor, H., R. Kühne, and W. Balz. Untersuchungen des Verkehrsablaufs im Bereich der Leistungsfähigkeit und bei instabilem Fluß, Forschung Straßenbau und Straßenverkehrstechnik, Heft 524, Bundesministerium für Verkehr, Bonn, 1988.

Greenshields' Legacy
Urban Streets

NATHAN H. GARTNER
University of Massachusetts, Lowell

While Greenshields is well known for his development of The Fundamental Diagram in Traffic Flow Theory, it is less known that he also made fundamental contributions to the study of traffic behavior at signalized intersections. This was done while he was associated with the famed Yale Bureau of Highway Traffic at Yale University in New Haven, Connecticut in the mid-1940s and culminated in a seminal report entitled Traffic Performance at Urban Street Intersections. Greenshields developed instrumentation for the study of traffic behavior at intersections. He then used this instrumentation to study the time–space relationships of vehicle movements at intersections. "It is the total time required to pass a given number of vehicles through the intersection that is of primary interest to the traffic engineer. In analytical categories, this time depends on the integration of individual patterns of reaction time, acceleration, speed and spacing." He went on to determine the famed "Greenshields Numbers" which measure the time required by successive vehicles to cross a signalized intersection when starting from a standing queue. This led to the determination of signal-controlled approach capacity. Greenshields also studied behavior patterns at unsignalized intersections: "In the absence of a signal to control his crossing, the driver selects the gaps in opposing traffic that he considers large enough to be safe. The rapidity with which the driver makes up his mind and the rate at which he accelerates once he has decided are factors to be considered." Whence, the well-known method of gap acceptance. Greenshields was also a pioneer in the application of probability theory to traffic problems. Among the typical problems he analyzed are: traffic delay caused by drawbridge, accident exposure due to obscured vision, size of temporary storage space at parking lot unnecessary stops at stop sign, and optimum signal timing (*1*). Another concept that he proposed was "the quality of flow index" *Q*. This index may be defined as a number expressing the desirable ratio of the flow factors of time, change of speed, change of direction and distance. Thus *Q* is a rate per unit distance. The quality of traffic flow may be expressed by the equation

$$Q = T \times S \times D/L$$

where

***Q* = quality,**
***T* = time,**
***S* = change of speed,**
***D* = change of direction, and**
***L* = distance.**

The smaller the *Q*, the better the travel.

REFERENCE

1. Greenshields, B. D., D. Schapiro, and E. L. Ericksen. Traffic Performance at Urban Street Intersections, *Technical Report No. 1*, Yale Bureau of Highway Traffic, Yale University, New Haven, Conn., 1947.

Treiterer's Legacy
Traffic Flow Measurements and Characteristics

RAHIM F. BENEKOHAL
University of Illinois Urbana–Champaign

The highlights of contributions of the late Professor Joseph Treiterer to traffic flow theory, characteristics, data collection methods, platoon dispersion, and automated vehicle control are presented in this paper. Treiterer was pioneer researcher and about 40 years ago tested the application of infrared radar for longitudinal control of vehicles on highways. His collection method for the trajectory data that captured the disturbance in traffic flow is still unmatched, after 35 years. Treiterer studied the traffic dynamics using aerial photogrammetry techniques and developed methods for data reduction and analysis of aerial photos. He continuously recorded the traffic movement on I-71 by following platoons of vehicles from a helicopter with an aerial camera. His data collections method was unique and enabled him to follow the propagation and dissipation of traffic disturbances through platoons of vehicles. He was able to capture the hysteresis phenomenon in traffic flow. Treiterer showed that there are loops in q-k curves (A loop and B loop) and showed that a rapid "jump" in flow rate (or speed) occurs when state of traffic was about to change from congested state to uncongested state. Then, traffic studies on I-70 were carried out by aerial surveys to obtain further data on the hysteresis phenomenon. He found that the shape of the q-k curves and thus the hysteresis phenomenon was not the same for single disturbance compared to multiple stop-and-go conditions. Treiterer also found that even the marginally safe traffic flow criterion was not satisfied for about 11% of the total observation time. Treiterer studied the application of infrared technology for longitudinal control of vehicles. He found that speed difference between the lead and trailing vehicles and the acceleration pattern of the lead vehicle could be the most useful information to improve traffic safety and increase traffic capacity. Treiterer studied the effects of signal spacing on platoon dispersion and characteristics as they passed through nine signalized intersection on two urban arterials. This comprehensive data included platoon size, lane distribution, lane change maneuvers, and traffic breakdowns on 28 platoons. He developed models to predict queue lengths and vehicle delays as a function of signal offset and spacing conditions. Treiterer found that platoon size effected platoon behavior. Platoon velocity seemed to decrease as platoon size increased. He also found that initial acceleration characteristics of the smaller-sized platoons (four to six vehicles) were higher that of the larger-sized platoons (10 to 13 vehicles). Treiterer also developed a unique technique for collecting simultaneous traffic data on all approaches of an intersection using time-lapse photography.

INVESTIGATION OF TRAFFIC DYNAMICS BY AERIAL PHOTOGRAMMETRY TECHNIQUES

Overview

Treiterer said, “Continued improvement of the design and operation of the highway system depends on the understanding of the dynamics of traffic using the system. This understanding can only be gained once the critical parameters of traffic flow have been identified and measured.” He used aerial photogrammetry technique to collect data on a traffic stream continuously, study its behavior and propagation of disturbances. A study of traffic dynamics began in 1966 and finished in 1975. It was sponsored by Ohio Department of Transportation in cooperation with the FHWA. The work was conducted by the research staff of the Transportation Research Center of the Ohio State University under direction of Dr. Joseph Treiterer. A main goal of research was to study traffic dynamics using aerial photography with emphasis on research in the theory of traffic flow, and utilizing research results to improve traffic flow and increase safety.

There are four interim reports and a final report for the study. Interim Report 1 and Interim Report 2 investigated the photogrammetric techniques to be used for automatic data reduction from aerial photographs. Computer programs to generate vehicle trajectories were developed as discussed in Interim Report 2. Interim Report 3 documented the results of two specific studies of traffic dynamics: the use of traffic energy as a parameter of traffic flow and characteristics of the traffic stream in the vicinity of traffic bottlenecks. Interim Report 4 presented the results of multilinear speed–flow models and energy models to describe traffic flow.

HIGHLIGHTS OF INTERIM REPORT 1

Interim Report 1 was published in August 1967. Aerial photographs were taken from a Bell Helicopter flying over Interstate 71. The northbound of I-71 changed from three lanes per direction to two lanes near N. Broadway Ave. The reduction led to disturbances in traffic flow and generated kinematic disturbances and shock waves that propagated in a southerly direction. Two cameras to take the aerial photos were evaluated: the first camera was the Maurer P-2 70-mm reconnaissance camera, which could cause unpredictable distortions due to any film warping since it was not equipped with a vacuum back. The second camera was the KA-62A aerial camera that used 5-inch wide film roll with 250-foot film capacity. The latter one was selected because it provided better resolution and higher quality photographs.

A total of six flights and five films were taken from I-71 and over 3,000 frames were obtained by August 1967. Initially, a control vehicle with two high-intensity beam lights on top of vehicle was used so the helicopter pilot can use it as a guide to follow the platoon and communicate with the pilot about the traffic disturbance generation; however, in later flights the control vehicle was not used. After a lot of effort, the data reduction process was automated and vehicle trajectories were obtained. To plot the trajectories, accumulative distance traveled by each vehicle over time was used. The report’s Figure 2.4 shows trajectories of 48 vehicles on the median lane of I-71 taken by the Maurer P-2 70-mm camera.

Treiterer looked at the spaces between vehicles in car-following situations to determine if drivers kept an absolute or marginal safety distance. The absolute safety is when the following driver keeps a distance from the leading vehicle that is at least equal to the stopping sight distance. The marginal safety is when the following driver leaves spacing equal to the distance traveled during his/her reaction time. The above mentioned trajectory data was used and it was found that assuming 0.7 s reaction time, about 90 % of the drivers maintained the marginal safety distance in car following.

HIGHLIGHTS OF INTERIM REPORT 2

Treiterer's Interim Report 2 was published in June 1969 and it highlights are as follows. Treiterer indicated instability in traffic flow can occur for apparent reason. On one of the morning peak period aerial surveys of southbound of I-71 he observed that a kinematic wave was generated on the median lane and traveled at 19 ft/s against traffic. The disturbance lasted for 110 seconds and forced nearly 40 vehicles to stop for a period of time ranged from a few seconds to 12 seconds. At the same time no such disturbance was observed on the adjacent outside lane, as shown in the report's Figures 3.2 and 3.3.

He said that volume–density relationship deduced from the aerial data "indicates a certain amount of support for the Lighthill-Whitham Theory describing the variation of the flow-concentration curve in a bottleneck."

Chandler, Herman, Montroll and Newell, as well as Kometani and Sasaki had proposed that the variance in the reaction of drivers to changes in behavior of the leading vehicle cause instability. Treiterer investigated the stability in car following using the trajectories for 24 vehicles on I-71. A disturbance caused drivers to slow down to less than 3.2 mph and then accelerate to about 30 mph. The stability analysis indicated that the sensitivity parameter λ, in the linear relative velocity follower model (GM 1 Model) alone is not a good criterion for determining the stability of traffic flow. He indicated that $2\lambda\tau$ value appears to provide a better measure of stability than λ alone. The report's Figure 2.9 shows trajectories for vehicles entering, within and leaving disturbance as well as for uniform flow (recovered). Plot of $2\lambda\tau$ values are shown in the report's Figure 2.11 and indicate that for the entering flow there is an increase in the frequency of values between 1 and 3, indicating instability in flow. The plot for within-disturbance shows a high concentration between 0 and 1, indicating stable conditions. The stability within the disturbance is expected since most of the vehicles had stopped. For the vehicles leaving the disturbance most of the values are greater than one, indicating instability. For the uniform flow after recovering from the disturbance, most of the values are one or less. He recommended that this was limited data and additional data need to be collected and analyzed.

He preferred lognormal model for headway distribution and stated that "it would seem that the use of either the composite exponential or Pearson Type III for predicting headways would be of no value when compared with the reliability and accuracy of the lognormal."

He examined the speed–density plot and concluded that "it is shown that the single continuous curvilinear function does not adequately describe the speed–density behavior of the traffic." He proposed a multilinear speed–density model by dividing it into five regions [with threshold <35, 70, 105, 140 and 210 (jam density)] (see the report's Figure 5.20). The multilinear

model provided good fit for speed–density, volume–density, and energy-–density relationships, and he suggested further investigation of this idea.

He identified two groups of vehicles: “before-within” and “within-after” the disturbance. Then for each group, speed–density and volume–density relationships were plotted and traffic density breakpoints were identified. In addition, plot of the standard deviations of the vehicle speeds versus density were obtained and it bore amazing similarity with speed–density plot. The standard deviation could indicate the interaction of a group of vehicles and thus a particular operating condition. For the free-flow region, there was no data available. At stable flow region it seems that density does not have much influence on speed variation and the standard deviation showed two dissimilar patterns. In the forced flow region, the σ begins to decrease rapidly for both groups of vehicles, indicating that bunching of vehicles restrict speed variation. In the forced flow region, σ attains a local minimum. It is because all vehicles are moving slowly and at near uniform speed. Traffic is unstable in this region and further increase in density creates more disturbance in traffic and higher variation in σ. In the disturbed flow region nearly half of the vehicles in platoon were forced to stop. As traffic reaches jam density, σ decrease to near zero since speed are uniform and close to zero (see the report’s Figure 5.16 and 5.17).

To study hysteresis phenomenon, two platoons (see the report’s Figure 5.22 and Figure 5.23) were followed for approximately two minutes, and average platoon speed and densities were calculated at 1-second intervals. The kinetic energy of each platoon was calculated and plotted versus the corresponding traffic density (see the report’s Figures 5.24 and 5.25). It was noticed that the traffic stream did not recover at the same rate with density as it had entered jam conditions, until a certain level of density is reached. That retardation effect was termed traffic hysteresis.

For each platoon σ was computed and plotted against density (see the report’s Figures 5.26 and 27) and these plots show an interesting relationship. Comparing energy–density plots with σ–density plots shows that the point of recovery coincides with the proposed breakpoint between stable and forced flow. Low σ values were generally associated with high kinetic energy values, thus, σ may be a good indicator of internal energy. Treiterer concluded that the energy-change parameter, though he did not quantify its parameters, holds much promise as a valid measure of the quality of freeway flow, and he suggested continued empirical investigations of it.

HIGHLIGHTS OF INTERIM REPORT 3

Interim Report 3 was published in June 1970 and its highlights are listed below. Data reduction system consisted of three basic components (the report’s Figure 2.1):

- Mann Type 829D Comparator,
- Mann Type 1945 Data Logger, and
- IBM 026 Printing Card Punch.

The Mann Type 829D Comparator was a high-precision device to measure distances on photographic films or plates to an accuracy of 0.001 mm. The x and y coordinate of the vehicles on the photographs were measured by moving plates of film. The Mann Type 1945 data logger was utilized to provide digital display and tabulation of coordinate values and identifying data.

Then, a hand-activated switch transferred the data from the logger to IBM 026 Printing Card Punch output device. With this system an aerial survey film containing 300 usable frames could be completed in 150 hours. This was a significant achievement in data reduction at that time.

The data was extracted from aerial traffic survey that was carried out on July 25, 1967 on southbound I-71 during morning peak hour (around 7:45 a.m.). The photos were taken from a helicopter at time intervals of once second. Data for eight platoons were processed and analyzed (see the report's Figure 3.1-3.4). Platoon 122 contained a group of vehicles that proceeded from low-density, high-speed conditions through a disturbance into rather stable conditions. Platoon 123 was unique since no vehicles entered or left the platoon. Platoon 126 had a large range of density in a rather short period of time. Platoon 127 showed the maximum observed density. Platoons 141 to 144 were selected because they contained mutually exclusive groups of vehicles. The report contains the analyses for Platoon 142 since it represented the entire study.

The best fit to the speed-density plot of "before-within" disturbance traffic was three linear regime model (one for each low, mid, and high density) for all eight platoons. For within-after condition, two-regime model was used since none of the platoons returned to low-density region. The breakpoints at 55 vpm and 95 vpm provided the optimum overall situation.

When a platoon was released from the maximum density conditions, the resulting values of speed, volume and kinetic energy ($\alpha * k * u^2$ where α is a dimensionless constant) were retarded compared to the "before" conditions at respective densities. The existence of the hysteresis phenomenon did not depend on the maximum density reached. The kinetic energy of a traffic stream was defined as $\alpha. * k. *u^2$ where α is a dimensionless constant, k is the density and u is the speed. Drew (1968) introduced the concept of energy into traffic flow analysis by considering the traffic stream as analogous to the flow of a compressible fluid in a constant-area duct. Drew proposed that the total energy of a traffic stream consist of kinetic energy and the internal energy. The internal energy was thought to be related to the interactions among vehicles and Drew suggested "acceleration noise" as a measure of internal energy. Acceleration noise is the standard deviation of acceleration distribution of one vehicle along a road. However, Treiterer using the aerial data proved that this term cannot represent the internal energy and proposed using coefficient of variation of speed as indicator of internal energy.

HIGHLIGHTS OF INTERIM REPORT 4

Interim Report 4 was published in Oct 1972. And it highlights are as follows. Two platoons (Platoon A and B) that went through a kinematic disturbance were selected from the trajectory data collected on I-71. This report investigated how multilinear model fit into the data and how different energy density models could explain the internal energy of traffic. Data for Platoons A and B confirmed that the hysteresis phenomenon happened. Then, they looked at the relationships for the queue-forming part (going in to congestion) of the two platoons. Similar to the study by Drake, Schofer, and May (1967), this study examined how different speed–density models fits the queue forming part of the data. The major difference between the two studies is that Drake's data came from measurements at point, but Treiterer's study was based on data for platoon that traveled a section of the road. Treiterer divided the speed–density curve into six regions and proposed a linear model for each region. The platoon data covered only four of the six regions. They concluded that it is possible to fit such linear models. Then they compared their six regions to the six levels of service (LOS) given in the 1965 Highway Capacity Manual. It

should be noted that Treiterer established the levels of service based on density whereas the LOS in the 1965 Manual were established based on speed. Treiterer proposed using density to determine quality of service and suggested the density threshold for those levels. For each platoon going into queue, they looked at speed, absolute safe spacing, average headway and standard deviation of speeds average spacing, headway, and speed dispersion for each region (see the report's Figures 1-28 and 1-29). There were no data points in the first two regions of the initially proposed six regions and data points in the fifth and sixth regions were rather scarce. So they verified their model for only four regions. The report also examined the idea of platoon speed change and platoon length change as indications of external and internal energy of traffic stream.

A new concept of level of service based on density as the fundamental parameter was proposed. Based on the results of the multilinear speed–density model, six subregions were classified according to the operating characteristics as follows:

- Free-flow region: 0 to10 vpm;
- Semi-free flow region: 10 to 30 vpm;
- Capacity flow region: 30 to 60 vpm;
- Restricted flow region: 60 to 105 vpm;
- Disturbed flow region: 105 to 145 vpm; and
- Forced flow region: 145 vpm and more.

The last three regions were combined and were labeled as Level D. Thus, the following levels were proposed:

- Level A: 0-10 vpm;
- Level B: 10-30 vpm;
- Level C: 30-60 vpm; and
- Level D: 60 vpm and more.

Based on the findings of the study, they implemented a ramp-metering scheme for three ramps on southbound of I-71. After the implementation, traffic volumes on I-71 increased by 6% without a noticeable decrease in overall travel speeds. This resulted in an 8,000 vehicle per hour increase in the throughput (q.u) of the metered section. Treiterer recommended similar analysis of traffic flow, focusing on after-disturbance behavior in hysteresis phenomenon.

HIGHLIGHTS OF FINAL REPORT

The data was collected on Interstate 71 from the downtown area of Columbus, Ohio, north to Route 161 in Worthington, Ohio, and on Interstate 70 from downtown to James Road. A KA 62A aerial reconnaissance camera and a 3-inch lens with an aperture of range f/4.5 to f/11 was used to provide a coverage of about 6,360 ft by 4,500 ft at an altitude of about 30,000 ft. The photographs were taken from a Bell 4752 helicopter at an altitude of 3,000 feet. Photographs were taken at nominal scale of 1:12,000 to provide an accuracy of measurement better than the required ± 1 foot. Photographs were taken as at intervals of 1 second. In order to construct the

vehicle trajectories from the aerial photographs, Treiterer utilized a new data reduction method and obtained the following information:

- Vehicle number,
- Photo coordinate,
- Ground coordinate,
- Accumulative distance,
- Spacing,
- Headway, and
- Velocity.

He used this information to plot trajectories of vehicles. The report's Figures 3.5 and 3.6 (of Final Report) are good examples.

He suggested dividing traffic flow into five operational regions based on density: free flow, stable flow, forced flow, disturbed flow, and jammed flow (see Figure 3.21, Final Report). The density thresholds for them are 35, 70 105, 140, and 210 (jam density).

He investigated different traffic patterns between queue forming and queue releasing conditions. The two platoons observed are shown in the report's Figure 3.23 and 3.24. Plot of the kinetic energy of the platoon versus density are shown in the report's Figure 3.25 and 3.26. He observed that traffic stream did not recover at the same rate as it entered the jam conditions. He called this retardation the hysteresis phenomenon of traffic flow. He also computed the standard deviation of the speed of vehicles in platoon (sigma) and plotted it against density (see the report's Figures 3.27 and 3.28). The energy–density and sigma–density relationships showed that "when the queue releases at a point where the hysteresis effect recovers, the platoon remains at that density and no further decrease in density occurs. Furthermore, the point of recovery coincides with the point proposed as a breakpoint, in the previous section, between stable and forced flow." They also noted that low sigma values were generally associated with the higher kinetic energy values, thus indicating that sigma can be representative of internal energy.

Treiterer used the aerial photogrammetric data to solve the problem of recurrent congestion on I-71 north of Columbus, Ohio (from North Broadway to Hudson Street). After determining the cause of the congestion, he proposed a ramp-metering technique for Broadway-to-Hudson Street section to alleviate congestion during the morning peak hour.

A constant-rate ramp metering was implemented on I-71, and data showed a 6% increase in traffic volume (equivalent to an increase of 8,000 vehicle miles per hour) without a noticeable decrease on speed on Interstate 71. He suggested a more sophisticated ramp-metering technique that used freeway traffic condition to determine the rate of ramp metering.

DATA FROM I-70

Treiterer indicated that density is a better measure of traffic flow conditions on a roadway. He investigated the feasibility of real-time measurement of traffic density using inductive loop detectors. He used a 158-foot-long multiple Figure 8 loop configuration with 18-feet section lengths in the field (see the report's Figure 5.13). The 158-foot-long loop detector was installed on the rightmost lane of I-70 where it splits from the northbound of I-71. The multiple Figure 8

loop detector performed very well in the field and provided speed and density data that can be used for real-time traffic control. However, the field tests produced some significant differences from the lab tests conducted at Ohio State University. The results from field test showed that the loop detector output varied in shape and magnitude for different types of vehicles particularly at low density condition, but at high density traffic it was not sufficient to evaluate the influence of traffic composition on density measurements. Another difference between the lab and field results was that the detector, consisting of six fields, produced only two distinctive peaks, enough for speed measurement, but more uniform and more pronounced peaks are desirable. Further research was recommended to understand the causes for the difference in field and lab results and to improve the system. Study of hysteresis phenomenon traffic studies on I-70 were carried out by aerial surveys to: a) obtain traffic density distributions for inbound and outbound traffic on I-70 during peak traffic hours; b) evaluate traffic conditions (travel times) for peak-hour traffic, c) collect more data on the hysteresis phenomenon of traffic flow that had been observed on I-71.

Data showed that traffic density during the evening peak hour was more uniformly distributed, and its peaks were less than the peak densities in the morning peak hour. Speed–distance profiles for peak traffic hours were obtained using moving vehicle method. The study section on I-70 was divided into 16 subsections and the most critical locations were determined. It was found that the morning peak traffic (inbound) posed more serious problems than the evening peak traffic that was outbound (westbound).

Conditions for aerial traffic survey on I-70 were different than those on I-71 because I-70 is close to Columbus Airport and that restricted the flying altitude to 2,000 ft. That in turn reduced the coverage to about 3,000 ft. I-70 had three lanes per direction while I-71 had two lanes per direction at the time of data collection. Trajectory data for two adjacent lanes on I-70 between Nelson Rd. and 18th S.t are shown in the report's Figure 7.27 and 28. Traffic flow was disturbed more on the middle lane than the shoulder lane, so the middle lane data was used to study hysteresis phenomenon. Traffic was operating in stop-and-go condition because of multiple disturbances in that section. Due to the tailwind, the helicopter could not follow a platoon of cars though the disturbances. Thus, the data collected on I-70 was not as extensive as the I-71 and it only represented small platoons of about seven to 11 vehicles over a limited distance.

Similar to I-71, the disturbance forced the vehicles to stop for 14 s on I-70. However, for the I-70 data presented only the A hysteresis loop, and not the Ba loop. Maximum density on I-70 was higher than on I-71 and it reached as high as 280 vph. The most important difference between I-70 and I-71 data is that the A loop is reversed. Traffic entered the disturbance at much lower speed than in the speed in recovery phase. The deterioration occurred at density of 125, but recovery happened at 80 vpm. The data seems to be a typical case of recycling around A-loop which resulted in stop-and-go operation. This was supported by the speed data for A-loop which was from zero to 25 mph. Data for other platoon supported this reversed A-loop observation as shown in the report's Figures 7.31 thru 7.35. One platoon in I-70 showed a gradual decrease in speed over a density range of 70 to 200 vpm (see the report's Figure 3.38).

I-71 data showed the hysteresis phenomenon for a cycle of a single disturbance, but I-70 data if a multiple disturbance that resulted in stop-and-go conditions. The drop in speed for multiple disturbances occurs at 120 to 150 vpm which is much higher than the density for single disturbance (around 80 vpm). Higher peak densities were observed for the stop-and-go operating

conditions. In stop-and-go condition there is no B loop. More data is required to have a better understanding of hysteresis phenomenon.

APPLICATION OF INFRARED TECHNOLOGY TO CONTROL VEHICLE MOVEMENT IN A PLATOON OF CARS

Treiterer published this report after 1966 (no exact date on the report). He was concerned with rear-end collisions on highways. He examined the spacing between vehicles using the aerial photos from I-71 and found that about 11% of the observation time the spacing was below the marginal safe distance (assuming a reaction time of 0.7 seconds for all drivers), indicating a safety concern for vehicles in platoon on freeways. Marginal safe distance was defined as the distance traveled during reaction time of a driver. He looked at the promising technologies to help with safety and capacity, and selected infrared as the most economical and promising technology. He determined that speed difference between the lead and following vehicles and the acceleration pattern of the lead vehicle should be used in longitudinal control of vehicles with infrared technology.

He developed two infrared systems (a self-contained one and a source-sensor system). He used the source-sensor one over 1,000 miles on freeway. The source-sensor system had two main components. The first component was the source which was placed at the rear of the leading vehicle. The source was an infrared beam with a pulse frequency that depended on speed of vehicle. The second component was the sensor unit that was mounted under the front bumper of the following vehicle. It detected pulses emitted by the source on lead vehicle and converted it speed. The speed was compared to speed of the following car and the speed difference between the lead vehicle and the following vehicle was displayed on a meter in the following vehicle.

The self-contained sensing system did not require any active equipment in the leading vehicle. An infrared source-sensor unit was mounted on the front bumper of the vehicle which emitted a frequency modulated infrared signal in the direction of travel. The emitted signal was reflected back to the source by the lead vehicle that had a special rear license plate with corner reflectors. Since license plates were changed every year in most of the states, the introduction of corner reflectors on license plates would not have imposed an unreasonable demand on public.

Treiterer suggested a step by step introduction of this system to public. In step one it would be a driver aid to improve efficiency and safety. Then it would provide semiautomatic control for car following situation in high density traffic. And finally it would provide fully automatic control in steering and spacing of vehicles on freeways. He suggested that the capacity would exceed 3,000 vph per lane by using this technology.

EFFECT OF SIGNAL SPACING ON PLATOON DISPERSION

This study determined the characteristics of platoon of vehicles as they traveled through a series of signalized intersections and used them in developing a simulation model. Aerial photographs of 28 platoons on two one-way parallel arterial (Summit and 4th Streets in Columbus, Ohio) were taken in April, 1968. Summit St. was a southbound and North Fourth Street was a northbound arterial each with nine signalized intersections. The trajectories of the vehicles of the 28 platoons were constructed using approximately 28,000 time-space positions. Treiterer found

that in addition to signal spacing and signal offset, the platoon size had appreciable effects on platoon behavior. He also found that platoon velocity tended to decrease with increasing platoon size. The initial acceleration characteristics of the smaller-sized platoons (four to six vehicles) were generally higher than those of the larger-sized platoons (10 to 13 vehicles). He concluded that the patterns of mean velocity or mean spacing, the coefficient of variation of velocity and traffic density could best characterize platoon movement along an urban arterial.

RESOURCES

Chandler, R. E., R. Herman, and E. W. Montroll. Traffic Dynamics: Studies in Car Following. In *Ops. Res.,* Vol. 6, 1958, pp. 165–184.

Chandler, R. E., R. Herman, and E. Montroll. Car Following Theory of Steady-state Traffic Flow. In *Ops. Res*., Vol 7, 1959, pp. 499–595.

Drake, J., J. Schafer, and A. May. A Statistical Analysis of Speed–Density Hypotheses. In *Vehicular Traffic Science*, *Proceedings of the Third International Symposium on The Theory of Traffic Flow* (L. Edie, R. Herman, and R. Rothery, eds.), Elsevier, New York, 1967.

Drew, D. R. Traffic Flow Theory and Control. McGraw-Hill, 1968.

Lighthill, M.J, and Whitham, G.B. On Kinematic Waves, II; A Theory of Traffic Flow on Long Crowded Roads. In *Proc., Royal Society of London*, Series A 229, 1955, pp. 317–345.

Gazis, D. C., Herman, R., and Rothery, R.W. Nonlinear Follow-the-leader Models of Traffic Flow. In *Ops. Res*., Vol. 9, 1961, pp. 545–567.

Highway Research Board. *Special Report 87: Highway Capacity Manual*. HRB, National Research Council, Washington, 1965.

Kometani, E., and T. Sasaki. Dynamic Behavior of Traffic with a Nonlinear Spacing. In *Theory of Traffic Flow* (R. Herman, ed.), Elsevier, Amsterdam, 1961, pp. 105–119.

Newell, G. F. Nonlinear Effects in the Dynamics of Car-Following. In *Ops. Res*., Vol. 9, 1961, pp. 209–229.

Treiterer, J. Application of Infra-Red Technology to Control Vehicle Movement in a Platoon of Cars. Civil Engineering Dept., Ohio State University, Columbus, 1966.

Treiterer, J., and H. Russell. A Study of Longitudinal Control Systems for Platoon Movement. Ohio State University Engineering Experiment Station, Project HR-63-3-3, 1966.

Treiterer, J., and J. I. Taylor. Traffic Flow Studies by Photogrammetric Techniques. In *Transportation Research Record 142*, HRB, National Research Council, Washington, D.C., 1966, pp. 1–12.

Treiterer, J. and research staff. Investigation of Traffic Dynamics by Arial Photogrammetry Techniques, Interim Report 3, EEB 278-1, Experiment Station, Ohio State University, Columbus, August 1967.

Treiterer, J. and research staff. Investigation of traffic Dynamics by Arial Photogrammetry Techniques, Interim Report 3, EEB 278-2, Experiment Station, Ohio State University, Columbus, June 1969.

Treiterer, J. and research staff. Investigation of traffic Dynamics by Arial Photogrammetry Techniques, Interim Report 3, EEB 278-3, Experiment Station, Ohio State University, Columbus, June 1970.

Treiterer, J. and research staff. Investigation of traffic Dynamics by Arial Photogrammetry Techniques, Interim Report 4, EEB 278-4, Experiment Station, Ohio State University, Columbus, October 1972.

Treiterer, J. Investigation of Traffic Dynamics by Arial Photogrammetry Techniques. Ohio-DOT-09-75, Final report. Experiment Station, Ohio State University, Columbus, February 1975.

Treiterer, J., Z. Nemeth, and R. Vecellio. Effect of Signal Spacing on Platoon Dispersion, Final Report. EES-311, Experiment Station, Ohio State University, Columbus, July 1973.

Modern Approaches to Basic Traffic Modeling
Three-Phase Traffic Theory

BORIS S. KERNER
Daimler AG

In this paper, a brief review of some of the hypotheses of three-phase traffic theory introduced by the author in 1996–2004 and their applications for traffic modeling are presented. The basis of this paper is the presentation made on "Greenshields 75 Symposium," organized by TRB's Traffic Flow Theory and Characteristics Committee (AHB45) July 8–10, 2008, in Woods Hole, Massachusetts.

The understanding of empirical traffic congestion occurring on unsignalized multilane highways and freeways is a key for effective dynamic traffic management, control, organization, and other applications of transportation engineering. However, the traffic flow theories and models (see references *1–6*) that dominate up to now in transportation research journals, scientific conferences, and teaching programs of most universities cannot explain either traffic breakdown or most features of the resulting congested patterns. These theories are also the basis of most dynamic traffic assignment models and freeway traffic control methods, which therefore are not consistent with features of real traffic.

For this reason, the author introduced in 1996–2002 an alternative traffic flow theory called three-phase traffic theory, which can predict and explain the empirical spatiotemporal features of traffic breakdown and the resulting traffic congestion. There are three phases in this theory:

1. Free flow (F),
2. Synchronized flow (S), and
3. Wide-moving jam (J).

The synchronized flow and wide-moving jam traffic phases are associated with congested traffic.

In this article, a brief review of hypotheses of three-phase traffic theory and their applications for traffic modeling is presented. A much more detailed account of three-phase traffic theory, traffic flow control methods based on this theory, and the criticism of earlier traffic flow theories, models, and traffic control methods can be found in the books *The Physics of Traffic* (*7*) and *Introduction to Modern Traffic Flow Theory and Control: The Long Road to Three-Phase Traffic Theory* (*8*) and reviews by Kerner (*9–11*).

HISTORY AND MOTIVATION

The history of three-phase traffic theory began in 1992 when the author joined the former Daimler-Benz Company (now the Daimler Company) in Stuttgart, Germany. At that time, the author finished previous investigations of complex spatiotemporal self-organization effects

associated with phase transitions and pattern formation observed in physical, chemical, and biological systems (*12*). This experience with complex systems of diverse nature helped very much in finding and understanding mathematical nonlinear solutions of earlier traffic flow models and their sense (*13*):

- A traffic flow instability of the General Motors model class associated with a driver reaction time, which should explain traffic breakdown in vehicular traffic (*1–6*), leads to the emergence of a wide-moving jam in free flow ($F \rightarrow J$ transition) (Figure 1).
- The Lighthill-Whitham-Richards (LWR) model class (*1–6*) cannot explain empirical features of traffic breakdown as well as wide-moving jam emergence found in real measured traffic data.

The authors found (*13*) that the wide-moving jam exhibits some characteristic parameters (e.g., the flow rate q_{out} in the jam outflow, if free flow is formed in this outflow and the mean velocity of the downstream front of the jam, v_g), which are the same for different wide-moving jams (Figure 2); however, this is only true under the same traffic parameters (e.g., weather, the percentage of long vehicles, etc).

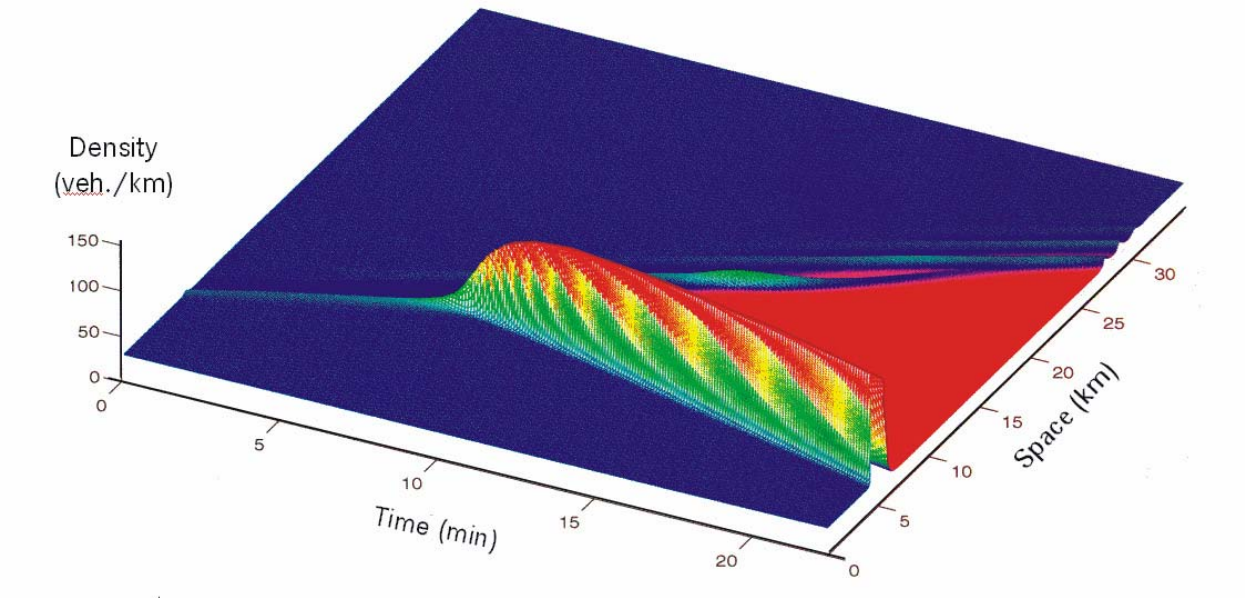


FIGURE 1 Wide-moving jam emergence in free flow in the GM model class (*13*).

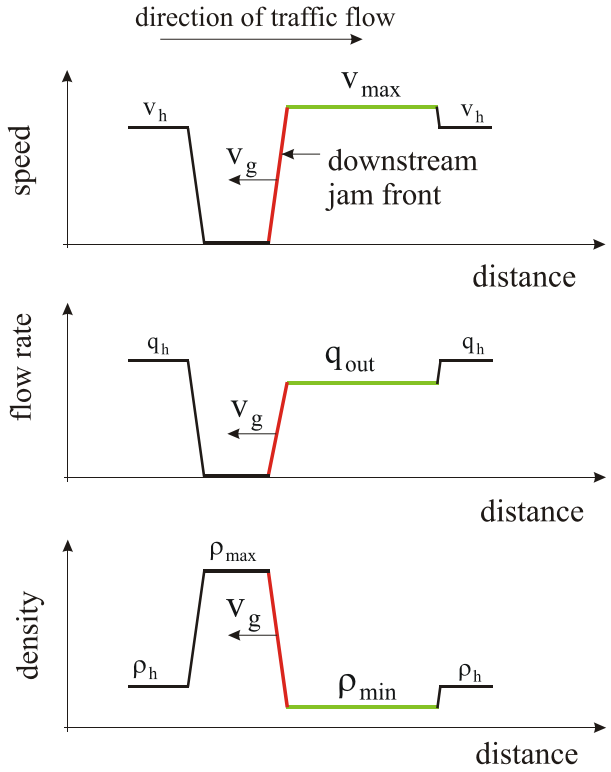


FIGURE 2 Explanation of jam characteristic parameters: Qualitative space distributions of speed, flow rate and density within a wide-moving jam shown in Figure 1 at a given time instant (*13*.)

The characteristic jam parameters can be presented in the flow-density plane by a line *J* whose slope is equal to the mean velocity of the downstream front of the jam v_g (Figure 3). The authors found (*13*) also that states of free flow (F) that are related to the flow rates (Figure 3)

$$q \geq q_{out} \quad (1)$$

are metastable states with respect to an $F \rightarrow J$ transition, i.e., to the emergence of a wide-moving jam in free flow. This means that if a local disturbance of a great enough amplitude occurs in an initial free flow, the jam can emerge in the free flow.

The study of earlier traffic flow models and theories made by Kerner and Konhäuser (*13*) and some later papers have forced the author to begin an empirical data analysis. This was possible to do because from 1995 to 1996 the author worked at the Heusch/Boesefeldt traffic consulting and software company, which dealt with the development of software for the management of traffic data measured on German and Dutch highways. Thus the author had an opportunity to work with a huge amount of real traffic data measured during many years on many different highways.

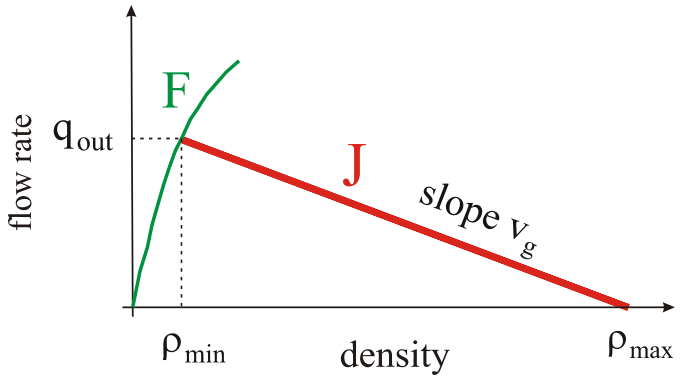


FIGURE 3 Explanation of the line *J* and metastable free flow with respect to wide-moving jam emergence (F – free flow, J – the line J) (*13*).

The empirical data study whose results can be found in the book *The Physics of Traffic* led to the following conclusions about the earlier traffic flow theories and models:

1. The line *J* firstly found in work by Kerner and Konhäuser (*13*) is the only result of earlier traffic flow models, which is in accordance with spatiotemporal features of empirical traffic congestion.
2. None of the known earlier traffic flow models and theories can explain and predict the set of fundamental empirical (measured) features of traffic breakdown.

This empirical data study inspired the author for an alternative traffic theory later called "three-phase traffic theory" that explains traffic breakdown and the resulting congested patterns (*7–11*, *14*).

TRAFFIC PHASE DEFINITIONS AND THEIR SENSE

The empirical study of real measured traffic data mentioned above shows that (*7*)

- There are common pattern features that are qualitatively the same independent of traffic parameters, i.e., they are qualitatively the same on different highways in various countries over many days and years of measurements and they are qualitatively the same for different network infrastructure and highway bottleneck types, weather, percentage of long vehicles, other road conditions, and vehicle technology. Moreover, these common spatiotemporal pattern features are qualitatively independent of day time, working day or weekend, whether the day is sunny or rainy or else foggy, dry or wet road, or even with ice and snow on road, etc.
- The empirical traffic phase definitions (S) and (J) for the synchronized flow and wide-moving jam phases in congested traffic made in three-phase traffic theory are these common empirical pattern features.

The definition of the wide-moving jam phase (J): A wide-moving jam is a moving jam that maintains the mean velocity of the downstream jam front v_g, even when the jam propagates through other traffic phases or bottlenecks.

The definition of the synchronized flow phase (S): In contrast with the wide-moving jam phase, the downstream front of the synchronized flow phase does not exhibit the wide-moving jam characteristic feature; in particular, the downstream front of the synchronized flow phase is often fixed at a bottleneck.

Recall that a moving jam is a propagating upstream localized structure of great vehicle density and very low speed spatially limited by two jam fronts. In the downstream jam front vehicles accelerate escaping from the jam; within the upstream jam front, vehicles slow down approaching the jam.

The empirical traffic phase definitions (S) and (J) are explained in Figures 4 and 5 in which real measured traffic data are presented. In Figure 4, traffic breakdown is induced by a moving jam propagating upstream through an on-ramp bottleneck (labeled by "bottleneck B_1"). The downstream front of congested traffic resulting from traffic breakdown is fixed at the bottleneck, i.e., in accordance with the definition (S) this congested traffic is associated with the

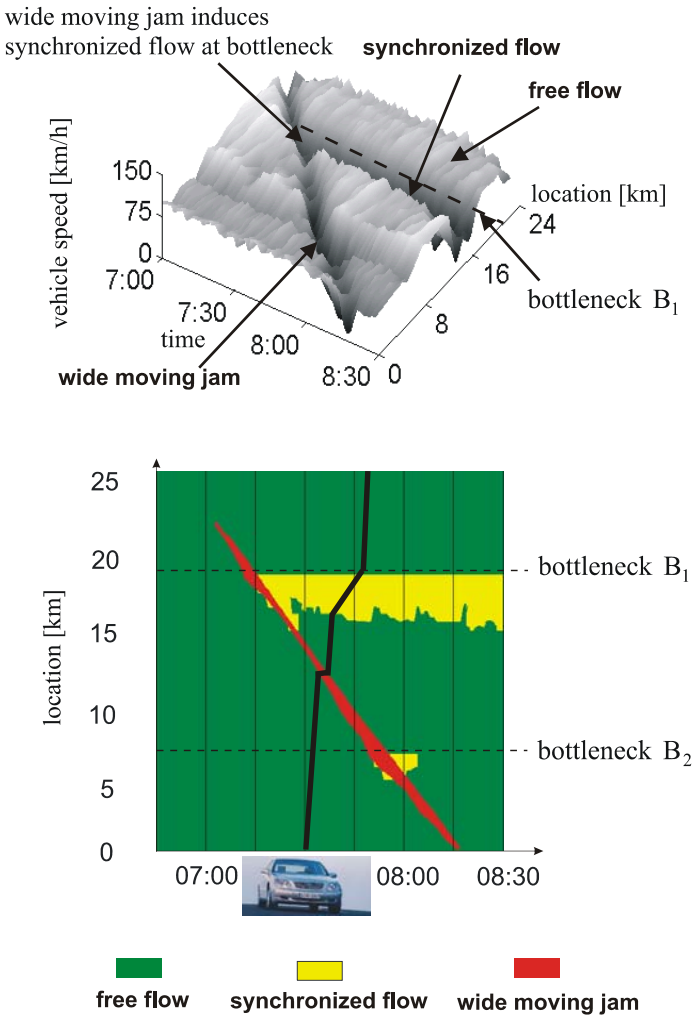


FIGURE 4 Empirical data explaining the traffic phase definitions (J) and (S) (*7*).

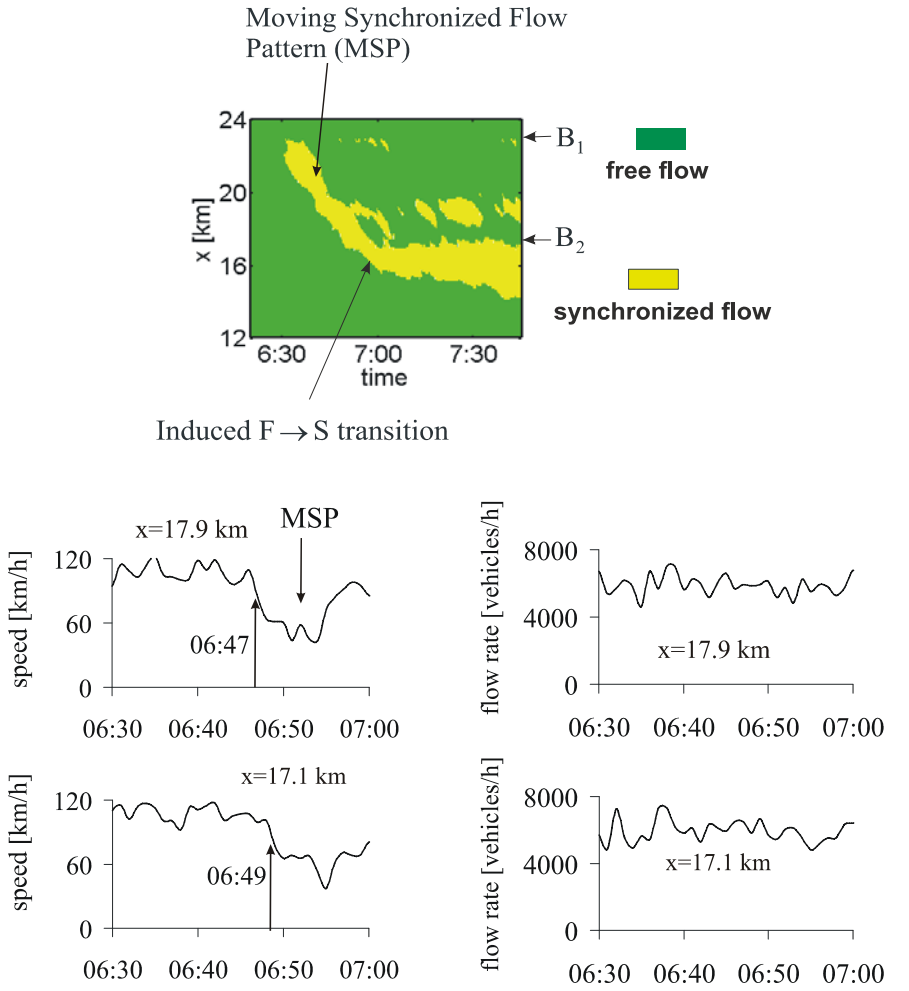


FIGURE 5 Empirical data explaining the traffic phase definition (S): Top: A moving synchronized flow pattern (MSP) that has initially occurred at an off-ramp bottleneck (labeled B_1) propagates upstream and causes an induced traffic breakdown at an upstream on-ramp bottleneck (labeled B_2). Bottom: Average speeds (left) and flow rates (right) as time-functions at two road locations, specifically downstream of the on-ramp bottleneck ($x = 17.9$ km) and at the bottleneck ($x = 17.1$ km) (*7*).

synchronized flow phase. Synchronized flow is self-sustaining for a very long time (more than an hour) upstream of the bottleneck. In contrast with synchronized flow, the moving jam propagates through this bottleneck with the mean velocity of the downstream jam front remaining unchanged (Figure 4). Thus in accordance with the definition (J), this moving jam is associated with the wide-moving jam phase.

Another empirical example of an induced traffic breakdown at an on-ramp bottleneck (labeled by B_2) is shown in Figure 5. In this case, rather than a wide-moving jam a moving synchronized flow pattern (MSP), i.e., a moving congested pattern consisting of the synchronized flow phase only propagates upstream of an off-ramp bottleneck (the off-ramp bottleneck is labeled by B_1 in Figure 5). Indeed, rather than propagating through the on-ramp bottleneck while maintaining the downstream front velocity, the MSP is caught at the bottleneck.

This catch effect is inconsistent with the traffic phase definition (J). In other words, the MSP satisfies the traffic phase definition (S), i.e., the MSP is indeed associated with the synchronized flow phase. Another important feature of the MSP is as follows: the speed within the MSP is lower than in free flow (bottom, left in Figure 5) but the flow rate is almost as great as in the free flow (bottom right in Figure 5). The downstream front of a congested pattern resulting from the induced traffic breakdown is fixed at the on-ramp bottleneck B_2, i.e., in accordance with the definition (S) the resulting congested pattern is also associated with the synchronized flow phase.

In some averaged, i.e., macroscopic traffic data of congested traffic, in particular which are measured outside of bottlenecks, the traffic phase definitions (S) and (J) cannot be applied to perform an accurate identification of traffic phases. Thus a question arises: What is the sense of the traffic phase definitions (S) and (J), if the definitions cannot be used for the identification of traffic phases in some real macroscopic data of congested traffic?

A response on this question is: Rather than to distinguish traffic phases in real measured macroscopic traffic data, the main sense of the traffic phase definitions (S) and (J) is that the definitions result from and, therefore, distinguish qualitatively different common spatiotemporal characteristics of congested traffic patterns found in real measured traffic data.

NATURE OF TRAFFIC BREAKDOWN AT BOTTLENECKS

In three-phase traffic theory, the nature of traffic breakdown at a bottleneck is explained by a competition between two opposing tendencies occurring within a random local disturbance localized at the bottleneck in which the speed is lower and density is greater than in an initial free flow at the bottleneck *(7)*:

1. a tendency toward synchronized flow due to vehicle deceleration associated with a speed adaptation effect (Figure 6*a*) and
2. a tendency toward the initial free flow due to vehicle acceleration associated with an overacceleration effect (Figure 6*b*).

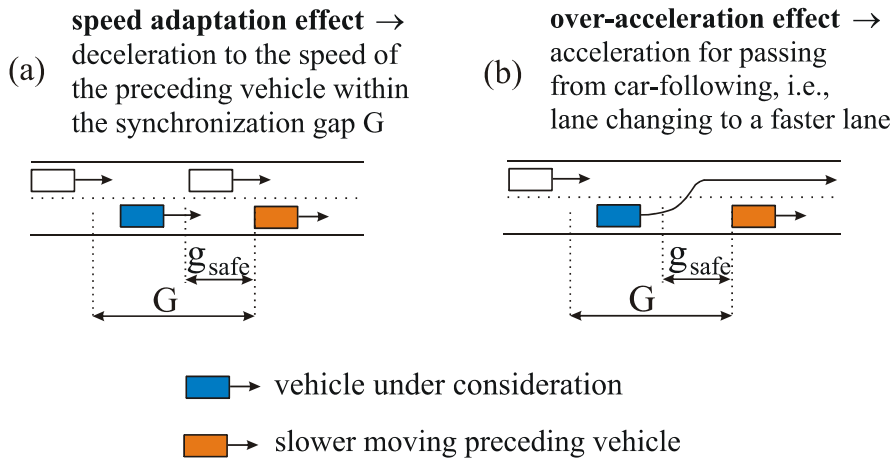


FIGURE 6 Explanation of (*a*) speed adaptation and (*b*) overacceleration (*7, 8*).

To explain this, the authors use the following two hypotheses of three-phase traffic theory *(14)*:

- Steady states of synchronized flow cover a 2-D region in the flow-density plane (Figure 7). The multitudes of free-flow states overlap steady states of synchronized flow in the vehicle density (Figure 7, left). On a multilane road, at each given density the synchronized flow speed is lower than the free-flow speed.
- At each given density ρ, probability of passing P from car following is smaller in synchronized flow than in free flow: $P(\rho)$ is a discontinuous density function, in particular it is a Z-shaped density characteristic (Figure 8).

The low boundary of the 2-D region for the steady states of synchronized flow (Figure 7, left) is associated with a synchronization space gap G between vehicles (Figure 7, right). Because in any steady state of synchronized flow the speed v is time-independent and $v > 0$, a space gap (net distance) between vehicles $g = v\tau$, where τ is a time headway, and therefore the space gap G is associated with the synchronization time headway $\tau_G = G / v$. The existence of this boundary of the 2-D region means that at the space gap $g > G$, i.e., at the time headway $\tau > \tau_G$ a vehicle accelerates (Figure 9).

The upper boundary of the 2-D region (Figure 7, left) is associated with a safe space gap g_{safe} (Figure 7, right), i.e., with a safe time headway $\tau_{safe} = g_{safe} / v$. The existence of this boundary of the 2-D region for steady states of synchronized flow means that at $g < g_{safe}$; i.e., at $\tau < \tau_{safe}$ the vehicle decelerates (Figure 9).

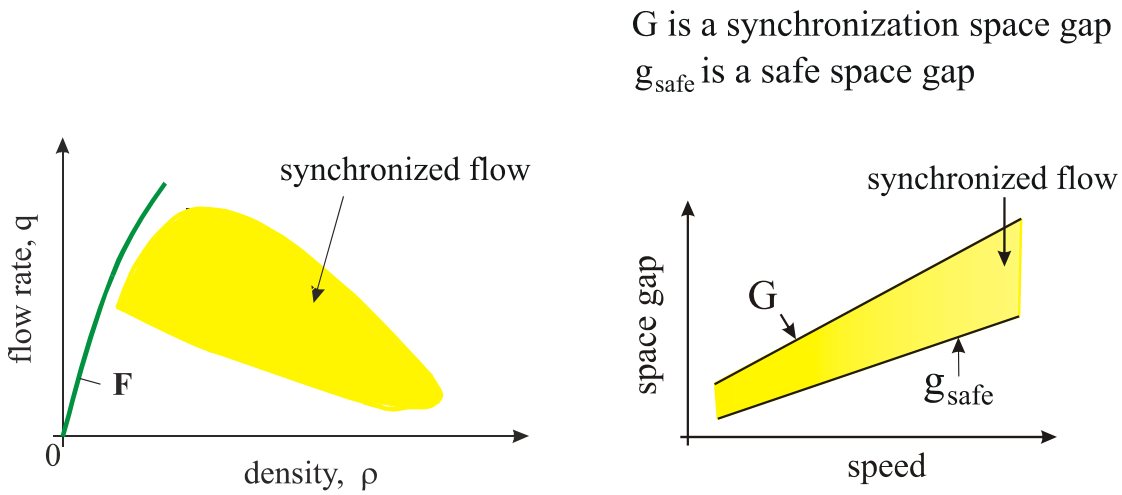


FIGURE 7 Fundamental hypothesis of three-phase traffic theory in the flow-density plane (left) (F – free flow) and in the gap-space-speed plane (right) (*7, 14*).

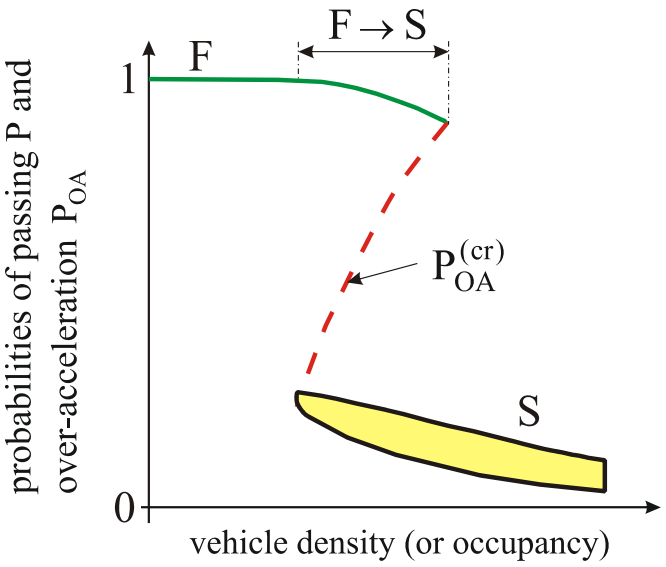


FIGURE 8 Qualitative probabilities of passing and overacceleration as a density (occupancy) function (*7*, *14*).

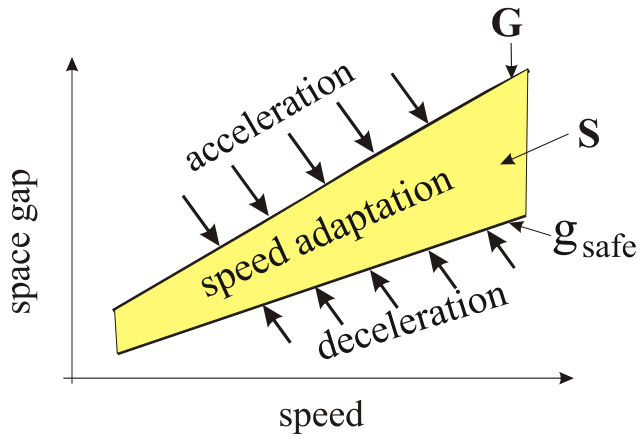


FIGURE 9 Explanation of speed adaptation effect (*7*).

At a given steady speed there is the infinite number of space gaps g within the range (Figure 7) (*7*)

$$g_{safe} \le g \le G \quad (2)$$

which is equivalent to

$$\tau_{safe} \le \tau \le \tau_G \quad (3)$$

The speed adaptation effect is vehicle deceleration within the gap G (i.e., within the time headway τ_G) occurring if a vehicle approaches the slower preceding vehicle and the vehicle

cannot pass it (Figure 6*a*). In a general case, when speeds of the vehicle and the preceding vehicle are not time-independent, the speed adaptation effect leads to car following without caring of the space gap g as long as it satisfies Equation 2: under Equation 2, the vehicle accelerates when it is slower than the preceding vehicle, and decelerates when it is faster than the preceding vehicle.

In general, the overacceleration effect is defined as follows: the overacceleration effect is driver maneuver leading to a higher speed from initial car-following at a lower speed occurring under Equation 2.

To understand the term overacceleration, the authors consider a scenario in which a vehicle moving in free flow on a multilane road approaches a slower moving preceding vehicle. If the vehicle cannot pass the preceding vehicle, then the vehicle decelerates due to speed adaptation leading to car-following of the slow preceding vehicle (Figure 6*a*). The authors assume that later the vehicle can pass this slow preceding vehicle. To pass the preceding vehicle, the vehicle should change lanes and accelerate. The vehicle acceleration takes place, even if the vehicle is not currently slower than the preceding vehicle; this explains why this vehicle acceleration is called overacceleration. Thus in the case, overacceleration is vehicle acceleration for passing from car-following, i.e., lane changing to a faster lane (Figure 6*b*). As a result, probability of overacceleration denoted by P_{OA} is equal to passing probability P from car following:

$$P_{OA} = P \tag{4}$$

i.e., probability for overacceleration is a Z-shaped density function (Figure 8).

The Z-shaped function (Figure 8) explains traffic breakdown as follows *(7)*. Free flow remains on a multilane road as long as the overacceleration effect, which describes the tendency toward free flow, is stronger than the speed adaptation effect that describes the tendency toward synchronized flow. However, the greater the density in free flow, the smaller the probability of overacceleration, i.e., the weaker the overacceleration effect.

There is a critical probability of overacceleration (dashed curves between states F and S in Figure 8) associated with a critical speed denoted by $v_{cr,FS}^{(B)}$ (critical density denoted by $\rho_{cr,FS}^{(B)}$) within a local disturbance in free flow at a bottleneck. This means that when probability of overacceleration is equal to the critical probability of overacceleration, then the tendency toward free flow due to overacceleration within the disturbance is on average as strong as the tendency toward synchronized flow due to speed adaptation.

When the speed within the disturbance is lower than $v_{cr,FS}^{(B)}$ (density is greater than $\rho_{cr,FS}^{(B)}$), then probability of overacceleration within the disturbance is smaller than the critical probability. This means that the overacceleration effect is weaker within the disturbance than the speed adaptation effect. In this case, the tendency toward synchronized flow due to speed adaptation overcomes the tendency toward free flow due to overacceleration that results in traffic breakdown.

Otherwise, if the speed within the disturbance is higher than $v_{cr,FS}^{(B)}$ (density is smaller than $\rho_{cr,FS}^{(B)}$), then probability of overacceleration within the disturbance is greater than the

critical probability. This means that the overacceleration effect is stronger within the disturbance than the speed adaptation effect. In this case, the tendency toward free flow due to overacceleration overcomes the tendency to synchronized flow due to speed adaptation. As a result, no traffic breakdown occurs and free flow remains at the bottleneck.

INFINITE NUMBER OF HIGHWAY CAPACITIES OF FREE FLOW AT A BOTTLENECK

As is well known, traffic breakdown at a bottleneck limits a highway capacity in free flow at the bottleneck (*1*, *2*). However, at the same bottleneck traffic breakdown can be either spontaneous or induced (Figures 4 and 5). The nature of traffic breakdown at a bottleneck disclosed in three-phase traffic theory leads to the following conclusion about highway capacity of free flow at the bottleneck (*7*, *8*): the infinite number of the flow rates in free flow downstream of the bottleneck at which traffic breakdown can be induced at the bottleneck are the infinite number of highway capacities of free flow at the bottleneck.

Thus in three-phase traffic theory there are the infinite number of highway capacities of free flow at the bottleneck between some minimum and maximum capacities (Figure 10). The definition of highway capacities and their sense can be found in *The Physics of Traffic* (*7*). A critical discussion of well-known capacity definitions as a particular fixed or stochastic (probabilistic) value can be found in *Introduction to Modern Traffic Flow Theory and Control: The Long Road to Three-Phase Traffic Theory* (*8*).

WIDE-MOVING JAM EMERGENCE IN SYNCHRONIZED FLOW

In real measured traffic data, wide-moving jams can emerge spontaneously in synchronized flow only. No spontaneous emergence of wide-moving jam in free flow has been observed (*7*). This is consistent with real measured data in which traffic breakdown is governed by an $F \rightarrow S$ transition, i.e., synchronized flow emergence in an initial free flow as explained above.

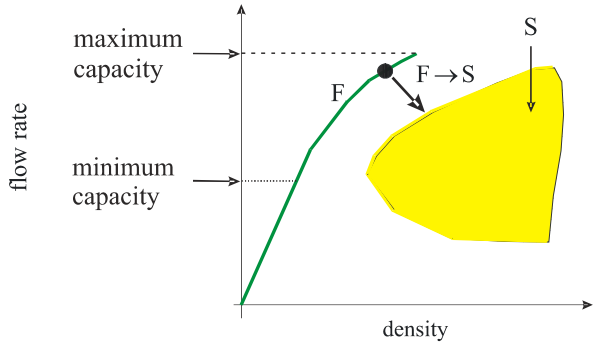


FIGURE 10 Infinite number of highway capacities of free flow at a bottleneck (*7*).

A spontaneous emergence of a wide-moving jam in synchronized flow (S → J transition) is explained in three-phase traffic theory as follows (*7, 14*):

The line *J*, which represents in the flow-density plane the propagation of a wide-moving jam while maintaining the mean velocity of the downstream jam front, divides the 2-D region of steady states of synchronized flow into two different classes (Figures 11–13):

- States below the line *J* are stable with respect to wide-moving jam formation. These states are associated with homogeneous synchronized flow in which no wide-moving jams can emerge or persist over time (Figure 12).
- States on and above the line *J* are metastable synchronized flow states with respect to wide-moving jam emergence, i.e., in metastable synchronized flow wide-moving jams can emerge and exist (Figure 13).

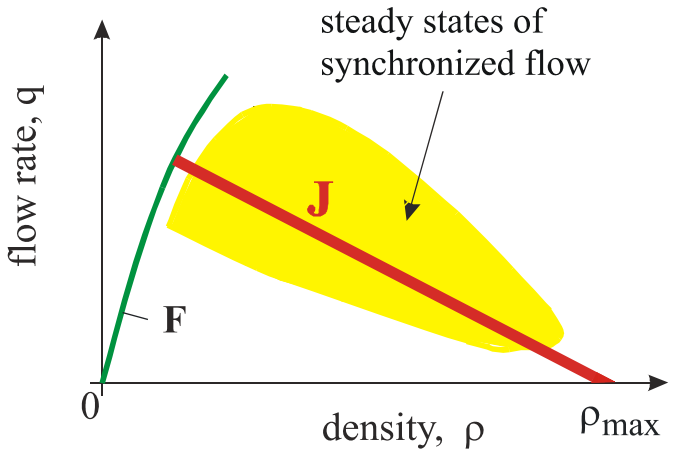


FIGURE 11 Fundamental hypothesis of three-phase traffic theory together with the line *J* (red line) (*7, 14*).

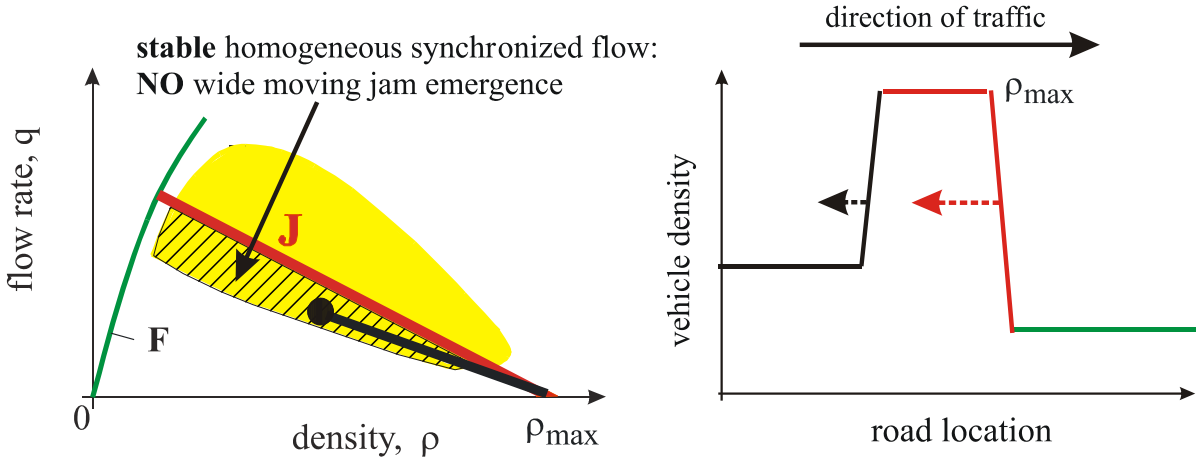


FIGURE 12 Explanation of stable homogeneous synchronized flow (*7, 14*).

To explain this hypothesis, the authors assume that a state of synchronized flow directly upstream of a wide-moving jam (black line upstream of the upstream jam front in Figure 12, right) is associated with a point in the flow-density plane that is below the line *J* (black point in Figure 12, left). The velocity of the upstream front of the wide-moving jam $v_g^{(up)}$ equals the slope of the line in the flow-density plane between this point below the line *J* and the point related to the jam density ρ_{max} (Figure 12, left). The absolute value of this velocity satisfies the condition

$$\left|v_g^{(up)}\right| < \left|v_g\right| \tag{5}$$

where v_g is the mean velocity of the downstream jam front (red in Figure 12, right) given by the slope of the line *J* (red line in Figure 12, left). Thus the width of the wide-moving jam shown in Figure 12, right, gradually decreases and the jam dissolves. This means that no wide-moving jams can persist continuously: i.e., all steady states of synchronized flow below the line *J* are stable with respect to wide-moving jam emergence. These states are homogeneous synchronized flow states, which exist at any density that is possible in synchronized flow, including very great density values. In three-phase traffic theory, stable homogeneous synchronized flow of great density is possible.

In contrast, the authors assume that a state of synchronized flow upstream of another wide-moving jam (Figure 13, right) is associated with a point in the flow-density plane, which is above the line *J* (black point in Figure 13, left). In this case, the velocity of the upstream front of the wide-moving jam $v_g^{(up)}$ equals the slope of the line in the flow-density plane between this point above the line *J* and the point related to the jam density ρ_{max} (Figure 13, left). The absolute value of this velocity satisfies the condition

$$\left|v_g^{(up)}\right| > \left|v_g\right| \tag{6}$$

i.e., it is always greater than that of the downstream jam front $\left|v_g\right|$. Therefore, the width of the wide-moving jam in Figure 13 (right) should gradually increase. For these reasons, wide-moving jams can be formed and persist in states of synchronized flow that lie on or above the line *J*.

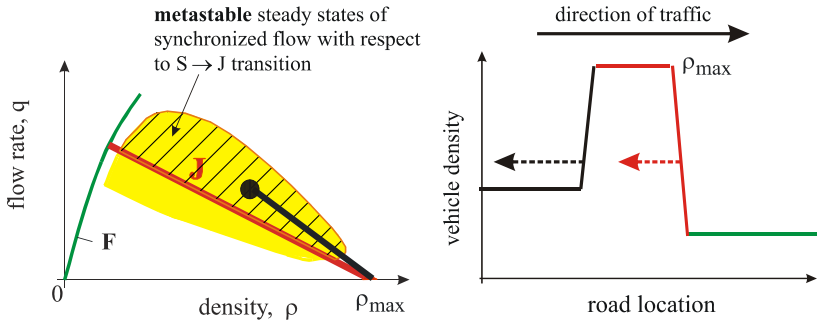


FIGURE 13 Explanation of metastable synchronized flow with respect to wide-moving jam emergence (S → J transition) (*7*, *14*).

Thus there should be some time-limited disturbances in the states of synchronized flow that can lead to the emergence of wide-moving jams: i.e., the states are metastable ones with respect to wide-moving jam emergence.

Thus in three-phase traffic theory at the same given vehicle density depending on the vehicle speed either homogeneous synchronized flow (at lower speeds), in which wide-moving jams do not emerge, or metastable synchronized flow (at higher speeds), in which wide-moving jams can emerge spontaneously, can exist. This explains a great complexity of real traffic congestion observed at highway bottlenecks (*7*, *15*).

FUNDAMENTAL HYPOTHESIS OF THREE-PHASE TRAFFIC THEORY AS THE RESULT OF THE TRAFFIC PHASE DEFINITIONS (J) AND (S)

The fundamental hypothesis of three-phase traffic theory (Figure 7) discussed above results from the phase definitions (J) and (S) as follows (*8, 9*):

- The line *J* in the flow-density plane results from the definition of wide-moving jam (J). Any point on the line *J* can be a final steady traffic state for vehicles accelerating at the downstream front of a wide-moving jam. If the speed in this state is lower than the minimum speed that is possible in free flow, the state is a synchronized flow steady state that lies on the line *J*. Thus there should be infinite synchronized flow states lying on the line *J* (Figure 11).
- The definition of synchronized flow (S) means that downstream fronts of synchronized flow regions do not exhibit the jam characteristic feature (J). For this reason, in addition to synchronized flow steady states lying on the line *J* there should be synchronized flow steady states that are outside of the line *J* in the flow-density plane. Indeed, only in this case the downstream front between two different states of synchronized flow does not exhibit the characteristic jam velocity given by the slope of the line *J*. Thus there should be also synchronized flow steady states outside of the line *J*, i.e., that there is a 2-D region of the steady states in the flow–density plane. This explains the fundamental hypothesis of three-phase traffic theory about the 2-D region of the steady states of synchronized flow (Figure 7).

DOUBLE Z–CHARACTERISTIC FOR PHASE TRANSITIONS

As explained above, in three-phase traffic theory, traffic breakdown is a phase transition from free flow to synchronized flow ($F \rightarrow S$ transition). Wide-moving jams can occur spontaneously in synchronized flow only ($S \rightarrow J$ transition), i.e., due to a sequence $F \rightarrow S \rightarrow J$ transitions. The $F \rightarrow S$ and $S \rightarrow J$ transitions are described in the speed-density plane by a double Z–characteristic (Figure 14) (*7*).

The first Z-characteristic consists of free-flow states (*F*), a branch for a critical speed required for breakdown (dashed curve between states *F* and *S*), and synchronized flow states (2-D region labeled by *S*).

The second Z-characteristic consists of the synchronized flow states, a branch for a critical speed required for an $S \rightarrow J$ transition (dashed curve between states *S* and *J*), and states within wide-moving jam (*J*).

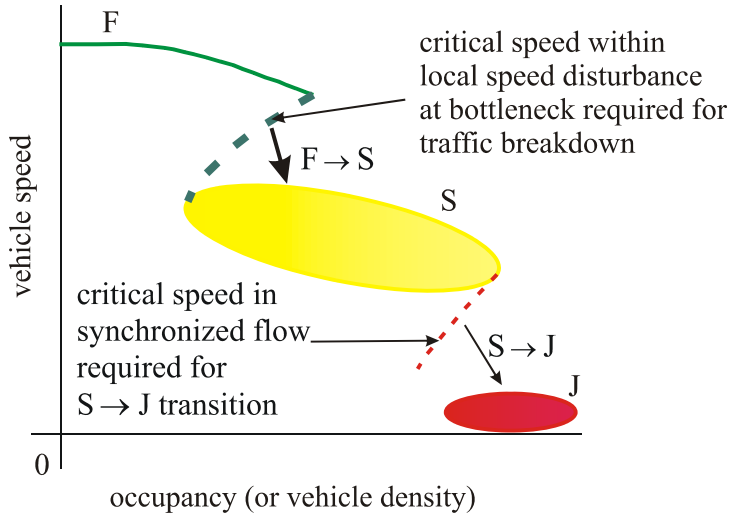


FIGURE 14 Qualitative double z–characteristic for phase transitions (*7*).

It must be stressed that the F → S and S → J transitions shown in Figure 14 occur at different road locations. Therefore, they cannot be distinguished from each other without knowledge of the whole spatiotemporal dynamics of a congested pattern. It can be concluded that, in general, a solely analysis of different congested traffic states and phase transitions in the flow-density plane (as well as in speed–density, space–gap–density planes, etc.) is not adequate with empirical features of congested traffic. This is because most of these empirical spatiotemporal features are lost in this congested traffic analysis. Thus an initial study of the whole spatiotemporal dynamics of the congested pattern is needed.

CLASSIFICATION OF EMPIRICAL SPATIOTEMPORAL CONGESTED PATTERNS

In three-phase traffic theory, in accordance with common empirical features of congested patterns, there is the following classification of congested traffic patterns at a bottleneck (*7*):

1. A synchronized flow pattern (SP) is a congested pattern that consists of the synchronized flow phase only (Figure 15*a*).
2. A general congested pattern (GP) is a congested pattern that consists of the synchronized flow and wide-moving jam phases (Figure 15*b*).

Because on real roads there are usually many adjacent bottlenecks at which traffic breakdown is possible, the authors use also the term expanded congested pattern (EP). An EP is a congested traffic pattern whose synchronized flow affects at least two effectual adjacent bottlenecks.

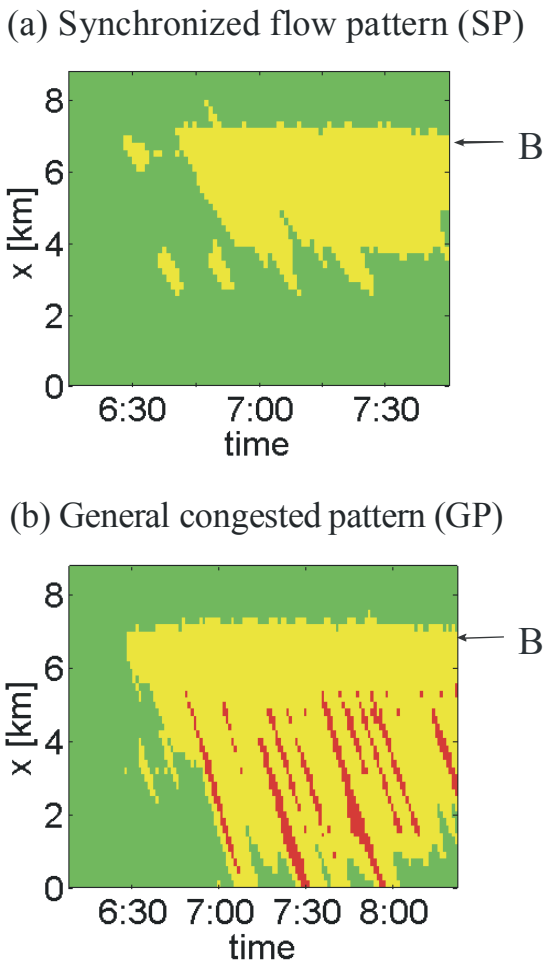


FIGURE 15 Empirical examples of (*a*) synchronized flow pattern (SP) and (*b*) general congested pattern (GP) at an on-ramp bottleneck labeled by B (*7*).

The terminology "expanded congested pattern" emphasizes only that rather than the SP or the GP occurring at an isolated effectual bottleneck, i.e., an effectual bottleneck at which traffic breakdown is observed on many different days (Figure 15), the synchronized flow of the SP or the GP affects two or more adjacent effectual bottlenecks. To emphasize this point, the authors use for these different EPs two terms—"ESP" and "EGP."

- ESP (an expanded synchronized flow traffic pattern) is an SP whose synchronized flow affects two or more adjacent highway bottlenecks.
- EGP (an expanded general congested traffic pattern) is a GP whose synchronized flow affects two or more adjacent highway bottlenecks. An empirical example of an EGP is shown in Figure 16.

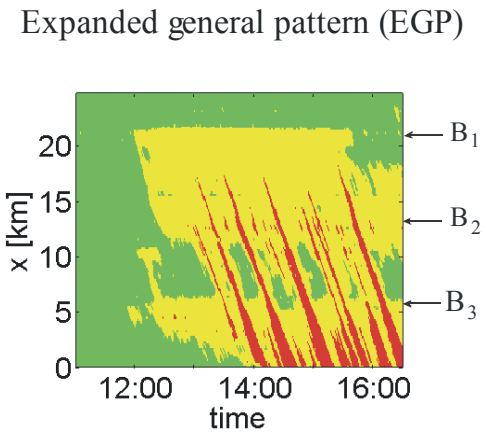


FIGURE 16 Empirical expanded general pattern (EGP) (*7*).

Thus a GP is always a general case of a congested pattern. This is because in three-phase traffic theory there are only the synchronized flow and wide-moving jam traffic phases in congested traffic and, in accordance with the GP definition made above, any GP consists of both phases. This is independent of whether a GP appears at an on-ramp bottleneck, an off-ramp bottleneck, or else the GP is caused by a combination of several on- and off-ramps, etc. A GP is a generic term for many species, i.e., subordinate terms of different types of GPs like an EGP whose synchronized flow affects two or more effectual bottlenecks, or a dissolving GP (*7*), or else a GP with a nonregular pinch region (*15*).

Respectively, an SP in which congested traffic consists of synchronized flow only is the generic term for many species terms of different types of SPs like a moving SP, a localized SP, a widening SP (*7*), an ESP, etc.

As mentioned, this classification of congested patterns in three-phase traffic theory is made based on common empirical features of congested traffic (*7*) found during many years of observations of real measured traffic data. Therefore, the classification of congested patterns into different types of GPs and SPs is well predictive and self-contained (see Sect. 2.4.10 of (*7*)).

Moreover, at a heavy bottleneck, which restricts the average flow rate in a congested pattern upstream of the bottleneck to very small flow rate values, rather these “usual” congested patterns, a mega-wide-moving jam (mega-jam for short) can be formed. A review of a theory of the mega-jam and other congested patterns upstream of heavy bottlenecks caused for example by bad weather or accidents derived in *Journal of Physics* can be found in other works by Kerner (*8*, *10*).

ABOUT DETERMINISTIC MICROSCOPIC THREE-PHASE TRAFFIC FLOW MODEL

The first mathematical traffic flow model based on the above hypotheses of three-phase traffic theory, which can show empirical features of traffic breakdown and resulting congested patterns, was developed in 2002 by Kerner and Klenov. Later, several different three-phase traffic flow models were developed (*8*, *9*). Here, the authors briefly discuss a microscopic deterministic traffic flow model (*16*).

In this model, driver acceleration (deceleration) $a(x, t)$ depends on the traffic phase in which a driver is as follows:

$$\frac{dx}{dt} = v, \quad \frac{dv}{dt} = a, \quad \frac{da}{dt} = \begin{cases} \dfrac{a^{(free)} - a}{\tau^{(del)}} \text{ at } g > G, g_{max}^{(jam)} \\ \dfrac{a^{(syn)} - a}{\tau^{(del)}} \text{ at } g_{max}^{(jam)} < g \le G, \\ \dfrac{a^{(jam)} - a}{\tau^{(del)}} \text{ at } 0 < g \le g_{max}^{(jam)} \end{cases} \quad (7)$$

where a driver time delay $\tau^{(del)}$ depends on the space gap, speed, and acceleration (deceleration); $g_{max}^{(jam)}$ the maximum space gap within a wide-moving jam. Steady states of synchronized flow cover a 2-D region in the space–gap–speed and flow–density planes (Figure 17). The overacceleration effect is simulated through the use of a speed gap between states of synchronized flow and free flow. The model can shown all known empirical features of traffic breakdown and resulting traffic congestion.

LINKING THREE-PHASE TRAFFIC THEORY AND FUNDAMENTAL DIAGRAM APPROACH

A link between three-phase traffic theory (*7*) and the fundamental diagram approach to traffic flow modeling (*1–6*) can be created through the use of the averaging of an infinite number of steady states of synchronized flow shown in Figure 17 to one synchronized flow speed for each density (Figure 18) (*16*). In this case, the authors should find rules for vehicle motion in a traffic flow model whose steady states are associated with a fundamental diagram, however, the model should show the free flow, synchronized flow, and wide-moving jam phases as well as to a sequence $F \rightarrow S \rightarrow J$ transitions (Figure 14). Examples of such three-phase traffic flow models are speed adaptation (SA) models whose steady states are shown in Figure 18.

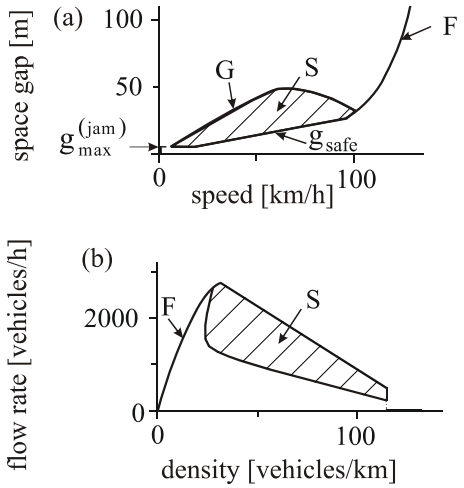


FIGURE 17 Steady states in the ATD model (*16*).

Traffic breakdown is simulated in the SA models through the use of one of the following features of states in a neighborhood of the maximum point on the fundamental diagram (*16*):

- A discontinuity of states on the fundamental diagram between free flow and synchronized flow states (Figures 18*a* and *b*).
- An instability of states on the fundamental diagram within some density range (dashed curve *FS* in Figure 18*c*).

An $S \rightarrow J$ transition is simulated in the SA models through instability of states of synchronized flow on the fundamental diagram whose density is greater than some critical density denoted by $\rho_{cr}^{(J)}$ (dashed parts of curves *S* at the density $\rho > \rho_{cr}^{(J)}$ in Figure 18).

In other words, in these models there are two separated density regions on the fundamental diagram in which phase transitions occur (Figure 19) (*16*):

- A region in a neighborhood of the maximum point on the fundamental diagram labeled by $F \rightarrow S$ in which traffic breakdown occurs and,
- A region of greater densities of synchronized flow labeled by $S \rightarrow J$ in which wide-moving jam emergence in synchronized flow is realized.

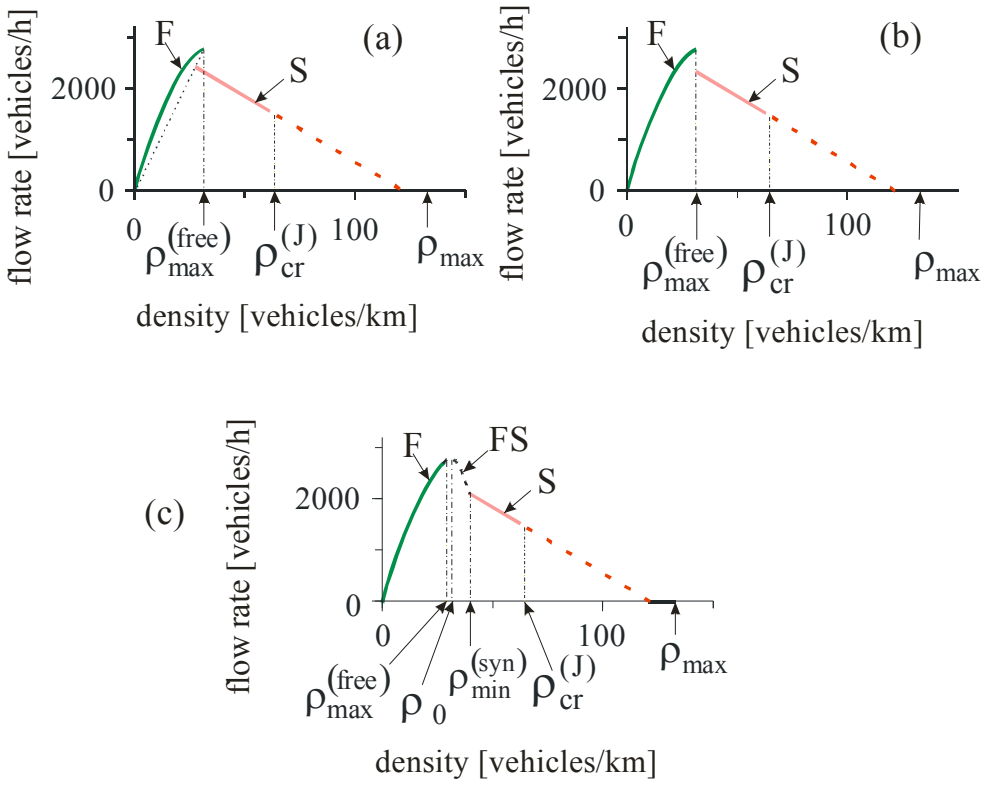


FIGURE 18 Steady states in the SA models (*16*).

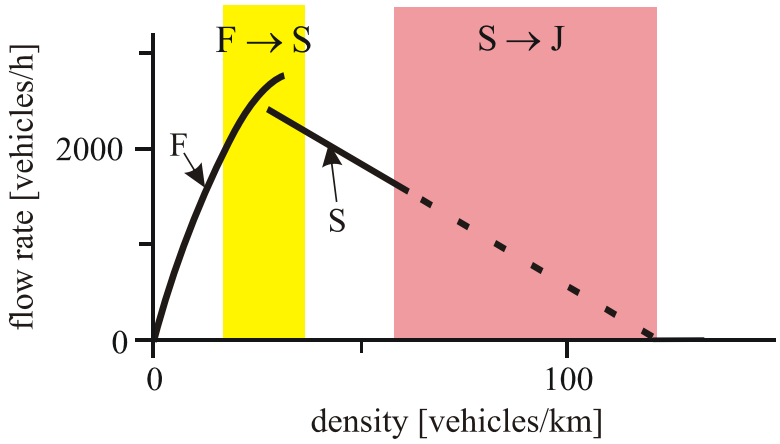


FIGURE 19 Two density regions for phase transitions in SA models (*16*). Steady states of traffic flow in the flow-density plane are the same as those in Figure 18*a*.

As shown by Kerner (*16*), the models are able to show and predict the F → S → J transitions. However, due to the averaging of the infinite number of steady states of synchronized flow to one synchronized flow speed for each density, these models are not able to show important features of synchronized flow as well as many features of coexistence of the free flow, synchronized flow, and wide-moving jam phases found in real traffic. These drawbacks of the models are associated with the ignoring of the 2-D-region of steady states of synchronized flow of three-phase traffic theory.

REFERENCES

1. May, D. *Traffic Flow Fundamentals*. Prentice-Hall, Inc., New Jersey, 1990.
2. *Highway Capacity Manual*. TRB, National Research Council, Washington, D.C., 2000.
3. Gartner, N. H., C. J. Messer, and A. Rathi (eds.) *Traffic Flow Theory. A State-of-the-Art Report*. Transportation Research Board, Washington, D.C., 2001.
4. Gazis, D. C. *Traffic Theory*. Springer, Berlin, 2002.
5. Haight, F. A. *Mathematical Theories of Traffic Flow*. Academic Press, New York, 1963.
6. Drew, D. *Traffic Flow Theory and Control*. New York, New York, McGraw Hill, 1968.
7. Kerner, B. S. *The Physics of Traffic*. Springer, Berlin, New York, 2004.
8. Kerner, B. S. *Introduction to Modern Traffic Flow Theory and Control: The Long Road to Three-Phase Traffic Theory*. Springer, Berlin, New York, 2009.
9. Kerner, B. S. Traffic Congestion, Modeling Approaches to. In *Encyclopedia of Complexity and System Science* (R. A. Meyers, ed.), Springer, Berlin, New York, 2009, pp. 9,302–9,355.
10. Kerner, B. S. Traffic Congestion, Spatiotemporal Features of. In: *Encyclopedia of Complexity and System Science*. Ed. by R. A. Meyers. Springer, Berlin, New York, 2009, pp. 9,355–9,411.
11. Kerner, B. S., and S. L. Klenov. Traffic Breakdown, Probabilistic Theory of. In *Encyclopedia of Complexity and System Science* (R. A. Meyers, ed.), Springer, Berlin, New York, 2009, pp. 9,282–9,302.
12. Kerner, B. S., and V. V. Osipov. *Autosolitons: A New Approach to Self-Organization and Turbulence.* Kluwer, Dordrecht, Boston, London, 1994.

13. Kerner, B. S., and P. Konhäuser. Structure and Parameters of Clusters in Traffic Flow. *Phys Rev E* 50, 1994, pp. 54–83.
14. Kerner, S. Congested Traffic Flow: Observations and Theory. In *Transportation Research Record*: *Journal of the Transportation Research Board, No. 1678*, TRB, National Research Council, Washington, D.C., 1999, pp.160–167.
15. Kerner, S. A Theory of Traffic Congestion at Heavy Bottlenecks. In *Journal of Physics A: Mathematical and. Theoretical,* Vol. 41, 215101, 2008.
16. Kerner, B. S., and S. L. Klenov. Deterministic Microscopic Three-Phase Traffic Flow Models. In *Journal of Physics A: Mathematical and. Theoretical,* Vol. 39, 2006, pp. 1775–1809

The Fundamental Diagram
From Theory to Practice

Traffic Flow Theory Historical Research Perspectives

S. L. DHINGRA
ISHTIYAQ GULL
IIT Bombay, India

Traffic flow is a kind of many-body system of strongly interacting vehicles. Traffic jams are a typical signature of the complex behavior of vehicular traffic. Various mathematical models are presented to understand the rich variety of physical phenomena exhibited by traffic. This paper provides an overview of what is currently the state of the art with respect to traffic flow theory. Starting with a brief history about vehicular traffic flows, this paper discusses the Greenshields, Greenberg, and Gurein, etc., models. This paper also discusses some basic relations between traffic flow characteristics, i.e., the fundamental diagrams; speed, volume, and density relationships; hydrodynamic analogies; and traffic hump formation (shock wave). It also sheds some light on the different points of view adopted by the traffic engineering community. Some performance indicators are reviewed that allow assessment of the quality of traffic operations. A final part of this paper gives the probabilistic description of traffic flow, distribution of vehicles on a road.

Traffic-flow theories seek to describe in a precise mathematical way the interactions among vehicles, drivers, and the infrastructure. The infrastructure consists of the highway system and all its operational elements, including control devices, signage, and markings. These theories are an indispensable element of all traffic models and analysis tools that are being used in the design and operation of streets and highways. The scientific study of traffic flow had its beginnings in the 1930s with the application of probability theory to the description of road traffic and with the pioneering studies conducted by Bruce D. Greenshields at the Yale Bureau of Highway Traffic on the study of models relating volume and speed and the investigation of performance of traffic at intersections. After World War II, with the tremendous increase in the use of automobiles and the expansion of the highway system, there was also a surge in the study of traffic characteristics and the development of traffic-flow theories. In December 1959, the First International Symposium on the Theory of Traffic Flow was held at the General Motors Research Laboratories in Warren, Michigan. This was the first of what has become a series of triennial symposia on the theory of traffic flow and transportation. The field of traffic-flow theory and transportation has become too diffuse to be covered by any single type of meeting, and numerous other symposia and specialty conferences about a variety of traffic-related topics are held on a regular basis. Yet, even as traffic-flow theory is increasingly better understood and more easily characterized through advanced computation technology, the fundamentals are just as important today as in the early days. They form the foundation for all the theories, techniques, and procedures that are being applied in the design, operation, and development of advanced transportation systems. This elementary and brief introduction to traffic-flow theory is included to extend the engineer's knowledge in this vital area and to relate the theory to other aspects of traffic engineering.

BASIC RELATIONSHIPS

Earlier, the major elements of traffic flow were outlined as: (1) composition or classification, (2) volume, (3) origin and destination, (4) quality, and (5) cost. Traffic flow is concerned particularly, with three of these elements: composition, volume, and quality.

One objective of traffic flow theory is to derive theoretical relationships between the various traffic variables so that the engineer can determine the characteristics of traffic streams and hence predict the consequences of alternative designs. Initial work is concentrated on the flow, density, and space–mean–speed variables as these three variables are of particular interest since flow describes how many vehicles are moving, and flow (U_s = resultant demand), density (D), and space–mean–speed (V) together describe the quality of service experienced by drivers, which can serve as a useful first approximation to travel costs. These three variables are related to each other by the equation

$$V = D * U_s \tag{1}$$

This relationship represents the fundamental equation of traffic flow. If a relationship is established between any two of the variables, the relationship of the third can be controlled by Equation 1. Obviously the possible combinations are volume and density, speed and density, and speed and volume, all of which have been investigated empirically.

At this point it is useful to consider some of the characteristics of the physical driver-vehicle-highway system and the manner in which it operates, and to set out the basic conditions that must be satisfied by any explanatory theory. The definitions clearly require volume to be zero when density is zero, and require volume equal to zero at maximum density (i.e., when vehicles are lined up end-to-end). Further, it is evident that volume decreases before density reaches a maximum value (Note that the maximum volume that can pass through a given section of highway is referred to as the highway "capacity" by traffic engineers). The relationship between volume and density, then, must have a general form similar to that illustrated in Figure 1.

This has been termed the "fundamental diagram of road traffic" by Haight. Clearly there are innumerable volume–density functions that will pass through the zero-volume points and have a maximum volume value falling in between, and several functions have been proposed or derived from empirical data or theoretical consideration.

Since space mean speed is volume divided by density, the slope of the line $\overline{OA}$ in Figure 1 represents the space mean speed corresponding to volume V_A and density D_A. It is also clear that for the volume level V_A, there is another possible speed, defined by the slope of the line $\overline{OB}$ and corresponding to density D_B.

Further, at the jam density D_j, speed is zero, and as the volume and density approach zero, the speed will be equal to the means free speed (i.e., the speed drivers will assume when free or virtually free from inference by other vehicles). The mean speed will be a function of the drivers, their vehicle characteristics, the highway characteristics (lane width, sight distance, etc.), and other factors such as lighting and weather.

In general, the relationship between speed and volume will be of the form shown in Figure 2. As volume increases, the space mean speed decreases, and travel time increases. If the input volume for the facility approaches V_m, the dynamics of traffic flow and shock wave action

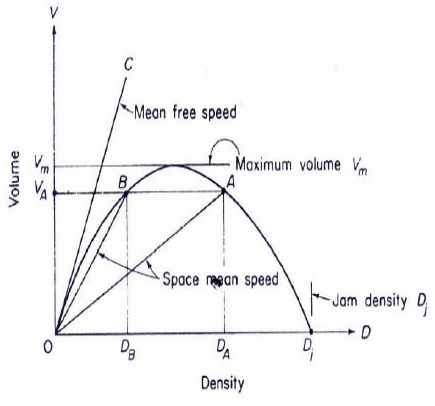


FIGURE 1 The fundamental diagram.

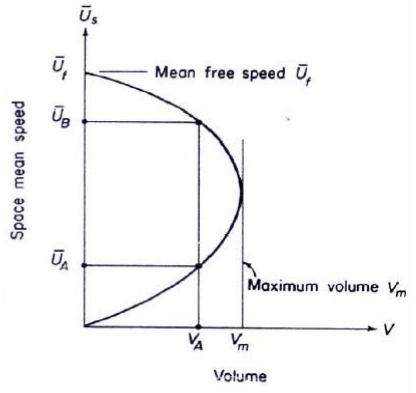


FIGURE 2 General form of road traffic volume-speed relationship.

may cause the capacity to be reduced below V_m, with speeds corresponding to lower-speed portion of the curve; in such an instance, the maximum volume that can pass through the roadway section might be reduced to V_A, say, and the speed to U_A. From the traffic engineers' and highway users' point of view, the reduction in speed and in capacity is clearly an undesirable situation, because only few vehicles can pass and because they must travel at lower speeds than appears reasonable for the section of highway. In the section that follows, the results of the more pertinent research on traffic flow theory will be summarized and applied to these sorts of problems.

EMPIRICAL STUDIES OF VOLUME, DENSITY, AND SPEED RELATIONSHIPS

One of the earliest recorded works is that of Greenshields in 1934. He found a linear relationship as shown in Figure 3 between average density and average speed of the form

$$U_s = U_f - \left(\frac{U_f}{D_j}\right) * D \tag{2}$$

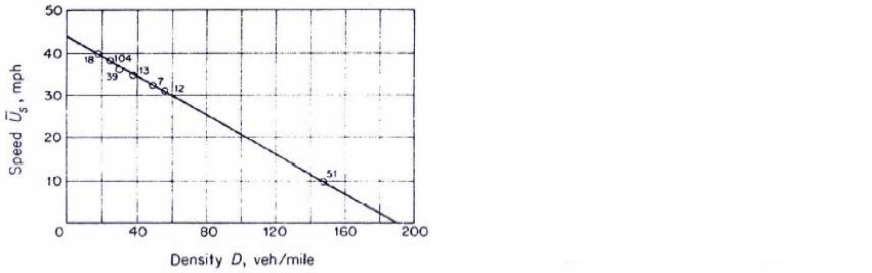


FIGURE 3 Speed–density relationship.

where U_f is the mean free speed, and D_j is the jam density. Thus for a Greenshields' data with the mean free speed U_f as 46 mph and the jam density D_j as 195 vehicle per mile; thus for his data,

$$U_s = 46 - 0.236 * D$$

If a linear relationship between speed and density is assumed, then the corresponding functions for volume and density and for volume and speed are parabolic. The volume density relationship can be derived by substituting V/D for U_s in Equation 2; the result is

$$V = U_f D - \frac{U_f}{D_j} D^2 \tag{3}$$

Similarly, by substituting V/U_s for D in Equation 2, the volume–speed relationship is

$$V = D_j U_s - \frac{D_j}{U_f} U_s^2 \tag{4}$$

These two relationships are illustrated in Figure 4. To determine (1) the density, and (2) the speed at which the flow or volume is at maximum, Equations 3 and 4 must be differentiated with respect to density and speed, respectively; the resulting differentials can then be set equal to zero and solved. Thus:

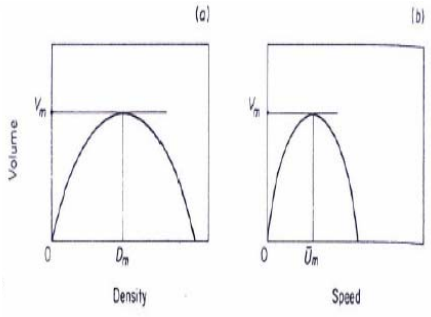


FIGURE 4 Parabolic volume based on 244 groups of 100 vehicles' density and volume–speed curves.

Density when volume is maximum (D_m) :

$$\frac{dV}{dD} = U_f - \frac{U_f}{D_j} 2D = 0 \text{ and } D = D_m = \frac{D_j}{2} \tag{5}$$

Speed when volume is maximum (U_m) :

$$\frac{dV}{dU_S} = D_j - \frac{D_j}{U_f} 2U_s = 0 \text{ and } U_s = U_m = \frac{U_f}{2} \tag{6}$$

From Equations 1, 5, and 6, the maximum volume is

$$V_m = D_m U_m = \frac{D_j U_f}{4} \tag{7}$$

Thus for the Greenshields' data with U_f = 46 mph and D_j =195 veh/mile, the D_m = 98 veh/mile, U_m = 23 mph, V_m = 2,240 veh/h. Greenberg analyzed data collected in the north tube of the Lincoln Tunnel and fitted an exponential function to the speed and density observations, which are shown in Table 1. His hypothesis that speed and density are related exponentially was based on a theoretical analysis in which high-density traffic flow was assumed to be analogous to continuous fluid flow. The form of the function was

$$D = Ce^{bU_s} \tag{8}$$

where e is the base of the Napierian or natural log system. The parameters C and b can be estimated by making a least-square regression fit to the data shown in Table 1. To express Equation 8 in a linear form, algorithmic transformation is made as follows:

$$\log_e D = \log_e C + b U_s$$

Putting Y = $\log_e D$, and c = $\log_e C$, the linear equation becomes

$$Y = c + b U_s \text{ or } Y = a + b (x - \bar{x}) \tag{9}$$

where x is U_s, and $\bar{x}$ is the mean value of U_s (that is, $\bar{x} = \sum\left(U_s/n\right)$).

From the data in Table 1, after transformation,

$\sum y = 81.098$, $\sum U_s = 285$, $\sum y^2 = 368.603$, $\sum U_s^2 = 5{,}453$, $\sum U_s y = 1{,}229.325$

From the above equations, b = –0.0581 or –1/17.2 and C = 5.425109, thus

$\log_e C$ = 5.425109, C = 227, and

$$D = 227e^{-(U_s/17.2)} \tag{10}$$

TABLE 1 Speed and Density Observations in Lincoln Tunnel

Volume *V, veh/h*	Speed U_s, *mph*	Density *D, veh/mile*
1,088	32	34
1,232	28	44
1,325	25	53
1,380	23	60
1,480	20	74
1,558	19	82
1,496	17	88
1,504	15	94
1,410	15	94
1,344	14	96
1,339	13	102
1,344	12	112
1,188	11	108
1,290	10	129
1,188	9	132
1,112	8	139
1,120	7	160
990	6	165

The estimated coefficient of determination calculated comes out to be $r^2 = 0.988$, and the standard error of estimate for density once calculated yields $s = 4.93$ vehicles/mi.

Figure 5 includes the observed data and the plot of Equation 10. The exponential form gives an infinite mean free speed and a jam density of 227 vehicles/mi.

In the former respect, the exponential violates the boundary condition for speed at zero density. The relationship between speed and volume follows directly from Equations 1 and 10:

$$V = U_s\,(227e^{-(\mathrm{Us}\,/17.2)} \tag{11}$$

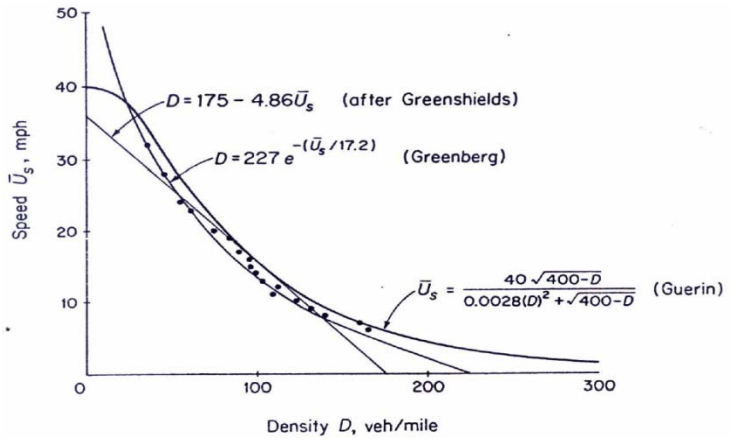


FIGURE 5 Density–speed relationships.

This volume–speed equation is shown in Figure 6, together with Greenberg's field data. The standard error of estimate and estimated coefficient of determination have been computed, giving s = 61.7 vehicles/h, and r^2 = 0.85, respectively. From both statistics and Figure 6, the derived form for the volume–speed relationship is not as suitable as the equation for speed–density. In particular, Equation 11 underestimates volume in the region of maximum volume. Violation of the boundary condition at zero density is also evident, because speed is infinite at zero volume. Using Equation 11, a maximum volume of 1,460 veh/h occurs at a speed of 17.2 mph.

Figure 7 shows the relationship between volume and density and the observed field data. The functional relationship is

$$V = 17.2\, D \log_e \frac{227}{D} \tag{12}$$

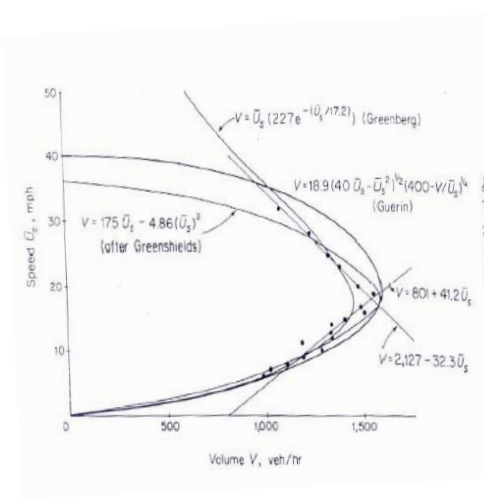


FIGURE 6 Volume–speed relationships.

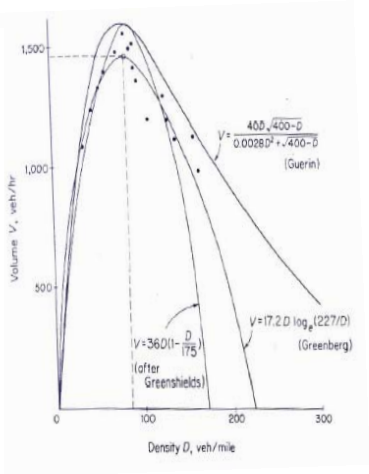


FIGURE 7 Volume–density relationships.

This equation satisfies the volume–density boundary conditions and has the general form postulated for the fundamental diagram in Figure 1. The density at maximum volume, as shown in Figure 7, is approximately 82 vehicles/mi.

For purpose of comparison, a linear equation has been fitted to Greenberg's density–speed data. A least-squares regression fit was made for an equation with the following general form:

$$D = C + bU_s$$

For the data in Table 1,

$\sum D = 1{,}766$, $\sum U_s = 285$, $\sum D^2 = 197.156$, $\sum U_s{}^2 = 5{,}453$, $D\,U_s = 23{,}375$, $C = 175$, $b = -4.8576$.

Thus

$$D = 175 - 4.86\,U_s, \quad \text{and } r^2 = 0.936,\ s = 8.5 \text{ vehicles/mi} \tag{13}$$

Equation 13 is shown in Figure 5, where it may be observed that the linear regression fit is not as suitable as Greenberg's equation. In particular, there are large deviations for densities less than 50 veh/mi and for densities larger than 140 veh/mi. This has resulted in a pattern of systematic errors for the regression line. However, as shown above, the r^2 value and the standard error of estimate were reasonably good for most engineering purposes.

The previous comparison illustrates an important aspect of curve fitting, and a distinction should be made between "good prediction functions" and the "true functions." If engineers are primarily interested in obtaining "good or satisfactory prediction" equations between speed and density for the same range of data values, the linear form may be preferable because it is easy to use. However, the linear form cannot be assumed as the true relationship between the variables, because of having a high r^2 value. In fact, there is every reason to assume that it cannot be the true relationship in this case because the errors appear to be systematically distributed about the regression line, indicating that some form of concave function is closer to the true relationship. In the present case, the exponential form of the function does have a higher value of r^2, and the errors do not show any obvious bias; thus the exponential form is more likely to be a true function than a linear form, although even it may not be the true form of relationship.

Thus, as discussed previously, the linear relationship between speed and density leads to a parabolic curve for the derived forms; they are

$$V = 175\,U_s - 4.86\,(U_s{}^2) \tag{14}$$

$$\text{And } V = 36D\left(\frac{1-D}{175}\right) \tag{15}$$

It is evident that all the boundary conditions specified earlier are satisfied by Equations 14 and 15. The maximum volume of 1,580 veh/h occurs at a speed of 18 mph and at a density of approximately 82 veh/mi.

Many empirical studies have been used on observations of travel time and volume with particular emphasis on developing working relationships. Guerin, in an analysis of data collected

in Chicago and New Haven, suggested the use of boundary curves. For speed and density, the boundary curve is defined as

$$U_s = \frac{U_f\sqrt{D_j - D}}{AU_fD^2 + \sqrt{Dj - D}}$$

For Greenberg's data, the following constants and parameter values are suitable:

$$U_f = 40,\ D_j = 400,\ A = 0.00007, \text{ giving } U_s = \frac{40\sqrt{400 - D}}{0.0028D^2 + \sqrt{400 - D}} \tag{16}$$

Equation 16 is plotted on Figure 5. The boundary curves for the other variables are

$$V = 18.9(40U_s - U_s^2)^{\frac{1}{2}}\left(400 - {}^{V}\!/_{U_s}\right)^{\frac{1}{4}} \tag{17}$$

$$V = \frac{40D\sqrt{400 - D}}{0.0028D^2 + \sqrt{400 - D}} \tag{18}$$

The equations shown in Figures 6 and 7 satisfy the four boundary conditions stated previously. Guerin pointed out that the form of the fundamental diagram defined by Equation 18 also satisfies a fifth possible boundary condition, as follows:

$$\lim_{D\to 0}\frac{dU_s}{dD} = 0 \tag{19}$$

This boundary condition implies that a small number of vehicles can be added to an empty system without a decrease in the mean free speed.

A number of empirical studies have postulated a linear relationship between space mean speed and volume. One of the first can be attributed to Normann and was later used in the *Highway Capacity Manual*. This functional form does not serve as a satisfactory basis for developing a theory of traffic flow, but it does seem to have useful engineering applications. As an illustration, linear equations have been fitted to Greenberg's speed–volume data (Table 1). A general equation of the following form was fitted by least squares of two groups of data:

$$V = c + bU_s$$

The first group of data included the first five entries in Table 1 and resulted in the following equation:

$$V = 2{,}127 - 32.3U_s \text{ for } U_s \geq 18.1 \text{ mph}, \tag{20}$$
$$r^2 = 0.998, \quad s = 7.5 \text{ veh/h}$$

The remaining 13 observations in Table 1 resulted in the following equation:

$$V = 801 + 41.2\ U_s \quad \text{for } U_s \leq 18.1 \text{ mph}, \tag{21}$$
$$r^2 = 0.939, \quad s = 41.5 \text{ veh/h}$$

Equations 20 and 21 are shown in Figure 6. Maximum volume and the associated speed are V = 1,541 veh/h and U_s = 18.1 mph; because V is equal to $U_s D$, D at maximum volume is 85 veh/mi.

Equations 20 and 21 are only applicable for a particular range of speeds since both give unrealistic values for volume when extended beyond the range of Greenberg's data. Within that range they appear to be better predictors than any of the curves discussed earlier. Thus a volume–speed curve consisting of three linear segments may prove satisfactory for many engineering problems.

HYDRODYNAMIC ANALOGIES

Analogies have been often drawn between the flow of fluids and the movement of vehicular traffic. However, the current evidence suggests that the equations developed from hydrodynamic analogies hold good only for high-density traffic and indicates that continuous and steady flow analogies almost totally obscure the fact that each vehicle is individually controlled. Even so, most important traffic-control problems occur only under high-density and other than "free movement" conditions; thus it might appear that better understanding of these analogies is worthwhile.

Principal contributions to this topic have been made by Greenberg, Lighthill and Whitham, and Richards. To develop the basic relations, Greenberg's approach will be used. The assumptions are

1. High-density traffic will behave like a continuous fluid, and the corresponding fundamental motion equation for one-dimensional continuous fluid is

$$\frac{dU_s}{dt} = -\frac{c^2}{D}\frac{\partial D}{\partial x} \tag{22}$$

where

U_s = fluid velocity or space mean speed, mph;
D = density, vehicles/mile;
X = distance, miles;
T = time to move distance x; and
C = roadway parameter.

Equation 22 can be rewritten by expressing speed in terms of distance and time:

$$\frac{\partial U_s}{\partial t} + \frac{U_s \partial U_s}{\partial x} + \frac{c^2}{D}\frac{\partial D}{\partial x} = 0 \tag{23}$$

2. For a fluid, mass-per-unit time is constant (this is normally termed the continuity of flow); thus, if traffic and fluid flow are analogous,

$$\frac{\partial D}{\partial t} + \frac{\partial V}{\partial x} = 0 \tag{24}$$

where V is the volume or flow rate in vehicles per hour. Also Equation 23 can be restated:

$$V = U_s D \tag{25}$$

3. If it is further assumed that speed is only a function of density, Equation 23 to Equation 24 may solved to yield speed in terms of density:

$$U_s = c \log_e \frac{D_j}{D} \tag{26}$$

where D_j is the jam density. Substituting speed from Equation 25 into Equation 26 gives the following result:

$$V = cD \log_e \frac{D_j}{D} \tag{27}$$

Equation 27 defines the form of the fundamental diagram shown previously in Figure 1. It can easily be verified that

1. When $V = 0$, $D = 0$ or $D = D_j$ and
2. When $D = 0$, $U_s = \infty$, $D = D_j$, $U_s = 0$.

Thus, as a complete theory of traffic flow, the formulation violates the boundary condition that at zero volume and density the speed is the free mean speed. This is particularly obvious in Figure 6, where volume and speed are graphed. This figure also illustrates that at high volumes (and hence densities) the hydrodynamic analogy is a good description.

Other useful concepts can be derived from fluid flow analogy to traffic flow. Lighthill and Whitham have shown that the speed of waves "carrying" continuous changes of volume through a vehicular flow is given by dV/dD or the slope of the fundamental diagram; thus

$$U_w = \frac{dV}{dD} = \frac{dU_s D}{dD} = U_s + D\frac{dU_s}{dD} \tag{28}$$

where U_w is the speed of the wave in miles/h.

As shown previously (e.g., Figure 5), the space mean speed of traffic decreases with increasing density, and thus dU_s/dD will be negative. Also, Equation 28, the speed of the wave (U_w) will be less than the mean speed of traffic (U_s), and the wave will move backward relative to the mean vehicular flow. At low densities, when interaction between vehicles is very small, dU_s/dD approaches zero, at which point the wave velocity equals the vehicle speed. At densities above the point of maximum volume, the waves will move backward relative to the road, and under the conditions for maximum volume, the wave is stationary relative to the road; at lower densities, the wave moves forward relative to the road (see Figure 8).

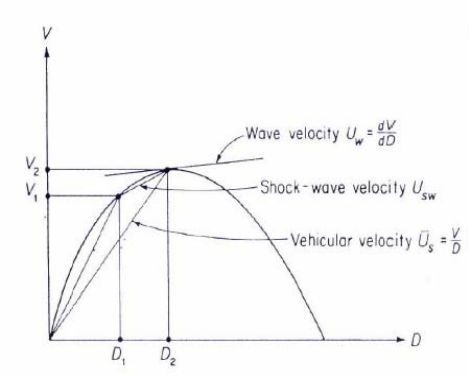


FIGURE 8 Velocity and volume.

Since the wave velocity changes with the vehicular density, it is possible to have different waves traveling through a vehicular stream. A particularly interesting case is a section of road with low-density flow immediately followed by a section with higher-density flow. This situation might be caused by an accident, a reduction in number of lanes, an entrance ramp, or other constricted circumstances. The wave in the lower-density traffic will travel forward (relative to the road) at a higher speed than the wave in the higher-density flow. When these waves meet there will be a change in flow and a new wave will form. Both the original waves and new wave will move forward relative to the road. The new wave has been termed a shock wave by Lighthill and Whitham, and its speed is given by the slope of the chord joining the respective points on the fundamental diagram (Figure 8); analytically, its speed U_{sw} is as follows:

$$U_{sw} = \frac{V_2 - V_1}{D_2 - D_1} \tag{29}$$

Lighthill and Whitham illustrate their work with several examples concerning conditions that sometimes occur in vehicular flow. The first example they term a "traffic hump," which is a period during which the flow of vehicles onto a highway suddenly increases. A typical example might be the release or discharge of a large number of vehicles from a parking lot. The highway is assumed uniform along its length, and hence the fundamental diagram will have the same form at all points. The wave pattern of flow may be illustrated conveniently on a space–time diagram as shown in Figure 9.

The light lines on Figure 9 show the paths of waves, each traveling at constant speed along the highway. During periods of normal flow input, the wave velocities are the same and hence are shown by parallel lines in Figure 9. As the input flow and densities increase, the wave velocity is reduced. As the period of increased flow ends, the wave velocities increase again until they reach normal value, as shown in Figure 9. The faster waves at the right will meet the slower, causing a discontinuity and the formation of a shock wave, as shown by the dotted lines in Figure 9.

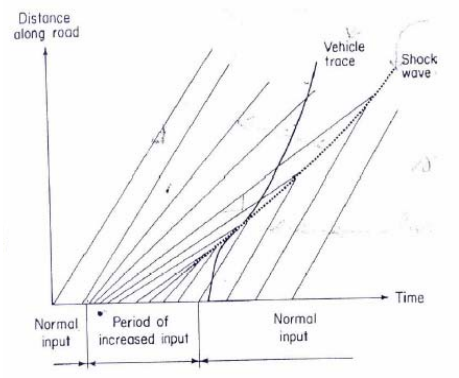


FIGURE 9 Shock wave density relationships.

CAR-FOLLOWING THEORY

The hydrodynamic analogy discussed in the previous section relied on the assumption that high-density streams of vehicles behave like an incompressible fluid. In other words, it was assumed that individual drivers react to the presence of other vehicles in such a way that only the stream or total flow characteristics need be considered. The character of intervehicle relationships was not specified or necessary to analogy. Car-following theory, by contrast, deals with the intervehicle relationship and from this basis builds up a description of character of total vehicular flow.

Car-following theories have been developed by numerous researchers, but they often are associated with the work of Herman et al. at the research laboratories of General Motors Corporation. Herman pointed out that the "follow the leader" types of problems basic to car following are the product of the psychological behavior of the drivers as they respond to certain stimuli. The first car-following model considered is one in which it is assumed that a driver will attempt to keep relative speed between his vehicle and the one immediately ahead as small as possible. The stimulus will be a change in the speed of the vehicle ahead, and the lagged response will be a change in the speed of the following vehicle, whereas the process is constrained to minimize differences in the relative velocity between the two vehicles. These ideas can be expressed in an equation of the form

$$x''_n(t) = \lambda\left[x'_{n-1}(t-T) - x'_n(t-T)\right] \tag{30}$$

where

$x''(t)$ = acceleration of the nth vehicle at time t, or $a_n(t)$;
$x'_n(t)$ = velocity of nth vehicle at time t, or $U_n(t)$;
T = time lag or stimulus-response time of driver-car system; and
Λ = driver-sensitivity coefficient.

Equation 30 can be rewritten as

$$a_n(t) = \lambda\left[U_{n-1}(t-T) - U_n(t-T)\right] \tag{31}$$

The driver-sensitivity coefficient λ is to account for the driver's awareness of and intensity of reaction to stimuli from the vehicle in front. As the vehicles get closer to one another, drivers pay more attention and hence their awareness of, and sensitivity to, the actions of the other vehicles increases. In other words, it is unlikely that λ would be constant, and in fact, experiments have shown that it is variant. Herman indicates that a good approximation to the sensitivity is a constant (λ_0) divided by the front-to-front distance between the vehicles. Thus Equation 31 becomes

$$a_n(t) = \frac{\lambda_0}{x_{n-1} - x_n}\left[U_{n-1}(t-T) - U_n(t-T)\right] \tag{32}$$

Again, T is the time lag or reaction–response time between a driver perceiving a change in the speed of the vehicle ahead and actually changing his own speed.

Equation 32 is a complex second-order differential equation, the solution of which involves the Laplace transform. By integrating Equation 32, an expression relating speed and spacing or speed and density, can be obtained. Gazis et al. showed that the result is

$$U_s = \lambda_0 \log_e \frac{D_j}{D} \tag{33}$$

where

D_j = jam density and
D = density.

Comparison of Equation 32 with Equation 33, which was derived earlier by using the fluid-flow analogy proposed by Greenberg, will show that the same form for the fundamental diagram of traffic flow has been derived from two independent approaches. However, as discussed earlier, this particular formulation does not satisfy the boundary condition of mean free speed at zero volume and density. In order to improve the car-following model in this respect, Edie has suggested a modification that expresses the driver sensitivity as a function of both vehicle speed and vehicle spacing. He assumes that the faster a driver is traveling, the greater will be his awareness of the behavior of other vehicles. The modified form of Equation 32 yields

$$a_n(t) = \frac{\lambda_2 U_n(t-T)}{(x_{n-1} - x_n)^2}\left[U_{n-1}(t-T) - U_n(t-T)\right] \tag{34}$$

where λ_2 is a constant. Integration of Equation 34 yields an expression relating speed and density as follows:

$$D = D_j \log_e \frac{U_f}{U_s} \tag{35}$$

where D_j is the jam density and is equal to the $1/\lambda_2$. It should be noted that Equation 35 has the following properties:

$$\text{When } D = 0\text{: } U_s = U_f$$
$$\text{When } U_s \rightarrow 0\text{: } D \rightarrow \infty$$

Therefore, the modified car-following equation satisfies one boundary condition at the expense of another. Edie, therefore terms equations having the form of Equation 33 as congested models, and equations having the form of Equation 35 as uncongested models. He further suggests that these models provide the basis for the development of more complex models.

PROBABILISTIC DESCRIPTION OF TRAFFIC FLOW

The theoretical techniques described previously have in some respects dealt with opposite ends of the traffic flow behavior spectrum. The car-following models described and kept track of individual car movements, while the fluid models described the movement of the total vehicle population. Specifically, these two models can be regarded as deterministic flow behavior models, one at the micro-level and the other at the macro-level.

In many traffic engineering problems, however, an intermediate description of traffic, which preserves some of its discrete features and deals with probabilistic aspects of flow behavior, can be particularly useful. Typical applications are studies of intersection control, left-turn storage areas, and other cases where engineer would like to estimate the traffic delay or queuing resulting from various control measures. A reasonably adequate intermediate description of traffic can be developed from the use of probability theory, statistics, and queuing theory.

As an initial example, consider a stretch of road free of traffic-control devices and on which the volume of vehicles is very light. An observer at one point on the road might note that over a period of time vehicles pass without any regularity being apparent. Therefore flow may be described as random. That is, the number of vehicles arriving in any interval of time is independent of the number of vehicles that arrived during any previous time interval. Earlier it was noted that this assumption is made for the Possion's distribution (as well as for gamma and exponential distributions). For Poisson-distributed arrival (n), the probability of exactly n vehicles arriving in any t-s interval, is given by

$$p(n) = \frac{\mu^n e^{-\mu}}{n!} \quad \text{for } n = 0, 1, 2, \ldots, \infty \tag{36}$$

where

$n!$ = n (n–1) (n–2) … (2) (1),
λ = V/T = mean rate of arrivals per unit time interval,
μ = λt = average number of vehicles arriving in time t, and
V = total volume of vehicles arriving during time T.

The probability of no vehicles arriving in the interval t (that is, $n = 0$) is

$$p(0) = \frac{\mu^0 e^{-\mu}}{0!} = e^{-\mu} = e^{-\lambda t} \text{ for } t \geq 0$$

If no vehicles arrive in time t, then there must have been a gap or time headway of at least t s. Thus the probability of a headway h being equal to or greater than t is

$$P(h \geq t) = e^{-\lambda t}, \text{ for } t \geq 0 \quad (37)$$

and the probability of a headway being less than t is

$$P(h < t) = 1 - e^{-\lambda t}, \text{ for } t \geq 0$$

Equation 37 is a continuous function because t can assume all values from zero to infinity. (This cumulative probability is termed the exponential distribution.) Since any volume of vehicles V will also have V gaps, the frequency of occurrence of a gap of size h in time T is given by

$$\begin{aligned} \text{Freq}(h \geq t) &= Ve^{-\lambda t} \quad \text{for } t \geq 0 \\ \text{Freq}(h < t) &= V(1 - e^{-\lambda t}) \quad \text{for } t \geq 0 \end{aligned} \quad (38)$$

If, for example 705 vehicles pass in a period of 2,691 s, and if random arrivals are assumed, the probability and the number of various size gaps may be computed as follows:

$$\lambda = 705/2{,}691 = 0.262$$
$$P(h \geq t) = e^{-0.262t}$$
$$\text{Freq}(h \geq t) = 705\, e^{-0.262t}$$

Because the headway or gap distribution equation, Equation 37, is continuous, the probability of headway of exactly t s is zero. However, for an interval, such as the interval from t_1 to t_2, the probability would be

$$P(t_2 > h \geq t_1) = e^{-\lambda t_1} - e^{-\lambda t_2} \quad \text{for } t \geq 0 \quad (39)$$

Thus, for an example with t_1 and t_2 equal to 2 and 3, respectively,

$$P(3 > h \geq 2) = 0.5921 - 0.4557 = 0.1364$$

The analysis so far has been based on the assumption that arrivals for light traffic are randomly distributed over time. If this is the case, the Possion's distribution can be used to predict vehicle arrivals and, in turn, the exponential distribution can be used to describe the headways. In order to test the hypothesis about random nature of light traffic, it is necessary to collect field data, to estimate the parameter value, and to analyze the empirical and theoretical data. Generally, various numerical or empirical studies have been conducted, which shows that for light traffic conditions, the Possion distribution can be used to approximately describe the flow. However, as the flow rate increases, Equation 37 becomes unacceptable, at least for the situations where statistical acceptance of order indicated is required. This conclusion only serves to confirm the

previous car-following results that relationships do exist between vehicles, particulary as the traffic density increases.

Haight and others have suggested use of the gamma or Erlang probability function. This probability function may be written as

$$f(t) = \frac{\lambda}{(k-1)!}(\lambda t)^{k-1} e^{-\lambda t} \qquad \text{for} \quad t \geq 0 \qquad (40)$$

where k and λ are parameters of probability function and k is a positive integer. For Possion-distributed events or arrivals, the gamma probability function may be used to represent the waiting time t until the kth arrival or event. That is, $f(t)$ is equal to the probability of waiting time t until the kth arrival. The special case of the gamma probability function for k being to 1.0 is called the exponential probability function; it is:

$$f(t) = \lambda e^{-\lambda t} \qquad \text{for } t \geq 0 \qquad (41)$$

where $f(t)$ is equal to the probability of waiting time t until the first arrival or event.

USE OF CELLULAR AUTOMATA

Dhingra et al. (2004, 2005) contributed research work on traffic-flow modeling concentrated on various areas like midblock, intersections, and ramps, and many of the works were on homogeneous traffic conditions. In developing nations such as India, the traffic scenario is even more chaotic, as the element of heterogeneity is brought in by varying vehicular, human, road and environmental factors. Simulation has been established as a very vital tool for the traffic analyst, in getting to know more about the traffic flow and to verify the performance of control systems. Cellular automata, a new entrant to this area, introduced a concept of minimal modeling, against the complex modeling procedures being in use until then. It has proven to be computationally efficient in analyzing midblock traffic for homogeneous and heterogeneous conditions. The model is used to study the behavior of heterogeneous traffic such as lane changing, incident occurrence, and the effect of different composition of traffic. The study of cellular automata gives a new computationally efficient model for heterogeneous traffic. However, these models can handle only unidirectional traffic at midblock. Therefore, further study is required to explore the vehicular behavior in bidirectional traffic at midblock as well as intersection.

CONCLUSION

This paper essentially gives the gist of various theoretical approaches of traffic flow that describe the way the vehicles move over parts of the highway system. The various empirical studies relating to volume, density, and speed involving approaches such as Greenshields,' Greenberg's, and Guerin's are presented. Macroscopic models such as hydrodynamic analogies including contributions from Lighthill and Whitham and Richards are highlighted. Car-following models consisting of treatment of traffic flow at microscale, involving intervehicle relationship

characteristics, and details of individual vehicles are shown here. The microscale treatment of traffic flow is inspired by General Motors approach which is generally associated with the work of researchers Herman et al. Apart from above deterministic approaches, an intermediate description of traffic, preserving some of its discrete features and dealing with probabilistic aspects of traffic behavior is explained at length.

RESOURCES

Dhingra, S. L., P. J. Gundaliya, and T. V. Mathew. Cellular Automata: An approach for Traffic Flow Modelling. Accepted in START 2005, at IIT Kharagpur.

Dhingra S. L., P. J. Gundaliya, and T. V. Matthew. Heterogeneous Traffic Flow Modelling Using Cellular Automata. 10th WCTR 2004, Istanbul.

Edie, L. C. Car Following and Steady-State Theory for Non-Congested Traffic. Tunnel Traffic Capacity Study Report VI, Port of New York Authority, New York, May, 1960.

Gazis, D. C (ed.). *Traffic Science*. John Wiley, New York, 1974.

Geetam, T. *Traffic Flow and Safety: Need for New Models for Heterogeneous Traffic.*

Gerlough, D. L., and D. G. Capelle (eds.). Definitions and Notations. In *Special Report 79: An Introduction to Traffic Flow Theory*, HRB, National Research Council, Washington, D.C, 1964, pp. vii–viii.

Greenberg, H. A Mathematical Analysis of Traffic Flow. Tunnel Traffic Capacity Study, Port of New York Authority, New York, 1958.

Greenberg, H. An Analysis of Traffic Flow. In *Operations Research*, Vol.7, No.1, 1959.

Greenshields, B. D., and F. M. Weids. *Statistics with Applications to Highway Traffic Analyses*. Eno Foundation for Highway Traffic Control, Saugatuck, Conn., 1952.

Guerin, N. S. Travel Time Relationships. In *Quality and Theory of Traffic Flow*, Bureau of Highway Traffic, Yale University, New Haven, Conn., 1961.

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

Herman, R., E. Montroll, R. Potts, and R. Rothery. Traffic Dynamics: Analysis of Stability in Car Following. In *Operations Research*, Vol. 7, 1959.

Highway Capacity Manual, Bureau of Public Roads, U.S. Department of Commerce, 1950.

Lighthill, M. J., and G. B. Whitham. On Kinematic Waves; II. A Theory of Traffic Flow and Long Crowded Roads. *Proc., Royal Society.* (London), Series A, Vol. 229, No. 1178, 1955, pp. 317–345.

May, A. D., *Traffic Flow Fundaments*, Prentice Hall, 1990.

Normann, O. K. Results of Highway Capacity Studies. In *Public Roads*, Vol. 23, No. 4, 1942.

Richards, P. I. Shock Waves on the Highway. In *Operations Research*, Vol. 4, No.1, 1956, pp. 42–51.

Traffic Flow Prospectives
From Fundamental Diagram to Energy Balance

CHRISTOF LIEBE
REINHARD MAHNKE
Rostock University, Rostock, Germany

REINHART KÜHNE
German Aerospace Center, Berlin

HAIZHONG WANG
University of Massachusetts Amherst

Congestion modeling is a topic considered in traffic flow theory since Greenshields' study of traffic capacity in 1934. Several levels of investigation deal with different categories: microscopic car-following dynamics, mesoscopic stochastic theories of car cluster formation (*1*,*2*) [for queuing description, see Wang et al. (*3*)], as well as macroscopic fluid-dynamical models study the characteristics of traffic states, given by speed, flow, and density, shown in the fundamental diagram. By this relationship, speed and flow can be represented as a function of density, usually as a steady-state dependence but also as function of time because traffic states change continuously. Distinguishing a collection of similar traffic states as phases, the phenomenon of phase transition from (nearly) free-flow to stop-and-go traffic is the emergence of a random fluctuation (e. g., of speed) developing into a vehicular congestion known as breakdown (*4*).

Usually a traffic breakdown is defined as a certain amount of speed drop in a dense traffic situation. It always has a reason, but the reason is hard to analyze. To describe these dynamics successfully a probabilistic model is chosen where the unpredictable influences are summarized by a stochastic force creating vehicular platoons (called vehicular clusters) out of the metastable free flow. This spontaneous formation of a new traffic phase is modeled as a stochastic cluster emergence process (*1*, *2*). A car cluster is a vehicular platoon larger than a predefined critical number of participating (or bounded) vehicles. Using this definition, the speed drop is translated into an overshot of the threshold given by the critical cluster size.

TRAFFIC INVESTIGATIONS

Figure 1 shows one particular time frame from a video observation done in 2003 in the NGSIM program (*N*ext *G*eneration *Sim*ulation program, http://ngsim.fhwa.dot.gov/). Cars were traced and their trajectories recorded in a data set. Data such as these can be used to create microscopic

FIGURE 1 Original pictures of traffic observation in the NGSIM project. Red rectangles belong to tracing software (*5*).

traffic models. To study the basic behavior of a certain model, it is necessary to simplify the boundary conditions. The first simplification is the one-lane model, which forbids overtaking maneuvers. The second simplification is periodic boundary conditions. Figure 2 shows the simplified model situation.

Figures 3 and 4 are generated from GA400 ITS data in the whole year of 2003 from empirical observations at stations numbered 4001115 and 4005103 on the inflow direction to the Atlanta city in Georgia, United States. Station 4001115 is located on the basic highway segment with three lanes, while station 4005103 is set up on a one-lane on-ramp to GA400 south (inflow direction). The original data for each station is 20 s aggregated, when used to generate the fundamental diagrams, the original data were aggregated to 5 min. Because only time mean speed is available, it is used to calculate the density instead of space mean speed. From the two fundamental diagrams, the pattern on basic highway segment and on- off-ramp looks similar but has distinguishable features.

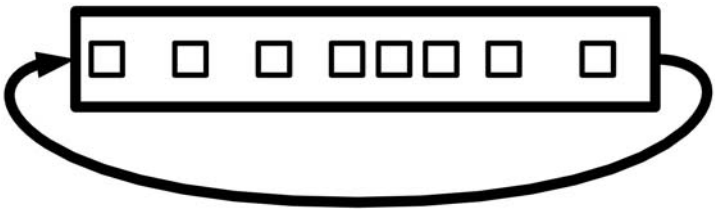


FIGURE 2 Simplified street for theoretical traffic flow analysis (only one lane with periodical boundary conditions).

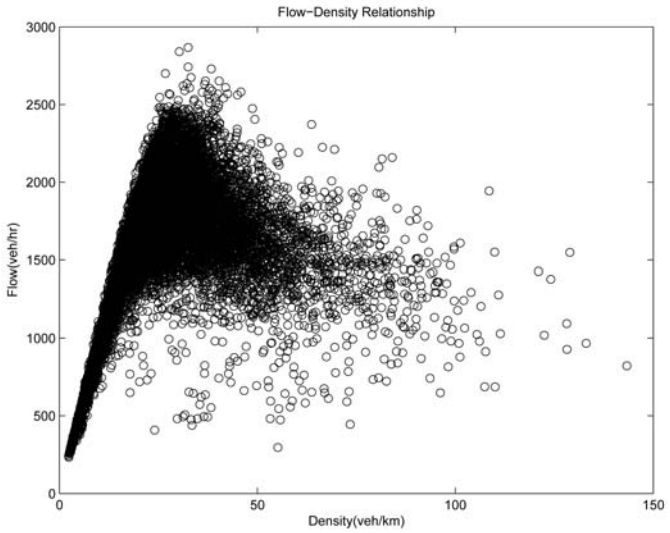


FIGURE 3 Flow–density relationship plotted from 1-year observations at Station 4001115.

FORCES

A physical picture of traffic can be achieved by defining forces between two cars and one car with the road. Figure 5 shows three kinds of forces in the one-lane model.

When a driver is alone on the road, he will accelerate until he reaches a certain maximum velocity. This can be translated into an accelerating force F_D driving the car exponentially to this maximum velocity. Most of the time a car is not alone on the street. So there must be an interaction between cars.

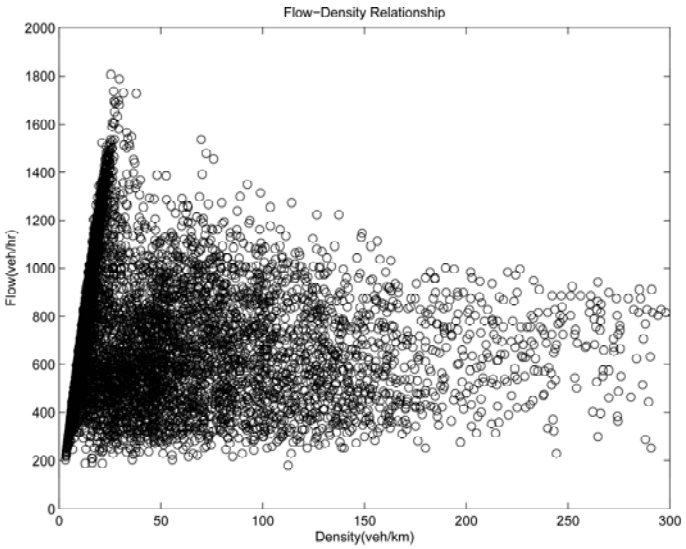


FIGURE 4 Flow–density relationship plotted from 1-year observations at Station 4005103 (on–ramp).

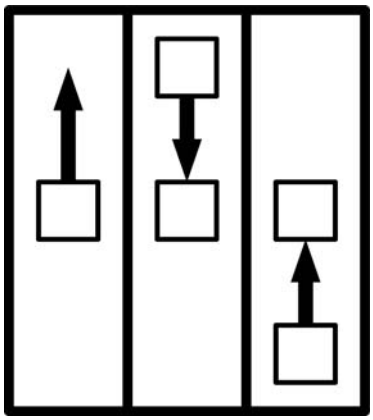


FIGURE 5 Different forces in traffic model. Left: An empty road leads to acceleration until the maximum velocity is reached. Middle: A car in front leads to a repelling force and as a consequence to breaking. Right: A car behind a driver could also lead to some forces.

This interaction can be described as a repelling force, F_F, depending on the distance to the leading car. A similar force, F_B, can describe the interaction to the car behind

$$F_D(v) = \frac{1}{\tau_D}(v - v_{\max}) \tag{1}$$

$$F_F(\Delta x_F) = \frac{1}{\tau_F}\left[v_{\max} - v_{\text{opt}}(\Delta x_F)\right] \tag{2}$$

$$F_B(\Delta x_B) = -\frac{1}{\tau_B}\left[v_{\max} - v_{\text{opt}}(\Delta x_B)\right] \tag{3}$$

Putting all interactions together determines a force $F(\Delta x_B; \Delta x_F; v)$, acting on a certain car

$$\begin{aligned} F(\Delta x_B, \Delta x_F, v) = &\frac{1}{\tau_F}\left[v_{\max} - v_{\text{opt}}(\Delta x_F)\right] \\ &- \frac{1}{\tau_B}\left[v_{\max} - v_{\text{opt}}(\Delta x_B)\right] \\ &+ \frac{1}{\tau_D}(v - v_{\max}) \end{aligned} \tag{4}$$

If point-like cars and next-neighbor interactions are considered as described, the headways Δx_F and Δx_B are the distances of a certain car (i) to the car next to it in driving direction ($i + 1$) and to the directly following car ($i - 1$)

$$\Delta x_F = x_{i+1} - x_i \tag{5}$$

$$\Delta x_B = x_i - x_{i-1} \tag{6}$$

Taking into account that the interaction with the following car is slightly weak, it can be neglected. Further on, the relationship can be made that $\tau_D = \tau_F = \tau$ if F_F is considered to be in the

same range as F_D. This leads to a final equation for the force used in the so-called optimal velocity model

$$F(\Delta x, v) = \frac{1}{\tau}\left[v_{\max} - v_{\text{opt}}(\Delta x)\right] + \frac{1}{\tau}(v - v_{\max}) \tag{7}$$

For the optimal velocity v_{opt} a function is needed which becomes zero at small distances and the maximum velocity at big distances. The function, proposed in Mahnke et al (*2*),

$$v_{\text{opt}}(\Delta x) = v_{\max}\frac{(\Delta x)^2}{D^2 + (\Delta x)^2} \tag{8}$$

fulfills these requirements.

Fundamental Diagram

If a homogeneous long-time regime with a density of $\rho_{\text{hom}} = N/L$ is considered, the velocity of all cars will be $v_{\text{opt}}(N/L)$. This directly leads to the fundamental diagram (see Figure 6 for dimensionless plot) flux over density ($j = \rho v$ over ρ). The nonlinear function represents the homogeneous situation. The linear functions represent heterogeneous situations after a breakdown.

Based on the NGSIM data, a fundamental diagram can be drawn (see Figure 7). If the dimensionless theoretical plot (Figure 6) is compared with the one from the data, a mismatch in small density region can be seen. The data indicate a linear relationship between the density and the flux. The model is in this region not linear. This is because of the density dependence of the optimal velocity in the small density region. The congested phase is represented quite well. The linear behavior found in the data also exists in the model after the breakdown.

The data diagram is widely scattered, which is a hint to the probabilistic character of traffic flow. But also the measuring process leads to some scattering because the number of cars is always a finite number and the length of the road segment is constant. When a car leaves the

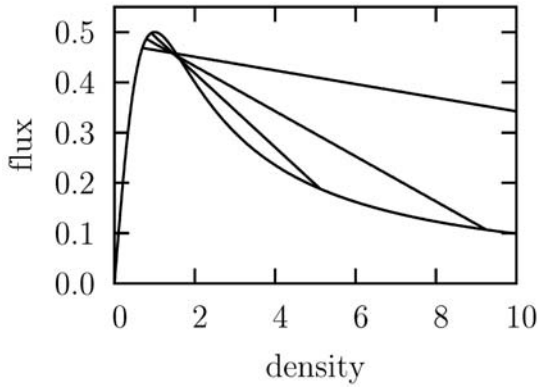


FIGURE 6 Fundamental diagrams for parameter values $b = 0.9$, $b = 1.0$ and $b = 1.1$. (the function with the smallest negative envelope belongs to $b = 0.9$, and further reduction of $b = \tau/(v_{\max}D)$ would lead to accidents).

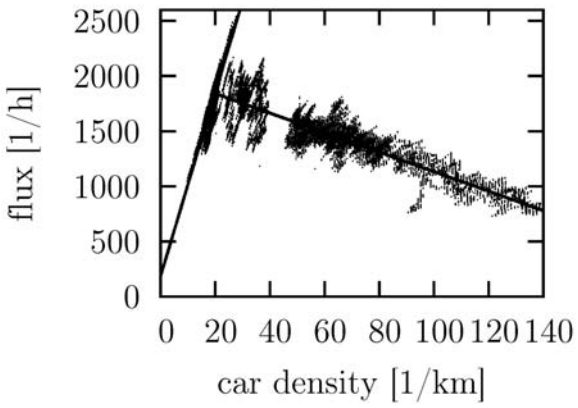


FIGURE 7 Fundamental diagram of real traffic data (linear fits of the two traffic regions are also shown).

observed road, the number of cars is decreased by one and the effective length of the road increases by the length of the car. Both lead to a decrease of the density. The model could be extended by an additive noise term, $\xi(t)$, to take the probabilistic behavior into account called Langevin equation

$$m\frac{dv}{dt} = F_{\text{det}} + \xi(t) \tag{9}$$

Nevertheless it is hard to distinguish between the systematic noise due to the measuring process and the real noise in a traffic situation. The following description presents the noise-free model.

Energy and Energy Flux

The definition of forces allows the definition of energy and energy flux. In physics, the kinetic energy is defined as

$$E_{\text{kin}} = \frac{m}{2}v^2 \tag{10}$$

The conservative force defines the potential energy, because the conservative force is the negative derivative of the potential energy with respect to the position

$$F_{\text{cons}} = -\frac{dE_{\text{pot}}}{dx} \tag{11}$$

$$E_{\text{pot}} = m\frac{v_{\max}D}{\tau}\left[\frac{\pi}{2} - \arctan(\Delta x)\right] \tag{12}$$

Because the energy of the system is now well defined, it is possible to investigate the energy flux Φ as the negative time derivative of the energy

$$-\Phi = \frac{d}{dt}\left(E_{\text{kin}} + E_{\text{pot}}\right) \tag{13}$$

$$-\Phi = F_{\text{D}}(v)v + F_{\text{F}}(\Delta x)v_{\text{F}} \tag{14}$$

The energy flux depends on the forces and the velocity on the own and the leading car. It can be split into two parts, an inflow and an outflow part

$$\Phi_{\text{in}} = -F_{\text{D}}(v)v \le 0 \tag{15}$$

$$\Phi_{\text{out}} = -F_{\text{F}}(\Delta x)v_{\text{F}} \ge 0 \tag{16}$$

Because there are always forces and velocities in the system, the two parts of the energy flow never become zero, although the sum is zero in the long-time regime. Only in the long-time regime is the energy in the system conserved.

Phase Transition

As was noted before, the homogeneous situation is not the only long-time solution of the system. Depending on the density, the homogeneous situation becomes unstable and a small perturbation leads to a phase separation. This is shown in Figure 8 at one specific density (*6*). The system starts in a (nearly) homogeneous situation without kinetic energy. After a short time of homogeneous flow, the energy of the system changes stepwise. A closer analysis shows that the jumps in the energy appear when regions of high density merge. If the cars are considered as particles, this would be called cluster formation and merging. This is interesting because there is not only one growing cluster; there are several clusters early in the system, but in the long-time result there is only one cluster left. When this last cluster is formed, the energy of the system is stable in time.

Figure 9 shows the long-time result for the whole interesting density range. Four densities marked with dotted lines have a special significance. The two outer densities separate the regions where a cluster formation is possible (between the densities) or not (left and right from the densities). The two inner densities separate the region where a homogeneous long-time result is possible (left and right of the densities) and where only the cluster solution is stable (between them).

A dimensionless description leads to only one parameter, $b = \tau/(v_{\max}D)$, which defines the behavior of the system depending on the density. Figure 10 shows the different phases of the system. That means it shows which stable long-time solution is possible at a certain density and a certain combination of the parameters D, τ, and $v_{\max}$.

CONCLUSION

Once this concept is established the authors ask what external principle lies behind the cluster formation as phase transition from free-flow to congested traffic. The vehicular flow as an open nonequilibrium system has energy sources such as the carried gasoline and energy sinks such as road friction and air resistance. The paper presents investigations about the energy flow analysis and its relationship to phase transitions. Based on the stochastic model about dynamics of jam formation (*2*) similar to nucleation (see schematic, Figure 2 with two contributions only, called attachment and detachment rate), the paper addresses the connection of these rates of jam

formation to the energy concept, which finally leads to an energy change during traffic breakdown.

Microscopic traffic models based on follow-the-leader behavior are strongly asymmetrically interacting many-particle systems. These well-known vehicular models include the fact that (1) the driver is mainly looking forward interacting with the lead vehicle and (2) the car travels on the road always with friction. Due to these realistic assumptions the moving car needs petrol for the engine to compensate dissipation by rolling friction (compare schematic Figure 11). Here we investigate the flux of mechanical energy to evaluate the energy balance out of the given nonlinear dynamical system of vehicular particles.

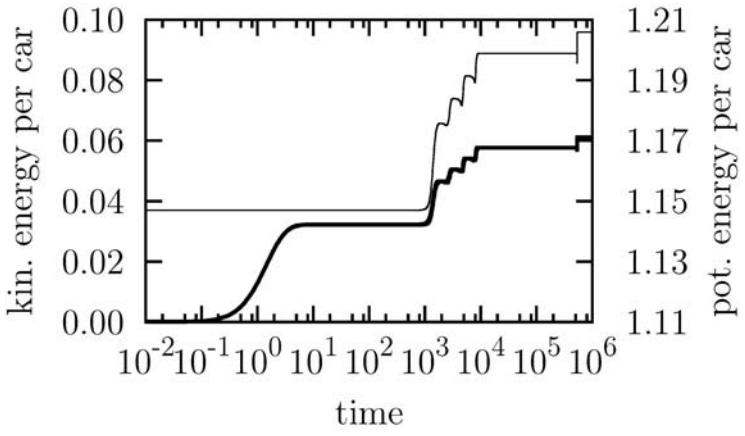


FIGURE 8 Dense traffic simulation with 60 cars. In the first instance several small jams appear, which join until only one jam is left in the final regime. The merging of jams leads to an increase of the energy. The thin line is the potential energy. The thick line is the kinetic energy.

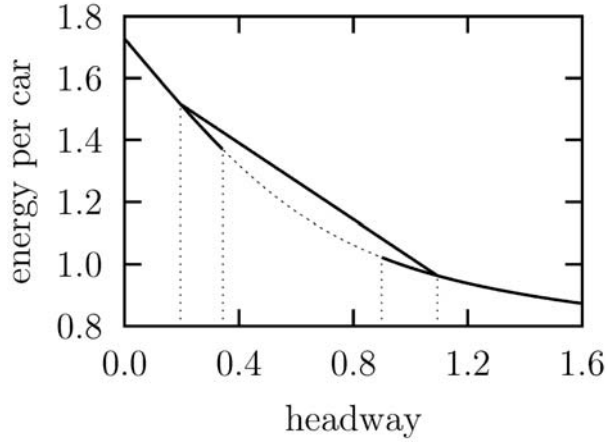


FIGURE 9 Mean energy per car with $b = 1.1$. The vertical dotted lines indicate the critical headways which separate the different phases. The thermodynamic limit is shown.

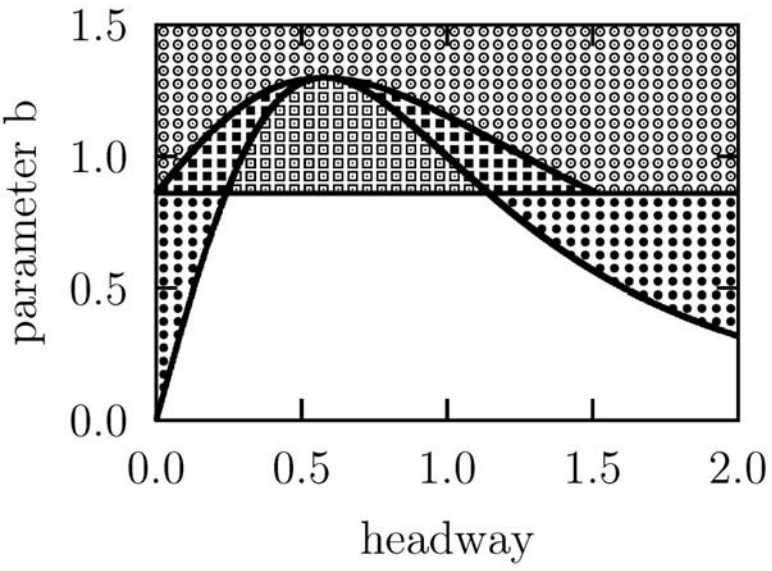


FIGURE 10 Different phases of the optimal velocity model. Empty circles: Only homogeneous situation is possible. Empty squares: Only heterogeneous situation is possible. Filled squares: Depending on the initial condition the homogeneous or the heterogeneous situation is stable. Filled circles: Homogeneous situation is stable, heterogeneous situation immediately leads to accidents. Empty area: Homogeneous situation is unstable, heterogeneous situation immediately leads to accidents.

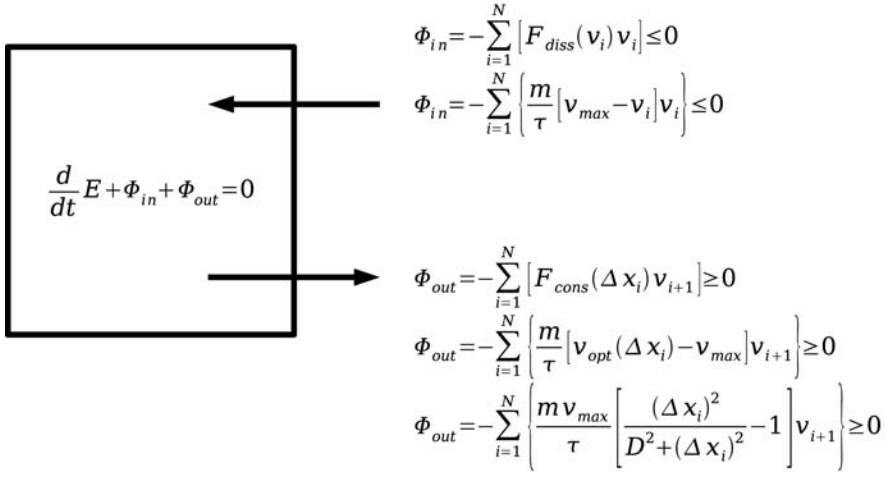


FIGURE 11 The energy flux in the system.

Traffic flow is a dissipative system of driven or active particles (*6*). It means that the total energy $E = E_{kin} + E_{pot}$ is not conserved, but we have an energy balance equation

$$\frac{dE}{dt} + \Phi = 0 \tag{17}$$

with the energy flux Φ following from the equations of motion and consisting of dissipation (due to friction) and energy input (due to burning of petrol). In the long-time limit, the many-car system tends to certain steady state. In the microscopic description it is either the fixed-point (free flow) or the limit cycle (stop-and-go traffic) in the phase space of velocities and headways depending on the overall car density and control parameters. The steady state is characterized by a certain energy value.

In order to understand the traffic breakdown as transition from free-flow to congested traffic, we estimate the total energy per car at low and high densities and observe the energy of jam formation.

REFERENCES

1. Mahnke, R., J. Kaupužs, and I. Lubashevsky. *Physics of Stochastic Processes: How Randomness Acts in Time*, Wiley-VCH, Weinheim, 2009.
2. Mahnke, R., J. Kaupužs, and I. Lubashevsky. Probabilistic Description of Traffic Flow. In *Physics Reports*, Vol. 408, 2005, pp. 1–130.
3. Wang, H., D. Ni, J. Li, K. Rudy, and R. Mahnke. A Markovian Approach to Predict Traffic Breakdown Probability and Link Flow Optimization. Submitted to *International Journal of Simulation and Process Modeling*, 2008.
4. Kühne, R., R. Mahnke, and J. Hinkel. Understanding Traffic Breakdown: A Stochastic Approach. *Transportation and Traffic Theory 2007* (R. E. Allsop, M. G. H. Bell, and B. G. Heydecker, eds.), Elsevier, Oxford, 2007, pp. 777–790.
5. Liebe ,C. *Stochastik der Verkehrsdynamik: Von Zeitreihen-Analysen zu Verkehrsmodellen*, diploma Rostock University, 2006. . http://rosdok.uni-rostock.de/ resolve?id=rosdok_thesis_000000000003
6. Mahnke, R., J. Kaupužs, J. Hinkel, and H. Weber. Applications of Thermodynamics to Driven Systems. In *European Physical Journal B*, Vol. 57, 2007, pp. 463–471.

A Stochastic Macroscopic Modeling Framework to Interpret the Fundamental Diagram

SERGE P. HOOGENDOORN
HANS VAN LINT
VICTOR KNOOP

Modeling breakdown probabilities or phase transition probabilities is an important issue when assessing and predicting the reliability of traffic flow operations. Looking at empirical spatiotemporal patterns, these probabilities clearly are not only a function of the local prevailing traffic conditions (density, speed), but also of time and space. For instance, the probability that a start–stop wave occurs generally increases when moving upstream away from the bottleneck location.

This contribution proposes a unifying stochastic modeling framework that allows us to model the dynamics of the breakdown or phase-transition probabilities in an intuitive manner using the kinematic wave model of Lighthill and Whitham (*1*) as a basis. Different researchers have considered the dynamic modeling of breakdown probabilities (*2*–6) commonly using (stochastic) queuing analysis and nucleation models; we refer to Kerner for a thorough discussion of these modeling approaches. In doing so, we propose using coupled set of partial differential equations describing the traffic dynamics and the dynamics of the phase-transition probabilities p(t,x). We argue that the proposed modeling framework can be considered as a straightforward generalization of the kinematic wave theory to three-phase theory.

The main result is that we can reproduce the main characteristics of the breakdown probabilities and related traffic flow operations. This is illustrated by means of two examples: free flow to synchronized flow (F-S transition) and synchronized to jam (S-J transition). We show that the probability of an F-S transition increases away from the on-ramp in the direction of the flow; the probability of an S-J transition increases as we move upstream in the synchronized flow area.

REFERENCES

1. Lighthill, M. J., and G. B. Whitham. On Kinematic Waves; II. A Theory of Traffic Flow and Long Crowded Roads. *Proc., Royal Society.* (London), Series A, Vol. 229, No. 1178, 1955, pp. 317–345.
2. Heidemann, D. Mathematical Analysis of Non-Stationary Queues and Waiting Times in Traffic Flow with Particular Consideration of the Coordinate Transformation Technique. *Transportation and Traffic Theory in the 21st Century* (M. P. Taylor, ed.), Pergamon, 2002, pp. 675–696.
3. Brilon, W., J. Geistefeldt, M. Regler. Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity. Flow, Dynamics and Human Interaction. *Proc., 16th International Symposium on Transportation and Traffic Theory* (H. Mahmassani, ed.).
4. Kühne, R., and R. Mahnke. Controlling Traffic Breakdowns. Flow, Dynamics and Human Interaction. *Proc., 16th International Symposium on Transportation and Traffic Theory* (H. Mahmassani, ed.).
5. Kerner, B. S. Probabilistic Breakdown Phenomenon at On-Ramp Bottlenecks in Three-Phase Traffic Theory. *In Transportation Research Record: Journal of the Transportation Research Board, No. 1965*, TRB, National Research Council, Washington, D.C., 2006, pp. 70–78.
6. Kerner, B. S. *The Physics of Traffic: Empirical Freeway Pattern Features, Engineering Applications, and Theory*. Berlin, Springer, 2004.

Analysis of LWR Model with Fundamental Diagram Subject to Uncertainty

JIA LI
QIAN-YONG CHEN
DAIHENG NI
HAIZHONG WANG
University of Massachusetts at Amherst

The LWR model is of interest because it can successfully reproduce some essential features of traffic flow, such as the formation and propagation of various waves. Recently, however, the disadvantageous aspects of the traditional LWR model are attracting more and more attention. Many researchers have made great effort to develop new models based on the LWR model, such that some prominent nonlinear characteristics of traffic flow, such as platoon diffusion, capacity drop, hysteresis, and spontaneous onset of congestion are captured. In this paper, the authors investigate the LWR model from an uncertainty perspective. Rather than developing a new explanatory model, they attempt to analyze how reliable the LWR model prediction will be if the fundamental diagram (FD) used is not specified accurately. Their work reflects the current debate on the nature of the FD. To be specific, the authors postulate a form of the flux function driven by random free-flow speed, which accommodates the scattering feature observed in the speed–density plot. The authors show the properties of the LWR model with the new flux. A third-order essentially nonoscillatory finite difference algorithm is devised to solve the model. An approach to evaluate the predictability of traffic disturbance propagation with this model is presented. The authors interpret the results and conclude that if the FD cannot be undoubtedly specified, the LWR model will deteriorate and only make a reasonable prediction in relatively short time scale.

INTRODUCTION

Background

The LWR model (*1*, *2*), as the simplest kinetic wave (KW) model, is central in many investigations associated with traffic dynamics and control. This is probably because some essential features of traffic flow, such as wave formation and propagation, can be qualitatively well reproduced with the LWR model. The reader is referred to works by Daganzo and Zhang (*3*, *4*), and the references therein for full discussion of this model and its variations.

At the same time, the deficiencies of the LWR model are well known. For example, the LWR model fails to generate capacity drop, hysteresis, relaxation, platoon diffusion, or spontaneous congestion. The existence of these traits has been confirmed in many field observations. In the LWR model, mass (i.e., the vehicles) conservation always holds true provided that no sinks and sources are involved. Thus to remedy the problem, one may either follow the rationale of hydrodynamics and add momentum conservation type equations, or question the initial assumptions regarding the fundamental diagram (FD), i.e., the smoothness and concavity, and propose other forms of models. These two strategies are, in spirit, consistent

with the discussions by Zhang (*4*), where they are named "higher-order" and "lower-order" extensions of LWR model, respectively.

Of the two strategies, the former seems much more debatable. Daganzo (*3*) describes the logical flaws in the argument of deriving high-order continuum models, and he shows that negative flow and speed could be unreasonably generated (this means anisotropy property will be violated). Inconsistency inherent in such model in a statistical perspective is also demonstrated by Heidemann (*5*). Zhang (*4*) holds that the two strategies are very close in nature, regarding the anisotropy issue as well as necessity to introduce behavioral laws, and he finds that the models investigated therein can all be reduced to a KW model endowed with an appropriate "effective fundamental diagram." Moreover, the plots of speed-concentration or flow-concentration usually exhibit structured randomness/scatterings, especially around congestion region. These observations, though not necessarily conflicting the initial assumption of the FD, are quite common and make thorough analysis possible. Therefore, it seems to be at least a good starting point to follow the latter strategy and see what may happen.

Different interpretations of the scattering have lead to many new models, either explaining this effect itself or incorporating this effect to obtain traffic flow models with new features. On one hand, Castillo and Benitez (*6*) and Cassidy (*7*) demonstrate that well-defined and reproducible relations of traffic variables, such as flow-occupancy, can be obtained only if the stationary traffic data are used. This implies the scattering is due to the existence of transient states. Zhang (*4*) shows, at the price of lowering data resolution, a well-behaved FD can also be constructed from locally smoothed transient traffic data. On the other hand, the hysteresis phenomenon has been well known (*8*, *9*). This indicates speed-concentration curve should consist of hysteresis loops rather than a single curve. Meanwhile, it is noted that Newell's extension of the FD (*10*) and the reversed-λ shaped FD have been in existence for a long time. Nonetheless, the FD with structured variability still seems to be a puzzling issue in recent years and continues triggering discussions in different dimensions. Treiber and Helbing (*11*) show that conventional measurement methods with the delay of driving behavior may result in the scattering of flow–density data in synchronized congested traffic. Zhang and Kim (*12, 13*) utilize car-following models and explain the occurrence of capacity drop and hysteresis with one variable of gap time. Wong and Wong (*14*) and Ngoduy and Liu (*15*) attribute nonlinear traffic phenomena such as hysteresis, capacity drop, and dispersion of traffic platoon to the distribution of heterogeneous drivers and they formulate this idea into multiclass continuum traffic models. Kerner et al. (*16–18*) criticize the FD approach, arguing that this approach fails to produce the spatial-temporal features of congestion correctly. The concept of synchronized flow is introduced and the three-phase traffic flow theory is developed. This list of works is not intended to be complete in any sense. One fact is evident: the concept of the FD, though seemingly basic, is hardly unequivocal since its birth. Many interesting findings regarding traffic flow are initiated with the fresh insight into the FD.

The aim of this paper, rather than resolve those existing controversies with FDs, is to investigate the potential influences brought by such ambiguity. To fulfill this end, we analyze the possible sources of randomness with the FD first, and then following the methodology of uncertainty analysis, develop a numerical procedure to evaluate how the randomness affects the faithfulness of the LWR model. Though not attempting to be explanatory, our study essentially unveils the robustness of LWR model to the imperfect specification of the FDs. Knowledge of this hopefully prompts further understanding of traffic flow features.

Paper Organization

The remainder of this paper is organized as follows. First, a brief account for uncertainty analysis is presented and the source of randomness associated with the speed–density plot is differentiated. In this context, the authors discuss the extension of the FD from a single curve to random function, which accommodates a distribution of flow for any given density value. Then a specific extension is proposed that assumes that the specification of free flow is subject to uncertainty. Reasons and implication of introducing such assumption are detailed in a perspective of interpolation. Based on the proposed random FD, the authors conduct a numerical study. The model to be solved is the LWR with a random FD, and an essentially nonoscillatory (ENO) finite difference scheme is devised to fulfill the purpose. Subsequently, an example is designed with local jam–vacuum initial data that attempts to find how predictable the propagations of these local disturbances are with the previously discussed uncertain setting. Finally, the authors briefly summarize the paper and show the gaps that were found.

UNCERTAINTY ANALYSIS: THE LWR MODEL IN STOCHASTIC SETTING

In system modeling literature, the uncertainty is commonly understood as the deviation from the unachievable ideal knowledge of the relevant system. Roughly speaking, the uncertainty analysis is to model the uncertainty with a system and understand the related influences. Risk and reliability assessment are among the most significant examples of uncertainty analysis, and find their applications in hydrology, structure engineering and economics, etc. The uncertainty analysis in the general context of conservation laws has been reported in literature for two decades. Many of them are in the spirit of randomizing the flux function, either spatially or temporally. For example, Holden and Risebro (*19*) and Holden (*20*) discuss the expectation and convergence of a stochastic Buckley-Leverett equation. Wehr and Xin (*21*) analyze the large-time property of Burger's equation. More recently, Bale et al. presented a general solution approach, without directly addressing the issue of randomness, for the conservation equation with flux function admitting spatial variation (*22*). All of them have focused on the spatial variability of the flux function. In the traffic field, Jou and Lo considered a nonlinear macroscopic traffic flow equation perturbed by a Brownian motion (*23*). Before unfolding our investigation, empirical evidence is presented. For any empirical speed–density relation, the mapping from density to speed is always multi-valued after rounding off the density data. This enables us to calculate the mean and standard deviation of the speed indexed by density. In Figure 1 the mean and standard deviation of the speed versus density at one station from GA-400 ITS data set (which are collected by virtual loop detectors on freeway GA-400) are shown. In the plot, one can see that the standard deviation is comparable to the mean of speed in quite wide range, indicating the existence of non-ignorable variability in speed–density relation. Moreover, the shape of the standard deviation is somewhat interesting, exhibiting a peak around 50 veh/mi. Our investigation begins from these observations. The LWR model assumes that the speed–density (v-k) relation is time independent, i.e., system equilibrium is already achieved (by equilibrium, we mean only transition between stationary traffic states is possible). The LWR model says that the density k(x,t) of traffic flow is the solution to the following equations:

$$k_t + (kv)_x = 0$$
$$v = v(k) \quad (1)$$
$$k(x,0) = k_0(x)$$

The first equation is the conservation law with the assumption of smoothness of involved functions up to a certain order, the second equation is the fundamental relation which holds under the equilibrium assumption, and the last equation provides the initial state of the solution. Flow q=q(k)=kv(k) is known as the flux function of k, which depicts the dynamics of the concerned quantity, i.e., traffic flow density. To correctly incorporate the exhibited randomness into Equation 1, we first need to analyze the possible sources of randomness. We begin by categorizing the involved randomness in two classes. The first type of randomness is the uncertainty during data collection and processing, e.g., inaccurate reading and data round-off that is reflected in the initial data setting and scatter plot of the *v*-*k* relation. In this case, the collective small additive errors would altogether obey the normal distribution law, due to the central limit theorem; alternatively, if the errors are small and multiplicative, they jointly behave by the lognormal law. This type of randomness is statistical and relatively well known. The second type of randomness is due to the inherent system dynamics. Taking the transportation system for example, the drivers' behaviors vary from one driver to another, thus the group is best described in distributional terms rather than deterministically. We claim that this randomness underlies the random *v*-*k* relation as aforementioned. An intuitive interpretation of Figure 1 would be in an analogous manner to that of Brownian Bridge. That is, taking the v as a random process indexed by *k*, at two points 0 and jam density k_{jam} the knowledge of *v* is relatively complete since the constraints are imposed by definition of the two states. This explains the smaller variances at two ends as shown. Based on the above analysis, we argue that the second type of randomness is essential, since the first type can usually be controlled reasonably well, for example, through improving the measuring techniques. Back to the problem of formulating the LWR model in a stochastic setting, we consider the second type randomness since it is dominant and inherent.

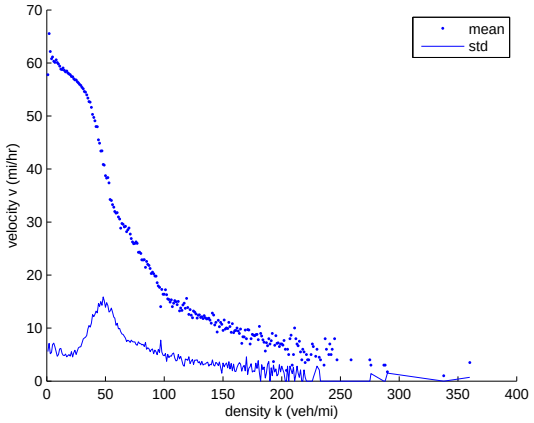


FIGURE 1 First and second order speed-density relation at one typical site based on GA400 ITS data.

Most generally, to account for this randomness, we express traffic speed as a positive valued multivariate function,

$$v = v(k, \omega(x,t)) : (R^+, M_{\Omega(x,t)}) \to R \quad (2)$$

where ω is an appropriately defined subset of sample space Ω (set of all such subsets is denoted as $M_{\Omega(x,t)}$, a σ-algebra defined on the sample space), the probability space equipped with measure $P_{x,t}$. This definition will lead to a stochastic flux function,

$$f = q = kv = kv(k, \omega(x,t)) \quad (3)$$

Substituting the Formula 3 back into Equation 3, we obtain the stochastic formulation of the LWR as,

$$K_t + (KV)_x = 0 \quad (4)$$

Equation 4 in general admits a random field *K* as solution. Before attempting to obtain useful calculation results, it is desirable to justify the validity of this equation. The minimum requirement is that with trivial probability measure it is consistent with usual deterministic equation.

FD DRIVEN BY RANDOM FREE-FLOW SPEED

To make the analysis and computation tractable, we need restrict our focus to a specific form of Equation 1 in this paper. This is equivalent to treat a specific form of Equation 3. In particular, we assume that the stochastic flux function is independent of space and time. This assumption greatly simplifies the matter since each realization of *f* would be a function of one variable *k* only. This reduces to the deterministic case we usually deal with. To address the features observed in Figure 1, we find the following form plausible. We postulate that the randomness of *f* is due to the uncertainty of the free flow speed. First, the free-flow speed is written as

$$v_f = \overline{v_f} + (sk + r)\varepsilon \quad (5)$$

where $\overline{v_f}$ is a constant, and ε represents the imperfect knowledge of actual v_f, with $(sk+r)$ being a scaling coefficient. Adopting this scaling coefficient implicitly assumes $E\varepsilon = 0$ and $Var(\varepsilon) = 1$. The reason of using a function of *k* as the scaling coefficient is as follows. If we adopt the *v*-*k* relation of,

$$\left(\frac{v(k)}{v_f}\right)^a + \left(\frac{k}{k_{jam}}\right)^b = 1 \quad (6)$$

then after substituting Equation 5 in, we obtain the *v*-*k* relation.

The meaning of $(sk+r)$ is interpreted as follows. Letting $s>0$, then as k increases, the variance of the first term in the product of Equation 5 becomes larger. This reflects the common perception that the free flow speed is less informative for the inference of $v(k)$ when density k increases, as in this case, the system state represented by (k,v) is moving away from the state $(0,v_f)$. Recall that speed–density relation can be regarded as a curve on k-v plane that interpolates two points $(0,v_f)$ and $(k_j,0)$.

While the speed–density relation depicted here has a lack of sound proof at microscopic level, it has three advantages. First, it is constructed in a heuristic manner as previously mentioned. Interpretation of such relation is quite intuitive. Second, the postulated v-k relation is consistent with the observation that a peak of standard deviation exists between 0 and k_{jam}, which also takes a maximum value in this region. Third, this relation leads to an easy-to-sample random flux function, namely, a family of curves governed by single random parameter ε.

NUMERICAL STUDY

Numerical Scheme

In this paper we adopt the finite difference methods with an ENO reconstruction to obtain the numerical solution of Equation 1. Roughly speaking, the ENO method reconstructs the cell boundary values through adaptively utilizing the local stencil information. In particular, the stencil with the minimal non-smoothness measure is selected. With ENO, high order finite volume or finite difference methods are immediately available. The order of accuracy depends on the size of the adopted stencil. The effectiveness of this method to solve the conservation equations of traffic flow (i.e., the LWR model) has been examined (*24*, *25*). For a detailed description of this method, the reader is referred to Shu (*26*) and its references. In summary, we devise the algorithm as follows to repeatedly generate random flux function and solve the corresponding LWR model. This algorithm consists of two parts, i.e., initialization and time marching. Detail is as follows:

* Initialization:

1. Load initial data $\{k_i^0, i = 1,\ldots,N\}$, where $k_i^0 = k_0(x_i)$ is the grid point value. Moreover, set $s = 1$, $\epsilon_0 = 1$;
2. Generate ϵ_s, independent of $\epsilon_1,\ldots,\epsilon_{s-1}$ and follows $U(-\sqrt{3},\sqrt{3})$. Generate the random flux function $f(k)$ following (11);
3. Split the flux function $f(k) = f^+(k) + f^-(k)$, following (13);

* Time Marching:

4. Identify $\{f^+(k_i^n), i = 1\ldots,N\}$ as cell averages and obtain $v_{i+1/2}^- = \hat{f}_{i+1/2}^+$ by ENO reconstruction. Similarly, get $v_{i+1/2}^+ = \hat{f}_{i+1/2}^-$;
5. If $n+1 \le T/\Delta t$, update the point value,

$$k_i^{n+1} = k_i^n - \frac{\Delta t}{\Delta x}(\hat{f}_{i+1/2} - \hat{f}_{i-1/2})$$

Let $s = s+1$, and go back to 2; else, stop.

Example

To make the model more realistic, the four parameters in Equation 6 need to be estimated. In this paper, we adopt the values $\alpha = 1, \beta = 1, s = 0.05$ and $r=3$ for the purpose of illustration. And the following are set: $\overline{v_f} = 60$ and $k_{jam} = 200$.

The free boundary conditions are imposed throughout the simulations in this section, i.e., $\partial k / \partial x = 0$ at the left and right boundaries of the computation region. We set $dt = 1$ s, $dx = 0.1$ mi.

We investigate the distributional properties of the solution of the proposed model, driven by the i.i.d. random sequence. In particular, the propagation of local disturbance on a one-lane freeway is studied. Here jam/vacuum is defined by significantly different values of local density compared with its neighborhood. We assume $\varepsilon_i \sim U(-\sqrt{3}, \sqrt{3})$ for the aforementioned reason. The settings for the cases studied are below (see Table 1).

The simulation results at $t = 600$ *s* are shown in Figure 2. It seems that the predictability of the output drops significantly when the initial randomness has a variance of one. We also list certain statistics of the solutions in Table 1. Two quantities, the maximum local fluctuation and its location, are particularly interesting. The former is defined as

$$k_* = \max\{|k(x_i) - 50|, 0 \le x_i \le 10\} \quad (7)$$

and the latter is defined by

$$x_* = \arg\max\{|k(x_i) - 50|, 0 \le x_i \le 10\} \quad (8)$$

We also calculate the coefficient of variation (COV) using the values listed in Table 1 and take its reciprocal as the measure of predicting accuracy. It turned out that $Cov_{a,k^*} > Cov_{b,k^*}$ and $Cov_{a,x^*} > Cov_{b,x^*}$. There are multiple implications from the above results. First, the LWR model is often used to predict the location of shock waves. Our example indicates that caution should

TABLE 1 The Statistics of Numerical Solution Under Initial Condition a and b

1. Local jam:

$$k(x,0) = 50 + 80\sin((x-2)\pi)I(2 \le x \le 3)$$

2. Local vacuum:

$$k(x,0) = 50 - 30\sin((x-2)\pi)I(2 \le x \le 3)$$

	Mean of k_*	Std of k_*	Mean of x_*	Std of x_*
Under initial condition a	16.09	0.50	5.99	0.75
Under initial condition b	26.46	1.92	8.63	0.63

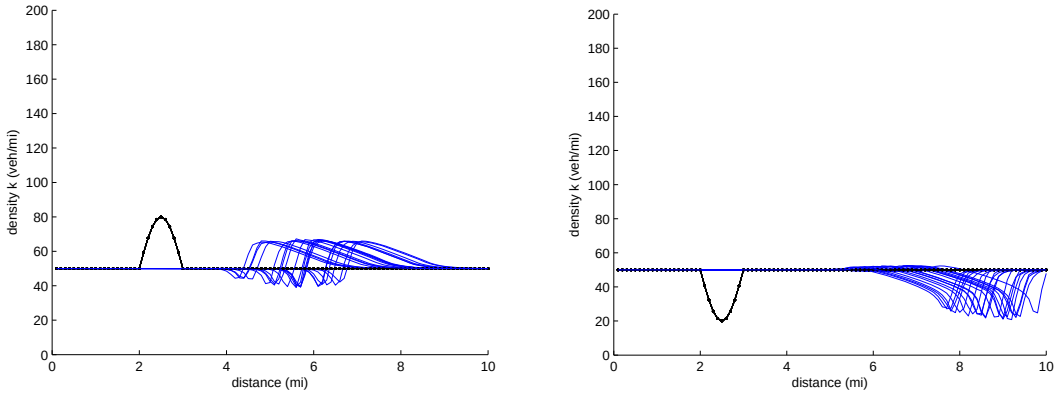


FIGURE 2 Propagation of local disturbance with random flux function (t = 600 s, number of realizations = 20). The dotted line is initial data and solid lines are simulation results.

be taken with such an application when uncertainty of the FD is present, because even with small randomness of the FD the final outputs (i.e., the density profile) may be distributed quite differently. Second, we observe that there is no trend of decay for the spread of model outputs as time passes. This confirms the general intuition that a model deteriorates when applied for a long time prediction. Third, quantitatively, the propagation of local vacuums seems more predictable than local jams, while the latter will be of more interest in application. To summarize, our observations are consistent with the common view point that the LWR model produces a qualitatively good prediction. We find that this statement is credible only if the specification of the FD is certain and a relatively short time scale is involved.

CONCLUDING REMARKS

In this paper, motivated by empirical observations from freeway traffic data and noticing the ongoing debate over the FD, we investigate the LWR model in a stochastic setting. In particular, we discuss the general concept of uncertainty and analyze the sources of randomness associated with the FD. A specific form of the stochastic flux function is postulated, which is driven by random free flow speeds. We provide the mathematical properties of the flux function that are essential for developing a flux splitting scheme. An ENO finite difference numerical scheme is devised and implemented to solve the proposed model. With a hypothetical example, we illustrate how the predictability of two types of local density disturbance travel can be evaluated. This example implies that that if FD cannot be undoubtedly specified, the LWR model will deteriorate and only make reasonable predictions in a relatively short time scale. The analysis framework in this article is hopefully extended to investigation of the general KW models. We remark that the model parameters need to be estimated in order to achieve realistic simulation results, which is a gap left by the current work. Also, some assumptions are possibly relaxed in later study. In this sense, our work in this paper is illustrative in nature. The merit of this study is to initiate the investigations of influences of inherent randomness on traffic modeling. We expect our study will be further advanced with availability of data of finer resolution in the future.

ACKNOWLEDGMENT

The authors wish to thank Steven Andrews, graduate student of civil and environmental engineering at University of Massachusetts, Amherst, for his careful proofreading of the entire draft. We also appreciate the three anonymous reviewers for their valuable comments.

REFERENCES

1. Lighthill, M. J., and G. B. Whitham. On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of Royal Society of London A*, Vol. 229, 1995, pp. 317–345.
2. Richards, P. R. Shock Waves on the Highway. In *Operations Research*, Vol. 4, 1956, pp. 42–51.
3. Daganzo, C. F. Requiem for Second-Order Fluid Approximations of Traffic Flow. In *Transportation Research Part B*, Vol. 29, 1995, pp. 277–286.
4. Zhang, H. M. New Perspective on Continuum Traffic Flow Models. In *Networks and Spatial Economics*, 1, 2001, pp. 9–33.
5. Heidemann, D. Some Critical Remarks on a Class of Traffic Flow Models. In *Transportation Research Part B*, Vol. 29, 1995, pp. 373–389.
6. Castillo, J. M. D., and F. G Benitez. On the Functional Form of the Speed–Density Relationship I: General Theory. In *Transportation Research Part B*, Vol. 29, 1995, pp. 373–389.
7. Cassidy, M. J. Bivariate Relations in Nearly Stationary Highway Traffic. In *Transportation Research Part B*, Vol. 32, 1998, pp. 49–59.
8. Treiterer, J., and J. A Myers. The Hysteresis Phenomenon in Traffic Flow. *Proc., Sixth ISTTT*, 1974, pp. 13–38.
9. Zhang, H. M. A Mathematical Theory of Traffic Hysteresis. In *Transportation Research Part B*, Vol. 33, 1999, pp. 1–23.
10. Newell, G. F. Instability in Dense Highway Traffic: A Review. *Proc., Second ISTTT,* 1965*,* pp. 73–85.
11. Treiber, M., and D. Helbing. Memory Effects in Microscopic Traffic Models and Wide Scattering in Flow-Density Data. In *Physical Review E*, 68, 2003.
12. Zhang, H. M, and T. Kim. A Car-Following Theory for Multiphase Vehicular Traffic Flow. In *Transportation Research Part B*, Vol. 39, 2005, pp. 385–399.
13. Kim, T., and H. M. Zhang. A Stochastic Wave Propagation Model. In *Transportation Research Part B*, Vol. 42, 2008, pp. 619–634.
14. Wong, G. C. K., and S. C. Wong. A Multi-Class Traffic Flow Model: An Extension of LWR Model with Heterogeneous Drivers. In *Transportation Research Part A*, Vol. 36, 2002, pp. 827–841.
15. Ngoduy, D., and R. Liu. Multiclass First-Order Simulation Model to Explain Non-Linear Traffic Phenomena. In *Physica A: Statistical Mechanics and its Applications*, Vol. 385, 2007, pp. 667–682.
16. Kerner, B. S. The Physics of Traffic. Springer, 2004.
17. Kerner, B. S. Three-Phase Traffic Theory and Highway Capacity. In *Physica A: Statistical Mechanics and its Applications*, Vol. 333, 2004, pp. 379–440.
18. Kerner, B. S., and S. L. Klenov. Deterministic Microscopic Three-Phase Traffic Flow Models. In *Journal of Physics A: Mathematical and General*, Vol. 39, 2006, pp. 1775–1809.
19. Holden, H., and N. H. Risebro. Stochastic Properties of the Scalar Buckley-Leverett Equation. In *SIAM Journal on Applied Mathematics*, Vol. 51, 1991, pp. 1,472–1,488.
20. Holden, L. The Buckley-Leverett Equation with Spatically Stochastic Flux Function. In *SIAM Journal on Applied Mathematics*, Vol. 57, 2000, pp. 1,443–1,454.
21. Wehr, J., and J. Xin. Front Speed in the Burgers Equation with a Random Flux. In *Journal of Statistical Physics*, Vol. 88, 1997, pp. 843–871.

22. Bale, D. S., R. J. LeVeque, S. Mitran, and J. A. Rossmanith. A Wave Propagation Method for Conservation Laws and Balance Laws with Spatially Varying Flux Functions. In *SIAM Journal on Scientific Computing*, Vol. 24, 2002, pp. 955–978.
23. Jou, Y. J., and S. C. Lo. Modeling of Nonlinear Stochastic Dynamic Traffic Flow. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1771*, TRB, National Research Council, Washington, D.C., 2001, pp. 83–88.
24. Zhang, M., C. W. Shu, G. C. K. Wong, and S. C. Wong. A Weighted Essentially Non-Oscillatory Numerical Scheme for a Multi-Class Lighthill-Whitham-Richards Traffic Flow Model. In *Journal of Computational Physics*, Vol. 191, 2003, pp. 639–659.
25. Zhang, P., S. C. Wong, and C. W. Shu. A Weighted Essentially Non-Oscillatory Numerical Scheme for a Multi-Class Traffic Flow Model on an Inhomogeneous Highway. In *Journal of Computational Physics*, 212, 2006, pp 739–759.
26. Shu, C. W. High Order ENO and WENO Schemes for Computational Fluid Dynamic. RTO educational notes, 1998, pp. 439–582.

A Macroscopic Fundamental Diagram of Urban Traffic
Recent Findings

NIKOLAS GEROLIMINIS

This paper presents recent findings from Nikolas Geroliminis and Carlos F. Daganzo (*1–2*). Various theories have been proposed to describe vehicular traffic movement in cities on an aggregate level. They fall short of creating a macroscopic model with variable inputs and outputs that could describe a rush hour dynamically. This paper presents recent findings showing that a macroscopic fundamental diagram (MFD) relating production (the product of average flow and network length) and accumulation (the product of average density and network length) exists for neighborhoods of cities in the order of 5 to 10 km^2. These findings demonstrate that, conditional on accumulation, large networks behave predictably and independently of their origin–destination tables. The results are based on analysis using simulation of large-scale city networks and real data from urban metropolitan areas. The real experiment uses a combination of fixed detectors and floating vehicle probes as sensors. The analysis also reveals a fixed relation between the space-mean flows on the whole network and the trip completion rates, which dynamically measure accessibility. This work also demonstrates that the dynamics of the rush hour can be predicted quite accurately without the knowledge of disaggregated data. This MFD can be applied to develop perimeter control and other macroscopic control strategies (pricing, allocation of urban space, etc.) based on neighborhood accumulation and speeds and improve accessibility without the uncertainty inherent in today's forecast-based approaches.

REFERENCES

1. Geroliminis, N., and C. F. Daganzo. Macroscopic Modeling of Traffic in Cities. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
2. Geroliminis, N., and C. F. Daganzo. Existence of Urban-Scale Macroscopic Fundamental Diagrams: Some Experimental Findings. In *Transportation Research Part B*, Vol. 42 No. 9, 2008, pp. 759–770.

Influence of Various Restrictions on Speed-Flow Models

MARIAN TRACZ
STANISŁAW GACA
Krakow University of Technology

The paper presents traffic speed models developed on the basis of numerous measurements. In studies of the speed–flow relationships the following research issues were analyzed:

- **impacts of the speed limit, traffic flow, traffic flow composition, and time of day on speed flow and**
- **impacts of qualitative road and roadside features on free-flow speed (cross sections, road function, accessibility, intensity of development, and roadside pedestrian traffic).**

The research included road sections with numerous constraints of traffic freedom. Such sections comprise sections of roads through built-up areas (small communities and towns). The regression analyses, with use of generalized multiple regression, was the basic tool in investigations. Additionally, in analyzing data from speed measurements, the classic form of the logit function for a multiple-valued regression was applied. The analyses showed that even at general and posted speed limits the characteristic traffic flow–speed relations can be identified. The shape of these relations is characteristic for empirical results in which spread of recorded values is typical for randomness and impacts of local factors. Average speed in traffic flow depends not only on geometrical alignment of a road cross section but also on local factors as well as speed limits. Speed in free-flow traffic is significantly influenced by qualitative road and roadside characteristics. Most important of these factors are: type of cross section, intensity of roadside development, and accessibility of the road.

Quality of Service Beyond the Traditional Fundamental Diagram

ANJA ESTEL
Ruhr University, Germany

In the German Highway Capacity Manual (HBS) the design and operation analysis of freeway sections is based on the volume-to-capacity ratio, which is used as a substitute for the average travel speed. For undersaturated flow the travel time can indirectly be calculated from speed-flow diagrams given within the HBS. However, the applicability of speed-flow diagrams is restricted to demand values less than the capacity. A differentiation of oversaturated traffic conditions with a demand beyond the speed-flow diagram is not possible so far. The paper presents approaches for a division of Level of Service F (LOS F) into differentiated degrees (F1 to F4) and proposes corresponding calculation methods. For the estimation of travel times as a function of traffic demand, different traffic flow models and queuing equations are considered. Four standardized methods are described in the paper, differing in the performance measures used. In the first three methods, the peak hour is retained as the analysis interval. In contrast to these concepts, the latter evaluation is based on a 24-hour analysis. This more sophisticated concept allows the consideration of spatial as well as temporal effects. The methods are analyzed and compared in terms of their significance, their sensitivity and the availability of input data. Predominantly, the evaluation methods developed in this research can be applied for assessing the quality of service on congested facilities depending on traffic demand. In the scope of decision making, funds can be spent more efficiently.

In the German HCM (HBS, FGSV 2005) the design and evaluation of freeway sections are made on the basis of the volume-to-capacity ratio as a substitute for the average travel speed and traffic density. For undersaturated flow [Level of Service (LOS) A–E] the most decisive criterion for users, the travel time, can be derived from speed-flow diagrams given within the HBS. Thus, the fundamental diagram is a practical instrument for evaluating the operation quality on freeway sections below saturation. Beside the Levels of Service A to E there is only one level “F” relating to oversaturated operating conditions. Various oversaturated situations cannot be differentiated so far.

The obvious increase of traffic volume more often results in serious congestions. A considerable growth of daily congestion hours requires a differentiated evaluation of oversaturated operating. Hence, the aim is to divide Level of Service F into differentiated sections (F1 to F4) and to develop calculation methods for their determination. This will help in making decisions to eliminate worst bottlenecks first. Additionally it permits the quantification of improvement capabilities of the operation quality by implementing corridor control systems or temporary unblocking of shoulders.

For such evaluations the estimation of travel times is needed as a function of traffic demand. Here the use of the fundamental diagram and speed-flow diagram are restricted for oversaturated flow conditions. The lower branch of the traffic flow model does not represent traffic demand but the possible flow during congestion. Moreover, the fundamental diagram reflects snapshots of traffic flow detached from time dependence and spatial propagation of traffic conditions. On this account there has been a search for calculation methods allowing evaluation beyond the speed-flow diagram.

Already within the 1990s the need for a differentiated evaluation of congestion was recognized in the United States. Various concepts were elaborated and published. Within this scope May (1996, 1997) recommended a differentiation into three congestion levels, F1 to F3, using a combination of the performance measures saturation and the duration of the congestion. An alternative concept exceeds the evaluation of the peak hour by evaluating the quality of service over a whole day. Cameron (1996) recommended the assessment of each hour of a day using the methods for undersaturated flow. The oversaturated operating conditions are divided into four congestion levels, F to I, and additional thresholds are defined. As a result the number of each congestion level over the whole day, like H1 G3 F7, can be obtained. Since such results cannot be compared easily with each other, Baumgaertner (1996) took up this concept and suggested a summarized evaluation of the whole day using a so-called congestion index. This value (0 for uncongested up to more than 40 for extremely congested) represents a weighted average of the 24 hourly levels of service.

These concepts haven't been included within the American HCM (TRB, 2000) so far. They will form a foundation for some of the concepts, developed within this paper. The following requirements should be fulfilled for a differentiated evaluation method:

- Demand patterns for several hours should be considered, to assess congestion effects on following time periods and segments like spatial and temporal shifting of flow (demand) as well as the phenomenon of the capacity drop.
- The method has to be feasible using only input data that are available in practice: these data are often rare and inaccurate.
- It has to be user friendly but realistic: congestions should be described as realistically as possible without losing clarity. For this purpose, traffic flow models should be implemented.
- It should be used without software solution (pen and paper methods).

DEFINITION AND DESCRIPTION OF CONGESTION

Definitions

On highways segments congestion often occurs at so called bottlenecks which can be lane reductions, ramps, or topographical changes like steep gradients. There are various definitions for congestion. The most common definition is

> The occurrence of a traffic demand that exceeds the capacity of the highway segment during an interval of a specific duration. The demand cannot completely pass during this interval.

Often, congestion is defined due to the effects caused by the congestion. For example these are the disturbance of flow (unstable flow resulting in stop-and-go traffic) as well as additional travel times (occurring during undersaturated flow conditions, increasing seriously during congestion). As another definition, exceeding upper or lower limit values of measures that describe operating conditions (given thresholds between LOS E and F in the manual) can be used.

For the development of evaluation methods for congested traffic, significant performance measures are required to describe and compare the effects of different congested situations. A literature research shows that there are numerous performance measures, more or less

appropriate for different purposes. Comprehensive overviews over possible measures are given by Shaw (2003), Bertini (2005) as well as Schrank, Lomax (2001, 2005). Different dimensions of congestion should be considered.

In order to integrate a performance measure in a method applicable for evaluation as well as for design, the measure has to be computable. Most of the performance measures, shown in Figure 1, can be calculated with the aid of traffic flow models and queuing solutions. The following performance measures, that seem appropriate for describing congestion effects, were chosen as a basis for evaluation methods:

- Volume to capacity ratio,
- Average delay,
- Sum of delays within a specific analysis period,
- Travel time or travel time index, and
- The probability of breakdown.

Macroscopic models for calculating those measures are described in the following paragraphs.

Traffic Flow Models

The correlation between traffic flow and density on a road segment is described within the fundamental diagram. In the speed-flow diagram there are two speed values belonging to each traffic flow value, one for fluent traffic and one for congested condition.

This results in an upper and a lower branch, marking steady traffic flow conditions in each case. The transition between both branches occurs suddenly and not continuously. The average speed that can be realized during a traffic flow below capacity can be determined with a traffic flow model.

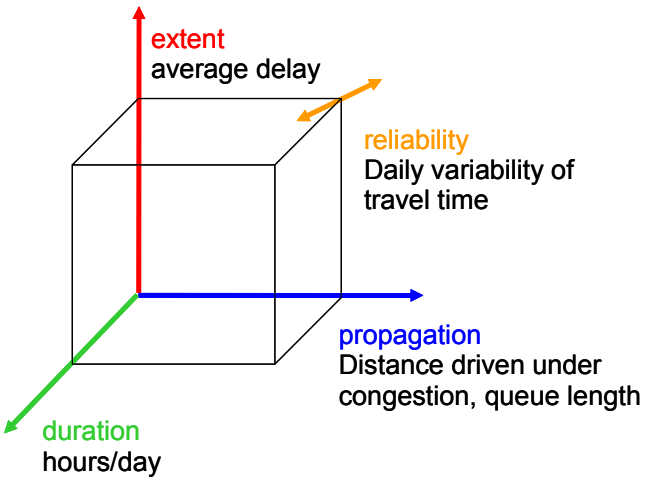


FIGURE 1 Dimensions of congestion (Lomax et al., 1997).

Lighthill and Whitham (1955) developed a theory, describing traffic flow along free road segments in a hydrodynamic analogy to the flow of one-dimensional, compressible fluids. The motion of waves on a river is taken as an analogy to the movement of density waves in traffic flow. Thus, traffic flow can be described at every location. The basic idea of this traffic flow model is based on the analysis of small segments in the space–time area as well as the assumption that no vehicles get lost within those segments. Thus, the following differential equation can be applied (continuity equation):

$$\frac{\partial q(s,t)}{\partial s} + \frac{\partial k(s,t)}{\partial t} = 0 \tag{1}$$

Assuming stationary flow conditions, the well-known equation of state can additionally be formed to describe traffic flow:

$$q = k \cdot \bar{v}_m \tag{2-a}$$

where

q = flow rate (veh/h),
k = density (veh/km), and
$\bar{v}_m$ = space mean speed (km/h).

If two different traffic densities meet each other a density wave moves upstream. Its speed corresponds to the secant between both points within the fundamental diagram. It can be calculated by

$$u = \frac{q_2 - q_1}{k_2 - k_1} \tag{2-b}$$

where

u = speed of the density wave (km/h),
$q_{1,2}$ = flow rate in both flow conditions (veh/h), and
$k_{1,2}$ = density in both flow conditions (veh/km).

Beside the hydrodynamic models queuing models can be used for the description of traffic flow. The bottleneck is treated as the server. The theory permits the calculation of several performance measures like the average delay in the queue as well as the expected number of waiting vehicles.

Brilon and Ponzlet (1995) developed a macroscopic traffic flow model describing the relation between traffic flow and average speed on highway segments operating below saturation, based on a simple M/M/1-queuing analogy. As long as the flow does not exceed the capacity, the average speed can be determined by

$$v(q) = \frac{v_0}{1 + \frac{v_0}{L_0 \cdot (C_0 - q)}} \tag{3}$$

where

$v(q)$ = average speed at a specific flow q (km/h),
v_0 = parameter of the model (km/h),
L_0 = parameter of the model (km),
C_0 = parameter of the model (veh/h), and
q = traffic volume (veh/h).

Kimber and Hollis (1979) developed methods for time-dependent queuing calculations to describe fluent and congested traffic conditions at road junctions. The recommended equations given in the appendix are based on a probabilistic approximation of a transformed M/M/1 queuing model. The advantage of this model consists in the consideration of the statistical character of traffic flow as well as of the time dependence of demand and capacity.

Since the assumption of random arrivals and service at a give-way junction can be adopted for describing traffic flow at a freeway bottleneck, the equations may be used for calculation of growth and decay of the average queue length depending on patterns of demand and capacity on a highway segment. This allows an evaluation of the effects of congestion (queue length, delay, duration) beyond the peak hour. The necessary input data are demand as well as capacity, which both are available in practice.

CONCEPTS FOR DIFFERENTIATED EVALUATION METHODS

Preliminary Remarks

Within a research project (Brilon and Estel, 2007) evaluation methods to assess the degree of congestion were developed. By combining performance measures and different time periods with specific evaluation methods, various potential concepts were elaborated for congested flow conditions. They differ concerning comprehensibility, complexity, data availability, and significance as well as user friendliness. By means of example evaluations data requirements and practicability were assessed. After an intensive discussion with German experts from science and practice, the following four concepts were selected upon which to base evaluation:

- The measures of effectiveness within the current guideline (HCM),
- The probability of breakdown,
- A travel time index, and
- The delay due to congestion.

The developed methods can be used for all kinds of road traffic facilities. The following statements are concerning freeway segments. Procedures describing oversaturated conditions are

more complex, because congestion on the one hand affects facility sections up- and downstream from the bottleneck and on the other hand influences the flow conditions in subsequent time intervals. Thresholds between four levels of congestion can be proposed. Since their values have to be defined by the responsible road authorities as a political decision, no thresholds are given within this paper.

Application of the Methods That Are Described in HBS for Undersaturated Traffic Flow Conditions

The application of the calculation procedures described for undersaturated flow on oversaturated conditions allows a simple approximate evaluation for sections of freeways during the peak hour. Depending on the volume-to-capacity ratio higher than one a congestion Level F1 to F4 can be determined. For this purpose additional thresholds have to be determined to differ between congested situations of different severity.

Some variations should be considered compared to the treatment of fluent traffic conditions. Since congestion influences the operation conditions in subsequent analysis periods as well as in adjacent highway segments the effects of congestion have to be evaluated. As long as an isolated peak hour is evaluated, the shift in demand is not considered. What must not be neglected are the capacity drop phenomenon as well as the limitation of volume during the peak hour for downstream highway segments.

The capacity drop, defined as the difference between the capacities in fluent and congested traffic flow, is caused probably due to downstream bottlenecks, the adjustment of driver behavior at the beginning of a bottleneck, as well as the limited acceleration capabilities of different vehicles (Brilon et al., 2005). The value of capacity drop has been analyzed for short analysis periods as 1- to 5-minute-intervals so far. Banks (1990) as well as Hall and Agyemang-Duah (1991) analyzed the capacity drop for different North American freeways. The measured capacity drop values averaged 3% to 6%. Ponzlet (1996) analyzed traffic flow on German freeways and determined a 6% drop for 5-minute flow rates and a 4% drop for 15-minute flow rates. Brilon and Zurlinden (2003) computed an average capacity drop of 24% for 5-minute flow rates, by comparing a stochastic capacity to flow rates during congested operation. A relationship between segment characteristics and the value of the capacity drop has not been found yet. However, since the existence of this phenomenon has doubtlessly been detected by several empirical studies, this effect should be considered within the evaluation of congested conditions to reproduce realistic congestion extends. As the average capacity in fluent traffic for 1-hour flow rates is considerably lower than for 5-minute flow rates, a 5% drop is recommended as input for the evaluation method.

The determination of the operating performance has to be performed separately for each segment. Segments are parts of a road section beyond the sphere of intersections that differ in several segment characteristics (gradient, number of lanes, speed limits…). Since there are no ramps between two segments a constant traffic volume can be assumed. However, the capacity can vary considerably due to the segment characteristics.

If the capacity of one of the segments is lower than the traffic volume, the number of vehicles that can pass this as well as the following segments during the peak hour is limited by this capacity and additionally restricted by the capacity drop. That is why the relevant traffic volume for each segment is determined by

$$q_{act}(i,t) = \min\begin{cases} q_B(i,t) \\ C_{act}(i-1,t) \\ q_{act}(i-1,t) \end{cases} \tag{4}$$

where

$q_{act}(i,t)$ = actual flow within the peak hour in segment i (veh/h),
$q_B(i,t)$ = design flow during the peak hour (veh/h),
$C_{act}(i-1,t)$ = capacity of the upstream segment i-1 within the peak hour (veh/h), and
$q_{act}(i-1,t)$ = actual flow of the upstream segment i-1 within the peak hour (veh/h).

In case of congestion the actual capacity considering the capacity drop has to be used for the calculation of the volume to capacity ratio:

$$C_{act}(i,t) = \begin{cases} C(i,t) & \text{fluent traffic} \\ C(i,t) \cdot c_{Drop} & \text{congested traffic} \end{cases} \tag{5}$$

where

$C_{act}(i,t)$ = actual capacity of the segment i during the peak hour (veh/h)
$C(i,t)$ = capacity in fluent traffic of the segment i during the peak hour (veh/h), and
c_{Drop} = capacity drop of segment (c_{Drop}=0,95) (-).

The described effects are explained in Figure 2.

The volume-to-capacity ratio than can be calculated to

$$a = \frac{q_{act}(i)}{C_{act}(i)} \tag{6}$$

where

a = volume-to-capacity ratio (-),
$q_{act}(i)$ = actual flow rate within the peak hour in segment i (veh/h), and
$C_{act}(i)$ = actual capacity of the segment i during the peak hour (veh/h).

Since in this method only the peak hour is considered, the effects can only be evaluated incompletely.

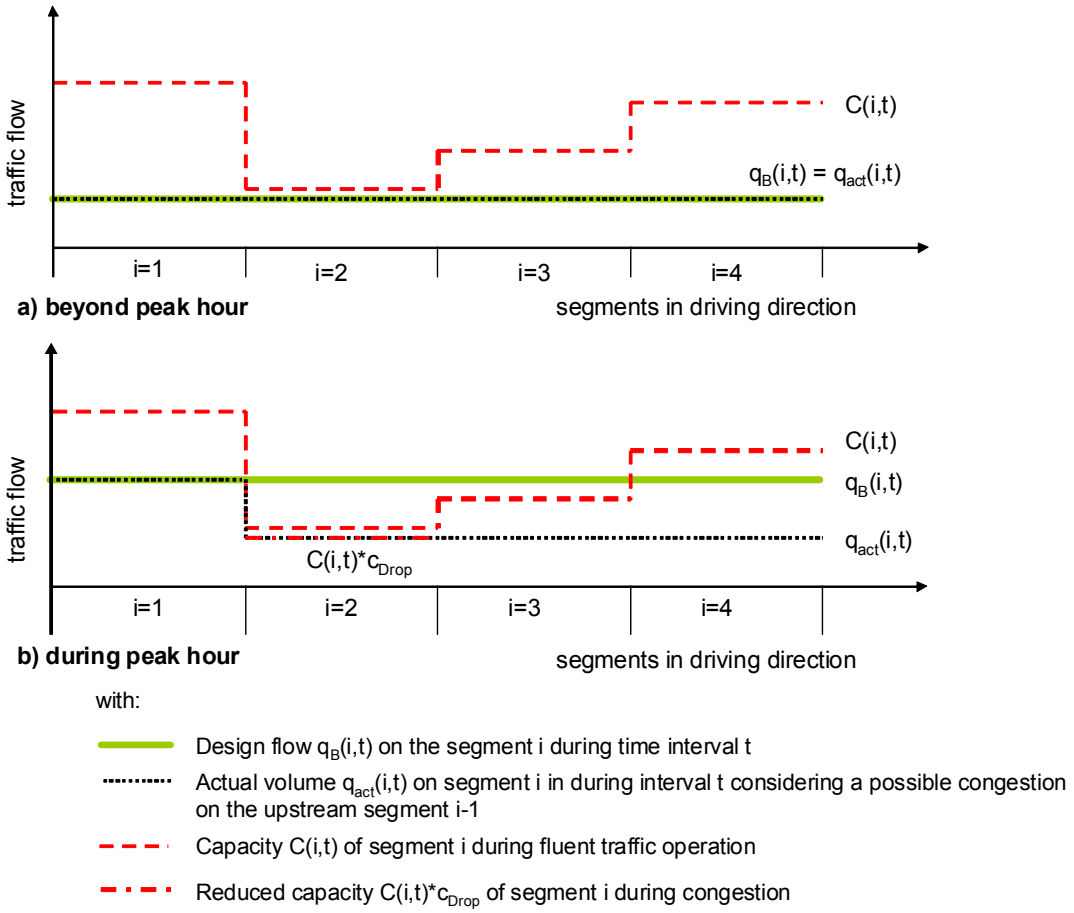


FIGURE 2 Effects on flow and capacity due to congestion on subsequent freeway segments.

Evaluation Based on the Probability of Breakdown

This evaluation method seems to represent another appropriate approach for a differentiated evaluation of traffic quality on freeways. It does not focus the effects in case of an oversaturated condition. Instead, the substantial question of this assessment is at which probability an oversaturated condition will occur at all. The question of whether or not congestion will occur is particularly important for the quality of service on freeways sections. Because of the capacity drop, oversaturated conditions dissipate rather slowly. At the same time a large total delay results from the high travel demand on freeways. Therefore a prevention of oversaturated conditions is essential.

The criterion of breakdown probability is based on the fundamental idea of a stochastic capacity. Instead of the assumption of capacity being a constant value, a distribution of varying capacities can be observed on freeway sections by detecting traffic breakdowns at different flow rates (Elefteriadou et al., 1995; Ponzlet, 1996; Minderhoud et al., 1997; Persaud et al., 1988; Kuehne and Anstett, 1999; Lorenz and Elefteriadou, 2000; and Okamura et al., 2000).

Defining the probability of a breakdown as the probability of reaching a specific capacity, the distribution function can be derived from empirical speed-flow relations by means of the product limit method (Brilon and Zurlinden, 2003, 2004; Regler, 2004; and Brilon et al. 2005, 2006). The resulting distribution functions can be fitted to Weibull distributions by means of maximum likelihood estimation (Figure 3). The criterion for determining the level of congestion is the probability at that demand exceeds capacity and congestion occurs. Within the evaluation method the breakdown probability is determined depending on the traffic volume as well as on the proportion of heavy vehicles. Both input data are available in most cases.

Since the distribution function of breakdown probability varies due to the segment characteristics, general design charts are prepared as a dimensioning or evaluation aid for the user. To provide such diagrams, empirically calculated distribution functions of freeway capacity have to be standardized and transformed into 1-hour-intervals. For this purpose, the probabilities for non-appearance of a single breakdown event within the 5-minute-intervals of the peak hour are multiplied. The probability that no breakdown will occur in any 5-minute-interval can be calculated as follows (Brilon et al., 2005):

$$p_Z(q_{60}) = 1 - \prod_{i=1}^{12} e^{-\left(\frac{q_i}{b}\right)^a} \tag{7}$$

where

$p_Z(q_{60})$ = breakdown probability relating to a 1-h volume q_{60} (-),
q_i = flow within a 5-minute interval i (veh/h), and
a, b = parameter describing Weibull distribution.

One can assume that the nonappearance of breakdowns in each of the 12 5-minute intervals is statistically independent of each other. The more important question for the transformation is how the traffic flow within the 5-minute intervals q_i is distributed in relation to the average hourly volume q_{60}. Constant flow rates as well as normal distributed flow rates may be assumed.

If the traffic flow during the peak hour is assumed to be nearly constant (q_i = const = q_{60}), Equation 7 can be reduced to Equation 8:

$$pz(q_{60}) = 1 - e^{-\left(\frac{q_{60}}{b'}\right)^{a'}} = 1 - e^{-\left(\frac{q_{60}}{b/\sqrt[a]{12}}\right)^{a'}} \tag{8}$$

where

$p_Z(q_{60})$ = breakdown probability relating to the volume q_{60} (-),
q_{60} = traffic volume during the peak hour (veh/h),
a' = a,
b' = $b/\sqrt[a]{12}$,
a = shape parameter (describing the variance of the Weibull distribution), and
b = scale parameter of the Weibull distribution.

Such an assumption allows a simple, approximate calculation of the breakdown probability. In Figure 4 a result for $p_Z(q_{60})$ is shown as an example for a three-lane carriageway with less than 2% gradient in urban areas.

For freeway segments without variable speed-control system the average capacity value given in the manual HBS (FGSV, 2005) corresponds to a breakdown probability of 40% if constant traffic flow is assumed, i.e., a volume equal to capacity will cause a considerable breakdown in 40% of all cases.

Geistefeldt (2007) analyzed the influence of the kind of distribution of traffic flow on the distribution function of capacity (i.e., mathematical form of 5-minute volumes fluctuating around the average 1-h volume). Here no significant differences could be determined between the assumption of a normal and a uniform distribution. Most decisive for the capacity distribution is the standard deviation.

Analyzing 36 sections of two- and three-lane freeways average standard deviations of 5.8% (three-lane) and 8% (two-lane) were obtained in the range of capacity. Due to this supposition there are 5-minute flow rates considerably higher than the average traffic volume q_{60}. The probability of nonappearance of a breakdown in these intervals is much lower than under the assumption of constant traffic flow. As a consequence the resulting distribution functions are shifted in the direction of lower traffic flow.

The determination of the distribution functions is based on Equation 7. For normal distributed q_i this equation can only be solved numerically or by simulation. Within this research a Monte Carlo simulation was chosen. The resulting functions conform to the functions obtained by Geistefeldt who used a numerical approach.

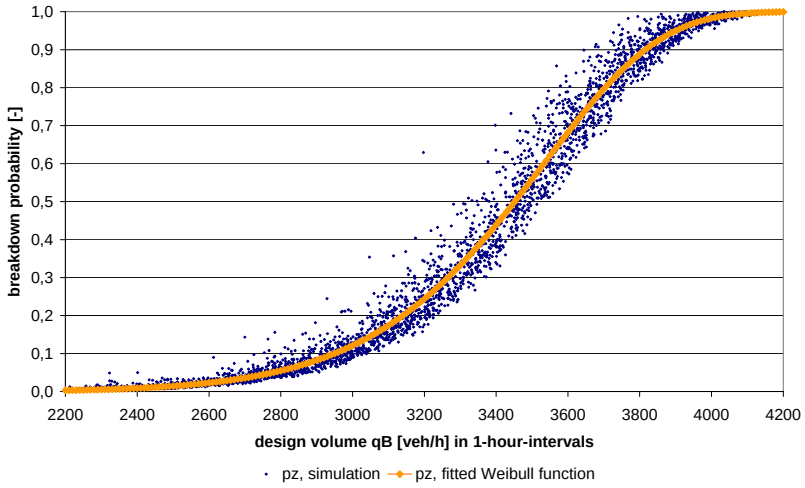


FIGURE 3 Calculation of the distribution function of breakdown probability assuming normal distributed flow rates.

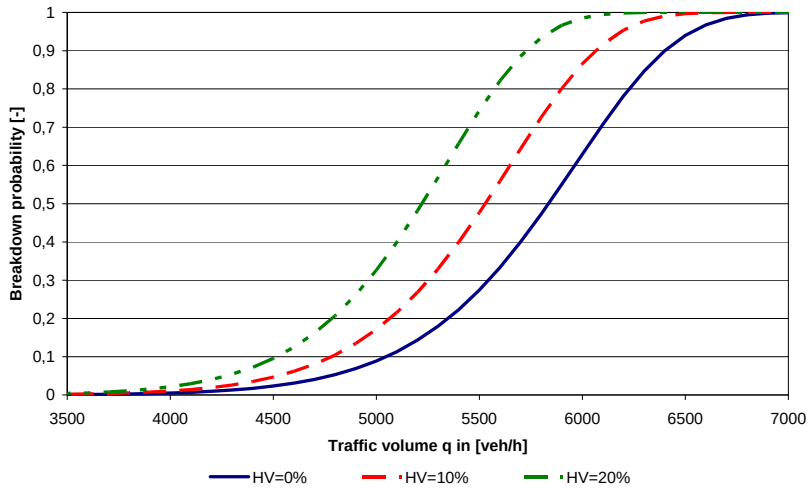


FIGURE 4 Breakdown probability assuming constant flow during the peak hour.

As a consequence of the steep slope of the breakdown probability function in the range of larger saturation, this evaluation method reacts very sensitively on small changes of the input volume. Due to the skewness of the Weibull distribution function the evaluation is more sensitive to under- than overestimation. The extensive sensitivity has to be considered through the proposal of thresholds between the congestion levels. Due to this reason, the thresholds should not be defined linearly dependent on the breakdown probability.

Evaluation Based on a Travel Time Index

The application of the travel time index allows an evaluation based on travel time performance. This characteristic is particularly suitable to describe the performance of the traffic flow. The travel time varies depending on the quality of service. Thus, the use of the travel time reliability as a measure of performance is frequently suggested (Schrank, Lomax, 2001; Chen, 2003; Shaw, 2003; and FHWA, 2006). As a component of the proposed measures a specific percentile of the empirical travel time data is used to express the reliability. Since such a percentile cannot be computed without measurement, it does not seem to be an appropriate measure of effectiveness within a standardized assessment method for both, design and evaluation. As an alternative a travel time index is suggested, which describes the ratio of an average travel time that can be observed during a specific quality of service and an acceptable travel time. The latter is the average travel time observed during 90% saturation. At this saturation the maximum efficiency of the transport facility is reached (Brilon, Zurlinden, 2003). For undersaturated flow conditions the travel time is described by the speed-flow diagram. For congested flow conditions the travel time can be calculated with a queuing model or on the basis of the kinematic wave theory. By generating an index, influences of segment characteristics as well as speed limits can be considered. Thus, comparative evaluations of segments with different characteristics become possible. At the same time this concept permits a consistent, summarizing evaluation of sequences of different traffic facilities.

Evaluating the peak hour, the ratio of an average travel time, which can be achieved at a certain traffic volume, and an acceptable travel time is calculated and used as performance measure. The actual travel time consists of two components: the delay due to congestion as well as the reduced time to pass the segment. The total sum of travel times within the peak hour can approximately be calculated using an extended deterministic queuing model (Geistefeldt, 2005). In this case, the queue length at the end of the peak hour equals to

$$N(i,t) = \begin{cases} 0 & q_{act}(i,t) - C_{act}(i,t) \le 0 \\ (q_{act}(i,t) - C_{act}(i,t)) \cdot T & \text{else} \end{cases} \quad (9)$$

where

$N(i,t)$ = queue length of segment *i* at the end of the peak hour (veh),
$q_{act}(i,t)$ = actual flow within the peak hour in segment *i* (veh/h),
$C_{act}(i,t)$ = actual capacity within the peak hour in segment *i* (veh/h), and
T = duration of time interval (h).

The delay due to congestion RZ_{cong} is equivalent to the area below the function $N(i,t)$. The total sum of travel times than is equal to

$$RZ(i,t) = \begin{cases} \frac{L}{v(q_{act}(i,t))} \cdot q_{act}(i,t) \cdot T & \text{during fluent traffic} \\ \frac{L}{v_{crit}} \cdot q_{act}(i,t) \cdot T + RZ_{cong}(i,t) & \text{during congestion} \end{cases} \quad (10)$$

where

$RZ(i,t)$ = total sum of travel times of segment *i* during the peak hour (veh * h),
L = length of the segment *i* (km),
)) $v(q_{act}(i,t$ = average speed depending on the actual flow q_{act} (km/h),
$q_{act}(i,t)$ = actual flow within the peak hour in segment *i* (veh/h),
T = duration of time interval (h),
v_{crit} = critical speed at which flow reaches capacity (q = C) (km/h), and
$RZ_{cong}(i,t)$ = total sum of delays due to congestion (veh * h).

The average travel time per vehicle is

$$rz_{act}(i,t) = \frac{RZ(i,t) \cdot 60\,\text{min}/h}{q_{act}(i,t)} \quad (11)$$

where

$rz(i,t)$ = average travel time per segment *i* and per vehicle (min/veh),
$RZ(i,t)$ = total sum of travel times of segment *i* during peak hour (veh * h), and
$q_{act}(i,t)$ = actual flow within the peak hour in segment *i* (veh/h).

Instead of this queuing model the more realistic approach of the kinematic wave model can be applied for the calculation of the actual average travel time per vehicle. In this case one needs the fundamental diagram of the bottleneck segment as well as of the upstream segment in which the queue will occur as a consequence of the congestion. From that fundamental diagram of the upstream segment the speeds and corresponding densities during congestion as well as fluent traffic can be extracted. It is more difficult to estimate the average speed during congestion. If unknown the use of v_{cong} = 20 km/h is suggested. The second fundamental diagram for the segment downstream of the bottleneck reveals the traffic flow conditions as well as the reduced average capacity at the bottleneck. The queue length at the end of the interval *t* then is

$$L_{cong}(t) = -\frac{\Delta q(t)}{\Delta k(t)} \cdot (t-(t-1)) = \frac{q_{act}(t) - C_{act}}{k(t) - k_{cong}} \cdot (t-(t-1)) \tag{12}$$

where

L_{cong} = queue length at the interval of peak hour (km),
$q_{act}(t)$ = actual flow within the peak hour in segment *i* (veh/h),
C_{act} = actual capacity of the segment *i* during the peak hour (veh/h),
$k(t)$ = density in the undisturbed inflow (veh/km),
k_{cong} = average density in the queue upstream the bottleneck (veh/km), and
t = time (h).

and the additional average delay due to congestion is

$$rz_{cong}(i) = L_{cong}(t) \cdot \left(\frac{1}{v_{cong}} - \frac{1}{v_{free}} \right) \tag{13}$$

where

$rz_{cong}(i)$ = average travel time per segment *i* and per vehicle during peak hour (h/veh),
L_{cong} = queue length at the interval of peak hour (km),
v_{free} = average speed during fluent flow upstream the bottleneck (km/h), and
v_{cong} = average speed in the queue upstream the bottleneck (km/h).

For the determination of the travel times both segments downstream and upstream the bottleneck are considered. The actual travel time that is consumed to pass both segments can be obtained by

$$rz_{act}(i) = \begin{cases} \frac{L_{up}}{v_{up}(q(t))} + \frac{L_{down}}{v_{down}(q(t))} & for\ L_{cong} = 0 \\ \frac{L_{cong}(t)}{v_{cong}} + \frac{L_{up} - L_{cong}(t)}{v_{up}(q(t))} + \frac{L_{down}}{v_{crit}(q = C_{cong})} & for\ L_{cong} < L_{up} \\ \frac{L_{cong}(t)}{v_{cong}} + \frac{L(down,t)}{v_{down}(q(t))} & else \end{cases} \quad (14)$$

where

$rz_{act}(i)$	=	average travel time per segment *i* and per vehicle during peak hour (h/veh),
L_{cong}	=	queue length at the interval of peak hour (km),
L_{up}	=	length of the segment upstream the bottleneck (km),
L_{down}	=	length of the segment downstream the bottleneck (km),
v_{cong}	=	average speed in the queue upstream the bottleneck (km/h),
$v_{up}(q(t))$	=	average speed upstream the congestion depending on the flow rate (km/h),
$))v_{down}(q(t$	=	average speed downstream the bottleneck depending on the flow rate (km/h), and
$v_{crit}(q = C_{cong})$	=	critical speed downstream the bottleneck (km/h).

This model reflects the consumed travel times more realistically than the theoretical consideration of stagnant waiting vehicles. Independent from the applied model the travel time index for the peak hour can be determined for each segment. To evaluate the whole freeway section, the actual and acceptable travel times of each segment can be summarized to obtain the ratio

$$TTI(i) = \frac{rz_{act}(i)}{rz_{accept}(i)} \quad (15)$$

where

$TTI(i)$	=	travel time index of segment *i* (-),
$rz_{act}(i)$	=	average travel time per segment *i* per vehicle during peak hour (min/veh), and
$rz_{accept}(i)$	=	acceptable travel time per segment *i* per vehicle during peak hour (min/veh).

In the case of congestion, the throughput is limited by the capacity. As a consequence the flow passing the downstream segment is also restricted. Considering the quality in the upstream segments, the actual traffic flow equates to Equation 4. Since the capacity within congested flow is reduced by the capacity drop compared to fluent traffic, the capacity has to be modified depending on the existence of a queue regarding to Equation 5.

Since the additional travel time due to congestion does not depend on the segment lengths, comparing the same congestion extent (same volume-to-capacity ratio) the ratio of actual and acceptable travel time becomes larger with decreasing segment length. To obtain a comparable evaluation criterion, the thresholds between the congestion levels have to be related to the segment length.

Here it should be pointed out that this kind of performance measure may also be used as measure of effectiveness in the range of fluent traffic flow (LOS A to E) as an alternative to the volume-to-capacity ratio.

Evaluation Based on Delay Due to Congestion

Using the delay due to congestion as measure of effectiveness allows a comparative evaluation of different facilities as well as the determination of a quantitative structure for an economic evaluation. Since delay is a time dependent value one should look at more than one peak hour. Additionally, the examination of several hours allows the consideration of effects caused by the congestion, such as capacity drop and time shift in demand. Some freeway sections are congested twice a day. If this should be observed, it is necessary to evaluate the traffic quality for almost all 24 h of a day.

There are different models to calculate the delays caused by congestion. Deterministic or stochastic queuing models as well as a hydrodynamic flow model can be used. Independent from the preferred calculation model, an adequate duration of the time intervals has to be chosen. To minimize the effort for the user, 1-h intervals are proposed. However this interval length tends to underestimate the duration of congestion since the highest traffic flow is lower due to the averaging. Using a spreadsheet program it is possible to heterodyne the hourly volumes with a minute-by-minute flow pattern whose averaged hourly means represent the input volumes. This procedure, developed by Brilon and Zurlinden (2003), is shown in Figure 5.

The delay can be calculated for every interval. To obtain a comparable measure of effectiveness only delays due to congestion should be considered, since this measure is independent from the length of the segment.

As already indicated, some effects of congestion have to be considered. For each segment as well as each interval it has to be analyzed, if a queue exists. Beside the volume within the current interval the existence of a queue depends on the operation quality in the bordering segments as well as on the existence of a queue within the previous interval. Thereby temporal shifts of demand can be considered. The effects on the spatial dimension are regarded through the segments, indicated by the letter i, while impacts on the temporal dimension are implemented by evaluating several intervals, indicated by the letter t.

Depending on the quality of traffic flow in the upstream segments, the actual flow and capacity values (Equations 5 and 6) have to be used as input for the traffic flow model. The calculation of the queue length differs according to the chosen model.

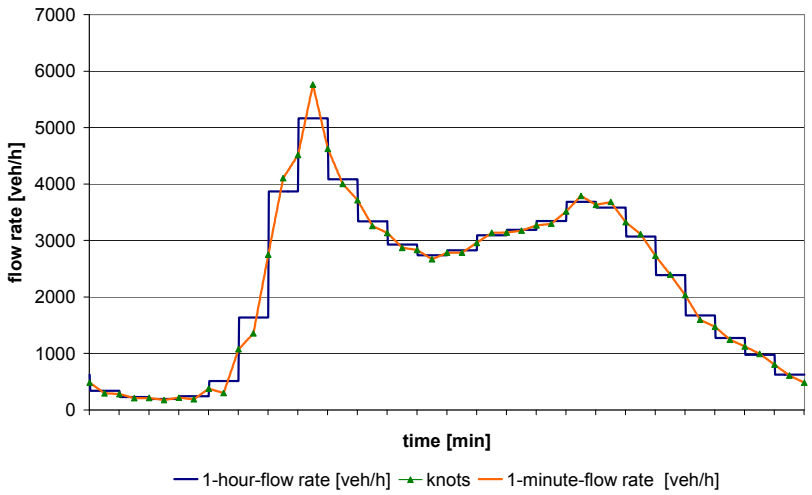


FIGURE 5 Development of demand patterns in 1-min intervals based on 1-h intervals.

Using the simple deterministic queuing model the number of queuing vehicles at the end of interval t is

$$N(i,t) = \begin{cases} 0 & N(i,t-i) = 0 \; ; \; q_{act}(i,t) - C_{act}(i,t) \le 0 \\ 0 & N(i,t-i) < 0 \; ; \; q_{act}(i,t) - C_{act}(i,t) \le 0 \\ N(i,t-1) + \left[q_{act}(i,t) - C_{act}(i,t)\right] \cdot T & else \end{cases} \quad (16)$$

where

$N(i,t)$ = queue length at the end of interval t on segment i (veh),
$N(i,t-1)$ = queue length at the end of previous interval t-1 on segment i (veh),
$q_{act}(i,t)$ = actual flow of the segment i within interval t (veh/h),
$C_{act}(i,t)$ = capacity of the segment i within interval t (veh/h), and
T = duration of the interval (h).

Applying the cinematic wave model, the length of the queue is given by

$$N(i,t) = -\frac{\Delta q(t)}{\Delta k(t)} \cdot (t - (t-1)) + N(i,t-1) \quad (17)$$

where

$N(i,t)$ = queue length at the end of interval t on segment i (km),
$N(i,t-1)$ = queue length at the end of previous interval t-1 on segment i (km),
$q(t)$ = flow rate (veh/h),
$k(t)$ = density (veh/km), and
t = time (h).

A transformed stochastic approach for a queue length equation can be adopted from queuing problems at road junctions under the assumption of random arrival and service at the bottleneck. The equation developed by Kimber and Hollis (1978) for the queue length as a function of demand, capacity and time is given in the appendix. Since the resulting functions of queue length hardly differ, the simple deterministic approach in Equation 16 is suggested.

The delay due to congestion again corresponds to the area below the function of queue length. It can be calculated by a numerical integration for each segment. All delays due to congestion are summarized to a total delay (for example over a whole day). Since this value is influenced notably by the demand pattern as well as the segment geometry, the absolute sum of delay does not seem appropriate as a comparable performance measure. With an increasing number of lanes the performance based on total delay becomes worse. In order to avoid this feature, the sum of delays has to be standardized to be comparable. There are some alternatives for this standardization.

Since the congestion effects should be evaluated, it seems reasonable to refer the delay to all vehicles that are affected by the congestion. However, using this standardized delay as evaluation measure showed that referring delay to affected vehicles can result in inconsistencies in cases of high daily traffic flow (Figure 6*a*). For example this issue occurs, if a second minor congestion arises additionally to a huge one, or both congestions merge. While the total sum of delays only marginally increases, the number of affected vehicles considerably rises. As a result the standardized delay is reduced despite the worse extent of congestion. Thus, the value used for the standardization has to be independent from volume as well as the duration of congestion. Additionally, segment characteristics that cause a capacity reduction have to be considered to allow a comparing evaluation of varying freeway sections. These conditions can be kept by using the capacity at 0% HV as one possible reference value. In this case the measure of effectiveness is defined as

$$vz_{cong}(i) = \frac{\sum_t VZ_{cong}(i,t) \cdot 60 \frac{\text{min}}{h}}{C_{SV=0\%} \cdot 24h} \tag{18}$$

where

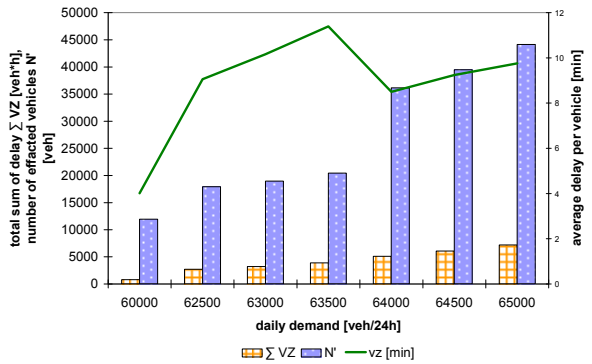
$vz_{cong}(i)$ = average standardized delay due to congestion on segment i (min/veh),

$\sum_i VZ_{cong}$ = total sum of delay due to congestion on segment i over the whole day (veh*h), and

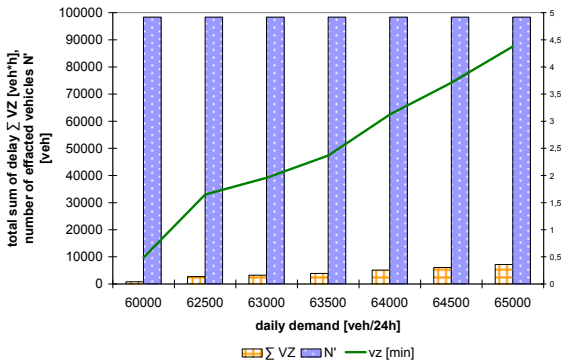
$C_{SV=0\%}$ = capacity of segment i assuming 0% heavy vehicle (veh/h).

With this standardization also two congestion periods within 1 day can be evaluated consistently and comparably (Figure 6*b*).

Depending on the standardized delay the level of service can be determined for every freeway segment. The evaluation based on a 24-h analysis offers the largest degree of information, since congestion effects are evaluated beyond the peak hour additionally. Beside the



a) Referring the total delay to affected vehicles



b) Referring the total delay to capacity at 0% HV

FIGURE 6 Total delay and standardized average delay depending on the daily traffic (example: two-lane carriageway, 1% gradient, 9% heavy vehicles).

resulting total standardized delay, which serves as service measure, the duration of congestions can be determined. This concept is mostly qualified to evaluate and compare oversaturated situations in a multidimensional way. Spatial as well as temporal effects of the congestion can be considered.

PRACTICAL CONCLUSIONS AND RECOMMENDATIONS

To conclude, it can be stated, that standardized methods were formulated which allow a differentiated evaluation of oversaturated flow conditions beyond a static consideration of traffic state. Although the fundamental diagram alone does not suffice for a differentiated evaluation of congestion, it builds the foundation for further approaches. Predominantly, the evaluation methods, developed within this research, can be applied for evaluating the quality of service on existing, congested facilities. In the scope of determining the priority for removals of bottlenecks, funds can be spent efficiently by the comparison of facilities and the quantification of improvement capabilities regarding operation performance.

Generally, the evaluation methods for congestion are much more sensitive due to uncertainties within the input data than those for undersaturated traffic flow conditions. Thus, a high quality of input data is needed. The application of software for the evaluation of oversaturated conditions is recommended. In consideration of the fact that level-of-service analysis in practice is mostly accomplished by using software, the 24-h analysis method should be short-listed in spite of its higher complexity. The obtained additional information justifies the effort.

RESOURCES

Banks, J. H. The Two-Capacity Phenomenon: Some Theoretical Issues. In *Transportation Research Record 1320*, TRB, National Research Council, Washington, D.C., 1991, pp. 234–241.

Baumgaertner, W. E. Levels of Service. Getting Ready for the 21st Century. In *ITE Journal,* Jan. 1996, pp. 36–39.

Bertini, R. L. *Congestion and Its Extent.* 2005, http://www.its.pdx.edu/pdf/bertini_congestion.pdf (17.03.2006).

Brilon, W. Traffic Flow Analysis Beyond Traditional Methods. In *Proc., 4th International Symposium on Highway Capacity*, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 26–41.

Brilon, W. and A. Estel. Differenzierte Bewertung der Qualitätsstufen im HBS im Bereich der Überlastung. Schlussbericht, Lehrstuhl für Verkehrswesen, Ruhr-Universität Bochum, 2007.

Brilon, W., J. Geistefeldt, and M. Regler. Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity. In *Transportation and Traffic Theory: Flow, Dynamics and Human Interaction*, *Proc., 16th International Symposium on Transportation and Traffic Theory*, Elsevier, Oxford, 2005.

Brilon, W., J. Geistefeldt, and M. Regler. Randomness of Capacity—Idea and Application. In *Proc., 5th International Symposium on Highway Capacity and Quality of Service*, Vol. 1, Yokohama, July, 2006, pp. 147–157.

Brilon, W., J. Geistefeldt, and H. Zurlinden. Implementing the Concept of Reliability for Highway Capacity Analysis. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2027*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 1–8.

Brilon, W., and M. Ponzlet. Application of Traffic Flow Models. In *Proceedings of the Workshop in Traffic and Granular Flow*, World Scientific, Singapore, 1995.

Brilon, W., and Zurlinden, H. Überlastungswahrscheinlichkeiten und Verkehrsleistung als Bemessungskriterium für Straßenverkehrsanlagen. Schriftenreihe Forschung Straßenbau und Straßenverkehrstechnik, Bonn, Heft 870, 2003.

Brilon, W., and H. Zurlinden. Kapazität von Straßen als Zufallsgröße. Straßenverkehrstechnik 4/2004, pp. 164–172.

Cameron, P. E. G3 F7—An Expanded LOS Gradation System. In *ITE Journal,* Jan. 1996, pp. 40–41.

Chen, C., A. Skabardonis, and P. Varaiya. Travel Time Reliability as a Measure of Service. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1855*, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 74–79.

Elefteriadou, L., R. P. Roess, and W. R. McShane. Probabilistic Nature of Breakdown at Freeway Merge Junctions. In *Transportation Research Record 1484*, TRB, National Research Council, Washington, D.C., 1995, pp. 80–89.

FGSV - Forschungsgesellschaft für Straßen- und Verkehrswesen, Hrsg. Handbuch für die Bemessung von Straßenverkehrsanlagen (HBS). Ausgabe 2001, Fassung 2005, Köln, 2005.

Federal Highway Administration (FHWA). Time Reliability: Making It There on Time, All the Time, 2006, http://www.ops.fhwa.dot.gov/publications/tt_reliability/brochure/index.htm (20.02.2006).

Geistefeldt, J. Ermittlung von Reisezeitverlusten auf Autobahnen unter Verwendung eines Ganglinienmodells. Diplomarbeit. Lehrstuhl für Verkehrswesen, Ruhr-Universität Bochum, 2003.

Geistefeldt, J. Schätzung von Reisezeiten auf Autobahnen unter Verwendung eines erweiterten Verkehrsflussmodells. Tagungsband zur HEUREKA ’05 – Optimierung in Verkehr und Transport. Forschungsgesellschaft für Straßen- und Verkehrswesen, Köln, 2005.

Geistefeldt, J. Verkehrsablauf und Verkehrssicherheit auf Autobahnen mit vierstreifigen Richtungsfahrbahnen. Schriftenreihe des Lehrstuhls für Verkehrswesen der Ruhr-Universität Bochum, Heft 30, 2007.

Hall, F. L., and K. Agyemang-Duah. Freeway Capacity Drop and the Definition of Capacity. In *Transportation Research Record 1320*, TRB, National Research Council, Washington, D.C., 1991, pp. 91–98.

Kimber, R. M., and E. M. Hollis. Traffic Queues and Delays at Road Junctions. *TRRL Laboratory Report 909*, 1979.

Kühne, R. D., and N. Anstett. Stochastic Methods for Analysis of Traffic Pattern Formation. In *Proc., 14th International Symposium on Transportation and Traffic Theory*, Jerusalem, Verlag Elsevier, 1999.

Lighthill, M. J., and G. A. Whitham. On Kinematic Waves, Part II. A Theory of Traffic Flow on Long Crowded Roads. *Proceedings Royal Society of London*, Vol. 229 A, No. 1178, 1955, pp. 317–345.

Lomax, T., S. Turner, and G. Shunk. *NCHRP Report 398: Quantifying Congestion. Volume 1, Final Report*. TRB, National Research Council, Washington, D.C., 1997.

Lomax, T., S. Turner, and G. Shunk. *NCHRP Report 398: Quantifying Congestion. Volume 2, User´s Guide*. TRB, National Research Council, Washington, D.C., 1997.

Lorenz, M. and L. Elefteriadou. A Probabilistic Approach to Defining Freeway Capacity and Breakdown. In *Transportation Research Circular E-C018: Proceedings of the 4th International Symposium on Highway Capacity*, TRB, National Research Council, Washington, D.C., 2000, pp. 84–95.

May, A. D. Performance Measures and Levels of Service in the Year 2000 Highway Capacity Manual. NCHRP 3-55 (4) Interim Report: Recommended Performance Measures and Levels of Service in the Year 2000 Highway Capacity Manual, 1996.

May, A. D. Performance Measures and Levels of Service in the Year 2000 Highway Capacity Manual. NCHRP 3-55 (4) Final Report, 1997.

Minderhoud, M. M., H. Botma, and P. H. L. Bovy. Assessment of Roadway Capacity, Estimation Methods. In *Transportation Research Record 1572*, TRB, National Research Council, Washington, D. C., 1997, pp. 59–67.

National Transportation Operations Coalition (NTOC): Performance Measurement Initiative. Final Report, Washington, 2005.

Okamura, H., S. Watanabe, S., and T. Watanabe. An Empirical Study on the Capacity of Bottlenecks on the Basic Suburban Expressway Sections in Japan. In *Transportation Research Circular E-C018: Proceedings of the 4th International Symposium on Highway Capacity*, TRB, National Research Council, Washington, D.C., 2000.

Persaud, B. N., and V. F. Hurdle. Some New Data That Challenge Some Old Ideas About Speed-Flow Relationships. *Transportation Research Record 1194*, TRB, National Research Council, Washington, D.C., 1988, pp. 191–198.

Ponzlet, M. Auswirkungen von systematischen und umfeldbedingten Schwankungen des Geschwindigkeitsverhaltens und deren Beschreibung in Verkehrsflussmodellen. Schriftenreihe des Lehrstuhls für Verkehrswesen der Ruhr-Universität Bochum, Heft 16, 1996.

Regler, M. Verkehrsablauf und Kapazität auf Autobahnen. Schriftenreihe des Lehrstuhls für Verkehrswesen der Ruhr-Universität Bochum, Bochum, Heft 28, 2004.

Schrank, D., and T. Lomax. The 2001 Urban Mobility Report. Texas Transportation Institute, 2001.

Schrank, D., and T. Lomax. The 2005 Urban Mobility Report. Texas Transportation Institute, 2005.

Shaw, T. *NCHRP Synthesis of Highway Practice 311: Performance Measures of Operational Effectiveness for Highway Segments and Systems*. Transportation Research Board of the National Academies, Washington, D.C., 2003.

Highway Capacity Manual. TRB, National Research Council, Washington, D.C., 2000.

Zurlinden, H. Ganzjahresanalyse des Verkehrsflusses auf Straßen. Schriftenreihe Lehrstuhl für Verkehrswesen, Ruhr-Universität Bochum, Heft 27, 2003.

APPENDIX

Queue Length at Road Junctions by Kimber and Hollis

The time depending queue length is determined by

$$Fn(x) = 0{,}5 \cdot \left(\sqrt{(Cx(1-g)+1)^2 + 4gCx} - (Cx(1-g)+1) \right)$$

where $l = \dfrac{g}{(1-g)}$

I. for $g \geq 1$: $\quad L(t) = Fn(t + t_0)$

where $\quad t_0 = \dfrac{L_0(L_0+1)}{C(g(L_0+1) - L_0)}$

II. for $g < 1$:

a) $0 < L_0 < l \quad L(t) = Fn(t + t_0)$

where $\quad t_0 = \dfrac{L_0(L_0+1)}{C(g(L_0+1) - L_0)}$

b) $L_0 = l \quad L(t) = l$

c) $l < L_0 \leq 2l \quad L(t) = 2l - Fn(t + t_0)$

where $\quad t_0 = \dfrac{(2l - L_0)(2l - L_0 + 1)}{C(g(2l - L_0 + 1) - (2l - L_0))}$

d) $L_0 > 2l \quad L(t) = \begin{cases} L_0 + \left(g - \dfrac{L_0}{(L_0+1)} \right) Ct & 0 \leq t \leq t_c \\ 2l - Fn(t - t_c) & t > t_c \end{cases}$

where $\quad t_c = \dfrac{(2l - L_0)}{C\left(g - \dfrac{L_0}{(L_0+1)} \right)}$

where

C = capacity and
g = saturation.

REFERENCE

1. Kimber, R. M., and E. M Hollis. Traffic Queues and Delays at Road Junctions. *TRRL Laboratory Report 909*, 1979.

Measurements and Characteristics of Traffic Flow

Measuring Traffic Flow Using Real-Time Data

B. G. HEYDECKER
J. D. ADDISON
University College London, England

The theory of traffic flow based upon speed, flow and density that vary only slowly in space and time is well established. However, matching field observations up to this theory and extracting estimates of quantities of interest is not always straightforward. Spatial density of traffic is not measured readily, and inductive loops are often used instead to measure the proportion ω of a sampling period for which a vehicle is present, which is known as occupancy: the relationship between occupancy ω and density k is $k = \omega / L$, where L is the mean effective length of a vehicle at the detector. According to this, correct interpretation of occupancy depends on the composition of the traffic that is measured, which will affect the value of L. Estimates of the capacity of a road are most useful when they are expressed in units that are independent of traffic composition. In this paper, we show how the value of L can be estimated from the 1-minute point observations of a kind that are available from the Highways Agency MIDAS data that are collected on the U.K. motorway network. The value of this quantity was found to vary substantially over time during the day, between lanes on the road, and according to the control status of the road thus reflecting variations in traffic composition, and variations in lane usage. The consequences of this are discussed for interpretation and use of traffic data of this kind in estimating the speed–density relationship, capacity and related properties of a road section.

Descriptions of traffic based upon the three quantities of speed (v), flow (q) and density (k) are well established, and analysis leads in the case of homogeneous traffic directly to the formula that has become known as the fundamental equation (*1*):

$$q = k\,v \tag{1}$$

where

q = flow, measured in vehicles per unit of time;
k = density, measured in vehicles per length of road; and
v = space-mean speed, measured in length per unit time (*2*).

This fundamental equation can be supplemented by adoption of a model relationship between speed and density of the form

$$v = f(k) \tag{2}$$

Whilst Equation 1 is an inescapable consequence of the definitions of the quantities that are involved, equations such as Equation 2 represent descriptive models of traffic behavior, which is based on an association between speed v and density k. Traffic flow can be analyzed effectively following on from these relationships (*2–5*).

Estimation of values of these quantities for use in models requires appropriate processing of measured data: in particular, traffic density cannot usually be measured conveniently but instead is estimated from other detector data that are readily available. Here, we consider how this can be undertaken and practical issues that arise in the process.

Measurement of the spatial density k of traffic is not always practical, as this requires instantaneous observation of a substantial length of road. Instead, as Hall (*6*) has pointed out, it is more usually estimated from the proportion ω of time for which a vehicle is present at a fixed detector, known as occupancy, which can be measured conveniently by repeated sampling of the state of the detector. The occupancy measured in this way provides a time-based estimate of the proportion of the road that is covered by vehicles. This is related to density k through the relationship

$$\omega = k\,L \tag{2}$$

where L is the mean effective length of a vehicle at the detector, representing the mean range of positions of a vehicle for which the detector will be occupied. According to this, proper interpretation of occupancy data from a detector requires knowledge of the mean effective vehicle length, which depends on the magnetic properties of vehicles, and the composition of traffic.

In the present paper, we investigate the relationship between occupancy ω and density k as revealed in traffic data from a U.K. motorway road. These will be required in order to interpret observations of occupancy for use in standard models of traffic flow. In particular, we develop different methods to estimate L and investigate their performance. This leads us to observations about typical values of L for each lane of the motorway in different traffic control conditions.

DATA

The data used for the present investigation are derived from the MIDAS (*7*) and associated Halogen data-logging systems. They represent measurements made between Junctions 10 and 16 on the 188-km M25 London Orbital Motorway in England. In this region, the road has four running lanes in each direction and carries up to 85,000 vehicles per day in each direction. The running lanes are numbered consecutively from 1 at the nearside (shoulder) to 4 at the offside (median). The rules of the road are that overtaking vehicles should pass to the offside of those being overtaken, and vehicles should return to the available lane that is furthest to the nearside as soon as safely possible (*8*). Under most circumstances, heavy goods vehicles and those drawing trailers are prohibited from using the most offside lane of a motorway (i.e., Lane 4 in the present case).

According to the traffic conditions, mandatory speed limits can be imposed at each of the national value of 70 mph, the reduced values of 60 mph and 50 mph that are used to manage flow under busy conditions, and the value of 40 mph that is used to slow traffic in advance of joining slow moving queues. The MIDAS system has arrays of detectors spaced at approximately 500-m intervals along the carriageway, with a separate detector in each lane. The data are aggregated over 1-minute intervals, each record containing

flow	the number of vehicles crossing the detector during that minute,
speed	the mean speed in km/h of vehicles crossing the detector, and
occupancy	the proportion of time that the detector is occupied.

The data used in the present analysis represent all of the available records for Wednesday, May 1, 2002, at location code 4757 on the clockwise (A) carriageway. Analysis of corresponding data from the same site on Wednesday, May 15, 2002, and from another Site 4747 between the same junctions on Wednesday, May 1, 2002 were found to yield similar results.

The number of observations available for analysis is reported in Table 1. There were over 1,000 minutes of the day during which the speed limit defaulted to the national one of 70 mph. By contrast, only 18 minutes of data were available for the 60-mph speed control condition. The number of available observations in each controls state varied substantially among lanes in the case of the 70-mph speed limit, with about half the number available in Lane 4 compared with Lane 1: this arises because in light traffic conditions such as occur late at night, the high availability of low-numbered lanes leads to Lane 4 being empty for extended periods of time.

Preliminary analysis of Equations 1 and 2 leads to the relationship between the quotient q/v and occupancy ω:

$$q / v = \omega / L \tag{3}$$

In order to investigate this correspondence, values of q/v were plotted against those of ω for each of the four running lanes. The results of this are given in Figure 1. According to Equation 3, the gradient of the chord from the origin to each observation corresponds to the reciprocal of the mean effective vehicle length L associated with that observation. The results in Figure 1 show that in each lane, there is a good relationship between these quantities while the occupancy is less than about 0.30, showing consistency in estimates of mean effective vehicle length in each lane. On the other hand, for occupancies above about 0.30, the relationship between q/v and ω becomes weak, showing substantial scatter. As can be seen from these plots, most of the observations of higher occupancy were made when the speed limit was set at 40 mph, indicating the presence of slow-moving queues.

TABLE 1 Number of Observations Available for Analysis

Speed Limit (mph)	Lane			
	1	2	3	4
70	1,015	893	733	541
60	18	18	18	18
50	97	97	97	97
40	239	239	240	241

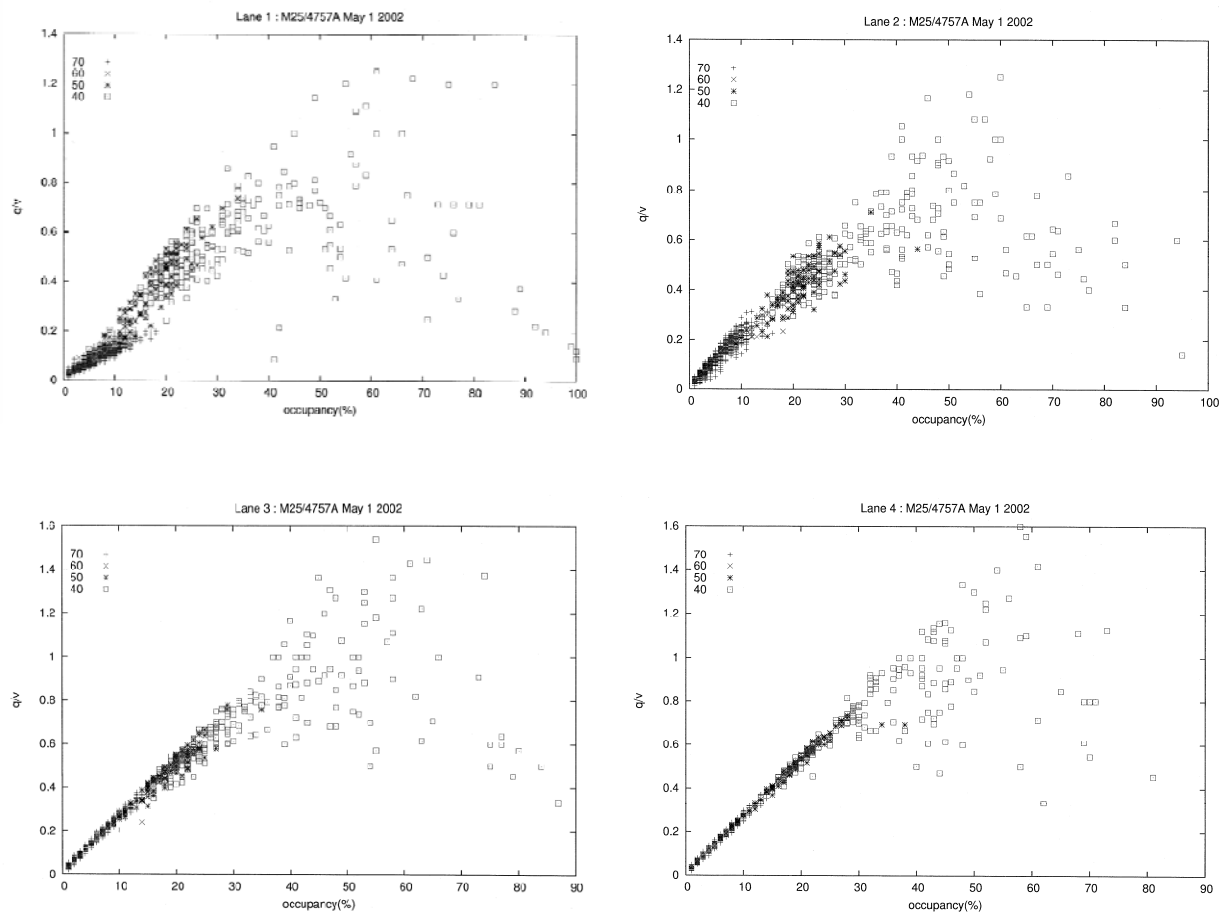


FIGURE 1 Plots of flow divided by speed against occupancy (q / v versus ω) for four lanes of the M25 motorway.

METHOD

Several methods are available to estimate values of the mean effective vehicle length L using speed, flow and occupancy data. These differ according to whether they provide estimates of L or instead they provide estimates of its reciprocal $\Gamma = L^{-1}$, and according to whether they arise as direct calculations of mean values or incidentally from statistical analysis. The key distinction between estimates based upon L and those based upon Γ arises from the weight associated with observations that deviate above and below the estimate: those that deviate substantially above the estimate of L will correspond to smaller deviations in the reciprocal space, in which case they will be carry less influence.

Here, we identify six distinct methods as follows, which are then tested using the M25 data and compared using their standard errors and their stability between circumstances.

DIRECT ESTIMATION

Equation 3 can be rearranged to give the expression

$$L = v\,\omega\,/q \tag{4}$$

according to which values of average vehicle length for each minute can be calculated directly from the observations of v, ω and q. The mean $\bar{L} = \frac{1}{n}\sum_{i=1}^{n} L_i$ of the resulting calculated values will yield an unbiased estimate of the mean vehicle length L. However, if the calculated value will be used to estimate traffic density k from observations of occupancy ω through the equation

$$k = \omega/L \tag{5}$$

then a more appropriate approach will be to calculate the average value of observations of the required quantity

$$\Gamma = q/(v\,\omega) \tag{6}$$

In each of these cases, the standard error of estimation σ_e can be calculated from the sample estimate of the variance $S^2 = \sum_{i=1}^{n}\left(L_i - \bar{L}\right)^2 \Big/ (n-1)$ of the data according to $\sigma_e = \sqrt{S^2/n}$, where n is the number of observations.

The advantage of direct estimation in this way is that it provides a methodology that is simple, transparent and independent of any model formulation, although it does depend on choice of quantity as between L and Γ.

REGRESSION ESTIMATES BASED ON OCCUPANCY

An approach to estimate values of L and Γ can be founded in regression analysis, and the variance of estimation quantified accordingly. First, we consider each observation of occupancy ω to consist of a systematic part corresponding to $L \times q/v$ together with an error ε that is measured on the scale of occupancy. Thus we can express

$$\omega = L\left(\frac{q}{v}\right) + \varepsilon \tag{7}$$

from which we can estimate L by regressing the observed values of ω on those of the quotient q/v : the fitted coefficient of q/v then provides the required estimate of L. Using the method of ordinary least squares (OLS) will minimize the sum of the squares of errors ε, and hence lead to minimum variance of estimation. Use of standard regression methods (*9*) also provides standard errors of estimation σ_L calculated as $\sigma_L = \sigma_e \Big/ \sqrt{S_{xx}^2}$, where $S_{xx}^2 = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ represents the sum of squared deviations of the explanatory variable $X = q/v$ from its mean.

As in the case of direct estimation, we may prefer a minimum variance estimate of Γ. This can be achieved by rearranging Equation 7 to

$$\left(\frac{q}{v}\right) = \Gamma\omega + \varepsilon \tag{8}$$

where the error ε is attached to observations of the quotient q/v rather than to the occupancy ω. In this case, Γ is estimated as the coefficient of ω in the regression of q/v.

The advantage of regression-based methods is that is provides a systematic approach to the estimation of the quantities of interest, L and Γ. The choice of errors ε on the scale of occupancy ω associates them with this macroscopic quantity and thus allows to some extent for deviations from model values that arise in partial observation of vehicles at the boundary of sample periods and when vehicles change lanes while they are passing over detectors.

REGRESSION ESTIMATES BASED ON FLOW

By rearranging Equation 3 between speed, flow and occupancy, we can express flow as

$$v\,\omega = L\,q + \eta \tag{9}$$

where η is an error that is measured on the scale of flow. Thus L can be estimated as the coefficient of q in a regression of the product $v\,\omega$.

Finally, this expression can be rearranged so that Γ can be estimated as a coefficient of the product $v\,\omega$ in the regression

$$q = \Gamma\,v\,\omega + \eta \tag{10}$$

These approaches share the advantage of regression-based methods with those described earlier. The choice in this case of errors η on the scale of flow q associates them with this quantity, which is observed in this context microscopically as a count of vehicles during the sampling period. This approach seems well suited to applications in which the focus will fall on estimates of flow and capacity.

SUMMARY

We have identified and presented six distinct methods for the estimation of the effective length L of vehicles at a detector for use in the interpretation of occupancy. Each approach can be formulated as estimation either of mean effective length L or of its reciprocal Γ. The direct methods embodied in Equations 4 and 6 use standard sampling approaches, but each relies on quantities derived from primary observations. The other approaches rely on regression analysis: methods 7 and 8 with errors on the scale of occupancy associated respectively with observations of occupancy and the quotient q/v, whilst methods 9 and 10 have errors on the scale of flow (hence weighting differently according to speed) associated respectively with observations of the product $v\,\omega$ and flow.

RESULTS

The six approaches described earlier to estimation of values of L for use in traffic flow modeling led to different estimates, with varying quality. In this section, we review and compare the results of these methods. In order to make comparisons between methods that estimate the length L explicitly and those that estimate its reciprocal Γ, we present the results in terms of length, calculating the estimate $L = \Gamma^{-1}$ with standard error $\sigma_L = \sigma_\Gamma / \Gamma^2$. These calculations were undertaken separately for each combination of lane and speed control condition.

DIRECT ESTIMATION

The results of direct estimation of L according to Equation 4 and that arising from direct estimation of its reciprocal Γ according to Equation 6 are presented in Table 2.

REGRESSION ESTIMATION ON THE SCALE OF OCCUPANCY

The results of regression estimation of L according to Equation 7 and that arising from regression estimation of its reciprocal Γ according to Equation 8 are presented in Table 3.

REGRESSION ESTIMATION ON THE SCALE OF FLOW

The results of regression estimation of L according to Equation 9 and that arising from regression estimation of its reciprocal Γ according to Equation 10 are presented in Table 4.

TABLE 2 Results of Direct Estimation of L and Its Reciprocal Γ

Lane	Speed Limit (mph)	Number of Observations	Estimate of L from Equation 4 (m)		Estimate of L as Γ^{-1} from Equation 6 (m)	
			L	Standard Error σ_L	L	Standard Error σ_L
1	70	1,015	11.76	0.09	10.94	0.09
	60	18	10.41	0.47	9.97	0.47
	50	97	8.27	0.17	7.99	0.17
	40	239	12.83	1.16	9.18	1.18
2	70	893	6.90	0.06	6.51	0.06
	60	18	8.63	0.36	8.39	0.35
	50	97	8.69	0.13	8.52	0.13
	40	239	11.24	0.54	9.60	0.55
3	70	733	5.76	0.03	5.62	0.03
	60	18	6.47	0.22	6.38	0.21
	50	97	6.57	0.05	6.54	0.05
	40	240	8.07	0.26	7.33	0.27
4	70	541	5.81	0.03	5.72	0.03
	60	18	5.81	0.16	5.71	0.16
	50	97	6.29	0.04	6.27	0.04
	40	241	7.39	0.20	6.87	0.20

TABLE 3 Results of Regression Estimation of *L* and Its Reciprocal Γ on the Scale of Occupancy

Lane	Speed Limit (mph)	Number of Observations	Estimate of L from Equation 7 (m)		Estimate of L as Γ^{-1} from Equation 8 (m)	
			L	Standard Error σ_L	L	Standard Error σ_L
1	70	1,015	11.98	0.38	12.45	0.48
	60	18	10.27	1.79	10.88	1.93
	50	97	7.62	0.29	7.79	0.23
	40	239	9.91	0.13	12.96	0.19
2	70	893	7.19	0.26	7.39	0.19
	60	18	8.65	1.05	8.94	0.94
	50	97	8.51	0.27	8.70	0.24
	40	239	10.29	0.13	12.24	0.17
3	70	733	6.10	0.21	6.13	0.13
	60	18	6.53	0.82	6.64	0.55
	50	97	6.56	0.25	6.60	0.16
	40	240	7.82	0.10	8.87	0.10
4	70	541	6.06	0.25	6.08	0.15
	60	18	6.20	0.82	6.25	0.52
	50	97	6.36	0.22	6.40	0.14
	40	241	7.31	0.10	8.02	0.08

TABLE 4 Results of Regression Estimation of *L* and Its Reciprocal Γ on the Scale of Flow

Lane	Speed Limit (mph)	Number of Observations	Estimate of L from Equation 9 (meters)		Estimate of L as Γ^{-1} from Equation 10 (meters)	
			L	Standard Error σ_L	L	Standard Error σ_L
1	70	1,015	11.87	0.08	12.34	0.13
	60	18	10.31	0.43	10.61	0.56
	50	97	7.71	0.10	7.91	0.09
	40	239	8.30	0.07	9.02	0.09
2	70	893	7.11	0.05	7.31	0.05
	60	18	8.63	0.25	8.83	0.27
	50	97	8.51	0.09	8.68	0.10
	40	239	8.86	0.08	9.40	0.09
3	70	733	6.08	0.04	6.11	0.03
	60	18	6.50	0.19	6.57	0.15
	50	97	6.54	0.08	6.57	0.06
	40	240	6.95	0.07	7.20	0.06
4	70	541	6.04	0.04	6.06	0.03
	60	18	6.17	0.19	6.19	0.14
	50	97	6.28	0.07	6.30	0.06
	40	241	6.58	0.06	6.80	0.05

DISCUSSION OF RESULTS

According to these results, the mean effective length of vehicles varies substantially between lanes and between speed control conditions. As well as this, differences arise in the estimates among the different approaches (estimation of L and of Γ) and different methods (direct, or regression analysis). The precision of the estimates varies as well, both among speed control conditions and among estimation approaches and methods. We discuss these aspects in turn.

All of the results presented in this section show that the mean effective vehicle length varies among lanes of the motorway and according to the traffic control state. There is a consistent tendency under each traffic control condition for the mean effective vehicle length to be less in the higher numbered lanes (i.e., those farther to the offside). This reflects the tendency for long and heavy vehicles preferentially to use the nearside lanes.

The effects of variations in speed limit condition differ among the lanes. In Lane 1, at the nearside shoulder, the mean effective vehicle length decreases as the speed limit is reduced from the national limit of 70 mph to 50 mph, with an increase in mean effective length when the speed limit is reduced from 50 mph to 40 mph. On the other hand, the mean effective length of vehicles in the other lanes tends to increase as the speed limit decreases, suggesting a migration of longer vehicles from lower to higher number lanes as the speed limit is reduced.

Figure 2 shows the cumulative distribution of 1-minute observations of effective vehicle lengths calculated for Lane 1 under different speed control conditions. This shows a substantial proportion of measurements that correspond to long vehicles under the 40-mph speed control condition. Figure 3 shows that the distribution of effective vehicles lengths under the national speed limit of 70 mph is almost identical between Lanes 3 and 4, but varies among the other lanes. However, the distribution of effective vehicle lengths seems to be influenced by some artifacts because the maximum permitted physical length of an articulated truck on U.K. roads without special permission is 16.5 meters, which corresponds only to the 85th percentile of the distribution in Lane 1 under 40-mph control conditions.

The different approaches to estimation—either as direct estimation of L or as its reciprocal Γ—lead to different values. In the case of direct estimation, the results given in Table 2 show that estimation of L leads to greater values than does estimation through the reciprocal.

By contrast, when regression methods are adopted, estimation of L as opposed to its reciprocal Γ leads to smaller values.

Comparison of the results of the three different methods (direct estimation and the two regression formulations) shows that the estimates of L vary substantially among them. Notable differences arise under 40-mph control conditions: the direct method 4 leads consistently to in estimates of L that are larger than those of the other methods (by up to 2 to 3 meters), whilst the regression method 8 with errors on the scale of occupancy leads consistently to estimates of L that are larger than those of the other methods.

The standard errors of the estimates shows that the mean effective vehicle length L can be estimated with greater precision under the national (70 mph) and 50-mph speed limit conditions than under other conditions. In the case of the 60-mph speed limit condition, relatively few observations were available so the estimated variation was not reduced substantially due to the sample size. However, despite the substantial sample size of 239 or greater for the 40-mph speed limit condition, the inherent variability in traffic under this

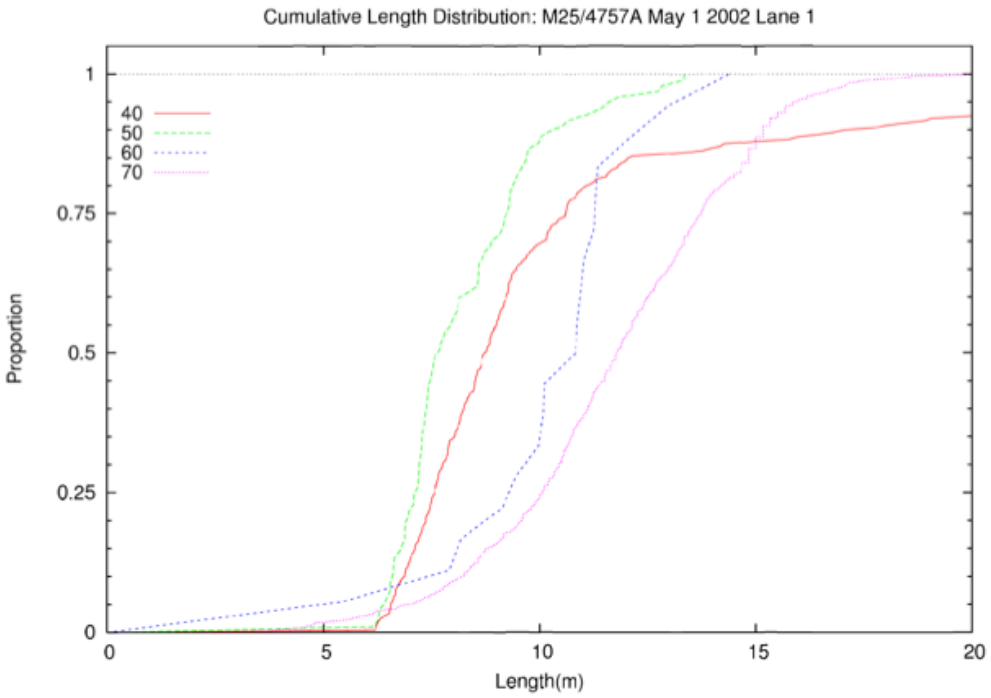


FIGURE 2 Distribution of effective length of vehicles in Lane 1 according to speed control condition.

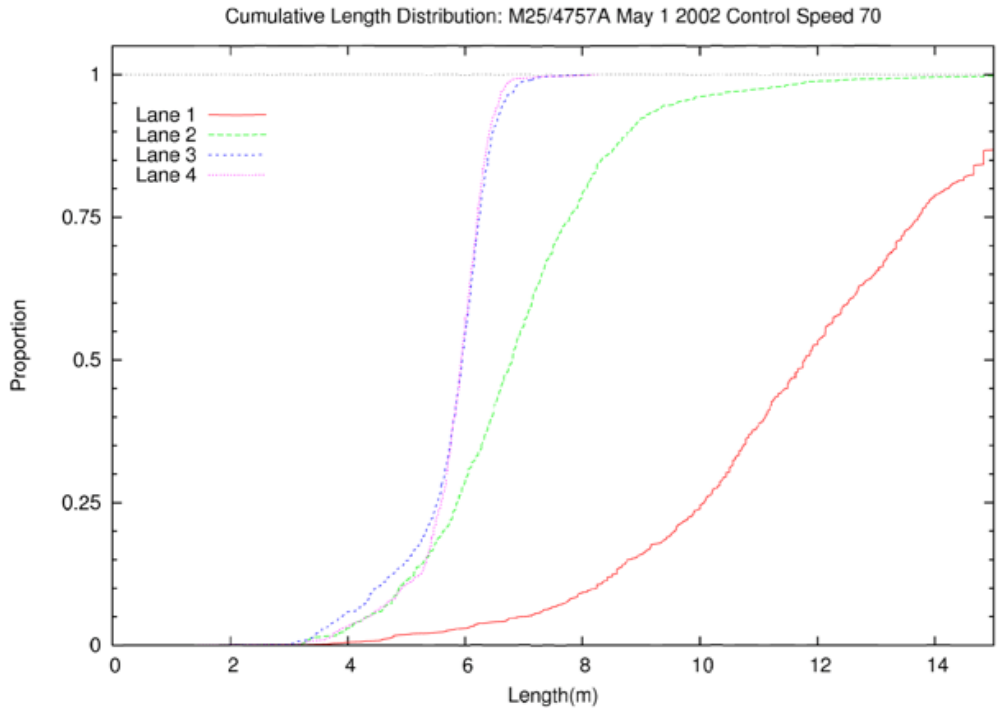


FIGURE 3 Distribution of effective length of vehicles under national speed limit according to lane.

condition led to large standard errors in direct estimation, especially in Lane 1. This seems to be due to the presence of a few observations with large values of L. The standard errors for estimates of L resulting from each of the regressions with errors on the scale of flow that are presented in Table 4 were generally small, and in most cases less than those of the other methods.

CONCLUSIONS

Measurement of detector occupancy as a proxy for traffic density has substantial currency, and confers substantial practical advantages of cost and convenience. In order to estimate traffic density from observations of detector occupancy, an estimate of mean effective vehicle length L is required. The analysis presented here shows that the value of L varies substantially among lanes and among traffic control conditions, and that there is no unique method to estimate it. Because this quantity is generally used to estimate values of density k from observations of detector occupancy ω using the relationship $k = \omega/L$,estimates of the reciprocal $\Gamma = L^{-1}$ are of interest as opposed to those of L itself.

The present research has considered six different approaches to the estimation of L, including two direct ones and four regression ones. From the point of view of consistency of estimates and precision of the values, the method preferred on the basis of the results presented here is the regression (10) of flow q on the product of speed and occupancy $v\,\omega$, which has errors on the scale of flow and yields an estimate of Γ as a coefficient. This approach automatically weighs observations according to the prevailing speed, so that observations made in low-speed conditions under which detectors might not operate as intended have reduced influence.

The finding that the mean effective vehicle length varies according to traffic conditions gives some cause for concern in the interpretation of traffic data. This results partly from the variation over time of the composition of traffic, and could arise because of heavy goods vehicle operators avoiding busy motorway roads at their busiest times, and partly from properties of the detector system. A consequence of this is that capacity of the road might be better described in terms of occupancy multiplied by speed than flow of vehicles, measured as density multiplied by speed. This will be the topic of further research.

ACKNOWLEDGMENTS

The work reported here was undertaken as part of a project funded by the U.K. Engineering and Physical Science Research Council under research grant EP/D 506840, and was supported by the U.K. Highways Agency.

REFERENCES

1. Greenshields, B. D. A Study of Traffic Capacity. *Highway Research Board Proceedings*, Vol. 14, 1935, pp. 448–477.
2. Wardrop, J. G. Some Theoretical Aspects of Road Traffic Research. In *Proceedings of the Institution of Civil Engineers Part II*, 1952, pp, 325–378.

3. Lighthill, M. J, and G. B. Whitham. On Kinematic Waves: II. A Theory of Traffic Flow on Long Crowded Roads. In *Proceedings of the Royal Society*, Vol. 229A, 1955, pp. 317–345.
4. Leutzbach, W. *Introduction to the Theory of Traffic Flow*. Springer, London, 1988.
5. Daganzo, C. F. *Fundamentals of Transportation and Traffic Operations*. Pergamon, London, 1997.
6. Hall, F. L. Traffic Stream Characteristics. In *Traffic Flow Theory: A State-of-the-Art Report* (N. H Gartner, C. J. Messer, and A. Rathi, eds.), Transportation Research Board of the National Academies, Washington, D.C., 2001.
7. Highways Agency. Motorway Incident Detection and Automatic Signalling (MIDAS) *Design Standard* TD 45/94. In *Design Manual for Roads and Bridges*, 9(1)2, Department for Transport, London, 1994.
8. Department for Transport. *The Highway Code*. The Stationary Office, London, 2007.
9. Petruccelli, J. D, B. Nandram, and M. Chen. *Applied Statistics for Engineers and Scientists*. Prentice Hall, London, 1999.

Airborne Traffic Flow Data and Traffic Management

MARK HICKMAN
PITU MIRCHANDANI
ATLAS Center, University of Arizona

Imagery collected from airborne platforms has long been used to document the evolution of traffic congestion over time and across extended areas. The authors have documented such published research dating all the way back to 1927, and these activities continued in various periods of activity up through the late 1980s. Beginning in the late 1990s, many important ideas, new and old, on the use of airborne imagery for traffic analysis were investigated and tested in the field by a number of research teams around the world. Through detailed image processing techniques, this imagery can be used to automatically determine traffic flow measures. Prototype software tools have been developed to automatically estimate queue lengths at intersections, to estimate vehicle speeds, and to estimate vehicle flows and densities. Similar research has integrated traffic flow data from airborne imagery into formal data collection programs, where the data from the imagery is fused with ground detector data to enhance the estimates and forecasts of traffic flows. Some of these research teams have also "architectured" approaches to automatically, in real-time, georeference images from remote cameras for managing traffic. This is done by integrating the imagery with information on the height, location, and orientation of the camera. Using these camera data, in combination with a geographic representation (latitude–longitude) of the area to be studied, lead to an explicit way to georeference the road and vehicle locations. Absolute values of vehicle positions, speeds, accelerations, decelerations, and lane changes can be determined. There have been a wide variety of experiments in this area; these tests have extracted individual vehicle trajectories from digital video imagery. Uses of the obtained airborne traffic data are many: (a) off-line analyses can be used for transportation planning purposes, (b) online, real-time analyses can be used for traffic management, especially in areas where ground sensors are not available and for large-scale incidents, and (c) one can develop data sets of vehicle trajectories for use in the calibration–validation of micro-, meso-, and macroscopic traffic flow models.

Measuring traffic flow and traffic behavior has been a subject of much research and practical application over more than 80 years. Interestingly, aerial photogrammetry and other methods of using imagery as a means of measuring traffic flow and traffic behavior have been practiced throughout this history. This is due in part to the common need to visualize traffic conditions and the challenges of capturing the movement of vehicles both in time and in space.

Airborne imagery has a number of major advantages over other forms of traffic measurement:

- The sensor itself can move and, even motionless, can capture a broader spatial extent of vehicle movements than traditional sensors at designated points in the road network.
- Because the sensor can move quickly from one area to another, or can move with the traffic flow, a single sensor can cover several locations.
- The sensor itself is managed remotely, without the need to disrupt traffic to maintain and operate the sensor. In addition, the sensor generally can be operated without any major effect on driving behavior in the observed traffic.

- The imagery can provide a wide variety of data on traffic flow and traffic behavior. This can improve the cost-effectiveness of data collection. Also, if the imagery is stored, it can be re-examined later for the same or other purposes.

Perhaps without saying, the primary disadvantage of airborne imagery is its cost: the need for stable imagery, over long periods of time and over larger geographic areas, requires aircraft and sensor platforms that can be quite expensive. Moreover, because of both the cost and aircraft requirements, the temporal coverage can be relatively short, in comparison to continual (24-hour) measurements from other types of traffic sensors.

Much can be said, however, for the possible integration of both airborne imagery and other more typical ground-based (or location-specific) traffic sensors, in achieving a more complete understanding of traffic flow. The combination of the persistence of measurement at a single location, with the broader spatial coverage provided by airborne imagery, can lead to a better understanding of both the spatial and temporal aspects of traffic flow and traffic behavior. Such integration of sensor data also illustrates the unique value added through the data from airborne imagery.

This paper provides a brief review of some of the major historical developments in the application of airborne imagery for traffic flow measurement. It also reviews some of the major recent developments, through the use of automated image processing techniques. To this end, the paper provides a review of data collection methods and applications in macroscopic traffic behavior, examining flows, speeds, densities, queuing, and related measures. It also provides a review of applications of airborne imagery for microscopic traffic flow analysis, specifically in the trajectories of individual vehicles. Finally, some thoughts on the future of airborne imagery in traffic data collection are provided in the last section.

MACROSCOPIC TRAFFIC BEHAVIOR

One of the earliest studies in which aerial photography was used to understand traffic flow is described about 80 years ago (Johnson, 1928). The study examined traffic along the one existing road between Baltimore, Md. and Washington, D.C. It included 127 aerial photographs, overlapping by about 50%, taken by a single airplane flying about 3,600 ft above ground. Curiously, the study also included traffic counts taken at four stations along the road, and six vehicles with a white sheet on the top were placed in the traffic stream (using what we now call the "floating car" technique) so as to be clearly visible in the aerial photographs. Average speeds of 20 to 30 mph were recorded by these floating cars. Johnson also used the imagery to determine the "density" of vehicles, taken as the number of vehicles observed in every quarter mile along the roadway. From these data, observations of vehicle clearance (space between vehicles, C, in feet) and velocity (V, in mph) were plotted on a log-log graph, shown in Figure 1*a*. The relationship was estimated to be

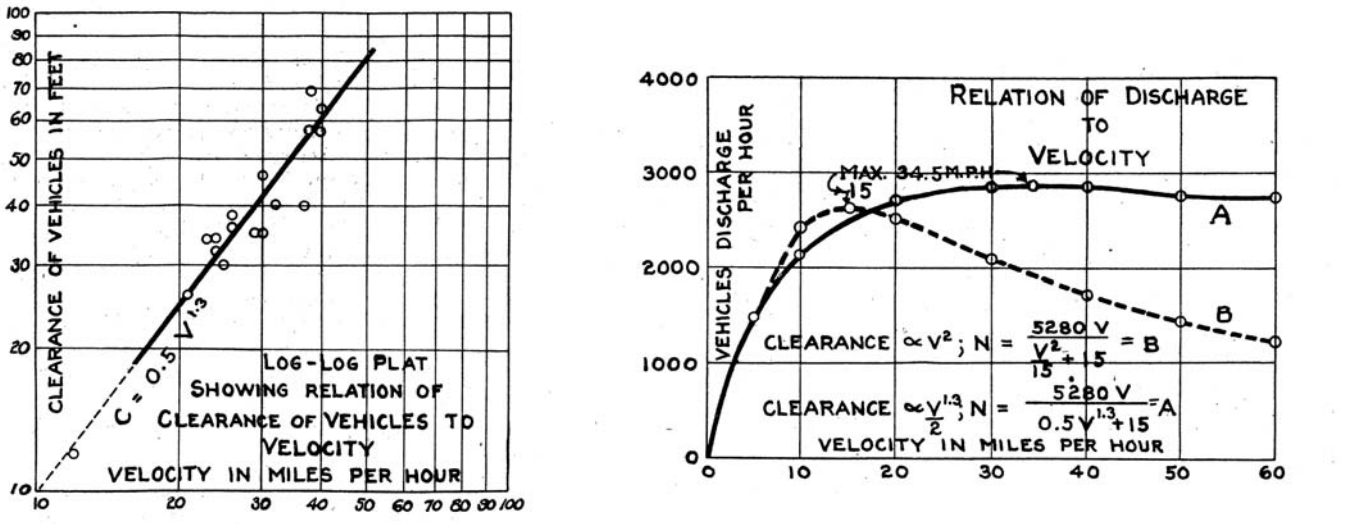
$$C = 0.5 \cdot V^{1.3}$$

In this way, it appears that Johnson actually was hypothesizing a relationship between vehicle density and speed. He further noted that the average flow rate, assuming a 15-ft vehicle length, is given by

$$N = \frac{5280 \cdot V}{C + 15} = \frac{5280 \cdot V}{0.5 \cdot V^{1.3} + 15}$$

where N is the flow rate. In this case, Johnson was using the fundamental relationship of traffic flow. What results from this relationship is a very curious speed–flow curve, where flow increases almost directly with speed when speeds are low, but the flow levels off as speed gets higher (over 25 mph). This is shown in Figure 1*b*.

Another set of studies examining the value of aerial photography for describing traffic behavior is given by none other than Greenshields (1947). He used a blimp flying at about 2,000 ft above ground to record traffic movements through a motion picture camera, covering a field of view of about ½ mile in each direction (1-mile square). Additional tests with a camera taking images every 2 to 5 seconds were conducted, from both a helicopter and an airplane. He concluded that "[H]igh altitude haze, shadows of buildings, trees, and the movement of the blimp are difficulties which, it is believed, are overshadowed by the complete and accurate record of all that happens within the area studied" (p. 291) and, "It is felt that several hours of such observations will reveal more than days of less complete data. From this standpoint it could well be that aerial photographs will prove comparatively cheap" (p. 297). This paper clearly indicates that Greenshields himself had strong and positive opinions on the potential value of airborne imagery for traffic analysis.



(*a*) (*b*)

FIGURE 1 Measures of (*a*) clearance versus speed, and (*b*) flow versus speed (Johnson, 1928).

A variety of studies since then have documented the use of aerial photography for macroscopic traffic analysis. Indeed, many of the more famous researchers in traffic flow theory have made extensive use of aerial photography, including Greenshields, Forbes (Forbes and Reiss, 1952), and May (Wagner and May, 1963). In the last paper, May introduces the "traffic density contour map," shown in Figure 2, that illustrates the spatial and temporal nature of congestion, as measured directly through the density from the aerial photography. Other examples of macroscopic flow studies with airborne imagery include Rice (1963), Jordan (1963), Cyra (1971), and Makigami et al. (1985), among several others in the academic literature.

To this day, much of the analysis of airborne imagery and aerial photographs has relied on manual analysis of the imagery. In addition to the previously mentioned studies, more comprehensive studies of traffic flows, densities, speeds, travel times, intersection delay, freeway congestion and level-of-service, origin–destination flows, and special event traffic management have been documented. Firms such as Skycomp (2008) make a full business out of the use of aerial photography for traffic measurement and analysis, for a variety of clients in the United States and Canada (e.g., MAG, 2000). A more extensive review of private sector involvement in the market for aerial photography is given in Murray (2002).

In addition, there is a fairly extensive literature which documents the analytic methods for such manual analysis. A more technical treatment of the subject is given in Peleg et al. (1973); a more recent summary is given in Mirchandani et al. (2002), Angel and Hickman (2002b), and Angel et al. (2002, 2003). In addition, specific methods for freeway level of service, arterial travel times, and intersection analysis are given in Angel and Hickman (2002c), Angel and Hickman (2002a), and Angel and Hickman (2003), respectively.

It is important to note here that most research and practical traffic studies have used fixed-wing airplanes, helicopters, or blimps as the means to carry personnel and photographic equipment, which in itself can be quite expensive. Only a few small studies to date have

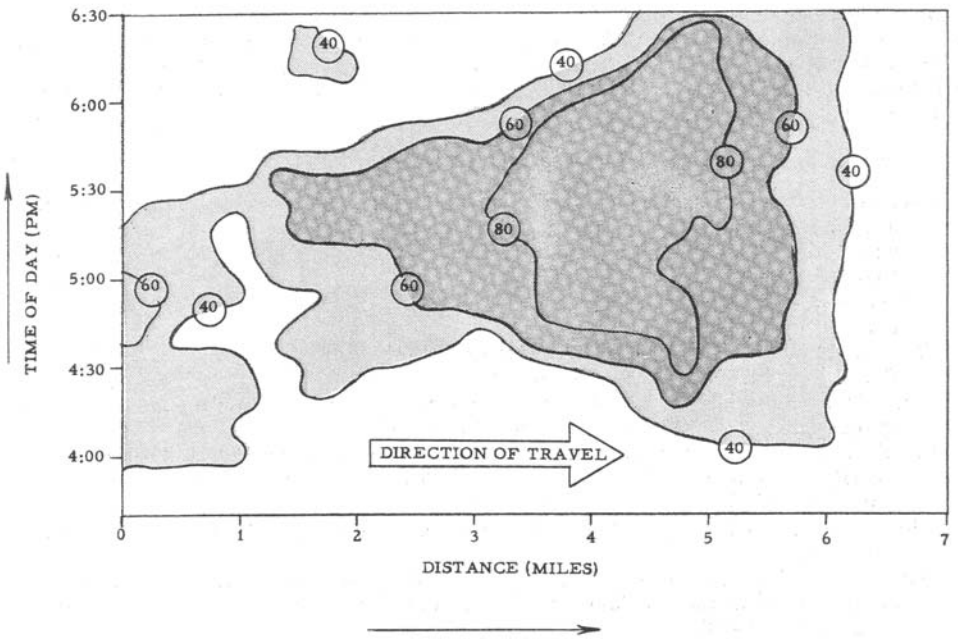


FIGURE 2 Traffic density contour map (Wagner and May, 1963).

examined the use of smaller, lightweight, low-cost unmanned aircraft systems (UASs, formerly unmanned aerial vehicles or UAVs) to record traffic imagery (e.g., Coifman et al., 2004 and Coifman et al., 2006). These studies have shown the potential use of UASs to collect data for freeway density measurement, queue measurement at intersections, origin–destination studies, and even parking lot utilization.

Finally, one may observe that there has been relatively little automation of these analysis methods for macroscopic traffic analysis. Most researchers have focused instead on the automation of detection and tracking of individual vehicles (described in detail in the next section). Of course, automated methods for microscopic traffic analysis can also provide data, at an aggregate level, that can be used for such studies. As an example, however, Agrawal and Hickman (2002, 2003) describe an automated technique to estimate queue lengths at signalized intersections, using some common image processing methods.

MICROSCOPIC TRAFFIC BEHAVIOR

At the simplest level, some research has examined the movement of individual vehicles, in the hopes of capturing some basic properties of traffic flows through the revealed behavior of specific vehicles. At present, airborne data collection provides one of only a couple proven methods of capturing such movements, as the data collection requires spatial and temporal persistence, and a relatively high sampling rate (up to 100% of vehicles) of vehicle movements on the ground.

There are many common issues that arise with airborne imagery that need to be addressed in this case. Briefly, these challenges include the following:

- **Correspondence**. Correspondence refers to the ability to identify one vehicle in an image as the same vehicle in a second, temporally and spatially separated, image. Depending on the frequency with which images are sampled, vehicles can remain stationary, or can move very short or very long distances, between images. The challenge is to find some method, manual or automated, that can match the same vehicle between these images. The use of different matching algorithms to determine vehicle speeds on a freeway, especially for on-line, real-time applications, is discussed by Chandnani and Mirchandani (2002). Another interesting study of the correspondence problem, using truck movements in and around a truck terminal, is described by O'Kelly et al. (2005).
- **Projection**. The vehicles identified in the imagery must somehow be related to specific locations in the real world, so that their positions and movements can be correctly interpreted. This usually requires some way of controlling for the movement of the airborne platform relative to the fixed ground. It also requires some method to project the location of vehicles in the image to actual coordinates of the vehicle on a specific roadway, and even in a specific lane on that roadway. This latter step of projection will commonly require knowing the scale of the image, so as to project distances measured in the image (e.g., millimeters or pixels) into ground measures (e.g., meters). In addition, as one alternative, this may be completed through more specific measurements of locations on the ground where aerial reconnaissance is expected. A second method uses specific ground control points to register the imagery to the real-world coordinates, using fairly standard image transformation techniques. In yet other cases, detailed camera information (x, y, z, roll, pitch, yaw) can be used to project the captured image onto some topographic map, allowing direct interpretation of vehicle locations in the image. A final step in

projection involves relating the vehicle position to specific roadway characteristics. For example, it becomes necessary to capture not only the location of the vehicle on the ground, but also its position relative to the lanes on a given roadway. In automated techniques, this step may make use of a "road mask" or high-resolution map to connect the vehicles to specific lanes.

In spite of these challenges, there remains a great need has been to establish traffic data at the microscopic level—at the point where the movements of specific vehicles, and hence the behavior of those vehicles' drivers, can be revealed. Such data commonly refers to individual vehicle movements on a roadway, or in a specific lane of traffic. To represent these movements, a time-space diagram is also commonly used, indicating the position of the vehicle in the direction of travel in the lane of traffic versus the time at which the vehicle is seen at that position. The data from a time–space diagram can indicate vehicle speeds, accelerations, decelerations, lane changing, car-following behavior, and many other driver behaviors of interest. When microscopic data can be collected from fixed cameras (e.g., NGSIM, 2008), the spatial coverage tends to be somewhat limited at 0.25 to 1.0 mile, although the temporal coverage can be fairly rich, covering 30 to 60 minutes at a time.

The original groundwork for such microscopic studies, using airborne traffic data, was completed at the Ohio State University in the late 1960s and early 1970s by Joseph Treiterer and colleagues (Treiterer and Taylor, 1966 and Treiterer, 1975). A helicopter was used to track vehicle movements on a roadway in Columbus, Ohio. Vehicle positions at various points in time were recorded through a high-resolution camera flying above the roadway in a helicopter. Ground control points were used to register (project) the images to specific coordinates in the real world, using a linear transformation of coordinates. Images were taken at a frequency of higher than one image per second, allowing fairly simple, but manually tedious, identification (correspondence) of the same vehicle in multiple images. An example of a resulting time–space diagram with individual vehicle trajectories is shown in Figure 3. Treiterer's studies led to more extensive understanding of microscopic traffic behavior, and clearly serve as a benchmark in the analysis of traffic flow.

Perhaps because of the cost of data collection and the challenge of data reduction from extensive sets of imagery, there was only limited work in this area until the late 1990s. In the mid-1980s, the U.S. Federal Highway Administration commissioned a study to generate microscopic traffic data for a number of freeway sections, using aerial photography from a fixed-wing aircraft (Smith and Roskin, 1985). The images were digitized manually, but generated several large data sets for microscopic traffic analysis.

Since then, a number of research groups around the world have experimented with automated methods to reduce vehicle trajectories from digital imagery, usually in the form of digital video. Video provides imagery at very high frequency (up to 30 frames per second) and, in many cases, at a high level of resolution with reasonable field of view (400 to 800 meters) in order to capture vehicle movements.

Through the work of the *National Consortium on Remote Sensing in Transportation–Flows*, the University of Arizona has been investigating the use of digital video taken from a helicopter to derive vehicle trajectories. This work was begun primarily through the work of Shastry (2002) and summarized by Shastry and Schowengerdt (2002). Their method relies heavily on existing digital image processing techniques. Processing the digital video, the individual frames are first registered to an original frame. Proceeding one frame at a time, consecutive images are then subtracted from each other. As long as vehicles are moving, small

frame-to-frame movements are then detected as blobs in the difference image; this directly solves the correspondence problem. The vehicle movements are projected into a relative world through a simple linear scaling technique, converting pixels to feet on the ground. In this case, however, the actual coordinates of the vehicles in the real world are only relative, but they do provide some sense of vehicle speeds, accelerations, and decelerations.

Subsequent enhancements of this same image processing technique are described in Kadam (2005). Furthermore, the use of the technique to determine vehicle trajectories has been outlined by Hickman and Mirchandani (2006), and is summarized in Figure 4. The video image processing step involves registering the imagery to a single point of view, correcting for the movement of the helicopter. Vehicle detection and tracking involves identifying interesting features of appropriate dimension in the image, and finding the correspondence between features from one image to the next. Generation of vehicle trajectories then requires knowing how to project the image into relative or absolute ground coordinates, using the image scale (e.g., meters per pixel) and some knowledge of the roadway and lane positions.

Over the same time, a team of researchers at the Technical University of Delft was developing a similar system (Hoogendoorn et al., 2003; Schreuder et al., 2003; Hoogendoorn and Schreuder, 2005; and Nejadasl et al., 2006). Their technique likewise uses digital video collected from a helicopter, with the objective of capturing microscopic traffic data. Their method differs from the Arizona approach in that it specifically uses an automated technique to generate and match ground control points in the imagery, in order to register the imagery. Vehicle detection is done by first generating a "background" image (the roadway without vehicle traffic), and vehicles are detected as differences from this reference image. Correspondence of vehicles from one image to the next is also achieved through a coarse matching algorithm, made easier since the imagery is taken at relatively high frequency. Finally, the conversion of vehicle coordinates in the image to vehicle coordinates in the real world appears to have been done manually, through specific lateral and longitudinal references on the roadway. More recent research by this same team has examined finer resolution of the vehicle trajectories through the use of a Kalman filtering technique (Hoogendoorn and Schreuder, 2005).

Both the Arizona and the Delft researchers have been examining microscopic traffic characteristics from the perspective of off-line analysis. Another major research effort has examined on-line methods of microscopic traffic analysis, led by researchers at the German Aerospace Center (DLR, Deutsches Zentrum für Luft- und Raumfahrt), and published by Ernst et al. (2003), Hipp (2006), Kühne et al. (2007), and Ruhé et al. (2007).

The experiments at the DLR have involved a variety of different aircraft, including helicopters, airships (zeppelins), and fixed-wing airplanes. A similar element in these aircraft is the data collection platform, called the Airborne Traffic Analyzer, or ANTAR. The on-board equipment includes a high-resolution digital camera that can be pointed in any direction, and an inertial measurement unit (IMU). The digital camera system is capable of capturing five images per second, with the restriction on the frame rate being dependent on the communication between the camera and the computer used to collect and forward these images. The various sensor platforms also include some combination of GPS antennas or differential GPS receivers, a separate wireless data downlink to transmit imagery to the ground, and/or a camera-mounted wireless downlink (Ruhé et al., 2007). The imagery from the camera is then georeferenced on the ground using a commercial road network database. With up to five frames per second, the location and speed of vehicles in each frame can be easily determined. In addition, the data from

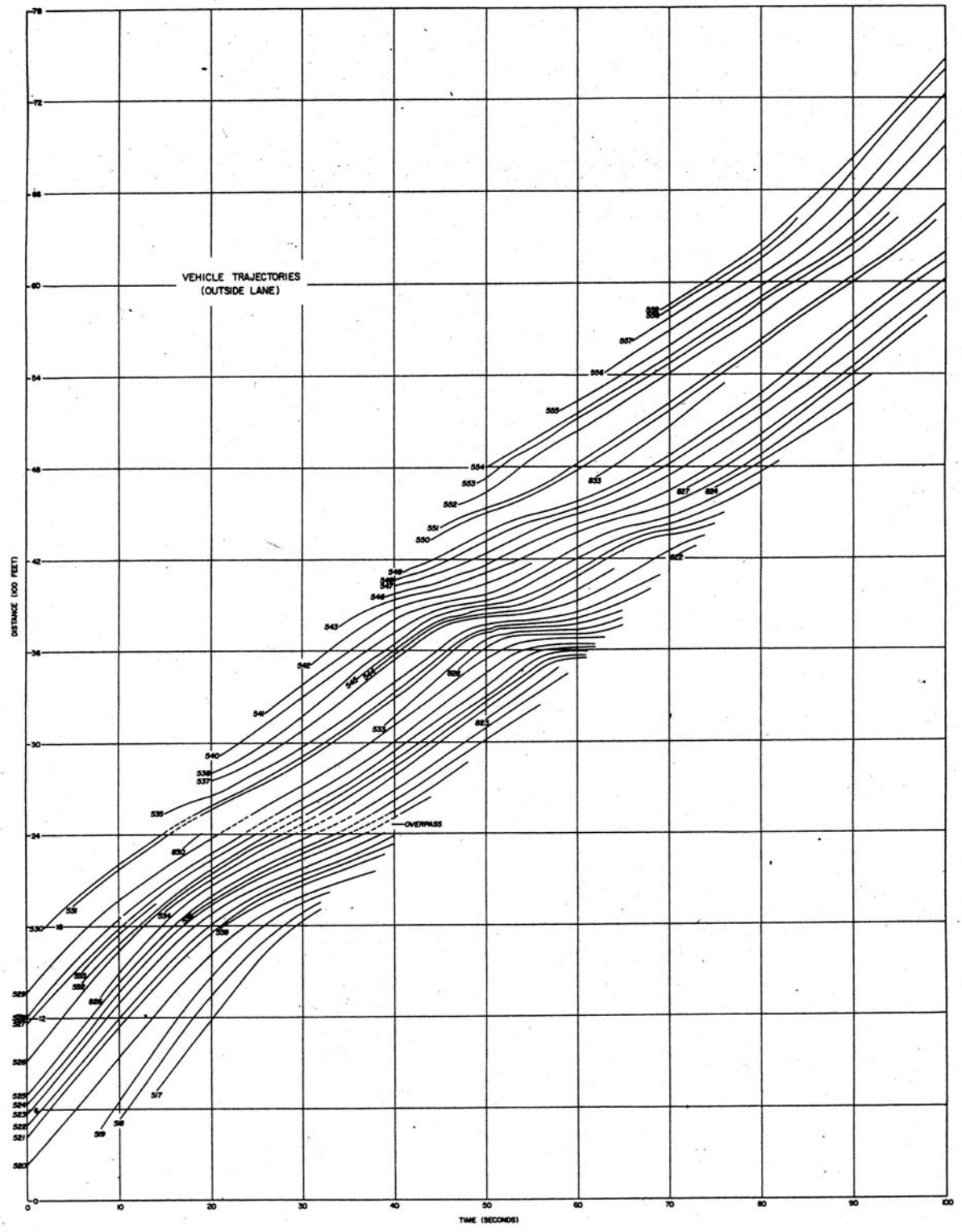


FIGURE 3 Vehicle trajectories (Treiterer and Taylor, 1966).

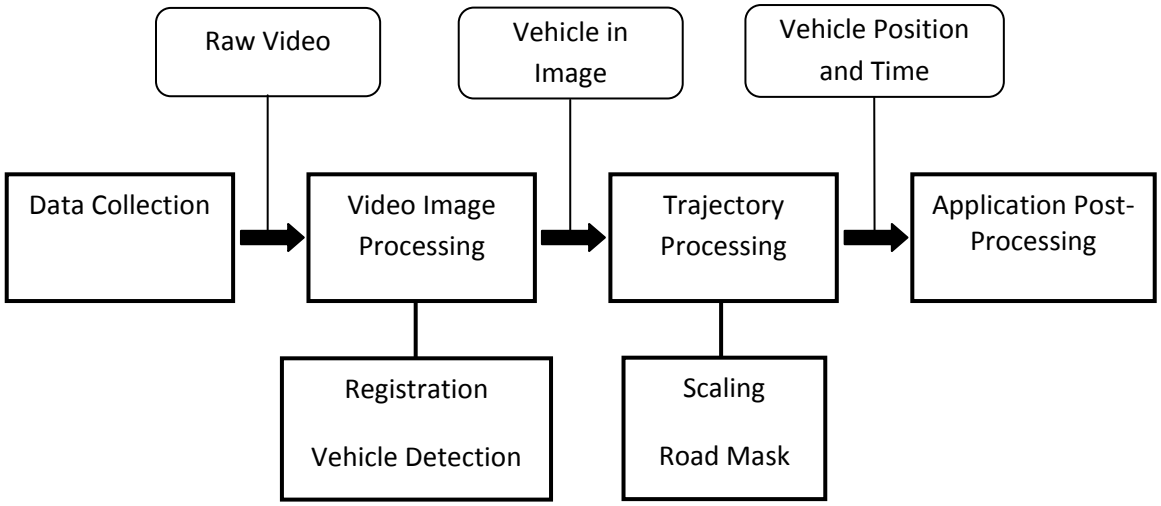


FIGURE 4 Process of traffic data collection and processing (Hickman and Mirchandani, 2006).

the airborne platform is then fused with data from other sources like inductive loops and vehicle probes. The software, called Traffic Finder, includes a graphical online presentation of the current traffic situation. The Traffic Finder software also includes a short term traffic "forecast" (up to 30 min into the future), based on a microscopic traffic simulation model or a bottleneck queuing model (Ernst et al., 2003).

An example of the types of vehicle trajectories available from this system is shown below in Figure 5. In this case, the trajectories are relatively short because of the movement of the helicopter during the data collection. Nonetheless, even these partial trajectories can be used in real time to do short-term forecasts of traffic conditions. The local speed, density, and flow estimates are determined from the individual trajectories; these are then translated through a fundamental diagram to determine the movement of shock waves in time and space along the roadway. With the shock wave analysis, vehicle movements are then projected into the future, accounting for decelerations and accelerations due to the shock waves. Hence, a short-term traffic forecast can be created (Hipp, 2006; Kühne et al. 2007; and Ruhé et al., 2007).

CONCLUSIONS

Based on the above review, it is clear that the history of traffic flow measurement and modeling from airborne imagery is very rich, where Greenshields played a very prominent role. Many researchers and traffic flow analysts have shown the potential advantages of this imagery for covering both the spatial and temporal movement of vehicles. Throughout the 80-year history, studies with this imagery have produced methods for obtaining macroscopic measures of speed, flow, density, and many other level-of-service variables of the observed traffic. Such measures are providing visually meaningful and quantitative congestion information, from which

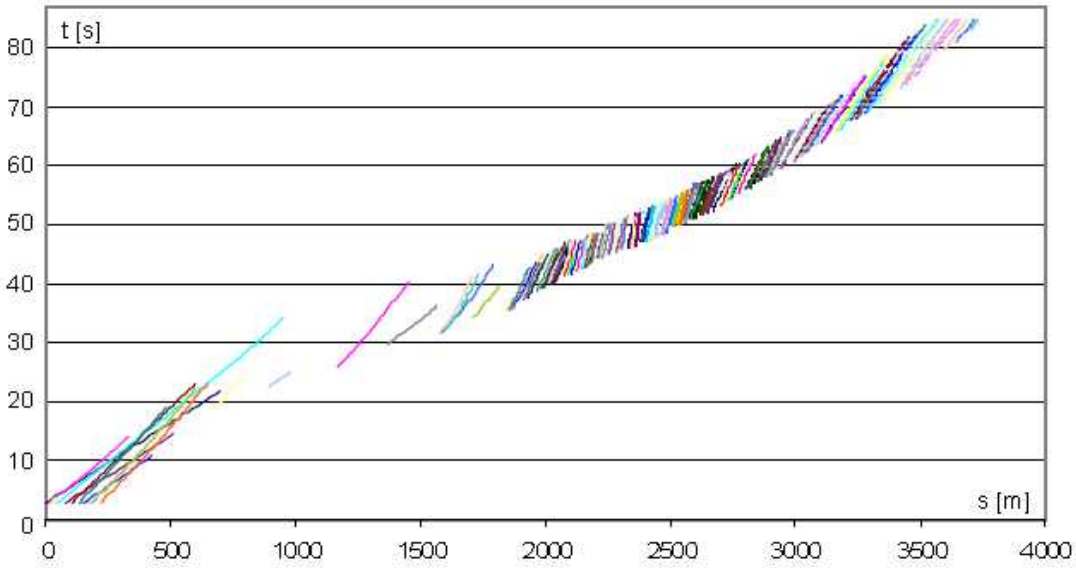


FIGURE 5 Vehicle trajectories (Hipp, 2006).

engineers, planners and policy makers can make more informed decisions on improvements to the transportation system. Moreover, several researchers have developed tools to automate the processing of this imagery, creating data sets of vehicle trajectories for a wide variety of roadway facilities. In turn, these data can be used for developing and testing new traffic flow theories and for calibrating and validating traffic simulation tools. Current research directions and improvements in image acquisition speed and quality will shortly result in applicable methods for real-time monitoring of traffic from airborne platforms which should be useful for traffic management purposes. It is only a matter of time that airborne imagery using fixed wing planes or helicopters, and even perhaps UASs, will be an accepted method by practitioners to collect traffic data for transportation planning and traffic management, and by researchers as a very appropriate method to collect traffic data for studying traffic flow theories and for calibrating and validating traffic simulation models.

RESOURCES

Agrawal, A., and M. Hickman. Automated Extraction of Queue Lengths from Airborne Imagery. Presented at IEEE Intelligent Transportation Systems Conference, October 4–7, 2004.

Agrawal, A., and M. Hickman. Automated Extraction of Queue Lengths from Airborne Imagery. Presented at 84th Annual Meeting of the Transportation Research Board, Washington, D.C, 2005.

Angel, A., M. Hickman, D. Chandnani, and P. Mirchandani. Application of Aerial Video for Traffic Flow Monitoring and Management. In *Applications of Advanced Technology in Transportation*, *Proceedings of the Seventh International Conference* (C. P. Wang, S. Madanat, S. Nambisan, and G. Spring, eds.), ASCE, Reston, Va., 2002, pp. 346–353.

Angel, A., M. Hickman, P. Mirchandani, and D. Chandnani. Methods of Analyzing Traffic Imagery Collected from Aerial Platforms. In *IEEE Transactions on Intelligent Transportation Systems*, Vol. 4, No. 2, 2003, pp. 99–107.

Angel A., and M. Hickman (2002a). Experimental Investigation of Travel Time Estimation Using Geo-Referenced Aerial Video. Presented at 81st Annual Meeting of the Transportation Research Board, Washington, D.C, 2002.

Angel A., and M. Hickman (2002b). *Considerations for Aerial Traffic Data Collection*. National Consortium on Remote Sensing in Transportation—Flows (NCRST-F), Cookbook #2, June 2002. Available at http://www.ncrst.org/research/ncrst-f/library/cookbook_reports.html.

Angel A., and M. Hickman (2002c). *A Method for Measuring Freeway Level of Service from Airborne Imagery.* National Consortium on Remote Sensing in Transportation—Flows (NCRST-F), Cookbook #3, June 2002. Available at http://www.ncrst.org/research/ncrst-f/library/cookbook_reports.html.

Angel A., and M. Hickman. Method for Analyzing Performance of Signalized Intersections from Airborne Imagery. Presented at 82nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2003.

Chandnani, D., and P. Mirchandani. *Real-Time Estimation of Travel Speeds with Digital Cameras*. Technical Report, ATLAS Center, University of Arizona, 2002.

Coifman, B., M. R. McCord, R. Mishalani, and K. Redmill. Surface Transportation Surveillance from Unmanned Aerial Vehicles. Presented at 83rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2004.

Cyra, D. J. Traffic Data Collection Through Aerial Photography. *Highway Research Record 375*, HRB, National Research Council, Washington, D.C, 1971, pp. 28–39.

Ernst, S., S. Zuev, K. Thiessenhusen, M. Hetscher, S. Rassmann, and M. Ruhé. LUMOS—Airborne Traffic Monitoring System. In *Proceedings of the IEEE 6th International Conference on Intelligent Transportation Systems*, Shanghai, China, 2003.

Forbes, T. W., and R. J. Reiss. 35-Millimeter Airphotos for the Study of Driver Behavior. *Bulletin 60*, HRB, National Research Council, Washington, D.C., 1952, pp. 59–66.

Greenshields, D. The Potential Use of Aerial Photographs in Traffic Analysis. In *Highway Research Board Proceedings,* Vol. 27, HRB, National Research Council, Washington, D.C., 1947, pp. 291–297.

Hickman, M., and P. Mirchandani. Uses of Airborne Imagery for Microscopic Traffic Analysis. In *Applications of Advanced Technology in Transportation, Proceedings of the 9th International Conference* (K. Wang, B. Smith, D. Uzarski, and S. C. Wong, eds.), ASCE, Reston, Va., 2006, pp. 238–243.

Hipp, E. Ansätze zur Modellierung einer Kurzzeitprognose auf Basis Luftgestützter Verkehrsdaten (in German). MS thesis. University of Stuttgart, Stuttgart, Germany, 2006.

Hoogendoorn, S. P., H. J. van Zuylen, M. Schreuder, B. Gorte, and G. Vosselman. Microscopic Traffic Data Collection by Remote Sensing. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1855*, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 121–128.

Hoogendoorn, S. P., and M. Schreuder. Tracing Congestion Dynamics with Remote Sensing. Presented at 84th Annual Meeting of the Transportation Research Board, Washington, D.C., 2005.

Johnson, N. Maryland Aerial Survey of Highway Traffic Between Baltimore and Washington. In *Highway Research Board Proceedings*, Vol. 8, HRB, National Research Council, Washington, D.C, 1929, pp. 106–115.

T.D. Jordan. Development of the Sky Count Technique for Highway Traffic Analysis. In *Highway Research Record 19*, HRB, National Research Council, Washington, D.C., 1963, pp. 35–46.

Kadam, K. Detection and Tracking of Vehicles in an Airborne Video Sequence. MS thesis. Department of Electrical and Computer Engineering, University of Arizona, May 2005.

Kühne, R., M. Ruhé, and E. Hipp. A Model for New Data—Using Airborne Traffic Flow Measurement for Traffic Forecast. Presented at TRISTAN VI, Phuket, Thailand, June 2007.

Maricopa Association of Governments (MAG). *1998 MAG Regional Congestion Study*. September 2000. Available at http://www.mag.maricopa.gov/.

Makigami, Y., H. Sakamoto, and M. Hayashi. An Analytical Method of Traffic Flow using Aerial Photographs. In *Journal of Transportation Engineering*, Vol. 111, No. 4, 1985, pp. 377–394.

Mirchandani, P., M. Hickman, A. Angel and D. Chandnani. Application of Aerial Video for Traffic Flow Monitoring and Management. In *Proceedings of the Pecora 15 Conference* (CD-ROM), November 2002.

Mishalani, R. G., B. Coifman, and D. Gopalakrishna. Evaluating Real-Time Origin–Destination Flow Estimation Using Remote Sensing Based Surveillance Data. In *Applications of Advanced Technology in Transportation, Proceedings of the Seventh International Conference* (C. P. Wang, S. Madanat, S. Nambisan, and G. Spring, eds.), ASCE, Reston, Va., 2002.

Murray, A. T. *Airborne Techniques for Estimating Traffic Flow in the Private Sector*. National Consortium on Remote Sensing in Transportation—Flows (NCRST-F), Cookbook #1, October 2002. Available at http://www.ncrst.org/research/ncrst-f/library/cookbook_reports.html

Nejadasl, F. K., B. G. H. Gorte, and S. P. Hoogendoorn. Robust Vehicle Tracking in Video Images Being Taken from a Helicopter. Presented at ISPRS Commission VII Mid-term Symposium, *Remote Sensing: From Pixels to Processes*, Enschede, Netherlands, May 2006.

NGSIM. *Next Generation SIMulation* website, http://www.ngsim.fhwa.dot.gov/. Accessed June 11, 2008.

O'Kelly, M., T. Matisziw, R. Li, C. Merry, and X. Niu. Identifying Truck Correspondence in Multi-Frame Imagery.In *Transportation Research Part C*, Vol. 13, No. 1, 2005, pp. 1–17.

Peleg, M., L. Stoch, and U. Etrog. Urban Traffic Studies from Aerial Photographs. In *Transportation*, Vol. 2, 1973, pp. 373–390.

Ruhé, M., R. Kühne, I. Ernst, S. Zuev, and E. Hipp. Airborne Systems and Data Fusion for Traffic Surveillance and Forecast for Soccer World Cup. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.

Shastry, A., and R. Schowengerdt. Airborne Video Registration for Visualization and Parameter Estimation of Traffic Flows. In *Proceedings of the Pecora 15 Conference* (CD-ROM), November 2002.

Shastry, A. Airborne Video Registration and Traffic Flow Parameter Estimation in a Multimedia GIS. MS thesis. Department of Electrical and Computer Engineering, University of Arizona, December 2002.

Schreuder, M., S. P. Hoogendoorn, H. J. Van Zuylen, B. Gorte, and G. Vosselman. Traffic Data Collection from Aerial Imagery. In *Proceedings of the IEEE Intelligent Transportation Systems Conference*, Vol. 1, 2003, pp. 779–784.

Skycomp. http://www.skycomp.com/. Accessed June 9, 2008.

Smith, S. A., and M. E. Roskin. Creation of Data Sets to Study Microscopic Traffic Flow in Freeway Bottleneck Sections. In *Transportation Research Record 1005*, TRB, National Research Council, Washington, D.C, 1985, pp. 121–128.

Treiterer, J. *Investigation of Traffic Dynamics by Aerial Photogrammetric Techniques*. Final Report EES 278, Ohio State University, Columbus, 1975.

Treiterer, J., and J. I. Taylor. Traffic Flow Investigations by Photogrammetric Techniques. In *Highway Research Record 142*, HRB, National Research Council, Washington, D.C., 1966, pp. 1–12.

Wagner, F. A., and A. D. May. Use of Aerial Photography in Freeway Traffic Operations Studies. *Highway Research Record 19*, HRB, National Research Council, Washington, D.C, 1963, pp. 24–34.

Bird's Eye Perspective on Traffic Flow

New Insights from Observing Traffic from a Helicopter

SERGE P. HOOGENDOORN
HANS VAN LINT
VICTOR KNOOP

In order to truly advance traffic flow theory, information regarding the spatiotemporal characteristics of macroscopic traffic flow characteristics is necessary but not sufficient. Detailed accounts of the time-space behavior of all individual vehicles in a region of interest are needed. The seminar work of Treiterer (*1*) already indicated the potential of these data, showing asymmetric acceleration and deceleration and the hysteresis phenomenon. Since then, different initiatives have been started in order to collect extensive data sets of vehicle trajectories, such as the NGSIM project and the Tracing Congestion Dynamics project (*2*). In the former, vehicle trajectory data are collected from fixed locations (high-rise buildings). In the latter, these data are collected using a mobile platform (a helicopter), the main advantage of which is the increased flexibility of the data collection location. This contribution provides a review of these new approaches to microscopic traffic data collection. The authors briefly describe the observation technique, as well as the algorithms used to derive the vehicle trajectory data from the camera images collected. The authors show a couple of examples of data analysis techniques and review the main new findings of applying these to the gathered trajectory data. This includes new insights from model identification at the individual driving level, such as interdriver heterogeneity, multileader behavior, lane changing and overtaking. The authors also present results of collecting and analyzing trajectory data to study driving behavior under specific circumstances (i.e., driving behavior during incident conditions). Finally, the authors discuss some of the remaining open questions and how these questions might be answered in the near future. In particular, they discuss the need to fly along with the traffic for a prolonged period of time to investigate intra-driver differences, as well as intensifying the empirical research on driving behavior under special circumstances.

REFERENCES

1. Treiterer, J. and J. Myers. The Hysteresis Phenomena in Traffic Flow. In *Proceedings of the Sixth Symposium on Transportation and Traffic Flow Theory (*D. Buckley ed.), Elsevier, 1974, pp. 13–38.
2. Hoogendoorn, S. P. and H. J. van Zuylen. *Tracing Congestion Dynamics with Remote Sensing. Management and Information Systems*. 2004, Malaga, Spain, WIS.

From Inductance Loops to Vehicle Trajectories

R. EDDIE WILSON
University of Bristol, United Kingdom

This paper describes ongoing work on the United Kingdom's M42 motorway, which has a uniquely high coverage of inductance loop detectors. The spacing of detectors is sufficiently small for one to use individual vehicle data to follow single vehicles down the highway. The paper gives a brief outline of the data collection work and sketches how the vehicle re-identification algorithms work. Sample data sets are available from the project website http://www.enm.bris.ac.uk/trafficdata.

Our aim is the collection of vehicle trajectory data from highways for the better understanding of traffic flow dynamics and the better calibration of microsimulation models. One path to this goal is to mount video cameras on tall buildings or gantries and apply computer vision techniques to the resulting video streams (*1*, *2*). However, a limitation is that the resulting data sets are quite small (order 15 minutes is typical) because a great deal of manual intervention is required for the computer vision algorithms to work correctly.

An alternative source of traffic data is the inductance loop detection infrastructure which is a common feature of highways in the developed world. The spacing between loop detectors varies, but in Western Europe it is typically 500 m (approx. 1,650 feet) and in the United States it is similar. Loop detectors may either be single, in which case they measure only flow and occupancy, or double, in which case they also measure vehicles' velocities and lengths. In their usual operation, these measurements are bundled into time averages (1 minute is a typical unit) and communicated back to a control office. The consequent spatiotemporal data has led to an intense discussion of macroscopic traffic patterns and the fundamental mechanisms which explain them (*3*–*6*).

The focus of our ongoing work is individual vehicle data (IVD) collected from inductance loop detectors. For this, one intercepts and stores the velocity, length, and timing information of individual vehicles before the time-average is applied. Some past studies of IVD have focused on data from one detector and the rich connections between headway statistics and lattice gases [see Neubert et al. (*7*) and many subsequent papers]. In contrast, Coifman and collaborators have pioneered vehicle reidentification techniques which apply pattern-matching methods to the IVD from a pair of detectors [see Coifman and Cassidy (*8*) and Coifman and Krishnamurthy (*9*) for the latest perspective]. The chief idea is that a vehicle's velocity at the upstream detector may be used to forecast its arrival time at the downstream detector. One then searches for a downstream length record which matches that recorded upstream and for which the arrival time is consistent. Unfortunately, with the typical loop spacing of order 500 m, traffic may shuffle significantly between detectors. Since length measurements are noisy, one in practice may only reidentify a vehicle with confidence if its length is sufficiently distinguished or if the traffic flow is sufficiently light. Coifman has thus been limited by the spatial resolution of

his data and has worked mostly on the reidentification of trucks for the purpose of monitoring segment journey times.

The new opportunity described here is provided by the English Highways Agency's (*10*) Active Traffic Management (ATM) system which operates on a 15-km (approx. 9-mile) stretch of the M42 motorway constituting part of the box of motorways around Birmingham (the United Kingdom's second largest city). In busy periods, variable message signs set reduced speed limits and open the emergency breakdown lane for ordinary driving. Because of the need to monitor traffic closely in this situation, inductance detectors have been installed much more densely than is usual, with a nominal spacing of 100 m (approx. 330 feet). However, in a 900-m section where queuing is common, this spacing is reduced to circa 30 m (approx. 100 feet). In normal operation, the ATM system captures the usual 1-minute average data, but the spatial resolution is such that the structure of stop-and-go waves may be examined in a level of detail that was not previously possible (see Figure 1).

In 2003, transport engineering consultancy TRL conducted a 2-day trial where IVD was collected from six consecutive ATM inductance detectors with a nominal spacing of 100 m. When displayed through an appropriate graphical user interface (see Figure 2), the intelligent human can identify the patterns of vehicles from detector to detector and thus it appears possible

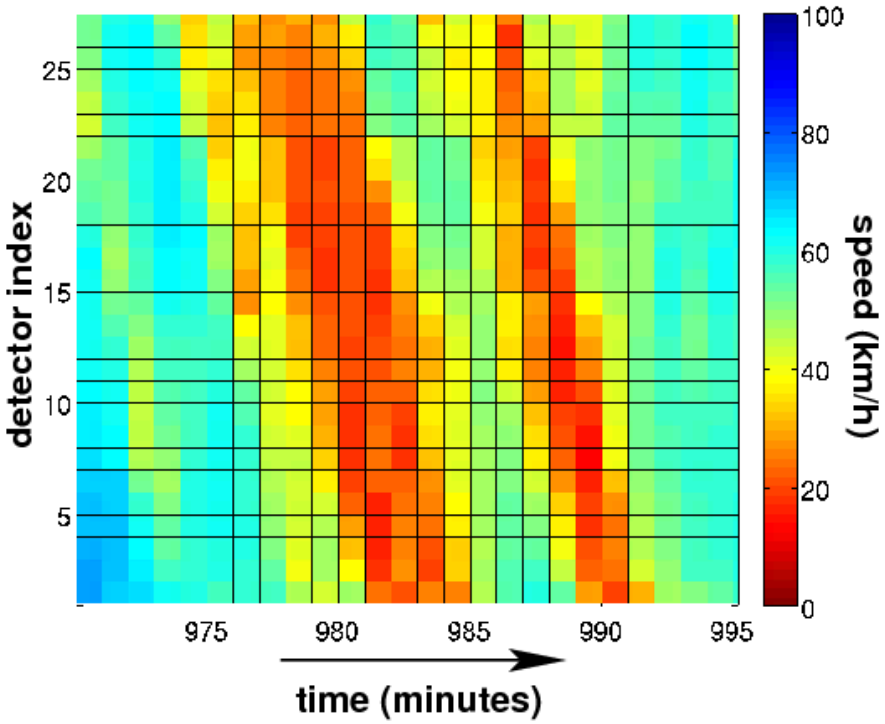


FIGURE 1 A spatiotemporal plot of speed (averaged across three running lanes) for the high coverage section of M42 ATM showing two stop-and-go waves. No interpolation has been used in the production of this picture. The vertical extent is approximately 900 m (about 3,000 feet). The authors believe that this degree of spatial resolution is unique.

to reidentify almost all vehicles with confidence—not just between a pair of detectors, but through the entire 500-m section. Thus in effect one coarsely reconstructs the trajectories of vehicles (although fine details of vehicles' accelerations cannot be captured). The technical challenge is then to devise algorithms which replicate the human pattern-matching process (*11*). In this respect, we developed algorithms which reidentify vehicles over 100 m with a success rate which on average exceeds 99% (using a human-matched set as the ground truth). Unfortunately, this success rate is due in part to the anomalously quiet traffic conditions experienced during the trial (in particular, flow breakdown did not occur). Since ongoing improvements in communications hardware and standards have reduced the need to access road-side hardware for IVD capture, there is now scope for a much more comprehensive data collection exercise.

In the work that we announce here, we have exploited a commercial equipment trial to capture the IVD from 16 consecutive detectors over 1.5 km (nearly 1 mile) of the northbound M42. The trial runs from January to October 2008, and because weekday traffic flows are approximately 70,000 vehicles, each detector will capture and store the IVD of over 15,000,000 vehicles in total. This data resource is being made available to the traffic research community over the summer and autumn of 2008 (*12*). The aims of this paper are twofold: To give further details of the data collection exercise and the basic calibration work and to give a brief outline of how the reidentification algorithms are being developed. Finally, conclusions are presented and ideas for possible joint projects are listed.

DETAILS OF THE INSTRUMENTED SECTION

We now give details of the ongoing IVD collection exercise. A kml file is available for download (*13*) which may be imported into Google Maps in order to display the instrumented section, which is approximately 1.5-km (nearly 1-mile) long. Figure 3 gives a schematic diagram of the layout. Note that the sections of Lane 0 labeled ATM are emergency breakdown lane in which vehicles do not normally drive. The Active Traffic Management system may open this lane for ordinary driving in peak periods, but at present this facility is not used within the instrumented section.

The key feature in Figure 3 is the midsection on-ramp, which contributes order 10% to 15% of the downstream traffic on average, divided roughly equally between its two lanes. Firstly, this presents new challenges for the vehicle reidentification algorithm between Detectors 4 and 9 where the majority of merges occur. Secondly, it introduces the possibility of interesting traffic dynamics. Indeed, an examination of 1-minute average data (which has been collected for several years) indicates that this section regularly induces flow breakdown and stop-and-go waves as well as experiencing large amplitude stop-and-go waves which have propagated back from highly congested junctions further downstream.

The data that each inductance detector records is listed in Table 1. A significant advance since our 2003 exercise is that arrivals are now quoted to 0.1 seconds accuracy (whereas previously they were quoted only to the nearest second, with in-lane headway quoted to 0.1 seconds). These new data are extracted as a by-product of the IDRIS waveform analysis system which is undergoing a commercial trial. It is a general observation that IDRIS appears to produce significantly cleaner IVD than the standard hardware used in the 2003 exercise. In particular, vehicles miscounts (due to either close following or lane straddling) are fewer than 1 in 1,000.

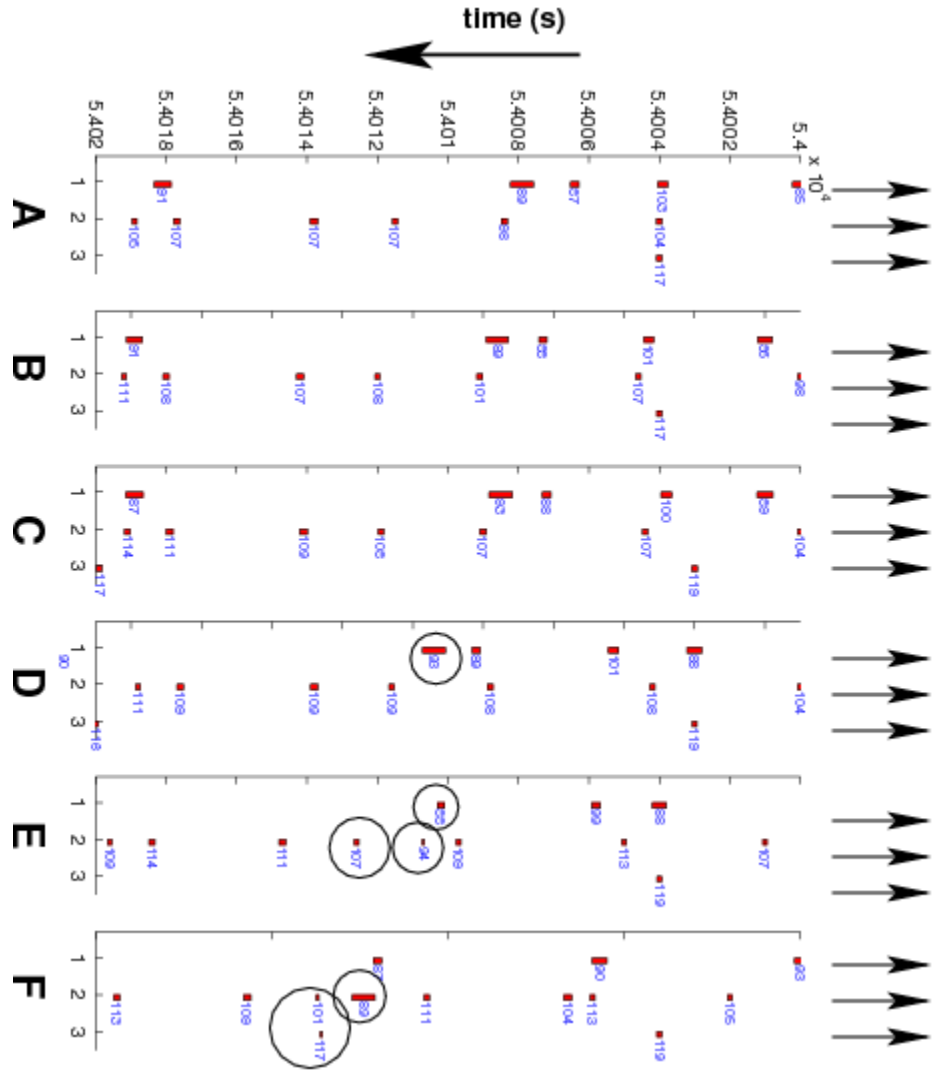


FIGURE 2 Visualization of 20 s of IVD captured during the 2003 trial. Panels A–F denote the six detector sites progressing in downstream order. In each panel, lane numbers 1–3 are plotted horizontally whereas time is plotted down the vertical axis and thus plays the role of a space-like coordinate in which vehicles drive up the axis (rather like a photo-finish camera). Vehicle records are illustrated by rectangles (whose size is derived from the vehicle's length) next to which the velocity in km/h is given. There is a time-offset of 3 seconds between each panel so that vehicles at 120 km/h (approx. 75 mph) maintain the same horizontal level. Vehicles may clearly be reidentified from panel to panel and the overall effect is similar to six helicopter views of the traffic showing how the relative configuration of vehicles changes down the highway. Note that in the United Kingdom, slow traffic (trucks, etc.) drives on the left. Some lane-changing events, where vehicles straddle detectors, have been circled.

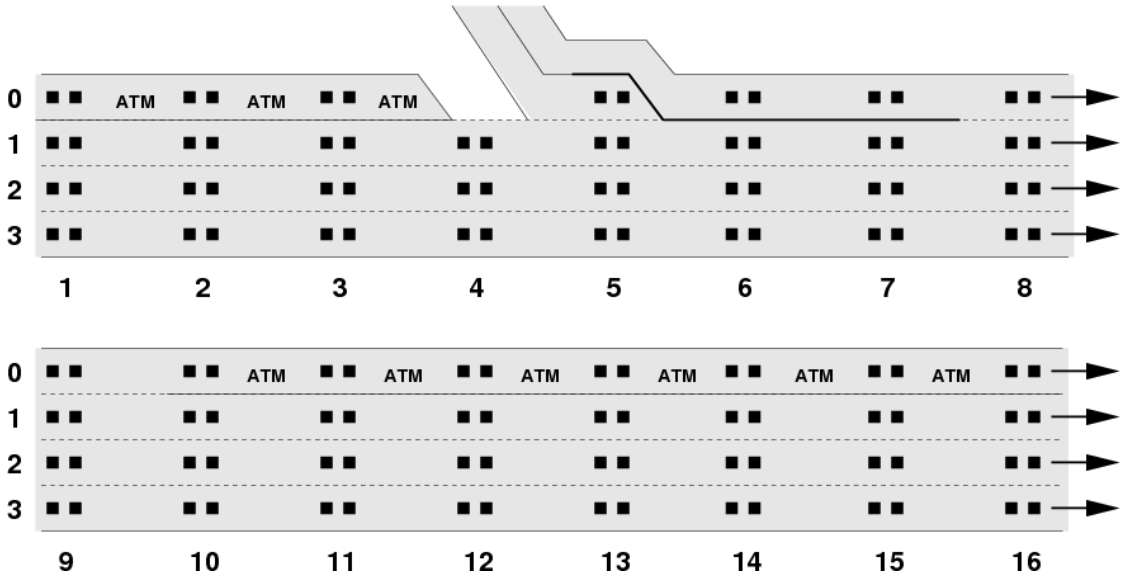


FIGURE 3 Schematic layout of the instrumented section of motorway. Loop detector sites 1–16 are separated by about 100 m (approx. 330 ft) longitudinally so that the whole section is approximately 1.5 km (nearly 1 mi). Lanes 1–3 constitute ordinary running lanes to which U.K. driving rules apply (slower vehicles should tend to drive on the left in lane 1). Lane 0 is made up of the emergency breakdown lane (labeled ATM, since it may be activated as an ordinary running lane) or parts of the on-ramp. The bold lane-marking on the on-ramp denotes a region of chevrons which vehicles should not normally cross.

Our first challenge is to calibrate the loop data. Whereas we may assume that vehicles' arrival times are correct to their quoted resolution, the velocity and length records are subject to measurement error. Furthermore, the spacing of consecutive detectors is only nominally 100 m and must be determined more accurately for the reidentification algorithms to work efficiently. Note that the question of velocity calibration can only be solved definitively by driving a probe vehicle repeatedly through the instrumented section, and cross-checking with its IVD — which we have yet to do.

In contrast, the calibration of length measurements, at least in relative terms, may be achieved directly from IVD. The technique is to consider consecutive pairs of detectors and to seek upstream–downstream pairs of vehicle records which must match, because conditions are sufficiently quiet that there are no other vehicle arrivals at either detector within any reasonable time tolerance. This requirement can be specified precisely and many thousands of such unique possible matches can be found during the night. Joint distributions of the pairs of measured lengths can then be analyzed. For private cars (length less than 5.5 m, approx. 18 feet) the

TABLE 1 Data Fields and Nominal Resolution for Each Individual Vehicle Record

Quantity	Arrival Time	Lane	Velocity	Length
Resolution	0.1 s	integer 0-3	0.01 m/s	0.01 m

NOTE: Velocity and length records are less accurate than their nominal resolution.

difference in length measurements between consecutive detectors is small — with mean order 1 cm (less than half an inch) and standard deviation order 7 cm (approx. 3 inches). Consequently, detectors' length measurements do not require calibration and moreover there is sufficient information in them to assist the reidentification of even private cars (whose lengths are not especially distinguished). Measured length differences for trucks are more widely spread but this is not a serious problem because of their relative scarcity. Unfortunately trucks' statistics do contain many outliers due to lane-changing events, where an anomalous length is recorded due to the straddling of detectors in adjacent lanes.

Unique possible matches can also be used to discover the true driving distance between detectors, by taking the time difference between upstream and downstream records and multiplying by the mean of their velocities. This calculation gives a distribution of distances with a spread which is due principally to the ±0.1-s accuracy of the time difference. By taking a large number of records, the true driving distance may be extracted from the statistics (see Table 2). Since this method assumes accurate velocity measurements, we checked the Table 2 distances with Google maps and found very close agreement. This indicates that the errors in velocity measurement have a very small mean component.

Now that driving distances have been determined accurately, we may use them to seek unique possible matches with much tighter tolerances than we used previously. We thus dramatically increase the number of unique possible matches which in turn leads to more accurate calibration of the loops. The drawback of this boot-strapping approach is that length-error statistics are built on portions of traffic data where there are lots of unique possible matches —that is, principally sparse and hence fast-moving traffic. Thus there are potential limitations in exploiting the length-error statistics to reidentify slow-moving traffic.

OUTLINE OF REIDENTIFICATION ALGORITHMS

We now sketch how the reidentification algorithms work. (The full details of the algorithms will be the subject of a forthcoming journal paper.) For simplicity, we may focus on matching the records for a single pair of detectors which do not overlap with the on-ramp, for example, Numbers 11 and 12. (The matching for detector pairs 12–13, 13–14, 14–15 and 15–16 is rather similar.) In fact, algorithms under development make use of the information from more than two detectors simultaneously, but are beyond the scope of this discussion. Since the reidentification of sparse traffic turns out to be rather trivial, it is necessary to test algorithms on a sufficiently congested day with strongly dynamic traffic patterns, for example January 24, 2008. See Figure 4.

TABLE 2 Actual Driving Distances Between Consecutive Pairs of Detector Sites

loops	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9
gap (m)	106.0	99.5	101.6	91.4	98.8	92.4	101.7	91.3

loops	9–10	10–11	11–12	12–13	13–14	14–15	15–16
gap (m)	92.7	92.9	79.1	109.9	103.2	98.6	99.9

The first step in reidentification is to partition the data so that when matches are sought, only relatively small numbers of vehicles need to be analyzed simultaneously. This partition exploits the possible match idea which we introduced earlier. Specifically, for each vehicle record at the upstream detector, we forecast the arrival time at the downstream detector using the distance x between the detectors and the velocity and arrival time at the upstream detector. We then find all records at the downstream detector whose actual arrival time is within a tolerance e of the forecast. The tolerance e may be designed in various ways, but must include 0.1 s to account for the nominal error in time measurements, as well as other components which model for velocity measurement error or the possibility of nonzero acceleration.

We then apply the procedure in reverse: that is, for each downstream record we forecast the earlier upstream arrival time based on downstream velocity and find all possible matches at the upstream detector. In this way we construct a bipartite graph of connections between upstream and downstream vehicle records. The data is then partitioned into sets of possible matches by finding the maximal connected components of the (symmetrized) bipartite graph. We call these match-sets. For example, a unique possible match corresponds to a match-set with a

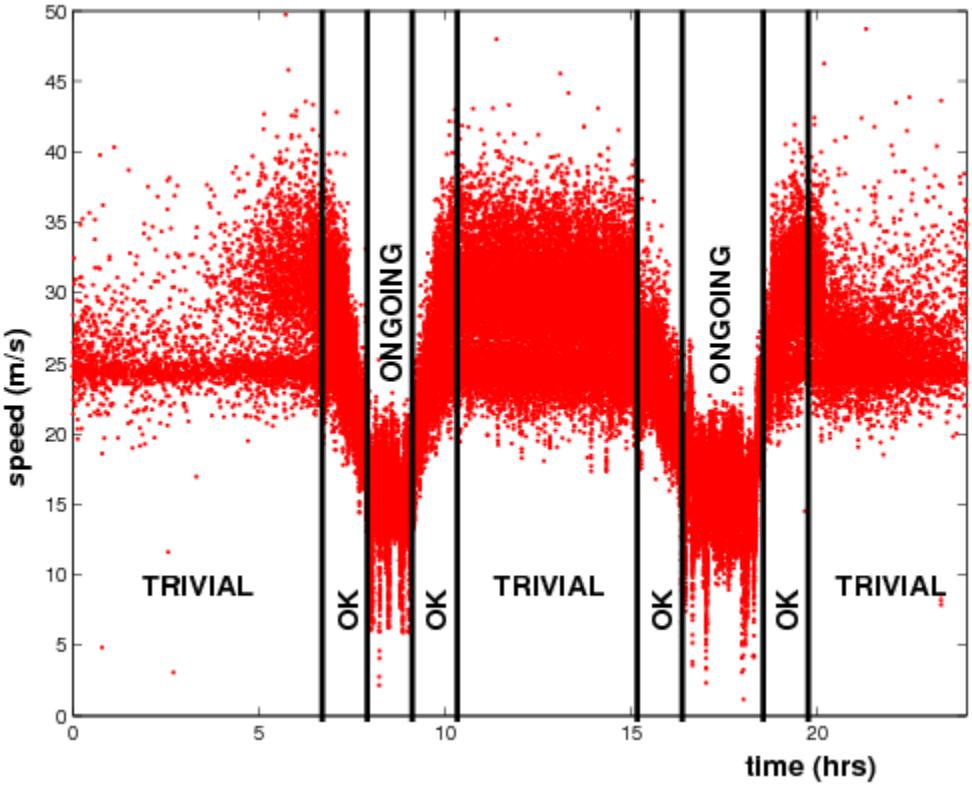


FIGURE 4 Individual vehicle velocity data from Detector 11 for January 24, 2008. In summary, we are presently able to reidentify traffic robustly except in the periods 08:00–09:30 hours and 16:30–18:30 hours when the mean speed drops sharply and stop-and-go waves nucleate at this location. Our success includes the period 07:00–08:00 hours when the flow rate at times approaches 6,000 vehicles/hour summed over the three running lanes. During 07:00–08:00 hours we observe a small velocity variance measured both temporally and between lanes.

single upstream and a single downstream record. In busy traffic, match-sets grow in size, because headways are smaller, so that one finds more and more vehicles within any given tolerance interval. Note that large match-sets tend to make reidentification more complicated and more computationally intensive, but they do not necessarily render it intractable.

We now consider the analysis of a single match-set. The simplest situation is the "square" case where it consists of an equal number of upstream and downstream records. To reidentify we then seek the best bijection between the upstream and downstream records. The simplest version of this technique defines a pairwise error score with (1) a component based on the compatibility of the upstream and downstream arrival times and velocities with the spacing between the detectors and (2) a component based on the difference of measured vehicle lengths upstream and downstream (whose design may be informed by the joint probably distribution of measured lengths for unique possible matches). The best bijection can then be defined as that which minimizes some norm of the score vector—in the case of the 1-norm, this problem may be solved by a standard numerical technique known as the Hungarian algorithm.

In practice however, pairwise scoring methods are confused by groups of vehicles with similar velocities and lengths who pass a detector at about the same time but in different lanes. Hence more sophisticated algorithms use lane information and work on the (guessed) relative likelihood of different reorderings of the vehicles. We developed these methods commercially during our original study (*11*) and tuned their parameters to perform optimally against human-matched sets. (Recall the clear patterns in Figure 2.) When compared to further human-matched data, these methods exceeded a 99% correct reidentification rate. Early indications are that this success rate is equaled in the 2008 data in all but the busiest and most strongly dynamic periods.

We now consider how the algorithms break in busy conditions. In addition to an explosion in the size of match-sets, a common problem we encounter is that match-sets are not square. Occasionally this is because a vehicle has straddled detectors during a lane change and hence has been counted twice at one site or missed completely at the other. This type of problem is in fact relatively mild and we have a variety of ad hoc solutions for dealing with it, although the merge section still presents challenges.

More seriously, in strongly dynamic traffic conditions, we presently obtain many non-square match-sets—this indicates that the search for possible matches is itself breaking down. This is because under harsh braking conditions, the velocity at the upstream detector does not give a reliable forecast of when the vehicle will arrive downstream. To solve this problem we are developing alternative methods which rely principally on matching the sequences of vehicle lengths recorded at each detector. However, in very slow traffic this technique will be subject to uncertainty for two reasons. First, the vehicles have time to substantially reorder themselves over 100 m if they are driving slowly enough. Secondly, the detectors themselves do not capture lengths reliably at very low speeds. In our favor, the behavior of slow-queuing traffic is not of especial interest.

CONCLUSIONS

We have summarized ongoing work in the collection and reidentification of IVD from inductance loops. This project uses the Active Traffic Management section of the M42 motorway, which is one of the most densely instrumented highways in the world. The section

from which we collect data includes a merge and is a good location for observing complex spatiotemporal patterns in detail.

At present, our reidentification algorithms work extremely well (>99% accurate) in all but the most congested and most strongly dynamic conditions. Our current work is focused on extending the algorithms to deal with these challenging situations and moreover to identify vehicle merges correctly.

By the end of our project in October 2008, we will have constructed the trajectories of in excess of 15,000,000 vehicles. This means that for each vehicle, we will determine a 16x4 array detailing its arrival time, speed, lane number and measured length at each of the 16 detector sites, and supply pointers to the arrays of its immediate neighbors. Small samples of the data will be freely available at http://www.enm.bris.ac.uk/trafficdata and the complete data set will be provided on application.

Our data unfortunately cannot describe dynamics which occur between the detectors (nominal spacing 100 m, approx 330 feet) and so in particular, the fine details of vehicles' accelerations are not directly accessible. However the advantage compared to camera trajectory data is the shear volume of our data set. Consequently, statistical inference might be used to develop descriptions of driver behavior which are much more detailed than the data appears to allow at first sight. In this respect it will be interesting to see to what extent camera trajectory information and inductance loop IVD can be fused.

With such a large volume of data, we may disaggregate in many different ways and yet retain statistically significant numbers of vehicle trajectories. Thus it seems that (highly parametrized) models of lane changing would benefit in particular, but one could also model the dependence of driver behavior on more exotic factors such as the weather. We are open to suggestions for joint projects that take the applications of this work forward.

ACKNOWLEDGMENT

This research is supported by an EPSRC Advanced Research Fellowship. Thanks also to Diamond Consulting, TRL and the English Highways Agency for access to inductance loop data collected from the M42 ATM system.

REFERENCES

1. NGSIM (Next Generation Simulation) project. http://www.ngsim.fhwa.dot.gov.
2. Hoogendoorn, S. P., H. J. van Zuylen, B. G. H. Gorte, and M. G. Vosselman. Microscopic Traffic Data Collection by Remote Sensing. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1855*, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 121–128.
3. Daganzo, C. F., M. J. Cassidy, and R. L. Bertini. Possible Explanations of Phase Transitions in Highway Traffic. In *Trans. Res. A*, Vol. 33, 1999, pp. 365–379.
4. Kerner, B. S. *The Physics of Traffic*. Springer-Verlag, 2004.
5. Schönhof, M. and D. Helbing. Empirical Features of Congested Traffic States and Their Implications for Traffic Modeling. In *Transportation Science*, Vol. 41, 2007, pp. 135–166.
6. Wilson, R. E. Mechanisms for Spatio-Temporal Pattern Formation in Highway Traffic. *Phil. Trans. R. Soc. A*, Vol. 366, 2008, pp. 2,017–2,032.
7. Neubert, L., L. Santen, A. Schadschneider, and M. Schreckenberg. Single-Vehicle Data of Highway Traffic: A Statistical Analysis. In *Phys. Rev. E*, Vol. 60, 1999, pp. 6480–6490.

8. Coifman, B. and M. Cassidy. Vehicle Reidentification and Travel Time Measurement on Congested Freeways. *Transportation Research Part A*, Vol. 36, 2002, pp. 899–917.
9. Coifman, B. and S. Krishnamurthy. Vehicle Reidentification and Travel Time Measurement Across Freeway Junctions Using the Existing Detector Infrastructure. In *Trans. Res. C*, Vol. 15, 2007, pp. 135–153.
10. Highways Agency. http://www.highways.gov.uk.
11. Lunt, G., M. Day and R. E. Wilson. Enhancing Motorway Traffic Data with Novel Vehicle Reidentification Algorithms. In *Proc., 13th World Congress on Intelligent Transport Systems and Services*, 2006.
12. M42 Individual Vehicle Data project website. http://www.enm.bris.ac.uk/trafficdata
13. Location of Inductance Loops Involved In The IVD Trial. KML file downloadable from http://www.enm.bris.ac.uk/staff/rew/kml/m42_ivd_loops.kml.

Empirical Observations of Traffic Flow Characteristics

Empirical Relation Between Stochastic Capacities and Capacities Obtained from the Speed-Flow Diagram

JUSTIN GEISTEFELDT
Institute for Transportation and Traffic Engineering, Ruhr-University

In conventional procedures for the assessment of traffic flow quality, the capacity of highway facilities is treated as a constant value. For freeways, the derivation of design capacities given in guidelines such as the HCM or the German HBS is based on speed-flow diagrams. More recent investigations, however, demonstrated that freeway capacity has to be regarded as a random variable. The capacity distribution function represents the probability of a traffic breakdown in dependence on the flow rate. The objective of the paper is to compare stochastic capacities with conventional capacity values. For a considerable number of data samples from freeway sections in Germany, the breakdown probability that corresponds to the capacity obtained from the speed-flow diagram is determined. In conclusion, suitable breakdown probabilities that could be used for the definition of design capacities are derived.

The capacity of a highway facility is usually defined as the maximum possible throughput that can be expected under specific geometric, traffic, and control conditions (*1*). In conventional procedures for the assessment of traffic flow quality, the capacity of a highway facility is treated as a constant value. For freeways, the derivation of design capacities given in guidelines such as the HCM (*1*) or the HBS (*2*) is usually based on speed-flow diagrams, which represent measured data. The observed data points are described by useful analytical functions. The volume at the apex of this function is treated as the capacity of the facility.

In contrast to the traditional understanding of capacity, recent research shows that freeway capacity has to be regarded as a random variable rather than as a constant value. The variability of capacities can be determined by analyzing the variance of volumes observed prior to traffic breakdowns.

The stochastic concept provides a new way to define the capacity of freeway facilities. As the stochastic capacity distribution function represents the probability of a traffic breakdown in dependence on the flow rate, capacity design values could, for example, be defined as specific percentiles of the capacity distribution function. Compared to slight speed differences in fluid traffic, a traffic breakdown entails significant delays for the road users. Thus, the probability of a traffic breakdown represents an important measure of traffic flow quality on freeways, particularly for heavily trafficked sections.

The objective of the paper is to compare stochastic capacities with conventional capacity values. The empirical analysis is based on data samples from a number of freeway sections in Germany. For each section, the breakdown probability that corresponds to the capacity obtained from the speed-flow diagram is determined. In conclusion, suitable breakdown probabilities that could be used for the definition of design capacities are derived.

CONVENTIONAL CAPACITY ESTIMATION

For the capacity estimation in the speed-flow diagram, a procedure that was developed for the revision of the design capacities for basic freeway segments given in the German Highway Capacity Manual HBS (*2*) is used [see Brilon and Geistefeldt (*3*)]. The volume at the apex of the speed-flow diagram is determined by applying the speed-flow-density relationship proposed by Van Aerde (*4*). With this traffic flow model, all traffic states in the fundamental diagram are described by a continuous function. Thus, in contrast to two-regime traffic flow models, the apex volume does not depend on the specification of a speed or density threshold between the fluid traffic regime and the congested flow regime.

Van Aerde's (*4*) approach is based on a simple car-following model, which describes the minimum desired distance headway between consecutive vehicles as the sum of a constant term, a term depending on the difference between the current speed and the free speed, and a term depending on the current speed. The speed-density relationship is

$$d(v) = \frac{1}{\Delta x} = \frac{1}{c_1 + \frac{c_2}{v_0 - v} + c_3 \cdot v} \tag{1}$$

where

d	=	density,
v	=	speed,
Δx	=	distance headway between consecutive vehicles,
v_0	=	free speed, and
c_1, c_2, c_3	=	model parameters.

The model parameters v_0, c_1, c_2 and c_3 can be calibrated by nonlinear regression in the speed-density plane. To receive an even approximation over the whole range of densities, it is useful to divide the empirical data into classes, for example, with a class width of 1 veh/km. The speed-density function in Equation 1 can then be fitted to the average speeds and densities of each class. To receive a good representation of the empirical speed-flow relationship, a sufficient number of data points in the congested flow regime is required. The presence of congested values also indicates that the data sample contains volumes up to the capacity of the analyzed cross section, which is an important prerequisite to obtain a realistic capacity estimate.

As an example, Figure 1 shows the application of Van Aerde's (*4*) traffic flow model to describe the speed-flow-density relationship of a two-lane freeway carriageway. The fitting of the model function to traffic data in 1-h intervals delivers a capacity of 3,530 veh/h.

The procedure for the assessment of traffic flow quality on freeways given in the HBS (*2*) is based on the analysis of 1-h intervals. Correspondingly, 1-h averages of flow rates and speeds are used for the empirical capacity estimation. If such large time intervals are analyzed, the data points in the speed-flow diagram result from aggregations of different traffic states. Particularly in case of a transformation between fluid traffic and congestion, 1-h averages may represent a traffic state that never existed in real traffic flow (*5*). As illustrated in Figure 2, 1-h averages tend to be located more in the center of the parabolic speed-flow scatter plot than e.g., 5-min observations. As this effect significantly influences the apex volume of the fitted

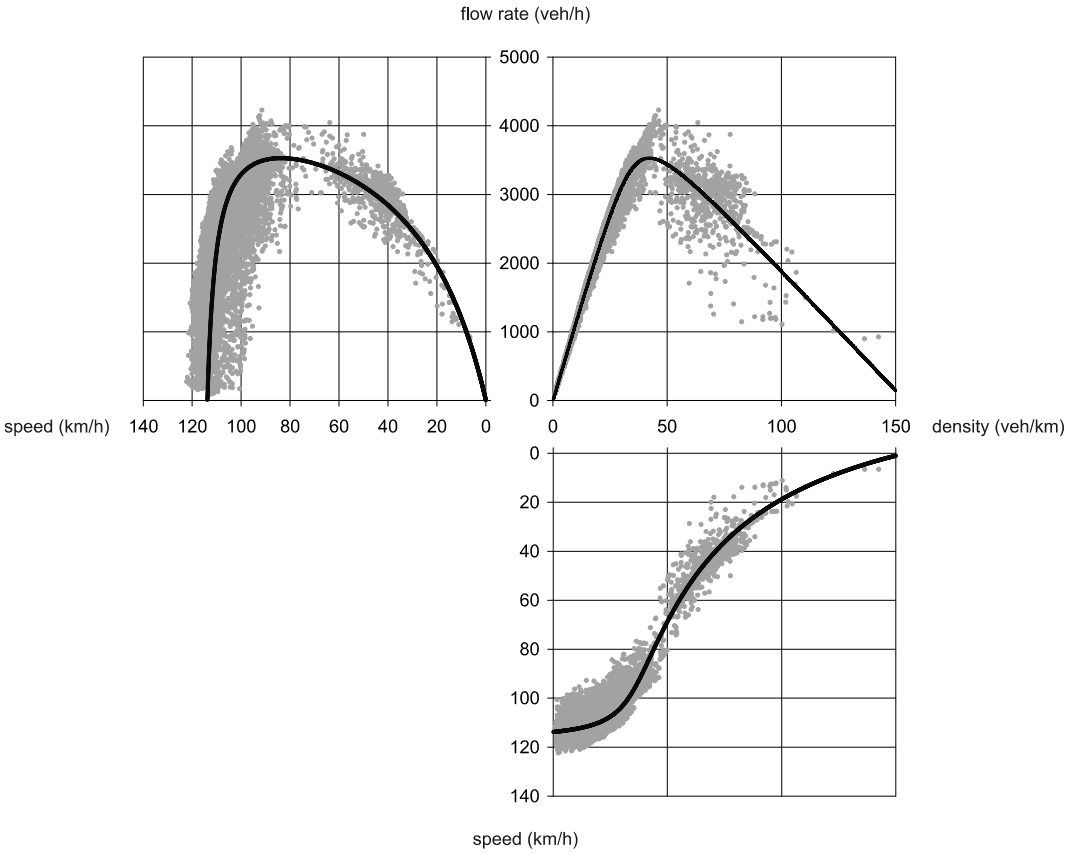


FIGURE 1 Speed-flow-density relationship of a two-lane freeway carriageway (1-h data).

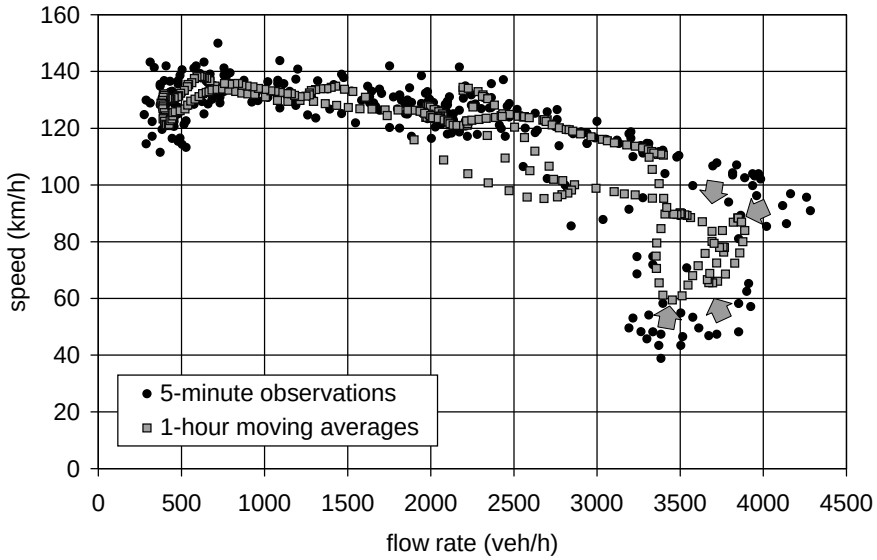


FIGURE 2 Impact of the aggregation interval on the flow-speed diagram in the presence of congested values: 5-min data versus 1-h moving averages (two-lane freeway section).

speed-flow curve, 1-h intervals with unsteady flow conditions are excluded from the analysis. A 1-h interval is considered as representing unsteady flow conditions if the root mean squared error of the 5-min speeds within the hour is greater than 10 km/h.

In some rare cases, the application of Van Aerde's (*4*) model delivers a capacity slightly below or even beyond the highest observed volumes. This mainly applies to speed-flow diagrams with a distinct gap between the fluid traffic regime and the congested flow regime. In order to avoid unrealistically high capacity estimates, the apex volume of the fitted speed-flow curve is compared with the 99th percentile of the distribution of all flow rates in the sample. If the apex volume exceeds this threshold, the percentile value is considered as capacity estimate.

STOCHASTIC CAPACITY ANALYSIS

The stochastic nature of freeway capacity has been a topic of several recent studies (*5–9*). In contrast to the traditional understanding of capacity as a constant value, these empirical investigations show that the maximum traffic throughput of a freeway facility varies, even under constant external conditions.

Empirical capacity distribution functions for specific roadway, traffic, and control conditions can be estimated by using mathematical methods for lifetime data analysis (*8*, *10*). The capacity, whose distribution function represents the probability of a traffic breakdown during a particular time interval, is considered as a lifetime variable. In this analogy, the breakdown of traffic flow represents the failure event.

Traffic flow observations on freeways deliver pairs of values of volumes and average speeds in particular time intervals. For capacity analysis, "uncensored" and "censored" intervals are distinguished. An interval i is classified as uncensored if the observed volume q_i causes a breakdown of traffic flow, thus the average speed drops below a specific threshold in the next interval $i + 1$. In this case, the volume q_i is regarded as a realization of the capacity c. If traffic is fluent in interval i and remains fluent in the following interval $i + 1$, this observation is classified as censored, which means that the capacity c in interval i is greater than the observed volume q_i. Intervals after a breakdown with an average speed below the threshold are not considered for analysis because volumes observed under congested flow conditions do not contain any information about the capacity in fluent traffic.

To estimate distribution functions based on samples that include censored data, both nonparametric and parametric methods can be used (*11*). The nonparametric product-limit method (PLM) by Kaplan and Maier (*12*) delivers a set of flow rates and corresponding breakdown probabilities, which form a discrete distribution function:

$$F_c(q) = 1 - \prod_{i\, q_i \le q} \frac{k_i - d_i}{k_i} \quad i \in \{B\} \tag{2}$$

where

$F_c(q)$ = capacity distribution function,
q = traffic volume (veh/h),
q_i = traffic volume in interval i (veh/h),
k_i = number of intervals with a traffic volume of $q \ge q_i$,
d_i = number of breakdowns at a volume of q_i, and
$\{B\}$ = set of breakdown intervals (Classification B, see above).

The distribution function will only reach a value of 1 if the maximum observed volume is an uncensored value, i.e., followed by a traffic breakdown. Otherwise, the distribution function terminates at a value of $F_c(q) < 1$.

For a parametric estimation, the function type of the distribution must be predetermined. The distribution parameters can be estimated by applying the maximum-likelihood technique. For capacity analysis, the likelihood function is (*10*):

$$L = \prod_{i=1}^{n} f_C(q_i)^{\delta_i} \cdot \left[1 - F_C(q_i)\right]^{1-\delta_i} \tag{3}$$

where

$f_c(q_i)$ = statistical density function of the capacity c;
$F_c(q_i)$ = cumulative distribution function of the capacity c;
n = number of intervals;
δ_i = 1, if interval i contains an uncensored value; and
δ_i = 0, if interval i contains a censored value.

An empirical comparison between different function types revealed that freeway capacity is Weibull distributed (*10*, *13*). The Weibull-type capacity distribution function is

$$F_c(q) = 1 - \epsilon^{-\left(\frac{q}{\beta}\right)^{\alpha}} \tag{4}$$

where

$F_c(q)$ = capacity distribution function,
q = flow rate (veh/h),
α = shape parameter, and
β = scale parameter (veh/h).

The shape parameter α determines the variance of the distribution function. The variance decreases with increasing α. The scale parameter β is proportional to both the mean value and the standard deviation of the distribution function. The scale parameter represents the systematic factors affecting freeway capacity, such as number of lanes, grade, and driver population.

The stochastic concept of capacity is based on the analysis of traffic breakdowns. The prebreakdown volumes represent the momentary capacity of the facility. As a breakdown of traffic flow is usually a sudden event, only short time intervals (5 min or even less) are suitable for the empirical capacity estimation. For greater intervals, the causality between the observed traffic volume and the occurrence of the breakdown is too weak. However, a theoretical approach can be used to transform capacity distribution functions between different interval durations. Based on a capacity distribution function estimated in 5-min intervals, the 1-h distribution can be estimated by applying the following relationship (*13*):

$$1 - F_{c,60}(q_{60}) = \prod_{i=1}^{12} \left[1 - F_{c,5}(q_{5,i})\right] \quad (5)$$

where

$F_{c,5}(q)$ = capacity distribution function estimated in 5-min intervals,
$F_{c,60}(q)$ = transformed capacity distribution function in 1-h intervals,
$q_{5,i}$ = flow rate in 5-min interval i (veh/h), and
q_{60} = 1-h average flow rate (veh/h).

With Equation 5, values of the capacity distribution function estimated in 5-min intervals can numerically be transformed into 1-h intervals. A Weibull distribution can then be fitted to the transformed values by a least squares estimation. The variance of the 5-min flow rates during 1 hour (denoted by $q_{5,i}$ in Equation 5) can be considered by using normal distributed values according to the following Equation 13:

$$q_{5,i} = z_{(2 \cdot i-1)/24} \quad (6)$$

where

$q_{5,i}$ = flow rate in 5-min interval i (veh/h), i = 1..12,
z_p = p-quantile of the N(q_{60}, σ_q),
q_{60} = 1-h average flow rate (veh/h), and
σ_q = standard deviation of the 5-min flow rates $q_{5,i}$.

The standard deviation σ_q, which represents the variability of the 5-min flow rates during one hour, has a significant impact on the transformation. The value can be estimated from empirical data. As only the highest volumes are relevant for the capacity analysis, the average value of σ_q for the upper percent of all 1-h flow rates in the sample is applied.

As an example, Figure 3 shows the capacity distribution functions of a two-lane freeway section in 5-min and 1-h intervals. Traffic data over 1 year, including a total of 152 traffic breakdowns, were analyzed. The 5-min capacity distribution function has a median of 4,708 veh/h. This means that with a probability of 50%, a demand of 4,708 veh/h in a 5-min interval leads to a traffic breakdown. The nonparametric product–limit estimation fits very well into the Weibull distribution function. The median of the transformed 1-h Weibull capacity distribution function amounts to 4,051 veh/h, which is 14% less than the median of the 5-min distribution.

COMPARISON OF DETERMINISTIC AND STOCHASTIC CAPACITIES

To compare deterministic capacity values with the corresponding stochastic capacity distribution functions, the distribution percentile that corresponds to the capacity estimated in the speed-flow diagram is determined. As illustrated in Figure 4, the conventional capacity estimate is compared with both the capacity distribution function in 5-min intervals and the transformed capacity distribution function in 1-h intervals.

The empirical relation between deterministic and stochastic capacities was analyzed for a total of 26 freeway cross sections in Germany (see Table 1). The data samples include two-lane and three-lane sections with different control conditions (no speed limit, variable speed limit, permanent speed limit). The samples contain traffic data measured over periods between 6 months and 6 years, including between 12 and 287 traffic breakdowns.

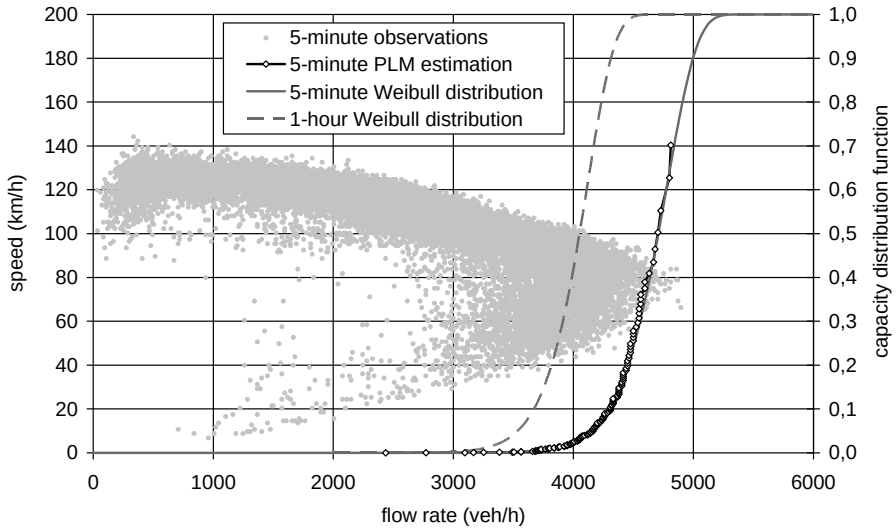


FIGURE 3 Estimated capacity distribution function in 5-min intervals and transformed distribution function in 1-h intervals (two-lane freeway Section A57/140S; see Table 1).

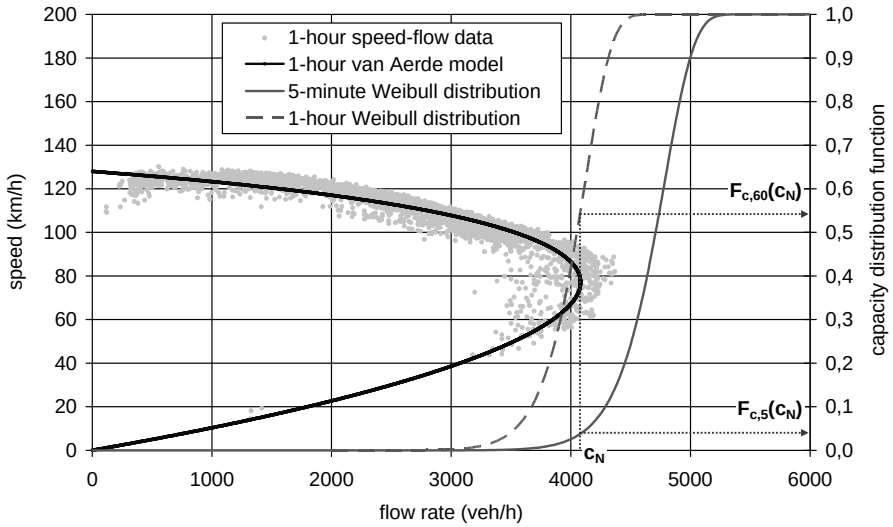


FIGURE 4 Relation between the nominal capacity c_N in 1-h intervals and the values $F_{c,5}(c_N)$ and $F_{c,60}(c_N)$ of the capacity distribution function in 5-min and 1-h intervals, respectively (two-lane freeway Section A57/140S; see Table 1).

In order to obtain realistic capacity estimates, it is important to account for systematic factors affecting the capacity of a freeway. In consequence, only those intervals measured during daylight and dry conditions were included in the analysis. Temporary influences such as work zones or accidents were excluded by analyzing the speed-flow diagram, the time series of speed and volume and the lane-flow distribution. Unusual values, e.g., periods with a reduced speed level (indicating a temporary speed limit due to a work zone) or intervals with a zero flow on one lane (indicating a lane blocking) were excluded. To account for the impact of heavy vehicles, the average heavy vehicle percentage at the highest volumes (denoted as p_{HV}) was determined. The value was rounded to a multiple of 2.5%. For the capacity estimation, all intervals with a heavy vehicle percentage in the range of $p_{HV} \pm 5\%$ were considered. For higher or lower heavy vehicle percentages, the maximum volumes were too low for a reliable capacity estimation.

For each traffic data sample, the 1-h capacity in the speed-flow diagram (denoted as "nominal" capacity c_N), the capacity distribution function in 5-min intervals and the transformed capacity distribution function in 1-h intervals were determined by applying the procedures presented in the previous chapters. The results are summarized in Table 1. The nominal capacity estimates are mostly comparable with the HBS (*2*) capacities for the section-specific geometric, traffic, and control conditions. This evidences that the results of the applied method for the conventional capacity estimation are consistent with the HBS (*2*) design concept. On the average, the nominal capacities correspond to the 3rd percentile of the capacity distribution function in 5-min intervals and the 40th percentile of the transformed capacity distribution function in 1-h intervals. This means that the conventional capacity estimate roughly represents a 3% probability of a traffic breakdown during a 5-min interval and a 40% probability during a 1-h interval. However, the section-specific breakdown probabilities $F_{c,5}(c_N)$ and $F_{c,60}(c_N)$ show a considerable variability.

Figure 5 shows the empirical relation between the nominal 1-h capacity estimates and the scale parameter β of the capacity distribution functions in both 5-min and 1-h intervals. Linear regression analysis yields that the scale parameter of the 5-min capacity distribution function can approximately be estimated by multiplying the corresponding nominal capacity by 1.26. The scale parameter of the 1-h capacity distribution function is on average 1.05 times higher than the nominal capacity.

CONCLUDING REMARKS

Conventional design capacities given in guidelines such as the HCM (*1*) or the HBS (*2*) are based on the analysis of speed-flow diagrams. The volume at the apex of the speed-flow relationship is treated as the capacity of the facility. In contrast, methods for stochastic capacity analysis deliver a capacity distribution function, which represents the probability of a traffic breakdown in dependence on the flow rate. For a total of 26 data samples from German freeways, the empirical relation between deterministic and stochastic capacities was analyzed by determining the percentile of the capacity distribution function that corresponds to the conventional capacity estimated in the speed-flow diagram. It was found that the conventional capacity in 1-h intervals roughly implies a 40% breakdown probability during 1-h and a 3% breakdown probability during a 5-min interval.

TABLE1 Empirical Relation Between the Nominal Capacity and the Stochastic Capacity Distribution Function for 26 Freeway Cross Sections

Freeway/ Cross Section	No. of Lanes	Grade (%)	Speed Limit	p_{HV} (%)	c_N (veh/h)	$F_{c,5}(c_N)$	$F_{c,60}(c_N)$
A1/190N	2	< 2	100 km/h	15.0	3,979	0.0558	0.6652
A3/140S	3	< 2	No speed limit	10.0	5,576	0.0344	0.4264
A3/24RN	3	< 2	No speed limit	10.0	6,393	0.0338	0.4389
A3/66GS	3	< 2	Traffic-adaptive	12.5	5,615	0.0278	0.5038
A3/110GN	3	< 2	Traffic-adaptive	15.0	5,598	0.0077	0.1755
A5/24GN	3	4	Traffic-adaptive	12.5	5,195	0.0394	0.4903
A5/100S	2	< 2	120 km/h	7.5	3,705	0.0368	0.4488
A8/Q73	3	6	Traffic-adaptive	12.5	4,302	0.0377	0.4377
A9/307	3	< 2	No speed limit	10.0	5,439	0.0164	0.1991
A40/0008W	2	< 2	Traffic-adaptive	10.0	3,967	0.0222	0.3721
A40/12E	2	< 2	Traffic-adaptive	7.5	3,818	0.0444	0.5683
A40/22E	2	< 2	Traffic-adaptive	7.5	4,071	0.0356	0.4686
A40/32W	2	< 2	Traffic-adaptive	7.5	4,019	0.0316	0.4621
A42/0457O	2	< 2	100 km/h	12.5	4,084	0.0198	0.2534
A42/0457W	2	< 2	100 km/h	15.0	3,986	0.0167	0.3098
A43/0265N	2	< 2	120 km/h	10.0	3,714	0.0211	0.4826
A43/0265S	2	3	120 km/h	10.0	3,465	0.0182	0.2630
A43/0295S	2	< 2	120 km/h	10.0	3,841	0.0127	0.1812
A44/1459O	2	< 2	Traffic-adaptive	15.0	3,738	0.0179	0.2756
A45/0271N	2	< 2	Traffic-adaptive	15.0	3,834	0.0554	0.6304
A45/1GIN	2	< 2	No speed limit	12.5	3,721	0.0115	0.1837
A52/120W	3	< 2	No speed limit	5.0	6,781	0.0298	0.4467
A57/140S	2	< 2	Traffic-adaptive	12.5	4,082	0.0383	0.5510
A57/150N	2	< 2	Traffic-adaptive	10.0	4,034	0.0376	0.5839
A66/9ZO	2	< 2	100 km/h	5.0	4,228	0.0214	0.3645
A661/1FS	2	< 2	100 km/h	7.5	3,950	0.0068	0.1034

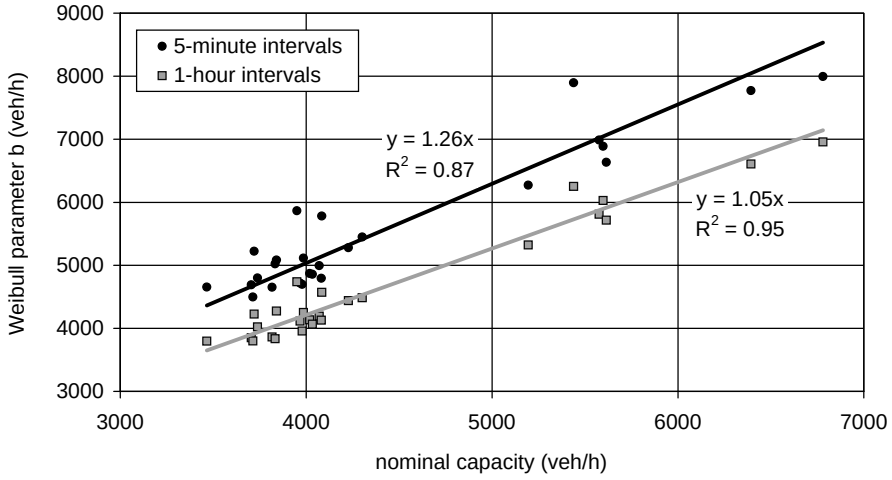


FIGURE 5 Nominal capacity versus median value of the capacity distribution function in 5-min and 1-h intervals.

Compared to the impact of speed differences in fluid traffic, a traffic breakdown entails significant delays for the users of a freeway. Hence, the breakdown probability is an important measure of effectiveness, because it represents the reliability of traffic operation. Defining a maximum acceptable breakdown probability could therefore be considered as an alternative way to derive design capacities. The results of the empirical analyses indicate that appropriate 1-h breakdown probabilities that could be used for the derivation of freeway design capacities are in a range between about 25% and 50%.

REFERENCES

1. *Highway Capacity Manual*. TRB, National Research Council, Washington D.C., 2000.
2. FGSV. *Handbuch fuer die Bemessung von Strassenverkehrsanlagen (German Highway Capacity Manual)*. Forschungsgesellschaft fuer Strassen- und Verkehrswesen, Cologne, 2001.
3. Brilon, W., and J. Geistefeldt. *Ueberpruefung der Kapazitaetswerte und q-v-Diagramme des HBS 2001 fuer Autobahnabschnitte ausserhalb der Knotenpunkte* (*Revision of the HBS 2001 Design Capacities and Speed-Flow Diagrams for Basic Freeway Segments*). Final Research Report, Institute for Transportation and Traffic Engineering, Ruhr-University Bochum, 2008.
4. van Aerde, M. A Single Regime Speed-Flow-Density Relationship for Freeways and Arterials. In *Proc., 74th Annual Meeting of the Transportation Research Board*. Transportation Research Board of the National Academies, Washington, D.C., 1995.
5. Elefteriadou, L., R. P. Roess, and W. R. McShane. Probabilistic Nature of Breakdown at Freeway Merge Junctions. In *Transportation Research Record 1484*, TRB, National Research Council, Washington D.C., 1995, pp. 80–89.
6. Minderhoud, M. M., H. Botma, and P. H. L. Bovy. Assessment of Roadway Capacity Estimation Methods. In *Transportation Research Record 1572*, TRB, National Research Council, Washington D.C., 1997, pp. 59–67.
7. Persaud, B., S. Yagar, and R. Brownlee. Exploration of the Breakdown Phenomenon in Freeway Traffic. In *Transportation Research Record 1634*, TRB, National Research Council, Washington D.C., 1998, pp. 64–69.
8. Lorenz, M. R., and L. Elefteriadou. Defining Freeway Capacity as Function of Breakdown Probability. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1776*, TRB, National Research Council, Washington D.C., 2001, pp. 43–51.
9. Brilon, W., J. Geistefeldt, and H. Zurlinden. Implementing the Concept of Reliability for Highway Capacity Analysis. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2027*, Transportation Research Board of the National Academies, Washington D.C., 2007, pp. 1–8.
10. Brilon, W., and H. Zurlinden. Ueberlastungswahrscheinlichkeiten und Verkehrsleistung als Bemessungskriterium fuer Strassenverkehrsanlagen (Breakdown Probabilities and Traffic Efficiency as Design Criterion for Freeways). *Forschung Strassenbau und Strassenverkehrstechnik, No. 870*. Bonn, 2003.
11. Lawless, J. F. *Statistical Models and Methods for Lifetime Data*. Wiley, New York, 2003.
12. Kaplan, E. L., and P. Meier. Nonparametric Estimation from Incomplete Observations. In *Journal of the American Statistical Association*, Vol. 53, 1958, pp. 457–481.
13. Geistefeldt, J. *Verkehrsablauf und Verkehrssicherheit auf Autobahnen mit vierstreifigen Richtungsfahrbahnen (Traffic Flow and Road Safety on Freeways with Four-Lane Carriageways)*. Doctoral thesis. Institute for Transportation and Traffic Engineering, Ruhr-University Bochum, No. 30, Bochum, 2007.

Fundamental Diagram for Signalized Arterials
An Empirical Analysis Using High-Resolution Traffic Data

XINKAI WU
HENRY X. LIU
NIKOLAS GEROLIMINIS
Department of Civil Engineering, University of Minnesota, Twin Cities

For uninterrupted traffic flow, fundamental diagram (FD) is well known as the description of the relationship between flow rate and density in steady state. Such FD for urban roads with signalized intersections, however, is an active research topic that deserves more discussion. In this paper, the authors study arterial fundamental diagram by using empirical observations from high-resolution, event-based traffic data collected from a major arterial in the Twin Cities area. The authors demonstrate that FD does exist on signalized roads, when the flow-occupancy data from loop detectors is aggregated at the time interval of one cycle length. However, raw flow-occupancy diagram from detector data for the morning and afternoon peaks may give misconceptions of arterial traffic flow. By exploring the impacts of *g/C* ratio, signal coordination, turning movements, and queue spillbacks, the authors especially answer the following two questions: 1) Why do different capacity values appear in the FDs for the morning and afternoon peaks? and 2) Do the scattered points with high occupancy values really mean congestion on signal link? From empirical data, the authors show that the impact of queue-over-detector (QOD) is the main cause for the appearance of "two capacity values" phenomenon for the morning and afternoon peaks and occupancy expansion on the FDs. Frequent QOD is mainly due to poor signal coordination, which leads to higher traffic flow arrival on red. The revised flow-occupancy diagrams, with removal of QOD impact, are more revealing to arterial congestion. The positions of the scattered points on revised FD, as the authors demonstrate, are good indicators for signal oversaturation. Finally, in this paper the authors also provide a mathematical description of both flow and occupancy fluctuations for the arterial FDs.

Fundamental Diagram on Urban Roads
Myth or Truth?

ELMAR BROCKFELD
ALEXANDER SOHR
PETER WAGNER
Institute of Transportation Systems, German Aerospace Centre

NATHAN H. GARTNER
Department of Civil and Environmental Engineering, University of Massachusetts

There is an ongoing discussion about the suitability of fundamental diagrams in urban networks. The authors have argued in favor of such a fundamental diagram, while some researchers deny that there is such a thing as a fundamental diagram on urban roads, because for any level of demand any travel time is possible. This contribution tries to sort this out by referring to new data sources as well as to simulation approaches. The authors hypothesize that in fact there is a fundamental diagram in urban roads; however, there is no one-to-one correspondence to the fundamental diagram on freeways.

Simulation and Calibration of Traffic Flow Models

A Simple and Pragmatic Representation of Traffic Flow

MICHAEL J. MACNICHOLAS
University of Ulster

Understanding the mechanics of traffic flow has been a fascination for engineers and others since the 1930s, starting with seminal papers by Greenshields (*1*) and Adams (*2*). There are very good practical reasons for this, as an understanding of the fundamentals of traffic flow is important for both the design of traffic facilities, and for the control of traffic operations.

However, no attempt is made here to chart progress in the development of traffic flow theory since its inception. Neither is any serious attempt made to discuss or differentiate between the different approaches or models that have been advanced as a means of predicting or explaining traffic flow. There is an abundance of literature on this topic, which is growing all the time, and consequently that particular task is best left to others.

Instead, the paper will concentrate on a particular approach that grew out of the development of a model to simulate transit times on long routes, MacNicholas (*3*). This model, which has both deterministic and stochastic components, is used by the National Roads Authority to simulate the mean and variance of journey times on Ireland's primary road network, and in particular to assess the impact of road improvements.

Obviously, this model is on a completely different scale, and can have a cruder representation of road infrastructure characteristics than is envisaged here. Nonetheless, it is thought that the basic approach, and especially the macroscopic and deterministic component of the model, has interesting features that would appear to be applicable outside its immediate practical use in simulating transit times for long routes.

FUNDAMENTAL CONSIDERATIONS

Any worthwhile model describing traffic flow must in the end be capable of being readily calibrated, using data collected in the field. And it must also replicate traffic behavior in situations where there is some experience of how traffic actually behaves.

While the above are not at first sight very demanding requirements, in practice the latter requirement does not appear to be as easy to achieve as one might initially imagine, judging by literature on the subject. To make any progress on this front it seems fairly obvious that one has to, firstly, borrow certain ideas or principles that have already been established by previous work in the area.

Secondly, one has to start with a basic uniform single-lane system with homogenous traffic, to which refinements can be added later. These could include nonuniform road characteristics, different vehicle classes, multilane systems with side road inflows and outflows. Additional reality could be introduced by incorporating random effects.

Consider then the first stage in the development of what would be described as a "deterministic single-lane continuous flow model," with the primary emphasis on determining average journey time through a road system. The basic traffic stream variables are traffic flow Q

(vehs/h), concentration or density C (veh/km), and space mean speed V (km/h), and these are related from first principles as given below:

$$Q = CV \tag{1}$$

If therefore one has a relationship between V and C (without thinking too much for the moment about precisely what it means) the system is completely specified. There is an abundance of field data obtained at many sites relating V to C, so that whatever this relationship represents it must have a central function in any realistic model of traffic flow. In other words in keeping with the theme of the Greenshields symposium one starts with a "fundamental diagram." This is the first major assumption of the development process.

The initial emphasis therefore is on identifying a suitable function for the V–C relationship that satisfies a specification with the following requirements:

1. Preferably, for convenience it should be a single-regime function.
2. It must satisfy any obvious boundary conditions.
3. It must have sufficient shape flexibility to be fitted accurately to measured field data, and to reflect the expected moderate slope of the V–C relationship when the concentration or density (C) is small.
4. The basic traffic stream variables should be simple explicit functions of each other.
5. In addition to its ability to predict V from C the function ought to, when transformed, offer some further insight into the mechanism of traffic flow.

None of the known functions, used in previous studies of traffic flow, appear to meet all of these requirements to a sufficient degree. Consequently, something new is needed if this relationship is to underpin a pragmatic model of traffic flow, when it should satisfy these requirements as closely as possible. As a result different potential functions were assessed to find the one that was deemed to be the most suitable, and the selected function is described below.

A PROPOSED SPEED-CONCENTRATION FUNCTION

It is obvious that in any real traffic situation there is a limiting value of average speed, generally called free speed V_0, that occurs when concentration is zero. And there is a limiting value of concentration, generally called jam concentration C_{jam}, that occurs when speed is zero. If therefore two new transformed (or normalized) variables x and y are introduced such that

$$x = \frac{C}{C_{jam}} \qquad y = \frac{V}{V_0} \tag{2}$$

then the proposed single-regime function, with shape parameters n and K, fulfilling all of the above requirements is given below:

$$y = 1 - x^n - Kx^n y \tag{3}$$

In general, $n \geq 1$ and $K \geq 0$ has been found for all data sets examined. Obviously, this four-parameter function satisfied the boundary constraints ($y = 1$ when $x = 0$ and $x = 1$ when $y = 0$) for all values of n and K.

Reorganizing the above expression (and utilizing the fundamental relationship in Equation 1 above) yields the following explicit relationships between the transformed variables. These can then be used to plot functions involving all the traffic stream variables, Q, C and V.

$$y = \frac{1 - x^n}{1 + Kx^n} \tag{4}$$

or

$$V = V_0[\frac{(C_{jam})^n - C^n}{(C_{jam})^n + KC^n}] \tag{5}$$

$$x = \sqrt[n]{\frac{1 - y}{1 + Ky}} \tag{6}$$

and

$$Q = (V_0 C_{jam})xy \tag{7}$$

In other words Q can be found as an explicit function of either C or V.

The space headway H in meters can be found as a function of speed V, simply by substituting Equation 6 above.

$$H = \frac{1000}{xC_{jam}} \tag{8}$$

$$H = \frac{1000}{C_{jam}}(\frac{V_0}{V_0 - V} + \frac{K}{V_0 - V}V)^{1/n} \tag{9}$$

If x* is the value of x, and y* is the value of y, when Q is a maximum it can be shown using elementary calculus that these values can be obtained as follows:

For convenience let

$$R = (K - n - 1 - nK) \tag{10}$$

then

$$x^* = \sqrt[n]{\frac{R + \sqrt{R^2 + 4K}}{2K}} \tag{11}$$

$$y^* = \frac{1 - (x^*)^n}{1 + K(x^*)^n} \tag{12}$$

and it follows that

$$Q_{max} = (V_0 C_{jam}) x^* y^* \tag{13}$$

Finally, it can be shown that slope of the V–C relationship is given by the expression

$$\frac{dy}{dx} = \frac{-nx^{n-1}(1+K)}{(1+Kx^n)^2} \tag{14}$$

It can be seen from the above expression that for certain values of the parameters K and n

$$\frac{dy}{dx} \to 0$$

when

$$x \to 0$$

CALIBRATION OF SPEED-CONCENTRATION FUNCTION

Fitting the function in Equation 5 above to a set of single-lane, speed and concentration values determined from field measurements is a fairly straightforward optimization problem. The parameters are varied, in some cases subject to practical constraints, to minimize the sum of the squares of the deviations of the data from the model.

However, this is not the main problem that has been identified to date in research. The amount of scatter found in published data obtained in many field trials is generally thought to be much greater than could be attributed to normal experimental error or random effects. For example, see work by Erlingsson et al. (*4*) or Hranac et al. (*5*) as an indication of the amount of scatter found. There are many other published results that exhibit similar scatter.

The most logical explanation for this is that measurements must actually reflect transient as well as equilibrium, or steady state, traffic conditions. The transient states are caused by non-uniform road conditions and/or disturbances. Consequently, a typical speed-concentration relationship for a site contains information on a variety of traffic states.

But there appears to be a partial consensus amongst the various commentators on this issue, and in some ways it is intuitively obvious, which could be interpreted as follows: it would not be unreasonable to assume that a model fitted to any set of data represents something close to

the equilibrium traffic states, with transient traffic states scattered about the line. This is the second major assumption of the development process.

Nonetheless, for the purpose of model calibration in Figure 1 below it is thought best to select a set of data (from the many possible data sets) that was collected under carefully controlled conditions. In other words, where the amount of scatter is minimal. Data obtained from some single-lane traffic experiments on straight sections by the Road Research Laboratory (*6*) was considered to be suitable for this purpose.

The input parameters for the model determined from the curve fitting process are shown in Table 1. In addition it also gives some important calculated output values, the speed and concentration at maximum flow, and the maximum flow.

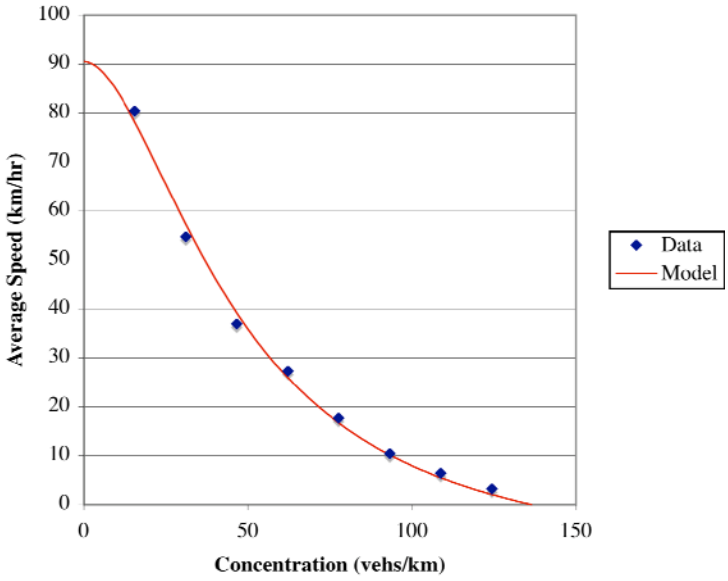


FIGURE 1 Proposed function fitted to field data.

TABLE 1 Function Parameters and Values

Input	V_0	90.58	km/hr
	C_j	136.40	vehs/km
	K	6.83	
	n	1.81	
Output	V*	46.00	km/hr
	C*	40.26	vehs/km
	Q_{max}	1851.9	vehs/hr

COMPARISON WITH THE VAN AERDE FUNCTIONS

The widely used four-parameter model first proposed by Van Aerde (*7*) has some basic features in common with the model proposed here, and a comparison is interesting. It satisfies the same boundary conditions at zero concentration and maximum concentration as described earlier. And it can assume the same variety of shapes for the relationships between the traffic stream variables V, C and Q.

However, it differs in the sense that V* and Q_{max} are input parameters to the Van Aerde model, instead of being calculated as the outputs (as in Table 1) from the model described here. This is not seen as a disadvantage for the proposed model, given the difficulty of measuring these quantities directly. Because of unstable conditions around this point in a speed-flow relationship values of V* and Q_{max} are probably found more accurately using model calibration as described earlier.

In addition there is a further price to be paid for using the Van Aerde formulation, in that the relationships between the traffic variables are much more cumbersome, in contrast with those given in Equations 4 to 14 above. This is particularly the case where V as an explicit function of C is required. In fact the function is indeterminate when C = 0.

Since it has already been decided that the V–C relationship is very important, and must underpin any explanation of how traffic actually behaves, the Van Aerde model is just not very convenient to use. Hence a more convenient alternative function has been suggested, and is used here.

Nonetheless, it is interesting to compare the functional relationships produced by the model proposed here, with the parameters specified in Table 1, and those produced by Van Aerde, using the calculated values of V* and Q_{max} obtained from Table 1. For all practical purposes the functions shown in Figures 2, 3, 4 and 5 for the two different approaches are very similar. In fact, they are almost identical.

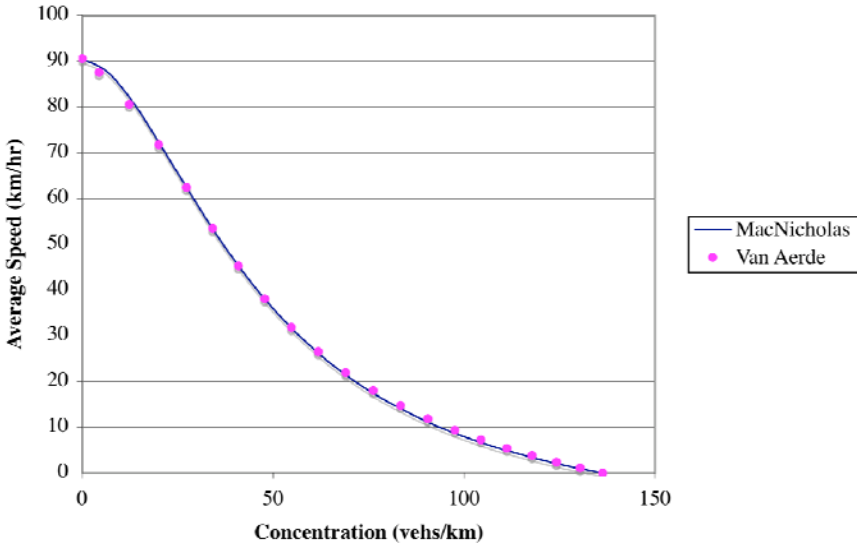


FIGURE 2 Comparison of the speed-concentration relationships.

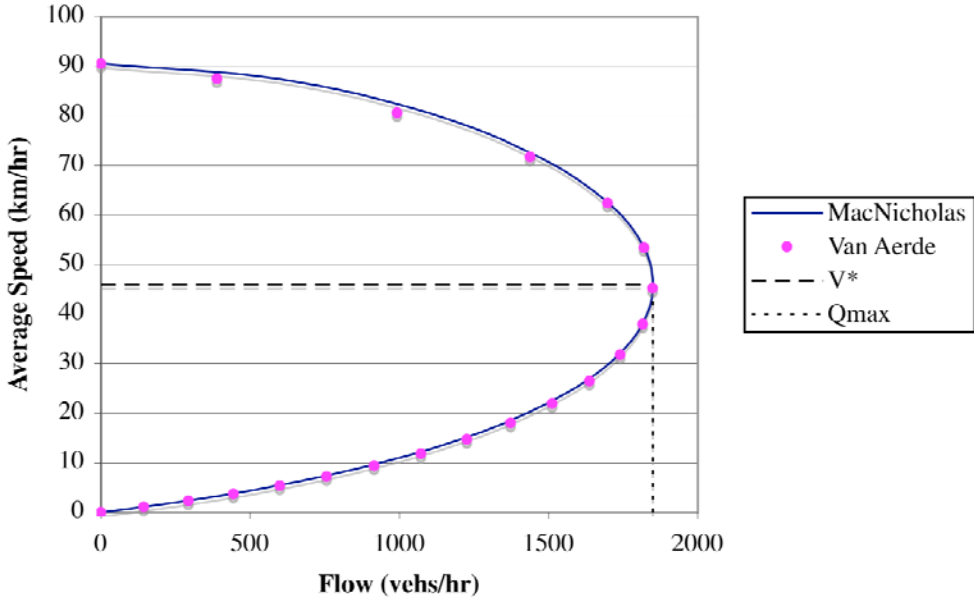


FIGURE 3 Comparison of the speed-flow relationships.

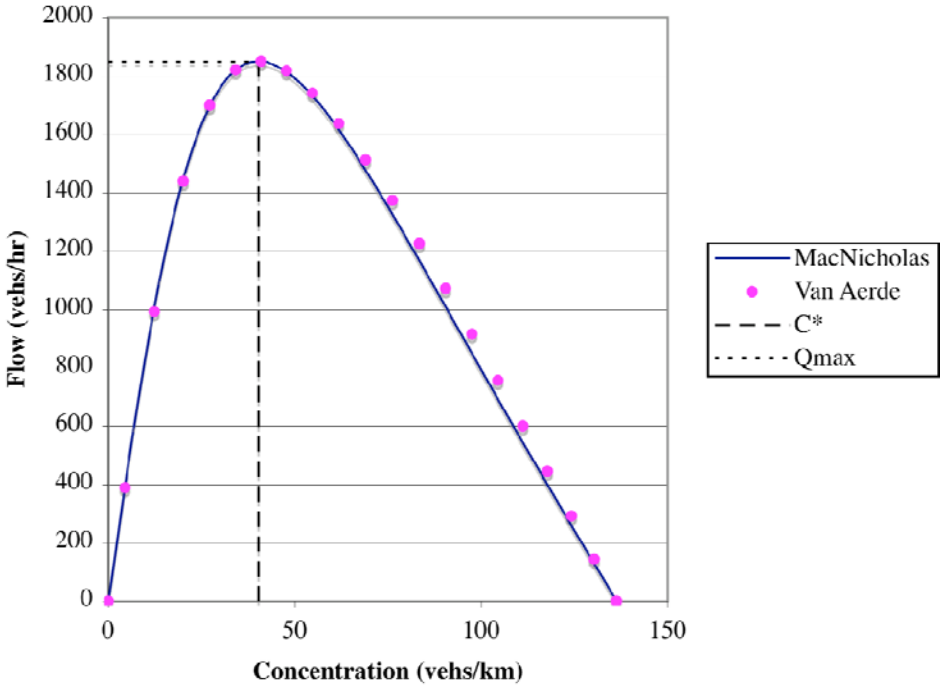


FIGURE 4 Comparison of the flow-concentration relationships.

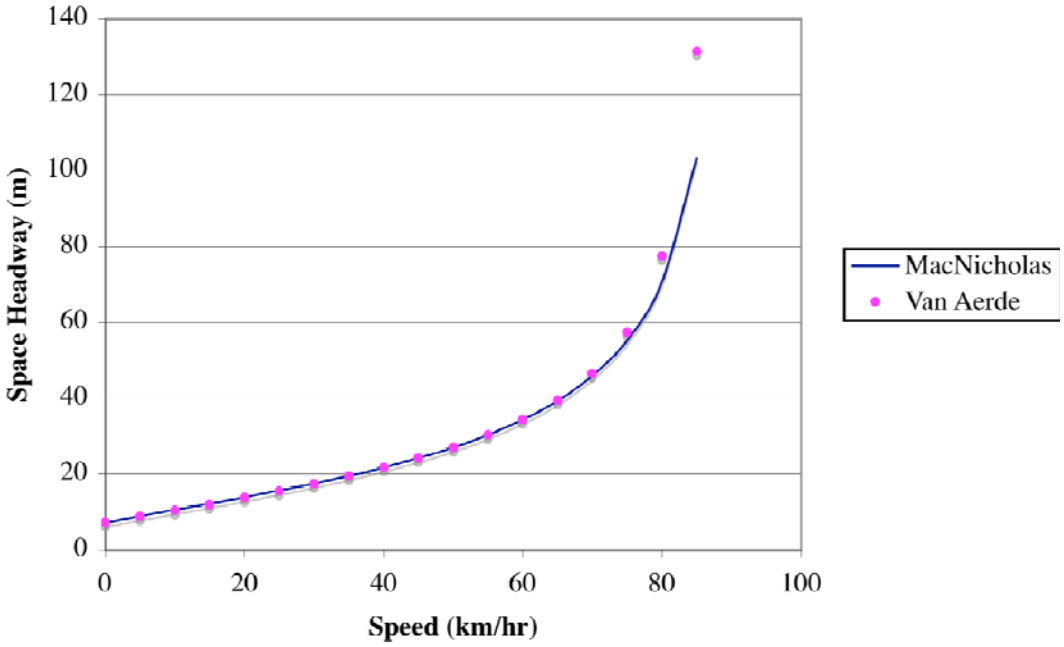


FIGURE 5 Comparison of headways.

And this leads one to speculate, is there anything more in common between the two sets of arbitrary equations, other than they produce overlapping relationships that are very similar for all values of the input parameters? Some algebraic manipulation of the Van Aerde expressions shows that for the special case when n = 1 in the proposed model, the relationships are identical for all values of K. However, this finding is not of great practical use, because values of n > 1 are normally required to achieve the required shape of the V–C relationship when C approaches 0.

A FRAMEWORK FOR SIMULATING TRAFFIC BEHAVIOR

For computational efficiency there are clear advantages in considering traffic behavior discretely. In other words a single-lane length of road is broken down into discrete sections of space (dL) and is considered at discrete intervals of time (dT), Apart from this it helps to understand how traffic might behave by considering the evolution of traffic states in these discrete intervals.

The computational sequence is illustrated in Figure 6 for a situation where each section is of equal length and has uniform characteristics. However, one can deal with sections of unequal length and nonuniform characteristics with little modification of the basic calculations. This is useful where more detailed modeling is required in the vicinity of a bottleneck.

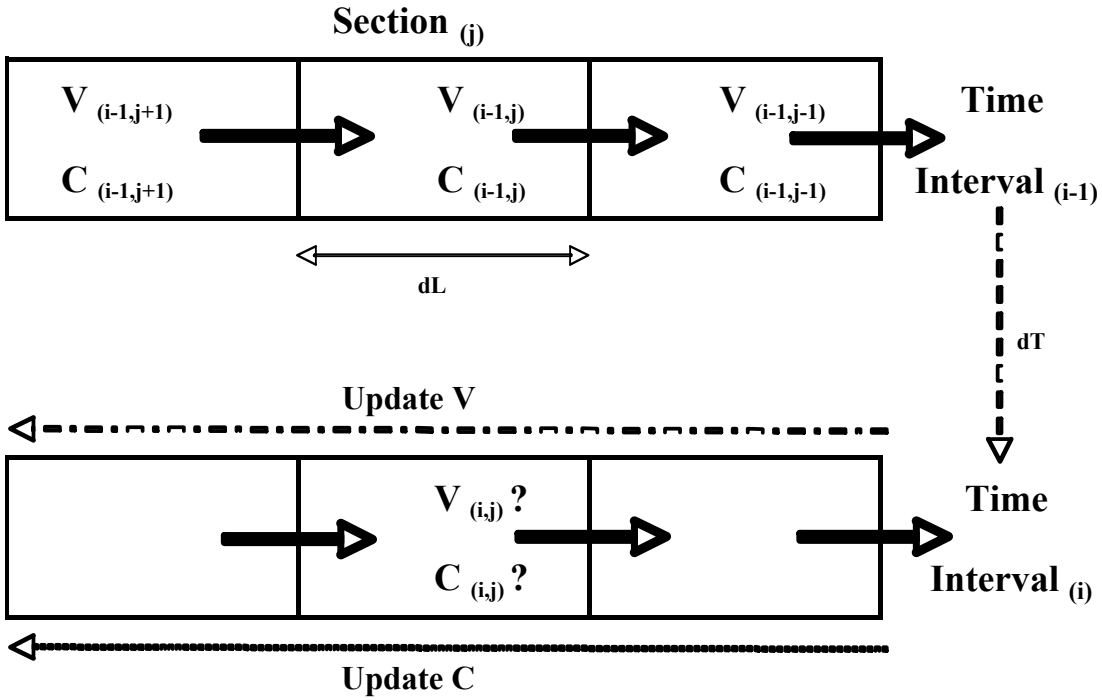


FIGURE 6 Calculation sequence.

The process begins with known values of V and C in all the sections at a particular time interval, which completely specifies the state of traffic since Q = VC. These values of V and C can be either initial assumed values, or the latest updated values. The next step is to update the value of V for the next time interval from known or historic values of V and C in the previous time interval.

Obviously, if updated values of speed can be obtained, then using known values of C, the revised values of Q and then C can readily be obtained for the next time interval using a version of the continuity equation. The real question to be resolved is as follows, what is a plausible value of V in any section for the situation described in the calculation sequence?

To answer this question one must look backwards at the evolution of thinking on how traffic is considered to behave. For example, it has been recognized for a considerable time that a continuous-flow model using the speed-concentration relationship as the sole determinant of speed at point is inadequate. What is happening to traffic as it flows along a lane is obviously more complicated than is explained by a simple relationship between speed and concentration.

A fundamental flaw with the naïve application of this relationship alone can easily be demonstrated. Once traffic gets into a high concentration state on any downstream section, flow is very much reduced as can be seen in Figure 4. The section then cannot always discharge vehicles at the same rate as input flow, and the system locks up.

The problem can be overcome by the use of a "look-ahead" term first suggested by Payne (*8*). This uses information on conditions in the downstream section as a factor in the determination of speed. It is accepted therefore that in any plausible description of traffic flow

there must be some mechanism for incorporating a look-ahead term. This is the third major assumption of the development process.

Further interesting work by Ross (*9*) provides a partial answer to the question posed earlier on a plausible value of V for a section in the calculation sequence. This work introduced the concept of the "desired change" in vehicle speed as being a function of actual speed V, desired speed V_0 and a constant relaxation time T. The suggestion is supported by actual field observations.

$$\frac{\partial V}{\partial t} = \frac{(V_0 - V)}{T} \tag{15}$$

From this expression, which is inherently logical, it would appear that in any situation acceleration is determined by the desire of drivers to move towards (if possible) free speed. But, of course this desire can be constrained by actual traffic conditions, particularly those ahead.

If the proposed speed-concentration relationship in Equation 4 is then reorganized as follows:

$$1 - y = 1 - \frac{1 - x^n}{1 + Kx^n} \tag{16a}$$

Or

$$\left[\frac{V_0 - V}{V_0}\right] - \left[\frac{1 + K}{(\frac{C_{jam}}{C})^n + K}\right] = 0 \tag{16b}$$

The above expression consisting of two parts suggests an idea on how it might be used as an element of a traffic model. Most importantly, it can be used in a way that is consistent with the earlier assumptions on the V–C relationship, and on incorporating a "look ahead" term.

The first part of the expression can be considered as a measure of the tendency of traffic on a section to accelerate, which is very similar to Equation 15 above. Then the second part of the expression (calculated for the section ahead) can be taken as a corresponding and plausible measure of the tendency to decelerate. In other words, it is a logical and effective look-ahead term for any particular section.

Clearly, the individual parts have values between 0 and 1, and a function of difference between them would have a value between –1 and +1. The fourth assumption is therefore that this is deemed to give a net effective acceleration/deceleration factor AD_{fac}, which ought to relax over time towards zero, and an equilibrium or steady state traffic flow.

Consider then how these ideas might be expressed to determine V for a discrete section of space at a discrete interval of time, for the single traffic lane shown in Figure 6. Let

$$AD_{fac} = \left[\{\frac{V_0 - V_{(i-1,j)}}{V_0}\} - \{\frac{1 + K}{(\frac{C_{jam}}{C_{(i-1,j-1)}})^n + K}\}\right]^w \tag{17}$$

where W is a calibration exponent. This is the simplest version of an expression to determine the acceleration/deceleration factor. More complicated versions exist and these will be discussed briefly at a later stage.

Then if A_{max} (m/sec^2) is the assumed maximum acceleration/deceleration (or there could be different maximum values for acceleration and deceleration) the speed in the section is determined from

$$V_{(i,j)} = V_{(i-1,j)} + 3.6A_{max}(AD_{fac})dT \tag{18}$$

The obvious equilibrium-seeking feature of this method of determining speed provides a neat and pragmatic answer to the creation of a behavioral model that incorporates all of the earlier assumptions, and in particular the idea that the speed-concentration relationship represents something close to equilibrium conditions.

However, it should be noted that the use of this approach is not confined exclusively to the model suggested here. But it is believed to be the only simple function that has the capability to be transformed in this way and at the same time has the required shape flexibility.

REALISTIC CONSTRAINTS IN UPDATING THE PROCEDURE

The primary determination of speed is by Equation 18 above in the calculation sequence outlined earlier. But a question has to be asked: does the determination of speed in this way imply that simulated traffic in the calculations ceases at any time to behave like real traffic?

There appears to be no simple analytical answer to this question. Consequentially, the calculation of $V_{(i,j)}$ must be subject to numerical constraints so as to ensure that simulated traffic behavior is always realistic. This is the final assumption in model development, and it remains to outline the form that these constraints should take.

In keeping with the earlier assumptions about what the V–C function actually represents (equilibrium states), some latitude must be allowed in setting constraints based on this function. It is reasonable therefore to assume that potentially at least, speed V and flow Q in a transient state might exceed V_0 and Q_{max} by a small amount e_1 or e_2. The constraints given in Equations 19 and 20 follow from this.

Maximum speed

$$V_{(i,j)} \leq V_o + e_1 \tag{19}$$

Maximum flow

$$V_{(i,j)} \leq \frac{Q_{max} + e_2}{C_{(i-1,j)}} \tag{20}$$

With realistic traffic flow the calculation of $V_{(i,j)}$ must be always ensure that concentration remains within the limits $0 \le C \le C_{jam}$, since these are physical limits. The constraints given in Equations 21 and 22 follow from this requirement, with calibration coefficients $k_1 < 1$ and $k_2 < 1$.

Forward packing

$$V_{(i,j)} \le \frac{k_1 \frac{dL}{dT}(C_{jam} - C_{(i-1,j-1)}) + V_{(i-1,j-1)}.C_{(i-1,j-1)}}{C_{(i-1,j)}} \tag{21}$$

Depletion

$$V_{(i,j)} \le \frac{k_2 \frac{dL}{dT}(C_{(i-1,j)}) + V_{(i-1,j+1)}.C_{(i-1,j+1)}}{C_{(i-1,j)}} \tag{22}$$

While in practice the percentage of times any of the above constraints are actually triggered appears to be quite small, they are still necessary to ensure realistic behavior at all times.

TESTING THE PROPOSED MODEL

As was stated earlier a simulation model purporting to replicate traffic behavior is only as good as the results it produces in situations where it is fairly well understood how traffic would actually behave. To assess the model a (closed system) circular single-lane track of 5-km length with 100 sections was used for a numerical experiment using a 1-second time interval.

The model parameters specified in Table 1 are used, and simulation starts with given initial conditions of speed and concentration. Plausible values were also chosen for the additional parameter specified in Equations 17 to 22 above.

A closed system with a fixed number of vehicles, and where there is no beginning or end to the traffic stream, has obvious advantages in understanding what is happening. Average concentration is known from the number of vehicles in the system, so that one can design a numerical experiment where all section concentrations should hopefully converge towards the equilibrium value irrespective of the starting position.

A fairly extreme and most testing example is chosen for initial concentration in the simulation, with a queue length of about 800 m at maximum concentration, whereas elsewhere concentrations are moderate or small. The average or equilibrium concentration is 26 vehicles/km.

The initial conditions specify the speed of each 50-m section at the start of simulation. While the process can commence with any permitted section speed, in practice it is perhaps more realistic to start with a situation where all vehicles are at rest. All vehicles then move, or attempt to move, on a signal indicating the start.

For the system outlined above there would appear to be three obvious questions that can be used to assess how realistic is the simulation process:

1. Does traffic reach an equilibrium state as could be expected where there is no external influence, and how quickly is this equilibrium reached?
2. What happens at the start, in terms of traffic flow and the direction and speed of the acceleration shock wave?
3. If measurements of speed and concentration are made for intervals of time and space during the simulation, how are these points (representing transient states) scattered about the speed-concentration function?

Of course, all the simulation output is dependent to some extent on the parameter values chosen earlier. But it was found that generally it is not very sensitive to the values used provided that these are within a plausible range.

Output shown in Figure 7 indicates the variation over time of section concentration and speed. Effectively, there is little variation after about 600 seconds, which is about the time it would take for a typical vehicle to make two loops of the circuit. Intuitively, this appears to be a reasonable replication of what might happen in practice.

Variation of section concentration over space, towards the end of the simulation period, is shown in Figure 8. It can be seen that as time goes on traffic behavior moves from what are clearly transient states towards a steady state, in other words where any memory of the initial state gradually disappears. Again after about 600 seconds there is very little spatial variation.

The rates at which the traffic system reaches stability, or steady state, appears to be influenced by different possible forms of the function to determine the acceleration–deceleration factor given in Equation 17. For example, one can use a weighted form of the deceleration element for several of the sections ahead. Or the parameter w could be a function of the expression within the brackets.

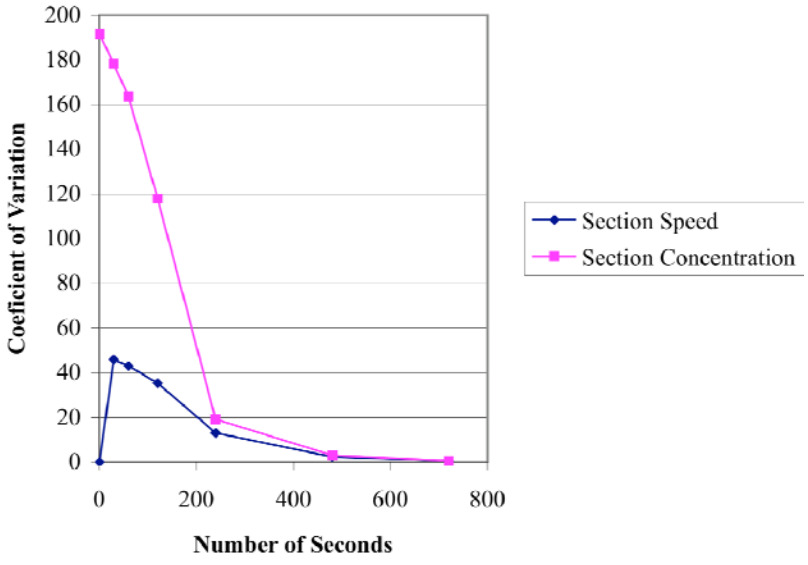


FIGURE 7 Reduction in variation over time.

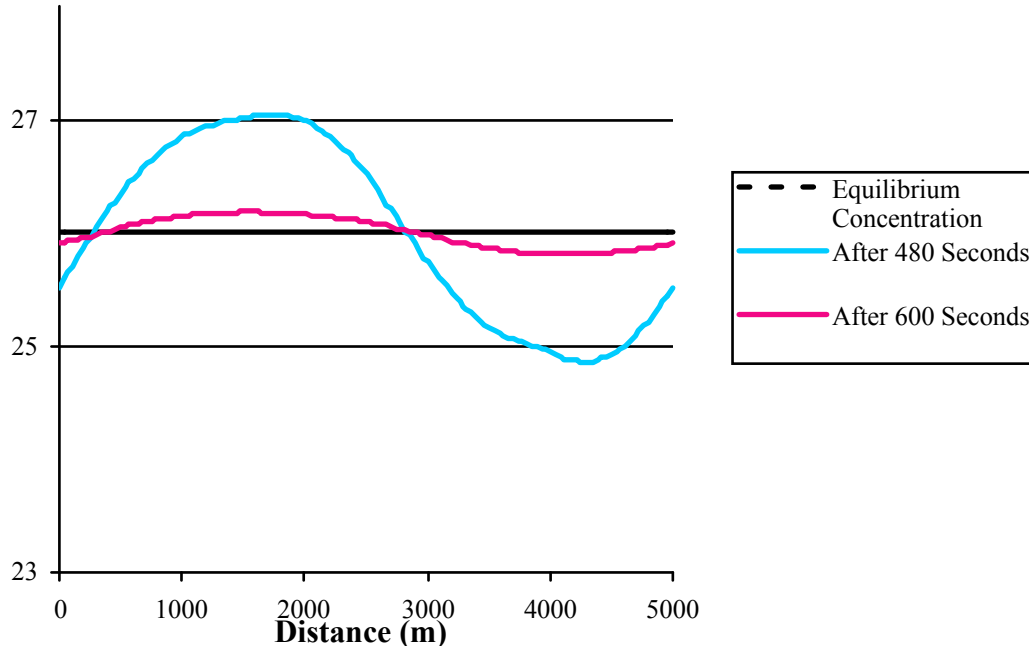


FIGURE 8 Reduction in variation over space.

In practice, actual traffic behavior would also include random effects or noise. And random effects can easily be added to the model outlined if necessary, where naturally they would affect the output seen in Figure 7 and Figure 8. But for the purpose of this exercise they are deliberately excluded so as to better understand the underlying deterministic behavior.

The next question is examined by looking in detail at what happens to the queue as it disperses after the start. It is found that traffic flow increases from 0, at the stationary queue, to about Q_{max} downstream of the queue within a short time as would be expected.

Furthermore in Figure 9, an acceleration shock wave can be seen to move upstream as expected, with a measured velocity of about 20 km/h. Interestingly, this velocity corresponds almost exactly with that predicted using the classical analysis of Gerlough and Huber (*10*), based on the work of Pipes (*11*). The final question is, what sort of a picture is presented by the transient values of speed and concentration, sampled during say the first half of the simulation period? As can be seen in Figure 10 there is significant scatter about the model even with a purely deterministic formulation.

If random effects, and experimental error with field measurements, were added then scatter would clearly increase. This corresponds with practical experience where many commentators have noted the amount of scatter found in plots of speed and concentration or density.

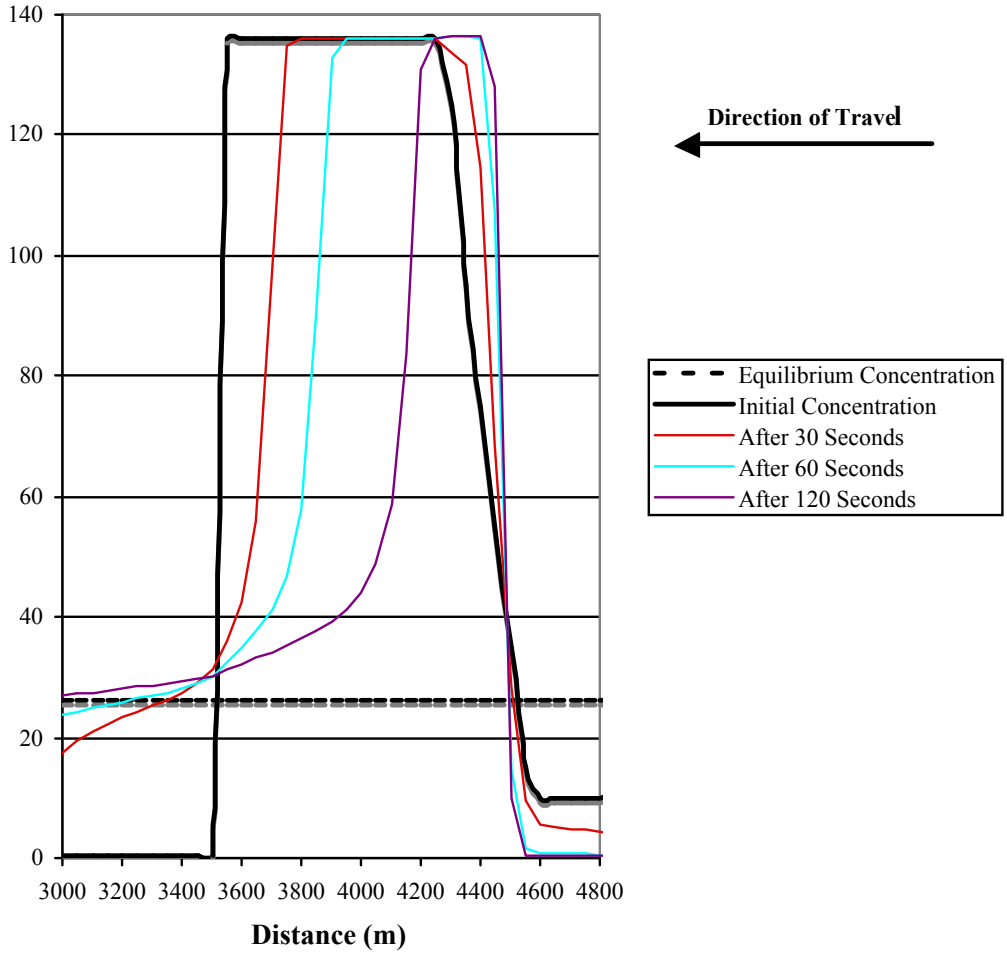


FIGURE 9 Acceleration shock wave.

While starting conditions for the simulation, in terms of the variation of concentration around the circuit, are probably more extreme than would be expected with any field experiment, the results would appear to validate the earlier assumption that a model fitted to any set of field data represents something close to a set of equilibrium traffic states, with transient traffic states and "noise" causing deviations about the line.

CONCLUSIONS

Simple empirical functions have been suggested for the relationships between traffic variables, speed, concentration and flow, as an alternative to Van Aerde functions. While both approaches have the shape flexibility to fit a variety of data, and produce almost identical shapes, it is believed that the proposed functions are significantly easier to use.

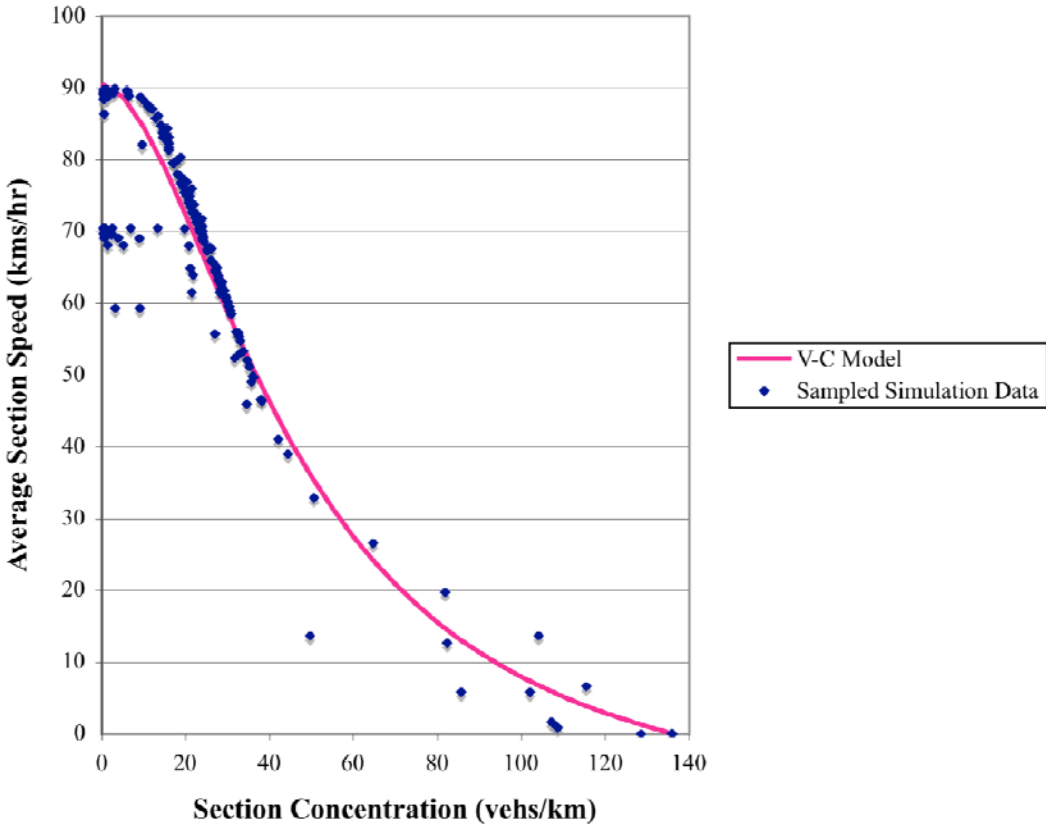


FIGURE 10 Scatter in section values.

In addition, a simple transformation of the proposed function, or fundamental diagram, appears to offer an insight to into how a traffic stream behaves, which takes the proposed model beyond a purely data fitting role. A numerical experiment demonstrates that it provides convincing output, very similar to that which would be expected in a real-life traffic situation.

Examination of the results of different model runs shows that in certain circumstances stop–start traffic flow conditions are created, which again reproduces what is observable in the field. However, the primary purpose of the model is to assess journey times through a road system, rather than to replicate particular traffic phenomena. Other models discussed at the symposium are probably more suited to this task.

A feature of the proposed model is that it can deal seamlessly with bottlenecks and free-flow sections along a route to obtain robust journey times. As is well recognized in many areas of engineering the aggregate behavior of a system can be realistic even if the modeling of individual components is not perfect.

REFERENCES

1. Greenshields, B. D. A Study in Highway Capacity. *Highway Research Board Proceedings*, Vol. 14, 1935, p. 458.
2. Adams, W. F. Road Traffic Considered as a Random Series. In *Journal of Institution of Civil Engineers*, Vol. 4, 1936, United Kingdom, pp 121–130.
3. MacNicholas, M. J. et al. Structure and Calibration of a Model for the Simulation of Route Transit Time. Technical Assistance Programme, Paper 7.7, 1993, Ireland.
4. Erlingsson, S., A. M. Jonsdottir, and T. Thorsteinsson. Traffic Stream Modelling of Road Facilities. Presented at Transport Research Arena, 2006.
5. Hranac, R., E. Sterzin, D. Krachmer, H. Rakha, and M. Farzaneh. *Empirical Studies on Traffic Flow in Inclement Weather*. Study for Federal Highway Administration, 2006.
6. Road Research Laboratory. *Research on Road Traffic* (Table 5.1, p. 193). HMSO, United Kingdom, 1965.
7. Van Aerde, M. Single Regime Speed-Flow-Density Relationship for Congested and Un-congested Highways. Presented at 74th Annual Meeting of the Transportation Research Board, Washington, D.C., 1995.
8. Payne, H. J. Models of Freeway Traffic and Control. In *Simulation Council Proceedings*, Vol. 1, No. 1, June 1972.
9. Ross, P. Modelling Traffic Flow. In *Public Roads*, Vol. 51, No. 3, Federal Highway Administration, 1987.
10. Gerlough, D. L. and M. J. Huber. *Special Report 165: Traffic Flow Theory,* TRB, National Research Council, Washington, D.C., 1975.
11. Pipes, L. A. *Special Report 79: Hydrodynamic Approaches-Part 1: An Introduction to Traffic Flow Theory*, HRB, National Research Council, Washington, D.C., 1961

Calibration of Steady-State Car-Following Models Using Macroscopic Loop Detector Data

HESHAM RAKHA
YU GAO
Virginia Tech Transportation Institute

The paper develops procedures for calibrating the steady-state component of various car-following models using macroscopic loop detector data. The calibration procedures are developed for a number of commercially available macroscopic traffic simulation software, including: CORSIM, AIMSUN2, VISSIM, Paramics, and INTEGRATION. The procedures are then applied to a sample data set for illustration purposes. The paper then compares the various steady-state car-following formulations and concludes that the Gipps and Van Aerde steady-state car-following models provide the highest level of flexibility in capturing different driver and roadway characteristics. However, the Van Aerde model, unlike the Gipps model, is a single-regime model and thus is easier to calibrate given that it does not require the segmentation of data into two regimes. The paper finally proposes that the car-following parameters within traffic simulation software be link specific as opposed to the current practice of coding network-wide parameters. The use of link-specific parameters will offer the opportunity to capture unique roadway characteristics and reflect roadway capacity differences across different roadways.

The rapid development of personal computers over the last few decades has provided the necessary computing power for advanced traffic microsimulators. Today, microscopic traffic simulation software are widely accepted and applied in all branches of transportation engineering as an efficient and cost effective analysis tool. One of the main reasons for this popularity is the ability of microscopic traffic simulation software to reflect the dynamic nature of the transportation system in a stochastic fashion.

The core of microscopic traffic simulation software is a car-following model that characterizes the longitudinal motion of vehicles. The process of car-following consists of two levels, namely modeling steady-state and non-steady-state behavior (*1*). Ozaki defined steady state as conditions in which the vehicle acceleration and deceleration rate is within a range of ±0.05g (*2*). Another definition of steady-state or stationary conditions is provided by Rakha (*3*) as the conditions when traffic states remain practically constant over a short time and distance. Steady-state car-following is extremely critical to traffic stream modeling given that it influences the overall behavior of the traffic stream. Specifically, it determines the desirable speed of vehicles at different levels of congestion, the roadway capacity, and the spatial extent of queues. Alternatively, non-steady-state conditions govern the behavior of vehicles while moving from one steady state to another through the use of acceleration and deceleration models. The acceleration model is typically a function of the vehicle dynamics while the deceleration model ensures that vehicles maintain a safe relative distance to the preceding vehicle thus ensuring that the traffic stream is asymptotically stable. Both acceleration and deceleration models can affect steady-state conditions by reducing queue discharge saturation flow rates.

Traffic stream models describe the motion of a traffic stream by approximating for the flow of a continuous compressible fluid. The traffic stream models relate three traffic stream measures, namely, flow rate (q), density (k), and space-mean-speed (u). Gazis et al. (*4*) were the first to derive the bridge between microscopic car-following and macroscopic traffic stream models. Specifically, the flow rate can be expressed as the inverse of the average vehicle time headway. Similarly, the traffic stream density can be approximated for the inverse of the average vehicle spacing for all vehicles within a section of roadway. Therefore every car-following model can be represented by its resulting steady-state traffic stream model. Different graphs relating each pair of the above parameters can be used to show the steady-state properties of a particular model, including the speed-spacing (u-s) and speed-flow-density (u-q-k) relationships. The latter curve is of more interest, since it is more sensitive to the calibration process and the shape and nose position of the curve determines the behavior of the resulting traffic stream.

A reliable use of microsimulation software requires a rigorous calibration effort. Because traffic simulation software are commonly used to estimate macroscopic traffic stream measures, such as average travel time, roadway capacity, and average speed, the state of the practice is to systematically alter the model input parameters to achieve a reasonable match between desired macroscopic model output and field data (*5*). Since the macroscopic flow characteristics are mostly related to steady-state conditions, this requires the user to calibrate the parameters of the steady-state relationship and therefore the knowledge of the steady-state behavior of the car-following model is necessary in this process. It should be mentioned that under certain circumstances, the non-steady-state behavior can also influence steady-state behavior (*3*); however since this is not the general case, the focus of this paper will be on steady-state conditions.

Over the past decade, several car-following models have been proposed and described in the literature. Brackstone and McDonald (*6*) categorized the car-following models into five groups, namely, Gazis-Herman-Rothery (GHR) models, safety distance models, linear models, psychophysical or action point models, and fuzzy logic–based models. However, as it was mentioned above the measures that are usually used by transportation engineers are those of macroscopic nature, which are mostly affected by car-following models. Consequently, calibrating these software using macroscopic data offers a significant appeal to modelers.

The goals of this paper are twofold. First, the paper identifies the steady-state car-following model for a number of state-of-practice commercial microscopic traffic simulation software. Second, the paper develops a procedure for calibrating these steady-state models using macroscopic loop detector data.

TRAFFIC SIMULATION CAR-FOLLOWING MODELS

The modeling of car-following and traffic stream behavior requires a mathematical representation that captures the most important features of the actual behavior (Table 1). In this treatment, the relationships obtained by observation, experimentation, and reasoning are given: the researcher attempts to express their steady-state behavior in a graphical form, and classify them based on their steady-state representation.

Typically, car-following models characterize the behavior of a following vehicle (vehicle n) that follows a lead vehicle (vehicle n-1). This can be presented by either characterizing the relationship between a vehicles' desired speed and the vehicle spacing (speed formulation), or

alternatively by describing the relationship between the vehicle's acceleration and speed differential between the lead and following vehicles (acceleration formulation).

Over the last few decades, several car-following models have been developed and incorporated within microsimulation software packages. This section describes the characteristics of six of the state-of-practice and state-of-art car-following models, including the Pitt model (CORSIM), Gipps' model (AIMSUN2), Wiedemann74 and 99 models (VISSIM), Fritzsche's model (PARAMICS), and the Van Aerde model (INTEGRATION). Subsequently, each model is characterized based on its steady-state behavior and procedures are developed to calibrate the model parameters.

TABLE 1 Software Car-Following Model Formulations

Software	Model	Formulation
CORSIM	Pitt Model	$u_n(t+\Delta t) = \min\left\{3.6\cdot\left[\frac{s_n(t)-s_j}{c_8} - b\left(u_n(t)-u_{n-1}(t)\right)^2\right], u_f\right\}$
VISSIM	Wiedemann74	$u_n(t+\Delta t) = \min\left\{\begin{array}{l} 3.6\cdot\left(\frac{s_n(t)-s_j}{BX}\right)^2 \\ 3.6\cdot\left(\frac{s_n(t)-s_j}{BX\cdot EX}\right)^2, u_f \end{array}\right\}$
	Wiedemann99	$u_n(t+\Delta t) = \min\left\{\begin{array}{l} u_n(t) + 3.6\cdot\left(CC8 + \frac{CC8-CC9}{80}u_n(t)\right)\Delta t \\ 3.6\cdot\frac{s_n(t)-CC0-L_{n-1}}{u_n(t)} \end{array}, u_f\right\}$
Paramics	Fritzsche	$u_n(t+\Delta t) = \min\left\{\begin{array}{l} 3.6\cdot\left(\frac{AD-A_0}{T_D}\right) \\ 3.6\cdot\left(\frac{AR-A_0}{T_r}\right) \end{array}, u_f\right\}$
AIMSUN2	Gipps	$u_n(t+T) = \min\left\{\begin{array}{l} u_n(t) + 3.6\left[2.5a_{max}T\left(1-\frac{u_n(t)}{u_f}\right)\sqrt{0.025+\frac{u_n(t)}{u_f}}\right], \\ 3.6\left[-bT + \sqrt{b^2T^2 + b\left(2\left[s_n(t)-L_{n-1}\right] - \frac{u_n(t)}{3.6}T + \frac{u_{n-1}(t)^2}{3.6^2\times b'}\right)}\right] \end{array}\right\}$
INTEGRATION	Van Aerde	$u_n(t+\Delta t) = \min\left\{\begin{array}{l} u_n(t) + 3.6\cdot\frac{F_n(t)-R_n(t)}{m}\Delta t, \\ \frac{-c_1' + c_8u_f + \tilde{s}_n(t) - \sqrt{\left[c_1' - c_8u_f - \tilde{s}_n(t)\right]^2 - 4c_8\left[\tilde{s}_n(t)u_f - c_1'u_f - c_2\right]}}{2c_8} \end{array}\right\}$ Where: $\tilde{s}_n(t) = s_n(t) + \left[u_{n-1}(t+\Delta t) - u_n(t)\right]\Delta t + 0.5a_{n-1}(t+\Delta t)\Delta t^2$

It should be noted again that this study only describes car-following behavior under steady-state conditions, when the lead vehicle is traveling at similar speeds and both the lead and following having similar car-following behavior, i.e., $s_n \approx s_{desired}$, $\Delta u_n \approx 0$, where s_n.is the spacing between the lead vehicle (vehicle n-1) and following vehicle (vehicle n) and Δu_n is the relative speed between the lead and following vehicle ($u_{n-1} - u_n$). In addition to these two conditions, we are capturing the average behavior given that driver behavior is stochastic in nature. The analysis of randomness was presented in an earlier publication (*7*) and thus is not considered further in this research effort.

CORSIM SOFTWARE

CORSIM was developed by the Federal Highway Administration (FHWA) and combines two traffic simulation models: NETSIM for surface streets and FRESIM for freeway roadways. The FRESIM model utilizes the Pitt car-following model that was developed by the University of Pittsburgh (*8*). The basic model incorporates the vehicle spacing and speed differential between the lead and following vehicle as two independent variables as demonstrated in and cast as

$$s_n(t) = s_j + c_3 \frac{u_n(t+\Delta t)}{3.6} + bc_3 \frac{\Delta u_n(t+\Delta t)^2}{3.6^2} \quad (1)$$

where $s_n(t)$ is the vehicle spacing between the front bumper of the lead vehicle and front bumper of following vehicle at time t (m), s_j is the vehicle spacing when vehicles are completely stopped in a queue (m), c_3 is the driver sensitivity factor (s), b is a calibration constant that equals 0.1 if the speed of the following vehicle exceeds the speed of the lead vehicle, otherwise it is set to zero (h/km), Δu is the difference in speed between lead and following vehicle (km/h) at instant $t+\Delta t$, and u_n is the speed of the following vehicle at instant t (km/h).

Given that steady-state conditions are characterized by travel at near equal speeds, the third term of the car-following model tends to zero under steady-state driving. Consequently, the steady-state car-following model that is incorporated within FRESIM can be written as

$$s_n(t) = s_j + c_3 \frac{u_n(t+\Delta t)}{3.6} \quad (2)$$

Introducing a constraint on the vehicle speed based on the roadway characteristics and roadway speed limit, the car-following model can be written as

$$u_n(t+\Delta t) = \min\left\{u_f, 3.6\left(\frac{s_n(t) - s_j}{c_3}\right)\right\}. \quad (3)$$

Rakha and Crowther (*9*) demonstrated that the steady-state car-following behavior is identical to the Pipes or the GM-1 model. Furthermore, if we assume that all vehicles are similar in behavior, the vehicle subscripts can be dropped from the formulation. The model then requires the calibration of three parameters, namely: the facility free-flow speed, the facility jam density, and a driver sensitivity factor (DSF) c_3. In the case of the NETSIM software the parameter is fixed

and equal to 1/3,600, however in the case of the FRESIM model Rakha and Crowther (*9*) showed that the DSF can be related to macroscopic traffic stream parameters as

$$c_s = 3600\left(\frac{1}{q_c} - \frac{1}{k_j u_f}\right) \tag{4}$$

where q_c is the mean saturation flow rate (veh/h), k_j is the mean roadway jam density (veh/km), and u_f is the space-mean traffic stream free-flow speed (km/h). For example, considering a freeway facility with an average lane capacity of 2,400 veh/h/lane, an average free-flow speed of 100 km/h, and an average jam density of 150 veh/km/lane; the DSF can be computed as 1.26 s, as summarized in Table 2. In other words, the modeler would need to input an average DSF of 1.26 s, a free-flow speed of 100 km/h, and an average vehicle spacing of 6.67 m in order to simulate a saturation flow rate of 2,400 veh/h/lane. The estimation of the three macroscopic traffic stream parameters q_c, u_f, and k_j using loop detector data is described later in the paper. Rakha and Crowther (*9*) demonstrated the calibration of the DSF can be achieved by changing a base networkwide parameter and changing link-specific adjustment parameters.

TABLE 2 Steady-State Model Calibration

Car-following Model	Steady-State Calibration
Pitt Model	$c_s = 3600\left(\frac{1}{q_c} - \frac{1}{k_j u_f}\right)$
Wiedemann 74	$E(BX) = 1000\sqrt{3.6}\sqrt{u_f}\left(\frac{1}{\alpha q_c} - \frac{1}{k_j u_f}\right)$ and $E(EX) = \frac{\frac{k_j u_f}{q_c} - 1}{\frac{k_j u_f}{\alpha q_c} - 1} \simeq \alpha$
Wiedemann 99	$CC0 = \frac{1000}{k_j} - \bar{L}$ and $CC1 = 3600\left(\frac{1}{q_c} - \frac{1}{k_j u_f}\right)$
Gipps	b=b′: $T = 2400\left(\frac{1}{q_c} - \frac{1}{k_j u_f}\right)$ b>b′: Invalid behavior with a non-concave car-following relationship b<b′: $b' = \frac{b}{\left(1 - \frac{25920b}{k_j u_c^2}\right)}$ and $T = 2.4\left(\frac{1000}{q_c} - \frac{1000}{k_j u_c} - \frac{u_c}{25.92b}\left(1 - \frac{b}{b'}\right)\right)$
Fritzsche	$A_0 = \frac{1000}{k_j}$; $T_D = 3600\left(\frac{1}{q_c} - \frac{1}{k_j u_f}\right)$; and $T_r = 3600\left(\frac{1}{q_c^{max}} - \frac{1}{k_j u_f}\right)$
Van Aerde	$c_1 = \frac{u_f}{k_j u_c^2}\left(2u_c - u_f\right)$; $c_2 = \frac{u_f}{k_j u_c^2}\left(u_f - u_c\right)^2$; $c_3 = \left(\frac{1}{q_c} - \frac{u_f}{k_j u_c^2}\right)$

The calibration procedure was applied to a sample arterial data set in which the traffic stream space-mean speed is sensitive to the flow rate in the uncongested regime, as illustrated in Figure 1. Because the Pipes model assumes that the traffic stream speed remains constant regardless of the flow rate in the uncongested regime, the model is not suitable for such applications. Furthermore, the model assumes that the speed-at-capacity is identical to the free-flow speed, which is not the case in this data set. It should be noted that the capacity for this example is fairly low given that it is measured upstream of a traffic signal.

The AIMSUN2 car-following behavior is modeled using the Gipps car-following model (*10–12*) and presented in Table 2. According to Gipps, the speed of the following vehicle is controlled by three conditions. The first condition ensures that the vehicle does not exceed its desired speed or a vehicle-specific free-flow speed (U_n). The second condition ensures that the vehicle accelerates to its desired speed with an acceleration rate that initially increases with speed and then decreases to zero as the vehicle approaches its desired speed. The combination of these conditions results in Equation 5 which controls the vehicle acceleration while vehicles are distant from each other (free-flow behavior). The equation coefficients were obtained from fitting a curve to field data collected on a road of moderate traffic.

$$u_n(t+T) = u_n(t) + 3.6\left[2.5a_nT\left(1-\frac{u_n(t)}{U_n}\right)\sqrt{0.025+\frac{u_n(t)}{U_n}}\right] \quad (5)$$

where $u_n(t)$ is the speed of vehicle n at time t (km/h); a_n is the maximum desired acceleration rate of vehicle n (m/s^2); T it the driver's reaction time (s); and U_n is the desired speed of vehicle n or the vehicle-specific free-flow speed (km/h).

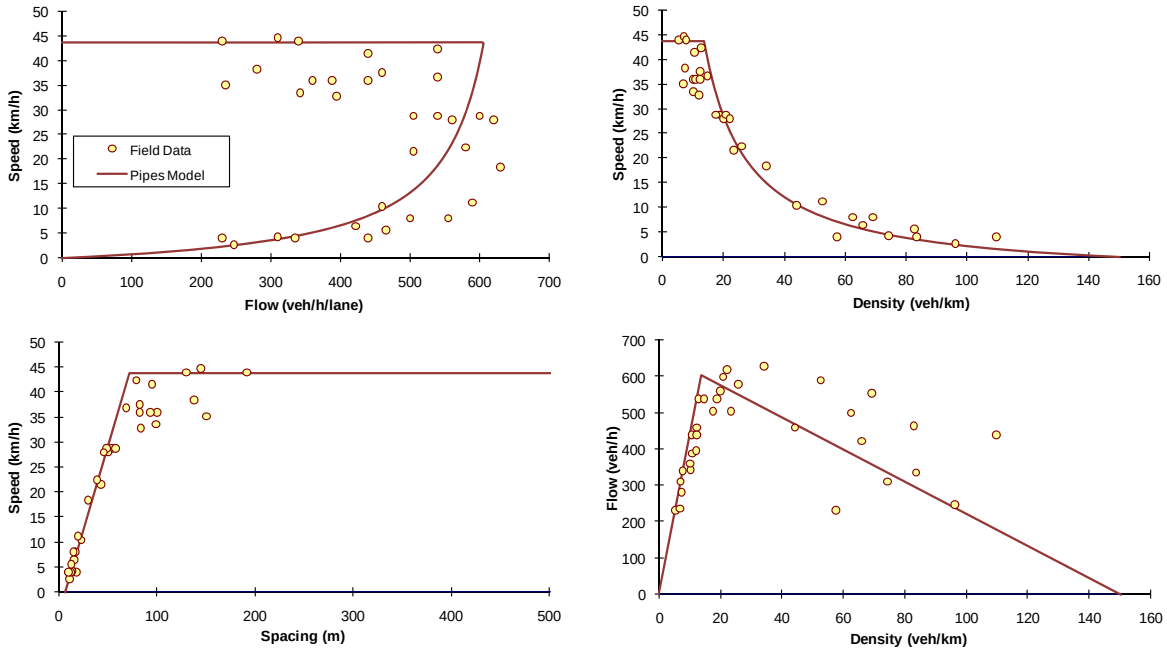


FIGURE 1 Example illustration of pipes model calibration AIMSUN2 software.

In a constrained traffic situation, when vehicles are traveling close to each other, the third condition becomes dominant and controls the behavior of the follower vehicle while decelerating. The speed of the follower vehicle (see Equation 6) is affected by the driver reaction time, the spacing between the leader and follower vehicles, the speed of the leader and follower vehicles, and the deceleration rates they are willing to employ. Gipps pointed out that a safety margin should be added to the driver's reaction time. The safety margin would assure the vehicle's ability to stop even when there is a delay to initiate its reaction for some reason. The safety margin was assumed to be constant in value and equal to $T/2$ (half the reaction time). This safety value is implicit in Equation 6.

$$u_n(t+T) = 3.6\left[-bT + \sqrt{b^2T^2 + b\left\{2\left[s_n(t) - L_{n-1}\right] - \frac{u_n(t)}{3.6}T + \frac{u_{n-1}(t)^2}{3.6^2 \times b'}\right\}}\right] \tag{6}$$

Here b and b' are deceleration parameters of vehicle n (m/s^2); b is the actual most severe deceleration rate the vehicle is willing to employ in order to avoid a collision; and b' is the estimated most severe deceleration rate the leader vehicle is willing to employ. It is an estimated value because it is impossible for the follower to evaluate the real intention of his/her leader; L_{n-1} is the effective length of vehicle n-1 (the actual length plus a safety margin); $s_n(t)$ is the spacing between vehicle n and n-1 at time t (m); and $u_{n-1}(t)$ is the speed of the preceding vehicle (km/h).

The parameters related to deceleration rates (b and b') are very important for the braking process modeling. These parameters influence the spacing between the follower and leader vehicles and thus affect the lane capacity.

Assuming the vehicles will travel as close to their desired speed as possible and considering the dynamics limitations, the speed of vehicle n at time $t + T$ can be computed as

$$u_n(t+T) = \min\left\{\begin{array}{l} u_n(t) + 3.6\left[2.5a_nT\left(1 - \frac{u_n(t)}{U_n}\right)\sqrt{0.025 + \frac{u_n(t)}{U_n}}\right], \\ 3.6\left[-bT + \sqrt{b^2T^2 + b\left\{2\left[s_n(t) - L_{n-1}\right] - u_n(t)T + \frac{u_{n-1}(t)^2}{b'}\right\}}\right] \end{array}\right. \tag{7}$$

According to the above formulation, once the road is unconstrained and the space headways between the vehicles are large enough to allow them to travel at their desired speed, the first argument of Equation 7 is applied. In this case, the following vehicle is able to accelerate according to the empirical equation of vehicle dynamics. Alternatively, in congested conditions, where short headways are typical, the second argument of Equation 7 is applied. In such a case, the speed is limited by the leader vehicle performance. Each vehicle establishes its speed in order to avoid a collision based on the assumption that the leader deceleration rate will not exceed b'.

A detailed mathematical analysis of Gipps' car-following model under steady-state conditions was presented in two earlier publications (*11*, *12*). Consequently, the paper will only summarize the major findings of these studies and then develop an analytical calibration procedure of the model. In his study, Wilson (*12*) presented a mathematical analysis of simplified scenarios and identified parameter regimes that deserve further investigation. The paper also showed the derivation of uniform flow solutions (steady-state) and speed-spacing functions under simplifying conditions concerning parameters b, b', and T, and an analysis of the

linear stability of the uniform flow, identifying stable and nonstable flow regimes. Wilson demonstrated that the steady-state car-following model can be cast as

$$s = s_j + \frac{1}{2.4}Tu + \frac{1}{25.92b}\left(1 - \frac{b}{b'}\right)u^2 \tag{8}$$

Rakha et al. (*11*) demonstrated that in the case that b and b' are identical the driver reaction time can be computed as

$$T = 2400\left(\frac{1}{q_o} - \frac{1}{k_j u_f}\right) \tag{9}$$

When b is set greater than b', Wilson (*12*) demonstrated that the car-following relationship may become unphysical and produce multiple solutions for some sets of parameters. Consequently, b should be set less than or equal to b'. In the case that b is less than b', Rakha et al. (*11*) demonstrated that the steady-state car-following relationship can be cast as

$$u = \min\left(u_f, \frac{5.4bT}{\left(1 - \frac{b}{b'}\right)}\left[-1 + \sqrt{1 + \frac{8000\left(\frac{1}{k} - \frac{1}{k_j}\right)\left(1 - \frac{b}{b'}\right)}{9bT^2}}\right]\right) \tag{10}$$

Starting with Equation 8 the speed-flow relationship can be derived as

$$q = \frac{1000u}{s_j + \frac{1}{2.4}Tu + \frac{1}{25.92b}\left(1 - \frac{b}{b'}\right)u^2} \tag{11}$$

Using the function of Equation 11 Rakha et al. developed lookup tables to estimate the facility capacity considering different microscopic car-following parameters. This paper extends the research by developing analytical expressions to estimate the microscopic car-following model parameters based on macroscopic traffic stream measurements.

Considering that the maximum flow rate occurs when the first derivative of flow with respect to speed equals to zero, the speed-at-capacity can be computed as

$$u_o = \min\left(3.6 \times \sqrt{\frac{2000b}{k_j\left(1 - \frac{b}{b'}\right)}}, u_f\right) \tag{12}$$

Consequently, we derive the relationship between the microscopic car-following and macroscopic traffic stream parameters as

$$b = \frac{1}{\left(\frac{1}{b'} + \frac{25920}{k_j u_o^2}\right)} \tag{13}$$

where $b < b'$, and

$$T = 2.4\left(\frac{1000}{q_o} - \frac{1000}{k_j u_o} - \frac{u_o}{25.92b}\left(1 - \frac{b}{b'}\right)\right) \tag{14}$$

The calibration of the model entails assuming the most severe deceleration rate the driver is willing to employ (b') and then computing b using Equation 13 for a desired facility-specific mean speed-at-capacity and jam density. The reaction time (T) can then be computed using Equation 14, as demonstrated in Table 2. It should be noted that in the case that $b = b'$ Equation 14 reverts to Equation 9 given that the speed-at-capacity equals the free-flow speed as computed using Equation 12.

The calibration procedure was applied to the same sample data set gathered along an arterial, as illustrated in Figure 2. The figure demonstrates a reasonable fit to the data, however given that the data demonstrate that traffic stream speed is sensitive to the traffic stream flow in the uncongested regime; the model offers a suboptimal fit to the field data for the uncongested regime with a good fit for the congested regime. The speed-at-capacity is different from the free-flow speed and thus the model is able to capture this phenomenon.

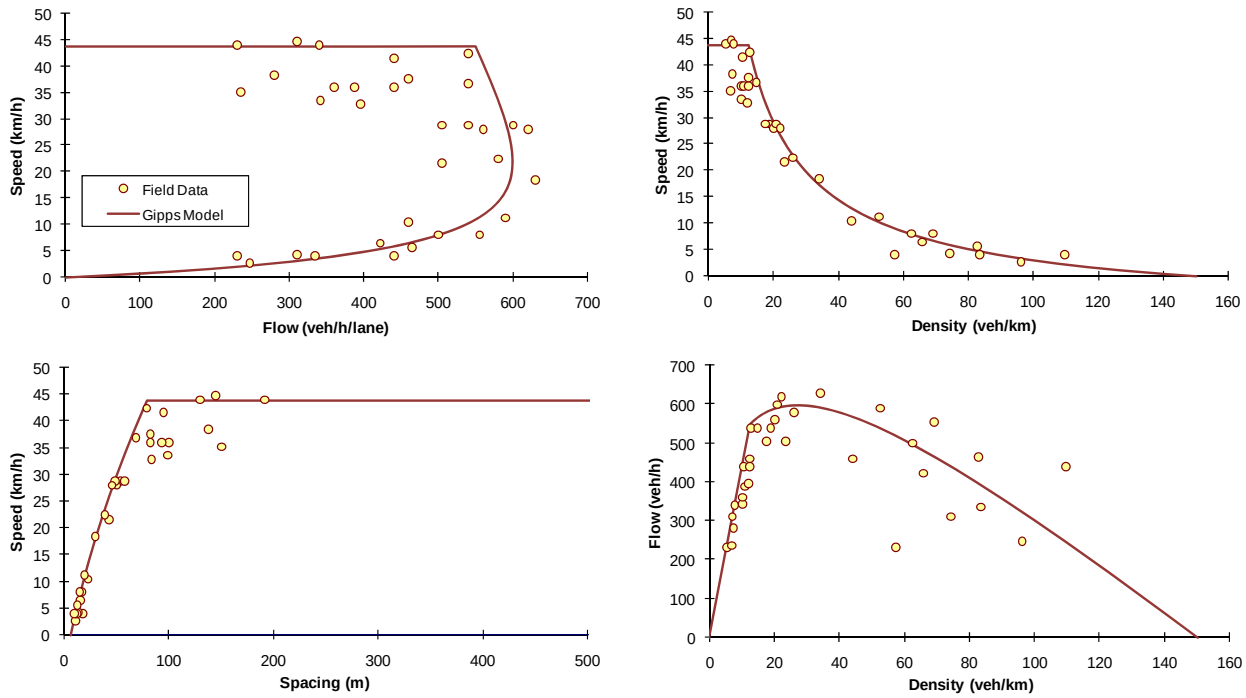


FIGURE 2 Example illustration of Gipps model calibration.

VISSIM SOFTWARE

The car-following model used in VISSIM is a modified version of two models developed by Wiedemann (Wiedemann74 and 99 models) and belongs to a family of models known as psychophysical or action-point models. This family of models uses thresholds or action points where the driver changes his/her driving behavior. Drivers react to changes in spacing or relative speed only when these thresholds are crossed. The thresholds and the regimes they define are usually presented in the relative speed/spacing diagram for a pair of lead and follower vehicles.

First the Wiedemann74 model is described followed by a description of the Wiedemann 99 model. For the purposes of this study only the area identified as steady-state is of interest. This area as was mentioned before has the following steady-state criteria ($s_n \approx s_{desired}$, $\Delta u_n \approx 0$). In the case of the Wiedemann74 model, the desired vehicle spacing is an interval ($ABX \le s \le SDX$) instead of a single value as was the case with previously mentioned models. Given that $\Delta u_n \approx 0$, only the boundaries of desired vehicle spacing interval (*ABX* & *SDX*) determine the steady-state characteristics of the VISSIM car-following model. The expected value of *ABX* and *SDX* parameters can be calculated as

$$E(AX) = s_j + AXadd + AXmult \cdot E(RND1_n) = s_j + 0.5 \approx s_j \quad (15)$$

$$E(ABX) = E(AX) + E(BX)\sqrt{u} = s_j + E(BX)\sqrt{u} \qquad u \le u_{desired} \quad (16)$$

$$E(SDX) = s_j + E(BX) \cdot E(EX)\sqrt{u}, \qquad u \le u_{desired} \quad (17)$$

where the *BX* and *EX* random variables are computed as

$$BX = BXadd + BXmult \cdot RND1_n \quad (18)$$

$$EX = EXadd + BXmult \cdot (NRND - RND2_n) \quad (19)$$

Here $RND1_n$ and $RND2_n$ are user specified vehicle-specific (where *n* is the vehicle index) normally distributed random variables with a default mean value of 0.5 and a standard deviation of 0.15. *NRND* is also a normally distributed random variable with a default mean value of 0.5 and standard deviation of 0.15. The expectation of *SDX* given as E(*SDX*) ranges between 1.5 to 2.5 times the expected value of *ABX* (E(*ABX*)), where BX_{add}, BX_{mult} , EX_{add}, and EX_{mult} are user-defined calibration parameters.

Equations 16 and 17 demonstrate that the parameters *ABX* and *SDX* are not internally constrained and thus an external maximum speed constraint ($u \le u_{desired}$) must be enforced. Given that the desired speed is insensitive to traffic conditions ($u_{desired} = u_c = u_f$), the uncongested steady-state behavior has a flat top, as illustrated in Figure 3.

In order to calibrate the steady-state Wiedemann74 model the following calibration procedure is developed. The calibration of the Wiedemann74 model can be achieved by deriving the speed-flow relationship for the congested regime as

$$q = \frac{1000u}{\frac{1000}{k_j} + \frac{E(BX)E(EX)}{\sqrt{3.6}}\sqrt{u}} \quad (20)$$

Here u is the traffic stream space-mean speed (km/h); q is the traffic stream flow rate (veh/h), and k_j is the traffic stream density (veh/km). By taking the derivative of flow with respect to speed the relationship is demonstrated to be a strict monotonically increasing function as shown in Equation 21.

$$\frac{dq}{du} = \frac{1000\left(\frac{1000}{k_j} + \frac{E(BX)\cdot E(EX)}{2\cdot\sqrt{3.6}}\sqrt{u}\right)}{\left(\frac{1000}{k_j} + \frac{E(BX)\cdot E(EX)}{\sqrt{3.6}}\sqrt{u}\right)^2} > 0 \tag{21}$$

Consequently, the maximum flow occurs at the boundary of the relationship and thus at the maximum desired or free-flow speed. As was the case with the Pipes' model, the speed-at-capacity equals the free-flow speed. By inputting the maximum flow (capacity) and free-flow speed in Equation 20, removing the E(*EX*) term to compute the capacity upper bound, and re-arranging the equation; the expected value of *BX* can be computed as

$$E(BX) = 1000\sqrt{3.6}\sqrt{u_f}\left(\frac{1}{\alpha q_c} - \frac{1}{k_j u_f}\right) \tag{22}$$

By considering that the expected value of *SDX* is α times the expected value of *ABX* (i.e. E(*SDX*) = α × E(*ABX*)), where the parameter α ranges from 1.5 to 2.5; the expected value of *EX* can be computed as

$$E(EX) = \frac{\frac{k_j u_f}{q_c} - 1}{\frac{k_j u_f}{\alpha q_c} - 1} \simeq \alpha \tag{23}$$

Given that $k_j u_f / q_c$ is typically very large, the expected value of *EX* is approximately equal to the parameter α.

The proposed calibration procedure was applied to the same arterial dataset and the fit is illustrated in Figure 3. Again, as was the case with the Pipes' model the fit to the field data is unable to reflect the reduction in traffic stream speed as the arrival rate increases in the uncongested regime. Furthermore, the curvature of the car-following model (speed-spacing diagram) contradicts typical driver behavior (curvature is convex instead of concave). The model does provide a range of behavior for the congested regime as illustrated by the two lines.

In an attempt to validate the calibration procedure a simple network was coded and simulated using the VISSIM software. The network was composed of two single-lane links in order to isolate the car-following behavior (i.e., remove any possible impact that lane-changing behavior might have on the traffic stream performance). Initially the capacity of both links was set equal using the proposed calibration procedure. The arrival rate was increased gradually until it exceeded the capacity of the entrance link. The traffic stream flow and speed were measured

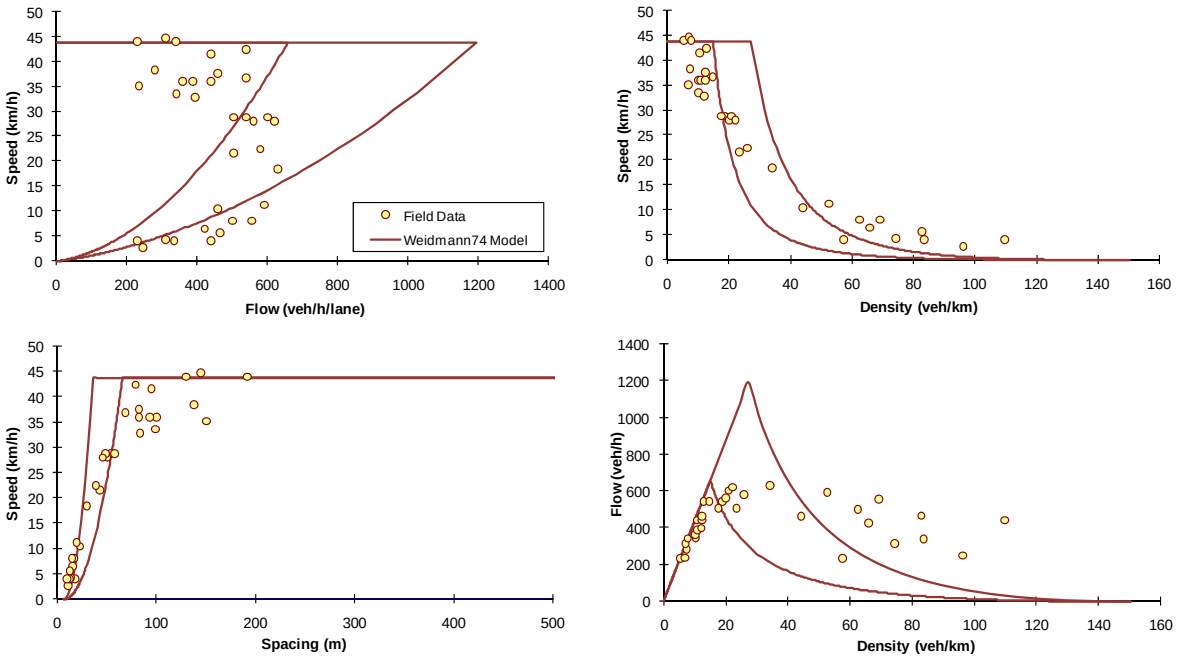


FIGURE 3 Sample calibration of the Wiedemann74 model.

using a number of loop detectors along the first link. The default model randomness was set (AXadd = 2, BXadd = 3, and BXmult = 4). The results demonstrate that the uncongested regime is flat as was suggested earlier, as illustrated in Figure 4. Subsequently, the capacity at the downstream link was reduced by selecting the input parameters using the calibration procedures presented earlier. The demand was fixed at the capacity of the upstream link and thus a bottleneck was created at the entrance to Link 2. The departure flow rate and speed were directly measured upstream of the bottleneck to construct the congested regime of the fundamental diagram. As demonstrated in Figure 4 the simulated data appear to initially follow the ABX curve and then move towards the *SDX* curve as the capacity of the downstream bottleneck increases. Given that the movement between the two regimes is not documented in the literature it is not clear how this is done. The figure clearly demonstrates that the proposed calibration procedures are consistent with the VISSIM model output.

The VISSIM software also offers a second car-following model, namely the Wiedemann99 model. The model is formulated as

$$u_n(t+\Delta t) = \min\left\{\begin{array}{l} u_n(t) + 3.6 \cdot \left(CC8 + \dfrac{CC8 - CC9}{80} u_n(t)\right)\Delta t \\ 3.6 \cdot \dfrac{s_n(t) - CC0 - L_{n-1}}{u_n(t)} \end{array}, u_f\right\} \quad (24)$$

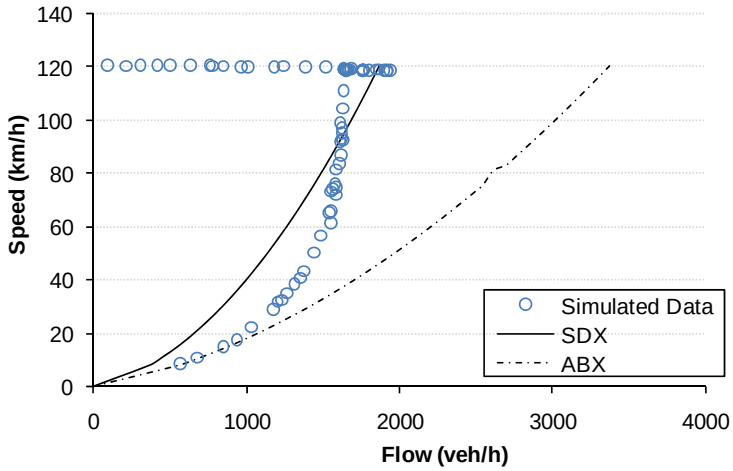


FIGURE 4 Wiedemann74 calibration procedure validation.

This model, as was the case with the Gipps model, computes the vehicle speed as the minimum of two speeds: one based on the vehicle acceleration restrictions and the other based on a steady-state car-following model. The model considers a vehicle kinematics model with a linear speed-acceleration relationship where *CC8* is the maximum vehicle acceleration at a speed of 0 km/h (m/s^2) and *CC9* is the maximum vehicle acceleration at a speed of 80 km/h (m/s^2). The VISSIM software also allows the user to input a user-specified vehicle kinematics model that appears to override the linear model. This user-specified relationship allows the user to modify the desired and maximum driver speed-acceleration relationship. The second term of Equation 24 computes the vehicle's desired speed using a linear car-following model and thus is identical to the Pipes model.

Consequently, as was done with the Pipes model, the model constants *CC0* and *CC1* (also known as the driver sensitivity factor) can be computed as

$$CC0 = \frac{1000}{k_j} - \bar{L} \tag{25}$$

and

$$CC1 = 3600\left(\frac{1}{\dot{q}_o} - \frac{1}{k_j u_f}\right) \tag{26}$$

where *CC0* is the spacing between the front bumper of the subject vehicle and the rear bumper of the lead vehicle. This equals the jam density spacing minus the average vehicle length. The DSF (*CC1*) can be calibrated using three macroscopic traffic stream parameters, namely, the expected roadway capacity, jam density, and free-flow speed.

PARAMICS SOFTWARE

The car-following model utilized in the Paramics software, as was the case with the VISSIM software, is a psychophysical car-following model that was developed by Fritzsche (*13*).

Fritzsche's model uses the same modeling concept as the Wiedemann74 car-following model. The difference between these two models is the way thresholds are defined and calculated. Figure 5 depicts the Fritzsche model's thresholds in the $\Delta u - \Delta x$ plane.

The area corresponding to steady-state conditions is almost identical to Wiedemann's car-following model. The vehicle spacing for this regime lies between the desired spacing (AD) and the risky spacing (AR). These two boundaries are determined as

$$AR = A_0 + T_r \cdot \frac{u_n}{3.6} \tag{27}$$

and

$$AD = A_0 + T_D \cdot \frac{u_{n-1}}{3.6} \tag{28}$$

where A_0 is the vehicle spacing at jam density, T_r is the risky time gap (usually 0.5 s), T_D is the desired time gap (with a recommended value of 1.8 s). The resulting steady-state car-following model can be written as

$$u_n(t+\Delta t) = \min\left\{\begin{matrix} 3.6 \cdot \left(\frac{AD - A_0}{T_D}\right) \\ 3.6 \cdot \left(\frac{AR - A_0}{T_r}\right) \end{matrix}, u_f\right\}$$

Similar to the Wiedemann car-following model, the desired speed constraint must be enforced externally. Again, as was the case with the Wiedemann74 model, the relationship provides a range of car-following behavior within the congested regime. Unlike the Wiedemann74 model,

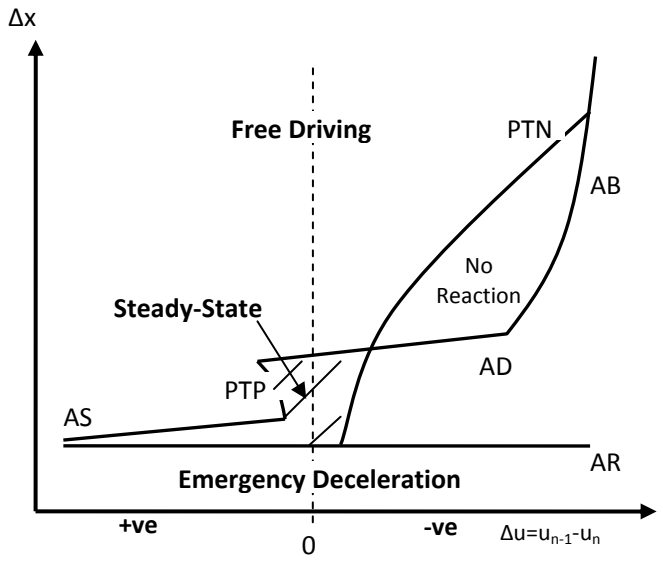


FIGURE 5 Fritzsche's car-following model.

the car-following model is linear and thus a Pipes model. Using similar calibration procedures, the various car-following model parameters are related to macroscopic traffic stream parameters as

$$A_0 = \frac{1000}{k_j} \tag{29}$$

$$T_D = 3600\left(\frac{1}{q_c} - \frac{1}{k_j u_f}\right) \tag{30}$$

and

$$T_r = 3600\left(\frac{1}{q_c^{max}} - \frac{1}{k_j u_f}\right) \tag{31}$$

The calibration procedure was applied to the same arterial data set and the results are similar to those of the Wiedemann74 model, as illustrated in Figure 6. It should be noted that the car-following model provides a range of data in the congested regime considering a linear car-following modeling.

INTEGRATION SOFTWARE

The steady-state functional form that is utilized in the INTEGRATION software is the Van Aerde nonlinear functional form that was proposed by Van Aerde (*14*) and Van Aerde and Rakha (*15*), which is formulated as

$$s_n(t) = c_1 + c_3 u_n(t + \Delta t) + \frac{c_2}{u_f - u_n(t + \Delta t)} \tag{32}$$

where c_1, c_2, and c_3 are model constants. Demarchi (*16*) demonstrated that by considering three boundary conditions the model constants can be computed as

$$c_1 = \frac{u_f}{k_j u_c^2}\left(2u_c - u_f\right); \quad c_2 = \frac{u_f}{k_j u_c^2}\left(u_f - u_c\right)^2; \quad c_3 = \left(\frac{1}{q_c} - \frac{u_f}{k_j u_c^2}\right). \tag{33}$$

As was demonstrated by Rakha and Crowther (*17*) this functional form amalgamates the Greenshields and Pipes car-following models.

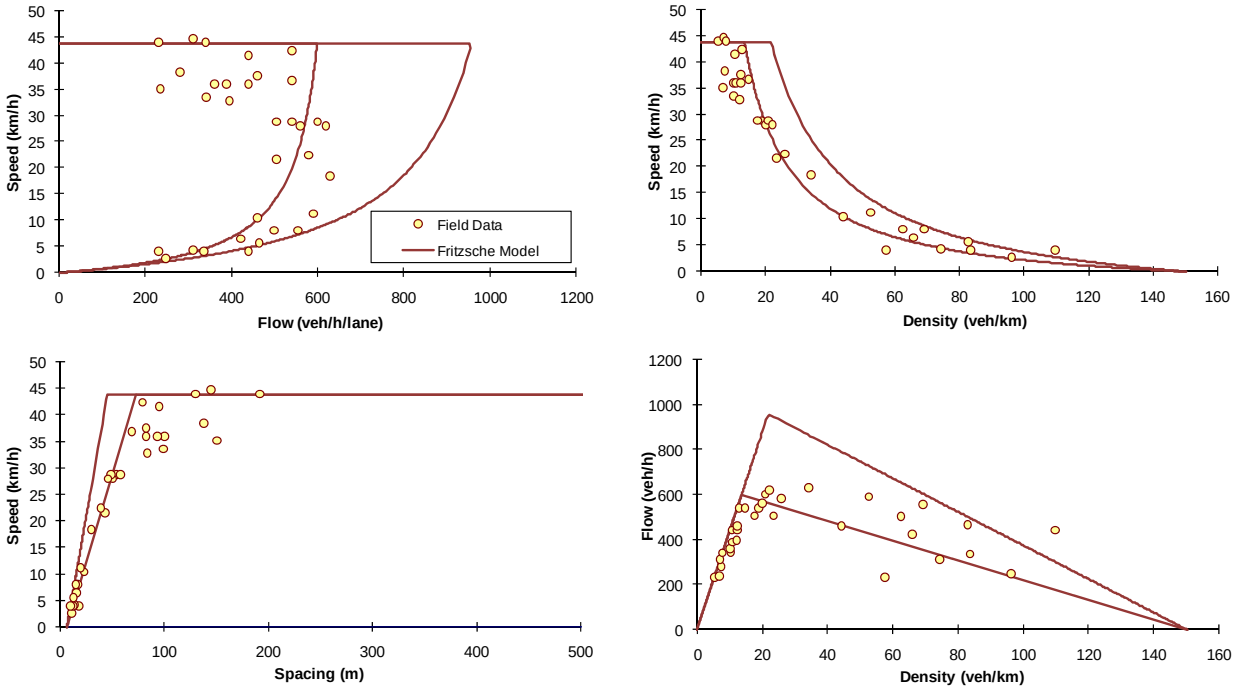


FIGURE 6 Sample calibration of the Fritzsche model.

Ignoring differences in vehicle behavior within a traffic stream and considering the relationship between traffic stream density and traffic spacing, the speed–density relationship can be derived as

$$k = \frac{1000}{c_1 + \frac{c_2}{u_f - u} + c_3 u} \tag{34}$$

Of interest is the fact that Equation 34 reverts to Greenshields' linear model, when the speed-at-capacity and density-at-capacity are both set equal to half the free-flow speed and jam density, respectively (i.e., $u_c = u_f/2$ and $k_c = k_j/2$). Alternatively, setting $u_c = u_f$ results in the linear Pipes model given that

$$c_1 = \frac{1}{k_j} = s_j; \quad c_2 = 0; \quad c_3 = \frac{1}{q_c} - \frac{1}{k_j u_f}$$

Rakha (*3*) demonstrated that the wave speed at jam density (denoted as w_j) can be computed by differentiating the speed–density relationship with respect to density at jam density, to be

$$w_j = k_j \left.\frac{du}{dk}\right|_{k_j} = -s_j \left.\frac{du}{ds}\right|_{s_j} \tag{35}$$

By applying Equation 32 to 34 and ignoring differences between vehicles Rakha derived

$$w_j = -s_j \left.\frac{1}{\frac{ds}{du}}\right|_{u=0} = -\frac{s_j}{c_3 + \frac{c_2}{u_f^2}} = -\frac{u_f}{k_j\left(c_3 u_f^2 + c_2\right)} = -\left[\left(\frac{k_j}{q_c} - \frac{u_f}{u_c^2}\right) + \frac{\left(u_f - u_c\right)^2}{u_f u_c^2}\right]^{-1} \quad (36)$$

Considering, a typical lane capacity of 2,400 veh/h, a free-flow speed of 110 km/h (which is typical of U.S. highways), and a jam density of 140 veh/km/lane, the wave velocity at jam density ranges between approximately -11.5 and -20.3 km/h, when the speed-at-capacity is varied from 80% to 100% the free-flow speed (which is typical on North American freeways).

As was demonstrated earlier, the Van Aerde model reverts to the Pipes linear model when the speed-at-capacity is set equal to the free-flow speed. Consequently, it can be demonstrated that under this condition the wave speed of Equation 36 reverts to

$$w = -\frac{q_c u_f}{k_j u_f - q_c} \quad (37)$$

which is the speed of the linear model. Furthermore, when $u_c = u_f/2$ and $k_c = k_j/2$ the wave speed at jam density is consistent with the Greenshields model estimates and is computed as

$$w_j = -u_f \quad (38)$$

Field observations demonstrate a concave speed-headway relationship. Consequently, the derivative of the speed-density relationship was computed as

$$\frac{du}{ds} = \frac{1}{c_3 + \frac{c_2}{\left(u_f - u\right)^2}} = \frac{\left(u_f - u\right)^2}{c_3\left(u_f - u\right)^2 + c_2} \quad (39)$$

Given that the c_2, c_3, and u_f parameters are always positive, Rakha (*3*) demonstrated the function is a strictly increasing monotonic function. Alternatively, the speed-density relationship is a strictly decreasing monotonic function as

$$\frac{du}{dk} = \frac{du}{ds} \times \frac{ds}{dk} = -\frac{\left(u_f - u\right)^2}{c_3\left(u_f - u\right)^2 + c_2} \times \frac{1}{k^2} \quad (40)$$

While a strict monotonic function is desired from a theoretical standpoint, it is not necessarily reflective of real-life driving behavior. For example, drivers might abide by a facility speed limit if they are the only vehicle on a roadway, however if other vehicles are present on the roadway slower drivers might be encouraged to follow faster vehicles recognizing the lower likelihood of being ticketed for over-speeding. This behavior may only hold when the traffic stream density is very low but contradicts typical traffic flow theory.

The Van Aerde model was calibrated to the same arterial data that were presented earlier, as illustrated in Figure 7. The figure demonstrates that the model is extremely flexible and thus is capable of providing a good fit to the field data for the entire range of data both in the uncongested and congested regimes. It should be noted that the fit provides the expected relationship. Differences in driver behavior can be captured by introducing differences in the four traffic stream parameters, namely: free-flow speed, speed-at-capacity, capacity, and jam density.

TRAFFIC STREAM MODEL CALIBRATION

The estimation of the four traffic stream parameters (u_f, u_c, q_c, and k_j) requires the calibration of a traffic stream model to loop detector data. This effort entails four decisions, namely: (1) define the functional form to be calibrated, (2) identify the dependent and the independent variables, (3) define the optimum set of parameters, and (4) develop an optimization technique to compute the set of parameter values. Van Aerde and Rakha (*15*) and later Rakha and Arafeh (*18*) developed a calibration approach that minimizes the orthogonal error about the 3-D fundamental diagram to estimate the expected value of the four traffic stream parameters. The model is briefly described here; however a more detailed description is provided elsewhere (*18*). The approach is unique because it does not require the identification of dependent and independent variables since it applies a neutral regression approach (minimizes the orthogonal error).

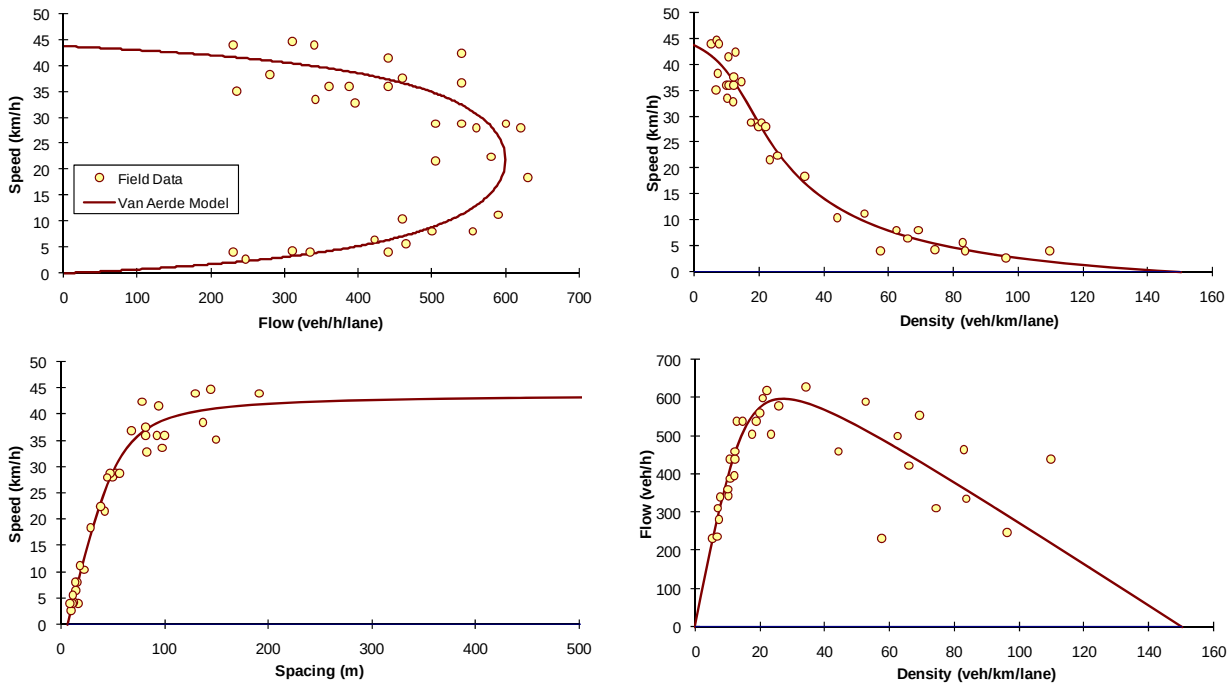


FIGURE 7 Sample calibration of the Van Aerde model.

If we consider the Van Aerde functional form given that it provides the highest level of flexibility, as was demonstrated in the previous section, the optimization model can be formulated as

$$\text{Min} \qquad E = \sum_i \left\{ \left(\frac{u_i - \hat{u}_i}{\bar{u}} \right)^2 + \left(\frac{q_i - \hat{q}_i}{\bar{q}} \right)^2 + \left(\frac{k_i - \hat{k}_i}{\bar{k}} \right)^2 \right\}. \tag{41}$$

S.T.

$$\hat{k}_i = \frac{1}{c_1 + \frac{c_2}{u_f - \hat{u}_i} + c_3 \hat{u}_i} \qquad \forall i$$

$$\hat{q}_i = \hat{k}_i \times \hat{u}_i \qquad \forall i$$

$$\hat{q}_i, \hat{k}_i, \hat{u}_i \geq 0 \qquad \forall i$$

$$\hat{u}_i < u_f \qquad \forall i$$

$$0.5u_f \leq u_c \leq u_f; \quad q_c \leq \frac{k_j u_f u_c}{2u_f - u_c} \tag{42}$$

$$c_1 = \frac{u_f}{k_j u_c^2}\left(2u_c - u_f\right); \quad c_2 = \frac{u_f}{k_j u_c^2}\left(u_f - u_c\right)^2; \quad c_3 = \frac{1}{q_c} - \frac{u_f}{k_j u_c^2}$$

$$u_f^{\min} \leq u_f \leq u_f^{\max}; \; u_c^{\min} \leq u_c \leq u_c^{\max}; \; q_c^{\min} \leq q_c \leq q_c^{\max}; \; k_j^{\min} \leq k_j \leq k_j^{\max}$$

where u_i, k_i, and q_i are the field observed space-mean speed, density, and flow measurements, respectively. The speed, density, and flow variables with hats (^) are estimated speeds, densities, and flows while the tilde variables (~) are the maximum field observed speed, density, and flow measurements. All other variables are defined as was done earlier in describing the Van Aerde functional form.

The objective function ensures that the formulation minimizes the normalized orthogonal error between the three-dimensional field observations and the functional relationship, in this case the Van Aerde functional form. The three error terms are normalized in order to ensure that the objective function is not biased towards reducing the error in one of the three variables at the expense of the other two variables. This data normalization ensures that the parameters in each of the three axes range from 0.0 to 1.0 and thus a minimization of the orthogonal error provides a quality of fit that is equivalent across all three axes.

The initial set of constraints, which is nonlinear, ensures that the Van Aerde functional form is maintained, while the second set of constraints is added to constrain the third dimension, namely the flow rate. The third and fourth set of constraints guarantees that the results of the minimization formulation are feasible. The fifth set of constraints, ensures that the four parameters that are selected do not result in any inflection points in the speed–density relationship (i.e., it ensures that the density at any point is less than or equal to the jam density). A detailed derivation of the final constraint is provided elsewhere (*3*). The sixth set of equations provides estimates for the three model constants based on the roadway's mean free-flow speed (u_f), speed-at-capacity (u_c), capacity (q_c), and jam density (k_j). The final set of constraints provides a valid search window for the four traffic stream parameters that are being optimized (u_f, u_c, q_c, and k_j).

The total number of independent decision variables equals twofold the number of field observations plus the four traffic stream parameters u_f, u_c, q_c, and k_j. For example a problem with 100 observations results in a total of 204 independent decision variables (2 x 100 + 4). The heuristic approach that was developed earlier was applied to the data to estimate the four traffic stream parameters (*15*, *18*). Once the four traffic stream parameters are estimated the individual car-following models can be calibrated using the equations provided in Table 2.

CONCLUSIONS

The paper developed procedures for calibrating the steady-state component of various car-following models using macroscopic loop detector data. The paper then compared the various steady-state car-following formulations and demonstrated that the Gipps and Van Aerde steady-state car-following models provide the highest level of flexibility in capturing different driver and roadway characteristics. However, the Van Aerde model, unlike the Gipps model, is a single-regime model and thus is easier to calibrate given that it does not require the segmentation of data into two regimes. An analysis of existing software demonstrated that a number of car-following parameters are network- and not link-specific and thus do not offer model users with the flexibility of coding different roadway capacities for different facility types. In some software, however, arterial and freeway roadway car-following parameters can be coded separately, as in the case of CORSIM and VISSIM. However, major roadway capacity differences can be observed within the broad range of facility categories. For example, the saturation flow rate may vary from 1,300 to 2,000 veh/h on an arterial depending on the roadway and driver characteristics. Consequently, the paper recommends that modifications be made to the various software to allow more flexibility in setting link-specific car-following parameters.

ACKNOWLEDGMENT

The authors acknowledge the valuable input from Dr. William Perez of Cambridge Systematics, Inc. and Roemer Alfelor and David Yang of the Federal Highway Administration. Finally, the authors acknowledge the financial support provided by the FHWA and the Mid-Atlantic University Transportation Center in conducting this research effort.

REFERENCES

1. Rakha, H., P. Pasumarthy, and S. Adjerid. The INTEGRATION Framework for Modeling Longitudinal Vehicle Motion. In *TRANSTEC*, Athens, Greece, 2004.
2. Ozaki, H. Reaction and Anticipation in the Car-following Behavior. In *12th International Symposium on Transportation and Traffic Theory*, Elsevier, 1993, pp. 349–366.
3. Rakha, H. A. Validation of Van Aerde's Simplified Steady-State Car-Following And Traffic Stream Model. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006.
4. Gazis, D., R. Herman, and R. Rothery. Nonlinear Follow-the-Lead Models of Traffic Flow. In *Operations Research*, Vol. 9, No. 4, 1961, pp. 545–567.
5. Dowling, R., A. Skabardonis, J. Halkias, G. McHale, and G. Zammit. Guideline for Calibration of Microsimulation Models: Framework and Applications. In *Transportation Research Record: Journal*

of the Transportation Research Board, No. 1876, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 1–9.
6. Brackstone, M., and M. McDonald. Car-Following: A Historical Review. In *Transportation Research Part F,* Vol. 2, No. 4, 1999, pp. 181–196.
7. Farzaneh, M., and H. A. Rakha. Impact of Differences in Driver-Desired Speed on Steady-State Traffic Stream Behavior. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1965,* Transportation Research Board of the National Academies, Washington, D.C., 2006. pp. 142–151.
8. Halati, A., H. Lieu, and S. Walker. CORSIM: Microscopic Traffic Simulation Model for Integrated Networks. Presented at 76th Annual Meeting of the Transportation Research Board, Washington, D.C., 1997.
9. Rakha, H., and B. Crowther. Comparison and Calibration of FRESIM and INTEGRATION Steady-State Car-Following Behavior. In *Transportation Research Part A*, Vol. 37, 2003, pp. 1–27.
10. Gipps, P. G. A Behavioral Car-Following Model for Computer Simulation. In *Transportation Research Part B*, Vol. 15, 1981, pp. 105–111.
11. Rakha, H., C. C. Pecker, and H. B. B. Cybis. Calibration Procedure for the Gipps' Car-Following Model. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1999,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 115–127.
12. Wilson, R. E. An Analysis of Gipps' Car-Following Model of Highway Traffic. In *IMA Journal of Applied Mathematics*, Vol. 66, 2001, pp. 509–537.
13. Fritzsche, H. T. A Model for Traffic Simulation. In *Traffic Engineering and Control*, Vol. 5, 1994, pp. 317–321.
14. Van Aerde, M. Single Regime Speed-Flow-Density Relationship for Congested and Uncongested Highways. Presented at 74th Annual Meeting of the Transportation Research Board, Washington D.C., 1995.
15. Van Aerde, M. and H. Rakha. Multivariate Calibration of Single Regime Speed-Flow-Density Relationships. In *Proc., 6th 1995 Vehicle Navigation and Information Systems Conference*, Seattle, Wash., 1995.
16. Demarchi, S. H. A New Formulation ror Van Aerde's Speed-Flow-Density Relationship (in Portuguese). In *XVI Congresso De Pequisa e Ensino em Transportes,* Natal, RN, Brazil, 2002.
17. Rakha, H., and B. Crowther. Comparison of Greenshields, Pipes, and Van Aerde Car-Following and Traffic Stream Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1802*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 248–262.
18. Rakha, H. and M. Arafeh. Tool for Calibrating Steady-State Traffic Stream and Car-Following Models. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.

Calibrating Speed–Density Functions for Mesoscopic Traffic Simulation

RAMACHANDRAN BALAKRISHNA
CONSTANTINOS ANTONIOU
HARIS N. KOUTSOPOULOS
YANG WEN
MOSHE BEN-AKIVA

Mesoscopic traffic simulation models combine macroscopic supply (e.g., link performance functions and capacities) with microscopic demand (e.g., individual drivers and disaggregate behavior models) to capture the time-varying evolution of congestion patterns, queues, and spillbacks on traffic networks. Such systems are being applied to solve a variety of off-line and online traffic management problems. Often, the link performance function takes the form of a speed–density relationship for each link or segment in the study network. These functions are based on the fundamental diagram and model the variation of average vehicle speed with traffic density. Since each network link or segment can have a separate speed–density function (each with a few parameters), the total number of parameters to be calibrated is potentially very large. This paper presents some experiences with calibrating speed–density functions on a variety of real networks. Flexible off-line and on-line calibration frameworks are discussed. The calibrated functions are embedded within the DynaMIT dynamic traffic assignment model, whose accuracy is evaluated against real data. The expected benefits from recalibrating speed–density functions during on-line operations using real-time sensor data are also illustrated.

Mesoscopic simulation tools such as DynaMIT (*1*) and DYNASMART (*2*) have been developed to model large-scale traffic networks. Such dynamic traffic assignment (DTA) tools combine detailed demand and supply model components together with their complex interactions to provide accurate depictions of queues, spillbacks and congestion evolution. Macroscopic simulation models also are used for modeling freeway corridors (*3*, *4*). The demand components of these models typically include time-varying origin–destination (O-D) matrices and route choice models, while the supply side captures link capacities, queuing models and macroscopic traffic dynamics relationships. Critical for modeling traffic dynamics are speed–density functions that play an important role in both mesoscopic and macroscopic traffic simulation models.

Numerous studies have focused on estimating various demand-side inputs and parameters. For example, many papers focus on the estimation of dynamic O-D flows from traffic counts (*5–10*). Route choice models have been estimated from surveys (*11*) and refined with aggregate data (*12*, *13*). The supply side seems to have received relatively less attention, though it plays a critical role in determining network performance. Speed–density functions are particularly challenging, as they must encapsulate a variety of effects including traffic dynamics, lane speed distributions, vehicle and driver mix, and weather conditions. Many of these aspects vary with location within the network, and require careful calibration against real-world sensor data.

This paper reviews methods for calibration of speed–density relationships and presents some experiences with calibrating speed–density functions in the context of dynamic traffic assignment (DTA) models, on a variety of real networks including Irvine and Los Angeles (California), Lower Westchester County (New York) and Southampton (United Kingdom). Both off-line and on-line calibration frameworks are discussed, that allow the incorporation of any available source of data, including data from automated vehicle identification (AVI) systems and probe vehicles. The calibrated functions are embedded within the DynaMIT dynamic traffic assignment model, whose accuracy is evaluated against real data. The expected benefits from re-calibrating speed–density functions during on-line operations using real-time sensor data are illustrated.

APPROACHES FOR CALIBRATION OF SPEED–DENSITY RELATIONSHIPS

The calibration of a DTA's speed–density functions involves a potentially large set of parameters. Recent studies have employed systematic algorithms for the calibration of DTA supply models, in particular speed–density relationships, with varying degrees of success. The typical data used for the calibration of these parameters are sensor records of at least two of the three primary traffic descriptors: speeds, flows (or counts) and densities (or detector occupancies). In this section we review the experience with optimization algorithms applied to this problem. We classify the applications as off-line (archived sensor data) and on-line (real-time sensor data and calibration).

Off-Line Calibration Approaches

Link performance functions in much of the literature are calibrated by fitting a curve to the observed traffic data. For example, Leclercq (*14*) estimates four parameters of a two-regime flow-density function with data from arterial segments in Toulouse, France. The function is comprised of a parabolic free-flow part and a linear congested regime. An interior point, conjugate gradient method is employed to optimize the fit to observed sensor flows, with the fitted flows obtained from the assumed flow-density function. Van Aerde and Rakha (*15*) describe the calibration of speed-flow profiles by fitting data from loop detectors on I-4 near Orlando, Florida. Links without sensors are allotted a speed-flow profile from a physically similar link that is instrumented (and for which a profile was fitted).

A major drawback of the above approach is one of localized fit. The estimated link performance functions reflect spot measurements at discrete sensor stations, and do not necessarily correspond to overall link dynamics (especially in the presence of congestion). The estimation procedure also does not enforce consistency across contiguous links or segments, stressing the need for an expanded approach that considers larger sections of the network. For example, it may be beneficial to relax the fit at a lightly traveled link so as to improve the fit at several links further downstream.

In the context of traffic simulation and DTA models, most calibration approaches focus on the independent estimation of subsets of supply parameters. Munoz et al. (*16*) describe a calibration methodology for a modified cell transmission model (MCTM), applied to a 14-mile westbound stretch of the I-210 freeway in Pasadena, Calif. Free-flow speeds are obtained through least squares, by fitting a speed-flow plot through each detector's data. Free-flow speeds

for cells without detectors are computed by interpolating between the available speed estimates. In the case of bad or missing sensor data, a default of 60 mph was assumed. Speed-flow functions are obtained through constrained least squares on sensor data from congested cells.

Many applications of macroscopic traffic models focus on freeway corridors or sections. Ngoduy and Hoogendoorn (*4*) calibrate 10 METANET parameters for a section of the A1 freeway in The Netherlands using the Nelder-Mead method (a gradient-free algorithm working directly with objective function evaluations). The calibrated terms include fundamental diagram parameters such as free-flow speed, minimum speeds and maximum density, and other coefficients that capture the effects of merging, weaving and lane drops. Ngoduy et al. (*17*) calibrate six of these METANET parameters for a freeway section with no ramps. An objective function measuring the fit to count and speed data is minimized.

Park et al. (*18*) apply DynaMIT to a network in Hampton Roads, Va. and estimate speed–density functions for segments. They adopt the procedure in Van Aerde and Rakha (*15*) and conclude that the initial calibration results need adjustments to improve DynaMIT's overall ability to estimate and predict traffic conditions.

Kunde (*19*) describes a three-stage approach to speed–density calibration. At the disaggregate level, segment speed–density relationships are estimated similarly to Van Aerde and Rakha (*14*). In the second stage, a suitable subnetwork is chosen, and the estimates from the previous stage are refined by accounting for interactions between the segments. The choice of a subnetwork depends on the structure of the network and the location of sensors. An ideal subnetwork would allow one to deduce the true O-D flows for the subnetwork from the available sensor count information, so that the supply parameters may be inferred under known demand conditions. The final stage utilizes the entire network to incorporate demand–supply interactions into the calibration process. The approach was demonstrated with the DynaMIT DTA model, using data from Irvine, Calif. A total of 1,373 segments were divided into 11 groups and a speed–density function fitted for each group.

A least squares objective function measuring the fit to count data was then minimized on a sub-network with known demands inferred from sensor counts. Simultaneous perturbation stochastic approximation (SPSA) was used to fine tune the speed–density function parameters. This algorithm approximates the components of the gradient vector from just two objective function evaluations, after perturbing all components of the parameter vector simultaneously (*20*). The Box-Complex algorithm (*21*) was applied with better success, though the numerical example was too small to draw general conclusions.

In the previous approaches, the calibration variables were limited to the speed–density function parameters. However, the optimization depends on several other DTA inputs such as O-D flows and route choice model parameters. The values selected for these other inputs and parameters thus impact the outcome of the supply calibration. One may thus iterate between demand and supply calibration steps until convergence (as defined by the modeler) is reached. This iterative procedure can be time consuming and inefficient, as only a subset of the available data is used in either calibration.

Balakrishna (*22*) and Balakrishna et al. (*23*) present a calibration framework that allows the simultaneous calibration of all supply and demand parameters and unknown inputs typical to DTA models (e.g., O-D flows, route choice parameters, capacities, and speed–density parameters) using any available data (e.g., counts, speeds, densities, and queue lengths). Thus all significant DTA inputs and parameters may be estimated simultaneously, providing the most efficient result. The problem is solved with the SPSA algorithm.

This calibration approach provides a unique advantage. Since the parameters of the speed–density functions for all segments are estimated simultaneously, the function parameters for each segment can be calibrated to better fit the traffic data at the network level.

On-Line Calibration Approaches

In the current DTA framework, only the O-D f ows are calibrated on-line. In most cases, the approach to the problem of calibration of the other parameters has been to calibrate the simulation models off-line using a database of historic information. The calibrated parameter values are then used in the on-line simulations. The calibrated model parameters, therefore, represent average conditions over the period represented in the data. Models that were calibrated this way may produce satisfactory results in off-line evaluation studies, which are concerned with the expected performance of various traffic management strategies. However, this may not be the case in real-time applications, which are concerned with the system performance on the given day. If the model that was calibrated off-line is used without adjustment, the system is not sensitive to the variability of the traffic conditions between days, which are the result of variations in the parameters of the system, such as weather and surface conditions. Such variations may cause traffic conditions to differ signif cantly from the average values. Thus, the predictive power of the simulation model may be reduced. To overcome this problem, real-time data can be used to recalibrate and adjust the model parameters on-line so that prevailing traffic conditions can be captured more accurately. The wealth of information included in the off-line values can be incorporated into this process by using them as a priori estimates.

Doan et al. (*24*) outline a framework for periodic adjustments to a traffic management simulation model to maintain an internal representation of the traffic network that is consistent with that of the actual network. A similar approach is proposed in Peeta and Bulusu (*25*), where consistency is sought in terms of minimizing the deviations of the predicted time-dependent path flows from the corresponding actual flows. He et al. (*26*) develop an on-line calibration process that complements their off-line approach and adjusts an analytical dynamic traffic model's output to be consistent with real-world traffic conditions.

Tavana and Mahmassani (*27*) use transfer function methods (bivariate time-series models) to estimate dynamic speed–density relations from typical detector data. Huynh et al. (*28*) extend the work of Tavana and Mahmassani (*27*) by incorporating the transfer function model into a simulation-based DTA framework. Qin and Mahmassani (*29*) evaluate the same model with actual sensor data from several links of the Irvine, Calif. network.

Wang and Papageorgiou (*30*) present a general approach to the real-time estimation of the complete traffic state in freeway stretches. They use a stochastic macroscopic traffic flow model and formulate it as a state-space model, which they solve using an extended Kalman filter (EKF). The formulation allows dynamic tracking of time-varying model parameters by including them as state variables to be estimated. A random walk is used as the transition equations for the model parameters. Wang et al. (*31*) present an extended application of this approach.

Antoniou et al. (*32*) formulate the problem of on-line calibration of a DTA model as a nonlinear state-space model that allows for the simultaneous calibration of all model parameters and inputs. The methodology is generic and flexible and does not make any assumptions on the underlying model structure, the parameters to be calibrated or the type of available measurements. Because of its nonlinear nature, the resulting model cannot be solved by the Kalman f lter, and therefore, nonlinear extensions are considered: the EKF; the limiting EKF

(LimEKF); and the unscented Kalman f lter. The solution algorithms are applied to the on-line calibration of the state-of-the-art DynaMIT DTA model, and their use is demonstrated in a freeway network in Southampton, U.K. The LimEKF shows accuracy that is comparable to that of the best algorithm but with vastly superior computational performance.

CASE STUDIES AND RESULTS

We now present results related to the application of the state-of-the-art methods reported in Balakrishna et al. (*23*) and Antoniou et al. (*32*). These studies respectively focus on the off-line and on-line calibration of speed–density parameters used in the DTA model DynaMIT. The focus of the discussion is on the impact of data used in the calibration and the importance of calibration with respect to the ability of the model to replicate actual conditions. In all studies the quality of the calibration was ascertained using the normalized root mean square error (RMSN) statistic to evaluate the fit to both count and speed data:

$$RMSN = \frac{\sqrt{S\sum_{i=1}^{S}(y_i - \ddot{y}_i)^2}}{\sum_{i=1}^{S} y_i}$$

where y_i are observed measurements, $\ddot{y}_i$ are simulated values and S is the total number of measurements.

Off-Line Calibration Results

The above method has been applied to the Los Angeles network (Figure 1), an area of heavy traffic throughout the year owing to commuters and the regularity of sporting and convention special events. The network has 740 segments.

Table 1 summarizes the numerical results. $RMSN^c$ represents the fit to counts while $RMSN^s$ is the fit to speeds. RMSN was used to evaluate the quality of the results. It is seen that the estimator S(c), denoting supply calibration using count data, results in a significant improvement in replicating the counts and traffic dynamics (speeds) in the area. Further, the increased accuracy is reflected on both freeway and arterial links. In the base case (Ref), the speed–density parameters were fitted at individual sensor locations and attributed to all segments in the respective groups.

Vaze (*33*) presented a case study, using the framework suggested by Balakrishna et al. (*23*), in which demand and supply parameters were simultaneously calibrated using not only loop detectors data but also AVI data. In the network of Lower Westchester County, N.Y. (shown in Figure 2), sensors deployed for an electronic toll collection system at various locations detect vehicles equipped with transponders and report, among other information, point-to-point travel time for specific locations. Using synthetic data for this setup, Vaze calibrated a total of more than 6,400 parameters (3,856 demand parameters, i.e., 482 O-D pairs for each of the eight intervals of simulation, and 2,624 supply parameters, including 10 groups of parameters for speed–density relationship). SPSA was used as the solution algorithm.

Vaze confirmed that simultaneous demand–supply calibration was found to be superior compared to the demand-only calibration and it increased the calibration accuracy substantially. He also found that the use of AVI data had improved the calibration accuracy, both in terms of sensor count error as well as travel time (Table 2).

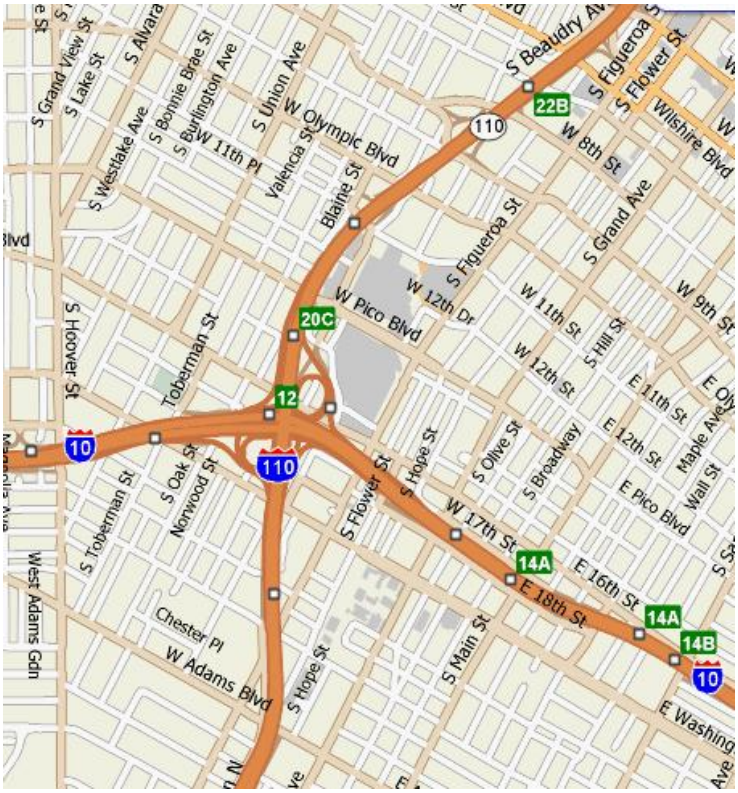


FIGURE 1 Los Angeles network (SOURCE: www.mapquest.com).

TABLE 1 RMSN Statistics for Los Angeles Study

Estimator	Fit to Counts (RMSN^c)		Fit to Speeds (RMSN^s)	
	Freeway	Arterial	Freeway	Arterial
Ref	0.218	0.239	0.181	0.203
S(c)	0.149	0.178	0.119	0.131

On-Line Calibration Results

Antoniou et al. (*32*) present an application of their on-line DTA calibration methodology using data from a freeway network in Southampton, U.K. (Figure 3). The study demonstrates the performance gains that can be obtained through the dynamic, simultaneous calibration of the speed–density relationships and other supply-side parameters.

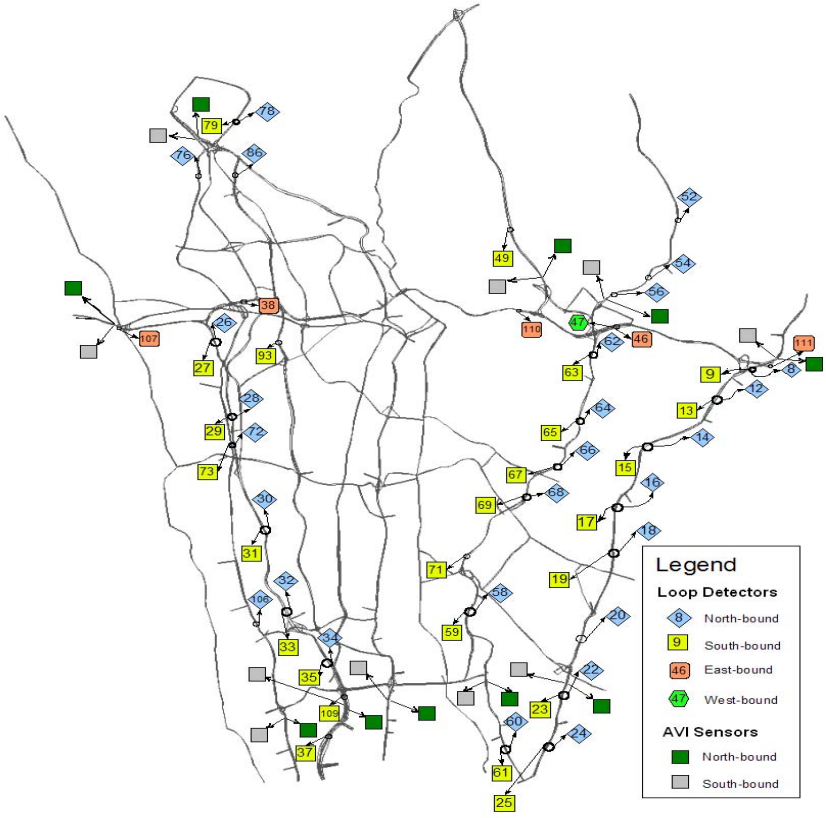


FIGURE 2 Lower Westchester County network.

TABLE 2 RMSN Statistics for Lower Westchester County Study

	Fit to Counts ($RMSN^c$)	Fit to Speed ($RMSN^s$)
A priori	0.253 0.	291
Count data	0.200 0.	222
Count+AVI data	0.182 0.	212

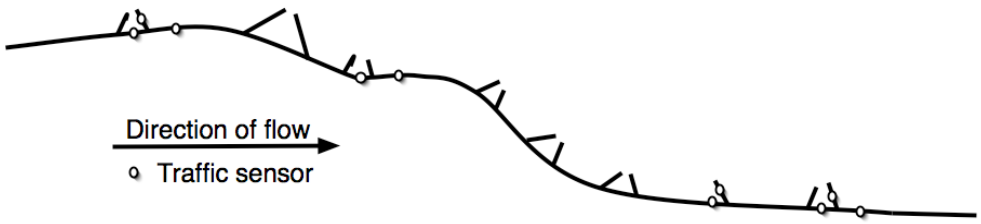


FIGURE 3 Lower Westchester County network.

Table 3 summarizes the DTA model estimation accuracy improvements due to the impact of the on-line calibration of the supply parameters. The base case in this table is the situation where only the demand parameters (O-D flows) are calibrated on-line. The initial off-line calibration of the DTA model has been based on data from several days with normal (dry) weather conditions. When the on-line calibration is performed for a day with similar environmental conditions, the simultaneous on-line calibration of both the demand and supply parameters results in an improvement of the model estimation accuracy of more than 10% both in terms of fit to counts and fit to speed (over the base case in which only the demand parameters are calibrated on-line).

The power of the on-line calibration procedure is further demonstrated by applying the same model (initially calibrated using data from days with dry weather) to a rainy day. Again, the on-line calibration allows the model to adapt and capture the prevailing conditions satisfactorily.

Figure 4 presents examples of the estimated speed–density relationships. As expected, the average curve falls between the relationships obtained for intervals with dry and wet weather conditions respectively. For the same density, lower speeds are experienced under wet weather conditions, while in general dry weather conditions allow for higher speeds for the same density values.

TABLE 3 RMSN Statistics for Southampton Study

	Demand-Only		Demand and Supply	
	Fit to Counts ($RMSN^c$)	**Fit to Speeds ($RMSN^s$)**	**Fit to Counts ($RMSN^c$)**	**Fit to Speeds ($RMSN^s$)**
Dry weather	0.128 0.	126 0.	109	0.112
Wet weather	0.115 0.	131 0.	102	0.117

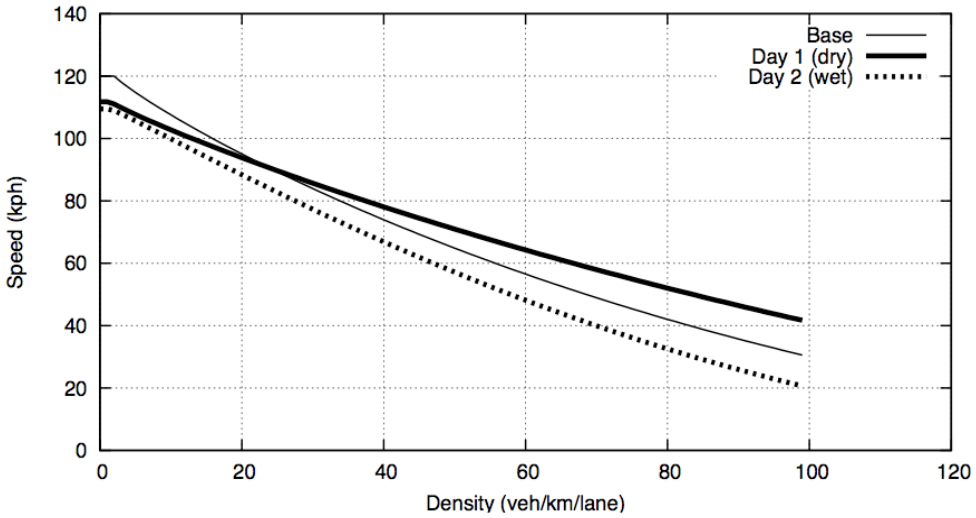


FIGURE 4 Speed–density relationships for average, dry and wet weather conditions.

CONCLUSION

Speed–density functions are important supply-side inputs to many DTA models, and help capture traffic dynamics. Owing to the wide variation in the factors that affect traffic dynamics (such as merging and weaving behavior, vehicle and driver mix, weather conditions, etc), these functions must be calibrated with traffic data. This paper reviews some applications of the DynaMIT DTA model in a variety of locations, focusing on the calibration of its segment-specific speed–density functions.

The local calibration of speed–density functions is generally straightforward, but has some potential drawbacks. The functions can be over-fit in each specific location resulting in a suboptimal calibration at the network level. Grouping of segments due to sparse sensor coverage can also result in the discarding of spatial differences. Further, the availability of modern data such as point-to-point travel times can help to better explain traffic dynamics. A flexible calibration approach that addresses these aspects is reviewed, and its benefits outlined. Case studies drawn from real-world applications of DynaMIT, a mesoscopic DTA model, are presented to illustrate that speed–density functions can be accurately and reliably calibrated with a wide range of data. The validity of these calibrations is confirmed through off-line and on-line tests of state estimation and prediction performance.

REFERENCES

1. Ben-Akiva, M., M. Bierlaire, H. N. Koutsopoulos, and R. Mishalani. Transportation and Network Analysis: Current Trends. In *Real-Rime Simulation of Traffic Demand-Supply Interactions Within DynaMIT* (M. Gendreau and P. Marcotte, eds.), Kluwer Academic Publishers, 2002, pp. 19–36. Miscellenea in honor of Michael Florian.
2. Mahmassani, H. S. Dynamic Network Traff c Assignment and Simulation Methodology for Advanced System Management Applications. In *Networks and Spacial Economics*, Vol. 1, No. 3, 2001, pp. 267–292.

3. Messmer, A., and M. Papageorgiou. Freeway Network Simulation and Dynamic Traffic Assignment with METANET Tools. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1776*, TRB, National Research Council, Washington, D.C., 2001, pp. 178–188.
4. Ngoduy, D., and S. P. Hoogendoorn. An Automated Calibration Procedure for Macroscopic Traffic Flow Models. In *Proc., 10th IFAC CTS*, Tokyo, 2003.
5. Okutani, I. The Kalman Filtering Approach in Some Transportation and Traffic Problems. In *Transportation and Traffic Theory* (N. H. Gartner and N. H. M. Wilson, eds.), Elsevier, New York, 1987, pp. 397–416.
6. Cascetta, E., Inaudi, D., and Marquis, G. (1993). Dynamic Estimators of Origin–Destination Matrices Using Traff c Counts. In *Transportation Science*, Vol. 27, No. 4, 1993, pp. 363–373.
7. Ashok, K. and M. Ben-Akiva. Dynamic O-D Matrix Estimation and Prediction for Real-Time Traff c Management Systems. *In Transportation and Traffic Theory* (C. Daganzo, ed.), Elsevier Science Publishing, 1993, pp. 465–484.
8. Ashok, K. and M. Ben-Akiva. Alternative Approaches for Real-Time Estimation and Prediction of Time-Dependent Origin–Destination Flows. In *Transportation Science*, Vol. 34, No. 1, 2000, pp. 21–36.
9. Ashok, K. and M. Ben-Akiva. Estimation and Prediction of Time-Dependent Origin–Destination Flows with a Stochastic Mapping to Path Flows and Link Flows. In *Transportation Science*, Vol. 36, No. 2, 2002, pp. 184–198.
10. Balakrishna, R. and H. N. Koutsopoulos. Incorporating Within-Day Transitions in the Simultaneous Off-Line Estimation of Dynamic Origin–Destination Flows Without Assignment Matrices. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2085*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp, 31–38.
11. Ramming, M. S. Network Knowledge and Route Choice. PhD thesis. Massachusetts Institute of Technology, 2001.
12. Balakrishna, R., Koutsopoulos, H. N., and Ben-Akiva, M. Calibration and Validation of Dynamic Traffic Assignment Systems. In *Proc., 16th International Symposium on Transportation and Traffic Theory: Transportation and Traffic Theory: Flow, Dynamics and Human Interaction* (H. S. Mahmassani, ed), Elsevier, 2005, pp. 407–426.
13. Tsavachidis, M. Aggregate Analysis of Driver Response to Collective Route Guidance and Implications for System Control. In *Proc., 10th International Conference on Road Transport Information and Control*, Commonwealth Institute, London, U.K, Institution of Electrical Engineers, U.K., 2000.
14. Leclercq, L. Calibration of Flow-Density Relationships in Urban Streets. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1934*, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 226–234.
15. Van Aerde, M., and H. Rakha. Travtek Evaluation Modeling Study. Technical report. Federal Highway Administration, U.S. DOT, 1995.
16. Munoz, L., X. Sun, D. Sun, G. Gomes, and R. Horowitz. Methodological Calibration of the Cell Transmission Model. In *Proceedings of the Annual Control Conference*, Boston, Mass., 2004, pp. 798–803.
17. Ngoduy, D., S. P. Hoogendoorn, and H. J. Van Zuylen. Comparison of Numerical Schemes for Macroscopic Traffic Flow Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1876*, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 52–61.
18. Park, B., D. M. Pampati, and R. Balakrishna. Architecture for On-Line Deployment of DynaMIT in Hampton Roads, VA. In *Proc., 9th International Conference on Applications of Advanced Technology in Transportation (AATT)*, Chicago, Ill., August 2006, pp. 605–610.
19. Kunde, K. Calibration of Mesoscopic Traffic Simulation Models for Dynamic Traffic Assignment. MS thesis. Massachusetts Institute of Technology, 2002.

20. Spall, J. C. Developments in Stochastic Optimization Algorithms with Gradient Approximations Based on Function Measurements. In *Proc., 1994 Winter Simulation Conference* (J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila, eds.), 2004.
21. Box, M. J. A New Method of Constrained Optimization and a Comparison with Other Methods. In *Computer Journal*, Vol. 8, No. 1, 1965, pp. 42–52.
22. Balakrishna, R. Off-Line Calibration of Dynamic Traffic Assignment Models. PhD thesis. Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 2006.
23. Balakrishna, R., M. Ben-Akiva and H. N. Koutsopoulos. Off-Line Calibration of Dynamic Traffic Assignment: Simultaneous Demand-Supply Estimation. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2003*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 50–58.
24. Doan, D. L., A. Ziliaskopoulos, and H. Mahmassani. On-Line Monitoring System for Real-Time Traffic Management Applications. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1678*, TRB, National Research Council, Washington, D.C., 1999, pp. 142–149.
25. Peeta, S., and S. Bulusu. Generalized Singular Value Decomposition Approach For Consistent On-Line Dynamic Traffic Assignment. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1667*, TRB, National Research Council, Washington, D.C., 1999, pp. 77–87.
26. He, R., S. Miaou, B. Ran, and C. Lan. Developing an On-Line Calibration Process for an Analytical Dynamic Traffic Assignment Model. *Proc., 78th Annual Meeting of the Transportation Research Board*, TRB, National Research Council, Washington, D.C., 1999.
27. Tavana, H., and H. Mahmassani. Estimation and Application of Dynamic Speed–density Relations by Using Transfer Function Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1710*, TRB, National Research Council, Washington, D.C., 2000, pp. 47–57.
28. Huynh, N., H. Mahmassani, and H. Tavana. Adaptive Speed Estimation Using Transfer Function Models for Real-Time Dynamic Traffic Assignment Operation. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1783*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 55–65.
29. Qin, X., and H. Mahmassani. Adaptive Calibration of Dynamic Speed–density Relations for Online Network Traffic Estimation and Prediction Applications. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1876,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 82–89.
30. Wang, Y., and M. Papageorgiou. Real-Time Freeway Traffic State Estimation Based on Extended Kalman Filter: A General Approach. In *Transportation Research Part B*, Vol. 39, No. 2, 2005, pp. 141–167.
31. Wang, Y., M. Papageorgiou, and A. Messmer. Real-Time Freeway Traffic State Estimation Based on Extended Kalman Filter: A Case Study. In *Transportation Science*, Vol. 41, No. 2, 2007, pp. 167–181.
32. Antoniou, C., M. Ben-Akiva, and H. N. Koutsopoulos. Nonlinear Kalman Filtering Algorithms for On-Line Calibration of Dynamic Traffic Assignment Models. In *IEEE Transactions on Intelligent Transportation Systems*, Vol. 8, No. 4, 2007, pp. 661–670.
33. Vaze, V. Calibration of Dynamic Traffic Assignment Models with Point-to-Point Traffic Surveillance. MS thesis. Massachusetts Institute of Technology, 2007.

RESOURCE

May, A. *Traffic Flow Fundamentals*. Prentice Hall, 1990.

Lane-Change Maneuver Detection with Differential Global Positioning System Data from Probe Vehicle

YIGUANG XUAN
BENJAMIN COIFMAN

The impact of lane-change maneuvers is fundamental to microscopic traffic flow theory. Due to the difficulty of tracking many vehicles over time and space, most of the published research in this area seeks to find lane-change maneuvers visually from wayside cameras. This paper presents a different approach, finding the lane-change maneuvers of a probe vehicle itself using differential global positioning system data. The authors first use multiple probe vehicle trajectories through a study corridor to establish a reference trajectory from the median of all trajectories, and this reference trajectory will be used to define the position of the current lane. This approach eliminates the need for high-resolution maps accurate enough to capture the exact position of the individual lanes. The lane change maneuver detection is then divided into two parts, controlling for the impacts of mandatory lane change maneuvers (MLC) and then for discretionary lane change maneuvers (DLC). MLC are detected by comparing the difference between the mean and median of lateral distance of all trajectories relative to a reference trajectory. After distinguishing all the MLC, the DLC are found by setting lateral thresholds around the reference trajectory, i.e., when a given trajectory leaves this virtual lane. In the process the authors control for the impacts of GPS errors, such as multipath, arising from obstructions. DLC are then found by comparing the out-of-threshold-line time and length to a threshold acquired empirically from data.

Effect of Lane-Change Maneuvers on a Simplified Car-Following Theory

CHAO WANG
BENJAMIN COIFMAN

This paper investigates the linearity of empirically observed spacing-speed relation for various drivers in the context of car-following theory and how lane change maneuvers perturb the relation. It is shown that the impacts of lane-change maneuvers are not balanced; the response time to an exiting vehicle is much longer than the response time to an entering vehicle. This accommodation imbalance will propagate upstream and as discussed, it appears to be a source of speed and flow fluctuations within a queue. The present work is motivated by Newell's simplified car-following theory, namely that during congested periods the trajectory of a given vehicle is essentially identical to the preceding vehicle's trajectory except for a translation in space and time. One of the basic assumptions in Newell's presentation is that spacing and speed are linearly related. While other researchers have found macroscopic evidence supporting Newell's theory, they have also found that it fails in the presence of frequent lane-change maneuvers. This paper takes a microscopic approach, employing vehicle trajectory data. This paper provides support for Newell's assumed linear relation between spacing and speed over a large range of speeds when vehicles are not impacted by lane-change maneuvers. It also offers a possible explanation for the degraded performance of Newell's theory in the presence of heavy lane-change maneuvers. Although the focus is on Newell's simplified car-following theory, the empirical results of this study have similar implications for many other car-following theories as well.

Strategies to Improve Dissipation into Destination Networks Using Macroscopic Network Flow Models

VINAYAK V. DIXIT
ESSAM A. RADWAN
Department of Civil & Environmental Engineering
University of Central Florida

Backups originating from destinations have been observed during evacuation. These backups usually occur due to congestion at the destination network, which results in spillbacks onto the evacuation routes. These spillbacks result in queuing and delays that hamper evacuation operations. This paper presents theoretical proofs for the fundamental flow–speed–concentration relationship and the speed–accumulation relationship (Greenshields, Greenberg, and bell-shaped model) at a network level. These relationships and the relationships between inflow-accumulation and outflow-accumulation at a network level are studied using microscopic simulation. A strategy is developed (called the network breathing strategy) to improve dissipation of vehicles into the destination network using these relationships between network level variables. A comparison of the network breathing strategy to a do-nothing strategy in a simulation network showed a statistically significant increase in the number of vehicles dissipated into the network. This indicates that the application of such strategies on the destination networks would help improve evacuation operations by clearing evacuation routes and reducing queuing.

Recent natural disasters like hurricanes Katrina and Rita in 2005 have highlighted the importance of efficient transportation strategies to ensure smooth and effective evacuation of people out of harm's way. Strategies such as improving capacity by starting contraflow operations and demand staging have been widely studied and implemented as effective plans to reduce evacuation time.

During the South East U.S. Regional Transportation Analysis Meeting in 2000, it was observed that "more than half the evacuees felt like it took them more than 5 hours longer to reach their destination than they thought that it would" (*1*). This is mainly due to the limited capacity of the exit ramps as well as congestion caused due to the large number of vehicles in the destination network. Evacuation routes usually terminate at large cities, and road networks in these cities are not designed to handle the large number of vehicles entering them during evacuation. This results in congestion which in turn leads to backups that can extend for miles on the evacuation route. Though such phenomena have been observed repeatedly, limited literature on evacuation seems to have addressed this issue of network congestion at the termination node. Most simulation studies tend to assume ideal destinations, where vehicles leave the system as soon as they reach the destination irrespective of the number of vehicles already present in the destination road network. This is a myopic perspective of analyzing evacuation routes. Therefore it is important to understand network-level properties of traffic variables.

Initial attempts to understand relationships between network-level variables consisted of Zahavi's (*2*, *3*) work on the α-relationship between network-level parameters of traffic intensity (*I*, the distance traveled per unit area), road density (*R*, length of road per unit area) and the

weighted space mean speed (v). Using data from England and the United States he arrived at the relationship in Equation 1.

$$I = \alpha R / v \tag{1}$$

Buckley and Wardrop (*4*) later showed that α was strongly correlated to the space mean speed. In a later field study, Ardekani (*5*) proved that the α parameter had a positive correlation to network concentration. This made the α parameter model highly inaccurate. Chapter 6 of the *Traffic Flow Theory* monograph revised in 1997 (*6*) contains a comprehensive review of these macroscopic flow models.

In order to characterize flow of vehicles in urban network Prirgogine and Herman (7) proposed the two-fluid theory. The two-fluid model assumes that vehicular traffic in an urban network can be differentiated as stopped vehicles and running vehicles. These models were constructed between the average travel time per mile (T) versus the average running time per mile (T_r) using regression (Equation 2). The parameters (k, T_m) involved in this two-fluid model were indicative of the quality of service of the networks.

$$T_r = T_m^{\frac{1}{k+1}} T^{\frac{k}{k+1}} \tag{2}$$

Mahmassani et al. (*8*) and Williams et al. (*9*) during their study of two fluid models using computer simulation showed that relationships between the three fundamental traffic variables speed-flow-concentration (Equation 3) at a network level were similar to those on individual road facilities.

$$Q = KV \tag{3}$$

In a later simulation study Mahmassani et al. (*8*) found that both the linear V-K model proposed by Greenshields, and the nonlinear bell-shaped function proposed by Drake et al. (*11*) were able to describe the relationship between V and K fairly well. In their paper they also studied the effect of length and width of links as well as various traffic controls (perfectly coordinated, isolated and simultaneous signal operation) on the speed-concentration relationships and the flow-concentration relationships.

Even though these studies showed interesting results, due to the very few (six) data points used for the analysis, the conclusions in the paper are prone to major skepticism. Also, each simulation run was done for constant concentration conditions, in which constant concentration was maintained by allowing vehicles to circulate in the network. Such concentration conditions generally do not prevail in real urban networks, where vehicles enter and leave, and concentrations in the network vary more dynamically.

Mahmassani et al. (*12*) conducted microscopic simulation experiments on larger urban networks than the ones studied in Mahmassani et al. (*8*, *10*) and Williams et al. (*9*, *13*). The experiments concluded that the relationship between speed and concentration remained significantly identical for various network sizes. This indicated that these relationships between various network-level variables were independent of network size and consistent. During their analysis they observed that the average network speed at a given concentration was lower when

the intersections were operated as an unsignalized (stop-sign control) as compared to signalized intersections.

Ardekani (*5*) studied the two -fluid characterizations urban road networks and proved the validity of these models on real urban road networks. Ardekani through field studies also concluded that the fundamental Equation 3 holds true.

Recently Daganzo (*14*), using average network flow and accumulation, suggested various recipes for improving city mobility through gridlock control. The paper proposed a relationship between the outflow (exit function (*G(n)*)) and the number of vehicles in the network. The paper derived a differential equation (Equation 4) describing the number of vehicles in the network, based on the inflow (*f(t)*) and outflow (*G(n)*).

$$\frac{dn}{dt} = f(t) - G(n(t)), \quad \text{for} \quad t \geq 0 \tag{4}$$

These relationships are used in the paper to determine an optimal control strategy (A–B strategy) to control inflows so as to maximize outflow. One of the practical drawbacks of this approach is that due to the stochastic nature of traffic flow there are periods where inflow is greater than the outflow, leading to eventual jam conditions, hence in the strategy proposed for efficient operations real-time monitoring and control of the network is required. In this paper the proposed strategy overcomes this drawback.

Geroliminis and Daganzo (*15*) as a continuation of Daganzo's (*14*) theoretical work conducted simulation experiments with the San Francisco network. They showed a linear dependence between the travel production in the network and the outflow from the network, and an inverted U-shaped relationship between the travel production and accumulation. In addition the paper also describes the behavior of inflow with respect to accumulation. They showed that inflow remained constant up to a certain degree of accumulation and then started decreasing. The paper also proposed control strategies based on real-time observation of accumulation and were tested using simulation.

The network breathing strategy proposed in this paper is a cyclic process of allowing vehicles to enter the network, and allowing the network to become congested, followed by closure of their entry into the network, in which the congestion in the network is allowed to dissipate. After this, entrance into the network is allowed again. This process will be referred to as "network breathing." The period during which vehicles are allowed to enter is referred to as "network inhalation" and when vehicles are held from entering the network is referred to as "network relaxation." The advantage of such an approach is that the times for network inhalation and network relaxation can be predetermined depending on the network properties and would not need real-time feedback.

This paper presents the derivation for the fundamental flow-speed-concentration relationship, Greenshields' model, Greenberg's model, and the bell-shaped function proposed by Drake, Schofer, and May (*11*) at a network level. It also compares the performance of each of these models in explaining network level speed–accumulation relationships. It defines the various variables that are used in the rest of the paper and presents the derivation for relationships between various network-level variables. It also outlines the methodology for the strategy and discusses the results of applying the developed strategies to a simulated network. It concludes by discussing the advantages of these strategies and ideas for expansion of this work for further research.

DEFINITIONS

This section summarizes all the relevant variables that are used throughout this paper.

Q_i Flow on link *I*;
Q_{Out} Total outflow from network;
Q_{in} Total inflow into the network;
K_i Concentration on link *i*;
K_j Network level jam concentration;
v_j Speed of the j[th] vehicle on the network;
V_i Average speed on link *i*;
V_f Inverse of the average minimum time taken to travel a mile in the network, at free flow conditions;
n_i Number of vehicles on link *i*;
n Total number of vehicles on the network (sum of all n_i) (accumulation);
n_c Number of vehicles in the network, when the outflow from the network is the maximum;
n_p Maximum number of vehicles the network can accommodate at a given time;
l_i Lane-mile of link *i*;
l Total lane-mile in the network (sum of all l_i);
q_p The average flow in the network when number of vehicles in the network is n_p; and
x Number of vehicles in the network after network relaxation.

The three fundamental traffic variables speed, concentration and flow at network level are defined as average over all vehicles during an observation period ε. These were defined by Ardekani (*5*) and Mahmassani et al. (*7*, *9*). Average speed in a network is defined as the total number of vehicle miles traveled divided by the total number of vehicle hours in the network during an observation period ε. This is consistent with the generalized definitions defined by Edie (*15*). If there are n vehicles in the network and the velocity of the j[th] vehicle in the network is v_j, then the average velocity of vehicles in the network represented by V is

Total vehicle mile = $\sum_{j=1}^{n} v_j \varepsilon$

Total vehicle hours = $n\varepsilon$

$$V = \frac{\sum_{j=1}^{n} v_j \varepsilon}{n\varepsilon} \tag{5}$$

$$\Rightarrow V = \frac{\sum_{j=1}^{n} v_j}{n} \tag{6}$$

The average concentration K in the network is defined as the total number of vehicles in the network per unit lane mile. l is the total lane miles of roadway in the network.

$$K = \frac{n}{l} \tag{7}$$

The average network flow Q is defined as the average number of vehicles that pass through a random point in the network per unit time. The average network flow is given by

$$Q = \frac{\sum l_i Q_i}{\sum l_i} \tag{8}$$

In Kalfastas and Peeta's work (*16*) on the cell transmission model they observed that on individual links the backward propagating wave speed is lower than the free-flow speed, indicating that the link can never reach maximum jam density, since only a part of the available space would be filled up. Since a network is a combination of individual links, it is fair to assume that during maximum congestion, there exists a maximum accumulation (n_p) in the network and it corresponds to some minimum outflow (q_p). In Ardekani's (*5*) field study a maximum concentration of 30 vehicles/lane mile was observed in the network.

With the next generation of technology of Vehicle Infrastructure Integration (V.I.I.), data of the network will be available during every time instant. Hence the average velocity, flow and density of the network would be known at every time instant. This will help provide real-time state of the network, enabling us to develop real-time strategies for the network.

MACROSCOPIC PROPERTIES

Measurements and relationships between macroscopic variables—flow, concentration, and speed—have been extensively studied for traffic streams both in theory and in field [Edie (*15* and Gazis (*17*)]. The relationships between these macroscopic variables at a network scale have been studied in simulation experiments (*7*, *9*, *8*, *12*) and field studies (*1*). This section provides theoretical proofs for the validity of relationships at a network scale. These relationships are then validated using microscopic simulation (VISSIM) for a network shown in Figure 1. The network is a small grid network of two-lane, one-way roads, and the entire length of the roadway is 2.12 miles. The intersections consisted of two-phase signals with a cycle length of 60 seconds.

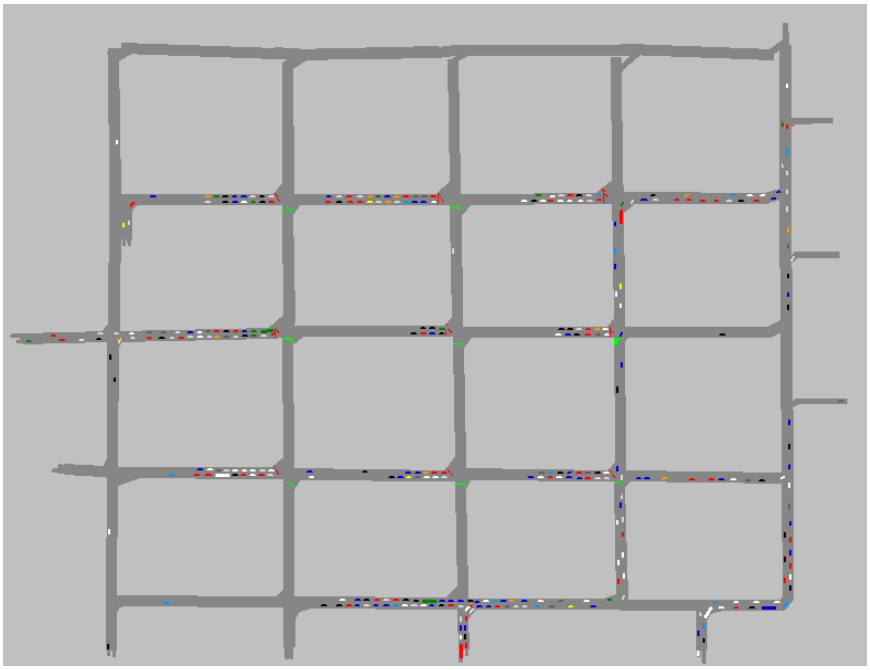


FIGURE 1 Simulation network considered for the study.

The simulation for the network consisted of multiple runs with each run of 3,240 seconds. An input of 2,000 veh/h was provided at each input. The average concentrations, speeds, flows, inflows and outflows were averaged over 120 seconds for the network. The demands were allocated to routes from origin to destinations. We also assume a fixed percentage of demand for each route on the network, to ensure that the average trip lengths are constant.

FUNDAMENTAL NETWORK SPEED-FLOW-CONCENTRATION RELATIONSHIP

This subsection derives the fundamental speed-flow-concentration relationship on a general network.

On an individual link i of length l_i in the network, it is known that

$$Q_i = K_i V_i \tag{9}$$

where V_i is the average velocity on link i defined by Equation 10

$$V_i = \frac{\sum_{\text{vehicle } j \text{ is in Link } i} v_j}{n_i} \tag{10}$$

Multiplying both sides of Equation 9 with l_i, we get

$$l_i Q_i = l_i K_i V_i \tag{11}$$

Summing both sides of Equation 11 over all links i in the network and dividing by the total lane miles in the network, we get

$$\frac{\sum l_i Q_i}{\sum l_i} = \frac{\sum l_i K_i V_i}{\sum l_i} \tag{12}$$

It is observed that the left hand side of Equation 12 is the definition for average network flow. It is also observed that $l_i K_i$ is the number of vehicles in link i. Substituting $l_i K_i$ with n_i, and the definition of average network flow in Equation 12.

$$Q = \frac{\sum (l_i K_i) V_i}{\sum l_i} = \frac{\sum (n_i) V_i}{\sum l_i} \tag{13}$$

$$\Rightarrow Q = \left(\frac{\sum n_i}{\sum n_i}\right)\left(\frac{\sum (n_i) V_i}{\sum l_i}\right) = \left(\frac{\sum n_i}{\sum l_i}\right)\left(\frac{\sum (n_i V_i)}{\sum n_i}\right) \tag{14}$$

Substitute network concentration and definition of V_i from Equation 10 in Equation 14.

$$Q = K\left(\frac{\sum\left(n_i\left(\frac{\sum_{\text{vehicle } j \text{ is in Link } i} v_j}{n_i}\right)\right)}{\sum n_i}\right) \tag{15}$$

$$\Rightarrow Q = K\left(\frac{\sum \sum_{\text{vehicle } j \text{ is in Link } i} v_j}{\sum n_i}\right) \tag{16}$$

In Equation 16, $\sum \sum_{\text{vehicle } j \text{ is in Link } i} v_j$ is the sum of velocities of all vehicles on the network. Hence by definition $\left(\frac{\sum \sum_{\text{vehicle } j \text{ is in Link } i} v_j}{\sum n_i}\right)$ is the average velocity on the network.

$$\Rightarrow Q = KV \tag{17}$$

Equation 17 is the fundamental network speed-flow-concentration relationship.

The interesting aspect of Equation 17 is that there are no inherent assumptions involved in the derivations, indicating that the fundamental network speed-flow-concentration relationship

holds for a network in any state as long as the variables are defined in the correct manner. The fundamental relationship for speed-flow-concentration should hold when the vehicles are non-homogenously loaded in the network or when the network is in a transient state.

Using data of network density and average network speed from the simulation runs the average network flow was calculated using Equation 17 and was compared to the observed average network flow from the simulation. A plot (Figure 2) between the calculated and observed average network shows a perfect regression fit for $y=x$.

SPEED–ACCUMULATION RELATIONSHIP AT NETWORK LEVEL

This section derives Greenshields,' Greenberg's and the bell-shaped model (*11*) to describe the relationship between average network speed and accumulation in a homogenous network (a network in which the concentration, jam concentration and free-flow speeds do not significantly differ between different links). The theoretical results are validated through simulation for the network described earlier. A plot was constructed for data points, such that vehicles are homogenously distributed.

Greenshields' Relationship

To derive the Greenshields' model for a network, it is assumed that the speed-density relationship on an individual link i in a homogenous network follows the Greenshields model.

$$V_i = V_f\left(1-\frac{K_i}{K_j}\right)$$

Since the free flow speed (V_f) and jam density (K_j) do not significantly vary between links of the network.

$$\Rightarrow \sum(n_i V_i) = V_f\left(\sum n_i - \frac{\sum(n_i K_i)}{K_j}\right)$$

$$\Rightarrow \frac{\sum(n_i V_i)}{\sum n_i} = V_f\left(\frac{\sum n_i}{\sum n_i} - \frac{\sum(n_i K_i)}{K_j \sum n_i}\right)$$

Since the link densities do not significantly differ between links and is approximately equal to the network density $K_i \sim K$.

$$\Rightarrow V = V_f\left(1-\frac{\sum(n_i K)}{K_j \sum n_i}\right) = V_f\left(1-\frac{K\sum n_i}{K_j \sum n_i}\right) = V_f\left(1-\frac{(n/l)}{(n_j/l)}\right)$$
$$\Rightarrow V = V_f\left(1-\frac{n}{n_j}\right) \quad (18)$$

This proves that Greenshields' relationship holds for average network speed and accumulation in a homogenous network.

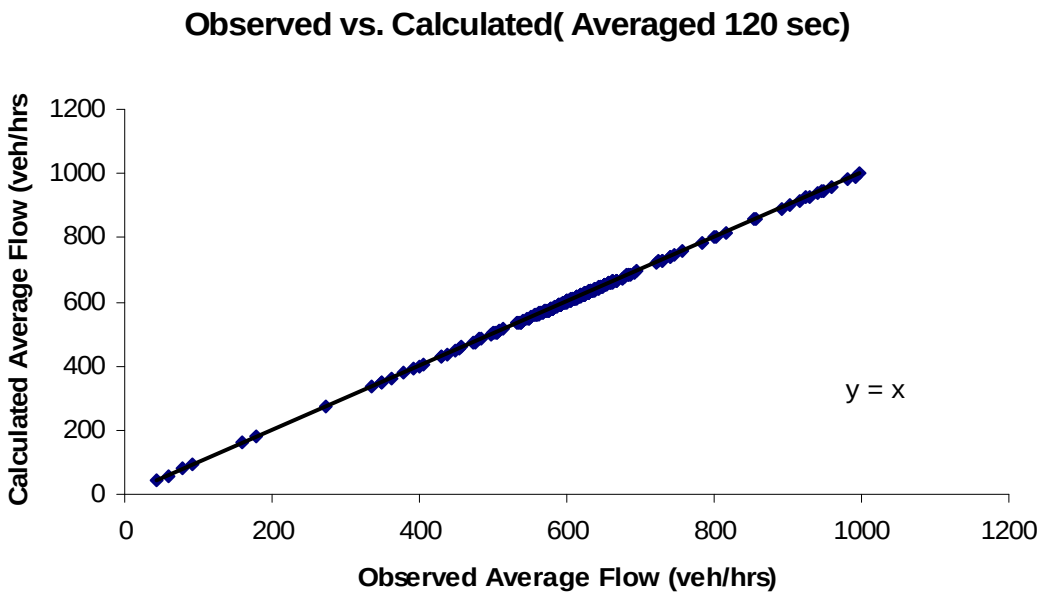


FIGURE 2 Observed flow versus flow calculated using speed and density in Equation 2.

Greenberg's Relationship

To derive the Greenberg model for a network, it is assumed that the speed-density relationship on an individual link *i* in a homogenous network follows Greenberg's model.

$$V_i = -V_f \ln\left(\frac{K_i}{K_j}\right)$$

Since the free-flow speed (V_f) and jam density (K_j) do not significantly vary between links of the network.

$$\Rightarrow \sum(n_i V_i) = -\sum n_i V_f \ln\left(\frac{K_i}{K_j}\right)$$

$$\Rightarrow \frac{\sum(n_i V_i)}{\sum n_i} = \frac{-\sum n_i V_f \ln\left(\frac{K_i}{K_j}\right)}{\sum n_i}$$

Since the link densities do not significantly differ between links and is approximately equal to the network density $K_i \sim K$.

$$\Rightarrow \frac{\sum(n_i V_i)}{\sum n_i} = \frac{-\sum n_i V_f \ln\left(\frac{K_i}{K_j}\right)}{\sum n_i} = -\frac{\sum n_i}{\sum n_i} V_f \ln\left(\frac{K}{K_j}\right) = -V_f \ln\left(\frac{(n/l)}{(n_j/l)}\right)$$
$$V_i = -V_f \ln\left(\frac{n}{n_j}\right) \quad (19)$$

This proves that Greenberg's relationship holds for average network speed and accumulation in a homogenous network.

Bell-Shaped Relationship

To derive the bell-shaped model for a network, it is assumed that the speed-density relationship on an individual link *i* in a homogenous network follows the bell-shaped model.

$$V_i = V_f \exp\left[-\alpha\left(\frac{K}{K_j}\right)^d\right]$$

Since the free-flow speed (V_f) and jam density (K_j) do not significantly vary between links of the network.

$$\Rightarrow \sum(n_i V_i) = \sum n_i V_f \exp\left[-\alpha\left(\frac{K_i}{K_j}\right)^d\right]$$
$$\Rightarrow \frac{\sum(n_i V_i)}{\sum n_i} = \frac{\sum n_i V_f \exp\left[-\alpha\left(\frac{K_i}{K_j}\right)^d\right]}{\sum n_i}$$

Since the link densities do not significantly differ between links and is approximately equal to the network density $K_i \sim K$.

$$\frac{\sum(n_i V_i)}{\sum n_i} = \frac{\sum n_i V_f \exp\left[-\alpha\left(\frac{K}{K_j}\right)^d\right]}{\sum n_i} = \frac{\sum n_i}{\sum n_i} V_f \exp\left[-\alpha\left(\frac{K}{K_j}\right)^d\right] = V_f \exp\left[-\alpha\left(\frac{(n/l)}{(n_j/l)}\right)^d\right]$$
$$V_i = V_f \exp\left[-\alpha\left(\frac{n}{n_j}\right)^d\right] \quad (20)$$

This proves that relationship proposed by Drake, Shofer and May (*10*) holds for average network speed and accumulation in a homogenous network.

The plot (Figure 3) for the simulation results were fitted with Greenshields', Greenberg's, and the bell-shaped model. When the results were fitted with a linear regression the *R*-square was found to be 0.79. A logarithmic fit showed to perform better with an *R*-square of 0.93. The fit for the bell-shaped model proposed by Drake, Shofer and May (*10*) performed the best with an *R*-square of 0.964. The bell-shaped fit for the results showed to perform significantly better in explaining the speed–accumulation relationship than Greenshields' model and Greenberg's model.

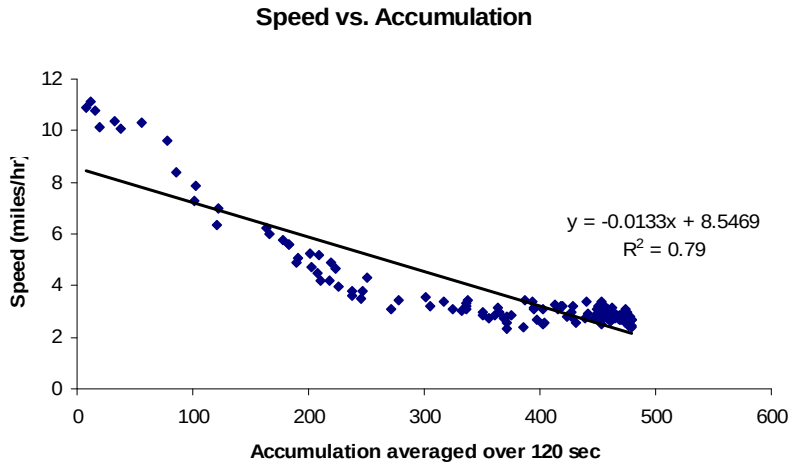


FIGURE 3*a* Model fit for Greenshield's model for speed–accumulation relationship.

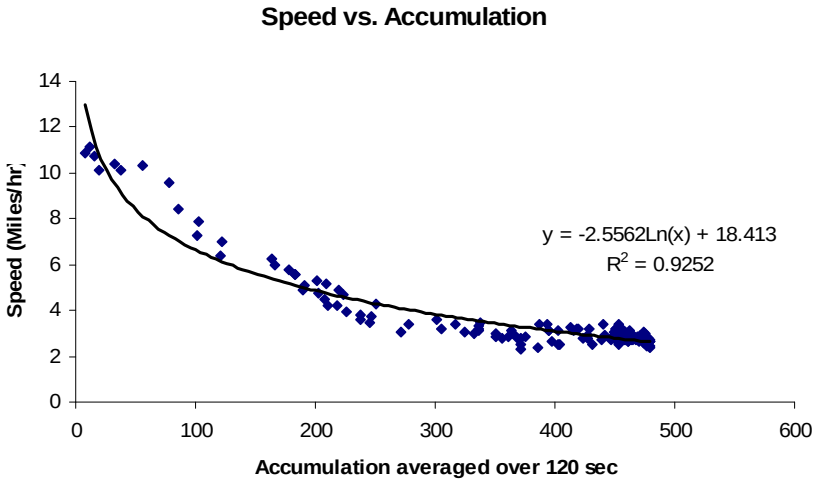


FIGURE 3*b* Model fit for Greenberg's model for speed–accumulation relationship.

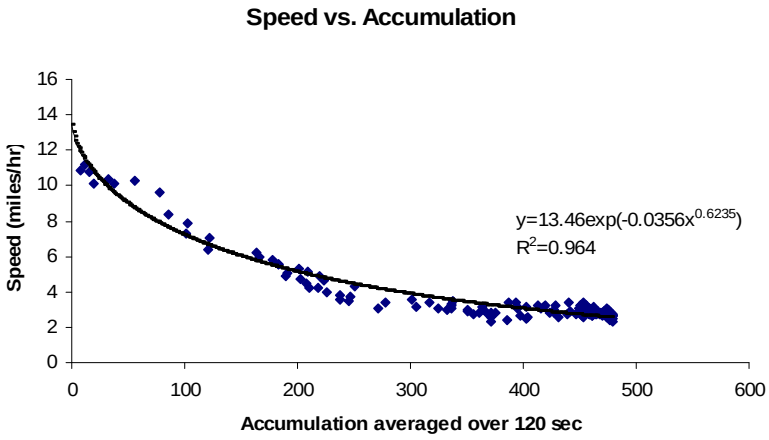


FIGURE 3*c* Model fit for Bell-shaped model for speed–accumulation relationship.

OUTFLOW-ACCUMULATION RELATIONSHIP AT NETWORK LEVEL

To understand the relationship between outflow and accumulation, the outflow and accumulation averaged over 120 seconds collected during the simulation of the network described earlier was plotted (Figure 4).

The plot shows two distinct regimes. The first regime corresponds to the constrained regime, where the outflow increases to a maximum value (q_c=178 veh/120 s) as the accumulation increases. This maximum outflow corresponds to an accumulation of n_c equal to 212 veh. The second regime corresponds to the constrained regime, where the outflow decreases from its maximum value as the accumulation increases. The reason for the reduction in outflow during the constrained regime is due to blockage of exits by vehicles accessing other exits. Observing the trend in Figure 4, it is fair to assume a piecewise linear relationship to describe the two regimes of the outflow-accumulation relationship (Figure 5). The piecewise formulation describing the outflow-accumulation relationship is written as

$$Q_{out}(n) = \begin{cases} \lambda n & \forall n \le n_c \\ -\alpha n + \beta & \forall n \ \ s.t.\ n_p \ge n > n_c \end{cases} \tag{21}$$

For this network λ was equal to 0.84, α was equal 0.084 to and β was found to be 179. It was observed that the maximum accumulation was 480 vehicles, the flow at which was found to be (q_p) 130 veh/120 s.

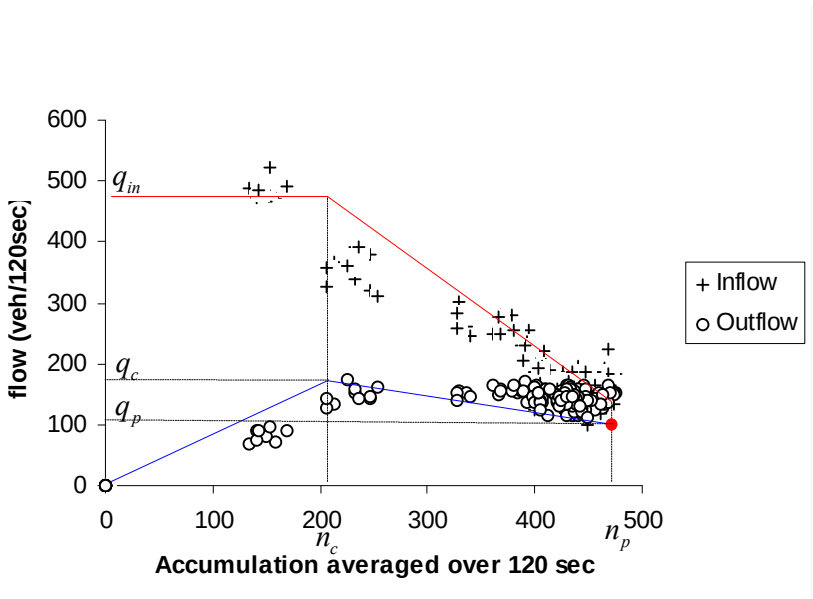


FIGURE 4 Plot of inflow and outflow versus accumulation and the trend assumed.

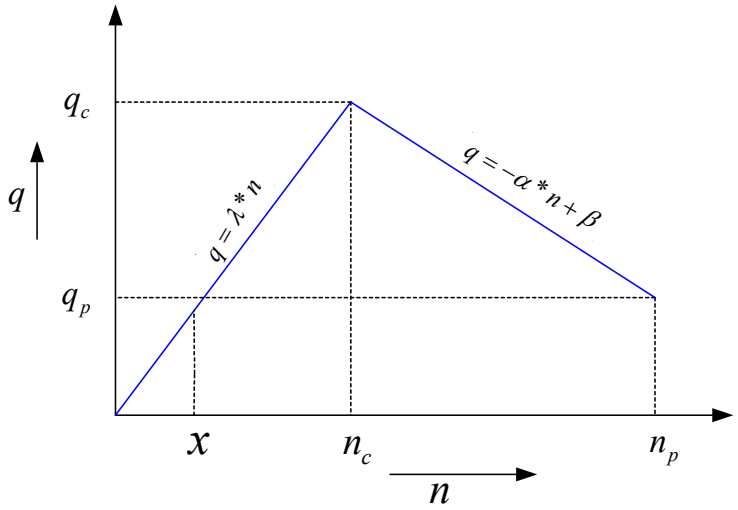


FIGURE 5 Conceptualization of relationship between outflow and accumulation.

INFLOW-ACCUMULATION RELATIONSHIP AT NETWORK LEVEL

Simulation studies by Geroliminis and Daganzo (*14*) showed that the dependence of inflow on accumulation had a trend similar to as seen in Figure 4. The inflow remains constant until the accumulation reaches n_c and then begins to decrease. The accumulation n_c corresponds to the boundary of the constrained and unconstrained outflow. During the unconstrained regime vehicles have an inflow equal the maximum possible flow q_{in}, since none of the entries into the network are blocked. The moment the constrained conditions begin to set in, the outflow reduces leading to queuing of vehicles and blocking of inflows. This results in the inflow being

constrained. Under the constrained regime the inflow keeps reducing, as the accumulation increases until the maximum accumulation (n_p) allowed in the network. The moment the number of vehicles in the network reaches n_p, the inflow drops to the outflow q_p. Under the assumption that the network can have a maximum of n_p accumulation, if the inflow is greater than outflow (q_p) the accumulation would increase to a value greater than n_p, resulting in a violation of the assumption. Therefore the inflow will be equal to the outflow at the accumulation n_p. Therefore there is a discontinuity for the inflow at n_p. It will be later seen in this paper that if the inflow has a continuous trend and the constrained regime of the inflow intersects the outflow at the maximum accumulation (n_p), then infinite time would be taken to reach congested conditions (accumulation of n_p). In reality networks do reach congested conditions and are regularly observed in road networks around the world. To get around this problem of infinite convergence to congestion, discontinuity was assumed at the maximum accumulation (n_p). The inflow into the network is greater than q_p just before an accumulation of n_p. Due to stochastic effects there are fluctuations in the accumulations, and may be periods when the total accumulation is just lower than the maximum accumulation (n_p). Therefore, the traffic inflow needs to stop for brief periods, so as to maintain an average inflow of q_p. These kinds of stop-and-go behaviors with large oscillations in flows (*6*) have been observed in congested conditions and can be explained by such a formulation. Therefore it is fair to assume a discontinuity. The relationship between the inflow and accumulation is shown in Figure 6, and can be formulated as

$$Q_{in}(n) = \begin{cases} q_{in} & \forall n \ \ s.t. \ \ n \le n_c \\ -\gamma n + \sigma & \forall n \ \ s.t. \ \ n_p > n > n_c \\ q_p & n = n_p \end{cases} \tag{22}$$

A plot describing the conceptualization of the trends of the outflow-accumulation relationship and the inflow-accumulation relationship are shown in contrast to the observed values in Figure 4. Since the inflow was large the network quickly progressed towards constrained condition. As can be seen in Figure 4, the vehicles remained in unconstrained regime for a brief period, during which the inflow was found to be 460 vehicles/120 s. The inflow of vehicles just before n_p was taken to be 155 veh/120 s and the value of outflow at accumulation n_p (q_p) was found to be around 138 veh/120 s, which was also the value of the inflow at an accumulation of n_p. γ was found to be 0.87 and σ was equal to 573. These values were calibrated for do-nothing strategy.

MODELING AND METHODOLOGY

This section presents a formal methodology for the network breathing strategy. In the network breathing strategy the inflow is allowed into the network until the network reaches jam conditions, after which the network is allowed to relax, implying that inflow into the network is shut down. The intuitive reason in adopting this cyclical approach is to ensure that no sustained constrained flow (q_p) at maximum accumulation (n_p) exists. This can be observed more clearly in Figure 7. A comparison between the area under the curve between Figure 7*a* and 7*b* shows that the total number of vehicles dissipated into the network is larger with network breathing. The important aspect to determine is that relaxing the network for how long would provide the benefit being looked for.

The dynamics of the number of vehicles (accumulation) in the network can be described by a differential equation (Equation 23). The differential equation basically states the rate of change of accumulation is the difference between the inflow and the outflow. If the outflow is greater than the inflow then the accumulation will decrease with time. If the outflow is less than the inflow the accumulation will increase with time. It is assumed that the maximum inflow (q_{in}) is greater than the maximum outflow (q_c).

$$\frac{dn}{dt} = Q_{in}(n) - Q_{out}(n) \tag{23}$$

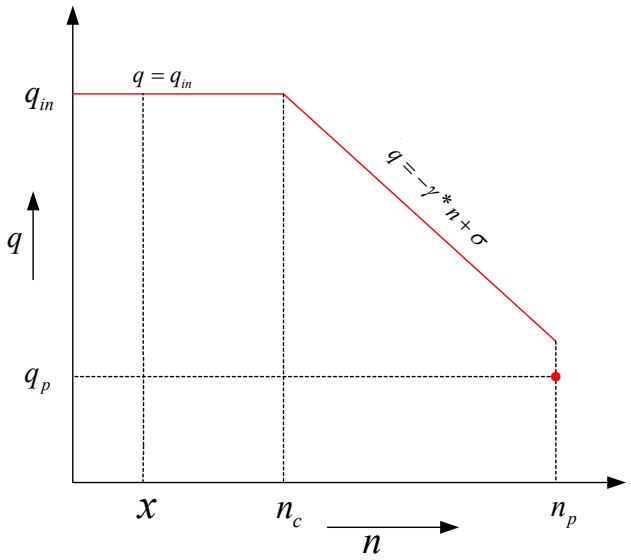


FIGURE 6 Plots describing relationships between inflow and accumulation.

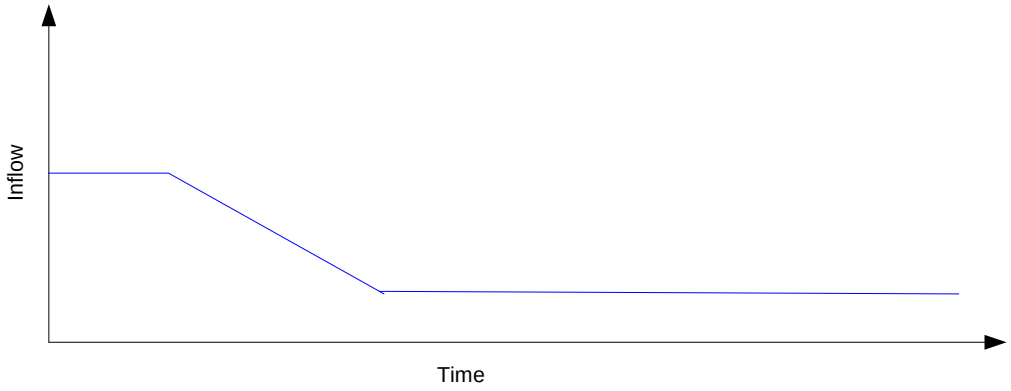


Figure 7*a* Inflow profile under normal conditions

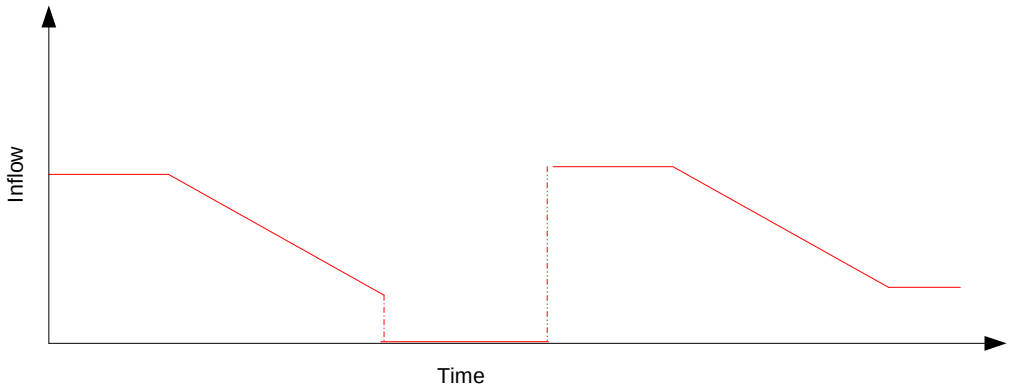


Figure 7*b* Inflow profile under Network breathing scheme.

When vehicles initially enter the network the outflow and inflow correspond to the unconstrained regime ($n < n_c$). Therefore the time taken to dissipate vehicles during the unconstrained regime when accumulation grows from 0 to n_c, t_s is shown in Equation 25.

$$\int_0^{n_c} \frac{dn}{(q_{in} - \lambda n)} = \int_0^{t_s} dt \tag{24}$$

$$t_s = -\frac{1}{\lambda} \ln\left(1 - \frac{\lambda n_c}{q_{in}}\right) \tag{25}$$

After the accumulation in the network has reached n_c, the equations describing the inflow and outflow shift to the constrained regime, where $n > n_c$. The maximum accumulation that can be reached in the network is n_p. The time taken for the accumulation in the network to grow from n_c to n_p is represented by t_d, in Equation 26.

$$\int_0^{t_d} dt = \int_{n_c}^{n_p} \frac{dn}{((-\gamma n + \sigma) - (-\alpha n + \beta))} = \int_{n_c}^{n_p} \frac{dn}{((\alpha - \gamma) n + (\sigma - \beta))}$$

$$t_d = \frac{1}{(\alpha - \gamma)} \ln\left(\frac{(\alpha - \gamma) n_p + (\sigma - \beta)}{(\alpha - \gamma) n_c + (\sigma - \beta)}\right) \tag{26}$$

If the network has an accumulation at some intermediary value x then the time taken in the constrained regime for the accumulation to reach the maximum accumulation n_p is represented by $t_d(x)$ and is shown in Equation 27. $t_d(x)$ is the time taken to reach n_p under constrained conditions, therefore the minimum value x can have during constrained conditions is n_c, but if $x \leq n_c$, then the time in the constrained region is t_d in Equation 26.

$$t_d(x) = \begin{cases} \dfrac{1}{(\alpha-\gamma)} \ln\left(\dfrac{(\alpha-\gamma)n_p + (\sigma-\beta)}{(\alpha-\gamma)n_c + (\sigma-\beta)}\right) & \forall x \le n_c \\ \dfrac{1}{(\alpha-\gamma)} \ln\left(\dfrac{(\alpha-\gamma)n_p + (\sigma-\beta)}{(\alpha-\gamma)x + (\sigma-\beta)}\right) & \forall x > n_c \end{cases} \quad (27)$$

It will be interesting to observe here the formulation of $t_d(x)$ in Equation 27. The numerator inside the logarithmic term of Equation 27 is the difference between the inflow and the outflow at the maximum accumulation (n_p). If the inflow describing the constrained region intersected the outflow at n_p without discontinuity, the value of the numerator would tend to 0, making the value of $t_d(x)$ at n_p infinite. This would indicate that it would take infinite time to reach jam conditions, which is obviously not true. For this reason it would be fair to assume a discontinuity at n_p for inflow.

According to the network breathing strategy proposed, once the network is congested with the maximum accumulation possible, the network relaxation is started. During this process inflow to the city is prohibited, allowing the accumulation in the network to reduce and, therefore, outflow would increase. The basic idea behind this approach is to maintain a total inflow to the city greater than the inflow at the maximum accumulation over a given period of time.

If the network is relaxed to an accumulation of x, the time taken to reach an accumulation of x from n_p depends on whether $x \le n_c$ *or* $x > n_c$ and is represented by $t_b(x)$. If $x < n_c$ then during the reduction of accumulation from n_p to n_c the outflow will correspond to the constrained regime, after which the outflow will correspond to the unconstrained regime. Therefore the time taken to reach x if $x \le n_c$ is the sum of the time taken in the two regimes. This is shown in Equation 28.

$$t_b(x) = \int_{n_p}^{n_c} \frac{dn}{(0-(-\alpha n+\beta))} + \int_{n_c}^{x} \frac{dn}{(0-\lambda n)} = \int_{n_p}^{n_c} \frac{dn}{(\alpha n-\beta)} + \int_{n_c}^{x} \frac{dn}{-\lambda n} \qquad \forall x \le n_c$$

$$t_b(x) = \frac{1}{\alpha} \ln\left(\frac{\alpha n_c - \beta}{\alpha n_p - \beta}\right) - \frac{1}{\lambda} \ln\frac{x}{n_c} \qquad \forall x \le n_c \quad (28)$$

If $x > n_c$, then the drop in accumulation in the network from n_p to x occurs in the constrained regime, therefore the outflow corresponds to the constrained regime. Hence the second term in the R.H.S. of Equation 28 is dropped, and since the network is relaxed to x. ($x > n_c$), n_c is replaced with x. Therefore the time taken to relax when $x > n_c$ is given by Equation 29.

$$t_b(x) = \frac{1}{\alpha} \ln\left(\frac{\alpha x - \beta}{\alpha n_p - \beta}\right) \qquad \forall x > n_c \quad (29)$$

$$\therefore t_b(x) = \begin{cases} \frac{1}{\alpha}\ln\left(\frac{\alpha n_c - \beta}{\alpha n_p - \beta}\right) - \frac{1}{\lambda}\ln\frac{x}{n_c} & \forall x \le n_c \\ \frac{1}{\alpha}\ln\left(\frac{\alpha x - \beta}{\alpha n_p - \beta}\right) & \forall x > n_c \end{cases} \quad (30)$$

After the network relaxation process, the inflow is allowed back into the network, due to which the accumulation begins to increase. $t_n(x)$ is defined as the time for the network accumulation to grow from x to n_c under the unconstrained regime. This is shown in Equation 31.

$$\int_x^{n_c} \frac{dn}{(q_{in} - \lambda n)} = \int_0^{t_n} dt$$

$$t_n(x) = -\frac{1}{\lambda}\ln\left(\frac{\lambda n_c - q_{in}}{\lambda x - q_{in}}\right) \qquad \forall x \le n_c \quad (31)$$

The time taken for the accumulation to grow back to n_p is the sum of the time taken for the accumulation to grow from x to n_c under unconstrained regime and the time taken for accumulation to grow from n_c to n_p under the constrained regime. If $x>n_c$, the accumulation to which network was relaxed lies in the constrained regime. Since $t_n(x)$ is defined for unconstrained conditions, $t_n(x)$ is taken to be *0* for $x > n_c$. The functional form for $t_n(x)$ is given in Equation 32.

$$t_n(x) = \begin{cases} -\frac{1}{\lambda}\ln\left(\frac{\lambda n_c - q_{in}}{\lambda x - q_{in}}\right) & \forall x \le n_c \\ 0 & \forall x > n_c \end{cases} \quad (32)$$

Since the inflow remains constant during the unconstrained regime the number of vehicles dissipated into the network during unconstrained conditions is the product of q_{in} and the time spent in the unconstrained region. To determine the number of vehicles dissipated into the network during the network breathing process, it is required to determine the increase in accumulation during constrained conditions.

From Equation 22, the inflow during congested conditions can be described by Equation 33.

$$Q_{in} = -\gamma n + \sigma \qquad \forall n > n_c \quad (33)$$

$$\frac{dn_{in}}{dt} = -\gamma n + \sigma \qquad \forall n > n_c \quad (34)$$

$$\frac{dn_{in}}{dn}\frac{dn}{dt} = -\gamma n + \sigma \qquad \forall n > n_c \tag{35}$$

$$\frac{dn_{in}}{dn} = \frac{-\gamma n + \sigma}{\left(\frac{dn}{dt}\right)} \qquad \forall n > n_c \tag{36}$$

Using Equation 23 and inserting the equations describing the inflows and outflows during the constrained regime, the rate of change of accumulation in the network during constrained conditions can be described by Equation 37.

$$\frac{dn}{dt} = (\alpha - \gamma)n + (\sigma - \beta) \qquad \forall n > n_c \tag{37}$$

To derive a differential equation (Equation 38) describing the dynamics of the number of vehicles dissipated into the network with respect to the accumulation, Equation 37 is replaced in Equation 36.

$$\frac{dn_{in}}{dn} = \frac{-\gamma n + \sigma}{\left((\alpha - \gamma)n + (\sigma - \beta)\right)} \qquad \forall n > n_c \tag{38}$$

To determine the number of vehicles dissipated into the network from an accumulation of x to n_p, ($n_{in}(x)$), under constrained conditions, is shown in Equation 39.

$$n_{in}(x) = \int_x^{n_p} \frac{-\gamma n + \sigma}{\left((\alpha - \gamma)n + (\sigma - \beta)\right)} dn \quad \forall n > n_c$$

$$n_{in}(x) = \frac{-\gamma(n_p - x)}{(\alpha - \gamma)} + \frac{\gamma}{(\alpha - \gamma)}\left(\frac{\sigma}{\gamma} + \frac{(\sigma - \beta)}{(\alpha - \gamma)}\right)\ln\left(\frac{n_p + \frac{(\sigma - \beta)}{(\alpha - \gamma)}}{x + \frac{(\sigma - \beta)}{(\alpha - \gamma)}}\right) \qquad \forall x \geq n_c \tag{39}$$

When the accumulation grows from an unconstrained region, the dissipation of vehicles will be $q_{in}t_n(x)$, after which the number of vehicles dissipated under constrained conditions is $n_{in}(n_c)$. Therefore, for $x<n_c$, $n_{in}(x)$ is defined as $n_{in}(n_c)$. The functional form for $n_{in}(x)$ is given in Equation 40.

$$n_{in}(x)=\begin{cases}\dfrac{-\gamma(n_p-n_c)}{(\alpha-\gamma)}+\dfrac{\gamma}{(\alpha-\gamma)}\left(\dfrac{\sigma}{\gamma}+\dfrac{(\sigma-\beta)}{(\alpha-\gamma)}\right)\ln\left(\dfrac{n_p+\dfrac{(\sigma-\beta)}{(\alpha-\gamma)}}{n_c+\dfrac{(\sigma-\beta)}{(\alpha-\gamma)}}\right) & \forall\, x< \eta \\ \dfrac{-\gamma(n_p-x)}{(\alpha-\gamma)}+\dfrac{\gamma}{(\alpha-\gamma)}\left(\dfrac{\sigma}{\gamma}+\dfrac{(\sigma-\beta)}{(\alpha-\gamma)}\right)\ln\left(\dfrac{n_p+\dfrac{(\sigma-\beta)}{(\alpha-\gamma)}}{x+\dfrac{(\sigma-\beta)}{(\alpha-\gamma)}}\right) & \forall x > n_c\end{cases} \tag{40}$$

Our objective is to dissipate the maximum number of vehicles into the destination network during time *T*. The total time can then be written as the sum of the initial time taken to reach an accumulation of n_p from 0 and the product of the number of network breathing cycles taken and the time spent in each network breathing scheme. The time taken for the network breathing scheme is the sum of time required to relax the network to an accumulation of x ($t_b(x)$), the time required to get back to n_c ($t_n(x)+t_d(x)$). The cycles might not be able to complete the entire time period, hence during the rest of the time (*del*) the outflow is equal to the inflow (q_p). Hence T is written as Equation 41.

$$T=t_s+t_d(n_c)+m(t_b(x)+t_d(x)+t_n(x))+del \tag{41}$$

The number of network breathing cycles in the process of dissipation is shown in Equation 42.

$$m=\left\lfloor\frac{(T-t_s-t_d(n_c))}{(t_b(x)+t_d(x)+t_n(x))}\right\rfloor \tag{42}$$

The excess time left (*del*) after the network breathing cycles is given by Equation 43.

$$del=T-\left(t_s+t_d(n_c)+\left\lfloor\frac{(T-t_s-t_d(n_c))}{(t_b(x)+t_d(x)+t_n(x))}\right\rfloor(t_b(x)+t_d(x)+t_n(x))\right) \tag{43}$$

During this time the accumulation in the network is at the maximum and the outflow is equal to the inflow at q_p. Therefore the number of vehicles dissipated into the network during this period (*del*) is given by Equation 44.

$$excessN=\begin{cases}del*q_p & del>0\\ 0 & del\le 0\end{cases} \tag{44}$$

The total number of vehicles dissipated into the network during time *T,* can be written as the sum of vehicles dissipated into the network during unconstrained conditions, constrained conditions, and the number of vehicles dissipated during each cycle and during time *del*.

$$N=q_{in}t_s+n_{in}(n_c)+m(q_{in}t_n(x)+n_{in}(x))+excessN \tag{45}$$

Replacing value of *m* in Equation 45 we get Equation 46.

$$N = q_{in}t_s + n_{in}(n_c) + \left\lfloor \frac{(T\text{-}t_s - t_d(n_c))}{(t_b(x) + t_d(x) + t_n(x))} \right\rfloor (q_{in}t_n(x) + n_{in}(x)) + excessN \quad (46)$$

To determine the strategy that maximizes the dissipation of vehicles into the network, *N*, in Equation 46, is maximized in order to determine the accumulation to which the network should be relaxed. Using this value the relaxation time and the time for network inhalation can be calculated, which can then be used to meter the inflow into the network.

RESULTS

To test the proposed above methodology the relationships between the outflows and inflows with the accumulations were calibrated for the network earlier described (Figure 1). The values observed during the analysis of the outflow versus accumulation and inflow versus. accumulation, were used to calibrate the equation describing these models.

To determine the accumulation to which the network should be relaxed during the network breathing strategy (the variable, *x*) so that the number of vehicles dissipated into the network is maximized, Equation 45 was maximized with respect to relaxation time. The time unit taken was 120 seconds, since all flows and number of vehicles were averaged over 120 seconds. The maximization of Equation 45 indicated that the network should be relaxed for a period of 2.25 units (2.25*120 s) to an accumulation of 148 vehicles. A plot (Figure 8) constructed between the relaxation time and the number of vehicles dissipated into the network indicates that the dissipation of number of vehicles drops significantly quickly after the maximum. Therefore to be on the conservative side the network was relaxed for 1 unit of time (120 s) after which the inflow was allowed for 360 seconds.

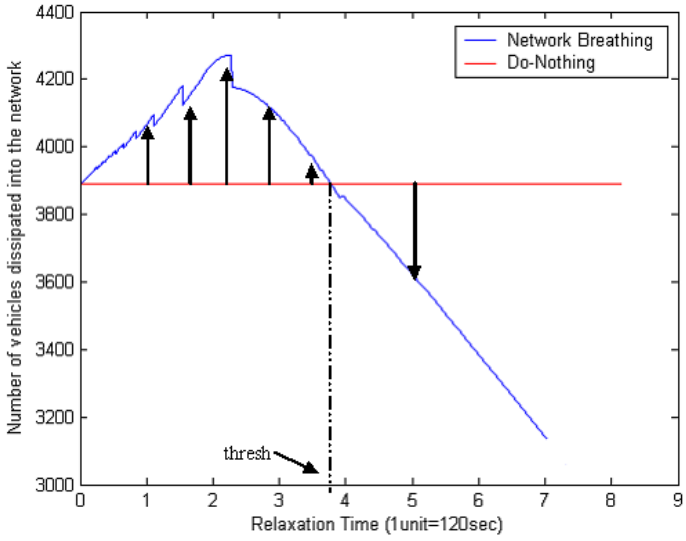


FIGURE 8 The plot between the number of vehicles dissipated into the network and time allowed for the network to relax.

The simulation was run for 3,240 seconds for multiple runs. It was found that the average number of vehicles dissipated during the do-nothing scenario for the same time period was 3,022 vehicles and for the network breathing scenario, where the network was relaxed for 120 seconds and inflow allowed for 360 seconds had a statistically significant increase in the dissipation of vehicles in the network to 4,076 an increase of 154 vehicles in 3,240 seconds. The simulation results showed that the network breathing strategy performs very well. Figure 9 indicates the results of the observed and predicted number of vehicles dissipated vs. the estimated. The errors printed above the observed and predicted number of vehicles in Figure 9, is less than 1%, indicating a good performance of our modeling methodology.

CONCLUSION

This paper provides a theoretical basis to various relationships observed previously between average network level variables in simulation and field experiments. It then utilizes these relationships to develop network scale strategies that increase the dissipation of the number of vehicles into the network.

Using the theoretical framework for the network breathing strategy, the relaxation time and the network inhalation times were determined and tested using simulation. The network breathing strategy considerably outperformed a do-nothing scenario. One of the main advantages of the network breathing strategy is that it does not required real time feedback. If the properties of the network are known earlier, then the methodology described earlier can be used to come up with a prescription for relaxation time and network inhalation time. Under emergency conditions when the real-time feedback might not be able to be implemented such an approach will be very useful. The network breathing strategy can be implemented using signals of cycle length determined by the relaxation time and the network inhalation time.

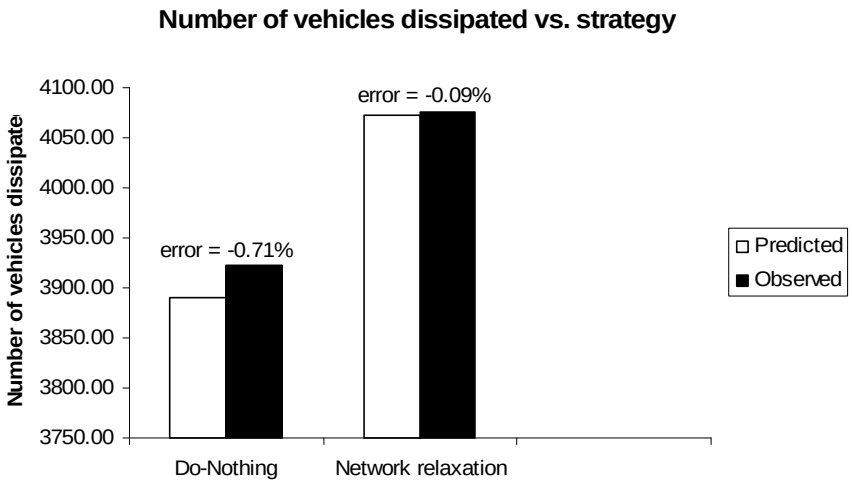


FIGURE 9 Results between prediction of number of vehicles dissipated into the network through theoretical methodology and actual observed number of vehicles dissipated.

Most of the microscopic simulation studies do not consider the effect of the destination network. This results in a myopic analysis resulting in a gap between the expectations from simulation analysis and reality. The macroscopic relationships and modeling discussed need to be used in conjunction with microscopic simulation, in order for the models to incorporate effects of destination nodes.

During evacuation it is crucial to keep the traffic moving at an acceptable rate and reduce delays. With present strategies vehicles wanting to exit onto a particular destination are constrained by the state of the destination network. This results in backups which result in reduction of flows on the mainstream evacuation route, hindering flow of vehicles wanting to go to a destination further downstream on the evacuation route. The network breathing scheme can be used to effectively increase the dissipation into the destination networks. During the restriction of the inflow into one destination network (network relaxation) the evacuating vehicles can be redirected to another destination, hence improving the flow on the evacuation routes.

ACKNOWLEDGMENTS

The authors acknowledge PhD student Vikash V. Gayah from the University of California, Berkeley, for his useful comments, which helped improve the quality of this paper immensely.

REFERENCES

1. Minutes of the SE U.S. Regional Transportation Analysis Meeting, Held in the FEMA Region IV Office in Atlanta, On March 7–8, 2000, Available at http://www.sam.usace.army.mil/pd/hes/minutes7-8Mar2000.htm, Accessed: July 6, 2007.
2. Zahavi, Y. Traffic Performance Evaluation of Road Networks by the α -Relationship. In *Traffic Engineering and Control*, Vol. 14, No. 5, 1972.
3. Zahavi, Y. Traffic Performance Evaluation of Road Networks by the α -Relationship, Part 2.In *Traffic Engineering and Control*, Vol. 14, No. 6, 1972.
4. Buckley, D.G., and J. G. Wardrop. Some General Properties of a Traffic Network. In *Australian Road Research*, Vol. 10, No.1, 1980.
5. Ardekani, S. A. The Two-Fluid Characterization of Urban Traffic: Theory Observation, and Experiment. Ph.D. dissertation. University of Texas at Austin, 1984.
6. Gartner, N., C. Messer, and A. Rathi (eds). Monograph on Traffic Flow Theory. TRB, National Research Council, 2001. http://www.tft.pdx.edu/docs/revised_monograph_2001.pdf. Accessed June 16, 2011.
7. Prigogine, I., and R. Herman. *Kinetic Theory of Vehicular Traffic*. American Elsevier, 1971.
8. Mahmassani, H., J. C. Williams, and R. Herman. Investigation of Network-Level Traffic Flow Relationships: Some Simulation Results. In *Transportation Research Record 971*, TRB, National Research Council, Washington, D.C., 1984, pp. 121–130.
9. Williams, J. C., H. S. Mahmassani, and R. Herman. Analysis of Traffic Network Flow Relations and Two-Fluid Model Parameter Sensitivity. In *Transportation Research Record 1005*, TRB, National Research Council, Washington, D.C., 1985, pp. 95–106.
10. Mahmassani, H., J. C. Williams, and R. Herman. Performance of Urban Traffic Networks. In *Proc., Tenth International Symposium on Transportation and Traffic Theory: Transportation and Traffic Theory* (N. H. Gartner and N. H. M. Wilson, eds.), Cambridge Mass., Elsevier, 1987.

11. Drake J., J. Schofer, and A. D. May. A Statistical Analysis of Speed-Density Hypotheses. In *Proc., Third International Symposium on the Theory of Traffic Flow*, American Elsevier, 1967.
12. Mahmassani, H. J., R. Jayakrishnan, and R. Herman. Network Traffic Flow Theory: Microscopic Simulation Experiments on Supercomputers. In *Transportation Research Part A*, Vol. 24, No.2, 1990, pp 149–162.
13. Williams, J. C., H. S. Mahmassani, and R. Herman. Urban Traffic Network Flow Models. In *Transportation Research Record 1112*, TRB, National Research Council, Washington, D.C., 1987, pp. 78–88.
14. Daganzo, F. C. Urban Gridlock: Macroscopic Modeling and Mitigation Approaches. In *Transportation Part B*, Vol. 41, 2007, pp. 49–62.
15. Geroliminis, N., and C. F. Daganzo. Macroscopic Modeling of Traffic in Cities. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
16. Edie, L. C. Discussion of Traffic Stream Measurements and Definitions. In *Proc., Second International Symposium on the Theory of Traffic Flow*, OECD, 1965, pp. 139–154.
17. Kalafatas, G., and S. Peeta. Exact Graph Structure for Dynamic Traffic Assignment: Formulation, Properties, Computational Experience. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
18. Gazis, D. C. *Traffic Science*. Wiley, New York, 1974.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **www.TRB.org**

www.national-academies.org

TRB

TRANSPORTATION RESEARCH BOARD
500 Fifth Street, NW
Washington, DC 20001