Highway Traffic Analysis and Design

Third Edition

R. J. Salter and
N. B. Hounsell

HIGHWAY TRAFFIC ANALYSIS AND DESIGN

Also by the late R.J. Salter
Highway Design and Construction, second edition
Traffic Engineering – Worked examples, second edition

Other Macmillan titles of interest to Civil Engineers
Soil Mechanics – Principles and Practice
G.E. Barnes
Understanding Hydraulics
Les Hamill
Civil Engineering Materials, fifth edition
edited by N. Jackson and R.K. Dhir
Reinforced Concrete Design, fourth edition
W.H. Mosley and J.H. Bungey
Reinforced Concrete Design to EuroCode 2
W.H. Mosley, R. Hulse and J.H. Bungey
Civil Engineering Contract Administration and Control
I.H. Seeley
Civil Engineering Quantities, fifth edition
I.H. Seeley
Understanding Structures
Derek Seward
Surveying for Engineers, third edition
J. Uren and W.F. Price

HIGHWAY TRAFFIC ANALYSIS AND DESIGN

R.J. Salter (deceased)
Formerly Reader in Civil Engineering,
University of Bradford

Revised by

N.B. Hounsell
University of Southampton

THIRD EDITION

palgrave

First edition 1974
Revised edition 1976
Reprinted 1978, 1980 (with amendments)
First paperback edition (with further amendments) 1983
Reprinted 1985
Second editon 1989
Third edition 1996

Published by
PALGRAVE
Houndmills, Basingstoke, Hampshire RG21 6XS and
175 Fifth Avenue, New York, N. Y. 10010
Companies and representatives throughout the world

PALGRAVE is the new global academic imprint of
St. Martin's Press LLC Scholarly and Reference Division and
Palgrave Publishers Ltd (formerly Macmillan Press Ltd).

ISBN 978-0-333-60903-3 ISBN 978-1-349-13423-6 (eBook)
DOI 10.1007/978-1-349-13423-6

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources.

A catalogue record for this book is available from the British Library.

10 9 8 7 6 5 4 3
08 07 06 05 04 03 02 01

Typeset by TTP International, Sutton Surrey

Contents

Preface

More than three decades have passed since transportation and highway traffic engineering became recognised as an academic subject in the United Kingdom. During this period, significant advances have been made in traffic engineering theory and practice, against a background of ever increasing mobility and travel demand. Considerable interest has also developed in the economic, social and environmental costs of highways and other transport systems.

Richard Salter's book on *Highway Traffic Analysis and Design* was first published in 1974 and has proved an invaluable text-book for students specialising in Transportation Planning and Engineering in many Institutions of Higher Education in the United Kingdom as well as for many practitioners. The attraction of the book has been the concise and thorough way in which the many aspects of traffic analysis and design are presented, supported by numerous examples and 'question and answer' material.

Sadly, Richard Salter died in 1992. For the book to reflect recent developments in the traffic engineering field, Nick Hounsell has now revised and updated relevant sections of the book, while maintaining its qualities and basic structure.

The book is presented in three parts. Part I covers traffic surveys and prediction. This provides an introduction to the transport planning process and descriptions of data collection methods and of the four-stage modelling process of trip generation, trip distribution, modal split and traffic assignment. Recent modelling software is also described. Part I concludes with a consideration of evaluation methods for transportation proposals.

Part II of the book covers traffic analysis and design, starting with explanations of traffic flow theory and relationships between the main elements of speed, flow, density and capacity on highway links. A new chapter on geometric design of highway links is included. Part II then covers non-signalised junction design in detail, including priority junctions, roundabouts and grade-separated intersections. Computer-aided design techniques are included. Part II concludes with a consideration of the environmental effects of traffic noise and pollution and the equally topical issue of traffic congestion and restraint.

Part III of the book is devoted to traffic signal control. All issues of design and evaluation for isolated signals and urban traffic control systems are described in detail, including recent developments in computer programs for traffic signal design.

Dr Salter was particularly grateful to the many colleagues and students who made valuable comments on the contents of the first and subsequently revised editions. This edition contains many of the suggestions they made. Dr Salter also expressed his appreciation to the following bodies for permission to reproduce their copyright material: Acer Consultants, Bedfordshire County Council, the

Building Research Station, the Department of the Environment, the Eno Foundation, the Greater London Council, the Institution of Civil Engineers, the Institution of Highway Engineers, National Research Council, Royal Borough of New Windsor, Traffic Engineering and Control, the Transport Research Laboratory and Wilbur Smith and Associates. Crown copyright is reproduced with the permission of the Controller of HMSO.

Nick Hounsell would also like to express his thanks to the Department of Transport and the Transport Research Laboratory, again for permission to reproduce their copyright material. In addition, he would like to thank Professor Mike McDonald and Dr Mazen Hassounah of the University of Southampton for their contributions to this third edition.

PART I
TRAFFIC SURVEYS AND PREDICTION

1

Introduction to the transportation planning process

Large urban areas have in the past frequently suffered from transportation congestion. It has been recorded that in the first century vehicular traffic, except for chariots and official vehicles, was prohibited from entering Rome during the hours of daylight. While congestion has existed in urban areas the predominantly pedestrian mode of transport prevented the problem from becoming too serious until the new forms of individual transport of the twentieth century began to demand greater highway capacity.

Changes in transport mode frequently produce changes in land-use patterns; for example, the introduction of frequent and rapid rail services in the outer suburbs of London resulted in considerable residential development in the areas adjacent to local stations. More recently the availability of private transport has led to the growth of housing development which cannot economically be served by public transport.

In areas of older development however the time scale for urban renewal is so much slower than that which has been recently experienced for changes in the transport mode that the greatest difficulty is being experienced in accommodating the private motor car. Before the early 1950s it was generally believed that the solution to the transportation problem lay in determining highway traffic volumes and then applying a growth factor to ascertain the future traffic demands.

Many of the early transportation studies carried out in the United States during this period saw the problem as being basically one of providing sufficient highway capacity and were concerned almost exclusively with highway transport.

During the early 1950s however it was realised that there was a fundamental connection between traffic needs and land-use activity. It led to the study of the transportation requirements of differing land uses as the cause of the problem rather than the study of the existing traffic flows. The late 1950s and early 1960s saw the commencement of many land use/transportation surveys in the United Kingdom and the era of transportation planning methodology could have been said to have commenced.

Because the planning of transportation facilities is only one aspect of the overall planning process which affects the quality of life in a developed society, the provision of transport facilities is dependent on the overall economic resources available. It is dependent on the value that is placed on such factors as environmental conditions: for some transport facilities are considered to detract from the quality of the environment and others can be considered to improve the environment. Land use and transport planning are also closely connected because the demand for travel facilities is a function of human land activity and conversely the provision of transport facilities has often stimulated land-use activity.

Because we are living in a society that is changing rapidly, and in which the rate of change appears to be increasing, it is important for some attempt to be made to develop economic, environmental, land use, population and transport planning policies. The fact that planning attempts in all these fields have not met with conspicuous success in the past decade should be taken as an attempt to improve the methodology rather than an indication that short term plans based on expediency or intuition should be employed.

Transportation studies may be carried out to determine the necessity or suitability of a variety of transport systems such as inter-city air-links, a new motorway or a combination of private and public transport modes such as is found in a large urbanised conurbation. The methodology of these surveys will vary in detail – but most transportation surveys that are based on land-use activity tend to be divis-ible into three major subdivisions:

(1) The transportation survey, in which an attempt is made to take an inventory of the tripmaking pattern as it exists at the present time, together with details of the travel facilities available and the land-use activities and socio-economic factors that can be considered to influence travel.
(2) The production of mathematical models, which attempt to explain the relationship between the observed travel pattern and the travel facilities, land-use activities and socio-economic factors obtained by the transportation survey.
(3) The use of these mathematical models to predict future transportation needs and to evaluate alternative transportation plans. These three stages of the transport planning process are illustrated in figure 1.1, which shows the procedure used to estimate future travel in the Greater London Area.

In the first stage, details of the existing travel pattern together with information on land use and transport facilities are obtained for the area of the study. This area is bounded by an external cordon and so that the origins and destinations of trips within the area can be conveniently described, the study area is divided into traffic zones.

Details of the existing travel pattern are obtained by determining the origins and destinations of journeys, the mode of travel and the purpose of the journey. Most surveys obtain information on journeys that have origins in the survey area by a household interview method, which records details of the tripmaking of survey-area residents. In addition there will be some trips that have origins outside the external cordon and destinations within the cordon and others that have neither origin nor destination within the survey area but pass through the study area. Details of these trips will be obtained by interviewing tripmakers as they cross the

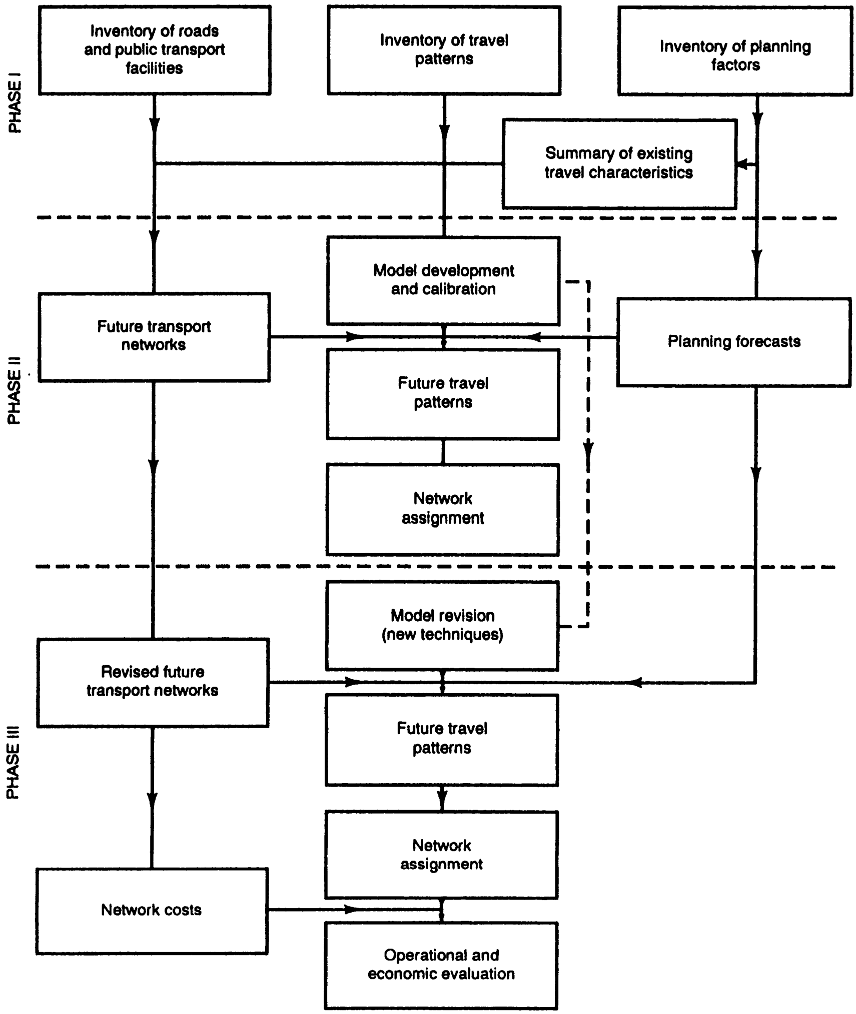


Figure 1.1 *Stages of the London Transportation Study*

cordon. Additional surveys will be necessary to obtain details of commercial vehicle trips originating in the survey area and in some circumstances trips made by means of taxis.

Information on transport facilities will include details of public transport journey times, the frequency of service, walking and waiting times. For the road network, details of traffic flows, journey speeds, the commercial vehicle content and vehicle occupancies are frequently necessary.

As land-use activity is the generator of tripmaking, details of land-use activity are required for each traffic zone. For industrial and commercial land use, floor space and employment statistics are necessary, while for residential areas, the density of development is frequently considered to be of considerable importance. At the same time, socio-economic details of the residents are obtained since many surveys have indicated a connection between tripmaking and income, and also between tripmaking and social status.

In the second stage of the transportation process, models are developed which attempt to explain the connection between the tripmaking pattern and land-use activity. This is the most difficult aspect of the transportation planning process in that failure adequately to determine the factors involved in tripmaking will result in incorrect predictions of future transport needs.

While the decision to make a trip is a complex process based on the availability of destinations, the travel facilities, the cost of travel and the journey purpose, it is usual to divide model building into the following interconnected processes which are described in detail in the subsequent sections:

(1) Trip generation, which attempts to determine the connection between tripmaking and land-use factors noted in the planning inventory.
(2) Trip distribution, which determines the pattern of trips between the zones. It is usually postulated that the number of trips between zones is proportional to the size of the zones of origin and destination and inversely proportional to some measure of their spatial separation.
(3) Traffic assignment, which decides on which links of either the public transport or the highway network a trip will be made. On the basis of the travel cost, the decision is made on the alternative routes between the origin and the destination. Trips are usually assigned to the route of least cost, frequently measured by travel time. More recently a proportion of trips has been assigned to the alternative links between an origin and a destination in an attempt to produce a realistic simulation of the real life situation.
(4) Modal split, by which a decision is made as to which travel mode a tripmaker will use. This decision may be made at the trip generation stage or frequently after trip distribution. The observed modal split relationship may be explained in terms of car ownership, relative travel costs between the alternative modes or on the basis of restrained private vehicle use.

After these interrelated processes have been set up, the existing land-use characteristics are inserted into the trip generation procedure. The trips are distributed, assigned and a modal split decision at the appropriate stage of the process carried out. This allows checks to be made on the accuracy of the whole process because the flows assigned to the present day network may be compared with those actually observed on the network.

With confidence in the ability of the developed models established, it is then possible to forecast the travel needs of future land use and transport plans. The evaluation of the effects of these alternative proposals is the object of the third stage of the transportation planning process. Arriving at an optimum solution is however an intuitive process because the planning process can only predict the likely tripmaking which will arise if the plan is implemented. Alternative plans may be evaluated, on a limited basis, by the estimation of the costs and benefits which arise if the plan is carried out.

Problem

During the period 1950–60 an important reappraisal of thinking regarding the estimation of future transportation requirements took place. Select the correct description of this reappraisal from the following and amplify the statement:

(a) It was appreciated that the highway network should be designed to accommodate at least three times as many motor vehicles as are observed on the existing network.
(b) It was appreciated that there was a basic connection between land-use activity and transportation requirements.
(c) It was appreciated that the existing highway network in the centres of large urban areas would require extensive modification.

Solution

(a) While a considerable increase in the number of motor vehicles in use can be expected in the future it is not considered desirable simply to scale up the existing highway network. The demand for transport facilities will increase at differing rates in differing situations and not all of these demands are likely to be met by the provision of enlarged highway facilities.
(b) The realisation that the demand for transport facilities was a function of land use caused an important reappraisal in the estimation of future transport needs. Instead of observing the end product and applying some growth factor to estimate future values it became possible to obtain the traffic requirements of differing land-use plans. It changed transport planning from a vehicle count to a land-use-based operation.
(c) While it was appreciated that the modification of the highway network in the centres of large urban areas would be required in the future it is considered desirable to restrain the use of motor cars in many central areas on environmental grounds.

2

The transportation study area

The area of a country covered by a transportation study will depend upon the geographical distribution of the trip patterns which are of particular interest to the study organisers. The study area has been defined as the area within which trip patterns will be significantly affected by the implementation of any resulting transport proposals. It is important that sufficient thought be given to the boundary of the area because within the boundary the transport information is collected in considerable detail and the resulting cost of data collection will be largely related to the location of the boundary.

Some surveys are concerned with individual towns of small size and in these cases the area may cover only the developed land together with land which may become developed during the economic lifetime of the transport proposals which may result from the study.

The boundary of the study area within which trip information is collected and modelled in considerable detail is termed the external cordon. It is not possible to confine transportation surveys to existing local government boundaries because tripmaking and transportation needs are common to a region. Most surveys are carried out on a regional basis.

The survey area within which travel is to be studied in detail is bounded by an external cordon. The position of the external cordon is fixed so that as far as possible all developed areas which influence travel patterns are included together with areas which are likely to be developed within the forecasting period of the study. Within the external cordon, trip information is collected in considerable detail and so the cost of the survey is largely related to the cordon location.

To permit the aggregation of data, the survey area is divided into internal traffic zones; the remainder of the country external to the cordon is also divided into considerably larger external zones. The Department of the Environment has divided the country into Standard Regions and this is useful for the external traffic zones.

Zones are selected after consideration of the following factors:

(1) They should be compatible with past or projected studies of the region or adjacent region.
(2) They should permit the summarising of land use and tripmaking data.

(3) They should allow the convenient assignment of trips, which are assumed to be generated at the centroids of traffic zones, to the transportation network. There is often conflict here between zones which are suitable for the highway network and zones suitable for the public transport network.
(4) While greater accuracy is possible with smaller traffic zones they should not be so numerous as to make subsequent data processing difficult.
(5) The size of the zones will be governed by the detail which is required in the modelling process and in the resulting transport proposals.
(6) Because of the increasing use of data from the Registrar General's Census of Population the zones should be selected after consideration of enumeration, district boundaries and the grouping of local planning authority data.
(7) As data is frequently aggregated by zone, the latter should be areas of predominantly similar land use activity in which similar rates of future growth are anticipated.

When the traffic zones are being determined, it is important to plan them with thought for the provision of one or more screenlines. A screenline is a line which divides the area within the external cordon and it is chosen so that the trips crossing it at the present day can be easily measured. For this reason a natural barrier to communication should, if possible, be used so as to make it possible to count the traffic movements across the screenline. It should be so positioned that it is unusual for trips to cross the screenline more than once. To avoid the complex traffic movements which take place in the central sector a screenline is frequently placed to intercept all the movements into the central area.

An illustration of a hierarchy of traffic zones is given in figures 2.1, 2.2, 2.3 and 2.4 showing the arrangements used in the Bedford–Kempston transportation Survey [1]. Figure 2.1 shows the internal zones varying in area from the small intensively devel-

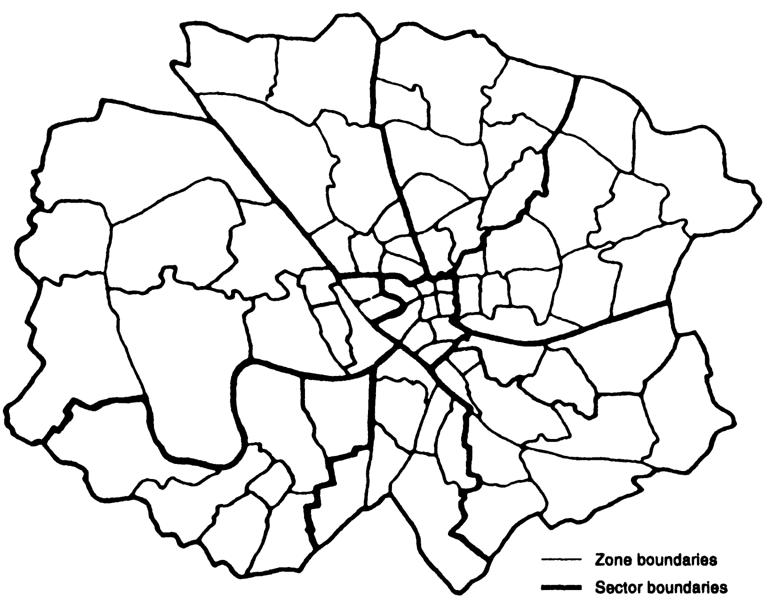


Figure 2.1 *Internal traffic zones*

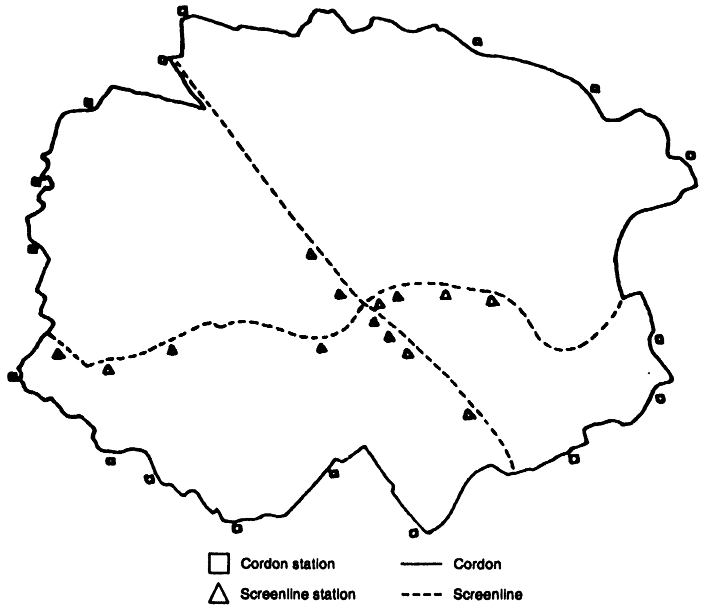


Figure 2.2 *External cordon and screenlines*

Figure 2.3 *External traffic zones (intermediate)*

Figure 2.4 *External traffic zones (national)*

oped Central Area zones to the larger zones in the outer area of Bedford. These internal zones are grouped into a central area sector and eight radial sectors. The sector numbers (0–8 inclusive) are the first digit in the internal zone code numbers.

Within the external cordon, two screenlines were set up for checking purposes and this is shown in figure 2.2. Also shown are the screenline counting stations and also the external cordon stations.

Outside the external cordon but immediately adjacent to the survey area the external zones are comparatively small, generally corresponding to local authority areas or combinations of areas. These zones are shown as intermediate external zones in figure 2.3. The remainder of the country is divided into national external zones as shown in figure 2.4. All the external zones are numbered in accordance with the Department of the Environment standard classification.

Reference

1. Bedfordshire County Council, *Bedford–Kempston Transportation Survey, Phase 1* (1965)

3

The collection of existing travel data

Existing tripmaking or travel data may conveniently be classified into four groups according to the origin and the destination of the trip being considered. These four classes of movement are:

(a) trips which have an origin within the external cordon and a destination outside the external cordon;
(b) trips which have both an origin and a destination within the external cordon;
(c) trips which have an origin outside the external cordon and a destination within it;
(d) trips which have neither origin nor destination within the external cordon but which pass through it.

Details of these trips are obtained in differing ways. Trips of type (a) and (b) are usually recorded by means of a home interview survey while trips of type (c) and (d) are normally recorded by means of an origin–destination survey conducted along the external cordon.

The method of obtaining survey information is the survey instrument. There are three main groups: (i) *observational* techniques, such as cordon/screenlines counts and car-parking surveys; (ii) *written* techniques, such as travel diary surveys and mail-back household questionnaires; and (iii) *oral* techniques such as personal and telephone interview travel surveys. The most appropriate technique for any application depends on the survey objectives, the population of interest (that is the complete group about which data is sought), the data sought, survey cost, the desired response rate and response quality. Oral surveys, for example, are relatively more expensive to conduct than written surveys and usually yield a lower response rate, but they produce a higher quality of response.

For travel surveys it is necessary to select a sample to represent the population of interest. A sample is a collection of non-overlapping units (for example an individual, a household or a vehicle movement). Required sample size then depends on the variability of response and the degree of precision required. In practice, survey cost is often a constraint. Either simple random sampling can be used, where the sample is obtained by selecting survey units at random from the whole population, or stratified random sampling can be adopted. With this method, the population is first divided into homogeneous groups and survey units are then selected at random from each stratum, using the same sampling rate or different sampling rates. All surveys are best preceded by a pilot survey on a small scale to test the survey methods and sampling procedures.

Before a survey is conducted, the length of time required to complete it should be estimated along with the number of people required to administer the survey. Survey personnel should be introduced by means of training sessions. During the actual conduct of the survey, survey personnel should be closely supervised and data should be checked while being collected. The collected information should be entered into a computer file using a database management system. The whole process along with the survey results and computer file layout should be well documented.

Survey data may contain errors and/or bias. In general, there are two types of survey bias: (i) random sampling errors due to a sample being used to estimate the population, which are influenced by sample size and method of sampling; and (ii) systematic errors, due to factors such as incorrect sampling, insufficient control of survey procedures and excessive non-response. These issues will now be examined in the context of home-interview surveys.

The home is the major source of trip generation and details of a considerable proportion of trips generated in an urban area may be obtained by this form of survey. A home-interview survey usually involves leaving a diary or similar document for the respondent to complete at a later stage.

One way in which an unbiased sample may be obtained is by sampling from the electoral roll or from the valuation list, both of which are held by local authorities. The Department of Transport in their *Traffic Appraisal Manual*[1] advise that the first step in selecting a sample is to define the 'population' of interest, the households in the area being surveyed. Normally all members of the household are of interest in the survey but it might be car drivers, adults, public transport users or some other category. Having decided upon the population of interest an approximation to the population can be obtained from the Electoral Register or the Valuation List.

Sample size must depend upon the errors in the data collection process and in the subsequent trip prediction process. Where approximate estimates of travel requirements are necessary, a smaller sample may be tolerated.

Population size	Sample size (%)
under 50 000	20
50 000–150 000	12.5
150 000–300 000	10
300 000–500 000	6.66
500 000–1 000 000	5
over 1 000 000	

The *Traffic Appraisal Manual*[1] gives the following equation for the calculation of sample size, n

$$n = \frac{P(1 - P)N^3}{\left(\frac{E}{1.96}\right)^2 (N - 1) + P(1 - P)N^2}$$

where N is the total number of households within the survey area,
E is the required accuracy expressed as a number of households,
P is the proportion of households with the attribute of interest.

For example in a household-interview survey it is estimated that there are approximately 35 500 households and it is also estimated that not more than 1 in 20 households will, on the survey date, make the particular journey of which the characteristics are to be modelled. If the required accuracy is 2 per cent of the population calculate the required sample size.

From the above equation

$$n = \frac{0.05 \times 0.95 \times 35500^3}{\left(\frac{710}{1.96}\right)^2 (35499) + 0.05 \times 0.95 \times 35500^2}$$

$$= 450$$

Care should be taken to obtain an unbiased sample and if a completed questionnaire cannot be obtained or if it is impossible to deliver a questionnaire then these questionnaires should be omitted from the survey and not delivered at adjacent homes. The trip information obtained must be increased to give all the trips in a zone by the use of a sampling factor

$$\text{sampling factor} = \frac{A - (C \times A/B)}{B - (C + D)}$$

where A = total households in zone,
B = number of households selected for sampling in zone,
C = number of households vacant or demolished,
D = number of households refusing to co-operate.

As it is desired to obtain relationships between tripmaking and such variables as land-use characteristics and socio-economic factors, many details other than those relating to the trips must be obtained. The cost of the survey is considerable. Hence great care must be taken in the preparation of these questionnaires both from the point of view of obtaining the required information and also of coding and analysis.

There are alternative approaches to the collection of this data. An interviewer may attempt to contact every member of a selected household individually and details of tripmaking and socio-economic characteristics of the tripmaker are recorded in the presence of the interviewer. Alternatively the interviewer may only collect details of the household characteristics, leaving the questionnaire on tripmaking to be completed at a future date. The interviewer will then collect the questionnaires in the future after checking them for completeness. While the first method should result in accurate detailing of tripmaking, it is expensive and there is a danger of interviewer prompting. The second method is less expensive in interviewer costs but will tend to be less accurate. Figure 3.1 shows a typical questionnaire used to record the characteristics both of the dwelling unit and the residents while details of the trips they make are recorded on the type of form shown in figure 3.2.

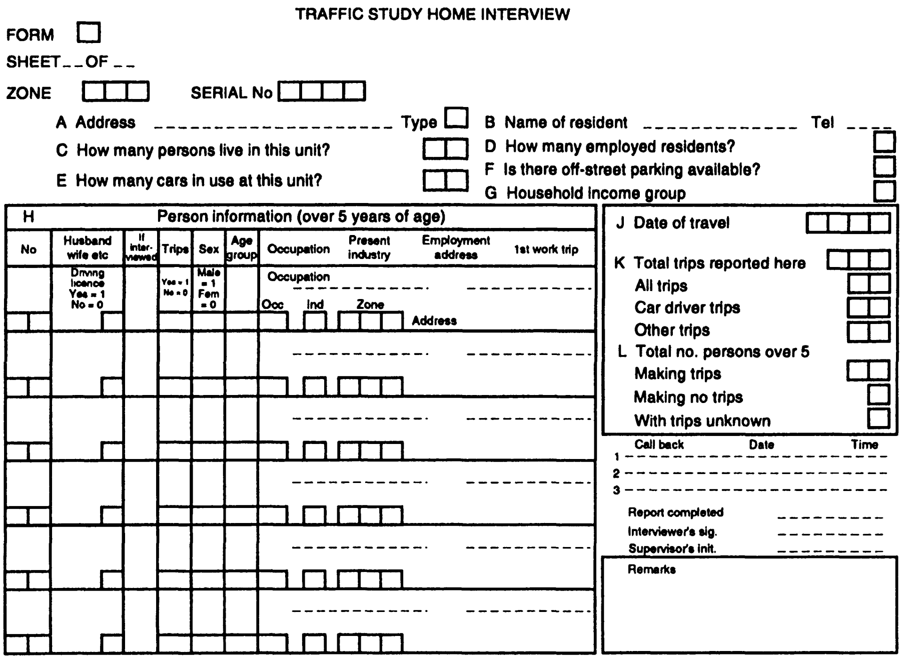
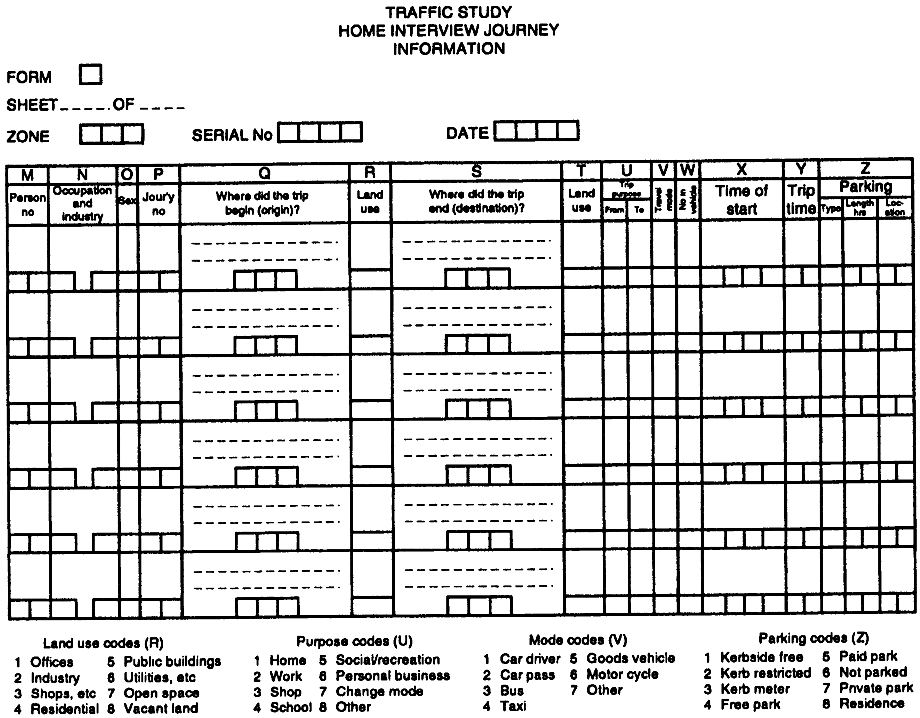


Figure 3.1 *Home interview survey form*

Advice on the successful conduct of home-interview surveys is given by the Department of Transport in the *Traffic Appraisal Manual*. It is stressed that as many home-interview surveys require the householder to recall past details of tripmaking it is essential that the interviewer ensures that trip details are accurately remembered by the respondent. A successful interviewer is also required to make sure that the respondent to the travel survey understands his part in the transportation survey. For any survey to be successful it is necessary for the respondents who supply the trip information to have a strong motivation to answer the questions correctly; hence the interviewer should aim to increase the respondent's motivation.

Cross-sectional and longitudinal surveys

Most transport planning studies have relied on information about a cross-section of population at a single point in time, from which models to predict future travel choices are calibrated. Such models rely on the sample size being sufficient, incor-



TRAFFIC STUDY
HOME INTERVIEW JOURNEY
INFORMATION

FORM

SHEET ____ OF ____

ZONE SERIAL No DATE

M	N	O	P	Q	R	S	T	U		V	W	X	Y	Z		
Person no	Occupation and industry	Sex	Joury no	Where did the trip begin (origin)?	Land use	Where did the trip end (destination)?	Land use	Trip purpose		Travel mode	No in vehicle	Time of start	Trip time	Parking		
								From	To					Type	Length hrs	Location

Land use codes (R)		Purpose codes (U)		Mode codes (V)		Parking codes (Z)	
1 Offices	5 Public buildings	1 Home	5 Social/recreation	1 Car driver	5 Goods vehicle	1 Kerbside free	5 Paid park
2 Industry	6 Utilities, etc	2 Work	6 Personal business	2 Car pass	6 Motor cycle	2 Kerb restricted	6 Not parked
3 Shops, etc	7 Open space	3 Shop	7 Change mode	3 Bus	7 Other	3 Kerb meter	7 Pnvate park
4 Residential	8 Vacant land	4 School	8 Other	4 Taxi		4 Free park	8 Residence

Figure 3.2 *Home interview journey information*

porating enough people changing their choices so that there is a large variation within the data. An alternative approach is to establish longitudinal or time-series data to develop more reliable transport forecasting models.

Longitudinal data is sometimes collected through repeated cross-sectional surveys covering measurements on samples from a population made without ensuring that any respondent is included in more than one round of data collection. Alternatively a panel survey can be used in which similar measurements are on the same population sample at different points in time. Panel surveys, although theoretically attractive, can be expensive to conduct and usually suffer from non-response problems.

Revealed and stated preferences

Transport models are often based on data about actual or observed choices made by survey units, that is revealed preference data. However, choice is not actually observed or measured with this approach, as the data reflects what people report that they have done. Further limitations are the difficulty in distinguishing between the effects of different attributes on choice and the difficulty in predicting the effect on a new option on choice.

As alternatives to revealed-preference surveys, stated-preference surveys were proposed in the late 1970s. In a stated-preference or attitude survey, individuals are asked directly about what they would do in a hypothetical situation. A key issue with this approach is data reliability (that is the extent to which individuals actually do what they state they would when the case arises).

Reference

1. Department of Transport, *Traffic Appraisal Manual*, London (1982)

4

The external cordon and screenline surveys

The object of the external cordon survey is to obtain information on the trips originating outside the external cordon having origins within the external cordon or passing through the survey area.

For highway trips whether by public or private transport this information is obtained by an origin–destination survey carried out at census points on the road side. For non-highway public transport trips a survey is carried out while the trips are being made or, alternatively, an examination of ticket records may give some of the required origin–destination information.

Origin–destination surveys at the external cordon may be carried out by interviewing drivers at census points. Alternatively, in certain limited circumstances, the survey may be accomplished by using questionnaire postcards handed to drivers at the census points for later completion and return by post.

The first method, that of direct interview of drivers, is probably one of the most widely used techniques and is commonly used not only to obtain details of trips which cross the external cordon in transportation studies but also for estimating the journeys which would be made on proposed highways.

Direct interviews at roadside census points are often carried out simultaneously at all the census points but a reduction in the number of staff required at any one time, but not in total, and also in the shelters, barriers and warning signs necessary is possible by dealing with different census points and directions of traffic movement at separate times.

Questions asked at the census points will, of course, vary according to the survey, but they should be carefully phrased and without ambiguity. To avoid bias in the results obtained, the questions may sometimes be printed on a card and shown to a driver but it is more usual for the interviewers to be given instructions to ask the exact question required by the traffic engineer in charge of the survey.

A form is usually completed for each interview, the enumerator noting the time, the census point and the direction being sampled, the class of vehicles, the origin and destination of the journey and any intermediate stops. If the economic benefits of the proposed highway are to be assessed then additional questions may be asked to determine the purpose of the journey and the number of occupants of the vehicles. A typical interview sheet used for recording details of vehicle trips is given in figure 4.1.

TRAFFIC STUDY
CORDON STUDY ROADSIDE
INTERVIEW FORM

FORM ☐

INTERVIEWER ----------

ZONE ☐☐☐ CORDON POINT ☐☐ DIRECTION INBOUND OUTBOUND DATE ☐☐☐☐ HOUR BEGINNING ☐☐☐☐

A	B	C	D	E	F	G	H		J	K	L		M
Serial no	Vehicle type	Persons in vehicle	Where did this trip begin? (origin)	Land use at (origin)	Where did this trip end? (destination)	Land use at destination	Trip purpose From	To	Occupation of driver	Industry of driver	Did you stop in X?	Purpose	Are you resident in X?
			Address Town ☐☐☐ County		Address Town ☐☐☐ County						Yes = 1 No = 0		Yes = 1 No = 0

Vehicle type codes (B)
1 Passenger car
2 Light goods
3 Heavy goods
4 Bus or coach
5 Motor cycle

Land-use codes (E and G)
1 Offices
2 Industry
3 Shops, etc
4 Residential
5 Public buildings
6 Utilities, etc
7 Open space
8 Vacant land

Purpose, cars (H)
1 Home
2 Work
3 Shop
4 School
5 Social/recreation
6 Personal business
7 Change mode
8 Other

Purpose (L)
1 Eat
2 Tourist
3 Shop
4 Business
5 Other

Purpose, trucks (H)
1 Pick up
2 Deliver
3 Pick up and deliver
4 Depot
5 Service
6 Personal business
7 Other

Figure 4.1 *Cordon survey interview form*

Under nominal traffic conditions it is of course not possible to interview the driver of each vehicle without causing undue delay or employing an excessive number of interviewers. For this reason only a proportion of the vehicles travelling past a census point are stopped for interview, a process known as sampling.

The selection of the vehicles for the sample must be free from bias and the sampling technique developed by the Transport and Road Research Laboratory has been widely used[1].

In this method a police officer, upon receiving a signal from either of the two interviewers that he is ready to commence the next interview, directs the first vehicle that arrives at the census point into the interview lane. All other traffic is allowed to pass the census point without being halted, a volume count of the number of vehicles passing the point being carried out by the enumerator.

If the first vehicle to arrive at the census point after an interviewer has indicated that he is ready to carry out a further interview is halted, then the sampling technique is free from bias. It has been found, however, that on derestricted lengths of highway the first vehicle may be travelling at a high speed and the police officer directing the traffic may halt a slower vehicle which is following. Should this tendency not be eliminated by slowing down the whole of the traffic stream then the sample is biased in favour of slower vehicles.

As no attempt is made to ensure that the number of vehicles halted is a constant proportion of the total number of vehicles passing the census point, a sampling factor is introduced for each class of vehicle which varies according to the hour. The sampling factor for a certain hour is simply the number of vehicles of a particular class passing the census point during that hour divided by the number of vehicles of the particular class included in the sample.

The growth of traffic that has occurred since the publication of reference 1 has led to the issue of detailed instructions for the conduct of traffic surveys by roadside interview[2].

General instructions are given on the location of census points with particular emphasis on minimising accident risk, carriageway obstruction and delay to vehicles. It is obviously important to site the census point away from junctions, on lengths of highway without excessive gradient and where the carriageway is as wide as possible. For single carriageway roads it is desirable to incorporate a bypass lane otherwise it will be necessary to operate stop/go or shuttle working. There are fewer difficulties when a census point is located on a dual carriageway road provided speeds and flows are not high. In the British Isles interviewing on motorways has not been allowed and only very occasionally has it been allowed on slip roads.

The information requested at roadside interview surveys carried out in accordance with these recommendations is standardised. The data collected from the drivers sampled comprises: the full postal address of the next stop of the vehicles; the reason for the journey to the next destination which is classified into the categories of home, holiday home, work, employer's business, education, shopping, personal business, visiting friends and recreation/leisure. Details are also required of the last stop of the vehicle and the reason for being at that address; the same categories as are used for destinations are employed. Car, motorcycle and van drivers are asked the full postal address at which the vehicle was garaged during the previous night.

In addition to interviewers it is also necessary for enumerators to undertake classified counts of vehicles passing through the census station during half-hour periods.

Registration number surveys

If it is desired to know the points at which vehicles enter and leave a survey area but it is not desirable to stop vehicles then registration number surveys are carried out. In these surveys the class of each vehicle and its registration number are noted as it passes each survey point together with the time the vehicle was observed. Subsequent comparison of the records will indicate entry and exit points into and from the survey area together with travel times for different classes of vehicles.

Records can be obtained by manual recording and the *Traffic Appraisal Manual* recommends that one observer be stationed at each station and note as many registration numbers as possible. If because of the traffic flow or for any other reason a number is missed, the observer places a tick on the recording sheet. In this way a record of the traffic flow is obtained. To facilitate recording it is recommended that only a portion of the registration number be noted (that is, for A405NJX record A405, for DNW405Y record 405Y). Vehicle arrival and departure times should be noted to the nearest minute and a continuous check be kept on the time. Classification of vehicles can be made by the use of an additional code letter.

It is also recommended that control cars be used to control the quality of the information; they pass through the survey area noting any unusual conditions and the exact time when they pass the survey stations. In this way the differing observers are connected together both by time and by the control cars.

A problem associated with partial registration number surveys is the possibility that matches may occur between different vehicles, causing spurious matches. This is particularly a problem when flows are heavy but few vehicles actually travel between entry and exit points of the survey area. An approximate formula for the number of spurious matches is given in the *Traffic Appraisal Manual*[3].

When traffic flows are heavy, manual recording becomes impracticable and use should be made of portable tape recorders. They are left permanently on and registration numbers together with vehicle class recorded by an observer. If a registration number is missed the passage of vehicle should be marked by the use of some word such as 'blank' so that a count of vehicles passing the point can be made. Frequent time references should be incorporated in the tape.

Using postcards

The use of prepaid postcards which are handed to drivers as they pass the census points is an inexpensive method of obtaining journey information. Delay to vehicles is small and a considerable amount of information may be obtained from the questions on the postcard[2]. A great deal of advance publicity however is required if a reasonable proportion of postcards is to be returned, and often the number of cards returned has been disappointing. There is also the danger that certain types of vehicle drivers are more likely to complete and return the postcards than others, so giving a bias to the results obtained.

Details of the trips crossing the screenlines within the external cordon are necessary to allow a check to be made on the information obtained by the home-interview, external cordon and goods movement surveys. The total existing trip information is assigned to the present day or base year transport networks and if a discrepancy is found then the survey information is adjusted. Screenline checks are also useful to compare sector-to-sector movements as a check on the trip generation, attraction and distribution models developed from the present day pattern of movement.

References

1. Transport Road Research Laboratory, *Research on Road Traffic*, Chapter 4, HMSO, London (1965)
2. Department of Transport, *Traffic Surveys by Roadside Interview*, Advice Note TA/11/81, London (1981)
3. Department of Transport, *Traffic Appraisal Manual*, London (1982)

5

Other surveys

Commercial vehicle surveys

This survey is designed to measure the trips made by commercial vehicles within the internal area. The data is obtained by issuing drivers with forms on which they record each trip, together with trip origin, destination, purpose and parking details. The size of the sample depends on the size of the survey area and also on the variability of goods-use activity, ranging from 100 per cent for a small town to about 25 per cent for a large area. The population may be obtained from the land-use survey, which will indicate addresses at which commercial vehicles may be kept or alternatively excise licence records may be consulted. A typical questionnaire is reproduced in figure 5.1.

Bus passenger surveys

Bus trips made by residents are obtained from the home-interview survey; trips made by public transport by non-residents are obtained by a bus passenger survey carried out at the external cordon. Duplication of trips made by residents within the survey area should be checked.

Questionnaires are usually distributed to bus passengers as the bus enters the survey area at the external cordon. The completed questionnaires are collected from the bus or passengers are requested to return the completed forms by post. The questionnaire used in the Leicester Traffic Survey[1] is given in figure 5.2 as an illustration.

In the larger urban areas there will also be a considerable number of non-home-based trips within the central area by tripmakers who are not resident within the external cordon. As these trips will not be recorded by the home-interview survey it is necessary to either conduct additional surveys within the external cordon or interview a sample of tripmakers leaving the survey area.

TRAFFIC STUDY
COMMERCIAL VEHICLE ORIGIN
AND DESTINATION SURVEY

FORM ☐ SHEET ______ OF ___
SERIAL No ☐☐☐☐ INTERVIEWER ____________
ZONE ☐☐☐ DATE ☐☐☐☐ TOTAL No OF TRIPS ☐☐ ESTIMATED DAYS MILEAGE ☐☐☐
OWNER NAME ______________ BUSINESS ADDRESS ______________ TELEPHONE _______
INDUSTRY OR BUSINESS ____________ ☐☐ REG No ______ VEH TYPE ☐

A	B	C	D	E	F		G		H	J	K
Trip No.	Where did this trip begin (origin)?	Land use	Where did this trip end (destination)?	Land use	Trip purpose		Trip times		Weight of cargo		Number of stops
					From	To	Type	Length in minutes			
☐☐	Address Town ☐☐☐ County	☐	Address Town ☐☐☐ County	☐	☐	☐	☐☐☐	☐☐☐	☐	☐	☐☐
☐☐	Address Town ☐☐☐ County	☐	Address Town ☐☐☐ County	☐	☐	☐	☐☐☐	☐☐☐	☐	☐	☐☐
☐☐	☐☐☐	☐	☐☐☐	☐	☐	☐	☐☐☐	☐☐☐	☐	☐	☐☐
☐☐	☐☐☐	☐	☐☐☐	☐	☐	☐	☐☐☐	☐☐☐	☐	☐	☐☐
☐☐	☐☐☐	☐	☐☐☐	☐	☐	☐	☐☐☐	☐☐☐	☐	☐	☐☐
☐☐	☐☐☐	☐	☐☐☐	☐	☐	☐	☐☐☐	☐☐☐	☐	☐	☐☐

Vehicle type codes
2 Light goods
3 Heavy goods
4 Bus or coach
6 Taxi

Land-use codes (C and E)
1 Offices
2 Industry
3 Shops, etc
4 Residential
5 Public buildings
6 Utilities
7 Open space
8 Vacant land

Purpose codes (F)
1 Pick up
2 Deliver
3 Pick up and deliver
4 Depot
5 Service
6 Personal business
7 Other

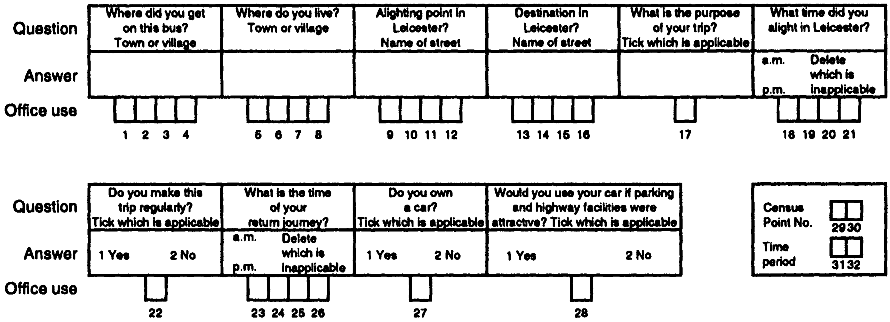
Cargo codes (H)
1 Empty
2 $\frac{1}{4}$ full
3 $\frac{1}{2}$ full
4 $\frac{3}{4}$ full
5 Full
6 Unknown

Figure 5.1 *Commercial vehicle survey form*

Train passenger surveys

Questionnaires are frequently distributed to train passengers as they board trains bound for the survey area. Questionnaires are collected on the train or passengers are requested to return them by post. When postal return is used the information obtained should be checked wherever possible when the return is less than 70 per cent. The information requested on a train passenger survey will vary according to the survey type. It will include questions requesting information on:

(a) the address of the traveller;
(b) the mode of travel from the home to the station;
(c) the scheduled departure time of the train;
(d) the mode of travel from the destination to the workplace;
(e) the address of the destination;
(f) the journey purpose;
(g) details of vehicle ownership.



Question	Where did you get on this bus? Town or village	Where do you live? Town or village	Alighting point in Leicester? Name of street	Destination in Leicester? Name of street	What is the purpose of your trip? Tick which is applicable	What time did you alight in Leicester?
Answer						a.m. p.m. Delete which is inapplicable
Office use	1 2 3 4	5 6 7 8	9 10 11 12	13 14 15 16	17	18 19 20 21

Question	Do you make this trip regularly? Tick which is applicable	What is the time of your return journey?	Do you own a car? Tick which is applicable	Would you use your car if parking and highway facilities were attractive? Tick which is applicable
Answer	1 Yes 2 No	a.m. p.m. Delete which is inapplicable	1 Yes 2 No	1 Yes 2 No
Office use	22	23 24 25 26	27	28

Census Point No.	29 30
Time period	31 32

Figure 5.2 *A typical bus passenger questionnaire*

Taxi surveys

If required a survey is carried out as for a commercial vehicle survey. Normally taxi trips form only a small proportion of tripmaking in most cities and a taxi survey is not normally necessary unless the number of trips by this mode of transport is considerable.

Reference

1. J. M. Harwood and V. Miller, *Urban Traffic Planning*, Printerhall, London (1964)

Problems

Select the correct answer to the following questions.

(1) The commercial vehicle survey obtains:
- (a) details of all trips made by commercial vehicles based within the survey area;
- (b) details of commercial vehicle trips which cross the external cordon;
- (c) an inventory of all commercial vehicles garaged within the external cordon.

(2) Bus and train passenger surveys obtain:
- (a) details of trips made by residents who live within the external cordon;
- (b) details of trips made into and through the survey area by tripmakers resident outside the external cordon;
- (c) details of journey times by public transport within the survey area.

Solutions

The correct answers to the above questions are:

(1) The commercial vehicle survey obtains:
   (a) details of all trips made by commercial vehicles based within the survey area.

(2) Bus and train passenger surveys obtain:
   (b) details of trips made into and through the survey area by tripmakers resident outside the external cordon.

6

Trip generation

Once the transportation survey has collected all the details of the existing trip-making pattern and the socio-economic, land-use and transportation-system characteristics of the survey area, the second stage in the transportation planning process is the development of relationships between the total number of trip origins and destinations in a zone and the total characteristics. It is assumed that these relationships will be true in the future and so, if land-use and socio-economic factors can be predicted, future trips can be estimated for any proposed transport system.

At this stage some basic definitions have to be made. A trip is a one-way person movement by one or more modes of travel and each trip will have an origin

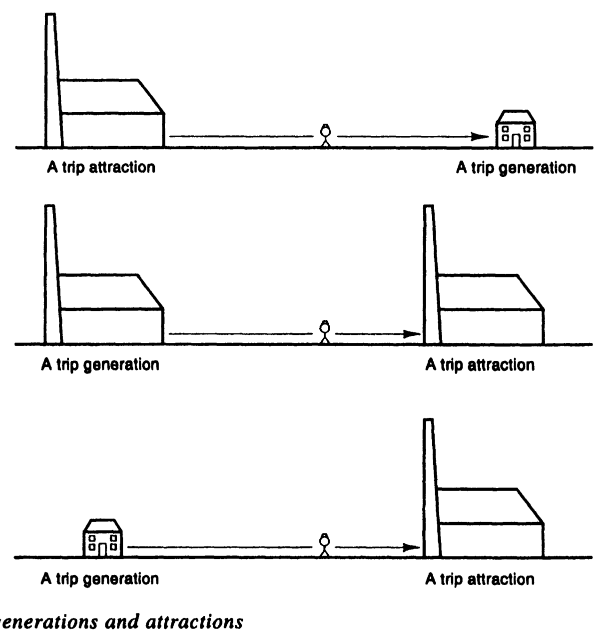


Figure 6.1 *Trip generations and attractions*

and a destination. Most surveys divide trips into home-based and non-home-based trips. All trips which have one end at the home are said to be generated by the home and the other end of the trip is said to be attracted to the zone in which it commences or terminates. For non-home-based trips the zone of origin is said to generate the trip and the zone of destination is said to attract the trip. Figure 6.1 illustrates the differences between generation and attraction.

The form of the relationship connecting trip generation and land use and socio-economic factors depends upon the basic framework of the transportation process. If the study is designed to use trip end modal-split models, which allocate trips to differing modes of transport before the distribution of trips between the traffic zones, then trip generation models are designed to predict person movement by differing modes of travel. On the other hand when trips are distributed between the zones before the decision is made regarding the mode of travel, then a trip interchange model is required, which predicts person movement in terms of total person movement by all modes of travel.

Because trips made for different purposes have different distribution and modal split characteristics, trip generation equations are generally stratified into trips for different purposes whether trip end or trip interchange models are being developed.

As the home is regarded as being the generator of the major proportion of trips, a considerable number of trips have the home end of the trip as part of the description; that is

home to work trips
home to recreation trips

and where the land use at the destination does not fall into any of the stated categories then the description of

home to other trips

is used.

Trip generation equations have as their dependent variable the number of trips generated per person or per household for different trip purposes, while the independent variables are the land-use and socio-economic factors that are considered to affect tripmaking.

Land use is of course a major consideration in the generation of trips for in residential areas the density of development and the type of housing can be expected to be major factors in the number of trips likely to be generated. Similarly in areas where commercial or industrial land use is of importance then the type of activity and the number of workers employed per unit area will influence trip generation.

Other factors connected with the home, that are considered to influence the trip generation rate, are family size, income, motor vehicle ownership and the social status of the head of the household being considered.

These relationships between trip ends, land-use and socio-economic factors may be obtained in two ways:

(1) by zonal least squares regression analysis;
(2) by category analysis.

Zonal least squares regression analysis

Past transportation studies have made extensive use of zonal least squares regression. For each of a number of zones a certain number of trip ends – the dependent variable – are observed and each zone has certain measurable characteristics to which this trip generation rate may be related. These characteristics X_1, X_2 etc. are referred to as the independent variables and are the land-use and socio-economic factors which have been previously referred to.

The equation obtained by least squares analysis is of the general form

$$Y = b_0 + b_1 \times X_1 + b_2 \times X_2 + \ldots + b_n \times X_n$$

where b_0 is the intercept term or constant.

$b_0, b_1 \ldots b_n$ are obtained by regression analysis,
$X_1, X_2 \ldots X_n$ are the independent variables.

In developing regression equations it is assumed that:

(1) All the independent variables are independent of each other.
(2) All the independent variables are normally distributed; if the variable has a skew distribution often a log transformation is used.
(3) The independent variables are continuous.

It is not usually possible for the transportation planner to conform to these requirements and regression analysis has been subjected to considerable criticism. Nevertheless regression analysis is a powerful tool when used in conjunction with a computer for handling considerable volumes of data.

It is important however to recognise that the regression process contains the likelihood of the future values of the dependent variable Y being in error when future values of the independent variables X_1, X_2, etc. are substituted into the equation.

The likely sources of error may be stated to be:

(a) Errors in the determination of the existing values of the independent variables owing to inaccuracy or bias in the transportation survey.
(b) Errors in the determination of the existing values of the dependent variables, also as a result of inaccuracy or bias in the transportation survey. This may be detected and corrected by adequate screenline checks.
(c) The assumption that the regression of the dependent variable on the independent variables is linear, a matter of some importance when future values of the independent variables are outside the range of observed values.
(d) Errors in the regression obtained owing to the scatter of the individual values and the inadequacy of the data.
(e) Difficulties in the prediction of future values of the independent variables, for the future value of the dependent variable will only be as good as the future estimates of the independent variables.
(f) Future values of the independent variable will be scattered as are the present values.

(g) The true regression equation may vary with time because factors that exert an influence on tripmaking in the future are not included in the present-day regression equation.

Most computer programs introduce or delete the dependent variables in a stepwise manner. Only variables that have a significant effect on the prediction of the dependent variables are included in the regression analysis.

It is usual to compute the following statistical values to test the goodness of fit of the regression equation:

(1) Simple correlation coefficient r which is computed for two variables and measures the association between them. As r varies from -1 to $+1$ it indicates the correlation between the variables. A value approaching ± 1 indicates good correlation.
(2) Multiple correlation coefficient R which measures the goodness of fit between the regression estimates and the observed data. $100R^2$ gives the percentage of variation explained by the regression.

Transportation studies have produced a considerable number of regression equations and their variety is often confusing. This is partly owing to variations in the form of the independent variables which have been used and also to variations in tripmaking.

A typical equation obtained in the Leicester transportation study[1] was

$$Y_8 = 0.0649X_1 - 0.0034X_3 + 0.0066X_4 + 0.9489Y_1$$

where Y_8 = total trips per household where the head of the household is a junior non-manual worker/24 h,
X_1 = family size,
X_3 = residential density,
X_4 = total family income,
Y_1 = cars/household.

If all the factors influencing the pattern of movement are correctly identified then it might be expected that trip generation equations obtained in one survey would be applicable to other surveys. Attempts to establish similarity between relationships observed in different surveys have not however met with any great success because of variations in both the dependent and independent variables chosen.

Category analysis

Difficulties with the use of regression equations for the study of trip generation have led to considerable support being given to the use of disaggregate models, that is, models based on the household or the person. Making use of the unexpanded sample data, these models make no reference to zone boundaries and allow considerable flexibility in the selection of alternative zone systems when future trip ends are being predicted.

This approach has become known as category analysis and has been largely developed by Wootton, Pick and Gill[2,3] and has been applied to a considerable number of transport studies. Category analysis uses the household as the fundamental unit of the trip generation process and assumes that the journeys it generates depend on household characteristics and location relative to workplace, shopping and other facilities. Trip generation is measured as the average number of one-way trips generated by a household on an average weekday.

Household characteristics that are readily measured and appear to account for variation in generation both at the present time and in the future are disposable income, car ownership, family structure and size. Location characteristics have proved more difficult to isolate and the characteristic that has found the greatest application is public transport accessibility.

Wootton and Pick[2] classified households according to:

1. Cars owned – (1) None
   (2) 1
   (3) More than 1

2 Income – Originally 6 classes were proposed from less than £500 to £2500 or more. These classes were modified in the West Midland Study to give an income up to £10,000. There is usually a difficulty in obtaining sufficient observations in higher income groups and yet it is in these ranges that future prediction is important.

3. Family structure

	Adults employed	Adults not employed
(1)	None	1
(2)	None	More than 1
(3)	1	0, 1
(4)	1	More than 1
(5)	More than 1	0, 1
(6)	More than 1	More than 1

These categories produced 108 household classes and associated with each class is a trip rate. It was proposed that trips be classified by mode of travel and trip purpose.

These are:

Modes 1. Drivers of cars or motor cycles
2. Public transport passengers
3. Other passengers (mostly car passengers)

Purpose 1. Work 2. Business 3. Education
4. Shopping 5. Social 6. Non-home based

where there are thus $6 \times 3 = 18$ mode and purpose trip combinations.

The basic assumption is that trip rates are stable over time and that the future behaviour of a household can be described by the category into which it falls. The number of households in each category can be estimated by the fitting of mathematical distributions to the observed values of income, car ownership and family structure.

A great advantage of the category analysis technique is that it is possible to estimate household categories from the Registrar General's Census Data using known relationships. Trip generation rates obtained from other surveys are then used subject to a small survey check on the accuracy of the rates. When the cost of large scale home interview surveys is considered, the advantage of this technique can be appreciated.

The computational techniques are also simpler than those required for zonal least squares regression. The use of disaggregate data may also be expected to simulate human behaviour more realistically than zonal values.

A disadvantage of the category analysis technique however is that it is assumed that income and car ownership will increase in the future. The categories with high incomes and car ownership are however the ones which are least represented in the base year data. Moreover they are the categories which are most likely to be used for future estimates of trip generation.

Income distribution

Pick and Gill have shown that income distribution may be represented by a continuous probability density function $Q(x)$ such that the number of households having income x, $a < x < b$, is given by

$$N \int_a^b Q(x)\,\mathrm{d}x$$

where N is the number of households in the zone.

The following distribution has been used for $Q(x)$:

$$Q(x) = \frac{\alpha^{n+1}}{n!} x^n \mathrm{e}^{-\alpha x}$$

where $\alpha = \bar{x}/s^2$,
$n = (\bar{x}^2/s^2)-1$,
$\bar{x}$ is the mean income,
s^2 is the standard deviation of income.

Future incomes were projected on the basis of

$$x^1 = x(1 + g)y$$

where g is the annual growth rate of income relative to the cost of living,
y is the number of years projected,
x^1 is the future income,
x is the present income.

Car ownership distribution

A conditional probability function $P(n/x)$ was derived from the West Midlands Transportation Study to give the probability of a household owning n cars if its income relative to the price of cars is x.

Then

$$P(0/x) = Ke^{-\beta x}$$

$$P(1/x) = Ce^{-\beta x} \times (\beta x)^n$$

where K and C are constants and β varies with bus accessibility.

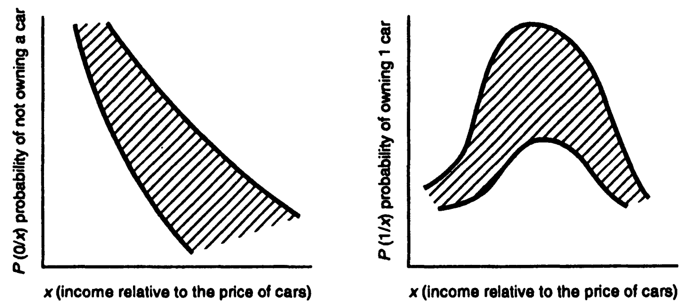


Figure 6.2 *Probability of car ownership (adapted from ref. 3)*

Variations of $P(0/x)$ and $P(1/x)$ within a study area as given by Pick and Gill are shown in figure 6.2. These variations have been explained by the following factors, but more detailed information can be obtained from ref. 3:

(a) An increase in residential density produces a higher $P(0/x)$ curve and a lower $P(1/x)$ curve. It has been found that the effects of high and low density are well defined but there is some variation in the medium density ranges.

(b) An increase of the public transport accessibility index of a traffic district causes an increase in the value of $P(0/x)$. It is defined as

$$\sum_{j} \sqrt{b_j}/\sqrt{a_i}$$

where a_i = area of district i (sq. miles),
b_j = number of buses on route j passing through the district per unit time.

This index has a value of 0 in a rural area without bus services and a value of approximately 60 in central London.

(c) The cost of housing has been found to influence car ownership: the greater the housing cost the higher the $P(0/x)$ curve.

(d) Lack of garage space inhibits multi-car owners and influences both the $P(0/x)$ and $P(1/x)$ curve.

(e) Because older household members are likely to have more money available for car purchase and younger household members are likely to be moped and motor cycles owners, age structure influences car ownership curves.

(f) Spatial relationships affect car ownership for if activities are closely related to housing then there is less necessity for the purchase of vehicles.

Work by the Department of Transport for the development of the Regional Highway Traffic Model[4] based on Family Expenditure Survey data for 1965, 1966, 1969–75 indicated that a relationship stable over time existed between household income deflated by an index of car price. Doubts concerning this relationship were raised when Family Expenditure Survey data for 1976–78 became available which indicated car ownership would fall because car prices rose more rapidly than real gross household incomes, but in actual fact car ownership continued to rise. For this reason a new model was proposed which did not include the car price index. Instead the independent variables defining car availability were replaced by the gross household income deflated by the Retail Price Index and driving licences per adult. It is hoped that the latter variable will reflect behaviour over time.

Family structure distribution

As previously stated households are split into 6 categories and the number of households in each category may be estimated using the following argument:

Let the probability of a household having n members be $Q(n)$.

Let $p(n)$ be the probability that a member of an n member household is employed. The probability that r members of an n member household are employed is

$$\frac{n!}{r!(n-r)!}\,p(n)^r(1-p(n))^{n-r}$$

assuming a binomial distribution. The probability that a household has n members of whom r are employed is then

$$P(n,r) = \frac{n!\,Q(n)}{r!(n-r)!}\,p(n)^r\,(1-p(n))^{n-r}$$

$$\text{where } Q(n) = \frac{e^{-x}x^{n-1}}{(n-1)!} \qquad n = 1, 2, \ldots$$

$$x = \text{average family size} - 1$$

$$p(n) = \frac{\text{employed residents}}{\text{total households}}$$

Trip generation and attraction models developed for the Regional Highway Traffic Model

The Department of Transport[4] developed models for two basic types of trip: home-based trips which have either origin or destination at the home, and non-home-based trips which have neither origin nor destination at the home.

Home-based trip end models were developed for three separate trip purposes: home-based work trips, home-based employer's business and home-based other. For non-home-based trips models were developed for employer's business and other trips.

For both home-based and non-home-based trip models two separate but related sub-models were calibrated. In the case of home-based purposes the models were for trip generation and trip attraction, for non-home-based trips models were developed for trip generation (or creation if the purpose was non-home-based) and trip attraction (or allocation if the purpose was non-home-based).

Both models for non-home-based trip creations had as the independent variables the proportions of households in the zone with zero, one, and two or more cars. The models for non-home-based trip allocations had as independent variables the households in the zone, total employment places, retail employment places, manufacturing employment places and service employment places in the zone and the average car ownership per household.

The generation models developed were:

$G_w = (a + bE_y)(1 - P)$	work car trip generations per household for car ownership group Y and zone type Z
$G_{eb} = (a + bE_y)P$	employer's business car trip generations per household for car ownership group Y and zone type Z
$G_o = c + dL + eN$	other purpose car trip generations per household for car ownership group Y and zone type Z
where E_y	is the average number of employed residents per household in car ownership group Y and zone type Z
P	is the proportion of work and employer's business trips which is employer's business in car ownership group Y and zone type Z
L	is the average number of driving licences per household in car ownership group Y and zone type Z
N	is the average number of non-employed residents per household
and a, b, c, d and e	are coefficients which are calibrated for car ownership group Y and zone type Z.

The average number of employed residents in one car and multi-car-owning households (E_1 and E_2) was determined by:

$$E_1 = E + B\,(1 - C)$$

$$E_2 = E + B\,(2.169 - C)$$

where E and C were respectively the zonal averages of Employed Residents and Cars per household and the form of this equation with $B = 0.62$ ensured compatibility with the overall planning data estimates of activity rates. The value of 2.169 was the average number of cars per household in multi-car-owning households.

Eight different types of zones (Z) were defined, comprising 2 types of rural zones, 3 types of urban zones and 3 types of London zones. Three levels of household car ownership (Y) were also defined, zero, one and two or more cars available per household.

The home-based attraction models which were developed were for work trips

$$A_w = f \times TE + g \times TE \times AX + h \times AREA$$

for employer's business trips

$$A_{eb} = i \times TE \times AX$$

for other trips

$$A_o = j \times RET \times AX + k \times HH \times AX + l \times CARS$$

where TE is the total employment in the zone,
AX is a measure of car accessibility,
$AREA$ is the zone area in hectares,
RET is the retail employment in the zone,
HH are the households in the zone,
$CARS$ is the number of cars in the zone,
f, g, h, i, j, k, l are calibration coefficients.

The category analysis method of modelling trip generation was employed by Scott Wilson Kirkpatrick and Partners for home-based trip productions in the Baghdad Transportation Study.

The household categories used were based on five variables; these were, number of persons per household, number of workers per household, number of cars available per household (0, 1 or 2 plus), income per household (low, medium or high) and household density (less than 1, and more than 20 and 1 to 20 households per hectare). The variables of number of workers per household and number of persons per household were combined into 10 worker/resident groups. The variable of household density was added to the analysis when it was found that low-density rural areas and also high-density urban areas had trip rates which were lower than the average.

Household trip rates were found to increase rapidly with increasing household income and with increasing car availability. On the other hand, the trip rate did not vary greatly with household size, especially in the range 5 to 11 residents per household.

Future numbers of households in each category were predicted from a household category generation sub-model and the future trip generation from these future households obtained by assuming present-day trip generation rates remained constant with time.

For non-home-based trips, stepwise multiple linear regression analysis was used to develop trip generation equations.

An example of a trip generation equation derived by stepwise linear regression analysis was for non-home-based zonal attractions which had the form:

$$\begin{aligned}\text{24 hour NHB attractions} = {} & 130.1 \\ & + 4.209 \text{ Employment in wholesale trade} \\ & + 0.907 \text{ Employment in community services} \\ & + 2.337 \text{ Number of restaurant seats} \\ & + 0.765 \text{ Employment in retail trade} \\ & + 0.729 \text{ Employment in construction}\end{aligned}$$

References

1. J. M. Harwood and V. Miller, *Urban Traffic Planning*, Printerhall, London (1964)
2. H. J. Wootton and G. W. Pick, Travel estimates from census data, *Traff. Engng Control*, **9** (1967), 142–5
3. G. W. Pick and J. Gill, New developments in category analysis, *PRTC Symposium*, London (1970)
4. Department of Transport, *Traffic Appraisal Manual*, London (1982)

Problems

Are the following statements true or false?

(a) A trip with an origin at the workplace and a destination at the home is said to be generated by the home.

(b) A journey from work to home made by walking to the bus, travelling by bus to the station and completing the journey by train is regarded as three trips.

(c) A trip end modal-split generation model predicts the trips generated by a traffic zone regardless of the mode of travel.

(d) In a trip generation equation the independent variables usually describe land-use and socio-economic factors.

(e) There would be no objection to the use of total household income and number of employed members in the household as independent variables in a trip generation equation.

(f) The use of trip generation equations to predict future trips depends on the ability to estimate future values of the independent variables.

(g) Category analysis uses disaggregate survey data while regression analysis employs zonal aggregate survey data.

(h) The examination of survey data by the use of category analysis techniques shows that the trip generation rate decreases with car ownership and increases with income class.

Solutions

Statement (a) is correct since all work trips with an origin or a destination at the home are said to be generated at the home.
Statement (b) is incorrect since a trip is a single journey between an origin and a destination.
Statement (c) is incorrect since a trip end modal-split model predicts trips classified by modal type; that is, modal split is carried out before trip distribution takes place.
Statement (d) is correct because trip generation equations have as their dependent variable the number of trips generated while the independent variables are the land-use and socio-economic factors which affect the generation of trips.
Statement (e) is incorrect in that the independent variables should be independent of each other while there is likely to be a strong correlation between the number of employed members of a household and the total household income.
Statement (f) is correct because future trips are estimated by assuming that the same correlation as exists today between tripmaking and land-use and socio-economic factors will exist in the future. Future trips are then estimated by substituting future land-use and other factors into the trip generation equations.
Statement (g) is correct in that category analysis attempts to predict tripmaking at the household level while least squares regression analysis derives equations from aggregated zonal data.
Statement (h) is incorrect since the analysis of the tripmaking habits of households shows that the number of trips made by households increases as the number of cars owned by the household increases and as household income increases.

7

Trip distribution

Trip distribution is another of the major aspects of the transportation simulation process and although generation, distribution and assignment are often discussed separately, it is important to realise that if human behaviour is to be effectively simulated then these three processes must be conceived as an interrelated whole.

In trip distribution, two known sets of trip ends are connected together, without specifying the actual route and sometimes without reference to travel mode, to form a trip matrix between known origins and destinations.

There are two basic methods by which this may be achieved:

(1) Growth factor methods, which may be subdivided into the

 (a) constant factor method;
 (b) average factor method;
 (c) Fratar method;
 (d) Furness method.

(2) Synthetic methods using gravity type models or opportunity models.

Trip distribution using growth factors

Growth factor methods assume that in the future the tripmaking pattern will remain substantially the same as today but that the volume of trips will increase according to the growth of the generating and attracting zones. These methods are simpler than synthetic methods and for small towns where considerable changes in land-use and external factors are not expected, they have often been considered adequate.

(a) *The Constant Factor Method* assumes that all zones will increase in a uniform manner and that the existing traffic pattern will be the same for the future when growth is taken into account. This was the earliest method to be used, the basic assumption being that the growth which is expected to take place in the survey area will have an equal effect on all the trips in the area. The relationship between present and future trips can be expressed by

$$t'_{ij} = t_{ij} \times E$$

where t'_{ij} is the future number of trips between zone i and zone j; t_{ij} is the present number of trips between zone i and zone j. E is the constant factor derived by dividing the future number of trip ends expected in the survey area by the existing number of trip ends.

This method suffers from the disadvantages that it will tend to overestimate the trips between densely developed zones, which probably have little development potential, and underestimate the future trips between underdeveloped zones, which are likely to be extremely developed in the future. It will also fail to make provision for zones which are at present undeveloped and which may generate a considerable number of trips in the future.

(b) *The Average Factor Method* attempts to take into account the varying rates of growth of tripmaking which can be expected in the differing zones of a survey area.

The average growth factor used is that which refers to the origin end and the destination end of the trip and is obtained for each zone as in the constant factor method. Expressed mathematically, this can be stated to be

$$t'_{ij} = t_{ij}\frac{(E_j + E_i)}{2}$$

where $E_i = \dfrac{P_i}{p_i}$ and $E_j = \dfrac{A_j}{a_j}$

t'_{ij} = future flow *ab*,

t_{ij} = present flow *ab*,

P_i = future production of zone i,

p_i = present production of zone i,

A_j = future attraction of zone j,

a_j = present attraction of zone j.

At the completion of the process attractions and productions will not agree with the future estimates and the procedure must be iterated using as new values for E_i and E_j the factors P_i/p'_i and A_j/a'_j where p'_i and a'_j are the total productions and attractions of zones i and j respectively, obtained from the first distribution of trips. The process is iterated using successive values of p'_i and a'_j until the growth factor approaches unity and the successive values of t'_{ij} and t_{ij} are within 1 to 5 per cent depending upon the accuracy required in the trip distribution.

The average factor method suffers from many of the disadvantages of the constant factor method, and in addition if a large number of iterations are required then the accuracy of the resulting trip matrix may be questioned.

(c) *The Fratar Method*[1]. This method was introduced by T. J. Fratar to overcome some of the disadvantages of the constant factor and average factor methods. The Fratar method makes the assumptions that the existing trips t_{ij} will increase in proportion to E_i and also in proportion to E_j. The multiplication of the existing flow by two growth factors will result in the future trips originating in zone i being greater than the future forecasts and so a normalising expression is introduced which is the sum of all the existing trips out of zone i divided by the sum of all the existing trips out of zone i multiplied by the growth factor at the destination end of the trip. This may be expressed as

$$t'_{ij} = t_{ij} \times \frac{P_i}{p_i} \times \frac{A_j}{a_j} \times \frac{\sum^{k} t_{ik}}{\sum^{k} (A_k/a_k) t_{ik}}$$

where t'_{ij} = future traffic flow from i to j,
t_{ij} = existing traffic flow from i to j,
P_i and A_j are total future trips produced by zone i and attracted to zone j respectively,
p_i and a_j are total existing trips produced by zone i and attracted to zone j respectively,
k = total zones.

The procedure must be iterated by substituting t'_{ij} for t_{ij}, $\sum_j t'_{ij}$ for p_i, $\sum_i t'_{ij}$ for a_j. Agreement to between 1 and 5 per cent is achieved by successive iterations.

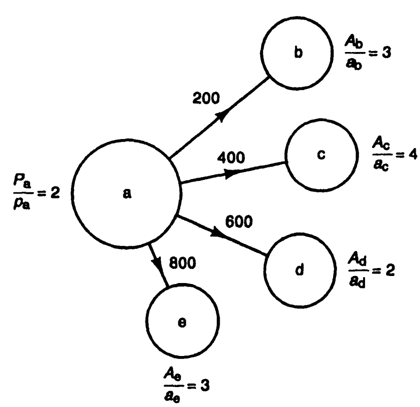


Figure 7.1 *The Fratar method of trip distribution*

The Fratar method can be illustrated by the simple example shown in figure 7.1 which shows the growth factors and the existing trip pattern.

Then

$$t'_{ij} = t_{ij} \times \frac{P_i}{p_i} \times \frac{A_j}{a_j} \times \frac{\sum^{k} t_{ik}}{\sum^{k} \left(\frac{A_k}{a_k}\right) t_{ik}}$$

that is

$$t'_{ab} = 200 \times 2 \times 3 \times \frac{(200 + 400 + 600 + 800)}{200 \times 3 + 400 \times 4 + 600 \times 2 + 800 \times 3}$$
$$= 414$$

$$t'_{ac} = 400 \times 2 \times 4 \times \frac{(200 + 400 + 600 + 800)}{200 \times 3 + 400 \times 4 + 600 \times 2 + 800 \times 3}$$
$$= 1103$$

$$t'_{ad} = 600 \times 2 \times 2 \times \frac{(200 + 400 + 600 + 800)}{200 \times 3 + 400 \times 4 + 600 \times 2 + 800 \times 3}$$
$$= 828$$

$$t'_{ae} = 800 \times 2 \times 3 \times \frac{(200 + 400 + 600 + 800)}{200 \times 3 + 400 \times 4 + 600 \times 2 + 800 \times 3}$$
$$= 1655$$

In this example the future trips produced by zone a meet the requirements that $P_a/p_a = 2$, but the requirements that

$$\frac{A_b}{a_b} = 3; \frac{A_c}{a_c} = 4; \frac{A_d}{a_d} = 2; \text{and } \frac{A_c}{a_c} = 3$$

are not met and in a practical example, where the number of zones would be considerably greater, further iterations would be required.

(d) *The Furness Method*[2]. In this method the productions of flows from a zone are first balanced and then the attractions to a zone are balanced. This may be expressed

$$t'_{ij} = t_{ij} \times \frac{P_i}{p_i}$$

$$t''_{ij} = t'_{ij} \times \frac{A_j}{\sum \text{trips attracted to j in first iteration}}$$

$$t'''_{ij} = t''_{ij} \times \frac{P_{ij}}{\sum \text{trips produced by i in second iteration}}$$

where the symbols are as previously stated.

Usually these simplified approaches to trip distribution are only suitable for smaller surveys where similar growth factors are applied to zones and where considerable areas of new development are not expected. The following extract from *Land*

Use / Transport Studies for Smaller Towns[3] illustrates the use of growth factors and the Furness method of trip.

The four steps in the calculation are:

(1) Total the outgoing trips for each zone and multiply by the zonal growth factor to obtain the predicted origin outgoing totals.
(2) Multiply lines in the matrix by the appropriate origin factor.
(3) Total the incoming trips into each zone and divide into the predicted incoming totals to obtain the destination factors.
(4) Repeat the iteration processes until the origin or destination factor being calculated is sufficiently close to unity (within 5 per cent is normally satisfactory).

Example

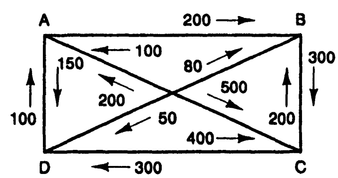


Figure 7.2 *The Furness method of trip distribution (initial flows)*

TABLE 7.1 *Present flows*

	Origin	Destination A	B	C	D	Present outgoing totals	Predicted outgoing totals	Growth (origin) factor
	1	2	3	4	5	6	7	8
9	A	0	200	500	150	850	2550	3.0
10	B	100	0	300	50	450	1125	2.5
11	C	200	200	0	300	700	1400	2.0
12	D	100	80	400	0	580	925	1.6
13	Present incoming totals	400	480	1200	500		6000	
14	Predicted incoming totals	480	720	3600	1200		6000	
15	Acceptance (destination) factor	1.2	1.5	3.0	2.4			

Notes: col. 7 = col. 6 col. 8
line 14 = line × 13 × line 15
Total col. 7 must equal approximately total line 14; that is future trip ends in system must balance, any adjustment to calculated acceptance factors necessary to secure balance being made at the non-residential end of the trips.

TABLE 7.2 *Iteration 1*

		A	B	C	D	
		16	17	18	19	
20	A	0	600	1500	450	Figures in this matrix are those in lines 9–12 multiplied by respective origin factors in column 8
21	B	250	0	750	125	
22	C	400	400	0	600	
23	D	160	128	640	0	
24		810	1128	2890	1175	
25		480	720	3600	1200	As in line 14
26	New destination factors	0.59	0.64	1.25	1.02	

TABLE 7.3 *Iteration 2*

	A	B	C	D		(as col. 7)	
A	0	384	1870	458	2712	2550	0.94
B	148	0	935	128	1211	1125	0.93
C	217	256	0	612	1105	1400	1.26
D	95	82	797	0	974	925	0.95

Figures in this matrix are columns 16–19 multiplied by respective new destination factors on line 26.

TABLE 7.4 *Iteration 3*

	A	B	C	D
A	0	362	1760	432
B	138	0	870	119
C	300	324	0	775
D	90	78	752	0
	528	764	3382	1326
	480	720	3600	1200
	0.91	0.94	1.07	0.91

TABLE 7.5 *Iteration 4*

	A	B	C	D			
A	0	441	1882	392	2615	2550	0.98
B	125	0	906	108	1141	1125	0.99
C	273	305	0	702	1280	1400	1.09
D	82	73	807	0	962	925	0.96

TABLE 7.6 *Iteration 5*

	A	B	C	D
A	0	333	1840	383
B	123	0	895	107
C	299	334	0	770
D	79	70	777	0
	501	737	3512	1260
	480	720	3600	1200
	0.96	0.96	1.02	0.95

When all the origin or destination factors being calculated in one iteration are within 5 per cent of unity, the result may be considered satisfactory.

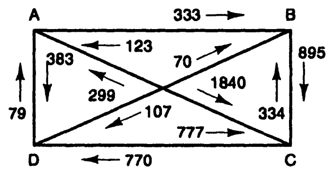


Figure 7.3 *The Furness method of trip distribution (distributed flows)*

General comment on growth factor methods

The usual application of growth factor methods is in updating recent origin–destination data where the time scale is short and where substantial changes in land use or communications are not expected or have not already taken place. The use of a growth factor method is largely dependent on the accurate calculation of the growth factor itself and this is a likely source of inaccuracy. It is however the lack of any measure of travel impedance which is the major disadvantage of these methods; without this it is impossible to take into account the effect of new and improved travel facilities or the restraining effects of congestion.

Trip distribution using synthetic models

The use of synthetic distribution models allows the effect of differing planning strategies and communication systems and, in particular, travel cost to be estimated, whereas growth factor methods base future predictions on the existing pattern of movement.

The models are usually referred to as synthetic models because existing data is analysed in order to obtain a relationship between tripmaking and the generation and attraction of trips and travel impedance. These models seek to determine the causes of present-day travel patterns and then assume that these underlying causes will remain the same in the future. The most widely used trip distribution model is the so-called 'gravity model'. It has been given this name because of its similarity with the gravitational concept advanced by Newton. It states that trip interchange between zones is directly proportional to the attractiveness of the zones to trips and inversely proportional to some function of the spatial separation of the zones.

The gravity model may generally be stated as

$$t_{ij} = kA_iA_jf(Z_{ij})$$

where A_i and A_j are measures of the attractiveness of the zones of origin and destination to tripmaking and $f(Z_{ij})$ is some function of deterrence to travel expressed in terms of the cost of travel, travel time or travel distance between zones i and j.

If $f(Z_u)$ is taken as $1/Z_{ij}^2$ then the formula is similar to the law of gravitational attraction and for this reason the model is referred to as the gravity model.

Tanner[4] has shown that this form of the deterrence function cannot give valid estimates of travel over large or small distances. He suggested a function of the form

$$\exp(-\lambda Z_{ij}) \times Z_{ij}^{-n}$$

where λ and n are constants.

For work journeys a typical value of λ was found to be 0.2 where Z_{ij} was measured in miles. For many purposes n may be taken as 1.

As Z_{ij} increases the effect of $\exp(-\lambda Z_{ij})$ decreases and the effect of Z_{ij}^{-n} increases.

This is the distribution of trip costs, times or distances which is observed in most studies of tripmaking. The trips that are least in cost, distance or time occur most frequently while trips that are greater in cost, distance or time are observed less frequently. There is considerable evidence that the value of the deterrence function varies with trip type and for this reason practically all studies use differing values for differing trip categories. It is also probable that the function varies with time, but this effect is at present largely ignored.

In practice tripmaking between zones is assumed to be proportional to the total trips generated by the zone of origin of the trip and the total trips attracted by the zone of destination of the trip, or

$$t_{ij} = KP_iA_jf(Z_{ij})$$

where K is a constant.

P_i is the total trip production of zone i, that is $\sum_i t_{ij}$

A_j is the total trip attraction of zone j, that is $\sum_j t_{ij}$

The value of K is frequently taken as

$$\frac{1}{\sum_j A_j f(Z_{ij})}$$

so that

$$t_{ij} = \frac{P_i A_j f(Z_{ij})}{\sum_j A_j f(Z_{ij})}$$

This is often referred to as a production constrained model because the use of this form results in the total trip production of zone i; that is $\sum_j t_{ij}$ is equal to P_i.

When it is desired to doubly constrain the distribution model, that is

$$\sum_j t_{ij} = P_i$$

$$\sum_i t_{ij} = A_j$$

then K may be replaced by $c_i d_j$, giving

$$t_{ij} = c_i d_j P_i A_j f(Z_{ij})$$

and

$$\sum_j t_{ij} = \sum_j c_i d_j P_i A_j f(Z_{ij}) \text{ or } c_i = \frac{1}{\sum_j d_j A_j f(Z_{ij})}$$

and similarly

$$d_j = \frac{1}{\sum_i c_i P_i f(Z_{ij})}$$

The distribution is constrained by setting d_j equal to unity, and obtaining

$$t_{ij} = \frac{d_j P_i A_j f(Z_{ij})}{\sum_j d_j A_j f(Z_{ij})}$$

If $\sum_i t_{ij}$ is not equal to A_j then d_j is made equal to $A_j / \sum_i t_{ij}$ and the procedure iterated until $\sum_i t_{ij}$ approaches A_j.

Past experience has shown that the exponent of travel time varies with travel purpose and also with travel time. For this reason the gravity model is usually presented in the revised form

$$t_{ij} = \frac{P_i A_j F_{ij} K_{ij}}{\sum_{j=1}^{n} A_j F_{ij} K_{ij}}$$

where F_{ij} = an empirically derived travel time or friction factor which expresses the average area wide effect of spatial separation on trip interchange between zones which are z_{ij} apart. This factor approximates $1/Z^n$ where n varies according to the value of Z expressed as the travel time between zones.

K_{ij} = a specific zone-to-zone adjustment factor to allow for the effect on travel pattern of defined social or economic linkage not otherwise accounted for in the gravity model formulation.

Standard computer programs are available for determining the most suitable values of F_{ij} and K_{ij} for the particular transportation study. These values are then assumed to remain constant in the future to allow future flows to be predicted; this procedure is known as calibration.

The travel time factors are calculated on a survey wide basis by assuming a value of F_{ij} for a given time range. t_{ij} is calculated for all the flows within the same time range and compared with observed values. The procedure is iterated until agreement is obtained and finally zone-to-zone agreement is obtained by the use of K factors.

Application of the gravity model to the London Transportation Study

The gravity model concept was used for trip distribution in the London Transportation Study[5]. It was assumed that the number of trips generated in one district and attracted to another was a function of trip generation and attraction of the two districts respectively and the travel time between them.

These functions, referred to as distribution functions, were calculated from the survey data and assumed to remain constant with time. Trips were stratified by purpose and mode into six types: work, other home-based and non-home-based, for both car owner and public transport trips. Because these functions may vary with trip length the trips were divided into thirteen trip time-length intervals so that approximately equal numbers of trips were in each interval.

Using travel data from the household interviews the distribution functions F_{kt} for each district of attraction k and time interval t were calculated:

$$F_{kt} = \frac{\sum_{i(t)} T_{ik}}{A_k \sum_{i(t)} G_i}$$

where i = district of generation,

T_{ik} = trips generated in district i and attracted to district k,

A_k = total trips attracted to district k,

G_i = total trips generated in district i,

$\sum_{i(t)}$ = summation over all districts of generation, falling in time interval t.

It was found that when these functions were plotted to a logarithmic scale against time, they showed an approximately linear relationship but with a discontinuity in gradient at a trip time length of around 25 minutes.

Forecasting of future trips was carried out separately for each trip purpose after the removal of a fixed percentage of intra-district trips. The first iteration of the distribution was obtained by the use of the model

$$T_{1ik} = \frac{F_{ik}A_kG_i}{\sum_k (F_{ik}A_k)}$$

where F_{ik} = distribution function for trips from districts i to k using 1981 trip times,

A_k = forecast attractions in district k,

G_i = forecast generations in district i.

A ratio was then obtained for each district of attraction, so that

$$R_{1k} = \frac{A_k}{\sum_i T_{1ik}}$$

Using this ratio, a second iteration was made

$$T_{2ik} = \frac{F_{ik}R_{1k}A_kG_i}{\sum_k (F_{ik}A_kR_{1k})}$$

followed by the calculation of

$$R_{2k} = \frac{R_{1k}A_k}{\sum_i T_{2ik}}$$

The procedure was carried out for a third time, the criterion for convergence being that the average (unsigned) error should be within the range 2 to 5 per cent.

References

1. T. J. Fratar, Vehicular trip distributions by successive approximations, *Traff Q.*,**8** (1954), 53–64
2. K. P. Furness, Time function iteration, *Traff Engng Control*, 7 (1965), 458–60
3. Ministry of Transport, Land Use/Transport Studies for Smaller Towns, *Memorandum to Divisional Road Engineers and Principal Regional Planners*, London (1965)
4. J. C. Tanner, Factors affecting the amount of travel, *DSIR Road Research Technical Paper* 51, HMSO, London (1961)
5. Greater London Council, *Movement in London*, County Hall, London (1969)

Problems

(1) Car driver work trips produced by the residents of zone 1 amount to 1000 trips. It is desired to distribute these trips to zones 1–4 which have the following characteristics:

Zone 1 has an intrazonal time of 7 minutes and has 1000 work trips attracted to it from all zones in the study area. (The intrazonal time is the average travel time of trips which have origins and destinations within the zone.) Terminal time 2 minutes.
Zone 2 is 15 minutes from zone 1 and has a total of 700 work trips attracted to it from all zones in the study area. Terminal time 3 minutes.
Zone 3 is 19 minutes from zone 1 and has a total of 6000 work trips attracted to it from all zones in the study area. Terminal time 2 minutes.
Zone 4 is 20 minutes from zone 1 and has a total of 3000 work trips attracted to it from all zones in the study area. Terminal time 3 minutes.

The travel time factors applicable to journeys of this type are given below. The terminal time is the average time to park and walk to the destination, or walk from the origin to the car park.

Travel time (minutes)	*F*
1	200
5	120
7	100
11	80
14	68
16	61
17	58
18	52
20	49
21	47
23	45
25	39

(2) The present and the future generated and attracted trips from four traffic zones of a transportation study are as given below, together with the present trip matrix.

Present and future generated and attracted trips

Zone	A	B	C	D
present generated trips	1500	900	1800	800
present attracted trips	1200	1000	1500	2000
future generated trips	3000	1200	2700	2400
future attracted trips	1800	3000	3500	4000

Present trip matrix

Origins	Destinations			
	A	B	C	D
A	–	400	400	300
B	200	–	300	200
C	400	300	–	600
D	200	100	300	–

(a) Calculate the first approximation to the future trips between the zones using the average factor method.

(b) Calculate the second approximation to the future trips between the zones using the Furness method.

(c) Trips between the traffic zones of a proposed new town are assumed to be proportional to the trips generated by the zone of origin and the trips attracted by the zone of destination of the trip and inversely proportional to the second power of the travel time between the zones. Details of three traffic zones and the value of the future trips from C to A are also given below.

Zone	Attracted trips	Generated trips
A	2400	3600
B	1600	2000
C	4000	5000

The travel time between the zones is 10 minutes.

		Zone of destination		
		A	B	C
zone of origin	A			X
	B	Y		
	C	208	Z	

What is the correct value of X, Y and Z in the above table?

Solutions

(1)

1	2	3	4	5	6	
Zones	A_j	Terminal times	Travel time	t_{ij}	F_{ij}	column 2 × column 6
1–1	1000	2 + 2	7	11	80	80 000
1–2	700	2 + 3	15	20	49	34 300
1–3	6000	2 + 2	19	23	45	270 000
1–4	3000	2 + 3	20	25	39	117 000
						$\sum$ 501 300

Using the trip distribution equation

$$T_{ij} = \frac{P_i A_i F_{ij}}{\sum_j A_j F_{ij}}$$

where A_j is given in column 2
and F_{ij} is given in column 6

$$\text{Trips } 1\text{–}1 = \frac{1000 \times 80\,000}{501\,300} = 160$$

$$\text{Trips } 1\text{–}2 = \frac{1000 \times 34\,300}{501\,300} = 68$$

$$\text{Trips } 1\text{–}3 = \frac{1000 \times 270\,000}{510300} = 539$$

$$\text{Trips } 1\text{–}4 = \frac{1000 \times 117\,000}{501\,300} = 233$$

(2) (a) Calculation of the trips between the zones using the average factor method.

Zone	A	B	C	D
Attraction factor	1.5	3.0	2.3	2.0
Production factor	2.0	1.3	1.5	3.0

then

$$t'_{AB} = 400 \frac{(2.0 + 3.0)}{2} = 1000$$

$$t'_{AC} = 400\,\frac{(2.0 + 2.3)}{2} = 860$$

$$t'_{AD} = 300\,\frac{(2.0 + 2.0)}{2} = 600$$

$$t'_{BA} = 200\,\frac{(1.3 + 1.5)}{2} = 280$$

$$t'_{BC} = 300\,\frac{(1.3 + 2.3)}{2} = 540$$

$$t'_{BD} = 200\,\frac{(1.3 + 2.0)}{2} = 330$$

$$t'_{CA} = 400\,\frac{(1.5 + 1.5)}{2} = 600$$

$$t'_{CB} = 300\,\frac{(1.5 + 3.0)}{2} = 675$$

$$t'_{CD} = 600\,\frac{(1.5 + 2.0)}{2} = 1000$$

$$t'_{DA} = 400\,\frac{(3.0 + 1.5)}{2} = 450$$

$$t'_{DB} = 100\,\frac{(3.0 + 3.0)}{2} = 300$$

$$t'_{DC} = 300\,\frac{(3.0 + 2.3)}{2} = 795$$

These future trip interchanges can now be tabulated in an origin–destination matrix:

Origins	Destinations			
	A	B	C	D
A		1000	860	600
B	280		540	330
C	600	675		1000
D	450	300	795	

If complete details of the trip interchanges between all the zones of the survey had been given it would be possible to note that the future attractions and generations of the zones as obtained in this first iteration did not agree with the values given. It would then be necessary to carry out the procedure again using new values of attraction and production factors based on the ratio of first iteration interchanges to future interchanges.

(b) Calculation of the trips between the zones using the Furness method.

Iteration 1
Using the production factors calculated in (a)

	A	B	C	D
A	–	800	800	600
B	260	–	390	260
C	600	450	–	900
D	600	300	900	–
Attractions	1460	1550	2090	760
Required future attractions	1200	2400	2300	2200
New attraction factor	0.82	1.55	1.10	1.25

Iteration 2

	A	B	C	D
A	–	1240	880	750
B	213	–	429	325
C	492	698	–	1125
D	492	475	990	–

(3) It is stated that

$$t'_{ij} = \frac{KP_iA_j}{t^2}$$

From the data given of the trip interchange from C to A

$$208 = \frac{K \times 5000 \times 2400}{10^2}$$

$$K = \frac{208}{120\ 000}$$

Then

$$X = t'_{AC} = \frac{208 \times 3600 \times 4000}{120\,000 \times 1000}$$
$$= 250 \text{ trips}$$

$$Y = t'_{BA} = \frac{208 \times 2000 \times 2400}{120\,000 \times 1000}$$
$$= 83 \text{ trips}$$

$$Z = t'_{CB} = \frac{208 \times 5000 \times 1600}{120\,000 \times 100}$$
$$= 139 \text{ trips}$$

8

Modal split

Trips may be made by differing methods or modes of travel and the determination of the choice of travel mode is known as modal split. In the simplest case when a small town is being considered the choice is normally between one form of public transport and the private car, with the car being used for all trips where it is available. In such a situation most trips on the public transport network are captive to public transport and very little choice is being exercised. In the larger conurbations however the effect of modal split is of very considerable significance and is greatly influenced by transport policy decisions.

Modal split should not be viewed as an entity; it is closely related in the real situation with trip generation and distribution. It has been shown in Chapter 6 that additional tripmaking occurs when a private car is available and it has been observed that the destination is often influenced by the ease with which the car can be used. If the use of the private car is restricted then it is likely that the number of trips generated will be decreased rather than be made by an alternative travel mode.

In the simulation of the real system which is referred to as the transportation planning process, modal split may be carried out at the following positions in the process.

(a) Modal split may be carried out as part of trip generation whereby the number of trips made by a given mode is related to characteristics of the zone of origin. This means that transport trips are generated separately from private transport trips.

(b) Modal split may be carried out between trip generation and distribution. Car-owning households in the zone of origin have a choice of travel mode depending upon the car/household ratio while non-car-owning household trips are captive to public transport.

(c) Modal split may be carried out between the trip distribution and the trip assignment process. Trip distribution allows journey times both by public and private transport to be estimated and then the modal split between public transport trips may be made on the basis of travel time and cost.

These varying approaches will now be considered in greater detail.

(a) Modal split considered as part of the trip generation process

The direct generation of public and of private transport trips was used in early work in the USA and has often been carried out in surveys of small urban areas in the UK. Usually the modal split is made on the basis of car ownership in the zone of origin, distance of the zone of origin from the city centre and residential density in the zone of origin. Sometimes the relative accessibility of the zone of origin to public transport facilities is also included.

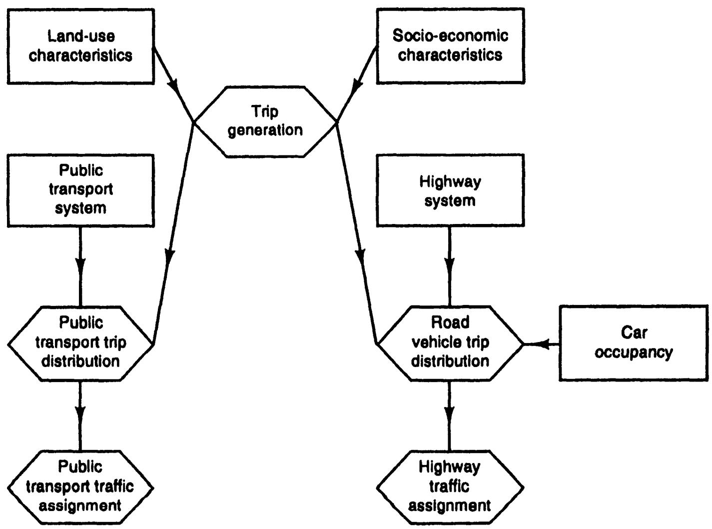


Figure 8.1 *A generalised trip end modal-split procedure*

Figure 8.1 illustrates the transport planning procedure graphically when public transport and car trips are generated separately.

This approach makes it difficult to take into account changes in the public transport network, improvements in the highway system and the restraint of private car use by economic means. Usually these models indicate a very high future car use and arbitrary modal split has to be imposed after the assignment process. For these reasons, modal split is now rarely considered at this early stage in the modelling process.

(b) Modal split carried out between trip generation and distribution

In this approach person trips are predicted and the percentage of these trips made by public and private transport estimated from such factors as socio-economic and land-use characteristics, the quality of the public transport system and the number of cars available. The assumption is made in this method that the total number of trips generated is independent of the mode of travel.

The transport planning procedure when modal split is carried out between generation and assignment is illustrated graphically in figure 8.2. This is not generally adopted because it has to be assumed here that tripmakers choose their mode of travel before deciding where to go.

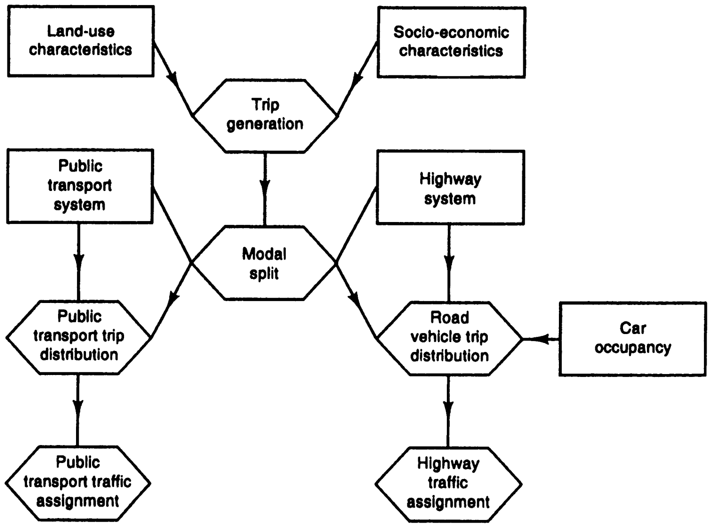


Figure 8.2 *The generalised model when modal split is carried out between generation and assignment*

(c) Modal split carried out between trip distribution and assignment

This approach is frequently used in transportation studies because it allows the cost and level of service of a trip to be used as the modal split criterion.

Because of the complexity of the transportation process, travel time alone is sometimes used as the cost criterion. Normally travel times based on road speeds are utilised to distribute the choice trips. These travel times together with travel time by public transport are then used to determine modal split and the public transport portion of these trips is added to the captive public transport trip ends as shown in figure 8.3 which is abstracted from *Movement in London*[1].

An early application of modal split based on travel time is illustrated in figure 8.4 taken from *Movement in London* where 54 diversion curves were developed from the survey data for 3 trip purposes (work, other home-based and non-home-based), long and short trips (divided at 15 minutes of travel time by private transport excluding parking) for each of the 9 sections of the transportation study area.

Original work on modal choice has been largely carried out by Quarmby[2] using observed travel characteristics in Leeds. Quarmby developed the concept of generalised cost, a linear combination of time and money costs which he referred to as disutility of travel, and showed that a good fit to observed modal split for trips between two zones could be obtained by an exponential cost difference equation of the form

$$\frac{n_x}{n_y} = \exp\left[-\alpha(c_x - c_y)\right]$$

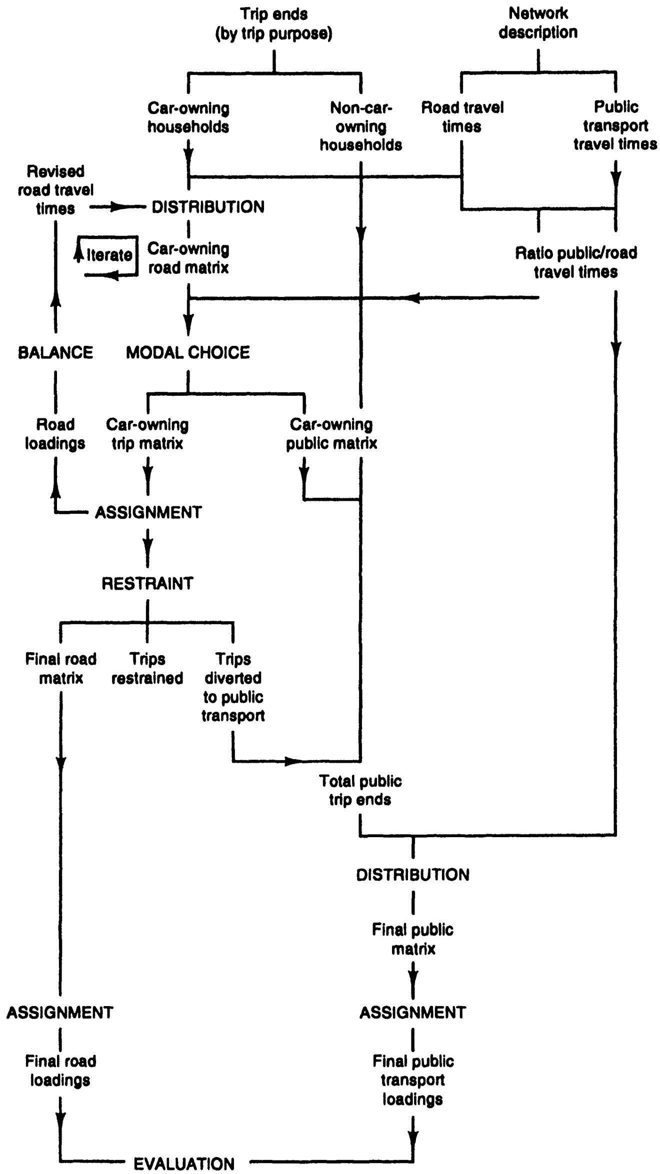


Figure 8.3 *Procedure when modal split is carried out between trip distribution and assignment*

where n_x is the proportion of trips by mode x, n_y is the proportion of trips by mode y, α is a calibration constant, and c_x and c_y are the generalised costs of travel between the zones by the two modes. Because n_x plus n_y must equal unity then

$$n_x = \frac{\exp^{-\alpha cx}}{\exp^{-\alpha cx} + \exp^{-\alpha cy}}$$

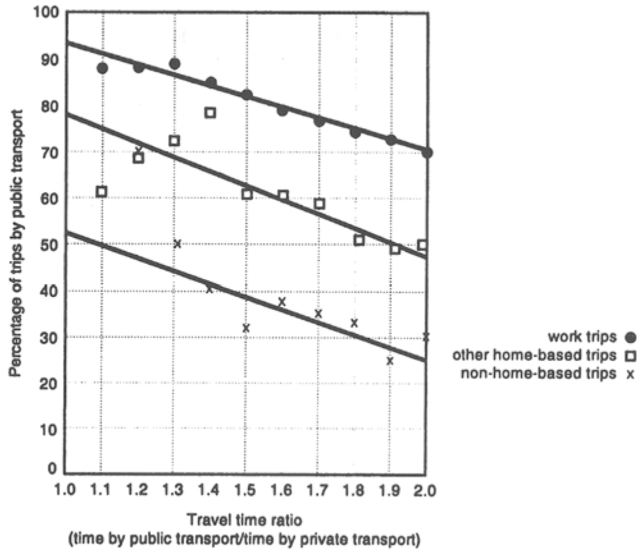


Figure 8.4 *The London Traffic Survey, specimen diversion curves (adapted from ref. 1)*

There are no particular reasons for selecting a particular modal choice procedure and the model developed will depend to a large extent on the survey information available; for this reason validation of the model against observed data is important.

The Department of Transport[3] has developed a modal choice model for new highways when a significant proportion of trips is not greater than 25 miles and where the scheme is not too large and is located where the effects of the scheme are unlikely to spread over a wide area of the network.

The procedure directly determines the proportion of trips made by public transport (mst) from

$$\frac{1}{1 + \exp(Z(I_t - I_c))}$$

where I_t is a measure of impedance for public transport trips from zone i to zone j,
I_c is a measure of impedance for car trips from zone i to zone j,
Z is a calibration constant which varies with trip purpose.

The ratio of public transport trips to car trips (r) is also

$$\exp(-Z(I_t - I_c))$$

When this relationship is used it is found that when the impedances I_t and I_c are equal the ratio r is not equal to 1 and to allow for this an additional 'modal preference factor' d is introduced into the public transport impedance so that I_t is equal to (in-vehicle time) + (w × out-of-vehicle time) + (trip cost)/(value of time) + d. The car trip impedance is given by (in-vehicle time) + (w × car access time) + (parking cost)/(value of time) + (highway trip distance × vehicle operating cost)/(value of time) where w is a conversion factor usually taken as 2.

The factor *d* differs at different study locations and if previous transport studies at the location of the new scheme have not indicated a suitable value from observed travel data then the default values in table 8.1 must be used.

TABLE 8.1
Default values of Z and d (from ref. 3)

Trip purpose	*Z*	*d*	
		Central areas	*Other*
Home-based work	0.03	−33	+17
Home-based other	0.03	−50	+33
Non-home-based	0.04	−25	+25

To apply this method it is necessary to determine the following values for all zone to zone movements: the highway distance, public transport links and costs, vehicle operating costs, parking costs, median income and access time both to public and private transport.

An illustration of the application of this method is shown in figure 8.5 where the percentage of public transport trips as a function of cost difference is given for various trip types.

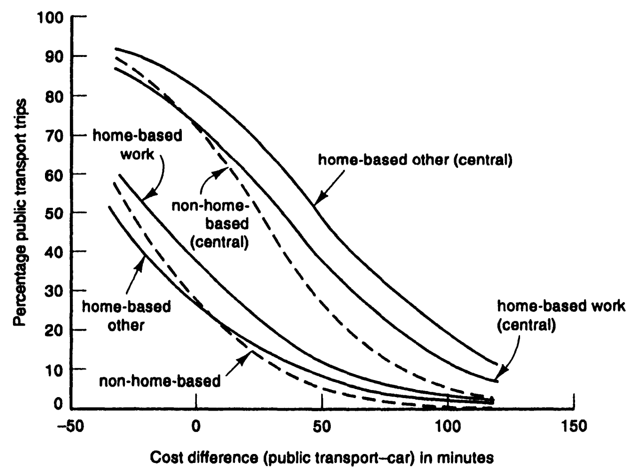


Figure 8.5 *Model split for highway schemes (based on ref. 3)*

Probabilistic models and utility functions

Models have been developed recently for mode choice based on probability concepts[4]. Each competing model is described in a utility (or disutility) function, and these utilities, expressed mathematically, describe the probability associated with a tripmaker's choosing of the competing alternatives. These models require the selection of a mathematical form and the calibration of appropriate utility functions, which must allow the selected model to reproduce the available base year data.

A utility function in this context describes the degree of satisfaction that tripmakers derive from their mode choices. A disutility function represents the generalised cost (or 'independence') associated with each choice. The magnitude of either depends on the attributes of each choice and of the individual making that choice (for example, their socio-economic status). The utility function requires a selection of appropriate variables in an appropriate functional form, at a suitable level of aggregation (for example, regional, zonal, household or individual), perhaps also varying by trip purpose and time of day.

The utility function is often expressed as the linear weighted sum of the independent variables, such as

$$U = a_0 + a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

where U is the utility derived from a choice defined by the magnitude of the attributes $X_1 \ldots X_n$ that are present in that choice, weighted by the corresponding model parameters $a_0 \ldots a_n$.

Separate utility models are sometimes calibrated for each mode, perhaps with similar attributes applying to each mode (such as cost, level of service or 'convenience') but with different model parameter values. This mode-specific formulation does present problems, however, when a new mode is introduced because the base year data necessary to calibrate its utility function would probably be unavailable.

A utility-based modal-choice enables estimation of the likelihood that a given mode will be selected or of the proportion of tripmakers who will select each mode of travel. The relationship between this proportion and the utilities of competing modes is usually expressed as a Logit model.

Logit models

The most commonly used model for modal choice is the Logit model (or multinominal Logit model). This model calculates the proportion of trips that will select a specific model K, $p(K)$, according to the following relationship:

$$p(K) = \frac{e^{U_K}}{\sum_i e^{U_i}}$$

where U_K is the utility of mode K.

This equation is often expressed in its expanded form:

$$\frac{T_k}{\sum_i T_i} = \frac{e^{-\lambda C_k + \delta C_k}}{\sum_i e^{-\lambda C_i + \delta C_i}}$$

where T_k = number of trips by mode k,
C_k = cost by mode k,
δC_k = mode-specific parameter not explicitly taken account of elsewhere,
λ is a general parameter to be estimated.

An alternative form is the hierarchical (or nested) Logit model, where a primary choice is made initially between two mode categories (such as public and private transport). A separate equation (with a different λ value) is then used for the choice between public transport modes.

References

1. Greater London Council, *Movement in London,* County Hall, London (1969)
2. D.A. Quarmby, Choice of travel mode for journey to work: some findings, *J. Transp. Econ. Policy,* **1** (1967), 273–314
3. Department of Transport, *Traffic Appraisal Manual,* London (1982)
4. C.S. Papacostas, *Fundamentals of Transportation Engineering,* Prentice-Hall, Englewood Cliffs, New Jersey (1987)

9

Traffic assignment

Previously the estimation of generated trip ends has been discussed together with the distribution of trips between the traffic zones. Modal split methods also have been reviewed in which the proportion of trips by the varying travel modes are determined. At this stage the number of trips and their origins and destinations are known but the actual route through the transportation system is unknown. This process of determining the links of the transportation system on which trips will be loaded is known as traffic assignment.

Apart from the very largest transportation surveys traffic assignment tends to deal with highway traffic. This is because it is usually not difficult to estimate the route taken by public transport users and also because the loading of trips on the public transport network does not materially affect the journey time.

The usual place of assignment in transportation planning synthesis is as illustrated in figure 8.3. Trip ends where there is no choice of travel mode, that is from non-car-owning households, are accumulated as public transport trip ends. Choice trips where a car is available are separated by the modal choice procedure into car trips and public transport trips, the public transport trips being accumulated and the car trips assigned to the network.

Usually it will then be found that the proposed road network is overloaded and some car trips will need to be restrained. If a car cannot be used then some trips will not be made at all, while other trips will be transferred to public transport and accumulated.

As the basis of assignment is usually travel time the travel times on the network links will vary the imposed loading. In addition as travel time is used in the trip distribution process it is necessary to carry out an iterative procedure between distribution, modal choice and assignment as illustrated in figure 8.3.

The change in speed with volume on a highway link is carried out using speed/flow relationships for the varying highway types and it is interesting to consider just what are the effects of a speed change. Firstly it affects the choice of route because assignment is made on the basis of travel times through the network. Secondly it affects the destinations of trips because trips are distributed to varying destinations on the basis of travel time when a gravity model is used. Finally it may affect the choice of travel mode because modal choice is often made by a comparison of travel times.

There are many problems associated with speed/flow relationships, considerable variation being observed between differing highways even of the same type. There is also the additional problem that most transportation studies are based on 24 hour flows so that it is necessary to know the hourly variation and the directional distribution of flows.

In assignment it is first necessary to describe the transport network to which trips are being assigned. The network is described as a series of nodes and connecting links; in a highway network the nodes would be the junctions and the links the connecting highways. Centroids of traffic zones, at which it is assumed that all zonal trips are generated and to which they are attracted, are either at nodes or connected to them by additional links. The cost of using a link and a junction, usually in the form of travel times and delays, is a key item of prediction in the assignment process.

A number of methods have been developed for undertaking traffic assignment:

(1) All-or-nothing assignment.
(2) Assignment by the use of diversion curves.
(3) Capacity-restrained assignment.
(4) Multipath proportional (or stochastic) assignment.
(5) Stochastic assignment with capacity constraint.
(6) 'Wardrop' equilibrium assignment.

All-or-nothing assignment

In this method an algorithm is used to compute the route of least cost, usually based on travel time between all the zone centroids. For each zone centroid selected as origin, a set of shortest routes from the origin to all the other zone centroids is referred to as a minimum tree. When the trips between two zones are assigned to the minimum path between the zones, then the assignment is said to take place on an all-or-nothing basis.

There are obvious difficulties with such a simplified approach, some of which are inherent in the other assignment methods. It is obviously incorrect to assume that all trips commence and terminate at a zone centroid. If the length of the links within the zones is small compared with the length of the remainder of the minimum link path, then the errors may not be so serious. Because of its simplicity, travel time is usually employed as a measure of link impedance, but travel times may not be precisely estimated by the traveller. The use of a cost function which reflects the perceived cost of travel is desirable. The loading on a link in this method is extremely sensitive to estimated link and node costs. If these have been incorrectly estimated, then the resulting assignment is open to question. There is also the problem that links with small travel costs will attract trips without any adjustment in link cost.

Assignment by the use of diversion curves

Originally diversion curves were used to estimate the traffic that would be attracted by a single new route or transport facility. It is thus necessary to compare travel cost with and without the new transport facility, the decision as to whether a trip would use the new facility being based on a cost ratio or difference between the with and the without situation. When diversion curves are used for assignment it is necessary to consider the network with and without the new facility or highway link as input to an all-or-nothing assignment. The travel costs from the two networks are then used with the diversion curve to determine the proportion of the trips that is diverted from the existing network and transferred to the network with the new facility.

An early application of this technique was used in the traffic studies for the first section of the M1 motorway[1] where 50 per cent of trips were diverted to the motorway if the cost (time) of the trip on the existing route was within ± 10 minutes of the trip cost on the motorway. It has often been considered that differences of trip cost more effectively modelled human behaviour than ratios. A diversion curve of this type is illustrated in figure 9.1.

Figure 9.1 *Early diversion curve*

It is also possible to employ diversion curves similar to the one shown in figure 9.2, which has been derived from observation of driver behaviour[2].

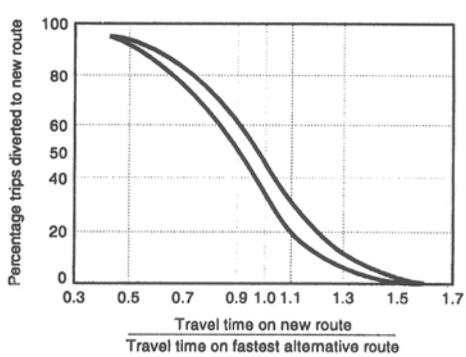


Figure 9.2 *Specimen highway diversion curves (adapted from ref. 2)*

In the USA indifference curves have been developed that take into account both time and distance savings. Figure 9.3 shows the form of the relationship proposed by Moskowitz[3] and based on observations of traffic diversion to new highways.

The mathematical form of the relationship is

$$\text{Percentage diverted to new route} = 50 + \frac{50(d + 0.5t)}{\sqrt{[(d - 0.5t)^2 + 4.5]}} \tag{9.1}$$

where d = distance saving on new route,
t = time saving on new route.

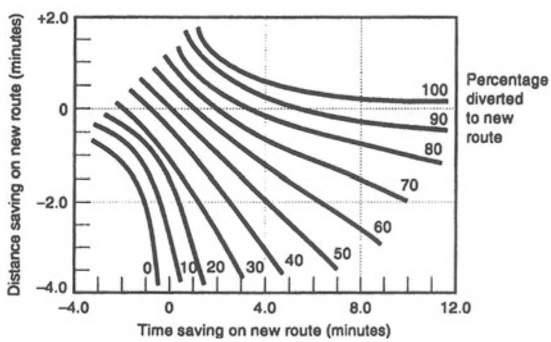


Figure 9.3 *Highway diversion curves based on time and distance savings (adapted from ref. 3)*

Drivers perceive costs but as they are able to measure speed and form an assessment of journey speed, speed can be used to assess the diversion of drivers to a new route. This may well be true of the present motorway system in the UK in that drivers are reasonably certain of being able to maintain a given speed but for certain journeys may not be aware if their journey time is greater on the new route. The diversion relationship shown in figure 9.4 developed for the Detroit Area Traffic Survey[4] illustrates the form of the assignment curves.

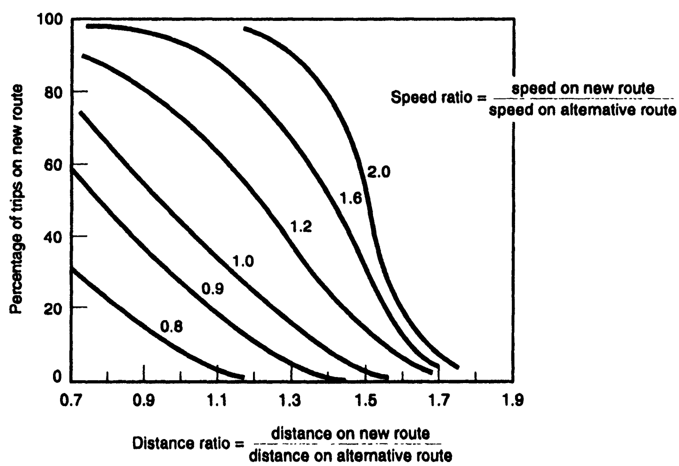


Figure 9.4 *Highway diversion curves based on distance and speed ratios (adapted from ref. 4)*

Several mathematical forms have been suggested for diversion curves and empirical curves can be derived. The *Traffic Assignment Manual*[5] gives an example of a diversion curve derived from the Burrell method where the proportion of traffic P_2 on route 2 is given by

$$P_2 = f(x)$$

where $f(x)$ is the normal integral

x is $\dfrac{(\bar{C}_1 - \bar{C}_2)}{I(\bar{C}_1 + \bar{C}_2)}$

$\bar{C}_1$ is the mean disutility on route 1 (usually travel time or generalised travel cost),

$\bar{C}_2$ is the mean disutility on route 2,

I is $V^2/3$,

and V is the Burrell variation term as a proportion of unity where V^2 has been found equal to 0.706 minutes from empirical evidence

Capacity-restrained assignment

All-or-nothing assignment results in a link with a favourable cost attracting a considerable number of trips, while links with unfavourable costs attract few trips. In practice this would result in the originally favourable link becoming overloaded, a situation which would not occur in real life.

The problem usually occurs when assigning trips to a highway network because in practice a balance exists between travel costs and flows. A number of routes between an origin and a destination are selected by tripmakers so that the perceived travel cost is approximately equal on all the routes. With a public transport network the choice of routes is more limited and the major difficulties occur when incorporating walking, waiting and interchange times.

Capacity-restrained assignment to highway networks attempts to tackle this problem by taking into account the relationship which exists between speed and flow on a highway. These relationships are discussed fully when considering the evaluation of transportation proposals.

The Department of Transport[5] recommends the use of capacity-restrained assignment when considering major road schemes when traffic congestion dissuades a significant number of drivers from using otherwise attractive routes. Because of the extra resources required they suggest the use of capacity restraint only when all-or-nothing or multi-routing would not adequately spread the traffic over the network.

Multipath proportional assignment

In urban areas there are many alternative routes between a given origin and destination and in actual fact tripmakers would be distributed over all these routes. This is because tripmakers will be unable to judge the route of least cost accurately and different tripmakers will make differing decisions. Multipath proportional assignment attempts to simulate this situation by assigning proportions of the trips between any two zones to a number of alternative routes.

Burrell[6] has described a model in which it is assumed that the tripmaker does not know the actual cost of using a link but that he associates a supposed cost with each link. This supposed cost is randomly generated from a distribution with the actual link cost as its mean and a given standard deviation representing the inability of a tripmaker to judge costs effectively. The assignment for one zone is then made on an all-or-nothing basis using the randomly determined link costs. Assignment of trips from the next zone considered will then be made on the basis of a further set of randomly determined link costs and the process repeated until all trips have been assigned to the network.

It has been reported that the assignment made by this model produces a satisfactory comparison with actual network flows giving more satisfactory results than all-or-nothing assignment.

Stochastic assignment with capacity constraint

This method combines the benefits of stochastic and capacity-constrained assignments. The procedure requires all trips to be initially assigned to the network using a stochastic assignment with fixed link costs. Current link volumes are updated by taking a weighted average of the flows between this and the previous iteration. Link costs are then adjusted to correspond to the new traffic flow and the process repeats in an iterative manner. Iterations continue until satisfactory compatibility between flows and speeds is achieved.

A problem which can arise with this, and other, iterative procedures is 'non-convergence', where traffic continues to switch between competing routes for successive iterations. Methods can be adopted to damp these oscillations down, but it can be difficult to choose a satisfactory end point.

Wardrop equilibrium assignment

Wardrop equilibrium assignment is commonly used in many of the computerised assignment procedures currently used in practice. In its usual form, it does not take account of drivers' different perceptions of times/costs. Its basis is that of route choice in which drivers are assigned to their minimum journey time/cost route taking account of the interaction between link flows and costs. This requires an iterative procedure to reach equilibrium.

Congested assignment models

Congested assignment models are now available and commonly used in practice. They include detailed modelling of link journey times, through speed/flow curves, and junction operations. In many congested networks, particularly in urban areas, junctions are the cause of bottlenecks and the main contributor to route journey times. It is therefore often important that assignment models contain an adequate representation of junction operations and delay. This requires modelling of different types of junction, including priority junctions, roundabouts and traffic signals with delays calculated using time-dependent queueing theory (see Parts II and III). Typical examples of models of this type are QVIEW[7]and TRIPS[8].

Two of the most detailed and comprehensive congested assignment models used in the United Kingdom are SATURN[9] and CONTRAM[10]. SATURN is a static equilibrium assignment containing linked simulation and assignment modules, enabling a good representation of co-ordinated traffic signal operations to be acheived. CONTRAM is a dynamic model incorporating a time-dependent demand and queueing process and an assignment based on the movements of individual vehicles or packets of vehicles. The structure of this model is illustrated in figure 9.5.

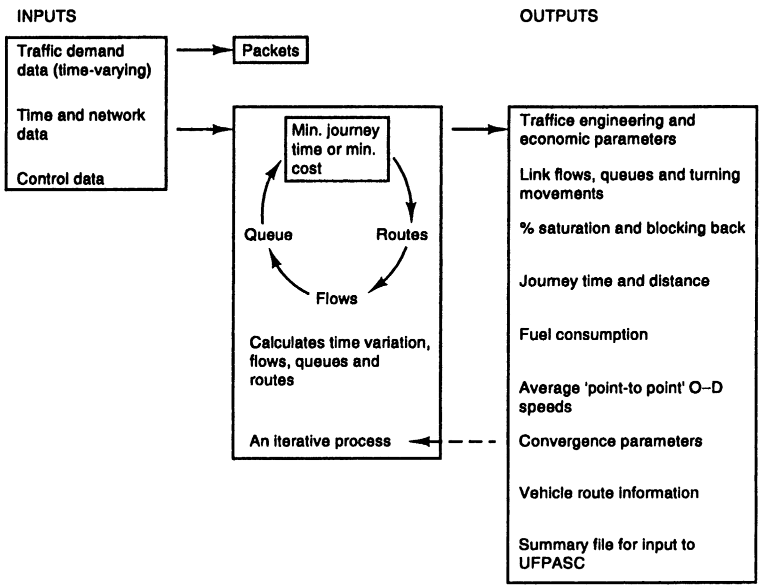


Figure 9.5 *CONTRAM – structure of the model (ref. 10)*

References

1. Road Research Laboratory, The London–Birmingham Motorway – traffic and economics, *Tech. Pap. Rd Res. Bd* 46, HMSO, London (1961)
2. R. E. Schmidt and M. E. Campbell, *Highway traffic estimation,* Eno Foundation for Highway Traffic Control, Saugatuck, Connecticut (1956)
3. K. Moskowitz, California method of assigning diverted traffic to proposed freeways, *Highw. Res. Bd Bull.* 130 (1956)
4. *Detroit Metropolitan Area Traffic Study*, Detroit, Michigan (1956)
5. Department of Transport, *Traffic Appraisal Manual,* London (1982)
6. J. E. Burrell, Multiple Road Assignment and its application to capacity restraint. Proceedings of the 4th International Symposium on the Theory of Traffic Flow, *Strassenbahn und Strassenverkehrstechnik* 86 (1969)

7. Wootton Jeffreys, *QVIEW Version 3.4 User Manual*, Wootton Jeffreys (1991)
8. MVA Systematica, *TRIPS Manual*, MVA Systematica (1982)
9. D. Van Vliet, SATURN – A Modern Assignment Model, *Traffic Engineering and Control*, 23, 575–81 (1982)
10. N.B. Taylor, CONTRAM 5 – An Enhanced Traffic Assignment Model, *Transport and Road Research Laboratory Research Report* 249 (1989)

10

The evaluation of transportation proposals

The object of the simulation of the land-use/transportation process is to estimate the trips that will be attracted to a proposed future transportation system for a given pattern of land-use development. To compare effectively different transportation proposals it is imperative that each proposal should be evaluated.

There are several grounds on which the end product of the simulation process must be judged. These are tabulated below:

(1) It is necessary to be certain that the numerical results output by the computer are realistic and all the computer programs are functioning correctly.
(2) It is necessary to check that the predicted future transportation requirements can be met by the proposed transportation system being evaluated.
(3) It is necessary to estimate the economic consequences of the provision and operation of the system under evaluation.
(4) It is necessary to consider the environmental effects of the operation of the proposed transportation system.

These sections will now be considered in greater detail.

Program checking to ensure that numerically correct results in accordance with the mathematical model are incorporated into the program is a very tedious process and even after experience with transportation programs doubt can sometimes be felt about the validity of the results obtained. A frequent source of error is incorrect input information and this should be carefully checked to ensure that the input data is correct.

The second stage in the evaluation of a transportation plan is to ensure that the proposed system is capable of dealing with the flows assigned to it by the simulation process. Many transportation simulation programs deal with 24-hour flows while critical flows usually occur in practice during the peak hours. It may be possible to estimate the peak hour flows from the 24-hour flows or alternatively the transportation programs may be run to deal with a peak hour flow.

Where highway networks are being evaluated it may be found that where there is no restriction on modal choice the highway system may be grossly overloaded. This problem may be overcome by the use of a modal choice relationship that incorporates a travel time by public transport ratio or by the arbitrary transfer of trips from the private transport to the public system.

Finally it should be remembered that the predictions of future trips on the transport system are dependent on the initial assumptions made. Where all-or-nothing assignment is used, for example, the network may have a section of highway where heavy traffic flows are predicted while an adjacent link, which has a longer travel time, may be lightly loaded. In practice the trips would be distributed between the adjacent links and predicted flows should be considered on the basis of corridors of movement.

In the past the environmental effects of proposed transport systems received less consideration than the effects of capacity or economic considerations. It is unfair to criticise transportation planners for ignoring environmental considerations in the past. Normally they were required to provide the least-cost solution and since damage to the environment had in the past not been given a high economic price, solutions that damaged the environment were preferred.

There are however considerable problems in placing a value on environmental factors and this has been partially responsible for the failure to consider them. The most important environmental effect at the present time is that of noise and although it is possible to predict the noise levels that will be caused by major highway proposals, the valuation of the effects of noise is less easy to make.

Transport facilities usually have an adverse effect on the environment: they cause noise and vibration, highway traffic pollutes the air and frequently intrudes both physically and visually. The disturbance to the environment must however be balanced against the result of doing nothing. A situation in which traffic is congested or there is a high accident risk may damage the environment to a greater extent than a new highway or other transport system.

The cost of reducing the damage to the environment caused by a new transport system may however be better spent providing alternative facilities that may improve the environment. This factor should be seriously considered at a time when any new transport proposal arouses intense opposition from those immediately affected.

Economic evaluation of transport schemes

It is however the economic consequences of carrying out transportation schemes which have received increasing attention during the last decade and it is becoming increasingly necessary to justify any new proposal on economic grounds.

This evaluation of a transport scheme on economic grounds is referred to as a cost–benefit analysis. As would be expected, the costs of carrying out a scheme are compared with the benefits that may be anticipated when the scheme is completed and in operation.

The usual method adopted in the United Kingdom for the economic evaluation of highway schemes when the benefits of the scheme occur over a considerable period of time is to discount the costs and benefits of the scheme so that they represent present-day values. In this way all the costs and benefits are converted to a net present value. Any scheme which has a positive net present value is economically worthwhile and when similar schemes are being compared on economic grounds, the scheme which has the greatest net present value is preferred.

Discounting is the reduction in value of future costs and benefits by dividing future values by $(1 + r)^n$ where r is the discount rate raised to the power n, the number of years hence when the cost or benefit is incurred. This means that when the discount rate is taken as 10 per cent and a benefit of £1000 will be obtained in 5 years' time, then the present value of this benefit is £621. If discounted benefits are considered positive and discounted costs are considered negative, then their sum over the period of time being considered is the net present value of the scheme.

For example, consider a proposal to upgrade the signal control at a highway junction at a cost of £15,000 and with associated future maintenance costs as shown in table 10.1. The benefits of upgrading the signal control will be a reduction in delay to vehicles passing through the junction and when this is valued it is estimated it produces the benefits also shown in table 10.1.

In table 10.1 the costs and benefits of the scheme are tabulated for a 10 year period; the discount rate selected is 7 per cent and the sum of the discounted net benefits is calculated as £42,747. As this is positive, the scheme is economically viable.

It is of course possible that with periodic maintenance the improvements to the signal control will have a life considerably longer than 10 years. It is usual in the United Kingdom to consider costs and benefits over a period of 30 years even though many highway works will have a residual value greater than this period of time. It is considered unimportant to consider costs and benefits beyond 30 years because of the uncertainty associated with long-term forecasts and also because

TABLE. 10.1
Calculation of net present value

Year of operation	*Capital and maintenance cost (£)*	*Value of annual reduction in delay (£)*	*Net benefit (£)*	*Discounted net benefit (%) (£)*
1	25 000	1 500	−23 500	−21 963
2	2 000	3 000	1 000	873
3	2 000	5 000	3 000	2 449
4	4 000	9 000	5 000	3 815
5	2 000	13 000	11 000	7 843
6	2 000	16 000	14 000	9 329
7	2 000	19 000	17 000	10 587
8	4 000	20 000	16 000	9 312
9	2 000	21 000	19 000	10 335
10	2 000	22 000	20 000	10 167
				42 747

the present value of a £1 cost or benefit in 30 years time is only 13p when discounted at 7 per cent.

There are usually several alternative schemes which can be proposed to alleviate transport problems, each with differing costs and associated benefits. When a choice is made purely on economic considerations, then the scheme with the highest net present value would normally be preferred.

Cost–benefit analysis applied to United Kingdom trunk road investments

Cost–benefit analysis was first used in the assessment of trunk road schemes in the United Kingdom in the 1960s; later, in the 1970s, the Department of Transport developed the computer cost–benefit analysis program COBA[1] where the benefits derived by road users are calculated and expressed in monetary terms.

When a scheme does not have a marketable output, as is the case when users are not charged directly for road travel, then cost–benefit analysis may be applied. It is not possible to value all the costs and benefits of a highway scheme and in most instances the benefits of a highway project have been limited to time saving benefits, savings in operating costs, and savings in accident costs. These benefits are related to capital and maintenance costs. As has been previously discussed, costs and benefits occur over a period of time and to relate them to the present time, discounting is used. In the valuation of costs and benefits, it is resource costs which should be used with taxes and subsidies, which are merely transfer payments, eliminated.

The first stage in the COBA appraisal program is to determine the alternative schemes which are to be subjected to cost–benefit analysis. Because of the pressures of public opinion there will normally be several alternatives for any road proposal, the minimum number is of course two: the 'Do-Minimum' and the 'Do-Something'.

The 'Do-Minimum' option is frequently the situation where road improvements are not to be carried out and traffic will have to travel along existing facilities. There are many cases however when some improvements will be carried out regardless of whether the scheme being appraised is constructed. An example could be where improvements to a signalised junction will provide better pedestrian facilities and will be carried out regardless of major road improvements. A further example might be where a relatively small highway improvement or low-cost traffic management would considerably improve traffic conditions; in this case the 'Do-nothing' option would offer an unrealistic alternative to the scheme being assessed.

When a highway improvement is carried out it will result in changes in the trip pattern. Existing traffic may take a new route; that is, reassignment of the trip pattern will take place. Existing traffic may change its destination because of the convenience of the new road; that is, redistribution may take place. New traffic may be attracted on to the new system because of the existence of the new road; that is, generation of traffic may take place. Trips may be attracted on to the new road which previously went by another mode of travel; that is, a modal split. Lastly, trips may be made at another time of day compared with the 'Do-Minimum' situation.

Because COBA deals with highway schemes whose effects may be largely restricted to a corridor of travel, only the effect of redistribution is considered. Where the road scheme is in a congested urban area, or if a complete long inter-urban route or an estuarial crossing is being appraised, then the other factors may have an important effect on the resulting traffic pattern.

The major benefit arising from the construction of a road scheme is reduction in travel time for vehicles travelling along links and through junctions. These time savings have to be assigned a monetary value so that they may be compared with construction and maintenance costs. The value of a time saving depends upon the purpose for which the trip was being made. The COBA program recognises two trip purpose and values time as 'working time' and 'non-working time'. As many trips in peak hours are for travel to and from work, this is often an important element of time saving; in the COBA program. it is classified as 'non-working time'.

Working time is valued as the cost to an employer of the travelling employee; it consists of the gross wage rate together with overheads. Variations in wage rates among the population of car drivers, car and bus passengers are taken into account by the use of a mileage-weighted average income taken from national statistics. The value of 'non-working time' has been derived from studies of the choice made by travellers between a slower cheaper mode of travel or a faster more expensive mode. These studies indicate that on average the value of in-vehicle non-working time is equivalent to 25 per cent of the hourly wage rate.

It has frequently been argued that small time savings do not have the same unit value as large time savings resulting from major road schemes. In the COBA program all time savings have the same unit regardless of their duration.

The COBA program calculates time user costs for the 'Do-Minimum' and the 'Do-Something' options by disaggregating the traffic flows by vehicle category and by flow group; the flows are then further divided into work and non-work journeys. The time cost per vehicle is a function of vehicle occupancy which varies with vehicle type and journey purpose. The hourly flow of vehicles is then converted into a flow of people whose time can be valued to produce time costs for links and junctions.

Savings in vehicle operating cost between the 'Do-Something' and the 'Do-Nothing' options are one of the benefits of road improvement schemes. They depend on link distance and on link speed; usually they are positive but depending upon the relative changes in distance and speed may possibly be negative. In the COBA program, vehicle operating cost is related to distance travelled and is a function of fuel, oil, tyres, maintenance and depreciation and the size of vehicle fleets. The resource cost of fuel consumption depends on average link speed and the hilliness of the link; the relationship gives a high cost at low speeds as would be expected with stop–start conditions.

The marginal resource costs of oil and tyres are a fixed cost per kilometre, maintenance is in part related to link speed and in part varies with speed.

Depreciation for vehicles, other than cars, is entirely related to mileage; for cars it is related to the passage of time as well as to mileage; only the latter is included in car vehicle operating cost. Increased link speeds may result in a reduction in the number of working cars, commercial vehicles and public service vehicles which operators need to provide; this is taken into account in vehicle operating cost.

An important element in the benefits which arise from road schemes is the change in the number and severity of accidents consequent on the change from the 'Do-Minimum' to the 'Do-Something' option. To allow accident reduction to be considered in the COBA program it is necessary to place a monetary value on accidents. It is assumed that the cost of an accident is composed of three elements. Firstly, the direct costs to individuals and organisations involved in the accident, that is, vehicle damage, police and medical costs. Secondly, the lost output of those killed or injured measured by the expected loss of earnings together with non-wage payments such as insurance contributions paid by an employer. Thirdly, an allowance is made for the 'pain, grief and suffering' which results from injury or death in a road accident; its value in COBA is stated to be largely notional and minimal.

Accident severity has been found to vary with road type because of differences in the number of casualties per accident and in the amount of damage per accident. As would be expected an accident on a rural road will result in more casualties per accident and greater damage than on an urban road where speeds may be expected to be lower.

With the monetary value of accidents having been established, it is necessary to know the accident rate for the particular road types being considered in the 'Do-Minimum' and the 'Do-Something' situations. The only complete records of road accidents in the United Kingdom are those where personal injury takes place, and for this reason accident rates in COBA are expressed as personal injury accidents per million vehicle kilometres and an allowance made in the valuation of these accidents for damage only accidents. It is preferable to separate accidents which take place on links between junctions and those which take place at or adjacent to the junctions. For accidents at junctions, COBA uses relationships between junction flows and accidents for various junction types. When the 'Do-Minimum' option is being considered and little change is contemplated, then this option becomes the 'Do-Nothing' option and existing accident data is preferred. For the 'Do-Something' option then, predicted values based on past accident experience on similar road types must be used.

The benefits of the road scheme in the COBA program are taken as the monetary value of the reduction in travel time and accidents and usually the reduction in vehicle operating cost, and these are compared with construction cost, delays to traffic during construction and the changes in maintenance cost due to the construction of the new road. The comparison between costs and benefits is made for a 30 year period on the basis of their present value.

The effects of redistribution, generation and modal split on economic assessment of road schemes

As has been previously stated, the COBA program only considers the redistribution of trips within the highway network assuming that the reduction in travel cost will not cause any change in the demand for travel. In many cases this will be incorrect and additional trips will be generated and a change in travel mode could take place if the scheme is of considerable magnitude. In addition, any increase in traffic flow will lead to a reduction in journey speed, an increase in travel time and an increased trip cost. These effects on the trip matrix are discussed in Chapter 27.

When economic assessment of a road scheme where these effects are considered important is carried out, complexities arise because car travellers usually perceive the cost of a trip to be less than the total user cost, normally only considering fuel costs which include taxation costs but which are excluded in cost–benefit analysis as being a transfer payment. The car traveller frequently ignores non-fuel costs and these have to be included in an economic assessment.

Urban road economic assessment

COBA was developed for the evaluation of interurban road schemes. Link traffic flows for input into COBA are usually obtained from the previous application of a traffic assignment model to the 'Do-Nothing' and 'Do-Something' network situations. COBA then recalculates link and junction delays and costs for input into its own economic evaluation processes. The application of COBA to urban areas is usually less satisfactory than in interurban situations, because the traffic modelling in COBA is less detailed than in most assignment models so that incompatibility can arise in delay estimates between the assignment model and COBA with COBA usually being less accurate. This is particularly the case in congested urban networks where small changes in flow can have a large effect on delays and costs.

The United Kingdom Department of Transport now recommends the use of the URECA[2] model for *U*rban *R*oad *E*conomic *A*ssessment. With this process, economic calculations from COBA are carried out directly on the traffic flow outputs from the assignment models for the forecast years. This avoids the intermediate step of journey time/delay calculations by COBA and the incompatibilities caused by this.

Forecasting future vehicle ownership and use

In the assessment of road schemes it is necessary to take into account the increase in the number of vehicles which can be expected in the future. In particular it is the growth in the number and use of private cars which has caused problems to the transport planner and the highway engineer working in urban areas. The rapid growth in the number of private cars in the United Kingdom during the 1950s and 1960s led to research on the prediction of vehicle ownership and use and in particular car ownership and use.

The earliest methods of forecasting car ownership were based on logistic growth curves of cars per head. In 1965 Tanner[3] postulated that, except for tractors, the forecasts of vehicles per head should lie on a logistic curve determined by the 1964 levels of ownership and future saturation levels of ownership when further increases in car ownership per person ceased.

If the logistic relationship is assumed, when the percentage increase per year is plotted as the ordinate and the corresponding cars per head is plotted as abscissa then a linear relationship is observed and the value of car ownership when the percentage increase per year is zero is the saturation level. Using data available at that time, Tanner estimated that the saturation level in Great Britain would be 0.45 cars per head with an alternative estimate of 0.40 if a moderately restrictive attitude to motoring was taken. With the logistic relationship defined, car ownership could be determined for any year in the future and when multiplied by population estimates produced future levels of car ownership. Car traffic was considered to increase with car ownership as it was assumed that average annual car travel would remain constant.

Such a simple relationship was criticised in that it took no account of economic or social circumstances. As a consequence and shortly after the economic changes of 1973, Tanner presented a revised model of car ownership[4] in *Transport and Road Research Laboratory Report* LR650 which although still adopting a logistic relationship introduced the economic variables gross domestic product and fuel prices. At the same time the kilometres per car per year were assumed to vary with gross domestic product and the length of the motorway network. This report introduced variability into forecasting by assuming gross domestic product growth rates of 2, 3 and 4 per cent and saturation car ownership levels of 0.42, 0.43 and 0.44 cars per head.

In response to continued criticism that car ownership and car travel did not fully take into account the underlying causes of growth and travel, Tanner proposed in 1977[5] a power growth model where the second half of growth takes longer to develop than the first half of the curve. The factors which determined growth in this model were the passage of time, income in terms of gross domestic product per person, and the cost of motoring. A range of saturation levels was used: 0.44, 0.46 and 0.48; gross domestic product growth rates of 2, 3 and 4 per cent; and fuel price rises of 4, 2 and 0 per cent.

At this time there was a growing interest in models which used the factors which predicted car ownership and travel rather than models which depended on the passage of time. A causal model of this form was proposed in Bates[6] in 1978 where models for the proportions of households with at least 1 car and for households with at least 2 cars were proposed in terms of the ratio of household income to the index of car purchase costs, the latter acting as an income deflator. This model was developed from information in the Family Expenditure Survey of Great Britain and the relationship was found to be reasonably constant over the period 1965–75. In the period 1975–78 however car prices rose more rapidly than real gross household income and the model indicated that car ownership would fall, but in fact car ownership continued to increase and the model underestimated the 1978 level by 16 per cent.

Because of difficulties with the Bates model the Department of Transport in their *National Road Traffic Forecasts*, 1984[7] adopted a model which related the probability of ownership of at least 1 car per household and at least 2 cars per household to income deflated by the retail price index and also by the number of driving licences per adult. When making the 1984 forecasts of road traffic the Department of Transport also considered the forecasts of the LR799 model but by using additional data was able to obtain new estimates of the relationship between car ownership, income and motoring costs based on time-series data. The new estimate of income elasticity is in a range of 0.15 to 0.18 (previously assumed to be 0.35 to 0.41) and the new estimate of elasticity with respect to motoring costs is in the range −0.13 to −0.17 (previously −0.22 to −0.27) indicating that the effect of time is more important than was previously assumed.

The environmental evaluation of transport schemes

Major transport proposals require much more than an engineering and economic appraisal and it is well established that consideration needs to be given to the environmental effects of projects. In addition, the increasing necessity for public consultation during the initial planning stage of transport schemes, and the emphasis placed on environmental issues by protest groups, have made the assessment of environmental planning increasingly important.

Environmental evaluation in the United States

Environmental issues became important in the United States during the 1960s and a procedure suggested by Leopold *et al.*[8] has been modified for use in the United States.

Leopold stated that an assessment of the probable impacts of the variety of the specific aspects of the proposed action upon the variety of existing environmental elements and factors should consist of three basic elements:

(a) a listing of the effects on the environment which would be caused by the proposed development and an estimate of the magnitude of each;
(b) an evaluation of the importance of each of these effects;
(c) the combining of magnitude and importance estimates in terms of a summary evaluation.

For each scheme being assessed these three elements are analysed using a matrix, on one axis of which are the actions which cause environmental impact and on the other axis are existing environmental conditions which might be affected.

There are a total of 100 environmental actions which may cause environmental impact, grouped into modifications of the regime, land transformation and construction, resource extraction, processing, land alteration, resource renewal, changes in traffic, waste emplacement and treatment, chemical treatment, and accidents. Listed vertically are 88 environmental characteristics grouped into physical and chemical characteristics, biological conditions, cultural factors, and ecological relationships.

For each action which is expected to have a significant interaction with an environmental condition the relevant square in the matrix is divided by a diagonal which runs from the upper right to the lower left.

After all anticipated actions have been checked with possible environmental effects, a decision is made as to the magnitude and also the importance of the interaction. Within each box, representing a significant interaction, a weighting factor is introduced ranging from 1 to 10. The number is placed in the upper left-hand corner to indicate the relative magnitude of the interaction and in the lower right-hand comer to indicate the relative importance of the interaction. A value of 1 indicates the least magnitude or importance of the interaction and a value of 10 the greatest magnitude or importance.

For an example, the circular considers the case of an engineering proposal which requires the construction of a highway and a bridge. The action will have environmental effects which may be classified under erosion, deposition and sedimentation. It may be that, because of poor consolidation of the soil in the region of the bridge, erosion is likely to be considerable and so the magnitude of the impact will be weighted with a factor of 6 or more. If, however, the river already carried large sediment loads and further erosion will not have undesirable effects, then the weighting of the importance of the interaction may be 2 or less.

If possible, assignment of numerical weights to the magnitude and importance of an interaction should be based on factual data. When all significant interactions have been considered then a simplified matrix can be presented containing the product of the magnitude and importance weightings, beneficial interaction products being shown as positive and detrimental interaction products shown as negative values.

Environmental assessment in the United Kingdom

Procedures for the environmental assessment of highway schemes in the United Kingdom are set out in Volume 11 of the Department of Transport's *Design Manual for Roads and Bridges*. This document, entitled 'Environmental Assessment'[9] supersedes the early DOT *Manual of Environmental Appraisal*[10]. This manual describes the assessment procedures for air quality; cultural heritage; disruption due to construction; ecology and nature conservation; landscape effects; land use; traffic noise and vibration; pedestrians, cyclists, equestrians and community effects; vehicle travellers; water quality and drainage; geology and soils; and impact of road schemes on policies and plans.

The Environmental Assessment manual describes how each of these potential effects of new road schemes can be quantified and assessed. This involves the use of environmental impact tables in a framework. This is simply the tabular presentation of data summarising the main likely direct and indirect impacts on people of the alternative options for a proposed highway scheme[11]. For trunk road scheme appraisal the Department of Transport states that each option for the construction of the road, which is technically feasible and considered to be an acceptable and clearly separate solution, will have a separate column in the framework. The minimum number of columns will be one containing details of the officially proposed route, one containing the forecast changes if a new route is not constructed (that is, the 'Do-Nothing' or the 'Do-Minimum' solution), and finally a column for comments.

For the data presented in the framework to be consistent from scheme to scheme it is recommended that information be presented in the following groups: the effects on users of facilities, the effects on policies for conserving and enhancing the area, the effects on policies for development and transport, and finally financial effects.

In the first group are the effects on travellers; defined as vehicle drivers and passengers, cyclists and pedestrians; in terms of time savings or delays, changes in vehicle operating costs, accident reductions, driver stress, view from the road, and amenity and severance.

In the second group, the effects on occupiers of property, occupiers are subdivided into occupiers of residential property, industrial and commercial property, schools and hospitals, public or special buildings, recreational space, and agricultural land. The impacts on occupiers which should be considered are: number affected by demolition, changes in noise level, visual effects, severance, disruption during construction, and landtake.

In the third group, users of facilities are subdivided into users of: shopping centres, public buildings such as churches, libraries or community centres, and users of recreational areas and facilities.

The impacts described in the fourth and fifth group are the formal views of the highway authority themselves, while in group six the economic costs and benefits are presented for each option in terms of the net present value.

The impact of traffic noise

Noise levels generated by highway traffic can be measured or calculated and the subject is considered in detail in Chapter 25. It is however the reaction of human beings to noise levels which is of importance in attempting to determine the impact of the noise. Different people have different reactions to the same noise level and it is necessary to determine the distribution of responses to noise by the use of attitude surveys.

Social surveys have been carried out in which respondents were asked to give their reactions to traffic noise levels experienced at home. Using a seven-point scale, they were asked to express their reactions. A score of 1 was labelled 'definitely satisfactory' and a score of 7 was labelled 'definitely unsatisfactory'.

It was found[12] that a family of curves could be derived, of the form

$$p_s = a + \frac{1}{(b + e^{-c(L-d)})}$$

where p_s is the proportion of people with scale score $\geqslant S$ at noise level L,
a, b, c, d are constants depending upon the value of S,
L is the 18 hour L_{10} noise level (dB(A)).

Score	Equation for proportion scoring as shown
$S > 1$	$p_1 = 1.0$
$S > 2$	$p_2 = 0.2 + \dfrac{1}{(1.25 + \exp[-0.104(L - 56)])}$
$S > 3$	$p_3 = 0.1 + \dfrac{1}{(1.11 + \exp[-0.116(L - 63)])}$
$S > 4$	$p_4 = 0.08 + \dfrac{1}{(1.09 + \exp[-0.129(L - 68)])}$
$S > 5$	$p_5 = 0.06 + \dfrac{1}{(1.06 + \exp[-0.144(L - 73)])}$
$S > 6$	$p_6 = 0.02 + \dfrac{1}{(1.02 + \exp[-0.113(L - 78)])}$

$$S > 7 \qquad p_7 = 0.0 + \frac{1}{(1.0 + \exp[-0.096(L - 86)])}$$

Using these equations, the proportions in various scale score ranges can be generated. If the numbers of people within the noise interval of which L is the mid-point are known, then the numbers with scale scores $> S$ at L can be calculated. To obtain the number of people with scale score n, the number with a scale score $\geqslant n+1$ is subtracted from the number with scale score $\geqslant n$

To aggregate these scale scores over an area subject to highway noise impact it is necessary to make a decision as to how they should be weighted. It is suggested that a scale score of 2 ('not at all bothered by the noise intrusion') represents a neutral attitude and that a progressively less favourable attitude is indicated by higher scores. The intensity-weighted scale score ranges from +1 (basic scale score =1) to −5 (basic scale score = 7). These weightings allow the weighted estimation of the number of people affected by a proposed highway scheme to be determined.

It is necessary to combine noise impacts for different land uses where these vary. It is suggested that the following procedure be used:

(1) Where the land use is such that speech communication is important (such as schools), 5 dB(A) should be added to the anticipated noise level and the distribution of scale scores for residential use should be applied to this modified noise level, together with a duration weighting (see below).
(2) Similarly, where a low level of noise is desirable (such as religious buildings, parkland), 10 dB(A) should be added to the anticipated noise level and a duration weighting applied.
(3) When considering pedestrians, 10 dB(A) is added to the anticipated noise level, because pedestrians receive a noise impact which is not attenuated by a building structure. As before, a duration weighting is applied.

Duration weightings for length of exposure to noise in various situations are as follows:

Residential land use	1.0
Educational establishment	0.5
Hospitals	1.5
Churches	0.2
Parks	0.2
Pedestrians	0.1

A measure of the noise impact of a highway proposal may thus be obtained by summing the products of the people affected by noise impact, their intensity-weighted scale score and the appropriate duration weight.

As an example of the above procedure consider a proposed highway scheme which results in a residential area containing 600 people being submitted to an 18 hour L_{10} noise level in the range 64–66 dB(A). In a school 300 children and staff will be subjected to an 18 hour L_{10} noise level in the range 62–64 dB(A) and a church with an average attendance of 200 people will be subjected to an 18 hour

L_{10} noise level in the range 64–66 dB(A). Calculate the intensity and duration-weighted number of people exposed to noise impact. Considering the residential development where the 18 hour L_{10} noise level is in the range 64–66 dB(A):

The proportion of people with a scale score > 1 is 1.0

The proportion of people with a scale score > 2

is $$0.2 + \frac{1}{(1.25 + \exp[-0.104(L - 56)])}$$

$$0.2 + \frac{1}{(1.25 + \exp[-0.104(65 - 56)])}$$

$$= 0.81$$

The proportion of people with a scale score > 3

is $$0.1 + \frac{1}{(1.11 + \exp[-0.116(L - 63)])}$$

$$0.1 + \frac{1}{(1.11 + \exp[-0.116(65 - 63)])}$$

$$= 0.63$$

The proportion of people with a scale score > 4

is $$0.08 + \frac{1}{(1.09 + \exp[-0.129(L - 68)])}$$

$$0.08 + \frac{1}{(1.09 + \exp[-0.129(65 - 68)])}$$

$$= 0.47$$

The proportion of people with a scale score > 5

is $$0.06 + \frac{1}{(1.06 + \exp[-0.144(L - 73)])}$$

$$0.06 + \frac{1}{(1.06 + \exp[-0.144(65 - 73)])}$$

$$= 0.30$$

The proportion of people with a scale score > 6

is $$0.02 + \frac{1}{(1.02 + \exp[-0.113(L - 78)])}$$

$$0.02 + \frac{1}{(1.02 + \exp[-0.113(65 - 78)])}$$

$$= \quad 0.21$$

The proportion of people with a scale score > 7

is $$0 + \frac{1}{(1.0 + \exp[-0.096(L - 86)])}$$

$$0 + \frac{1}{(1.0 + \exp[-0.096(65 - 86)])}$$

$$= \quad 0.12$$

The proportions with scale scores equal to 1, 2, 3, 4, 5, 6 and 7 are calculated by subtraction and are entered in table 10.2. In a similar manner the proportions of people in the school and the church are calculated and also placed in table 10.2. The intensity and the duration of the weightings are also placed in the table and finally by multiplication and by addition the number of people affected weighted by intensity and duration is obtained. Similar calculations can be made for other highway alignments and a comparison made between the noise impact of the options which are being evaluated.

A simpler approach is suggested by the Department of Transport in the *Manual of Environmental Appraisal* where it is suggested that in the preliminary stages of scheme preparation it is sufficient to list the number of properties that are wholly or partly within distance bands of 0–49, 50–99, 100–99 and 200–300 m from the road centre line of rural roads. For urban roads where shielding from noise due to adjacent properties and the distorting effects of gaps between properties and reflection will contain the noise only, properties within 0–49 and 50–99 m should be listed.

The visual impact of transport works

There can be little doubt that major transport works, and particularly urban highways, cause a considerable change in the visual scene. In some circumstances and to some people this causes a considerable loss of amenity.

In all aspects of transport, economic considerations are becoming increasingly important. Thus the cost of installation and operation of any transport scheme has to be weighed against alternatives, such as schools or welfare facilities, to which the resources could be allocated. In such a situation the extra cost of avoiding impact has to be balanced against the value of the amenity which is lost.

While the cost of improving the visual image of transport works or reducing impact by alternative routes or by tunnelling can be readily calculated, the reduction in impact which results has been difficult to estimate.

In the context of trunk roads the Standing Advisory Committee on Trunk Road Assessment[13] distinguished two types of visual impact. When a road structure

TABLE 10.2

Calculation of intensity and duration-weighted number of people subjected to noise impact

	Scale score S	*Proportion with scale score S*	*No. of people*	*Intensity weighting*	*Duration weighting*	*Weighted number of people*
Residential development	1	0.19	114	+1	1.0	+114
	2	0.18	108	0	1.0	0
	3	0.16	96	−1	1.0	−96
	4	0.17	102	−2	1.0	−204
	5	0.09	54	−3	1.0	−162
	6	0.09	54	−4	1.0	−216
	7	0.12	72	−5	1.0	−360
School	1	0.15	45	+1	0.5	+22
	2	0.16	48	0	0.5	0
	3	0.13	39	−1	0.5	−20
	4	0.18	54	−2	0.5	−54
	5	0.12	36	−3	0.5	−54
	6	0.11	33	−4	0.5	−66
	7	0.15	45	−5	0.5	−112
Church	1	0.08	16	+1	0.2	+4
	2	0.08	16	0	0.2	0
	3	0.09	18	−1	0.2	−4
	4	0.14	28	−2	0.2	−11
	5	0.18	36	−3	0.2	−22
	6	0.17	34	−4	0.2	−27
	7	0.26	52	−5	0.2	−52
						−1320

obstructs the view then this impact was referred to as visual obstruction. The more subjective effect of the road on the landscape was referred to as visual intrusion.

Some initial research into the measurement of visual obstruction has been carried out by Hopkinson[14] who has proposed a tentative method by which an assessment of visual obstruction can be made.

It is possible to measure some forms of obstruction without too much difficulty. If, for example, a large bridge structure is erected immediately in front of a dwelling house, then the loss of daylight is a reasonable measure of the obstruction. In a great many cases, however, there is no loss of daylight, and it is simply the view which is changed. Where a new highway or public transport interchange replaces derelict urban property, many residents may consider the intrusion an improvement in amenity. In a suburban area the same construction, requiring the demolition of residential properties, would usually be considered to be an intrusion with a considerable loss of amenity.

When considered in this way it would seem that the problem defies rational solution but the problem of noise impact presents similar difficulties. The effect of intruding noises varies according to the frequency and level of the intrusion and in this situation measurement scales have been calibrated against human response, as obtained from social surveys.

An attempt to assess the visual consequences of transport works was first made by Hopkinson, who used the principle that the extent of visual obstruction can be quantified by the amount of the field of view taken up by the obstruction. A unit by which this field of view may be measured is the solid angle expressed in terms of the steradian, the angle subtended at the centre of a sphere of unit radius by a unit area on its surface. In practice the millisteradian (msr) is the unit used.

An additional factor which must be considered in assessing the obstruction of an element in the field of view is the position of the object in the visual field. This is measured by the position factor: an object subtending a given angle at the centre of the field of view is considered to be more obstructive than an object subtending the same angle near the periphery of the visual field. The central zone up to 6° from the centre of the field of view has the greatest significance and is given a weighting of 100; between 6° and 20° the position factor is 30; from 20° to 50°, the outer limit of which covers most of the binocular field, the position factor is 10. From 50° to 90° is that part of the visual field seen by only one or other of the eyes, and has a position factor of 1. The solid angular subtense of an object may then be multiplied by the position factor to give a weighted solid angle.

For the simplified case of a straight elevated highway or railway embankment, or similar structure, with a uniform height H above the surrounding ground, it can be shown that the solid angle subtended by the transport facility is

$$S_\theta = \frac{H}{d_p}(\cos\theta_1 - \cos\theta_2)$$

where d_p is the perpendicular distance in plan from the point at which the visual obstruction is being measured to the line of the transport facility; θ_1 and θ_2 are the angles subtended between the rays to the limits of vision of the works and a datum line parallel to the road. This is only approximately correct; where the viewpoint is close to the intrusion, so that the vertical angle subtended at the viewpoint is greater than approximately 30°, a spherical correction is required.

The height H of the works depends largely on the height of the structure when it is elevated, but in the case of depressed facilities the height may be the barrier wall or, for open cuttings, the perceived height of the open excavation.

Determination of the solid angle may be made by calculation after measurement from plans and sections of the proposed works, by the use of special full-field cameras or by the use of specially devised protractors which make the employment of non-specialist staff possible.

Calculation from measurements of H, d_p, $\cos\theta_1$ and $\cos\theta_2$ made on plans and sections is tedious. Where a preliminary impact analysis is being carried out it may be sufficient to calculate the solid angle subtended by the central 40° of the field of view and ignore the effect of the position factor.

The impact of air pollution

Motor vehicles propelled by engines using petrol, petroleum gas or diesel as a fuel emit a wide range of gaseous and particulate materials some of which have a potential to be harmful to human beings, the amount of the pollution depending on the condition and type of the engine and on operating conditions.

There is increasing public disquiet regarding the effects of motor vehicle air pollution and this has resulted in strict control of exhaust emissions in the United States. From 1971 progressively more stringent regulations have been introduced in Europe governing the emissions of passenger cars. These regulations cover carbon monoxide, hydrocarbons and oxide of nitrogen levels emitted under standard test conditions. Over the same period the total amount of lead emitted has been reduced by controlling the proportion added to petrol as a means of increasing engine efficiency.

Public reaction to air pollution likely to be caused by proposed highway schemes is often expressed by residents of communities adjacent to the line of new road schemes. This is a growing problem as many highway proposals are, in the future, likely to be in urban areas.

This concern has been noted by the Advisory Committee on Trunk Road Assessment[13] which recommended that "where air pollution is likely to be a problem a special air quality report should be prepared: otherwise it should be excluded from the assessment." This view was endorsed by the Standing Advisory Committee on Trunk Road Assessment, which commented that the impact of air pollution should be assessed where it is a particular problem.

For these reasons the Department of Transport in their *Manual of Environmental Appraisal*[10] recommend that where appropriate an air quality report should be included in the environmental assessment of alternative proposals. The subject of air pollution due to road traffic is considered in detail in Chapter 26.

References

1. Department of Transport, *The COBA Manual*, London (1987)
2. Department of Transport, *URECA Version 2.3, Urban Road Economic Assessment*, London (1992)
3. J. C. Tanner, Forecasts of vehicle ownership in Great Britain, *Roads and Road Construction*, **42** (1965), 515, pp 341–7; **43** (1965), 516, pp 371–6
4. J. C. Tanner, Forecasts of vehicles and traffic in Great Britain, 1974 revision, *Transport and Road Research Laboratory Report* LR 650, Crowthorne (1974)
5. J. C. Tanner, Car ownership trends and forecasts, *Transport and Road Research Laboratory Report* LR 799, Crowthorne (1977)
6. Departments of the Environment and Transport, A disaggregate model of household car ownership, *Research Report* 20 (1978)
7. Department of Transport, *National Road Traffic Forecasts* (1984)
8. L. B. Leopold, F. E. Clarke, B. B. Hanshaw and J. R. Balsey, A procedure for estimating environmental impact, *U.S. Dept. of Interior Geological Survey Circular No. 645* (1971)
9. Department of Transport, *Environmental Assessment Design Manual for Roads and Bridges*, Vol. 11, London (1993)
10. Department of Transport, *Manual of Environmental Appraisal*, London (1983)
11. Department of Transport, Frameworks for trunk road appraisal, *Departmental Standard* TD/12/83, London (1983)
12. Department of the Environment, *The Environmental Evaluation of Transport Plans*, HMSO, London (1976)

13. Department of Transport, *Report of the Advisory Committee on Trunk Road Assessment. Chairman: Sir G. Leitch*, HMSO, London (1977)
14. R. G. Hopkinson, The evaluation of visual intrusion in transport situations, *Traff. Engn. Control*, 387–91, 395

PART II

ANALYSIS AND DESIGN FOR HIGHWAY TRAFFIC

11

The capacity of highways between intersections

The capacity of a highway may be described as its ability to accommodate traffic, but the term has been interpreted in many ways by different authorities. Capacity has been defined as the flow which produces a minimum acceptable journey speed and also as the maximum traffic volume for comfortable free-flow conditions. Both these are practical capacities while the Highway Capacity Manual[1] defines capacity as the maximum hourly rate at which persons or vehicles can reasonably be expected to traverse a point or uniform section of a lane or roadway during a given time period under prevailing roadway, traffic and control conditions. The time period used in most capacity analysis is 15 minutes which is considered to be the shortest interval during which stable flow exists.

Highway capacity itself is limited by:

(1) The physical features of the highway, which do not change unless the geometric design of the highway changes.
(2) The traffic conditions, which are determined by the composition of the traffic.
(3) The ambient conditions which include visibility, road surface conditions, temperature and wind.

A term used in the Highway Capacity Manual to classify the varying conditions of traffic flow that take place on a highway is 'level of service'. The various levels of service range from the highest level, which is found at a flow where drivers are able to travel at their desired speed with freedom to manoeuvre, to the lowest level of service, which is obtained during congested stop–start conditions.

The Highway Capacity Manual approach

The level of service afforded by a highway to the driver results in flows that may be represented at the highest level by the negative exponential headway distribution when cumulative headways are being considered, by the double exponential distribution as the degree of congestion increases, and by the regular distribution in 'nose-to-tail' flow conditions.

To define the term level of service more closely, the Highway Capacity Manual gives six levels of service and defines six corresponding volumes for a number of highway types. These volumes are referred to as maximum service flow rates and may be defined as the maximum hourly rate at which persons or vehicles can reasonably be expected to traverse a point or uniform section of a lane or roadway during a given time period under prevailing roadway, traffic and control conditions while monitoring a designated level of services.

Levels of service are based on one or more operational parameters which describe operating quality for a particular type of facility. These parameters are referred to as measures of effectiveness and differ according to the type of facility, as detailed in table 11.1.

TABLE 11.1
Measures of effectiveness used for determination of level of service[1]

Facility	*Measure of effectiveness*
Freeways	
basic freeway segments	Density (pc/mi/ln)
weaving areas	Average travel speed (mph)
ramp junctions	Flow rates (pcph)
Multi-lane highways	Density (pc/mi/ln)
Two-lane highways	Percentage time delay (%) Average travel speed (mph)
Signalised intersections	Average delay (s/veh)
Unsignalised intersections	Reserve capacity (pcph)

Freeway operating characteristics include a wide range of rates of flow over which speed is relatively constant. For this reason, speed alone is not an adequate measure of performance for the definition of level of service. Drivers however are sensitive to the freedom to manoeuvre and proximity to other drivers, traffic flow characteristics which are related to the density of the traffic flow. Importantly, rate of flow is related to density throughout the range of traffic flow. For this reason density is now the parameter used to define level of service for basic freeway segments.

Descriptions of operating conditions for the six levels of service are as follows.

Level of service A is primarily free-flow operation with average travel speeds near 60 mph on 70 mph freeways and vehicles have almost complete freedom to manoeuvre. Minor traffic incidents and breakdowns are easily absorbed and traffic quickly returns to level of service A. This level of flow gives drivers a high level of physical and psychological comfort. Average spacing between vehicle gives a maximum density of 12 passenger cars per mile per lane (pc/mi/ln).

Level of service B also represents reasonable free-flow conditions and speeds of over 57 mph are maintained on 70 mph freeways, and vehicles have only a slightly restricted freedom to manoeuvre. Minor traffic incidents and breakdowns are still easily absorbed although local deterioration in flow conditions would be more marked than in level of service A. The psychological and physical level of service is still good. Maximum density is 20 passenger cars per mile per lane.

Level of service C gives stable operation but flows approach the range in which a small increase in flow will cause a marked reduction in service; average speeds over 54 mph are maintained. Freedom to manoeuvre is considerably restricted and lane changes require care. Incidents may still be absorbed but local deterioration in service will be substantial with queues forming at major incidents. Drivers experience a major increase in tension because of the increased density of traffic flow which has a maximum value of 30 passenger cars per mile per lane.

Level of service D borders on unstable flow and small increases in flow result in substantial falls in the level of service; average speeds are in the region of 4–6 mph. Freedom to manoeuvre is severely limited with drivers experiencing drastically reduced physical and psychological levels. Maximum density is 42 passenger cars per mile per lane.

Level of service E – the boundary between level of service D to level of service E – describes capacity operation; traffic flow at this level is very unstable, vehicles are spaced at approximately uniform headways, any disruption due to a vehicle entering the traffic stream forms a disruptive wave which moves upstream, and average speeds are in the region of 30 mph. Any incident produces extensive queueing. Physical and psychological conditions afforded to drivers are extremely poor. Maximum density has a maximum value of 67 passenger cars per mile per lane. Level of service F describes breakdown or forced flow, conditions behind points on the highway where demand exceeds capacity.

The Highway Capacity Manual gives values of maximum service for freeways and these are given in table 11.2. These rates represent the maximum service flow rates for traffic flow in ideal conditions for 70, 60 and 50 mph design speed freeways.

The values for speed given in table 11.2 reflect the influence of a 55 mph speed limit and, unless enforcement is stringent, average travel speeds are expected to be slightly higher than the speed limit. The maximum service flow rates are given in units of passenger cars per hour per 12 ft lane (pcpp).

The service flow rate for a particular roadway element is then obtained by multiplying by the number of lanes in one direction of the freeway, a factor to adjust for the effects of restricted lane width and lateral clearance, a factor to adjust for the effect of heavy vehicles, and a factor to adjust for the effect of driver population.

The factor which accounts for restricted lane width and/or lateral clearance takes account of lane widths narrower than 12 ft and/or for objects closer to the edge of the travel lanes than 6 ft either at the roadside or in the central reservation. Abridged details are given in table 11.3 and in the use of this table it is recommended that judgement be made regarding the effect of crash barriers which do not produce an effect on traffic flow.

Adjustment for vehicle type in the Highway Capacity Manual is made by determining the passenger-car equivalent for each truck, bus or recreational vehicle. The Highway Capacity Manual gives values of the passenger car equivalents of trucks and recreational vehicles depending upon the severity and length of the gradient, the percentage of trucks and recreational vehicles in the traffic stream and the highway type. For freeways, trucks are divided into three classes depending upon the power-to-weight ratio. When capacity analysis is being carried out for considerable lengths of freeway, then grade conditions can be generally classified as level, rolling or mountainous terrain. The definition of rolling terrain is where any combination of horizontal or vertical alignment causes heavy vehicles to reduce their speed substantially below those of passenger cars, but does not cause

TABLE 11.2
Maximum service flow rates[1]

Design speed (mph)	*Level of service*	*Density (pc/mi/ln)*	*Speed (mph)*	*Maximum service flow rate*
70	A	≤12	≥60	700
	B	≤20	≥57	1100
	C	≤30	≥54	1550
	D	≤42	≥46	1850
	E	≤67	≥30	2000
	F	>67	<30	unstable
60	B	≤20	≥50	1000
	C	≤30	≥47	1400
	D	≤42	≥42	1700
	E	≤67	≥30	2000
	F	>67	<30	unstable
50	C	≤30	≥43	1300
	D	≤42	≥40	1600
	E	≤67	≥28	1900
	F	>67	<28	unstable

heavy vehicles to travel at crawl speeds for any significant length of time. Mountainous terrain, on the other hand causes heavy vehicles to travel at crawl speeds for significant distances or at frequent intervals. Using this broad classification, the passenger car equivalents given in table 11.4 may be used.

The adjustment for driver population takes account of the observed differences in characteristics between regular weekday and commuter drivers and weekend drivers especially in recreational areas. It is recommended that a factor of between 0.75 and 0.90 be used to account for the different characteristics of drivers who are not regular weekday or commuter drivers.

The United Kingdom approach

In Great Britain, carriageway provision both in type and width depends upon an assessment of traffic flow forecasts, the composition of the traffic, variations in traffic flow referred to as 'peaking', the costs of delays during traffic incidents and maintenance, and environmental effects[2].

To determine appropriate carriageway widths for rural roads, the Department of Transport[3] have made extensive assessments of the economic benefits of providing different carriageway widths on new rural roads. Time, vehicle operating and accident costs were assessed using cost–benefit analysis. In this analysis, estimates of total maintenance costs were made by assuming typical work schemes and associated delays; costs of traffic incidents in terms of delay to other vehicles were also included.

Economic assessment indicated the lower bound of a flow range, the lowest at which a given carriageway width was likely to be preferred to a lesser width. The

TABLE 11.3

Adjustment factor for restricted lane width and lateral clearance[1]

Distance from travelled carriageway (ft)	*Obstruction on one side of road*			*Obstruction on both sides of road*		
	Lane width (ft)					
	12	11	10	12	11	10
	4-lane freeway (2 lanes each direction)					
≥6	1.00	0.97	0.91	1.00	0.97	0.91
4	0.99	0.96	0.90	0.98	0.95	0.89
2	0.97	0.94	0.88	0.94	0.91	0.86
0	0.90	0.87	0.82	0.81	0.79	0.74
	6 or 8-lane freeways (3 or 4 lanes each direction)					
≥6	1.00	0.96	0.89	1.00	0.96	0.89
4	0.99	0.95	0.88	0.98	0.94	0.87
2	0.97	0.93	0.87	0.96	0.92	0.85
0	0.94	0.91	0.85	0.91	0.87	0.81

TABLE 11.4

Passenger car equivalents for extended freeway segments[1]

Passenger-car equivalent	*Terrain*		
	level	*rolling*	*mountainous*
For trucks	1.7	4.0	8.0
For buses	1.5	3.0	5.0
For recreational vehicles	1.6	3.0	4.0

upper bound of a flow range is however determined not only by economic assessment but also by operational assessment. This form of assessment indicates the maximum traffic flow which a given width can accommodate under some stated conditions.

Two factors were considered when economic assessments were modified by operational assessment. These were: the duration of those periods within the traffic peaks when the flow was greater than the road 'capacity' and the diversion of flows during maintenance.

In the assessment process the projected growth of traffic in the future can result in traffic flows being reached which exceed the maximum flow levels which have actually been observed on the highway. In the economic assessment it is assumed that when this takes place the costs which are incurred are as a result of traffic diverting to other routes or being suppressed. This is considered satisfactory for short-lived peak flows but is unrealistic for longer periods of time. For this reason, in the operational assessment 'over-capacity' was confined to peak flow periods in the thirtieth year of assessment. It was found that this limitation reduced the upper bound of dual carriageway and motorway flow ranges.

The second factor considered was the limitation on flows caused by a maximum possible flow which could be diverted during maintenance works. For single carriageway roads the maximum divertable flow was assumed to be approximately 2000 vehicles per day, and 10 000 vehicles per day for dual carriageways and motorways.

As a result of these assessments, the flow levels for rural road assessment have been published by the Department of Transport[3] and are given in table 11.5.

The unit which is used to measure traffic flow in this table is the 24-hour Annual Average Daily Traffic (AADT) which is the total annual traffic on a road divided by 365. The traffic flow is expressed in vehicles in table 11.5; an allowance for the composition of the traffic flow is made in the economic assessment of the adoption of differing carriageway types. Essential to this is the effect of vehicle type on the speed/flow relationship; this will be discussed separately. As indicated in table 11.5, traffic growth is taken into account by designing for the traffic flows which are expected in the 15th year after opening of a road.

The flow ranges given in table 11.5 are not intended to be used inflexibly; they are a starting point for the assessment of a road proposal and merely meant to be a guide as to which road layouts are likely to be operationally and economically acceptable in normal circumstances. In the United Kingdom, the Department of Transport does not stipulate a minimum level of service in road design; decisions on road capacity must reflect economic and environmental considerations. Where construction and environmental costs are high, then the incremental cost between two alternative road widths may result in a lower standard of provision than might otherwise have been made.

Alternatives for carriageway provision should be tested for assumptions of high and low traffic growth by considering the effects on travellers in terms of using

TABLE 11.5
Flow levels for the assessment of rural roads[3]

Road class	*AADT 15th year after opening*	*Access treatment*
Normal single 7.3 m carriageway	up to 13 000	Restriction of access, concentration of turning movements, clearway at top of range
Wide single 10 m carriageway	10 000 to 18 000	Restriction of access, concentration of turning movements, clearway at top of range
Dual 2-lane all-purpose carriageway	11 000 to 30 000	Restriction of access, concentration of turning movements, clearway at top of range
	30 000 to 46 000	Severe restriction of access, left turns only, clearway
Dual 3-lane all-purpose carriageway	40 000 and above	Severe restriction of access, left turns only, clearway
Dual 2-lane motorway	28 000 to 54 000	Motorway regulations
Dual 3-lane motorway	50 000 to 79 000	Motorway regulations
Dual 4-lane motorway	77 000 and above	Motorway regulations

travel cost savings, accident savings, user costs during maintenance, reconstruction, accidents and breakdowns, and driving conditions. Occupiers of properties within 300 m which are subject to increased visual impact, the number of properties requiring demolition, and agricultural land acquisition should be included in the consideration. Transport development policy decisions which affect carriageway width are taken into account and the financial implications considered.

Design flows for urban roads are based on peak hour flows and specifically apply to roads which function as traffic links and are independent of the capacities of junctions. There may be circumstances where traffic management policy requires lower values of design flows for environmental and safety reasons. In determining the peak hour flow, this is defined as the highest flow for any specific hour of the week averaged over any consecutive 13 weeks during the busiest period of the year. In the United Kingdom the busiest 3-month period is likely to be June to August but considerable differences occur, particularly in urban areas; the weekday peak hour is normally 5–6 p.m. on Friday, but once again local variations

TABLE 11.6
Design flows of two way urban roads[3]

Road type	*Carriageway*		*Peak hourly flow (veh/hour)*
	type	*width (m)*	
Urban	2-lane dual	7.3 dual	3600
motorway	3-lane dual	11 dual	5700
All purpose,	2-lane carriageway	7.3	2000*
no frontage access,			
no standing veh.,	2-lane carriageway	10	3000*
negligible cross			
traffic	4-lane undivided	12.3	2550
	4-lane undivided	13.5	2800
	4-lane undivided	14.6	3050
	2-lane dual	6.75 dual	2950†
	2-lane dual	7.3 dual	3200†
	2-lane dual	11 dual	4800†
All purpose,	2-lane carriageway	6.1	1100*
frontage	2-lane carriageway	6.75	1400*
development,			
side roads,	2-lane carriageway	7.3	1700*
pedestrian	2-lane carriageway	9	2200*
crossings,			
bus stops,	2-lane carriageway	10	2500*
waiting			
restrictions	4-lane undivided	12.3	1700
throughout day,	4-lane undivided	13.5	1900
loading restrictions	4-lane undivided	14.6	2100
at peak hours	6-lane undivided	18	2700

*Peak hourly flow, both directions of flow.

†60/40 directional split assumed.

are possible. Practically, measurements over 13 weeks are not essential and a period of 5 to 7 weeks is normally considered adequate.

Design flows for urban roads given by the Department of Transport[3] are detailed in tables 11.6 and 11.7. In adopting the values, the following points should be kept in mind except when dual carriageway links are being considered. Firstly, in most cases the design flows of existing roads will be dependent on the capacity of terminal junctions. Secondly, although the road types given cover a range of conditions which are difficult to define precisely, extreme conditions can be met where the design flows given cannot be attained. Thirdly, the design flows are only appropriate when the road is used solely as a traffic link.

The design flows given in tables 11.6 and 11.7 allow for a heavy vehicle content in the flow of 15 per cent, and no adjustment is necessary for lower heavy vehicle contents. Where the heavy vehicle content exceeds 15 per cent, then a correction to the design flow as given in table 11.8 should be made.

TABLE 11.7

Design flows for one way urban roads[3]

Road type	*Carriageway width (m)*	*Design flow, one direction of flow (veh/hour)*
All purpose road,	6.75	2950
no frontage access,	7.3	3200
no standing veh.,	11	4800
negligible cross traffic		
All purpose road,	6.1	1800
no frontage development,		
side roads,	6.75	2000
pedestrian crossings,	7.3	2200
bus stops,	9	2850
waiting restrictions		
throughout day,	10	3250
loading restrictions at peak hours	11	3550

TABLE 11.8

Corrections to tables 11.6 and 11.7 for heavy vehicle content[3]

Road type	*Heavy vehicle content (%)*	*Total reduction in flow (veh/hour)*
Motorway and dual carriageway all purpose roads	15–20	100
Motorway and dual carriageway all purpose roads	20–25	150
10 m wide and above single carriageway roads	15–20	150
10 m wide and above single carriageway roads	20–25	225
Below 10 m wide single carriageway roads	15–20	100
Below 10 m wide single carriageway roads	20–25	150

References

1. Transportation Research Board, Highway Capacity Manual, *Special Report* 209, Washington DC (1985)
2. Department of Transport, Roads and Local Transport Directorate, Choice between options for trunk road schemes, *Departmental Advice Note* TA 30/82, London (1982)
3. Department of Transport, Highways and Traffic, Traffic flows and carriageway width assessment for rural roads, *Departmental Advice Note* TD 20/85, London (1985)

Problems

Are the following statements true or false?

(a) The selection of carriageway width for a given design flow should in the United Kingdom be determined by a consideration of economic factors.
(b) The derivation of maximum service flow rates in the Highway Capacity Manual is based on average travel speeds.
(c) In the United Kingdom a preliminary starting point for the assessment of carriageway width in rural areas is the maximum observed peak hour flow.
(d) The Highway Capacity Manual defines level of service A as representing congested nose-to-tail driving conditions.
(e) Design flow ranges recommended for used in the United Kingdom were derived from a consideration of minimum travel cost.
(f) Crash barriers adjacent to the travelled carriageway inhibit the behaviour of drivers and reduce design flow levels.
(g) The effect of traffic composition is taken into account when calculating design flow levels by the use of passenger car equivalents.

Solutions

(a) In United Kingdom practice the recommended design flow ranges will normally allow a highway designer a choice of types and widths. The choice of option is made by a consideration of user travel costs, accident costs, user costs during maintenance, reconstruction, after accidents and vehicle breakdowns, and also driving conditions on the highway. In addition environmental effects, property demolition and land acquisition must be considered. All these effects must be evaluated for both high and low traffic growth forecasts.
(b) In United States practice the Highway Capacity Manual relates design flows to the expected level of service which is to be provided by the highway to the road user. For freeways these levels of service are related to traffic density, a traffic flow parameter which is related to traffic conditions as observed by drivers.
(c) In United Kingdom practice the starting point for the assessment of alternative carriageway widths is based on design flow ranges which are expressed by the Annual Average Daily Traffic, that is, the total flow of all vehicles in a year divided by 365. The fact that flows vary considerably

both during the day and the year is taken into account when user costs are estimated by the Department of Transport COBA program. This program uses flow groupings, varying for differing road categories, which take into account the variations of flow throughout the year and the differing proportions of vehicle types as the flow varies.

(d) The Highway Capacity Manual level of service A represents free-flow, not congested, traffic conditions. Vehicles have almost complete freedom to manoeuvre, giving drivers a high level of physical and psychological comfort.

(e) Design flow ranges used for determining carriageway width in the United Kingdom are determined for the lower end of the range from economic considerations using cost–benefit analysis while at the upper end economic considerations are modified by operational considerations, that is, traffic delays during maintenance, the effects of traffic incidents and accidents, and the maximum flow of traffic which can be diverted from a highway during obstructions to flow.

(f) While the Highway Capacity Manual gives corrections for design flows that arise because of continuous obstructions adjacent to the carriageway, the presence of crash barriers does not appear to inhibit driver behaviour and it is recommended that in this case judgement be used in applying a correction.

(g) The effect of different vehicle types on design flows is taken into account in the Highway Capacity Manual by assigning an equivalent passenger car value to each vehicle type. In United Kingdom practice the cost–benefit analysis which assists in carriageway width choice divides vehicles into cars, public service vehicles, light goods vehicles and two classes of other goods vehicles. Design flows are expressed in vehicles and are not converted into passenger car units.

12

Headway distributions in highway traffic flow

The concept of level of service in highway traffic flow illustrates the differences in the characteristics of the flow which may be examined by a study of the headways between vehicles. Time headways are the time intervals between the passage of successive vehicles past a point on the highway. Because the inverse of the mean time headway is the rate of flow, headways have been described as the fundamental building blocks of traffic flow. When the traffic flow reaches its maximum value then the time headway reaches its minimum value.

If time headways are observed during any period of time the individual values of time headway vary greatly. The extent of these variations depends largely on the highway and the traffic conditions.

On a lightly trafficked rural motorway, where vehicles can overtake at will, a range of headways will be observed from zero values between overtaking vehicles to the longer headways between widely spaced vehicles.

When flow conditions are observed on more heavily trafficked highways there are fewer opportunities to overtake and fewer more widely spaced vehicles. When overtaking opportunities do not exist there is an absence of very small headways and, under very heavily trafficked conditions, all vehicles are travelling at uniform headways as they follow each other along the carriageway.

There are two general approaches to the method of measurement of headways. They may be measured by a device that registers the successive arrivals of vehicles at a fixed point. Alternatively headways may be recorded by aerial photography which records at one instant of time the distribution of headways between successive vehicles.

By the first method it is the time headway distribution that is obtained and in the second method it is the space headway distribution. Because of the ease of observation it is the time headway distribution that has been extensively researched and reported.

When the arrival of vehicles at a particular point on the highway is described, the distribution may either describe the number of vehicles arriving in a time interval or the time interval between the arrival of successive vehicles. The first type is the counting distribution and the second type is the gap distribution.

As great variability in all types of headway may be recorded, it has been described as the most noticeable characteristic of vehicular traffic. For this reason attempts to understand headways have employed statistical methods and probability theory to find theoretical distributions to represent the observed headway distributions.

Haight, Whisler and Mosher[1] studied the relationship between the counting and the gap distribution but usually it is the gap distribution that is studied. It requires a shorter period of observation to collect data for the investigation of gap distributions than for an investigation into counting distributions. In the study of intersection capacity, gaps in the major road flow are used by minor road vehicles to enter the major road and, once again, it is the gap distribution that is of importance.

One of the earliest headway distributions proposed for vehicular traffic flow was proposed by Kinzer[2] and Adams[3] who suggested that the negative exponential distribution would be a good fit to the cumulative gap distribution. Adams illustrated the validity of the negative exponential distributions by observations of traffic flow in London. When this distribution represents the cumulative headway distribution then arrivals occur at random and the counting distribution may be represented by the Poisson distribution. This type of flow may be found where there are ample opportunities for overtaking, at low volume/capacity ratios.

The negative exponential headway distribution

If the traffic flow is assumed to be random then the probability of exactly n vehicles arriving at a given point on the highway in any t second interval is obtained from the Poisson distribution which states

$$\text{probability } (n \text{ vehicles}) = (qt)^n \exp(-qt)/n! \tag{12.1}$$

where q is the mean rate of arrival per unit time.

Often this distribution is referred to as the counting distribution, because it refers to the number of vehicles arriving in a given time interval. It is the negative exponential distribution however which is most commonly used when describing headway distributions.

The negative exponential distribution can be obtained from the Poisson distribution if there are no vehicle arrivals in an interval t. In this case there must be a headway greater than or equal to t:

$$\text{probability } (\text{headway} \geqslant t) = \exp(-qt) \tag{12.2}$$

The mean rate of arrival q is also the reciprocal of the mean headway which can be computed from the observed headways.

It is possible to demonstrate the use of the Poisson and negative exponential distributions by taking observations of headways on a highway where traffic is flowing freely. This flow condition is likely to be found on two-way two-lane highways when the traffic volume in each direction does not exceed 400 veh/h or 800 veh/h on one-way two-lane carriageways and there are no traffic control devices within a distance of approximately 1 km upstream of the point of observation. The time interval between the passage of successive vehicles is noted using a stopwatch and the resulting headways are placed into classes with a class interval of 2 seconds. A smaller class interval is desirable but it is not normally justified unless more accurate means of measuring the headways are available. It will normally be necessary to continue observations for 30 minutes to obtain sufficient headways.

Observations, derived and theoretical values may be tabulated as shown in table 12.1.

TABLE 12.1

Row 1	Headway class
Row 2	Observed frequency of headways in class
Row 3	Observed frequency of headway ⩾ lower class limit
Row 4	Row 3 expressed as a percentage
Row 5	Theoretical percentage of headways ⩾ lower class limit (equation 12.2 multiplied by 100 where $t = 0, 2, 4 \ldots$)
Row 6	Theoretical frequency of headways ⩾ lower class limit (row multiplied by total number of headways observed)
Row 7	Theoretical frequency of headways in class (obtained by difference between successive values of row 6)

It should be noted that the following rows represent observed and theoretical values

row 2 and row 7

row 3 and row 6

row 4 and row 5

An example of observed and theoretical values is given in table 12.2. This table is incomplete as the greatest headway was in the class 64 to 65.9 seconds, but for the sake of brevity only a limited number of values are given.

TABLE 12.2

Row 1	0	2	4	6	8	10	12	14	16
Row 2	34	27	16	13	13	10	6	7	5
Row 3	168	134	107	91	78	65	55	49	42
Row 4	100	80	64	54	46	39	33	29	25
Row 5	100	84	70	58	49	41	34	29	24
Row 6	168	140	117	98	82	68	57	48	40
Row 7	28	23	19	16	14	11	9	8	7

Note that differences between successive values in row 3 give values in row 2; similarly differences in row 6 give values in row 7. This is because the number of headways greater than t_1 second minus the number of headways greater than t_2 second gives the number of headways in class $t_1 t_2$ second.

The closeness of fit of the negative exponential distribution to the observed values can be demonstrated graphically by drawing a histogram of the observed and theoretical number of headways in each class as in figure 12.1.

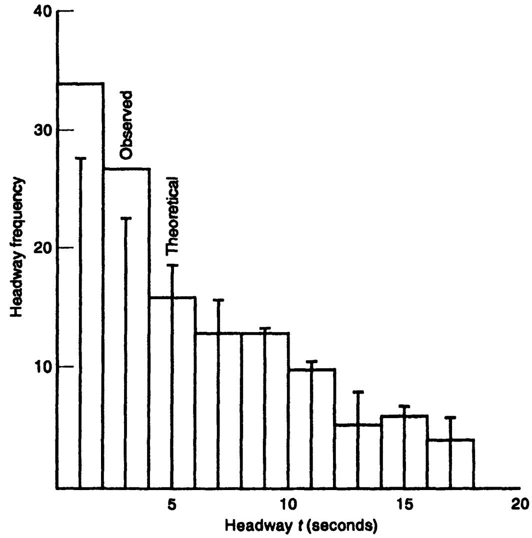


Figure 12.1 *A comparison of observed and theoretical headway frequencies*

However a statistical test of the closeness of fit of the theoretical distribution can be made by means of the chi-squared test, where chi-squared is the sum for each headway class of the following statistic

$$\frac{(\text{observed frequency} - \text{theoretical frequency})^2}{\text{theoretical frequency}}$$

that is

$$\frac{(\text{row } 2 - \text{row } 7)^2}{\text{row } 7}$$

Where the theoretical or observed frequency is less than 8 in any class, that class is combined with subsequent classes until the combined theoretical and observed frequencies are greater than 8.

Chi-squared is calculated in table 12.3.

TABLE 12.3
Calculation of closeness of fit

Row 2	*Row 7*	*(Row 2–Row 7)² /Row 7*
34	28	1.28
27	23	0.69
16	19	0.47
13	16	0.56
13	14	0.07
10	11	0.09
18*	24*	1.50 = 4.66

*Last three classes combined.

Examination of tables of chi-squared shows that at the 5 per cent level the value of chi-squared for 5 degrees of freedom is 11.07. The calculated value is 4.66 showing that there is no significant difference between the observed and theoretical headway distribution at the 5 per cent level.

Note that the degrees of freedom are equal to the number of combined classes, in this case 7, minus the number of parameters that were used from the observed to calculate the theoretical values. These two parameters are the total number of observed headways and the mean headway; the former was used in the calculation of row 6 and the latter in the calculation of row 5.

Another way in which the fit of the observed distribution to the theoretical distribution can be demonstrated is by plotting the theoretical and observed cumulative percentage frequency. These values are given in rows 4 and 5 respectively. If equation 12.2 is considered and natural logarithms are taken of both sides of the equation, then

$$\ln\,[\text{probability (headway} \geqslant t)] = -qt \qquad (12.3)$$

When plotted on semi-log paper equation 12.3 is a straight line; the nearer the observed distribution is to a straight line the better is the fit. This is illustrated graphically in figure 12.2.

This experiment shows, for the traffic volume observed and for the class interval chosen, that the distribution of headways may be represented by the exponential distribution. It should be noted that one of the difficulties of the exponential distribution, the surplus of small headways less than 1 second, is hidden by the use of a class interval of 2 seconds.

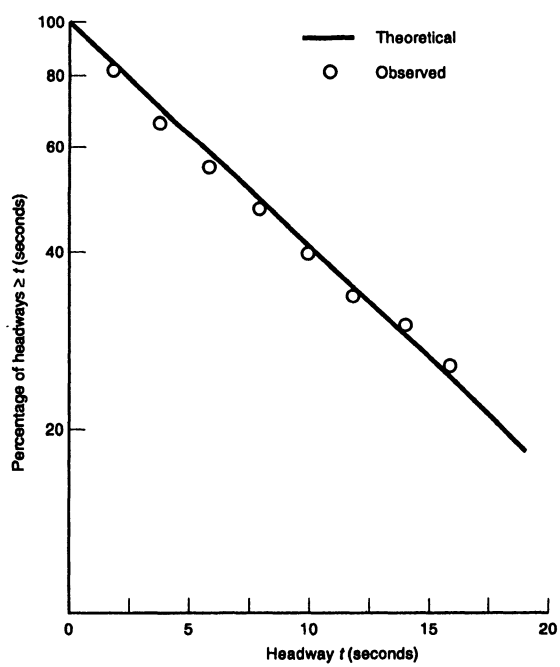


Figure 12.2 *A comparison of observed and theoretical cumulative headway distributions*

Congested vehicular headway distributions

While free-flowing vehicular headway distributions may be approximated by the negative exponential distribution in the case of the cumulative headway distribution and by the Poisson distribution in the case of the counting distribution, it is obvious that the distribution of headways will depend on the traffic volume and also on the capacity of the highway. If drivers cannot maintain their desired speed by overtaking slower moving vehicles then free-flow conditions no longer exist and the highway is beginning to show signs of congestion.

Highway congestion may increase until finally all vehicles are travelling at the same speed and following each other at their minimum headway.

In this case vehicles are regularly distributed along the highway and the headway distribution is determinate so that

$$\text{probability (headway} \geqslant t) = 0 \text{ when } t \neq \bar{t}$$
$$= 1 \text{ when } t = \bar{t}$$

where $\bar{t}$ is the mean headway between vehicles; the counting distribution is also given by

$$\text{probability } (n \text{ vehicles in time } t) = \frac{t - N\bar{t}}{\bar{t}}$$

where $n = N + 1$, and

$$= 1 - \frac{t - N\bar{t}}{\bar{t}}$$

where $n = N$.

N is the number of headways of time headway $\bar{t}$ contained in time interval t.

It is not however so much the two extremes of regularity or randomness which are of interest in the study of traffic flow as the distribution of these headways in the intermediate conditions between these extremes. The study of traffic flow requires a knowledge of the distribution of headways in a wide variety of traffic conditions if realism is to be achieved.

A difficulty of the use of the negative exponential distribution even under free-flow conditions is that the probability of observing a headway increases as the size of the headway decreases. As vehicles have a finite length and a minimum following headway this presents a problem when only a limited number of overtakings are observed. For this reason traffic flow has been described by the use of the displaced exponential distribution. This is a suitable distribution for the description of flow in a single lane of a multilane highway.

Where a small number of low-value headways are observed, such as is the case when a limited amount of overtaking is possible, the Pearson type III or the Erlang distribution may be used to represent the headway frequency distribution.

The Pearson type III distribution was developed by Karl Pearson to deal with a wide variety of statistical data. A random variable such as a headway (t) is said to be distributed as the type III distribution if its probability density function is given by

$$f(t) = \frac{b^a}{\Gamma(a)} t^{a-1} e^{-bt} \quad 0 < t < \infty \tag{12.4}$$

where $\Gamma(a)$ is known as the gamma function and is defined by

$$\Gamma(a) = \int_0^\infty z^{a-1} e^{-z} \, dz$$

Where headways in a single lane are being observed then any headway less than the minimum following distance cannot be observed and the Pearson type III distribution has to be modified to give

$$f(t) = \frac{b^a}{\Gamma(a)} (t - c)^{a-1} e^{-b(t-c)} \quad c < t < \infty \tag{12.5}$$

For equation 12.4 the theoretical distribution may be obtained from

$$\bar{t} = \frac{a}{b} \quad \text{and} \quad s^2 = \frac{a}{b^2}$$

For equation 12.5 the theoretical distribution may be obtained from

$$\bar{t} = \frac{a}{b}, \quad s^2 = \frac{a}{b^2}$$

and c is the minimum observed headway.

The Erlang distribution is a simplified form of the Pearson type III distribution in which the parameter a is an integer. This may be written

$$f(t) = \frac{(qa)^a}{(a-1)!} \times t^{a-1} e^{-aqt} \quad a = 1, 2, 3 \ldots \tag{12.6}$$

where q is the rate of flow, the reciprocal of the mean time headway.

The fitting of Pearson III and Erlang distributions to headway data

An example of the fitting of Pearson III distributions is given below for headway data collected on the Leeds–Bradford highway at Thornbury, a one-way, two-lane highway. The grouped data is shown in table 12.4.

The mean and the variance of the headway distribution is first calculated, giving

$$\bar{t} = 3.4 \text{ s} \quad \text{and} \quad s^2 = 7.9 \text{ s}$$

where $\bar{t}$ is the mean time headway and s^2 is the variance of the headways.

As headways are observed in the range 0–0.9 s the frequency curve may be assumed to pass through the origin, a reasonable assumption when it is considered that headways were observed on a one-way two-lane carriageway and frequent overtaking was possible. Using equation 12.5 to calculate the theoretical frequency distribution it is necessary to calculate the value of the parameters a and b.

From the relationship $\bar{t} = a/b$ and $s^2 = a/b^2$ then

$$a = 1.499 \quad \text{and} \quad b = 0.435$$

The value of $\Gamma(a)$ was then calculated using a standard gamma function program, giving a value of 0.923. The value of the theoretical frequency may then be

TABLE 12.4
Observed and fitted headway distributions

Headway class (seconds)	Observed frequency	Theoretical frequency: Equation 12.4 Pearson III	Theoretical frequency: Equation 12.6 Erlang
0–0.9	57	75.3	51.0
1–1.9	99	83.6	85.5
2–2.9	69	69.7	80.0
3–3.9	58	53.3	62.4
4–4.9	44	39.4	45.1
5–5.9	25	27.9	30.8
6–6.9	14	19.7	20.4
7–7.9	13	13.9	13.2
8–8.9	11	9.4	8.4
9–9.9	6	6.6	5.2
10–10.9	3	4.5	3.2
11–11.9	3	2.9	2.0
12–12.9	3	2.1	1.2
13–13.9	0	1.2	0.7
14–14.9	1	0.8	0.4
15–15.9	4	0.4	0.2

calculated using equation 12.5 by substituting for t the successive values of the mid-class marks.

Where a desk calculator or tables of the gamma function are not available the Erlang distribution, equation 12.6, may be used as a simpler approximation. The value of the parameter a in this distribution reflects the distribution of headways for a range of traffic flow conditions. When $a = 1$ the distribution becomes negative exponential and when a is infinite complete uniformity of headways results, the value of a thus reflecting flow conditions between free-flowing and congested states.

For this example the value of a is chosen as 2 and using equation 12.6 the theoretical frequency distribution is evaluated and given in table 12.4, together with the observed headway frequencies.

The double exponential headway distribution model

Schuhl[4] proposed a headway distribution in which vehicles travelling along a highway could be considered to be composed of two types, firstly those who were unable to overtake and were restrained in their driving performance and secondly those drivers who were unrestrained by other vehicles on the highway.

Drivers who are restrained by the action of the driver in front can approach to within a minimum time headway of e and their cumulative headway distribution may be represented by

$$\text{probability (headway} \geqslant t) = L \exp(-(t-e)/(\bar{t}_1 - e)) \quad t \geqslant e \qquad (12.7)$$
$$= L \qquad t \leqslant e$$

where L is the proportion of restrained vehicles in the traffic stream and $\bar{t}_1$ is the mean headway between restrained vehicles.

Similarly drivers who are not restrained by the vehicle in front have no limitation on the minimum headway, as they are able to overtake, and their cumulative headway distribution may be represented by

$$\text{probability (headways} \geqslant t) = (1 - L) \exp(-t/\bar{t}_2) \quad t \geqslant 0 \qquad (12.8)$$

where $\bar{t}_2$ is the mean headway between unrestrained vehicles.

On the highway both restrained and unrestrained vehicles are present and so the observed headway distribution is represented by the sum of equations 12.7 and 12.8

$$\text{probability (headways} \geqslant t) = L \exp(-(t - e)/(\bar{t}_1 - e))$$
$$+ (1 - L) \exp(-t/\bar{t}_2) \quad t \geqslant e \qquad (12.9)$$

Some of the earliest research into the fit of the double exponential distribution to observed headways on two-lane urban streets was carried out by Kell[5]. The theoretical cumulative headway distribution chosen for fitting to observed values was

$$\text{probability (headways} \geqslant t) = \exp(a - t/K_1) + \exp(c - t/K_2)$$

and the relationships of the parameters to the traffic volume were found to be

$$K_1 = 4827.9/V^{1.024}$$

$$a = -0.046 - 0.000\,448V$$

$$K_2 = 2.659 - 0.0012V$$

$$c = \exp(-10.503 + 2.829 \ln V - 0.173 (\ln V)^2) - 2$$

The fitting of a double exponential distribution to headway data

The use of this double exponential distribution may be demonstrated by an analysis of headways on a highway where traffic is experiencing some degree of congestion. Normally this can be expected to occur on a two-lane two-way highway when the traffic volume in each direction exceeds approximately 600 veh/h or on a two-lane one-way highway when the traffic volume exceeds approximately 1000 veh/h.

The observed headways given in table 12.4 will be used to illustrate the fitting of the double exponential distribution using a graphical technique. Observed headway frequency, cumulative frequency and percentage cumulative frequency distributions are given in table 12.5.

TABLE 12.5
Observed headways

Headway class (seconds)	*Observed frequency*	*Number of headways greater than lower class limits*	*Percentage of headways greater than lower class limit*
0–0.9	57	410	100.0
1–1.9	99	353	86.1
2–2.9	69	254	62.0
3–3.9	58	185	45.3
4–4.9	44	127	31.1
5–5.9	25	83	20.4
6–6.9	14	58	14.4
7–7.9	13	44	11.0
8–8.9	11	31	7.8
9–9.9	6	20	5.1
10–10.9	3	14	3.7
11–11.9	3	11	2.9
12–12.9	3	8	2.2
13–13.9	0	5	1.5
14–14.9	1	5	1.5
15–15.9	4	4	1.2

The distinctive graphical plot on semi-log paper of the cumulative headway distribution when the traffic flow is partly restrained is illustrated in figure 12.3, which should be compared with figure 12.2 (page 105) where the cumulative headway distribution for free-flowing traffic is plotted.

The graphical fitting of the double exponential curve may be described by reference to the form of the theoretical curve shown in figure 12.4. The theoretical cumulative headway distribution for both free-flowing and restrained vehicles is shown by a solid curve and, if the observed cumulative headway distribution is plotted, it will approximate to this line if the underlying theoretical concept is correct. The straight line portion of this curve represents the headway distribution of free-flowing vehicles as the effect of restrained vehicles is negligible at the larger values of headway. This portion of the curve may then be represented by equation 12.8. If the straight line portion of the curve is extended to the vertical axis, as shown by a broken line, the percentage of headways greater than t represents the value of $100(1 - L)$ so allowing L to be determined.

If a point is taken on the line and a value of the probability of a headway $\geqslant t$ and the corresponding value of t are substituted in equation 12.8 then the value of $\bar{t}_2$ may be found.

The vertical difference between the straight line and the cumulative curve represents the headway distribution of restrained drivers. Since this difference represents equation 12.6 it is again a straight line. The value of t, when the proportion of restrained vehicles is L, on this straight line gives e. A point may then be selected on the line and a value of the probability and the corresponding value of t substituted in equation 12.7 to give $\bar{t}_1$.

Using the parameters L, e, $\bar{t}_1$ and $\bar{t}_2$ the theoretical cumulative headway distribution may be calculated using equation 12.9 and plotted. For instance the theoretical

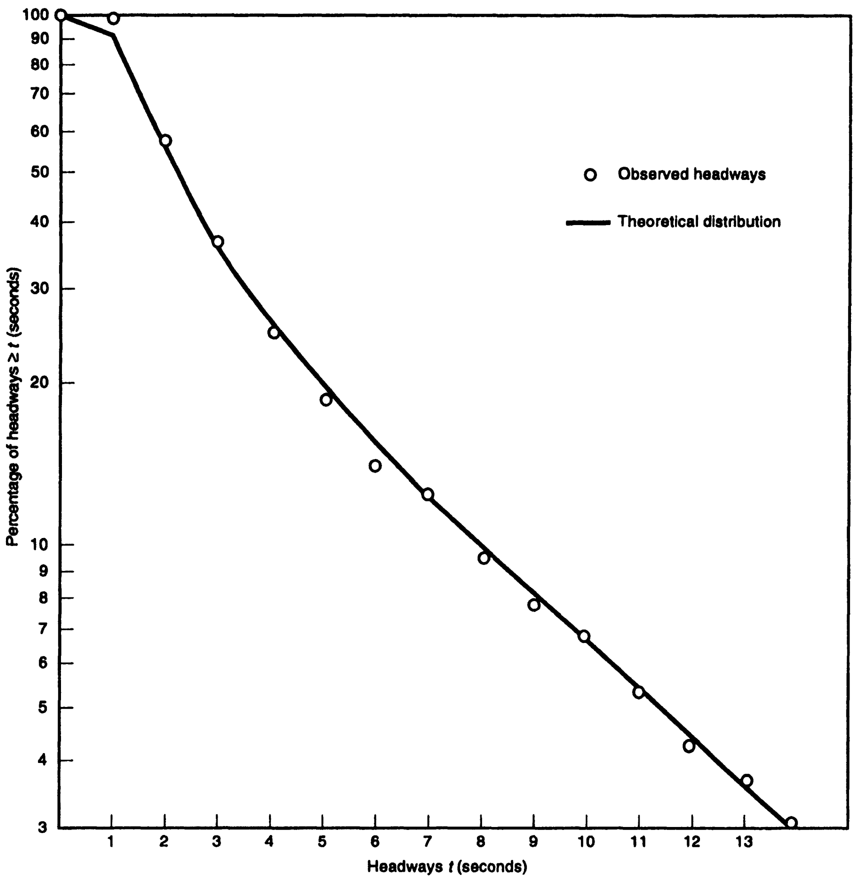


Figure 12.3 *Observed headways and the fitted double exponential distribution*

values were obtained using the least squares technique and the computer program developed by R.J. Salter, giving the following values of the parameters

L 0.66

t_1 3.0 seconds

t_2 4.4 seconds

It is interesting to note that at the traffic volume sampled, approximately 1000 veh/h, considerably more than 50 per cent of drivers appear to be restrained by the preceding vehicles. Also at a time headway in the region of 7 seconds drivers appear to cease to be influenced by the vehicle in front.

The observed and the theoretical headway frequencies obtained from the double exponential distribution using the derived parameters are compared in figure 12.5. It can be seen that once again the exponential distribution gives too high a frequency of headways in the class 0–1 second when compared with the observed frequency.

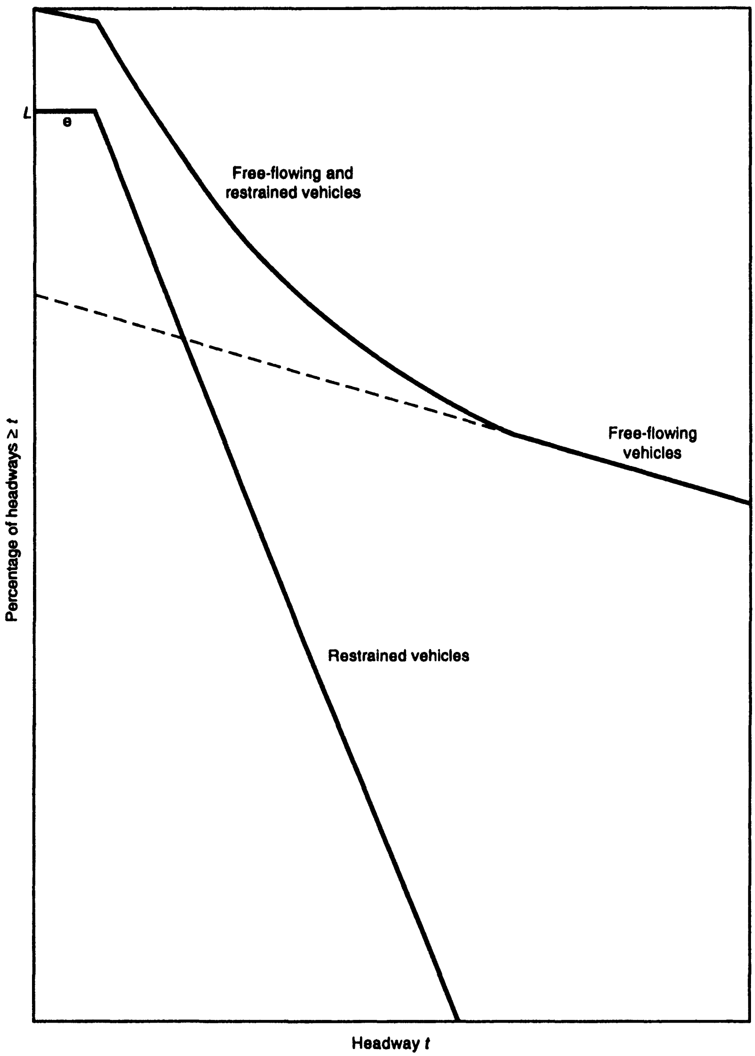


Figure 12.4 *The theoretical cumulative headway distribution for free-flowing and restrained vehicles in a traffic stress*

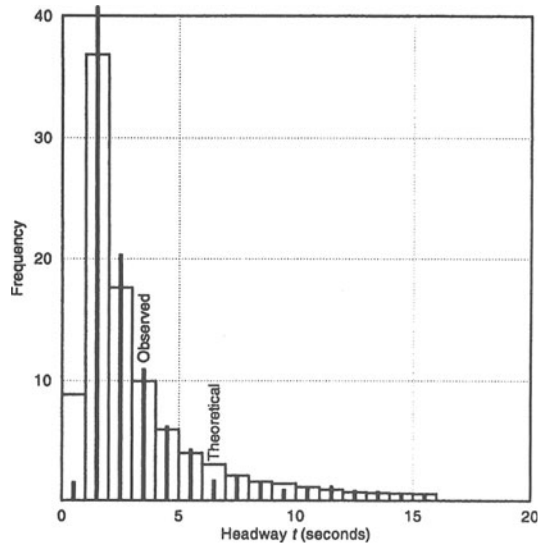


Figure 12.5 *Observed and theoretical headway frequencies for congested flow*

References

1. F. A. Haight, B. F. Whisler and W. W. Mosher, New statistical method for describing highway distribution of cars, *Proc. Highw. Res. Bd*, **40** (1961), 557–64
2. J.P. Kinzer, Applications of the theory of probability to problems of highway traffic, *Proc. Inst. Traffic Engrs*, (1934), 118–23
3. W. F. Adams, Road traffic considered as a random series, *J. Instn Civ. Engrs* (1936), 121–30
4. A. Schuhl, The probability theory applied to distribution of vehicles on two-lane highways, *Poisson and Traffic*, The Eno Foundation (1955), pp. 59–75
5. J.H. Kell, Analysing vehicular delay at intersections through simulation, *Highw. Res. Bd Bull.*, **1** (1962), 28–9.

Problems

(1) Select the correct completion of the following statements:
(a) The time headway distribution between successive highway vehicles may be obtained
(i) from an aerial photograph of a highway,
or (ii) by an observer equipped with a stopwatch standing at the side of the highway.
(b) The counting distribution as applied to highway traffic flow indicates
(i) the distribution of the numbers of vehicles arriving at a point on the highway during fixed periods of time,
or (ii) the distribution of time intervals between the arrival of vehicles,
or (iii) neither.

(c) Gap distributions are normally observed by highway traffic researchers because
(i) the counting distributions require a considerable time to collect adequate data,
or (ii) the gap distribution may be easily obtained by the use of aerial photographs.

(d) The time headway distribution is of importance to a study of highway traffic flow because
(i) the probability of observing a headway increases as the size of the headway decreases,
or (ii) the inverse of the mean time headway is the rate of traffic flow.

(e) The negative exponential distribution represents
(i) the cumulative headway distribution of freely flowing highway traffic,
or (ii) the cumulative headway distribution of heavily congested highway traffic,
or (iii) the counting distribution of vehicle arrivals on a traffic signal approach.

(2) The headway distribution on a two way urban highway is given in table 12.6.
(a) Assume the traffic is free flowing and graphically fit a theoretical distribution to the observed values. Calculate the theoretical frequencies.
(b) Use a histogram to show the closeness of fit of the theoretical and observed distributions and state what conclusions may be drawn from this comparison.

TABLE 12.6

Headway class (second)	*Observed frequency*
0–0.9	19
1–1.9	67
2–2.9	58
3–3.9	29
4–4.9	26
5–5.9	14
6–6.9	17
7–7.9	7
8–8.9	9
9–9.9	6
10–10.9	5
11–11.9	8
12–12.9	4
13–13.9	4
14–14.9	4
15–15.9	3
16–16.9	3
17–17.9	0
18–18.9	2
19–19.9	1

TABLE 12.7

No. of vehicles arriving in 2 minute intervals	*No. of times this number of vehicles was observed*
8	4
7	1
6	4
5	9
4	10
3	11
2	12
1	11
0	1

(3) The number of vehicles arriving in two minute intervals at a check point on a highway is given in table 12.7.
Show that the traffic flow on this highway is random.

(4) The traffic flow on a highway is composed of free-flowing and following vehicles. The free-flowing vehicles have a mean time headway of 8.0 seconds while the following vehicles do not approach closer to the one in front than a time headway of 0.5 second. When the traffic volume was 1000 vehicles per hour it was noted that there were 25 per cent of free-flowing and 75 per cent following vehicles.
Calculate:
(a) the proportion of vehicles with a time headway greater than 6 seconds;
(b) the time headway at which only 5 per cent of vehicles are restrained;
(c) the proportion of following vehicles which have a headway greater than 0.25 second.

Solutions

(1) (a) The time headway distribution between successive highway vehicles may be obtained by an observer equipped with a stopwatch standing at the side of the highway.
(b) The counting distribution as applied to highway traffic flow indicates the distribution of the numbers of vehicles arriving at a point on the highway during fixed periods of time (table 12.8).
(c) Gap distributions are normally observed by highway traffic researchers because the counting distributions require a considerable time to collect adequate data.
(d) The time headway distribution is of importance to a study of highway traffic flow because the inverse of the mean time headway is the rate of traffic flow.
(e) The negative exponential distribution represents the cumulative headway distribution of freely flowing highway traffic.

(2) To graphically fit a theoretical distribution to the observed headway distribution the percentage cumulative headway distribution will be calculated. If the traffic is free flowing then the arrival of vehicles at a point on the highway will be random and the negative exponential

TABLE 12.8

Class (seconds)	*Observed frequency*	*Observed cumulative frequency*	*Percentage observed cumulative frequency*	*Percentage theoretical cumulative frequency*	*Theoretical cumulative frequency*	*Theoretical frequency*
0–0.9	19	286	100.0	100.0	286	56
1–1.9	67	267	93.4	80.3	230	46
2–2.9	58	200	69.9	64.4	184	36
3–3.9	29	142	49.7	51.7	148	29
4–4.9	26	113	39.5	41.5	119	23
5–5.9	14	87	30.4	36.4	104	19
6–6.9	17	73	25.5	26.7	76	15
7–7.9	7	56	19.6	21.4	61	12
8–8.9	9	49	17.1	17.4	50	10
9–9.9	6	40	14.0	14.0	40	8
10–10.9	5	34	11.9	11.2	32	6
11–11.9	8	29	10.1	9.0	26	5
12–12.9	4	21	7.3	7.2	21	4
13–13 9	4	17	5.9	5.8	17	3
14–14.9	4	13	4.5	4.6	13	3
15–15.9	3	9	3.1	3.7	11	2
16–16.9	3	6	2.9	3.0	9	2
17–17.9	0	3	1.0	2.4	7	1
18–18.9	2	3	1.0	1.9	6	1
19–19.9	1	1	–	1.5	4	0

distribution will fit the observed cumulative headway distribution. This can be demonstrated by plotting the observed percentage cumulative distribution on semi-log paper. The relationship will be linear and the best straight line can be fitted to the data to obtain the theoretical cumulative distribution from the straight line. The theoretical cumulative distribution can be converted to a cumulative distribution and the differences between successive values will give the theoretical frequencies.

The observed and theoretical frequencies are compared with the help of a histogram. It can be seen that the first class is a particularly poor fit as can be expected with the exponential distribution in which the probability of a headway increases as the headway size decreases. This does not occur on the highway where there is a minimum headway between vehicles unless these vehicles are overtaking.

(3) The calculation can be carried out in tabular form as shown in table 12.9. The number of degrees of freedom is equal to the number of groups (6) minus the constraints (the mean number of vehicles arriving per 2 minute interval, and the total number of vehicle arrivals).

From tables of chi-squared it can be seen that the value of chi-squared for 4 degrees of freedom at the 5 per cent level of significance is 9.49. There is thus no significant difference at the 5 per cent level of significance between the observed and the theoretical distributions indicating that vehicle arrivals are random.

(4) The cumulative headway distribution of the freely flowing vehicles in the stream is given by

$$\text{Probability } (h \geqslant t) = (1 - L) \exp(-t/\bar{t}_2)$$

TABLE 12.9

(1) No. of vehicles per 2 minute interval	*(2)* No. of times observed (f_0)	*(3)* (1) × (2)	*(4)* Theoretical probability*	*(5)* Σ (2) × 4 Theoretical frequency (f_t)	*(6)* x^2 $\frac{(f_0 - f_t)^2}{f_t}$
8	4	32	0.015	0.95	
7	1	7	0.036	2.27	0.18
6	4	24	0.073	4.60	
5	9	45	0.128	8.06	0.11
4	10	40	0.187	11.78	0.27
3	11	33	0.218	13.78	0.54
2	12	24	0.191	12.03	0.00
1	11	11	0.110	6.99	1.04
0	1	0	0.032	2.02	
	Σ63	Σ216			Σ2.14

*Probability (n vehicles arriving per 2 minute interval)

$$= \frac{\exp(-m)m^n}{n!} \quad \text{where } m = \Sigma((1) \times (2))/\Sigma(2) = 3.43 \text{ vehicles}$$

while the cumulative headway distribution of the following vehicles in the stream is given by

$$\text{probability } (h \geqslant t) = L \exp(-(t - e)/(\bar{t}_1 - e))$$

From the information given in the question

$$\text{traffic volume} = \frac{3600}{L\bar{t}_1 + (1 - L)\bar{t}_2}$$

$$\bar{t}_2 = 8.0 \text{ seconds}$$

$$L = 0.75$$

Hence

$$1000 = \frac{3600}{0.75\bar{t}_1 + 0.25 \times 8.0}$$

that is

$$\bar{t}_1 = 2.1 \text{ seconds}$$

(a) The proportion of vehicles, both free flowing and following, with a time headway greater than 6 seconds is given by

$$\begin{aligned} \text{probability } (h \geqslant 6) &= (1 - 0.75) \exp(-6/8) + \\ &\quad 0.75 \exp(-(6 - 0.5)/(2.1 - 0.5)) \\ &= 0.25 \times 0.4724 + 0.75 \times 0.1385 \\ &= 0.1181 + 0.2389 \\ &= 0.3570 \end{aligned}$$

(b) The probability of a headway relating to a following vehicle may be calculated from

probability $(h \geqslant t) = L \exp(-(t - e)/(\bar{t}_1 - e))$

In this case the equation is solved to obtain t when the probability is 0.05

$0.05 = 0.75 \exp(-(t - 0.5)/(2.1 - 0.5))$

$0.067 = \exp(-(t - 0.5)/1.6)$

Taking natural logarithms of both sides of the equation

$$-2.7031 = -\frac{t - 0.5}{2.7}$$

$t = 4.8$ seconds

(c) As all following vehicles have a headway greater than 0.5 second then all following vehicles have a headway greater than 0.25 second. The proportion of following vehicles is 0.75 and so the proportion of following vehicles with a headway greater than 0.25 second is 0.75.

13

The relationship between speed, flow and density of a highway traffic stream

Theoretical relationships between speed, flow and density

When considering the flow of traffic along a highway three descriptors are of considerable significance. They are the speed and the density or concentration, which describe the quality of service experienced by the stream; and the flow or volume, which measures the quantity of the stream and the demand on the highway facility.

The speed is the space mean speed; the density or concentration is the number of vehicles per unit length of highway and the flow is the number of vehicles passing a given point on the highway per unit time.

The relationship between these parameters of the flow may be derived as follows. Consider a short section of highway of length L in which N vehicles pass a point in the section during a time interval T; all the vehicles travelling in the same direction.

The volume flowing $Q = N/T$

The density $D \quad = \dfrac{\text{average no. of vehicles travelling over } L}{L}$

The average number of vehicles travelling over L is given by

$$\frac{\sum_{i=1}^{N} t_i}{T}$$

where t is the time of travel of the ith vehicle over the length L; then

$$D = \frac{\sum_{i=1}^{N} t_i}{T} \bigg/ L = \frac{\frac{N}{T}}{\frac{L}{\frac{1}{N}\sum_{i=1}^{N} t_i}}$$

or

$$\text{density} = \frac{\text{flow}}{\text{space mean speed}} \tag{13.1}$$

Numerous observations have been carried out to determine the relationship between any two of these parameters for, with one relationship established, the relationship between the three parameters is determined. Usually the experimenters have been interested in the relationship between speed and flow because of a desire to estimate the optimum speed for maximum flow.

Greenshields[1] is one of the earliest reported researchers in this field and in a study of rural roads in Ohio he found a linear relationship between speed and density of the form

$$\bar{V}_s = \bar{V}_f - \left(\frac{\bar{V}_f}{D_j}\right) D \tag{13.2}$$

where $\bar{V}_s$ is the space mean speed,

$\bar{V}_f$ is the space mean speed for free flow conditions,

$\bar{D}_j$ is the jam density.

With this relationship determined the flow–density relationship can be obtained by substitution of equation (13.1) in equation (13.2) to give

$$Q = \bar{V}_f D - \frac{\bar{V}_f}{D_j} D^2 \tag{13.3}$$

and similarly the relationship between flow and speed may be obtained as

$$Q = D_j \bar{V}_s - \frac{D_j}{\bar{V}_f} \bar{V}_s^2 \tag{13.4}$$

The density and speed at which flow is a maximum can be obtained by differentiating equations 13.3 and 13.4 with respect to density and speed. To obtain the density when flow is a maximum, from equation 13.3

$$\frac{dQ}{dD} = \bar{V}_f \left(2 \times \frac{\bar{V}_f}{D_j} D\right) = 0 \text{ for a maximum value}$$

$$D = D_{max} = \frac{D_j}{2} \tag{13.5}$$

To obtain the speed when flow is a maximum, from equation (13.4)

$$\frac{dQ}{d\bar{V}_s} = D_j - \left(\frac{2 \times D_j}{\bar{V}_f} \bar{V}_s\right)$$

$$\bar{V}_s = \bar{V}_{max} = \frac{\bar{V}_f}{2} \tag{13.6}$$

Substituting these maximum values in equation 13.1 gives $\bar{Q}_{max}$

$$\bar{Q}_{max} = D_{max}\bar{V}_{max} = D_j\bar{V}_f/4 \tag{13.7}$$

Observations obtained by Greenshields gave the following values

$\bar{V}_f = 74$ km/h

$D_j = 121$ vehicles/km

from equation 13.5 $D_{max} = 61$ vehicles/km

from equation 13.6 $\bar{V}_{max} = 37$ km/h

from equation 13.7 $Q_{max} = 2239$ vehicles/h.

Figures 13.1, 13.2 and 13.3 show the relationships between speed and flow, density and speed and flow and concentration respectively.

A considerable number of relationships have been proposed between speed and density and the fit of some of these hypotheses to observed data has been investigated by Drake, Schofer and May[2]. Each hypothesis then affects the relationship between speed and flow and also between density and flow.

Greenberg[3] observed traffic flow in the north tube of the Lincoln Tunnel, New York City. He assumed that high density traffic behaved in a similar manner to a continuous fluid for which the equation of motion is

$$\frac{d\bar{V}_s}{dT} = -\frac{C^2}{D} \times \frac{\partial D}{\partial L}$$

where C is a constant and the remaining symbols are as previously defined. Then

$$\frac{d\bar{V}_s}{dT} = \frac{\partial \bar{V}_s}{\partial L} \times \frac{dL}{dT} + \frac{\partial \bar{V}_s}{\partial T} \times \frac{dT}{dT}$$

$$= \frac{\partial \bar{V}_s}{\partial L} \times \bar{V}_s + \frac{\partial \bar{V}_s}{\partial T}$$

or

$$\frac{\partial \bar{V}_s}{\partial L} \times \bar{V}_s + \frac{\partial \bar{V}_s}{\partial T} + \frac{C^2}{D} \times \frac{\partial D}{\partial L} = 0$$

From the equation of continuity of flow

$$\frac{\partial D}{\partial T} + \frac{\partial Q}{\partial L} = 0$$

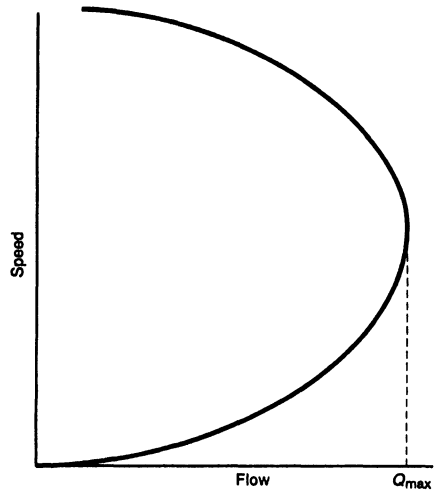


Figure 13.1 *The relationship between speed and flow for highway traffic flow*

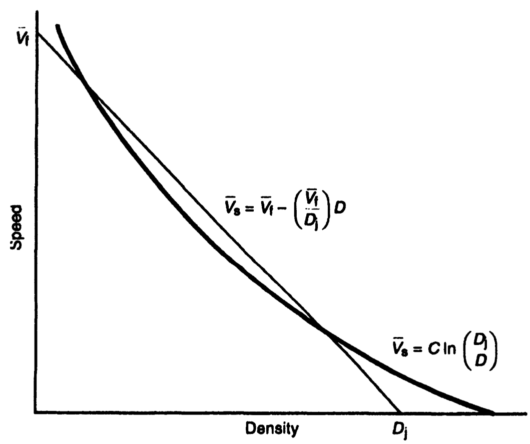


Figure 13.2 *The relationship between speed and density for highway traffic flow*

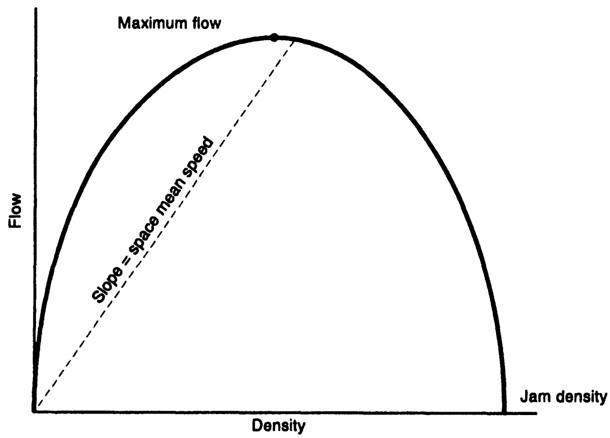


Figure 13.3 *The relationship between flow and density for highway traffic flow*

and

$$Q = \bar{V}_s D$$

then

$$\bar{V}_s = C \ln (D_j/D)$$

Using equation 13.1 gives

$$Q = CD \ln \frac{D_j}{D} \tag{13.8}$$

and

$$D = \frac{CD}{\bar{V}_s} \ln \frac{D_j}{D} \tag{13.9}$$

It is interesting to note that the Greenberg model does not give $\bar{V}_s = \bar{V}_f$. When $D = 0$ the boundary conditions are:

(1) When $Q = 0$, $D = 0$ or $D = D_j$.

(2) When $D = 0$, $\bar{V}_s = \infty$.

(3) When $D = D_j$, $\bar{V}_s = 0$.

The relationship between flow and density illustrated in figure 13.3 has been referred to as the fundamental diagram of traffic[4]. Its form has received considerable attention from those interested in highway flow and it can be described as follows.

The flow is obviously zero when the density is zero and at the jam density the flow may also be assumed to be zero. Between these limits the flow must rise to at least one maximum, often referred to as maximum capacity, to give a shape of the approximate form shown in figure 13.3. At any point on this curve the slope of the line joining that point to the origin is the space speed. The slope is obviously greatest at the origin and decreases to zero at the jam density.

Lighthill and Whitham[5] using a fluid-flow analogy have shown that the speed of waves causing continuous changes of volume through vehicular flow is given by dQ/dD which is the slope of the fundamental diagram, that is

$$V_w = \frac{dQ}{dD}$$

where V_w is the speed of the wave. From equation 13.1

$$V_w = \frac{d(\bar{V}_s D)}{dD} = \bar{V}_s + D \frac{d\bar{V}_s}{dD} \tag{13.10}$$

The space mean speed decreases with increasing density so that $d\bar{V}_s/dD$ is negative. Hence the wave speed is less than that of the traffic stream.

At low densities the second term of equation (13.10) approaches zero and wave velocity approaches the speed of the stream. At the maximum flow the wave is stationary relative to the road since dQ/dD is zero and at higher densities the waves move backwards relative to the stream.

When movement is viewed relative to the road, the wave moves forward at densities less than that of the maximum flow while the wave moves backwards at densities greater than that of the maximum flow.

Changes in density in a traffic stream will result in the production of differing waves travelling at differing velocities through the stream. An example of this could be a length of highway with a heavily trafficked entrance ramp. The wave in the low density length of highway downstream of the entrance ramp will travel forward relative to the highway at a greater speed than the wave in the higher density length of highway upstream of the entrance ramp. When these waves meet, a shock wave will form which will have a velocity

$$V_{sw} = \frac{Q_2 - Q_1}{D_2 - D_1}$$

where Q_1, Q_2 and D_1, D_2 are the respective points on the fundamental diagram for the low and high densities of flow.

Observations of the relationship between speed, flow and density

The sample observations shown in table 13.1 of the traffic flow in the fast lane of the Lincoln Tunnel, New York City by Olcott[6] have been adapted and are used to illustrate the fundamental relationships between density, flow and speed. Observations were made with an event recorder, the speed of vehicles being estimated by time of travel over a measured baseline. Summarised details of the observations are given in table 13.1.

TABLE 13.1

No. of vehicles observed in a 5 min. period	*Space mean speed* $\bar{V}_s$ *(km/h)*	*Flow* Q *(veh/h)*	*Density* D *(veh/km)*	D^2	$\bar{V}_s D$
97	27.0	1164	43.1	1857.6	1163.7
108	25.4	1296	51.0	2601.0	1295.4
104	30.7	1248	40.7	1656.6	1249.5
100	25.6	1200	46.9	2200.0	1200.6
113	34.8	1356	39.0	1521.0	1357.2
116	41.4	1392	33.6	1129.0	1391.0
116	30.2	1392	46.1	2125.2	1392.2
110	40.4	1320	32.7	1069.3	1321.1
115	39.7	1380	34.8	1211.0	1381.6
91	51.2	1092	21.3	453.7	1090.6
	Σ 346.4		Σ 389.2	Σ 15824.4	Σ 12842.9

The fit of this data to the relationship between speed and density proposed by Greenshields and given in equation 13.2 may be shown by estimating $\bar{V}_f$ and $(\bar{V}_f/D_j)$ using the method of least squares. In this method

$$\Sigma\, \bar{V}_s = n\, \bar{V}_f + (\bar{V}_f/D_j)\, \Sigma D$$

$$\Sigma\, \bar{V}_s D = \bar{V}_f \Sigma D + (\bar{V}_f/D_j)\, \Sigma D^2$$

Substituting the values obtained by summation in table 13.1 gives

$$\bar{V}_f = 71.4 \text{ km/h}$$

$$\frac{\bar{V}_f}{D_j} = -0.94$$

giving

$$D_j = 76 \text{ vehicles/km}$$

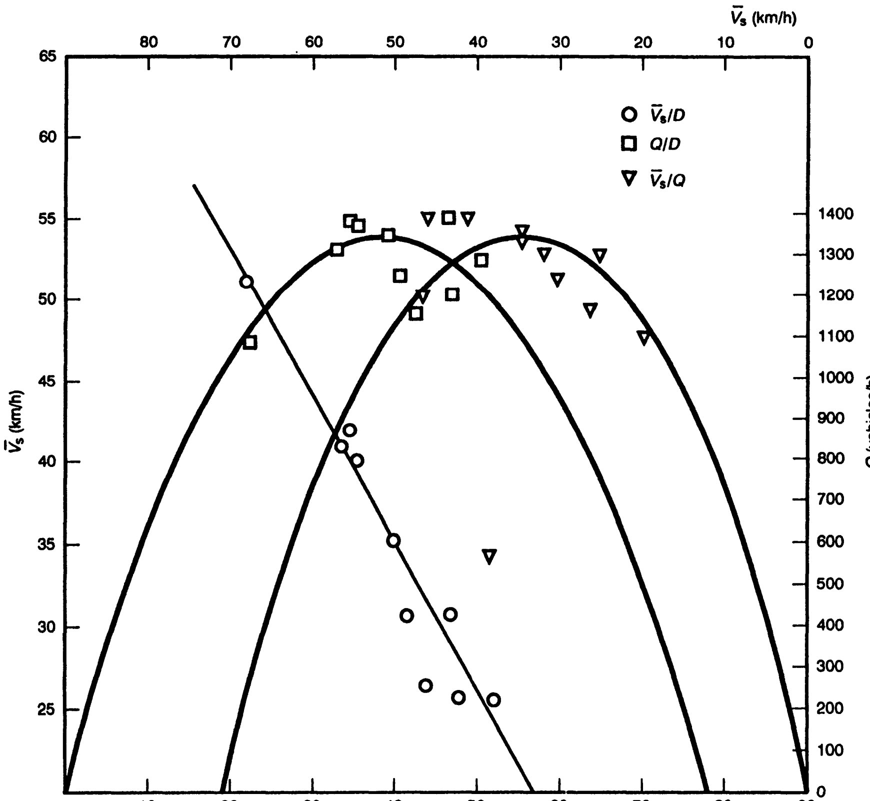


Figure 13.4 *Observed and fitted relationships between speed/density, flow/density and speed/flow*

The relationship is

$$\bar{V}_s = 71.4 - 0.94D$$

also

$$Q = 71.4D - 0.94D^2$$

and

$$Q = 1.06\,\bar{V}_s(71.4 - \bar{V}_s)$$

These relationships are illustrated graphically in figure 13.4 where observed values and the fitted speed/flow, speed/density and flow/density curves are shown.

From inspection of the graphical plot of flow and space mean-speed, or from differentiation, the maximum flow is found to be 1351 vehicles/h; it occurs when the space mean-speed of the stream is 35.7 km/h.

Empirical speed/flow relationships

Substantial studies have been undertaken in the United Kingdom and elsewhere to determine speed/flow relationships on different types of road, based on measurements. These relationships are fundamental for a variety of traffic/transport applications, such as traffic assignment modelling and in the economic evaluation of new road schemes; an accurate estimate of average speed is essential here for journey time calculations and economic evaluation. For example the Department of Transport's COBA program[7] has 12 road classifications, ranging from multi-lane motorways to central urban streets, each classification having its own speed/flow relationship obtained from empirical observations.

Empirical speed/flow relationships in the United Kingdom typically take the form shown in figure 13.5.

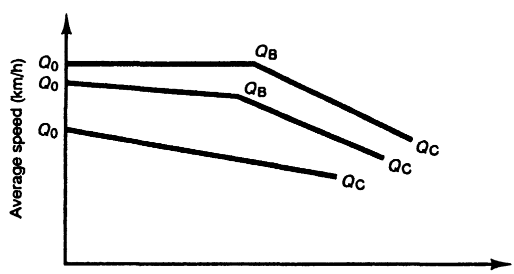


Figure 13.5 *Typical forms of speed/flow relationships*

These curves are defined by:

(1) the free speed of traffic, Q_0 (at zero or low flow);
(2) the capacity flow, Q_C;

(3) the location of 'break points', Q_B, where applicable;
(4) the influence of flow on speed, as represented.

Within each of the twelve COBA road classes, the values of Q_0, Q_B, Q_C and speed/flow slope have been found to be related to a number of factors concerned with specific road layout geometry, visibility, numbers of junctions and accesses, and so on.

For example, for rural single carriageway roads, the factors influencing travel speed were found to be:

(a) Bendiness, the total change of direction per unit distance (deg/km).
(b) Hilliness, total rise and fall per unit distance (m/km).
(c) Net gradient, net rise per unit distance, used only for one-way links (m/km).
(d) Total flow, all vehicles per standard lane (veh/hour/3.65 m lane).
(e) Average carriageway width (nominal 7.3 m or 8.5 m and above).
(f) Average verge width, both sides, including any metre strips (m).
(g) Total number, both sides, of lay-bys, side roads and accesses excluding houses and field entrances per km (no./km).
(h) Average sight distance (harmonic mean) (m).

For rural all-purpose dual carriageways and motorways the factors which influence travel speed are:

(a) Bendiness, the total change of direction per unit distance (deg/km).
(b) Sum of rises per unit distance (m/km).
(c) Sum of falls per unit distance (m/km).
(d) Total flow, all vehicles per standard lane (veh/hour/3.65 m lane).

For central area urban roads, average off-peak speed of all vehicles varies with the frequency of major intersections per km. For non-central area urban roads, average off-peak speed of all vehicles varies with the degree of development, which is defined as the proportion of the non-central road network that has frontage development. Shops, offices and industry are counted as 100 per cent and residential development as 50 per cent. The figure used should be the weighted average for all links in the non-central area.

Suburban roads are defined as major suburban routes in towns and cities where the speed limit is generally 40 mph. For these roads the average journey speed of all vehicles includes delays at junctions and varies with the following variables:

(a) Frequency of major intersections (no./km).
(b) Number of minor intersections and private drives per km.
(c) Percentage of development.
(d) Percentage of heavy vehicles.
(e) Total flow, all vehicles per standard lane (veh/hour/3.65 m lane).

References

1. B. D. Greenshields, A study of traffic capacity, *Proc. Highw. Res. Bd*, **14** (1934), Pt 1, 448-74
2. J. Drake, J. Schofer and A. May, A statistical analysis of speed–density hypotheses. Vehicular Traffic Science, *Proc. 3rd International Symposium on the Theory of Traffic Flow* (1965), Elsevier, New York (1967)
3. H. Greenberg, An analysis of traffic flow, *Ops. Res.*, **7** (1959), 1, 79–85
4. F. A. Haight, *Mathematical Theories of Traffic Flow,* Academic Press, New York (1963)
5. M. J. Lighthill and G. B. Whitham, A theory of traffic flow on long roads, *Proc. R. Soc.*, **A229** (1959), 317–45
6. E.A. Olcott, The influence of vehicular speed and spacing on tunnel capacity, Presented at the Informal Seminar in Operations Research, Johns Hopkins University, November 1954
7. Department of Transport, *The COBA Program*, London (1987)

Problems

(1) Greenshields proposed a linear relationship between speed and density. Using this relationship it was noted that on a length of highway the free speed V_f was 80 km/h and the jam density D_j was 70 vehicles/km.

(a) What is the maximum flow which could be expected on this highway?
(b) At what speed would it occur?

(2) Observations of the speed and flow through a highway tunnel showed that the relationship between speed and density was of the form

$$\overline{V}_s = 35.9 \ln \frac{180}{D}$$

where speeds are measured in km/h and densities in vehicles/km.

What is the jam density on this highway?

(3) The diagram below (figure 13.6) shows the relationship between volume and density for traffic flow on a highway. Four points on the curve are marked A, B, C and D; indicate for the traffic flow situations given below those which could be represented by point(s) A and/or B and/or C and/or D.

(a) Traffic flow conditions where the wave velocity was moving backwards relative to the roadway.

(b) Traffic flow conditions where the wave velocity was less than the space mean-speed of the traffic stream and the wave was moving forward relative to the roadway.

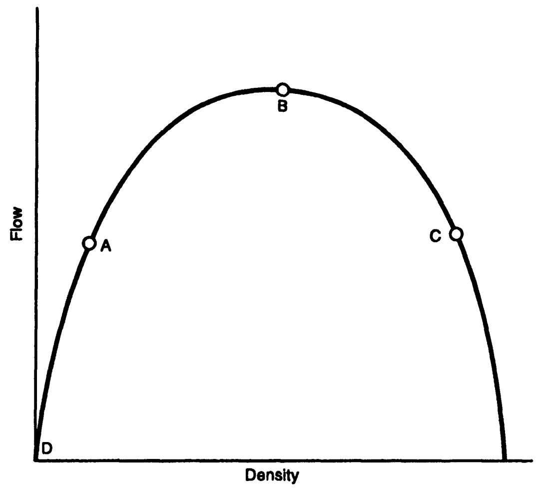


Figure 13.6

(c) Traffic flow conditions where the wave velocity and the space mean-speed stream velocity were equal.

(d) Congested traffic flow conditions where level of service E prevailed.

(e) Free flow conditions where level of service A prevailed.

Solutions

(1) (a) The maximum flow $\bar{Q}_{max}$ is given by

$$D_j \bar{V}_f / 4$$

when there is a linear relationship between speed and density, that is

$$\bar{Q}_{max} = \frac{70 \times 80}{4} = 1400 \text{ veh/h}$$

(b) The speed $\bar{V}_{max}$ of which the flow is a maximum is given by

$$\frac{\bar{V}_f}{2} = 40 \text{ km/h}$$

(2) The jam density D_j on a highway is given by

$$\bar{V}_s = C \ln \frac{D_j}{D}$$

Hence in the equation given

$$\bar{V}_s = 35.9 \ln \frac{180}{D}$$

the jam density Dj is 180 veh/km.

(3) (a) Traffic flow conditions where the wave velocity was moving backwards relative to the roadway (C).

(b) Traffic flow conditions where the wave velocity was less than the space mean speed of the traffic stream and the wave was moving forward relative to the roadway (A).

(c) Traffic flow conditions where the wave velocity and the space mean-speed velocity were equal (D).

(d) Congested traffic flow conditions where the level of service E prevailed (C).

(e) Free-flow conditions where level of service A prevailed (D).

14

Traffic speed distributions and estimation

Space mean and time mean-speed

One of the fundamental parameters for describing traffic flow is the speed, either of individual vehicles or of the traffic stream. It is of importance in work connected with the theory of traffic flow because of the fundamental connection between speed, flow and concentration when the movement of a traffic stream is being considered.

In traffic management a knowledge of the speed of vehicles is required for the realistic design of traffic signs, the layout of double white lines and the assessment of realistic speed limits. Geometric design of highways also requires that deceleration and acceleration and sight distances, superelevation and curvature must be related to an assumed design speed.

Because speed measurements are used for many purposes in highway traffic engineering and highway design, several differing definitions of speed are commonly used. Where vehicle speed is used to assess journey times, in connection with transportation studies that are concerned with travel over an area, it is the average journey speed that is important. On the other hand if vehicle speed is being used to investigate the accident potential of a section of highway then it is the spot speed which is required. The difference between these two types of speed can be illustrated by a simple example.

Five vehicles travel at 40, 50, 60, 70 and 80 km/h over a distance of 1 km. The mean of the spot speeds is 60 km/h. The mean travel time of the five vehicles is however 0.018 h so that the mean journey speed is 55.6 km/h. When speeds are measured at one point in space over a period it is the time mean-speed that is obtained. These individual speeds are often referred to as spot speeds. It is defined as

$$\overline{V}_t = \frac{\sum V_t}{n}$$

where $\overline{V}_t$ is the time mean-speed,
V_t are the individual speeds in time,
n is the number of observations.

Where speeds are averaged over space as is the case when the mean journey speed is calculated then it is the space mean-speed that is calculated, that is

$$\overline{V}_s = \frac{n}{\sum 1/\overline{V}_t}$$

where $\overline{V}_s$ is the space mean-speed,
n is the number of observations.

Wardrop[1] has shown that time and space mean-speeds are connected by the relationship

$$\overline{V}_t = \overline{V}_s + \frac{\sigma_s^2}{\overline{V}_s}$$

where σ_s is the standard deviation of $\overline{V}_s$.

Speed measurements

Most of the speed measurements made by highway traffic engineers produce a time mean-speed because they are obtained by the use of radar speedmeters or timing devices using short baselines. If in the latter case however the mean travel time is used to calculate the mean speed then it is the space mean-speed that is obtained. Speed measurements derived from successive aerial photographs can however be used to derive space mean-speed distributions directly by noting the travel distance of vehicles between successive exposures.

The simplest and cheapest method of obtaining speed data is to measure the time of travel over a measured distance or baseline. The accuracy of the method depends on the method of timing and where a stopwatch is used it is often assumed that a skilled observer can read to 0.2 s but difficulties of a constant reaction time often mean that the level of accuracy is approximately 0.5 s.

Where details of journey times are required then the time of travel over distances of 1 to 2 km may be obtained by the use of observers stationed at each end of this extended baseline and having synchronised stopwatches. The time of arrival of each vehicle into the measured length and the time of its departure are recorded together with some portion of the registration number. Subsequent comparison of times and numbers allows the journey times to be calculated. It is difficult for one observer to read and record more than one registration number every 5 s, even when only a portion of the registration number is recorded. For this reason only registration numbers ending with given digits are recorded by both observers according to the sample size required.

A more realistic form of device for use in present-day traffic conditions makes use of two pneumatic tubes attached to the road surface over a short baseline. Most of these instruments are transistorised allowing operation from portable power sources and some allow sampling of vehicle speeds to be carried out, a necessity in heavy traffic flow conditions. As in any observations sampling of the traffic flow should be carried out with care to avoid obtaining a biased sample. Care should be exercised to avoid selecting the first vehicle in a platoon and conversely selection of free-flowing single vehicles will tend to bias the resulting speed distribution towards the faster vehicles. Where speed observations are not classified into vehicle type then the combined observations should be in proportion to the traffic composition.

Pneumatic tubes are mostly used for temporary surveys of traffic speeds. Their vulnerability to breakage through the action of traffic often makes them unsuitable for more permanent installation. An alternative solution is to use a pair of inductive wire loops buried in the road surface. Traffic speed can be calculated by comparing the relative states of occupancy of the two loops. This technique is commonly used on United Kingdom roads for speed measurements. Examples include: (i) permanent installations at 0.5 km spacings on United Kingdom motorways, for monitoring speeds and flows; and (ii) speed loops on high-speed roads on the approaches to signal-controlled junctions or pelican-type pedestrian crossings. (Green durations may be extended on detection of a fast moving vehicle, to reduce accident risk at the junction.)

A speed measuring device that does not rely on timing the passage of a vehicle over a base line is the radar speedmeter. These devices operate on the fundamental principle that a radio wave reflected from a moving target has its frequency changed in proportion to the speed of the moving object. Radar meters have been extensively employed both for police and for research purposes. In these meters the radar beam may be transmitted directly along the highway making it necessary to site the meter on a curve or on the central reservation or, as is more usual in the newer models, a beam is transmitted at an angle across the carriageway.

Analysis of speed studies

Because in any speed study a considerable number of speeds are observed, statistical techniques are used to analyse the data obtained. Depending upon the accuracy of the data, the use to which the derived results are to be put and the number of observations obtained, a suitable class interval is chosen.

Table 14.1 shows speed observations obtained on a major traffic route. Individual speeds have been grouped into 4 km/h classes given in column 1 – an interval which reduces the data into an easily managed number of classes yet does not hide the basic form of the speed distribution. In the selection of class intervals thought should be given to the dial readings when visual observation of the speed is made. Most speeds will be recorded to the nearest dial reading and these form convenient mid-class marks.

The number of observations in each class, or the frequency, is given in column 2 and converted into the percentage in each class by dividing the individual values in column 2 by the sum of column 2. This percentage frequency is given in column 3. The cumulative number of observations or cumulative frequency is given in column 4. This column represents the number of vehicles travelling at a speed greater than the lower class limit. In column 5 the percentage cumulative frequency

TABLE 14.1

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Speed class (km/h)	*Frequency*	*Percentage frequency*	*Cumulative frequency*	*Percentage cumulative frequency*	*Deviation*	*(2) × (6)*	*(2) × (6)²*
44–47.9	1	0.286	1	0.286	–9	–9	81
48–51.9	2	0.571	3	0.857	–8	–16	128
52–55.9	2	0.571	5	1.429	–7	–14	98
56–59.9	4	1.143	9	2.571	–6	–24	144
60–63.9	11	3.143	20	5.714	–5	–55	275
64–67.9	24	6.875	44	12.571	–4	–96	384
68–71.9	40	11.429	84	24.000	–3	–120	360
72–75.9	48	13.714	132	37.714	–2	–96	192
76–79.9	63	18.000	195	55.714	–1	–63	63
80–83.9	40	11.429	235	67.143	0	0	0
84–87.9	34	9.714	269	76.857	1	34	34
88–91.9	29	8.286	298	85.143	2	58	116
92–95.9	25	7.143	323	92.286	3	75	225
96–99.9	13	3.714	336	96.000	4	52	208
100–103.9	5	1.429	341	97.429	5	25	125
104–107.9	3	0.857	344	98.286	6	18	108
108–111.9	1	0.286	345	98.571	7	7	49
112–115.9	2	0.571	347	99.143	8	16	128
116–119.9	2	0.571	349	99.714	9	18	162
120–123.9	1	0.286	350	100.000	10	10	100
	Σ 350					Σ –180	Σ 2980

is given. It is obtained by dividing the value in column 4 by the total number of speeds observed.

It is often assumed that speeds are normally distributed. To test this hypothesis it is necessary to estimate the mean speed and the standard deviation of the observed speeds. Both the mean and standard deviation can be calculated by the use of coding to reduce the arithmetic manipulation necessary. A class is selected which is considered likely to contain the mean speed, although it is not essential that the mean does lie within the class. The number of class deviations from this selected class is given in column 6. The sign of the deviation should be particularly noted.

These deviations are multiplied by the corresponding frequency, given in column 2, and the resulting value is entered in column 7.

The mean speed is then given by

$$\text{mid-class mark of selected class} + \frac{\text{class interval } \sum(\text{column 7})}{\sum(\text{column 2})}$$

$$82 - \frac{4.180}{350} = 79.9 \text{ km/h}$$

The standard deviation is given by

$$\text{class interval} \sqrt{\left[\frac{\sum (\text{frequency (deviation)}^2)}{\sum (\text{column 2})} - \left(\frac{\sum (\text{frequency deviation})}{\sum (\text{column 2})}\right)^2\right]}$$

The value of $\sum$(frequency × deviation) has already been calculated in column 6 and it is now necessary to calculate the frequency (deviation)^2 for each speed class. These values are given in column 8.

$$4\sqrt{\left[\frac{2980}{350} - \left(\frac{-180}{350}\right)^2\right]} = 11.6 \text{ km/h}$$

The two parameters of the normal distribution have now been determined and it is possible to calculate the theoretical values of speed frequency in each class by either of two methods. These are:

(1) by the use of a table of areas under the normal probability curve;
(2) by the use of the probit method as developed by Finney[2].

Using tables of the area under a normal probability curve it is possible to calculate the theoretical frequency assuming a normal distribution.

This is performed in a tabular manner in table 14.2. In column 1 the upper class limits used in the original speed observations of table 14.1 are given. The deviation of these class limits from the previously calculated mean are given in column 2 and then converted into standard deviations from the mean by dividing the value in column 2 by the previously calculated standard deviation. These values are given in column 3. From tables of the normal probability curve, the area under a normal curve between these class limits and the mean can be obtained and these values are given in column 4. The difference between successive values in column 4 gives the theoretical area under the normal curve between the class limits. This is the theoretical probability of a speed lying between the class limits and is given in column 5. When this probability is multiplied by the observed total frequency given in table 14.1 the theoretical frequency is obtained. This is given in column 6.

The second method, which uses the probit analysis approach, is based on the percentage cumulative frequency, given in column 5, table 14.1. This approach is described by Finney[2] and uses the relationship that when speeds are distributed normally the cumulative speed distribution may be written.

Percentage of vehicles travelling at a speed equal to, or less than

$$V = 1/(\sigma \times 2\pi) \int_{-\infty}^{V} \exp\left(-(V - \bar{V})^2 / 2\sigma^2\right) \mathrm{d}V$$

where $\bar{V}$ is the mean speed,

σ is the standard deviation of the speeds.

A demonstration of the fit of the observed cumulative speed distribution to a cumulative normal distribution may be obtained by plotting the probit of the percentage of vehicles travelling at or less than a certain speed, against the speed upper class limit. Values of probits may be obtained from the suggested reading or can be obtained from figure 14.1. The use of this technique converts a cumulative normal curve into a straight line whose equation is

$$\text{Probit of percentage of vehicles travelling at a speed} < V = 5 + \frac{1}{\sigma}(V - \bar{V})$$

Using the derived values of σ and $\overline{V}$ this gives

$$\text{Probit of percentage of vehicles travelling at a speed} < V = 5 + 0.0862(V - 79.9)$$
$$= 0.0862V - 1.6887 \qquad (14.1)$$

TABLE 14.2

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Upper speed class limit (km/h)	*Column 1 minus mean speed*	*Column 2 divided by standard deviation*	*Normal area*	*Probability*	*Theoretical frequency*	*Observed frequency*	$\frac{((6)-(7))^2}{(6)}$
44	−35.9	−3.10	−0.499				
48	−31.9	−2.75	−0.497	0.002	0.7	1	
52	−27.9	−2.40	−0.492	0.005	1.8	2	2.27
56	−23.9	−2.06	−0.480	0.012	4.2	2	
60	−19.9	−1.72	−0.457	0.023	8.1	4	
64	−15.9	−1.37	−0.415	0.042	14.7	11	0.93
68	−11.9	−1.025	−0.349	0.066	23.1	24	0.04
72	−7.9	−0.680	−0.252	0.097	33.9	40	1.10
76	−3.9	−0.336	−0.132	0.119	41.9	48	0.89
80	+0.1	0.009	0.004	0.137	48.0	63	4.69
84	+4.1	0.354	0.138	0.134	46.9	40	1.02
88	+8.1	0.70	0.258	0.120	42.0	34	1.52
92	+12.1	1.04	0.351	0.093	32.6	29	0.40
96	+16.1	1.39	0.418	0.067	23.4	25	0.11
100	+20.1	1.74	0.459	0.041	14.3	13	0.12
104	+24.1	2.08	0.481	0.022	7.7	5	
108	+28.1	2.42	0.492	0.011	3.8	3	
112	+32.1	2.76	0.497	0.005	1.8	1	0.01
116	+36.1	3.11	0.499	0.002	0.7	2	
120	+40.1	3.46	0.500	0.001	0.4	2	
124	+44.1	3.81	0.500	0.000	0	1	
							Σ 13.10

This line is plotted in figure 14.1 together with the observed cumulative frequency distribution obtained from column 5 of table 14.1. The close agreement except at high speeds can be seen.

Using the equation above it is possible to calculate the theoretical probit of vehicles travelling at a speed > V, and using either the scale given in figure 14.1 or the tables given by Finney it is possible to calculate the theoretical percentage acceptance. This is done in columns 2 and 3 of table 14.3. The difference between successive values in column 3 gives the percentage frequency, which is given in column 4. When this column is multiplied by the total number of speeds observed the frequency is obtained and this is given in column 5. The values given in this column differ slightly from those derived in column 6 of table 14.2. This is due to errors involved in the interpolation of values in the tables used in the calculation.

It is often necessary to calculate whether the observed speed frequencies differ significantly from the expected speed frequencies. It is possible to estimate whether

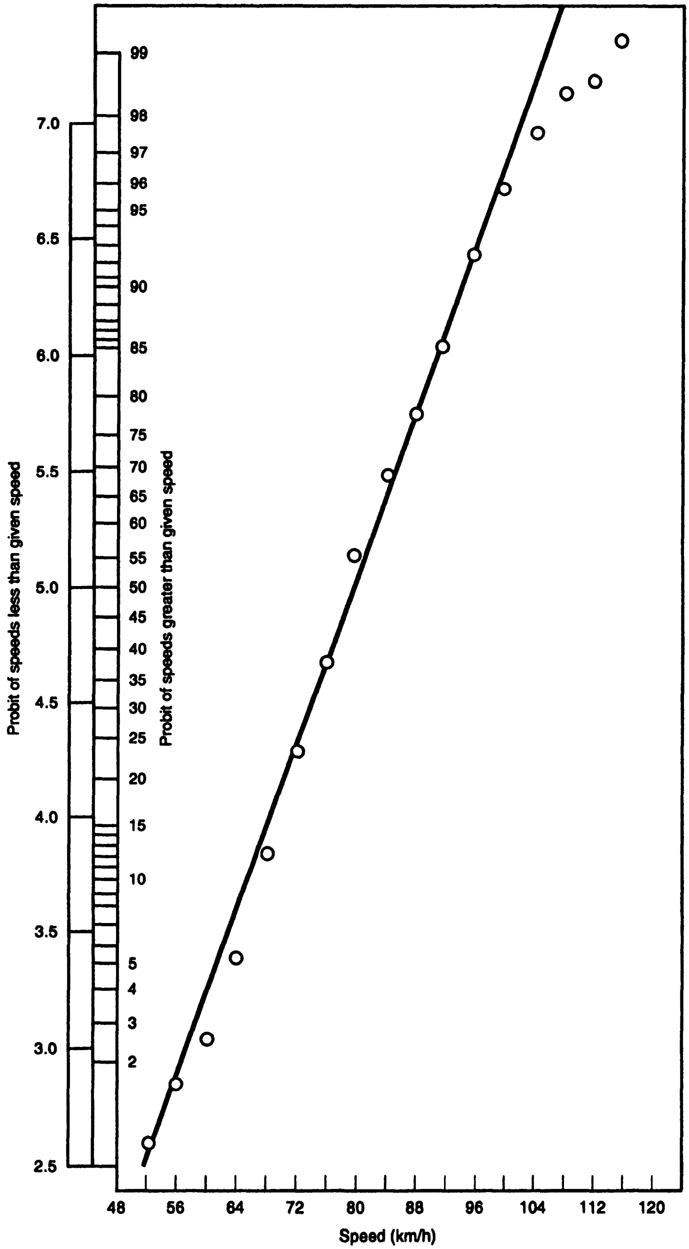


Figure 14.1 *The observed and fitted speed distribution*

the observed and expected results differ significantly by the use of the chi-squared test of significance where chi-squared is given by

$$\chi^2 = \sum \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}}$$

If χ^2 exceeds the critical value at the 5 per cent level of significance then it is said that there is a significant difference between the observed and the expected values. The value of chi-squared is calculated in column 8 of table 14.2 and summed at

TABLE 14.3

1	*2*	*3*	*4*	*5*
Speed class limit (km/h)	*Probit cumulative percentage*	*Cumulative percentage*	*Percentage frequency*	*Frequency*
44	2.1031	0.0		
48	2.2502	0.3	0.3	0.1
52	2.5950	0.8	0.5	1.8
56	2.9398	2.0	1.2	4.2
60	3.2746	4.3	2.3	8.1
64	3.6294	8.6	4.3	15.1
68	3.9742	15.3	6.7	23.5
72	4.3190	24.8	9.5	33.3
76	4.6638	36.9	12.1	42.4
80	5.0086	50.3	13.4	46.9
84	5.3534	63.8	13.5	47.3
88	5.6982	75.8	12.0	42.0
92	6.0430	85.2	9.4	32.9
96	6.3878	91.7	6.5	22.8
100	6.7326	95.9	4.2	14.7
104	7.0774	98.1	2.2	7.7
108	7.4222	99.2	1.1	3.9
112	7.7670	99.7	0.5	1.8
116	8.1118	99.9	0.2	0.7
120	8.4566		0.1	0.4
124	8.8014			

the bottom of the column to give a value of 13.10. It should be noted that where an individual frequency is less than 8 both the observed and the expected frequencies have been combined with other classes.

The critical value of chi-squared which must not be exceeded depends on the number of degrees of freedom. This is given by the number of rows in column 8 (since the greater the number of individual values forming the sum of chi-squared the greater will be its expected value) minus the constraints or the number of parameters which were taken from the observed frequency values to calculate the theoretical frequency values. In this case the constraints are the mean speed, the standard deviation of speeds and the total frequency of speeds. This means the number of degrees of freedom of the calculated value of chi-squared is $12 - 3 = 9$.

Tables of chi-squared given in most statistical text books show that the critical value of chi-squared for 9 degrees of freedom at the 5 per cent level of probability is 15.5. The calculated value is 13.10 and so it can be assumed that the observed speed distribution may be represented by a normal distribution.

Having established that speeds are normally distributed it is possible to use the well-known properties of the normal distribution to obtain detailed information on the speed distribution. For example, it is often desirable to calculate the 85 percentile speed and this can be obtained by the use of equation 14.1. The probit of 85 per cent is 6.0364 giving

$$6.0364 = 0.0862V - 1.6887$$

or

$$V(85 \text{ percentile}) = 89.6 \text{ km/h}$$

It is also a property of the normal distribution that 68.27 per cent of all observations will lie within plus or minus one standard deviation of the mean value. This means that approximately two-thirds of all vehicles will be travelling at a speed between 79.9 − 11.6 km/h and 79.9 + 11.6 km/h, that is between 68.3 and 91.5 km/h. Examination of column 2 of table 14.1 shows that this is correct.

A knowledge of the mathematical form of the speed distribution is also of considerable importance in many theoretical studies of traffic flow and in the simulation of driver behaviour in many highway situations.

Macroscopic determination of speed and flow of a highway traffic stream

The relationships considered in Chapter 13 between speed, flow and density of a traffic stream considered traffic flow in a microscopic sense in that the speeds of, and headways between, individual vehicles have been used to obtain relationships for the whole traffic stream. An alternative approach to the determination of speed/flow relationships is the use of macroscopic relationships obtained by the observation of stream behaviour.

Speed and flow of a moving stream of vehicles may be obtained by what has become known as the moving car observer method. Wardrop and Charlesworth[3] have described how stream speed and flow may be estimated by travelling in a vehicle against and with the flow. The journey time of the moving observation vehicle is noted, as is the flow of the stream relative to the moving observer. This means that when travelling against the flow the relative flow is given by the number of vehicles met and while travelling with the flow the relative flow is given by the number of vehicles that overtakes the observer minus the number that the observer overtakes. Then

$$Q = \frac{(x + y)}{(t_a + t_w)}$$

and

$$\bar{t} = t_w - y/Q$$

where Q is the flow,

$\bar{t}$ is the mean stream-journey time,
x is the number of vehicles met by the observer, when travelling against the stream,
y is the number of vehicles that overtakes the observer minus number overtaken,
t_a is the journey time against the flow,
t_w is the journey time with the flow.

These relationships are determined as follows. Consider a stream of vehicles moving along a section of road, of length l, so that the average number Q passing through the section per unit time is constant. This stream may be regarded as consisting of flows

Q_1 moving with speed V_1

Q_2 moving with speed V_2

etc. etc.

Suppose an observer travels with the stream at speed V_w and against the stream at speed V_a, then the flows relative to him, of vehicles with flow Q_1 and speed V_1 are

$$Q_1\,(V_1 - V_w)/V_1$$

and

$$Q_1\,(V_1 + V_a)/V_1$$

respectively.

If t_w is the journey time $1/V_w$, t_a is the journey time $1/V_a$ and t_1 is the journey time $1/V_1$, etc., then

$$Q_1\,(t_a + t_1) = x_1 \quad \text{and} \quad Q_2\,(t_a + t_2) = x_2, \text{ etc.}$$

$$Q_1\,(t_w - t_1) = y_1 \quad \text{and} \quad Q_2\,(t_w - t_2) = y_2, \text{ etc.}$$

where x_1, x_2, etc., are the number of vehicles travelling at speeds V_1, V_2, etc. met by the observer when travelling against the stream and y_1, y_2, etc. are the number of vehicles overtaken by the observer minus the number overtaking the observer.

Summing over x and y

$$x = x_1 + x_2 + x_3 + \ldots$$

$$y = y_1 + y_2 + y_3 + \ldots$$

$$x = Q_1(t_a + t_1) + Q_2(t_a + t_2) + \ldots$$

$$y = Q_1(t_w - t_1) + Q_2(t_w - t_2) + \ldots$$

$$x = Qt_a + \sum(Q_1t_1 + Q_2t_2 + \ldots)$$

$$y = Qt_w - \Sigma(Q_1t_1 + Q_2t_2 + \ldots)$$

and

$$\bar{t} = \frac{(Q_1t_1 + Q_2t_2 + Q_3t_3 + \ldots)}{Q}$$

so that

$$x = Qt_a + Q\bar{t}$$

$$y = Qt_w - Q\bar{t}$$

or

$$Q = (x + y)/(t_a + t_w)$$

and

$$\bar{t} = t_w - y/Q$$

so allowing the stream speed to be calculated.

The application of this method of measuring speed and flow is illustrated by the following field data, which was obtained to estimate the two-way flow on a highway. To reduce the number of runs required by the observation vehicles, details of the relative flow of both streams are obtained for each run of the observer's vehicle. For this reason, details are given of two relative flows in the data obtained when the observer is travelling eastwards and when he is travelling westwards. These observations are given in table 14.4.

TABLE 14.4

Line	Time of commencement of journey	Journey time	No. of vehicles met	No. of vehicles overtaking observer	No. of vehicles overtaken by observer
Observer travelling to east					
1	09.20	2.51	42	1	0
2	09.30	2.58	45	2	0
3	09.40	2.36	47	2	1
4	09.50	3.00	51	2	1
5	10.00	2.42	53	0	0
6	10.10	2.50	53	0	1
Observer travelling to west					
7	09.25	2.49	34	2	0
8	09.35	2.36	38	2	1
9	09.45	2.73	41	0	0
10	09.55	2.41	31	1	0
11	10.06	2.80	35	0	1
12	10.15	2.48	38	0	1

Length of highway test section 1.6 km.

The information relating to the *highway flow to the east* is abstracted and tabulated in table 14.5.

TABLE 14.5

Time of commencement of journey	Relative flow rate: with observer	Relative flow rate: against observer	t_a (min)	t_w (min)	Q (veh/min)	$\bar{t}$ (min)	V (km/h)
09.20	1			2.51			
09.25		34	2.49		7.0	2.37	41
09.30	2			2.58			
09.35		38	2.36		8.1	2.33	42
09.40	1			2.36			
09.45		41	2.73		8.3	2.24	43
09.50	1		3.00				
09.55		31		2.41	6.3	2.84	34
10.00	0			2.42			
10.05		35	2.89		6.6	2.42	40
10.10	−1			2.50			
10.15		38	2.48		7.4	2.36	41

Similarly, the information relating to the *highway flow to the west* is abstracted and tabulated in table 14.6.

TABLE 14.6

Time of commencement of journey	Relative flow rate: with observer	Relative flow rate: against observer	t_a (min)	t_w (min)	Q (veh/min)	$\bar{t}$ (min)	V (km/h)
09.20		42	2.51				
09.25	2			2.49	8.8	2.38	41
09.30		45	2.58				
09.35	1			2.36	9.3	2.25	43
09.40		47	2.36				
09.45	0			2.73	9.3	2.62	37
09.50		51	3.00				
09.55	1			2.41	9.6	2.31	42
10.00		53	2.42				
10.05	−1			2.89	9.8	2.79	35
10.10		53	2.50				
10.15	−1			2.48	10.4	2.39	40

If the mean stream speed and flow is required then these individual values may be averaged, to give

stream velocity = 40 km/h

stream flow = 9.5 veh/min

References

1. J. G. Wardrop, Some theoretical aspects of road traffic research, *Proc. Instn Civ. Engrs* 2 (Pt 1) (1952), 2, 325–62
2. D.J. Finney, *Probit analysis, a statistical treatment of the signed response curve*, Cambridge University Press (1947)
3. J.G. Wardrop and G. Charlesworth, A method of estimating speed and flow of traffic from a moving vehicle, *J. Instn Civ. Engrs*, **3** (Pt 2) (1954), 158–71

Problems

(1) Drivers in a vehicle testing programme travel around a race track at a constant speed measured by radar speedmeters and by measuring their travel time around the track. The mean speed of all the vehicles was obtained by the two methods and compared.

(a) If all the vehicles travel at the same speed as indicated by their corrected speedometers, will the
(i) two mean speeds be the same?
(ii) the radar speedmeter mean be the higher?
(iii) the radar speedmeter mean be the lower?

(b) If the vehicles travel at differing speeds as indicated by their corrected speedometers will the
(i) two mean speeds be the same?
(ii) the radar speedmeter mean be the higher?
(iii) the radar speedmeter mean be the lower?

(2) The speed distribution on a rural trunk road was noted to be normally distributed with a mean speed of 90 km/h and a standard deviation of 20 km/h. Using the probit technique estimate the 85 percentile speed on the highway.

(3) In a stream of vehicles, 30 per cent of the vehicles travel at a constant speed of 60 km/h, 30 per cent at a constant speed of 80 km/h and the remaining vehicles travel at a constant speed of 100 km/h. An observer travelling at a constant speed of 70 km/h with the stream over a length of 5 km is passed by 17 vehicles more than he passes. When the observer travels against the stream at the same speed and over the same length of highway, the number of vehicles met is 303.
(a) What is the mean speed and flow of the traffic stream?
(b) Is the time mean or the space mean-speed obtained by this technique?
(c) How many vehicles travelling at 100 km/h pass the observer, while he travels with the stream?

Solutions

(1) When the speeds are measured by a radar speedmeter and averaged it is the time mean-speed $\overline{V}_t$ that is obtained whereas when speeds are measured from travel time it is the space mean-speed $\overline{V}_s$ that is calculated. The relationship between the two mean speeds is given by

$$\overline{V}_t = \overline{V}_s + \frac{\sigma_s^2}{\overline{V}_s}$$

(a) In this case all the vehicles travel at the same speed and the standard deviation of the space mean-speeds σ_s is zero so that the two mean speeds are the same.
(b) In this case the vehicles travel at differing speeds and since the standard deviation of speeds is positive the mean speed obtained from the radar speedmeter will be higher than the mean speed obtained from journey times.

(2) When speeds can be represented by the cumulative normal distribution then the following relationship may be used.

Probit of percentage of vehicles travelling at a speed $> V = 5 + \frac{1}{\sigma}(V - \overline{V})$

Using the values given

Probit of percentage of vehicles travelling at a speed $> V = 5 + \frac{1}{20}(V - 90)$

The probit of 85 per cent is 6.03, that is

$$6.03 = 5 + \frac{1}{20}(V_{85} - 90)$$

or

$$V_{85} = 110.6 \text{ km/h}$$

(3) (a) The flow Q of a traffic stream may be obtained from

$$\frac{x + y}{t_a + t_w}$$

The time of travel t_a and t_w is 5 km at 70 km/h. Then

$$Q = \frac{303 + 17}{5/70 + 5/70}$$

$= 2240$ vehicles /h

and

$$\bar{t} = \frac{5}{70} - \frac{17}{2240}$$

$$= 0.0714 - 0.0076$$

$$= 0.0638$$

The mean speed of the stream = 5/0.0638 km/h
= 78.3 km/h

(b) The speed obtained by the moving observer technique is calculated from a journey time and hence it is the space mean-speed that is obtained.

(c) It has been shown that

$$Q_1(t_w - t_1) = y_1$$

where Q_1 is the flow of vehicles with a speed of 100 km/h. In this case it is 0.4 × 2240 or 896 veh/h,
t_w is the travel time of the observer, in this case, 0.071 h,
t_1 is the travel time of the vehicles travelling at 100 km/h, that is 0.05 h,
y_1 is the number of vehicles travelling at 100 km/h that overtake the observer.

Then

$$896\,(0.071 - 0.05) = y_1$$
$$= 18 \text{ or } 19 \text{ vehicles}$$

15

Highway link design

The geometric design of new road links in the United Kingdom is now undertaken in accordance with the Department of Transport Departmental Standard TD9/93 'Highway Link Design'. The overall requirement is that an alignment is provided for a design speed which is consistent with the anticipated vehicle speeds on the road.

Design speed

The first requirement in geometric design is to select an appropriate design speed for the type and function of road involved. The approach adopted in the United Kingdom is to specify design speed bands and to produce an alignment so that 85 per cent of drivers will travel at or below the selected design speed. The concept of design speed bands arose from a consideration of speed distributions, which have been shown to remain reasonably constant on roads with similar geometric properties. From measurements, it has been found that the relationship between the 99th percentile, the 85th percentile and the 50th percentile speeds is given by

$$\frac{V_{99}}{V_{85}} \approx \frac{V_{85}}{V_{50}} \approx 4\sqrt{2} = 1.19$$

This consistency leads to a structured framework of design speeds, and the concept of design speed bands, emerging from table 15.1.

TABLE 15.1

50th %ile speed (km/h)	*85th %ile speed (km/h)*	*99th %ile speed, (km/h)*
100	120	145
85	100	120
70	85	100
60	70	85

With this framework, a design speed of 100 km/h would indicate that the 85th percentile speed lies in the range 85 to 99.9 km/h; thus, the mean speed could be as low as 70 km/h and in other cases the 99th percentile speed could be as high as 119.9 km/h. It is clearly important here that the road alignment is safe for the 99th percentile driver, even though the road is designed for the 85th percentile speed.

Speed models

The unified model used in TD9/93[1] for the prediction of speeds on links is

$$\bar{V}_{50\text{wet}} = 110 - L_c - A_c \text{ km/h} \tag{15.1}$$

where $\bar{V}_{50\text{wet}}$ = mean wet weather speed (km/h),
L_c = layout constraint,
A_c = alignment constraint.

As the terms suggest, A_c is the degree of constraint on speed incorporated by the geometric alignment while L_c is the degree of constraint on speed imposed by the road cross-section, verge width and frequency of junctions and accesses.

Equation 15.1, derived from empirical observations, is used with reference to table 15.2 for L_c and the following equations for A_c:

$$A_c = 6.6 + B/10 \quad \text{for dual carriageways}$$
$$= 12 - VISI/60 + 2B/45 \quad \text{for single carriageways}$$

where B = bendiness (degrees/km),
$VISI$ = harmonic mean visibility (m)

TABLE 15.2
Layout constraint L_c (km/h) (ref. 1)

Road type	S2				WS2		D2AP		D3AP	D2M	D3M
Carriageway width (ex. metre strips)	6 m		7.3 m		10 m		Dual 7.3 m		Dual 11 m	Dual 7.3 m & hard shoulder	Dual 11 m & hard shoulder
Degree of access at junctions	H	M	M	L	M	L	M	L	L	L	L
Standard verge width	29	26	23	21	19	17	10	9	6	4	0
1.5 m Verge	31	28	25	23							
0.5 m Verge	33	30									

There is no research data available for 4-lane single carriageway roads between 12 and 14.6 m width (S4). In the limited circumstances for their use, design speed should be estimated assuming a normal D2AP with a layout constraint of 15–13 km/h

L = Low access numbering 2 to 5 per km
M = Medium access numbering 6 to 8 per km
H = High access numbering 9 to 12 per km

$$VISI = \frac{n}{\frac{1}{v_1} + \frac{1}{v_2} + \ldots + \frac{1}{v_n}}$$

where n = number of visibility observations along the road,
v_1 = visibility (sight distance) at point 1, etc.

A further constraint on speed arises from mandatory speed limits, which are:

	mph	km/h
Motorways and dual carriageways	70	112
Single carriageways	60	96

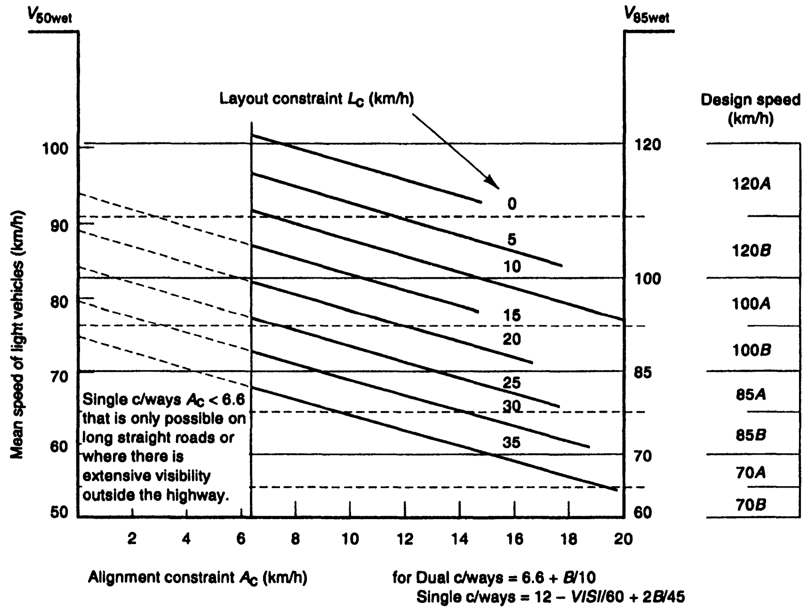


Figure 15.1 *Selection of design speed (rural roads) (ref. 1)*

For new rural roads, design speed is derived from figure 15.1 which shows the variation in speeds for a given L_c against A_c. The design speeds are arranged in bands, within which suffixes A and B indicate the higher and lower categories of each band. The design process is to:

(1) select trial design speed;
(2) draw up an initial alignment;
(3) measure A_c over sections of at least 2 km;
(4) calculate design speed from figure 15.1 and compare with trial design speed;

(5) as necessary, relax alignment to achieve cost or environmental savings, or upgrade alignment to achieve required design speed;
(6) recheck resulting speeds following any changes in alignment.

For urban roads, design speeds are selected with reference to the speed limits envisaged for the road, permitting a small margin for speeds in excess of the speed limit. The speeds shown in table 15.3 apply.

TABLE 15.3

Speed limit		*Design speed*
mph	*km/h*	*(km/h)*
30	48	60*B*
40	64	70*A*
50	80	85*A*
60	96	100*A*

Design speed-related parameters

The design speed adopted for a road dictates the minimum geometric parameters for the design, in terms of horizontal curvature, stopping sight distance and overtaking sight distance, as shown in table 15.4. These elements of design are introduced later in this chapter. Table 15.4 also illustrates the three-tier hierarchy of

TABLE 15.4

Design speed (km/h)	*120*	*100*	*85*	*70*	*60*	*50*	V^2/R
Stopping sight distance (m)							
Desirable minimum	295	215	160	120	90	70	
One step below desirable minimum	215	160	120	90	70	50	
Horizontal curvature (m)							
Minimum R* without elimination of adverse camber and transitions	2880	2040	1440	1020	720	510	5
Minimum R* with superelevation of 2.5%	2040	1440	1020	720	510	360	7.07
Minimum R* with superelevation of 3.5%	1440	1020	720	510	360	255	10
Desirable minimum R with superelevation of 5%	1020	720	510	360	255	180	14.14
One step below desirable minimum R with superelevation of 7%	720	510	360	255	180	127	20
Two steps below desirable minimum R with superelevation of 7%	510	360	255	180	127	90	28.28
Vertical curvature							
Desirable minimum* crest K value	182	100	55	30	17	10	
One step below desirable min. crest K value	100	55	30	17	10	6.5	
Absolute minimum sag K value	37	26	20	20	13	9	
Overtaking sight distances							
Full overtaking sight distance (FOSD)(m)	*	580	490	410	345	290	
FOSD overtaking crest K value	*	400	285	200	142	100	

*Not recommended for use in the design of single carriageways.

geometric design – desirable minimum standards, relaxations and departures. The desirable minimum standard provides an alignment appropriate for the 85th percentile driver; a relaxation provides an alignment one or more design speed bands below this according to road type, while a departure produces an alignment two or more design speed bands below. Relaxations may be acceptable in certain situations where the strict application of desirable minimum standards would lead to disproportionately high construction costs or severe environmental impact on people, properties and/or landscapes. DOT Standard TD9/93 provides advice on the circumstances where relaxations may be acceptable and states the situations where relaxations are not permitted (such as on the immediate approaches to junctions).

Departures from standards may be allowed in situations of exceptional difficulty. All such proposals have to be submitted to the Department of Transport for scrutiny.

Stopping sight distance

The stopping sight distance (SSD) is the theoretical sight distance required by a driver to stop a vehicle when faced with an unexpected obstruction in the road. SSD is made up of

(1) Perception–reaction distance, which is the distance travelled from the time the driver sees the obstruction to the start of braking. A (safe) perception–reaction time of 2 seconds is used, although this varies in reality with age, fatigue, distraction, etc.

(2) Braking distance, which is the distance travelled during braking. Maximum comfortable and absolute maximum values of $0.25g$ and $0.375g$ are used for design standards. (Note that these do not relate to the times and distances which can be achieved by good drivers in emergency braking conditions on dry roads.)

SSD is measured between a minimum driver's eye height of between 1.05 m and 2.00 m to an object height of between 0.26 m and 2.00 m above the road surface. It is necessary to provide forward visibility in both horizontal and vertical planes between points in the centre of the lane nearest the inside of the curve.

Full overtaking sight distance

On single carriageway roads, safe overtaking manoeuvres require the provision of adequate forward visibility, termed full overtaking sight distance (FOSD). Four elements have been identified in the overtaking manoeuvre which control FOSD requirements:

(1) perception–reaction distance, which is the distance travelled by the vehicle while the driver decides whether or not to overtake;
(2) the overtaking distance (D1), which is the distance travelled by the vehicle to complete the overtaking manoeuvre;

(3) the closing distance (D2), which is the distance travelled by the oncoming vehicle while the overtaking manoeuvre is taking place;
(4) the safety distance (D3), which is the distance required for clearance between the overtaking vehicle and the oncoming vehicle at the instant the overtaking vehicle has returned to its own lane.

These elements (D1–D3) are illustrated in figure 15.2. The time taken to complete an overtaking manoeuvre is variable, depending particularly on the relative speeds of the vehicles involved. However, 85 per cent of manoeuvres have been found to take 10 seconds or less, and this value has been used to produce the design speed related values of FOSD in table 15.4.

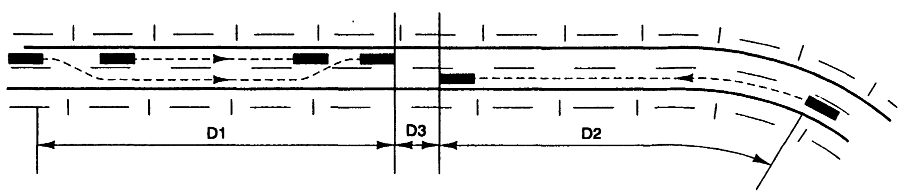


Figure 15.2 *Elements of the overtaking manoeuvre*

On single carriageway roads it is necessary to provide sufficient sections with FOSD to allow safe overtaking. This requires the identification of overtaking sections which could be:

(1) 'level' overtaking sections, which are straight, or nearly straight, sections;
(2) climbing lane sections, where an additional lane is provided on uphill sections, particularly to allow slower vehicles to be overtaken;
(3) single-lane downhill sections, constrained by a solid/broken white line where the combination of visibility and horizontal curvature provides clear opportunities for overtaking when the opposing traffic permits;
(4) sections of dual carriageway, or single carriageway four-lane roads, which provide overtaking opportunities in an otherwise S2 road.

These requirements have led to recommendations *against* medium/large radius curves for single-carriageway horizontal curve design, as they inhibit the design of clear overtaking sections. They produce long, dubious overtaking conditions for vehicles travelling in the left-hand curve direction and simply reduce the length of overtaking straight which could otherwise be achieved.

Horizontal alignment

The horizontal alignment of a road consists of a series of interconnected straights, transition curves and circular curves, although cubic spline layouts may also be designed using computer-aided design packages. The chosen alignment for a road depends on a variety of economic, operational, environmental and other factors, although the two issues of driver comfort and safety are of particular importance in the design of individual elements of the alignment. Horizontal curves are usually circular in plan and their radius is controlled by the design speed of the road, sight distance requirements and superelevation.

Superelevation

The detailed design of the horizontal curvature stems from a consideration of vehicle dynamics. As a vehicle moves round a curve it is subject to a centrifugal force which causes it to slide outwards or to overturn. This force is resisted by the side friction between the wheels and the road surface, but it is also possible to raise (or superelevate) the outside of the curve, to counteract these sideways forces. Figure 15.3 illustrates the forces involved.

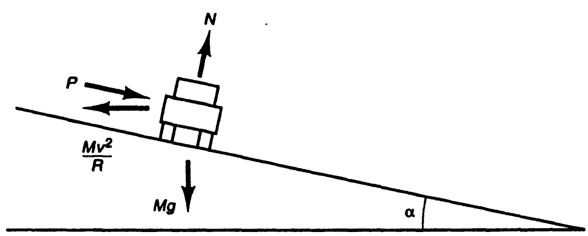


Figure 15.3 *Superelevation and forces on a vehicle travelling around a curve*

Resolution of these forces at right angles to the road surface gives

$$\left[\frac{Mv^2}{R}\right] \cos\alpha = Mg \sin\alpha + P = N \tag{15.2}$$

Resolving parallel to the road surface gives

$$P = \mu \; [Mg \cos\alpha + (Mv^2/R) \sin\alpha] \tag{15.3}$$

We therefore have

$$\frac{v^2}{gR} = \tan\alpha + \mu + (\mu v^2/gR) \tan\alpha \tag{15.4}$$

The last term of this expression is very small and can be ignored. If the angle of superelevation, α, is expressed as a slope, e, we have

$$\frac{v^2}{gR} = e + \mu \tag{15.5}$$

If v (m/s) is replaced with V (km/h) and g = 9.81 m/s^2, we have

$$\frac{V^2}{127R} = e + \mu \tag{15.6}$$

If the mean speed of vehicles is taken as 67 per cent of the design speed and the term μ is omitted, the application of superelevation, equation 15.6 becomes

$$e = \frac{V^2}{283R}$$

or with e as a percentage

$$e\ (\%) = \frac{V^2}{2.828R} \tag{15.7}$$

Table 15.4 shows the application of equation 15.7 in terms of minimum radii of horizontal curves required for given design speeds and superelevation.

Where superelevation is not provided or required, such as on straight or nearly straight sections of road, a crossfall, or camber is provided to aid drainage. This usually applies to roads where the combination of design speed, V, and curve radius, R, provides a lateral acceleration (V^2/R) of less than 5. The crossfall, which is typically 2.5 per cent, applies across the width of the carriageway or, more usually, the centre line (or crown) of the road is the highest point, with a crossfall of 2.5 per cent to each kerbline. This road cross-section can cause 'adverse camber' to occur if maintained on curved sections of road. In this case, a higher radius curve is required for a particular design speed, as shown in table 15.4, to maintain a similar level of sideways force.

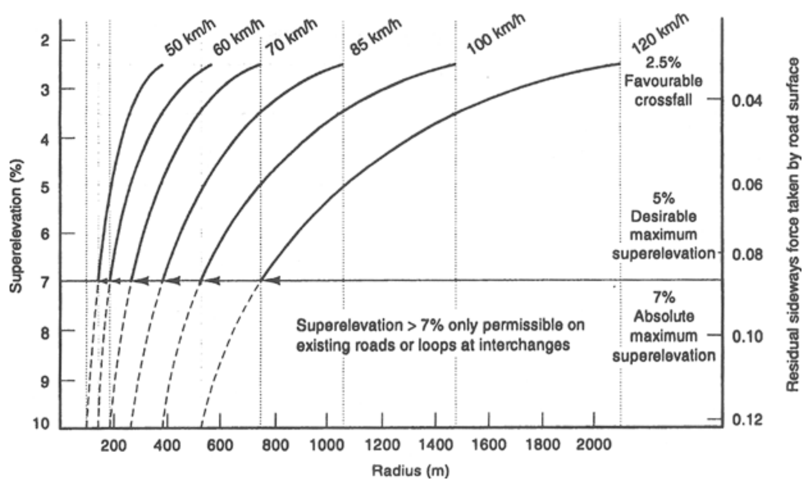


Figure 15.4 *Superelevation, curve radius and design speed*

Superelevation can be provided up to 7 per cent. Higher values are not allowed because of the requirements of slow moving traffic, or stationary traffic on icy roads. A design chart relating superelevation, curve radius and design speed is given in figure 15.4.

Transition curves

Transition curves are provided to allow a gradual introduction or reduction of centrifugal forces between straight and curved sections of road, or between two curved sections with different radii. Transition curves may not be provided at slow speeds on large radii curves, but are important for driver comfort and safety at the entry to, or exit from, high speed/sharper radii curves. The curve which provides a constant rate of change of curvature is a transition spiral or 'clothoid'. The equation of the clothoid is:

$$\frac{1}{r} = \frac{l}{RL} \tag{15.8}$$

where r = radius of curvature at any point, P, along its length,
l = length of transition curve between the origin (O) and point P,
R = radius at the end of the transition curve (such as point S, figure 15.5),
L = total length of the transition curve (OS, figure 15.5).

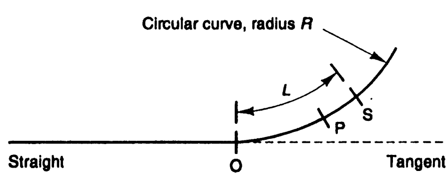


Figure 15.5 *The transition spiral*

Figure 15.5 illustrates the clothoid and the measurements in equation 15.8.

Another form of transition curve sometimes adopted is the cubic parabola. This curve readily allows points on the transition to be calculated in terms of co-ordinates. The transition is then defined by the equation

$$x = y^3 /6RL \tag{15.9}$$

The length of transition curve (L) for any design speed (V) and radius of circular arc (R) is given by the formula

$$L = V^3/46.7qR \tag{15.10}$$

where q = the rate of change of lateral acceleration (m/s^3).

Values for q of between 0.3 and 0.6 m/s^3 are normally used in design, to ensure a comfortable rate of change of lateral acceleration without introducing an excessively long transition curve.

Vertical alignment

The vertical profile of a highway, or its vertical alignment, consists of a series of straight gradients and circular or, more usually, parabolic vertical curves. Figure 15.6 illustrates the elements involved. Vertical curves have to be provided at all changes of gradient.

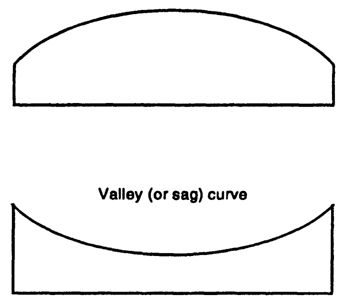


Figure 15.6 *Vertical curves*

The desirable maximum gradients for design vary according to road classification (and hence the road's design speed). Current values are as follows:

	Desirable maximum grade (%)
Motorways	3
AP dual carriageways	4
Single carriageways	6

However, in hilly terrain, steeper gradients will often be required, particularly on single carriageways where traffic volumes are relatively low. Where steeper gradients are considered, an economic evaluation is often undertaken to determine the trade-off between construction/environmental cost savings of steep gradients and the disbenefits to traffic.

For effective drainage with kerbed roads, a minimum gradient of 0.5% is recommended.

Parabolic curves are often preferred to circular curves in vertical alignment because they provide a constant rate of change of curvature, leading to a smooth alignment. The general equation for the parabola illustrated in figure 15.7, is

$$y = kx^2 \tag{15.11}$$

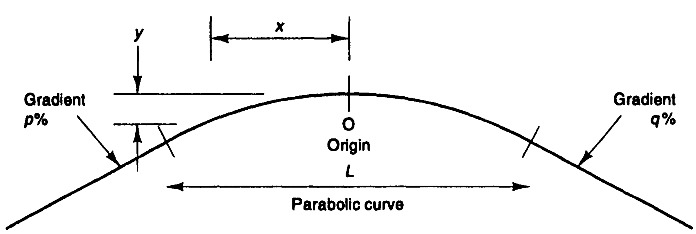


Figure 15.7 *Vertical parabolic curve*

Equation 15.11 can be developed to provide a formula to determine the levels along the curve. One such formula is

$$y = [(q - p)/2L]x^2 \tag{15.12}$$

where q and p are gradients leading into/out of the vertical curve, with rising gradients (positive) and falling gradients negative,
x, y and L are as illustrated in figure 15.7.

The required length of a vertical curve (L) is given by the formula

$$L = K(q - p) \tag{15.13}$$

where K is a value taken from table 15.4, which varies according to (i) design speed, (ii) whether the curve is a crest or a sag curve (figure 15.6) and (iii) whether (for a crest curve) FOSD is to be provided.

The term $(q - p)$ represents the algebraic change of gradient expressed as a percentage (for example, $+3$ per cent rise to -2 per cent fall would indicate a grade change of 5 per cent).

K values for crest curves on roads with design speeds of 50 km/h and above are controlled by visibility (that is the need to provide minimum SSD or FOSD). Below 50 km/h design speed, driver comfort would become the dominating constraint. For sag curves, a comfort criterion of a maximum rate of vertical acceleration of 0.3 m/s^2 is used in design to specify the K values (table 15.4).

Road layouts and cross-sections

The relationship between road classification, design flows and capacity was discussed in Chapter 11, where table 11.5 listed seven road classes. This description is continued in Department of Transport Standard TD 9/93, where recommendations for rural road layouts are given, according to their cross-sectional dimensions, edge detail requirement and restrictions on accesses and junction provision/type. The cross-sectional details of these seven road classes are summarised in table 15.5.

Dual carriageways and motorways

These roads are designed to permit light vehicles to maintain the design speed. Subject to traffic conditions, light vehicles can overtake slower moving vehicles throughout, without conflict with opposing traffic. Unlike single carriageways, therefore, there is no limitation upon the use of horizontal or vertical curves in excess of absolute minimum values, and the co-ordination of design elements mainly involves the design and optimisation of aesthetic alignments.

Tables 15.5 includes the layout requirements/considerations for dual carriageways and motorways. In addition, it is desirable to provide a smooth flowing alignment by suitable co-ordination of the different geometric elements of the alignment. The following principles help to achieve such an alignment:

TABLE 15.5
Typical rural road cross-sections

Carriageway type*	Typical cross-sections (widths) (m)					
	C/way	Hard shoulder	Verge†	Central reserve†	Total	
					C/way	Road‡
S2	7.3	–	3.5	–	7.3	14.3
WS2	10.0	–	3.5	–	10.0	17.0
D2AP	2 × 7.3	–	3.5	4.5	14.6	26.1
D3AP	2 × 10.95	–	3.5	4.5	21.9	33.4
D2M	2 × 7.3	3.3	1.5	4.0	14.6	28.2
D3M	2 × 10.95	3.3	1.5	4.0	21.9	35.5
D4M	2 × 14.6	3.3	1.5	4.0	29.2	42.8

* S2 Single carriageway, 2 lanes
WS2 Wide single carriageway, 2 lanes
D2AP Dual carriageway, 2 lanes, all purpose
D3AP Dual carriageway, 3 lanes, all purpose
D2M Dual carriageway, 2 lanes, motorway
D3M Dual carriageway, 3 lanes, motorway
D4M Dual carriageway, 4 lanes, motorway.

† Includes 1 metre hardstrip(s).

‡ Between outer verge lines.

(1) embankments and cuttings should not make severe breaks in the natural skyline;
(2) gradual curves should be used to maintain an unbroken background;
(3) short curves and straights should not be used, while adjacent curves should be similar in length;
(4) small changes of direction should be avoided, to prevent a disjointed appearance;
(5) curves which are visible from one another should not be joined by a short straight;
(6) changes in horizontal and vertical alignment should be phased to coincide wherever possible;
(7) large-radius curves should be used in preference to straights;
(8) horizontal and vertical curves should be generous at interchanges to enhance sight distances;
(9) sharp horizontal curvature should not be introduced near the top of a pronounced crest – this is especially hazardous at night as drivers cannot see the change in horizontal alignment;
(10) the view of the road ahead should not appear distorted by sharp horizontal curvature introduced near the low point of a sag curve.

Computer-aided design

The geometric design of highways is now normally undertaken by Consultants and Highway Authorities using computer-aided design (CAD) software packages. Key advantages of CAD are that it greatly speeds up the highway design process, allows many more design options to be investigated and provides improved visual representations of the design.

The most common system of CAD for the design of major highways is MOSS[2]. This is a three-dimensional surface modelling system used for a range of civil engineering applications, including highway design. In addition to MOSS, there are an increasing number of CAD packages appearing which can aid the highway design process.

References

1. Department of Transport, Highway Link Design, *Departmental Standard*, TD9/93, London (1983)
2. J.M. Houlton, The role of the computer in highway design, *Sino–British Highways and Urban Traffic Conference*, Beijing (1986)

16

Intersections with priority control

Intersections are of the greatest importance in highway design because of their effect on the movement and safety of vehicular traffic flow. The actual location of intersection is determined by siting and design and the act of intersection by regulation and control of the traffic movement. At an intersection a vehicle transfers from the route on which it is travelling to another route, crossing any other traffic streams which flow between it and its destination. To perform this manoeuvre a vehicle may diverge from, merge with, or cross the paths of other vehicles.

Priority control of traffic at junctions is one of the most widely used ways of resolving the conflict between merging and crossing vehicles. The universal adoption of the 'Give Way to traffic on the right' rule at roundabouts together with the use of 'Give Way' and 'Stop' control at junctions have considerably increased the number of occasions at which a driver has to merge or cross a major road traffic stream making use of gaps or lags in one or more conflicting streams.

Both urban and rural motorways of the future will make use of priority control at grade-separated junctions where merging vehicles have to enter a major traffic stream and whilst some authorities plan to use traffic signal control for the junctions on their proposed inner ring roads, the relatively low volumes of flow during non-peak hours and the difficulty of dealing with a large proportion of right-turning vehicles at traffic signals leads to the roundabout being preferred in many cases.

Types of priority intersections

There are three basic types of major–minor priority junctions. These are: simple junctions, ghost island junctions and junctions with single lane dualling which may be used for cross-roads, T-junctions, right–left stagger and left–right stagger junctions. In various forms these junctions may be used on single carriageway roads.

Simple major–minor junctions are used for T-junctions or staggered junctions without any ghost islands in the major road and without any channellising islands in the minor road approach. For junctions in the United Kingdom the Department of Transport[1] states that this type is appropriate for most accesses and minor junctions on single carriageway roads. For new construction they are only recommended when the design flow on the minor road is not expected to have a two-way annual average daily traffic flow exceeding approximately 300 vehicles. For existing junctions the improvement of a simple junction to a ghost island junction should be considered when the minor road two-way annual average daily traffic flow exceeds 500 vehicles or when a right-turning vehicle accident problem exists.

Ghost island junctions have a painted hatched island in the middle of single carriageway roads which provides a diverging lane and a waiting space for right-turning vehicles from the major road into the minor road. Vehicles which turn right from the minor road into the major road are also assisted because in heavy major road flow conditions the minor road vehicle can enter the major road in two stages. The additional construction cost compared with a simple junction is small and is amply justified by the increased safety of the junction.

Single lane dualling at major–minor junctions is achieved by the use of a physical island in the middle of a major single carriageway road. This provides an offside diverging lane for right-turn major road vehicles and a safe waiting area for right-turn vehicles from both the major and minor roads. An important safety feature is the provision of only one through lane in each direction so preventing overtaking and reducing speed in the vicinity of the junction.

On continuous dual carriageway roads major–minor junctions are provided by increasing the width of the central reservation. This allows a diverging lane and a waiting space to be constructed and assists right-turn drivers from the minor road by allowing them to wait within the central reservation before completing this turning movement.

Where major–minor roads intercept to form crossroads then priority control is only appropriate when the minor road flows are low. Crossroads should not be constructed in rural areas or used with dual carriageways or single lane dualling. The staggered junction is preferable with vehicles leaving the minor road turning right and after travelling along the major road turning left into the minor road. This form of staggered junction is recommended rather than the left–right staggered junction.

From the point of view of road safety it is recommended[1] that the elimination of lightly trafficked junctions and the collection of these traffic movements at a single junction be carried out wherever possible.

Accidents at priority intersections

Where these merging, diverging and crossing manoeuvres take place a potential, if not an actual, collision may take place between vehicles. The point of a potential or actual collision and the zone of influence around it has been defined as a conflict area.

The number and type of conflict areas may be taken as a measure of the accident potential of a junction, as shown in figures 16.1 and 16.2. It is shown in figure 16.1 that for a four-way priority type junction there are 16 potential crossing conflicts, 8 diverging conflicts and 8 merging conflicts. When this same intersection is converted to a right–left stagger as shown in figure 16.2, then there are only 6 crossing conflicts, 6 diverging conflicts and 6 merging conflicts. While there are objections to such a simplified approach to intersection safety, this method does give an indication of the increase in safety achieved by a staggered crossing.

Intersections may be divided into the following types: (a) priority intersections; (b) signalised intersections; (c) gyratory or rotary intersections; (d) grade-separated intersections.

In the first three types the conflicts between vehicles are resolved by separation in time while in the fourth type separation is used to resolve the main vehicle conflicts. In the selection of intersection type the following factors should be considered: traffic volumes and delay design speeds; pedestrian movements; cost and availability of land; and the accident record of the existing intersection.

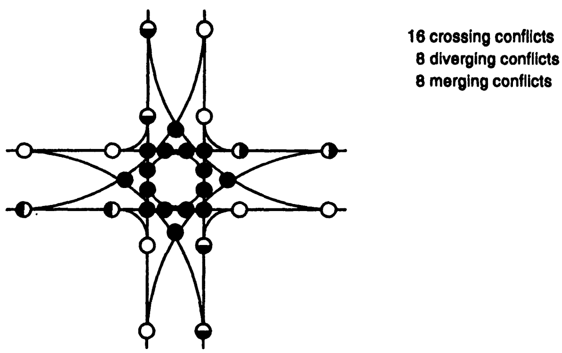


Figure 16.1 *Conflict areas at a crossroads*

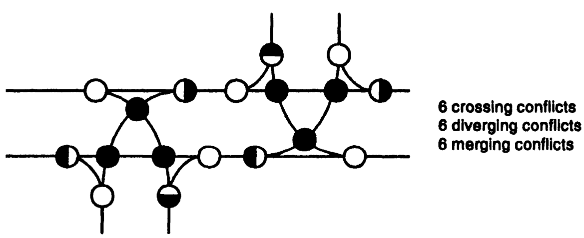


Figure 16.2 *Conflict areas at a staggered crossroads*

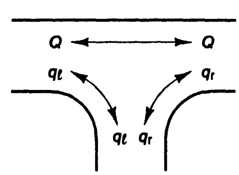


Figure 16.3 *Vehicle streams at a three-way intersection*

An early investigation into accidents at rural three-way junctions was carried out by J. C. Tanner of the Transport and Road Research Laboratory[2]. Traffic flows entering or leaving the minor road were divided into flows around the left and right-hand shoulders of the junction respectively as shown in figure 16.3 making the assumption that flows in opposite directions in the same paths were equal.

By means of regression analysis of accident data obtained from 232 rural junctions it was found that an approximate relationship between the annual accident rate around the left and right-hand shoulders of the junction, A_ℓ and A_r respectively, could be given by the equations

$$A_\ell = R_\ell \sqrt{q_\ell Q}$$

$$A_r = R_\ell \sqrt{q_r Q}$$

where q_ℓ, q_r, Q are the August 16-hour flows around the left and right-hand shoulders and on the major road in each direction respectively. For the data studied R_r and R_ℓ were found to be 4.5×10^{-4} and 7.5×10^{-4}, respectively.

A more recent investigation into junction accident rates has been used to provide data for the Department of Transport COBA program used for the assessment of highway schemes[3].

The approach used in forecasting future accident numbers at new or existing junctions is that the annual number of accidents occurring within 20 m of each junction is estimated from formulae which are of two types, both having the form

$$\text{Annual number of accidents} = x(f)^y$$

where f is a function of traffic flow, and x and y vary according to the type of junction.

The two models are referred to as the cross-product model and the inflow model. In the cross-product model the value of the function f is obtained by multiplying the combined inflow from the two major opposing links by the sum of the inflows on the other one or two minor links. These flows are expressed as thousands of vehicles per annual average day. In the inflow model the value of the function f is the total inflow from all links expressed once again as thousands of vehicles per annual average day.

Junction classification for these models is very coarse and for priority major–minor junctions the designation of junction type depends upon the number of arms of the junction and the nature of the links forming the junction. This classification is shown in table 16.1. For each type of major–minor junction the coefficients x and y vary; these coefficients are given in table 16.2.

TABLE 16.1
Major–minor junction classification for accident estimation

Junction no.	*No. of arms*	*Highest link standard (single or dual)*	*Formula type*
1	3	single	cross-product
2	3	dual	cross-product
3	4	single	inflow
4	4	dual	cross-product

A more detailed study of accidents at rural T-junctions has been carried out by Pickering *et al.*[4]. When considering total junction accidents within 20 m of a junction they found the major–minor flow product model supported previous accident research. The model was of the form

Annual number of accidents $= 0.195\ (f)^{0.49}$

where f is the average annual daily traffic expressed as thousands of vehicles.

TABLE 16.2
Coefficients for use in accident estimation models

Junction no.	*x*	*y*
1	0.195	0.460
2	0.195	0.460
3	0.361	0.440
4	0.240	0.710

This relationship which was obtained from a study of 299 T-junctions with considerable variation in junction type indicated that the relationships presented in COBA under-predicted accidents by 20 per cent. It was considered that the difference could be due to the COBA relationship having been derived from studies of new or recently modified junctions.

Considerably more detailed results are also presented by Pickering *et al* which classified accidents by the movements of the vehicle involved in the accidents and related them to geometric features of these junctions for accidents at distances of between 0 and 20 m and between 20 and 100 m from the junction.

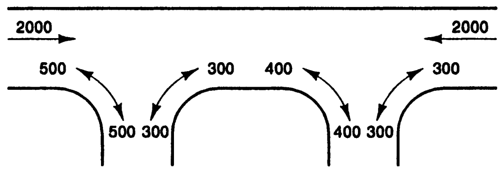
References

1. Department of Transport, Junctions and Accesses: the layout of major/minor junctions, *Departmental Advice Note* TA 20/84, London (1984)
2. M. G. Colgate and J. C. Tanner, Accidents at rural three way junctions, *Road Research Laboratory Report* LR 87, Transport and Road Research Laboratory, Crowthorne (1967)
3. Department of Transport, *The COBA program*, London (1981)
4. D. Pickering, R. D. Hall and M. Grimmer, Accidents at rural T-junctions, *Research Report* 65, Transport and Road Research Laboratory, Crowthorne (1986)

Problems

(1) At a T-junction the estimated annual average daily traffic is 3500 and 4000 vehicles on the major road and 1500 vehicles on the minor road. Calculate the expected annual number of accidents using the COBA model.

(2) A road layout has the AADT flows shown on page 164. Show that the combination of these two junctions into a single junction will reduce the expected number of annual accidents.



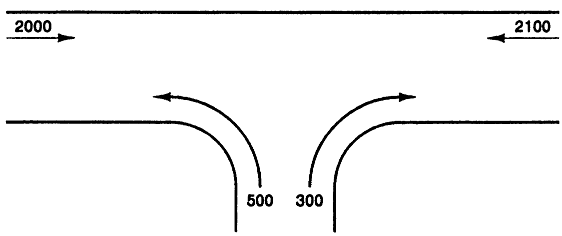
Solutions

(1) The COBA model is the cross-product model and the coefficients x and y are 0.195 and 0.460 respectively. The function f is given by (3.5 + 4) 1.5.

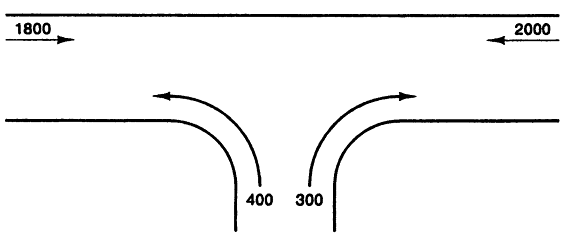
Annual number of accidents $= 0.195\ (11.25)^{0.46}$

$= 1.4$

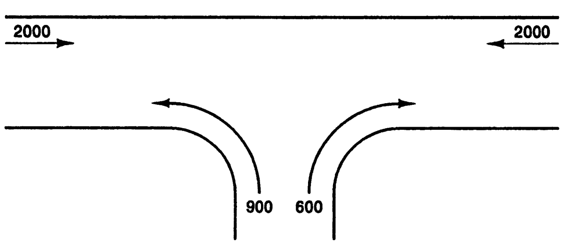
(2) The inflows at junction A are shown below.



The inflows at junction B are shown below.



The inflows at junctions A and B combined are shown below.



Annual accidents = $x\ (f)^y$; from table 16.2 x is 0.195 and y is 0.46 and f is the cross-product of the inflows expressed as AADT/1000.

Annual accidents at junction A $= 0.195\ (4.1 \times 0.8)^{0.46}$

$$= 0.34$$

Annual accidents at junction B $= 0.195\ (3.8 \times 0.7)^{0.46}$

$$= 0.31$$

Annual accidents at combined A and B $= 0.195\ (4 \times 1.5)^{0.46}$

$$= 0.44$$

It can be seen that the combination of the two junctions reduces the estimated annual accidents from 0.34 + 0.31 or 0.65 to 0.44.

17

Driver reactions at priority intersections

Interaction between traffic streams is an important aspect of highway traffic flow. It occurs when a driver changes traffic lane, merging with or crossing a traffic stream. Probably it takes place most frequently when priority control is used to resolve vehicular conflicts at highway intersections.

When a minor road driver arrives at an intersection he may either enter the major road by driving into a gap in the major road traffic stream or he may reject it as being too small and wait for a subsequent gap. It is important to differentiate when taking observations between lags and gaps. A lag is an unexpired portion of a gap which remains when a minor road driver arrives at the junction of the major and minor roads.

A minor road driver may make only one acceptance decision but he may make many rejections. If every decision of each driver is included in the observations then the resulting gap or lag acceptance distribution will be biased towards the slower driver.

To avoid bias in the observations it is possible to observe only first driver decisions. In this way most of the observed values will be lags and acceptances will be made by drivers from a rolling start. Second, third and subsequent driver decisions will be gap acceptances and will normally be higher in value because of the stationary start.

The usual hypothesis of minor road driver behaviour at priority intersections is the time hypothesis. It is assumed that a gap and lag acceptance judgement is formed as if a minor road driver made a precise estimate of the arrival time of the approaching major road vehicle. The minor road driver is thus able to estimate the time remaining before the approaching vehicle reaches the intersection and merges or crosses on the basis of the time remaining.

Observation of gap and lag acceptance

Lag acceptance behaviour can be studied by observing driver reactions at a priority intersection where the flow on the minor road exceeds 100 veh/h and where the major road traffic stream or streams exceed 400 veh/h.

Observations should be taken of one class of minor road vehicle making one type of turning movement, preferably observing left-turning cars to reduce the time required to complete the observations.

Using a stopwatch the time is noted at which a minor road vehicle arrives at the intersection and also the time at which the next conflicting major road vehicle arrives at the intersection. It is noted also whether the minor road driver accepts the lag and drives out into the major road or rejects the lag and remains in the minor road. After the first driver decision all further actions of the driver are ignored to reduce bias in the observations.

The difference between the two recorded times is the accepted or rejected lag. When observations are taken with a stopwatch it is normally only possible to record lag acceptance to the nearest 0.5 second and so it is convenient to group the observed lag acceptances and rejections into one-second classes.

A set of observations for left-turning passenger cars is shown in table 17.1. Notice that observation was continued until at least 25 observations were obtained in each class.

Column 1 contains the classes that run between zero and 100 per cent acceptance. Columns 2 and 3 give the numbers of observed lags that were rejected and accepted in each class. Column 4 is obtained by dividing column 3 by the sum of columns 2 and 3, the resulting value being expressed as a percentage.

TABLE 17.1
Observed rejected and accepted lags for left-turning vehicles

1	*2*	*3*	*4*
Lag class (s)	*Number of observed rejections*	*Number of observed acceptances*	*Percentage observed acceptance*
0.5–1.4	25	0	0
1.5–2.4	82	2	2
2.5–3.4	56	23	29
3.5–4.4	47	23	33
4.5–5.4	28	41	59
5.5–6.4	17	41	71
6.5–7.4	5	30	86
7.5–8.4	2	48	96
8.5–9.4	0	30	100

If the percentage acceptance is plotted against lag class mark as in figure 17.1 then it can be seen that the curve is approximately of cumulative normal form. This is to be expected in any action that is dependent on human reaction.

It is often necessary in theoretical traffic flow studies and in simulation work to find a mathematical distribution that approximates to an observed phenomenon.

When lag acceptance is of a cumulative normal form then the distribution may be written

Proportion of drivers accepting a lag of t second =

$$100/(\sigma\sqrt{2\pi})\int_{-\infty}^{t} \exp(-(t - \bar{t})^2/2\sigma^2 dt$$

where $\bar{t}$ and σ are the mean and standard deviation.

A demonstration of the fit of the observed lag acceptance to a normal distribution may be obtained by plotting the probit of the acceptance against the lag class mark. Values of probits may be obtained from ref. 1 or can be read from figure 17.2. The use of this technique converts a cumulative normal curve into a straight line and allows the mean and the standard deviation and also the theoretical values of acceptance to be estimated.

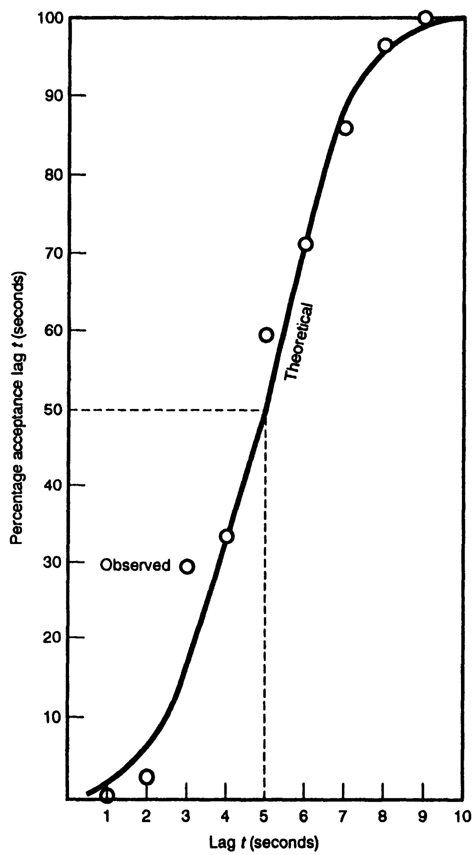


Figure 17.1 *Lag acceptance distribution at a priority intersection*

In figure 17.2 the best straight line is drawn through the observed point values. The mean lag acceptance t is the value when the probit is 5 and the standard deviation of lag acceptance is equal to the reciprocal of the slope of the line. From figure 17.2 these values are

$\bar{t} = 5.0$ seconds

$\sigma = 1.8$ seconds

The theoretical percentage acceptances can also be read from the graph and are given in table 17.2. The theoretical lag acceptance curve and the observed values are shown in figure 17.1 but a statistical test of the closeness of fit can be made by the chi-squared test. Column 8 is calculated from columns 3 of table 17.1 and column 7. The first three rows in table 17.2 have been summed to give an adequate number of acceptances in each group. The summed value of chi-squared has four degrees of freedom: the seven groups less the three constraints; the mean and standard deviation of the lag acceptances; and the total number of decisions.

From statistical tables it can be seen there is not a significant difference at the 5 per cent level between the observed and theoretical distribution.

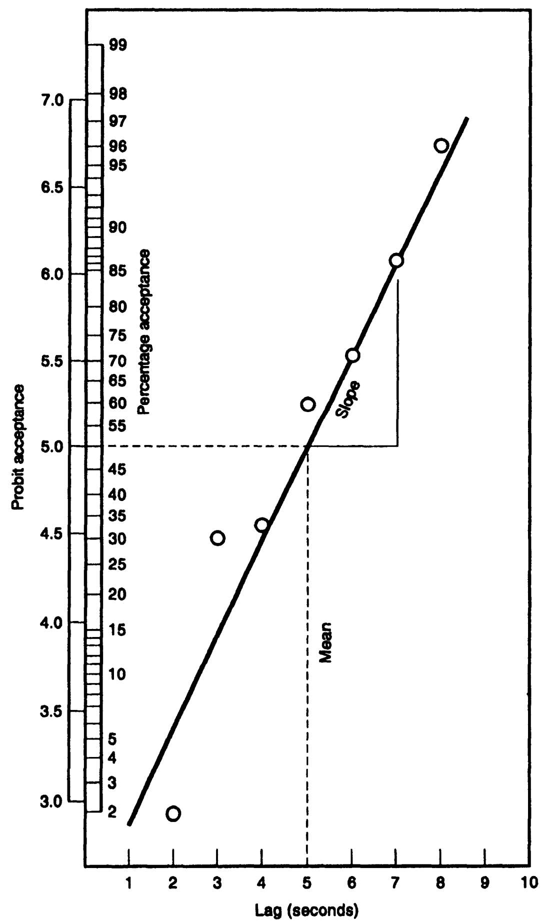


Figure 17.2 *Transformation of cumulative normal distribution to a linear relationship using probit acceptance*

TABLE 17.2
Theoretical rejected and accepted lags for left-turning vehicles

5	6	7	8
Lag class (seconds)	*Percentage theoretical acceptance*	*Number of theoretical acceptances*	*Chi-squared*
0.5–1.4	2	1	
1.5–2.4	6	5	
2.5–3.4	16	13	1.90
3.5–4.4	32	22	0.05
4.5–5.4	54	37	0.43
5.5–6.4	74	43	0.09
6.5–7.4	88	31	0.03
7.5–8.4	96	48	0
8.5–9.4	99	30	0
			Σ 2.50

Many observations of merging performance of vehicles have indicated that a cumulative log-normal distribution of lag acceptance is a better fit than the cumulative normal distribution. This can be tested using the probit transformation by the graphical fitting of a straight line to the observed lag acceptance with the log lag class-mark replacing the lag class-mark. A chi-squared test can then be used to compare the closeness of fit with the cumulative normal distribution.

A different value of lag was used by Raff and Hart[2] in investigations into driver behaviour in New Haven, Connecticut. They defined a critical lag as that lag of which the number of accepted lags shorter than it is equal to the number of rejected lags longer than it.

This critical lag is obtained graphically in figure 17.3 where the critical lag is determined graphically to be 4.6 seconds. This approach is only useful in giving a value of lag acceptance that may be used in traffic studies that use average values of driver performance. It was used by Tsongos and Weiner[3] to compare the effect of darkness on gap acceptance, a subject of considerable interest because of the proportion of peak hour flows that occurs during the hours of darkness.

It has been shown that the critical lag and the mean of the lag acceptance distribution are related in the case where the lag acceptance distribution is normally distributed. This relationship is

$$\text{critical lag} = \bar{t} - \sigma^2 q/2$$

where $\bar{t}$ and σ are the mean and standard deviation of the lag acceptance distribution and q is the flow of vehicles on the major road.

References

1. D. J. Finney, *Probit analysis, a statistical treatment of the signed response curve*, Cambridge Unversity Press (1947)
2. M. S. Raff and J. W. Hart, *A volume warrant for urban stop signs*, The Eno Foundation for Highway Traffic Control, Saugatuck, Connecticut (1950)
3. N. G. Tsongos and S. Weiner, Comparison of day and night gap acceptance probabilities, *Publ. Rds, Wash.*, **35** (1969), 7, 157–65

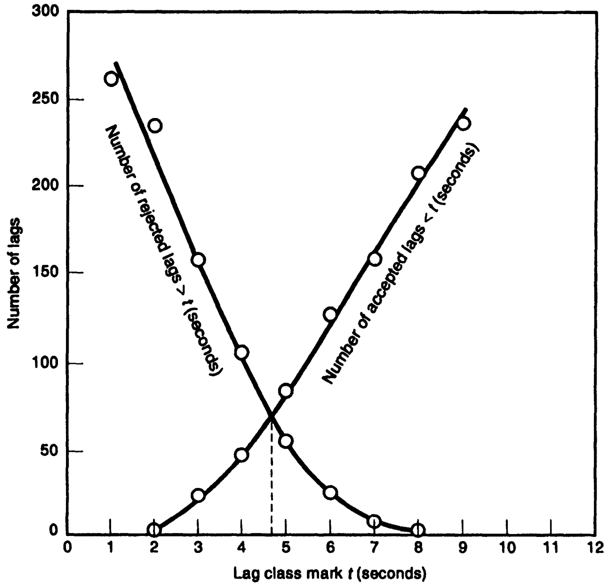


Figure 17.3 *Graphical determination of the critical lag*

Problems

Observations of driver lag acceptance at an intersection showed that it was distributed with a mean of 5.0 s and a standard deviation of 1.5 s.

(a) Is the percentage of drivers requiring a lag greater than 7.0 s, 9 per cent, 15 per cent or 31 per cent?

(b) If it is noted that the mean of the combined lag and gap acceptance distribution is 6.0 s would this indicate
(i) that drivers entering the major road without stopping are likely to enter the major road more readily than those drivers who have come to a halt before entering?
(ii) that the second decision gap acceptance of drivers has a lower mean value than first decisions?
(iii) neither of these?

(c) What is the value of the critical lag at this intersection when the major road flow is 600 veh/h?

Solutions

(a) The percentage of drivers accepting a lag greater than t seconds may be obtained from the straight-line relationship between the probit of the percentage of lag acceptance and the time lag, which may be stated as

$$\text{probit percentage acceptance} = a + b \times \text{lag } t \text{ s}$$

where b = the reciprocal of the standard deviation of lag acceptance. From the value of the mean and standard deviations given and probit tables

$$5 = a + \frac{5}{1.5}$$

and

$$a = 1.67$$

so that

$$\text{probit percentage acceptance} = 1.67 + \frac{\text{lag}}{1.5}$$

When the lag is 7.0 s

$$\text{probit percentage acceptance} = 1.67 + \frac{7.0}{1.5}$$
$$= 6.34$$

From probit tables (or from figure 17.2) percentage acceptance = 91 per cent. The percentage of drivers who require a lag greater than 7.0 s is therefore

$$100 - 91 \text{ per cent} = 9 \text{ per cent}$$

(b) As the combined lag and gap acceptance distribution has a mean of 6.0 s and lag acceptance distribution has a mean of 5.0 s this indicates that the gap acceptance distribution has a larger value of mean acceptance than the lag acceptance. Hence (1) is correct since drivers who enter the major road without stopping and hence accept lags do so more readily than those who come to a halt.

(c) From the relationship

$$\text{critical lag} = \bar{t} - \sigma^2\, q/2$$

$$\bar{t} = 5.0 \text{ s}$$

$$\sigma = 1.5 \text{ s}$$

$$q = 600/3600 \text{ veh/s}$$

$$\text{critical lag} = 5.0 - 1.5^2/6.2$$
$$= 4.8 \text{ s}$$

18

Capacities and delays at priority intersections

In contrast to traffic signal-controlled junctions, where saturation flows and resulting junction capacities can be easily calculated, the estimation of the practical capacity of priority type junctions presents considerable difficulties.

At the present time the growth of traffic on a largely unimproved urban highway system has resulted, even in the smaller urban areas, in considerable congestion and delay at intersections during peak hours. These symptoms of traffic growth were reached many years ago in the United States and it is not surprising that some of the first work on the capacity of priority type junctions was carried out in the USA. A measure of the practical capacity of an intersection is the average delay to minor road vehicles. At volumes beyond the practical capacity the average delay increases considerably with only small increases in volume.

Morton S. Raff of the Bureau of Highway Traffic, Yale University and Jack W. Hart[1] of the Eno Foundation for Highway Traffic Control made observations on the delays to vehicles at four intersections in the built-up areas of New Haven, Connecticut. They hoped to obtain an empirical equation connecting minor and major road volumes with average delays to minor road vehicles. Observations

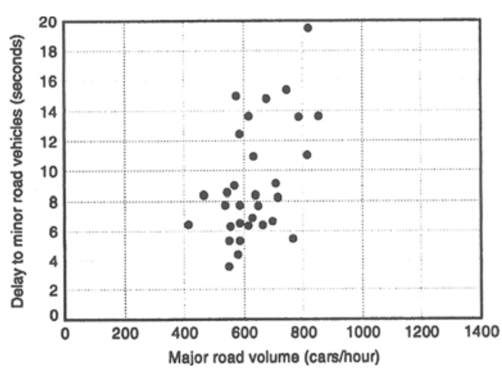


Figure 18.1 *Observed variations in delay to minor road vehicles (adapted from ref. 1)*

were made using a multi-pen recorder to note the time at which vehicles passed through the intersection and it was possible by this means to note traffic volumes and delays.

Typical of their observations is figure 18.1, which shows the average wait for all minor road vehicles. The considerable scatter of the results could not be explained by the investigators and they were, not surprisingly, unable to obtain an empirical relationship connecting main road volume and average delay.

An analytical approach to delay

It was not until 1962 that a mathematical formula or model was proposed which connected the many variables inherent in calculating the average delay to side road vehicles. J. C. Tanner[2] proposed a formula for the average delay $\overline{w}_2$ to minor road vehicles.

The variables used to calculate $\overline{w}_2$ in this formula are

The major road flow q_1.
The minor road flow q_2.
The minimum time headway between major road vehicles β_1.
The minimum time headway between minor road vehicles emerging from the minor road β_2.
The average lag or gap α in the major road traffic stream accepted by minor road drivers when entering the major road traffic stream.

These variables and their determination for a particular junction will now be explained.

No explanation of the calculation of q_1 and q_2 is required but it should be noted that the formula for delay was propounded for a T-junction where the merging or crossing movement is into or through a single unidirectional traffic stream. In the formula these flows are normally substituted as vehicles per second.

The introduction of a minimum time headway β_1 between major road traffic introduces an element of bunching or platooning into the flow caused by the inability of vehicles to overtake. Time headway is the interval of time between the fronts of successive vehicles passing a point on the highway. The usual method of estimating β_1 is to note the time headway between the front and rear vehicle of a bunch or platoon of vehicles which are following each other at their minimum separation. The average time headway between them is then calculated. The process is repeated for a considerable number of bunches until an average value representative of the traffic flow is obtained.

The minimum time headway β_2 between minor road vehicles emerging into the major road represents the time between successive minor road drivers who perceive the gap or lag available to them in the major road flow and follow each other into the major road. It is estimated by observing the average time headway between minor road vehicles following each other into the major road.

The last variable needed to estimate the average delay at a priority intersection is the average gap or lag α in the major road stream that is accepted by the minor road driver.

Observations are made of the time at which a minor road vehicle arrives at the junction with the major road and also of the time at which the next major road vehicle arrives at the junction. The gap or lag is the difference between these two times. It is also observed whether the minor road driver enters the first gap or lag that is available to him in the major road traffic stream. If he enters the major road traffic stream, then he is said to have accepted the first available gap or lag, while if he waits until a subsequent gap to enter the major road traffic he is said to have rejected the first available gap or lag. To avoid bias in favour of slower drivers, once a driver has rejected a gap his subsequent performance is neglected.

A lag is an unexpired portion of a gap that is presented to a minor road driver arriving at the junction after the commencement of a gap.

Gaps and lags are divided into classes of suitable size and the percentage of acceptance for each class calculated. When the percentage acceptance is plotted against the gap or lag class mark the curve obtained is of a cumulative normal form and it is possible by statistical means to estimate the most likely value of the 50 percentile acceptance.

With these variables known it is possible to estimate the average delay to minor road vehicles at varying major and minor road traffic volumes using the formula

$$\bar{w}_2 = \frac{\frac{1}{2}E(y^2)/Y + q_2 Y \exp(-\beta_2 q_1)[\exp(\beta_2 q_1) - \beta_2 q_1 - 1]/q_1}{1 - q_2 Y[1 - \exp(-\beta_2 q_1)]} \tag{18.1}$$

$$E(y) = \frac{\exp[q_1(\alpha - \beta_1)]}{q_1(1 - \beta_1 q_1)} - \frac{1}{q_1}$$

$$E(y^2) = \frac{2\exp[q_1(\alpha - \beta_1)]}{q_1^2(1 - \beta_1 q_1)^2}\{\exp[q_1(\alpha - \beta_1)] - \alpha q_1(1 - \beta_1 q_1) - 1$$

$$+ \beta_1 q_1 - \beta_1^2 q_1^2 + \tfrac{1}{2}\beta_1^2 q_1^2/(1 - \beta_1 q_1)\}$$

$$Y = E(y) + 1/q_1$$

In addition Tanner derived an expression for the maximum discharge from a minor road and this is given below, where the symbols have the same meaning as before.

$$q_2(\max) = \frac{q_1(1 - \beta_1 q_1)}{\exp[q_1(\alpha - \beta_1)][1 - \exp(-\beta_2 q_1)]} \tag{18.2}$$

An indication of likely values of these function flow parameters can be obtained from information that was given in a previous Department of Transport junction capacity assessment method when β_1 was taken as 2 s for minor road vehicles that merge with or cross one major traffic stream and as 1 s when they cross two major road traffic streams. β_2 is taken as 3 s and the value of α varies between 4 s and 12 s according to intersection layout and design speed.

The interactive approach to capacity

Whilst the previously discussed gap acceptance approach describes an important feature of the operation of priority intersections there are a number of difficulties in the use of these methods for estimating the capacity of a junction. Accurate measurement of the parameters that influence capacity are not easy to observe and so there are inaccuracies in the estimation of capacity. Also under heavily trafficked conditions some major road vehicles give way to the more aggressive minor road drivers and gap acceptance operation breaks down.

An investigation into the relationship between the capacity of non-priority flows and priority flows at major/minor junctions has been carried out by the Transport and Road Research Laboratory and Martin and Voorhees Associates[3]. Subsequently the results of the work were incorporated into a Department of Transport advice note on the design of major/minor junctions[4].

In this approach to design, the capacity of a junction is taken to be the traffic situation when there is a continuous queue feeding one or more turning movements at the junction. Not all the movements need be in this state for the junction to be considered to be at capacity.

Consider the traffic streams shown in figure 18.2 when a major road has arms A and C and a minor road is denoted by arm B. The traffic flows at the junction are represented by q_{c-a} etc. as shown and the streams for which it is required to calculate capacity are

q_{b-c}, q_{b-a} and q_{c-b}

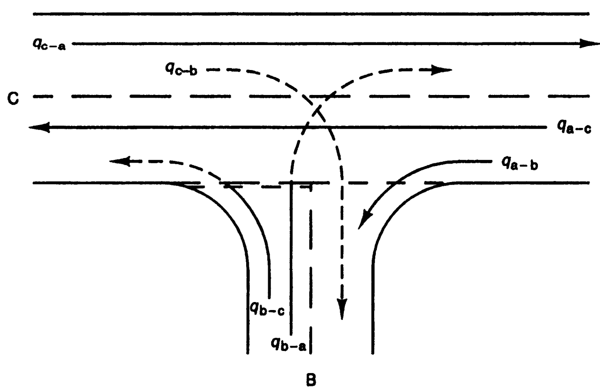


Figure 18.2

In this interactive approach the capacities of these streams depends on the flows of one or more of the priority streams and also on the geometric feature of the junction.

The best predictive equations for turning stream capacity (sat. q) when the major road has approach speeds not exceeding 85 km/h are given as:

$$q_{b-a} = D(627 + 14w_{cr} - Y[0.364q_{a-c} + 0.114q_{a-b} + 0.229q_{c-a} + 0.520q_{c-b}]) \quad (18.3)$$

$$q_{b-c} = E(745 - Y[0.364q_{a-c} + 0.144q_{a-b}]) \tag{18.4}$$

$$q_{c-b} = F(745 - 0.364Y[q_{a-c} + q_{a-b}]) \tag{18.5}$$

where $Y = (1 - 0.0345W)$ and W is the width of the major road (in metres).

In the above equations for the saturated capacities the geometric characteristics of the junction are represented by D, E and F and are stream-specific:

$$D = [1 + 0.094(w_{b-c} - 3.65)]\,[1 + 0.0009(V_{rb-a} - 120)] \times [1 + 0.0006(V_{lb-a} - 150)]$$

$$E = [1 + 0.094(w_{b-c} - 3.65)]\,[1 + 0.0009(V_{rb-c} - 120)]$$

$$F = [1 + 0.094(w_{c-b} - 3.65)]\,[1 + 0.0009(V_{rc-b} - 120)]$$

where w_{b-a} denotes the average lane width over a distance of 20 m available to waiting vehicles in the stream b–a, and V_{rb-a}, V_{lb-a} are the corresponding visibilities to the right and left measured in the normal manner for visibility splays. In these equations capacities and flows are in pcu/hour where a heavy goods vehicle is equivalent to 2 pcu. Detailed information on the measurement of geometric characteristics is given in refs 3 and 4.

Queues and delays at oversaturated junctions

In the study of junctions it is frequently important to determine queue lengths and delays. When demand is less than capacity then delay and consequently queue length can be predicted using a steady-state approach such as that theorised by Tanner[2].

Using steady-state theory, delays increase as demand approaches capacity and become infinite when demand reaches capacity. In the real-life situation when demand is close to capacity or even when the capacity is exceeded for short periods, the queue length growth lags behind that predicted by steady-state theory.

Deterministic queueing theory in which the number of vehicles delayed is the difference between capacity and demand ignores the statistical nature of vehicle arrivals and departures and seriously underestimates delay. Using deterministic theory, delay is zero when demand equals capacity.

In many traffic situations demand is close to capacity and even exceeds it for short periods of time and a combination of both steady-state and deterministic theory has been proposed by Kimber and Hollis[5]. Using a co-ordinate transformation technique they developed the following expressions for queue length and delay:

$$\text{Queue length} = \frac{1}{2}\left((A^2 + B)^{1/2} - A\right) \tag{18.6}$$

where

$$A = \frac{(1-\rho)(\mu t)^2 + (1 - L_0)\mu t - 2(1 - C)(L_0 + \rho\mu t)}{\mu t + (1 - C)}$$

$$B = \frac{4(L_0 + \rho\mu t)\,[\mu t - (1 - C)\,(L_0 + \rho\mu t)]}{\mu t + (1 - C)}$$

$$\text{Delay per unit time} = \frac{1}{2}\,((F^2 + G)^{1/2} - F) \qquad (18.7)$$

where

$$F = \frac{(1 - \rho)\,(\mu t)^2 - 2(L_0 - 1)\mu t - 4(1 - C)\,(L_0 + \rho\mu t)}{2(\mu t + 2(1 - C))}$$

$$G = \frac{2(2L_0 + \rho\mu t)\,[\mu t - (1 - C)\,(2L_0 + \rho\mu t)]}{\mu t + 2(1 - C)}$$

where $C = 1$ for random arrivals and service,
$C = 0$ for regular arrivals and service,
μ = capacity,
$\rho = q/\mu$,
q = demand,
t = time,
L_0 = length of queue at start of time interval.

Demand on both major and minor roads increases during peak periods and then decreases, resulting in a decrease and an increase in capacity.

For any non-priority stream it is possible to calculate capacity by taking into account the levels of flow in the other interacting traffic streams and the geometric characteristics of the junction. If the variation of flow with time is known then the changes in queue length and delay can be calculated.

An example of the manual computation of queue length and delay using the results of interactive theory follows. At a junction of the type shown in figure 18.2 the demand flows shown in table 18.1 (in pcu) are noted and the capacity is then calculated from equation 18.4.

TABLE 18.1

Time	*Demand* *b−c*	*Demand* *a−c*	*Capacity* *b−c*
0815	200	900	508
0820	500	1000	481
0825	300	1500	350
0830	300	1400	376
0835	200	1100	455
0840	200	900	508

For each time interval the values of q, μ and ρ are calculated and given in table 18.2 and the queue length and delay calculated from equations 18.6 and 18.7.

In the calculation of queue length and delay it was assumed that the queue was zero at 0800 hours and that vehicles arrived and departed at random. The calculated values are entered in table 18.2.

In the design of junctions the Department of Transport recommends[4] that a Design Reference Flow be selected to allow the detailed design of alternative

TABLE 18.2

Time interval commencing	q_{b-c} *(pcu/s)*	μ_{b-c} *(pcu/s)*	ρ_{b-c}	*Queue length at end of interval*	*Delay per unit time (s)*
0800	0.056	0.141	0.397	0	1.6
0815	0.139	0.134	1.037	40.5	13.0
0830	0.083	0.097	0.856	34.3	26.4
0845	0.083	0.104	0.798	13.6	11.5
0900	0.056	0.126	0.444	1.70	1.0
0915	0.056	0.141	0.397		0.7

feasible junctions to be carried out, the performance of which can then be assessed. When a peak hourly flow is selected to represent the Design Reference Flow the traffic function of the road should be considered. If a junction is designed to be adequate for the highest peak hour flow in a future design year then it is likely that the design will not be economically viable.

If the road has a recreational facility where peak summer flows are very much higher than at other times of the year, then it is suggested that the 200th highest hourly flow might be appropriate. On the other hand on an urban road where there are not likely to be very great variations in flow then the 30th highest flow is suggested as being suitable. The 50th highest flow is considered to be most appropriate for inter-urban roads.

PICADY

A computer program called PICADY (Priority Intersection Capacity and Delay) has been developed by the Transport Research Laboratory as an aid to the design of priority intersections. The latest version of the program, PICADY3[6], is a fully interactive PC-based model. It allows the modelling of three-arm and four-arm major–minor junctions, including cross-roads, left–right and right–left staggers. Other modelling facilities include flared minor road approaches, the prediction of geometric delays (see Chapter 24), pedestrian (zebra) crossings and the modelling of the blocking effects of opposed right-turning traffic on the major road.

The program models the build-up and decay of demand, queueing and delays which occur during peak periods, using discrete time intervals of typically 10–15 minutes. Traffic demand is constant within each interval, but usually varies between intervals. The procedure in each interval involves:

(i) calculating the capacity for each non-priority stream, using equations 18.3 to 18.5 above;
(ii) using these capacities and relevant demand flows to derive queues and delays, using equations 18.6 and 18.7 above.

The basic inputs to the model include traffic demand, composition and turning proportion data, geometric data required to calculate stream capacities and information relating to pedestrian crossings, where relevant. The program outputs predict capacities, queues and delays per link and per time interval as well as aggregate statistics. PICADY3 also provides accident predictive modelling facilities (for example, see Chapter 16) so that the safety implications of different layouts can be assessed.

References

1. M. S. Raff and J. W. Hart, *A volume warrant for urban stop signs*, The Eno Foundation for Highway Traffic Control, Saugatuck, Connecticut (1950)
2. J. C. Tanner, A theoretical analysis of delays at an uncontrolled intersection, *Biometrika*, **49** (1962), 163–70
3. R. M. Kimber and R. D. Coombe, The traffic capacity of major/minor priority junctions, *Transport and Road Research Laboratory Supplementary Report* 582, Crowthorne (1980)
4. Department of Transport, Junctions and accesses: determination of size of roundabouts and major/minor junctions, *Departmental Advice Note* TA 23/81 (1981)
5. R. M. Kimber and Erica M. Hollis, Traffic queues and delays at road junctions, *Transport and Road Research Laboratory Report* 909, Crowthorne (1979)
6. Transport Research Laboratory, PICADY3: Priority Intersections Capacity and Delay, *Transport Research Laboratory, User Guide PICADY3* (1993)

Problems

(1) State the traffic flow characteristics defined by Tanner in the estimation of delay to non-priority vehicles at a priority junction.
(2) Describe the difference between lag and gap acceptance when observations are made at a priority junction.
(3) Delays at priority junctions can be estimated by probabilistic or deterministic approaches. Discuss the differences in these methods.

Solutions

(1) The five traffic-flow characteristics used to calculate the theoretical delay to road vehicles at a priority intersection are
 q_1 the major road flow with which minor road vehicles conflict,
 q_2 the minor road demand flow,
 β_1 the mean minimum headway between bunched vehicles on the major road,
 β_2 the mean minimum headway between bunched vehicles on the minor road entering the major road when the major road flow is zero,
 α the mean gap or lag accepted by minor road drivers.
(2) A lag is a portion of a gap; when a minor road vehicle arrives at the stop or giveway line, then normally the minor road vehicle will accept or reject a portion between successive major road vehicles. If the driver rejects the offered lag it is a gap between major road vehicles which the minor road driver must either accept or reject. First decisions of drivers are normally lags whilst subsequent decisions are gaps.

(3) A probabilistic approach to determining delay at priority junctions takes account of the random or semi-random nature of vehicle arrivals at the junction. A well-known approach using probabilistic theory is that due to Tanner which predicts delays rising rapidly as the minor road demand flow approaches capacity and infinite delays at capacity. In contrast, deterministic approaches predict zero delay and queue length when the minor road demand flow equals capacity with queue length and delay increasing with the passage of time. In practice a compromise between the two approaches is used, the relationship between queue length and delay and demand flow being derived by an axis transformation technique.

19

A simulation approach to delays at priority intersections

Prediction of capacities and delays at highway intersections under priority control may be made by the use of empirical, mathematical or simulation models.

The dependent variable in these models is usually the average delay to minor road vehicles and measures the effectiveness of the junction in allowing conflicting traffic streams to merge or intersect. The performance of the junction may be measured and alternative geometric designs evaluated by consideration of average delays to minor road vehicles.

Empirical models attempt to predict highway capacity on the basis of past observations and probably the best known example of this approach is the Highway Capacity Manual[1] calibrated by traffic data collected throughout the United States. Other researchers[2-5] have studied traffic flow at priority type intersections and used regression analysis methods to obtain relationships between average delay and traffic volumes.

At the present time the average delay to minor road vehicles and the maximum discharge from the minor road at a priority intersection are calculated using mathematical models given by Tanner[6].

Both mathematical and empirical models have limitations in their use. Traffic flow at an intersection is such a complex phenomenon that even the most complex mathematical model has to adopt a macroscopic approach in that all vehicles and drivers generally have the same characteristics.

An empirical model would require a considerable amount of data for its calibration and it is often very difficult to obtain this data with the precision, in the quantity or at the required traffic volume levels. Often the traffic volumes required may, when found, not persist long enough for sufficient observations to be obtained.

If the effect of geometric features is being investigated it will also be necessary for a junction having these features to be located or else constructed in the field or on a test track before traffic observations can be obtained.

While the development and verification of a simulation model is a complex process it is not subject to many of the disadvantages of empirical and mathematical models. Simulation models of intersection traffic flow can be flexible enough to cover a wide range of highway and traffic conditions. Inputs to the model can be specified to any distribution and the form of traffic control and driver characteristics varied. Simulation time may be as long as desired and several figures of merit may be printed out during the simulation process.

A simulation model requires the formation of a model system which represents the real situation at the site being studied. Two forms of simulation have been used in the study of highway traffic flow, analog simulation and digital simulation. In analog simulation an analogy is made between the real world situation and an analog physical system, the components of the analog physical system interacting in the same manner as in the real situation.

In digital simulation the state of the simulated system is stored in digital form and the situation is updated in accordance with stored instructions or rules of the model. Updating of the system may be carried out in two ways, by regular time scanning or by event scanning.

When regular time scanning is used the state of the traffic system is examined at equal scan intervals and all required vehicle movements calculated for the next interval in time. Updating of the traffic situation is carried out for all components of the model. The process is then repeated for the next time scan period.

The length of the time scan period selected will affect the precision of the model. It must be small enough to include all vehicle actions of significance but if the scan interval selected is too small the computer run time will be increased. The time scan interval selected will depend on the computer resources available, the objects of the simulation process and the method of generation of vehicular headways.

When event scanning is used the events that have the greatest importance in the operation of the program are used to build the program. The time at which the next significant event occurs is determined by the program and the traffic situation updated to that event.

To evolve a simulation program, it is usually most convenient to draw up a flowchart showing the steps in the simulation process. Figure 19.1 shows the flowchart for the simulation of the traffic system where a single stream of left-turning minor road vehicles intersect or merge with a single stream of major road vehicles. In this program regular time scanning is used and the figure of merit employed is the average delay to minor road drivers.

An appreciation of digital computer simulation can be obtained by a manual simulation. While this is laborious to carry out for any but a very limited real time period it illustrates the use that can be made of mathematical representations of headway and gap acceptance distributions in simulation.

As an illustration, the traffic flow at the intersection of a single stream of left-turning minor road vehicles and a stream of major road vehicles may be simulated. The major road flow–headway distribution may be assumed to be represented by the double exponential distribution with the following parameters

$$\bar{t}_1 = 2.5 \text{ s}$$

$$\bar{t}_2 = 5.5 \text{ s}$$

$$L = 0.5$$

where $\bar{t}_1$, $\bar{t}_2$ and L are the mean time headway of free-flowing vehicles, the mean time headway of restrained vehicles and the proportion of restrained vehicles in the traffic flow respectively.

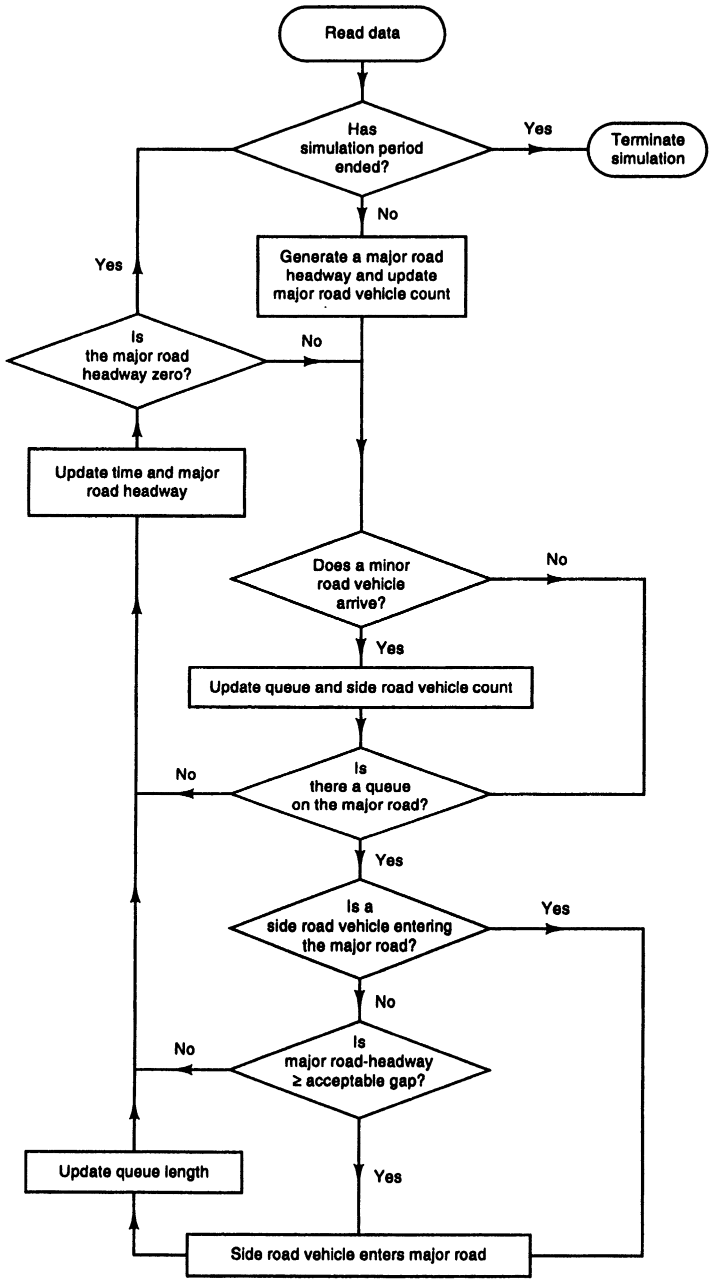


Figure 19.1 *A flowchart for the simulation of a single minor road stream with a single major road stream*

The minor road flow headway-distribution may be assumed to be represented by the negative exponential distribution, but no vehicle may arrive with a time headway of less than one second between it and the preceding vehicle. The single parameter in this distribution is

$$\bar{t} = 9 \text{ s}$$

where $\bar{t}$ is the mean headway. In addition minor road vehicles exit from the minor road at a minimum headway of 1.0 s.

Gap and lag acceptance for minor road drivers is assumed to be similar and to be represented by the normal distribution with the parameters

$$\bar{t} = 2.5 \text{ s}$$

$$\sigma = 0.9 \text{ s}$$

where $\bar{t}$ and σ are the mean gap and lag acceptance and the standard distribution of gap and lag acceptance respectively.

Furthermore it is noted that no drivers will accept or reject a lag or gap less than the mean minus one standard deviation or greater than the mean plus one standard deviation respectively.

Generation of major and minor stream headways will be carried out by Monte Carlo processes making use of a series of random numbers. Such a series is shown in table 19.1 and could be easily prepared from a series of ten identical cards, each

TABLE 19.1
A table of random numbers

81	77	72	64	42	31	29	46	62	21
45	93	60	17	35	78	25	42	41	16
11	64	30	58	60	21	33	75	79	74
15	63	47	59	51	13	59	85	27	62
50	53	22	54	96	95	65	24	25	73
27	89	64	72	81	74	40	09	19	61
82	52	75	59	55	79	17	14	24	33
88	66	54	64	70	52	85	50	13	63
23	80	45	68	42	93	67	03	97	42
03	02	59	34	49	77	70	40	75	22
34	95	49	12	91	51	95	80	07	36
51	85	12	18	18	75	06	16	32	84
60	97	67	18	93	92	23	09	41	64
32	78	34	17	83	90	90	51	62	55
58	71	19	68	23	46	76	30	18	79
45	11	28	93	45	58	10	79	31	21
47	56	38	09	90	43	25	27	74	03
52	64	10	66	39	49	93	74	48	15

carrying a number from 0 to 9, by picking cards at random and noting the numbers obtained.

To generate major road headways, successive two digits are selected from table 19.1 and regarded as the first two decimal places of a random fraction. The random fraction may then be used to solve the following equation for t

$$\begin{aligned}(\text{random fraction}) &= L \exp\left[-(t - e)/(\bar{t}_1 - e)\right] + (1 - L) \exp(-t/\bar{t}) \\ & \quad t \geqslant e \\ &= L + (1 - L) \exp(-t/\bar{t}_2) \qquad (19.1) \\ & \quad 0 < t < e\end{aligned}$$

While the rapid solution of this equation to obtain t, the major road headway, would not present any difficulty when using a digital computer a graphical solution will be used in this example.

TABLE 19.2
Generated major and minor road headways

Random fraction	*Major road headway (seconds)*	*Cumulative major road headway (seconds)*	*Minor road headway (seconds)*	*Cumulative minor road headway (seconds)*
81	1.2	1.2	2.6	2.6
77	1.3	2.5	3.1	5.7
72	1.5	4.0	3.6	9.3
64	1.8	5.8	4.5	13.8
42	3.0	8.8	7.9	21.7
31	3.8	12.6	10.3	32.0
29	4.0	16.6	10.8	42.8
46	2.7	19.3	7.1	49.9
62	1.9	21.2	4.8	54.7
21	5.3	26.5	13.4	68.1
45	2.8	29.3	7.3	75.9
93	0.7	30.0	1.5	76.9
60	2.0	32.0	5.0	81.9
17	6.2	38.2	15.0	96.1
35	3.5	41.7	9.3	105.4
78	1.3	43.0	3.0	108.4
25	4.6	47.6	12.0	120.4
42	3.0	50.6	7.9	128.3
41	3.1	53.7	8.0	136.3
16	6.6	60.3	15.6	151.9
11	8.5	68.8	18.6	170.5
64	1.8	70.6	4.5	175.0
30	4.0	74.6	10.5	185.5
58	2.1	76.7	5.3	190.8
60	2.0	78.7	5.0	195.8
21	5.3	84.0	13.4	209.2
33	3.7	87.7	9.8	219.0
75	1.4	89.1	3.3	222.3
79	1.2	90.3	2.8	225.1

To obtain the major road headways by graphical means equation 19.1 is plotted as shown in figure 19.2. A series of two random numbers is then taken successively from table 19.1 and inserted in figure 19.2 to obtain the headways which are shown in table 19.2.

Successive headways on the minor road are next simulated by substituting successive random fractions in the cumulative negative exponential distribution (random fraction) $= \exp[-(t-1)/(\bar{t}-1)]$

As with the major road headways, the successive solution of this equation will be made by a graphical method. The cumulative minor road headway distribution is

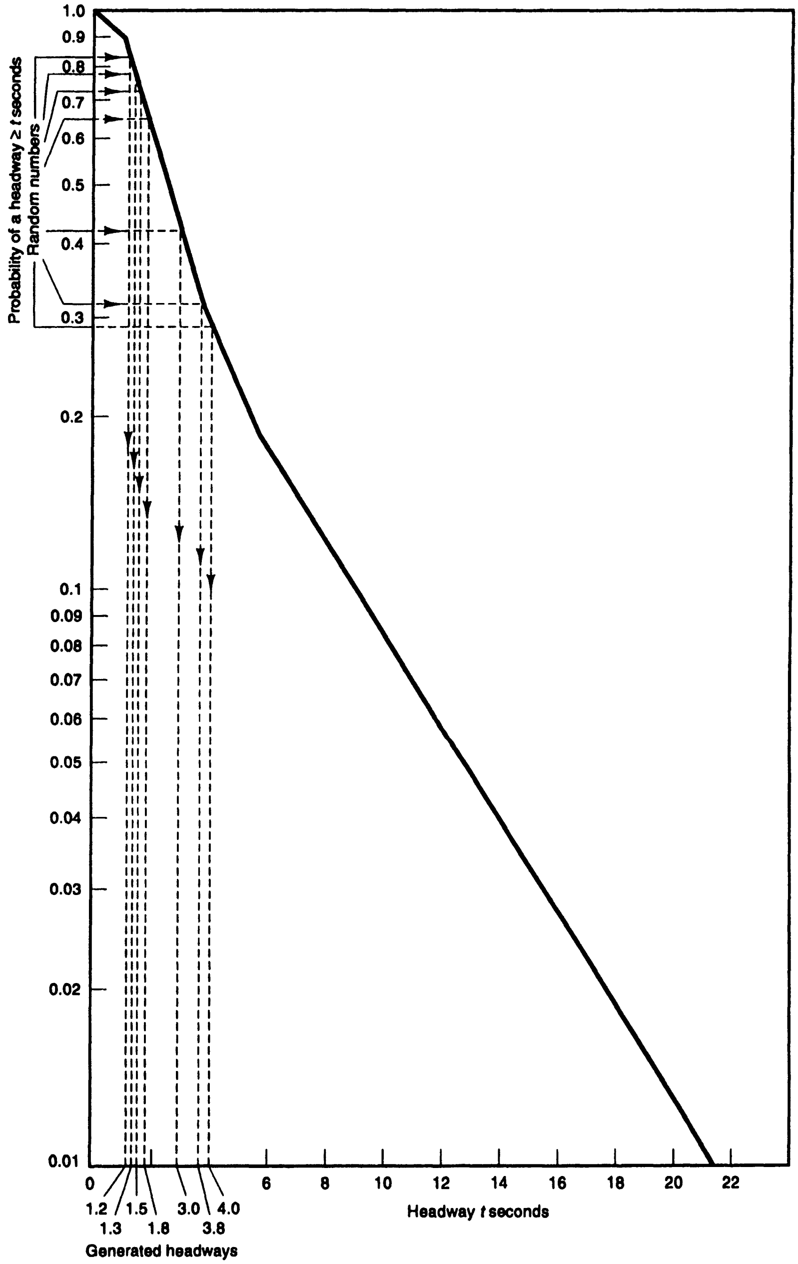


Figure 19.2 *The generation of major road headways*

shown in figure 19.3, and successive series of two random numbers are used as random fractions to generate headways as previously. The minor road headways generated in this way are given in table 19.2.

The third set of variables to be simulated are the lag and gap acceptances of minor road drivers. The same technique as has been used previously for generat-

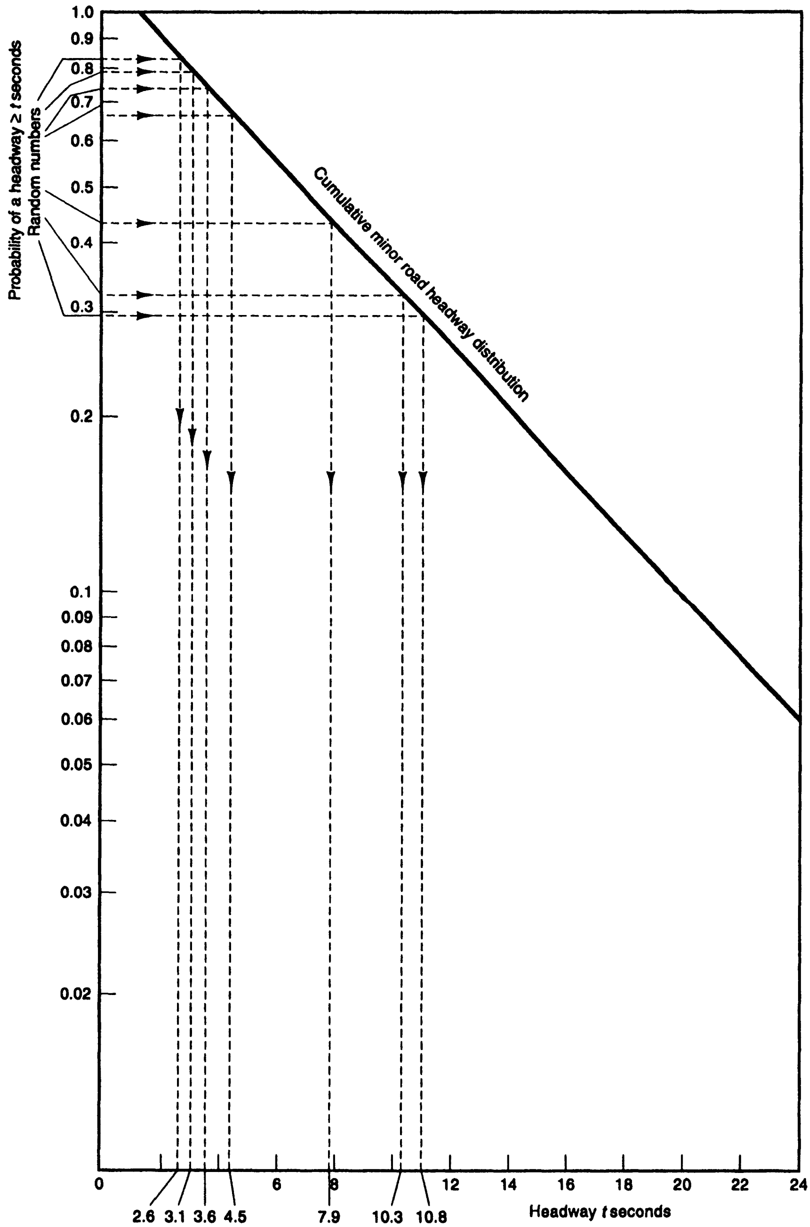


Figure 19.3 *The generation of minor road headways*

ing headways will be used and successive random fractions are entered into the cumulative acceptance distribution as shown in figure 19.4. The acceptable lags obtained in this manner and the successive random fractions used to generate them are given in table 19.3. It should be noted that the generated acceptable lags cannot exceed the mean lag plus two standard deviations; nor be less than the mean lag minus two standard deviations. Maximum and minimum values are thus 3.4 s and 1.6 s, respectively, and it is thus not necessary to amend any of the values generated.

TABLE 19.3
Random fractions and generated lags and gaps

Random fraction	*Acceptable gap or lag (seconds)*
0.81	3.3
0.77	3.2
0.72	3.0
0.64	2.8
0.42	2.3
0.31	2.0
0.29	1.9
0.46	2.4
0.62	2.8
0.21	1.7
0.45	2.4
0.93	3.8
0.60	2.7

All the variables required for the simulation of traffic flow have now been generated and it is possible to set up a table of events in which successive features of importance are arranged in order of their occurrence. This simulation table is shown as table 19.4 and initially the columns (1) and (2) are introduced from tables 19.2 and 19.3. Column (3) has the same value as column (2) if a queue does not exist. The value in column (3) is the lowest value of the time of arrival or the exit time of the previous vehicle in the queue plus one second.

Column (4), the required lag, is obtained from the previously generated lags in table 19.3 and the values are entered when a vehicle arrives at the stop or give way line.

A time value is entered in column (5) as soon as the gap or lag in the major road stream is equal to or greater than the required lag or gap. It is assumed that a vehicle leaves the minor road instantly but a further vehicle cannot leave the minor road for a further one second.

Column (6) records the queue length while column (7) is the difference between time of arrival of a minor road vehicle and the time of exit. This column is summed and when divided by the number of minor road vehicles entering the intersection gives the average delay to minor road vehicles.

For the traffic situation simulated the value of average delay is 2.4 s: the period of simulation is however extremely short and the simulation should be extended to a minimum period of 600 s. At high traffic flows the effect of assuming that a queue does not exist at the commencement of the simulation period tends to underestimate delays unless the simulation period is sufficiently long.

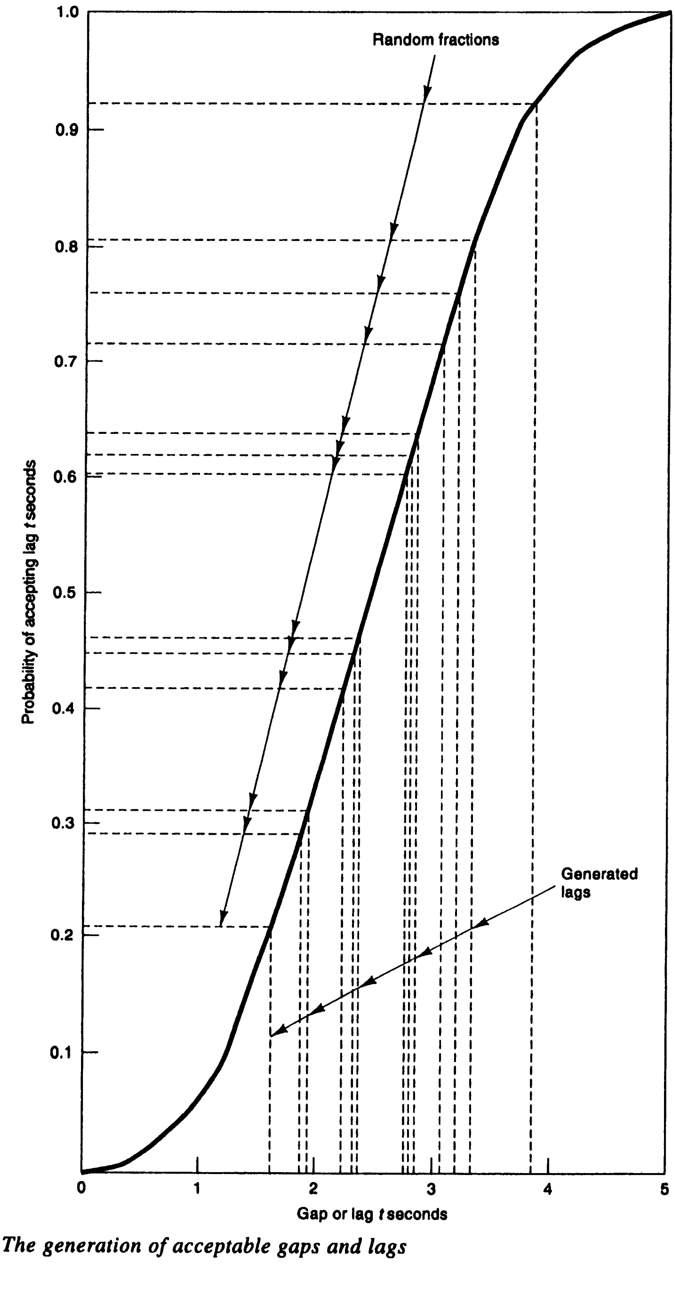


Figure 19.4 *The generation of acceptable gaps and lags*

TABLE 19.4
The simulation table

1	2	3	4	5	6	7
Major road vehicle arrives at junction	*Minor road vehicle arrives at end of queue*	*Minor road vehicle arrives at stop or give way line*	*Required gap or lag (seconds)*	*Minor road vehicle enters major road*	*Queue length*	*Delay of exiting vehicle (seconds)*
1.2						
2.5						
	2.6	2.6	3.3	no	1	
4.0				no		
	5.7			no	2	
5.8				no	2	
8.8				8.8	1	6.2
		9.8	3.2	no	1	
	9.3			no	2	
12.6				12.6	1	6.9
		13.6	3.0	13.6	0	4.3
	13.8	14.6	2.8	no	1	
16.6				no	1	
19.3				no	1	
21.2				21.2	0	7.4
	21.7	22.2	2.3	22.2	0	0.5
26.5						
29.3						
30.0						
32.0						
	32.0	32.0	2.0	32.0	0	0
38.2						
41.7	42.8	42.8	1.9	no	1	
43.0				43.0	0	0.2
47.6						
	49.9	49.9	2.4	no	1	
50.6				50.6	0	0.7
53.7						
	54.7	54.7	2.8	54.7	0	0
60.3						
	68.1	68.1	1.7	no	1	
68.8				68.8	0	0.7
70.6						
74.6						
	75.4	75.4	2.4	no	1	
	76.9			no	2	
76.7				no	2	
78.7				78.7	1	3.3
		79.7	3.8	79.7	0	2.8
	81.9	81.9	2.7	81.9	0	0
84.0						
				cumulative delay		33.0

average delay = cumulative delay/number of minor road vehicles leaving intersection
= 33.0/13 s
= 2.5 s

References

1. Highway Capacity Manual, *Highway Research Board Special Report* 87 (1965)
2. W. N. Volk, Effect of type of control on intersection delay, *Proc. Highw. Res. Bd* (1956), 523–33
3. DeLeuw, Cather and Co., The effect of regulatory devices on intersection capacity and operations, *Final Report Project 3-6*, National Co-operative Highway Research Project, August (1966)
4. J. G. Wardrop and J. T. Duff, Factors affecting road capacity, *International Study Week in Traffic Engineering*, Stresa (1956)
5. F. V. Webster and B. M. Cobbe, Traffic signals, *Tech. Pap. Rd. Res. Bd* 56, HMSO, London (1966)
6. J. C. Tanner, A theoretical analysis of delays at an uncontrolled intersection, *Biometrika*, **49** (1962), 163–70