Department for Transport

Provision of Technical Advice and Support for Matrix Building Guidance

Technical Report

Rep_20161219

Final| 19 December 2016
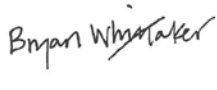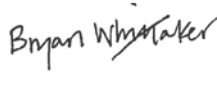
Job number   246022

ARUP

Document Verification

ARUP

Job title		Provision of Technical Advice and Support for Matrix Building Guidance	**Job number**	
			246022	
Document title		Technical Report	**File reference**	
			04-05-08	

Document ref					

Revision	**Date**	**Filename**	20151209 Interim Report Draft Issue1.docx		
Interim Report	09 Dec 2015	**Description**	Interim Report for Discussion		
			Prepared by	Checked by	Approved by
		Name	Various Authors	Matthew Sinnett	Bryan Whittaker
		Signature			
Final Draft	5 Feb 2016	**Filename**	20160205 Matrix Guidance Final Draft.docx		
		Description	Incorporating DfT Feedback on Interim Report		
			Prepared by	Checked by	Approved by
		Name	Various Authors	Matthew Sinnett	Bryan Whittaker
		Signature		*Msinnett*	*Bryan Whittaker*
Final	19 Dec 2016	**Filename**	20161219 Matrix Guidance Final_Dft Rev1.docx		
		Description	Incorporating DfT Feedback on Draft Report		
			Prepared by	Checked by	Approved by
		Name	Various Authors	Matthew Sinnett	Bryan Whittaker
		Signature		*Msinnett*	*Bryan Whittaker*
		Filename			
		Description			
			Prepared by	Checked by	Approved by
		Name			
		Signature			

Issue Document Verification with Document ✓

Contents

Tables

Table 1: Incremental model types

Table 2: Summary of Sources of Error and Sources of Variability in Traffic Counts

Table 3: Worked Example of Count Error Estimation

Table 4: Worked example of merging trip estimates

Table 5: Case Study Examples

Table 6: Incremental Model Structures

Table 7: Eightcases' method for applying synthetic growth (taken from Daly et al, 2012)

Figures

Figure 1: Alternative Matrix Forms

Figure 2: Model Structure Type 1-A - Traditional (4-stage) Model

Figure 3: Alternative Incremental Model Structure Types

Figure 4: Building Matrices from RSI Surveys

Figure 5: Overview of Synthetic P/A Matrix Creation Process for Home Based Purposes

Figure 6: Matrix Adjustment Procedure

Appendices

Appendix A

Eight Cases Method for Applying Synthetic Growth

Appendix B

Merging Data from Different Sources

Executive Summary

Background

The construction of base matrices is one of the key stages in building a transport model. This is obvious if the general guidance in WebTAG towards incremental models is followed, since the matrices become the 'pivot' about which changes are predicted. However, even when 'absolute' or fully synthetic models are being used in forecasts, the base year matrices are an essential aspect of model validation.

To date, however, there has been little comprehensive guidance on best practice in preparing base year matrices. Existing guidance and practice concentrates mainly on the development of highway origin-destination (O-D) matrices used for assignment purposes rather than the development of production/attraction (P/A) and multi-modal matrices that are used in demand modelling. There is a recognised need to bring these elements closer together.

DfT has commissioned this report on *Technical Advice for Matrix Building Guidance*. It is complementary to a separate commission to provide advice on the use of mobile network data for matrices (undertaken for the DfT by the Transport Systems Catapult). The aim, following consultation, is to use both sets of advice as a basis for preparing a new WebTAG module on base matrix development.

The primary intention of the authors has been to provide advice on good practice in preparing trip matrices. This project has considered varying techniques for matrix building for differing applications and also the use of suitable data sources when it is increasingly difficult to collect survey data. An important aim was to ensure that matrices are theoretically sound and fit for purpose. The advice in this report is intended to be efficient and practical, but also durable as policies, methods and data sources evolve.

In order to set the advice in context, it has been necessary to first consider the purpose for which the matrices will be used and how they are to be used. This will allow practitioners to appreciate the basis for the recommended methods, some of which require a considerable departure from common current practice.

Key recommendations

1. Any matrix building exercise should begin with a statement of the usage objectives for the matrices and the approach should reflect these aims.

2. A full data statement should be produced, explaining what is known about all input data, its provenance, definitions, processing history, quality, suitability and how the data features in the model. Data should not be used when unsure about its suitability without strong caveats.

3. Surveyed (so called 'observed') data should not be considered as any more valid than synthetic/modelled data as a matter of course. Each source should

be evaluated for its value, errors and biases. It is not good practice to freeze and retain surveyed cells during matrix development as if they are specially trusted without thorough justification.

4. Validation of base matrices is an important aspect of model development, but an evolution of current practice and guidance is proposed. Generally, all relevant available data should be used in developing the matrix with concurrent calibration and validation. It is not recommended to hold data back specifically for the purpose of what is sometimes referred to as independent validation, except in special circumstances, but rather to concentrate on how well the model fits all of its inputs, recognising the mutual inconsistency that usually exists among the data sources.

5. Generally (except for minor schemes, those where variable demand is not a feature or early scheme business case development), matrix construction and modelling should be in both P/A form and O-D form for the purpose of demand and assignment models respectively. By preference, changes to O-D form during calibration and validation should feed back to P/A form.

6. A suitable balance should be sought between design of the matrices for good forecasts (segmentation, spatial accuracy and constraints) and achieving the best possible validation for the present day. In many cases this means that constraining matrices to present day traffic or passenger counts will not be an overarching priority, but that counts will be part of a more holistic matrix construction process. Validation effort should be concentrated in areas where schemes are to be tested. Explanations are needed when validation fit is below ideal levels, but this does not necessarily mean that the model is unfit.

7. The technique of matrix estimation to counts should be used appropriately. It should not usually be applied in order to over-fit matrices to agree with count data as a final step of adjustment. Nonetheless, counts can be very valuable data and matrix estimation can be part of a process to help introduce them to matrices in a balanced way alongside other data.

1   Introduction

1.1   Background

Arup was selected by the Department for Transport (DfT) in October 2015 to provide *Technical Advice for Matrix Building Guidance*. The Arup project team has been supported by two transport modelling specialists, Dr John Bates of John Bates Services and Prof Hugh Neffendorf of Katalysis Limited. In addition, Ian Williams, another independent expert, has provided valuable internal technical assurance.

The emphasis of the project is on a) highway trip matrices and b) trip matrices by all modes in a multi-modal modelling context. We have not addressed specifically what might be required for uni-modal models (other than highway), although our discussion in the multi-modal context will have some relevance for these.

Later chapters of this report are intended for a technical audience and assume a familiarity with transport modelling terminology and an understanding of transport modelling theory or practice. The DfT aim, following consultation, is to use this document as a basis for preparing a new WebTAG module on base matrix development.

Existing WebTAG Guidance (Unit M2, paragraph 1.3.1) notes that, "The Department's long-established preferred approach is to use an incremental rather than an absolute model, unless there are strong reasons for not doing so." We will define the terms 'incremental' and 'absolute' more carefully in Chapter 2, but essentially an incremental model predicts a forecast change to an accepted base matrix while an absolute model predicts a future travel demand independent of the base. In the recommended incremental case, base trip matrices are a fundamental element of transport models, whether for local schemes or large-scale strategic and multi-modal studies.

In spite of this, there has been little specific guidance on best practice in preparing base matrices. Existing guidance and practice concentrates mainly on the development of highway matrices used for assignment purposes: in this respect the guidance (M3.1) goes back to Chapter 8 of the *Design Manual for Roads and Bridges (DMRB)* Volume 12 Section 1 Part 2 (August 1991), also known as the *Traffic Appraisal Manual* (TAM) and to §4.3 of DMRB Volume 12 Section 2 Part 1 (May 1996). On the public transport side the guidance (M3.2) essentially follows the line of M3.1, although with different data sources. As far as demand modelling is concerned, the relevant section in the guidance (M2) is 2.5.4-2.5.10 with an illustration given in Appendix B of that guidance.

As we will discuss, not only has the availability of data[1] changed in the last 20 years, but the emphasis in DMRB is on **origin-destination** (O-D) matrices rather than the development of **production/attraction** (P/A) and multi-modal matrices

---

[1] Of particular importance is the increasingly greater difficulty of conducting roadside interview surveys, which are a major component of the DMRB guidance.

which are used in demand modelling. We discuss this further in Chapter 2. There is a recognised need to bring these elements closer together.

This project has considered varying techniques for developing matrix building for differing applications and also the use of suitable alternative approaches that rely less on roadside interview (RSI) survey data. An important aim was to ensure that matrices are theoretically sound and fit for purpose. The advice in this report is intended to be efficient and practical, but also durable as policies, methods and data sources evolve.

The project has been concerned with person trip matrices in a conventional sense, as reflected in typical transport models and in WebTAG guidance. Such transport models work with a simplified version of current trip-making choices and assume that the underlying behaviour will hold in the future, although it is of course understood that this will not always be the case. In studies that might consider very substantially different social, policy or technical environments from today, the principles of good practice described in this report will be relevant, but other sources of data and model structures are likely to be needed. The project has not extended to freight movements or to the matrices implied in land use/transport interaction models. Similarly, it was not intended to consider very different forms of modelling such as micro-simulation or household activity models. This review is complementary to the separate project to support guidance on use of mobile phone data for matrices that has been undertaken for the DfT by the Transport Systems Catapult.

1.2   The role of trip matrices

As noted, trip matrices are the basic building blocks from which conventional transport models are developed. They represent the demand for travel between geographic areas, which are generally referred to as zones. While matrices have valuable roles in supporting operational or environmental evaluation, it can be argued that their most important application is to support the economic assessment of alternative schemes or policies, generally for some future time. To achieve this, the demand represented needs to be sensitive to factors under consideration, including changes in land use, demographics and the economy (e.g. fuel prices, vehicle ownership) or supply (e.g. network, parking, fare) costs that might vary between schemes and policies. This implies that the forecast matrices should be segmented and available in a form that can react to the variations under test.

In most cases, the trip matrices are also assigned to the networks under consideration to give flows to support design, operational and environmental evaluation and as part of an iterative cycle to determine how demand loaded onto the infrastructure might cause supply costs to change.

An early stage of transport model development is the construction of base or current year matrices from synthetic models, available or specially commissioned data sources or, generally, a combination of the two. These base matrices, which are the subject of this report, should be built primarily to support their future uses but they are also used in the important process of current year validation to give

assurance that they are a fair representation of the present day and will be suitably responsive to the intended applications.

Of course, the fact that a model has a satisfactory base year validation does not guarantee that it will form a suitable basis for forecasting. Nonetheless, a suitable degree of validation is a necessary and important step that will improve the level of confidence in the forecasting model.

Achieving a good base year validation from models and sample surveys is not an easy task, and methods (e.g. matrix estimation from counts – ME) have evolved which modify the matrices in order to fit more closely with count survey data. And, as described later in the report, achieving the best current fit can sometimes be at odds with obtaining a good matrix structure for demand forecasts. The report contains advice on achieving a suitable balance for differing applications.

The distinction between the P/A matrix form that is required for demand models and the O-D form[2] for assignment models is known to most transport modellers but is still worth emphasising since it forms a central part of the advice in this report. For matrices to be sensitive to land use changes, including socio-demographic change, they need to indicate the home end of trips (the most important source of the **production** of trip making). In addition, the modelling of destination choice (or 'distribution') only makes sense on a P/A basis. But it is the O-D matrices used for assignment that form, to a large part, the basis of current model validation, typically comparing with traffic or passenger counts, which are directional and usually contain no P/A intelligence.

As we will discuss, it is relatively easy to transform P/A matrices to O-D form. However, in this process the P/A information is lost and it is then much more difficult to relate any subsequent O-D matrix adjustments made in order to improve validation (generally through ME) back to the underlying P/A form.

The significance of this particular issue is substantial. Modellers today do not usually attempt to address the dilemma that the O-D and P/A forms have become inconsistent, once the O-D matrices have been altered through matrix estimation for example: instead they try to apply demand changes for a future year to the post-ME O-D form, with only a crude link to the underlying P/A form that should be used for forecast demand studies. In this way, the performance of the model in reflecting traveller behaviour under future conditions can be seriously weakened.

1.3   Matrix construction and current practice

Section 2.6 describes a brief history of matrix construction in the UK. An appreciation of this history provides some useful understanding of how approaches and requirements have changed over time and why this review is taking a broad and durable view.

Current WebTAG guidance essentially focuses on unimodal methods, and predominantly on highway matrices where, as noted, current practice has tended to retain an O-D basis as the essential outcome of base matrix validation.

---

[2] These terms are discussed in more detail in Section 2.2

In considering the need for new and updated guidance, it is useful to reflect on the wide range of users who rely on WebTAG modelling units. Clearly, these include consultants who support DfT studies but the guidance is also important to major scheme sponsors such as Highways England or Network Rail and to local authorities, including those in major metropolitan areas, or even to private developers. In view of this range of applications, all types of conventional trip matrix development have been considered to be in scope for this review.

The relative balance of strengths required in matrix development between spatial accuracy and segmentation detail will differ between types of applications. At the extremes this indicates that:

a) For small to medium scale network enhancements, the accuracy of the modal matrices for those trips that may pass in the vicinity of the eventual scheme will be critical, whereas variable demand and the need for segmentation by trip purpose and person type may not be as great;
b) For studies of policy measures relating to transport pricing or regulation, and for longer-range policy and major scheme appraisals, the requirements for local present day accuracy of the matrices may have to be balanced against the need to be able to segment these matrices across person type, trip purpose, etc. coupled with the need to include all-mode matrices.

Some small scheme studies (local area models), where rerouting is the main issue, will only need a present day O-D matrix in terms of vehicles for assignment since it is unlikely that a variable demand model will be required. For such studies, specific survey data will usually be collected with which to construct the O-D matrix with less emphasis on underlying traveller behaviour.

More typically, much larger study areas are involved (strategic, wide area, regional models), that require more involved methods. The DfT undertook a review of current matrix building practice in 2013 and found a variety of methods in application. There was a clear call for guidance from the survey respondents, some on specific technical issues and also on more general methodology.

The DfT review identified a number of notable issues of concern with the approaches most frequently described:

- There has been an over-reliance on RSI survey data, as if it is a truth (often referring to it as 'observed' rather than sampled and sometimes treating it as fixed data in spite of its inherent biases);
- Furthermore, it was noted that it is becoming increasingly difficult to collect such surveys and will become more problematic in future;
- Supplementary sources of data, such as the Census Journey to Work, National Travel Survey or mobile phone and GPS matrices, are sometimes used, in some cases without due consideration of their own definitional inconsistencies and biases;
- In spite of Guidance to the contrary, there was a reliance on matrix estimation to meet validation targets where the process seems to be largely uncontrolled while the targets are overly controlled;
- The use of the model for forecasting appears often secondary to the pursuit of base matrix validation.

It seems that many practitioners consider WebTAG validation against traffic count targets to be mandatory rather than, as intended, guidance. This report comments further on these issues and provides advice on appropriate methods in varying circumstances.

1.4   Structure of this Report

Base matrix development is a complex area, and we have taken the view that it is necessary to explain the background thoroughly. Accordingly, in Chapter 2, we begin with a general exposition of conventional transport models, giving a clear outline of the difference between P/A and O-D matrix forms, and explaining the different kinds of incremental forecasting approaches. In the course of doing this we also discuss validation issues. These explanations are important background to understanding the good practice advice that follows later.

Having set out the basic theory, Chapter 3 discusses the different forms of data that can be used for constructing matrices, distinguishing between those which are essentially zonal, those which relate to the movements between zones, and other elements such as traffic counts which can also provide information about movements. A key component of the data discussion relates to its reliability. At the end of the Chapter, we give a short assessment of current practice.

Chapter 4 then provides advice as to how appropriate base matrices can be built from available data, retaining the basic structure imposed by the underlying transport model, but modifying this according to the statistical reliability of the data. As noted at the outset, the requirements for the matrices will vary with the application.

Chapter 5 summarises key points of the advice, draws out the main implications for the Guidance that will be based on this report and also makes recommendations for research and development.

2   Background to Models and Base Matrices

2.1   Overview

In this Chapter we set out some of the principles of transport modelling, emphasising the important and fundamental role that base matrices contribute to the successful development of transport models. The objective is to highlight model assumptions and techniques that are relevant to later discussion of good practice in the development of base matrices. We also discuss how the current state of affairs, in which base matrices emerge from the process principally on an O-D basis, with significant emphasis on validation rather than forecasting, has come to pass. In particular, we explain the way in which O-D matrices are transformed generally through matrix estimation during validation such that the link back to the P/A form is lost, so that we can recommend what to do about it.

The construction of base P/A matrices, which in our view should represent the default approach to trip modelling, is significantly under-researched, and presents a number of challenges. What is required is a method which allows these matrices to be built using **all** relevant available data sources, while respecting their statistical properties and potential biases. Ideally, it is these P/A matrices, rather than O-D matrices, which should be used as the base for forecasts.

In the remainder of this chapter there follows a discussion of model structures and forecasting that sometimes goes well beyond the practice of constructing the base matrices. These are explained in some considerable detail since it is necessary to have a full appreciation of how the base matrices are to be used in order to appreciate how they should be built.

2.2   Traditional transport model matrices

The traditional four stage transport model contains both demand and supply elements, both being founded on generalised cost. The demand model estimates trip generations, their distribution and mode of transport given an estimate of travel costs for all possible journeys. The supply (network assignment) model answers the question (where congestion/crowding is an issue): what would the generalised cost be if the estimated demand were loaded on to the transport system? It also provides the routeing through the transport networks.

Trips or tours?

As WebTAG Unit M2 paragraph 2.5.1 notes, "A choice can be made by the model developers between trip-based and tour-based approaches." Trips (movements between two zones) are classified by mode and journey purpose: there is a further convention of distinguishing person trips between home based and non-home based purposes – HB/NHB. A tour is a series of linked trips returning to the original point of departure and tour based models offer the possibility of modelling trip chains, thus dealing with NHB (as well as other forms of tour). In practice, however, tour based models are mainly comprised of

primary destination, simple binary (outward and return) paired trips[3]. From diary surveys[4] (e.g. London Area Travel Survey, National Travel Survey) it is known that at least 70% of all tours are binary, and for some trip purposes the percentage is higher. Of those tours which do contain NHB trips, the majority contain only one, so that they are of triangular form. Whether working with tours or trips, the treatment of NHB movements remains a weakness for the model. Based on analysis of the National Travel Survey, NHB trips are about 15% of all trips.

A further claimed advantage of tour modelling is that modal constraints can be handled properly – this is primarily an issue of accounting for car movements (if the car is moved from the home, it must usually be returned; in addition, few car NHB trips are made unless a car has been used for the prior leg). But this is also dealt with in trip based modelling of binary tours as long as it is on a P/A basis, where car availability relates to the home (production) end.

On this basis, practical tour models can be looked upon as a minor variant, and in what follows we describe conventional trip based models. That said, a better understanding of tours/trip chains and, with them, better NHB and serve passenger[5] modelling would be desirable. This is particularly the case when considering less conventional household activity models. For now, we will simply say that these modelling issues are fruitful areas for research.

Modelling of trips

Demand modelling on a **synthetic trip basis** for HB purposes thus begins at the home end and conventionally with an estimate of the number of trip productions by purpose (and possibly further segmentations). As a default these are derived from TEMPro (or, if more detail is required, by direct use of the NTEM methodology[6] which underlies the TEMPro forecasts). Thus the productions are given by $T_i^{h(s)}$, where i is the production zone, h is the home based purpose, and s represents further possible segmentations (e.g. car ownership). Typically these trips relate to the whole weekday; weekend travel is normally excluded.

The next stage of the modelling is to allocate the trip productions to attraction zones, and to deal with mode choice (except where we have a unimodal model). These two processes can usefully be considered in one step[7] (Destination and

---

[3] Although most tour-based models do attempt to model secondary and even higher order destinations.

[4] One of the issues with tour modelling is the need to process diary surveys in some depth, so that successive trips are linked. This is a task not available in standard survey software, so the details of tours are only known when special analysis has been carried out.

[5] Serve passenger trips – a particular type of tour or trip chain, which is awkard to model, is when a driver drops off a passenger (perhaps a schoolchild) and then returns home or goes on to their own destination. There is a parallel in the latter case when a driver stops to buy a small item or fuel, for example en route to elsewhere. The complexity of these trips includes the fact that they might already be duplicated by the passenger's trip or the driver's longer trip in the matrices. In general, if local data indicates that such trips are not a major consideration, it is recommended that the model should not be complicated further by their additional inclusion.

[6] This is described in Section 2 of TAG Supplementary Guidance: NTEM Sub-Models. See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/370447/TAG_Supplementary_-_NTEM_Sub-Models_November2014.pdf.

[7] Although the way the two choices are modelled is often performed hierarchically.

Mode choice – DMS), so that we can write $p_{mj|i}^{h(s)}$, where m is the mode, j the attraction zone, and p indicates the proportionate allocation. Advice on how to estimate the DMS model is given in Chapter 4: the allocation p will be a mathematical function of the generalised cost matrix, which we write as **C**.

This stage therefore produces a (modal) matrix:

$$T_{ijm}^{h(s)} = T_i^{h(s)} \cdot p_{mj|i}^{h(s)}(\mathbf{C}) \qquad (2.1)$$

which is in Production/Attraction (P/A) form. We illustrate the distinction between this and an O-D form in **Figure 1**.

Note that if a tour based model is used, the matrix $T_{ijm}^{h(s)}$ represents the number of tours produced in i with attraction ('primary destination') in j: in practice each tour will typically involve 2 trips, outbound from i to j, and return from j to i. The convention with trip based models is that, as illustrated in **Figure 1**, the (P/A) matrix holds 2 trips for each such movement both having production zone i and attraction zone j.

The full range of demand responses (in particular, trip generation and the modelling of destination choice) can only be modelled sensibly on a P/A (or tour) basis. However, when we consider the impact on the network, and move to the assignment, we transform to an O-D basis, because for network analysis it is irrelevant for any particular movement which zone is the production and which the attraction: the direction of travel is the issue.

To convert to O-D format, the all-day P/A matrix needs to be allocated between outward and return movements and transposed for the return movement.

In most cases, the assignment is on a period-specific basis (usually an hour) rather than for the whole day. This means that further factoring by time period (t) is required. While some models will include time period choice within the demand model, in which case this allocation will already have been done, typically the time period allocations are held fixed. Note that the time period factors are fairly general ones, usually applied as averages across the whole matrix and one would therefore expect those hourly outputs to be quite approximate in detail – more so than is implied by attempting to validate flows closely at that level. There is a case for these and other global factors in the models (e.g. car occupancy) to receive more attention in terms of how they can vary across a study area, with trip length, etc. but that is not an issue for this project. It does, however, raise an important question about the relevance of an intensive focus on detailed count validation.

Figure 1: Alternative Matrix Forms

There are two alternative formats for transport demand matrices, and the distinction is of major importance. The **Origin-Destination** format relates to trips starting in zone i and ending in zone j, and thus indicating the direction of travel:

		Destination zone			
		1	2	…	N
Origin zone	1				
	2				
	…				
	N				

By contrast, the **Production/Attraction** format relates to trips produced in zone i and attracted to zone j. An alternative formulation is the tour, which is a chain of linked trips beginning and ending at the zone of production. In most cases the zone of production is taken as the zone of residence, although some work-related journeys can be produced from the zone of workplace:

		Attraction zone			
		1	2	…	N
Production zone	1				
	2				
	…				
	N				

To see the difference, consider the following simple two zone example. Zone 1 has 10,000 residents who all work in zone 2, and zone 2 has 1,000 residents who all work in zone 1. Each person travels once to work and back in a day. The total daily volume of travel can thus be represented as:

		Attraction zone		
		1	2	Total
Production zone	1		20,000	
	2	2,000		
	Total			22,000

		Destination zone		
		1	2	Total
Origin zone	1		11,000	
	2	11,000		
	Total			22,000

O-D matrices when taken over a whole day tend to be **symmetric**. This is not true of P/A matrices. The totals are the same, but the distribution over cells is quite different. It is the P/A form that allows trip making to be linked to zonal information (e.g. land uses, demographics).

Hence the process of P/A to O-D conversion can be described as follows. Introduce a further argument (d) to indicate direction: $d = 1$ for outward and $d = 2$ for return. Apply direction and time of day factors π_t^{hd}: these will certainly vary by purpose and might also vary to some extent by sub-area zones. Then the corresponding period-specific O-D matrices are:

$$A_{ijmt}^{h(s)} = T_{ijm}^{h(s)} \cdot \pi_t^{h,d=1} + T_{jim}^{h(s)} \cdot \pi_t^{h,d=2} \qquad (2.2)$$

[note the transpose in the 2nd term].

For non-home based purposes (n), a simpler approach is commonly used. Origin and destination trip ends are again derived as a default from TEMPro (or, if more spatial detail is required, by direct use of the NTEM methodology which underlies the TEMPro forecasts). These trip ends are essentially attraction trip rates that relate to land use values such as jobs or floor-space in a zone. By means of an appropriate DMS model, these origins and destinations are linked on an O-D basis for the whole day, and it is then only necessary to allocate by time period. In some cases, a more sophisticated NHB approach has been used, with some linking between the trips, which may mean that the time allocation is achieved automatically.

In the case of the car mode, these person trip O-D matrices need to be further converted to vehicles, since this is the basis of highway assignment. Typically, this involves assumptions of average car occupancy, which will vary by purpose and possibly time period. As with time period factors mentioned above, these occupancy factors are usually averages across the matrix and will, again, give only approximate outputs. In cases where, by contrast, the mode choice includes a choice between car driver and car passenger, the O-D matrices for the car driver mode provide the required vehicle matrices directly.

There are also questions relating to the definition of 'car', which are particularly relevant for later comparisons with traffic counts. The issue here is that some data sources include private travel in light vans as person trips whereas vans are typically separated from cars in traffic counts. This is mentioned further in Chapter 4.

Assignment to network

Given the O-D matrices, we now consider assignment. Although assignment is treated as a single model stage, it in fact relates to a number of separate processes, which may be described as:

1) choice of route (or path) for each *i–j* combination;
2) aggregating *i–j* flows on the links of the chosen paths (loading);
3) dealing with supply-side effects (capacity restraint or public transport overcrowding) as a result of the volume of link[8] flows relative to capacity; and
4) obtaining the resulting cost for each *i–j* combination.

---

[8] For public transport assignment, a link should be interpreted as a section of a line between adjacent stations or stops.

In practice iteration of the above steps is required to achieve convergence.

It is not usually necessary to maintain the same demand model detail of segmentation for the assignment, and the combinations of purpose and other segmentations are typically aggregated into user classes (e.g. business trips), where the essential characteristic is the generalised cost formulation, and value of time in particular, since this can influence route choice. In addition, other user classes not within the compass of the demand model may be introduced to make up the total modal network demand (such as goods vehicles in the highway assignment).

Item (4) produces the required cost matrices **C** for further iterations of the demand model (note that these may need to be converted back appropriately to P/A form – see Chapter 4). However, item (2) also permits comparisons to be made with vehicle counts on the network (and, in principle, with corresponding counts on public transport networks). We discuss this further below.

Figure 2 indicates what might be described as the traditional form of the model, which we will henceforth refer to as Type 1-A, in which an **absolute (synthetic) demand** model is used, in P/A form, to produce the matrices. They are then converted to O-D format, and assigned directly to the network, and the costs are then taken from the network, converted back to P/A format as required, and the demand model is updated. The dark black arrows apply in all cases, but the dashed arrows may be replaced in some of the model type variants that we discuss below.

Figure 2: Model Structure Type 1-A - Traditional (4-stage) Model
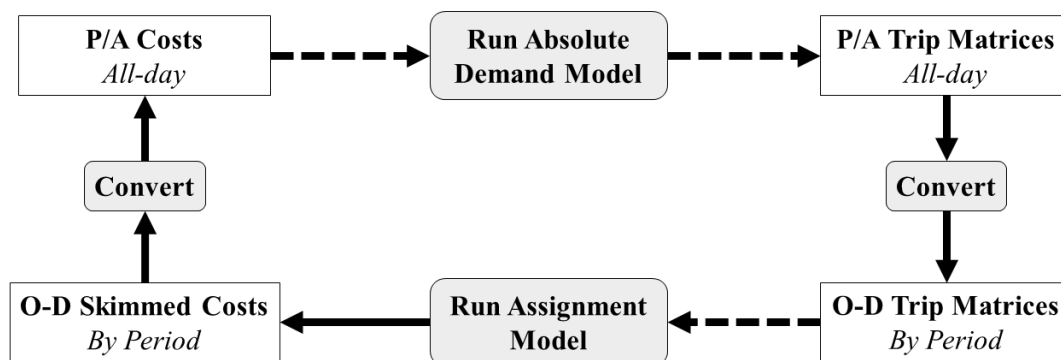
Note: While the flow chart is circular and suggests iteration, it can be viewed as starting in the top central box (Absolute Demand Model).

2.3 Validation and incremental modelling

The fit of these absolute synthetic models to the surveyed or observed base year data can often be quite poor, or particularly challenging and time consuming to calibrate satisfactorily. General experience is that, even when considerable effort has been expended in constructing the DMS model, the resulting synthetic[9]

---

[9] By this term we mean matrices which have been derived using a mathematical function of the network cost **C**.

matrices of travel that are assigned to a network do not give acceptable levels of link flows, compared with external evidence. For this reason, in line with WebTAG, it has become normal practice in application to make much more explicit use of so called 'observed' partial matrices (more correctly, matrices which are built by making substantial use of sample intercept surveys collecting the origins and destinations of trips). In contrast to absolute models, we discuss the application of common types of **incremental** models below.

Of course, it is always possible to improve the fit of the model to surveyed data by introducing arbitrary additional parameters. The aim is, however, to find a principled way of doing this that satisfies statistical rigour while still involving appropriate levels of detail. Whether the demand model is absolute or incremental in form, there will be a need to validate the base matrix to a suitable extent at the network level. In practice, this means that the conversion from P/A to O-D is carried out, and the resulting matrix assigned to the network. Then the assignment process is validated according to the procedures given in the existing Guidance on Highway Assignment (WebTAG M3.1).

At this stage, it is useful to make some remarks about the assignment procedure since this is a key element of model performance, yet often taken for granted when considering validation results. Although there are many variants on this, it is most common for highway assignment to use a User Equilibrium approach based on Wardrop's Principle[10], which may be described as follows:

With (O-D) demand represented by the matrix A_{ij}, the equilibrium loadings may be obtained by minimising the objective function:

$$Z(V) = \sum_a \int_0^{V_a} c_a(W)\, dW \qquad (2.3)$$

subject to the following conditions:

$$A_{ij} = \sum_r R_{rij} \qquad \text{(total demand)}$$

$$V_a = \sum_{i,j,r} R_{rij} \cdot \delta_{ra|ij} \qquad \text{(link flows)}$$

$$C_{rij} = \sum_a c_a \cdot \delta_{ra|ij} \qquad \text{(path costs)}$$

where r represents a path, a is a link, and i and j are zones. The (0,1) matrix $\delta_{ra|ij}$ denotes whether route r contains link a. R_{rij} is the volume of demand between i and j allocated to path r. c_a is the link cost, and is a function of the link flow V_a.

The outcome flows V_a on any link a can be compared with observed counts. Based on the above equations, we can write:

---

[10] The Wardrop condition implies that at equilibrium all used paths should have the same (minimum) cost.

$$V_a = \sum_{i,j} A_{ij} \cdot \varepsilon^*_{a|ij} \qquad (2.4)$$

where ε^* gives the optimum proportion[11] of A_{ij} allocated to link a. In practice, the reliability of ε^* depends on: a) the underlying theory of user equilibrium, b) the link cost functions, c) the coding of the network, and d) the level of demand as represented by A_{ij}, since this affects the level of congestion. In addition, we have simplified here for the purposes of presentation by omitting any reference to user class (which relates both to the demand matrix, the link costs, and in some cases differential access to the network).

Given all this, including the generality of the modelling processes, it would be unwise to expect a high degree of consistency between the outcome flows and the observed counts, even if there was a substantial level of confidence in both the demand matrices and the counts themselves. WebTAG Unit M3.1 recognises this in principle, and proposes three separate tasks (Network Data, Coding & Checking; Network Calibration & Validation; and Route Choice Calibration & Validation) with the aim of removing possible errors. However, these tasks are unlikely to impact much on either a) or b).

Problems occur when, after reasonable adjustments to the network, it is concluded that significant validation differences remain which are essentially attributable to the matrix. Ideally, further data should be introduced to the whole modelling procedure in such a way that the base P/A matrix is modified to be consistent with any changes to its O-D counterpart. Unfortunately, there is very little documented experience of how to do this.

An alternative approach, widely used, is to modify the assignment (O-D) matrix, using various techniques that can be referred to as 'matrix estimation from counts' (these are discussed in more detail in §4.5). At its most general, this consists of a model form for the (estimated) matrix $\tilde{A}_{ij}$, the parameters of which are estimated under a selection of constraints, which may relate to row and column totals, totals within different sub-matrices or total traffic flows across defined screenlines (thus introducing network elements relating to paths). Furthermore, they may range from 'hard' equality constraints, which must be satisfied, to 'soft' inequality constraints, which may be downgraded by assigning them lower weights or bounds (taken to reflect the modeller's estimate of confidence in them).

With the current state of practice, if this position is encountered, such that the base matrix is now in O-D form, the most common approach to forecasting is to use a 'pivot' version of the assignment model. Essentially, after converting the forecast output of the demand model from P/A to O-D, the resulting matrix is not directly assigned, but is compared with a base case matrix from the demand model also converted to O-D, and the implied changes are used to adjust the independently validated assignment matrix. While this can be expected to improve the

---

[11] These correspond to the PIJA factors in SATURN software terminology. It should be noted that while in Wardrop equilibrium the link flows are unique, the path flows are not. Hence the ε^* factors are just one possible solution out of an infinite number.

representation of congestion, there is a danger that it gives too much weight to the base year validation process.

If the O-D matrix is modified in this way to improve the quality of the assignment, P/A insight is lost and there will be an on-going discrepancy between the demand model and the assignment model. There is then no obvious way in which these adjustments can be conveyed back to the P/A-based demand model. Chapter 4 includes a proposed method to deal with this.

2.4 Different approaches to pivoting

There are thus two types of incremental modelling, a) on the demand side, on a P/A basis, and b) on the assignment side, on an O-D basis. These two incremental variants are entirely independent of each other, so that there are four possible combinations, which can be described according to **Table 1** (note the coloured numbers-letters for a later diagram).

In these types, the number (1 or 2) refers to whether the demand model is absolute (1) or incremental (pivot) (2), while the letter (A or B) refers to whether the assignment matrix is taken directly from the demand model (A) or merely uses the demand model to adjust an independent base assignment matrix (B).

Table 1: Incremental model types

Assignment→ Demand ↓	Matrix converted direct from demand P/A matrix	Independent matrix adjusted for predicted O-D change
Absolute	**1-A**	**1-B**
Pivot	**2-A**	**2-B**

Type 1-A is the traditional approach already described, with an absolute demand model and direct assignment (after conversion to O-D format), although this is now seldom used since, while it might represent a valid model of behaviour, it does not usually give a satisfactory level of base year validation.

Type 1-B takes the traditional (absolute) approach to the demand model, but uses the predicted changes (resulting from a scheme or policy) to modify an independently derived (and validated) O-D assignment matrix, as briefly discussed at the end of the last section. This adjustment could be done in a number of ways, proportionally, additively, or by a mixture of the two – discussed further below. This is used by a number of models (for example, the London Transport Studies model).

Type 2-A (in our view the generally preferred method) replaces the demand model by an incremental or pivot form, relative to a P/A base matrix, in a way which is described below. The resulting P/A matrix, after conversion to O-D, is directly assigned to the network.

Type 2-B is a combination of Types 1-B and 2-A, where, even after pivoting from a base matrix in the demand model, it is still considered necessary to make further

modifications to the resulting O-D matrices for assignment (for example, because the base pivot matrix fails to provide satisfactory network flows).

In **Figure 3**, these possibilities are sketched out. The red text and arrows indicate an incremental or marginal Type 2-A demand model where the model pivots off a base matrix (in P/A form) and is driven by the change between the scheme costs and the base costs. The blue text and arrows indicate an incremental or modified Type 1-B assignment process, in which the demand matrix, after conversion to O-D format, is compared with a base case, and the implied (O-D) changes are used to modify an independently validated assignment matrix.

Figure 3: Alternative Incremental Model Structure Types
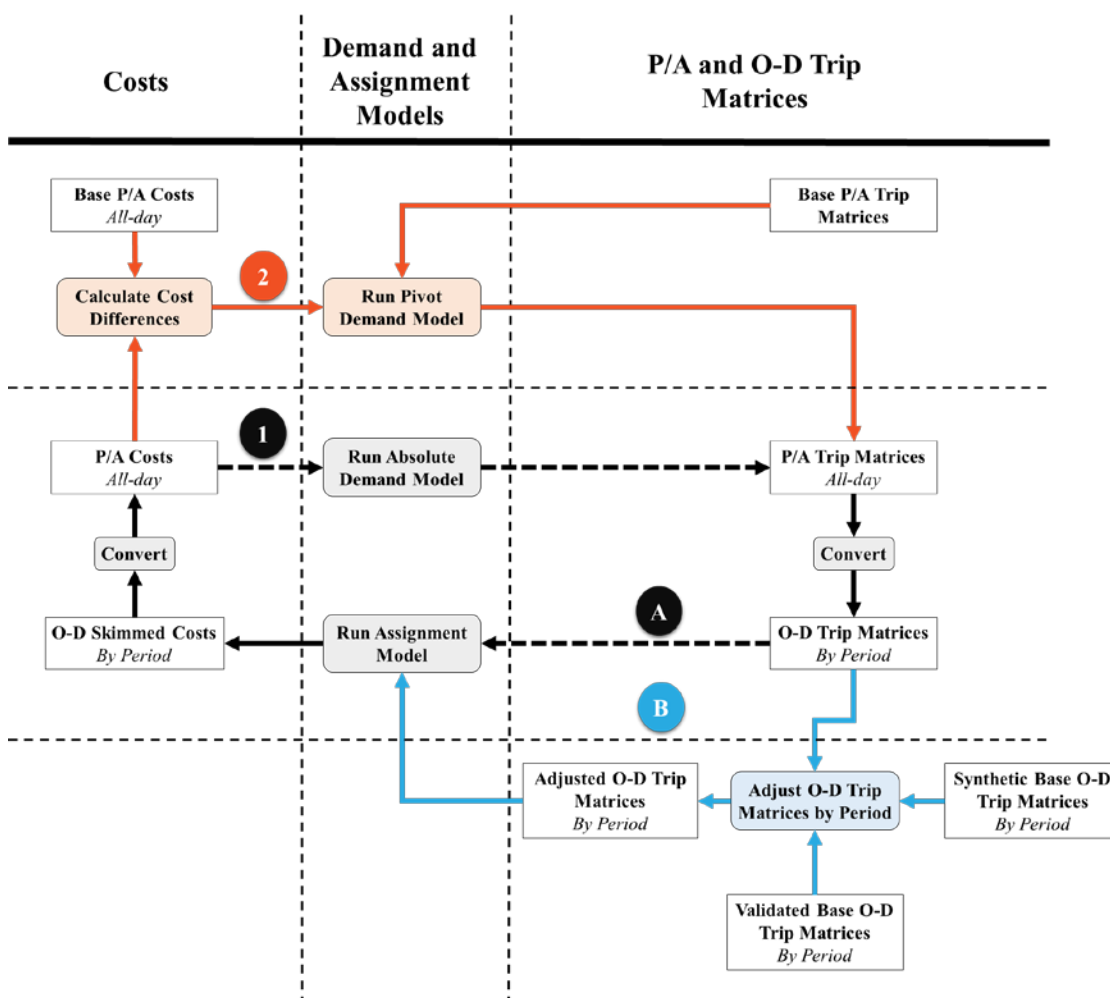
There has been some confusion about the terms incremental, pivot etc. A useful paper by Daly et al (2012)[12] describes various approaches, noting at the outset that, "The methods used for pivoting are diverse, sometimes giving substantially different results for the same inputs." They note that WebTAG distinguishes the following kinds of models [M2, paragraph 4.3.1]:

---

[12] Daly, A., Fox, J., Patruni, B. and Milthorpe, F. (2012) Pivoting in travel demand models, presented at ATRF, Perth, Western Australia

- absolute models, that use a direct estimate of the number of trips in each category;
- absolute models applied incrementally, that use absolute model estimates to apply changes to a base matrix; and
- pivot-point models, which use only the cost changes to estimate the changes in the number of trips from a base matrix.

Unfortunately, this does not distinguish between whether the pivoting relates to the demand (P/A) matrix or the assignment (O-D) matrix. As Daly *et al* note, the third method 'pivot-point' can only be applied validly on a P/A basis.

Demand-side pivoting

On the demand side, for Types 2-A and 2-B, there are two main possible options: method (i) adjusts the demand model so that it produces the base matrix $\mathbf{T}^0$ given an associated base cost matrix $\mathbf{C}^0$, while method (ii) continues to use an absolute model, but applies the predicted change to the independently derived base matrix $\mathbf{T}^0$. Method (i) thus falls into the *pivot-point* category, while method (ii) is an 'absolute model applied incrementally' [*AMAI*]. These two methods are referred to now and later by these Type codes. As we discuss below, under certain conditions the approaches give the same answer, but are not generally equivalent.

There are three ways in which the pivoting can be achieved for method (i), given an absolute functional specification for the demand model $\mathbf{T} = f(\mathbf{C})$. These are essentially equivalent mathematically, and therefore the choice is motivated by computational convenience:

a)   add constants $\{k\}$ to the model to ensure $\mathbf{T}^0 = f(\mathbf{C}^0,\{k\})$: maintain these in the specification;

b)   calculate additional cost matrix $\mathbf{Q}$ such that $\mathbf{T}^0 = f(\mathbf{C}^0 + \mathbf{Q})$: maintain the addition of these costs for all future runs (note: this is the 'residual disutility' specification of Williams & Beardwood, 1993);

c)   incorporate $\mathbf{T}^0$ directly into the model specification so that $\mathbf{T} = f(\mathbf{T}^0; \mathbf{C} - \mathbf{C}^0)$ (note: this is the incremental hierarchical logit specification of Bates, Ashley & Hyman (1987)) – see Appendix E of WebTAG Unit M2. This is pivot-point in the sense implied by WebTAG M2 paragraph 4.3.1, cited above.

The alternative method (ii) [AMAI] is used extensively in the models built by RAND Europe and described in the paper by Daly *et al* (2012). A prominent example of its application is in the PRISM model for the West Midlands. The basic idea is to run the unadjusted (absolute) demand models to give synthetic estimates $\mathbf{S}^0 = f(\mathbf{C}^0)$ and $\mathbf{S} = f(\mathbf{C})$. For each element of the matrix (*ij*, say), the ratio of the estimates S_{ij} / S_{ij}^0 is then multiplied by the corresponding base P/A matrix element T_{ij}^0. Hence the outcome result can be written as:

$$T_{ij} = T_{ij}^0 (S_{ij} / S_{ij}^0) \qquad \text{or equivalently} \qquad T_{ij} = S_{ij}(T_{ij}^0 / S_{ij}^0) \qquad (2.5)$$

As the authors note, this can give rise to problems in certain cases, according to the presence of zero or small values in the matrix elements. They therefore suggest recommended treatments for cases defined according to the presence of zeros in the cells T_{ij}^0, S_{ij} and S_{ij}^0. We discuss this further in Appendix A.

In current guidance, both methods (i) and (ii) are suggested as being appropriate. They are variants of the Type 2-A red arrow sequence in **Figure 3**. It is shown by Daly et al (2012: Appendix) that for a single level logit model (MNL), provided the modifications to the base values in Eq. (2.5) are 'normalised' so that the adjusted probabilities sum to 1, the two methods are mathematically identical, and the nature of the proof implies that this also applies to a nested logit [NL] structure, provided the normalisation is carried out at each level of the nest (thus ensuring once again that the probabilities add to 1 in each case). However, current guidance does not require this 'normalisation', so that the methods will not in general give the same result.

Assignment-side pivoting

On the assignment side (Types 1-B and 2-B), there is currently no **guidance** on how to implement the adjustment to the assignment matrices illustrated by the blue arrow sequence in **Figure 3**. As noted, the best approach would be to incorporate the adjustments in the demand model so that there is no need for the blue arrow sequence. This would maintain the desired consistency between the demand and assignment models. However, a method for achieving this is currently lacking, and is one of the issues addressed later in this report.

The need for this assignment adjustment arises when the base P/A matrix $\mathbf{T}^0$, after converting to O-D form and making any other required corrections (e.g. for time period, car occupancy) to obtain the corresponding $\mathbf{A}^0$, is found not to be compatible with the validated assignment base matrix ($\mathbf{B}^0$, say). If there is no way of further adjusting $\mathbf{T}^0$, how should a shift from the base matrix to $\mathbf{T}$ (with corresponding assignment matrix $\mathbf{A}$) be treated?

As Daly *et al* note, the AMAI method can be used in this case as well[13]. As with the demand-side variant, the unadjusted (absolute) demand models are run to give synthetic estimates $\mathbf{S}^0 = f(\mathbf{C}^0)$ and $\mathbf{S} = f(\mathbf{C})$. In this case, however, rather than applying the ratio to a P/A matrix, these matrices are converted to O-D form $\mathbf{A}^0$ and $\mathbf{A}$, so that the actual matrix $\mathbf{B}$ which is assigned has the default value:

$$B_{ij} = B_{ij}^0 (A_{ij} / A_{ij}^0) \qquad \text{or equivalently} \qquad B_{ij} = A_{ij} (B_{ij}^0 / A_{ij}^0) \qquad (2.6)$$

with the same eight cases treatments to allow for exceptional circumstances.

Discussion of pivot methods

As is well known, the sensitivity of discrete choice models is strongly affected by the shares of the various alternatives (modes, destinations). Hence, in principle, there is virtue in bringing the demand model as close as possible to the best

---

[13] Some modellers have modified the assignment matrix $\mathbf{B}^0$ by adding the **difference** (A_{ij}-A^0_{ij}).

estimate of the base position, and this is the philosophy behind the 2-A type model, which will be recommended for most applications.

Nonetheless, it is accepted that many current studies have proceeded to focus on the base O-D matrices $\mathbf{B}^0$, and thus the AMAI method can be viewed as a viable way of applying the results of an independent demand model to such base matrices that have emerged in O-D form after adjustment for counts validation. While the results of this type 1-B method will in general be different from the preferred 2-A approach, the closer that $\mathbf{A}^0$ and $\mathbf{B}^0$ are, the smaller will be the differences in practice.

An advantage of the AMAI method is that, by relating the forecast changes explicitly to the (validated) base assignment matrix, it can be expected to produce a) more credible estimates of flows and b) more credible estimates of generalised cost for the next round of the demand model. However, because **multiple** adjustment ratios (by direction and by time period) are being applied to the same underlying demand element $\tilde{T}_{ij,m=cd}$ (refer to Eq. (2.2)), there is no obvious way in which the adjustments can be conveyed to the demand model. This in turn means that the adjustment will not have any effect on model elements that are not subject to pivoting. It is similarly unclear what role normalisation can play in this case. The importance of this will depend on whether the model is primarily a highway model or is intended to represent all modes.

In the case of a multi-modal model, suppose that for a particular ij movement the highway matrix is substantially over predicted in all periods. This might suggest that the car mode split is being over predicted for this movement. While this will be corrected for the purposes of assignment, the likelihood is that the mode share for PT and active modes is being under predicted, and the AMAI method does nothing to correct this. Thus while the AMAI method has apparently the significant advantage that it has more general applicability (and in particular can be applied to the assignment matrices), it is not currently clear whether it can have any impact on the wider outcomes from the demand model.

The different types of pivots and their associated model types have different implications (and validity) for their use in **forecasting** with respect to base matrices, as we now go on to discuss.

2.5   Forecasting issues

While this project focuses on Base Matrices, the earlier DfT consultation and review noted that in practice the pursuit of base matrix validation often dominates the use of the model for forecasting. We noted in §1.2 that a suitable degree of validation is a necessary and important step that will improve the level of confidence in the forecasting model. Ultimately, however, it is the implications of the way the base matrices are used for forecasting that is critical, and this is the topic of this section.

It is useful to make a conceptual distinction between changes resulting from exogenous developments (broadly, land-use and socio-demographic changes) and those associated with changes in generalised cost. The guidance therefore

—

recommends the establishment of a 'reference case[14]' in which adjustments are made to the trip ends, usually on the basis of TEMPro forecasts (unless a local trip end model is available). On the assumption that generalised costs do not change, these changes can be implemented on the base year matrices by the Furness process (method of bi-proportional fitting).

Demand based pivot

If the demand model pivots off a base P/A matrix (types 2-A and 2-B), and method (i) [pivot-point] is being used, the pivot matrix needs to be adjusted for the forecast year. Note that if the pivot is being achieved by means of either approach a) [adding constants] or b) [residual disutility] in the previous section, it is not necessary to develop the reference matrix explicitly, as long as the model's generalised cost parameters (i.e. the weightings of the various components) do not change[15].

For pivot-point approach c), the adjustment to the P/A matrix should ideally be done at the full level of segmentation and spatial detail for the base matrix $\mathbf{T}^{0:h(s)}$. Thus the future year reference case (*f*) is defined as:

$$T_{ijm}^{0f:h(s)} = a_i^{f:h(s)} b_j^{f:h} T_{ijm}^{0:h(s)} \qquad (2.7)$$

where *a* and *b* are balancing factors designed to ensure that the estimated growth in productions and attractions is reflected. The resulting matrix is then used as the new pivot for the demand model in the future year.

The change in the pivot matrix is likely to have consequences on generalised cost via the supply model. In addition, there will be other changes (e.g. to the network, or because of changes to the components of generalised cost such as fares, parking and fuel prices) which will need to be reflected in modelling the Do-Minimum (DM) situation. Having obtained a converged DM, there will usually be some computational advantages in terms of faster convergence from using the DM as the pivot for subsequent Do-Something (DS) model runs since DS is closer to DM than it is to the reference case (see DIADEM Manual for some discussion). This will not affect the model outputs, however.

For method ii) [AMAI], there is no need to construct the reference matrix: the land-use changes affecting the trip ends and any changes to generalised cost can be implemented in one step, provided that the (future year) synthetic matrices S_{ij}^f are multiplied by the ratios T_{ij}^0 / S_{ij}^0 before converting to O-D form. In this case, there are more options for 'normalisation': for the two methods to be equivalent, it

---

[14] NB The terminology is not always consistent, as 'reference case' is also sometimes used to denote a Do-Minimum case. What is intended here is a hypothetical case where the generalised cost remains constant but the trip matrix is modified to reflect exogenous changes: this is consistent with the use in WebTAG M2 paragraph 2.5.5. The resulting trip matrix will not normally be a **realistic** outcome: it is merely an interim stage for the modelling of future years.

[15] For most models, the default assumption on the highway side will be of no change (though strictly the coefficients on time and distance should be modified in line with assumptions about the value of time). However, if they do change, then the reference matrix will need to be explicitly derived, and then approaches a) and b) re-applied to the reference matrix.

would be necessary for the normalisation to reflect the changes in the row and column totals as well as ensuring that the probabilities at each level sum to 1.

Assignment based pivot

If the assignment model is adjusted for a base O-D matrix (Types 1-B and 2-B), then one way to achieve this is by, after converting the (synthetic) matrix output from a demand model to O-D form, multiplying the resulting matrices A_{ij}^f by the ratios B_{ij}^0 / A_{ij}^0. More generally, the AMAI approach described in the previous section can also be used.

Note that TEMPro also provides growth factors on an O-D basis, raising the possibility of creating a reference case O-D matrix. However, this should only be done in the exceptional cases where it is considered appropriate to model the future on an O-D rather than P/A basis, as is discussed in Chapter 4. Any adjustment of demand in response to congestion cannot be done adequately on an O-D basis.

Discussion

There are, in our view, advantages in terms of clarity from distinguishing the exogenous changes from those associated with cost changes. There may also be special cases (greenfield sites) where particular attention is needed. Hence it is recommended that, whatever approach to pivoting is being used, initial matrices which assume no changes in cost be created, as a starting point in the forecasting process. As well as providing a good opportunity for checking the application of the land-use changes, this is a key step on the way to obtaining a properly converged Do-Minimum scenario.

2.6   A brief history of matrix building

This section (summarised briefly in Chapter 1) is not intended to be a definitive account, but more an impressionistic and informative description of how the current state has been reached so that this review project can be seen in context. The aim is to show how the approaches to requirements change and, with them, the role of base matrices. It appears here, since it relates to terminology and methods that have been explained already in this chapter.

In the earliest days of UK practice (for example, the Greater London Transport Plan model in the 1960s), reliance was placed on the conventional four-stage model, and trip matrices were essentially the product of the distribution model component. There was little emphasis on calibration/validation, more on derived mathematical descriptions of behaviour (a pure form of model), and application was mainly in the urban context. This continued through the 1970s with such models in all metropolitan areas. The Department was actively involved in steering and technical committees for these projects and sponsored considerable work on travel behaviour research.

In the latter part of the 1970s, the Department's role transformed to be predominantly in the field of interurban highway investment. This led to an increased emphasis on collecting traffic counts and roadside survey data and

diminished interest in the role of the demand model (other than for growth forecasting). It also heightened awareness of the often-encountered inadequacies of the distribution model within standard practice and hence led to the greater role for roadside interviews (RSIs). Ignoring the demand implications also allowed the attention to be focused on O-D rather than P/A matrices. RSIs were carried out in the vicinity of the scheme, and the unobserved parts of the matrix were infilled by a variety of methods (including the 'partial matrix' technique[16], until this was shown to be inappropriate in some cases and its use became discouraged).

This highways emphasis continued through the 1980s. At the same time, the increasing cost and administrative problems relating to RSIs gave rise to potentially promising methods of obtaining 'matrices from counts' (matrix estimation – ME). Emphasis was also placed on current year validation of highway matrices, and significant guidance appeared in the form of the Traffic Appraisal Manual – later incorporated within the Design Manual for Roads and Bridges (DMRB).

Recognition of the deficiencies of this uni-modal approach brought the demand issues back into scope (culminating in the 1994 SACTRA Report). The 1990s saw a renaissance of interest in urban models while, within highway appraisal, concern that ME could be used as a self-fulfilling process for satisfying the validation criteria, which were being increasingly strictly applied, led to advice to control its excessive application. The attempt to introduce a new emphasis on multi-modal studies in 1999, while not generally considered a success, led to the development of enhanced guidance (GOMMMS) and culminated in the WebTAG guidance over the course of the 2000s. Among other things, this suggested limits on the amount of change to the matrices which should be brought about by ME.

In addition, model developers historically faced a chicken and egg problem where they could not develop and calibrate good demand models without an input of reliable congested speed estimates for routes along highways but in turn could not develop reliable congested speed estimates without good car matrices output from the demand model. The consequent development of methods for the creation of base matrices that were not heavily dependent on the current quality of the synthetic demand matrix (relying on matrix estimation) was an understandable response to this conundrum.

However, the recent widespread availability of independently-sourced speed data provides an alternative to the awkward traditional method that required at an early stage of the model development process, high quality traffic flow estimates by link (plus good link capacity estimates) in order to ensure that the estimated congested speeds were realistic across the network. This change in approach could lessen network flow-cost dependencies and risk and could improve the quality of the overall model without requiring increased work.

---

[16] Partial matrix techniques are those in which gravity models are fitted to a partially observed matrix of trips and journey costs, and then used to **infer** the trips in the unobserved cells. See http://eprints.whiterose.ac.uk/2413/1/ITS315_WP111_uploadable.pdf . That paper suggests that the technique might be helpful if applied carefully, and further research might be fruitful.

There is now a general view that forecasts should be developed incrementally from good quality base year matrices. While WebTAG makes some remarks about the distinction between P/A and O-D, the emphasis in terms of matrix construction remains on uni-modal methods, and predominantly on highway matrices, where practice has tended to retain an O-D basis as an outcome of ME. For interurban public transport, the existence of the LENNON rail matrices (and the CAA air matrices) means that matrix construction is less of an issue (although the coach market presents a continuing problem). Within urban areas, the construction of matrices for public transport remains challenging. Only a few recent applications have made serious attempts to construct (all mode) P/A matrices and then carry those through to forecasts after base year O-D assignment adjustments.

Throughout this history, there has been relatively little attention to improving methods for freight[17] and goods vehicle matrices. These are a necessary component of traffic studies and of course are needed for highway assignment: separate attention outside this current project is needed to their development.

2.7   Concluding Remarks

In this chapter we have set out the basic theory of transport modelling as it relates to base matrices, making a clear distinction between O-D and P/A forms. We have mentioned that the traditional absolute model form (1-A) typically fails to produce base matrices which are compatible with current year validation data, particularly those that are more relevant on a network basis. This has led to the current official recommendations for forecasts that use incremental modelling.

This term, however, requires greater clarification, and we have introduced the distinction between pivoting on the P/A demand side and methods which use the demand model to adjust a validated O-D assignment matrix.

The chapter has also included a brief account of how the current situation has been reached, which serves as something of a justification for this current project.

> **A key principle is that the base matrices need to be fit for the purpose for which they will be used. In our view, the overriding perspective is that of the demand model. Given an incremental approach, there is thus a need for good P/A matrices to deal with a) exogenous (land-use) changes and b) changes in distribution/mode split resulting from cost changes.**

Having provided the relevant background, in the following two chapters we will first discuss in this context the available data sources and their properties, and then the ways in which they can be used to produce appropriate matrices.

---

[17] Note DfT's 2009-11 Base Year Freight Matrices project – see https://data.gov.uk/dataset/base-year-freight-matrices

3 Sources of Data

3.1 Introduction

Data is discussed in WebTAG Unit M1.2 Data Sources and Surveys.

In this chapter we provide comment and advice on the data that is relevant to creating base matrices, given the context presented earlier. The sources are covered under three headings:

- Zonal data – e.g. land use and demographic information
- Matrix data – mainly from surveys, plus some other sources
- Network-related – principally counts and journey times

But first, the next two sections mention the important consideration of data quality and certain principles of calibration and validation that impact on data usage.

3.2 Attention to error and bias

In order to construct base matrices, data is required that represents the demand for trip making. The quality of the matrix will be dependent on both the processes used to construct them and the quality and suitability of the data used. There are always constraints to the collection of suitable good quality data, and this problem is growing, so that there are understandable tendencies to use anything that might be considered relevant. With this trend, there are risks that matrix quality might suffer. This chapter discusses data sources with the intention to identify good practice and opportunities and to provide advice on their use but also to encourage practitioners to look at all data inputs with 'eyes open'.

Almost all data, collected or modelled, has bias and error. Bias in this sense refers to systematic differences between data estimates and true values, arising from data gathering, sampling, weighting (expansion) and processing methods. Error, by which we intend statistical error, is a measure of the estimated difference between the observed or calculated value of a quantity and its true value. Some data has errors in the more prosaic sense arising from mistakes or misjudgements. It is important to be aware of this and to attempt to identify or quantify the effects, particularly when using third party data or when including data from sample surveys, from models or from new types of resources.

Data is also not always what it appears to be. Definitional differences are a regular feature of data sources, sometimes arising from the nature of the data source and sometimes from item categorisations or forms of questions in data collection. It is important to look specifically at issues of compatibility.

There is more on specific types of data later but, in general, data from third party sources or from surveys should be treated with caution until its value is demonstrated. It would not be good practice, for example, to make use of a matrix from a commercial supplier or another study without understanding clearly how it was created, while recognising the purposes for which it was created. Similarly when using survey data, it has to be recognised that it often contains

significant bias, missing values, lumpy content from sampling, erroneous processing and the like. To refer to survey data as 'observed' when it can be some way from reliable truth, can tend to overstate its quality. This is not to deny the potential value of surveys, only to emphasise that the provenance and processes by which the data has been assembled need to be understood and examined critically. Essentially, this involves a sound appreciation of data issues and is a plea for thorough metadata to accompany the sources.

Good practice would require that all data inputs (and indeed modelling processes) should be subject to rigorous consideration of bias and error. To support this aim, it is recommended that all base matrix developments should have a quality statement regarding each input data source. A formal template should be provided in Guidance to allow users to describe:

- **The provenance of a data source – ownership and history**

- **Availability and transparency of reports or metadata to describe the data**

- **What is known and unknown about the collection or processing steps**

- **Definitional issues that might impact on compatibility with other sources**

- **Issues of bias that are known or suspected and how these have been treated**

- **Spatial and age inconsistencies of the data and their treatment**

- **Overview of suitability of the data and its impact on the model for its intended purpose**

Some examples of such issues are mentioned later where they apply specifically to certain data types. Once template entries have been completed for any specific data source, that statement can then remain connected to that source as descriptive metadata.

3.3   Calibration and Validation

Before discussing the various items of data, there is a point of principle which needs to be raised in respect of the use of such data. Section 4 in WebTAG 3.1 makes the distinction between calibration and validation. The basic principle that a model should be tested against independent data is, of course, worthy, but its relevance is debatable in the context of transport models. Transport modelling is not usually awash with data. In many cases, the data is so sparse that 'holding back' data simply makes the model less reliable.

There is some difference of opinion among modellers as to the extent that some data should be held back from model development and calibration in order to be used as an independent validation check. One school of thought is that models that can reproduce independent data well are desirable. Another view is that all valid data should go into producing the best model and that its main test of veracity is its ability to reproduce its (sometimes mutually inconsistent) source data. The authors of this report tend to the latter view. That is not to deny the

importance of model validation, but to suggest that there is a different way of viewing it.

Validation then relates to the process of how well you fit all of it, including understanding of typical inconsistency among all data. Independent tests would only apply when special data has been collected for that purpose or is available in a way that is not intended to be part of the model estimation and calibration process.

Independent validation is certainly sometimes a feature of statistical model estimation. It is most often thought of in fairly straightforward models, such as regression equations. In addition to looking at residuals to check the basic model fit, one can test performance against a separate dataset. A good, simple, example is to estimate a model using data for even years and then to test it on data for odd years. If it does not fit the odd years well, one should reduce confidence in or try to improve the model. One would not usually improve the model by then bringing in the separate odd-year data.

But transport models are far from conventional statistical models, although they contain separate sub-models that can be estimated in a conventional sense. Between those is a varied set of processes, factors, adjustments and approximations that we refer to collectively as a 'model'. It cannot be described completely as a mathematical entity, only some of its components. Its base matrices are also a blend of sub-models and processes, somewhat akin to a Bayesian approach of introducing added value to a prior model, but seldom with the statistical rigour that that would imply.

Current guidance contains considerable advice on base year model validation, particularly for highway flows. The focus of validation advice and professional contention, of course, has been on traffic counts. The emphasis of such advice is predicated upon the view that more accurate flows should, in principle, result in a good reproduction of network journey times. However, it should also be considered that, in order to influence network times, excessive application of matrix estimation might reproduce acceptable flows, but is likely to distort the matrices beyond being useful. We shall have more to say on this in Chapter 4. In addition, as we note elsewhere, the advent of quite comprehensive journey time information available from GPS and other technologies, makes these a better source of time data, as well as a further source of network calibration.

Nevertheless, a model that is considered to validate well will give some confidence, so it is proper that due attention is given to the guidance on validation. The narrative to accompany the validation report should make clear the purpose of the model and what levels of validation are considered appropriate. Where the levels appear to fall below guidance measures, the modellers should explain why this might be the case and the extent to which it is of concern.

Overall, the recommended perspective on validation is that one should report on how well the model relates to a whole range of mutually incompatible measures and then consider the results in the context of the objectives for the model. The objectives will only rarely be about use of the model to give good traffic flows in the base year, so the times when count validation is paramount should be few. In general, it is not recommended that some count data should be held back for independent validation. Thus, model calibration and validation can be considered to be a combined process, with reporting on validation still a key element.

3.4 Zoning systems and external zoning

In discussing the various items of data, we need to start briefly with the zoning system, since these are the essential spatial units by which most data can be referenced. Detailed zoning guidance is provided in WebTAG M3.1 and will not be repeated here.

In general, zone boundaries should take account of the need for internal screenlines for trip matrix building, calibration and validation. The zones are likely to be smallest in the area of detailed modelling, and it may be found that, in some key geographical areas of the model, there is even a need to disaggregate Census Output Areas (OAs). This should be done with reference to the network topology but also might relate to differences in trip-generating land uses and availability of data, perhaps from local planning authorities. It is also worth mentioning that, since the 2011 Census, specially created Business Output Areas have been available and these can be useful in areas of substantial employment, where population based OAs are often too large.

The expectation is that zones will be based on larger spatial units within the rest of the fully modelled area and progressively much larger for the external model area.

3.4.1 External zones

If the model is to include a synthetic element (in P/A form) it will be necessary for the external zones to extend to the whole of Great Britain as a closed system. This is because some of the internal productions, especially those near the periphery of the internal study area, will be satisfied by external attractions in the distribution model, and vice versa for external productions attracted into the study area.

The external zones will be more granular near the study area and less so at a distance. In all cases, they should be made up of official statistical areas (e.g. Census OAs, LSOAs and MSOAs[18]) nearby and administrative areas (e.g. Wards, Districts, Counties, Regions) further away so that demographic data will be readily available. The topology of the zones should also reflect how the networks that access the study area are configured, splitting external areas into separate zones when they will access the study area in differing ways. The external

---

[18] Lower and Mid-level Super Output Areas

network structure for the whole country should be consistent in density with the zoning.

There is comment in Chapter 4 on external-to-external cell movements that do not impact on the study area as possible through trips. If a model is entirely in O-D form, which should be rare, it will not be necessary to extend external zoning to the whole country, and cordon survey values can be used instead.

3.5   Zonal data - land-use/demographic

The data discussed here is essentially used on a zonal basis to support the trip end model and is therefore relevant to base matrices when a synthetic model is involved. The two main elements are population (and its arrangement into households) and employment (and sometimes floorspace) by various categories. The DfT's TEMPro methodology provides much relevant data for base and forecast years, although this will usually need to be disaggregated to local zoning systems and, sometimes, supplemented to support the desired model design.

Population forecasts take the most recently available decennial Census as a base and these are updated from time to time using Office for National Statistics mid-year estimates. According to the TAG Supplementary Guidance note[19]: "The demographic data in the TEMPro system is derived via the Scenario Generator. Further information on this can be obtained from the DfT web site." The figures are made available through the TEMPro program, typically for 5-yearly intervals.

The population and household information (together with some additional variables) are input to the DfT car ownership program (NATCOP), which segments the zonal households into different levels of car ownership (in practice, this includes light vans for private use – a complication for comparison with traffic counts, mentioned later). These in turn are then input, together with the zonal employment data, to the trip end model (NTEM), to generate zonal productions and attractions by purpose and mode[20]. By means of the conversion procedures within the NTEM program, these can also be translated into origins and destinations: trip ends for both matrix formats are also available by six time periods (representing both time of day and day of week[21]).

The maximum zonal definition is in the process of being updated to 2011 MSOAs (with the comparable geography in Scotland).

The current methodology produces eight home based purposes and seven non-home based, as well as six modes, as set out in the Supplementary Guidance note.

---

[19] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/370447/TAG_Supplementary_-_NTEM_Sub-Models_November2014.pdf
[18] The supplementary guidance note has further details on these models.
[21] The definitions are as follows:
    D1 Weekday AM peak period (0700 - 0959)
    D2 Weekday Inter peak period (1000 - 1559)
    D3 Weekday PM peak period (1600 - 1859)
    D4 Weekday Early or Late (0000 - 0659) and (1900 - 2359)
    D5 Saturdays (all times of day)
    D6 Sundays (all times of day)

Note that 'serve passenger' or escort trips, such as when a driver drops off a passenger (perhaps a schoolchild) and then returns home or goes on to their own destination, are coded according to the purpose of the person being escorted (this is done to preserve consistency of the attraction trip ends).

While TEMPro restricts further segmentation to different levels of household car ownership, the underlying methodology contains more detail by traveller type (for full information see Section 2.5 of the Supplementary Guidance mentioned above). In addition, information about income is available from the car ownership program NATCOP, although this is not currently used within NTEM. While it is likely that this level of detail will be (more than) adequate for most applications, users would need to gather further data if they want to model outside these constraints (for example, some models distinguish blue and white collar workers).

On the trip attraction side, total employment is subdivided into twelve categories, based on an earlier version of the Standard Industrial Classification [SIC92]. Any further proposed indicators (such as school places, large special generators, etc.) would need to be specifically collected.

It should be noted that the underlying TEMPro procedures have recently been updated.

In the base year, there are some other data sources for population and employment, which can be considered. Examples include the Annual Population Survey, Labour Force Survey, Annual Business Enquiry, Valuation Office Agency data and locally available land use data. All sources that are used should be mentioned in model reports.

3.6   Matrix data

A brief overview of sources of transport data that can be reconfigured in matrix form is given here.

Roadside Interview (RSI) Surveys – Traditionally RSI intercept surveys have been used as a primary source of highway matrix data. Although they only relate to the particular trip being intercepted, and are thus not well suited for models based on the concept of tours, they provide data suitable for both P/A and O-D matrices, since they normally collect the purpose at both ends of the trip. They can also provide information on vehicle occupancy (and, in principle, person type, though this is rarely collected). And if carried out in accordance with sound statistical survey principles they can provide relatively unbiased estimates of the longer distance movements which are more likely to be intercepted.

Unfortunately, there are a number of difficulties in collecting such data, and this increasingly limits its use as a foundation data source. As traffic and congestion levels rise it is becoming more difficult to find suitable sites where it is safe to conduct the surveys, without causing excessive congestion. Undertaking them requires agreement from the highway authority and the police, and such agreements are becoming more difficult to obtain. They are not suitable to be undertaken on links with high vehicular flows and they cannot be undertaken on motorway links, although they can be undertaken on on-slip roads. Consequently

it is particularly problematic to develop watertight cordons or screenlines in the desired locations. The sample of observed traffic compared to all traffic is generally not high (about 10%), often resulting in 'lumpy' outputs[22], with much lower sample sizes typically achieved for goods vehicles.

Household Interview (HI) Surveys – HI surveys provide the most complete picture of travel by residents within a study area, including walking and cycling. The travel data is usually collected by means of travel diaries, typically for a defined day. Outputs from these surveys can be segmented by the key variables of household type, person type, trip purpose, mode and time period, and provide essential information for synthetic models. HI surveys are normally carried out by trained interviewers, which makes them expensive, and likely to be confined to major urban areas, though with emerging survey methods using the Internet this could change.

However, building of meaningful trip matrices directly from household surveys is not generally practical due to the small sample sizes that are achievable with reasonable budgets: this applies particularly to the less frequent longer distance movements. Response bias in HI surveys is an important consideration, although there are established correction methods.

Census Journey to Work - For some time, transport modellers have used the Population Census Workplace tables and Mode of Journey to Work tables to supplement other information on the journey to work. Potentially, this is appealing, due to the large sample size – almost the complete population. But this data has often been used without close attention to the suitability of the data sources. Essentially, the problem is one of definitions. The Census asked two related questions, "What is your usual place of work?" and "What is your usual method of travel to that workplace?" It is easy to see that this will give results that can be quite different to that observed in a survey on any specific day or implied in a model of a typical day's travel. Many employees do not go to the same place every day or by the same mode. Some will be off ill or on holiday or absent for some other reason. The Census data is good data but tends to overstate a typical day's work trips and main modes, particularly in areas of employment concentration where most transport problems are focused.

The DfT sponsored development of a national set of matrices from the 2001 Census tables and later, when the definitional issues were pointed out, arranged for some factoring for the main inconsistencies. But it remains the case that this data should be used cautiously since it is not wholly compatible with the requirements of transport models to relate to movements in an average day. Furthermore, in the 2011 Census (with data already five years old as we write in 2016), there has been much greater disclosure control applied to the tables to deal with confidentiality issues and the information now has less content and spatial detail available than in the 2001 Census. It should be said, however, that the Census does give as good a land use estimate (10-yearly snapshot) of usual place of employment as is available.

---

[22] Lumpy infers that sampled observations may be concentrated in some zone pairs, leaving others that were observable empty. Treatment for lumpiness usually involves aggregating zones into logical units (sectors).

National Travel Survey – While this data is not coded at sufficient detail to provide reliable matrix data for movements between zones, it represents a more or less unique source of data at an aggregate level. By combining data appropriately over several years, it is possible to produce reliable and unbiased information of movements between (for example) standard economic regions, with some disaggregation by journey purpose and mode.

With travel data being based on a 7-day household diary from a continuing survey on a consistent basis since 2002[23], it would be useful if a standard dataset could be prepared for the purposes of dealing with high-level controls, especially for long-distance movements.

Automatic Number Plate Recognition (ANPR) and Bluetooth Data – ANPR and more latterly Bluetooth data can be used to develop cordon matrices for small areas. Typically ANPR cameras are positioned at entry and exit points of a network to form a watertight cordon. The resultant matched number plate records provide a matrix of movement between the camera locations. The time stamp of each record allows a detailed temporal distribution to be developed together with a record of the point to point journey time. In recent years survey companies have collected similar data by detecting the Bluetooth signal from devices within each vehicle. While this data requires less processing, not all vehicles contain a Bluetooth device, with the result that sample sizes are smaller than ANPR and the data will be biased towards typical users of Bluetooth enabled devices. The limitations with either type of such data are that:

- The data does not provide any information on journey purpose;
- The data does not provide the true origin and destination of each trip, as there is no way of knowing where the trip started or ended outside of the cordon. The modeller is therefore not able to relate this data to any zonal data. The modeller is also unable to ascertain whether the unmatched records either started or ended their journeys within the cordon, or are simply the result of errors inherent in the matching process.

This data would be most appropriate to use if building a model limited to a small number of junctions (typically up to around a dozen). While in principle it could be useful in validating a larger matrix in a particular area of interest, in conjunction with network (routeing) information, appropriate techniques have not yet been developed.

Mobile Network Data - Recently, a great deal of attention has been given to the potential of matrix data derived from mobile network data (and SatNav GPS) records, as captured by the mobile network operators. This is in the realm of 'big data', with vast volumes of information being captured, and represents an alluring data source.

The DfT commissioned the Transport Systems Catapult to undertake a parallel project to support guidance on use of mobile data. That guidance is intended to be a subsidiary element to the general base matrix guidance reflected in this report.

---

[23] The survey has been continuous since 1988, but there were some administrative changes in 2002 which make it preferable not to take detailed analysis further back.

The two project teams have liaised to seek a consistent approach, but it has been agreed that advice on mobile data will only appear in the Catapult report.

As indicated in Section 3.2, there are general guiding principles to be followed when considering a new data source, notably concerning bias, statistical error and definitional consistency. These are essentially the same as for any data source and should not be ignored in the development of matrices from mobile data.

Public Transport Data – As for highway data, public transport (PT) matrix information is required on a P/A basis. However, there is much less standardisation, guidance and consistency of practice for PT data and studies that have considered PT modelling have generally adopted a variety of approaches.

There are a few national data sources, particularly for rail travel. As an input to some PT models, mainly interurban, it is possible to obtain rail passenger matrices from the LENNON rail ticket information database, which is updated daily. As an initial stage, the demand estimate derived from the processing of LENNON data is a station-to-station trip matrix, with information by ticket type but not journey purpose. With some assumptions, it can be viewed as a P/A matrix source and estimates of true origins and destinations can be approximated. It is of more limited value in urban studies, where ticket information can be confused by season tickets, travel cards, etc.

The National Rail Travel Survey (NRTS) was a survey of passenger trips on the national rail system in Great Britain on weekdays outside school holidays. It was carried out for the London area as part of the London Area Travel Survey (LATS) in 2001 and throughout the rest of Great Britain in 2004 and 2005. In contrast to LENNON, this contains information on ultimate origins and destinations as well as stations used: it also contains information on journey purpose and ticket types, thus allowing a bridge to be constructed to the LENNON data. Unfortunately, significant changes have occurred in the ticket type mix since the survey was carried out and the data is now very old.

There is no nationally available matrix data on bus or coach movements. Although private sector operators may hold some relevant information, it is generally not released. At a more local level, some operators do make data available, for example TfL has provided Oyster ticket data for research purposes and its continuous RODS/BODS rail and bus passenger monitoring data is available.

For urban multi-modal studies, where PT is an important feature, local data is usually assembled or collected specially. Examples of such data include:

- Previously developed PT trip matrices – sometimes sourced from quite different models;

- Outputs from continuous monitoring interview surveys – examples are in Tyne and Wear and in London;

- On-board passenger interviews – administered interview or postal return;

- Passenger interviews at stops, station entrances or on platforms – administered interview or postal return. This approach was applied extensively for the PRISM model;

- Household interview surveys; sometimes rolling programmes, as in Tyne and Wear and in London – these are insufficient samples for matrix building but are very useful to verify general patterns and trip-making characteristics and to support adjustment as control totals (for internally produced trips);

- Ticket data – often very good for station-station matrices but only O-D data, no ultimate origin or destination information (some similarity to ANPR data for highways) and can be distorted by undetected ticket types.

There are examples of very good survey practice but PT interviews are notoriously difficult to achieve successfully. Notably, full coverage, sample control and expansion can be problematic. Without aiming to give advice on the design and conduct of PT surveys, we will just say here that the principles are similar to those that should apply to highway matrix data, such as:

- Consider carefully issues of coverage, definitions, representativeness and bias in all sources

- Seek verification data; adjust items or discard as appropriate

- Think about using the data in the model – is it in the right form and what does it contribute?

Because of the difficulties and expense of obtaining good PT matrix data, there is considerable hope that new sources, such as mobile phone tracking, will help to supplement or replace other surveys. Much work is on-going but the results so far have been rudimentary and there is much to be done.

3.7   Link Data - Counts

Public Transport Data. Analogous to the use of count data in highway matrix building, an important aspect of PT matrix development and validation is the use of surveyed link flow and station/stop count data. TAG unit M3.2 section 7 provides guidelines for public transport matrix validation checks. Typically these counts are used to inform matrix adjustment, even including variants on matrix estimation. In general, however, PT link-flow data are generally considered less reliable than highway count data – this is because data collection methods are often difficult to apply reliably and comprehensively.

Information on national rail passenger counts is available via the DfT Rail Statistics webpage: https://www.gov.uk/government/organisations/department-for-transport/series/rail-statistics. DfT also publishes annual PiXC statistics presenting passenger counts at London termini and other major rail centres as well as separate statistics on light rail and trams. The information on bus travel is much more limited, and usually confined to the number of passenger journeys on local bus services, classified by local authority (England only).

At a local level, some studies collect their own bus and rail passenger counts by various methods. These include:

- Station entry and exit counts

- Boarding and alighting counts

- Observation counts from outside the vehicles

- Ticket control counts (again, considering the nature of ticketing)

As with highway data, such counts will have survey error and should be considered in relation to their proposed usage. If used for survey expansion or for validation (sometimes on a cordon basis), it is important to question their reliability in relation to the emphasis on their value.

Traffic Count Data. Traffic count data is a key source which is used in survey expansion, matrix estimation and model validation. The Department has an extensive database of traffic count information available. The following geographical website, www.gov.uk/government/organisations/department-for-transport/series/road-traffic-statistics, provides Annual Average Daily Flow (AADF) and traffic data for every junction-to-junction link on the A-road and motorway network in Great Britain. Highways England also has a large amount of traffic data available, namely the Highways England Traffic Information System (HATRIS). Study-specific counts may also be collected by automatic means (Automatic Traffic Counts - ATCs) or manually (Manual Classified Counts - MCCs). While this project is not concerned with matrices for HGVs and LGVs, counts of car traffic will be required, and therefore the ability to classify vehicles appropriately is of relevance.

Traffic counts need to be related to the defined 'average' day for which the model matrices are constructed, so that ideally one would collect all traffic counts for a model in the same base year, for every day of the year. However, even with permanent ATC sites, problems will arise, and it is therefore necessary to define the level of statistical error (as well as correcting for biases) in line with what is practically available and affordable.

If the highway model is defined as being for a 'typical' weekday then the most suitable value might be the average over days in school term-time, say. A more specific definition such as 'average working day in May 2015' might allow a better estimate of the true value where data was collected in that month, but makes use of any data not collected in May more problematic.

In practice, data may be available from several different years and different times of year, and from relatively short periods of collection of counts using methods that are subject to measurement and sampling error. Then various processing procedures might be used to mitigate these errors and to control for known variability. For example, scaling data to the base year may be the 'best practice' way to mitigate the mean error, but it is likely to increase the uncertainty, i.e. it may increase the variance of the error. We need to quantify the likely error that is induced in each of the processing steps, in order to establish a confidence interval for the site counts.

Some kinds of errors will be unbiased random variations and will therefore tend to cancel out when averaged over larger time periods or large numbers of counts, while other kinds of errors may introduce systematic bias.

Some possible mechanisms are listed in **Table 2**. Any 'scale of effect' is an estimate by the authors and may depend on the site or the nature of the count. The remainder of this section discusses some of these in more detail, and some possible ways to measure them.

Table 2: Summary of Sources of Error and Sources of Variability in Traffic Counts

Issue	Description
Errors	
Measurement error in raw ATC counts	Probably small for ATCs in general, though individual sites can have malfunctions or missing data.
	However, some proportion of sites may be suffering from a bias or malfunction which has not been identified. The DMRB requirement is that 85% of sites meet certain criteria but the remainder have no requirement.
	At ATCs the classification by type of vehicle (e.g. light or heavy) is known to be subject to errors: both because individual machines on a specific route may inadvertently have been set up with different criteria to distinguish lights from heavies; and because for some types of vehicles they do not make the distinction in a form that is consistent with the vehicle classification used for MCCs or other forms of survey.
Measurement error in raw MCC counts	MCCs are known to suffer from larger measurement errors; even when scaled to match ATC totals, the results by vehicle class will retain larger error margins due both to small sample sizes and to inconsistencies in vehicle type classification.
Scaling counts from a different season or year	To account for differences in trends observed in count samples, and the effect that applying scaling factors will have on the resultant confidence interval of the adjusted count.
Variability	
Difference between ATC data average and 'true' average.	Need to estimate a confidence interval for the observed mean, especially where ATC is only temporary. See remarks below.

3.7.1   Scaling

Statistics will invariably show different trends in the rate of traffic decline/growth through recent years across different road types and in different areas and these trends further differ by vehicle type and may differ between different times of day.

There are some existing rules of thumb about how much this might increase the confidence intervals of the counts (e.g. in the ERICA manual) but currently DMRB validation standards do not make any allowance for this.

3.7.2   Errors in Classified counts

When an ATC is subdivided using a one-day MCC, for example when producing totals for expanding RSI surveys, there will be additional error introduced by the MCC. This is partly because the day-to-day variability of heavy goods vehicle traffic may be different from the variability of car traffic or light goods vehicles, and partly because MCC surveys are known to have non-trivial observation errors.

Again, these will be only partially mitigated by using a long-run ATC average to control the total number of vehicles.

Some ATC counters can identify heavy vehicles (often based on vehicle length) but this introduces a different classification not easily reconciled with MCC surveys.

In general, MCC counts are only done for one day per site because of cost. This means that any associated error estimate would need to be based on general rules, e.g. heavy vehicles having a confidence interval of X% of total flow. If there are any examples of multi-day counts (e.g. as for the turning count analysis in the QUORAM report[24]) these could be considered directly, otherwise wider literature about MCC counts might help.

3.7.3   Blending Count Data

TAG M1-2 Section 3.3.32 describes the errors associated with different count types. However, factors are often required to convert count data to a common base. For example, the factors required to convert to a neutral month and to the model year can be calculated from the available data, however each factor will have an error associated with it. A worked example is shown in **Table 3**.

Table 3: Worked Example of Count Error Estimation

A 16 hour manual classified count has been undertaken to estimate the 24 hour flow of cars on the day of the count. The 16 hour MCC gave 20000 vehicles. What is the 95% confidence interval of the 24 hour estimate?

From TAG Unit M1.2, the 95% confidence interval is given as $\pm$ 10% for cars, (i.e. the error is 10% of 20000). The variance of the estimate of flow V(Q,) of 20000 vehicles is given by.

$$V(x) = \left(\frac{E}{1.96}\right)^2 = (2000/1.96)^2 = 1020^2$$

Suppose the factor for expansion of a 16 hour to a 24 hour count is 1.08 and the coefficient of variation is 2.8%.

The estimate of 24 hour flow is thus:

$$20000 \times 1.08 = 21600$$

The variance of a product of variables is given by

$$V(x_1\,x_2) = V(x_1)V(x_2) + \overline{x_1}V(x_2) + \overline{x_2}V(x_1)$$

In this case;

$x_1 = 20000,$

$x_2 = 1.08,$

$V(x_1) = 1020^2,$

---

[24] QUORAM (2008). Quality of Regional Assignment Models, MVA report to TfL

$V(x_2) = (2.8\% \text{ of } 1.08)^2 = (0.03)^2$

$$V(x_1 \, x_2) = 1020^2 . \ 0.03^2 + 20000 . 0.03^2 + 1.08 . 1020^2 \ = 1574459$$

We now have E, the estimate of error range at 95% confidence.

$$E = \bar{x} \pm 1.96 \sqrt{v(x)} = 21600 \pm 1.96 \sqrt{1574459} = 21600 \pm 2459$$

The 24 hour flow estimate of cars on a given day from a 16 hour MCC is $21600 \pm 11.4\%$. The process can be further repeated to estimate AADT or AAWT etc.

3.7.4   Time periods and effects between time periods

There are likely to be differences of definition that can exist between the model and surveyed data, particularly with respect to the period of time covered. As a typical example, consider the standard one-hour static assignments from SATURN. In defining the appropriate matrix to use, it would be possible to classify according to the start time, the mid-time or the end time. Provided that the kinds of trips taking place at the start of the period are similar to the trips taking place at the end (i.e. steady state) this may not matter much. This might be a reasonable assumption in the inter-peak but seems more problematic in the peaks: the beginning and end of the AM peak and PM peak hours may neither be 'quiet' nor necessarily in 'steady state' with the peak average. A particular problem arises with long trips which cannot be completed within the defined period.

In the AM peak there may actually be a tidal movement from outer to inner areas so that the vehicles passing through the outer cordon in this period will disproportionately pass through inner or central cordons in a later period, and those passing through the central cordon in this period will have a high proportion of trips that started prior to the period. Since SATURN, for example, makes a distinction between 'actual' and 'demand' flow, to deal with the phenomenon of flow metering[25] in a static assignment, there is an issue as to how the traffic 'transferred' from an earlier period should be considered in the way in which the outputs from a model are compared to observed counts.

3.7.5   Cordon and Screenline totals

The models will have a number of enclosure cordons and barrier screenlines which are used in matrix building, calibration and validation. Since the total number of vehicles crossing a cordon will have less variability than those at individual links, both because any day-to-day route switching will tend to cancel out and because of the larger sample, the assignment matrices should, as a test of calibration, match fairly closely to the cordon totals, and DMRB properly includes more stringent requirements for cordon totals than for links.

---

[25] Some congested assignment models also incorporate procedures for estimating the effects of capacity restrictions on downstream traffic flows, sometimes referred to as **flow metering**. This is an important feature of many congested road networks, and failure to take it into account can lead to serious over-estimation of queues and delays at downstream junctions and poor estimation of overall network delays.

The questions are then:

- How do the various issues described above at link level affect model and surveyed counts at cordon/screenline level?
- What are the effects when the screenline is not watertight? In particular, what variability is introduced by omitting motorway links from the outer cordon?
- How should time period interactions across cordons, which are affected by the trip lengths, be dealt with?

These are all issues which need to be taken into account when considering the validation of the assignment model against counts and, together with generalisations implicit in the modelling, should influence reasonable expectations of the extent of validation matching that is achievable.

3.8   Link Data - Speeds

Conventionally, modellers have obtained sampled network speeds from special surveys, such as moving car observer runs or spot speed information from various sources. This has been used as input to models and to help validate speeds output from base year models through the iterative process of assignment and speed adjustment. Such methods have been awkward in practice and will not be discussed further here. Instead, this section concentrates on emerging data sources that hold more promise as a starting point and source for validation.

There is increasing availability of good remotely sensed link speed data (e.g. Trafficmaster, Inrix, TomTom, HERE/Navteq) that can be used to generate suitable time estimates for highway links and routes for the base year of a model.

For each observation the following attributes should be provided: link reference, date and time period, link length and vehicle type. The time period chosen should coincide with the month(s) for which other data has been collected. The link should be geo-referenced, for instance it may be based upon the Ordnance Survey Integrated Transport Network layer.

The journey time database will be large and will require manipulation using appropriate database tools. The links within the journey time database may need to be reconciled with the traffic model links. Software suppliers are now typically providing tools that allow models be developed using geo-referenced network databases. This allows consistency to be maintained between the journey time dataset and modelled network, aiding model calibration and validation. If the journey time data is to be used to generate complete skim matrices then use of the geo-referenced model network is not strictly necessary; skims can be calculated directly from the journey time link database using a routing algorithm.

For each link record it is anticipated that an average journey time will be provided from a number of observations. Information relating to the number of observations together with a measure of the spread of the data should be provided. Experience of such data has shown in many cases that the standard deviation associated with mean journey time calculated from these datasets can be extremely high. This is in part due to a 'long tail' associated with random events, etc. that cause significant delays for a relatively small number of observations. This has the potential to

increase the mean journey times and hence may cause an issue in model usage as these 'outlying' journey times will not necessarily be reflected by the traffic counts. Consideration could be given to the use of median journey times if this was found to be a significant issue.

3.9 Current highway matrix building practice

In what follows, we give an account of our assessment of what appears to be reasonably common practice, generally following the advice in DMRB. This is not to imply that the practice is universal, but many of the steps are found in a number of studies. In the course of the description, we note some of our concerns.

In some cases, the matrices for a study are partly built from existing matrices, occasionally of unknown provenance, and are merely 'factored up', usually to new traffic counts at the old roadside interview sites or screenlines. More generally, a significant programme of new RSIs is carried out (although as noted, this is becoming more difficult), and in what follows we will assume this is the case.

Base year (O-D) trip matrices are usually assembled using the following sequence of procedures:

- using the RSI data to calculate the 'observed' movements of the trip matrix;
- 'infilling' the unobservable trip movements, either by using parts of another matrix, output from another model, or by gravity modelling;
- combining the observed and infilled parts of the trip matrix, together with any other matrix manipulations required to obtain O-D matrices for assignment; and
- matrix updating using factors or count data.

While the earlier steps can be carried out separately for different trip purposes and/or vehicle types, as well as direction reflecting the difference between P/A and O-D, matrix updating based on count data can only be applied to vehicle types.

3.9.1 Developing Matrices from RSI Data

Developing matrices from RSI data requires a significant amount of processing. Separate trip matrices will be required for the different time periods and trip purposes that are to be represented in both the demand and assignment processes in the model. In what follows, it is assumed that the general guidance in DMRB on the design and appropriate sampling rates has been followed.

Trip matrices for individual time periods are constructed directly from origin to destination data so that the characteristics of each time period are as realistic as possible, and model periods should be as long as possible and include the whole of each peak period (not just the peak hour).

Matrices should be built independently for individual trip purposes. All home based trip matrices should be built in 'from-home' and 'to-home' format so that the P/A information is retained. Whether the matrices are to be built for the full

time periods, or to be built as peak hour matrices depends on the context. In large urban areas, the long distance trips to the city centre often arrive early to avoid facing the peak congestion for the whole duration of their journey, whereas the short trips often start late as they can live with peak congestion for a short distance. This implies that the shape of the matrix of vehicles crossing the London Orbital M25, say, from 07:00 to 07:30 could be very different to that crossing from 08:30 to 09:00. Data should not be pooled across the period in such contexts. Purpose splits, and their associated trip lengths also differ by micro time – few school trips occur after 09:00.

The basic process to develop matrices from RSI data is given in **Figure 4** below.

Figure 4: Building Matrices from RSI Surveys
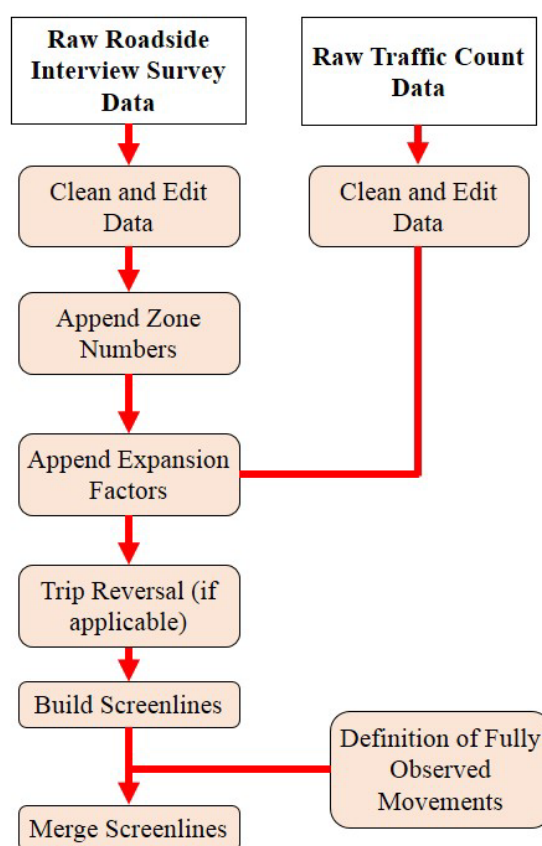
It will be necessary in the cleaning and editing to remove/edit interviews with key data missing or incorrectly recorded. Standard range checks should also be carried out on the survey responses (e.g. vehicle occupancy, identification of home-to-home trips) and basic checks should be performed on the reasonableness of address data. Origin, destination and survey site grid references can be used to check that a trip with this origin and destination was likely to have passed through the survey site. The need to correct errors of these types is fairly common.

The origin and destination zone should then be appended to each record. Where precise addresses are not known and only a partial postcode is provided, a method of random allocation can be used to give every valid record at least a full unit postcode.

Checks should be made for the completeness and miscoding of MCC data, and for consistency in ATC data collected at each RSI site. This assessment will highlight any anomalies in traffic flows encountered specific to the day of the RSI, which may either be as a consequence of a particular occurrence such as a road traffic accident, roadworks, adverse weather or an event, for instance. Alternatively, anomalies may occur because of the presence of the RSI itself, often through 'site avoidance' by drivers. Average weekday (Monday-Thursday) vehicle flows, by hour and direction, should be calculated from the ATC data for each interview site.

Expansion factors should be calculated as the ratio of counts to interview sample for each site, day and time period. Sometimes time periods have to be combined if it has not been possible to collect a good sample in all periods. Hence, the vehicle counts for expansion of the RSI sample data should ultimately be calculated as follows:

Count by hour and vehicle type = ATC for hour*[(hourly MCC by vehicle type)/(hourly MCC for all vehicles)]

Survey period expansion factors should be appended to each of the trip records. In the first instance interview records should be expanded to 60 minute counts. It is anticipated that records relating to trip-making by commercial vehicles should often require expansion to longer time periods (one hour or more) due to the usual smaller sample sizes.

The aim of this step is to build an expanded sampled 'observed' matrix which will only be partial because not all cell movements will have been intercepted. In this context it can be difficult to define what has been fully and partially observable. Only fully observable trips (i.e. those trips in any i-j pair which would be certain to pass through the screenline) should be added to the partial matrix. Partially observed trips, i.e. those i-j trips that could have passed through the screenline, or equally could have travelled on an alternative route around the screenline or cordon should not be included within the partial matrix.

In addition, care needs to be taken when dealing with trips that have been derived from multiple sources. Partial matrices will be derived by some averaging of duplicate or multiple recordings. Averaging should be weighted using variances (see Section 3.8.3) if there are major differences between sampling errors at different sites. The calculation of weighted variances should take into consideration the source of the data used, age, and reliability.

Because of the tendency to 'lumpiness' of the data, mentioned earlier, it is often necessary to group the RSI survey cells into zonal sectors for processing and control adjustments before disaggregating back to zones for the modelling.

A good appreciation of the errors and approximations that have been introduced through these, and other, processes should be developed and documented by the analysts. The quality of the results should contribute to a later assessment of the levels of model validation that can be expected.

3.9.2 Treatment of transposed trips

For practical reasons, RSIs are usually carried out in only one direction. Since, as noted in Chapter 2, the majority of 'tours' involve an outward and a return movement for the same purpose, this means that only one of the P/A movements will be intercepted. It is common, if questionable, practice to synthesise the characteristics of the reverse flow in the following manner.

First, the survey records are duplicated, and the origin and destination zones and trip purposes will be swapped to represent reversal. On the reversed trip records, AM peak period interviews are changed to PM peak period and vice versa. Inter-peak interviews are left unchanged, but 'from home' purposes are reversed to 'to home' and vice versa. The purpose for non-home based trips will be unchanged.

The allocation of time periods for the reverse trips is based on a two-stage process. First, the time is taken as a mirror image across the midpoint of the survey period (for example for a 12 hour survey starting at 07:00 a trip observed in the interview direction at 08:00 (second hour of the survey) would be given a time of 17:00 (second last hour of the survey) for the non-interview direction).

Secondly the expansion factor for each record is adjusted in line with the expected number of trips of that journey purpose in that time period by estimating journey purpose proportions for each time period based on records from other sites of a similar type (i.e. based on interviews for trips made in the same general direction). These proportions are applied to the traffic counts, thus enabling the expansion factors to be calculated at a journey purpose level.

While the sometimes questionable assumption has generally been that transposed trip records are a more reliable estimate of reverse direction movements than would be obtained from a synthetic source, there are two conflicting arguments;

- If the purpose of partial matrices is to control the synthetic matrices or to replace the synthetic matrices, then controls should only be applied using the trips which were truly surveyed.
- Statistical analysis of transposed data and surveyed data may well show that the transposed data is within the sampling error, if working on a P/A basis.

In terms of producing O-D matrices, there may be many trips in one direction that would not necessarily exist in the other direction. For example, experience of LGV and HGV trips shows there may be a strong asymmetry between the total trips in two directions in the 12 hour period. The problem is likely to be distance-dependent and not consistent throughout the matrix. Asymmetry also applies to NHB trips.

One can see that there are several dubious assumptions in the foregoing. It is therefore recommended that the modeller ought to be very cautious when dealing with transpositions, due to problems with the implied trip patterns, which may vary by trip purpose, by trip length and by area.

3.9.3   Screenline Matrices

Building screenline matrices is mainly a matter of combining matrices from individual RSI sites. Screenline matrices will be built separately for the interview and non-interview direction. This is because the statistical uncertainty is greater for the non-interview direction, which should be reflected in the matrix building process (as mentioned above).

Fully observable movements should be defined using a sector system and indicator matrices, as discussed in Section 3.8.1. In this context, 'fully observed' refers to sector to sector movements that must pass through an RSI screenline so all trips are observable (at least once). Partial matrices should then be developed containing only fully observable movements. This will mean that any ambiguously-observable O-D movements can be removed before the RSI matrices are combined with any other sources. For any RSI screenline, partial matrices can then be added to create screenline matrices.

TAM recommends that trip matrices, and the associated indices of dispersion, are merged in such a way as to minimise the coefficient of variation of the combined cell value. The theory is that all observations are valid but their accuracy depends upon the sample upon which they are based. Appendix B discusses the TAM formulae in more detail, and shows that under specified assumptions the merged trip estimate T_m is equivalent to the sum of the number of records divided by the sum of the sampling fractions, as shown in **Table 4** below.

Table 4: Worked example of merging trip estimates

> **The estimate of the number of HBW records between two zones fully observed in two screenlines can be merged as follows.**
>
> - At screenline 1: 7 HBW observations have been recorded out of 120 interviews. The traffic count reveals 1000 vehicles crossed the screenline.
> - At screenline 2: 6 HBW observations were made from 85 interviews, with 500 vehicles crossing the screenline.
>
> Using the equation above, the merged trip estimate is therefore
>
> $$Tm = \frac{7 + 6}{\frac{120}{1000} + \frac{85}{500}} = 44.8$$
>
> The total estimate of HBW tips between the two zones would be 44.8

To implement the above calculation, a method is needed to calculate the sample size for each cell (i.e. the number of trip records used to obtain the trip estimate). This can be obtained by running most of the matrix building process again with the expansion factor replaced by a 1.0. This yields a matrix of the number of records rather than the number of trips. Note that the number of records used in creating the separate interview direction and non-interview direction trip matrices is calculated.

Many cells have an observed zero. In these cases the variance formula cannot be used, as it results in a divide by zero. The variance of these cells could be based on the average for the matrix. This can be derived by calculating the average number of trips and the average number of records used in generating that trip

matrix and then using the standard variance formula. This should be done separately for each journey purpose/vehicle type and time period.

In order to prioritise the weighting given to interview direction RSI data, the sampling fractions of the non-interview direction data should be reduced (by possibly 0.5) within the merging process.

3.9.4   Infilling

A gravity model or similar modelling method is then used to produce the synthetic data to populate the cells that are unobservable by RSI data. Gravity models are typically developed on a 24hr, defined year, average weekday basis. The purposes to be considered are defined including home based and non-home based purposes.

Car mode trip ends can be derived from TEMPro with the TEMPro purposes suitably aggregated to the required purposes. Generalised costs can be obtained from model time and distance skims, or preferably now from journey time information available from GPS and other sources (see Section 3.8), combined with the latest WebTAG parameters for Value of Time (VoT) and Vehicle Operating Costs (VoC).

For purpose k, denote generalised costs for each zone pair ij as C_{ijk}, productions at zone i as O_{ik}, attractions at zone j as D_{jk}, and trips from zone i to j as T_{ijk}. For each purpose, define a deterrence function $f_k(C)$. Then T_{ijk} is given by:

$$T_{ijk} = A_{ik} B_{jk} O_{ik} D_{jk} f_k(C_{ijk}) \qquad (3.1)$$

where A_{ik} and B_{jk} are balancing factors calculated iteratively to ensure the production and attraction constraints.

For each purpose, the deterrence function can be determined by calibrating against the trip length distribution (TLD) derived from NTS (or locally available) data. NTS data is processed to give proportions of demand across a number of distance bands. Calibration is undertaken by systematically varying the parameters of the deterrence function to obtain the best fit using standard chi square goodness-of-fit statistics.

The model can further be improved by comparing the resultant data against other observations. For example, data is also available from NTS giving weekday region-to-region movements for car trips (car driver and car passenger). These are sampled observations, but can be expanded using population counts. Following TLD calibration, checks can be made on region-region movements.

The demand can also be calibrated to this data by standard multiplicative 'K-factors', separately for each purpose k:

$$T_{ijk} = K_{IJk} A_{ik} B_{jk} O_{ik} D_{jk} f_k(C_{ijk}) \qquad (3.2)$$

for all zones i and j in regions I and J.

The estimation of the K-factors is undertaken iteratively: at each iteration, the worst fitting region-region movement is corrected by calculating the relevant K-factor as the ratio of required demand to the modelled demand. This process can be repeated until a pre-determined overall level of goodness-of-fit is achieved. Following calibration of the region-region movements, a final check is made to ensure that the relevant deterrence function parameters remain the optimal choice.

3.9.5   Current ME practice

Matrix Estimation is currently used to ensure that models achieve the link and journey time validation guidelines set out in WebTAG M3.1. The changes brought about to the matrix by ME are usually reported in terms of:

- Changes to zonal cell values
- Changes to matrix trip ends
- Changes to trip length distributions
- Changes to sector-to-sector movements.

WebTAG M3.1 places severe restrictions on the application of Matrix Estimation through limiting levels of change in the above, detailed in Table 5 of the guidance. Experience suggests that many modellers tend to prioritise good link and journey time validation at the expense of distorting the matrix beyond the limits set.

Various elements of current guidance on matrix estimation need to be revisited. There is further discussion of matrix estimation in Chapter 4 in the context of recommended methods that place traffic counts in a more holistic matrix development process rather than using ME to 'force' a fit to counts.

3.10   Developing PT matrices

It was agreed with the DfT that the advice in this report should focus on matrices for public transport in a multi-modal context (principally in urban and regional studies and related to the form of guidance in WebTAG).

Some guidance for the calibration and validation of a Public Transport Assignment Model is available in WebTAG unit 3.11.2 (M3.2). This is mentioned as a setting for the text that follows and provides criteria for three types of validation:

- Validation of the trip matrix;
- Network and service validation; and
- Assignment validation.

Validation of the trip matrix involves comparing modelled flow and observed count values across complete screenlines. The criterion states that in 95% of cases modelled flows should be within 15% of 'observed' [26] counts. Network and service validation refers to checks completed on the link geometry and comparisons of modelled and observed values on individual services. Since there are no specific guidelines on these checks, best practice is sometimes followed by

---

[26] Observed is a fairly strong word for the often-approximate way that counts are usually derived.

checking and validating coding and modelled journey times against those timetabled. Assignment validation refers to comparison of flows on links on screenlines with 'observed' data and also comparison of boardings and alightings in urban centres. In this case, it is advised that flows across modelled screenlines should be within 15% of counts and, on individual links within the network, flows should be within 25% of counts except where the screenline counts are below 150 during the assignment period.

It is also relevant to consider modelled PT matrices against independent sources, such as published statistics or revenue information. In principle the P/A versus O-D conversion, time-of-day profiles, segmentation and journey purpose disaggregations could all be informed by TEMPro data. However there are known issues with using TEMPro at such a local area as stated in TAG unit 1.2. paragraph 2.2.5: "*Users should note, however, that TEMPro trip ends by mode are based on average rates over a wide area and do not necessarily take into account the accessibility of each zone by each transport mode.*"

With reference to the earlier sections on PT matrix and link count data, the development of PT matrices can follow a similar pattern to that described for highway matrices. Thus, a synthetic all-modes model can be used as a starting point with involvement of trip matrices from previous studies, matrices based on interview data and count information. As stated, it is essential that all sources are clearly understood and described to ensure their proper incorporation and weighting. The proposed methodology that is described later can then be applied to build PT matrices in a multi-modal setting. Care is needed to obtain reliable data to allow for modal split (given the caveat about TEMPro trip ends by mode). This will usually require special attention to validation processes to adjust the modelling to achieve the required modal balance.

4 Building and Using Base Matrices

4.1 General requirements and background

We begin by recalling some of the background discussed in the previous two chapters. In most cases, base matrices will have a role to play in forecasting. Where a non-uniform growth is forecast at either the production (home) end or the attraction end, or where changes in network costs will change travel behaviour differentially by home based segments of the demand, forecasts produced using O-D form will be less suitable than those produced using P/A based matrices. For this reason P/A matrices should normally be used, even if no explicit trip distribution modelling is performed. There may, however, be circumstances in which it can be satisfactory to use O-D based matrices for forecasting, where the model is simple enough or used for a sufficiently specific purpose that the analyst can be confident that the forecasts will not be unduly biased.

While it is necessary to carry out demand modelling in P/A form, it is also necessary to convert to O-D for the purpose of assignment. The assignment model splits the trips according to the route they take through a network, and then calculates the cost of travelling via each route. The cost calculations are needed not only for the assignment model itself but also as feedback for the demand model. Therefore, the performance of the assignment model is important, and this includes whether the assignment of the base year matrices is realistic. Of particular interest is whether the predicted link[27] flows are in agreement with available observations. To the extent that they are not, consideration must be given as to what useful information this implies for the development or improvement of the matrices.

The involvement of sample survey and traffic count data in the model development is an important issue. In relation to a greater emphasis in this report on the use of synthetic modelling, it must nevertheless be appreciated that not all real-world travellers make rational decisions based on generalised cost considerations. There are often local circumstances that account for traffic flows that cannot be captured through the generalities of modelling based on simple generalised cost with all the average assumptions and approximations that are involved. It is therefore necessary to involve the traffic flows in the model and to be concerned about how well the resultant model reproduces flows. It is, however, equally important to recognise the weaknesses of traffic count data, not to expect an unrealistic correspondence between the model and counts and not to allow attention to flows to just replicate a current pattern of travel and overwhelm the predictive quality of the model.

As we discuss further in section 4.5, the use of matrix estimation (ME) techniques has led to considerable concern. Because ME can be used inappropriately to allow some of the suggested validation criteria to be satisfied, the current Guidance, while not prohibiting use of ME, places severe restrictions on its potential impact. But what should be done is to allow the observed flows to

---

[27] As in Chapter 2, for public transport assignment, a link should be interpreted as a section of a line between adjacent stations or stops.

modify the matrix having regard to the statistical information that they contain. This alternative approach may lead to more matrix adjustment than is currently advised, while not necessarily achieving tight agreement with the counts that broadly represent observed flows. We note that those flows are sometimes, themselves, inadequate measures or mutually inconsistent for a variety of reasons.

In principle, any matrix adjustments made along these lines need to be conveyed back to the underlying P/A base matrices, as they are potentially important for the demand model (and for destination choice in particular). A desirable aim is to have compatible P/A and O-D matrices in line with the Type 2-A structure discussed in Chapter 2. Nonetheless, this may not be needed in all cases, and some compromises may be appropriate where it is difficult in practice to achieve complete compatibility. We discuss these special cases in Section 4.6.

A further point is that even if the current recommended validation criteria are viewed in a more general context in order to achieve a better behavioural demand model, there may be technical advantages in ensuring that, within the assignment, the base year flows do agree reasonably with the best estimate from reliable count data. This is particularly because of the role of the counts in representing congestion and crowding, via the relevant link cost functions. This could be achieved by further application of ME techniques, in this case viewed only as a device to improve the assignment model, with the implication that any such changes to the matrix would **not** need to be conveyed to the P/A matrix but that the resulting speeds (and hence costs) will be more realistic. Essentially, this would constitute a Type 2-B structure. There are other sources of network speeds available, and this point is mentioned elsewhere.

Ideally, given an appropriate model form based on theory, the parameters would be populated by a holistic estimation process which takes account of all the available data, with due recognition for any inconsistencies, biases and the statistical accuracy of the different sources. While some attempts have been made along these lines (e.g. the EU OPUS project (2007) and Lindveld (2006) whose work has been used for the Dutch National Model), they involve both theoretical and computation effort which is currently too advanced for most cases of standard practice. There is therefore a need to develop approaches which are practical but which nevertheless maintain some of the necessary rigour.

In line with this discussion, this chapter will proceed by first setting out an approach for the construction of synthetic P/A matrices (as well as the associated NHB matrices – on an O-D basis) for the demand model, based on conventional TEMPro data and other national datasets. It then goes on to consider how these synthetic matrices could be improved by comparison with other matrix based data information – including RSIs, Census Journey to Work data and potentially mobile phone data. An important principle is that, as far as possible, all data sources need to be brought to consistent definitions and date and corrected for intrinsic errors and bias in advance. Then, given the best estimate of the P/A matrix available from all such data sources, the final step is to address the issues associated with assignment and link flows.

The chapter is generally similar in outlook to Appendix B of WebTAG M2, which "contains one approach which could be adopted" [paragraph 2.5.10], and some of

the text is used here. However, our recommendation will be that the current Appendix should be removed, and that a more detailed unit specifically associated with Base Matrix Construction be prepared, based on the discussion in this chapter.

Having set out a recommended procedure for the most general (multi-modal) case, the later sections of this chapter will discuss how the procedure could be modified to deal with particular, usually simpler, cases.

Case studies involving recommended practice

The following case study examples illustrate some principles of general application of the practice that is reflected in the advice in this chapter.

Table 5: Case Study Examples

---

1. Tyne and Wear Multi-modal Model, 2007

The client required a versatile model that would validate well but one that would prove relatively sensitive in forecast years. A great deal of survey data was available, including a rolling household interview programme, many RSIs and traffic counts (over different years), continuous bus and Metro surveys and journey time surveys.

As well as adopting best current practice in the processing of the survey data, the study adopted a novel approach to development of base matrices. Essentially, the process began with as good a synthetic multi-modal model as could be developed and the matrix cell values were then improved where survey (both RSI and public transport intercept) data suggested a statistically significant better quantity (rather than starting, more typically, with the surveys and infilling synthetically). An iterative process followed, including matrix estimation for car traffic. The study derived a new mathematical process that would allow adjusted O-D results to feed back to the P/A matrices. The trip rates, with their connection to land use data, were taken as a fix in the iterative process, as strong as the traffic counts, and this gave both a good validation but also an outcome for forecasting that remained in P/A form.

There are important lessons from that experience that were used in the next two examples and that are explained in this report.

2. M6 Highway Model, 2008

The aim of this project was to undertake a strategic assessment and appraisal of improvement schemes to the M6 motorway in the Birmingham to Manchester corridor. Following the Tyne and Wear procedure (above), a multi-modal synthetic model was prepared, and representative base year highway demand matrices were required to ensure model validation.

Due to the size and nature of the model area it was impracticable to obtain survey data which could be fully representative. RSIs were collected on motorway on-slips, together with ATCs and MCCs, as it was not possible to draw effective survey cordons. From these surveys, partially observed matrices were built using conventional best practice.

As with Tyne and Wear, the partially 'observed' survey data was then allowed to adjust the synthetic matrix as a function of the quantified precision of the survey data. The proportional change was then applied back to the detailed zone-based matrices and the new 24-hour P/A matrix was then re-constrained to match TEMPro trip ends.

The output from this process was thus a complete matrix consistent with TEMPro and surveyed data. Standard matrix estimation was undertaken and applied back to the prior P/A matrix. The P/A matrix was then re-constrained to TEMPro trip ends. The whole process was repeated until convergence.

While ME was employed to allow additional data (e.g. count data on motorway links) to contribute to the base demand matrices, the highly controlled form in which it entered the process ensured that the matrix building was principally driven by the synthetic trip ends and the RSI data – considered to be a substantial improvement on conventional survey infilling methods.

3. Reading Borough Council Multi-modal TIF Model, 2008

This model's objectives were set by the requirements of the then Transport Innovation Fund (TIF) for multi-modal models which could examine price-based policy options, including road user charging, taking particular account of the different income levels of travellers.

As with Tyne and Wear, the starting point for the base matrices were synthesised P/A matrices matching the DfT National Trip End Model (NTEM) trip rates, which were used as a control during all subsequent processing, including generation of O-D matrices. A distinctive aspect of the PA matrix synthesis was the dual synthesis of 'national' and 'local' matrices. The national matrices were synthesised using singly- and doubly-constrained gravity models calibrated using NTS data for SE England.

The local matrices calibrated extra parameters (calibration area constants) associated with pairs of 'origin and destination area types'. The area types reflected predominant land uses in each zone: origin area types mainly reflected residents' average income characteristics, while the destination area types reflected employment types and activities. The aim of this local synthesis was to reflect the diversity of Reading's employment and education patterns, including various high-tech industries and businesses in the M4 corridor, and its two universities. The basis of the calibration was enhanced Census journey to work information and RSI, bus, and rail surveys conducted by the study. These surveys were successful in capturing statistically useful quantities of household income data, which was used to produce calibrations by income level, required for TIF models.

O-D matrices were derived from the P/A matrices and were then adjusted to reflect traffic count data using matrix estimation. These adjusted O-D matrices were used, in turn, to modify the P/A matrices which were then constrained to match trip productions. This process was repeated until a balance was observed with only limited changes occurring between each iteration.

4.2 Demand model matrices

The construction of appropriate synthetic demand model matrices is closely aligned to what would be done in designing an absolute demand model. In principle, this is a combination of theory and empirical evidence which aims at providing a statistically robust account of the base year travel pattern and its underlying drivers of traveller behaviour.

A first requirement is to define the base year to which the model/matrices will relate. Typically, this will be the year for which the majority of the relevant data is available. As noted in Chapter 3, any data which relates to different years (and this could, in principle, include previously constructed 'prior matrices', provided

their source and derivation is adequately documented) should be appropriately adjusted to be as consistent as possible, with all adjustments also documented.

A second requirement is to define the kind of day to which the matrix should relate. In most cases, this will be a weekday (thus excluding weekend travel), and typically a representative annual average day, or relating to a neutral month. It may also be restricted in terms of hours (e.g. 0600-2200). All subsequent work needs to bear these definitions of the model scope in mind.

A central challenge for the construction of a matrix is caused by the zonal definition (discussed in §3.4). Based on NTS data, we can conclude that in an average day (NB including weekends) some 140 million trips are made in Great Britain by all modes. Of these, some 62% are less than 5 miles, and these will predominantly be found within intrazonal movements or in movements between adjacent zones. A further 32% are in the range of 5-25 miles, with only 6% being longer than 25 miles.

Most study models will have smallest zones in the 'Area of Detailed Modelling', becoming larger for the rest of the fully modelled area and progressively much larger for the external area. As an example, consider the London Transport Studies model. This has 1,279 zones within the fully modelled area, and a further 23 external zones. Disregarding the latter, which are of course essentially for long-distance movements into and out of the fully modelled area, there are $1,279^2 = 1,635,841$ cells in the matrix. Of these, only 46,588 cells (2.8%) have a distance of under 5 miles, and this is probably high compared with most models (because of the high London density). Typically, therefore, the majority of the travel movements will be concentrated into a small minority of the possible zone-to-zone movements catered for within the matrix. In other words, most of the matrix consists of zeroes or relates to relatively infrequent trips.

Clearly, this presents a difficulty. While relatively representative survey data such as household interviews might be preferred for general model development, it will correspondingly contain small numbers of the longer distance movements. It is this which makes intercept surveys an appealing source of data for matrix construction, although they have some undesirable statistical properties.

Given the current infeasibility of a holistic process that brings together all sources in a consistent way, a key consideration is how to reconcile different estimates of the same quantity. For example, the estimate for a particular i-j movement from a synthetic model may be different from one based on a direct sample survey. As we shall discuss, a reasonable way to proceed is to combine the two estimates using an inverse variance weighting. However, quite apart from the fact that the appropriate variances may not be obvious, there is also the question of how to convey the outcome to other parts of the matrix (to ensure, for example, that trip end constraints are properly represented). This suggests that some kind of iterative heuristic process may be needed.

The construction of an independent absolute demand model (types 1-A, 1-B) will require substantial behavioural (e.g. household interview) data, and in practice this will rarely be available at the local level. This report does not presume to offer explicit advice on how this should be done when adequate data **is** available: examples of such models can be found in the Dutch National Model or the

London Transport Studies model. It should be noted that even in such cases, it is likely that the final outcome will be a model of Type 1-B rather than Type 1-A – i.e. some adjustment will still be required to achieve compatibility with assignment matrices that have been adjusted in the process of validation.

The aim here, by contrast, is to present a general default approach to the construction of a synthetic demand model which can serve as the first stage of an iterative process, aimed at achieving a preferred Type 2-A (or Type 2-B) model. This will broadly follow the outline of Chapter 2 using nationally available data. However, it is assumed that local data will be used where available and considered more reliable for local conditions. Examples of variables that might benefit from local data include trip lengths, trip rates or time of day factors.

> **The recommended overall approach may be summarised as one in which an initial set of base matrices is synthesised using national or local sources of information and associated modelling procedures. These synthetic matrices are then modified in two stages: first to reflect cell values obtained from traveller intercept surveys, and second further adjusted according to traffic count data using matrix estimation techniques. An important element of the approach is the key role given to trip production (and, in some cases, attraction) rates, which are used to ensure that any adjustments remain consistent with nationally or locally established values. Thus, the resulting outcome matrices will remain consistent with land use and demographic data, which should be an important constraint, considering its importance in forecasts.**

A particular requirement of the approach is that if the P/A matrix is to be (at least in part) informed by count data, then both directions (outward/return), all purposes, times of day and person segments need to be handled simultaneously (because the count data cannot distinguish these). This increases the challenge of dimensionality, since a large number of matrices are involved. However, it is a tractable method that has been proven in several applications (as in the case studies cited above).

4.3   The default synthetic demand model

The synthetic modelling described in this section is based on conventional practice using current WebTAG guidance. Other model forms can be substituted if considered preferable, although it will be good practice to explain the reasons for adopting other approaches. The synthetic modelling methods are generally well known but are described here in some detail since they form an important element of recommended base matrix development. The description has a multi-modal context in mind, but could also be used for a uni-modal model.

It is assumed that the zoning system, the scope of the model, modes, times of day, journey purposes and further segments to be modelled have been decided, and that the appropriate land use data for the base year is available. If the zones and the segmentation are compatible with TEMPro output, then it should be possible to obtain the trip productions, $T_i^{h(s)}$, and the origins for NHB purposes directly, as well as the corresponding attractions and NHB destinations. At the least, the

productions need to be segmented by household car availability[28]. In most circumstances the appropriate TEMPro time period to use will be the average weekday.

If the zones and the segmentation are incompatible with TEMPro output, usually involving more detail, it will be necessary to download the programs NATCOP and CTRIPEND[29] and run them at the zonal level. If the base year coincides with a decennial Census year, the zonal car ownership data should also be available: however, it will still be necessary to produce an allocation of traveller type by household car availability. In addition, the NTEM trip ends relate to the 7-day week, and need adjusting for the weekday/weekend distinction. This can be done by means of the rho (ρ) factors described below.

It should be checked that the NTEM trip rates (which vary by area type) are compatible with local data, where available. In particular, it may be possible to improve on TEMPro trip attractions for the education purpose[30], since these can preferably involve known school or college locations, student numbers and catchment areas, as well as specific start times. The National Pupil Database can provide local primary and secondary school places or they might be available from local authorities. See https://www.gov.uk/government/collections/national-pupil-database for more detail.

In addition, if TEMPro output is not being used, it will be necessary for the doubly constrained purposes to balance productions and attractions by mode at some appropriate spatial level. The aim is to ensure that within each 'balancing area' the number of productions and attractions are equal: see §2.4 of the Supplementary Guidance for more information.

The next stage is to move to a matrix description and allocate the productions to the attractions and split by mode, separately for each purpose. There are many ways in which this can be done, but chiefly two options:

a) split the productions by mode and allocate to attractions by mode; or
b) allocate (distribute) all productions to attractions and then split the matrices by mode.

The former option is now typical as it corresponds with the standard hierarchy recommended in WebTAG: however, both possibilities will be described since

---

[28] A minimum level of detail should be: no car / part car (fewer cars than adults) / full car (as many cars as adults.

[29] Information is available in https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/370447/TAG_Supplementary_-_NTEM_Sub-Models_November2014.pdf. The programs can be obtained from DfT.

[30] NB Although the current NTEM/TEMPro combine all trips associated with education into a single purpose, they might require separate attention. Given that school trips often take place in the peak hour of the AM peak period, that the car share has increased over time and that education policies are leading to larger catchment areas and so indirectly to greater car demand, their relevance to highway and to PT modelling in cities may be high despite their shorter than average distance. Furthermore, college or university trip making can be quite different to school journeys. The advice here is to think about the importance of education trips in the model and then to consider the best ways to represent the trip making.

there are differences of professional opinion as to which is best for different journey purposes.

In either case, it is necessary to introduce more information about average trip length, segmented by mode and purpose to ensure that the behavioural basis is sound. Although the demand model should be based on generalised cost (GC) specified in time units, there are two advantages from calibrating the synthetic matrix to distance: first, information about the trip length distribution by purpose is available from NTS (or local data), and second the distance is independent of the level of demand (supply effects). The distance matrix D_{ij} can be skimmed from the (unloaded) highway network.

For convenience throughout the modelling, we assume a negative exponential form while recognising that this will not always give a good fit to behaviour. In line with §3.3 of WebTAG M2, it is implied that possible further allowance will be made for 'cost damping' or possibly other ways of improving realism.

Note that, given the complete nature of the trip rates, the matrices will relate to the whole country, although the external zones will be spatially aggregate, and are unlikely to be reliable. The role of the external zones is a) to account for the movements between the study area and the external area and b) to ensure that, for the purposes of assignment, that any movements between external zones that have the potential to route through the study area are adequately modelled. The remainder of the trips **within** the external area are not of interest, and are merely included to maintain the integrity of the trip rates.

In line with NTEM, TEMPro provides data on trips for the following modes:

- on foot,
- by bicycle,
- by motor vehicle (both as a driver and passenger),
- by rail, and
- by bus.

The TEMPro guidance warns users that, "trip ends by mode are based on average rates over a wide area. They do not necessarily take into account the accessibility of each zone by each transport mode." While the active (walk and cycle) modal shares in NTEM can be considered reliable, the shares for mechanised modes – and rail in particular (because rail based accessibility to a reasonable variety of destinations is highly heterogeneous across space and area types) – will require more support from local data, except where PT usage levels are low.

A further complication is due to the NTEM definition of the car mode: this will also include (according to the NTS definition) private trips made by light vans. While this is generally compatible with the TEMPro car ownership definition, the chief issue here is for comparison with counts. We do not attempt to resolve this incompatibility issue here but highlight it as a topic that requires attention.

Although active modes are not usually assigned to the network, and are often intrazonal, they feature in the all-day trip rates by journey purpose from NTEM.

Once modelled (with care in representing their travel costs, typically related to distance, with appropriate distinction between walking and cycling), they can either be considered as residuals to be ignored in network studies or can be a focus in some planning contexts where these modes might be alternatives to vehicular trips under certain policies. In general, such resulting active mode matrices will not be suitably accurate for detailed study in their own right – for that purpose they should have their own studies with data gathering and analysis techniques that are best suited to the topic of interest.

a) Productions split by mode, with separate modal distribution

If TEMPro is being used, then the split of productions by mode $T_{im}^{h(s)}$ is already available, together with the corresponding origins for NHB: the same is true of the attractions and NHB destinations. Otherwise, the modal splits need to be achieved using the NTEM methodology. In either case, the caveats made above should be respected.

Separately for each purpose *p* (covering both home based *h* and non-home based *n*), the 24-hour trip ends by mode are distributed over distance using an appropriately fitted deterrence function. An example functional form for the deterrence function would be the negative exponential[31] – with form $\exp(-\lambda_m^p C_{ijm})$ – calibrated to mean trip lengths $\overline{D}_m^p$ from NTS data (or, where available, from a local household survey if it can support the required information). The aim is to choose λ_m^p so as to ensure that:

$$\overline{D}_m^p = \frac{\sum_{ij} T_{ijm}^p D_{ij}}{\sum_{i'j'} T_{i'j'm}^p} \qquad (4.1)$$

Note that there is evidence from NTS (see Tables NTS9910, 9912) that the trip lengths by purpose may differ substantially between dense urban and low-density rural areas. Hence it may be considered whether the calibration should be done separately for an NTS area type analogous to the production zone: this would mean, of course, having a value of λ_m^p which varies according to the classification of the production zone.

For the purpose of the calibration of the deterrence function, it will usually be necessary to exclude the external-external movements, as they have a disproportionate influence on the mean trip length calculation, especially given the typically arbitrary calculation of intra-zonal distances in the external zones. However, the external zones cannot be dropped completely because the integrity of the trip rates depends on the possibilities for journeys with an internal end to be made to or from external zones, and this assumption is particularly important near the boundary of the study area. A subsequent check can be made that the resulting patterns of estimated external movements is realistic or preferably matches a calibration target defined only for external trips, possibly using a separate parameter.

---

[31] Although others including the log-normal may be used if appropriate.

For all purposes and modes we calculate balancing factors a_{im}^p, b_{jm}^p such that:

- $T_{ijm}^p = a_{im}^p b_{jm}^p T_{i\bullet m}^p T_{\bullet jm}^p \exp(-\lambda_m^p C_{ijm})$

- $\sum_j T_{ijm}^p = T_{i\bullet m}^p$ for all i, m, p (4.2)

- $\sum_i T_{ijm}^p = T_{\bullet jm}^p$ for all j, m, p

(The dot notation in the subscripts indicates the sum over the relevant index i.e. $T_{i\bullet m}^p$ denotes the total productions at zone i of purpose p using mode m.)

Note that the presence of the third equation (destination constraint) and the associated balancing factor b_{jm}^p is in line with the implication that the NTEM based attractions are viewed as the actual number of trips with destination j. This should certainly be the case for the home based work and education purposes, but for other purposes could be relaxed.

b) All modes distribution model, with separate modal split

The alternative approach develops the distribution model at an all-mode level. It is proposed that in the first place[32], separately for each purpose p (covering both home based h and non-home based n), the all day trip ends are distributed over distance using an appropriate deterrence function – e.g. a negative exponential $\exp(-\lambda^p D_{ij})$ – calibrated to mean trip lengths $\overline{D}^p$ from NTS data (or, where available, from a local household survey if it can support the required information). The aim is to choose λ^p so as to ensure that:

$$\overline{D}^p = \frac{\sum_{ij} T_{ij}^p D_{ij}}{\sum_{i'j'} T_{i'j'}^p} \qquad (4.3)$$

Once again, for the purpose of the calibration of the deterrence function it will usually be necessary to make special arrangements for the external-external movements.

For all purposes we calculate balancing factors a_i^p, b_j^p such that:

- $T_{ij}^p = a_i^p b_j^p T_{i\bullet}^p T_{\bullet j}^p \exp(-\lambda^p D_{ij})$

- $\sum_j T_{ij}^p = T_{i\bullet}^p$ for all i, p (4.4)

- $\sum_i T_{ij}^p = T_{\bullet j}^p$ for all j, p

The remarks about the destination constraint in Eq (4.2) apply here as well.

---

[32] This will be re-visited once the mode choice model is calibrated.

With the modal split in mind, a further segment c relating to car availability for the productions should be included. In this case, separate λ values could be chosen for each car ownership segment, to reflect different average trip lengths.

The next stage is to develop a mode choice model of the form:

$$p_{m|ij}^{p} = \frac{\exp(-\beta^{p}[\alpha_{m}^{p} + C_{ijm}^{p}])}{\sum_{m'}\exp(-\beta^{p}[\alpha_{m'}^{p} + C_{ijm'}^{p}])} \qquad (4.5)$$

where C uses a standard WebTAG formulation, and the parameters α_m (these are the 'modal constants', one of which should arbitrarily be set to zero) and β are chosen so as to reproduce to a reasonable extent available evidence on modal shares and the mean modal distances.

Having done this, the all-mode distribution model should be re-visited so that, rather than distance, it is based on the logsum or composite cost (see WebTAG M2 for more details) from the mode choice model. As before, the logsum parameter should be estimated to ensure that mean trip lengths $\overline{D}^{p}$ are reproduced. Note that a logsum parameter value greater than 1 is unacceptable on theoretical grounds.

Outcome matrices by mode

Having used one or other of the two possible DMS methods we now have constructed the all-day synthetic modal P/A matrices T_{ijm}^{h}, together with the NHB O-D matrices T_{ijm}^{n}.

For the purposes of future mode choice, it is desirable to retain a segment c relating to car ownership. This can be done by factoring the home based matrices according to the production proportions:

$$\frac{T_{i\bullet m}^{hc}}{\sum_{c'} T_{i\bullet m}^{hc'}}$$

The resulting matrices can be written as T_{ijm}^{hc}. The NHB matrices cannot be segmented in this way and are left as T_{ijm}^{n}.

Time of day allocation

While the NHB allocations by time period can be obtained from TEMPro, this cannot be done for home based purposes, since TEMPro does not distinguish by direction. It is thus necessary to make use of the NTEM methodology to achieve this. There are two elements[33] to the process.

The first is the set of 'rho' factors which allocate productions to modes and time periods for the outbound leg. These can be re-formed to yield time of day factors by mode, but they also show some variation by: a) car availability and b) traveller

---

[33] These are the factors rho (ρ) and phi (ϕ) defined in TAG Supplementary NTEM Sub-Models paragraph 2.2.1.

type. As noted earlier, these should also be used to convert from a 7-day basis to a weekday basis.

The second is the set of 'phi' factors where $\phi_{th|rq}$ gives the proportion of trips with outward purpose q and period r which return with purpose h in time period t. It is implied that:

$$\sum_{h,t} \phi_{th|rq} = 1 \qquad \text{for each } (r,q) \text{ combination.}$$

The factors allow for a change of purpose between the outbound and return movements, so that to obtain the return movements ($d = 2$) we have to take account of all possible outward purposes:

$$T_{ijmt}^{hc,d=2} = \sum_{r,q} \phi_{th|rq} T_{ijmr}^{qc,d=1} \qquad\qquad (4.6)$$

The outcome of this is the full set of synthetic matrices. The home based matrices are segmented by Purpose, Mode, Time, Car Availability and Direction to give T_{ijmt}^{hcd}, while the non-home based matrices are segmented by Purpose, Mode and Time to give T_{ijmt}^{n}. If TEMPro is being used for all other steps, it should be checked that the total productions and attractions by time period are consistent for each zone and purpose.

Overall, the creation of the synthetic P/A matrix can be represented by **Figure 5**.

Figure 5: Overview of Synthetic P/A Matrix Creation Process for Home Based Purposes
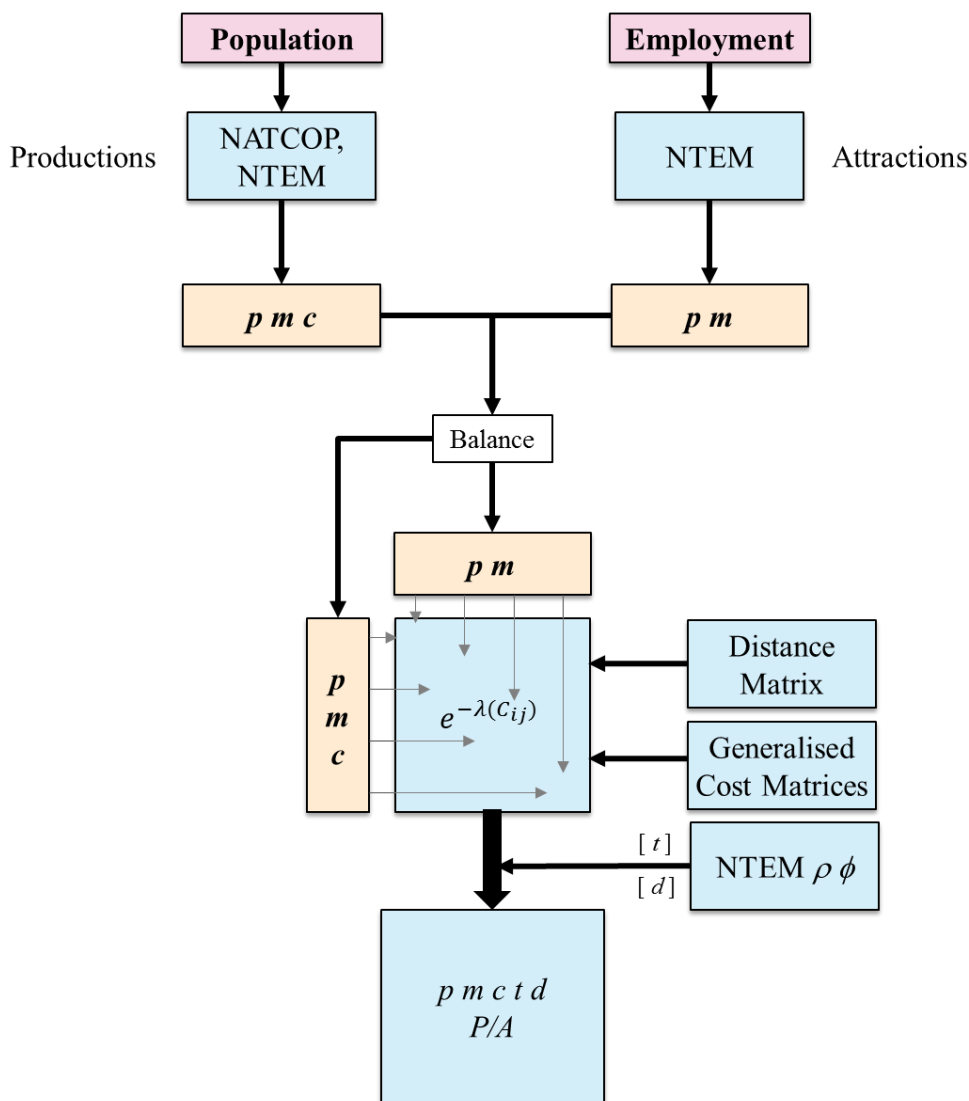
p = Purpose m = Mode c = Car Availability t = Time of Day d = Direction (P/A)

Concluding Remarks

The foregoing has been entirely concerned with creating internally consistent synthetic matrices that are a first step in developing the base matrices. It will be clear that some elements used in the construction of the synthetic matrices are more reliable than others. The greatest reliability probably lies in the trip productions, while the time of day and mode factors may vary much more by location than the national figures from NTEM. Hence the next step is to confront the synthetic estimates with other forms of data, regarding the synthetic estimates as a prior.

The chief data sources to supplement the synthetic model which we will consider are: a) matrices (principally from surveys – see §4.4) and b) link counts (see

§4.5). In both cases, the additional data should be compatible with the model scope (i.e. the year and the definition of the day), and should be derived from a documented procedure. By matrix information we include any data which allows the zones at both ends of a movement to be identified, in some cases, available initially at a more aggregate level.

It is important to note that the information will not always be at the same level of detail – for example, it may be aggregated over some of the segmentation categories in the synthetic matrix. In the absence of more information, any adjustment made to the synthetic matrix will need to be conveyed proportionately to **all** constituent elements.

4.4 Introducing other data sources - matrices

The process of introducing additional matrix data to the synthetic starting point must make allowance for the statistical accuracy of that data, based essentially on sampling theory (in the context of roadside interviews (RSIs) see guidance in DMRB volume 12, section 1 part 1, Appendix D13). The proposal made by Skrobanski *et al* (2012, Equation 6) is that for any two alternative matrix estimates T^a and T^b the best estimate of the fused matrix is:

$$\widetilde{T} = \left(\Omega_{T^a}^{-1} + \Omega_{T^b}^{-1}\right)^{-1} \cdot \left(\Omega_{T^a}^{-1} T^a + \Omega_{T^b}^{-1} T^b\right) \qquad (4.7)$$

where $\widetilde{T}$ is the fused matrix, and Ω represents the covariance matrix. Note that, in this notation, T is being treated as a vector[34] containing every ij cell in a trip matrix representing k zones; hence, the diagonal elements in Ω represent the variances of the k^2 trip estimates T_{ij}, while the off-diagonal terms represent the covariances between different cells (e.g. T_{ij}, T_{rs}). If there are k^2 cells in the trip matrix T, there are k^2 x k^2 elements in the covariance matrix Ω. This is therefore a substantial computational challenge.

However, if Ω contains no off-diagonal terms, then the equation in fact represents the standard result that for any cell the two independent estimates should be linearly combined with weights in inverse proportion to their variances. In this case, the above equation can be written out, for any matrix cell ij, as:

$$\widetilde{T}_{ij} = \kappa_{ij} T_{ij}^a + (1 - \kappa_{ij}) T_{ij}^b \qquad (4.8)$$

Where $\kappa_{ij} = \dfrac{V_{ij}^b}{V_{ij}^a + V_{ij}^b}$, $1 - \kappa_{ij} = \dfrac{V_{ij}^a}{V_{ij}^a + V_{ij}^b}$ and V are the variance terms, so that the mean and the variance of the output fused estimate $\widetilde{T}_{ij}$ are given by:

$$\widetilde{T}_{ij} = \frac{V_{ij}^a . V_{ij}^b}{V_{ij}^a + V_{ij}^b} \cdot \left(\frac{T_{ij}^a}{V_{ij}^a} + \frac{T_{ij}^b}{V_{ij}^b}\right) \quad \text{and} \quad \widetilde{V}_{ij} = \frac{1}{\dfrac{1}{V_{ij}^a} + \dfrac{1}{V_{ij}^b}} \qquad (4.9)$$

---

[34] In this context the trip matrix T is in fact a vector of 1 x k^2 elements, corresponding to a square matrix of k zones.

While this is certainly feasible computationally, and it is useful here to illustrate a desirable approach, it runs up against a key problem. If the new matrix source is based on a sampling process (as would apply to RSIs, for example – although there are complications when the RSI sampling is not uniform), it should be possible to make an estimate of the variance. It is much less clear, however, how this can be done for the synthetic matrix[35] described in the previous section, and even if all the necessary data – relating to model estimation statistics, and the properties of the underlying data – were available, it would be a very substantial task. In the light of this, a more heuristic method may be appropriate: among other approaches, this could be done along the following lines[36] as an example.

For each 'observed' intercept survey cell of the matrix (i.e. representing production zone, attraction zone, purpose, mode, time period as well as possibly some segmentation data relating to car availability, etc.), the prior (synthetic) value is tested as to whether it lies outside of the confidence region of the surveyed data: if so, the prior data will be modified. Note, in this context, that the surveyed data may be more aggregate than the synthetic matrix: as noted at the end of Section 4.3, any aggregate adjustment made to the synthetic matrix will need to be conveyed proportionately to **all** constituent elements. In other words, where the surveyed data provides no additional useful information, the synthetic matrix continues to play an important role.

The basic aim is to establish the extent to which the new matrix data should be allowed to alter the synthetic values, which have the important virtue of being internally consistent and have largely been based on national (or local) data relationships. This will mainly depend on the quality of the new observations based on considerations of:

- Statistical representativeness – reflecting random error, considered as mainly a matter of sampling rates;
- Systematic error – typically reflecting the age of data but also spatial variations;
- Measurement error – this includes survey bias, missing data and anomalous recording.

The default width of the bounds can be taken, for example, to be twice the assessed/assumed standard deviation of the surveyed data. Insofar as some aspects of the data are not subject to rigorous statistical assessment, it may be necessary to adjust this default width, either greater or smaller, on the basis of judgment. A more rigorous statistical bounds testing would be possible, but the extra complexity and computation is generally not warranted.

---

[35] This is a significantly under-researched area. There is very early work by Haskey (Zonal Trip Flow Error Analysis, J Haskey, MAU Note 241, 1972. A General Investigation into the efficiency of trip distribution models, J Haskey, MAU Note 244, 1973) and slightly later work by Gunn *et al* (1980) [Gunn, H.F., Kirby, H.R., Murchland, J.D. and Whittaker, J.C. (1980), The RHTM Trip Distribution Investigation, Proceedings 8th PTRC Summer Annual Meeting, University of Warwick July 1980]. A recent review on the more general topic of model uncertainty is to be found in de Jong *et al* (2007) [de Jong G, Daly A, Pieters M, Miller S, Plasmeijer R & Hofman F, Uncertainty in traffic forecasts: literature review and new results for The Netherlands, Transportation July 2007, Volume 34, Issue 4, pp 375-395].
[36] The method follows that used in the Tyne and Wear study and the M6 study (earlier box).

We begin with a complete synthetic matrix T_{ijmt}^{pcd} as described above. This matrix then gets modified in the light of surveyed matrix data, of various kinds. With the aim of avoiding confusion, we refer to this as **matrix modification** (MM) rather than matrix estimation (ME), the latter carrying with it the more restricted idea of adjusting an O-D matrix in the light of traffic counts.

External to external cell values, which do not impinge on the study area as possible through trips, can be disregarded in terms of transport impacts, once they have been verified as reasonable in terms of their accuracy (and share of trips). After the synthetic elements of the base matrices have been created, it can be appropriate (as with the internal cells) to replace cell values with more reliable P/A values from surveys or other data sources. Validation of the base matrices at the study area boundary will be of particular interest for external trips.

The basic concept behind the matrix modification described here is that we confront a prior matrix (which, in the first iteration, corresponds with the synthetic matrices as just derived) with intercept survey data, subject to confidence intervals. For each data item (which may be combinations of items in terms of the dimensions *i, j, m, t, p, c* and *d*), we calculate the appropriate prior quantity, and if it lies outside the confidence bounds of the observed data, it will be adjusted so as to lie on the boundary or some other suitable point as decided: otherwise it is unchanged.

For convenience, we will write the confidence bounds as $\Psi[X]$ where X is the observed quantity, and we use the notation: $\tilde{T}_{ijmt}^{pcd} \in \Psi[X]$ to imply the pair of constraints:

$$\tilde{T}_{ijmt}^{pcd} \geq X^L \qquad \text{and} \qquad \tilde{T}_{ijmt}^{pcd} \leq X^U \qquad\qquad (4.10)$$

where X^L and X^U are respectively the lower and upper bounds of the confidence interval $\Psi[X]$, and $\tilde{T}$ is the outcome (fused) matrix. In all cases, the confidence intervals will reflect the sample size, using appropriate formulae.

In the course of such modification, certain key features of the prior matrix may be lost: for example, the modifications are likely to change the total productions, or the average trip length. There will therefore be a need for an iterative process which attempts to re-impose some features as constraints. The re-balancing can typically be done by means of a (Furness-style) proportionate fitting method. Note that, provided that journey purpose is available in all the observed data, this adjustment process can be carried out independently for each purpose. If not, uniform factoring across purposes will apply.

In the original studies on which this proposal is based, the synthetic trip ends were taken as the key binding constraints throughout, so that we require:

$$\sum_{j,m,t} \tilde{T}_{ijmt}^{pc,d=1} = \sum_{j,m,t} \tilde{T}_{ijmt}^{pc,d=2} = \sum_{j,m,t} T_{ijmt}^{pc,d=1} = \sum_{j,m,t} T_{ijmt}^{pc,d=2} = T_{i\bullet}^{pc} \;, \quad \forall i, p, c \qquad (4.11)$$

However, it would also be possible to specify appropriate confidence bounds for these trip end quantities, so that in a similar way the modified total would only be brought to the nearest boundary if it lay outside the bounds.

For the doubly-constrained purposes there is also the possibility of a corresponding attraction constraint, although we note that attraction trip rates are less often as stable and reliable as productions:

$$\sum_{i,m,t} \tilde{T}_{ijmt}^{pc,d=1} = \sum_{i,m,t} \tilde{T}_{ijmt}^{pc,d=2} = \sum_{i,m,t} T_{ijmt}^{pc,d=1} = \sum_{i,m,t} T_{ijmt}^{pc,d=2} = T_{\bullet j}^{p} \,, \quad \forall j, p \tag{4.12}$$

The source of the matrix information will be varied. While traditionally much of it would be expected to come from RSIs, the approach outlined here is applicable to other data forms. In particular, it is recommended that use be made of spatially aggregate NTS data, with a view to improving the treatment of external zones. As noted earlier, when the comparison data is at a more aggregate level of detail, any adjustment made to the synthetic matrix will need to be conveyed proportionately to **all** constituent elements.

4.5   Introducing other data sources – counts

Other modifications to the matrices relate to the comparison with counts. Here there are further considerations, in addition to the statistical properties of the counts. To convert the matrices to counts, having transformed them from P/A to O-D (see Eq 2.2), a route choice procedure is needed. The standard User Equilibrium approach was set out in Eq 2.3, where it was noted that the outcome flows V_a on any link a can be obtained (Eq (2.4)) as $V_a = \sum_{ij} A_{ij} \varepsilon_{a|ij}^{*}$, where ε^{*} gives the optimum proportion of A_{ij} allocated to link a. As mentioned previously, the reliability of ε^{*} depends on: a) the underlying theory of user equilibrium, b) the link cost functions, c) the coding of the network and d) the level of demand.

Traffic counts can be a very valuable resource, particularly if available for recent years, when collected over several days and when related to a longer series of automated counts. It is valid to use them to improve matrices from synthetic modelling and interview surveys. This is certainly the case when confidence in prior matrices or surveys is lower than faith in the counts. Indeed, on the HE network it is increasingly the case that the count data is available 24/7 for many past years: hence, when suitably averaged, it will have no appreciable sampling error and accordingly (excepting count machine breakdown and classification errors, etc.) be more certain than other data sources.

In addition, given a matrix, however derived, together with an assignment network, it is reasonable to want to carry out a suitable level of validation, and counts are one of the relatively easy quantities to obtain (even if their statistical properties may be questioned). Ideally, a set of appropriate base year counts should be developed, making allowance for different collection dates and removing conflicts/inconsistencies to the extent possible.

Given such a set of counts, it is of interest how well the current model reproduces these. We noted above that even in a perfect world where both the counts and the matrix are correct, the success of the model in doing this will depend on factors determining the reliability of ε^{*}. It is assumed here that the processes for removing possible network errors set out in the current highway guidance M3.1 have been carried out, but we may still expect substantial errors in the $\varepsilon_{a|ij}^{*}$ values.

The question then remains – how and to what extent should differences between the modelled and observed counts be conveyed to the trip matrix?

The standard approach to this problem is to use some form of matrix estimation (ME). The earliest substantial work on matrix estimation from traffic counts is attributed to Willumsen (1978) at the University of Leeds[37]. His ME2 (maximum entropy for matrix estimation) technique produced a most likely trip matrix from a set of counts. The derivation does not require any underlying structure to the matrix, but is compatible with a version attributed to van Zuylen (1978) which also makes allowance for a prior matrix. The essential methodology appears in the SATURN ME procedure, which is widely used. In the general case, there is an indeterminate set of matrices which would be compatible with a typical set of counts, and the maximum entropy criterion is introduced to resolve this problem.

In spite of this, the derived solution:

$$A'_{ij} = A_{ij} \prod_a X_a^{\varepsilon^*_{a|ij}} \qquad (4.13)$$

(where X_a is a balancing factor for each observed link count, to be estimated) relies on the $\varepsilon^*_{a|ij}$ factors ('PIJA', in SATURN terminology) which are themselves indeterminate since, in the user equilibrium, only the link flows are unique, not the allocation to paths.

Further, note that the set of PIJA values is dependent on the assigned matrix A_{ij}, and is therefore likely to change as a result of the updated matrix, because of changes to link costs and hence routeing. As shown in Figure 13.2 of the SATURN Manual, an iterative approach is followed for the SATURN ME, "whereby an assignment is used to derive the route choice/PIJA factors which are in turn used to estimate a revised trip matrix. This is then reassigned and the process continued until stable values are found."

Since the PIJA factors are not unique, neither will the revised trip matrix be. It therefore makes sense to impose additional constraints on the process, but SATURN does not offer such a facility. Other routines for ME have been proposed: a substantial evolution of the concept, by Logie and Hynd[38], extended the method to consider multiple data sources, including confidence levels for the prior matrices. This was developed into the MVESTM software (now re-packaged as CUBE ANALYST) and the theory has become the essential basis of several of today's matrix estimation tools. The functional form for the matrix is given as:

$$A'_{ij} = a_i b_j A_{ij} \prod_a X_a^{\varepsilon^*_{a|ij}} \qquad (4.14)$$

---

[37] Willumsen, L.G., 1978, *Estimation of an O-D Matrix from Traffic Counts – A Review.* Working Paper. Institute of Transport Studies, University of Leeds, UK.
[38] Logie, M. and Hynd, A., 1989, *Trip matrix estimation from varied data sources.* 17th PTRC Summer Annual Meeting, England

which has an obvious similarity with the ME2 form given earlier, with the additional factors a_i, b_j to allow for trip end constraints. Although MVESTM[39] takes a maximum likelihood approach to estimation, the difference between this and the SATURN maximum entropy objective function is probably not especially significant.

The MVESTM program claims to be able to produce estimates which are weighted towards data of greater reliability, given an assessment of different reliability among the sources. Nevertheless, since it functions on an O-D basis, this makes the trip end constraints and an assessment of their reliability of limited value. Surprisingly, Cube Analyst does not allow for different user classes[40]: if this was done, the introduction of 'direction' as a 'user class' would allow the program to make a material improvement on the SATURN ME.

Another problem with the ME approach is that it is invariably carried out at the time period level, and thus is not well placed to take account of the inherent stability of the trip ends over the entire day. With some effort, this could also be developed along the lines of Cube Analyst, but in the interim the simpler approach of operating separately at the period level will prevail, and this will obviously reduce the confidence in the period-specific productions and attractions.

But the most important deficiency of these ME approaches is that the essentially mechanistic and indeterminate process can lead to considerable distortions of the prior matrix. For this reason, current guidance (WebTAG M3.1) places severe restrictions on its application. In the first place, it is emphasised (paragraph 8.3.1) that the "primary purpose of matrix estimation is to refine estimates of trips **not intercepted in surveys**, which have been synthesised" (emphasis added). Although it is not explicitly stated in the Guidance, this could be taken to suggest what was explicit in the earlier DMRB (vol 12 §2 Part 1) that, "When infilling, fully observed cell values must not be replaced by estimated values from other sources" [4.3.17] and that matrix estimation methods should "not be allowed to adjust observed cell values" [4.3.35]. These statements are very questionable as guidance and should be challenged.

Second, "The changes brought about by matrix estimation should not be significant" (WebTAG M3.1 paragraph 8.3.14), and the Table 5 which then

---

[39] It accepts as input:
- A set of route choice probabilities indicating how likely a trip from a given origin to a given destination uses a given link
- A set of (origin-destination[a]) trip ends with associated confidence levels
- A set of traffic counts, organised into screen lines, with confidence levels associated with each count
- A prior matrix[b] with confidence levels associated with each matrix cell

Given this information, the software applies statistical maximum likelihood methods to find an output matrix.

[a] although the software description does not make this explicit, Cube Analyst assumes that all matrices and hence the row and column totals are in O-D form rather than P/A form

[b] again, O-D rather than P/A

[40] This was understood to be the position as at 2008, though it was confirmed in discussion that the additional programming to achieve this would be minor. We have not seen any evidence that the position has changed.

follows sets clear limits to what changes are acceptable. These limits are considered arbitrary and do not take account of the data errors.

The problem is that these points of caution are at odds with current requirements for validation of the highway model. Although there are a number of elements to highway validation, the aspect that involves the most focus is how well assigned flows fit with available traffic counts. This concentrates substantially on the 'GEH statistic', a modified chi-squared statistic developed by Geoffrey E Havers at the Greater London Council in the 1970s. He used it to illustrate how well assigned flows in the highway modelling for the Greater London Development Plan would fit to available traffic counts, aiming to give an estimate of confidence in the modelling. It is and has become a useful method for reporting on traffic flows in model validation, although not intended by Havers at the time to be adapted to become a rigorous target.

In many cases it has been treated by practitioners as a metric that **must** be achieved for a model to be valid, although that is not the DfT's intention in the WebTAG guidance. WebTAG does not use the term 'compliant' in reference to models and DfT has made it clear that guidance is just that – an assembly of current views on good practice. While there are suggestions that modellers should aim to obtain certain levels of GEH comparisons, these are only a guideline.

The tension lies in the fact that, while the available ME methods have some flexibility, they are all based on a mathematical formula which aims to reproduce the counts as far as possible (subject to potential inconsistencies between counts). Although in some cases the counts may be considered extremely accurate, in principle the estimation process should make allowance for lower levels of reliability.

An alternative approach (data fusion) advocated by Skrobanski *et al* (2012) [41] aims to combine the estimates of matrices and counts according to statistical properties, which in principle is to be preferred. The essential theory can be briefly described as follows:

Both the prior matrix A_{ij} and the observed counts V_a^{obs} have associated variance matrices Ω^A and Ω^V. In practice these are diagonal. The fusion is achieved by means of minimising the (least squares) criterion:

$$(A'-A)^T(\Omega^A)^{-1}(A'-A)+(pA'-V^{obs})^T(\Omega^V)^{-1}(pA'-V^{obs}) \qquad (4.15)$$

where the matrices A_{ij} are interpreted as one-dimension vectors, and p is a (PIJA) matrix of the form a^*_{ij}, equivalent to the elements of $\varepsilon^*_{a|ij}$. The minimisation is with respect to the elements of the 'fused matrix' A'.

The solution is shown to be:

$$A'=\left[(\Omega^A)^{-1}+p^T(\Omega^V)^{-1}p\right]^{-1}\left[(\Omega^A)^{-1}A+p^T(\Omega^V)^{-1}V^{obs}\right] \qquad (4.16)$$

---

[41] NB A number of different methods are described in the paper, but we are here confining to the analytical WLS [weighted least squares] method.

The two Ω matrices, being diagonal, are readily inverted. The inversion of the first term in square brackets can be simplified by use of the Sherman-Morrison-Woodbury formula whereby:

$$M = \left[(\Omega^A)^{-1} + p^T (\Omega^V)^{-1} p \right]^{-1} = \Omega^A - \Omega^A p^T (\Omega^V + p\Omega^A p^T)^{-1} p\Omega^A \qquad (4.17)$$

The matrix M can also be shown to be the covariance matrix of the revised estimate A'.

Unless the counts are considerably more accurate than the matrices, it will not be expected to obtain such a good fit to the counts as can be achieved by other ME methods. This of course has consequences for the GEH statistics.

In addition, in the same way as the revised matrix can result in changes to the SATURN PIJA factors, an iterative procedure is in principle also required for the Skrobanski *et al* (2012) approach (though not discussed by the authors). It is important to note that, apart from some MSA (method of successive approximations) procedure, there is no way in which successive estimates of the matrix can be combined using the statistical theory of Skrobanski *et al* (2012). This is because the process assumes that the off-diagonal terms in the variance matrix Ω^A for the prior are zero, an assumption which can no longer be maintained once a modified prior has been obtained as a result of taking the counts into consideration using the theory outlined above.

The aims of the Skrobanski method are valid but it can be seen that there thus remain substantial practical problems in making significant improvements to the currently available ME software.

A related question is how far to translate the changes to the matrix brought about by (some version of) ME to the all-day P/A matrix. In the light of the difficulty of doing this, a simpler, but potentially less satisfactory approach, would be to accept a type 1-B approach as noted in Chapter 2 in line with the method of Daly *et al* (2012). This can be interpreted as: work with the synthetic matrix on the demand side, but when assigning convert to O-D and pivot changes on the 'validated' base O-D (using essentially M3.1 techniques for validation). As noted earlier, however, there is no known procedure for converting these adjustments back to the demand model. While this may be considered acceptable for a uni-modal model, there must be concerns about its use in a multi-modal context.

There are thus two substantial issues:

- how do we improve on use of ME within current or revised validation guidance?
- how far do we go in updating the P/A matrix?

This depends to a considerable extent on how closely the original or modified synthetic matrix is considered to be validated to counts at the network level.

4.5.1 Using ME to improve the P/A matrices

If it is desired to use the ME results to further update the base P/A matrix, the following approach, based on experience from the Tyne and Wear and M6 studies, can be taken. This illustrates a process that works in practice and that deals with the desire to reflect the results of O-D validation back to P/A so that the model is best suited for forecasting. It is not claimed to be the only way of doing this, nor is it as developed as it could be, but it represents a concept and method to build on.

For each time period t in which the assignment is carried out, calculate the ratio of the final matrix (after ME) to the initial matrix (as input to the assignment). Writing the final matrix as $\tilde{A}_{ijmt}$, calculate the adjustment factors $E_{ijmt} = \tilde{A}_{ijmt} / A_{ijmt}$. Since the ME procedure contains no information about either purpose or, for home based purposes, direction, the same ratio has to be applied to each purpose and direction (in the latter case, after transposing for the return portions of the P/A matrices). In other words:

$$\tilde{T}'^{pc,d=1}_{ijmt} = E_{ijmt} \tilde{T}^{pc,d=1}_{ijmt} \quad \text{and} \quad \tilde{T}'^{pc,d=2}_{jimt} = E_{ijmt} \tilde{T}^{pc,d=2}_{jimt} \quad\quad (4.18)$$

[note the transpose in the subscripts of the second equation].

After once again applying the production constraints (and the attraction constraints if desired in the case of doubly constrained purposes), the whole process can be repeated until acceptable convergence is achieved. Note that, as a variant, it would be possible to apply these constraints on the same basis as discussed in the matrix modification section earlier, making use of the confidence interval for the constraints (Ψ). The resulting P/A matrices can then be used as the more satisfactory pivot for the incremental demand model.

Throughout this section we have viewed the general approach of ME as a method of introducing the information from counts in a sensible way to the process of matrix construction, not only as a means of achieving model validation. An outstanding matter of concern is that current forms of ME will result in the solution that the matrix is actually constrained (distorted) to reproduce the counts. The combined method of MM and ME, within an iterative system, goes some way to alleviating this problem.

As a corollary, the current guidelines for validation based on counts need to be reassessed. The role of matrices in the transport model is much wider than their performance in matching a set of counts, and in this respect, matrices that do not give a good 'validation' fit to counts should not be viewed as inadequate (provided, of course, they reflect all available evidence in an appropriate way).

In summary, the following is the recommended outlook on the use of matrix estimation:

- **ME should seldom be used as a last step in base matrix construction to make the matrices fit with counts. This will distort the model such that it is less suited for forecasting.**

- **The ME tools will constrain the matrix (in response to assignment) such that the counts will be replicated as near as they can. This is a suitable technique, with well-trusted counts, if the objective is to improve network speeds to feed back into the model. Note, however, that new and better estimates of speeds are becoming available from various sources of remote sensing.**

- **ME does have a role to play as part of an approach in which the valuable information that is present in counts is brought into the matrix creation along with other inputs and constraints.**

- **Thus, it is propose that validation should be treated holistically, with attention to a range of factors that will make the model fit for purpose, and that attention to counts should not normally be paramount.**

4.6　　　Overview of proposed approach

Overall, the various procedures can be illustrated by **Figure 6**.

Figure 6: Matrix Adjustment Procedure
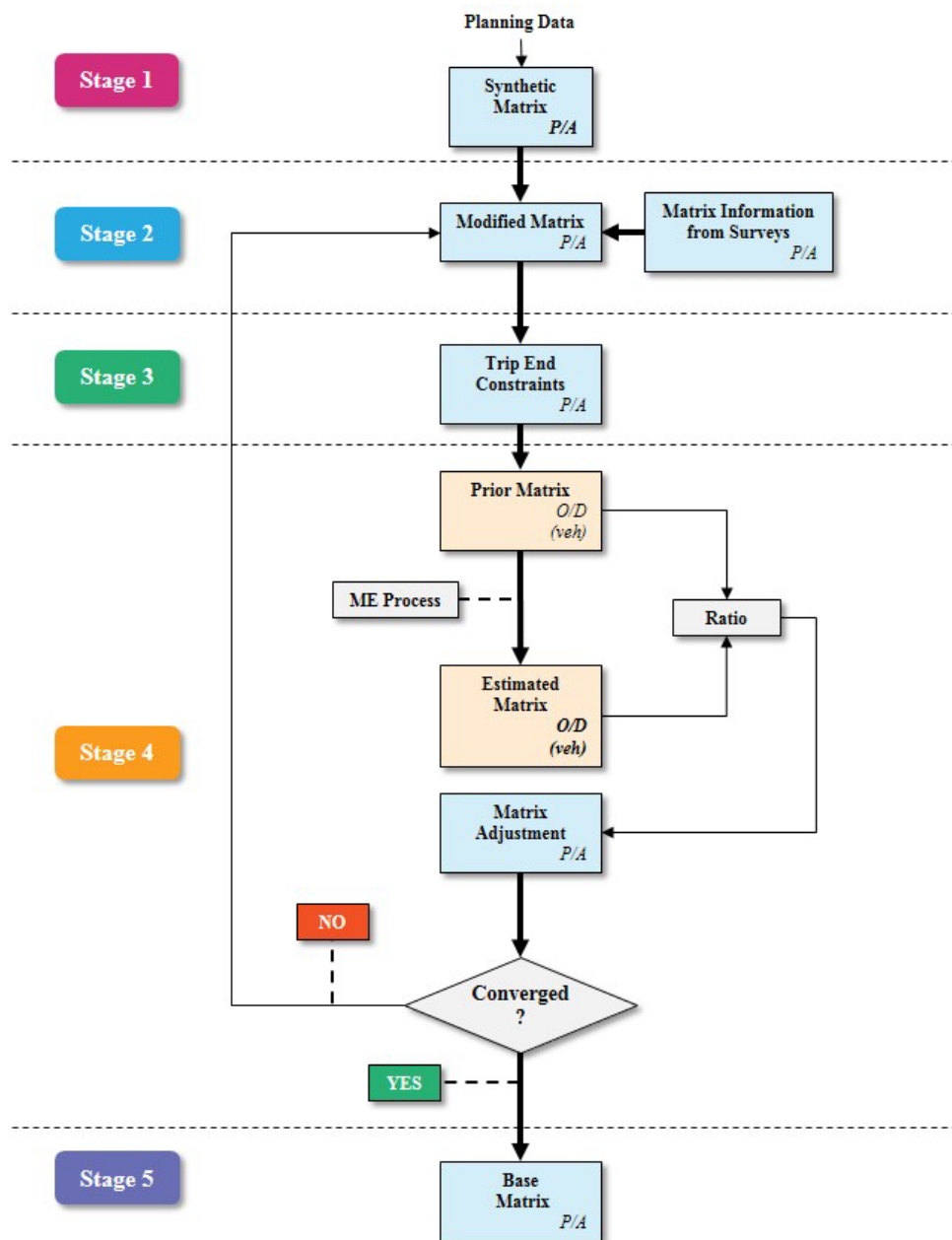
If, given all these adjustments, it is considered that the final derived base P/A matrix can, when converted to O-D form, be satisfactorily validated at the assignment level, then the form of the overall model will correspond to type 2-A in **Figure 3**. If the final base P/A matrix cannot produce a satisfactory matrix for assignment, then it will be acceptable in cases when that is a priority to carry out further matrix estimation work on the resulting O-D matrix, to produce an assignment matrix which can be satisfactorily validated but which is not compatible with the base P/A matrix: in this case the overall model will be of type

2-B form, and predicted changes in demand (after conversion from P/A to O-D) will be used to adjust this assignment matrix as an approximation.

Finally, when the base P/A demand matrix and the overall model structure are deemed satisfactory, the standard WebTAG 'realism' tests should be performed, in line with §6.4 of WebTAG M2.

While this is the preferred approach, there will be cases (primarily those where the modelling interest relates to a single mode) where the assignment-based pivot (1-B) will be acceptable, typically using the AMAI method. This also applies to the cases in the following section. **Table 6** provides an assessment of the key model structures discussed.

Table 6: Incremental Model Structures

Model Structure	Description	Approach to Use of Base Matrix	Advantages and Disadvantages	Likely Areas of Recommended Application
1-A	Traditional 4-stage without pivoting	Not needed	**Pro:** Straightforward application of 4-stage model. **Con:** Unlikely to deliver acceptable validation.	Scenario tests where validated assignment is not a priority
1-B	Pivots off base assignment (O-D) matrix (demand model 'synthetic')	AMAI	**Pro:** Modification only to validated assignment matrices; improved supply effects for modified mode. **Con:** No obvious way to convey modifications to demand model.	Predominantly uni-modal models, and possibly simple applications (Section 4.7)
2-A	Demand model pivots off base P/A matrix, assigned without further pivoting	Pivot-point **or** AMAI	**Pro:** Demand model fully reflects all relevant data, and produces acceptable assignment. **Con:** Currently unfamiliar process to some practitioners.	Multi-modal models
2-B	Demand model pivots off base P/A matrix, assignment pivots further off base assignment matrix	Demand: pivot-point **or** AMAI; supply: AMAI	**Pro:** Demand model fully reflects all relevant data. **Con:** Currently unfamiliar process PLUS further modifications required to secure satisfactory assignment (so demand model is not used directly).	Multi-modal models

NB: as discussed in Section 2.4, pivot-point models use only the cost **changes** to estimate the changes in the number of trips from an independently derived base P/A matrix, while AMAI (absolute models applied incrementally) models use absolute model estimates to apply changes to an independently derived base matrix which may be either P/A or O-D.

4.7   Special cases

In some instances, the relative complexities of a consistent P/A based demand model are not really necessary. This is likely to be the case when small local schemes are being considered, and/or when the expected improvements in generalised cost are not likely to result in significant changes in demand. Hence it may be acceptable to take a uni-modal approach, and concentrate on the assignment outcomes.

In such cases, an O-D based approach can be retained, and the general methods outlined as 'current practice' in §3.8 may be used, subject to the following general caveats:

- while the use of 'trip reversal' is generally to be discouraged, it may be considered acceptable, subject to appropriate checks, in this context;
- cells in the screenline matrix classified as fully observed should not be considered fixed, either in the infilling process or in any subsequent ME application.

As noted, if a P/A demand model is available, the AMAI process could be used with the O-D matrix.

5 Improving Best Practice

5.1 Summary of general advice

This list of recommendations also appears identically in the Executive Summary.

1. Any matrix building exercise should begin with a statement of the usage objectives for the matrices and the approach should reflect these aims.

2. A full data statement should be produced, explaining what is known about all input data, its provenance, definitions, processing history, quality, suitability and how the data features in the model. Data should not be used when unsure about its suitability without providing strong caveats.

3. Surveyed (so called 'observed') data should not be considered as any more valid than synthetic/modelled data as a matter of course. Each source should be evaluated for its value, errors and biases. It is not good practice to freeze and retain surveyed cells during matrix development as if they are specially trusted without thorough justification.

4. Validation of base matrices is an important aspect of model development, but an evolution of current practice and guidance is proposed. Generally, all relevant available data should be used in developing the matrix with concurrent calibration and validation. It is not recommended to hold data back specifically for the purpose of what is sometimes referred to as independent validation, except in special circumstances, but rather to concentrate on how well the model fits all of its inputs, recognising the mutual inconsistency that usually exists among the data sources.

5. Generally (except for minor schemes, those where variable demand is not a feature or early scheme business case development), matrix construction and modelling should be in both P/A form and O-D form for the purpose of demand and assignment models respectively. By preference, changes to O-D form during calibration and validation should feed back to P/A form.

6. A suitable balance should be sought between design of the matrices for good forecasts (segmentation, spatial accuracy and constraints) and achieving the best possible validation for the present day. In many cases this means that constraining matrices to present day traffic or passenger counts will not be an overarching priority, but that counts will be part of a more holistic matrix construction process. Validation effort should be concentrated in areas where schemes are to be tested. Explanations are needed when validation fit is below ideal levels, but this does not necessarily mean that the model is unsuitable for its intended purpose.

7. The technique of matrix estimation to counts should be used appropriately. It should not usually be applied in order to over-fit matrices to agree with count data as a final step of adjustment. Nonetheless, counts can be very valuable data and matrix estimation can be part of a process to help introduce them to matrices in a balanced way alongside other data.

5.2　Implications for guidance

As noted in §1.1, existing guidance for base matrix construction is contained in TAG units M3.1, M3.2 and M2. On the understanding that a new unit will be prepared (drawing from material and recommendations in this report) specifically relating to base matrix construction, we suggest the following changes to existing Guidance.

M3.1

In 1.2.1 add reference to a new unit, changing the last sentence to: "Specific advice on trip matrix development is given in [new unit]". In 1.2.2 remove final parenthetical remark. In 1.2.3 add new unit as a 'companion'. In 3.2.5 reflect on advice in the light of this Report: certainly it should be made clear that the existing advice relates to **matrices used only for assignment**. Reflect on §4.2 in the light of this report. In 8.1.1 replace Section 1 by a reference to the new unit. In Appendix F the items under "7 Trip Matrix Development Methodology" will need re-consideration. Finally, note that there are potentially some more 'philosophical' inconsistencies between the general advice in M3.1 and that given in this report concerning a) the roles of calibration and validation, and b) matrix estimation: these will need to be carefully considered.

M3.2

In §6.1 insert at an early stage a reference to the new unit: most of the material in this section and §6.2 could be removed, but retaining 6.1.3 and 6.1.4.

M2

2.5.7 can make explicit reference to the new unit, and 2.5.8-2.5.10 can be dropped, as can Appendix B (where, as noted, some of the text has already been used in this report). There is some inconsistency within 2.5.3 as to how important O-D based methods might be ("a number of circumstances"…"very uncommon"). Note also that in 2.5.6, the requirement for base matrices has hardly been discussed previously. 2.5.13 is rather out of place, and could be dropped (given the discussion in this report), while 2.5.14 would be better following 2.5.3. At the end of 4.3.1, the sentence about choice of model could perhaps be strengthened, while the whole section could be re-considered in the light of Chapter 2 of this report. 4.3.13 can probably be removed, as can 4.3.16-17. 4.4.2 should mention the need to transpose the matrix for the return trip, and 4.4.3 could be re-drafted making use of the model types defined in this report. The second part of 4.9.2 should be deleted, and the reference in 4.9.6 to "largely observed" should be re-considered. In 6.2.4, the unstated but critical issue is that the existing software operates on an O-D rather than P/A basis.

5.3   Recommendations for research and development

In preparing this report the authors have noted aspects where current practice has limitations or where investigation could bring benefits. This section contains brief comments on the main points identified.

- **Investigate O-D adjustments feedback to P/A to improve methods**

  The advice in Chapter 4 is based around an example of a method to maintain the base matrices in P/A form after any adjustment to matrices in O-D form. The method can be enhanced in a number of ways and demonstration projects are advocated.

- **ME software enhancement**

  While user classes in ME are seen as only relating to identifiable distinction of **vehicles**, introducing further facilities would allow some of the other potential factors (trip-ends, etc.) to be more fruitfully reflected, particularly in relation to productions and attractions. The development of available techniques within the general area of ME which could use ANPR and other such data relating to part of the overall journey to improve the matrix would be a considerable advantage. An improved methodology is required where the extent to which the outcome matrix is constrained to counts depends on their reliability.

- **Treatment of intrazonal costs**

  The advice in this report includes the modelling of all person trips (including active/slow modes) since standard trip rates are on that basis. It also recommends that the whole of Britain should be represented in models to allow for trips in and out of study areas. Both require estimation of intrazonal costs that cannot be derived from networks. It would be helpful to develop guidance on calculating such costs.

- **Light vans used for personal travel**

  It is noted that trip rates tend to include personal travel made in vans along with cars, whereas they are separated in traffic counts. Guidance on dealing with this inconsistency would be useful.

- **Statistical properties of synthetic matrices**

  The report recommends that data should be blended in ways that reflects statistical characteristics of the sources. But estimating errors for synthetic matrices is complex and there is limited advice on how to do this. Further research is warranted.

- **Holistic process that brings all sources together in a consistent way**

  Examples of projects that have attempted to blend disparate data sources have been encouraging but inconclusive (e.g. Bayesian methods in the EU OPUS project). This is a topic for longer term research.

- **Metadata protocols for data and models**

  Preparation of a standard data information template is proposed in Chapter 3. This could usefully be taken further to specify how metadata on models

should be presented. Both sets of information would help greatly with the documentation of modelling projects and with the archiving and handover of models. An experimental method and tools for describing statistical model metadata were demonstrated in the EU OPUS project and could be used as a guide for further development.

- **Use of mobile network data in trip matrix development**

    Preparation of advice for guidance is underway and is mentioned here for completeness. Much more research and development is anticipated.

- **Disaggregation of gross factors (e.g. car occupancy, time of day)**

    Most models use area-wide single factors (perhaps distinguished by journey purpose) to convert all day and car person trip matrices to time periods and to take account of car occupancy. This is a coarse aspect, incompatible with the level of detail that might be sought in assignment and validation. Further advice should be developed.

- **Freight and goods vehicle models**

    Freight matrices were outside the scope of this review, yet goods vehicle traffic is an extremely important aspect of highway assignment (and of transport policy in some areas). Current data sources and models are remarkably weak and a concerted effort at improvement is advocated, separating light van research from freight vehicles.

- **DMS hierarchy**

    There are differences of opinion among modellers as to the most appropriate hierarchy for distribution and mode choice. WebTAG currently recommends that mode be above (or at the same level as) destination choice (M2, paragraph 4.5.14), but this is in fact based on fairly limited evidence. One school of thought is that it should vary by journey purpose. Further, targeted research is required, in particular to develop an improved understanding of how best to model trip distribution, where little focus has been placed. The poor performance of the distribution model may be strongly influenced by the lack of adequate segmentation by person type and car ownership category, particularly for journeys to work and school and perhaps could also be improved by greater and more imaginative use of doubly constrained models.

- **Tours/NHB/Serve passenger**

    Some aspects of demand modelling could be better represented as tours, in particular to improve the representation of NHB and serve passenger trips. Research could be based on a variety of household surveys that are available.

- **Household activity models**

    There is a resurgence of interest in household member activity as a better representation of causal effects on behaviour, particularly when household members undertake activities together or compete for/share family cars. Major metropolitan planning organisations (MPOs) in the USA have research programmes in place. The Transport Systems Catapult in the UK is considering activity modelling as a basis for person trip simulation models in

response to a need for increasing understanding of the uncertainty of forecasts caused by disruptive influences in society and technology.

- **Partial matrix technique**

  As indicated in Chapter 2, use of the technique was discouraged when it was found that results could be unstable. However, there are indications that it could be usefully applied to help populate unobserved cells of matrices and estimation of synthetic gravity models from partial data. Research into its appropriate use is suggested.

Appendix A

Eight Cases Method for
Applying Synthetic Growth

A1   The eight cases distinguished by Daly et al (2012)

This Appendix sets out the eight cases based on Table 1 in Daly *et al*, but using the notation developed in this report. As can be seen, the cases depend on testing the various quantities against zero, and the authors suggest that in practice a test value of 10^{-3} could be used. The standard case given in Eq 2.6 of the main text corresponds with case 8, but even here an exception is made for 'extreme growth'. The idea (which applies in case 4 as well) is that beyond a defined cut-off point, the predicted absolute growth is applied, so that there are in fact 10 cases.

Table 7: Eight cases method for applying synthetic growth (taken from Daly et al, 2012)

Case	"Validated base" B^0_{ij}	Synthetic base (O-D) A^0_{ij}	Synthetic test (O-D) A_{ij}	Output B_{ij}	
1	$= 0$	$= 0$	$= 0$	0	
2	$= 0$	$= 0$	> 0	A_{ij}	
3	$= 0$	> 0	$= 0$	0	
4	$= 0$	> 0	> 0	Normal growth $(A_{ij} \leq X_1)$	0
				Extreme growth $(A_{ij} > X_1)$	$A_{ij} - X_1$
5	> 0	$= 0$	$= 0$	B^0_{ij}	
6	> 0	$= 0$	> 0	$B^0_{ij} + A_{ij}$	
7	> 0	> 0	$= 0$	0	
8	> 0	> 0	> 0	Normal Growth $(A_{ij} \leq X_2)$	$B^0_{ij}.(A_{ij} / A^0_{ij})$
				Extreme Growth $(A_{ij} > X_2)$	$B^0_{ij}.(X_2 / A^0_{ij}) + (A_{ij} - X_2)$

The cut-off values X_1 and X_2 for cases 4 and 8 are defined as follows:

$$X_1 = k_2 A^0_{ij} \qquad \text{and} \qquad X_2 = k_1 A^0_{ij} + k_2 A^0_{ij} \cdot \max[(A^0_{ij} / B^0_{ij}), (k_1 / k_2)]$$

where it is noted that, "Common values for the parameters k_1 and k_2 are $k_1 = 0.5$, $k_2 = 5$" (with the implication that these might be chosen by the analyst). However it appears that the most recent recommendation is that X_2 should have the same value as X_1.

Appendix B

Merging Data from Different Sources

B1 Merging Data from Different Sources

TAM (DMRB Vol 12 S1 P1 1991) section 8.7 discusses principles for merging data from different sources. The section identifies two different cases namely:

- Where the accuracy is known; and
- Where it is not.

Where the accuracy is known, the theory uses the fact that all observations are valid but their accuracy depends upon the sample upon which they are based. It recommends that trip matrices, and the associated indices of dispersion, are merged in such a way as to minimise the coefficient of variation (ratio of standard deviation to mean) of the combined cell value. Note that this is a slightly more complex requirement than requiring the merging to minimise the variance of the combined cell value, which leads to a straightforward result noted below.

Suppose we have two estimates of the same quantity T_1 and T_2. The corresponding variances are σ_1^2 and σ_2^2. We adopt the TAM approach of working with the 'index of dispersion' (ratio of the variance to the mean) so that we have $I_1 = \sigma_1^2 / T_1$ etc. The aim is to find an appropriate way for combining the two estimates, and in general terms we can consider the weighting approach $T_m = \lambda.T_1 + (1-\lambda).T_2$.

Note that it is easily shown that to obtain the minimum variance estimate of T_m we should set $\lambda = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$ and this can be further expanded to:

$$T_m = \frac{\sigma_2^2.T_1 + \sigma_1^2.T_2}{\sigma_1^2 + \sigma_2^2} = \frac{\sigma_1^2.\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.\left[\frac{T_1}{\sigma_1^2} + \frac{T_2}{\sigma_2^2}\right] \qquad \text{(B1)}$$

This is the *inverse-weighting procedure*, which extends to multiple observations.

If we require the minimum coefficient of variation, then it can be shown that we should set $\lambda = T_1.\sigma_2^2 /(T_2.\sigma_1^2 + T_1.\sigma_2^2)$ and this can be further expanded to:

$$T_m = \frac{\sigma_2^2.T_1^2 + \sigma_1^2.T_2^2}{T_2.\sigma_1^2 + T_1.\sigma_2^2} = \frac{I_2.T_1^2.T_2 + I_1.T_2^2.T_1}{I_1.T_2.T_1 + I_2.T_1.T_2} = \frac{I_2.T_1 + I_1.T_2}{I_1 + I_2} \qquad \text{(B2)}$$

This can also be extended to multiple observations. This is the equation used in MATVAL (see Wyley et al, 1981)[42].

While TAM provides one example of merging in Appendix D3, it merely amounts to adding the scaled-up trip estimates to obtain the merge estimate. In that example this is appropriate because the two estimates are actually disjunct – trips between i and j which route through different observation sites.

---

[42] Wyley, W.J., Hardcastle, D.J. and Fearon, E.J. (1981) Traffic data variability and its incorporation into model validation procedures. Traffic Engineering and Control, 22(8/9) pp. 473-477.

TAM includes further statistical information within appendix D13. The appendix gives the following formula for the variance of a trip estimate Q_a having a certain attribute *a* derived by sampling a proportion of trips:

$$Q_a = \frac{q_a}{q} Q \equiv \frac{Q}{q} q_a \qquad \text{TAM Equation 1}$$

$$Var(Q_a) = \frac{Q.(Q-q)}{q^2(q-1)}.q_a.(q-q_a) \qquad \text{TAM Equation 2}$$

Where:

Q is the counted flow within a period (e.g. 1000 cars),

q is the total number of vehicles interviewed (e.g. 100 cars interviewed),

so that

Q/q is the expansion factor

q_a is the number of records interviewed having the required attribute (e.g. 10 HBW interview records)

The variance formula comes from the hypergeometric distribution (though this is not stated!), based on the fact that the same vehicle cannot be sampled twice ("without replacement").

There are various ways in which this formula can be simplified. If we use the binomial (or Bernoulli) distribution – thus ignoring non-replacement – the

corresponding formula for the variance is: $$Var(Q_a) = \frac{Q^2}{q^3}.q_a.(q-q_a)$$

(B3),

which differs from TAM Eq (2) by a factor of $\frac{(Q-q).q}{Q.(q-1)}$. When q is a reasonable

size, the factor $\frac{q}{(q-1)}$ is approximately 1 and as long as q is small relative to the

total population, the factor $\frac{(Q-q)}{Q}$ will not be much less than 1. In practice, this

last condition will not always be fulfilled, depending on the sampling fraction.

MATVAL gives an alternative approximation:

$$Var(Q_a) = \frac{Q.(Q-q)(q-q_a)}{q^3}.q_a \qquad (B4),$$

which differs from TAM Eq (2) by a factor of $\frac{(q-1)}{q}$. As before, when q is a

reasonable size, the factor $\frac{q}{(q-1)}$ is approximately 1. In fact, noting that q_a/q

will normally be very small, so that $(1-q_a/q) \approx 1$, in implementation MATVAL uses the further approximation:

$$Var(Q_a) = \frac{Q.(Q-q)}{q^2}.q_a \qquad\qquad \text{(B5)},$$

But in any case DMRB goes on to recommend ("when Q is large and q_a is small") the even greater approximation:

$$Var(Q_a) = \frac{Q^2}{q^2}.q_a \qquad\qquad \text{TAM Equation 3}$$

Note that if this TAM Equation is used, then the index of dispersion I is given as Q/q, i.e. the expansion factor.

Working this through Equation B2 leads to the following:

$$T_m = \frac{T_1\frac{Q_2}{q_2} + T_2\frac{Q_1}{q_1}}{\frac{Q_1}{q_1} + \frac{Q_2}{q_2}} \qquad\qquad \text{(B6)}$$

Substituting for T_1 and T_2 using $T_1 = \frac{Q_1}{q_1}.q_a^{(1)}$ etc. and tidying up leads to:

$$T_m = \frac{\frac{Q_1}{q_1}.\frac{Q_2}{q_2}.q_a^{(1)} + \frac{Q_1}{q_1}.\frac{Q_2}{q_2}.q_a^{(2)}}{\frac{Q_1}{q_1} + \frac{Q_2}{q_2}} = \frac{Q_1}{q_1}.\frac{Q_2}{q_2}.\frac{q_a^{(1)} + q_a^{(2)}}{\frac{Q_1}{q_1} + \frac{Q_2}{q_2}} = \frac{q_a^{(1)} + q_a^{(2)}}{\frac{q_1}{Q_1} + \frac{q_2}{Q_2}}$$

(B7)

Thus on these assumptions the merged value is equivalent to the sum of the number of records divided by the sum of sampling fractions.