ME&P Ref: 10551148

PPAD 9/134/05

September 2002

DfT Integrated Transport and Economic Appraisal

Review of Freight Modelling

Final Report

Study Team

ME&P - WSP – Project Leader

University of Westminster
ITS, University of Leeds
Rand Europe
MDS-Transmodal
Katalysis Ltd.
Oxford Systematics
Parsons Brinckerhoff
Imperial College

ME&P
49 – 51 High Street
Trumpington
Cambridge
CB2 2HZ

Tel: +44(0)1223 840704
Fax: +44(0)1223 840384

http://www.wspgroup.com

Issue/revision	Issue 1	Revision 1	Revision 2	Revision 3
Remarks	Final Version			
Date	25 September 2002			
Prepared by Signature	Ian Williams			
Checked by Signature	Neil Raha			
Authorised by Signature	Ian Williams			
Project number File reference	10551148 Orignls\Wa_c\ Model_Design_Final.doc			

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The aim of this study commissioned by the ITEA Division of DfT is to review current freight modelling techniques, and through assessing the suitability of the options potentially available, to make recommendations on the most appropriate techniques for use in Great Britain. The review includes road, rail and other freight modes and also includes the modelling of light goods vehicles (LGVs).

Within this study the team has carried out a wide-ranging review:

- of the requirements for freight models in GB, and

- of the approaches to modelling that have been used in the UK and in other parts of the world.

This understanding has been used to develop a framework for the modelling of freight movements at the national level and then to outline how non-freight related LGV movements might also be modelled. The means by which the national level model might be utilised to provide data to improve highway modelling at finer spatial scales is outlined, together with other recommendations for improving these smaller scale models.

Considerable progress has been achieved in developing a comprehensive overall framework for freight modelling. As part of this framework the report introduces a coherent and integrated representation of many important features that have in the past been introduced infrequently, if at all, in freight modelling in the UK:

1. Various options of increasing sophistication and realism (and unfortunately of increasing associated data requirements) for linking growth in the economy to growth in the demand for freight transport.

2. A distinction between:

    - the pattern of economic trade in goods from the initial producer to a final consumer, which is termed the *Production/Consumption (P/C) zone pair matrix of trade,* and

    - the actual set of physical transport movements generated by the logistics structure that is used to distribute and transport these P/C trades in practice, which is termed the *Origin/Destination (O/D) zone pair matrix of shipments*

It is only through retaining this distinction within the model, that policies which impact on logistics behaviour can be represented in a coherent fashion with an appropriate spatial representation.

3. An innovative general purpose spatial input-output methodology that could be used to build up the P/C patterns of transport of goods from the original producers to the eventual consumers, whether these be industries or households. This same mathematical methodology could likewise be applied in a linked fashion to build up the actual transport O/D movements for each logistical stage that arises in transporting these goods from the producer to the consumer.

4. A focus on changes through time in the average length of vehicle trips – since this has been the primary past source of freight traffic growth - and not solely on growth in the number of vehicle trips.

5. The adoption of an incremental modelling approach to ensure that the heterogeneity inherent in freight demand patterns can be retained within the models.

6. An approach to modelling the different types of trips that use LGVs, which integrates with freight and passenger modelling and data sources in a consistent fashion.

7. Improvements in methods for local freight modelling, and clear procedures for making best use of forecasts from the national model to improve local estimates of freight demand.

The Terms of Reference for this study included the requirement to review options for modelling at the different levels from the national down to small scheme models. However, subsequent advice from the Steering Group made it clear that (while modelling and structural effectiveness at all levels needed consideration) the first priority was **to focus on the national level model options**. Therefore this report has a strong orientation towards the most promising possibilities for the further development of the freight component (NFTM) of the GB National Transport Model (NTM). It also incorporates model design features within lower-level models, and explicitly considers the modelling requirements for applications below the national level.

The NFTM would constitute the first step in a top-down hierarchical structure and would be able to act as a guide to the lower level models. There is recognition that the treatment of freight within scheme modelling is relatively poor at present and needs significant improvement through this feedback. The dependence of the lower level models on substantive progress at the upper level provides an additional urgency with regard to the practical implementation of the NFTM. Some elements that would be ideally included within the national model may have to be omitted in the short term where lack of data or basic methodological techniques prohibits their early incorporation. This report highlights such elements where they occur, but recommends paths to facilitate their development and integration into the NFTM at the appropriate juncture.

Freight modelling is acknowledged by most of those who have experienced it, to be substantially more complex than passenger transport modelling. It has certainly been under studied in the UK up to the present. Furthermore, freight does not have an equivalent database to the NTS for passengers. The understanding of overall passenger demand patterns available from analysis of the NTS has enabled rapid progress to be generated recently in developing a national passenger model. An equivalent comprehensive database for freight would be more complex to develop, but its absence necessarily slows down the pace of progress in developing a robust national freight model.

Accordingly, there are major challenges that have been described in this report, which need to be addressed subsequent to this study if suitable road and rail freight models are to be put in place in a reasonable timeframe. The fact that the overall framework and the methodological approach by which this framework could be implemented have been successfully outlined above, should not divert attention from the further research, data collection and implementation requirements that are needed to achieve successful models in a timely and cost effective fashion.

In summary, the challenges and risks lie in: the relative complexity of the modelling structure that is required in order to have an adequate representation of the operations and response of the freight markets themselves, coupled with the inadequacy of the existing data sources to facilitate the research and implementation details to underpin some of the important components of the model.

Particular challenges include the following:

a) The comprehensive spatial input-output modelling framework that is recommended, both for representing overall freight demand and the logistics stages of transport within it, will require careful experimentation and validation to ensure that it performs satisfactorily with the limited data that will initially be available to support it.

b) Because of the costs and difficulties in assembling large sample surveys of HGVs, and especially of LGVs, the methodology to develop incremental models based on effective interconnections of link counts, RSIs and survey matrices will prove very important. It will take resources plus some time and experimentation to make sure that it is fully effective.

c) LGV modelling starts from an existing position of having virtually no data on demand patterns, while the sample sizes envisaged for use in the surveys currently in preparation by DfT are relatively small. This implies that it could be many years before sufficiently detailed data would be available to enable a robust model of LGV movements to be calibrated.

However the existence of an overall modelling framework provides a coherent structure within which the development tasks can be formulated. The proposed way ahead would make effective use of the current building blocks already in place, in order to provide short-term substitutes for some of the more challenging elements of development. The multi-pronged approach to creating the national freight model provides time for the research and data collection to be developed that would provide the required support for such elements.

1 INTRODUCTION

The aim of this study commissioned by the ITEA Division of DfT is to review current freight modelling techniques, and through assessing the suitability of the options potentially available, to make recommendations on the most appropriate techniques for use in Great Britain. The review includes road, rail and other freight modes and also includes the modelling of light goods and commercial vehicles (LGVs). It is the freight movements of businesses that are considered in this study. Business trips that are not in LGVs or heavy goods vehicles (HGVs) are not within the scope of the study.

The study is divided into three main work areas summarised as:

Work area A What is needed? - Market structures and issues in freight

Work area B What is available? – Review of models and data

Work area C The way ahead – specification of models and data requirements

In parallel with carrying out the review of existing models it was decided to clarify the specific features of the GB freight system that are of special interest with regard to the aims of this study. The reports already produced in the work area A provide a requirement specification outlining which topics should ideally be included within the models. This has been achieved through analysing freight transport behaviour to identify the major behavioural driving forces and the main freight activities to be distinguished in the models.

This report covers work area C. By contrasting the review of requirements for models, which were presented in the work area A reports, against the review of available models and techniques in the work area B reports, the options for the way ahead are defined within this work area C. It presents an overall freight modelling framework for potential longer-term development, together with more specific detail on the immediate way ahead.

1.1 The aims of this study

The Terms of Reference for this study included (in Paragraph 2.1(i)) the requirement to review options for modelling at five different levels:

(a) National
(b) Urban and Regional
(c) Multimodal Studies

(d) Scheme models
(e) Small scheme models

However, subsequent advice from the Working Group and Steering Group made it clear that (while modelling and structural effectiveness at all levels needed consideration) the first priority was **to focus on the National level model options**. Therefore the current report is primarily oriented towards the most promising possibilities for the development of a national level freight model to satisfy the objectives discussed in the brief. Nevertheless, references are made throughout the report to the incorporation of model design features within lower-level models, and Chapter 8 explicitly considers the modelling requirements for applications below the national level.

The proposed freight component (NFTM) of the GB National Transport Model (NTM) would constitute the first step in a top-down hierarchical structure and would be able to act as a guide to the lower level models. There is recognition that the treatment of freight within scheme modelling is relatively poor at present and needs significant improvement through this feedback. The dependence of the lower level models on substantive progress at the upper level provides an additional urgency with regard to the practical implementation of the NFTM. Some elements that would be ideally included within the National Model may have to be omitted in the short term where lack of data or basic methodological techniques prohibit their early incorporation. This report highlights such elements where they occur, but recommends paths to facilitate their development and integration into the NFTM at the appropriate juncture.

There are two main uses of models that it is helpful to distinguish, though of course they are not wholly independent of each other:

- For a future reference case situation, to make **projections** of traffic supply and demand, based on a series of external assumptions on demographic and economic trends (e.g. NTEM, NRTF).

- To make **conditional policy tests** of some transport initiative (e.g. PLANET, APRIL) – often these tests will be carried out for a future year, because it generally takes time before such an initiative can be made operational.

Detail is given in the brief (see Chapter 9) to determine the minimum acceptable range of policy levers to which the proposed model should be able to respond. Subsequent advice from the Steering Group emphasised that this list was not to be considered exhaustive or prescriptive; instead, **the key requirement for policymakers was flexibility within the model system**. The ability to handle as broad a range of policy options as possible, mediated through an explicit and comprehensive set of model mechanisms and responses (levers), was the primary goal.

The design of this framework has been driven by the requirement for the models to be able to forecast the impacts of a wide range of transport policy initiatives. However, the investigation of suitable assessment systems to use together with these models lies outside the scope of this commission and so is not considered within this report.

This report does not attempt to provide a precise unique blueprint for each step in the freight modelling process. For some modelling stages there may be various appropriate methods by which they could successfully be achieved. This report concentrates on identifying the requirements that need to be satisfied in each of the modelling stages and in demonstrating how these stages may be linked together and made operational. It explains why some methods used in the past are not appropriate and it outlines the relative advantages and disadvantages of the methods that are considered to be of potential use in constructing freight models in GB.

Chapter 2 draws together the main findings from work area A to produce a list of the components that ideally should be included within the modelling framework for freight movements. Chapter 3 reviews the foundations that are common across model components, while Chapter 4 then draws heavily on the work area B to provide an outline of the freight modelling structure itself that would encompass these components. The modelling of LGVs is covered in Chapter 5. Chapters 6 and 7 review the software and data requirements for this modelling system, while Chapter 8 steps down from the national level and discusses how to apply such modelling techniques at the urban and sub-regional scale. Chapter 9 discusses the policy issues that the modelling framework is designed to address. Chapter 10 presents the set of steps to be carried out to implement the set of recommendations of this study, then the final Chapter 11 summarises the main conclusions from this study.

2 THE COMPONENTS AND GENERAL FRAMEWORK FOR A NATIONAL MODEL

The modelling of freight transport has in the past within the UK been less well developed than the modelling of passenger transport. Some topics (e.g.: modelling LGVs or the linking of logistics models with more conventional strategic transport models) have little or no track record of success in the UK. Accordingly there is a need to understand the operation of specific sectors of the freight industry in order to distinguish those actors and aspects of behaviour that should be central to the model, from those aspects that can be averaged out or downplayed when modelling at a specific spatial scale.

This chapter identifies the main ingredients that should be included within a comprehensive freight model at the national scale in Great Britain. It organises these ingredients together into a general modelling framework that provides a suitable order in which to connect them within an operational model. It takes a relatively comprehensive approach in order to develop a broad framework for use in longer term model development (the shorter term tasks are subsequently distinguished in Chapter 10).

Table 2.1 Summary of ingredients in national scale freight modelling framework

Ingredient	**Level of detail in the representation**
Modes	**Full representation**: road (Large, medium, light goods/commercial vehicles), rail (bulk, intermodal, other), including connecting ferries/tunnel to Europe. **Exogenous representation**: waterways, coastal & deep-sea shipping, air, pipeline
Type of freight	Relatively detailed segmentation of commodity type and logistics stage
Space	All of Great Britain – zoning to be defined hierarchically – at least at district level
Time horizon	10 to 15 year time horizon – no explicit representation of short term/cyclic trends
Responses to policy	Changes in length of haul, zone pair distribution, mode, route, vehicle size, distribution centre location/structure, handling factors, loading factors (shipment size/number of drops/empty running)[1]. **Possible**: Changes in volume of trade, in location of firms, time of day
Service characteristics	Monetary cost, travel time, reliability/quality/flexibility, capacity
Infrastructure/ network supply characteristics	Distance, charges, speed restrictions, capacity, allowed vehicles, traffic load/congestion, reliability, terminals
Vehicle characteristics	Pollutant emission rates, fuel type, operating cost structure, fleet size, vehicle size - carrying capacity and congestion factor
Usage of model	• Projections of future traffic • Conditional testing of policy initiatives (in both base and future years)

[1] See Appendix I for a definition of these terms.

This current chapter concentrates on explaining why the particular elements that are discussed need to be included within the freight modelling framework and so is based on the results of work area A. The methodology through which each element can actually be modelled is then outlined in Chapter 4, which is strongly based on the results of work area B. The main ingredients envisaged for inclusion within the model are summarised in Table 2.1.

The term "exogenous representation" used for certain modes in Table 2.1 implies that some independent external modelling or scenario building exercise would estimate their future volumes of movement at the GB terminals (ports and airports). The resulting domestic legs on road and rail to and from these terminals would then be included explicitly within the freight model.

2.1 Background to work area A

The aim of the work area A has been to provide a clear understanding of the primary factors that influence the behaviour of the various actors within the freight transport system. This is an important input to the design of review of existing models and techniques in work area B, since it provides a foundation that enables:

- any important gaps to be identified for which suitable techniques are currently unavailable,

- an assessment of the likely usefulness of individual models and techniques, through checking the extent to which they address the behavioural relationships and trends that are of greatest importance in practice in Great Britain.

Work area A considers the problem of representing the freight transport industry and its wider context, within the limitations of a computational model supplied with imperfect data. It poses the question, "What is needed?"

Conventionally, there are three fundamental objectives. A model should:

- Provide a realistic micro-economic framework – a behavioural foundation to underpin the responses that are forecast by the model.

- Anticipate the policy or industry initiatives – provide a means to introduce these external inputs into the model.

- Establish valid links between policy and the responses of the markets.

2.2 Overall framework for the model

The work of the REDEFINE project (NEI et al., 1999) provides a suitable starting point from which the design of a framework for GB freight modelling can be developed. Figure 2.1 presents the framework developed by REDEFINE. It shows the relationship between, at the top of the Figure, the value of the goods produced, through various stages, down to the eventual estimates of the road freight traffic demand, shown at the bottom of the Figure. The elements on the right hand side of the Figure are a series of "key ratios"[2] that represent measurable indicators, which taken together encompass the main influences on the future levels of road freight vehicle movements. The boxes on the left hand side denote the quantities that are forecast. They are measured in appropriate units for each stage: the top box is in monetary units, the next three in tonnes, then tonne-kilometres and finally vehicle kilometres. This highlights the need to consider different dimensions at different stages of a freight forecasting procedure.

If none of these key ratios were to change, then future road freight traffic levels would be perfectly correlated with changes in the value of the goods produced by the national economy. In practice, however, some or all of these key ratios may vary, either due to exogenous trends through time or in response to specific policy initiatives. These in turn would modify the future road freight traffic patterns.

The purpose of this current study is to design a framework for freight modelling in GB, and so to achieve this it is necessary to extend this REDEFINE framework in two main ways:

1. to adopt a multimodal approach (the orientation of REDEFINE was solely to road traffic)

2. to introduce an explicit detailed spatial dimension (REDEFINE focused on the aggregate country level estimates).

Figure 2.2 summarises this extended framework. It introduces these extensions through various changes both to the order of the stages and to the connections between these stages. In particular the modal split stage was moved to below the spatial distribution stages in which the tonne-kilometres would be estimated. The shaded boxes represent model stages while the unshaded boxes represent the main input and output data files that interconnect these model stages. In the sections in the rest of this Chapter each of the main stages within the framework of Figure 2.2 is discussed in order from the top to the bottom of the Figure. Each section explains what features the set of boxes represents and why they need to be explicitly represented within the modelling framework.

---

[2] See Appendix I for more detail on the REDEFINE approach and for definitions of these key indicators.

Figure 2.1 Framework to link economic activity to road freight traffic (REDEFINE)

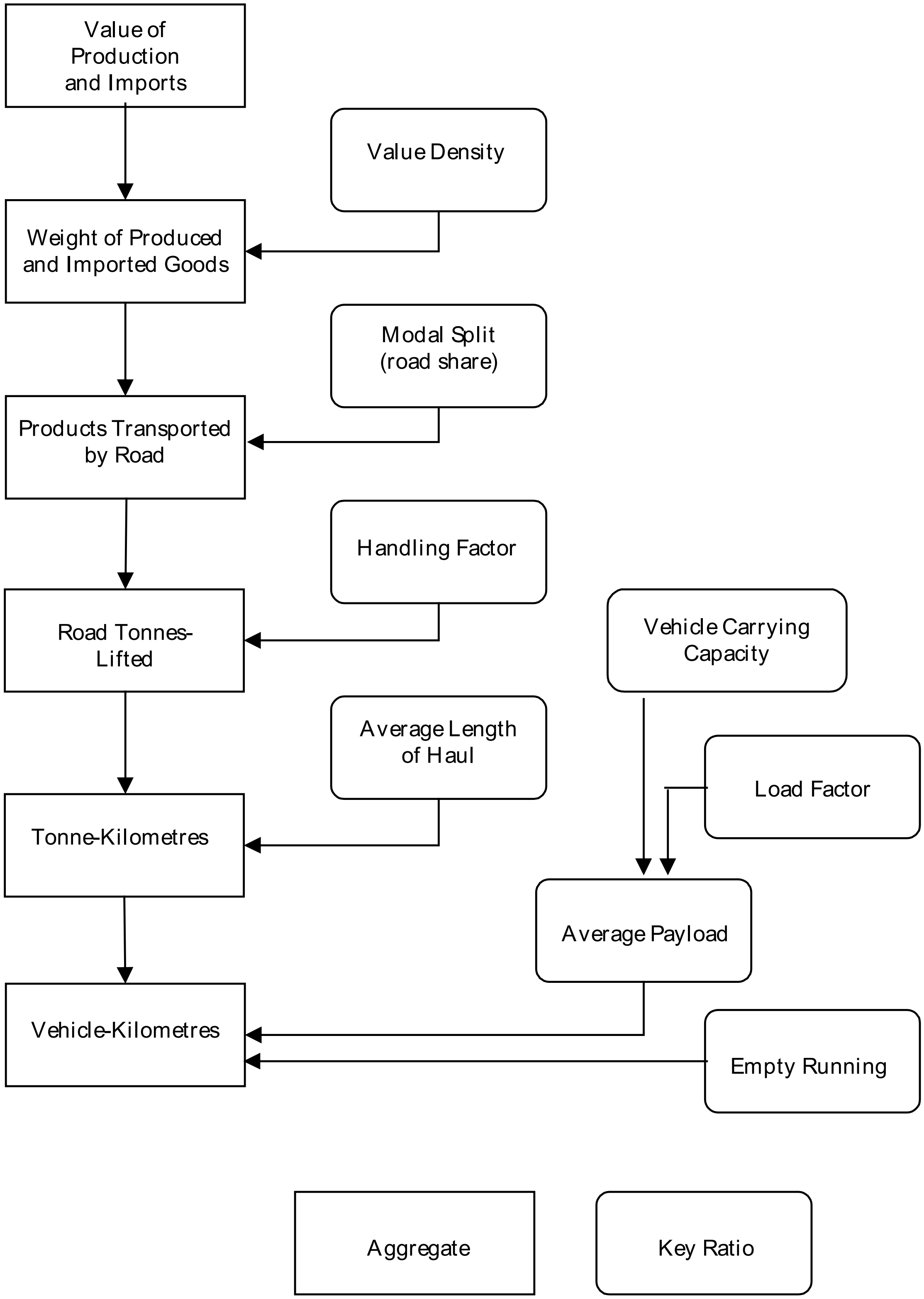
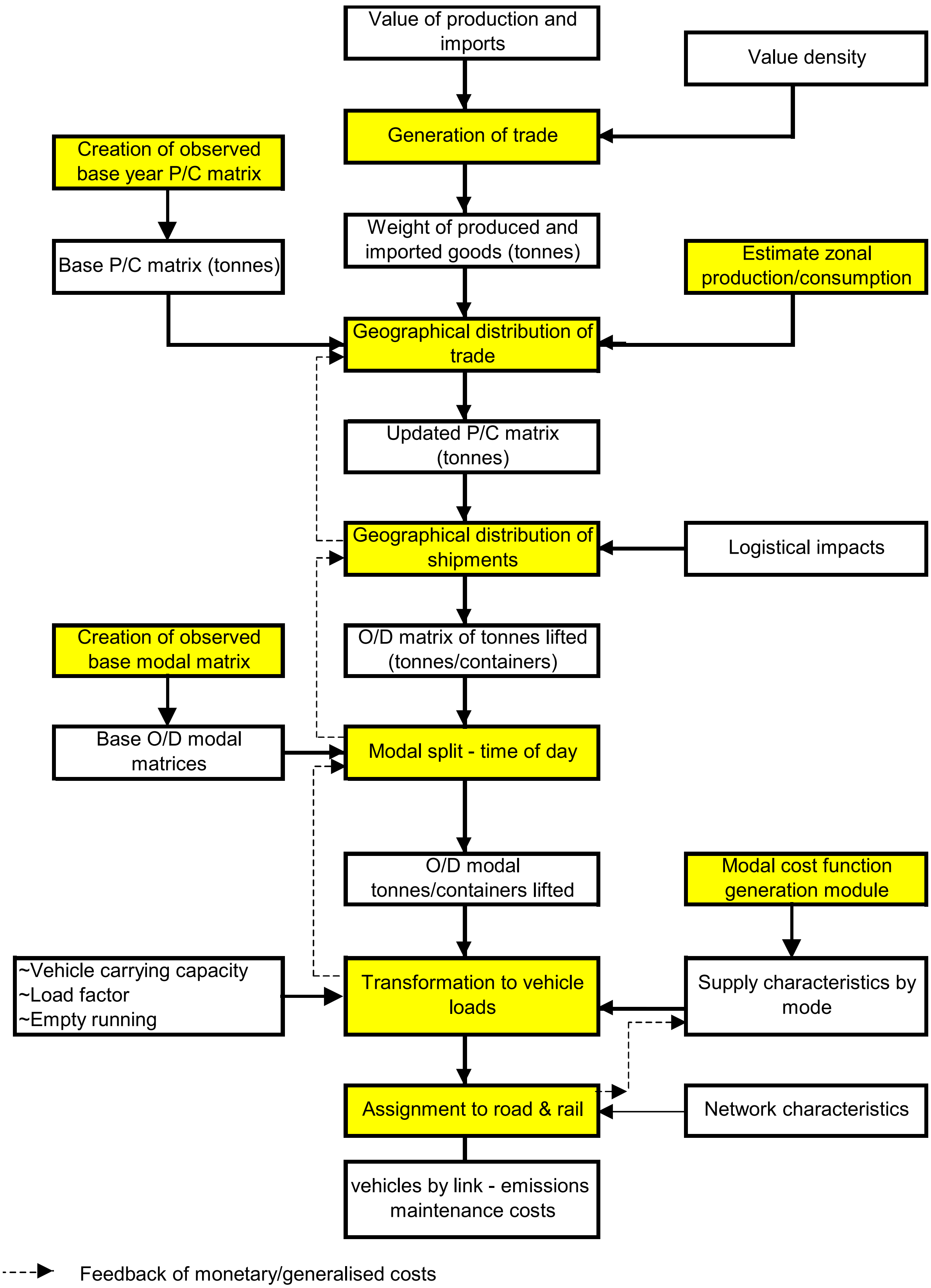


Figure 2.2 Framework to link economic activity to freight traffic on modal networks



The purpose of this extended framework is to provide an *idealised* overview of the modelling stages that would be required in constructing a set of freight models for GB. It does not imply that it is currently feasible to tackle all of these steps to the desired level of detail within a single operational model. Instead it provides a unifying framework within which the different modelling elements could potentially be integrated in a modular fashion. In this manner, improvements to certain stages could be introduced at a later date if necessary, without invalidating the coherence of the overall design.

By comparing the various stages shown in Figure 2.2 against the list of existing models reviewed in work area B, the framework can also be used to illustrate those stages where there may be potential weaknesses in current experience in model construction or data availability. The framework is set at a fairly general level, which means that a number of different technical options may exist for modelling a specific stage within this framework. The eventual choice as to which modelling option to adopt within a specific stage would be dependent on:

- the man-power and computing resources available for model development and running
- the range of data available, its consistency and its accessibility
- the experience, preferences and software availability (assuming that the software was suitable) of the team implementing this stage
- the need to interface with the other prior and subsequent stages of the overall modelling framework.

The segmentation of freight is not made explicit for the stages of the model shown in Figure 2.2. For reasons of computational efficiency and data availability it is likely that this segmentation will differ between different model stages. At the top stage a rather detailed commodity classification may be required, whereas at the assignment stage a very limited separation between bulks, medium value and high value goods may suffice. The segmentation into the three assignment streams shown is purely an option and explicit assignment to e.g. rail may not be required. The appropriate segmentation to adopt for each of the separate model stages is discussed in Section 3.2 below.

2.3 Forecasting trade patterns – the value and volume of freight

The demand for freight shipments does not exist in a vacuum. Transport is a *derived demand* that results from the operation of the economy as a whole. There is a strong but complex linkage between the performance of specific economic sectors and the level of freight traffic. A logical approach to the estimation of the future level of freight demand is to derive this demand from expectations on how the economy itself will evolve.

There is a significant body of economic research and forecasting on the future evolution of the UK economy and so it appears worthwhile to maximise the use that can be made of this knowledge. Estimates of the future growth in the different sectors of the economy (including the balance between final consumption, intermediate consumption, imports and exports) could be used to estimate the future value of trade by commodity type in each of the economic sectors. The modelling work of Cambridge Econometrics which adopts a detailed industrial segmentation and which includes regional breakdowns of the forecasts of production and consumption would be an approach that would appear to meet many of the requirements.

The work area B1 has reviewed in detail the series of Commodity Flow Studies initiated in 1975 by the Department of the Environment, which are summarised in the review report (Pike and Gandham, 1980). These studies have identified the most important explanatory variable in relation to the future generation and length of freight movements in the commodities studied, and have classified them under six main driving forces, namely:

1. **Economic**: demand factors arising out of the performance of the UK and, sometimes, the international economy.

2. **Industrial**: decisions taken by firms affecting the structure of their industry, output levels and product mix.

3. **Political**: actions of central government and EEC affecting the structure and conduct of an individual industry in general.

4. **Technological**: technical innovation and development leading to changes in methods of production.

5. **Social**: changes in social habits leading to altered patterns of consumption and production.

6. **Climatic**: effect of weather upon customer demand and agricultural output.

The key issue to note here is that the list does not stop after item 1. Although economic modelling is an important ingredient in forecasting the future level of demand for freight transport, there are a variety of other equally important factors that will play an important role, particularly at the local scale. Unhelpfully, a number of these factors do not easily lend themselves to being forecast with any confidence. This should influence the degree of certainty that can be placed on future forecasts of freight demand.

There is a requirement to forecast future growth in both the volume moved and in the average distance over which this volume is moved. These have quite separate trends through time as discussed in the Inception Report. Tonnes lifted have had a strong cyclic pattern but only a modest growth trend over the last thirty years. In contrast average lengths of haul have increased consistently throughout this period. Similar long term patterns through time of relative stability (having excluded the cyclic pattern) in tonnes lifted and of consistent growth in tonnes moved have also been evident in many other EU countries.

However, just because lengths of haul have grown consistently in the past thirty years across the EU it does not automatically imply that this will continue *in perpetuity* into the future. Similarly, the absence of a strong past trend in overall growth in tonnes lifted, does not guarantee similar stability in the future. **This is a fundamental topic on which research to increase understanding is urgently needed**. This research should aim to understand:

- the relative importance of each of the various factors that have combined to generate this past growth in freight lengths of haul and tonnes lifted, and

- the extent to which each of these factors is likely to continue into the future to accelerate or to diminish the growth in freight lengths of haul and tonnes lifted.

This research study should pay particular heed to the findings of the Commodity Flow Studies that were introduced above. It should include consideration of issues such as:

- supply effects – increases in road speed and reliability due to the motorway network expansion, reorganisation of rail services, trends in unit transport costs per tonne kilometre by mode

- industrial demand/structural economic effects – greater product specialisation, economies of scale, logistics developments, handling factors, outsourcing

- economic trends in domestic and in international freight – globalisation

- Detailed segmentation by type of commodity – how important is the growth in length of haul in each individual commodity type and how important is the change in the overall mix of freight that is moved. In Great Britain the finished goods, which tend to travel the longest distances, have increased in tonnage, whereas the bulk goods, which frequently are only shipped over shorter distances (other than by sea), have had little or no growth in tonnage.

This research would enable one of the weakest features of the NRTF 1997 freight model to be improved and is of major importance in improving confidence in the ability to carry out longer term forecasting of freight demand. Ideally this research should distinguish the specific policy levers that could impact on this trip lengthening process, because it is a major determinant of traffic growth.

2.3.1 Volume to value conversions

Past experience has demonstrated that it is unwise to expect for any specific commodity that the relationship of its volume in tonnes to its value in constant prices will remain unchanged through time. Accordingly, any value-based forecasts of the future growth rates in the production, imports or exports of a commodity would need to be converted from value units to volume units before being used to generate the growth in physical movements. In some of the bulk goods sectors (such as extraction of minerals, coal and petroleum products) future estimates of production and/or consumption may already be available in physical volume units and so could be used directly. However, in the finished products sectors, such as manufactured goods, foodstuffs, pharmaceuticals, etc. it is likely that any forecasts would be in value rather than volume terms.

There do not appear to be either comprehensive data sources or a strong track record of research within the transport field on the evolution of value density trends (i.e. volume to value ratios) for different types of commodities. It is a topic which has been identified as one of the most sensitive influences on freight vehicle forecasting. **Value density is a topic in which future effort on data collection and analysis is needed**. The data that has been available in the past has mainly been for imports and export goods, but these tend often to not be representative of the characteristics of the domestic goods within the same commodity class (e.g. imported grapes and wine versus local potatoes and beer). However, trade data does provide a considerable source of information about value to volume ratios and their trends. Exports, measured in tonnes and £s should provide a reasonable basis for analysing UK production.

Uncertainty still exists over the merits of including within the modelling framework this step of explicitly deriving future freight demand patterns from future economic forecasts in value terms. On the one-hand:

1. There is no doubt that the demand for freight is governed strongly by the state of the economy

2. SACTRA (1999) has emphasised that there would be a real benefit within transport assessment to being able to measure the impact on the economy of policy initiatives in the

freight transport sector. The linkage described above would help in achieving this (acknowledging that there are further important issues identified by SACTRA that would also need to be resolved to achieve this measurement)

However on the other hand:

3. Many economists have been hesitant about the reliability of forecasts of trade patterns by detailed commodity over the long term periods of 15 years into the future that are the norm for transport planning.

4. Forecasting the future evolution of the volume to value ratios for each commodity type is a challenging task, which is not made easier by the apparent absence of comprehensive past or even present data on this ratio on which to base time series analysis of trends.

The issue is not whether in principle this economic forecasting step should be included in the modelling framework - *in principle* it definitely should. But rather the issue is one of *practicality*. Are the data and the modelling experiences that are currently available sufficient to guarantee robust and realistic results from this step? Would it provide more reliable forecasts than those that might be produced by alternative forecasting methods that are based solely on trends in physical units (tonnes) rather than on the monetary value of trade?

On balance it appears worthwhile to proceed ahead with caution to develop methods of forecasting freight demand that are tied to forecasts of the growth in the value of the economy. However, further research into this topic would be beneficial, and closer links to the production of economic data would be one aspect of this work, as discussed further in Section 7.3.

2.4 The initial production to consumption zone matrix

In order to understand the relationship between transport and the economy, it is helpful to draw a distinction between two separate matrices of movements that are used in the analysis:

- The pattern of economic trade in commodities from the initial producer to a final consumer is termed the *Production/Consumption (P/C) zone pair matrix of trade* and is discussed in this section. Changes in this matrix are strongly influenced by changes *outside* the transport and distribution sectors. The question of whether these trades are denominated in monetary value or in physical tonnes is left open here (in accordance with the principle of this report to present a number of options within a high-level design framework, rather than be unnecessarily prescriptive).

- The actual set of physical transport movements generated by the logistics structure that is used to distribute and transport these P/C trades in practice is termed the *Origin/Destination (O/D) zone pair matrix of shipments* and is discussed in Section 2.6. Changes in this matrix are strongly influenced by changes *within* the transport and distribution sectors.

The analysis of freight and logistics that was reported in the work area A has already highlighted that international trade is growing much more rapidly than production itself. Changes in industrial structures, the sourcing of inputs and the marketing of outputs, have all tended to lead to trade being shipped over longer distances than before. Clearly, these changes have been facilitated both by past improvements in quality, and reductions in the unit costs of freight transport. However, many industrial and financial influences that are external to the transport and distribution sectors also strongly influence them.

The reason for having a P/C matrix in addition to an O/D matrix for the same commodity is that the impact of trends in logistics and of the responsiveness of logistics to policy initiatives can best be modelled in a realistic fashion if these impacts are applied to a P/C matrix. For example, the lengthening of P/C trades does not automatically imply a lengthening of O/D shipment lengths since there may be a succession of separate intermediate shipment legs between the initial production and the final consumption of a good, with the good being warehoused in distribution centres between these legs. An industry led policy to increase the number of distribution centres could, perhaps, generate a larger number of shorter legs even at a time when there is an overall lengthening of trade. However, in the past the broad trend has certainly been towards a smaller number of larger distribution centres, which has had the effect of increasing the average length of haul on O/D shipments.

Although in principle the future year P/C matrix could be synthesised, in practice there are sound reasons to construct it as much as possible on any good observed data sources that are available from the base year. Because of the greater heterogeneity inherent in freight movements than in passenger movements it is crucial to include the major observed peculiarities in the patterns of movement within the base matrices used in the model. Even when a major producer and a consumer of the same commodity type are located within the same zone, it is not guaranteed that there will be a major freight movement from one to the other. Either if they are in rival firms or if the detailed characteristics of the commodity that is produced imply that it does not match that which is consumed, then there may be little or no flow between them and the actual trades may take place over much longer distances.

In order to take account of such local peculiarities, the base P/C matrix of trade for each type of commodity should be derived to the maximum extent possible from good observed matrices of movements. It would be constructed in physical units – tonnes for most segments, but may best be in units of containers for some segments, particularly the import and export of higher value finished goods. The means by which this initial base P/C matrix is created is outlined in Section 2.6.1. This base matrix would then be used as one of the foundations for the estimation of future year P/C matrices and of the impacts of policy initiatives on patterns of demand within these matrices.

2.5 Updating the production to consumption zone matrix

The traditional procedure in past widespread use in GB of factoring up base year matrices of goods vehicle trips by a *single* overall measure of *traffic growth* has been identified as being "fundamentally flawed" by SACTRA (1999). Applying such a measure based on the change in *vehicle-kilometres* in order to scale up matrices of *vehicles* is far from ideal.

Zonal estimates of the future levels of production and consumption of commodities would be used as row and column totals to provide future estimates of the P/C matrix by type of commodity. These zonal estimates would be based on the future spatial pattern of location of firms/employees (see Section 2.5.1) and on the future overall levels of growth in production and consumption for each type of commodity that have been estimated in the top stage of the modelling framework (see, Section 2.3 above).

This matrix estimation procedure should also take account of the strong trend to increases in the length of trade in many commodity categories, as well as the extent to which this is likely to be influenced by future changes in transport costs and characteristics. This implies that the procedure should not be a simple growth factor or maximum entropy based matrix expansion since these procedures do not directly address the issue of the lengthening of vehicle trips, which has contributed significantly more to the growth in traffic than increases in the number of vehicle trips as shown by the CSRGT (DTLR, 2001).

Hence it would be preferable to adopt a methodology that can respect the main features of the observed base year pattern of movement, while adjusting to take account both of structural economic/industrial trends (most influential for high value goods) and of the influence of transport and logistics costs on the spatial distribution patterns of trade in the P/C matrix (most influential for low value goods).

When modelling road freight for relatively small scale local studies, in the interests of simplicity it may be necessary to ignore the impacts of trip lengthening. Even here however, it should be borne in mind that long distance trips tend to spend most of their journey on the trunk road network, whereas short distance trips will spend much of their journey on the minor road network. Accordingly, the differential growth in trip lengths has significant impacts on the local growth rates on specific types of roads.

2.5.1 Zonal estimates of tonnes produced and consumed

The estimation of the detailed zonal pattern of production of a specific commodity would rely on such zonal statistics as are available to identify the locations where, and at what scale, production takes place. Direct information at a detailed spatial scale on the transport or production volumes is not common, so that instead reliance would be placed on alternative indirect sources of information that have been discussed in the data review in Task B4. The two most promising potential sources are:

- employment data by SIC from the Inter Departmental Business Register (IDBR)
- Rating List and floorspace data by type of property from the Valuation Office Agency.

Zonal estimates of the consumption of commodities are similarly required. These are particularly difficult to estimate based on current transport data, since at present there appears to be an absence of clear, comprehensive, spatial information on consumption rates and patterns of freight tonnes lifted and dropped.

There is a compelling need for good observed data to link the volume of commodities lifted and dropped to the type and scale of activity that produces or consumes these goods. This information would be used in an analogous fashion to the National Travel Survey (NTS) data for passengers, which has been used to develop unit trip generation and attraction rates by trip purpose for different types of person/employee. It would be of immense value to obtain analogous estimates per employee of each industry type Y, of the average number of tonnes lifted and of tonnes dropped for each commodity type X. **A data collection and analysis exercise to generate such estimates would be of high value.** A large sample size might be required since previous work in the USA has found that such estimates have very high variances. Furthermore, since many goods are delivered in palletised units, knowledge of the contents is often not available to the carrier so there will be limits to the detail in commodity types that can be obtained.

One alternative possible source of existing information is from the analyses underlying the construction of input/output (I/O) tables of the economy, which are based on the Annual Business Enquiries and related sources ONS (1997). I/O tables[3] contain indirect information for each economic sector on the proportion of its production that is consumed by each industrial sector, the proportion that is exported, and the proportion that goes to meet the final demand by households, investments etc.

---

[3] See Appendix III for more detail on I/O tables and their data sources.

This, in principle, could provide a foundation for the local estimation of zonal consumption rates. However, there are certain complications that need to be considered in any such approach based on input/output data sources.

1. This information is in units of monetary value rather than physical volume so that it would require the application of the value to volume ratios, discussed above, to convert it into the dimension of tonnes per employee that is required. One possible solution would be to obtain volume as well value data within the underlying surveys, though the feasibility in practice of this may be questionable.

2. The categorisation of this I/O data is in terms of detailed economic rather than transport oriented sectors and so some transformation and aggregation would be required to convert the trade of these sectors into commodity categories that are consistent with definitions used in the CSRGT/rail freight data sources. As explained in Appendix III.2.1 the commodity based rather than the economic sector based categorisation of I/O data is the more appropriate for freight transport modelling purposes.

3. The spatial detail within the I/O tables is very limited. For the most part it is only available at the national level, with some limited further information available via the regional accounts for the UK. Accordingly, strong assumptions of homogeneity in consumption rates across area types and regions would be required, since any survey program to collect the data needed to overcome this lack of spatial differentiation problem would be prohibitively expensive.

4. The degree of access that would be available to the background analyses used in the creation of the I/O tables is unknown. It appears likely that the published I/O tables themselves are as much as may be available. Accordingly, this tantalisingly attractive data source may ultimately prove unhelpful.

Having created reasonable estimates for each commodity type of both the number of tonnes produced and the number of tonnes consumed in each zone, there is still the significant task of converting this into a P/C matrix that is based as closely as possible on observed data on freight movements. This task involves unravelling the observed O/D matrices of shipments, back to the actual P/C zone pair trade that generated these movements. The broad principles of this task are discussed further in Section 2.6 and the overall design of the methodology to be used to carry out the updating of this P/C matrix is outlined in Section 4.5.

2.6 Distribution and handling factors – the O/D matrix of tonnes lifted

The Origin/Destination (O/D) zone pair matrix of shipments is estimated through applying the influence of logistics and commercial distribution activities to the basic pattern of trade in the P/C matrix.

An example will illustrate the task. One sale of office supplies from a producer in zone A to a consumer in zone E may give rise in practice to a number of separate shipment legs:

1. from the original factory in zone A to the producer's regional warehouse in zone B

2. from the producer's warehouse to the wholesaler's warehouse in zone C

3. from the wholesaler's warehouse to the retailer in zone D

4. from the retailer to the office where the supplies are eventually consumed in zone E.

In this case the single trade has a *handling factor* of 4, in that it gives rise to four individual O/D shipments, which would each figure separately in the transport statistics as: tonnes lifted and tonnes dropped in each of the zones in which it has stopped off. Furthermore, the sum of the distances from A to B to C to D to E, may be substantially more than that from A to E directly, so that this also will increase the number of tonne-kilometres transported.

Most freight demand data that is collected in transport surveys relates actually to O/D matrices, rather than to P/C matrices. **There are not adequate comprehensive national sources of data that facilitate relating the legs of these O/D matrices back to the original P/C matrices from which they derive, so further data collection and research is needed** (see Section 7.1 for details). There is still the caveat that many of the patterns observed would be company-specific. However, even a simple synthetic approach that is informed and validated by a sufficient quantity of actual observed data would be valuable.

Observed O/D data for most commodity types show that increases in the average length of haul of shipments are a much greater source of traffic increase than increases in the number of tonnes lifted. This has been true in the UK (CSRGT: DTLR 2001) and in other European countries (REDEFINE: NEI et al, 1999). One obvious conclusion to be drawn from the rapid past growth in average length of haul is that the use of a "fixed matrix" modelling methodology, based on factoring up of base year freight movements to match future row and column totals, will be of limited validity except over the relatively short term future. The influence of changing trade patterns and of changing logistics structures cannot safely be ignored, except in short-term or localised studies.

In order to provide a coherent logical structure for modelling the impacts of logistics developments on freight demand, it will be necessary to develop a module that converts the P/C matrix into an O/D matrix in a manner that takes appropriate account of the location of the distribution centres that are relevant to particular logistical families of goods. In this way, policies that impact on logistics costs and organisation will feed through to influence the patterns of freight transport demand on the network. The impact of such policies will feed from their initial application via the cost functions developed for the networks, vehicles and terminals, through to the generalised costs of different logistics options for the shipment of goods from a producer in a zone to a consumer in another zone. The choices could include:

- direct door-to-door shipment from the producer to the ultimate consumer,
- use of a single intermediate warehouse,
- use of multiple warehousing for different agents within the system, etc.

The choice procedures here could be represented using the same methodological approaches that have been used for modal choice.

2.6.1 Creating the base P/C matrix

The initial creation of the base year P/C matrix in effect would involve reversing the procedure described above. It would commence from the observed base year O/D zone pair matrix of shipments in units of tonnes (or in containers for some classes of goods) derived from the CSRGT, rail, port, airport and other sources of statistics on freight movements, segmented by commodity type.

The following sequence of steps would then be required:

- *Merge intermodal movements*: connect together into a single movement the individual modal legs of intermodal shipments. For example, a container of metal products moved by road from Sheffield to Doncaster rail freight terminal, then from there on rail to Felixstowe, and then on ship to North America should be converted from these three separate modal legs (each identified in separate modal statistical sources) into a single intermodal export movement.

- *Merge legs of the distribution chain*: connect together into a single movement from the initial producer to the final consumer the set of individual shipment legs via distribution centres or other intermediate storage locations. These shipment legs may not necessarily all be on the same mode. For example, high volume movements of bulk goods between a large producer and their major distribution centre may well be economic to send by rail using direct rail sidings at each end of the shipment. However, dispersed low volume movements from this distribution centre to final consumers would be on road in almost all cases. The multi-drop pattern of distribution to final consumers also needs to be taken account of at this stage.

The distinction between the transport activity and the storage and distribution activity is one which is maintained in the estimation of value added within the official National Accounts data sources (ONS, 1997). In some instances where goods are stored as part of a change of mode the distinction may be ambiguous. However, this ambiguity does not complicate the execution of the two steps above.

Although the principles underlying these two steps may appear straightforward, the data required to implement them in practice is at present far from ideal. These implementation issues are addressed in Section 4.6.

2.7 Estimation of cost functions and supply characteristics

The analysis in Task A4 of the rail freight system has highlighted the need to base the behavioural responses within the model on realistic estimates of the costs that face the actors within the transport system. Likewise, the main mechanism by which government policy measures will actually influence the operation of the transport system is through the responses of actors to changes in the direct or indirect operating costs of transport that they face (or the tariffs, in cases where these are not representative of the underlying operating costs).

This implies that the modelling system that is designed here should include an explicit module in which the various operating costs of the freight transport system for each modal service being offered can be built up. This would be sub-divided into their main component parts, namely: track/network infrastructure, vehicles (both traction and wagons/trailers/containers) and terminals. A representation of the operating costs of distribution/storage centres is also required. The valuation placed on goods in transit, based on the capital charges involved for the owners of the goods, also needs to be included. In this way changes in government policy, whether they be regulation, taxation/subsidy or investment, can be traced through to calculate their implications in terms of the costs of use of the different parts of the transport and distribution system. These cost changes in turn, once they have been aggregated to a suitable level, would feed into the model to influence the behaviour of producers, consumers, shippers and vehicle operators in terms of their choices at the various stages of the model. The changes in the organisation of firms within the transport and logistics sectors, likewise, can result in changes in operating costs that may influence transport choices.

The issue of the quality of the service on offer, and not just its monetary price, is also important. There is ample evidence that the choice of service (particularly for higher valued, finished goods) is influenced by characteristics such as security in transit, guaranteed delivery times, rapid response, etc.

The paper by Bruzelius (2001) provides a recent survey of results from a variety of empirical studies in different countries on the valuation of freight transport time and quality factors. He discusses the considerable difficulties that are faced in deriving robust results from revealed- and stated-preference studies. The intrusion of logistics factors into stated preference valuation approaches in the automotive industry supply chain Australia has been noted by Wigan (2001). There is a significant body of UK evidence measuring the impacts of freight service quality characteristics available from the studies carried out in ITS, Leeds over the last decade (e.g. Fowkes et al., 1993) which could be used to introduce them into modelling procedures in a suitable manner.

2.8 Modal split of tonnes and containers

Whereas in the original REDEFINE framework in Figure 2.1, the modal split is higher up the chain of modelling stages than handling factors, in the extended framework in use here (Figure 2.2) it is placed after the distribution/handling factor stage. This is primarily to take account of the influence of the stage in the distribution chain when estimating the modal split.

One of the major influences on rail's ability to compete with road lies in the size and regularity of the shipments to be moved. If large volumes need to be move regularly from a single origin location to a single destination location then rail can be cost-effective, even when the distances are not so long. The costs of private sidings can be spread over a sufficient load to make them economic, and the costs of organising and providing the rail service can be streamlined, especially when full trainloads are being moved. Alternatively when there are irregular shipments to a wide variety of destination locations it is unlikely that rail can be cost effective except over very long distances, given the inevitable need to organise the final delivery to these locations to be by road and given the costs and complexities that this extra stage may entail. This highlights that whereas rail may be able to compete strongly on movements between ports and major distribution centres, or from a major distribution centre in Europe to one in GB, its ability to compete with road on shipments to retailers or to final consumers will be minimal, due to their diverse locations and small drop sizes for shipments.

Therefore, there is a need to identify the stage in the distribution chain as a dimension to be considered as a function of industry characteristics, prior to the application of the modal split procedure. Indeed, broad groups of commodities at a similar stage in the distribution chain may lose little from being aggregated together, prior to the application of the mode split procedure, so that computing resources could be freed up for use elsewhere. The typical magnitude and regularity of shipments between pairs of points may also be a dimension to include in the segmentation (or through some other means) when carrying out the modal split. However, this would not necessarily apply to separate establishments aggregated within zones. Closely grouped competitors might in total have large enough flows to make some mode other than truck viable, but they will only rarely consolidate loads on a single carrier. This raises an interesting paradox, in that several firms in one zone might collectively give rise to large flows, but in which none of them ship enough volume to consider anything other than motor freight.

Because of the intrinsic heterogeneity of freight, there are advantages at the modal split stage to avoid a synthetic matrix approach to modal split and instead to use some form of incremental (or marginal modelling) approach that takes direct account of the current observed pattern of freight transport by mode. As has been previously discussed with respect to the spatial distribution modelling stage, the incremental approach needs sufficient sophistication to enable the spatial balance of movements to evolve through future years in line with trends towards increases in lengths of haul. The danger with the incremental approach is that if it is too crude it may struggle to take sufficient account of major changes in the costs or characteristics of a mode.

An incremental modelling approach would also take account of the fact that there is a large part of the freight market, which comprises short distance movements of small shipments to dispersed locations, in which rail is unlikely to be able to compete cost effectively. It is important from a modelling point of view that that those components of the overall freight market in which rail has a possibility of competing are not averaged in with other components within which rail would not compete. Given the very large share of road freight compared to rail, quite small percentage changes on road could lead to apparently major percentage growth on rail (e.g. a 5% switch in road tonne kilometres would create a 50% increase in rail tonne kilometres). Hence there is a need to carry out the modal split at an appropriate level of segmentation within which each of the main markets in which rail can compete is separated out. This would separate out the main movements of the bulk flows as well as the intermodal traffic and a few other segments in which mode split between road and rail would be modelled explicitly. Care would be needed to ensure that the estimated pattern of rail share was correctly concentrated on the longer distance movements within each of these segments.

The output of this process would be a set of matrices, segmented by aggregate commodity/logistics/shipment regularity type and OD pair, that are moved wholly by road, wholly by rail or as intermodal traffic. In the latter the longer part of the journey is by rail, but with collection and/or distribution at one or both ends of the movement being by significant road stages. The road and rail components of these intermodal movements should be correctly accounted back to the networks of their parent modes when estimating the system-wide statistics on transport supply and demand.

It must be stressed that the *mode* referred to here and elsewhere in this report is the ***main mode***, that is the mode for which the primary operating characteristics and costs determine the selection of means of shipment. The actual modelled path might be multimodal (depending on the detailed implementation and software platform). Thus a ferry leg for road transport would simply be part of a multimodal path of *main* mode *road*, but road freight costs and characteristics would predominate in the choice mechanism. Similarly, a main mode *rail* trip might contain road feeder legs to and from rail terminals and indeed include a rail ferry leg.

2.8.1 Time of day choice

Within the road mode there is a further issue of the time of day of the movement. The hourly pattern of freight vehicles on the road network contains significant differences through the day from that of passengers. Road freight movements in the early morning are numerous, then in congested areas they may tail off during the morning peak, and then increase again in the inter-peak period. By the time of the evening peak and subsequently, the road freight volumes are at a much lower level than earlier in the day. Some of these patterns are caused by a wish to avoid congested running (drivers may break for breakfast). Some are caused by the need to carry out deliveries early in the day, and some are a consequence of enterprises such as the supermarket chains having switched to 24-hour distribution that entails the use of the vehicles throughout this 24-hour period.

Certain policy issues are directly related to the time period of travel. Urban areas may be subject to lorry bans at night or to bans on loading/unloading in the middle of the day. Parts of the Continent ban goods vehicle movements on Sundays. Major improvements in logistics efficiency have been gained in some sectors through reductions in the fixed cost component of vehicles, purely through intensifying their usage of vehicles from 8 hours to 24 hours per day.

In order to model the impacts of such policies, **research would be required into the valuation placed on "turn-up-and-go" access to the network. What is the cost of delaying a collection or delivery, and what are the determinants?** Then there is the issue of different cities having different loading/delivery times, and also the links between un-scheduled networks (road) interfacing with scheduled networks. This could be a major, but potentially valuable, undertaking.

For rail the choice of the time at which goods are shipped is dependent on the scheduled train patterns and on other aspects that are likely to be largely outside the control of the shipper. Knowing when a delivery is required is important in calculating the generalised cost of a rail service with fixed departure/arrival times, and hence for determining mode-split. Likewise, SRA need to know whether to push for maintenance to be carried out at night as opposed to at weekends. Spending on rail diversionary routes needs to be justified by the need to move consignments at particular times on particular days.

Care would be needed in segmenting trips between time periods when a trip is longer than the time period being modelled, or when a tour covers many periods of the day.

2.9 Transformation to vehicles

Some segmentation of road vehicles by size would be highly desirable in order to take appropriate account of their impacts: on congestion, on pollution emissions, on the operating costs of the vehicles, and of the impact of policies that might ban certain types of vehicles from certain locations or types of road. This also would tie in with the need to represent the LGV market and its contribution to the carriage and delivery of freight, and the overlaps and interactions between vehicles above and below the 3.5 tonne gross vehicle weight (gvw) cut-off. These are particularly important when commodity value densities rise - an established trend in many areas of consumer goods and electronics. Some form of vehicle stock model should be used to provide a profile of the fleet composition and to enable suitable aggregate vehicle operating cost, and pollution emission functions to be calculated for future years. In past years there has been a trend for the fleet of medium sized lorries to decline while the fleets of the smaller and larger vehicles have increased.

An explicit sub-modal split between lorries by size is one option that can be considered. Alterations in loading bans, changes in time widows by major receivers, and changes in access regulations are all policies which affect the total logistics decisions, and can also alter the types of vehicles as well as timings for shipment collection and delivery in the areas affected. A simpler alternative is to apply specific size profiles of lorries to each segment of shipments. Those shipments to final consumers would have a greater proportion of use of smaller vehicles, whereas movements to warehouses would mainly use the largest vehicles.

The process of transforming the tonnes moved into the vehicles required for this movement is a rather more complex stage than the equivalent estimation of vehicle occupancy factors for passengers. The REDEFINE project in Figure 2.1 has identified three key ratios that influence this transformation and their approach is used as the foundation for the methods proposed below.

The first key ratio is the vehicle carrying capacity. In the REDEFINE project this key ratio is the ratio of "the number of tonnes the lorry fleet is plated to carry" to "the number of lorries". For the purposes of the proposed freight model framework, this ratio would be extended to represent not just the maximum tonnage that can be loaded onto the vehicle, but also the maximum volume it could carry. Increasingly, with the lighter, higher valued goods (e.g. processed foodstuffs, tobacco, electronic equipment, etc.) that are growing in importance, it is the inability to fit more into a vehicle, rather than the weight of the shipment that determines the maximum load that could be carried. Furthermore, in order to streamline delivery times the use of cages and other reusable packaging devices may further reduce the effective space available for carriage within a vehicle. Having segmented the freight by load type so that loads were either weight or volume constrained, then define vehicle carrying capacity differentially by segment so that

- for volume constrained traffic it denotes the total volume available

- while for weight constrained traffic it denotes the sum of plated carrying capacities.

Load factors would then be defined analogously as the ratio of used capacity to the available capacity

On rail the precise details of gauge are of great importance. Increases over the years in the sizes of containers have meant that nowadays the larger containers cannot be carried on standard wagons over much of the UK rail network. Major investments are planned or underway by Railtrack to ensure that the main freight corridors to the deep-sea ports and to the continent are cleared to an appropriate gauge for future container requirements. Gauge restrictions are an important aspect to be considered within rail freight modelling as they can impact on the costs of the service or even the viability of a service along a specific route.

The second key ratio is the loading factor, which represents the degree to which a vehicle is fully loaded on average. It depends on the size of the shipment to be moved to each customer. For single drop trips, the shipment may be insufficient to fill the vehicle. For multi-drop journeys the vehicle may become successively less loaded after each call, except in the unlikely case where there are symmetric loads in both directions at every call. The use of sophisticated routing and scheduling packages by transport firms originally concentrated on minimising distance, but is ever more likely to effect an increase in the loading factors towards the maximum economic level. Depending on the size of the consignments, it may be more efficient to carry out shorter distance deliveries in small rather than large goods vehicles.

The third key ratio is the proportion of empty running which is employed within the proposed framework as a variant on the second ratio. It is desirable to attempt to associate the empty running with commodity types and logistic practices as far as possible, particularly where consignments of a commodity are typically shipped in single-commodity vehicles. There may also be a set of commodities which are typically shipped in conjunction with other commodities. For these it may not be possible to associate empty vehicles with a particular commodity and instead regard the empty running ratio as that of “empty vehicle kilometres” to “all vehicle kilometres” over that set of commodities. That would be more analogous to the usage in REDEFINE where the key ratio makes no reference to commodity type. For those bulk commodity types requiring specialised vehicles (e.g. petroleum products) empty running is unavoidable. In contrast intermodal transport may achieve low degrees of empty running due to the variety of goods that potentially may be carried, provided that the overall demand for freight movements is in balance for both directions (which has certainly not been the case between the UK and mainland Europe). Intermodal technologies have traditionally been regarded as having lower load factors than road freight but changes in logistic processes may permit significant enhancements here. The recent improvements in information technology and the Internet have facilitated the exchange of information to enable empty running to be reduced. However, available road freight data suggests that empty running is highly variable, even for a single truck tracked across several days.

These three key ratios need to be combined within an appropriate modelling procedure of vehicles by size to enable the movements of commodities within a stage of the distribution chain to be converted from units of tonnes into O/D movements of vehicles by vehicle type. Different stages in the distribution chain are likely to have rather different values of these key ratios. It is likely to be more feasible to maintain high load factors for long hauls between major warehouses, than it would be when delivering small consignments to dispersed customers. Accordingly, the economically optimal vehicle size to use for this latter stage is likely to be smaller than that for the long haul stage.

2.10 Assignment to the modal networks

There is certainly a benefit to assigning the road vehicles to a highway network for GB in a manner that can take direct account of the impact of congestion. Carrying out an equilibrium assignment would be the preferred approach but this implies that the network must also contain (implicitly or explicitly) the passenger vehicles that create most of the congestion. A discriminating representation of congestion effects on the network would need some means of taking account of the fundamentally different demand characteristics in the different time periods. Alternatively, it could explicitly include separate assignments by time period and accept the associated extra computational burden that this would imply for a large scale or national model. A consistent procedure would also be needed to take account of the fact that most of the longer freight journeys will spread over more than one time period.

The response of road freight operators to congestion is more complicated than that of passengers. Road freight is often priced on distance travelled, and so the cost paid by shippers may not in the short term reflect congestion. Operators are sensitive to congestion particularly through measures of network reliability (they equate congestion with higher variances in travel times, which in turn signals lower reliability and less cost-effective operations). The varied effects of congestion on shippers and on operators are not clear-cut, and depend much on the ownership and operating characteristics attributed to different classes of goods vehicles. At the overall system level however, increases in road congestion will lead to higher transport costs, due to the higher employment, larger fleets and increased fuel usage required to deliver the necessary goods.

A major benefit from carrying out a conventional network assignment would lie in its ability to generate discriminating statistics on the spatial patterns of pollution emissions, noise and other forms of nuisance, as well as the congestion effect of goods vehicles and measures of wear-and-tear on the road surface. For many of these indicators the locations and road types on which the indicators are growing or declining may be a matter of considerable interest to various parties. In any case when moving down from the national to the sub-regional scale and below, the network assignment component becomes a more important feature of the whole modelling exercise. Also the estimation of vehicle movement matrices, which needs to be checked against observed classified counts, demands a consistent assignment process (which is embedded in many of the matrix estimation procedures employed).

The need to carry out a direct assignment of the rail freight to the rail network is less clear-cut. The capability to do this certainly exists, since many of the features required to model passenger rail services carry across to rail freight. Added complications relate to gauge restrictions and to the definition of the supply of train paths that may be made available to rail freight on different parts of the network. The latter is how the competition between passengers and freight is resolved in practice, but the decision processes behind it are extremely complex and difficult to attempt to model endogenously at this level. Therefore, such a model component might be better incorporated fully at a later date, or carried out in the shorter term with simplifying assumptions about rail freight routing and path availability.

When assigning intermodal traffic it will be necessary to ensure that the road leg of an intermodal movement is assigned to the road network, as well as assigning the rail leg to the rail network. In this way undercounting of trips is avoided.

3 THE FOUNDATIONS OF THE OVERALL MODEL DESIGN

This chapter presents the foundations on which the designs of the various stages of the freight and LGV models are constructed. It concentrates on those aspects such as zoning, networks, segmentation and the broad principles of base matrix creation, that are common to many model stages. Aspects that are more specific to a particular model stage are then discussed in later chapters. Likewise, the primary focus here is on modelling at the national level. Aspects that relate only to more detailed spatial scales are discussed in Section 8.

3.1 The study area and zoning system

The internal study area for the national freight model covers Great Britain but not Northern Ireland. The model will be focused on domestic traffic and on the domestic legs of international traffic.

In principle a model with more zones and smaller zones will provide more refined results than a model that covers the same study area with fewer and larger zones. In practice, however, the computational effort required in a model tends to increase proportional to the square of the number of zones so that the inclusion of too large a number of zones will greatly reduce the degree of segmentation that can be introduced in other equally important aspects of the behavioural basis of the model. The recent development by Williams and Lindsay (2002) of zone pair based sampling techniques does provide a means of adopting large numbers of zones, without the corresponding overwhelming computational burden. The most appropriate trade-off between these conflicting consumers of resources is a matter of judgement that will be influenced by many of the issues discussed in this chapter.

The availability of existing data provides a potential constraint that may reduce the degree of spatial detail available within a model. For road freight data the current sampling rates in the CSRGT and IRHS mean that their observed O/D matrices cannot safely be disaggregated below the county level, without creating the danger of significant sampling errors.

The rail O/D matrix data that has been assembled for the GBFM model has been coded to the county level, but the raw data is coded at the individual station level. So in principle it could be categorised more finely, subject to avoiding problems of confidentiality. Sampling errors are not a significant issue for rail freight data.

Given that there appears to be broad agreement within the Freight Review team that some form of network based model is appropriate, it is difficult to see how this could be implemented effectively with a zoning system that is spatially coarser than at the county scale. In fact a local authority (LA) district based zoning system would appear to be a more appropriate level for national freight modelling than the county level. At the district level there is some distinction between the significant towns within a county, and the major conurbations, including London, can be subdivided. The NTEM zoning level (see Section 3.1.2 below) would be better again than the district level because it is designed to distinguish specific urban zones, and so it would provide a more discriminating representation of the origins and destinations of freight flows. However, it would increase the computational burden and data requirement still further.

Unfortunately, existing road freight O/D data is not directly available either at the LA district or at the NTEM scale (for either origins or destinations). To address this shortcoming, either very much larger sample sizes would be needed in the CSRGT and IRHS, or else innovative methods that are discussed below would need to be devised that would enable the existing data to be disaggregated spatially in a robust fashion.

3.1.1 Zones for ports and airports

In order to represent the important freight movements of imports and exports, a further set of point zones is needed to represent the freight ports and airports in an explicit fashion. In this way a distinction can be maintained between domestic production/consumption in a zone and the imports/exports that enter/leave GB via that zone. This distinction is important in order to enable appropriate scaling factors to be applied to these different traffics. Typically in the past the import/export traffic has been growing in volume far more rapidly than domestic production.

For the national freight model the primary focus is on road and rail borne freight so that petroleum ports or other large bulk ports that cater solely for products processed adjacent to the port need not be included. It is the set of ports that specialise in unitised traffic or RoRo freight ferries that are of special importance, since these are the major generators of large volumes of longer distance road and rail freight transport demand. It may be necessary to extend this list of point zones to included significant port expansions that are in the pipeline (Thamesport, Mostyn, Dartford) and major new air freight hubs.

The Channel Tunnel should be included as a specific zone, due to its major freight flows on road and rail. The main west coast ferry ports between Mainland Britain and Northern Ireland plus the Irish Republic also need to be included, together with the ports carrying foreign trade to the continent and the rest of the world.

3.1.2 Zoning systems for freight models operating below the national scale

The zoning systems in use in other freight models that operate at a finer spatial scale than the national level, should if at all possible be consistent with the zoning in use in the Department's main other modelling initiatives, namely either:

- The approximately. 10,000 zone system of 1991 frozen wards (or any future update of this based on the 2001 Census) which is currently used as the basis for both the National Trip End Model (NTEM) and for the Pass3 road assignment in the passenger component (NPTM) of the overall GB National Transport Model (NTM).

- The zoning system that is used for the provision of demographic and passenger trip information from the NTEM via the TEMPRO system (approx. 1,000 zones). Each NTEM zone is a contiguous group of 1991 Census frozen wards. In general, each local authority district (defined as 1999 unitary authorities, metropolitan boroughs, and shire districts) is split into named urban zones and rural district remainders to create homogeneous areas

  1. In shire districts, zones distinguish urban and rural areas

  2. In conurbations, zones distinguish the major cities, and divide the remainder into rural and urban areas

  3. In London, zones distinguish areas of higher and lower population density.

  Some zones are whole districts in those cases where the entire district is of the same area type. A typical district is divided into about three zones, with some districts (where there are several distinct urban areas) divided into 6 or 7 zones.

The standardisation of the zoning systems in use in different modelling studies should be reviewed, since it is an important issue in terms of making best overall use of resources. It facilitates the reuse of data in later studies, thus freeing up resources for improved data collection on aspects that are really needed. However, differences in orientation between local studies with different objectives may make this difficult to achieve in practice below the district level.

The nature of freight activities and commodity flows is such that there will be many highly concentrated point sources (e.g. mines, factories, warehouses and other enterprises), which will have to be linked appropriately to the representations used for the transport networks. Accordingly, a spatial database (GIS) is clearly the appropriate entity for data management, provided that data on individual enterprise can be obtained initially. The point data could subsequently be aggregated into zones when required, while accurate point data would be used for other purposes such as intraurban and/or micro-simulation modelling.

3.2 The modal networks

3.2.1 The highway network

The highway network that is currently in use for the NPTM covers the whole of GB and includes the minor roads in considerable detail, and not just the main road network. It would provide a suitable network for use in the national freight model for the following reasons.

Firstly, it would enable the national freight and passenger models to be implemented using a common representation of the road supply. This would facilitate calculating the impacts on freight vehicles of congestion caused by passenger traffic, and vice versa.

Secondly, this network already has a significant amount of current network supply and demand information linked to it within a GIS and so would save repeated tasks and minimise inconsistencies in data sources within the Department.

Finally, constructing all DfT's various network related activities on the basis of a single common network should lead to substantial efficiencies whenever it is necessary to update such information. The consistent use of a single network should generate sufficient resources to enable it to be carefully checked and maintained to a higher standard than would be the case with a proliferation of different networks in use in different models.

It is not necessarily the case that this NPTM network is also the most appropriate for use in sub-regional and urban models. It is a link based network that does not at present contain information on junctions. For this reason it will have limited usefulness at the detailed spatial level where local schemes are being assessed.

The Department and the Highways Agency should consider the case for an initiative to standardise in a single GIS the use of networks across different local scale studies. The aim would be a service that creates and maintains a database that is gradually expanded within a harmonised framework until the whole of the national network is covered. New studies would make use of and improve the parts they require of this network and then in a standardised fashion reinsert their improved updated version in the GIS. Clearly the design and control of such a system (e.g. to exclude redundant or inconsistent data) would require significant resources initially to provide a design sufficient to cater for the different remits of different studies. It would subsequently require a lower level of resources to provide a continuing network management, maintenance and promulgation service within the GIS. In the longer term it is likely both to be very cost effective and to lead to improvements in quality, through avoiding the waste of resources in repeating the task of network construction and checking in overlapping study areas. It would also facilitate the comparison of results from different studies, since all would estimate traffic on the same set of links. This would facilitate improved methods of monitoring the quality of model forecasts.

3.2.2 The rail network

Although a spatially detailed rail network is in use in the rail passenger module of the NPTM, it is not automatically the case that this would be the most appropriate choice for use also for rail freight.

A rail freight network is already is in use in the GBFM freight model created by MDS-Transmodal and in use by the SRA– this is not the same network as that developed by Oscar Faber for the detailed rail passenger module within the NPTM. There are in fact two kinds of networks available to the GBFM model. One is a set of pairs of loading and unloading points acting as hyperlinks, e.g. Daventry-Muizen which carries containers and swapbodies. The other is a link and node network that could be used for assignment but as yet has not been.

There is not a major need to simultaneously assign passengers and freight to a single rail network. The competition for capacity between them is mediated by administrative decisions on the allocation of train paths, and is not resolved through the mechanism of continual direct competition for network capacity in real time that occurs on road.

The network supply characteristics relevant to freight trains are not the same as for passenger trains. Many parts of the network may be unsuitable for use for certain types of freight movement for reasons of gauge restrictions. The existing scheduled freight services (e.g. those provided by Freightliner from the deep-sea ports) would need to be included explicitly in a network coding of services.

3.2.3 Intermodal and distribution networks

Because of the requirements to examine intermodal traffic, either where road feeds rail, or where road/rail feed shipping, the networks should have the capability to represent intermodal movements in a consistent fashion. For intermodal movements, the road legs to feed rail terminals or ports should be assigned as traffic on road links in the same manner as would a movement wholly by road.

An intermodal representation implies that major road/rail interchange centres should be explicitly coded, as well as the more important freight ports for both ferries and for conventional shipping. The representation of the (un)loading costs and delays is also needed at such interchanges. These requirements would necessitate that the assignment procedure eventually adopted should be designed in a form that enables all stages of intermodal movements to be represented explicitly. A traditional road assignment package is unlikely to be sufficient in isolation.

There may also be a case for the explicit inclusion of major distribution centres on the network, as this provides one means by which distribution chains can be represented. However, other means of representing distribution may be adopted as an alternative to a purely network based representation.

3.3 Creating good quality base matrices

Because of the intrinsic heterogeneity of freight transport patterns, there are benefits to be gained throughout the model from maximising the use of good observed base matrices. Achieving these benefits requires the development of two main innovative sets of procedures:

1. Procedures to construct realistic base matrices, which make consistent and effective use of all relevant observed data – these procedures are discussed in the current section

2. Procedures that are constructed around these base matrices (variants are known as incremental, marginal, pivot point or residual disutility based models), which model behavioural responses in a flexible fashion – these procedures are discussed in Section 4.1.

The construction of base matrices is discussed below from a general standpoint, which is equally applicable to passenger traffic. More specific considerations relevant to constructing a specific matrix for a specific mode or for a particular stage in the modelling process are discussed in later chapters.

The task of estimating matrices of vehicle trips from link based observations is a much studied task for which a variety of approaches have been developed over the years. Earlier approaches to this problem focused on making best use of data from one specific observed source. For example, Echenique and Williams (1982) showed how to maximise the useful information that could be extracted from a set of roadside interview surveys (RSIs) that collect data on the origin and destination of trips. Nguyen (1984) showed how to make efficient use of road vehicle counts that do not have accompanying data on origins and destinations. The ME2 method of Van Zuylen and Willumsen (1980) is a vehicle count based approach that has been widely used in practice within the UK to expand or update O/D matrices.

Research in the 1990s has created a more comprehensive approach where observed data from a variety of sources is all included as an input to the matrix production. Logie (1993) discusses one particular package that has been developed using maximum likelihood, MVESTM, and also provides a useful bibliography of other approaches. This is an evolving field with a number of alternative approaches such as the Combined Calibration Method (Gunn et al. 1997) and the Improved Matrix

Estimation Technique (IMEST) method (Heydecker et al 1995). The modelling approach that was proposed by the ASH Research Group (1997) in the National Transport Model Feasibility Study – Project 2 is along the lines of these developments. The New York City model (List, *et al*. 2001) described in Report B2 appears to be of particular interest as it provides an optimisation method of matrix estimation that makes efficient use of a wide variety of data sources, and provides pointers to potential major inconsistencies in observed data sources in use.

This more recent research is concerned with the development of matrix construction methods that:

(i) make efficient use of **all** of the observed data (e.g. prior matrix, traffic counts, RSIs) that is available

(ii) but assume that the different sources of observed data are not necessarily all fully consistent with one another nor equally accurate (e.g. some RSIs may have larger sampling fractions than others, some data may be less accurate than the norm)

(iii) and consequently allow different weights to be assigned to different sources of observed data in order to take account of the expected reliability of each source, relating to the accuracy of measurement, sampling errors, inherent variability, etc.

The goal is to ensure that the resulting estimated "observed" matrices are as close as possible to observed aggregate O/D patterns and traffic counts. A quotation from Gunn et al (1997) summarises some of the distinctive features of this approach:

> "A basic difference between the method described here and standard US/UK practice is that none of the data items is taken as a *constraint,* rather it is a piece of statistical evidence to be weighed against others. The accuracy of the data item (be it trip-end total, overall trip-length proportion, or total of aggregated O/D flows) determines the weight that it has in determining the final estimates of the matrix."

3.3.1 The ideal requirement – hybrid data expansion and model construction procedures

The ideal requirement for base matrix creation within the GB freight modelling study is ambitious and hence may only be fully achievable over the longer term. There are however feasible and useful intermediate positions that could be achieved. In total it would involve creating a linked set of matrices that ideally would include all of the following constituent components.

1. Movements are in units of tonnes and have a detailed segmentation: by mode (and within the mode by vehicle type), by broad commodity type and by stage in the distribution chain.

2. Origins, destinations and intermediate interchanges are at an appropriate level of spatial detail in its zoning system – probably LA district but possibly down to NTEM zone level (see Section 3.1 above). Zonal indicators of local levels of trip generations and attractions are used to help in the spatial disaggregation process.

3. Movements at the most detailed level of the matrix represent modal O/D shipment legs. These legs have explicit procedures:

    - to enable intermodal movements to be constructed that contain direct information on the intermediate interchange points at which specified changes of mode take place

    - to enable P/C trade matrices to be constructed that contain direct information on the intermediate distribution points at which the stages of the distribution chain take place

This base matrix is constrained to ensure that its partial totals and its derivative information (e.g. the link flows obtained through assigning the O/D shipment matrix to a modal network) are matched to an appropriate degree of fit to the observed control data values, which would include:

1. CSRGT county O/D matrices of domestic road shipments, segmented by type of commodity

2. IRHS county to port matrices of international road shipments on UK registered vehicles, segmented by type of commodity

3. Paladin based rail station to station matrices, segmented by type of commodity

4. RSI data on O/D patterns of HGVs by vehicle size and possibly broad commodity type

5. Road count data of HGV traffic, possibly segmented by vehicle size

6. Airport, port and Channel Tunnel data on road and rail movements that (un)load goods at the terminal or in the case of RoRo or shuttle services that actually pass through the terminal.

It will be clear that the base matrix creation procedure that would be required to meet the above set of challenging requirements, is one which is a great deal more complex than traditional procedures such as ME2 or Fratar/Furness matrix adjustments. The procedure would blur the distinction between the traditionally somewhat separate activities of:

- statisticians expanding survey data to produce observed matrices,

- modellers constructing and calibrating a model based on these observed matrices.

A very large sample size would be required (as found by the RHTM), together with a complex survey design and a long questionnaire in order to create purely through direct observation a matrix of the type specified above. Accordingly the scale of resources required in order to implement fully this traditional approach of direct observation would almost certainly be prohibitive.

Alternatively a hybrid matrix synthesis approach could be adopted in which the twin tasks of model construction / calibration, and of sample expansion / harmonisation across surveys, are integrated into an intertwined co-ordinated set of tasks. This type of approach has already been used successfully in the past in the construction of models. One such example is the APRIL model of road pricing in London (Williams and Beardwood, 1993) where detailed base matrices of tours by mode, time of day, traveller type and trip purpose were synthesised based on observed LATS data.

3.3.2 The approach to base matrix creation

A formal overview of the general approach to be adopted in the creation of the base matrix is now presented. Define the following notation:

T_{ijk+}^{m} denotes the estimated base O/D matrix of shipment legs in units of tonnes from the origin zone *i* to the destination zone *j*, segmented by shipment type *m* (denoting commodity type, stage in the distribution chain, time of day and any other categorisation that may be of relevance to the study), and by mode *k* (denoting vehicle size and type, stage in the intermodal chain, etc.). This is the matrix to be estimated by an appropriate mixture of observed data and model calibration. Here the notation "+" denotes summation over the corresponding index.

p_{ijkl}^{m} for a specific shipment leg which is of type *m* and mode *k* from zone *i* to *j,* it denotes the probability that it will pass through the link *l* in the course of that journey – this probability is an output from the assignment process for the mode *k*.

$\widetilde{M}_{ijkl}^{m}$ denotes a set of directly observed data on the movements. In general, observed data values (denoted by "~" here) are only available for a limited number of partial, aggregated slices of this 5 dimensional matrix (e.g. county to county O/D matrices, traffic counts, RSIs, etc.).

These observed values may be subject to errors in observation. They may be in different units (e.g. vehicles not tonnes) to those in the base O/D matrix.

w_{ijkl}^{m} is a set of weights denoting the relative reliability of the corresponding observed data element. A value of zero denotes that there is no worthwhile observation available.

$Q_{ijkl}^{m}(\,)$ denotes a transformation that is applied to the elements T_{ijk+}^{m} of the estimated base matrix in order for them to be in the same units as in the corresponding set of observed data (e.g. shipments transformed from tonnes to vehicles or containers).

The general mathematical specification of the problem is to estimate the set of values for T_{ijk+}^{m}, which minimises the error between estimated and observed values as measured by the general constraint expression

$$\sum_{ijklm} F\left[\tilde{M}_{ijkl}^{m}, Q_{ijkl}^{m}\left(T_{ijk+}^{m} p_{ijkl}^{m}\right), w_{ijkl}^{m}\right] \qquad (3.1)$$

The functional form of *F[ ]* might be defined to be: the absolute difference, least squares, maximum likelihood or some other such form. The estimation process chooses values for the set of parameters that are embedded in the model used to estimate the matrix T_{ijk+}^{m} such that the constraint function (3.1) is minimised.

Although in principle this formulation can be solved by general non-linear optimisation routines, in practice a more tailored approach that relies on the particular structure of the model and on decomposition techniques is almost certain to be needed. Otherwise the process may not be computationally feasible for models with large numbers of zones.

3.3.3 Examples of formulations of base matrix creation

Some examples are now introduced to illustrate the notation, starting with the simplest fixed matrix approach. Then some extra features are introduced to show some of the complexities that may arise in a more comprehensive formulation.

The first example is the standard fixed trip matrix estimation problem. This takes a prior O/D vehicle matrix from some source (perhaps, an observed matrix T_{ij++}^{+0} from an earlier year, where the superscript '0' denotes 'prior'.) and adjusts this matrix so as to make it match an updated set of origin and of destination totals. In this instance the updated matrix is estimated using row and column balancing factors r_i and s_j that are applied to the prior matrix. This can be denoted by

$$T^{+}_{ij++} = r_i s_j T^{+0}_{ij++} \tag{3.2}$$

The constraints from equation (3.1) can be written in a simplified form as

$$\sum_i \left[\widetilde{M}^{+}_{i+++} - T^{+}_{i+++}\right]^2 + \sum_j \left[\widetilde{M}^{+}_{+j++} - T^{+}_{+j++}\right]^2 \tag{3.3}$$

which when minimised with respect to the balancing factors provides, through its first order conditions, a set of equations to evaluate r_i and s_j which ensure that all row and column totals are exactly matched. In this simple example the terms p^{m}_{ijkl} w^{m}_{ijkl} and Q^{m}_{ijkl} are not needed.

This procedure can be extended to the case where:

- a synthetic generalised cost based distribution model replaces the observed prior matrix or is used to provide greater spatial discrimination than is available within the observed matrix,
- additional to the origin and destination totals that are to be matched in eqn. (3.3), there is also a set of observed O/D matrix elements that the estimated matrix is controlled to match.

The details of a formulation that meets these requirements are presented in Appendix II.

The procedure can further be generalised to cover the case where the observed control O/D matrix is available only at a more aggregate zoning system and/or a more aggregate segmentation of type of movement than is used when estimating the synthetic model matrix.

Another type of extension to this procedure is discussed in Gunn et al. (1997). This introduces the influence of the uncertainty that can arise from, say, sampling errors when the observations are only a sample from the total population, and hence are inappropriate to use as exact constraints when they have been scaled up to match the population total. The weights w^{m}_{ijkl} represent the expected reliability of different sets of observations, and the functional form of *F[ ]* represents the probability distribution of the errors in observations.

Another strand is covered in the work of Zhang et al. (1999), which examines methods for the solution of the combined trip matrix estimation and stochastic user equilibrium assignment problem. This uses the p^{m}_{ijkl} term to take account of the probability of a specific O/D movement passing through a specific link *l*. The resulting matrix is that which best matches a set of observed traffic on links.

In summary, it can be seen from the examples above that the various ingredients needed for a general approach to base matrix estimation have all been developed. The estimation process may require elements of several of these approaches to be simultaneously implemented in the estimation of a large and complex base matrix of the type implied by the requirements outlined in Chapter 2. The challenge is one of scale and of cost effectiveness. It is certainly not a trivial challenge, especially given that such an approach has not as yet been implemented in a comprehensive fashion for passenger modelling – generally an easier task than freight modelling.

3.4 The segmentation of freight transport demand

The reason why freight demand is segmented into detailed categories in various stages of the model is to create a set of categories of shippers/shipment types within each of which the behavioural responses would be reasonably homogeneous within the corresponding model stage. Unless all the individual elements within a category have reasonably homogeneous responses, the ability of the model to respond realistically to policy initiatives will be compromised. This requirement for homogeneity in behaviour creates a pressure to adopt a large number of detailed categories within the model.

In contrast the limitations on the range of data available, the existence of sampling errors in survey data, and the inconsistencies between the data definitions used in different data sources to categorise data, all limit the level of detail in the segmentation that can safely be adopted. Furthermore, the greater the number of categories that are distinguished within the model, the longer it will take to run and the more expensive it will be to operate. This means that there are a series of trade-offs to be resolved within the model design process, in order to finalise the most cost-effective segmentation to be adopted within each model stage.

Table 3.1 Potential segmentation by type of freight at different stages of the model

Stage	Number of segments	Segments
Generation – value to volume	<123	Aggregation of the 123 I/O table industry categories
Spatial dist. of P/C trade	17	CSRGT NST based commodities
Spatial dist. of O/D shipments	<17x5	CSRGT NST based commodities, subdivided by stage in the logistics chain, consignment size/regularity
Modal split	<17x5x2	Commodity type by logistics stage, consignment size/regularity by road, rail,
Transformation to trains, assignment to rail	3x2	Intermodal, bulk, others by trainload, wagonload
Transformation to road vehicles, assignment to road/ferry	3x2	LGV, small HGV, large HGV by high value, other
Number of zones	500	LA district level for GBr

The most efficient approach is within each model stage to use the specific segmentation that is most appropriate for that stage. In this way the segmentation adopted will be different in different stages of the model. Implementing this in practice is not always straightforward since data that needs to be transferred through a number of model stages may create difficulties. If the model design only requires that categories be successively aggregated at each transfer to the next lower model stage then no major issues arise. However, when at any stage a category needs to be disaggregated then the complexity increases unless such categories have already been carried through from previous stages.

Matching the CSRGT to the UK I/O categories

The matching of economic data to transport demand data is one task that is likely to need to be carried out within the national level freight model. The most accurate way to achieve this is by an initial low level matching at the most detailed level of categorisation and then to aggregate as required. The Table 3.2 presents an approximate match between the 19 CSRGT commodity categories and the 84 freight generating production industries contained within the UK standard 123 I/O industry categories (ONS, 1997). The remaining 39 I/O categories are mainly service industries, which are not producers of commodities, but may indeed consume them. Because these two categorisations have originally been defined for very different uses and so are not inherently consistent, considerable detail needs to be maintained in the *economy* oriented segmentation of I/O industry categories to facilitate the match to the CSRGT's *transport* oriented commodity categories. The code numbers for the I/O categories have been defined such that activities with similar economic characteristics will have adjacent numbers to facilitate their aggregation to coarser aggregate industry categories. However, the existence of some wide ranging combinations of numbers in the final column of Table 3.2 demonstrates the clear need for the matching of the I/O categories to the CSRGT to be carried out using these detailed categories rather than the coarser aggregate categories, where possible. It is also important to note that around 20% of the CSRGT data falls within the category 19, Miscellaneous Articles, about which little information is available to enable them to be matched back to the real commodity type carried, much less to the I/O categories. Other data sets may also use incompatible definitions of the "Miscellaneous" category making comparison and summation complicated.

An analogous task arises for rail freight movements, which need to be classified into a set of categories that can be matched to the I/O categories. The matching of the rail and CSRGT categories can be based on the GBFM model, which has addressed this task as part of the construction of combined road and rail freight O/D matrices for 1999 as described in work area B1.

Table 3.2 Proposed matching of UK I/O group categories to the CSRGT commodities

	CSRGT commodity category	I/O Groups of Commodities
1	Agricultural Products	1,3,12,13,15
2	Beverages	18,19
3	Other Foodstuffs	8,9,10,11,14,16,17,20
4	Wood, Timber & Cork	2,31
5	Fertiliser	39
6	Sand Gravel & Clay	7*
7	Other Crude Minerals	7*
8	Ores	6
9	Crude Materials	21,22,23,32,46
10	Coal & Coke	4
11	Petrol & Petroleum Products	5,35
12	Chemicals	36,37,38,40,41,42,43,44,45
13	Cements	52
14	Other Building Materials	49,50,51,53
15	Iron & Steel Products	54,55,56
16	Other Metal Products (not elsewhere specified)	57,58,59,60,61
17	Machinery & Transport Equipment	62 to 66, 68 to 80
18	Miscellaneous Manufactures	24 to 30, 33,34,47,48,81,82,83,84
19	Miscellaneous Articles (not elsewhere specified)	67

*The I/O category "7 - other mining and quarrying" can produce goods in either of the CSRGT categories 6 and 7 above.

In particular, since transport models typically pass matrices of demand down through a chain of model stages, but pass matrices of composite generalised costs/disutilities in reverse up this same chain of model stages, there is inevitably a problematic requirement for disaggregation in at least one of the two directions. The easiest way to avoid inconsistencies is to maintain a sufficiently detailed segmentation to avoid the need to disaggregate, but this can become prohibitive in the resources that it would consume.

Table 3.1 presents some approximate estimates of the segmentation detail and trade-offs for various components of the proposed model at the national level. This highlights that the upper stages of the model that were defined in the framework in Figure 2.2 contain the greater level of detail in types of freight. To enable the various trade-offs to be resolved efficiently in Table 3.1 the eventual segmentation should be finalised once all of the other model details have been specified.

3.5 The representation of various types of movements within the set of models

The discussion in the Report A3 on the 3-dimensional classification of: trip type (freight or non-freight), trip purpose (commercial or private) and vehicle type (Light goods/commercial vehicle – LGV or private) has used the work of Wigan and Rockcliffe (1998) to identify the definitional complexities that may arise when analysing freight. Table 3.3, which is adapted from the Report A3, provides examples to explain the types of trips in each category. This is used as a conceptual aid to ensure all LGV types are identified, to help create a clear definition of what trips are to be included within the LGV model, and how these relate to the trips to be included within the freight model as outlined in Table 3.4 and then discussed in Chapter 4. It should be noted that in this report all service trips are defined as non-freight (unlike report A3 in which those service trips involving the movement of goods were defined as freight) and they would be treated in the proposed LGV model rather than in the freight model described in Chapter 4. Further consideration needs to be given to understanding the relative importance of service trips that involve the movement of goods, and to deciding whether these are best treated in the freight model or in the proposed LGV model.

The proposed specification of where different types of trips are represented within an overall national modelling system is summarised in Table 3.4. This expands the segmentation from Table 3.3 by making an explicit distinction between road trips on LGVs and on HGVs. The passenger model is the passenger component (NPTM) of the Department’s National Transport Model which has been developed primarily using trip data from the NTS. Other passenger transport models at a finer spatial scale are likely to have a broadly similar level of coverage of passenger trip types. The “freight” model is as described in Chapter 4 and the “LGV” model covers the modelling of passengers and freight as described in Chapter 5.

Table 3.3 Segmentation of road trips by freight type, purpose and vehicle type

Segmentation	Examples	Relevance to LGV modelling	Relevance to freight transport modelling
Segment 1.1 Freight trip for commercial purpose using LGV	• Parcel delivered in Royal Mail LGV • Grocery home delivery by LGV • Computer field engineer using LGV to deliver and install replacement CPUs • Plumber using LGV to deliver and install new pipes and boiler in customer's building • Retail or manufacturing delivery from warehouse	Yes	High
Segment 1.2 Freight trip for commercial purpose using private vehicle	• Mail order goods delivered to customer's home by agent in car • Take-away meal delivered to customer's home on motorbike • Shopkeeper using car to transport goods from Cash and Carry • Computer field engineer using car to deliver and install replacement CPUs • Plumber using car to deliver and install new pipes and boiler in customer's building • Fashion clothing delivered in time-critical but small consignments	No	Medium
Segment 2.1 Non-Freight trip for commercial purpose using LGV	• Empty running of LGV used for grocery home deliveries • Repositioning between depots of LGV used for parcels deliveries • Electrician driving to customer in LGV to make quotation	Yes	None
Segment 2.2 Non-Freight trip for commercial purpose using private vehicle	• Electrician driving to customer in car to make quotation • Salesperson visiting premises to negotiate contract in car • Consultant visiting client in car	No	None
Segment 3.1 Freight trip for private purpose using LGV	• Private individual transporting personal shopping home from shops in LGV • Private individual purchasing second-hand goods from residential address in LGV • Private individual going to rubbish trip with garden waste in LGV	Yes	Low/Medium
Segment 3.2 Freight trip for private purpose using private vehicle	• Private individual transporting personal shopping home from shops in car • Private individual purchasing second-hand goods from residential address on motorbike • Private individual going to rubbish trip with garden waste or junk in car	No	Low/Medium
Segment 4.1 Non-Freight trip for private purpose using LGV	• Plumber making personal trip to visit relatives in LGV • Telephone engineer making personal trip to sporting event in LGV • Delivery driver going home for lunch on working day in LGV	Yes	None

The reason why some trips are marked in the final column as not modelled is because these trips are not included within the main surveys that are available on which to set up the models initially. In general these categories of trips are believed to be numerically small and so it is not a problem to ignore them. The row 1 implies that those movements of LGVs that are primarily to carry freight are included as a component of the freight model. Based on the documentation of the NTS data definitions, the existing NPTM is likely to include some small proportion of the total number of trips of the type in row 3, freight trips for commercial purpose using private vehicle.

Table 3.4 Summary of representation of road trips within proposed set of models

Segment	Trip	Relevance to freight modelling	Where modelled
1.1a	Freight trip for commercial purpose using LGV	High	LGV in NFTM
1.1b	Freight trip for commercial purpose using HGV	High	HGV in NFTM
1.2	Freight trip for commercial purpose using private vehicle	Medium	Part in NPTM
2.1a	Non-Freight trip for commercial purpose using LGV	None	LGV
2.1b	Non-Freight trip for commercial purpose using HGV	None	Not
2.2	Non-Freight trip for commercial purpose using private vehicle	None	NPTM
3.1a	Freight trip for private purpose using LGV	Low/Medium	LGV/NPTM
3.1b	Freight trip for private purpose using HGV	Low/Medium	Not
3.2	Freight trip for private purpose using private vehicle	Low/Medium	NPTM
4.1a	Non-Freight trip for private purpose using LGV	None	LGV/NPTM
4.1b	Non-Freight trip for private purpose using HGV	None	Not
4.2	Non-Freight trip for private purpose using private vehicle	None	NPTM

Table 3.4 is for illustrative purposes and should not be taken to imply that the passenger, freight and LGV passenger model components of the NTM would all be separate. The extent to which they are linked will depend on the particular options chosen as explained later in this report. In particular any separation would not necessarily extend to the road assignment stages of these models. The appropriate manner in which to obtain a converged equilibrium assignment is to assign all of the resulting car, HGV and LGV matrices as separate user classes within a single equilibration procedure. Then the resulting congested travel disutilities should be fed back up the choice hierarchy to influence mode choice and the spatial distribution pattern within each separate model.

In practice however this global modelling procedure would be unlikely to be the norm for the running of most policy tests. Instead it is likely when modelling, say passengers within the NPTM that an appropriate set of base LGV and HGV freight matrices for the year would be loaded onto the network to add to the congestion. Within the run, these LGV and HGV trips would not change behaviour other than their routings in response to local congestion. Likewise, it would not in general be envisaged that the full NPTM would also be run as part of a standard freight model run. The modularity of the models is discussed further in Chapter 6.

This Table 3.4 also demonstrates that in the proposed system there would not be significant double counting of trips between models. However, in the absence of an LGV model, the present version of the NPTM has been set up to include within the category termed car mode, those van trips that are captured within the NTS. **Accordingly, as soon as a national LGV model is fully operational, the NPTM should be revised such that the trips that are by vans are separated out, in order to avoid double counting**. The coverage of the NTS is discussed in more detail below and in Section 5.1.

3.5.1 The reasons for locating trips within specific models

This section explains the underlying influences that have governed the set of decisions on which models should include which types of trips, which are summarised in Table 3.4. It focuses only on the road vehicle flows, since ambiguities analogous to the LGV/HGV split are not an issue on rail. The ideal is that each trip should be included fully within one, and only one model. In practice although double counting can largely be avoided, some trip types of minor importance are not available from existing or proposed data sources and so are excluded.

The freight model is designed to make effective use of existing data sources (though some enhancements and extensions to these are also suggested in the later Chapter 7). The freight model's HGV road traffic component is denoted in Table 3.4 by the label *HGV* and comprises:

> movements on goods vehicles above 3.5 tonnes gvw for freight related purposes (i.e. including empty running to reposition the vehicle but excluding non-freight trip purposes) as defined in the CSRGT

This is the size cut-off that is used in the survey design for the CSRGT, which will be a major data source, and it also coincides with the categorisation of much of the published data on traffic counts by vehicle type. It is of interest to note that within the 3.5 to 7.5 gvw tonnes category this definition includes van-type vehicle bodies and lorry-type vehicles in almost equal proportions. It is the maximum carrying capacity measured as gvw, not the vehicle body-shape, which defines the category HGV.

The CSRGT does not include any movements for private travel purposes in HGVs. There is no good comprehensive source of information available at present on private trips made in HGVs, though some approximate information on the overall scale of these trips could be derived from the National Travel Survey (NTS). It is unlikely that this is a substantial source of overall vehicle kilometres, except perhaps for the vans at the bottom end of the size range.

The LGV road traffic component is denoted in Table 3.4 by the label *LGV* and comprises:

> movements on light goods/commercial vehicles of up to 3.5 tonnes gvw for all passenger and freight related purposes as defined in the forthcoming LGV surveys (LGVS) by the Department.

These surveys will be the major data sources for LGV movements and will be complemented with link traffic count data of LGVs. Unlike the CSRGT, the LGV surveys will be concerned with all trips made, irrespective of whether they are for freight, commercial or for private purposes. The number of private trips in LGVs, especially in the smaller car-based vans, is likely to be considerable and so it would have been unwise to neglect them. Those movements of LGVs that are primarily to carry freight should be included within the freight model. LGV trips for non-freight purposes might optionally be in a standalone LGV model or be integrated as part of a more general passenger model. In Section 5.3 it is recommended that LGV trips for commercial purposes are modelled in a separate module from the passenger model.

As discussed in Appendix I of the Report A3, the NTS is potentially a useful source of aggregate data on some types of LGV trips. One disadvantage to the use of the NTS in the context of HGV or LGV modelling is that the NTS does not use a vehicle type definition that is directly compatible with other major data sources. Within the NTS the variable V3, Type of Vehicle, (as defined since 1992) distinguishes "light van" as not more than 1.5 tons *unladen weight*, whereas most other national data sources generally use a definition of not more than 3.5 ton *laden weight.* The NTS has a second relevant vehicle type "Other Van or Lorry" of more than 1.5 tons *unladen weight*. The splitting of NTS trips between LGVs and HGVs will not be an exact match to the category definition proposed for use in the LGV model. This explains the background as to why it is difficult to produce a robust estimate of the number of HGV trips that are made for purposes not related to freight. It also implies that the development of an LGV model cannot merge NTS data with LGVS data in a manner that is entirely consistent.

Consideration should be given to whether the NTS separation between these two vehicle categories could be revised to adopt a categorisation that matches directly to other major data sources. As yet it has not been possible to ascertain whether the current mismatch in definitions is significant or minor in its impacts.

4 THE MODEL DESIGN FOR FREIGHT TRAFFIC

This chapter focuses on the modelling options that could be adopted to make operational each of the individual stages that were presented in Chapter 2. As explained in Section 1.1 above the discussion does not seek to be too prescriptive, but rather it endeavours to present the different approaches that might be of value and to provide some broad guidance on the relative advantages and disadvantages of each approach. It builds heavily on the material on models in use in the UK and abroad that was assembled in the work area B1 and B2 model review reports and tries to not duplicate these reviews. The reader is advised to consult these reports for further details and bibliography on the functioning of the various models that are mentioned below.

The level of detail in the coverage in the sections below will vary. In sections where past UK modelling experience is limited, such as 4.1 on incremental modelling or 4.6 on distribution and logistics, significant detail is included below. In other topics that have a longer track record, such as mode choice, much of the relevant material has already been presented in the reviews in B1 and B2. To avoid overburdening the text with mathematical detail, the mathematical specification of the approach has been mainly relegated to Appendix III. This Appendix nevertheless is important in that it provides a more complete and formal description of the precise manner in which the approach discussed below could be implemented within a feasible model structure.

This chapter concentrates on the creation of a national level model. It is important to appreciate that a model that operates at the national level is not synonymous with a model of just long distance freight travel. Over 50% of tonnes lifted on road goods vehicles have a length of haul of less than 50kms but these correspond only to 11% of total tonne kilometres moved on road (CSRGT, DfT, 2001). The national level freight model should include within it *all* freight movements, whether they be long or short distance. Otherwise it would only provide an incomplete measurement of the impacts of national policies that have ubiquitous effects, such as fuel taxation, vehicle pollution emission controls, etc. However, the representation of shorter distance local freight movements in the national model may well be rather cruder and more aggregate than that which could be developed in a smaller scale model of a specific individual conurbation.

The current chapter adopts a macro-simulation approach that broadly covers the Upper Level Model or strategic tier of the proposed LATS freight model (Neffendorf et al, 2001). Accordingly, the material below will provide an important input also to models that are constructed at the more detailed spatial scale. Other more specialised issues related to freight modelling at the more detailed spatial scale are discussed in the later Chapter 8, including those related to the use of micro-simulation as proposed for the Lower Level LATS Model. Although the micro-simulation approach could be applied beneficially at the national scale, it is likely to take time before this is feasible in practice and it may be better to first develop and use it successfully at a smaller scale before initiating its use at the national level.

4.1 Incremental modelling procedures that use base matrices

A fundamental aspect of the freight model design being proposed is the use of some form of incremental modelling approach of the type described in more detail below. The use of incremental models was also a strong recommendation of the earlier National Transport Model Feasibility Study – Project 1 (MVA, 1997). It should be emphasised here that *incremental* modelling does NOT refer to time stepping towards the future year, NOR does it imply the use of a lagged-response interactive land use/transport model. Instead, it simply means that if costs or other factors do not change at all relative to the base year then the base year matrices will be reproduced. However, when costs do change from those of the base year the model will predict the consequential change in freight traffic. This is sometimes known as pivoting about the base case, and alternative methods can be applied to achieve this which are discussed below.

Methods for the construction of base matrices that are closely tied to observed transport patterns have already been discussed in Section 3.3. These base matrices are then used as a foundation on which to construct the model in order to ensure that in the base situation the model results will validate well when compared to observed values.

However, the usefulness of a model does not reside primarily in its ability to match the observed base situation, but rather in its ability to forecast how this base situation would evolve in future years or in response to policy initiatives that are to be tested. In the past the users of models have often been faced with an unhelpful choice between either using a *synthetic* matrix model in which O/D trip distribution matrices by mode are constructed endogenously within the model (typically using a hierarchical logit based distribution and mode split procedures), or else using *observed* base year modal matrices that are exogenous to the model. In general when assigned to the modal networks *in the base year* the synthetic matrix approach is likely to validate less well than would do an observed matrix. However, when *policy testing in a future year,* the demand pattern exhibited by a synthetic matrix may well prove to represent travel demand responses more accurately than would a fixed matrix model, since it allows a much wider range of behavioural responses to the policy when moving the pattern of travel demand forward through time. In freight models the responsiveness to cost will be of particular importance when representing the choice of logistics options.

The performance of synthetic freight models in the UK that has been reviewed in the work area B1 illustrates the difficulties that they encounter in trying to provide a high level of explanation of current patterns of distribution or modal split. Accordingly, the preferred way forward is to obtain the best of both worlds through combining the relative strengths of the synthetic and the base matrix approach. This is what the category of incremental models sets out to achieve.

4.1.1 Why are incremental-type models needed?

A mathematical model of the choice processes of individual shippers is an attempt to represent real world behaviour through focusing attention on the more important relationships of interest, and simplifying or discarding other aspects that are either more difficult to model or are less relevant to the issues being analysed. For example, in a standard aggregate logit model of mode choice the elements represented explicitly for each mode are typically:

- the cost of travel
- the time of travel
- the quality of service (e.g. reliability, flexibility, security).

A calibrated model which includes just these three elements will not give a perfect estimate of the observed modal split pattern, since there will be many extra local aspects which influence shippers choice behaviour in each specific zone pair, and there may be local peculiarities introduced by the spatial aggregation into zones.

In some situations it might be reasonable to assume that these local aspects would be little affected by the types of policies to be tested within the transport model, so that rather than ignoring them completely, it would be desirable to carry their influence as a constant element through the modelling process. This is what an incremental modal split model seeks to achieve. (The term "pivot-point techniques" has also been used by Cambridge Systematics (1976) to describe this approach). The basic approach has been described in Bates et al (1987) as

> 'The incremental logit model as described by Kumar (1980), operates by implicitly estimating a modification to the modal constant, <u>appropriate for the particular movement under consideration</u>, which ensures that the model corresponds with the observed proportion. It is not, however, necessary to calibrate this adjustment explicitly, since a formula is available based directly on the observed values.'

Because the patterns of demand for freight transport generally appear to be more heterogeneous than those for passenger transport, there is benefit to adopting an incremental approach in order to lessen the impact of this heterogeneity on the ability of the freight model to match reality.

4.1.2 Variants on incremental models

Having established the need to use incremental-type models and having previously explained how the required good quality base matrices might be estimated, the next step is to outline the options available to make incremental models operational.

There are two main ways in which the incremental approach can be implemented. These can be illustrated using the logit modal split model as an example (noting that these methods can equally be generalised to represent other model stages).

1. The *incremental disutility change* approach was introduced in the papers by Kumar (1980) and by Bates et al (1987). To test a policy alternative, *a*, for the zone pair *i,j* and mode *k,* use the base case matrix of trips $\widetilde{M}_{ijk}$ and the *increment* in travel disutility *u*

   $$\Delta u_{ijk} = u_{ijk}^{a} - u_{ijk}^{b} \tag{4.1}$$

   of the alternative less the base case, *b*, in the (previously calibrated) logit model to estimate the modal demand in that policy as:

   $$T_{ijk}^{a} = T_{ij+}^{a}\left(\widetilde{M}_{ijk}\, e^{-\lambda \Delta u_{ijk}} \Big/ \sum_{k'} \widetilde{M}_{ijk'}\, e^{-\lambda \Delta u_{ijk'}} \right) \tag{4.2}$$

   Any disutility components that are common to the base and alternative policy case would cancel out in the equation (4.1) and so need not be included explicitly therein.

2. The *residual disutility* approach was introduced by Williams and Beardwood (1993). After the modal split model has been fully calibrated globally, the trips by mode are constrained to match the base case modal pattern for each zone pair. This constraint is implemented through adding to the travel disutility a value, z_{ijk}, called the residual disutility adjustment. The value is calculated such that an exact match of the modal flows to the base matrix is extracted for each zone pair and mode combination (it is an *explicit* representation of that which is *implicit* in the incremental disutility change approach). In subsequent model runs to test policy initiatives, this matrix of residual disutility elements is input to the model and added to the currently computed travel disutility for each zone pair and mode combination.

   $$T_{ijk}^{a} = T_{ij+}^{a}\left(e^{-\lambda\left(u_{ijk}^{a} + z_{ijk}\right)} \Big/ \sum_{k'} e^{-\lambda\left(u_{ijk'}^{a} + z_{ijk'}\right)} \right) \tag{4.3}$$

> In this approach the base trip matrix elements values and the disutility value for the base policy are not needed when running policy tests – the relevant information has all been summarised within the residual disutility value.

The work of Bates et al (1987) generalises Kumar's work on the single level multinomial model to cover the case of the hierarchical (or nested) logit model. This enables a set of linked logit models to be operated in an incremental form with information on the incremental change in the disutility of choices being passed up the decision tree in a consistent fashion using a special form of the log-sum formula for compositing disutilities.

In general there are not fundamental underlying differences between the two styles of approach to incremental modelling, rather they have operational differences whose implications relate more to: ease of set up, the flexibility of the model, and the design of the software on which it is run.

The initial calculation of the residual disutility values is a relatively complex estimation task if models have many hierarchical levels. Nevertheless, this method has been successfully implemented for the APRIL model (Williams and Beardwood, 1993) which is highly segmented and contains a multi-level choice process of destination, mode, time of outward and time of return trip for tours. This demonstrates its inherent feasibility even for complex model structures. It integrates well with the process of creation of base matrices that was discussed in Section 3.3 since the residual disutility values can be estimated directly as part of the base matrix estimation process, in a manner that takes account of the relative accuracy of different sources of observed data. It is relatively easy to introduce into standard modelling packages since it slots directly into standard procedures for passing composite utility terms up through the choice hierarchy.

The incremental disutility change approach avoids the need for the relatively complex initial step of calculating the set of residual disutility values. Accordingly it can be used straightforwardly in situations where a complete set of base demand matrices is directly available from observations. This advantage may be lost in more complex base matrix creation procedures that merge various data sources and require the use of synthetic models to implement segmentation of categories or the smoothing of sampling errors.

In summary both approaches are likely to be suitable to meet the needs for incremental modelling in the hierarchical structure for the national freight model that is presented in Figure 2.2.

4.2 Forecasting future trade patterns – the value and volume of freight generated

This task is concerned with forecasting the overall growth in the demand for freight in future years. It includes growth both in the volume moved and in the average distance over which this volume is moved. In this section the emphasis is on growth trends through time, taking a top-down approach. The local changes are discussed separately in Section 4.3 below which looks at zonal estimates of freight generation and attraction. This top down forecasting approach is operated where possible using a relatively detailed segmentation of commodity types. Individual commodity types have exhibited distinct transport patterns and growth trends in the past and are likely to continue to do so in the future. This heterogeneity cannot safely be ignored.

A number of the past forecasts of future growth in freight demand in the UK that are described in the work area B1 report were carried out independently for road and for rail, for some or all commodities. This is likely to have been partly because in certain of these studies the commissioning client only had responsibility for forecasting one of the modes. **Forecasting the demand independently for each mode is NOT an approach that is recommended here**, as is shown in Figure 2.2 by the placement of modal choice some stages below the demand generation stage.

The main modelling options for the generation stage that have been identified in the reviews B1 and B2 are summarised as follows:

1. Physical trend/time series based macro forecasts of *tonnes lifted* and of average lengths of haul for a set of commodity types. These have been used in various past UK studies and elsewhere.

2. Systems dynamic models.

3. Macroeconomic or input/output based approaches that estimate the *value* of domestic, import and export trade for a set of industry sectors and then convert this trade from units of monetary value to tonnes/containers by commodity type. Spatial computable general equilibrium (SCGE) models are another variant of this group. These approaches have been widely used outside the UK.

Many physical trend models simply extrapolate historical trends into the future, and therefore lack a behavioural basis, and in particular a link to the economy. More sophisticated time-series models can be developed, incorporating explanatory variables such as GDP, the tonnage or the value of sectoral output as described in Report B1 under the heading WP399 (Fowkes, Nash, Toner and Tweddle, 1993) and the SKM Additional Freight Market Model. The approach of WP399 was adopted in the NRTF 1997 study DETR (1998) which used time series to estimate the GB *road tonnes lifted* for each of 15 commodity sectors as a function of two explanatory terms: the *value* of domestic output (including exports), and the *value* of imports, applying lags where justified by the data. Future year forecasts were then based on exogenous estimates of economic growth in each of these sectors. This approach contains embedded within it a trend in volume to value ratios, which is estimated implicitly from the time series for each sector. The experience with these and other models (such as the international trade model of MDS-Transmodal that is used to feed their GBFM) suggests that estimating individual models at the commodity sector level and then aggregating them, provides better overall results than a model in which the sectors have been aggregated prior to the estimation.

The generation of future freight demand is best estimated at the segmented sectoral/commodity level and not as a total estimate of tonnes, to be subsequently disaggregated by commodity type.

The emphasis within this study is on forecasting the longer term freight growth, rather than on shorter term, year on year changes in total traffic. These latter short term aggregate traffic trends are likely to be best modelled by methods such as Box-Jenkins auto-regressive time series analysis, which are specifically designed to represent cyclic phenomena. Such methods have a good track record in estimating *short term* macroeconomic trends. However, because they tend to be operated in isolation from the underlying behavioural factors creating these trends, they are not necessarily suitable for use over the longer (10 to 15 year) time horizon in use in transport planning. Because of their lack of a behavioural basis these are unlikely to be of great use in policy testing.

The usage of systems dynamic models has appealed to some modellers (e.g. the ASTRA European model), but often systems dynamics models have had significant question marks over their calibration and validation. The methodology facilitates neither the detailed segmentation of space and of demand nor the rigorous statistical estimation of parameter values, which are typical in other approaches. It is not obvious that systems dynamics would be the most promising approach for the GB NFTM.

The third group of models, which forecast the value of trade, includes a number of different strands. These are not very easy to separate entirely from each other, given the overlaps in their methodologies and the benefits to be obtained from combining the best features from each.

(i) **Aggregate macroeconomic models.** Here the future value of production is estimated at an aggregate level *with little distinction by economic sector*. These aggregate growth rates are subsequently used to influence the growth in trade within detailed economic sectors in order to guide the growth in physical tonnes by commodity type.

(ii) **National or regional input/output models**. Here the future value of production is estimated at a detailed (20 to 100+) economic sector level using an I/O model (Leontief, 1986) *which does not represent the trade patterns between zones*, other than perhaps as a global set of sectoral imports and exports for a nation or a region as a whole. It then needs to be coupled with a spatial distribution model of some sort in order to create the spatial pattern of trade between zone pairs. The TEM-II, SMILE and SAMGODS models are examples of this type.

(iii) **Multi-regional input/output models**. Here the future value of production is estimated at a detailed (20 to 100+) economic sector level using an I/O model *that explicitly represents the trade patterns between a small number of regions*. This methodology developed by Leontief and Strout (1986) is rather cumbersome to implement and so has not been widely used in the field of transport.

(iv) **Spatial input-output models**. Here the future value of production is estimated at a detailed (20 to 100+) economic sector level using an I/O model *that explicitly represents the trade patterns by sector between all zone pai*rs. In contrast to the item (ii) above this methodology includes the spatial distribution model in a form that interfaces to the input/output calculations in every zone and so is able to estimate the geographic patterns of trade directly. The SCENES, EUNET and Sao Paulo State (Williams and Echenique, 1978) models are examples of this type.

(v) **Spatial Computable General Equilibrium (SCGE) models**. This contains a range of model formulations ranging from: the relatively abstract models of Venables and Gasiorek (1998) which only have a limited geography through introducing 2 or 3 regions, but treat imperfect competition and agglomeration effects explicitly, via the models of Bröcker (1998) which have a large number of zones and few sectors, through to experimental variants of the spatial input-output models that contain within them substitution effects and explicit production price determination. In essence the SCGE models replace the fixed I/O coefficients by more sophisticated production functional forms.

The manner in which these various I/O related model types are structured is explained in detail in Appendix III.4. There are some key aspects to note about the usage of I/O based methods:

- In general the I/O coefficients that are used will be identical for all zones/regions unless specific sets of regional coefficients have previously been estimated. No such recent comprehensive set of regional I/O coefficients appears to be available for the UK.

- The appropriate set of I/O coefficients to adopt is that from the I/O tables that are specifically designed to focus on commodity types, rather than on economic sectors[4]. This improves the ability to match to the commodity based categorisation within the CSRGT.

- When moving the model forward through a 15 year time period, there is a strong case for the I/O coefficients to be updated exogenously in order to take account of likely future structural trends in the economy, such as; higher proportions of imports and exports, and the increasing importance of the service sector. However, there does not appear to be widespread past experience by economists in carrying out such updating in practice as part of the forecasting of the future UK economy over the longer term.

The SACTRA (1999) report has recommended that further research be carried out into the use of I/O and of CGE models, in conjunction with transport models.

For purposes of providing a practical implementation in the immediate future of a zone based demand generation procedure, it may be premature to adopt the SCGE approach since it is still largely at the experimental stage. The IASON research project for DG TREN of the EU is currently exploring the practical use of SCGE models among others. Its ultimate usefulness in the longer term may be considerable so that **it would be valuable to carry out further research that addressed the issue of how to retain the theoretical richness of the SCGE approach within models that are a realistic representation of the current world patterns of location and transport.**

For the present some combination of the spatial I/O models and the more sophisticated and segmented versions of the physical trend models appear to be the most likely to meet requirements for generating the future overall demand for transport.

4.2.1 Volume to value conversions

In the approach based on the physical trend models the future evolution of the volume to value ratios is contained implicitly within the estimated model parameters. The exogenous estimates of the growth in the value of trade by sector (which incidentally are likely to have been generated through the use of I/O models in most cases), directly give rise to the growth in tonnes lifted by commodity type.

Within I/O model based approaches the future evolution of the volume to value ratios can be adjusted explicitly. Furthermore the conversion from the pattern of trade by detailed I/O commodity type to

---

[4] See Appendix III for an explanation of "Make" I/O tables and of the distinction between commodity and economic sector based categorisations

the physical volume by CSRGT commodity type (See Section 3.4 for a discussion on these categories) can be carried out in a flexible fashion. A specific I/O commodity can be matched to more than one CSRGT category if relevant.

In the absence of sources of domestic data on total volumes of production, the most appropriate source of information from which to estimate domestic value to volume ratios is to use the exports from the UK for which both value and physical volume data is published. The import based volume to value ratios are known, based on experience in Australia, to take different values and logically are appropriate only to apply to import trades. The use of combined rail plus CSRGT data as a surrogate for data on the volume of production is not ideal, other than for bulk commodities for which it is known that movements from initial production to final consumption do not involve intermediate distribution points. At present in the CSRGT there is no means of distinguishing the first haulage leg from the producing firm, from subsequent movements of the same goods further along the distribution chain, or to or from a rail terminal. Hence the CSRGT is not a consistent measure of the volume of production of goods.

4.3 Zonal estimates of tonnes produced and consumed

Some form of estimation of the volume of freight loaded and unloaded in individual locations or zones is required as a component of any network based freight model.

In a bottom–up approach (in contrast to the top-down approach of Section 4.2) the future volume of freight loaded/unloaded is determined directly from summing the individual zonal estimates. Given that there is at present insufficient direct data to provide good base estimates of freight flows at the zonal level, the bottom-up approach is not of widespread use for most commodity types. At the national level the exceptions arise for a few of the bulk sectors where there may be data available on the local capacity for production or consumption. When modelling at the more local scale it may be feasible through local surveys of large enterprises to obtain information on their productive capacity and future expansion plans. The manner in which the modelling of construction materials can be improved at the local level is discussed in Section 8.1.3.

The minerals, power and construction industries are the sectors where volume based zonal estimates of production and consumption of commodities are most likely to be feasible. The recent report (DTLR, 2002) on the *Survey of Land for Mineral Workings in England, 2000* will provide spatially coarse information on the future major locations for mineral extractions. Estimates of future dwelling construction and on major other construction sites (e.g. new road or rail links, or major urban redevelopment sites) can be used to estimate the demand for aggregates and other bulk inputs to construction which comprise a significant proportion of all tonnes lifted. Rubbish tips and landfill sites attract large volumes of HGV movements. The National Land Use Database, NLUD, may eventually be of use for estimating the location of production of agricultural production, but as yet it is some way from being complete.

For the non-bulk commodity types a top-down approach will generally be adopted. Appropriate zonal variables are used to disaggregate a control total volume of freight among the various zones. This spatial disaggregation may operate in various ways.

One task is the regionalisation of production or consumption trade values that are estimated at a national or regional scale using I/O or time series models. Typically the regionalisation will make use of employment data disaggregated by SIC (available from the IDBR down to ward level) since this is likely to be the main source of current local data in which detailed distinctions by industry type are available in order to match back to the economic sector of that trade.

At the production end of the movement of a particular type of good, the appropriate category of employment data to use will be that for the industry which produces the good. At the consumption end the estimation is more complex, since many different types of industrial, service and household sectors may each consume some amount of that good. The I/O coefficients can be used to estimate these amounts. This task is internalised in a consistent fashion as part of the standard procedure that is adopted within the spatial I/O models. This is one of their main advantages.

When disaggregating county level CSRGT/rail totals to a more detailed zoning system it will be necessary to introduce also the intermediate loading/unloading facilities appropriate to that particular type of good/stage in the distribution chain, and not solely the zonal levels of production and consumption. Rather than only using the employment in the distribution industry as the basis for this estimate, the use of the Valuation Office data on the floorspace for warehousing and storage may provide more discriminating estimates.

In the longer term if data on land use has been collected in the CSRGT a genuine zonal goods generation and attraction approach could be adopted as outlined in Section 2.5.1.

4.4 Creating the base P/C matrix of trade

The P/C trade matrix measures the volume of goods moved from the producer of each good to its eventual consumer. It is estimated for each of the commodity types individually. The matrix does not include an explicit representation of the actual logistical stages in which the transport and warehousing of the goods will occur along the way. Because there is currently no data source available that provides direct observations of this matrix, it is built up indirectly as an output from combining the observed data that is available within an appropriate model structure that explicitly includes the logistical stages. This spatial I/O model structure would provide this required functionality. It is explained in detail within Appendix III to which it has been relegated only because of its mathematical nature – it is of central relevance to the methods recommended for this study.

The main steps are as follows.

1. Estimate the zonal levels of production and consumption of the commodity using the methods described above.

2. Assemble observed O/D matrices of modal shipments by type of commodity from the relevant data sources: CSRGT, IRHS, RoRo, Rail freight movements, etc.

3. Merge the intermodal shipments. Multiple movements of a commodity that are measured as separate modal O/D shipment legs (e.g. a container moved on HGV from a factory to a rail terminal, then on rail to a deep-sea port) are collapsed into a single intermodal shipment from the factory to the port. If data later becomes available from the CSRGT on the land use at the origin and at the destination of consignments, this step will be simpler and more accurate to perform.

4. Decide on the number and the variations in distribution legs that are typical for the distribution chain of this commodity type, based on the literature on the logistics structures typical for this class of goods. If data later becomes available from the CSRGT on the land use at origin and destination of consignments, this analysis will become more robust.

5. Merge shipment legs within the distribution chain. This is the most complex step and would require more detailed research and development to finalise its mathematical structure. In outline it requires the following steps:

   (a) Set up the multi-leg distribution model for all commodity types as described in Section 4.6.2 below and in greater mathematical detail in Appendix III.5.

   (b) Run this model, aggregate the flows across distribution leg types and detailed zones, and compare the resulting estimated O/D totals against the observed O/D totals by type of commodity of step 3.

   (c) Adjust the model's parameters to ensure that the model's fit to the observed totals is commensurate with the expected accuracy of this observed data.

(d) When the calibration and performance of the model is satisfactory, convert it into an incremental form through applying an approach of the type discussed in Section 3.3 above.

The output from this set of steps is the required base P/C zone pair matrix of trade for each commodity type. The matrix is obtained by aggregation, through discarding the intermediate legs of the distribution chain, leaving only the initial origin and the ultimate destination, which represent the location of the producer and the consumer respectively.

The dimensions of the O/D matrices in use in the step 5 above will be large say, 600 x 660 (zones) x 20 (commodities) x 3 (logistical stages) x 2 modes. The inclusion of the intermediate locations of the warehousing used between logistical stages does not add to the dimensional scale of the task, since these are explicitly included through means of the “logistical stage” category of the movement. For example, the destination zone of a movement of logistical type: from producer to warehouse, is automatically a warehouse. The spatial I/O structure, as explained in Appendix III.5, does not retain memory for a good that is leaving a warehouse of precisely from which zone this shipment of the good was originally sourced. In this manner the excess dimensionality is avoided without loss of realism.

When data on the freight carried on LGVs becomes available from the LGVS it can be used to supplement the CSRGT. The LGVS does not directly measure the tonnage carried (only the proportion of the vehicle that is loaded) or the commodity type of what is carried, so that the ability to match it finely to the CSRGT categorisation of commodities will be limited. This is not necessarily a major issue since at the final leg of the distribution chain to consumers, where most LGV activity will take place, there may be little benefit to maintaining much segmentation of goods other than between high value/express and other.

This whole procedure of estimating and understanding the linkages within the distribution chains of commodities would be greatly facilitated if there were a different type of survey available that tracked goods through from the initial producer to the final consumer. This would accumulate information on any storage, assembly, repackaging, mode change or other activities along the way. This is discussed further in Section 7.2.

It is also clear that this process of model construction will need to be automated as far as possible, as new sources and survey cycles become available of the data that was used for model calibration.

4.5 Updating the production to consumption zone matrix

The production of a future year P/C matrix involves two main steps that are discussed below:

- forecasting the zonal pattern of production and consumption

- taking account of changing trends in patterns of trade.

The zonal pattern of production will be dependent on the future location of the industries that produce that specific commodity. The development of an explicit industrial location model is a substantial task in its own right. The EUNET Trans-Pennine model provides an example of how to model changes in the location of industry in response to changes in the transport costs of labour and of the goods that are input to production (see Appendix III.8 for more details). This method could now be extended to include other components of location costs such as land/floorspace, because the Valuation Office has recently introduced the publication of the floorspace statistics that are required.

It is argued by some that the location of industry has a relatively small elasticity to freight transport costs, so neglecting this influence may perhaps be justified on grounds of simplicity. On the other hand there have been major shifts in production patterns, with many metropolitan centres for traditional industries declining strongly, whereas high tech. industry is booming in many areas in the South of England that have had little or no previous industrial tradition. Accordingly, the accuracy of future estimates of the pattern of production of freight transport will be influenced by the ability to forecast future changes in industrial location patterns. This will be of particular importance at the finer spatial scale, since models at the national scale will be less sensitive to local inaccuracies in the forecast locations of firms.

If no industrial location model is created, a simple trend based approach could be applied which uses the TEMPRO employment growth forecasts to extrapolate into the future the present pattern of employment by type. This approach in isolation would not provide a very accurate foundation for future local estimates of freight growth. It should be complemented by the use of forecasts from other data sources for bulks (e.g. coal, aggregates, etc) and construction as discussed in Sections 4.3 and 8.1.3 respectively.

The zonal pattern of consumption of goods would be dependent on the location of industries, offices and households. Information on zonal numbers of employees and households could be provided by TEMPRO. The methods described in Section 4.3 would then be used to generate the zonal production and consumption volumes by type of commodity. These would be used as control totals for the rows and columns of the future year P/C matrix.

The second part of the matrix updating procedure is to adjust the base year matrix of trade to take account of the influence of:

- changes in the costs and quality of service of transport
- lengthening of trips in response to structural changes in industry or changing tastes of consumers.

The main changes in freight transport demand are likely to be as a result of the changes in mode and in logistical distribution patterns in response to transport supply changes. The representation of such responses is covered in later sections.

However, there are other structural aspects that also govern demand. Increases in incomes encourage consumers to broaden or refine their tastes, so that a much wider variety of goods is consumed than previously (beers from India, Japan, Italy - cars from all parts of the globe – regional delicacies from around Britain: cheeses, clotted cream, etc.). These involve longer patterns of supply to consumers, with many cross movements due to differences in the characteristics of the goods provided (e.g. Rover cars to Japan, and Toyota cars to the UK).

Changes in the structure of industrial production have also created longer movements. Globalisation and specialisation of manufacture to achieve economies of scale in production have meant that firms are sourcing their goods from other branches at considerable distances. The automotive industries now source many materials at the continental rather than country level. When modelling the future average length of P/C movements it will be necessary to make some assumptions on the continuation of the past increase in these lengths for each commodity type.

The mathematical methodology by which this can be achieved is presented in Appendix II. This shows how trip length patterns from a base matrix can be modified, either endogenously in response to changes in the generalised cost of transport for the commodity, or exogenously to represent the influence of structural changes. The resulting future year P/C matrix is then input to the procedure that divides it up into the legs of the distribution chain, as explained in the following section.

The methodology in Appendix II enables the influence of changes in the number of tonnes lifted, to be distinguished from changes in average lengths of haul, and so it would meet the requirements of SACTRA for improvements in methods for the creation of future matrices from base observed matrices.

4.6 Distribution and handling factors

This stage aims to convert from P/C matrices segmented by commodity type to O/D matrices segmented by shipment type by aggregate commodity type. The relevant past modelling approaches that appear to be suitable for use are from applications outside the UK. The proposed approach represents the impact of handling factors in a rather different form to the key ratios approach used in

GB in the past. It builds them up through explicit choice mechanisms rather than through aggregate numerical ratios.

4.6.1 The current state of the art

A model that has been important in successfully bridging the gap between strategic transport models and an explicit representation of logistics issues is the Dutch Strategic Model for Integrated Logistics Evaluation (SMILE) model (Tavasszy et al. 1998). This does not imply that SMILE itself directly provides a model platform that is recommended for the GB National Freight Model, but rather that it introduces many of the logistics concepts and features that need to be considered for inclusion within this model. The SMILE model has a very detailed representation of commodity types and logistics stages, a detailed representation of time (it is run at yearly intervals) but only a coarse representation of space and of transport network supply characteristics.

SMILE makes a distinction between product logistics and transport logistics (Tavasszy et al. 1998):

> "*Product logistics* has to do with the control of goods flows from basic products, via inventories and intermediate production activities till the physical distribution of final products to the customers."
>
> "*Transport logistics* involves the optimisation of the organisation of freight flows so that the utilisation of transport equipment is optimised, considering costs and quality elements such as reliability and speed."

The product logistics activity encompasses the two stages of: spatial distribution of trade ("production phase" in SMILE), and the spatial distribution of shipments ("inventory phase") in Figure 2.2, while the transport logistics activity contains the stages of mode choice, and assignment to vehicles and routes that are part of a typical freight transport model. SMILE introduces the concept of *logistical families,* which are categories of products that have homogeneous distribution and transport requirements, namely value density, packing density, delivery time, shipment size, demand frequency.

The Spatial Logistics Appended Module (SLAM – also developed by TNO), within the SCENES (2001) passenger and freight model of the EU, provides useful pointers on how to transfer key features from SMILE into an operational, spatially detailed, network based model. It transforms the P/C matrices of trade that are estimated by the SCENES regional economic model, into O/D matrices of shipments that incorporate alternative distribution chains. SLAM calculates the number and potential location of distribution centres (DCs) at a regional level throughout Europe by re-assigning tonnes per O/D to different transport chains. If a high number of tonnes pass though a region, then that region will be more attractive as a location for a DC.

Some of the features of the SLAM model have similarities to the freight model for the State of Sao Paulo, Brazil developed by Williams and Echenique (1978). This was one of the first operational applications of a spatial input-output modelling approach to transport demand estimation. The basic mathematical formulation in that model is the same as that currently in use in models such as EUNET and SCENES but the Sao Paulo model had two distinctive features:

- it used an input/output structure denominated in physical units (tonnes) rather than in monetary value units, because much of the Sao Paulo economy in the 1970s was focused on bulk rather than on finished goods

- it represented stages in the distribution chain through explicitly representing major distribution centres and subdividing the transport movements to and from them, for example, agricultural products were modelled as: farm to major wholesale, then wholesale to local retail, and petroleum products had three stages: to refinery, then to major distribution and then to final retailers.

This representation of distribution chains was carried out directly using a standard spatial I/O model mathematical formulation, which proved to be a simple and flexible means of representing distribution and warehousing activities (see Appendix III). The extension of this formulation to provide an economic framework for the estimation of zonal location patterns for activities was presented in Williams (1979).

4.6.2 The recommended approach

The aim of this spatial distribution and logistics module is to convert from P/C matrices segmented by commodity type to O/D matrices segmented by shipment type and aggregate commodity type.

It would include the following steps.

1. Identification of the zonal locations by commodity type for major warehousing and distribution centres. This would primarily use industrial sector information for large retail chains, petroleum, etc. supplemented by employment data by SIC from the IDBR and floorspace data from the Valuation Office. The recent study by the Highway Agency and the SRA of intermodal freight terminals may provide some relevant information also.

2. Identification of logistical families and definition of the categories to use within this stage of model including number and type of distribution chain stages, categories of consignment sizes, etc. for each commodity type.

3. Implementation, calibration and validation of a model of choice between logistics options (e.g. direct door-to-door, single intermediate warehouse, …) for each relevant commodity type. This would require the assembly of survey data on current patterns of distribution and of average handling factors for each commodity type. The CSRGT does not currently collect relevant information on the type of land use at which goods are loaded or unloaded, but could potentially be extended to do so in future years (see Sections 7.1 and 7.2).

4. Subdivide the movements into proportions by consignment size as appropriate for that category of commodity types and distribution stage. This can be a set of fixed proportions estimated using observed data from the CSRGT on size of drop, and from industry based information. The ideal source would be a survey of consignments, of the type that has been carried out in some other countries, such as in Italy for the construction of its national freight model, and in France and Germany as part of the MYSTIC study.

5. Output the O/D matrices of shipment legs that are segmented by consignment size, by stage in the distribution chain, by broad groups of commodities.

A decision will need to be finalised on the level of detail that is feasible in the representation of the logistics options in the step 3 above. It is not feasible to explicitly represent every distribution centre for every type of good within the UK. There would be too many of them and many would also be too small to be of any interest to be included explicitly within the model.

However, certain major distribution centres are large enough to be of considerable interest in their own right. These may also be locations where major road/rail interchange takes place so there may be other reasons also for wishing to represent them explicitly. This suggests that whereas a few major distribution centres may be explicitly located and connected to the physical network, the majority of the smaller distribution centres would effectively be treated as shadow distribution centres. These would be represented indirectly within this stage of the model in order to take appropriate account of the average diversions from the direct P/C movements that are created by the use of intermediate distribution centres. Data on their locations could be assembled from industry sources, as well as from national data sets on employment by SIC such as the IDBR and the Population Census of 2001.

4.6.3 Segmentation and local distribution

Although the proposed combined segmentation by consignment size, by stage in the distribution chain, and by broad group of commodities may appear to be excessively ambitious, it merits serious consideration. It is desirable in order to improve the robustness of subsequent stages of the transport model. The modelling of modal choice behaviour, the size of vehicle used and the transformation to vehicle loads are all closely related to these dimensions, and accordingly this segmentation has the potential to simplify and increase the accuracy of these later model stages.

It also relates to aspects that are discussed in more detail in later stages of this report. The local distribution of consignments to small retailers and to final consumers is an important stage, both for the modelling of the freight component of the LGV sector and for the modelling of freight within urban areas, where such movements are likely to be of much greater importance than on the interurban networks. It is also important to distinguish these local deliveries to final consumers, from the long haul trunk movements. The former are often constrained to operate by day when shops and offices are manned, whereas the latter can travel by night to avoid congestion where required.

For some movements, such as non-bulk flows to and from deep-sea ports, it may be helpful to convert from units of tonnes to units of containers at this model stage. This would enable a wide range of non-bulk products to be aggregated together into a single category (or possibly two categories: high value and other) with homogeneous requirements.

4.7 Estimation of cost functions

Ideally it is the user tariffs that are actually charged to shippers that should be utilised in the transport model to influence their choice of transport facilities. However the commercial sensitivity of this information makes it unlikely to be available in a comprehensive manner. Given the competitive nature of the freight industry (at least on the road haulage side) it has been commonplace to input operating cost functions instead to these models to represent the monetary influence on the choices of shippers. The assumption is that if the operating cost characteristics change significantly, then this will eventually filter through to modify the tariffs to shippers. Own Account transport is still important, and costs here will typically be less than charged by third party hauliers.

For reasons of computational efficiency many freight transport models adopt forms for vehicle cost functions that aggregate the various components of cost into a single equation that may be a function of the travel distance, travel time, vehicle type, etc. Likewise aggregate cost functions may be defined on the network for freight related activities such as warehousing and storage, transhipment, loading and unloading, etc.

This aggregation of the various underlying components of cost, such as capital, taxation, fuel, labour, maintenance and parts, administration, etc. into a single overall function for use in the model is not in principle a problem. What is required is that there exists outside the model some procedure to identify the relative contribution of each of these component cost elements. The DfT's Transport Economics Note (TEN) provides current information on a number of these cost components for road goods vehicles. The UNITE fifth Framework research study for the EC is a potential source of cost information by mode for the UK and other EU member states.

The key requirement is an explicit identification of the means by which the operating costs are built up through combining all of these individual cost components. In this manner there is a direct mechanism to estimate the extent to which a transport policy initiative can feed through to change the parameters of the aggregate operating cost function. The example of the influence of the EU Working Time Directive that was introduced in Chapter 2 of the Inception Report illustrates the approach by which a policy initiative may impact directly or indirectly on many separate cost components. These changes in unit costs should be quantified, aggregated and then input to the transport model through modifying the existing cost function parameters.

Cost models are needed for each of the vehicle and train types modelled to calculate and aggregate the cost components. The Vehicle Market Model (VMM98) provides considerable information on the stock of road goods vehicles and their operating characteristics. Cost models are also needed to cover the operations of all of the infrastructure facilities that have charges that filter through to the shippers. Section 1.4 of the Report A4 on rail freight provides a more detailed framework that can be used to structure the main cost centres into: track/road, traction, wagons, terminals/depots, and service providers. The various cost components can then be identified for each of these cost centres. These cost models could be developed as separate linked spreadsheets that have as outputs aggregate cost functional forms and their associated parameter values. These are then input directly into the transport model in order to take account of policy initiatives or external trends that would influence the operating cost structures of vehicles or infrastructure. The overall impacts of a policy initiative are then measured through rerunning the model and processing its results through a suitable assessment procedure.

The representation of the quality of the transport services provided is an important feature in freight modelling. It has been included explicitly on the network in the Swedish SAMGODS model as a set of quality-related costs to represent the risk of delay, a frequency related term and the value of time. These may be associated with transfers and storage or with the haulage of the goods. This model like other models that have been implemented using the STAN package carries out a joint mode and route choice procedure.

In many of the other models reviewed the quality of service is represented as part of the modal split procedure through the utility of the various characteristics of the modal services on offer.

4.8 Modal split

The modal split stage of the model takes as input the set of O/D matrices of shipment legs in units of tonnes/containers, which are segmented by shipment type (e.g. by consignment size, by stage in the distribution chain, by broad groups of commodities). It then splits these by mode and then by vehicle type within a mode, on the basis of the suitability and cost of each of the competing services.

A number of these shipment type categories will effectively be captive to road so that the mode split is simply between road vehicles of different sizes. For example, the deliveries of small consignments, those to dispersed final consumers, and those over short distances (except for regular, large consignment size, movements of bulk goods) are all likely to be captive to road. Accordingly, a number of these segments may be aggregated together at this and later stages in the transport model in order to save computing resources. In other instances, particularly over longer distances, rail may be a serious competitor to road, such as for containers from large deep-sea ports or for regular large-scale movement of bulk goods. The key issue is to start with a suitably homogeneous segmentation of shipment types that enables the choice sets to be realistically defined.

Freight mode split modelling is one of the model components in which there is already substantial experience both in the UK and abroad as outlined in the B1 and B2 reports. The preferred approach to the calibration of mode split models is through use of disaggregate calibration techniques as used in a number of the European models (STEMM, EUFRANET and Fehmarnbelt). This may be based on revealed and/or stated preference data on the responses of individual shippers. This would however require a significant investment in data collection in order to provide the necessary information, since existing data sources in GB do not cover this requirement, other than in a few specific studies (Fowkes et al. 1993, Fowkes & Toner, 1998). It may not be easy to obtain accurate, detailed information on the costs facing shippers, given the commercial sensitivities that are involved. Accordingly, a careful pilot study would first be required that tested the feasibility of such a major data collection exercise.

In the absence of suitable disaggregate choice data being collected the fallback position is to carry out an aggregate calibration at the zone pair level as has been done in calibrating many past models. However, this is inherently less satisfactory than a properly specified disaggregate calibration that is based on a sufficiently large sample to be statistically reliable for the large number of segments that are required. The Italian national model, SISD, uses a disaggregate mode split model to simulate mode choice for *each shipment*. This probably represents an upper bound on the sophistication that might be feasible in the GB model.

A number of models (e.g. EUNET and the Italian national model) have adopted separate mode and route choice procedures. Others (e.g. the Walloon and the Swedish and Norwegian national models) have adopted simultaneous mode and route choice procedures. In both cases intermodal transport is represented through the use of multimodal networks on which road feeder movements to rail are included. Moreover, the separate mode and route choice procedures typically are iterated until a joint equilibrium mode and route solution is achieved for the congested networks, so that they do not differ from the simultaneous procedures in this regard.

The GBFM adopts a slightly different approach by performing two assignments. The first is onto a spatially coarse multimodal network and serves just to determine the split between road and rail legs. The second then assigns the resulting road legs only, onto a detailed road network using a much finer zoning system for the O/D pairs and with more vehicle categories explicitly represented.

The main benefits from the use of a modal split procedure that is separated from the assignment lie in:

- its ability to take account of the stochastic variations in choices that are inherent in freight - this can partly be achieved if the assignment step is a stochastic (e.g. in EUNET) rather than a deterministic (e.g. in the Swedish and Norwegian national models) user equilibrium solution

- it facilitates maintaining a substantial degree of segmentation of shipment types at the modal split stage where it is most needed, without requiring that this number of segments be carried through to the assignment stage where it could overload the computational burden. The aggregation of some of the segments, after the modal split and conversion to vehicle loads has been completed, but prior to assignment, can be done with little loss in the quality of the model results.

There has recently been discussion (Transek et al. 2001) on the SAMGODS national model of Sweden to include an explicit mode choice model to substitute the simultaneous mode and route choice currently in place.

For road freight there is merit to distinguishing at least three vehicle types: LGV (below 3.5 gvw tonnes), small/medium HGV (over 3.5 but below 25 gvw tonnes) and large HGV. The CSRGT shows that this category of small/medium HGV comprises about 45% of the vehicle kilometres, less than 20% of the tonnes lifted and around 13% of the tonne kilometres moved within HGVs as a group. These percentages show that although they may not carry a large proportion of the total freight movements, the small/medium HGVs do generate a substantial amount of vehicular traffic. Within this category the vehicle kilometres of the smaller HGVs (3.5 to 7.5 gvw) have stayed broadly constant over the last decade, whereas the medium HGVs have lost market share to large HGVs. The average vehicle size within the large HGV category has also been increasing through time. These trends should be taken into account when forecasting, although an explicit distinction between small and medium HGVs may require undue computing resources. While an alternative segmentation between rigid and articulated lorries could be adopted, it would generate a greater variance in the average load per vehicle which would not be ideal.

However, the decision on the most suitable form of segmentation by vehicle type is not clear cut. One of the principal policy levers that might be used to discriminate between vehicle types is Vehicle Excise Duty (VED). The VED bands cover a wide range of rigid, articulated, gross vehicle weight, and axle combinations. At the moment, policies discriminating between VED bands are being tested. So perhaps about 20 combinations should be used (to include LGVs), in a form that would aggregate to up to at most 6 vehicle type categories for modal split. The model results could then be post processed into more detailed vehicle type categories.

For non-bulk commodity types the choice of size of HGV will be intimately linked to the stage in the distribution chain. For long hauls of large consignments the large HGVs will be used, whereas for the final distribution stage with its small drops to dispersed locations the use of smaller vehicles becomes more prevalent.

Unfortunately the CSRGT is not designed to provide any information on LGVs so that such data will need to be obtained from the Light Goods Vehicle Surveys that will soon be underway. In the early stages of freight model development it may be necessary to exclude LGVs until a suitable quantity of the required data has become available. However, suitable slots should be included within the model to enable LGVs to subsequently be inserted without major upheaval. Their main role on the freight side is likely to be in the shipment of high value, low volume consignments and in local distribution of consignments in small drops. The manner in which the modelling of freight moved on LGVs could be integrated with the modelling of HGVs is specified in Section 5.2 below.

On rail the distinction would be between the different types of service on offer, which may be linked to the types of goods being carried. A distinction between: conventional train-load, conventional wagon-load and intermodal services may be sufficient. Conventional services provide for railway shipments from one private siding/rail connected facility to another without any use of road. The wagon-load services have the disadvantage of potential delays at marshalling yards in the course of their journey. Intermodal rail services transport standard unit loads such as ISO containers, swap bodies or "piggyback" road trailers by rail for a trunk haul between two intermodal terminals, but local distribution is generally by road. This provides an added cost component that tends to make them uncompetitive over short lengths of haul.

4.8.1 Summary of the recommended approach

To enable the existing spatial pattern of modal split to be represented accurately, an incremental modelling approach is recommended.

Provided that resources are available to collect the necessary data, a disaggregate calibration procedure should be used. Quality of service characteristics should be included to the extent possible, and not solely the cost and travel time of the service. The validation of the model should ensure that the modal split is correct in total and that it correctly represents the changing competitiveness between modes by broad distance band as the length of haul increases.

Many of the studies reviewed have shown that there are strong differences between the service requirements of different freight market segments. Accordingly, the modal split model should be segmented into a sufficient number of shipment types to ensure homogeneous requirements and choice sets between modes. The Anglo-continental, the NERA et al. and the SKM studies reviewed in the B1 Report all provide potentially relevant insights for this task.

A modal split procedure that is separated from, rather than part of the route choice procedure appears to be the more appropriate.

Subdivision by size of HGV and type of rail service is desirable. This should be carried out using a nested logit structure or some other such procedure to avoid the red-bus blue-bus problem discussed in the B1 Report.

Although in principle it would also be of interest on road to further subdivide road services between own account and hire and reward providers, this increase in resource requirements would be significant. Model developers will have to take a decision on whether or not to include this split at the detailed model specification stage. If the subdivision is not felt appropriate, it should be feasible to handle it adequately instead as an external post-processing routine appended to the model.

4.9 Time of day choice

The set of issues that have been discussed in Section 2.8.1 point to the considerable benefits to be gained from including within the model a representation of the choice of time period of travel for road vehicle movements.

In principle the inclusion of time period choice can be carried out in a manner not dissimilar to the modal split procedure. Separate networks are created for each time period and the congested generalised costs are then obtained based on the traffic conditions in that period. This implies that the passenger and LGV traffic also needs to be known on each link for each time period in order for realistic congested speeds to be calculated. Vehicle operating cost functions for each time period are needed, taking note of any differences in labour costs that may be relevant. The traffic demand matrices need to be further segmented (on top of any other categorisations already in use) into a set of categories: that are captive to each specific time period, and those which have various degrees of propensity to switch time period. The relative cost/attractiveness to the shipper of goods departing/arriving in the different periods needs to be measured. A modelling methodology that represents time of day modelling of tours within a highly segmented incremental model structure has been successfully implemented in the APRIL passenger model as outlined in Bates and Williams (1993) and Williams and Beardwood (1993).

In practice, the inclusion of time period choice would have major impacts on the size and speed of the model, due to the need for further segmentation of the commodity/logistics/shipment regularity type to represent the specific requirements for travel within specified time periods. More pertinently, the data requirements to calibrate and validate a time of day choice component would be significant, since appropriate information for this purpose is not available at present. The CSRGT does not currently collect information on the start or end time of freight movements, though this is an item of extra data that is amongst those currently under consideration for the CSRGT.

Given the major added complexity from introducing time of day choice into the model it is not recommended for explicit inclusion in the national level freight model for the present.

At a later stage, depending on progress on data collection and in studies such as the ITeLS funded by the Department, time of day choice could be introduced to the model. It will be important that the model design ensures that such subsequent enhancements are not awkward to introduce.

What may be rather more feasible initially is to construct separate road goods vehicle matrices for different times of day. The key information for this purpose would be to combine both RSI and traffic count data, which do contain information on the time of travel, with the CSRGT data which does not. In this way base road vehicle matrices that are segmented by time of day could be created for use as appropriate. This may be best carried out outside the main modelling procedure to minimise the added computational burden.

This matrix assembly would also require the adoption of methods to deal with long distance and multi-drop movements that would by virtue of their duration span more than one time period.

4.10 Transformation to vehicle loads

The output from the modal split stage of the model is a set of O/D matrices of shipment legs in units of tonnes, which are segmented by consignment size, by stage in the distribution chain, by broad groups of commodities for each vehicle type within a mode. These tonnes then need to be converted into vehicle loads prior to being assigned to the appropriate modal network.

The conversion from tonnes to vehicles will depend on a combination of:

- The type of commodity – dense bulk commodities are restricted by weight, finished goods are often restricted by volume – the CSRGT provides useful information to estimate these values.

- The size of vehicle – again CSRGT provides relevant information.

- The stage in the distribution chain and the related size distribution of consignments – when delivering to final consumers and small retailers, the consignment sizes and requirements for multi-drops may militate against full loads – information on stage in the distribution chain is not currently available from CSRGT.

- The degree of empty running of vehicles – relevant information is available from CSRGT.

The Vehicle Market Model (VMM98) can be used to identify the likely future composition of different sizes and types of vehicles within the three road goods vehicle categories that are used. In this way trends in the average load that can be carried in each vehicle type can be estimated.

A perennial problem in freight modelling is the representation of empty vehicles, which form a non-trivial part of the total traffic and of the total costs of operation of both road and rail. The Reviews B1 and B2 have uncovered a number of approaches. The approach of balancing out and return vehicle movements that has been adopted in the Fehmarnbelt model seems interesting, but while it may be appropriate for an international corridor with limited connectivity, in the UK context triangular and more complex combinations of trips would lessen its robustness. The alternative approach suggested by ASH (1997) that each outward vehicle movement generates a proportion of empty return vehicles (dummy commodities that are calibrated by commodity type using the CSRGT) seems more helpful. It has been noted in many studies (WP399, NRTF 97) that the proportion of empty running varies substantially between commodity types. The work on key ratios by McKinnon also provides useful information on the nature and trends in empty running of road vehicles.

4.11 Assignment to the modal networks

The main options for consideration at this stage of the model are:

1. Simultaneous mode split and route choice or separate? As discussed in Section 4.8 above the separate approach is preferred.

2. The need to explicitly assign rail traffic to the network is not clear cut. It may be sufficient to simply use the network to accumulate the required travel supply characteristics. The EUFRANET model provides a detailed assignment of freight to the rail network, while the SCENES model includes an explicit representation of the schedules of major intermodal freight services, with other rail freight being assigned more simply.

3. Unimodal or multimodal assignment – even in the option to separate the mode split stage from route choice there is still a need to carry out the assignment to a multimodal network for any intermodal flows that use feeder modes as a stage in the journey. This includes both road as a feeder to rail terminals, and ferries/Channel Tunnel shuttle as feeders for international movements on road. The resulting feeder movements on road should be included in the equilibration of road congestion in the same way as would pure road flows.

4. The equilibrium assignment on road requires that passenger, LGV and HGV traffic are all assigned simultaneously. This would introduce the need for a loose linkage with the NPTM Pass3 assignment model so that matrices of vehicle movements can be transferred between models using a suitable common zoning system.

5. What type of equilibrium assignment should be used? There are advantages to the use of a stochastic user equilibrium (SUE) assignment especially for multimodal assignment as it avoids sending all traffic through a single path between a zone pair. Without this the choice of ferry for international traffic might be unrealistic, and tests of road pricing policies are less realistic. The use of a distribution of values of time as in the Copenhagen model is a potentially interesting approach to this. The choice between the use of SUE and deterministic user equilibrium assignment (DUE) is not critical because selecting the latter may facilitate access to a greater range of assignment packages.

6. What optimum function should be chosen for the assignment – user or system optimum? The STAN package that has been used in the Norwegian and Swedish national models calculates

the system optimum solution for mode and route choice, whereas most other models calculate the user optimum solution. This latter is the more relevant for modelling patterns of freight transport since movements are determined by the decisions of a large number of individuals concerned only with their own shipments and not with the overall system.

4.11.1 Multistop trips:

The movements that are used for local collection and delivery will often take the form of multistop tours/trips. Data on trips/tours with 5 or more stops are summarised in both the CSRGT and the LGVS. For these multistop trips/tours the individual collection and delivery locations are not collected, only the start and end base locations (if these are the same location it is a tour not a trip) and the total kilometres travelled. A means needs to be developed to assign these to the road network in an appropriate fashion until such time as better data is collected.

The following approach can be used:

1. Determine the most distant location reached. With the present data make this a location selected at a distance of x% of the total measured distance travelled in the tour. Typically x might be 30 to 40%. **In future the CSRGT and LGVS questionnaires should each be extended to request the location of the furthest stop on the multistop tour from the start base (or from the start and end bases if it is a multistop trip which does not return to the initial base)**.

2. Assign this multistop traffic from the start base to the furthest point and also from this furthest point to the end base. This should be a stochastic assignment with a low value for the path choice parameter. In this way the assigned traffic will not be concentrated on the shortest path but will spread onto less direct paths as would be expected to be the case for diversions to collect and deliver consignments.

An alternative approach using micro-simulation, as discussed in Section 8.2, could also be developed to assign actual individual multistop tours to the network.

5 THE MODELLING OF LIGHT GOODS VEHICLES

Although some limited aspects of the modelling of LGVs have already been discussed in previous chapters, other aspects of the modelling LGVs do not fall neatly within the earlier framework for modelling freight. The manner in which it is proposed to model LGVs is discussed now. It should be stressed that this chapter is more speculative than the rest of this report, due to the absence of a good database of LGV transport. There will be a need for a detailed analysis of the LGVS data as soon as it has been collected and processed, in order to finalise the formulation of the various components of LGV modelling. Some of the assumptions that underlie the model specification proposed below may turn out to be inappropriate in the light of a more complete picture of the operation of the LGV sector.

The modelling of the freight movements that are carried on LGVs should eventually be integrated with the modelling of HGVs within the freight model. It should become part of the overall freight modal split of road movements into different sizes of vehicles as described in Section 4.8 above.

The options of whether the modelling of LGV passenger flows should be standalone or be integrated as part of a more conventional passenger mode split model are discussed in Section 5.3. The assignment of the LGVs to the road network in all cases would be a conventional user equilibrium assignment of the type described in Section 4.11.

5.1 The classification of LGV trips

The definition of an LGV movement and the specification of where these movements are to be included in the modelling system have already been presented in the earlier Section 3.4 and are discussed further in Browne et al. (2002) . It is proposed that the following four groups of purposes should be distinguished for LGVs:

1. Freight related movements including empty running - this category of movement is explicitly excluded from the NTS and data can only be obtained in the future LGVS.

2. Commercial/business/service trips – a significant proportion of these are already captured in the NTS, but it excludes certain trips that are likely to be intensive users of LGVs. The NTS explicitly excludes trips for the following purpose categories:

   (i) to deliver goods,

   (ii) to convey a vehicle or passengers,

(iii) travel as drivers or crew of public transport vehicles,

(iv) travel in industrial or agricultural equipment,

(v) travel in specially equipped vehicles used in the course of a person's work, (including AA/RAC repair vehicles, and Post Office vans), and

(vi) journeys in course of work by people paid to walk or cycle (such as policemen on the beat, traffic wardens, leaflet distributors, messengers, postmen or roundsmen).

3. Commuting to normal place of work – in NTS

4. Other private (e.g. education, shopping, personal business, etc.) - in NTS.

The primary data source for the LGV generation stage will be the LGVS, since as explained above the NTS excludes important LGV trips. Furthermore, if it is only those trips within LGVs that are being analysed, the NTS has a rather small sample size, even when data is accumulated across years. The NTS could certainly provide some useful overall estimates of certain flow types and added background information, but it would not be the major source of data for the modelling of LGVs. The LATS 2001 will also provide potentially valuable information, though this will only be for the area around London rather than for GB as a whole. However, it could be difficult to clearly separate "goods trips" and "service trips that involve goods movement" in the LGVS survey. This may place data limitations on the distinction between group 1 (freight) and group 2 (commercial/business/service trips).

The definitions of trip purpose/land use type/vehicle type/ and of any other items that overlap between the LGVS and NTS or between the LGVS and CSRGT surveys, should be created in a consistent manner, sufficient to enable the data sets to be matched subsequently for these variables.

It is clear that for many years to come the cumulative size of the LGVS sample will be comparatively small so that the scope for detailed segmentation will be limited if sampling error is to be avoided. In particular it will at best only have a limited ability to distinguish spatial effects and would not produce a reliable detailed national O/D matrix of trips. Instead it will need to be supplemented with good land use data and traffic count data in order to expand the limited sample up to the national level.

5.2 Integration of LGV freight movements between models

The representation of freight traffic needs to be linked between the freight model itself, as described in Section 4, and the freight purpose within an LGV model. In practice there is no sharp break in the logistics chain between: the light goods vehicles that are to be captured in the LGVS, and the slightly larger vans that would be included instead within the CSRGT. Accordingly, the distinction that is convenient for data collection purposes between freight carried on LGVs and on HGVs should not maintain undue influence over the design of a suitable overall model for the movement of freight.

Until such time as the LGVS data has been collected, processed and analysed, there is no systematic source of data on the nature and scale of the freight movements on LGVs. It is clear that in overall performance terms (tonne kilometres) they are likely to be small relative to HGV movements, but in terms of vehicle kilometres or of value kilometres this imbalance may be reduced due to LGV's propensity to carry smaller loads of higher value products. In particular within urban areas, the movement of freight by LGVs plays an important role and one which may be susceptible to a number of policies on the pricing and permissions for different type of vehicles to move within urban areas at different times of the day. In this way the LGVs need to be considered as an alternative mode to HGVs for certain types of freight movements as is now explained.

Two types of market for freight movements on LGVs are distinguished.

1. The collection/delivery legs of small consignments of goods that are already included as conventional HGV/rail movements in other legs of the distribution chain. In this case there is a requirement to match these goods carried on LGVs back to the original P/C movement that gave rise to them. It is the same task as that already discussed in Section 4.6 with respect to the data from the CSRGT

2. the specialised (often high value) parcels/courier/post market where over shorter distances all legs of the consignment may take place in LGVs – though over the longer distances the trunk haul may be by HGV, rail or air. The high value and low density of this segment implies that its overall importance in terms of the economy or of vehicle demand is greater than might appear from a simple measure of the tonne kilometres that it generates.

The ability to distinguish this important movement of parcels/courier/post within the published CSRGT data will be very limited since it is included within the "miscellaneous articles" category. This comprises about 20% of all goods carried and contains a very wide mixture of goods, ranging from household waste and empty pallets through to arms and ammunition. These in total tonnes will dwarf the volume of parcels that are carried. Accordingly, it may be safer to derive the scale of this market from the LGVS directly, with supporting information from the CSRGT to the extent that it is useful. This is a topic on which data from within the industry for the larger operators might also prove useful, since these are the carriers that are most likely to operate on a scale that makes a significant use of HGVs to supplement the LGVs.

Care would be required in the definition of parcels etc. because the term parcels refers at least as much to the logistics operators/operations that handle these goods and the packaging of the goods as it does to similarity between the goods or their size/weight. The parcels handled by parcels carriers contain a wide range of products and come in widely varying sizes and weights. Therefore in transport terms one parcel is not necessarily similar to another. It is difficult to see how the type/weight of products carried on their final transport leg by parcel carriers can be estimated from published data. A survey would probably be necessary to obtain this data.

The processing of data in the CSRGT should be examined to ascertain whether a more differentiated classification of the "miscellaneous articles" category could be published for past and/or future years.

The approach to trip generation for freight in LGVs is the same as that proposed in Section 4.3 for freight modelling. It would require primary data collection of the collection and delivery of goods by LGVs and as such should be part of the same surveying exercise as for HGVs, since both will be in competition for at least some of these goods. The need is for access to sufficient reliable data to enable generation and attraction factors to be determined in the form already discussed in Section 2.5.1.

5.3 Options for the integration of LGV passenger movements between models

In the previous section it was explained that there is a need to integrate the representation of freight movements between models, it is logical to also consider the integration of passenger movements between an LGV model and existing passenger models such as the NPTM.

In principle it would of course be beneficial to include LGVs as an explicit mode in competition with car for passenger trips. For many tradesmen the choice between using an estate car or a van for their service trips may be a fine balance, so that minor changes in costs/taxation/tolling etc. could lead to significant shifts in the size of the LGV fleet. However, much of this choice between car and LGV takes place at the vehicle purchase stage rather than at the stage of making an individual trip. In this sense it is not simply a traditional trip based choice of mode (as for example between walk and car for a local shopping trip) and so would require to be dealt with as part of a vehicle stock model as well as within a trip based modal split procedure that takes account of the availability of cars as an alternative to the use of an LGV for a trip.

In practice the creation of a joint LGV and car passenger model would not be straightforward to achieve given the data sources that are available. The current NPTM development is based entirely on travel demand data that has been collected from the NTS. However, as has already been explained, the coverage of LGV trips within the NTS would not be easy to merge with the LGVS database, due to differences in definitions and due to the coverage of certain LGV trip types within the NTS being partial rather than comprehensive.

Three options merit consideration:

1. separate passenger models – here the underlying data used to calibrate the current NPTM is modified so that any trips associated with LGVs or HGVs are entirely excluded from the database for the model. The passenger model is then recalibrated where necessary based on this revised input data. An independent LGV passenger model is also produced which is based on LGVS rather than on NTS data. The main disadvantage to this approach is that there would be no elasticity of demand between cars and LGVs in response to policy changes. Furthermore the modelling of the private trips on LGVs independently from trips on all of the other modes raises problems of inconsistency in choice sets. Hence this approach is not recommended.

2. combined passenger and LGV model – here the data from the LGVS database is merged with the NTS to provide a set of combined trip rates and aggregate modal matrices in which the car and the LGV trips are distinguished throughout. Then the NPTM model is recalibrated with this combined data. The disadvantages to this approach is that it requires extending an already complex model with significant extra features and it requires integrating two data sources which have significant inconsistencies in definitions. It would be a substantial task.

3. intermediate combination – here the assumption is that the primary passenger role of the LGV lies within one specific segment of the current NPTM, namely the trips on the purpose 'in course of business', especially those by manual workers. The model would provide a special representation of this trip type. For other traveller type/trip purposes a simpler factoring of trips between car and LGV is applied.

The option 3 appears to be the most cost-effective approach and so is now discussed in more detail. There would be a separation of business travel between manual and non-manual workers. For the business trips by manual workers the NTS data would be supplemented by that collected from the LGVS, and a new joint model of car and of LGV movements would be created. It is assumed that business trips on LGVs by non-manual workers would be few enough to ignore. The validity of this assertion could be checked using the NTS data.

For each other trip purpose/traveller type combination a global proportion (calibrated using NTS and LGVS data) would be applied to each O/D movement by road, to reallocate trips between car and LGV prior to the assignment procedure. Other than this step the current NPTM model calibration would be maintained for these trip types. The accuracy of this step would be greatly improved through adopting a segmentation of travellers in which the non-manual, the skilled manual and the semi/unskilled manual workers/households are distinguished. There are likely to be strong differences in the proportion of households with vans between these three categories. This type of distinction by socio-economic group already exists in a number of land use models used to estimate the transport demand in the multimodal corridor studies, such as the LASER/LOIS, TAMMS and CHUMMS models, as well as in the EUNET regional freight and passenger model.

Where there is reliable current zonal vehicle stock data available, which provides the split between the number of cars and of LGVs available to households in each socio-economic group, it could be used to provide a zonal proportion, rather than a global proportion, for the split of O/D road trips between car and LGV for home-based trip purposes. The current national information on vehicle licensing may not be adequate for this purpose because in a significant number of cases for company vehicles, the address at which a vehicle is licensed may not be related to the address at which it may be available for use by households. This is likely to be particularly problematic for LGVs.

Prior to adopting this option of an intermediate combined model, the LGVS data should be analysed in order to ascertain that the overall characteristics of private trips on LGVs are not very different from the corresponding private trips by car as measured in the NTS. Specifically, for each of the traveller type/trip purpose combinations in the NPTM, the distribution of trip lengths should be compared between car and LGV to see whether there are major behavioural differences. If there are differences, then it would imply that for this trip type a simple factoring procedure is inadequate and that some form of modal split procedure should be used to take account of these differences in behaviour. For example it may be that for short distance social trips an LGV may indeed be used, whereas for longer distance social trips a more comfortable car is more likely to be used if it is available to the travellers.

It would be of interest to examine from the NTS the extent to which households with access to LGVs are without access to other cars. In this situation there would be strong incentives to use the LGV for private trips. However, where the household has an alternative car, the use of LGVs for private trips might be much less, other than for commuting to work. Since the NPTM already segments travellers by household size and number of vehicles available, this distinction would be feasible to introduce into the factoring procedure.

The modelling of the impact of Internet shopping could be implemented by adopting a more sophisticated mode split model for trips for the purpose, shopping. A new mode, LGV shopping delivery, could be introduced as an alternative to the travellers using their own car/LGV and spending time in travelling to the shop. Stated preference surveys could be used in order to calibrate such a model. To avoid double counting it would be necessary to keep the trips on this new LGV shopping delivery mode separated out from the general category of LGV freight movements. The reason why they are considered here instead of in the freight category is because they are substituting directly for private passenger shopping trips.

Some caution is needed here to identify the trips that fall into this category. It isn't necessarily the case that all LGV deliveries to home substitute directly for private passenger shopping trips. Some products delivered to homes by LGV are still viewed and purchased by customers in shops. Also a sizeable proportion of home deliveries are carried out by the Royal Mail, and it is not necessarily the case that individuals would make private trips to collect these items in the absence of Royal Mail deliveries (especially in the case of items sent by friends and relatives).

5.4 The spatial distribution of LGV trips

The primary data sources for the LGV trip distribution stage will be the LGVS and the traffic count data for LGVs. The LGVS will be used to identify the distribution pattern of trip lengths for each category of LGV movement. Initially a set of synthetic LGV matrices will be estimated for each of the passenger and the freight trip purposes. These matrices will take appropriate account of the specific patterns of trip length distributions and of origins and destinations relevant to each trip type.

The combined set of LGV matrices will then be scaled and adjusted in order to match to the traffic count data that is available, using the approach that has been discussed in Section 3.3. Because the LGV traffic count data provides no explicit information on the purpose of the vehicle movement, there is no alternative to combining all of the LGV freight and passenger movements together at this stage. This is in line with the earlier proposal that all of the different model components would need to combined within a single highway assignment stage, even though there may be strong arguments for keeping the other modelling steps separate between passengers and freight in order to lessen complexity and computational overload.

5.4.1 Counts of LGV traffic

Care will be needed in the use of the LGV traffic counts due to the problems that arise in visually separating between the categories of LGV and HGV as defined in the LGVS and CSRGT respectively. It is generally accepted that both visually based manual classified counts and automatic counts using vehicle length do not provide sufficiently accurate information to enable LGVs to be distinguished unambiguously from cars or small HGVs. The biases will differ between the methods used to carry out the counts. Accordingly, it should be accepted that the time-series count data currently available on LGVs is likely to be subject to a significant level of error.

Because the LGVS sample sizes are not very large and because past experience with surveys of LGVs has demonstrated that high response rates and reliable results are not easy to achieve, there is a particular premium on having a reliable alternative data source to enable the results to be scaled up to the national level. In their current form the existing traffic counts are unlikely to provide this reliable source but better candidates for this role are not obvious, other than the use of RSI data where it is available (e.g. LATS). Again, the RSI data sources are not without their own definitional problems for LGVs, even with respect to counts.

There would be major benefits from a detailed study that endeavoured to measure the source and level of the biases in the existing traffic counts of LGVs. This study should provide sufficient information to correct the categorisation by vehicle type of existing traffic count data to enable it to be broadly consistent on average with the definitions in the LGVS and CSRGT. It would require a variety of case studies of counts carried out in the traditional manner in different types of location. Each would be carried out in conjunction with some precise means of identifying each vehicle's correct characteristics (say, a combination of video number plate matching to DVLR records of vehicle make and automatic vehicle counters that have been carefully calibrated). By running the traditional and the precise methods in parallel at a variety of locations, a set of factors could be produced that would enable corrections to be applied. The degree of segmentation required for these correction factors is likely to be significant, since the proportions of cars, LGVs and HGVs are quite different on different road types and in different area types.

5.5 Forecasting into the future and for policies

The absence in the past of data on LGVs has meant that quantitative data sources are at present inadequate to provide an understanding of recent growth trends. These past trends have been identified through the growth in the LGV vehicle stock and through traffic count data. However, these two data sources by their nature provide no information on what the vehicles are actually being used for and so are unhelpful in understanding the behavioural forces underlying the growth.

There is an overriding need for further research to understand and quantify the set of underlying reasons for the rapid recent growth in LGV trips. This understanding would then aid in forecasting whether the rapid growth is likely to continue or to abate in future years.

An initial step in such research would be to analyse the LGVS data in order to ascertain which types of trips make the major contributions to the total vehicle kilometres. Then these trip types would be analysed to investigate the extent to which their growth can be linked to broader trends in the economy, such as:

1. increased household income leading to greater consumption of low volume/high value goods and related services (installation/repairs/servicing)

2. the rapid growth in office activity (relative to manufacturing) with its requirements for deliveries of small consignments of supplies, and demand for courier, parcel and mail services

3. the growth of a wide variety of new and out-sourced services, many of which are most naturally provided through use of LGVs to carry the tools of the trade and any small consignments that may need to be delivered

4 changes in the value density of goods, total logistics chain issues, more just-in-time (JIT) deliveries and constraints on HGV access to loading bays

5 growth in service industries and changes in goods supply and delivery associated with them.

These and other factors with a potential influence on the growth in demand for LGV movements have been discussed in more detail in the Report A3. They could be analysed for use in creating a behaviourally based model to estimate the future stock of LGVs. This model should be sensitive to the range of policy influences that might change the balance of purchase between cars, LGVs and HGVs. It should build on the procedures already in place within the Department's VMM98 model.

The future estimates of the stock of LGVs could be used to adjust the factors that split vehicle movements between cars and LGVs in the LGV passenger model described in Section 5.3. In this way changes through time and in response to policies would be represented.

Changes in the operating costs and the conditions of operation of LGVs relative to HGVs, would be introduced to the freight model to represent the introduction of policies that pertain to the use of LGVs for the carriage of freight.

6 THE SOFTWARE REQUIREMENTS

This chapter discusses the software issues that arise in making operational the freight modelling framework that has been developed in Chapter 4. It makes no attempt to analyse the precise software requirements that might arise for each specific option within each modelling stage. Instead it takes a broader approach and focuses on what could be achieved within the main commercial software packages that are available, and on the relative merits of the use of such commercial packages as opposed to specially written software for specific parts of the model.

6.1 Overall model size and hardware requirements

The hardware platform upon which the model is to be run will have an impact on the degree of segmentation that the model can support if it is to operate within an acceptable time. To illustrate the computational burden:

> An inexpensive modern desktop computer, running Windows, can be used reasonably efficiently to process matrices of about 500 megabytes in memory. If these matrices require 4 byte integers or 4 byte floating point numbers, (and in fact 8 byte floating point precision would be preferable) this would allow 125 million cells. If the zoning system is of the order of 3000 zones, a fully expanded OD matrix will occupy 9 million cells. This will allow a further segmentation (e.g. commodity) by a factor of just 13. An alternative structure might be:
>
> 500 zones *500 zones * 25 commodities * 20 vehicle types = 125 million

However, the degree of segmentation and the ratio of geographical to sectoral detail may be different within different processing modules; more geographical detail is required at the assignment stage, and more commodity segmentation at the matrix calculation stage. Zone pair sampling may be employed to maintain geographical resolution but decrease the computational burden in a similar way to that already developed to enable the NPTM equilibrium car assignment to operate at the 10,000 zone level on PC (Williams and Lindsay, 2002).

The second big computational constraint relates to shortest-path algorithms which are used frequently within models, and which tend to be proportional to the number of network vertices squared. If different products, vehicle types and so on have different shortest paths (using generalised cost as the measure of distance), there is a practical limit to (either) the degree of segmentation or the degree of network detail. In general however, the need for segmentation by commodity type is limited at the assignment stage, and many vehicle types can be grouped together unless vehicle specific tolling policies are being tested, so that the computational burden can be eased.

A third constraint is the volume of data generated by the results. A model, as proposed, could easily produce gigabytes of results. Again there would be a practical limit, either in terms of disk space required to store archives of results, or in the need to use an external database package to simplify the outputs.

There would appear to be no benefit to moving away from PC technology, because of the major overheads in cost and loss of flexibility from use of other systems. Accordingly, it will be necessary to compromise between the various requirements for detailed segmentation, both through space and for demand characteristics, in order to reach a feasible solution that has acceptable computer resources requirements. The model should be structured to run efficiently on the largest and fastest PCs available. The construction of the modelling system should ensure that it is feasible without major cost to extend the model size in the future to move beyond the current 2 gigabyte storage limit in Windows technology as soon as this becomes universally adopted for software products. This would enable finer detail to be included later, either to represent time of day, or smaller zones, or greater segmentation of demand.

6.2 Software

Most of the models reviewed in work area B use a single software platform: spreadsheet/statistics package, transport modelling package, or all-purpose programming language. To an extent they are all limited by this choice, and their respective strengths and weaknesses are evident. Although this aids internal consistency, a modular approach would give the best chance of flexibility, selecting the right tool for the job, and linking to external software. It would be essential to determine an appropriate set of self-contained modules into which the framework might sensibly be divided. Care would have to be taken in clearly specifying the data flows and ensuring definitional consistency between such modules. The use of the GTF to standardise data transfers between independent modules is discussed in Section 7.3 below.

The main software options for developing the NFTM are:

- use an existing package (e.g. TRANSCAD, TRANUS, MEPLAN, STAN)

- create specially designed software

- a combination of these.

These options overlap in practice since given the wide ranging nature of the model there would almost certainly be a need to create some study specific software, even if an existing package was used for most parts of the model (e.g. the Oregon model developed its own software for micro-simulation, but used TRANUS for regional modelling and Emme/2 for networks & assignment). Along with the software eventually used for the running of the model itself, there would be a need when initially constructing the model to use:

- specialised software tools to manipulate the data in the creation of good quality base matrices, as discussed in Section 3.3
- statistical software (e.g. A-Logit, Limdep, etc.) for the estimation of model parameters.

At the highway assignment stage of freight modelling there are a wide variety of software packages that are already in use around the UK that can carry out multi-user class assignments and so these are not discussed further here. They are able to address most of the requirements for LGV and HGV assignment at the more detailed spatial scale.

The software field narrows greatly when the requirement is for assignment procedures that can represent national and international intermodal freight routes in a form that takes realistic account of transfers between road, rail and ferries along the route. Appropriate cost functions need to be applied to the vehicle for each stage of the journey, and the costs and the time delays at modal interchange points must be represented. Furthermore this assignment would need to be linked to the other modelling stages shown in Figure 2.2, and each of these stages would require its own software module.

6.2.1 General purpose freight modelling packages

There are three main general purpose packages that are commercially available for freight model development.

TransCAD is a general purpose transport modelling package that is integrated with its own set of GIS tools and data structures. It has a wide range of capabilities for modelling passenger and freight transport in an integrated fashion. It was developed by the Caliper Corporation, Boston (2002) and has been widely used in the US over the last decade. It includes tools for vehicle routing and logistics planning at the level of the firm, along with tools for strategic transport planning.

MEPLAN is a general purpose package for transport, land use and regional economic modelling. It has a wide range of capabilities for modelling passenger and freight transport in an integrated fashion. It was developed by ME&P, Cambridge and has been used in many national and regional freight studies around the world. MEPLAN is currently in use by the DfT for the strategic and the detailed highway assignment components of the NPTM. The network manipulation and presentation in the NPTM is carried out in MapInfo GIS which is integrated with the MEPLAN file structures. The freight component of the EUNET Trans-Pennine model (ME&P, 2000a) developed for the Department using MEPLAN illustrates the operation of many of the modelling features that have been discussed in Section 4. **TRANUS** developed in Venezuela is an alternative package that is similar in methods to MEPLAN.

STAN was developed by INRO, Montreal (2002) and has been used in national models of Norway and Sweden as described in Report B2 as well as in other freight studies around the world. It differs from the other two packages in a number of ways. It is a freight specific package that does not include passenger movements. It uses a multimodal, multi-product assignment procedure that *minimises the total system* transportation cost, rather than the user cost. It does not contain stochastic user equilibrium assignment, it provides deterministic minimum cost solutions.

All three packages enable multimodal, multi-product models to be created. They provide a wide ranging set of modules for different stages of the transport modelling process. These modules can be combined in a reasonably flexible fashion to design the form of model required for a specific study. All can be run on PCs, while STAN is also available on workstations. All packages are maintained and updated on a regular basis by their creators.

A longer term alternative to the use of commercial software packages is for the Department to encourage the creation of an open-source transport modelling consortium (Wigan and Drain, 2003), analogous to developments in some other fields, such as the successful Linux computer operating system. This could build on the development of common data transfer procedures such as the GTF that is discussed in Section 7.4.

6.2.2 Software for a complete freight and passenger modelling system

It would be feasible to combine modules from more than one package if this were what was required. The current NPTM interfaces INRO's EMME/2 package that is used for the passenger rail modelling, with the MEPLAN package that is used for most of the rest of the model. In a similar manner it would appear to be feasible to separate out the rail freight assignment from other parts of the freight modelling process. In this way a variant on the rail component of the existing GBFM in use by the SRA might have the potential to contribute directly to the modelling system.

It becomes less satisfactory to separate out the highway assignment from the modal split since the need to iteratively recalculate the congested highway costs and the resulting modal volumes means that in principle these modules should be closely linked in order to guarantee efficient convergence to a joint equilibrium solution. However, an even more critical element on this front would be to ensure that the passenger demand on the highway network is also assigned to the network. Accordingly, the design of a suitable software architecture for the NFTM should be considered in tandem with that already in use for the NPTM.

This certainly would not imply that both these large models would always need to be run interdependently. It is simply to ensure that the interchange of data between them is streamlined and enables the impacts from one model to be easily fed to the other model to the extent that is required at any one time. A simple way to achieve this would be through using a common network of links in both models, and then preloading the links of this network with passenger cars when running the freight model, and instead preloading it with LGVs and HGVs when running the passenger model. In situations where more precise solutions are required, such as when testing road charging policies, all these traffic components could be loaded simultaneously as separate user classes and iterated to convergence. It is primarily at the highway assignment stage that the freight and passenger models need to interact so this stage would benefit strongly from the use of a common software platform.

At the national level the other model stages can be carried out separately between passengers and freight, so that for these other stages the use of separate software platforms between passengers and freight would not necessarily be problematic. The avoidance: of needing to learn how to use too many systems, of needing to pay for too many licenses and of having to introduce a succession of separate software upgrades, all provide lesser reasons for discouraging an undue proliferation of separate software packages within the overall national freight and passenger model system.

7 THE DATA REQUIREMENTS

Having outlined the options for the design of the models in earlier chapters, here the implications for data collection are brought together. Many of these have previously been alluded to in the context of the models themselves. It summarises the main new data collection and data access initiatives (in certain instances the required data may exist but may be unavailable for use in transport modelling for various institutional reasons) that would need to be carried out in order to enable particular model features to be implemented successfully. The review in the Data Report B4 contains greater detail on many of these topics.

7.1 Enhancements to the CSRGT, LGVS and related surveys

The sample sizes envisaged for the LGVS are comparatively small and hence the database on which the LGV models will be constructed will be weak until sufficient years have gone by to enable sampling errors to be reduced through the pooling of data across years. While the data expansion methods based on LGV link counts will help, the definitional issues at the counts in distinguishing LGVs unambiguously from cars and from HGVs are sufficient to introduce significant uncertainty into this process.

Major increases in the size of the LGVS samples would be highly desirable in order to provide a more robust database for LGV modelling in the medium term future. Given the inherent difficulty in surveying the LGV sector, there may however, be benefit to first carrying out the surveys for one year in order to check the effectiveness of the proposed survey methodology. Guided by experience from the first year, then increase the expenditure in the following year in order to obtain a sample large enough to put the LGV modelling and analysis on a firmer footing.

The recent LATS 2001 RSI survey may provide some useful extra information on LGVs since it contains information on both the trip purpose and the types of goods carried for both LGVs and HGVs with some segmentation by vehicle size.

Increases in the size of sample of the CSRGT would also be very desirable, both to provide more robust estimates through space and to enable the matrices segmented by commodity type to be more reliable**. Doubling the annual sample size and a once-off sample for next year that is 5 times the current size would be very beneficial in CSRGT**. In particular this would enable a significant body of new information to be assembled that is critical to obtaining successful model development. It would ensure that the answers to the new/changed questions listed below would be available for an adequately large sample in two years without having to wait a much longer period for an equivalent sample size to be accumulated through combining CSRGT results for many future years.

Other changes to the CSRGT that are needed are:

1. Include a question on the land use type (e.g. factory, office, warehouse, retail, etc.) at every loading and unloading point. **This is crucial to implement immediately, as it is the primary means whereby the CSRGT data could be linked to the stage in the logistic distribution chain, enabling P/C matrices to be distinguished from O/D matrices**. The published data must include an accurate representation also of intermodal transfers at ports, airports and rail terminals. Currently, there are not adequate estimates of HGV movements in and out of ports, since the RoRo only covers those vehicles and trailers that continue onwards by ferry, and not those that load and/or unload within the port itself. The CSRGT does ask hauliers to record whether journeys are to/from docks and railheads but the resulting sample sizes are small at the level of individual ports.

2. Code the geographical locations of trip ends, both for new and for recent past surveys at the LA District level or finer. Postcodes would be the ideal.

3. Improve the recording of the multi-drop trips/tours. Explicitly code the location of that drop which is furthest from the origin and destination of the trip/tour. The current data, which excludes this item is unhelpful for modelling purposes. This change is also required in the LGVS.

4. Survey foreign HGV activity within the UK, for one year at least.

5. Code the start and end times of journeys as a precursor to the introduction of time of day into the model. This when combined with the land use information and the commodity type carried would provide valuable insights into which types of movement appear to be captive to certain times of day.

6. Sub-divide the current miscellaneous commodity category, to enable more detail to be retained.

7. Investigate the apparent under-reporting in the CSRGT relative to count based estimates of HGV traffic.

Many of these topics above are being considered as part of the Quality Assurance Review of freight statistical sources currently underway at DfT.

For all surveys related to goods vehicle movements, whether they be at the national scale or below, common definitions should be adopted for all categorisations in order to facilitate combining data from different sources at a subsequent stage. In particular:

1. The definitions of LGV and HGV used within the NTS should be made consistent with the LGVS and CSRGT

2. Definitions, including trip purpose, land use at origin/destination, etc., should be common across the LGVS, CSRGT, IRHS and RSIs, and other more localised data collection exercises.

Provided that the data definitions are compatible, the data from these various sources can later be combined to help assemble a larger national database and checked one against another.

It should be noted that enhancements to the CSRGT and LGVS are *vital* to the development of sound national freight models. It is recommended strongly that the proposals above be implemented if at all possible.

7.2 New data sources required

This section reviews the potential for new data sources and data assembly initiatives beyond those currently in operation at DfT.

A detailed survey of the number of vehicles (LGVs and HGVs) arriving and departing from firms (segmented by SIC and size of firm in terms of number of employees) and households would be of great benefit in order to provide a reasonable calibration of the zonal freight production and attraction rates. These rates would equally be of use in studies below the national level. This data would need to be connected to an analysis of which vehicles do single drop, and which do multi drop tours, and how this varies by logistics stage, by commodity type carried and by types of firms visited. The scale of the survey required to achieve adequate coverage in such an exercise would be large, so that the cost may militate against obtaining this information. However, a once-off exercise would probably be adequate to assemble the required data. It might be possible to undertake this survey in conjunction with TRICS.

The acquisition of good data on land use at origin and destination within the CSRGT may provide a cheaper, though coarser, solution to this requirement to link transport demand to the activities that generate it. The TRICS database also can provide useful pointers on how to proceed and some initial data on the scale of the GV movements associated with some types of land use. In summary, one way or another **a substantial volume of new data is needed that connects goods vehicle movements to the activities and land uses that generate and attract these movements**.

The land use data itself that would be used in tandem with the freight generation and attraction rates could be provided by various data sources that have recently started to become available as discussed in the Data Report B4. The Inter Departmental Business Register (IDBR) or the related Central Business Directory (CBD) can provide spatially detailed current information on employment by type, and the Valuation Office can provide information on floorspace by type. **The Department should explore the means by which these data sources could be accessed, integrated and updated efficiently**. Their potential usefulness is much wider than just for freight modelling – the NTEM already makes use of the IDBR to estimate zonal trip attractions.

In order to maximise the value from such disparate data sources, more systematic standardised means of organising and accessing the data, using XML based definitions, should be considered along the lines of the GTF discussed in Section 7.4. There is a need for an information management strategy to complement the freight model development strategy discussed in the rest of this report.

An alternative approach to understanding the relationship between freight movements and the activities that generate them is the creation of a shipment based equivalent to CSRGT. This would trace goods (rather than vehicles) through the logistics distribution legs from production along the logistic chain to final consumption. The approach used in the MYSTIC (2000) study provides guidance on how this type of survey could be carried out. The MYSTIC project identified two candidate methodologies for collecting freight origin-destination data, as follows:

- The freight shippers survey carried out by the French and Dutch governments sampled shipments of freight and traced the shipment through each organisation that handled the goods until it arrived to its final destination. This method involved a series of telephone calls to each company involved in the shipment chain asking detailed information including behavioural information about the alternatives that the company could have used. The 'shippers' survey methodology can also be used to obtain details on the logistic organisation of the shipper, thereby considerably improving our ability to understand and forecast the movements of goods and their modal split. Eliciting this wider information can be costly and if the primary objective is to measure the origin/destination legs of a production/consumption matrix then slimming the interview down to the essential data items can reduce costs. For behavioural and detailed supply data an additional questionnaire can be administered to a sub sample.

- The freight EDI survey which was tested in the UK sought to investigate the possibility of intercepting the Electronic Data Interchange (EDI) messages, which pass over the Internet from one company to the next in the freight commodity chain, in advance of the goods reaching them. This EDI survey showed that the information required (i.e. origin, destination and other data items) did indeed exist in the computer systems of the originator of the consignment although the data items required usually spanned more than one organisation and spanned more than one computer system. This made automating the process difficult to achieve in the short term but a paper-based system could be possible.

A variant on the above approach would be a case study approach that surveyed in depth a limited number of major HGV and LGV operators, such as the Royal Mail, major supermarket chains, petroleum companies, parcels/courier firms, logistics/distribution firms, etc. Each of these would be analysed in order to understand the scale and pattern of the traffic they generate and to understand the underlying mechanisms that govern these patterns. There would also need to be some study of a selection of smaller firms in order to understand any fundamental differences in behaviour between the larger and smaller firms. This case study approach would provide two main outputs:

- an increased understanding of the forces that create the current transport patterns and of how these may respond to policy changes

- quantitative data on a significant part of the overall freight transport industry, and of the linkages between different parts of the distribution chain.

The future use of telematics data to provide population estimates on movements by vehicle type may be of major benefit when a sufficiently wide coverage of the vehicle stock eventually becomes available. This data source could be used to refine the expansion of survey data and to understand the under-reporting of movements within the CSRGT. **In the short term the department should support trials of telematics systems to ascertain their potential to substitute/complement other data sources**. In the longer term when allied to detailed spatial data on the types of activities in place at the locations at which the vehicles stop, this might eventually provide a cost effective supplement to the LGVS and CSRGT, enabling their sample sizes to be reduced. The telematics based data tends to be weakest in terms of providing an increase in understanding of the structure of the demand for transport, since it will provide little information on what is being carried. In the short term it is this increase in quantified understanding of the transport system that is the highest priority as an input to the development of good freight models.

7.3 National Accounts – Transport Satellite Accounts

In order to measure the linkage between transport and the economy, there is a need for improvement in the representation of transport within the UK National Accounts. Transport Satellite Accounts (TSA) have been introduced in the U.S., France and elsewhere as a means of improving the accuracy of measures of transport activity. For example, before the introduction of the TSA, U.S. domestic transport services were thought to have accounted for about 6% of the total cost of U.S. agricultural products, whereas in fact they account for 14% according to the work of the U.S. Dept. of Transportation et al. (1999) in the definitive report on TSA.

The recent UNITE EC research project (Mayers et al., 2001) has concluded that:

> "transport activities are not very well represented in most national accounts. The value added of the transport sector in the national accounts does not take into consideration the transport services organised by non-transport sectors and by the households themselves. For this reason the present national accounts are not a useful basis for transport policy and need to be extended by means of Transport Satellite Accounts before they can be of use."

Within the UK there has been progress in the measurement of the value of the output of the household production of transport (Short, 2001) as part of the more general development of household satellite accounts for the UK. However, there does not appear to have been corresponding developments in creating TSA for the UK. Without the development of TSA any macroeconomic analysis of the transport sector is likely to contain serious measurement and estimation errors. The problem is as follows.

Within the UK National Accounts own-account transport is generally treated as an "ancillary activity" which is not separately identified from the principal and secondary activities of an economic sector. Consequently own-account transport is not included within the transport sector, but is intermingled within each of the other industrial sectors to which it pertains. Accordingly, if there were a general shift by industries towards outsourcing more of their transport activities, this would increase the proportion of hire-and-reward transport relative to own-account transport within the UK economy. This in turn would *appear* to generate growth in the measured transport sector activity within the National Accounts, although in reality the total number of freight vehicle kms need not have changed at all. Since the ratio of own-account to hire-and-reward transport differs by mode and is not stable through time, the current transport accounts have a serious potential to mislead. Accordingly, **the creation of Transport Satellite Accounts for the UK is a critical ingredient for improving the data inputs to freight models and to provide a sound basis for policy analysis**.

7.4 Standardisation of data

A central theme throughout all of the discussions on data within this study relates to the need for standardisation in data definitions and in the processing of surveys across all data sources, modes and spatial scales. The statisticians in the Department have already achieved much with the data sources under their direct control at the national level, but standardisation continues to require constant vigilance as illustrated by the various action points identified earlier in this report. This standardisation is the key to obtaining best value from the large investments in transport data collection across the country.

The design for the NFTM that has been presented is a wide ranging design that encompasses all modes and that operates at a detailed level of segmentation throughout. This necessarily will involve combining data from different sources, which is a complex and inefficient task unless the data definitions are common across these sources. To enable the results of the NFTM to be used as an input to studies at a finer spatial scale, the local data collection for such studies needs to be standardised to match the national sources.

This standardisation entails more than just ensuring that the same number of categories are in use and that these have common names throughout. Appropriate meta-data that clarifies precisely what is measured and how it is measured is required for each data source. For traffic count data for example, not alone do the vehicle type definitions need to be common across surveys, but the associated information about each count itself needs also to be available for inspection. This meta-data would define: whether they are manual or automated counts, and for which hours of the day / days of the week the counts have been carried out. Although standard factors can be applied to bring counts to a common basis, it is important when gauging the relative accuracy of apparently contradictory "observed" data values, to be able to understand which factors might have exacerbated sampling errors between different sources of observations. The methodology for model development that has been put forward in Sections 4 and 5 entails combining data from surveys, RSIs and traffic counts in a sophisticated integrated manner. This will place particular pressure on the need for standardisation in these sources and in understanding the nature of subtle differences between their data collection procedures.

The standardisation of data transfers between models is a topic that has been studied in some detail as part of the EC project SPOTLIGHTS (Mandel & Ruffert, 2001). It proposed the use of a refined version of the Generalised Transport interchange Format (GTF) that was originally developed in the BRIDGES (2000) EC research project. The GTF ensures that data is passed consistently between models using an XML representation. This guarantees that the meta-data that is required in order to understand and make best use of the data remains available to those who will use the data. **The Department should consider the adoption of the XML based GTF as the means whereby datasets are interchanged between different spatial scales or different models** (e.g. the NPTM already has a number of separately created modules and some of these may eventually interchange data with parts of the NFTM). The initial costs of adopting an XML based system such as the GTF are significant, but in the longer term the degree of standardisation that can be achieved will encourage economies in scale and substantial increases in productivity in model development and in efficient usage of expensively collected data. The longer term benefits would outweigh the short term costs.

7.4.1 Coordination of Data

There is much duplication of data activity and inconsistency in applying data among transport modelling activities in Britain. The Department could usefully investigate this issue and develop mechanisms to support a more efficient approach to data availability, harmonisation and understanding. This could include experimentation with emerging data sources and development of best practice.

8 URBAN AND SUB-REGIONAL SCALE MODELLING

This chapter explains how the methodology that has been specified at the national scale can be used in order to improve freight modelling at the more detailed spatial level. The focus is on ensuring that the improved insights into the overall changes in future patterns of road and rail freight movements at the national scale, can be successfully filtered down to the more localised spatial scale at which most transport modelling actually takes place in practice. The application of national and regional control totals on growth in freight traffic is also discussed below.

Because the rail freight system competes most cost-effectively for movements over long distances, and because the rail freight network is sparse by comparison with the road network, the national level rail modelling tools will generally also be appropriate to address issues at the more local scale. At worst, some extra local detail might need to be included as part of the freight rail component of the national model. In fact it could be misleading to model rail freight only within a limited localised study area, since the main markets are likely to be in distant, rather than in adjacent areas.

The recommendation is that local rail freight models are only likely to be required in specialised circumstances, and that a local customisation of the freight rail component of the national model is likely to perform better and to be more cost-effective in most instances.

For this reason the focus in the rest of this chapter is only on road freight, because in contrast to rail, much of the road freight does indeed have a pronounced sub-regional or local character, which could get lost at the overall national scale. The longer distance road freight into and out of a local area, which is potentially in competition with rail, would be derived primarily through use of the multimodal national model, so that road-rail competition is not ignored in this proposed approach at the sub-national level.

It is generally accepted that the past level of performance of goods vehicle models has not been adequate. SACTRA (para. 10.50, 1999) has stated that the standard method of simply factoring up fixed base year vehicle matrices by a factor based on growth in vehicle kilometres is "seriously deficient" and that better procedures are required. The modelling methods proposed in Chapter 4 above explain how to include a much wider set of influences within freight models at the national scale. For smaller scale urban and sub-regional studies what will generally be required is an intermediate position that represents a significant step forward from the previous fixed matrix approach, but which can nonetheless be implemented in a relatively short time scale, using existing data, and requiring modest budgets. This approach is outlined in the next section and is followed in

Section 8.2 by an interesting and more sophisticated approach using agent-based micro-simulation models.

8.1 Improved local goods vehicle models

The various methodological developments at the national level to improve freight modelling will also provide useful insights into road freight modelling at the sub-national scale. In an ideal world the comprehensive methods already put forward for national scale modelling, should be applied in an analogous fashion at each of the other finer spatial scales. In practice, however, it is unlikely that the resources, or perhaps even the skill base, would be available in many such studies to achieve this. Accordingly, the focus here is to demonstrate how the most important of the methodological improvements can be passed down the spatial scale in a cost-effective fashion. This occurs in a number of ways that are first summarised here, and then expanded in more detail in subsequent sections.

1. Provide control information from the forecasts of the national model. This is used to ensure that local models are broadly consistent with the expected national/regional trends. This role has been played in the past by the NRTF growth rates.

2. Provide the background long distance external and through traffic matrices.

3. Enable future year vehicle matrices to be forecast in a more realistic fashion. As will be outlined in Section 8.1.2, the recommended methodology for the future growth of freight traffic needs to take appropriate account of changes in trip lengths and not just in trip totals.

4. Provide best-practice guidance on modelling features such as the segmentation of vehicle sizes, the construction of flexible vehicle operating cost functions, the generation and attraction of freight, the conversion from tonnes to vehicles, etc.

While the steps 3 and 4 above could be adopted immediately in sub-national modelling, the steps 1 and 2 would only come into operation, when suitable outputs are available from the NFTM.

8.1.1 Relationship of national level forecasts to sub-national modelling

The recent experience of the results from a number of the multimodal corridor studies has demonstrated the reluctance of their modellers to adopt anything other than the simplest of approaches to freight forecasting. This strongly suggests the need to provide a solid foundation at the national scale so that the maximum relevant information can be provided to those operating at the sub-national level, whether it is regional or local. The quantitative information would be passed down to sub-national studies in two ways.

The national level HGV and LGV models should provide default O/D vehicle matrices for long distance external and through traffic as well as control totals for the growth of traffic within the study area. The demand pattern for long distance trips will be strongly influenced by various factors that are likely to lie outside the study area of the sub-national study itself, and so might be difficult to deal with in isolation. The local pattern of origins and destinations of these long distance external trips within the study area could then be adjusted in order to increase the level of zoning detail and to make use of local knowledge on the likely generators and attractors of such movements. The resulting adjusted matrices would be assigned, together with the shorter distance locally modelled trips, to the highway networks within the study area.

These long distance external vehicle matrices would also provide a means by which studies in adjacent areas would avoid inconsistent assumptions in the LGV/HGV traffic that passes through them both. They would also help in mode split modelling between road and rail freight for such external movements, through lessening the need for an independent local rail freight model.

For the shorter distance GV movements it is expected that the ability to assemble suitable local classified traffic counts and possibly RSI's would imply that locally created base matrices should be more reliable than those available from the coarser national model. The methods that have been reviewed in Section 3.3 for using a variety of data sources to improve the estimation of base year vehicle matrices could equally be applied at the sub-national scale to improve the quality of the base HGV and LGV matrices that are constructed. The means by which these local base year matrices are forecast into the future in line with national trends is explained in the next section.

In order to ensure that these improved freight modelling procedures are taken up widely within the profession, **clear guidance to practitioners will need to be prepared and then published by the Department.**

8.1.2 Improved methods for forecasting future year O/D matrices

The standard method of HGV traffic forecasting that has been used in the past was to factor up a fixed base year HGV O/D matrix, through increasing the zonal numbers of vehicle *trip ends* to a level that enables the target growth in *vehicle kilometres* to be matched. In essence, the resulting future year matrix would retain a pattern of trip lengths that is broadly similar to that in the base year. However, based on published UK statistics, it was shown in the Inception Report of this study that substantially more of the growth in HGV traffic arises from trip length increases, than from increases in the number of vehicle movements. Accordingly, forecasting methods that ignore this effect are inherently flawed. Furthermore, such growth assumptions have no sensitivity to policy levers that might be tested.

As has been proposed in Section 4.1, incremental spatial distribution models are the appropriate tools to adopt for purposes of forecasting future freight matrices in a manner that allows trip lengths to vary. One way in which this can be carried out is through the use of a residual disutility based procedure. The precise steps by which a future year vehicle matrix can be forecast are presented in Appendix II, using a residual disutility approach. This methodology is scale independent and so is entirely suitable for sub-national modelling needs.

There are two main information inputs from the national level model to this vehicle matrix forecasting approach for the forecast period:

1. Growth factors for the increase in the number of vehicle trips

2. Growth factors for the increase in the average vehicle trip length.

These factors would be likely to differ between regions and should be those factors appropriate to the trips that are internal to the study area, excluding the long distance external trips that have already been provided directly from the national model. A major source of the differences in growth factors between regions would arise from the industrial structure of the region and of the typical commodity types that are moved within it. Many sub-national vehicle based freight models are unlikely to have information on the type of commodity carried (with the exception of London for which the LATS 2001 survey does ask for this information). This means that the only means whereby the strong differences in growth patterns between commodity types can be introduced locally is via the national model estimates.

Clearly in situations where it is feasible to build up the demand for freight transport at a more segmented commodity level as is done at the national scale, then the local estimation of growth patterns could be significantly improved.

The estimation of the zonal patterns of generation and attraction of freight can adopt the procedures that have been outlined in the Section 4.3.

8.1.3 Local issues

There are certain features of HGV traffic that by their nature have a much more important influence at the local scale than at the national scale. Accordingly they merit special interest when modelling at the finer scale.

Because of their overall volumes of tonnes lifted (about 30% of the total on road), their major localised impacts and nuisance, and their relatively short lengths of haul, there are benefits to considering a specialised representation at the local scale of the bulk movements of aggregates and other freight associated with the construction industry. At the local scale the locations of the main sources and sinks for such movements are not difficult to pin down. The local planners should be able to provide comprehensive spatial information on these from their statistics on housing starts, planning permissions, road and rail improvements/investment schemes, land-fill sites, municipal dumps, quarries, etc. This information should cover not just the present, but also the scale at which these sources and sinks may grow or decline in the future in each location, as the land available for development in some areas becomes exhausted while new developments start to come on-stream elsewhere. Use of this specific information, including where necessary spot surveys of the major sources and sinks for vehicles, can greatly improve vehicle generation and attraction estimates at the local level.

There is merit to **carry out some case studies in selected local areas to examine the success of the use of the above data sources to improve the procedures for estimating freight generation and attraction**. This would be allied to the use of TRICS data as discussed in Section 7.2. The results would then be published as guidelines for use in other local studies.

The other major topic in which local traffic patterns tend to differ strongly from the national level is LGV movements, since these are particularly concentrated in and around urban areas for the delivery of goods and services to industry, offices and households. However, the current size of the LGVS samples is not large enough to facilitate much spatial differentiation initially, so that it will be some years before a sufficiently large database is available to enable an improved quantitative understanding of their demand patterns to be obtained.

When modelling at the detailed scale it is feasible to test a number of area specific policies in a more direct fashion than at a larger scale. The inclusion on the network of initiatives such as lorry bans, low-emission zones, etc. can be used to estimate the resulting traffic impacts. However, care is needed to ensure that the full repercussions of such initiatives are correctly represented in the model, and this may require a considerable increase in the sophistication of the model structure required and move it closer to that in use at the national level.

An example with respect to lorry bans will illustrate the point. If a proposal is introduced to ban lorries above a certain size from entering certain urban areas, then the full knock-on effects on freight vehicle movements need to be estimated. Those activities that produce or consume goods within the area of the ban can be assumed:

1. either to continue in the area and to move the goods on smaller vehicles,

2. or to relocate outside the area.

To represent the first response it is necessary to switch the modelling from a simple vehicle based approach to a volume based approach that could correctly account of for the major increase in full and empty movements of smaller vehicles that would be required to replace the banned larger vehicles. The total number of vehicle kilometres might increase strongly (albeit in smaller vehicles on average), both within the area of the ban and outside it. The overall logistics structure would be unlikely to remain unchanged in this instance.

To represent the relocation response a model of industrial and service location would be required and this is not a trivial undertaking, since it would require a substantial amount of good data to calibrate and validate the estimation of relocation responses (see Appendix III.8).

Accordingly, when testing at the national level those policies that inherently have a localised effect there is a need to have a small number of best-practice local freight models that could act as a test-bed for exploring these types of policies. These would be models that have been created with larger budgets for data and for model development than would be the norm. London would be one obvious candidate for consideration, but some non-London sites that are more typical of the country at large would also be needed.

8.2 A micro-simulation based approach

In studies where there are sufficient resources available to enable a more complete model of freight flows to be developed, the use of an agent-based micro-simulation of goods movements is an interesting option. The approach that was proposed for the LATS model for London has been described in some detail in the feasibility study by Neffendorf et al. (2001) and the key components of both the Upper and Lower Level Models that this would include have been summarised in the Report B1. Many of the features of its Upper Level model overlap with the design of the national model that has been presented in Chapter 4 above.

Much of the design of the lower level micro-simulation component of this proposed model for London is derived from the pioneering work in Oregon in the USA as part of the Travel and Land Use

Model Integration Program of the Oregon Department of Transportation. As explained in the Report B2 this model (see Donnelly, 2002 for further details) has a sophisticated representation of many desirable features such as: shipments segmented by size, point locations of individual establishments at origin and destination, explicit transhipment points, itinerary generation and optimisation to represent multi-drop movements. The resulting vehicles are subsequently assigned, together with passenger vehicles to a conventional network using a multiclass equilibrium assignment. The validation of this model in Portland has produced encouraging results.

The particular attraction of the freight micro-simulation approach is that because it models choices at level of the *individual* consignment, it provides a means of representing the inherent heterogeneity of freight in an appropriate fashion. The alternative aggregate model approach that has been proposed in the rest of this review study relies on segmentation to tackle this issue of heterogeneity. Unfortunately there are limits to the degree of segmentation detail that can be maintained before a model and its data requirements become overwhelming, so that micro-simulation appears to provide greater potential in the future.

Accordingly, **the experimental development of a similar micro-simulation model somewhere within the UK should be considered seriously**. Given the scale of the data requirements to make the system operational, it would be unwise to initiate this at the national scale. It would be faster and cheaper to start with a smaller study area. The existence of the current LATS survey of freight and passenger transport, together with the general scale of congestion on road, **makes London a strong candidate** for the initial testing of this sophisticated approach.

This sub-national testing could proceed ahead in the immediate future, given that the feasibility study outlining the approach has already been produced and that the required LATS data will soon be processed and available. Once the system has been demonstrated to work well in London, its extension to cover other parts or even the whole of GB should then be considered.

9 THE USES OF THE MODELS

The SACTRA (1999) report "Transport and the Economy" has made recommendations covering the improvement of modelling procedures:

"11.29 We recommend that the Department conducts a thorough review of past work in the modelling of freight responses to changes in the transport system and initiates research to develop sound techniques for modelling goods vehicle responses"

11.30 We recommend that the Department initiates research to develop better procedures for forecasting the growth in the demand for goods vehicle movements"

In terms of transport models this translates into two different types of model usage that it is helpful to distinguish, though of course they are not wholly independent of each other:

1. To make *conditional policy tests* of some transport initiative (e.g. the NAOMI and LTS models) – often these tests will be carried out for a future year, because it generally takes time before such an initiative can be made operational. In SACTRA terms these correspond to movements along a single demand curve (at a point in time)

2. To make *projections* of traffic supply and demand for a future reference case situation, based on a series of external assumptions on demographic and economic trends (e.g. the NTEM and NRTF models). In SACTRA terms these correspond to shifts in the demand curve over time.

One important issue to be resolved is the relative merit from: addressing both of these usages within the same model, or from developing separate models for these separate purposes. Given the progress that has been made in the specification it appears that in principle a single set of models could satisfactorily be used for both of these purposes. The challenge is to obtain sufficient data and other resources to enable the proposed modelling system to be developed to a useful level within an acceptable period of time. The way forward that is proposed in Section 10.1 maximises the use of the currently existing model components to enable the newer developments to be focused on at an early stage.

Additional specifications from the study brief (Para 2.1(iii)) detailed the functional requirements of the proposed modelling options with relevance to *conditional policy testing*:

"The expected options proposed will need to meet the following criteria, as well as being suitable for use for policy testing:

(a) Provide a satisfactory model calibration for existing freight and vehicle movements
(b) Be sensitive to economic and land use changes
(c) Be sensitive to changes in travel costs (fuel and non-fuel), including intermodal differences. The review study of freight models would need to consider the possibility of incorporating such features as time-of-day scheduling, responses to congestion and assignment issues. Other considerations in responsiveness to costs (both vehicle operating costs and congestion costs) include the nature of response (which may require segmentation by industry type), how they are determined, what values of behavioural responses are used and what kind of data they are based on.
(d) Be able to reflect changes in the characteristics of freight transport operation
(e) The output from the proposed model would need to be combined with outputs from the NPTM to provide information on traffic and environmental impacts of all goods vehicle movements."

Explicit detail is in the brief (above) to determine the minimum acceptable range of policy levers to which the proposed model should be able to respond. Subsequent advice from the Steering Group emphasised that this list was not to be considered exhaustive or prescriptive; instead, **the key requirement for policymakers was flexibility within the model system**. The ability to handle as broad a range of policy options as possible, mediated through an explicit and comprehensive set of model mechanisms and responses (levers), was the primary goal.

An example of the range of policies which would be of interest, in the context of the 1998 Transport Policy White Paper include:

- Changing carbon emissions targets
- Impacts of changes in logistical patterns (rise in e-commerce and associated LGV activity)
- Volume to value changes leading to new patterns of distribution
- Consolidation and break-up of shipments at the edge of town (possibly mediated by the imposition of curfews, tolls or other regulatory/fiscal measures).
- Sensitivity of LGV traffic to different tolling regimes
- Potential impact of 24-hour shopping and 24-hour deliveries
- A possible charge on embedded miles in goods (similar to a carbon tax)
- Freight facilities grants for short sea traffic

There is also great interest in the impact of land use policies and trends on patterns of freight transport. However, there is great difficulty in deriving good relationships between land use and freight because practices and responses are often the result of internal company dynamics even when a high degree of segmentation is applied.

Owing to the great breadth of possible policies that might be of interest to policymaker either now or in the future, and to the necessity to be flexible rather than prescriptive, it is essential to specify a model structure that will allow the impact of many policy levers to be examined. The process is:

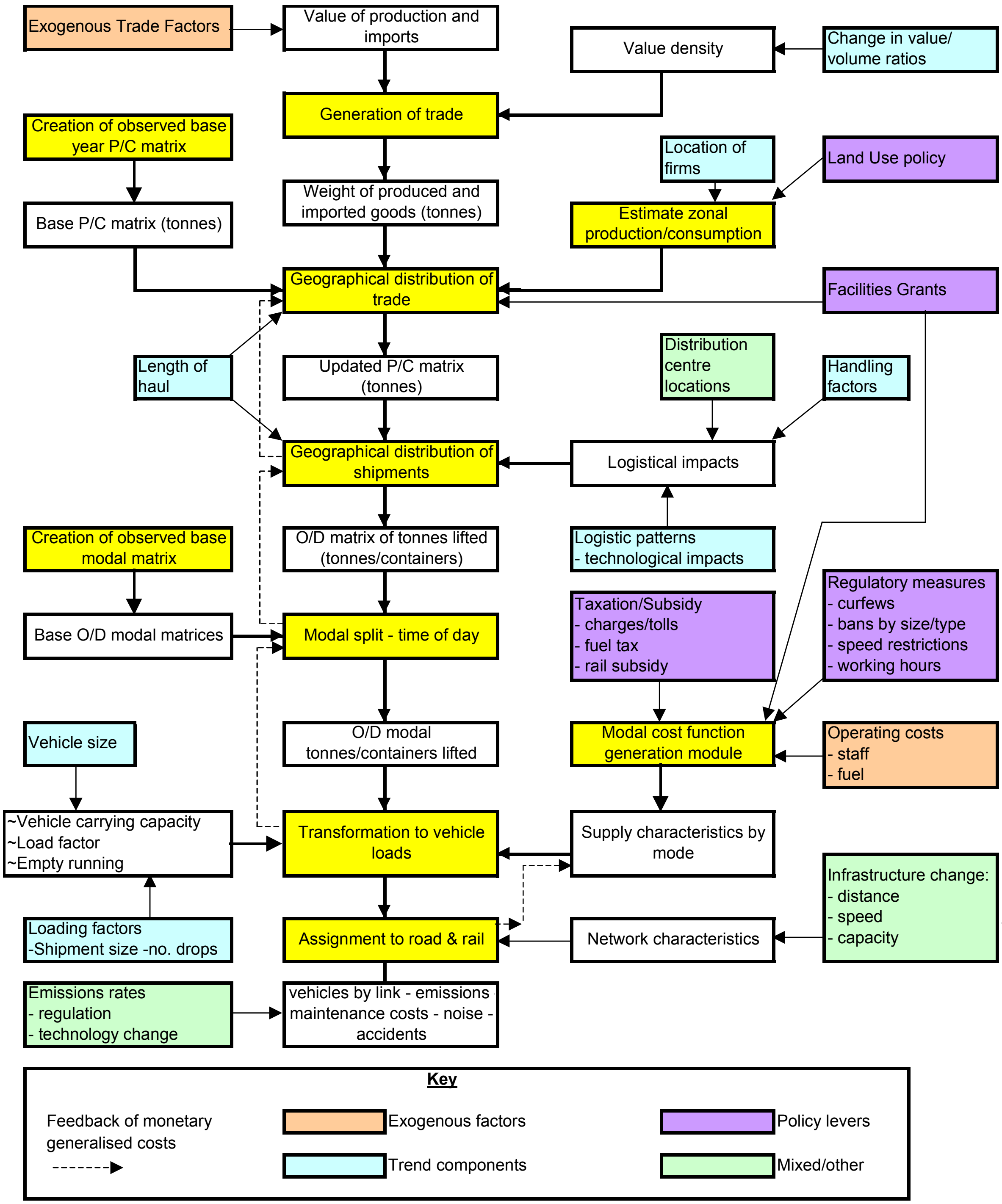
POLICY → **LEVER** → **IMPACT** → **ASSESSMENT**

The key requirement is an explicit identification of the means by which the operating costs are built up through combining the represented cost components. In this manner there is a direct mechanism to estimate the extent to which a transport policy initiative can feed through to change the parameters of the aggregate operating cost function. The example of the influence of the EU Working Time Directive that was introduced in Chapter 2 of the Inception Report illustrates the approach by which a policy initiative may impact directly or indirectly on many separate cost components. These changes in unit costs should be quantified, aggregated and then input to the transport model through modifying the existing cost function parameters.

There are many points within the framework at which policy levers can feed in to the system. Figure 9.1 indicates the modules into which several key policy levers can provide input. For example, regulatory measures such as lorry bans, and pricing measures such as road tolling, each feed into the module: modal cost function generation, in which they modify the operating costs of the vehicles affected. The exogenous factors and trend components are in separate boxes and represent those aspects, such as patterns of trade, spatial locations of firms and increases in lengths of haul, that are influenced by various industry specific factors and so they have a momentum that is somewhat independent of the actions of government.

The design of this model framework has been driven by the requirement for the models to be able to forecast the impacts of a wide range of transport policy initiatives. However, the investigation of suitable *assessment* systems to use together with these models lies outside the scope of this commission and so is not considered within this report.

Figure 9.1 Interface between policy levers and model structure



10 THE WAY AHEAD

This chapter outlines the main recommendations on how to proceed with the implementation of the modelling framework for freight and LGVs within Great Britain. It prioritises those actions that should be carried out in the short term (i.e. over the next 3 years) and provides broad guidance on the preferred structure for model development and data assembly.

The primary issues to be addressed in this chapter both for freight and LGV modelling are:

1. What steps can be carried out and what model structure should be adopted in the short term future in order to maximise the availability of an operational national freight model (NFTM) at all times?

2. What model structure is appropriate for use in the medium term? What other data collection and research tasks are needed to underpin it? How can it be created cost effectively through maximising the reuse of model components that have previously been created?

3. What steps should be adopted below the national scale in order to improve standards in freight modelling.

10.1 The short term development of a national freight model

The main aims are:

- for the Department to have available at all stages an operational NFTM that can be used both to forecast future freight growth and to test a gradually increased range of policy measures

- to minimise expenditure on stopgap tasks that are subsequently discarded and replaced by future enhancements

Accordingly, there are benefits at the early stages to the Department from continuing to make use of the Great Britain Freight Model (GBFM) that was created by MDS-Transmodal. The earlier Figure 2.2 presents an overview of the entire national freight modelling framework that has been proposed. The lower part of this figure, specifically from the box “O/D matrix of tonnes lifted” downwards, presents a set of model features that are already represented to a greater or lesser extent in the existing GBFM. However, many of the finer details of the structure that have been proposed are not necessarily the same as those in GBFM so that significant further development will eventually be required.

The upper part of the Figure 2.2 represents the demand modelling components and contains many of the more innovative features that have been proposed. Few of these features are contained in the existing GBFM.

Table 10.1 National freight model development tasks

	TASK SETS	Reference
A	**Audit and limited upgrade to existing GBFM**	
1.	Carry out an audit of the existing GBFM model in use in DfT	10.1.1
2.	Introduce a highway transport network that is consistent with the NPTM	3.1, 3.2
3.	Carry out other short-term enhancements that arise from the GBFM audit	10.1.1
B	**Research and implement improved demand modelling**	
1.	Data tasks Q1, Q3, Q7, Q11	Table 10.2
2.	Research tasks R1, R2, R4, R5	Table 10.2
3.	Create procedure for freight growth linked to economic change	4.2
4.	Create zonal freight production and consumption estimates	4.3
5.	Create an incremental modelling system for spatial distribution of trade by P/C and for subdivision into distribution legs	4.1, 4.6, App. III
6.	Estimate the base observed P/C matrix	3.3, 4.4
C	**Implement mode split and supply modelling**	
1.	Data tasks Q2, Q4, Q5, Q6, Q9, Q10	Table 10.2
2.	Research task R3	Table 10.2
3.	Create base O/D matrices for modal and intermodal movements	3.3
4.	Create an incremental O/D mode choice model (and possibly time of day split)	4.1, 4.8, 4.9
5.	Create modal cost function generation system	4.7
6.	Create system for splitting by vehicle type and transformation to vehicle loads	4.10
7.	Assign to modal and intermodal networks	4.11
8.	Calibrate and validate the interlinked components of the freight model	The above
D	**Run NFTM through the future, linked to the economy**	
1.	Generate future year forecasts of economic growth, run model in the future	4.5
E	**LGV model development**	
1.	Data tasks Q2, Q4, Q6, Q8	Table 10.2
2.	Research task R6	Table 10.2
3.	Integrate the freight LGV model within the freight model system	4.10, 5.2
4.	Develop the LGV model for private passenger trips within the NTPM	5.3
5.	Develop the LGV model for commercial trips	5.3
F	**Sub-national tasks**	
1.	Select and develop some best practice local models	8.1.3
2.	Prepare and publish guidelines on improved freight modelling procedures	8.1.2
3.	Carry out local case studies to improve procedures for estimating freight generation and attraction	7.2, 8.1.3
4.	Disseminate forecast control volumes derived from NFTM	8.1.1
5.	Develop agent-based micro-simulation model for London	8.2

The development and implementation of the national freight model can accordingly be split into a number of task sets that are summarised in Table 10.1, together with the reference to the section in which each task is described. Each task set comprises a set of linked tasks, though it should be noted that there may well be a requirement in practice for some linkages also between activities in one task set and another.

10.1.1 Task set A – Audit and limited upgrade to existing GBFM

The GBFM currently in use in the Department can perhaps provide functionality to tide the system over, while improved modules are being put in place. However, at present there appears to be only limited information available to the Department on its precise underlying structure and on the quality of its performance. GBFM is the property of MDS-Transmodal who created it themselves, rather than on behalf of the Department. Accordingly, an important first step would be to commission, in co-operation with MDS-Transmodal, a small independent study to carry out an audit of: the structure, the input data sources/processing and the quality of output results from the current GBFM. This would generate an appreciation within the Department of:

- the strengths and weaknesses from continuing to use GBFM for the present

- the extent to which existing GBFM components (e.g. the rail freight data, modal cost modules, etc.) could successfully be reused within later versions of the national model

- the minor short-term enhancements that should be carried out to improve its usefulness until a version of the proposed model is in a position to be used, for example:

    1. converting it to run on the same GB highway network as is currently in use in the NPTM

    2. deciding whether a common rail network should be in use in both the freight and the passenger rail component of the NPTM. Separate information on passenger and on freight rail service provision and characteristics would certainly need to be kept for the separate modules, but the underlying physical network should perhaps be maintained in common.

While this approach of maximising the use of the current GBFM is attractive for the present, care will be required to ensure that it does not act as a longer term brake on achieving the wider goals of the Department. Ultimately achieving an effective national modelling system in which the LGV, passenger and HGV traffic can all be assigned to an equilibrium solution on the same detailed highway network is an important goal. The continuing role that road congestion and pricing policies are likely to play in policy decisions highlights its importance. The recent successful implementation of zone-pair sampled assignments in the NPTM has ensured that such functionality is already in operation for passengers, so that it is important that the freight modelling software and functionality can also take this on board eventually and be directly interfaced to the NPTM.

A careful balance will need to be struck between the funding of the short term enhancements to GBFM and the funding of the wider developments required to put in place a model that will meet the more general national modelling aims of the Department. The dangers from this potential conflict in resourcing will already be well understood from the persistence of the FORGE system, which was originally introduced as a stopgap in the NRTF modelling system.

10.1.2 Task set B – Research and implement improved demand modelling

This set of tasks is innovative in terms of the modelling methodology that is proposed, and parts of it would greatly benefit from data (e.g. future CSRGT information on type of land use at origin and destination of trip) that as yet has not been collected and processed. Accordingly, even assuming that development of some of its components were to start soon, a fully operational model representing all of the tasks here would still only become fully available some time in the future.

There is a benefit to running various of the tasks here in parallel so that the methodological and software developments can be in place when the required data and analysis tasks have been completed.

In the same way that the GBFM can provide a useful starting point for the task set C work, the EUNET Trans-Pennine model provides a useful starting point for this task set B work. It covers the whole of GB (albeit with large zones outside the Trans-Pennine Corridor) and contains within its Regional Economic Model, operational versions of many of the features that are required for spatial I/O modelling. However, the demand database underlying EUNET is by now out-of-date, and the model contains no representation of logistic stages.

The task B3 here would also provide inputs to the task set C, while the completion of task B5 would depend on outputs from task set C.

10.1.3 Task set C – Implement mode split and supply modelling

This comprises two main interlinked activities:

- Developing and implementing base matrix estimation techniques in task C3 to make best use of a range of sources of data on demand. Success in this task is crucial to many of the other tasks, given the limited spatial detail available in practice from CSRGT/LGVS by virtue of their sample sizes.

- Using this enhanced data to aid in model development and calibration

This latter set of tasks mirrors many of the components already in the GBFM. Here the emphasis will be on:

- using the most up-to-date information currently available and refining the processing of the data into a form suitable for the model

- increasing the spatial detail and refining various aspects of segmentation and of the representation of market based behaviour

- taking account in the calibration of the increased segmentation of freight by logistic stage that is generated within the task set B activities

- ensuring that the operational modelling system (primarily the highway assignment) can ultimately be interfaced directly to the NPTM system.

The inclusion or not in this model of an explicit representation of time of day of travel is a matter to be decided eventually based on whether the types of policies of most interest in the short term to the Department would require this feature. Representing time periods themselves is not difficult, though modelling the choice between them is not straightforward. Explicit time of day representation does however consume significant computing resources to set up and run. Many of the most interesting impacts relate to the influence on the degree of congestion on the highway network that would result from retiming of freight trips. However, until such time as the NPTM also contains a network based representation of time of day of travel, there would only be limited benefit to introducing it in isolation into the HGV assignment.

10.1.4 Task set D – Run NFTM through to the future, linked to the economy

This task would commence once the task sets B and C had largely been completed and would run the complete NFTM system through future years and test its usage for policy testing and for general traffic projections.

10.1.5 Task set E –LGV model

The growth of light commercial and goods vehicle is likely to continue, and the role of the service industries in this growth is a key part of this emerging issue. The gaps in knowledge and modelling in this area need to be addressed, and ideally would involve one or more higher degree students to create expertise in this area and to research its links between logistics and light vehicle movements.

When a sufficient volume of processed data eventually becomes available from the LGVS, this task could be commenced in earnest. This postponement is fundamentally unsatisfactory however, and consideration should be given to synthesising the LGV freight traffic when the freight model is being initially implemented, in order to provide a more holistic treatment of the freight system and of the substitution between HGV and LGV.

The subsequent use of the LGVS data when it becomes available will involve some repetition of the activities that will have already have been completed for HGVs in the task set C, but this time for the freight that is carried in LGVs. Because the LGVS sample size will be much smaller than the ideal, it will be necessary to maximise the use of other data sources (link counts, RSIs) to increase the effective detail of the information on LGV movements. This task will require some limited recalibration of the whole system to ensure that best use is being made of the LGV freight information.

10.2 The medium to longer term development of freight models

A key source of the complexity that arises in the development of freight models lies in the inherent diversity of the range of goods that is produced in the modern economy as well as in the diversity of the consumers of these goods. This heterogeneity leads to a need to create a large number of segments at each stage of the modelling process to represent both the goods that are transported, and the range of transport and distribution services that are used to move these shipments of goods around the country. This plethora of segments when allied to the need to maintain a reasonably fine level of spatial detail then leads to onerous requirements for computer storage and power.

An alternative to the aggregate modelling approach described above is the agent based micro-simulation of the transport and distribution of individual freight shipments as described in Section 8.2. Note that this application of micro-simulation takes the consignment as its primary unit, as such it should not be confused with traffic micro-simulations of vehicles on networks, as produced by software packages such as Visum and Paramics. For example, the Portland freight micro-simulation model (Donnelly, 2002) uses a conventional user equilibrium assignment model to put vehicles onto modal networks.

While freight micro-simulation is also computationally onerous, it provides a natural representation of the diversity in behaviour based on individual decision making, rather than on decisions averaged over a demand or supply segment as a whole. In this manner it facilitates the introduction of a more focused picture of the large number of micro markets that together determine overall travel patterns. Through identifying individual producing firms and individual consumers, and then through sampling the range of consignments that pass between them, it enables the operation of each part of the freight transport system to be represented more realistically. This finer grain in representation in turn provides benefits in enabling transport policies to be tested in a more transparent fashion. Because it operates at the individual firm and consignment level, the software tools (e.g. scheduling and routing optimisation) that individual firms use themselves can be reflected directly within the system to represent operations more realistically.

Why then has this approach been currently consigned to the longer term development category? Given the scale of the data and model development requirements to make the system operational, it would be unwise to initiate experiments with this approach at the *national* scale. It would be faster and cheaper to start *soon* with a smaller study area, such as London. Once a fully operational model is tested and in place there, it could then be extended to cover GB as a whole. The micro-simulation model would replace large parts of the Phase I model, but would require most of the Phase II model to remain in place. In this way postponing the *national* level development of the micro-simulation model is not an inefficient use of resources.

10.3 The modelling of LGVs

Given the current absence of any suitable data source, the modelling of both the freight movements and of the service movements on LGVs should be postponed until the LGV surveys have been completed and the results have been fully checked and processed.

The partitioning of the car trips from the van trips in the NPTM need not be delayed and should be carried out as part of the next round of development of the NPTM. Currently the car mode includes that part of the LGV movements that are within the NTS survey. This partitioning is needed for two reasons:

1. to enable the private movements in LGVs to the be modelled explicitly within the NPTM. This segment would complete the representation of the LGV sector, when eventually combined with the results of the other two models of freight and of service movements on LGVs.

2. to ensure that the expansion methods discussed previously, which depend heavily on the use of link based LGV count data, will include within the LGV matrices that are being expanded all of the component LGV movements, whether they be for private, commercial or freight purposes.

The freight LGV model once its underlying database becomes available needs to be integrated into the overall freight modelling system. The separate model of commercial trips in LGVs would be developed in parallel.

10.4 Freight modelling below the national scale

The initiatives at the national level will provide useful insights into freight modelling at the sub-national scale, together with control totals for use when forecasting as already outlined in Section 8.1.

The main recommendations are

1. to improve the quality of the base year local vehicle matrices of LGVs and HGVs through making effective use of a wider range of observed data

2. to ensure that future year GV matrices take appropriate account of trip length increases and are not simply scaled up fixed matrices

3. in parallel with the standard relatively simple freight models, to create a small number of more sophisticated and well resourced sub-national freight models that would act as a test-bed both for new techniques (e.g. freight micro-simulation models) or for the testing of policies such as lorry bans, low emission zones, etc. that intrinsically have a local rather than a ubiquitous effect.

10.5 Recommendations

The key recommendations from the report are briefly itemised in Table 10.2 and Table 10.3 below. A reference is given for each, which shows the location within the main text of this report where further detail on the recommendation and supporting background can be obtained. Table 10.2 is structured under the following categories:

- Policy Recommendation to UK decision makers (UK Govt., Local Authorities, etc)
- Research Basic research requirement (probably to be instituted by UK Government).
- Data Data collection/analysis task (Statistical Agencies)

Each task is further classified: by whether it is a short, medium or long term task, and by whether it is of high or of normal priority. Many of these tasks have previously been included in Table 10.1 in which their linkage to other related pieces of work has been illustrated.

Table 10.2 Key recommendations in Policy, Research and Data for the Way Ahead

Prior-ity	Recommendation	Ref.
	Policy	
S, H	P1 - Standardisation of zoning systems in use in different modelling studies to optimise resource use	3.1.2
S, N	P2 - Standardisation of network use across different local scale studies (DfT, devolved authorities & Highways Agency)	3.2.1
M, N	P3 - Adopt the XML based GTF as the standard means whereby datasets are interchanged between different spatial scales or different models	7.4
	Data	
S, N	Q1 - Value density requires improved domestic data collection/analysis	2.3.1
S, H	Q2 - Larger sample sizes are urgently required for both LGVS and CSRGT	7.1
S, H	Q3 - Collect land use data at collection and delivery points in the CSRGT to relate legs of O/D matrices back to original P/C matrices from which they derive.	2.6, 7.1, 7.2
S, N	Q4 - In future the CSRGT and LGVS questionnaires should each be extended to request the location of the furthest stop on the multistop tour from the start base (or from the start and end bases if it is a multistop trip which does not return to the initial base). – Data on start/end time of trip is needed for time of day modelling – Under reporting of traffic from CSRGT should be investigated	4.11.1, 7.1
M, N	Q5 - Processing of CSRGT data should be examined to see if more differentiation within "miscellaneous articles" category is publishable for past and/or future years- ditto for coding of locations to the district rather than just county level.	5.2, 7.1
S, H	Q6 - Definitions of all common or overlapping factors in LGVS, CSRGT & NTS should be revised to be consistent.	3.5.1, 5.1, 7.1
S, H	Q7 - Primary data collection/analysis required to link volume of commodities lifted/dropped to type and scale of production/consumption activity.	2.5.1, 7.2
S, N	Q8 - Commission detailed study to measure source and level of biases in existing traffic counts of LGVs	5.4.1
S, N	Q9 - Survey foreign HGV activity within the UK for one year, at least	7.1
S, N	Q10 - Support trials of telematics systems to ascertain their potential to substitute/complement other data sources	7.2
S, H	Q11 - Create Transport Satellite Accounts	7.3
	Research	
S, H	R1 - Research into factors responsible for past trends in freight lengths of haul and tonnes lifted.	2.3
S, H	R2 - Research into the trends, forecasting and freight modelling implications of value density and handling factors is needed.	2.3.1
S, H	R3 - Research based on observed data to improve understanding on how best to relate legs of O/D matrices back to original P/C matrices from which they derive.	2.6
S, H	R4 - Case studies of the major companies involved in generating freight movements would allow better identification of market structures and assist in correctly specifying model processes.	7.2
M, N	R5 - Research how to retain theoretical richness of SCGE approach within realistic models	4.2
S, H	R6 - Research to understand and quantify underlying reasons for rapid recent growth in LGV trips.	5.5

Key to codes: S, M, L: short, medium, long term task. H, N: high, normal priority

Table 10.3 summarises the recommendations on model structure and content for guidance to the modellers commissioned to develop specific modules of the NFTM.

Table 10.3 Key methodological recommendations for modelling freight at the national level

Recommendation	**Reference (Section no.)**
Internal Study Area will cover Great Britain but not Northern Ireland. The model will focus on domestic traffic and the domestic legs of international traffic.	3.1
The national model zoning scheme must be at least at County level, but Local Authority District level would be the preferred option if feasible.	3.1
A further set of point zones is needed to represent freight ports and airports, including a zone for the Channel Tunnel. May need expanding in future years for new facilities.	3.1.1
Zoning systems in lower tier models should be consistent with the zoning in use within the Department's other main modelling initiatives.	3.1.2
Use NPTM highway network in the National Freight Model	3.2.1
Networks must represent intermodal movements consistently: assign road legs to feed rail terminals or ports in same manner as a movement wholly by road.	3.2.3
Develop procedures to construct realistic base matrices to use all relevant data	3.3
Use specific demand segmentation most appropriate to each model stage	3.4
Use of incremental modelling approach to be considered	4.1, 4.8.1
Do NOT forecast demand independently for each mode	4.2
Estimate future freight demand at the segmented sectoral/commodity level and NOT as a total estimate of tonnes to be subsequently disaggregated by commodity	4.2
Use spatial I/O model or sophisticated segmented physical trend model to estimate future overall demand	4.2
Create P/C matrix by combining available observed data within model structure	4.4
Produce future year P/C matrix by: forecasting zonal pattern of production and consumption, and taking account of changing trends in patterns of trade.	4.5
Convert from P/C matrices segmented by commodity type to O/D matrices segmented by shipment type and by aggregate commodity type.	4.6.2
Explicitly locate major distribution centres, but treat most smaller ones as shadow distribution centres.	4.6.2
Explicitly identify means by which operating costs are built up, through combining individually identified cost components.	4.7
Exclude LGVs from early stages of freight model development until data becomes available but provide suitable slots for their later integration.	4.8
Use disaggregate calibration procedures for modal split including quality of service characteristics.	4.8.1
Time of day choice should NOT be included in the short term due to major added complexities – the model design should facilitate its introduction when data and resources become available	4.9
Use user optimum assignment function (NOT system optimum)	4.11

11 CONCLUSION

Within this study the team has carried out a wide-ranging review:

- of the requirements for freight models in GB, and
- of the approaches to modelling that have been used in the UK and in other parts of the world.

This understanding has been used to develop a framework for the modelling of freight movements at the national level and then to outline how non-freight related LGV movements might also be modelled. The means by which the national level model might be utilised to provide data to improve highway modelling at finer spatial scales was outlined, together with other recommendations for improving these smaller scale models.

Considerable progress has been achieved in developing a comprehensive overall framework for freight modelling. As part of this framework the report introduces a coherent and integrated representation of many important features that have in the past been introduced infrequently, if at all, in freight modelling in the UK:

1. Various options of increasing sophistication and realism (and unfortunately of increasing associated data requirements) for linking growth in the economy to growth in the demand for freight transport.

2. A distinction between:

   - the pattern of economic trade in goods from the initial producer to a final consumer, which is termed the *Production/Consumption (P/C) zone pair matrix of trade,* and
   - the actual set of physical transport movements generated by the logistics structure that is used to distribute and transport these P/C trades in practice, which is termed the *Origin/Destination (O/D) zone pair matrix of shipments*

   It is only through retaining this distinction within the model, that policies which impact on logistics behaviour can be represented in a coherent fashion with an appropriate spatial representation.

3. An innovative general purpose spatial input-output methodology that could be used to build up the P/C patterns of transport of goods from the original producers to the eventual consumers,

whether these be industries or households. This same mathematical methodology could likewise be applied in a linked fashion to build up the actual transport O/D movements for each logistical stage that arises in transporting these goods from the producer to the consumer.

4. A focus on changes through time in the average length of vehicle trips – since this has been the primary past source of freight traffic growth - and not solely on growth in the number of vehicle trips.

5. The adoption of an incremental modelling approach to ensure that the heterogeneity inherent in freight demand patterns can be retained within the models.

6. An approach to modelling the different types of trips that use LGVs, which integrates with freight and passenger modelling and data sources in a consistent fashion.

7. Improvements in methods for local freight modelling, and clear procedures for making best use of forecasts from the national model to improve local estimates of freight demand.

However, freight modelling is acknowledged by most of those who have experienced it, to be substantially more complex than passenger transport modelling. It has certainly been under studied in the UK up to the present. Furthermore, freight does not have an equivalent database to the NTS for passengers. The understanding of overall passenger demand patterns available from analysis of the NTS has enabled rapid progress to be generated recently in developing a national passenger model. An equivalent comprehensive database for freight would be more complex to develop, but its absence necessarily slows down the pace of progress in developing a robust national freight model.

Accordingly, there are major challenges that have been described in this report, which need to be addressed subsequent to this study if suitable road and rail freight models are to be put in place in a reasonable timeframe. The fact that the overall framework and the methodological approach by which this framework could be implemented have been successfully outlined above, should not divert attention from the further research, data collection and implementation requirements that are needed to achieve successful models in a timely and cost effective fashion.

In summary, the challenges and risks lie in: the relative complexity of the modelling structure that is required in order to have an adequate representation of the operations and response of the freight markets themselves, coupled with the inadequacy of the existing data sources to facilitate the research and implementation details to underpin some of the important components of the model.

Particular challenges include the following:

a) The comprehensive spatial input-output modelling framework that is recommended, both for representing overall freight demand and the logistics stages of transport within it, will require careful experimentation and validation to ensure that it performs satisfactorily with the limited data that will initially be available to support it.

b) Because of the costs and difficulties in assembling large sample surveys of HGVs, and especially of LGVs, the methodology to develop incremental models based on effective interconnections of link counts, RSIs and survey matrices will prove very important. It will take resources plus some time and experimentation to make sure that it is fully effective.

c) LGV modelling starts from an existing position of having virtually no data on demand patterns, while the sample sizes envisaged for use in the surveys currently in preparation by DfT are relatively small. This implies that it could be many years before sufficiently detailed data would be available to enable a robust model of LGV movements to be calibrated.

However the existence of an overall modelling framework provides a coherent structure within which the development tasks can be formulated. The proposed way ahead would make effective use of the current building blocks already in place, in order to provide short-term substitutes for some of the more challenging elements of development. The multi-pronged approach to creating the national freight model provides time for the research and data collection to be developed that would provide the required support for such elements.

12 REFERENCES

ASH Research Group (1997). *Framework Implementation Manual, Technical Note 6.3*. National Transport Model Feasibility Study, Project 2 for UK, DETR. WS Atkins Planning Consultants.

Bates J.J., Ashley D.J. and Hyman G. (1987). The nested incremental logit model: theory and application to modal choice. *Proceedings of 15th PTRC Summer Annual Meeting, 1987*. PTRC Education and Research Services Ltd, London.

Bates J.J. and Williams I.N. (1993). APRIL - a strategic model for road pricing. *Proceedings of Seminar D, PTRC Summer Annual Meeting, 1993*. PTRC Education and Research Services Ltd, London.

BRIDGES (2000). *The BRIDGES Final Report*. http://www.mcrit.com/BRIDGES/Bridges.htm

Browne, M., Wigan M., Allen J. and Anderson S. (2002). Understanding the growth in service trips and developing transport modelling approaches to commercial, service and light goods movements. *Proceedings of the European Transport Conference 2002, Cambridge*. PTRC Education and Research Services Ltd, London.

Bröcker J. (1998). How would an EU-membership of the Visegrâd countries affect Europe's economic geography? *The Annals of Regional Science,* vol 32 no. 1, pp 91-114.

Bruzelius N. (2001). *The Valuation of Logistics Improvements in CBA of Transport Investments: A Survey*. Paper prepared for the Swedish Institute for Transport and Communications Analysis (SIKA).

Caliper Corporation (2002). Planning and Travel Demand Modeling with TransCAD. http://www.caliper.com/TCTravelDemand.htm

Cambridge Systematics (1976). *Guidelines for Travel Demand Analysis of Program Measures to Promote Carpools, Vanpools and Public Transportation*. US Department of Energy.

CSO (1995). Input-output Tables for the United Kingdom 1990. HMSO, London.

David Simmonds Consultancy and Marcial Echenique and Partners Ltd. (1999). Review of Land-Use/transport Interaction Models. Report to the Standing Advisory Committee on Trunk Road Assessment (SACTRA), DETR, London.

DETR (1998). *National Road Traffic forecasts (Great Britain) (1997), Working Paper No. 3, Non-car traffic: Modelling and forecasting*, DETR, London

Donnelly, R. (2002). *Development of the TLUMIP commercial travel (CT) component.* Working paper prepared by Parsons Brinckerhoff for the Oregon Department of Transportation, available from http://www.odot.state.or.us/tddtpau/modeling.html .

DTLR (2001). *Transport of Goods by Road in Great Britain, 2000 - CSRGT.* The Stationery Office, London.

DTLR (2002). *Survey of Land for Mineral Workings in England, 2000.* The Stationery Office, London.

DMRB (1996). *Design Manual for Roads and Bridges*, Volume 12 Section 2.2

Echenique M.L. and Williams I.N. (1982). O-D matrix production from cordon survey data. *Traffic Engineering and Control*, vol. 23, no. 12, pp 584-589.

Fowkes A.S., Nash C.A., Toner J.P. and Tweddle G. (1993). *Disaggregated Approaches to Freight Analysis: A Feasibility Study*, ITS University of Leeds, Working Paper 399, prepared for DoT

Fowkes A.S. and Toner J.P. (1998). *Freight Mode/Route Choice Modelling with Limited Data*, Proceedings of PTRC 26th Summer Annual Meeting, Loughborough, Seminar E: Transportation Planning Methods, Volume P424, pp 147-164, PTRC Education and Research Services Ltd, London

Gunn H, Mijjer P, Lindveldt K and Hofman F (1997). Estimating base matrices: the combined calibration method. *Proceedings of the 25th European Transport Annual Meeting. Transportation Planning Methods,* pp 375-385. PTRC Education and Research Services Ltd, London..

Heydecker B G, Verlander N.Q., Bolland J.D. and Arezki Y. (1995). Improved matrix estimation under congested conditions. *Traffic Engineering and Control*, vol. 36, no. 12, pp 667-672.

INRO (2002) STAN. http://www.inro.ca/products/stan.html

Jin Y. and Williams I.N. (2000). A new regional economic model for European transport corridor studies. *Proceedings of the 28th European Transport Annual Meeting. Transportation Planning Methods, Seminar F*. PTRC Education and Research Services Ltd, London.

Kumar, A (1980). Use of incremental form of logit models in demand analysis. *Transportation Research Record*, Vol. 775, pp 21-27.

Leontief W. (1986). *Input-Output Economics*, 2nd ed. Oxford University Press, New York.

Leontief W. and Strout A. (1986). Multiregional input-output analysis. In *Input-Output Economics*, 2nd ed., editor W. Leontief, pp 129-161. Oxford University Press, New York.

List, G., L. Konieczny, C. Dumford and V. Papayanoulis. (2001). *A best practice truck flow estimation model for the New York City region.* Proceedings of the Eighth TRB Conference on Transportation Planning Methods, Corpus Christi, Texas, 22-26 April.

Logie M (1993). Advances in estimating O-D trip matrices. *Traffic Engineering and Control*, vol. 34, no. 9, pp 441-445.

Mandel, B. and Ruffert, E. (2001). *Technical Note 5:GTF Conceptual Model (Specification)* version 0.6 revisions 5 & 6 in *SPOTLIGHTS Work package 4 Deliverable D13*. Last updated 10/12/01. Revision 5 is also available at: http://gtf.mkm.de/

Mayers, I., Proost S., Quinet, E., Schwartz, D. and Sessa C. (2001). Alternative Frameworks for the Integration of Marginal Costs and Transport Accounts. UNITE Deliverable 4. Funded by DG TREN of the EC in the 5th Framework RTD Programme. ITS, University of Leeds, UK. http://www.its.leeds.ac.uk/projects/unite/

McFadden D. (1978). Modelling the choice of residential location, in Karlqvist, A., Lundqvist L., Snickars F. and Weibull J.W., eds., *Spatial interaction theory and planning models*, pp 75-96. North Holland, Amsterdam.

ME&P (2000a). The EUNET Trans-Pennine Model, Final Report. Prepared for Contract No. PPAD 9/65/28 for UK DTLR.

ME&P (2000b). *Equilibrium Modelling in NRTF2000*. Final Report – Phase 1, prepared for Contract No. PPAD 9/65/68 for UK DTLR.

MVA (1997). *National Transport Model Feasibility Study Project 1: Policy Model*. Final report prepared for Contract No. 02/C/1993 for UK DETR.

MYSTIC (2000) MYSTIC Towards Origin-Destination Matrices for Europe, Final Report, Contract no ST-97-SC-2101 for DG TREN of the EC http://europa.eu.int/comm/transport/extra/final_reports/strategic/MYSTIC.pdf

Neffendorf H., Wigan M., Donnelly R., Williams I., and Collop M. (2001). The Emerging Form of Freight Modelling, *Proceedings of the European Transport Conference 2001, Cambridge*. PTRC Education and Research Services Ltd, London.

Netherlands Economic Institute et al., (1999). *REDEFINE: Relationship between Demand for Freight-transport and Industrial Effects*, Final Report, Contract No. RO-97-SC.1091, http://europa.eu.int/comm/transport/extra/rep_road.html .

Nguyen S. (1984). Estimating origin-destination matrices from observed flows. In M Florian (ed.) *Transportation Planning Models*. Elsivier, North-Holland.

ONS (1997). *National Accounts Concepts, Sources and Methods*, First Draft. Edited by E A Doggett, Office for National Statistics. The Stationery Office; London.

Pike J. and Gandham R. (1980). *Review of the Commodity Flow Studies* (1975-1979) Transport Operations Research Group Research Report 36, University of Newcastle

SACTRA (1999). *Transport and the Economy.* Report of the Standing Advisory Committee on Trunk Road Assessment, DETR. The Stationery Office; London.

SCENES (2001). *Transport Forecasting Model: Calibration and Forecast Scenario Results*. Deliverable D7 of the SCENES project, funded by the European Commission (DG TREN). ME&P, Cambridge. http://www.iww.uni-karlsruhe.de/SCENES/

Tavasszy, L.A., M. van de Vlist, C. Ruijgrok and J. van de Rest (1998). *Scenario-wise analysis of transport and logistic systems with a SMILE;* paper presented at Session B3, 8th WCTR conference, Antwerp, Belgium.

Transek, CERUM, Temaplan, INRO, TØI and NEA (2001). *Ideas for a new Swedish Freight Model*. Report for SAMPLAN, Sweden. http://www.sika-institute.se/utgivning/sam01_1.html

U.S. Department of Transportation, Bureau of Transportation Statistics (1999). *Transportation Satellite Accounts: a New Way of Measuring Transportation Services in America*. Washington DC. www.bts.gov/transtu/tsa92.pdf

Van Zuylen H.J. and Willumsen L.G. (1980). The most likely trip matrix estimated from traffic counts. *Transportation Research B*, vol. 14, no. 3, pp 281-293.

Venables, A. and Gasiorek, M. (1998), *The Welfare Implications of Transport Improvements in the Presence of Market Failure.* Report to the Standing Advisory Committee on Trunk Road Assessment (SACTRA), DETR, London.

Wigan, M.R. (2001). Valuation of supplier choice factors by Australian shippers . *6th Logistics Research Network Annual Conference* Sept 2001, Heriot-Watt University Edinburgh, Institute of Logistics and Transport.

Wigan M.R. and Drain P. (2003 - submitted). Open Transport Modelling Networking for transportation applications. *82nd Transportation Research Board Conference Forum, January 2003*, Washington DC, Transportation Research Board.

Wigan M.R. and Rockliffe N, 1998, *Freight Survey Requirements for Urban Areas*, Proceedings 19th ARRB Conference, Sydney, 7-11 December 1998, Volume 1, pp.109-131.

Williams H.C.W.L. (1977). On the formulation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A*, vol. 9, pp 284-344.

Williams H. and Yamashita Y. (1992). Equilibrium forecasts of travel demand and investment benefits measures for congested transport networks. *Proceedings of Seminar E, PTRC Summer Annual Meeting, 1992*. PTRC Education and Research Services Ltd, London.

Williams I.N. (1979). An approach to solving spatial allocation models with constraints. *Environment and Planning A*, vol. 11, pp 3-22.

Williams I.N. (1994). A model of London and the South East. *Environment and Planning B*, vol. 21, pp 535-553.

Williams I.N. and Beardwood J.E. (1993). A residual disutility based approach to incremental transport models. *Proceedings of Seminar D, PTRC Summer Annual Meeting, 1993*. PTRC Education and Research Services Ltd, London.

Williams I.N. and Echenique M.H. (1978). A regional model for commodity and passenger flows.. *Proceedings of PTRC Summer Annual Meeting, 1978,* Stream F, pp 121-128. PTRC Education and Research Services Ltd, London.

Williams I.N. and Lindsay C. (2002). An efficient design for very large transport models on PCs. *Proceedings of the European Transport Conference 2002, Cambridge. Transportation Planning Methods*. PTRC Education and Research Services Ltd, London.

Zhang X., Maher M.J. and Van Vliet D. (1999). Methods for the solution of the combined trip matrix estimation and stochastic user equilibrium assignment problem. *Proceedings of the 27th Association of European Transport Annual Meeting. Transportation Planning Methods,* pp 109-120. PTRC Education and Research Services Ltd, London.

This report makes extensive reference to a series of eight review reports that were written as part of the work areas A and B in the course of this study. They describe the current situation in a variety of topics related to freight modelling and to provide more detailed information on many of the topics discussed in this final report. These are:

Inception Report (incorporating A2 - Trends in Freight Transport)
A1 - Issues in Logistics and the Freight Industry
A3 - Review of the Light Goods & Commercial Vehicle Sector
A4 - Rail Freight
A5 - Other Modes
B1 - Review of GB Freight Models
B2 - Review of Models in Continental Europe and Elsewhere
B4 - Data Sources

APPENDIX I: RELATIONSHIP BETWEEN LOGISTICS TRENDS AND ROAD FREIGHT TRANSPORT ACTIVITY FROM THE REDEFINE PROJECT

This section contains extracts from the REDEFINE project final report.[5] This information has simply been copied from this document and appears here for information purposes.

The REDEFINE project (Relationship between Demand for Freight-transport and Industrial Effects) was a Transport RTD project funded by the European Commission that ran for 16 months from March 1997.

From page 3 of REDEFINE report:

The framework for this analysis is illustrated in Figure 2.1 which shows the relationship between the value of goods produced and road freight traffic demand as a series of key ratios.

- Value density: This ratio is used to convert the value based data recording economic activity into a weight based measure of production output. The ratio is derived from trade data. Though it bares some similarity to the density (i.e kg/m^3) of goods, it does not, necessarily, indicate any change in the physical density of goods in the economy.

- Handling factor: The ratio of tonnes-lifted to the weight of goods produced. Each item of production may be loaded onto a transport means several times during its movement along the supply chain (no doubt, on some occasions as part of a consignment): the item can therefore be recorded in transport statistics several times. The handling factor is a measure of this multiple counting and can be regarded as a rough index of the number of separate links in the supply chain. An item transported from a factory to a warehouse, from where a second truck delivers it to a retailer, would illustrate (part of) a supply chain with two links, and a production item with a handling factor of 2.

- Modal split: The division of the tonnes transported by the various modes of transport (e.g. sea, air, rail, and road) is called the modal split. A change in road transport's share affects the relationship between economic growth and road freight traffic demand. Road's share was expressed as the tonnes of production moved by road against the weight of all production.

- Average length of haul: This is the average distance of a loaded lorry's journey and, therefore, the average distance that each unit of freight moves on a single journey. This ratio is considered to be an average measure of the length of each link in the supply chain.

- Load factor (or lading factor): This is the ratio of what a lorry actually carried (in terms of weight) compared with what it could have carried if it was loaded to its maximum weight, weighted by the distance the lorry covered while carrying any load.

- Empty running: This is the proportion of vehicle-kms run empty against total vehicle-kms.

If each of these ratios remained stable, road freight traffic would be perfectly correlated with changes in the value of goods produced. In practice, each of these ratios can vary independently. By estimating changes in each of

[5] Netherlands Economic Institute et al., 1999, *REDEFINE: Relationship between Demand for Freight-transport and Industrial Effects*, Final Report, Contract No. RO-97-SC.1091.

the key ratios through time, it should be possible to establish how much of the growth of lorry traffic is a function of economic growth and how much is attributable to logistical changes.

APPENDIX II: MAKING ROAD FREIGHT MILEAGE RESPONSIVE TO COSTS

This Appendix contains updated extracts from Chapter 5 of the Final Report – Phase 1 (ME&P, 2000b) for the project Equilibrium Modelling in NRTF2000 carried out in 1999 by ME&P for ITEA at DTLR. It illustrates a procedure that may be used to match a synthetic mode to an observed O/D matrix in a form that is elastic to cost changes and which allows exogenous trends in trip lengthening to occur in future years.

The main steps in the task are summarised as follows:

1. Obtain base national and international O/D matrices of heavy goods vehicle (HGV) movements – possibly classified by vehicle size, and possibly classified by broad commodity types to enable different future growth patterns and different elasticities of response to be introduced for these segments.

2. Assign these vehicles to the national road network and check against link traffic counts to ascertain whether the O/D matrices, the classified link traffic counts and the national statistics on vehicle kilometres by vehicle type are consistent with one another. Resolve any incompatibilities between these different estimates.

3. Based on the travel costs and times derived from this assignment, run a doubly constrained trip distribution model for each vehicle type segment and produce a synthesised O/D vehicle matrix. Apply a residual disutility (RD) procedure. This will generate RD values that ensure that the synthesised matrix exactly matches the observed matrix for every observed zone pair.

4. Adjust the parameters of the trip distribution function so that for each segment the elasticity of response of the trip lengths in the matrix after the inclusion of the RD values, is consistent with those estimated from external studies.

5. Apply policy based cost and time changes as required to the road network, and recalculate the resulting set of O/D travel disutilities. Rerun the distribution models for each vehicle segment and assign the vehicles to the network to estimate the revised patterns of movements and mileages. This step could be carried out with base year matrices. Alternatively, the vehicle trip ends could be factored up to expected future year levels prior to running this distribution model stage.

The more complex steps 3 to 5 above are now explained in more detail.

II.1 Making freight O/D matrices elastic to transport costs

This section outlines an approach to trip distribution modelling that enables trip lengths to be elastic to cost changes for this base matrix.

Even though the notation and presentation are quite distinct between say Huw Williams and Yamashita's (1992) constrained incremental distribution model and Ian Williams and Beardwood's (1993) hybrid spatial distribution model based on the use of residual disutilities, the underlying mathematical structures are very similar in the two approaches. Both are extensions to the incremental distribution model. The following extract from this latter paper (where the term "travel disutility" denotes the generalised cost of a trip) describes the model's use in more detail.

"Hybrid spatial distribution models

Another advantage from using the residual disutility approach to trip distribution is that it provides an elegant generalisation of the two main types of distribution models that have been used widely in the past.

The oldest distribution models are matrix updating models which use some form of scaling (e.g. Frater, Furness) to update a base year observed matrix, so that it matches the current year pattern of origin and destination zone trip totals. These methods however, do not easily take account of the way in which localised improvements in transport infrastructure will tend to change the spatial pattern of trip distribution, or of the trends towards increasing trip lengths for passengers or of increasing lengths of freight haulage in some economic sectors.

To avoid this implied inelasticity in demand with respect to changes in the supply of transport, gravity type models were developed, which estimate the spatial distribution pattern of trips based on the generalised cost of transport between zones. This approach, while it may work well for regular urban passenger travel purposes, such as journeys to work, tends to be of more limited value when applied to long distance passenger or freight movements.

Accordingly, what ideally is required is a model that can combine the best features of both the gravity type and the matrix updating approaches to modelling the spatial distribution of travel. This can be achieved by use of a residual disutility based incremental gravity type model. If the model is run without any interzonal travel disutilities, it essentially simplifies to a matrix updating model. Otherwise, it maintains the strengths of both traditions, with the balance in the influence of the two approaches being an output from the calibration process.

If the observed distribution pattern is largely explained by a gravity model, then the variance of the travel disutilities will be large and the variance of the estimated residual disutilities will be small. If the generalised costs of transport play a less important role, then the relative magnitudes of these variances will be reversed. The spatial distribution of the pattern of trips in all cases will be elastic to changes in the characteristics of the supply of transport."

II.2 Setting up the model for the base year

To explain the approach in more detail, the method used to calculate the residual disutility terms that are integral to the hybrid distribution model is explained initially. Then the steps required in calculating future year matrices are specified.

The mathematical form of a synthetic doubly constrained model can be written using the notation of Annex E of DMRB v12.2.2 as

$$T_{ij}^{p} = T_{i.}^{p} T_{.j}^{p} e^{-\mu^{p}\left(c_{ij}^{p} + z_{ij} + s_{j}^{p} + r_{i}^{p}\right)} \qquad \text{(II.1)}$$

T_{ij}^{p} is the number of trips from zone i to zone j

c_{ij}^{p} is the generalised cost or time of travel from zone i to zone j

Here the superscript p denotes that the variables refer to the base year (pivot case) matrix of trips and travel costs. The origin and destination constraint adjustment terms are defined as:

$$r_{i}^{p} = (1/\mu)\ln \sum_{j} T_{.j}^{p} e^{-\mu^{p}\left(c_{ij}^{p} + z_{ij} + s_{j}^{p}\right)} \qquad \text{(II.2)}$$

$$s_{j}^{p} = (1/\mu)\ln \sum_{i} T_{i.}^{p} e^{-\mu^{p}\left(c_{ij}^{p} + z_{ij} + r_{i}^{p}\right)} \qquad \text{(II.3)}$$

It can easily be demonstrated that the definitions (II.2) and (II.3) ensure that the destination and origin zone constraints $T_{i.}^{p}$ and $T_{.j}^{p}$ are exactly satisfied for every zone. The values of the variables r_{i}^{p} and s_{j}^{p} are calculated by a standard iterative balancing procedure, which starts with initial values of zero for each variable.

The variable z_{ij} denotes an OD matrix of residual disutility terms that are calculated in a manner that ensures that the synthetic matrix that is produced by the equation (II.1) will be identical to the pivot case matrix for every zone pair. Unlike the other elements in equation (II.1) the residual disutility values do not get updated through time since they are determined purely from the base year pattern of travel. These residual disutility values are calculated as follows.

If the total number of zones is n, then there will be n^2 elements, z_{ij}, to be evaluated. However, because of the origin and destination zone constraints of equations (II.2) and (II.3) the number of degrees of freedom in z_{ij} is only $(n-1)^2$, so that $2n-1$ of the residual disutility elements can be arbitrarily set to zero. Accordingly, arbitrarily set

$$z_{i1} = 0 \quad z_{1j} = 0 \qquad \text{for all origin zones i and destination zones } j \qquad \text{(II.4)}$$

After some algebra the required formula to estimate the residual disutility for all elements i,j except where either $i=1$ or $j=1$ can be written as

$$z_{ij} = c_{i1}^{p} + c_{1j}^{p} - c_{11}^{p} - c_{ij}^{p} + (1/\mu^{p})\ln\left[\left(T_{i1}^{p}T_{1j}^{p}\right)/\left(T_{ij}^{p}T_{11}^{p}\right)\right] \qquad \text{(II.5)}$$

A formula that is similar in structure to that of equation (II.5) can be derived for the calculation of the residual disutility for the origin or destination singly constrained model.

The formulation can be generalised straightforwardly to cover the case where the pivot matrix is available only at a more aggregate zoning system and/or a more aggregate segmentation of type of movement than is in use when estimating the synthetic matrix.

II.3 Creating future year matrices

Because the model is in a synthetic matrix form (II.1), it is a straightforward task to adjust the value of the concentration parameter μ to ensure that the average length of haul of the matrix matches some exogenously specified value. The steps to create the trip matrix for the required future year *t* are as follows:

1. use the information from the pivot case matrices of trips and costs in equation (II.1) to calibrate the value for the concentration parameter μ^p which matches the observed mean trip length $\bar{c}^{p}$.

2. calculate the values of the residual disutility terms using equation (II.5)

3. apply suitable local growth factors, to the trip ends to generate future year zonal trip end totals as $T_{i.}^{t}$ and $T_{.j}^{t}$ and then determine the future year mean trip length $\bar{c}^{t}$ that provides an increase in trip mileage in the study area which matches the reference case growth rates.

4. in equations (II.1), (II.2) and (II.3), use these future year trip ends and the future year reference case matrix of costs to adjust the value of the concentration parameter μ^t, so as to match the expected future year mean trip length $\bar{c}^{t}$. The future year trip matrix is then produced by equation (II.1).

This matrix will remain close in pattern to the pivot case matrix, in the sense that if there is no change in the matrix of costs or in the value of the concentration parameter, then the results will be equivalent to a Furness procedure applied to the new trip ends. The degree of departure from this simple Furnessed matrix will increase smoothly, broadly in line with the degree of change in the cost matrix and with the rate of change required in the mean trip length.

APPENDIX III: MATHEMATICAL STRUCTURE OF SPATIAL I/O MODELS

This Appendix presents a mathematical formulation of the spatial input/output (I/O) model, and outlines the ways in which it can be used to meet a wide range of the needs identified within this freight model review. It endeavours to provide a precise mathematical specification of a range of topics that have been discussed in a less formal manner in the main part of the report. Its primary aim is to present a complete and consistent specification of all of the main component elements of the demand modelling required for this freight study.

The structure of a standard Leontief (1986) I/O model is explained firstly below, with the main terminology being introduced in italics. Then this structure is generalised to a spatial I/O model, which in turn is used to show how the distribution chain of the successive movement of goods within a logistics process can be represented effectively. Finally the manner in which this spatial I/O model approach can be used as a foundation for modelling the location of industries, services and distribution centres is described.

III.1 Background

The original formulation of the spatial I/O model was presented by Williams and Echenique (1978), based on a freight model they had developed for the State of Sao Paulo. The solution algorithm was then generalised in Williams (1979) to take account of zonal constraints, and to demonstrate how the method could be used to model the spatial competition governing the location patterns of firms, of households or of agricultural crops. It has been used in a variety of operational models since then including:

- to model the location patterns of households and services, including an endogenous estimation of the zonal rents for residential and commercial floorspace, within the LASER integrated land use and transport model (Williams, 1994)

- to model the transport of passengers and freight in the EUNET Trans-Pennine corridor model (Jin and Williams, 2000)

- to model the transport of freight throughout all of the EU in the SCENES model, which includes separate I/O tables for each of the EU member states (SCENES, 2001).

The spatial I/O model provides a flexible system that can be used to represent in an integrated fashion:

- the functional linkages between activities – production and consumption of goods and services and their associated costs

- the spatial competition for location of activities – land and building stock, planning policies and rents.

- the spatial linkages between activities – transport and communication

III.2 The mathematical structure of the I/O model

The standard manner in which to relate the economic linkages between different economic sectors is through use of the input/output modelling framework first developed by Leontief in 1936. The fundamental concept is that the production of each commodity, an *output*, consumes a variety of other commodities as *inputs*. Each such input itself needs to be produced, and so in turn they also will each consume inputs of other commodities in the course of their *chain of production*, and so on. The term "commodity" here denotes both goods and services (e.g. transport, warehousing, insurance, etc.) The mathematical formulation of the Leontief I/O model is quite simple

$$D^m = Y^{m0} + \sum_n a^{mn} X^n \qquad \forall m \qquad \text{(III.1)}$$

D^m denotes the total consumption of the commodity m and is the sum of the *final demand* plus the *intermediate demand*

Y^{m0} denotes the amount of *final demand* of the commodity m. It represents the outputs that are consumed by households, government and exports, rather than by producing industries.

X^m denotes the total amount of production (output) of the commodity m. A market clearing equation is required which specifies for each commodity m that the requirements for consumption are satisfied by the quantity produced.

$$X^m = D^m \qquad \forall m \qquad \text{(III.2)}$$

a^{mn} denotes the amount of input of the commodity m required to produce one unit of output of the commodity n. It is generally referred to as the *technical coefficient* or the *I/O coefficient*. The elements in each column n break down the inputs to production among the different commodities m. The matrix of coefficients is estimated through use of surveys of the purchasing patterns of each of the economic sectors producing these commodities and is periodically updated and published in national or regional I/O tables (CSO, 1995).

$\sum_n a^{mn} X^n$ denotes the *intermediate demand* which is the amount of the commodity *m* that is consumed as part of the process of the production of all types of commodities *n* (including the commodity *m* itself).

A simple example will explain the logic of the I/O structure. Households consume final demand for electricity for a variety of needs in the home. This electricity in the power station in which it is produced, in turn will consume various intermediate inputs of coal or gas, of labour, spare parts, services, etc., including some electricity itself. The production (mining) of this coal will in turn consume labour, electricity, transport, etc. The technical coefficients measure the amount of each such intermediate input that is consumed in the production of one unit of output for that commodity.

I/O models of the national economy are generally implemented using monetary units to represent the *value* of production (e.g. dollars, sterling, euro) rather than in physical units (e.g. tonnes) to represent the *volume* of production. The set of UK I/O national tables for 1990 (CSO, 1995) contained 123 separate commodity types for the domestic production of a complete range of goods and services within the economy, as well as a separate representation of imported goods and services.

III.2.1 The choice of a commodity or of an economic sector based classification

The I/O tables of coefficients may be set up using one of two different forms of classification:

- by the *economic sector* producing or consuming the transactions, or
- by the *commodity type* of the transactions that are produced or consumed.

Because a specific economic sector may produce more than one type of commodity, these two classifications are not synonymous. For example, a car manufacturer (economic sector: vehicle manufacturing) that uses its own fleet of lorries to deliver automotive components between its factories, is **producing** the commodity road transport, as well as producing the main commodity vehicles. However, if instead it hires a public haulage firm to deliver these components, then the economic sector, vehicle manufacturing, is **consuming** the commodity road transport, which is in this case produced by the economic sector road transport. Accordingly, road freight transport as considered in freight models corresponds more completely to the **commodity** classification, road transport, rather than to the **economic sector** classification, road transport.

Within the set of published I/O tables for the UK, the "Make Matrix" presents the interconnection between the economic sector based and the commodity based classification. The columns of the make matrix classify the producers by the economic sector to which they belong, whereas the rows classify the outputs from production by the commodity type of the good or service that is produced. Although much of this matrix is concentrated on the diagonal, it can also be seen (CSO, 1995) from the values in the row for the service commodity, road transport, that most *industrial* economic sectors produce road transport as an output commodity. This refers to the own-account road transport within each industrial sector.

In practice within the 1990 make matrix for the UK (CSO, 1995), the relatively small magnitude of the proportions shows that it is only be a part of the total UK own-account transport that has actually been captured in this data. In particular, the coefficients for own-account road transport have been universally set to zero for all columns of the agricultural, construction and service producing economic sectors. In order to meet the data requirements for freight modelling in this current study, there would need to be enhancements to the data collection and processing underlying the UK I/O tables. The creation of Transport Satellite Accounts, as discussed in Section 7.3, would enable the inaccuracies in the measurement of the transport activity to be corrected. This would ensure that the commodity type measurement of total freight transport activity would include not just the hire-and reward services currently measured in the transport sector itself, but also the own-account transport services which are currently hidden within the other industrial and non-industrial economic sectors. The split between these two forms of transport service should then broadly match the split that is published in the statistics of the CSRGT (DTLR, 2001).

To summarise, for the reasons explained above the most appropriate table of I/O coefficients to utilise for purposes of freight modelling is the commodity type by commodity type (rather than the economic sector by economic sector) matrix, since this will match best to the commodity based concept that underlies the CSRGT and other transport data sources. The commodity road transport, is produced both by the economic sector road transport, which represents the road haulage firms, and also is produced as a secondary or ancillary output from other economic sectors, which represent own-account road freight services. Similar definitions apply to other freight modes. The national accounts for transport need to be corrected through use of Transport Satellite Accounts.

III.3 The mathematical structure of the spatial input/output model

The original development of I/O models mainly concentrated on macroeconomic applications and contained no treatment of the detailed spatial locations of production or consumption. The initial attempt to introduce a spatial dimension was through what is termed the *multiregional I/O model*

(Leontief and Strout, 1986). This methodology is cumbersome to implement and use at the spatially detailed scale and so has not been widely adopted for transport models.

The spatial I/O model (Williams, 1979) is an alternative generalisation of the I/O model that integrates within an I/O modelling framework, the features of a discrete choice based spatial distribution model for the location of production. In essence it uses the same approach as in the standard I/O model but it explicitly recognises that the production of commodities will frequently take place in a different location to that in which they are consumed so that the process of the transport of these commodities from production to consumption should be represented explicitly. In this way differences in the cost of production and the cost of consumption between different locations can be calculated explicitly. It thus provides a natural structure for the modelling of the location decisions of manufacturing and service industries.

The generalisation of the original mathematical formulation of equation (III.1) to create the spatial I/O model is as follows:

D_c^m denotes the amount (the units of measurement are discussed in Section III.4 below) of consumption of the type of commodity *m* within the consumption zone *c*. It includes both the final demand Y_c^{m0} in zone *c* and the intermediate demand that is consumed in the production of other commodities within the zone *c*. It is calculated as

$$D_c^m = Y_c^{mo} + \sum_n a_c^{mn} X_c^n \qquad \forall c, m \qquad \text{(III.3)}$$

a_c^{mn} is the I/O coefficient defined in the same manner as in Section III.2, except that a subscript *c* is introduced to denote that different production technologies may exist in different zones, which in turn would create different patterns of inputs to production in these zones. For simplicity in practice, often the technical coefficients matrix is assumed to be constant across all zones within a region or even a country.

X_c^n denotes the amount of the commodity *n* that is produced as output in the zone *c*.

The total consumption summed over all of the consumption zones *c* in the study area is matched by an equal total production summed over the production zones *p* in the study area: over all zones market clearing is assumed of each individual commodity type *m*:

$$\sum_p X_p^m = \sum_c D_c^m \qquad \forall m \qquad \text{(III.4)}$$

In practice the formulation of the model is simplified if an identical zoning system is adopted for both the production and the consumption zones. However, within any one specific zone (*c=p)* it is not expected that the zonal consumption D_c^m would be equal to the output from production X_p^m in that same zone.

The total demand D_c^m for commodity *m* in consumption zone *c* must be met by an appropriate allocation of outputs from the set of production zones.

T_{pc}^m denotes the amount of the commodity *m* that is transported from the production zone *p* to meet the demand for consumption within zone *c*. It is calculated as

$$T_{pc}^m = D_c^m \, p_{p|mc} \qquad \forall c, m, p \qquad \text{(III.5)}$$

$p_{p|mc}$ denotes the probability that the commodity *m* that is consumed within the consumption zone *c*, will have been transported there from the production in the zone *p.* This probability can be calculated using a standard logit type formulation of a discrete choice location model as

$$p_{p|mc} = S_p^m \, e^{-\lambda u_{pc}^m} \Big/ \sum_{p'} S_{p'}^m e^{-\lambda u_{p'c}^m} \qquad \forall c, m, p \qquad \text{(III.6)}$$

S_p^m is a zonal size term that denotes the capacity of the zone *p* for the production of the commodity *m*. It represents the relative scale of the attraction of each of the competing zones.

Because of market clearing

$$X_p^m = \sum_c T_{pc}^m \qquad \forall m, p \qquad \text{(III.7)}$$

The set of equations (III.3) to (III.7) defines an interlinked set of relationships across space and across commodity types that pertain to the production and consumption of goods and services. Ultimately they are used to estimate the **amount** of each commodity that is transported from each production zone to each consumption zone.

An analogous set of equations is presented below that defines the set of **unit costs** and of **unit disutilities** that pertain to the different stages of this production and consumption system.

III.3.1 Chains of costs and disutilities in the spatial I/O structure for production

Before commencing the presentation of the disutility equations, it is helpful to clarify some notational issues.

Because of the inherently recursive nature of the spatial I/O formulation (e.g. electricity will be consumed in the production of electricity) the creation of an unambiguous and simple notation is a challenge. In the equation (III.3) a variety of input commodities, denoted by the superscript *m* are consumed in the zone *c* in the course of the production of the output commodity *n* in that zone. Now consider a specific one of these input commodities *m*, that is itself produced in a production zone denoted by *p*. This production in the zone *p* of the commodity *m* will in turn consume a variety of input commodities which are denoted by the superscript *k* as used in the equation (III.8) below. These input commodities *k* themselves are produced in the set of zones denoted by *q*.

To summarise, within the commodity notation the superscripts *k, m* and *n* each refer to the same collection of types of commodities but the differentiation is needed to clarify relationships of inputs to outputs in what in reality is a recursive formulation of linkages between commodities. Likewise, within the zonal notation the subscripts *c, p* and *q* each refer to the same collection of zones, with the differentiation being used to distinguish the location of production from the location of consumption within a recursive spatial formulation.

Within the equation (III.6) for the location of production, the term

u_{pc}^{m} denotes the zone pair disutility (generalised cost) of the production of one unit of the commodity *m* in the production zone *p*, plus the disutility of transporting it from there to the consumption zone *c*. It is calculated as

$$u_{pc}^{m} = \left(\sum_{k} a_{p}^{km} \tilde{u}_{p}^{k} \right) + \hat{u}_{pc}^{m} \qquad \forall c,\ m,\ p \qquad \text{(III.8)}$$

$\hat{u}_{pc}^{m}$ denotes the *transport disutility* associated with transporting one unit of the commodity *m* from the production zone *p* to the consumption zone *c*. It is an output from the transport model.

The transport disutility (which is a function of the cost, time and quality of service) for each mode between the pair of zones is initially calculated along the paths through the modal networks. These disutilities are next composited across modes, and then across the distribution legs available in the logistics chain, as described in Section III.5 below. Through taking account of the typical patterns of consignment sizes that are associated with specific types of distribution legs, the relationship between unit transport costs and the size of consignment is explicitly taken into account in the formulation in that section.

The zonal *production disutility* (generalised cost) of one unit of the commodity *m* in the production zone *p* is denoted by

$$\sum_k a_p^{km} \tilde{u}_p^k \qquad \forall m, p \tag{III.9}$$

It is calculated by summing the total *consumption disutility* (cost) of all of the input commodities that are consumed in the course of the production of one unit of the output commodity *m* in the zone *p*. This total consumption disutility is the product of how much of each input is consumed, times the unit consumption disutility of this commodity within the zone *p*.

$\tilde{u}_p^k$ denotes the composite *consumption disutility* of one unit of the commodity *k* in the consumption zone *p*. It is a composite (rather than an arithmetic average) over all of the production zones *q* from which the commodities have been sent for consumption in the zone *p*. It is calculated using the log-sum formulation as

$$\tilde{u}_p^k = (-1/\lambda)\ln \sum_q S_q^k e^{-\lambda u_{qp}^k} \qquad \forall p, k \tag{III.10}$$

In summary, this consumption disutility (price at point of final consumption in zone *c*) includes both the initial production disutility (the factory gate price in zone *p*) of the commodity and the disutility of transporting the commodity from the zone *p* where it was originally produced to the zone *c* in which it is finally consumed. Furthermore, it also includes recursively the second and higher order transport costs from zones *q* from which the second and higher order intermediate inputs to production have been sourced.

III.4 Relationship of the spatial I/O and SCGE models to the original I/O model

In the original economic modelling context formulated by Leontief a number of the concepts that were introduced in Section III.3 did not occur explicitly. However, it will be shown below that this more general spatial I/O model formulation simplifies naturally to the original formulation in a consistent fashion.

The units of measurement: the amount of commodities produced, consumed and transported can be measured either in units of monetary value or of physical tonnes. The applications of the Leontief model to Western economies are in units of value, although some early usages elsewhere to represent the bulk inputs of iron ore, coal, etc. consumed in heavy manufacturing industries, were in physical units.

Location: there is no explicit concept of location in the original Leontief formulation so that the explicit distinction is missing there between the production (factory gate) price calculated via equation (III.9) in the zone of production, and the consumption cost calculated via equation (III.10) in the zone of consumption. Instead transport is treated as a commodity like any other, and so transport costs are a uniform, ubiquitous input from the transport sector to the overall costs of production of each economic sector. In the spatial I/O model, in contrast, the precise transport costs relevant to each sector of production in each zone are correctly accounted using the equations (III.10) and (III.8). Moreover, all of the upstream (e.g. second, third order, etc.) costs of transporting intermediate input commodities at earlier stages of the chain of production are also passed through to the later stages of production, by virtue of the equation (III.10) above.

Disutilities: in the spatial I/O model the disutilities are a generalisation of the simpler monetary costs used in the original value-based Leontief formulation, in which there is no representation of the quality of a good or service or of the heterogeneity in their characteristics. If all goods and services are fully homogeneous then the parameter λ in the equations (III.6) and (III.10) would be set to infinity to denote that all goods consumed in a zone would be transported from the lowest cost location only, without any variance of modes or of production zones. Cost minimising models would then replace discrete choice logit based models and the disutility terms would contain pure monetary costs.

In practice however, the observed heterogeneity in the markets for goods, services and transport, together with the coarseness of the segmentation in the aggregate data, have led the developers of strategic freight models to discard simplified cost minimising formulations in favour of stochastic choice models that enable this observed heterogeneity to be represented. Leontief and Strout (1986) reached a similar conclusion: "after having experimented with linear programming models, we now avoid explicit use of the cost minimisation or revenue maximisation principle in the basic formulation of the multiregional scheme". The use of disutilities in parallel with monetary costs enables the influences of heterogeneity and of quality to be represented within a stochastic framework.

Production functions: the use of fixed I/O coefficients represents the simplest form of production function in economic modelling. The value of output is a simple linear sum of the values of all relevant inputs. If required, the production technologies may differ between production zones as shown by the use of a zonal subscript in the coefficient a_c^{mn} in the equation (III.3). However, there is no representation of economies of scale or of agglomeration, so that the amount of output produced is measured simply as the sum of the costs of each of the inputs, including also a profit margin and tax.

In this simplified case the cost/disutility of production can be set to 1 in all zones, as is the norm in the original I/O model. Since the unit of measurement of quantity is already in value (monetary) terms, the price per unit of output is by definition 1.

III.4.1 SCGE and other generalisations beyond the fixed Leontief coefficients

To represent economic systems more realistically, as well as to represent the competition for the location of industries and services, there are benefits to replacing the fixed I/O coefficients by more sophisticated production functions for firms, and by explicit consumption functions for households. This would generalise the spatial I/O model into what is often termed a *spatial computable general equilibrium* (SCGE) model. In the SCGE model the production function may be a constant elasticity of substitution (CES) form or some other more complex form in which features such as: increasing returns to scale, agglomeration economies (using a Dixit-Stiglitz formulation), imperfect competition, etc. can be represented. With these more complex production function forms there is a need to distinguish explicitly between the costs of the inputs and the amount of the resulting output that is produced. Price determination in this case is endogenous rather than being fixed as in the Leontief I/O model.

The distribution of trade between the production and consumption zones within the spatial I/O model is based on discrete choice models which act by endeavouring to minimise costs, while taking account of the uncertainty and heterogeneity that is inherent in the real world. These use a logit discrete choice structure to represent both the competition between producers in different locations, and the diversity of products within an industry. In particular, the log-sum composite cost form pioneered by Huw Williams (1977) is used in place of the constant elasticity of substitution (CES) sub-utility function that is used in many of the SCGE models, in order to provide a price index that represents the diversity of products supplied from a specific industry to a single location for consumption. McFadden (1978) has shown that the CES and multi-level logit forms can each be considered as simplifications of the same General Extreme Value family of choice models.

Another element that could be introduced into the production function is to explicitly represent the competition for location in constrained space through endogenously estimating location rents by zone (Williams, 1979). This would enable land use policies to be introduced into the models through varying the costs and availabilities of land or floorspace between locations. Finally, in SCGE models consumption functions for households are generally introduced to close the loop between the income passed to labour and the resulting expenditure of their households that then generates the final demand for goods and services.

Examples of models that include various of these more advanced features are found in the SCGE models of Venables and Gasiorek (1998) and of Bröcker (1998), and within the UK in the regional economic model component of the EUNET Trans-Pennine model of Jin and Williams (2000). A general review of these methods is provided in the report to SACTRA by David Simmonds Consultancy & ME&P (1999).

The calculation of converged solutions to more sophisticated spatial I/O models that have a variety of zonal constraints, transformations and endogenous price determinations may be computationally challenging. The algorithm presented in Williams (1979) has proved to be efficient for this purpose and has worked effectively in a wide range of different contexts for complex models. Although it may not be always possible to prove mathematically that there is only a single unique solution to the set of equations, in practice for genuine applications in the real world, multiple solutions have never been found.

III.5 Applications to represent distribution chains

Some of the inherent complexity in the formulation below can be better understood through noting that there are in effect three distinct but intertwined recursive structures that are to be solved simultaneously:

1. a production chain connecting the consumption of a set of intermediate input goods to the production of a specific output good – this was introduced in Section III.2

2. a spatial structure connecting origin zones to destination zones and thus generating transport movements – this was introduced in Section III.3

3. a distribution chain connecting a specific type of good from its original zone of production through DCs to its ultimate zone of consumption via different types of distribution legs – this is introduced in the current section

In Section III.3 the mathematical structure of I/O relationships was used to represent the manner in which a variety of input goods would be consumed in the process of being transformed into new types of output goods through the *production chain*. The production chain commences with the initial crude primary input goods and then passes through many intermediate production stages, to arrive eventually to the production of complex, finished manufactured goods for consumption as the final demand.

In this current section the mathematical structure of I/O relationships is used in a different manner to represent the way in which a specific type output good would be transported along the *logistics distribution chain* from the initial producer through to the final consumer. Unlike in the production process, here the good is not transformed but is merely repackaged as it follows the distribution chain from the original zone of production through various shipment legs and distribution centres to the final zone of consumption. The physical manifestation and context is different between the production chain and the distribution chain, but the mathematical representation can be common to both. To aid understanding, it is convenient to extend the notation in this section to make the separate logistics stages more explicit. However, this extended notation is primarily to aid understanding and does not represent any significant extension to the general mathematical formulation of the model already presented in Section III.3 above.

There is, however, a change in concept below in the spatial distribution model, which supplies goods to the zones of consumption. The equations (III.5) and (III.6) which estimate the transport flow from the production zone *p* to the consumption zone *c* in Section III.3, are replaced below by a more complex procedure that takes explicit account of logistic distribution stages. It is not that the transport flow T_{pc}^{m} of equation (III.5) is disaggregated. Instead it is replaced by a more general procedure that adds a further recursive layer to the model formulation. This builds up the transport flows in a manner that reflects the logistics options that are available at each transport stage, connecting the original production zone through to the ultimate consumption zone for a specific commodity.

The logic behind the approach is first illustrated, and then the mathematical formulation presented previously is extended to show how it can be introduced in practice into a formal model.

The demand for consumption of some good of type *m* in the zone *c* can be satisfied either by sending goods directly from a production factory (as in Section III.3) or else from a distribution centre (DC) in which the goods have been warehoused between distribution legs. To represent this latter option an extra dimension is introduced to denote the type of distribution leg *l* within the choice set facing a consumer located in the zone c. For example, if appropriate goods are actually produced in the zones that are nearby, then the most cost-effective form of delivery from these nearby zones to the consumer would be the distribution leg of type: direct from producer to consumer. In contrast, if the required goods are only produced in distant locations, they are most likely to be delivered to the final consumer from a DC located adjacent to the consumer, using the distribution leg type: DC to consumer.

This approach is generalised into a recursive procedure that commences with the set of ultimate consumption zones and meets their demands by transporting goods from producers or DCs, via the appropriate distribution leg type. The goods that are output from each DC will first have needed to be delivered to the DC, either from original producers or from other DCs, using the appropriate distribution leg types. The solution algorithm continues until all distribution chains have been traversed back to the original production zones. The dimensional scale of this solution algorithm is not overwhelming because within a DC the goods are pooled and the exact origin and distribution leg type of each input shipment is not remembered. The model uses only the overall probability distribution as estimated by the equation (III.12) below.

Define the following extensions to the existing notation (noting as previously explained in Section III.3.1 that the recursive nature of the formulation leads to a proliferation of letters in the subscripts and superscripts that are used to denote the recursive chain structures explicitly):

T_{jc}^{ml} denotes the volume of the good *m* that is shipped on a distribution leg of logistics type *l* from a DC or production factory in the zone *j* to meet the demand for consumption (either for ultimate consumption or for intermediate warehousing) within zone *c*.

X_j^{ml} denotes the volume of the good *m* that is transported on a distribution leg of logistics type *l* from the DC or factory in zone *j* to meet all of the demands for consumption elsewhere. It is calculated as

$$X_j^{ml} = \sum_c T_{jc}^{ml} = \sum_c D_c^m p_{jl|mc} \qquad \forall j, l, m \qquad \text{(III.11)}$$

$p_{jl|mc}$ denotes the probability that a consignment of the good *m* that is consumed within the consumption zone *c*, will have been transported there on a shipment on a distribution leg of logistics type *l* from the zone *j*. This probability can be calculated using a standard logit type discrete choice model as

$$p_{jl|mc} = S_j^{ml} e^{-\lambda u_{jc}^{ml}} \Big/ \sum_{j'l'} S_{j'}^{ml'} e^{-\lambda u_{j'c}^{ml'}} \qquad \forall c, j, l, m \qquad \text{(III.12)}$$

S_j^{ml} is a zonal size term that denotes the capacity of the zone *j* for the delivery of the commodity *m*. Depending on the type of distribution leg *l* it will represent either the size of the relevant warehousing in the DCs in the zone or the capacity of the factories producing the good there.

The final two equations that complete the extension of the set of original equations (III.3) to (III.7) of Section III.3 to represent distribution chains are:

$$D_c^m = Y_c^{mo} + \sum_{nl} a_c^{mnl} X_c^{nl} \qquad \forall c, m \qquad \text{(III.13)}$$

$$\sum_{pl} X_p^{ml} = \sum_c D_c^m \qquad \forall m \qquad \text{(III.14)}$$

By comparing the set of equations (III.11) to (III.14) with the corresponding set of equations (III.3) to (III.7) it can be seen that the former has merely introduced an extra dimension *l* for each occurrence of a transported input from a production to a consumption zone. This shows that the differences between the two formulations are simply notational and do not imply any differences in their fundamental mathematical structures.

The estimation in this more general distribution chain formulation of the total amount of a good *m* that moves from a specific production zone *p* to a consumption zone *c* requires an extra calculation to combine the different distribution legs that may be used between this pair of zones.

$$T_{pc}^{m*} = T_{pc}^{m1} + \sum_j \left(T_{jc}^{m2} p_{p3|mj} \right) + \sum_j \left(T_{jc}^{m2} \sum_i p_{i4|mj} p_{p3|mi} \right) + \ldots \qquad \text{(III.15)}$$

Here the terms on the right hand side denote respectively the direct, two leg, three leg, etc. shipment combinations where the distribution leg types *l* are numbered as follows:

1. direct from the initial production zone to the final consumer
2. from a DC to a final consumer
3. from initial production to a DC
4. from DC to another DC.

III.5.1 Chains of costs and disutilities in the spatial I/O structure for distribution

To complete this presentation the required disutility/cost equations for the extended distribution chain formulation are now presented:

u_{jc}^{ml} denotes the zone pair disutility (generalised cost) for one unit of the good *m*, which comprises both the disutility/cost of delivery to and of storage at the DC within the zone *j,* plus the subsequent transport disutility of the goods via a distribution leg of logistics type *l* through to consumption within zone *c*. It is calculated as

$$u_{jc}^{ml} = \left(\sum_k a_j^{kml} \tilde{u}_j^k \right) + \hat{u}_{jc}^{ml} \qquad \forall c, j, l, m \qquad \text{(III.16)}$$

$\hat{u}_{jc}^{ml}$ denotes the *transport disutility* associated with transporting one unit of the good *m* from a DC or a factory in zone *j* to the consumption zone *c* on a distribution leg of logistics type *l*. It is an output from the transport model.

This transport disutility (which is a function of the cost, time and quality of service for a mode) between the pair of zones is initially calculated along the paths through the modal networks. The transport disutility also includes the cost of loading and unloading at each end of the modal shipment.

The reason why the logistics type *l* is distinguished as a characteristic of the shipment and of its associated transport disutility $\hat{u}_{jc}^{ml}$ for a zone pair, is that the requirements for transport may differ between different distribution leg types when shipping goods between a given pair of zones. For example, yoghurts being transported from a factory to a warehouse of a major supermarket chain in a zone would almost certainly be delivered in a large HGV as a single drop consignment, quite possibly utilising the vehicle on a 24 hour basis for such deliveries. In contrast, deliveries of yoghurts from this factory to a variety of small retailers that are located in the same zone as the warehouse would typically be carried out in smaller vehicles, probably as part of multi-drop routes, and with little scope for delivery outside the shops' opening hours. Accordingly, in this latter case the cost of transport per yoghurt might be significantly higher due to the absence of opportunities for economies of scale in the transport.

There may also be a need to distinguish between shipments to large concentrated consumers and those to a range of small consumers in dispersed locations throughout a zone, and likewise to distinguish between major national DCs and smaller regional/local DCs. This distinction would enable the typical size and regularity profile of the consignments that are transported, to be defined in a more discriminating fashion for each distribution leg type. This in turn will have a major impact on the modal competition, on the most appropriate sizes of vehicle to use for the transport on each type of distribution leg, and hence also on the unit monetary costs of transport for each type of distribution leg.

The modal disutilities for a distribution leg are then composited across the modes using the standard log-sum formula to provide the required overall value for the transport disutility of the distribution leg. These shipment legs correspond to the O/D movements that are observed on road in the CSRGT and on rail in the rail freight survey data, provided that the rail movement is door-to-door rather than an intermodal shipment with collection and/or delivery by road.

The zonal *distribution disutility* (generalised cost) of one unit of the good *m* that is available in a DC in the zone *j* is denoted by

$$\sum_k a_j^{kml} \tilde{u}_j^k \qquad \forall j, m \qquad \text{(III.17)}$$

It is calculated for the good *m* by summing the total *delivery disutility* (cost) of the various shipments of that good that are received in the DC, plus the disutility/cost components associated with storing and processing one unit of the good at this DC.

The superscript *l,* to denote the distribution leg type within the technical coefficient, is introduced in order to distinguish distribution type activities from production type activities. If the distribution leg type denotes that the shipment commences from a producer, then the production formulation and interpretation of the technical coefficients that has been outlined in Section III.3 persists.

If however, the distribution leg type *l* denotes that the shipment commences from a DC then the distribution formulation and interpretation now presented is what is appropriate.

To ensure that the intermediate costs of transport and production are correctly passed through the distribution system, the value of the coefficient a_j^{mml} is set to unity in equation (III.17) in this case where it is representing a distribution process, as opposed to the production processes of Section III.3.

$\tilde{u}_j^m$ denotes for the case $k = m$ the composite *delivery disutility* of one unit of the good *m* in the consumption zone *j*. It is a composite (rather than an arithmetic average) over all combinations of production and/or DC zones *i* and of distribution leg types *l* via which the goods may be sent to the DC in the zone *j*. It is calculated using the log-sum formulation as

$$\tilde{u}_j^m = (-1/\lambda) \ln \sum_{il} S_i^{ml} e^{-\lambda u_{ij}^{ml}} \qquad \forall j, m \qquad \text{(III.18)}$$

In summary, this delivery disutility (price at point of distribution) includes:

- the initial production disutility (the initial factory gate price) of the good
- the cumulative disutility of all upstream stages of transporting the good from the zone where it was originally produced, via any intermediate DCs, through to the current DC
- the cumulative disutility of all associated operations in all of the upstream DCs through which the goods have passed up to this current logistics stage.

The value of the coefficient a_j^{kml} for $k \neq m$ represents the amount of the input of type *k* required per unit of *m* when using the DC in the zone *j*.

$\tilde{u}_j^k$ denotes for this case $k \neq m$ the unit disutility/cost of each component input of type *k* associated with the operation of the DC in the zone *j*. These input types would include the administration and operating costs of the DC, any costs of repackaging the goods, and the capital costs of the goods *m* tied up during their dwell time at the DC.

The set of equations (III.11) to (III.14) can be solved repeatedly, moving down the distribution chain starting from the consumption zones, through all of the distribution leg options until the initial production zones are reached. In this way all of the zonal consumption requirements will ultimately be met by production in some zone (or by imports). In the opposite direction, the cumulative set of production, transport and warehousing disutilities/costs of equations (III.16) to (III.18) are pushed up the distribution chain from the initial production zone in a manner that enables the choices between logistical options to be calculated at each stage on the basis of the full cumulative set of costs and disutilities associated with that option. In this manner there is a transparent process by which the network based transport and warehousing costs are passed consistently up through each choice process in the logistics chain.

To summarise the overall picture, it can be seen from comparing the mathematical formulation above in this section, which describes the distribution chain, against the formulation in Section III.3 above, which describes the production chain, that both are representations of the same general spatial I/O mathematical process - they are merely applied in different contexts. This demonstrates two major benefits from this general formulation:

- it simplifies the software development that would be required in order to represent this overall modelling system

- it guarantees that the production chains and distribution chains can be intertwined within a single mathematical formulation without introducing extra mathematical complexity in the solution system.

Because the solution of the production chain system of Section III.3 has been shown to converge in practice in a variety of operational models such as EUNET and SCENES, there are strong theoretical grounds to expect it to exhibit similar convergence behaviour in the more general combined logistical distribution chain and production chain system.

III.6 Implementation issues in practice

The main challenge in implementing the above structure in practice is not the modelling software or the coherence of the approach. The existing MEPLAN modelling package contains all of the functionality to represent this complete mathematical formulation. In fact most elements (other than the distribution chain) have already been implemented in operational MEPLAN based models such as EUNET and SCENES.

The greatest complication at present lies in the significant gaps in the availability of the data that would facilitate the introduction of this structure. The most important data issues are as follows:

- The data sources on the functional relationships between commodity types are only available in value units (e.g. the I/O coefficients) whereas the spatial patterns of freight movements are only available in volume units.

- There is no direct observed zonal data on the total value of production or of consumption segmented by type of commodity – such zonal values can only be approximated by disaggregating national or regional totals for the commodity through making use of a proxy variable such as the zonal employment in the relevant economic sector, which data in general is available.

- The observed data on physical freight flows by commodity type on either road or rail, do not contain the dimension, type of distribution leg.

The upshot is that the inherent complexity in implementing the freight modelling formulation discussed here is significantly increased by these and other lesser gaps in data. However, these issues are not insurmountable within GB as has been shown by the successful implementation of the production chain methodology of Section III.3 within the EUNET model using existing data sources.

III.7 Representing how transport costs influence the location of firms

One of the most appealing features of the mathematical structure of the spatial I/O model is that it provides a natural interface from transport characteristics through to industrial location models. The zonal production disutility as defined in equation (III.8) in Section III.3 for the producing industries, and the distribution disutility as defined in equation (III.17) in Section III.5 for the distribution sector, each encapsulate the full set of direct and indirect upstream transport disutilities and costs. In this manner the zonal production/distribution disutility provides a complete and consistent measure for use to represent the intrinsic transport advantage of that zone for the production industry or DC of interest. Accordingly this term is the appropriate measure to introduce into the zonal production function that is in use in the model.

The manner in which this measure of transport costs is introduced into the Leontief fixed I/O coefficient production function is now outlined. In the standard published I/O coefficient matrices, transport provision is represented as one or more commodity types (or more strictly, service types) denoting the different modal transport services provided (see Section III.2.1). The coefficient specifies the proportion of each transport service that is consumed per unit of output from each industry in turn. This proportion is calculated from purchasing surveys within each industry as: the national average expenditure on transport per unit of production for that domestic industry type.

In the spatial I/O model a more precise approach is adopted instead to represent the cost of the transport inputs to production. Instead of using an exogenous national average transport cost input that is constant across all zones, as contained in the I/O table of commodity based coefficients for the transport sectors, the actual network based transport costs are endogenously calculated for each zone in the model. This significantly improves the accounting of transport costs and the measurement of their impact upon producers and consumers.

In order to ensure that money is consistently accounted when using a Leontief I/O production function, it is only the monetary transport cost (not the transport disutility) that enters the I/O cost accounting. The calculation of the zonal monetary costs of the transport inputs to production uses a set of equations that are the monetary cost based equivalents of the disutility equations (III.8) to (III.10) presented previously. The corresponding coefficients from the I/O table for the consumption of transport services are reset to zero for every industry type to avoid double counting. These input transport costs have already been accounted directly within the model, via the monetary cost based equivalents of equations (III.8) and (III.10). This enables the exact zonal incidence of these costs to be measured consistently.

The transport disutility (rather than the monetary cost) is always used in the spatial distribution model. This ensures that the wider influences on location decisions are fully represented there. When using more sophisticated production functions than the Leontief I/O coefficients, the inclusion of a distinction between quantities and their unit costs enables more complete use to be made of the disutility values within the production function.

III.8 Industry and distribution centre location models

This section outlines the structure of the industrial location model that has been implemented in the EUNET Trans-Pennine model (ME&P, 2000a). It provides an illustration of how the spatial I/O approach lends itself naturally to represent the spatial influences on the location of firms or of DCs.

The model assumes that the pattern of industrial location is influenced by competition as manifested by the spatial pattern of the costs of production that would be faced in each zone by an industry.

Production in a zone that has a lower total cost in meeting the demand of consumers is assumed to grow more rapidly than production in other zones that would generate higher consumer prices. In the industrial location model, the I/O relationship maintains a spatial dimension as it builds up the production costs for each type of industry in each zone. This is best explained separately for the goods/services consumed by the firm, and for the employees who commute to work there.

For a consumer in a given model zone, the quantity of goods/services demanded from each producing zone is governed by the probability estimated by the single level multinomial logit model in equation (III.6). The following cost/disutility components are used therein to calculate the probability of a zone being chosen:

(a) the factory gate price of the goods/service produced in that zone, which includes all of the input costs/disutilities of production and of labour as calculated in equation (III.9). It also includes any zone specific taxes or subsidies and the rent for the land and floorspace used.

(b) the transport disutility of the good/service from that production zone to the consumer's zone

(c) the residual attractiveness of that zone for production – estimated as part of the incremental model structure, if used

(d) the residual trading disutility from that zone to the consumer's zone – estimated as part of the incremental model structure, if used.

Items (c) and (d) are estimated during the calibration using respectively, the observed amount of goods/services produced in the zones, and the observed amount of trade taking place between the zones. They represent the influence on industrial location and trade patterns of all aspects other than those explicitly included via (a) and (b) above.

The volume of production being located in a zone is determined in equations (III.5) and (III.6) by the cost and disutility elements listed above as well as by the parameter λ that represents the consumer's sensitivity to disutility variations amongst the zones. This parameter is estimated such as to match the modelled pattern of trade to the observed mean trip length of the goods or service in question.

As well as including the zonal costs of all of the input goods and services to production in the term (a) above, the zonal cost of labour also needs to be included since it may be a major determinant of locational attractiveness, especially for labour intensive service industries. Commuting travel is derived from labour demand. In a sense labour is treated as an input to the production process, in a way analogous to the inputs from service industries. Three steps are involved in the model to estimate journeys to work and their associated travel costs and disutilities.

(a) The I/O model estimates the number of jobs created in the zones of work for the various SEGs (the split between SEGs depends on the industry/service type) and their associated labour costs, including commuting disutilities and costs.

(b) These jobs demand employees in the zones of residence, through a distribution model for employees.

(c) The movements of employees are converted into commuting trips, using an average trip rate per employee. These can be later be split by mode and assigned to the modal networks in order to calculate the commuting travel disutilities and the impact of commuters on network congestion.

In this manner the cost and disutility of transporting the employees to the place of work are estimated in each zone for each industry type. These feed into the overall costs of production in the zone.

The third major zonal element of costs for inclusion in the industrial location model is the location cost associated with the land and floorspace in which production is located. Although this location cost element was not included within the EUNET Trans-Pennine model, it has been included in a number of other spatial I/O models as described in general in Williams (1979), and specifically for the LASER model in Williams (1994). The location rent is estimated endogenously within the model by adjusting the rent in a zone so as to ensure that the competition by firms for location in the space available in the zone does not exceed the capacity of that space.

This section has explained how the full direct and indirect production cost elements in a zone can be estimated in a precise and transparent fashion. These zonal cost differentials will clearly play a role in the location decisions of firms. However, there may be other local or historical aspects that do not relate to these transport cost and rent related elements, which are as important, or often more important, in the actual location decisions of firms. Provided that good observed base year data are available these other residual influences can be included within an incremental model structure. Accordingly, the approach presented above provides an excellent framework to examine the influence of transport cost changes on location patterns. However, it does not provide a complete industrial location model, since there are aspects other than transport and location costs that may also be significant in future years in determining location patterns. Further research and development on these non-transport elements would be needed in order to create a complete and robust industrial location model.

The approach described above is equally applicable to modelling the location of DCs. Simply use the equations of Section III.5 in place of the equations of Section III.3. In the case of DCs, because they are intimately tied up with the transport sector, their locations are likely to be much more sensitive to transport cost variations than would be the case for many of the other types of industry. Accordingly, this modelling structure should be sufficient to produce reasonable estimates of the likely future location patterns of DCs. The key to success will lie in ensuring that the unit transport and warehousing costs are correctly estimated for each logistics shipment leg type.

APPENDIX IV: GLOSSARY

AES	Annual Employment Survey
CSRGT	Continuing Survey of Road Goods Traffic
DC	Distribution Centre – site of warehousing and distribution activities
DfT	UK Department for Transport
gvw	Gross Vehicle Weight – the maximum permitted weight of the vehicle *plus load*
GBFM	Great Britain Freight Model – created by MDS-Transmodal
GTF	Generalised Transport interchange Format for transfer of data between models
HGV	Heavy Goods Vehicles – road vehicles above 3.5 tonnes gvw
IDBR	Inter-Departmental Business Register – a detailed database of employment numbers
I/O Table	An Input/Output matrix measuring the pattern of inter-relationships of trade between economic sectors
IRHS	International Road Haulage Survey
ITEA	Integrated Transport Economic Appraisal Division of DfT
ITS	Institute for Transport Studies, the University of Leeds
LA	Local Authority
LATS	London Area Transport Surveys – carried out in 1971, 1981, 1991 and 2001
LGV	Light Goods Vehicles – light goods/commercial vehicles of up to 3.5 tonnes gvw
LGVS	Light Goods Vehicle Survey – to be carried out for DfT
LoLo	Load on / Load off – container operations at ports
LTS model	TfL's multimodal transport model for the London area
NAOMI model	The Highways Agency's detailed highway assignment model for London and the surrounding regions
NARNAS	A database of traffic counts and other road link characteristics, assembled by DfT
NFTM	The National Freight Model component of the overall GB National Transport Model (NTM)
NOMIS	An online database of employment and other data sources
NPTM	National Passenger Model component of the overall GB National Transport Model (NTM)

NST	Nomenclature Statistique de Transport – standard European commodity classification with codes 1 to 99
NTEM	GB National Trip End Model
NTM	GB National passenger and freight Transport Model components
NRTF	National Road Traffic Forecasts of the UK Department of Transport
NTS	National Travel Survey of passenger travel in GB
O/D matrix	A matrix of movements from an origin zone to a destination zone
P/C matrix	A matrix of freight movements from the original zone in which a good was produced to the final destination zone in which it is consumed, ignoring any intermediate modal transfers or warehousing and distribution locations
pcu	Passenger Car Unit – used to convert vehicles into equivalent cars when measuring the impact of large vehicles on the congestion on a road
RHTM	Regional Highway Traffic Model – A GB national model created in the 1970s
RoRo	Roll on / roll off – ferries carrying passenger or freight vehicles - may also refer to the RoRo Enquiry, a DfT census at international ports of road goods vehicles and of unaccompanied trailers passing abroad on RoRo ferries and the Channel Tunnel
RSI	Roadside Interview – at which survey data on the type, origin and destination of vehicle movements is collected
SACTRA	Standing Advisory Committee on Trunk Road Assessment
SCGE	Spatial Computable General Equilibrium models
SIC	Standard Industrial Classification of types of employment
SLAM	Spatial Logistics Appended Module, part of the SCENES EU model
SMILE	Strategic Model for Integrated Logistics Evaluation, Netherlands
SRA	Strategic Rail Authority
TEMPRO	A software package for the presentation of the presentation of the results from NTEM, as well as of projections of household, population and employment growth by zone
TEN	Transport Economics Note from DfT defining standard vehicle operating cost values and other economic characteristics
TfL	Transport for London