3 Urban Travel Forecasting in the USA and UK[1]

David E. Boyce[1] and Huw C. W. L. Williams[2]

[1] Northwestern University, Evanston, Illinois, USA;
[2] Cardiff University, Cardiff, UK

3.1 Introduction

Nearly five decades ago in the United States, urban transport was first subjected to a systematic analysis that became the forerunner of modern transport planning. The reasons for the emergence of this new kind of planning activity are well known and will not concern us here. Within ten years both the approach and the embedded travel forecasting models had been implemented at many major urban centres of the world including several in Europe. Although subject to refinements and reinterpretation, the models in widespread contemporary use in urban, regional, national and international transport studies are recognisably the descendants of those four-stage procedures defined in the late 1950s.

There have always been criticisms of travel forecasting models. In the early 1960s criticism was muted, but by the late 1970s voices in opposition had turned into a clamour. Following the celebrated San Francisco Bay Area lawsuit and the subsequent Clean Air Act Amendments of 1990 (Garrett and Wachs 1996) the critique of "conventional" models took on new urgency. And yet, nearly 50 years after its inception, the four-stage travel forecasting procedure is still in widespread use in large metropolitan areas.

The purpose of this chapter is to examine some of the tensions and conflicts between the multiple states of the art and the states of practice. We interpret the former as comprising the theory of models, conceptions of model systems, model prototypes, and possibly advanced model systems being applied in practical studies which may be on the point of wider acceptance. The latter are characterised by widely available practical applications.

We shall structure our discussion of state of the art developments around two related themes: firstly, those evolutionary approaches in which the multi-stage procedure has been significantly refined; and secondly, those attempts beginning in the early 1970s to model travel behaviour "from the ground up." In Sect. 3.2 we

---

[1] Presented at the 43rd Congress of the European Regional Science Association, Jyväskylä, Finland, August 27–30, 2003.

Reproduced from *Methods and Models in Transport and Telecommunications: Cross Atlantic Perspectives*, A. Reggiani and L. Schintler (eds), Springer, Berlin, 2005, 25-44.

offer a view of the historical development focusing on model specification and equilibration, noting the consequences of some lost opportunities at the very outset of the discipline. This discussion forms an important point of contact with the various states of practice. In Sect. 3.3 we survey those attempts to rethink the field by constructive aggregation over microbehavioural relations, extending early microeconometric approaches based on the theory of discrete choice.

Models are not, of course, constructed in a vacuum but are strongly conditioned by their context, the planning and policy framework, the evaluation system and precision of information required. The selection of the United States (US) and the United Kingdom (UK) to exemplify our discussion in Sects. 3.4–5 is rather more than for the authors' convenience, as it illustrates the influence on model design of differences in historical development, institutional contexts, evaluation and information requirements, and the resources available to maintain and upgrade models. An assessment of progress, current problems and challenges is presented in Sect. 3.6, particularly in relation to what are key requirements for model validity and design. A brief conclusion follows.

3.2 Specification, Estimation and Solution Methods for Integrated Equilibrium Models

Throughout history, traffic congestion has been a phenomenon that mankind finds wasteful and offensive. So it was in the early 1950s when mathematical economics, a new and rapidly developing field, sought to tackle practical problems. A team of young economists (Beckmann et al. 1956) took up the problem of congestion in a transport network, and succeeded in devising a mathematical model of travel and route choices that contributed in a fundamental way to this new field. Assuming that travel between each pair of origins and destinations decreases with increasing cost, that used routes connecting each pair have minimal and equal travel costs, and that the travel costs on the links of the network are increasing functions of the total link flow, Beckmann devised an optimization problem whose solution simultaneously satisfies these three conditions. Although the solution properties of the problem were thoroughly analyzed, no solution method was devised. Concurrently, Wardrop (1952) succinctly described the properties of such equilibria.

This work, completed in 1954, and published in book form only in 1956, did not appear in academic journals. It was summarized in only one professional journal as late as 1967. It did not impact the urban transportation studies that began in those same years in the United States, perhaps because the mathematical treatment was not accessible to the engineers and planners who staffed those agencies. As a result, one of the most important innovations of this field was effectively lost for over ten years.

In place of this integrated model of travel and route choices proposed by Beckmann, a four-stage travel forecasting procedure evolved from 1956 onwards, as described by Martin et al. (1961), in which the authors depicted the procedure in a complex, multi-page diagram. At its heart lay four stages or steps: trip generation (G); trip distribution (D); modal split (MS) and traffic assignment (A). Each was depicted as a separate stage that received inputs from the former and provided outputs to its successor.

Following the acceptance of the four-stage paradigm, most researchers and professionals became engaged in the improvement of the models and methods described in the individual stages. Of particular relevance here, household-based category analysis replaced zone-based regression models for trip productions (Wooton and Pick 1967), and utilities or generalised costs derived in early studies of modal choice with models specified and estimated at the individual level (Quarmby 1967; Warner 1962) were embedded in assignment, modal split and distribution models (Wilson et al. 1969). The incorporated generalised costs, specified as linear functions of objectively measured attributes with travel time suitably scaled to money units, served as an interface between policies, behavioural response and benefit evaluation. The numerical estimate of the “value of time” has, arguably, proved to be one of the most important parameters in the whole of planning.

From an analytic viewpoint, the earliest distribution and modal split models, which involved apportioning trips between different locations and modes, adopted empirically derived functions – sometimes referred to as deterrence functions (for spatial interaction), and diversion curves (for modal shares) – and these were determined through “goodness-of-fit” criteria. By the late 1960s these started to be replaced by analytic functions, and share models of the multinomial logit form were widely adopted (Manheim 1979; Wilson 1970; Wilson et al. 1969). These were conceptually appealing, analytically tractable and consistent with several theoretical constructs that were starting to be used for interpreting dispersion associated with travel (Erlander and Stewart 1990; McFadden 1973; Wilson 1970).

A problem that exercised the earliest modellers was the ordering of the G, D, MS and A segments and how they should be “linked” together. There were, from the start, informal behavioural assumptions underpinning the four-stage approach in terms of a sequence of decisions that mapped onto the individual submodels. However, the correspondence between the Generation (G), Distribution (D), Modal Split (MS) and Assignment (A), with frequency (f), location (l), mode (m) and route (r) choice of the trip, respectively, remained tenuous.

Various alternative structures for the demand model were proposed reflecting, it was assumed, the conditionality of a sequence of decisions, the most popular being whether the distribution submodel preceded (G/D/MS/A), followed (G/MS/D/A) or was combined (G/D-MS/A) with the modal split submodel. The first two constructions involved the formation of “composite costs” that represented “average costs” at what were referred to as “later stages of the models.” As late as the mid 1970s no detailed theoretical basis for the entire model existed; for a given ordering, which was suggested *a priori* from behavioural assumptions, the form of the composite cost was regarded as an extra degree of freedom for achieving improved “goodness-of-fit;” see Senior and Williams (1977) for a review of model structures adopted in practice.

The derivation of the nested logit model within discrete choice theory provided one resolution to these ambiguities (Daly and Zachary 1978; McFadden 1978; Williams 1977). This development endowed the whole model with a behavioural rationale in which the analytical structure of the demand function reflected underlying utility functions, imposing two important restrictions on the overall model. Firstly, the composite costs that interfaced the different submodels needed to be formulated in a particular way; for logit-type models these were in the form of a "log sum" function, a form that had already been implemented with microdata by Ben-Akiva (1974). Secondly, the parameters that determined the sensitivity of trip choices to changes in times or costs, had to decrease as one progressed from route choice, through mode to locational and frequency selection in the G/D/MS/A structure. Only then would it be ensured that the *estimated* direct- and cross-elasticity parameters had the appropriate sign, requiring the demand for an alternative to fall when its cost rose or the cost of a substitute fell. The nested logit model thus provided a consistent way of combining the various constituent choices with differential cross substitution between alternatives, and made the ordering of associated logit share functions subject to empirical test. It is important to note that the specification of the demand model with empirically derived functions for locational and/or modal shares is not immune from this strict requirement for appropriate response properties derived from the calibrated model.

Williams and Senior (1977) reconfigured the four-stage procedure as a nested logit structure, experimented with different orderings of the distribution and modal split models, and showed that many models in UK practice did not satisfy the necessary parameter inequalities implied by the chosen structure. The specific implication of this finding was that such models could have produced counter-intuitive results. The more general implication was that calibration and the traditional notion of validation based on goodness-of-fit and held-back samples was not a sufficient test for the validity of such cross-sectional models for policy appraisal.

A fundamental problem that confronted theoreticians and practitioners since the earliest days was the design of solution procedures to generate equilibrium states of required precision, in which the costs (or times) of travel through the networks are consistent with the demands that created them. The practical difficulties of this problem almost certainly have resulted from the mathematical statement of the problem that Beckmann had produced years earlier being unknown to the professionals who initially proposed the *ad hoc* four-stage procedure.

Until 1975 the problem of equilibrium was almost entirely seen to be confined to the assignment of fixed trip matrices. Several *ad hoc* approximate procedures were devised before rigorous solution algorithms based on the Beckmann formulation emerged. In the morass of numerical detail involved in handling several trip matrices and large urban networks, the additional complexity of seeking self-consistency throughout the entire procedure tended to be seen as an unnecessary luxury or was simply unrecognised or ignored.

Where it was considered, the notion of "feedback" of costs from the assignment to other segments of the model began to be discussed. The congested costs were simply recycled back to the modal split and possibly to the distribution model, and amended modal matrices returned to the assignment process. A few scholars, however, had already begun to investigate ways to combine the trip distribution, mode split and traffic assignment stages, and eventually rediscovered Beckmann's formulation. The first to embark on this line of thinking was Murchland (1970), but his efforts were largely unsuccessful. Subsequently, Evans (1976) and Florian and Nguyen (1975, 1978) proposed formulations, which were special cases or elaborations of Beckmann's original formulation, and solution algorithms. Upon evaluation, only the algorithm of Evans proved to be practical for solving problems of realistic size (Boyce et al. 1988).

Even with this advance in understanding, these combined models were largely a research curiosity and effectively unknown to practitioners. Initial efforts to implement such models occurred in the early 1980s and continued through the 1990s. These early combined, or integrated, models had serious limitations as compared with professional practice, representing only a single class of travel or trip purpose, and being much smaller in scale than the models used in practice.

During the past ten years, several multiclass, integrated models have been implemented and applied. The first multiclass integrated model was implemented by Lam and Huang (1992); however, since classes in their model correspond to modes, it is not comparable to other multiclass models. The second model represents the work of de Cea et al. (2003), and is arguably the most advanced model available today. The third example is by Boyce and Bar-Gera (2003), a two-class research model for the Chicago Region implemented at the same scale and detail used in professional practice. Each of these models was estimated from available travel surveys and validated against census data. In the case of Santiago, Chile, the model has evolved to the status of commercially available software, called *ESTRAUS*. Each model is solved with algorithms that may be traced to Evans (1976). Subsequent to these developments, Bar-Gera and Boyce (2003) integrated the origin-based assignment algorithm of Bar-Gera (2002) into a single class integrated model. For a review of operational multiclass models, see Boyce and Bar-Gera (2004).

There is a close correspondence between the Evans algorithm for solving an integrated model and the four-stage procedure; in effect the four-stage procedure is actually a primitive algorithm, or solution procedure, for solving an unstated, integrated model. Current understanding of integrated models is providing guidance for solving the four-stage procedure with feedback, as is now required in the US. The gist of this insight is to recognize that the trip tables and road link flows are the solution variables of the problem, which need to be adjusted from iteration to iteration to drive the solution towards equilibrium. One difficulty from the four-stage perspective is the lack of a well-defined objective on which to base this adjustment procedure.

3.3 Specification, Estimation and Solution Methods for Microanalytic Models

In a rallying call to accompany the early stages of the Travel Model Improvement Program, Wachs expressed a view that resonated with many: “the state of practice in transportation planning consists of an obsolete approach to modelling which has been marginally updated and adapted by clever technicians but which has not been fundamentally rethought from the ground up” (Wachs 1996, p 213). The four-stage procedure was deemed to belong to another age, essentially one of large infrastructure development, quite inappropriate for a new generation of travel demand management policies. There was a widely held view that what was needed was a finer representation and understanding of both the demand for and supply of transport services more closely attuned to the contexts of application.

In fact, attempts to model travel and transport systems “from the ground up” started to appear in the early 1970s in the work of, among others, Ben-Akiva (1974) and Domencich and McFadden (1975). Systems of disaggregate demand models were specified and estimated at the individual level and were then subjected to aggregation to generate the desired forecasts. This approach contrasted with the traditional one of specifying models for broad groups and then subjecting them to parameter estimation.

The basic motivation for this style of work was to consider the full variability in individual behaviour and avoid or ameliorate aggregation bias, to provide efficient sampling procedures for model estimation, and above all to furnish the model system with a behavioural theory based on individual choice that would, it was hoped, yield a more suitable and stable basis for forecasting. We shall discuss this “ground up” approach and its emergence as an operational competitor to models of the traditional form. Further developments based on activity-travel frameworks are then considered briefly.

By the late 1970s the microeconomic theory of discrete choice had taken root and provided a behavioural basis for travel forecasting. For the last 25 years this theory has been the dominant paradigm within which predictions of travel behaviour have been made; see Ben-Akiva and Lerman (1985), Oppenheim (1995), Ortúzar and Willumsen (2001), as well as McFadden (2001) for a 30-year retrospective review of the approach. The discrete choice framework, based on random utility theory, proved to be a particularly powerful one capable of wide application, and many of the locational and travel-related models were reinterpreted from this viewpoint. Spatial interaction models were explicitly treated as models of locational choice within housing markets (Anas 1981; McFadden 1978; Williams and Senior 1978). Modal choice models were studied in great detail and the traditional two or three way choices were extended to a variety of private, public transport and slow (walk and cycle) modes. For these purposes, the multinomial and nested logit models were widely used, the latter incorporating “similarity” and differential substitution between alternative choices, and being preferred to the more general but computationally less tractable multinomial probit function. More general specifications, such as General Extreme Value (McFadden 1978) and mixed logit (McFadden and Train 2000) models are now increasingly applied in academic studies, but it may be some time before the multinomial and nested logit are replaced in practice.

While the revealed preference (RP) approach dominated the early studies of discrete choice, stated preference (SP) methods emerged in the 1980s as a powerful and widely used technique; see, for example, Bates (1988) and Hensher (1994). The use of micro-computers to implement stated preference experiments to study and forecast travel behaviour dates from the mid-1980s. *MINT* is a prominent example of software developed at that time for the generation of different SP designs. The version *WinMINT* resulted from a collaboration between Hague Consulting Group and Accent Marketing and Research (Hague Consulting Group 2001). The relative strengths of the revealed and stated preference approaches are now widely recognised and mixed data designs are increasingly used to exploit the strengths and avoid the weaknesses of each.

Much theoretical and applied developmental work on specification, estimation and aggregation was undertaken in the 1970s and 1980s by Ben-Akiva and by McFadden, and their colleagues. In Europe the contribution of Andrew Daly was particularly significant, not least in the development of the software *ALOGIT* for nested logit estimation (Daly 1987).

In its selective review of world practice in the late 1990s, Hague Consulting Group (1997) critically discussed, among others, applications of the microeconometric approach in the San Francisco Bay Area, The Netherlands, Stockholm and Sydney. The authors advanced the claim that “the evidence from the Review indicates that disaggregate models now provide a practical alternative to aggregate models in terms of providing trip tables to load onto networks. In terms of ability to capture aspects of behaviour that may affect forecasts of demand, they consistently outperform their aggregate counterparts.” They also note: “Representations of travel patterns as trips or tours have both been able to form the basis for operational systems; they are to be deemed ‘available technology. ’”

Research over the last twenty years aimed at relaxing some restrictions involved in the formation of simple utility models, by incorporating imperfect information, limited choice sets and satisficing behaviour – which are clearly present to differing degrees in all choice contexts (see, for example, Bhat 2002; McFadden 2001; Williams and Ortúzar 1982). These developments led to the question of how to refine existing models to improve the representation of behaviour, and the more philosophical problem of how much “understanding” is required for satisfactory prediction, questions that continue to divide social scientists.

For the last 25 years the major question confronted by travel behaviour theorists has been how to embed spatial and temporal choices and preferences within a rich constraint-based environment formed by household interdependencies, activity scheduling and the transport supply. Although the activity-travel framework has flourished as a *descriptive* basis for travel demand analysis since the late 1970s (see, for example, Jones 1979; Jones et al. 1983), synthesis of the activity-travel framework based on constraints and the discrete choice paradigm for *forecasting* travel had to await the late 1980s and 1990s before significant development was achieved. The broad distinction between activity-travel models under development, and in particular between econometric and hybrid simulation approaches, has been provided by, among others, Wachs (1996), Rossi and Shiftan (1997) and Bowman and Ben-Akiva (2001).

3.4 States of Practice in the USA

Urban travel forecasting may be regarded as beginning in 1955 with the Chicago Area Transportation Study, the first to implement a procedure that resembles the four stages described above. Significantly, it was the first study to assign a trip table (origin-destination flows) to a road network taking into account the effects of congestion. Concurrently, engineers at the Bureau of Public Roads were devising other procedures and computer programs for assigning trips to congested road networks.

Model and computer code development in the US prior to 1970 may be summarized as follows: use of household surveys to estimate daily trip frequencies (trip generation); development and application of two trip distribution models, the doubly-constrained gravity model and the intervening opportunities model; use of simple diversion curves to allocate trips to modes, either before or after trip distribution; implementation of heuristic procedures for assigning trips to congested road networks with travel times increasing with flow, but without solving a well-specified equilibrium problem; development and application of main-frame computer programs to solve this sequence of models, but without taking account of inconsistencies among the models.

This fragmented activity led in the early 1970s to a more coordinated effort within the US Department of Transportation. Some of the achievements of the Urban Transportation Planning System (*UTPS*) may be summarized as follows: introduction of the multinomial logit function for forecasting mode choice; introduction of a user-equilibrium algorithm (effectively, the Frank-Wolfe method) for assigning auto trips to congested road networks; improvements in the coding of transit networks and the assignment of transit trips; recognition of the need to achieve consistency among the four stages, with some concern about solving the procedure with feedback. The creation and distribution of *UTPS* also led to the preparation of manuals and training courses (US DOT 1977).

Changes in federal policy in 1981 led to a decision to terminate the development of *UTPS*. Comsis, a private organization that had been involved in its development, then launched their product, *MinUTP*. A somewhat similar product called *TranPlan* had been under development for some years, initially by Control Data Corporation. These two software systems, either directly or indirectly encompassing the model and code development efforts of US DOT, were the initial versions of PC-based travel forecasting models and software. Other software systems from the same period have effectively not survived.

In parallel with these developments, researchers at the University of Montreal, Canada, had implemented an equilibrium-based two-mode urban travel forecasting model called *EMME*, drawing on research findings cited above (Florian et al. 1979). Subsequently, this method became the basis for *EMME/2*, a commercial software system first released in the late 1980s for linking and solving the models of the four-stage procedure (INRO 2004). In addition to a rigorous implementation of a user-equilibrium road assignment algorithm and a probabilistic transit route choice algorithm, *EMME/2* includes tools for solving the doubly-constrained trip distribution model, stochastic mode choice models, network coding procedures and related utilities.

Building on the capabilities of the emerging field of Geographical Information Systems, a group of travel modellers assembled a system known as *TransCAD* based on PC technology (Caliper Corporation 2004). This system sought to embrace and incorporate various research advances throughout the US and beyond, including an early version of the combined model described above. Another US-developed research-based software system, which has been found to be useful for smaller regions, is *Quick Response System II* (*QRS II*), developed by AJH Associates (2004). *EMME/2* and *TransCAD* have enjoyed considerable success throughout the US during the 1990s; *EMME/2* also succeeded in developing an international following in the UK, Sweden, Canada, Australia and New Zealand, and South Africa, to name several.

The late 1980s was a period of change in the US, as environmental-based interest groups successfully sought to challenge the status quo of transportation planning in the courts (Garrett and Wachs 1996). The Bay Area lawsuit led to the Clean Air Act Amendments of 1990, and the requirement of conformity analysis for MPOs. To qualify a transportation plan for federal aid, each MPO had to demonstrate that building the proposed system would not result in the deterioration in air quality. One implication of this requirement is the need to determine road link flows and speeds by time periods of the typical weekday, in order to forecast atmospheric emissions. Since available travel forecasting methods were based on a 24-hour period, an ad hoc factoring method, based only on surveys, was devised to allocate flows by time periods of the day (Deakin et al.1993).

Attempts to reform transportation planning practice also led to new provisions in the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA). Many new requirements were placed on planners and MPOs by this legislation, but the one most pertinent to this discussion is the requirement to solve the four-stage procedure with “feedback.” During the early 1990s many MPOs updated their travel forecasting models for the first time in many years, which often resulted in adoption of a different software system.

In response to these legislative mandates, model updates and related technical requirements, US DOT and US Environmental Protection Agency (US EPA) created the Travel Model Improvement Program, as mentioned earlier. The agencies had begun to devise a series of short and longer run model improvement tasks, when they were directed by the US Congress to fund a large-scale systems simulation project at the Los Alamos National Laboratory dubbed TRANSIMS. This ambitious effort set out to apply microsimulation techniques to represent interrelated travel and location-based decisions of each inhabitant of a metropolitan area throughout the 24-hour weekday. The result was not a new, integrated model so much as a system of computer programs related to choice of when, where, and how to travel, and the associated computation of atmospheric emissions. An attempt to implement the system was initially made for the Portland, Oregon, metropolitan area, but was not completed by 2001, as intended. At that time a transportation modelling consultant was retained to make a final attempt to implement and validate the system. This effort remains a work in progress.

3.5 States of Practice in the UK

By the middle of the 1970s the era of the large urban transport studies was coming to an end and few city authorities had the resources or inclination to maintain large transport models. With few exceptions, most notably in London, model systems constructed a decade or more earlier, and the databases that supported them, were allowed to atrophy. Much local expertise dispersed, and the under-resourced and lonely task of local authority modellers fell on fewer and fewer shoulders. Where necessary, international consultants, some with the founders of the discipline at their helm, were called upon to address strategic and tactical issues, often in the form of urban development projects and traffic studies. Large urban transport projects were relatively few in number. Partly in an attempt to wrest control in a deregulated public transport environment, however, renewed interest was shown in Light Rapid Transit schemes for which the estimation of decongestion benefits was a significant component of the external benefit of such schemes and central to mounting the case for Central Government funding support. Throughout the 1990s there was also an increasing interest in demand restraint; several metropolitan areas conducted modelling exercises involving public transport, traffic restraint and, in some cases, limited highway investment.

For such purposes many software packages were available in the private sector. Indeed, a key feature of the 1980s and early 1990s was one of fragmentation of the transport planning software suite, and its organisation within management information systems as a collection of submodels that were refined and integrated on a “pick-and-mix” basis (see Williams 2004 for an overview of developments). In turn, large consultancies found it in their interest to join forces with smaller specialist companies, particularly in the context of implementation of land use models, microeconometric studies of discrete choice (typically multimodal studies), and stated preference exercises.

A prominent example, and one of the most widely used and innovative, was the *TRIPS* suite (now *CUBE*) of Martin, Voorhees and Associates (MVA 2004). The various submodels of the four-stage approach had been enhanced and available in both synthetic and incremental (pivot point) forms (Bates et al. 1987). In the former travel behaviour is modelled at the cross section and elasticity parameters estimated prior to forecasting, while in the latter, changes from a given state (e.g. the base state) are estimated utilising given elasticity parameters. Both are available for application at the micro (individual data) or aggregate (grouped data) level.

Since the mid-1980s *SATURN* (Van Vliet 1982) was extensively applied in and beyond the UK. Matrix updating was a standard part of the package and was widely applied to breathe new life into dated trip matrices. Initially promoted as a "modern" assignment program, it was the first to incorporate a rigorous approach to equilibrium assignment in UK transport planning applications. *SATURN* not only had the capability of working at different levels of network resolutions, requiring different and compatible specifications, but in the 1990s became a framework within which research was conducted on consistently integrating demand and assignment models. A further important development in the 1990s was the extension of microsimulation models from their traditional junction applications to treat wider urban networks. In the UK the *PARAMICS* model system developed by SIAS Limited is an example of a microsimulation approach which has been applied to a wide range of towns and cities (Druit 2000).

In the late 1980s and early 1990s there was considerable interest in both the US and UK in the effect of congestion on travel, and of additional traffic (and vehicle miles) that might be induced by capacity expansion and new roads. In the US the effect of induced traffic on atmospheric emissions and energy consumption was a matter of particular concern (TRB 1995), while in the UK it was the effect on congestion, traffic growth and economic benefits (SACTRA 1994). The report of the Standing Advisory Committee for Trunk Road Assessment had a substantial impact on the methods by which travel forecasting and the appraisal of schemes were conducted in the UK (Williams 2004).

Congestion problems on interurban networks were traditionally addressed within a single mode planning framework. Within an essentially "predict and provide" approach, construction of more highway capacity almost inevitably resulted. By the mid-1990s this approach was officially questioned and, in an attempt to establish sound environmental credentials, the 'New Labour' Government elected in 1997 sought to re-examine the problems of development and transport in the major corridors, drawing in the large urban areas or conurbations of which they were part. The policy terms of reference of the 22 Multimodal Studies set up in 1998 embraced all modes, and demand restraint in addition to road and public transport investment. The studies were accompanied by an enhanced evaluation framework that sought to establish a "level playing field" between the modes and is based on environment, safety, economy, accessibility and integration criteria.

The Department for Transport issued modelling and evaluation advice to accompany the Multimodal Studies and for Local Transport Planning (DETR 2000; DfT 2002). The methodological framework is not narrowly prescriptive and allows considerable discretion on the part of consulting teams according to the nature of the local areas and their problems. The wide range of approaches applied so far for the Multimodal Studies has included: both integrated land use-transport models and independent estimates of trip ends; four- and five-stage structures down to two-stage models with a simple bridge between modal networks; and synthetic and incremental (pivot point) approaches.

Both the Multimodal Studies and Local Transport Planning are being conducted within the overall framework of the Government's Ten Year Plan (which includes national targets for congestion, accidents and pollution), and each has the requirement to examine a whole range of initiatives which will encourage more sustainable transport arrangements and particularly improve integration, both between modes and between land use and transport. Land use, infrastructure investment, fiscal and regulatory policies, as well as a range of “soft” measures, such as workplace travel plans, school travel plans, teleworking, cycling and walking strategies, are to be considered. The evidence base for travel forecasting for some of these policies is not strong and transferability of experience will likely be considered of more value than model-based prediction as a guide to implementation.

As part of the guidance issued by the Department for Transport in relation to infrastructure investment appraisal, the advice on variable demand analysis, to be released in 2005 (Department for Transport, personal communication) is likely to have significant implications for both strategic and tactical model development in the UK. Following earlier advice (Highways Agency 1997), the emphasis will be on model design, and will be much enhanced in terms of the discussion of: model complexity, proposed structures, model parameters, design and convergence of equilibrium-seeking algorithms, model validation and sensitivity analysis. Of particular interest in the present context is the research conducted on the model system DIADEM (Development of Integrated Assignment and Demand Modelling) and the development of improved equilibrium models and solution algorithms. Further information about this research on variable demand modelling in general, and DIADEM in particular, can be found in Department for Transport (2003).

As part of the research underpinning the on-going scrutiny of impact models and appraisal methods by the Department for Transport, Bly et al. (2001) have reviewed the structure, parameters and validation of 24 of those models that have been implemented in the last ten years or so, drawing largely but not exclusively from UK applications. Some have been applied in Multimodal Studies, others in strategic assessments in major cities. The models exhibit wide variation in their explicit representation of responses and range from five-level nested structures in the form G/D/MS/T/A (with T representing the journey timing choice) to 2-stage models. Many models are in the form of nested (logit) functions, some in incremental form (Bates et al. 1987). Following our discussion in Sect. 3.2, it is interesting and sobering to note the finding by Bly et al. (2001) that several of the synthetic models, with nested structures, are endowed with estimated parameter values that are *inconsistent* with the structures selected. We would suggest that model validation is here deficient and some of the predictions may have been unreliable. Where incremental nested logit forms have been adopted, parameter values have been inserted that are consistent with the selected specifications. This UK experience suggests that current US practice, in which distribution models often include empirically derived deterrence functions, be scrutinised to confirm that the demand functions are endowed with acceptable elasticity properties.

3.6 Assessment of Progress, Problems and Prospects

The experience of the past 50 years, both in practice and research, offers many insights and lessons for the future development of our field of urban travel forecasting. In this section, we offer a very brief assessment of progress, problems and future prospects and some directions. To give some structure to the discussion we present our points under a number of key themes as shown below.

3.6.1 Contexts of Model Development

Although many of the core issues of our field, particularly congestion mitigation, accessibility provision and urban growth management have long been key concerns, the sustainability agenda, including air quality and environmental justice, have imposed new information and appraisal considerations. Future developments will be driven in part by these changes and information requirements. While we have cited much progress in the development of theory, method and technique over the last fifty years, the lead time of new ideas is uncomfortably long and we are even now questioning issues such as the impact of that most common of policies, highway development, that were once taken for granted. While infrastructure investment and pricing remain prominent policy instruments, new initiatives are now widely advocated. The field is still not well placed to deal with the full range of demand management policies. Although the generalised cost function is very flexible, and much ingenuity is shown in representing the “policy-model interface,” the knowledge base in many areas is currently weak. Indeed, some estimates for the implication of lifestyle changes, technological initiatives (such as e-commerce and telecommuting), and even some price-based demand management policies, are often little better than guesses and rely on minimal empirical evidence. Transferability of experience will be particularly important in this context.

3.6.2 Distinct Alternatives for Practice

We agree with the view expressed by the Hague Consulting Group (1997) that there is no generally accepted state of the art or state of practice in travel forecasting, but rather large classes of models and model systems, some at distinct stages of development and in a process of continual refinement. These classes of models may achieve particular significance in different periods of the development of the field. Thirty years ago “aggregate” and “disaggregate” models and “simultaneous” and “sequential” structures were distinctions that stimulated considerable debate. Nowadays, the former remains - whether aggregation precedes or follows estimation - but the unit of demand, be it trip-based, tour-based and activity-travel based, has become a prominent distinction. In the future, microanalysis specifications and estimates with cross sectional or longitudinal (panel) data may well become a further key distinction. Although the four-stage procedure has dominated applications over this period, they have long since absorbed different “disaggregate” concepts, becoming hybrid forms, and have been applied in anything from two- to five-stages or levels. Such is the variation in this approach, both in detail and quality of specification, that care is needed in applying a single collective description to the genus.

3.6.3 Theoretical Developments

Although still a technical and rather esoteric specialisation, travel forecasting has long since emerged from its exclusive traffic engineering, operational research and regional science roots, and now widely embraces economic, statistical, geographical and psychological theories and constructs. The award of a Nobel Prize to Daniel McFadden in 2000 affirmed both the contribution of transport applications to consumer econometrics, and the ascendancy of the discrete choice paradigm as a basis for describing, understanding and predicting travel behaviour. That said, from a behavioural standpoint, the field of practice is still theoretically immature. Although activity-based travel demand models are now developing rapidly, hitherto they have had little impact on forecasting in practice.

One of the original motivations for the development of the field, to gain an improved understanding of the interrelationship between land use and transport has been achieved and several operational, if cumbersome, integrated land use-transport models based on inter-related markets have emerged. Within the activity-travel framework, uncovering the detailed relationship between the travel behaviour of individuals and households and the structure of neighbourhoods and cities, which is showing considerable advances, (see, for example, Kitamura et al. 1997) will have important implications for our understanding of the sustainability of cities and regions.

The equilibrium paradigm remains dominant in our field and this is likely to continue. The analytical framework set out by Beckmann was rediscovered and is now being exploited more fully for the specification and solution of models. The representation of travel behaviour as a system of constraints, both equalities and inequalities with embedded choice functions, allows for the derivation and solution of families of models that may be widely adapted to different contexts. By focusing attention on the identification of this system and its relationships to benefits of engaging in activities, as well as travel costs, broadly defined, we can extract the essence of accumulated knowledge from multi-stage procedures.

3.6.4 Specification of Models and Model Design

Travel forecasting models are now characterised by greater precision in their representation of behaviour. We point to the increased stratification of traditional models by passenger groups, trip purposes, and time periods, and the greater refinement of the supply side. Time-of-day modelling and time-switching are now key aspects of our understanding of behavioural responses to congestion and transport policies. Experience with temporally stratified models, and current policy concerns, strongly suggest that 24-hour forecasts with factoring to time-of-day periods should no longer be used.

Highway network coding procedures have much improved the representation of supply, and may relate to demand in the form of individual vehicles, vehicle packets and flow groups. In particular they provide much greater appreciation of the significance of interacting flows at junctions. Public transport networks also now offer crowding effects. Transport networks should be represented in terms of road and transit link functions that depend upon the conflicting flows encountered on that link. Simpler functions that depend only on the links own flow may suffice for more aggregated representations suitable for long range, general models, but more detailed functions are needed for depicting the interactions of person and vehicle flows. The implications of network specification for solution algorithms and solution properties are important, and should be clarified for practitioners.

More information is needed about the costs and benefits of implementing models at different levels of resolution and the balance between over-elaboration and mis-specification, given current knowledge and the information requirements. In many areas of our field, and in particular, the treatment of the inter-relationships between land use and transport, and the detailed specification of network models, the value added by greater resolution must be subject to detailed design criteria.

3.6.5 Estimation of Response Parameters

Stated preference methods have emerged as a powerful technique to augment revealed preference studies of individual choice behaviour. A current issue of major importance is the extent to which such methods yield useful results when the time or cost changes representing policies are small (less than 5 minutes). See the discussions in Mackie et al. (2003) and Hague Consultant Group (1996). This is a significant and unresolved problem, given the value of typical travel time changes accompanying many transport policies. A considerable amount of effort is now being directed towards reconciling demand responses from different types of data, specifically longitudinal and cross-sectional data sets. Contributions to impacts from different short and long run response mechanisms require clarification, given their importance to the economics of projects.

3.6.6 Solution of Models

The need to measure small differences between equilibrium states, particularly associated with economic benefits and emissions, imposes strict requirements for the precise solution of models. Considerable progress has been achieved in the design of algorithms to solve models in an internally consistent manner. The concept of solving for travel choices in stages should be de-emphasized, and replaced with convergent solution methods now becoming available. These solution procedures should be imbedded in future software systems in ways that can be readily applied by practitioners. Software vendors need to incorporate these new approaches, allowing users to specify a system of functions and constraints to be solved by specific or general purpose solution procedures.

3.6.7 Validation of Models

Validation is often and, in our view, incorrectly viewed simply as constructing models that satisfactorily replicate aspects of travel behaviour in a base year. This approach is necessary but far from being sufficient for forecasting the *response* to policies that requires adequate elasticity measures. Many models have been applied that do not attain the necessary standards for the purposes specified. As long as this notion continues to dominate the field, inappropriately specified models will slip through the validation net and little progress will be made in the comparison of model systems.

3.7 Conclusions

In 50 years travel forecasting has evolved from a fledgling field in North America to a worldwide activity in the support of public and increasingly private policy decision making. It is also a greatly expanded academic discipline in which understanding of travel behaviour is an end in itself, which requires different assessment criteria than for travel forecasting.

Perhaps travel forecasting models have failed to deliver their initial promise; we have pointed to lost opportunities in the early years and to inappropriate notions of validity. Theoretical deficiencies, predictive inaccuracies, and a general battering by those with a different approach to promote are evident. Travel forecasting activity has itself been undermined by a discredited “predict and provide” approach to planning within which it was housed. The remarkable longevity of those models of the “traditional form,” however, is due to their capacity to absorb innovations; it has been suggested many times that in spite of known deficiencies, professionals trained in the use of these methods are, on the whole, comfortable with their results. There is also perceived to be a lack of a clear alternative for generating the required information. Assessment of the models is rightly seen as a relative not absolute activity, in relation to the next best alternative.

In conclusion we offer the following recommendations.

1. All models should be subjected to a detailed audit, particularly in relation to their descriptive, forecasting and policy testing characteristics. In particular, equilibration methods and qualitative and quantitative response properties should be assessed in order to screen for perverse behaviour.
2. Urban travel model development should be viewed as an ongoing process, requiring continual funding. It may well be context specific, and that one size will not necessarily fit all MPOs. It should also be seen as a technical matter. Legislative and administrative mandates are unlikely to be productive.

3. We, like many others, regard enrichment of the knowledge base in the form of case studies and before-and-after studies to be a priority for the development of the field. Case studies combining experience of transport system design and management with modelling of travel choices on these systems will be most useful in understanding both policies and how to model them.

We believe that the field is best served by taking a more eclectic view than is fashionable, avoidance of exaggerated claims and resisting model comparisons at different stages of their development. In this comparative exercise, nothing is obsolete until it is replaced by a model system fully tested in the arena of practice. The lack of standardisation of the models, and the protracted stages of development (from concepts, through prototypes to applications in the field) all suggest careful scrutiny of available alternatives for undertaking a specific task and caution against hasty and possibly ill-directed comparisons between models.

Acknowledgements

The contributions of the first author to this paper were initiated while he was an Erskine Fellow at the University of Canterbury, Christchurch, New Zealand, April–June 2003; revisions were made while he was a Visiting Professor at the University of Pennsylvania during the Fall Term 2003. The authors wish to acknowledge the helpful comments of two anonymous referees.

References

AJH Associates (2004) Quick response system II. Available at http://my.execpc.com/~ajh/ (accessed 19 September 2004)

Anas A (1981) Residential location markets and urban transportation: economic theory, econometrics and policy analysis with discrete choice models. Academic, New York

Bar-Gera H (2002) Origin-based algorithm for the traffic assignment problem. Transportation Science 36**:** 398–417

Bar-Gera H, Boyce D (2003) Origin-based algorithms for combined travel forecasting models. Transportation Research 37B: 405–422

Bates JJ, Ashley DJ, Hyman G (1987) The nested incremental logit model: theory and application to modal choice. Proceedings 15th PTRC Summer Annual Meeting, Seminar C, University of Bath

Bates JJ (ed) (1988) Stated preference methods in transport research. Journal of Transport Economics and Policy 22: 1–137

Beckmann M, McGuire CB, Winsten CB (1956) Studies in the economics of transportation. Yale University Press, New Haven

Ben-Akiva ME (1974) Structure of passenger travel demand models. Transportation Research Record 526: 26–42

Ben-Akiva ME, Lerman S (1985) Discrete choice analysis: theory and applications to travel demand. MIT Press, Cambridge

Bhat CR (2002) Recent methodological advances relevant to activity and travel behaviour analysis. In: Mahmassani HS (ed) In perpetual motion: travel behavior research opportunities and application challenges. Pergamon, Oxford, pp 381–414

Bly P, Emmerson P, van Vuren T, Ash A, Paulley N (2001) User-friendly multistage modelling advice. Phase 2: Modelling parameters, calibration and validation. Project report for ITEA Division, DTLR

Bowman JL, Ben-Akiva ME (2001) Activity-based disaggregate travel demand model system with activity schedules. Transportation Research 35A: 1–28

Boyce DE, LeBlanc LJ, Chon KS (1988) Network equilibrium models of urban location and travel choices: a retrospective survey. Journal of Regional Science 28: 159–183

Boyce D, Bar-Gera H (2003) Validation of multiclass urban travel forecasting models combining origin-destination, mode, and route choices. Journal of Regional Science 43: 517–540

Boyce D, Bar-Gera H (2004) Multiclass combined models for urban travel forecasting. Network and Spatial Economics 4: 115–124

Caliper Corporation (2004) TransCAD. Available at http://www.caliper.com (accessed 19 September 2004)

Daly AJ, Zachary S (1978) Improved multiple choice models. In: Hensher DA, Dalvi MQ (eds) Determinants of travel choice. Saxon House, Westmead

Daly AJ (1987) ALOGIT 3.2 user's guide. Hague Consulting Group, The Hague. The latest version ALOGIT4 is described at www.hpgholding.nl/software/welcsoft.htm (accessed 19 September 2004)

Deakin, Harvey, Skabardonis, Inc. (1993) Manual of regional transportation modeling practice for air quality. National Association of Regional Councils, Washington, DC.

De Cea J, Fernandez JL, Dekock V, Soto A (2003) ESTRAUS: a computer package for solving supply-demand equilibrium problems on multimodal urban transportation networks with multiple user classes. Presented at the Annual Meeting of the Transportation Research Board, Washington, DC.

Department of the Environment, Transport and the Regions (DETR) (2000) Guidance on the methodology for multi-modal studies (GOMMMS). Available at http://www.dft.gov.uk (accessed 19 September 2004)

Department for Transport (2002) Major scheme appraisal in local transport plans, parts 1, 2, 3 detailed guidance on public transport and highway schemes. Available at http://www.local-transport.dft.gov.uk (accessed 19 September 2004)

Department for Transport (2003) Variable demand modelling consultation advice background papers. Available at http://www.dft.gov.uk/stellent/groups/dft_econappr/ /documents/divisionhomepage/032181.hcsp (accessed 4 February 2005)

Domencich TA, McFadden D (1975) Urban travel demand: A behavioral analysis. North-Holland, Amsterdam

Druit S (2000) An introduction to PARAMICS. Available at http://www.sias.com/sias/ /paramics/articles/article1.html (accessed 19 September 2004)

Erlander S, Stewart NF (1990) The gravity model in transportation analysis. VSP, Utrecht

Evans SP (1976) Derivation and analysis of some models for combining trip distribution and assignment. Transportation Research 10: 37–57

Florian M, Nguyen S, Ferland J (1975) On the combined distribution-assignment of traffic. Transportation Science 9: 43–53

Florian M, Chapleau R, Nguyen S, Achim S, James-Lefebvre L, Galarneau S, Lefebvre J, Fisk C (1979) Validation and application of an equilibrium-based two-mode urban transportation planning method (EMME). Transportation Research Record 728: 14–23

Florian M, Nguyen S (1978) A combined trip distribution, modal split and trip assignment model. Transportation Research 12: 241–246

Garrett M, Wachs M (1996) Transportation planning on trial. Sage, Thousand Oaks

Hague Consulting Group (1997) A review of current world practice: a report for New South Wales Department of Transport. The Hague

Hague Consulting Group (2001) WinMINT 2.1. Available at http://www.hpgholding.nl/ /software/mint.htm (accessed 19 September 2004)

Hague Consulting Group, Accent Marketing and Research (1996) The value of travel time on UK roads. Prepared for the Department of Transport. Available at www.hpgholding.nl/publica/vot-uk.htm (accessed 19 September 2004)

Hensher DA (ed) (1994) Stated preference methods: Special issue. Transportation 21: 135–288

Highways Agency (1997) Design manual for roads and bridges, Volume 12, Traffic appraisal of road schemes. Section 2, Part 2: Induced traffic appraisal. The Stationery Office, London

INRO Consultants, Inc. (2004) EMME/2. Available at http://www.inro.ca/ (accessed 19 September 2004)

Jones PM (1979) New approaches to understanding travel behaviour: the human activity approach. In: Hensher DA, Stopher PR (eds) Behavioral travel modelling. Croom Helm, London

Jones PM, Dix MC, Clarke MI, Heggie IG (1983) Understanding travel behaviour. Gower, Aldershot

Kitamura R, Mokhtarian PL, Laidet L (1997) A microanalysis of land use and travel in five neighborhoods in the San Francisco Bay Area. Transportation 24: 125–158

Lam WHK, Huang H-J (1992) A combined trip distribution and assignment model for multiple user classes. Transportation Research 26B: 275–287

Mackie PJ, Wardman M, Fowkes AS, Whelan G, Nellthorp J, Bates J (2003) Values of travel time savings in the UK. Report to the Department for Transport. Institute for Transport Studies, University of Leeds, Leeds

Manheim ML (1979) Fundamentals of transportation systems analysis. MIT Press, Cambridge

Martin BV, Memmott 3rd FW, Bone AJ (1961) Principles and techniques of predicting future demand for urban area transportation. Research Report No. 38, M.I.T., Cambridge

Martin, Voorhees and Associates (MVA) (2004) CUBE. Available at http://www.mva-group.com (accessed 19 September 2004)

McFadden D (1973) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) Frontiers in econometrics. Academic, New York.

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

McFadden D (2001) Disaggregate behavioral travel demand’s RUM side: a 30-year retrospective. In: Hensher D (ed) Travel behaviour research: The leading edge. Pergamon, Amsterdam, pp 17–64

McFadden D, Train K (2000) Mixed MNL models for discrete response. Journal of Applied Econometrics 15: 447–470

Murchland JD (1970) Road network traffic distribution in equilibrium. In: Henn R, Kunzi HP, Schubert H (eds) Mathematical models in the social sciences 8. Anton Hain Verlag, Meisenheim am Glan, Germany, pp 145–183 (in German translation)

Oppenheim N (1995) Urban travel demand modeling, Wiley, New York

Ortúzar JD, Willumsen L (2001) Modelling transport, 3th edn. Wiley, Chichester

Quarmby DA (1967) Choice of travel mode for the journey to work: some findings. Journal of Transport Economics and Policy 1: 273–413

Rossi TF, Shiftan Y (1997) Tour based travel demand modelling in the US. Transportation systems 1997. In: Papageorgiou M, Pouliezos A (eds) Proceedings of the 8th IFAC/IFIP/IFORS symposium, Chania, Greece, Vol. 1. Pergamon, Oxford, pp 381–386

Standing Advisory Committee for Trunk Road Assessment (SACTRA) (1994) Trunk roads and the generation of traffic. The Stationery Office, London

Senior ML, Williams HCWL (1977) Model-based transport policy assessment; Part I: The use of alternative forecasting models, Traffic Engineering and Control 18: 402–406

Transportation Research Board (1995) Expanding metropolitan highways: implications for air quality and energy use. Special Report 245, Washington, DC.

US DOT (1977) An introduction to travel demand forecasting: A self-instructional text. Federal Highway Administration and Urban Mass Transportation Administration, Washington, D.C.

van Vliet D (1982) SATURN: A modern assignment model. Traffic Engineering and Control 23: 575–581

Wachs M (ed) (1996) A new generation of travel demand models: special issue. Transportation 2: 213–352

Wardrop JG (1952) Some theoretical aspects of road traffic research. Proceedings of the Institution of Civil Engineers, Part II 1: 325–378

Warner SL (1962) Stochastic choice of mode in urban travel: a study in binary choice. Northwestern University Press, Evanston

Williams HCWL (1977) On the formation of travel demand models and economic evaluation measures of user benefit. Environment and Planning A 9**:** 285–344

Williams HCWL (2004) Themes in the development and application of transport planning models. In Lee D-H (ed) Urban and regional transportation modeling. Edward Elgar, Cheltenham Northampton, pp 1–24

Williams HCWL, Ortúzar JD (1982) Behavioural theories of dispersion and the misspecification of travel demand models. Transportation Research 16B: 167–219

Williams HCWL, Senior ML (1977) Model-based transport policy assessment; part II: removing fundamental inconsistencies from the models. Traffic Engineering and Control 18: 464–469

Williams HCWL, Senior ML (1978) Accessibility, spatial interaction and the spatial benefit analysis of land-use transportation plans. In: Karlqvist A, Lundqvist L, Snickars F, Weibull JW (eds) Spatial interaction theory and planning models. North-Holland, Amsterdam, pp 253–287

Wilson AG, Hawkins AF, Hill GJ, Wagon DJ (1969) Calibrating and testing the SELNEC transport model. Regional Studies 3: 337–350

Wilson AG (1970) Entropy in urban and regional modelling. Pion, London

Wooton HJ, Pick GW (1967) A model for trips generated by households. Journal of Transport Economics and Policy 1: 137–153