1

Introduction

1.1 Transport Planning and Modelling

1.1.1 Background

The world, including transport, is changing fast. We still encounter many of the same transport problems of the past: congestion, pollution, accidents, financial deficits and pockets of poor access. We are increasingly becoming money rich and time poor. However, we have learnt a good deal from long periods of weak transport planning, limited investment, emphasis on the short term and mistrust in strategic transport modelling and decision making. We have learnt, for example, that old problems do not fade away under the pressure of attempts to reduce them through better traffic management; old problems reappear in new guises with even greater vigour, pervading wider areas, and in their new forms they seem more complex and difficult to handle.

We now have greater confidence in technical solutions than in the previous century. This is not the earlier confidence in technology as the magic solution to economic and social problems; we have also learnt that this is a mirage. However, Information Technology has advanced enough to make possible new conceptions of transport infrastructure (e.g. road transport informatics), movement systems (e.g. automated driverless trains) and electronic payment (e.g. smartcards, video tolling). Mobile phones and GPS services are changing the way to deliver useful traveller information, facilitating payment and charging for the use of transport facilities. Of particular interest to the subject of this book is the advent of low-cost and high-speed computing; this has practically eliminated computing power as a bottleneck in transport modelling. The main limitations are now human and technical: contemporary transport planning requires skilled and experienced professionals plus, as we will argue below, theoretically sound modelling techniques with competent implementations in software.

Emerging countries are becoming more significant in the world stage but they suffer serious transport problems as well. These are no longer just the lack of roads to connect distant rural areas with markets. Indeed, the new transport problems bear some similarities with those prevalent in the post-industrialised world: congestion, pollution, and so on. However, they have a number of very distinctive features deserving a specific treatment: relatively low incomes, fast urbanisation and change, high demand for public transport, scarcity of resources including capital, sound data and skilled personnel.

The birth of the twenty-first century was dominated by two powerful trends affecting most aspects of life and economic progress. The stronger trend is *globalisation*, supported and encouraged by the other trend, cheap and high-capacity *telecommunications*. The combination of the two is changing the way we perceive and tackle many modern issues; their influence in transport planning is starting to be

Modelling Transport, Fourth Edition. Juan de Dios Ortúzar and Luis G. Willumsen.

felt. Some of these influences are the role of good transport infrastructure in enhancing the economic competitiveness of modern economies; a wider acceptance of the advantages of involving the private sector more closely in transport supply and operations; the possible role of telecommunications in reducing the need to travel.

Important technical developments in transport modelling have taken place since the mid-1970s, in particular at major research centres; these developments have been improved and implemented by a small group of resourceful consultants. However, many of these innovations and applications have received limited attention outside the more academic journals. After these years of experimentation there is now a better recognition of the role of modelling in supporting transport planning. This book attempts a review of the best of current practice in transport modelling; in most areas it covers the 'state of the art' but we have selected those aspects which have already been implemented successfully in practice. The book does not represent the leading edge of research into modelling. It tries, rather, to provide a survival tool-kit for those interested in improving transport modelling and planning, a kind of bridge or entry-point to the more theoretical papers that will form the basis of transport modelling in the future.

Transport modelling is not transport planning; it can only support planning, and in a few cases it may have the most important role in the process. We have known many good professionals who have developed sophisticated transport models but are frustrated because their work has apparently been ignored in many key planning decisions. In truth, planning and implementation have the power to change the world and transport modelling can only assist in this if adopted as an effective aid to decision making. This requires wise planners and, above all, better modellers.

1.1.2 Models and their Role

A *model* is a simplified representation of a part of the real world–the system of interest–which focuses on certain elements considered important from a particular point of view. Models are, therefore, problem and viewpoint specific. Such a broad definition allows us to incorporate both physical and abstract models. In the first category we find, for example, those used in architecture or in fluid mechanics which are basically aimed at design. In the latter, the range spans from the mental models all of us use in our daily interactions with the world, to formal and abstract (typically analytical) representations of some theory about the system of interest and how it works. Mental models play an important role in understanding and interpreting the real world and our analytical models. They are enhanced through discussions, training and, above all, experience. Mental models are, however, difficult to communicate and to discuss.

In this book we are concerned mainly with an important class of abstract models: mathematical models. These models attempt to replicate the system of interest and its behaviour by means of mathematical equations based on certain theoretical statements about it. Although they are still simplified representations, these models may be very complex and often require large amounts of data to be used. However, they are invaluable in offering a 'common ground' for discussing policy and examining the inevitable compromises required in practice with a level of objectivity. Another important advantage of mathematical models is that during their formulation, calibration and use the planner can learn much, through experimentation, about the behaviour and internal workings of the system under scrutiny. In this way, we also enrich our mental models thus permitting more intelligent management of the transport system.

A model is only realistic from a particular perspective or point of view. It may be reasonable to use a knife and fork on a table to model the position of cars before a collision but not to represent their mechanical features, or their route choice patterns. The same is true of analytical models: their value is limited to a range of problems under specific conditions. The appropriateness of a model is, as discussed in the rest of this chapter, dependent on the context where it will be used. The ability to choose and adapt models for particular contexts is one of the most important elements in the complete planner's tool-kit.

P1: TIX/XYZ P2: ABC
JWST054-01 JWST054-Ortuzar February 24, 2011 8:23 Printer Name: Yet to Come

This book is concerned with the contribution transport modelling can make to improved decision making and planning in the transport field. It is argued that the use of models is inevitable and that of formal models highly desirable. However, transport modelling is only one element in transport planning: administrative practices, an institutional framework, skilled professionals and good levels of communication with decision makers, the media and the public are some of the other requisites for an effective planning system. Moreover, transport modelling and decision making can be combined in different ways depending on local experience, traditions and expertise. However, before we discuss how to choose a modelling and planning approach it is worth outlining some of the main characteristics of transport systems and their associated problems. We will also discuss some very important modelling issues which will find application in other chapters of this book.

1.2 Characteristics of Transport Problems

Transport problems have become more widespread and severe than ever in both industrialised and developing countries alike. Fuel shortages are (temporarily) not a problem but the general increase in road traffic and transport demand has resulted in congestion, delays, accidents and environmental problems well beyond what has been considered acceptable so far. These problems have not been restricted to roads and car traffic alone. Economic growth seems to have generated levels of demand exceeding the capacity of most transport facilities. Long periods of under-investment in some modes and regions have resulted in fragile supply systems which seem to break down whenever something differs slightly from average conditions.

These problems are not likely to disappear in the near future. Sufficient time has passed with poor or no transportation planning to ensure that a major effort in improving most forms of transport, in urban and inter-urban contexts, is necessary. Given that resources are not unlimited, this effort will benefit from careful and considered decisions oriented towards maximising the advantages of new transport provision while minimising their money costs and undesirable side-effects.

1.2.1 Characteristics of Transport Demand

The demand for transport is *derived*, it is not an end in itself. With the possible exception of sight-seeing, people travel in order to satisfy a need (work, leisure, health) undertaking an *activity* at particular locations. This is equally significant for goods movements. In order to understand the demand for transport, we must understand the way in which these activities are distributed over space, in both urban and regional contexts. A good transport system widens the opportunities to satisfy these needs; a heavily congested or poorly connected system restricts options and *limits* economic and social development.

The demand for transport services is highly *qualitative* and *differentiated*. There is a whole range of specific demands for transport which are differentiated by time of day, day of week, journey purpose, type of cargo, importance of speed and frequency, and so on. A transport service without the attributes matching this differentiated demand may well be useless. This characteristic makes it more difficult to analyse and forecast the demand for transport services: tonne and passenger kilometres are extremely coarse units of performance hiding an immense range of requirements and services.

Transport demand takes place over *space*. This seems a trivial statement but it is the distribution of activities over space which makes for transport demand. There are a few transport problems that may be treated, albeit at a very aggregate level, without explicitly considering space. However, in the vast majority of cases, the explicit treatment of space is unavoidable and highly desirable. The most common approach to treat space is to divide study areas into zones and to code them, together with transport networks, in a form suitable for processing with the aid of computer programs. In some cases, study

P1: TIX/XYZ P2: ABC
JWST054-01 JWST054-Ortuzar February 24, 2011 8:23 Printer Name: Yet to Come

areas can be simplified assuming that the zones of interest form a corridor which can be collapsed into a linear form. However, different methods for treating distance and for allocating origins and destinations (and their attributes) over space are an essential element in transport analysis.

The spatiality of demand often leads to problems of lack of coordination which may strongly affect the equilibrium between transport supply and demand. For example, a taxi service may be demanded unsuccessfully in a part of a city while in other areas various taxis may be plying for passengers. On the other hand, the concentration of population and economic activity on well-defined corridors may lead to the economic justification of a high-quality mass transit system which would not be viable in a sparser area.

Finally, transport demand and supply have very strong *dynamic* elements. A good deal of the demand for transport is concentrated on a few hours of a day, in particular in urban areas where most of the congestion takes place during specific peak periods. This time-variable character of transport demand makes it more difficult–and interesting–to analyse and forecast. It may well be that a transport system could cope well with the *average* demand for travel in an area but that it breaks down during peak periods. A number of techniques exist to try to spread the peak and average the load on the system: flexible working hours, staggering working times, premium pricing, and so on. However, peak and off-peak variations in demand remain a central, and fascinating, problem in transport modelling and planning.

1.2.2 Characteristics of Transport Supply

The first distinctive characteristic of transport supply is that it is a *service* and not a good. Therefore, it is not possible to stock it, for example, to use it in times of higher demand. A transport service must be consumed when and where it is produced, otherwise its benefit is lost. For this reason it is very important to estimate demand with as much accuracy as possible in order to save resources by tailoring the supply of transport services to it.

Many of the characteristics of transport systems derive from their nature as a service. In very broad terms a transport system requires a number of fixed assets, the *infrastructure*, and a number of mobile units, the *vehicles*. It is the combination of these, together with a set of rules for their operation, that makes possible the movement of people and goods.

It is often the case that infrastructure and vehicles are not owned nor operated by the same group or company. This is certainly the case of most transport modes, with the notable exception of many rail systems. This separation between supplier of infrastructure and provider of the final transport service generates a rather complex set of interactions between government authorities (central or local), construction companies, developers, transport operators, travellers and shippers, and the general public. The latter plays several roles in the supply of transport services: it represents the residents affected by a new scheme, or the unemployed in an area seeking improved accessibility to foster economic growth; it may even be car owners wishing to travel unhindered through somebody else's residential area.

The provision of transport infrastructure is particularly important from a supply point of view. Transport infrastructure is 'lumpy', one cannot provide half a runway or one-third of a railway station. In certain cases, there may be scope for providing a gradual build-up of infrastructure to match growing demand. For example, one can start providing an unpaved road, upgrade it later to one or two lanes with surface treatment; at a later stage a well-constructed single and dual carriageway road can be built, to culminate perhaps with motorway standards. In this way, the provision of infrastructure can be adjusted to demand and avoid unnecessary early investment in expensive facilities. This is more difficult in other areas such as airports, metro lines, and so on.

Investments in transport infrastructure are not only lumpy but also take a long time to be carried out. These are usually large projects. The construction of a major facility may take from 5 to 15 years from

P1: TIX/XYZ P2: ABC
JWST054-01 JWST054-Ortuzar February 24, 2011 8:23 Printer Name: Yet to Come

planning to full implementation. This is even more critical in urban areas where a good deal of disruption is also required to build them. This disruption involves additional costs to users and non-users alike.

Moreover, transport investment has an important political role. For example, politicians in developing countries often consider a road project a safe bet: it shows they care and is difficult to prove wrong or uneconomic by the popular press. In industrialised nations, transport projects usually carry the risk of alienating large numbers of residents affected by them or travellers suffering from congestion and delay in overcrowded facilities. Political judgement is essential in choices of this kind but when not supported by planning, analysis and research, these decisions result in responses to major problems and crises only; in the case of transport this is, inevitably, too late. Forethought and planning are essential.

The separation of providers of infrastructure and suppliers of services introduces economic complexities too. For a start, it is not always clear that all travellers and shippers actually perceive the total costs incurred in providing the services they use. The charging for road space, for example, is seldom carried out directly and when it happens the price does not include congestion costs or other external effects, perhaps the nearest approximation to this being toll roads and modern road-pricing schemes. The use of taxes on vehicles and fuels is only a rough approximation to charging for the provision of infrastructure.

But, why should this matter? Is it not the case that other goods and services like public parks, libraries and the police are often provided without a direct charge for them? What is wrong with providing free road space? According to elementary economic theory it does matter. In a perfect market a good allocation of resources to satisfy human needs is only achieved when the marginal costs of the goods equal their marginal utility. This is why it is often advocated that the price of goods and services, i.e. their perceived cost, should be set at their marginal cost. Of course real markets are not perfect and ability to pay is not a good indication of need; however, this general framework provides the basis for contrasting other ways of arranging pricing systems and their impact on resource allocation.

Transport is a very important element in the welfare of nations and the well-being of urban and rural dwellers. If those who make use of transport facilities do not perceive the resource implications of their choices, they are likely to generate a balance between supply and demand that is inherently inefficient. Underpriced scarce resources will be squandered whilst other abundant but priced resources may not be used. The fact that overall some sectors of the economy (typically car owners) more than pay for the cost of the road space provided, is not a guarantee of more rational allocation of resources. Car owners probably see these annual taxes as fixed, *sunk*, costs which at most affect the decision of buying a car but not that of using it.

An additional element of distortion is provided by the number of concomitant- or *side-effects* associated with the production of transport services: accidents, pollution and environmental degradation in general. These effects are seldom *internalised*; the user of the transport service rarely perceives nor pays for the costs of cleaning the environment or looking after the injured in transport related accidents. Internalising these costs could also help to make better decisions and to improve the allocation of demand to alternative modes.

One of the most important features of transport supply is *congestion*. This is a term which is difficult to define as we all believe we know exactly what it means. However, most practitioners do know that what is considered congestion in Leeds or Lampang is often accepted as normal in London or Lagos. Congestion arises when demand levels approach the capacity of a facility and the time required to use it (travel through it) increases well above the average under low demand conditions. In the case of transport infrastructure the inclusion of an additional vehicle generates supplementary delay to all other users as well, see for example Figure 1.1. Note that the contribution an additional car makes to the delay of all users is greater at high flows than at low flow levels.

This is the external effect of congestion, perceived by others but not by the driver originating it. This is a cost which schemes such as electronic road pricing attempt to internalise to help more reasoned decision making by the individual.

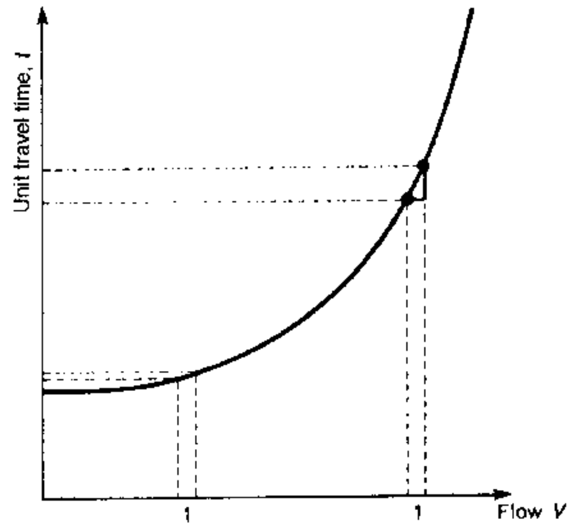

Figure 1.1 Congestion and its external effects

1.2.3 Equilibration of Supply and Demand

In general terms the role of transport planning is to ensure the satisfaction of a certain demand **D** for person and goods movements with different trip purposes, at different times of the day and the year, using various modes, given a transport system with a certain operating capacity. The transport system itself can be seen as made up of:

- an infrastructure (e.g. a road network);
- a management system (i.e. a set of rules, for example driving on the right, and control strategies, for example at traffic signals);
- a set of transport modes and their operators.

Consider a set of volumes on a network **V**, a corresponding set of speeds **S**, and an operating capacity **Q**, under a management system **M**. In very general terms the speed on the network can be represented by:

$$\mathbf{S} = f\{\mathbf{Q}, \mathbf{V}, \mathbf{M}\} \tag{1.1}$$

The speed can be taken as an initial proxy for a more general indicator of the *level of service* (LOS) provided by the transport system. In more general terms a LOS would be specified by a combination of speeds or travel times, waiting and walking times and price effects; we shall expand on these in subsequent chapters. The management system **M** may include traffic management schemes, area traffic control and regulations applying to each mode. The capacity **Q** would depend on the management system **M** and on the levels of investment **I** over the years, thus:

$$\mathbf{Q} = f\{\mathbf{I}, \mathbf{M}\} \tag{1.2}$$

The management system may also be used to redistribute capacity giving priority to certain types of users over others, either on efficiency (public-transport users, cyclists), environmental (electric vehicles) or equity grounds (pedestrians).

As in the case of most goods and services, one would expect the level of demand **D** to be dependent on the level of service provided by the transport system and also on the allocation of activities **A** over space:

$$\mathbf{D} = f\{\mathbf{S}, \mathbf{A}\} \tag{1.3}$$

Combining equations (1.1) and (1.3) for a fixed activity system one would find the set of equilibrium points between supply and demand for transport. But then again, the activity system itself would probably change as levels of service change over space and time. Therefore one would have two different sets of equilibrium points: short-term and long-term ones. The task of transport planning is to forecast and manage the evolution of these equilibrium points over time so that social welfare is maximised. This is, of course, not a simple task: modelling these equilibrium points should help to understand this evolution better and assist in the development and implementation of management strategies **M** and investment programmes **I**.

Sometimes very simple cause-effect relationships can be depicted graphically to help understand the nature of some transport problems. A typical example is the car/public-transport vicious circle depicted in Figure 1.2.

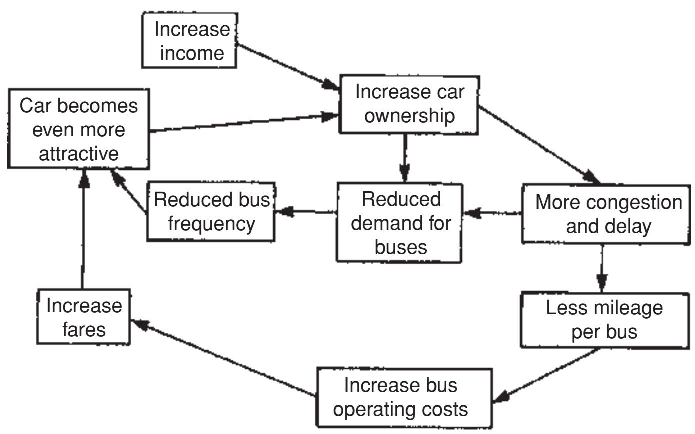

Figure 1.2 Car and public-transport vicious circle

Economic growth provides the first impetus to increase car ownership. More car owners means more people wanting to transfer from public transport to car; this in turn means fewer public-transport passengers, to which operators may respond by increasing the fares, reducing the frequency (level of service) or both. These measures make the use of the car even more attractive than before and induce more people to buy cars, thus accelerating the vicious circle. After a few cycles (years) car drivers are facing increased levels of congestion; buses are delayed, are becoming increasingly more expensive and running less frequently; the accumulation of sensible individual decisions results in a final state in which almost everybody is worse off than originally.

Moreover, there is a more insidious effect in the long term, not depicted in Figure 1.2, as car owners choose their place of work and residence without considering the availability (or otherwise) of public transport. This generates urban sprawl, low density developments that are more difficult and expensive to serve by more efficient public transport modes. This is the ‘development trap’ that leads to further congestion and a higher proportion of our time spent in slow moving cars.

P1: TIX/XYZ P2: ABC
JWST054-01 JWST054-Ortuzar February 24, 2011 8:23 Printer Name: Yet to Come

This simple representation can also help to identify what can be done to slow down or reverse this vicious circle. These ideas are summarised in Figure 1.3. Physical measures like bus lanes or other bus-priority schemes are particularly attractive as they also result in a more efficient allocation of road space. Public transport subsidies have strong advocates and detractors; they may reduce the need for fare increases, at least in the short term, but tend to generate large deficits and to protect poor management from the consequences of their own inefficiency. Car restraint, and in particular congestion charging, can help to internalise externalities and generate a revenue stream that can be distributed to other areas of need in transportation.

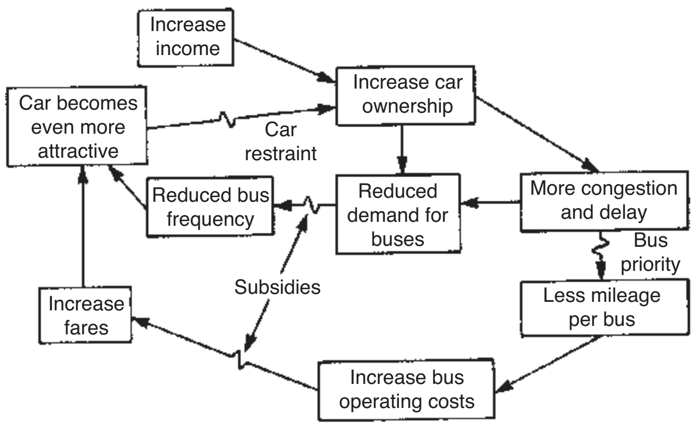

Figure 1.3 Breaking the car/public-transport vicious circle

The type of model behind Figures 1.2 and 1.3 is sometimes called a *structural model*, as discussed in Chapter 12; these are simple but powerful constructs, in particular because they permit the discussion of key issues in a fairly parsimonious form. However, they are not exempt from dangers when applied to different contexts. Think, for example, of the vicious circle model in the context of developing countries. Population growth will maintain demand for public transport much longer than in industrialised countries. Indeed, some of the bus flows currently experienced in emerging countries are extremely high, reaching 400 to 600 buses per hour one-way along some corridors. The context is also relevant when looking for solutions; it has been argued that one of the main objectives of introducing bus-priority schemes in emerging countries is not to protect buses from car-generated congestion but to organise bus movements (Gibson *et al*. 1989). High bus volumes often implement a *de facto* priority, and interference between buses may become a greater source of delay than car-generated congestion. To be of value, the vicious circle model must be revised in this new context.

It should be clear that it is not possible to characterise all transport problems in a unique, universal form. Transport problems are context dependent and so should be the ways of tackling them. Models can offer a contribution in terms of making the identification of problems and selection of ways of addressing them more solidly based.

1.3 Modelling and Decision Making

1.3.1 Decision-making Styles

Before choosing a modelling framework one needs to identify the general decision-making approach adopted in the country, government or decision unit. It must be recognised that there are several decision-making styles in practice and that not all of them use modelling as a basic building block. Previous editions of this text have characterised decision-making styles following the ideas of Nutt (1981); in practice, no decision-making style fits any of these categories exactly. This time, we would just like to distinguish two different paradigms: 'substantive rationality' and 'muddling through', following the lines of the very important book by Kay (2010).

The *substantive rationality* view of the world assumes that we know what our objectives are and we can envisage all alternative ways of achieving them and, with some luck, quantify the costs and benefits associated to each approach. This would apply to important decisions like choosing a place to live and less important ones like choosing a place to eat. This is the rational or normative decision-making approach implicit in most textbooks about transport planning. It is sometimes referred to as the 'systems approach' to planning. Here, quantification is essential. The decision problem is seen as one of choosing options from a complete set of alternatives and scenarios, with estimates on their probability of occurrence; the *utility* of each alternative is quantified in terms of benefits and costs and other criteria like environmental protection, safety, and so on.

In some cases it may even be possible to cast a decision problem into a mathematical programming framework. This means that the objective function is well understood and specified, and that the same applies to the constraints defining a solution space. However, for most real problems some elements of the objective function or constraints may be difficult to quantify or to convert into common units of measurement, say money or time. It may also be difficult to include some of the probabilistic elements in each case, but a good deal about the problem is learnt in the process. Modelling is at the core of this approach. The evaluation of plans or projects using Cost Benefit Analysis or a Multi-Criteria Framework is also based on this view of reality.

Some of the problems of applying normative decision theory are:

- Difficulties in actually specifying what the objectives are beyond generalities like reducing congestion or improving accessibility; as soon as we develop a measure or indicator for that objective, we find that it is actually misleading in respect of the things we want to achieve.
- The accusation of insensitivity to the aspirations of the public; people do not actually care about 'optimised' systems, they just want to see progress that is sustained along lines that are difficult to identify: they ask for speed but when it is delivered they are dissatisfied with the associated noise and emissions.
- Its high costs; substantive rationality is expensive to implement, requires advanced models and many runs for alternative arrangements and sensitivity analyses; efforts to apply this approach often overrun in time and budget; and
- The alienation of decision makers who may not understand, nor accept, the analytical treatment of the problem. This is a common complaint in our profession; the recurrent requisite to demonstrate the usefulness of our simulations may be irritating but reflects a real need to make our models and results relevant and communicable.

Moreover, there is very limited evidence that countries or organisations that do not follow this approach fare worse that those who do. Kay (2010) argues that many of the companies that were once hailed as paragons of good rational decision making failed spectacularly a few years later; there seem to be plenty of examples of this.

The main alternative approach to substantive rationality is what Lindblom (1959) called *muddling through*. The name, misleadingly self-deprecating, is not meant to imply that intuitive and unstructured decision making is desirable. On the contrary, in Lindblom's eyes, muddling through is a disciplined process but not one based on the substantive rational handling of defined objectives. The approach uses a combination of high-level (often unquantifiable) objectives, intermediate goals and immediate actions or experiments. Muddling through, or what Kay calls 'oblique or indirect approach', is characterised by:

- The use of high level objectives that are only loosely defined with no attempt to quantify them.
- Abandoning any clear distinction between objective, goals and actions; we learn about high-level objectives by adopting goals and implementing actions.
- Recognising that the environment is uncertain and that we cannot even know the range of events that might take place in the future, and
- Accepting that we can never identify, nor describe, all the range of options available; we can only deal with a limited set without aspiring to exhaust the search.

The following table, adapted from Kay's ideas, identifies additional contrasts between the two basic approaches:

Substantive rationality	**Issue**	**Indirect approach**
Interactions with others are limited and their response depend on our actions alone	Interactions	The outcome of interactions with others depend on context and their interpretation of our intentions
The relationships between objectives, states, goals and actions are understandable	Complexity	Our understanding of the relationships between objectives, states, goals and actions is imperfect but can be improved by experience
The problem and context can be described by a well specified and estimated analytical model	Abstraction	Appropriate simplification of complex problems must rely on judgement and understanding of context
What happens is what we intended to happen	Intentionality	What happens is the result of complex processes whose totality nobody fully understands
Decisions are made on the basis of the fullest possible information	Information	Decisions are recommended and made acknowledging that only limited knowledge is or can be available
The best outcome is achieved through a conscious process of maximisation	Adaptation	Good results are obtained through continual adaptation to constantly changing conditions
Rules and guidelines can be defined that allow people to find the correct solutions	Expertise	Experts can do things that others cannot – and can only learn with difficulty

In practice, no organisation relies on (attempts to) substantive rationality alone. Most apply an eclectic mixture of approaches using models, narratives, political context and sources of evidence. Modelling plays an important role in each of these approaches and the professional modeller should be ready to offer flexibility and capacity for adaptation, including new variables as required and responding quickly in the analysis of innovative policies and designs.

1.3.2 Choosing Modelling Approaches

This book assumes that the decision style adopted involves the use of models but it does not advocate a single (i.e. a normative) decision-making approach. The acceptability of modelling, or a particular modelling approach, within a decision style is very important. Models which end up being ignored by decision makers not only represent wasted resources and effort, but result in frustrated analysts and planners. It is further proposed that there are several features of transport problems and models which must be taken into account when specifying an analytical approach:

1. **Precision and accuracy required**. These concepts are sometimes confused. *Accuracy* is the degree to which a measurement or model result matches true or accepted values. Accuracy is an issue pertaining

P1: TIX/XYZ P2: ABC
JWST054-01 JWST054-Ortuzar February 24, 2011 8:23 Printer Name: Yet to Come

to the quality of data and model. The level of accuracy required for particular applications varies greatly. It is often the case that the accuracy required is just that necessary to discriminate between a good scheme and a less good one. In some cases the best scheme may be quite obvious, thus requiring less accurate modelling. Remember, however, that common sense has been blamed for some very poor transport decisions in the past.

Precision refers to the level or units of measurement used to collect data and deliver model outputs. One may measure travel times between two points in fractions of a second, but individuals may estimate and state the same much less precisely in five minute intervals. Precision is not accuracy and it is often misleading. Reporting estimates with high precision is often interpreted as confidence in their accuracy, whereas transport modellers often use precise numbers to report uncertain estimates. There is a difference between stating that 'traffic on link X was measured as 2347 vehicles between 8:00 and 9:00 AM yesterday' and saying that 'traffic on link X between 8:00 and 9:00 AM in five years time will be 3148 vehicles': the first statement may be both precise and accurate where the second is equally precise but certainly inaccurate. It is less misleading to report the second figure as 3150. As in the quote attributed to John Maynard Keynes 'it is much better to be roughly right than precisely wrong'.

2. **The decision-making context**. This involves the adoption of a particular *perspective* and a choice of a *scope* or coverage of the system of interest. The choice of perspective defines the type of decisions that will be considered: strategic issues or schemes, tactical (transport management) schemes, or even specific operational problems. The choice of scope involves specifying the level of analysis: is it just transport or does it involve activity location too? In terms of the transport system, are we interested in just demand or also on the supply side at different levels: system or suppliers' performance, cost minimisation issues within suppliers, and so on? The question of how many options need to be considered to satisfy different interest groups or to develop a single best scheme is also crucial. The decision-making context, therefore, will also help define requirements on the models to be used, the variables to be included in the model, or considered given or exogenous.
3. **Level of detail required**. The level of resolution of a model system can be described along four main dimensions: geography, unit of analysis, behavioural responses and the handling of time.

   Space is very important and it can be handled in an aggregate way, as a few zones with area-wide speed flow curves, or at the detailed level of the individual addresses for trips with links described in detail. There is a wide range of options in this field and the choice will depend on the application in hand: if the issue is a detailed design for traffic in a small area, highly disaggregated zones with an accurate account of the physical characteristics of links would be appropriate in a microsimulation model. Strategic planning may call for a more aggregate zoning system with links described in terms of their speed-flow relationships alone.

   The unit of interest for modelling may be the same zone with trips emanating and ending there or, at the other end of the spectrum, sampled or synthesised individuals; somewhere in between there will be different household or person strata as representative of the travelling population.

   The behavioural responses included may vary from fairly simple route choice actions in a traffic model to changes in time of travel, mode, destination, tour frequency and even land use and economic activity impacts.

   Time, in turn, can be treated either as a discrete or a continuous variable. In the first case the model may cover a full day (as in many national models), a peak period or a smaller time interval: all relevant responses will take place in that period although there may be interactions with other periods. Alternatively, time may be considered as a continuous variable which allows for more dynamic handling of traffic and behavioural responses like the choice of time of travel. Considering discrete time slices is a common option as treating time as a continuous variable is much more demanding.
4. **The availability of suitable data**, their stability and the difficulties involved in forecasting their future values. In some cases very little data may be available; in others, there may be reasons to

suspect the information, or to have less confidence in future forecasts for key planning variables as the system is not sufficiently stable. In many cases the data available will be the key factor in deciding the modelling approach.

5. **The state of the art in modelling** for a particular type of intervention in the transport system. This in turn can be subdivided into:
   - behavioural richness;
   - mathematical and computer tractability;
   - availability of good solution algorithms.

   It has to be borne in mind that in practice all models assume that some variables are exogenous to it. Moreover, many other variables are omitted from the modelling framework on the grounds of not being relevant to the task in hand, too difficult to forecast or expected to change little and not influence the system of interest. An explicit consideration of what has been left out of the model may help to decide on its appropriateness for a given problem.
6. **Resources available for the study**. These include money, data, computer hardware and software, technical skills, and so on. Two types of resource are, however, worth highlighting here: time and level of communication with decision makers and the public. *Time* is probably the most crucial one: if little time is available to make a choice between schemes, shortcuts will be needed to provide timely advice. Decision makers are prone to setting up absurdly short timescales for the assessment of projects which will take years to process through multiple decision instances, years to implement and many more years to be confirmed as right or wrong. On the other hand, a good level of communication with decision makers and the public will alleviate this problem: fewer unrealistic expectations about our ability to accurately model transport schemes will arise, and a better understanding of the advantages and limitations of modelling will moderate the extremes of blind acceptance or total rejection of study recommendations.
7. **Data processing requirements**. This aspect used to be interpreted as something like ‘how big a computer do you need?’ The answer to that question today is ‘not very big’, as a good microcomputer will do the trick in most cases. The real bottleneck in data processing is the human ability to collect, code, input the data, run the programs and interpret the output. The greater the level of detail, the more difficult all these human tasks will be. The use of computer-assisted data collection and graphics for input–output of programs reduces the burden somewhat.
8. **Levels of training and skills of the analysts**. Training costs are usually quite high; so much so that it is sometimes better to use an existing model or software that is well understood, than to embark on acquiring and learning to use a slightly more advanced one. This looks, of course, like a recipe for stifling innovation and progress; however, it should always be possible to spend some time building up strengths in new advanced techniques without rejecting the experience gained with earlier models.
9. **Modelling perspective and scope**. Florian *et al*. (1988) formalise decision-making contexts using a two-dimensional framework: the *level of analysis* and the *perspective*. The levels of analysis may include six different groups of *procedures*, where a procedure centres on one or more models and their specific solution algorithms. These are:
   - *activity location* procedures **L**;
   - *demand procedures* **D**;
   - *transport system performance* procedures **P**, which produce as output levels of service, expenditure and practical capacities, and depend on demand levels and on transport supply conditions;
   - *supply actions* procedures **S**, which determine the actions taken by suppliers of transport services and infrastructure; these depend on their objectives (profit maximisation, social welfare), institutional environment, their costs and estimates of future states of the system;
   - *cost minimisation* procedures **CM**;
   - *production* procedures **PR**.

The last two have more to do with the microeconomic issues affecting the suppliers in their choice of input combinations to minimise costs.

The perspectives dimension considers the six level procedures **L, D, P, S, CM, PR** and three perspectives: a *strategic* perspective **STR**, a *tactical* perspective **TAC** and an *operational* perspective **OPE**. These are, of course, related to the planning horizons and the levels of investment; however, in this context they must be seen as generic concepts dealing with the capacity:

- to visualise the levels **L, D, P, S, CM, PR** in their true and relative importance;
- to choose, at any level, what is to be regarded as fixed and what as variable.

Figure 1.4 summarises the way in which different perspectives and levels usually interact. The largest and most aggregate is, of course, the strategic level; analysis and choice at this level have major system-wide and long-term impacts, and usually involve resource acquisition and network design. Tactical issues have a narrower perspective and concern questions like making the best use of existing facilities and infrastructure. The narrowest perspective, the operational one, is concerned with the short-term problems of suppliers of transport services which fall outside the scope of this book; nevertheless, the actual decisions on, for example, levels of service or vehicle size, are important exogenous input to some of the models discussed in this book, and this is depicted in Figure 1.4.

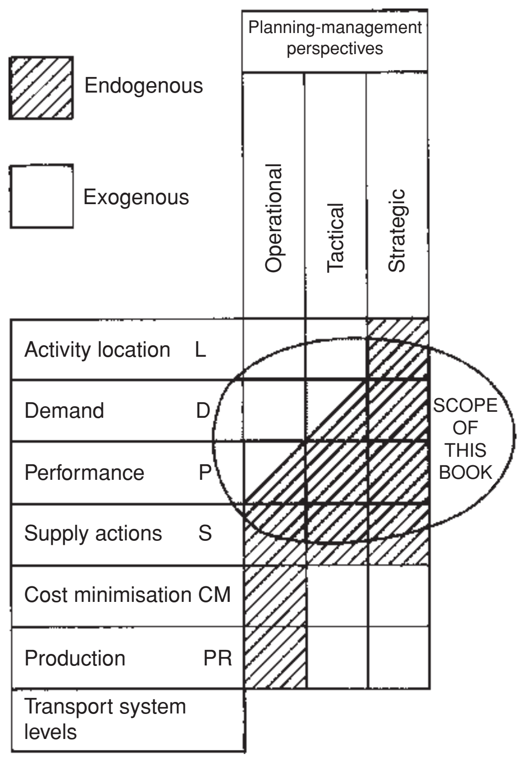

Figure 1.4 The two-dimensional conceptual framework

This is, of course, a rather abstract and idealised way of visualising planning problems. However, it helps to clarify the choices the analyst must face in developing a transport modelling approach. In this book we are mainly concerned with strategic and tactical issues at the demand and performance procedure levels. Nevertheless, some of the models discussed here sometimes find application outside these levels and perspectives.

1.4 Issues in Transport Modelling

We have already identified the interactions between transport problems, decision-making styles and modelling approaches. We need to discuss now some of the critical modelling issues which are relevant to the choice of model. These issues cover some general points like the roles of theory and data, model specification and calibration. But perhaps the most critical choices are those between the uses of aggregate or disaggregate approaches, cross-section or time-series models, and revealed or stated preference techniques.

1.4.1 General Modelling Issues

Wilson (1974) provides an interesting list of questions to be answered by any would-be modeller; they range from broad issues such as the *purpose* behind the model-building exercise, to detailed aspects such as *what techniques* are available for building the model. We will discuss some of these below, together with other modelling issues which are particularly relevant to the development of this book.

1.4.1.1 The Roles of Theory and Data

Many people tend to associate the word ‘theory’ with endless series of formulae and algebraic manipulations. In the urban transport modelling field this association has been largely correct: it is difficult to understand and replicate the complex interactions between human beings which are an inevitable feature of transport systems.

Some theoretical developments attempting to overcome these difficulties have resulted in models lacking adequate data and/or computational software for their practical implementation. This has led to the view, held strongly by some practitioners, that the gap between theory and practice is continually widening; this is something we have tried to redress in this book.

An important consideration on judging the contribution of a new theory is whether it places any meaningful restrictions on, for example, the form of a demand function. There is at least one documented case of a ‘practical’ transport planning study, lasting several years and costing several million dollars, which relied on ‘pragmatic’ demand models with a faulty structure (i.e. some of its elasticities had a wrong sign; see Williams and Senior 1977). Although this could have been diagnosed *ex ante* by the pragmatic practitioners, had they not despised theory, it was only discovered *post hoc* by theoreticians.

Unfortunately (or perhaps fortunately, a pragmatist would say), it is sometimes possible to derive similar functional forms from different theoretical perspectives (this, the *equifinality issue*, is considered in more detail in Chapter 8). The interpretation of the model output, however, is heavily dependent on the theoretical framework adopted. For example, the same functional form of the gravity model can be derived from analogy with physics, from entropy maximisation and from maximum utility formalisms. The interpretation of the output, however, may depend on the theory adopted. If one is just interested in flows on links it may not matter which theoretical framework underpins the analytical model function. However, if an evaluation measure is required, the situation changes, as only an economically based theory of human behaviour will be helpful in this task. In other cases, phrases like: ‘the gravitational pull of this destination will increase’, or ‘this is the most probable arrangement of trips’ or ‘the most likely trip matrix consistent with our information about the system’ will be used; these provide no help in devising evaluation measures but assist in the interpretation of the nature of the solution found. The theoretical framework will also lend some credence to the ability of the model to forecast future behaviour. In this sense it is interesting to reflect on the influence practice and theory may have on each other. For example, it has been noted that models or analytical forms used in practice have had traditionally a guiding influence on the assumptions employed in the development of subsequent theoretical frameworks. It is also well known that widely implemented forms, like the gravity-logit model we will discuss in Chapters 6 and 7, have been the subject of strong *post hoc* rationalisation:

> theoretical advances are especially welcome when they fortify existing practice which might be deemed to lack a particularly convincing rationale (Williams and Ortúzar, 1982b).

The two classical approaches to the development of theory are known as *deductive* (building a model and testing its predictions against observations) and *inductive* (starting with data and attempting to infer general laws). The deductive approach has been found more productive in the pure sciences and the inductive approach has been preferred in the analytical social sciences. It is interesting to note that data are central to both; in fact, it is well known that data availability usually leaves little room for negotiation and compromise in the trade-off between modelling *relevance* and modelling *complexity*. Indeed, in very many cases the nature of the data restricts the choice of model to a single option.

The question of data is closely connected with issues such as the type of variables to be represented in the model and this is, of course, closely linked again to questions about theory. Models predict a number of dependent (or endogenous) variables given other independent (or explanatory) variables. To test a model we would normally need data about each variable. Of particular interest are the *policy variables*, which are those assumed to be under the control of the decision maker, e.g. those the analyst may vary in order to test the value of alternative policies or schemes.

Another important issue in this context is that of aggregation:

- How many population strata or types of people do we need to achieve a good representation and understanding of a problem?
- In how much detail do we need to measure certain variables to replicate a given phenomenon?
- Space is crucial in transport; at what level of detail do we need to code the origin and destination of travellers to model their trip making behaviour?

1.4.1.2 Model Specification

In its widest and more interesting sense this issue considers the following themes.

Model Structure Is it possible to replicate the system to be modelled with a simple structure which assumes, for example, that all alternatives are independent? Or is it necessary to build more complex models which proceed, for example, to calculate probabilities of choice conditional on previous selections? Disaggregate models, such as those discussed in Chapters 7 to 9, usually have parameters which represent aspects of model structure and the extensions to methodology achieved by the mid-1980s have allowed the estimation of more and more general model forms. However, as Daly (1982b) has remarked, although it might be supposed that ultimately all issues concerned with model form could be resolved by empirical testing, such resolution is neither possible nor appropriate.

Functional Form Is it possible to use linear forms or does the problem require postulating more complex non-linear functions? The latter may represent the system of interest more accurately, but certainly will be more demanding in terms of resources and techniques for model calibration and use. Although theoretical considerations may play a big role in settling this question, it is also possible to examine it in an inductive fashion by means of ‘laboratory simulations’, for example in stated intentions/preferences experiments.

P1: TIX/XYZ P2: ABC
JWST054-01 JWST054-Ortuzar February 24, 2011 8:23 Printer Name: Yet to Come

Variable Specification This is the more usual meaning attached to the specification issue; which variables to use and how (which form) they should enter a given model. For example, if income is assumed to influence individual choice, should the variable enter the model as such or deflating a cost variable? Methods to advance on this question range from the deductive ('constructive') use of theory, to the inductive statistical analysis of the data using transformations.

1.4.1.3 Model Calibration, Validation and Use

A model can be simply represented as a mathematical function of variables X and parameters θ, such as:

$$Y = f(\mathbf{X}, \boldsymbol{\theta}) \tag{1.4}$$

It is interesting to mention that the twin concepts of *model calibration* and *model estimation* have taken traditionally a different meaning in the transport field. Calibrating a model requires choosing its parameters, assumed to have a non-null value, in order to optimise one or more *goodness-of-fit* measures which are a function of the observed data. This procedure has been associated with the physicists and engineers responsible for early aggregate transport models who did not worry unduly about the statistical properties of these indices, e.g. how large any calibration errors could be.

Estimation involves finding the values of the parameters which make the observed data more likely under the model specification; in this case one or more parameters can be judged *non-significant* and left out of the model. Estimation also considers the possibility of examining empirically certain specification issues; for example, structural and/or functional form parameters may be estimated. This procedure has tended to be associated with the engineers and econometricians responsible for disaggregate models, who placed much importance on the statistical testing possibilities offered by their methods. However, in essence both procedures are the same because the way to decide which parameter values are better is by examining certain previously defined goodness-of-fit measures. The difference is that these measures generally have well-known statistical properties which in turn allow confidence limits to be built around the estimated values and model predictions.

Because the large majority of transport models have been built on the basis of *cross-sectional* data, there has been a tendency to interpret model *validation* exclusively in terms of the goodness-of-fit achieved between observed behaviour and base-year predictions. Although this is a *necessary*, it is by no means a *sufficient* condition for model validation; this has been demonstrated by a number of cases which have been able to compare model predictions with observed results in *before-and-after* studies (see the discussion in Williams and Ortúzar, 1982a). Validation requires comparing the model predictions with information *not used* during the process of model estimation. This obviously puts a more stringent test on the model and requires further information or more resources.

One of the first tasks a modeller faces is to decide which variables are going to be predicted by the model and which are possibly required as inputs to it. Some will not be included at all, either because the modeller lacks control over them or simply because the theory behind the model ignores them (see Figure 1.5). This implies immediately a certain degree of error and uncertainty (we will come back to this problem in Chapter 3) which of course gets compounded by other errors which are also inherent to modelling; for example, sampling errors and, more important, errors due to the unavoidable simplifications of reality the model demands in order to be practical (see Figure 1.5).

Thus, the main use of models in practice is for *conditional forecasting*: the model will produce estimates of the dependent variables given a set of independent variables. In fact, typical forecasts are conditional in two ways (Wilson 1974):

- in relation to the values assigned to the policy variables in the plan, the impact of which is being tested with the model;
- in relation to the assumed values of other variables.

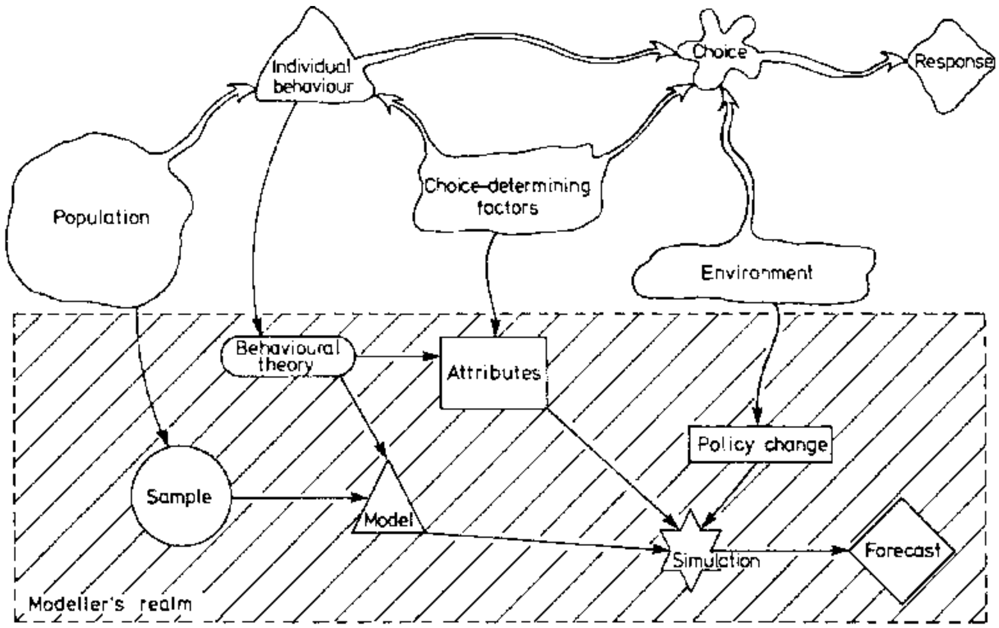

Figure 1.5 Modelling and sampling

A model is normally used to test a range of alternative plans for a range of possible assumptions about the future value of the other variables (e.g. low- and high-income scenarios). This means that it might be 'run' many times in the context of examining a particular problem. For this reason it may be of crucial importance that its specification allows for quick turn-around time in a computer; this is not an easy task in the case of a full-scale transportation model which involves complex processes of equilibration between supply and demand, as we will discuss in Chapter 11.

1.4.1.4 Modelling, Forecasting and Judgement

There is a subtle difference between modelling and forecasting. Modelling focuses on building and applying appropriate tools that are sensitive to the choices of interest and respond logically to changes in key policy instruments. The successful modeller will provide useful and timely advice to the decision-making process, even if the data and timescales are limited. In this case, it is important that the model produces consistent results for all expected interventions, policies and projects, such that they can be ranked fairly, even if the correspondence to reality is not perfect.

Forecasting is an attempt to envision and quantify future conditions. It normally involves estimating future travel demand and the resulting multimodal flows and costs over time. In the case of private sector projects, see Chapter 16, these projections are usually accompanied by revenue forecasts and investors will take considered risks based on these forecasts. Forecasting is usually based on formal models, but they alone cannot provide the full picture; it is necessary to incorporate other analyses and assumptions. Given the uncertainty about the future, several complementary approaches might be used in forecasting. For example a formal model may be supported by consideration about the main economic drivers of future travel activity in a region; in that way it is made clear how forecasts are dependent on the

P1: TIX/XYZ P2: ABC
JWST054-01 JWST054-Ortuzar February 24, 2011 8:23 Printer Name: Yet to Come

future of these activities. The success of forecasts can only be objectively measured through before and after studies.

The importance of formal models increases as the interventions under consideration diverge further from what is on the ground and known today. For example, when introducing a mode not currently available in a city, the model will often have to rely on stated preference data, information from other regions, or rational decision making theory. The same is true when evaluating any sort of policy not currently in existence (congestion charging) or when considering fuel prices or congestion conditions radically different than at present. In general, good advice on these issues cannot be given only on the basis of good modelling, however excellent. This requires intelligent consideration of other factors and assumptions, in particular about the limitations of any modelling approach.

Given the nature of analytical models, interpretation of their output is essential. Interpretation requires good judgement and this is only acquired with experience and a thorough understanding of the theories underpinning models and their limitations. For instance, most of the models described in this text are supported by random utility theory (see Chapter 7) that in turn assumes rational decision making on the part of travellers. However, there is an increasingly solid body of evidence, provided mostly by Behavioural Economics and Psychology, that humans are neither entirely rational nor consistent in their choices. This evidence (see Ariely 2009) punctures the theory underpinning our models–even the most advanced activity based approaches–and makes the application of judgement in the interpretation of model outputs even more important.

1.4.2 Aggregate and Disaggregate Modelling

The level of aggregation selected for the measurement of data is an important issue in the general design of a transportation planning study. Of central interest is the aggregation of exogenous data, that is, information about items other than the behaviour of travellers which is assumed endogenous (i.e. the model attempts to replicate it). For example, throughout the years it has been a cause for concern whether a given data item represents an average over a group of travellers rather than being collected specifically for a single individual. When the model at base aims at representing the behaviour of more than one individual (e.g. a population segment like car owners living in a zone), such as in the case of the *aggregate* models we will examine in Chapters 5 and 6, a certain degree of aggregation of the exogenous data is inevitable. But when the model at base attempts to represent the behaviour of individuals, such as in the case of the *disaggregate* models we will study in Chapters 7 to 9, it is conceivable that exogenous information can be obtained and used separately for each traveller. An important issue is then whether, as is often the case, it might be preferable on cost or other grounds to use less detailed data (see Daly and Ortúzar 1990).

Forecasting future demand is a crucial element of the majority of transport planning studies. Being able to predict the likely usage of new facilities is an essential precursor to rational decision making about the advantages or otherwise of providing such facilities. It may also be important to have an idea about the sensitivities of demand to important variables under the control of the analyst (e.g. the price charged for its use). In most cases the forecasts and sensitivity estimates must be provided at the aggregate level, that is, they must represent the behaviour of an entire population of interest. Therefore, the analyst using disaggregate models must find a sound method for aggregating model results to provide these indicators.

Aggregate models were used almost without exception in transportation studies up to the late 1970s; they became familiar, demanded relatively few skills on the part of the analyst (but required arcane computer knowledge) and had the property of offering a ‘recipe’ for the complete modelling process, from data collection through the provision of forecasts at the level of links in a network. The output of these models, perhaps because they were generated by obscure computer programs, were often considered more accurate than intended, for example predicting turning movement flows 15 years in the future. Aggregate models have been severely (and sometimes justifiably) criticised for their inflexibility, inaccuracy and cost. Unfortunately, many disaggregate approaches which have adopted sophisticated treatments of the choices and constraints faced by individual travellers have failed to take the process through to the production of forecasts, sometimes because they require data which cannot reasonably be forecast.

Disaggregate models, which became increasingly popular during the 1980s, offer substantial advantages over the traditional methods while remaining practical in many application studies. However, one important problem in practice is that they demand from the analyst quite a high level of statistical and econometric skills for their use (in particular for the interpretation of results), certainly much higher than in the case of aggregate models. Moreover, the differences between aggregate and disaggregate model systems have often been overstated. For example, the disaggregate models were first marketed as a radical departure from classical methods, a 'revolution' in the field, while eventually it became clear that an 'evolutionary' view was more adequate (see Williams and Ortúzar 1982b). In fact, in many cases there is complete equivalence between the forms of the forecasting models (Daly 1982a). The essential difference lies in the treatment of the description of behaviour, particularly during the model development process; in many instances the disaggregate approach is clearly superior to the grouping of behaviour by zones and by predefined segments.

Attempts to clarify the issue of whether disaggregate or aggregate approaches were to be preferred, and in what circumstances, have basically concluded that there is no such thing as a definitive approach appropriate to all situations (see Daly and Ortúzar 1990). These attempts have also established the need for guidelines to help the despairing practitioner to choose the most appropriate model tools to apply in a particular context. We have striven to answer that call in this book.

1.4.3 Cross-section and Time Series

The vast majority of transport planning studies up to the late 1980s relied on information about trip patterns revealed by a cross-section of individuals at a single point in time. Indeed, the traditional use of the cross-sectional approach transcended the differences between aggregate and disaggregate models.

A fundamental assumption of the cross-sectional approach is that a measure of the response to incremental change may simply be found by computing the derivatives of a demand function with respect to the policy variables in question. This makes explicit the assumption that a realistic *stimulus-response* relation may be derived from model parameters estimated from observations at one point in time. This would be reasonable if there were always enough people changing their choices, say of mode or destination, in *both* directions and without habit or time-lag effects.

However, the cross-sectional assumption has two potentially serious drawbacks. First, a given cross-sectional data set may correspond to a particular 'history' of changes in the values of certain key variables influencing choice. For example, changes in mode or location in time may have been triggered by a series of different stimuli (petrol prices, life-cycle effects, etc.) and the extent to which a system is considered to be in *disequilibrium* (because of, say, inertia) will depend on these. The trouble is that it can be shown (see Chapter 7) that the response of groups with exactly the same current characteristics, but having undergone a different path of changes, may be very different indeed. Second, data collected at only one point in time will usually fail to discriminate between alternative model formulations, even between some arising from totally different theoretical postulates. It is always possible to find 'best-fit' parameters from base-year data even if the model suffers severe mis-specification problems; the trouble is, of course, that these do not guarantee good response properties for a future situation. As we saw in section 1.4.1, a good base-year fit is not a sufficient condition for model validation.

Thus, in general it is not possible to discriminate between the large variety of possible sources of dispersion within a cross-sectional data set (i.e. preference dispersion, habit effects, constraints, and so on). Real progress in understanding and assessing the effectiveness of forecasting models, however, can

P1: TIX/XYZ P2: ABC
JWST054-01 JWST054-Ortuzar February 24, 2011 8:23 Printer Name: Yet to Come

only be made if information is available on response over time. From a theoretical point of view, it is also desirable that appropriate frameworks for analysis are designed which allow the eventual refutation of hypotheses relating to response. Until this is achieved, a general problem of potential misrepresentation will continue to cast doubts on the validity of cross-sectional studies.

The discussion above has led many people to believe that, where possible, longitudinal or time-series data should be used to construct more dependable forecasting models. This type of data incorporates information on response by design. Thus, in principle, it may offer the means to directly test and even perhaps reject hypotheses relating to response.

Longitudinal data can take the form of *panels* or more simply *before-and-after* information. Unfortunately, models built on this type of data have severe technical problems of their own; in fact, up to the end of the 1990s progress in this area had been limited. We will discuss some of the issues involved in the collection and use of this type of information in Chapters 3 and 7.

1.4.4 Revealed and Stated Preferences

The development of good and robust models is quite difficult if the analyst cannot set up experiments to observe the behaviour of the system under a wide range of conditions. Experimentation of this kind is neither practical nor viable in transport and the analyst is restricted, like an astronomer, to make observations on events and choices they do not control. Up to the mid-1980s it was almost axiomatic that modelling transport demand should be based on information about observed choices and decisions, i.e. *revealed-preference* data. Within this approach, project evaluation requires expressing policies in terms of changes in attributes which ‘map onto’ those considered to influence current behaviour. However, this has practical limitations basically associated with survey costs and the difficulty of distinguishing the effects of attributes which are not easy to observe, e.g. those related to notions such as quality or convenience. Another practical embarrassment has been traditionally the ‘new option’ problem, whereby it is required to forecast the likely usage of a facility not available at present and perhaps even radically different to all existing ones.

Stated-preference/intentions techniques, borrowed from the field of market research, were put forward by the end of the 1970s as offering a way of experimenting with transport-related choices, thus solving some of the problems outlined above. Stated-preference techniques base demand estimates on an analysis of the response to *hypothetical choices*; these, of course, can cover a wider range of attributes and conditions than the real system. However, these techniques were severely discredited at the start because it was not known how to discount for the over enthusiasm of certain respondents, e.g. not even half of the individuals stating they would take a given course of action actually did so when the opportunity eventually arose.

It took a whole decade for the situation to change, but by the end of the 1980s stated-preference methods were perceived by many to offer a real chance to solve the above-mentioned difficulties. Moreover, it has been found that, in appropriate cases, revealed-and stated-preference data and methods may be employed in complementary senses with the strengths of both approaches recognised and combined. In particular, they are considered to offer an invaluable tool for assisting the modelling of completely new alternatives. We will examine data-collection aspects of stated-preference methods in Chapter 3 and modelling issues in Chapter 8.

1.5 The Structure of the Classic Transport Model

Years of experimentation and development have resulted in a general structure which has been called the classic transport model. This structure is, in effect, a result from practice in the 1960s but has remained more or less unaltered despite major improvements in modelling techniques since then.

The general form of the model is depicted in Figure 1.6. The approach starts by considering a zoning and network system, and the collection and coding of planning, calibration and validation data. These data would include base-year levels for population of different types in each zone of the study area as well as levels of economic activity including employment, shopping space, educational and recreational facilities. These data are then used to estimate a model of the total number of trips generated and attracted by each zone of the study area (*trip generation*). The next step is the allocation of these trips to particular destinations, in other words their *distribution* over space, thus producing a trip matrix. The following stage normally involves modelling the choice of mode and this results in *modal split*, i.e. the allocation of trips in the matrix to different modes. Finally, the last stage in the classic model requires the *assignment* of the trips by each mode to their corresponding networks: typically private and public transport.

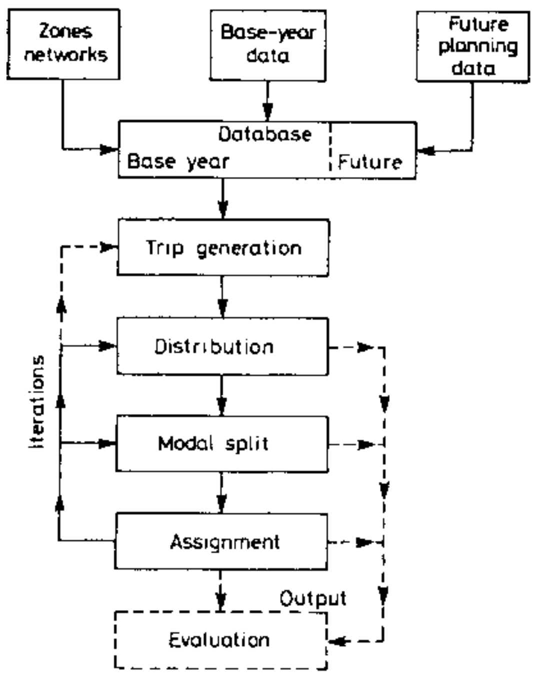

Figure 1.6 The classic four-stage transport model

The classic model is presented as a sequence of four sub-models: trip generation, distribution, modal split and assignment. It is generally recognised that travel decisions are not actually taken in this type of sequence; a contemporary view is that the 'location' of each sub-model depends on the form of the utility function assumed to govern all these travel choices (see Williams 1977). Moreover, the four-stage model is seen as concentrating attention on only a limited range of travellers' responses. Current thinking requires an analysis of a wider range of responses to transport problems and schemes. For example, when faced with increased congestion a trip maker can respond with a range of simple changes to:

- the **route** followed to avoid congestion or take advantage of new links; this includes choice of parking place or combination of services in the case of public transport;
- the **mode** used to get to the destination;
- the **time** of departure to avoid the most congested part of the peak;

P1: TIX/XYZ P2: ABC
JWST054-01 JWST054-Ortuzar February 24, 2011 8:23 Printer Name: Yet to Come

- the **destination** of the trip to a less congested area;
- the **frequency** of journeys by undertaking the trip at another day, perhaps combining it with other activities.

Furthermore, other more complex responses take place in the longer term, for example changes in jobs, residential location, choice of shopping areas and so on; all of these will respond, at least partially, to changes in the accessibility provided by the transport system.

Despite these comments, the four-stage sequential model provides a point of reference to contrast alternative methods. For example, some contemporary approaches attempt to treat simultaneously the choices of trip frequency (trips per week), destination and mode of travel thus collapsing trip generation, distribution and mode choice in one single model. Other approaches emphasise the role of household activities and the travel choices they entail; concepts like sojourns, circuits, and time and money budgets are used in this context to model travel decisions and constraints. These modelling strategies are more difficult to cast in terms of the four main decisions or sub-models above. However, the improved understanding of travel behaviour these activity based models provide is likely to enhance more conventional modelling approaches, see Chapter 14.

The trip generation–distribution–modal split and assignment sequence is the most common but not the only possible one. Some past studies have put modal split before trip distribution and immediately after (or with) trip generation. This permits a greater emphasis on decision variables depending on the trip generation unit, the individual or the household. However, forcing modal split before the destination is known requires “averaging” the attributes of the journey and modes in the model. This detracts policy relevance from the modal-split model. Another approach is to perform distribution and mode choice simultaneously, as discussed in Chapter 6. Note also that the classic model makes trip generation inelastic, that is, independent of the level of service provided in the transport system. This is probably unrealistic but only recently techniques have been developed which can take systematic account of these effects.

Once the model has been calibrated and validated for base-year conditions it must be applied to one or more planning horizons. In order to do this it is necessary to develop *scenarios* and plans describing the relevant characteristics of the transport system and planning variables under alternative futures. The preparation of realistic and consistent scenarios is not a simple task as it is very easy to fall into the trap of constructing futures which are neither financially viable nor realistic in the context of the likely evolution of land use and activities in the study area. Despite these difficulties, scenario writing is still more of an art than a technique and requires a good deal of engineering expertise combined with sound political judgement; unfortunately these are scarce resources seldom found together in planning teams.

Having prepared realistic scenarios and plans for testing, the same sequence of models is run again to simulate their performance. A comparison is then made between the costs and benefits, however measured, of different schemes under different scenarios; the idea is to choose the most attractive programme of investment and transport policies which satisfies the demand for movement in the study area.

An important issue in the classic four-stage model is the consistent use of variables affecting demand. For example, at the end of the traffic assignment stage new flow levels, and therefore new travel times, will be obtained. These are unlikely to be the same travel times assumed when the distribution and mode choice models were run, at least when the models are used in the forecasting mode. This seems to call for the re-run of the distribution and modal-split models based now on the new travel times. The subsequent application of the assignment model may well result in a new set of travel times; it will be seen that in general the naive feed-back of the model does not lead to a stable set of distribution, modal split and assignment models with consistent travel times. This problem will be treated in some detail in Chapter 11; its particular relevance is in the risk of choosing the wrong plan depending on how many cycles one is prepared to undertake.

1.6 Continuous Transport Planning

Transport planning models on their own do not solve transport problems. To be useful they must be utilised within a decision process adapted to the chosen decision-making style. The classic transport model was originally developed for an idealised normative decision-making approach. Its role in transport planning can be presented as contributing to the key steps in a 'rational' decision-making framework as in Figure 1.7:

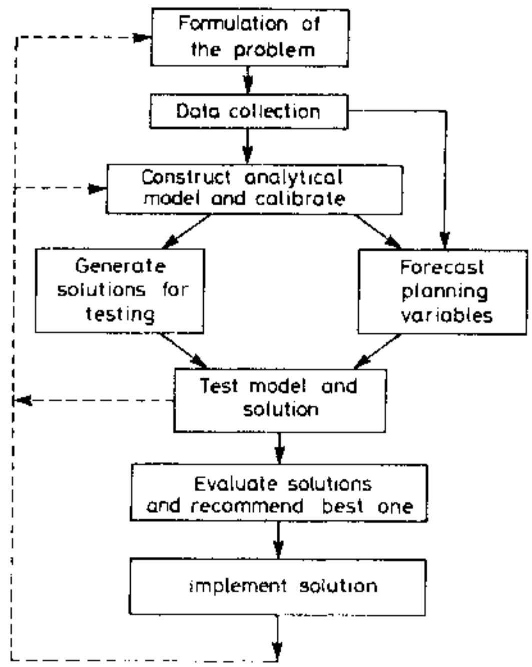

Figure 1.7 A framework for rational decision making with models

1. **Formulation of the problem**. A problem can be defined as a mismatch between expectations and perceived reality. The formal definition of a transport problem requires reference to objectives, standards and constraints. The first reflect the values implicit in the decision-making process, a definition of an ideal but achievable future state. Standards are provided in order to compare, at any one time, whether minimum performance is being achieved at different levels of interest. For example, the fact that many signalised junctions in a city operate at more than 90% degree of saturation can be taken to indicate an overloaded network. Constraints can be of many types, financial, temporal, geographical, technical or simply certain areas or types of building that should not be threatened by new proposals.
2. **Collection of data** about the present state of the system of interest in order to support the development of the analytical model. Of course, data collection is not independent from model development, as the latter defines which types of data are needed: data collection and model development are closely interrelated.
3. **Construction of an analytical model** of the system of interest. The tool-set provided in this book can be used to build transport models including demand and system performance procedures from a

P1: TIX/XYZ P2: ABC
JWST054-01 JWST054-Ortuzar February 24, 2011 8:23 Printer Name: Yet to Come

tactical and strategic perspective. In general, one would select the simplest modelling approach which makes possible a choice between schemes on a sound basis. The construction of an analytical model involves specifying it, estimating or calibrating its parameters and validating its performance with data not used during calibration.

4. **Generation of solutions** for testing. This can be achieved in a number of ways, from tapping the experience and creativity of local transport planners and interested parties, to the construction of a large-scale design model, perhaps using optimisation techniques. This involves supply- and cost-minimisation procedures falling outside the scope of this book.
5. In order to test the solutions or schemes proposed in the previous step it is necessary to **forecast the future values of the planning variables** which are used as inputs to the model. This requires the preparation of consistent quantified descriptions, or scenarios, about the future of the area of interest, normally using forecasts from other sectors and planning units. We will come back to this issue in Chapter 15.
6. **Testing the model and solution**. The performance of the model is tested under different scenarios to confirm its reasonableness; the model is also used to simulate different solutions and estimate their performance in terms of a range of suitable indicators. These must be consistent with the identification of objectives and problem definition above.
7. **Evaluation of solutions** and recommendation of a plan/strategy/policy. This involves operational, economic, financial and social assessment of alternative courses of action on the basis of the indicators produced by the models. A combination of skills is required here, from economic analysis to political judgement.
8. **Implementation of the solution** and search for another problem to tackle; this requires recycling through this framework starting again at point (1).

Although based on the idea of a normative decision theory approach, this framework could also be used within behavioural decision-theory styles, to formulate master plans or to provide ammunition in the bargaining involved in adaptive decision making. It implicitly assumes that the problem can be fully specified, the constraints and decision space can be defined and the objective function identified, even if not necessarily completely quantified.

However, one of the main arguments of this book is that real transport systems do not obey the restrictions above: objective functions and constraints are often difficult to define. With hindsight these definitions often turn out to be blinkered: by narrowing a transport problem we may gain the illusion of being able to solve it; however, transport problems have the habit of ‘biting back’, of reappearing in different places and under different guises; new features and perspectives are added as our understanding of the transport system progresses; changes in the external factors and planning variables throw our detailed transport plans off course. A strong but fixed normative decision-making framework may be suitable for simpler, well-defined and constrained problems but it hardly helps to deal with richer, more complex, many-featured and multi-dimensional transport issues.

How can we improve this general approach to cope with an ever-changing world? It seems essential to recognise that the future is much more tenuous than our forecasting models would lead us to believe. If this is the case, master plans need revising at regular intervals and other decision-making strategies need supporting with the inclusion of fresh information regularly collected to check progress and correct course where necessary. Adaptive or mixed-mode decision-making styles seem more flexible and appropriate to the characteristics of transport problems. They recognise the need to continually redefine problems, arenas and goals as we understand them better, identify new solution strategies, respond to political and technological changes and enhance our modelling capabilities through training, research and experience.

The introduction of a monitoring function is an important addition to the scheme in Figure 1.7. A monitoring system is not restricted to regular data collection; it should also facilitate all other stages in the decision-making framework, as highlighted in Figure 1.8. There are two key roles for a monitoring system. First, it should provide data to identify departures from the estimated behaviour of the transport system and of exogenous key variables such as population and economic growth. Second, the data collected should be valuable in further validating and enhancing the modelling approach followed in preparing the plans.

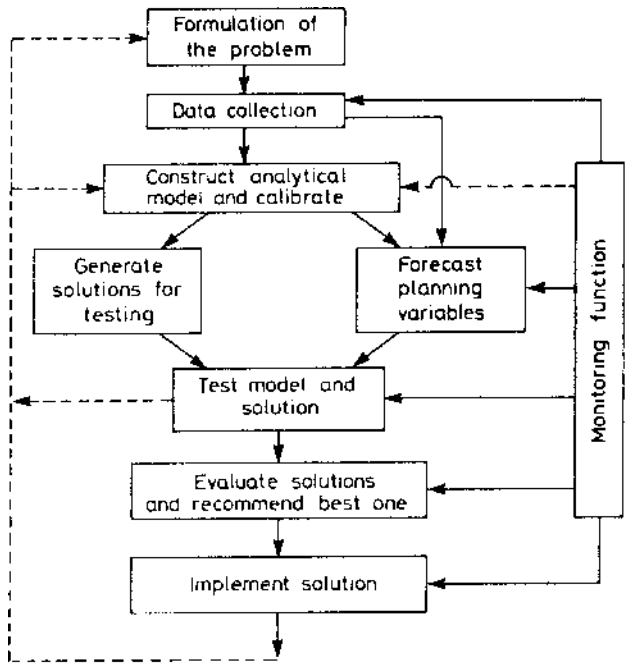

Figure 1.8 Planning and monitoring with the help of models

A good monitoring system should also facilitate learning by the planning team and provide ideas on how to improve and modify models. In this sense, major disruptions to the transport system, like public-transport strikes, short-term fuel shortages or major roadworks which may temporarily change the network structure and its characteristics, should provide a major source of information on transport behaviour to contrast with model predictions. These unplanned experiments should enable analysts to test and enhance their models. A monitoring system fits very well with the idea of a regular or continuous planning approach in transport. If the monitoring system is not in place, it should be established as part of any transportation study.

Monitoring the performance of a transport system and plans is such an important function that it deserves to influence the choice of transport models used to support planning and policy making. The use of models which can be re-run and updated using low-cost and easy-to-collect data, seems particularly appropriate to this task. As we shall see in subsequent chapters, these simpler models cannot provide all the behavioural richness of other more detailed approaches. However, there is scope for combining the two techniques, applying the tool with the highest resolution to the critical parts of the problem and using coarser tools that are easier to update to monitor progress and identify where and when a new detailed modelling effort is needed. We have made an attempt to identify the scope for trade-offs of this kind in the remainder of this book.

P1: TIX/XYZ P2: ABC
JWST054-01 JWST054-Ortuzar February 24, 2011 8:23 Printer Name: Yet to Come

The adoption of a monitoring function enables the implementation of a continuous planning process. This is in contrast to the conventional approach of spending considerable resources over a period of one or two years to undertake a large-scale transport study. This burst of activity may be followed by a much longer period of limited effort in planning and updating of plans. Soon the reports and master plans become obsolete or simply forgotten, and nobody capable of running the models again is left in the planning unit. Some years later a new major planning and modelling effort is embarked upon and the cycle is repeated. This style of planning with the help of models in fits and starts is wasteful of resources, does not encourage learning and adaptation as a planning skill, and alienates analysts from real problems. This approach is particularly painful in developing countries: they do not have resources to waste and the rapid change experienced there speeds up plan and data obsolescence. The use of models that are simpler and easier to update is advocated in Chapter 12 to help the implementation of a sound but low-cost monitoring function.

1.7 Theoretical Basis Versus Expedience

One of the recurring themes of transport modelling practice is the distance, and some would say mistrust, between theoreticians and practitioners. The practitioner would often refer to the need to choose between a theoretically sound but difficult to implement set of models, and a more pragmatic modelling approach reflecting the limitations of the data, time and resources available for a study. The implication is that the ‘pragmatic’ method can deliver the answers needed in the time period available for the study, even if shortcuts must be taken.

The authors have nothing against pragmatic approaches provided they deliver the answers needed to make sound decisions. There is no point in using sophisticated and expensive (but presumably theoretically sound) models for the sake of winning some credit in the academic fraternity. However, there are several reasons to prefer a model based on a sound theoretical background:

1. To guarantee stable results. The recommendations from a study should not depend on how many iterations of a model were run. Prescriptions like ‘always start from free flow costs’ or ‘iterate twice only’ are not good enough reasons to assume stable results: next time somebody will suggest running a couple more iterations or a different, and quite justifiable, starting point; this should not be able to change the recommendations for or against a particular scheme.
2. To guarantee consistency. One should be careful about using a particular model of travellers’ choice in one part of a model system and a different one in another. Pragmatic models sometimes fail to pass this test. Model consistency is necessary to pass the test of ‘reasonableness’ and public scrutiny.
3. To give confidence in forecasting. It is almost always possible to fit a model to an existing situation. However, there are plenty of examples of well-fitting models that make no sense, perhaps because they are based on correlated variables. Variables which are correlated today may not be so tomorrow; for example, a strong correlation between banana production and car ownership in a particular country may disappear once oil is discovered there. Therefore models should be backed by some theory of travel behaviour so that one can interpret them consistently and have some confidence that they will remain valid in the future.
4. To understand model properties and develop improved algorithms for their solution. When one is able to cast a problem in mathematical programming or maximum likelihood terms, to mention two popular approaches to model generation, one has a wealth of technical tools to assist in the development of good solution algorithms. These have been developed over the years by researchers working in many areas besides transport.
5. To understand better what can be assumed constant and what must be accepted as variable for a particular decision context and level of analysis. The identification of exogenous and endogenous

variables and those which may be assumed to remain constant is a key issue in modelling economics. For example, for some short-term tactical studies the assumption of a fixed trip matrix may be reasonable as in many traffic management schemes. However, even in the short term, if the policies to be tested involve significant price changes or changes to accessibility, this assumption no longer holds valid.

On the other hand practitioners have often abandoned the effort to use theoretically better models; some of the reasons for this are as follows:

1. They are too complex. This implies that heuristic approaches, rules of thumb, and *ad hoc* procedures are easier to understand and therefore preferable. This is a reasonable point; we do not advocate the use of models as 'black boxes'; quite the contrary. Model output needs interpretation and this is only possible if a reasonable understanding of the basis for such a model is available. Without ignoring the important role of academic literature in advancing the state of the art, there is a case for more publications explaining the basis of models without recourse to difficult notation or obscure (to the practitioner) concepts. Most models are not that complex, even if some of the statistics and computer implementations needed may be quite sophisticated. Good publications bridging the gap between the practitioner and the academic are an urgent need.
2. Theoretical models require data which are not available and are expensive to collect. This is often not entirely correct; many advanced models make much better use of small-sample data than some of the most pragmatic approaches. Improvements in data-collection methods have also reduced these costs and improved the accuracy of the data.
3. It is better to work with 'real' matrices than with models of trip making behaviour. This is equivalent to saying that it is better to work with fixed trip matrices, even if they have to be grossed up for the planning horizon. We will see that sampling and other data-collection errors cast doubts on the accuracy of such 'real' matrices; moreover, they cannot possibly respond to most policies (e.g. improvements in accessibility, new services, and price changes) nor be reasonable for oversaturated do-minimum future conditions. Use of observations alone may lead to 'blinkered' decision making, to a false sense of accuracy and to underestimating the scope for change.
4. Theoretical models cannot be calibrated to the level of detail needed to analyse some schemes. There may be some truth in this statement, at least in some cases where the limitations of the data and time available make it necessary to compromise in detail if one wishes to use a better model. However, it may be preferable to err in this way than to work with the illusion of sufficient detail but undermined by potentially pathological (predictions of the wrong sign or direction) or insensitive results from *ad hoc* procedures.
5. It is better to use the same model (or software) for most problems because this ensures consistency in the evaluation methods. This is, in principle, correct provided the model remains appropriate to these problems. It has the advantage of consistent approach, ease of use and interpretation, and reduced training costs. However, this strategy breaks down when the problems are not of the same nature. Assumptions of fixed trip matrices, or insensitivity to mode choice or pricing policies, may be reasonable in some cases but fail to be acceptable in others. The use of the same model with the same assumptions may be appropriate in one case and completely misleading in another.

The importance of these criteria depends, of course, on the decision context and the levels of analysis involved in the study. What we argue in this book is for the use of the appropriate level of resolution to the problem in hand. Our own preference is for striving to use good, sound models as far as possible even if some level of detail has to be sacrificed. One has to find the best balance between theoretical consistency and expedience in each particular case and decision-making context. We have striven to provide material to assist in this choice.