TAG UNIT M3.2
Public Transport Assignment

May 2020

Department for Transport

Transport Analysis Guidance (TAG)

https://www.gov.uk/transport-analysis-guidance-tag

This TAG unit is guidance for the **MODELLING PRACTITIONER**

This TAG unit is part of the family **M3 – ASSIGNMENT MODELLING**

Technical queries and comments on this TAG unit should be referred to:

Transport Appraisal and Strategic Modelling (TASM) Division
Department for Transport
Zone 2/25 Great Minster House
33 Horseferry Road
London
SW1P 4DR
tasm@dft.gov.uk

Contents

1 Public Transport Assignment Modelling

1.1 Introduction

1.1.1 This TAG unit provides detailed advice on assignment methods for public transport models and covers the following topics:

- the development of public transport passenger assignment models and their relation to highway assignment models in sections 1.3 to 1.6

- assignment methods in section 2

- generalised cost in section 3

- path building in section 4

- cost skimming in section 5

- convergence and validation of public transport passenger assignment models, in section 6

1.1.2 This unit is relevant to all public transport sub-modes. However, its focus is on public transport assignment in a multi-modal context and most usually for bus or light rail schemes. The same principles apply to modelling of heavy rail schemes. However, where inter-urban heavy rail is the primary focus of the scheme then a uni-modal modelling approach, using tools such as MOIRA, will often be the preferred approach. In these cases, contact the Department's Rail Analysis team.

1.2 Relationship of this Unit to Other Advice

1.2.1 This unit **excludes** advice on the following topics which are related to public transport assignment modelling:

- travel demand data collection, including public transport intercept surveys, in-vehicle passenger counts and boarding / alighting counts

- the development of base or prior trip matrices for public transport assignment models

- preparing forecasts using public transport assignment models

1.2.2 This unit is a companion to the following TAG units:

- M1.2 – Data Sources and Surveys

- M2.1 – Variable Demand Modelling

- M2.2 – Base Year Demand Matrix Development

- M3.1 – Highway Assignment Modelling.

1.3 Public Transport Passenger Assignment Models

1.3.1 The key role of public transport assignment models in demand forecasting is the provision of levels of service, the travel times, distances and costs associated with trips between origin-destination pairs, distinguishing components such as transfer and wait times, and, where relevant, different transport modes. These are used in scheme appraisal and demand modelling.

1.3.2 In simple terms, the key issues in assessing public transport assignment models are:

- assignment method

- generalised cost definition

- path building

- cost skimming

These issues are discussed in sections 2 to 5.

1.3.3   A key issue which will influence the form of the public transport assignment model is whether trips are allocated between public transport modes at the mode choice stage or at the assignment stage. The key factors which govern the choice of method are (a) the proportion of passengers who use more than one public transport mode; (b) the numbers of different mode combinations used to a significant extent; and (c) the level of competition between different public transport modes. Although allocation as part of the assignment process will generally be less reliable than an explicit mode choice model where competition between modes exists, mixed modes are generally easier to handle at the assignment stage, especially if there are a number of mode combinations to consider.

1.3.4   If the allocation of trips between public transport sub-modes is to be made at the mode choice stage, a separate network model is required for each public transport sub-mode included in the mode choice process and at least an implicit network for each combination of modes. If the allocation of trips between some public transport sub-modes is to be made at the assignment stage, a combined network is required which contains all the available public transport sub-modes and linkages between them.

1.3.5   If separate network models are used, the sophistication of the assignment stage may be considerably simplified compared to the combined network approach. The assignment process is likely to become a simple matter of loading a matrix of single mode trips onto single paths through the corresponding mode network. Within the bus network, unless the allocation of passengers between services is of interest, it is usually sufficient to represent the whole network with combined frequencies. Some cases may be so simple (for example, a service between an airport and a city centre) that assignment software is not required.

1.3.6   Advice given in TAG unit M5.1 will be pertinent where it is expected that the main mixed mode travel will be via car access to PT. Where this is the case, it may be more appropriate to model the P&R trip as part of a highway assignment.

1.3.7   If a combined network model is used (in the case where there are sufficient mixed mode trips between the various PT sub-modes available), the assignment process may be much more complicated, especially if the model is to produce accurate allocations of passengers to the various modes. In a combined network, there is likely to be a significant amount of direct competition between modes and the analyst needs to be sure that the assignment model distributes trips between the modes in a realistic manner (see section 6 on model validation).

1.3.8   It is important to avoid double-counting trips where mixed-mode trips will have been observed from surveys for more than one mode, for example, bus and rail. It is also important to build trips from the traveller's true ultimate origin to their true ultimate destination.

1.3.9   Mixed-mode trips should be loaded onto the appropriate networks for all legs of their journey. For example a trip that uses bus to access rail should appear in the bus flows between its origin and the railway station and in the rail flows from the railway station to its destination. This may be achieved either with a single network model with certain modes and routes banned to certain segments, or with separate network models with zones at mode interchange points.

1.4   Relation to Highway Assignment Models

1.4.1   Public transport schemes may **reduce** highway congestion by attracting travellers from car or **increase** localised road traffic congestion through measures to secure priority for public transport over general traffic. The impacts of public transport schemes on road traffic congestion should

therefore be estimated, unless schemes are aimed at serving low car ownership areas (and where modal transfers are therefore likely to be small) and which would not impact adversely on the capacity of the road system for general traffic.

1.4.2   Models for the appraisal of major public transport schemes which are expected to have an impact on road traffic congestion will require a validated highway (road traffic) assignment model. See TAG unit M3.1 for details on how to do this. Indeed, there are some cases where highway-based public transport schemes may be appraised satisfactorily using a highway assignment model alone, such as a bus priority scheme.

1.4.3   For sub-modes that run on-street (mainly bus, but some LRT schemes) it is important that journey times in the public transport (PT) assignment model are consistent with the level of traffic congestion. This will require some linkage of on-street PT mode link times in the PT network to assigned journey times from a highway assignment model. In congested urban situations it may not be appropriate to take bus times from a published timetable. In any case this will not be possible when forecasting and a link to a highway assignment model will be necessary to estimate PT on-street journey times for forecast years. The presence of bus or no-car lanes needs to be taken into account.

1.4.4   In all cases such as these, the highway assignment model should be sufficiently detailed to model both the road capacity changes required by the public transport scheme and the effects of those changes on road traffic congestion.

1.5   Network Representation

1.5.1   The structure of the network of existing heavy and light rail services can be derived by scaling the track layout from GIS (or large scale OS maps). Details of the services, including frequency, in-vehicle time, stops served, and fares, can be obtained from the operators, journey planners and data sets such as the Government's National Public Transport Data Repository or ATCO-CIF timetable files. For proposed light rail services, the network structure can be obtained by scaling from engineering layout plans, and details of the services, including their frequency, stops served and fares, can be postulated by the system design team. Of considerable importance, though, will be the operating speed. This should be available from operational modelling undertaken by the system design team. For street-running sections of light rail systems, it is important to reflect the interactions between road traffic and the light rail vehicles in the coding.

1.5.2   The structure of the network covered by existing bus services should be available from the Local Authority or Passenger Transport Executive. The geometry of the network should be derived from the road network coded for the highway assignment model. However, the number of nodes may need to be increased in order for the bus stops to be represented adequately. This will create a disparity between the highway and public transport networks which may cause difficulties where the effects of changes in road traffic congestion are to be transferred from the road network. Consideration therefore should be given to the introduction of "dummy" nodes to represent bus stops in the road traffic model network to ensure compatibility between highway and PT network representation, bearing in mind that this may increase the model run time significantly.

1.5.3   Of particular importance will be the accuracy with which in-vehicle times are represented. Schedules may not be adhered to, of course, and for bus services in direct competition with a proposed new public transport service, in-vehicle times should be obtained by direct observation, with an allowance for dwell time at stops.

1.5.4   Depending on the complexity of the public transport network to be modelled, there needs to be the facility to interchange between modes and services, and therefore the representation of bus stop to station connections and interchange walk links will be necessary where a combined approach to assignment is used.

1.5.5   Once networks and services have been defined, zone centroids should be connected to stops or stations in a way that realistically represents how people access the available public transport

services. In adding zone connectors it is important to avoid any bias towards a particular sub-mode or service. It is recommended that zones adjacent to rail services reflect the expectation that people would walk to the nearest stop or station whereas for zones further away some other mode of access, principally car, would be used. This of course is dependent on zone size and realistic expectations of walking, cycling, or using other modes of access. The times and costs allocated to the centroid connectors should, in principle, reflect the average times and costs for the people in the zone who travel by different access modes to get onto public transport.

1.5.6   The use of a car to access a public transport service may take two general forms:

- kiss-and-ride, where the public transport user is driven to the station and picked up again on the return journey

- park-and-ride, which may involve either the use of a designated park-and-ride site, parking at stations, or informal parking in streets surrounding stations

1.5.7   In a base year model, it will generally be acceptable to represent these complexities through the allocation of average access times and costs to the centroid connectors. It will be important, therefore, to ensure that the survey data contains sufficient information for these averages to be calculated. Advice on the modelling of formal park-and-ride schemes as part of the forecasting for the scheme being appraised can be found in TAG unit M5.1.

1.6   Design of the Zoning System

Design Principles

1.6.1   The design principles for the zoning system and modelled area for assignment models in general is discussed in TAG unit M3.1. This section discusses specific considerations relevant to public transport assignment models.

1.6.2   It is important to avoid the temptation to allow the zoning system for the public transport model to be distorted from the ideal by partially adapting a zoning system designed for some other model. The zoning systems appropriate for highway assignment modelling and public transport modelling may be quite different and the former should not unduly influence the latter in the case of models for the appraisal of major public schemes.

1.6.3   It is possible, within a single public transport model, to employ different zoning systems for the public transport assignment model, the highway assignment model, and the demand model. The primary constraint is that all the more detailed zoning systems should be subsets of the more aggregate zoning systems. A simple approach is to have the same zoning system between models, although the needs of the public transport assignment model should not be neglected and should be key in guiding the design of the zoning system if this approach is adopted. The fundamental requirement for zoning systems is that they are all compatible with each other.

Detailed Considerations

1.6.4   In general terms, there are two kinds of zone to be considered:

- zones through which a public transport service either does pass now (no station or stop is present) or may pass with the new scheme

- zones that are not within normal walking distances of an existing or future public transport service

1.6.5   Zones in the first category should be focused on stops or stations. They should generally be sufficiently small for it to be realistic to expect people to walk from anywhere in the zone to the nearest public transport stop or station. This distance can be assumed to be less than a one kilometre radius in urban areas. Generally, there should be one zone of this kind for each stop or

station, although, in central areas where bus stops are frequent, it may be acceptable for there to be more than one stop in any individual zone. This guidance applies not only to existing services but also to the new scheme to be appraised. Thus, it is important to have some idea, at the model design stage, of where the stops and stations for the new services could be located.

1.6.6   In zones in the second category, the expectation is that people would not normally access the public transport system by walking. People accessing facilities in these zones are much less likely to use public transport than people in the first category of zone and, as a consequence, the modelling of these zones can be more approximate, that is, the sizes of these zones can be geared to highway modelling which are often larger. Access to public transport in these zones is often likely to be car; further detail is provided in TAG unit M5.1. There may also be the case of zones that may lie between these two categories, with some choosing to access public transport services by walking (a potentially longer distance), cycling or by car. In any case, it is important for the accuracy of the forecasts of patronage on the existing and new public transport services that access costs are realistically represented in the model.

1.7   Representing Responses of Public Transport Operators

1.7.1   An issue for forecasting is that, in principle, public transport operators can respond to changes in demand by either changing the capacity they provide or by changing the fares they charge. This requires due consideration in order to appropriately represent PT services in future years. Further information on the inclusion of competitive responses of public transport operators in forecasting, particularly in the context of rail, is in TAG unit M4.

2   Assignment Method

2.1   Introduction

2.1.1   The two different assignment methods that are in common use in public transport assignment software packages are **frequency-based** and **schedule-based** approaches, whilst a further distinction can be made between **stochastic** and **deterministic** assignment. The approaches are described in sections 2.2 and 2.3, with advice on deciding on which approach to adopt given in section 2.4.

2.2   Frequency and Schedule-Based Approaches

2.2.1   In the case of frequency-based assignment scheduled times are not considered explicitly, but modellers refer to the service headways, or to their inverse (the service frequencies). Therefore it is not possible to calculate explicitly attributes that users consider in relation to individual route options, but only the average values that relate to that line. Frequency-based modelling constitutes the classical approach as it is usually simpler, requiring less input data and less computational power than schedule-based approaches. It corresponds to the steady state approach to user equilibrium highway assignment and therefore allows the use of some of the same techniques.

2.2.2   Schedule-based approaches reflect the actual clock face vehicle arrival/departure times at the time when users make their choices. This approach allows modellers to take into account the dynamics of supply and demand, and calculate the dynamics and variation of level of service attributes.

2.3   Deterministic and Stochastic Assignment

2.3.1   The basic assumption underlying many assignment tools is the user equilibrium principle: 'All used paths are minimum (generalised) cost paths and all paths that are not minimum cost are not used'. In a deterministic user equilibrium (DUE) assignment it is assumed that passengers behave as if they share a perfect and equal perception of the generalised travel costs to their destination and all choose the cheapest option. In deterministic frequency-based assignment, multiple-routeing across the acceptable paths between OD pairs is achieved via the service frequencies. In the absence of

crowding, deterministic frequency-based assignment allocates passengers to acceptable paths in proportion to the relevant line frequencies which can collapse into All-or-Nothing assignment.

2.3.2   In contrast, stochastic user equilibrium (SUE) assignment recognises individual variations in generalised cost perception. All passengers are still choosing their perceived cheapest option, but this may not be the same for all since they do not necessarily agree on what the cheapest option is. A further random element can be added to the assignment by not only assuming that passenger cost perception includes an 'error term' but also by assuming that the vehicle departure times are slightly random. Both assumptions will lead to traffic being split between more paths than in the deterministic case, which better reflects what happens in real life.

2.3.3   In many cases the level of taste variation amongst travellers and imperfect information are likely to require the use of a stochastic route choice method for frequency-based assignment. For schedule-based methods the minimum cost route will depend on departure time, leading directly to multi-routing behaviour. However, the point about taste variations and imperfect information also applies here and a stochastic method may still be appropriate. Note that not all software offers the option of stochastic schedule-based assignment.

2.4   Deciding on an Assignment Approach

2.4.1   The choice between frequency-based and schedule-based approaches, and between deterministic and stochastic models is driven by practical questions. The impact of these considerations on the choice of the model to be used is summarised in Table 1:

- Is the PT system operating with high or low frequency?

- How punctual is the system?

- How regular is the system?

- What kind of passenger information is available?

- Does the demand vary significantly over the modelled period?

- How detailed is the demand information (by day, by hour, or even more specific)?

- Does the system experience capacity problems?

- How big is the network to be modelled?

- Is the network complex, so that regular users behave differently compared to occasional users?

- How homogenous is the likely user group? For example, is there a large difference in perception or valuation of travel time?

- What are the levels of interchange between services?

- How many different sub modes are there?

- What fare structures are used and do they differ between services/modes/operating companies?

- Is the necessary data available for schedule-based modelling?

Practical considerations

2.4.2   Compared with frequency-based models, an advantage of schedule-based approaches is that vehicle loadings can be predicted for specific services at specific points in time. However, if passenger arrivals and/or vehicle departures are highly variable, frequency-based approaches may

give more realistic results, whilst the extra data and calculation efforts of schedule-based approaches may be unnecessary with high-frequency systems.

2.4.3 Despite their theoretical advantages schedule-based methods suffer from a number of practical disadvantages:

- results can be very sensitive to the actual timetable specified, although, in this case, it could be argued that only a timetable-based assignment can give realistic results

- it can be difficult to predict the schedule accurately a) for a scheme which does not yet exist or b) for several years into the future

- the way in which unreliable services should be handled is not clear

- run times are much higher than for frequency-based approaches

- there are greater data and associated resource requirements

2.4.4 A schedule-based approach can be considered to more accurately capture the complexities within a public transport network and only in specific circumstances will a frequency-based approach be fully adequate. However, frequency-based approaches may be proposed where practical considerations mean that schedule-based approaches may be disproportionate or impractical, or in cases where it is judged that the model will be fit for purpose even in the absence of a full representation of the complexities. Where this is the case, a thorough appreciation for the approximations being made in the model and the impact these have on the outcomes (e.g. the appraisal) is of clear importance and should form part of the model specification agreed in the Appraisal Specification Report.

High and low frequency services – wait times

2.4.5 Passengers typically dislike waiting as part of their journey, hence it is weighted more highly than IVT and passengers will typically seek to time their arrival at a station or stop to minimise the amount of time that they wait. Where services operate at a high frequency, a commonly accepted threshold of which is services of less than 10-15 minutes headway, passengers are more likely to arrive without consulting timetables (i.e. 'turn up and go'), and the differences between desired and actual departure time can be treated as a constant, for example half the headway. However, if the service operates with a lower frequency than this suggested 10-15 min threshold[1], travellers will arrive at the station for specific services. From this it follows that frequency-based models are less suitable for services that operate with headways larger than the threshold. On low frequency services, passenger wait times should either be capped to a maximum value or derived from a wait curve (see section 3.2).

Passenger information and service punctuality

2.4.6 The more information a traveller has and the more reliable this information is, the more the choice will be service-based rather than route-based and hence a schedule-based approach will be more valid. Frequency-based models will be more suitable if services operate with low punctuality and/or a low level of user information. Delays and irregularity have to be treated implicitly or explicitly in schedule-based models. An implicit treatment is possible by adding error terms to the path choice model. A Monte Carlo technique allows the explicit treatment of delays, but availability of this may be limited in commercial packages.

2.4.7 Only in the case of a high frequency service that is unreliable or has poor passenger information is a frequency-based model fully sufficient. If the service is not yet operating, estimates will have to be made; this reduces the advantages a schedule-based approach might have. In the case of low

---

[1] Combined headway of all used services for the OD movement

frequency services or high frequency services that are reliable and have good provision of passenger information, schedule-based approaches are advised.

Service regularity

2.4.8 Service regularity is a separate issue from punctuality. In this case it is the regularity in the headway or scheduled intervals between the arrivals of the vehicles rather than unplanned delays. Frequency-based models assume an equal headway per service and thus an equal share of passengers between the runs of this service. If a service is not scheduled to arrive with regular headways, (for example 00, 15, 30, 45 after the hour), but say 10, 15, 40, 45 after the hour, this might lead to line loading errors in frequency-based models. Further, a schedule-based approach might be required if there is a major influx of passengers during a certain period (like an underground station connected to a train station that brings a large number of passengers to the underground network once every hour) in order to show overloading of certain services. An additional consideration in such a case is that a schedule-based approach is better equipped to estimate the correct average wait times.

Crowding and capacity constraints

2.4.9 In principle, if in-vehicle crowding is, or is expected to be, so severe that demand for the mode concerned is, or would be, constrained, some means of representing the costs of the crowding for use in the demand model would be required. In practice, crowding is more likely to be of importance in the allocation of trips between alternative routes through a combined network model than in models of separate networks. Section 3.5 offers advice on how to represent in-vehicle crowding costs.

2.4.10 Congestion in highway assignment and capacity restraint in public transport assignment are not the same. This is for two reasons. Firstly, the cost function in the case of PT is not increasing continuously, but the finite capacity of public transport vehicles will lead to a step function; either a traveller can board the arriving vehicle or not, in which case the waiting time will increase by one headway. It is worth noting that in practice this can have implications for model convergence. Secondly, this capacity restraint will only be experienced by boarders. Passengers on-board have priority and do not perceive the same increase in cost, although they may experience some increase in discomfort due to crowding. In frequency based-models it is possible to handle capacity restraint implicitly through a concept referred to as **effective frequency**. The idea is to increase the perceived costs of boarders through a local reduction in service frequency, reflecting the fact that the passenger may not be able to board a vehicle at a particular point because of overcrowding. This approach can be criticised for two reasons: a) a cost increase based on the number of passengers wanting to board and spaces available is still a continuous cost function; b) an increase in cost does not prevent line capacities being exceeded, leading to inaccuracies elsewhere in the network. Additionally, it is not clear how the correct wait time can be extracted for demand response modelling and appraisal. Schedule-based models can treat capacity restraint explicitly and the modeller can see which services suffer from capacity problems.

Scale of network

2.4.11 Because of the more detailed network description and because of the dynamic representation of supply and demand, schedule-based approaches are computational more demanding and data hungry, particularly in larger networks.

Variation in user behaviour

2.4.12 If the variation in user behaviour is an important issue, models using Stochastic User Equilibrium (SUE) assignment are needed. A 'dispersion factor' can be used to model the different cost perception of different travellers. SUE assignment can be applied to schedule-based as well as frequency-based models. SUE assignment should also be applied if one wants to reflect the behaviour of occasional users in complex networks. Occasional users might not know about all

available routing options and therefore the route choice might not be restricted to the least generalised cost path only. For low frequency services it is of less importance to distinguish frequent and occasional users, firstly because the route choice is in most cases not as complex and secondly because in low frequency services passengers will not often change their path en-route. The SUE models differ in their assignment assumptions. Logit, nested logit or probit models are most common. Where paths overlap significantly and hence path utilities are positively correlated (in practice usually the case), it is advisable to use the nested logit, C-logit or probit model. Logit models tend to be more tractable than probit models.

Summary

2.4.13 Table 1 summarises which assignment models are advised, depending on network characteristics and (to a lesser extent) passenger behaviour and the options to be modelled.

Table 1 Summary of recommendations for PT assignment model applicability			
		Schedule-based (SB) or frequency-based (FB)	**Stochastic User Equilibrium (SUE) or Deterministic User Equilibrium (DUE)**
Service frequency	High		SUE
	Low	SB	DUE
Passenger information & service punctuality	High	SB	
	Low	FB	SUE
Transfer choice-making by travellers	Pre-trip	SB	
	En-route	FB	
Regular schedule	Yes		
	No	SB	
Crowding/ Congestion	Yes	SB	
	No		
Capacity problems	Yes	SB	
	No		
Scale of network	Large	FB	
	Small		
Day-by-day variations	Yes	SB	
	No		
Significant dispersion of behaviour	Yes		SUE
	No		DUE

NB Blanks indicate that either option is appropriate

3   Generalised Cost Definition

3.1   Introduction

3.1.1 Generalised costs are used in the calculation of the utility of paths as perceived by travellers and therefore in determining the assignment of passenger flows to the paths. It is a combination of a number of different attributes of a path with each attribute being given its own weight or coefficient. The coefficients convert components to common units (time or monetary) and are chosen to ensure that the relative importance of each component for passengers is reflected. These attributes will normally be a subset of the following list:

- access time (from trip origin to PT stop)

- egress time (from PT stop to trip destination)

- transfer time (between PT stops)

- origin wait time (time spent waiting for first service on path)

- transfer wait time (time spent waiting for subsequent services)

- in vehicle time (weighting may vary by mode/vehicle type)

- fare

- transfer penalty (based on number of transfers times a fixed penalty, possibly differentiating between different transfer types)

- distance₂

- overcrowding

- quality of service and facilities at interchanges

3.1.2 Ideally a PT model would have the capability to model and build each of these measurements into the generalised cost function. However, this is not always the case and commercial software differs widely in the extent to which these are incorporated. The calculation of some specific components of generalised cost is considered below.

3.1.3 Where these features are used in a public transport assignment model that includes capacity restraint, a process will be required to ensure that the costs used in the iterative process of path loading and generalised cost estimation converge.

3.1.4 Where a variable demand model is used, compatibility (convergence) between the assignment and the demand model(s) is very important. To optimise processing time and ensure true converged solutions the travel cost formulations used in both should contain the same ratio of weights of in-vehicle, walking and waiting times, and the same ratio of weights of time and fare.

3.1.5 The various components of generalised cost are weighted in order to reflect the perceived time spent at each step of the public transport journey. IHT's Guidelines on Developing Urban Transport Strategies (May 1996) and ITS and John Bates's review of value of time savings in the UK in 2003 suggest:

- value of walk time = 1.5 to 2.0 times in-vehicle time

- value of wait time = 1.5 to 2.5 times in-vehicle time

- interchange penalty = 5 to 10 minutes of in-vehicle time per interchange

3.2 Waiting Time

3.2.1 The simplest assumption for the calculation of the mean wait time is to assume that it is half the headway. This assumes that passengers arrive randomly at the stop and that the service is reliable. This may be a reasonable assumption for services with short headways but for long headways it is more realistic to assume that passengers will try to time their arrival at the stop to minimise waiting time. For this reason some packages allow the definition of 'wait curves'. These define the waiting time as a function of headway. An example is shown in Figure 1 below. This gives the waiting time

---

2 Either distance-based fare differentials or if cost damping is employed (see TAG unit M2.1).

as half the headway for headways up to 15 minutes, after which the wait time is capped at 7.5 minutes.

3.2.2   Often it is only appropriate to use wait curves for the first service that is boarded, i.e. the one where the passenger has the most control over the time they arrive at the stop. For subsequent services on the path (i.e. interchanging onto subsequent services) it may still be appropriate to use half the headway, however long that may be – in effect, this assumes random arrivals by passengers.
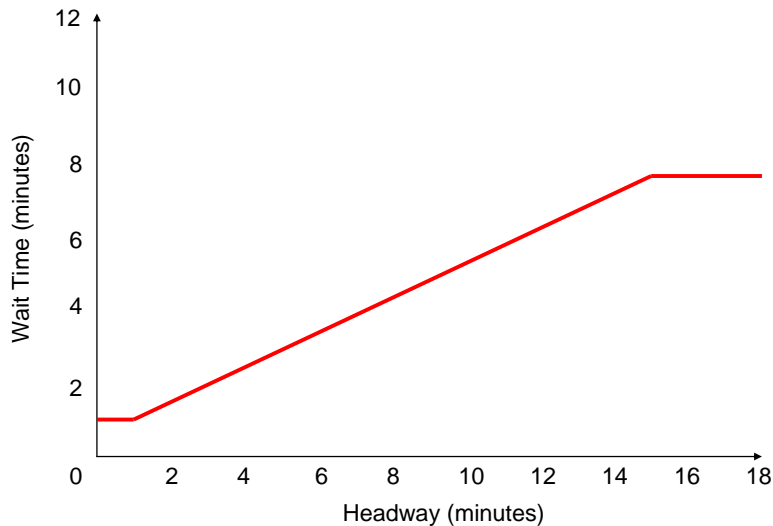

Figure 1 Example wait curve

3.2.3   It should be noted that when services are irregular (either planned or a result of poor punctuality), half the mean headway may actually be an underestimate of the mean waiting time. In this situation it is worth considering using wait curves where the waiting time is greater than half the headway. Reduced waiting times for a given headway can be used to model the effect of improved reliability and better passenger information, although both effects can be hard to quantify.

3.2.4   In schedule-based assignment systems it is possible to calculate intermediate waiting times exactly, due to knowledge of the timetable. For the first wait time, an arrival time must be estimated, or a generic time or boarding penalty may be assumed. In principle this will relate to the characteristics of individuals (e.g. genuine knowledge of the timetable, attitudes towards risk of missing services, etc.), the reliability of services and other factors that are required to be estimated as part of the model calibration process.

3.2.5   In the situation where there are "common lines" (i.e. a choice of services that will take users from a stop to the same destination), (first) waiting times are often calculated based on the combined headway of all used services, rather than calculating a waiting time separately for each service.

3.3   Fare Structures

3.3.1   Fares need not be included in the assignment provided that they do not influence route choice, and some software might not capable of modelling that. However, matrices of fares can be added to the generalised cost after the assignment and before passing cost matrices to a demand model or appraisal package. Where fares can influence route choice then it is essential to include them in the assignment. It is accepted that the complexity of some fare systems may prevent them from being represented exactly in the assignment model, but the model representation needs to be 'acceptable'. Acceptability can be gauged from whether the assignment model validates or not (see section 6).

3.3.2   There are a number of fare schemes that are used in PT modelling. Fares are converted to time using a value of time parameter in the assignment process. The various types of fare structures are listed below.

3.3.3 Distance-based. This usually consists of a fixed amount (boarding penalty/cost) plus a cost per km:

$$fare = a + b * distance$$

where a and b are user-defined. A variation of this definition allows b to vary by distance band.

3.3.4 Stop-to-stop fare. The fare is defined explicitly for each stop-to-stop combination.

3.3.5 Stage-based**.** For stage based fares, the fare depends on the number of fare stages passed by a trip.

3.3.6 Zone-based. This is similar to stage-based, but depends on the number of fare zone boundaries crossed (note that these are not likely to correspond to model zones).

Discounted Fares

3.3.7 A travel card typically permits unlimited travel within the area and for the time that the card is valid. Typically travel cards offer significant discounts to off-peak travellers or travellers who need to make many trips during a day. For assignment purposes travel card holders can be modelled as not having to pay a fare. However, a separate model may be required to forecast how many users will choose to purchase the card in the future.

3.3.8 The fare schemes may also be mixed together in some models. In some cases different structures and/or parameters can be used for different modes or sub-modes in the model. Note here that, if traveller types pay different fares (such as concessionaries) they may need to be treated as different user classes. Also, if fares differ between periods (e.g. off-peak fares) different models need to be constructed for each fare period. The most appropriate software package will depend on its ability to reflect the fares structure for the system in question.

3.3.9 Fares may not be important for holders of some kinds of travel passes, in which case they would need to be assigned as a separate user class. This is in addition to any other segmentation (e.g. by purpose or car availability) that might otherwise be required by demand modelling or appraisal.

3.3.10 Whatever levels of segmentation are used, it is important to ensure that the average fare paid across all segments in the model broadly correspond with what is paid in reality. For example, having one segment only that pays the full adult fare will not give an accurate representation of the true fares paid and hence potential operator revenues.

3.4 Boarding Times/Interchange Times

3.4.1 Boarding times, expressed in minutes, represent the time taken to board a particular service, taking into account factors such as demand at a stop or station, or number of doors on a vehicle. These can be defined as mode specific parameters and calibrated to reflect the relative attractiveness of one mode compared to another, and as such applied across a network.

3.4.2 Interchange times reflect the fact that walk time is incurred by a passenger when transferring from one service to another.

3.4.3 Passengers typically seek to minimise the number of times they interchange, as this incurs additional boarding time, interchange time and wait time, which is considered to be less desirable than in-vehicle time.

3.5 Capacity Constraints and Crowding

3.5.1 Section 2.4 discussed how the assignment package can deal with capacity constraints by locally reducing the effective frequency. Additionally, an overcrowding factor can be applied to the in-vehicle time. For instance the actual journey time might be 20 minutes but with an overcrowding factor of 1.5 the perceived journey time would become 30 minutes. The overcrowding factor

represents the additional discomfort and inconvenience to passengers – all passengers and not just boarders. It makes the crowded service less attractive to travellers, and will reduce the general attractiveness of PT in a mode choice model. Care must be taken when deriving costs for demand modelling and appraisal to ensure that the appropriate times (i.e. with or without the factor) are used.

3.5.2   The overcrowding factor is typically a continuous function of the flow to capacity ratio on the service, as shown in Figure 2. Taking into account of the two effects of crowding/discomfort/standing and physically not being able to board a service can avoid undesirable biases in the results when capacity constraints are active.
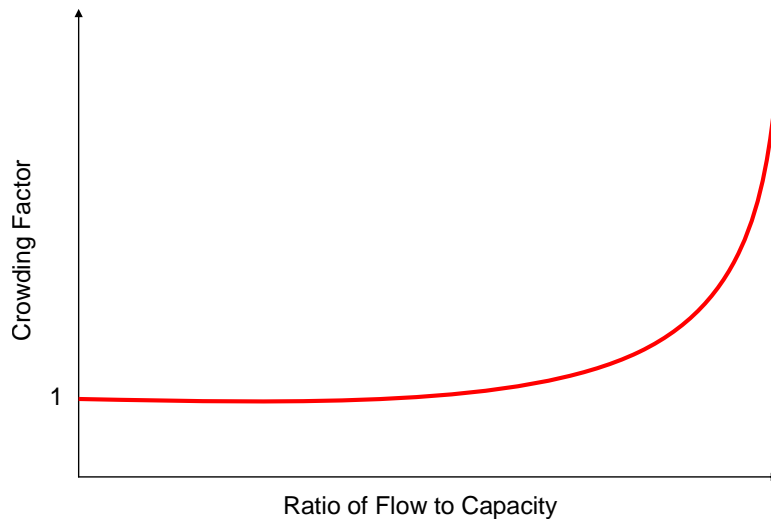

Figure 2 Typical crowding curve.

3.5.3   The number of vehicles that can use a road/track becomes important when capacity limitations cause delays to PT services. Current commercial software can only model this effect for road-based modes (bus and possibly LRT) but not heavy rail. They usually require a link to a highway assignment model in which PT and private vehicles compete for road space. Congested travel times for PT vehicles can then be fed back into the PT model. Modelling the effect of capacity constraints in public transport has the effect of making the generalised cost of travel time dependent on passenger flows (for example, more passenger flow, more crowding, but less car traffic, less congestion, faster travel time, more passenger flow, etc.).

3.5.4   The introduction of crowding has significant practical implications for PT assignment, namely the need for assignment to be an iterative procedure with a consequent impact on run times, the need to achieve convergence, and the need to calibrate crowding curves. For these reasons crowding should only be modelled where it is likely to have a significant effect on traveller behaviour or where an effect on crowding is one of the objectives of the scheme. Where crowding is not modelled it is still important to monitor volume to capacity ratios when forecasting to determine whether crowding will become a problem in the future.

3.6   Quality Factors

3.6.1   Quality factors, often referred to as "soft measures", incorporate values for comfort, security, information provision, ease of interchange, etc., into the costs of using public transport[3]. For example, where a bus service is particularly comfortable, with good passenger information systems and well lit interchange facilities, passengers may elect to use this service as opposed to a potentially quicker service that is of a lower quality.

---

[3] See TAG unit A3.1 - Social Impact Appraisal for a more complete description of quality factors.

3.6.2   This behaviour can be represented in a transport model in the assignment stage or in the mode choice model (see TAG unit M2.1). Within the assignment there are a number of ways to incorporate these into generalised cost, for example as reductions in wait times or factors to apply to in-vehicle times.

3.6.3   Valuations for bus quality factors are presented in Data Book Table M3.2.1 for bus users and car users, in generalised minutes:

M3.2.1: Segmented Values of Bus Quality Interventions (generalised minutes)

3.6.4   The following points should be noted when considering these quality values:

- Only overall figures have been presented for audio announcements, climate control and new bus shelters because the segmented bus and car user figures obtained from the models were not significantly different. Overall valuations obtained from an unsegmented model are also presented (they are an average of the bus and car user figures).

- The valuations are in generalised minutes. Introduction of a quality measure does not represent a time saving as such, but will increase attractiveness and can therefore be modelled as a reduction in generalised time.

- Valuations may not remain constant over time. The figures have been calculated as additive factors on the basis of respondents' current journeys, but as different aspects of these journeys change with time, the quality measure valuations may also change, requiring updated research.

- There is not a direct correspondence between the car and bus user categories and segmentation on the basis of car availability. However, it is suggested that the car user figures could be used as a proxy for the car available segment, with the bus user figures used as a proxy for the non-car available segment.

3.6.5   The valuations relate to commuting trips. Valuations for other trip purposes are likely to be different and valuations may be obtained by using the ratio of values of time (VOTs) for other trip purposes and the VOT for commuting. It is recommended that local VOTs are used where these are available, or VOTs from the Data Book Table A1.2.1 otherwise.

3.6.6   When evaluating the impact of introducing a number of measures, quality values can simply be summed. This is because research found little evidence of a "package effect", i.e. where the introduction of a package of soft measures was valued by travellers differently from the sum of the individual measures.

3.6.7   It is recommended that where buses are less frequent than every ten minutes, or where there are known problems with punctuality or reliability, the quality values should be reduced to account for this. This is because qualitative surveys did indicate the existence of a "service provision" effect, linking the perceived quality of features to the harder measures of service frequency and reliability.

3.6.8   Where a quality factor is already partially in place but is to be enhanced, the quality values should also be reduced. Additionally, if the scope of the proposed measures is narrower relative to those evaluated during the research, then again the values should be reduced. These reductions should be in proportion to the relative quality that exists before and after the enhancements are made (relative to the 'maximum' values presented).

3.6.9   Relative valuations attributed to different levels of information provision are also available. These are presented in Table 2 as percentages relative to the bus stop Real Time Passenger Information (RTPI) values presented in Data Book Table M3.2.1.

Table 2 Valuation of Information Provision	
Attribute	**Valuation**
Real Time Information in City Centre	83%
Real Time Information at Bus Station	85%
Real Time Information at Bus Stops	100%
SMS Real Time Information (Free)	64%
SMS Real Time Information (10p cost)	31%
SMS Timetable Information (Free)	13%
Web Based Information	29%

3.7 Mode Specific Values of In-Vehicle Time

3.7.1 In some instances, factors may be applied to the in-vehicle times to reflect people's preferences for the various modes. This is most likely to be relevant where the influence of fare on the choice of routes and services is likely to be quite weak and, as a result, the fare term may be excluded from the generalised cost formulation in the assignment. These in-vehicle time factors may be interpreted as mode-specific values of in-vehicle time. Thus, instead of an in-vehicle value of time of unity being used, non-unity values of in-vehicle time are used to represent the inherent, relative attractiveness of the various modes.

3.7.2 In-vehicle time factors or mode-specific values of time are best obtained from a properly designed and conducted stated preference experiment. An alternative method sometimes used is to calibrate in-vehicle run time factors, by a process of trial and error, during the assignment model development. However, this is a less precise way of identifying correctly the relative attractiveness of the alternative modes and a properly designed and conducted stated preference survey is to be preferred.

4 Path building

4.1 Introduction

4.1.1 The objective of path identification is to find all potentially attractive paths and calculate their cost. This consists of three stages:

- Identification of least cost paths between specific OD-pairs

- Establishing connectivity between PT paths

- Selection of acceptable PT paths
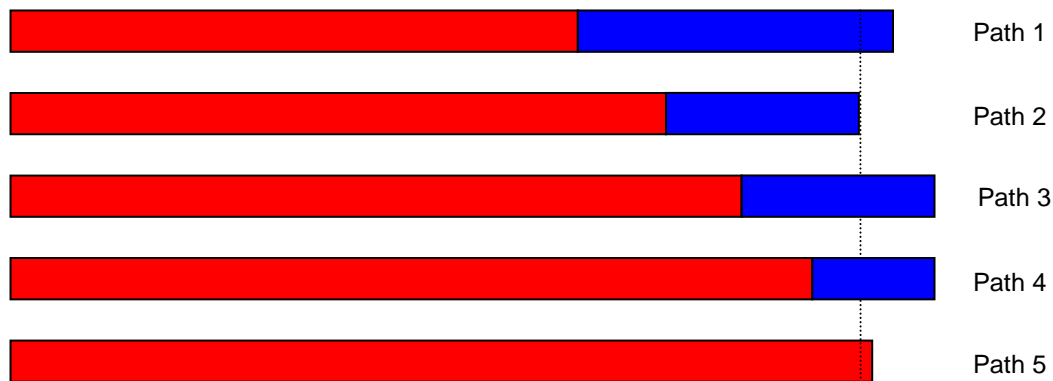
4.2 Identifying Acceptable Paths

4.2.1 There are two common methods for identifying acceptable paths.

Method 1 (frequency-based methods)

4.2.2 In this method the first step is to identify the shortest path excluding the waiting time for the first service from the generalised cost function. This represents the best option if the passenger can time their arrival at the stop/station to avoid any waiting. In practice however the passenger may well have to wait for this 'best path' and it may be possible to get to the destination sooner by taking an earlier service, albeit one which has a slightly higher journey time. In other words the passenger may choose to take a path with a longer journey time, provided that in return they benefit from a short waiting time. Any path whose generalised cost excluding wait time is higher than the

generalised cost of the best path plus the headway of the best path will not be used, i.e. the passenger will reach the destination sooner by waiting for the service on the best path[4].

4.2.3   The process is illustrated in Figure 3. The best path (excluding the origin wait time) is Path 1. On the other hand the path with the lowest maximum generalised cost (generalised cost excluding waiting time plus headway (headway=maximum waiting time)) is Path 2. Paths 2-4 all have a generalised cost (excluding the initial wait) that is less than the generalised cost of best path plus the headway of the best path; they are therefore considered acceptable paths, i.e. in some circumstances it would be better to take one of these paths rather than wait for the service on Path 1. Path 5 is not acceptable because the generalised cost is too high – it would always be preferable to wait for the service on Path 2.

KEY

Generalised cost excluding waiting at start of journey

Headway

Figure 3 Identification of acceptable paths (Method 1).

Method 2

4.2.4   Minimum generalised cost paths are identified between OD pairs to establish base path costs. A set of paths between each OD pair is then generated using simple network connectivity rules. These represent sets of links a traveller could use to get from an origin to a destination. These may be constrained to some computationally practical maximum number.

4.2.5   These are then constrained further to a set within some cost range, i.e. all paths with a cost a certain amount higher (expressed either as an absolute or percentage difference) than the minimum cost path will be discarded. This may also involve limiting number of transfers and interchanges for example (although these may already be built into the generalised cost function) and the total length of walk segments. This path set may be further reduced at the path choice stage, for instance if less than X% of flow uses a path it may be discarded for computational reasons.

4.3   Path Choice Methodology

4.3.1   Ultimately the best test of the adequacy of a particular algorithm is its ability to reproduce observed routing behaviour. Several methods are described in this section that should be considered in order to achieve this.

---

4 There is an assumption here that passengers are fully aware of the timetable.

4.3.2   Where multiple paths are identified, some mechanism for allocating flow to each path is required, usually as a function of the generalised cost on each path, including all aspects of time (including access and necessary further interchanges), fares and comfort (journey quality). Ideally, explicit consideration would be given to common/overlapping and parallel paths (i.e. where the 'common line' dilemma occurs) and some way of including the representation of individual preferences may be necessary through probabilistic or stochastic methods. Path choice is governed by calculations of 'probability of use' of each of the acceptable paths between OD pairs. As noted earlier a useful distinction can be made between deterministic and stochastic methods.

All-or-nothing (deterministic) assignment

4.3.3   In an all-or-nothing assignment all flow is loaded onto the single minimum cost route for each OD pair. With frequency-based methods there is therefore no multi-routing. This may be an adequate reflection of reality in some cases, particularly in schedule-based models or simple networks. In others, e.g. complex urban networks, there is likely to be observed multi-routing which would require a more complex assignment method to model accurately.

4.3.4   The all-or-nothing assignment is a deterministic method. Methods below are probabilistic or stochastic.

Simple discrete choice

4.3.5   In these stochastic methods no consideration is made of whether paths are overlapping or in parallel. Only the generalised cost on each path is considered. The following discrete choice functions are used:

- Logit: the most commonly used discrete choice model where passengers are distributed over a set of paths according to the absolute difference in cost between the paths

- Power function (Kirchoff): passengers are distributed over paths according to the power of the ratio of the costs of alternative paths

- Box-Cox: a flexible model form that includes power and logit as special cases

- Lohse: uses the ratio of path costs relative to the minimum cost

- Probit: similar to logit, although error terms are normally distributed rather than a logistic distribution

4.3.6   In each case the 'spread' of the path choice can be controlled by a user-defined parameter. This determines how strong the preference is for the minimum cost path. This will depend on the level of taste variation among passengers and how complete their knowledge of services is ('errors in perception').

4.3.7   With the exception of probit (which is not actually used in commercial packages) all of the above have theoretical shortcomings regarding their ability to deal with a choice between correlated alternatives. Path utilities will be correlated if, for example, they share a common segment.

Models with 'Independence'

4.3.8   The choice models given above in their basic form do not cater adequately for schedule-based stochastic assignment. Temporal factors are therefore incorporated into the models in order to make them more suited to schedule-based PT routeing. In order to do this, interactions between different connections are defined:

- the temporal proximity of the connections with regard to departure and arrival

- perceived journey cost differences between connections

- fare differences between connections

4.3.9   These factors are combined to derive an **independence of connection** factor which defines the attractiveness of a particular connection relative to all others. They ensure that identical alternatives are assigned same volumes of passengers if no other connections with temporal proximity have an effect.

Service frequency model

4.3.10   Passengers are assigned to a path according to the frequency (or a function of frequency) of services along available paths, i.e. the probability of using a path is proportional to its frequency or a function thereof. This is a simple approach where travellers are assumed to possess no knowledge of timetables or journey times and take the first reasonable service from the stop.

Service frequency and cost model

4.3.11   In this extension of the service frequency model the path choice probability is modified to reflect the difference in costs between the paths. Passengers are assumed to have some knowledge of the frequencies and journey times of alternative services and will decide whether to take the first feasible service from the stop or wait for a faster one.

4.3.12   In all but the simplest public transport networks, travellers between certain OD pairs are likely to be split between different paths and services. Therefore a multi-routing algorithm must be used to reproduce this behaviour. Most path identification methods are acceptable; the crucial part of the algorithm is how the flow is allocated to the used paths. Methods that take into account generalised costs, rather than just frequencies are likely to produce better-validated results. Where there are overlapping routes methods that consider the degree of independence between competing routes should, ideally, be used. However, at the time of writing this (probit) was not available in the commercial assignment packages that have been reviewed.

5   Cost Skimming

5.1   Introduction

5.1.1   The skimming of costs from assignment is important as skimmed costs are used in demand modelling and in economic appraisal (for instance as input into TUBA). Calculating costs along a particular path is straightforward. However, packages differ in the way that the costs on individual paths are combined to provide a single skimmed cost for each origin-destination pair.

5.2   Methods

5.2.1   A number of different methods of skimming costs are available, with some packages offering more than one option. It is common practice to used flow-weighted average costs of paths used (as used in TUBA), available as either total generalised cost or for individual components of cost. Other methods are available and may be more appropriate given certain circumstances and the requirements for their subsequent use:

- costs on minimum cost path; usually available as either total generalised cost or for individual components of cost

- straight average over all used paths; usually available as either total generalised cost or for individual components of cost

- frequency-weighted average costs; usually available as either total generalised cost or for individual components of cost

- composite cost (logsum where path choice is based on a logit model); available only for total generalised cost, not individual components

5.2.2 For input into TUBA it is important to be able to skim the individual components of generalised costs separately, particularly travel time and fares. For example, times for business trips will need to be unweighted total OD travel times; for consumer trips they need to be weighted OD travel times, using the weights for waiting and walking time (see paragraph 3.1.5).

5.2.3 For input into TUBA and the demand model costs need to be skimmed for each demand segment (e.g. combination of purpose, car availability and/or income). It is worth noting that it is possible to have fewer segments in assignment than in the demand model, provided the demand segments map unambiguously onto the assignment segments.

5.2.4 The results of different skimming procedures may lead to rather different results. The method used for skimming costs should be consistent with that used to split flow between routes5. For instance, if a logit model of route choice is used then the skimmed cost should ideally be the logsum measure, although this is not possible in many existing packages and TUBA requires different components of costs rather than a total generalised cost only. The current TUBA recommendation is to use flow-weighted average route costs. For input to TUBA these need to be separated between fares and the different components of time-related costs.

5.2.5 The skims also need to feed the demand model which itself may require skims of individual cost components and apply coefficients that vary by purpose and/or person type. This can be a problem if in the skimming procedure the model aggregates over routes. Any inconsistency here can lead to counter-intuitive results.

5.2.6 Overall, it is recommended that the assignment package's skimming capabilities are assessed before committing to its role in the modelling structure, to ensure that a robust interface between assignment and demand model can be achieved.

5.2.7 A final issue in skimming is the implications of using biased networks (also identified in TAG unit M2.1). To enable the skimming of levels of service for each of the public transport sub-modes as input to the choice model, the assignment networks may need to be manipulated or biased to favour one or more of the modes. This may also be done to obtain consistency between the sub-modal split in the demand and assignment models. The preferred approach is of course to avoid the need to do this, with network models producing sufficiently accurate costs to allow a suitably calibrated demand model to reproduce observed sub-mode shares.

5.2.8 If biased networks cannot be avoided, the analyst should ensure that the amount of bias introduced is as little as possible and recorded. A quantitative assessment should be made of the level of bias introduced, and its acceptability. Care should be taken that the biased networks used in application are also used for model estimation, and in future forecasting, the biases should be retained.

6   Model Validation

6.1.1 TAG unit M2.2 provides guidance on the development of prior trip matrices. If prior matrices do not validate satisfactorily compared to the guidelines set out in this section, the analyst may:

- reconsider the development of the prior trip matrices with a view to produce a new version which, when assigned, yields modelled flows that accord more closely with the counts

- refine the trip matrices using **matrix estimation**, as discussed below in this section

- explain the model performance and potential limitations for its use

---

5 Mott MacDonald, TUBA project.

6.1.2   Matrix estimation may be used to adjust the trip matrices to accord more closely with the counts. Matrix estimation generally has to be done at the most aggregate level (as passenger counts are unlikely to be available reliably segmented by any aspect of traveller type), although matrix estimation software may allow input proportions of trips by type to be preserved in the adjusted or output matrix.

6.1.3   The changes brought about by the matrix estimation process should be relatively small and must be examined to check for particular distortions. If distortions have been introduced, the count data being used as constraints should be checked for consistency. Use of the technique is likely to be most suspect when it produces large changes from the prior trip matrix, especially if the changes are, in proportional terms, uneven across the matrix. TAG unit M3.1 describes the assessment of the changes brought about by matrix estimation to highway matrices. The same principles should apply to PT matrices.

6.1.4   The **validation** of a public transport passenger assignment model should involve three kinds of check, which are:

- validation of the trip matrix

- network and service validation

- assignment validation

6.1.5   Validation of the **trip matrix** should involve comparisons of assigned and counted passengers across complete screenlines and cordons (as opposed to individual services). **At this level of aggregation, the Department's suggested guideline is that the differences between assigned and counted flows should, in 95% of the cases, be less than 15%.** Where models do not achieve the guidelines the analyst should review the assumptions and quality of data used to develop the trip matrices, but should not impose constraints just to improve the base year accuracy of the model. In reporting the analyst should explain why the model does not reproduce counts to these tolerances and should indicate the scale and nature of potential forecasting uncertainty and suitability of the model for its intended purpose.

6.1.6   Validation of the **network** should involve checks on the accuracy of the coded geometry and times/speeds in the model (i.e. for in-vehicle, access and interchange times).

6.1.7   Validation of the **services** should involve comparing the modelled flows of public transport vehicles with counts (as well as other features such as stopping patterns for rail, etc.).

6.1.8   Validation of the **assignment** should involve comparing modelled and observed:

- passenger flows across screenlines and cordons, usually by public transport mode and sometimes at the level of individual bus or train services

- passengers boarding and alighting in urban centres

6.1.9   The Department's recommendation is that across modelled screenlines, modelled flows should, in total, be within 15% of the observed values. On individual links in the network, modelled flows should be within 25% of the counts, except where observed hourly flows are particularly low (less than 150 passengers per hour).

6.1.10   The validation of assignment models of separate modes should be comparatively straightforward if the network, services and trip matrices have validated satisfactorily. The validation and subsequent recalibration of an assignment model of a combined network may be considerably more problematic.

6.1.11 Wherever possible, a check should be made between the annual patronage derived from the model and annual patronage derived by the operator. Precise comparisons may be difficult but may be sufficiently accurate to provide a cross-check on the general scale of patronage, bearing in mind that operator patronage is likely to be boardings and not trips.

6.1.12 If the validation fails to meet the required standards, the assignment model may be recalibrated by one or more of the following means:

- adjustments may be made to the zone centroid connector times, costs and loading points

- adjustments may be made to the network detail, and any service amalgamations in the interests of simplicity may be reconsidered

- the in-vehicle time factors may be varied

- the values of walking and waiting time coefficients or weights may be varied

- the interchange penalties may be varied

- the parameters used in the trip loading algorithms may be modified

- the path building and trip loading algorithms may be changed

- the demand may be segmented by person (ticket) type

6.1.13 The above suggestions are generally in the order in which they should be considered. However, it would be hard to argue that the priority between suggestions, which are adjacent in the above list, should be maintained in all instances. In all cases, any adjustments must remain plausible. Where models do not achieve the guidelines the analyst should review the assumptions and quality of data used to develop the trip matrices, but should not impose constraints just to improve the base year accuracy of the model. In reporting the analyst should explain why the model does not reproduce passenger counts to these tolerances and should indicate the scale and nature of potential forecasting uncertainty and suitability of the model for its intended purpose.

7 References

Institution of Highways and Transportation (1996) Guidelines on Developing Urban Transport Strategies.

Institute of Transport Studies, University of Leeds, in association with John Bates Services (2003), Values of travel time savings in the UK.

8 Document Provenance

This restructured unit is based on former TAG unit 3.11.2, which itself was based on Major Scheme Appraisal in Local Transport Plans Part 3: Detailed Guidance on Forecasting Models for Major Public Transport Schemes.