1. Dynamic Traffic Assignment: Recent Advances and New Theories Towards Real Time Applications and Realistic Travel Behaviour (Editorial)

Francesco Viti and Chris M.J.Tampère

1.1 BACKGROUND

Dynamic Traffic Assignment (DTA) is a set of criteria through which the demand for mobility is distributed over time and space on a transport network. Underneath this synthetic definition, there is a wide range of models and theories, which have been developed with the aim of solving this important transportation problem. Accordingly, this problem has been interpreted and solved in many ways, among which the most popular is the fixed-point approach derived from the two equilibrium principles of Wardrop (1952). These principles translate, in the transport context, the economic and game-theoretical principles of cooperative and non-cooperative equilibria in games (Nash, 1951). Traditionally, traffic assignment has been developed exclusively in the static context, before the seminal works of Merchant and Nemhauser (1978a, 1978b). This approach is still widely accepted in planning and design problems, or in general when congestion dynamics are of minor importance. The need for models able to capture in a more realistic way the dynamic features of traffic has been acknowledged since the growing application of dynamic management strategies, real-time traffic control and information systems etc., and also because of the increasing congestion levels worldwide. The simplifying assumptions characterizing static assignment approaches (e.g., steady-state conditions, time independency of the demand and the travel costs) are under these conditions unacceptable as they would fail in capturing important dynamic aspects and in assessing the impact of some management measure. The goal of Dynamic Traffic Assignment is therefore to represent more correctly the dynamic character of traffic and to capture the temporal effects of congestion. It is thus not surprising that DTA formulations bring to assignment problems many challenges in terms of
mathematical and computational tractability. Furthermore, in traditional traffic assignment models drivers are assumed to be fully rational, and to have perfect information about the costs in the network, for any possible alternative for travelling and distribution of the travel demand. As a consequence, they choose the most convenient alternative of transport, and, if no change is observed in the system, they will not be better off by shifting to other alternatives (Deterministic User Equilibrium, DUE). The assumption of perfect information and perception of travel costs has been addressed as a major weak point in this theory from the early developments and, although computationally advantageous, it has been criticized as being rather unrealistic. To incorporate errors in drivers’ perception and imperfect information Daganzo and Sheffi (1977) introduced in traffic the principle of Stochastic User Equilibrium (SUE), whereby heterogeneities in perceived travel costs and travel choices are explicitly modelled. Thereby the principle of perceived costs was adopted successfully in the Dynamic Traffic Assignment context (Ran and Boyce, 1996). Despite the large interest shown in the last 30 years and its application to a variety of transportation problems, DTA theory is still relatively undeveloped, as thoroughly discussed in the excellent literature review of Peeta and Ziliaskopoulos (2001). Many approaches, models and algorithms have been proposed, e.g., to better suit the various application domains or to incorporate dynamisms and behaviours that cannot be explained with classical principles of equilibrium, and ultimately to trade off computational tractability with real observations.

Figure 1.1 Framework of Dynamic Traffic Assignment.

The paramount role played by assignment processes can be described with the help of Figure 1.1. The demand for mobility is determined at the socio-demographic level, i.e. by looking at the number and location of activities and
households, and their relative accessibility and connectivity. These parameters determine the demand from each origin and toward any destination, which is loaded onto the network through the available and convenient travel alternatives. DTA theory deals with the most likely distribution of flows, determined at the decisional level, onto the set of travel alternatives, i.e. which destination, mode of transport, departure time and route each traveller chooses depending on his/her objective and perception of travel costs. These flows are loaded onto the network: traffic propagates according to the imposed traffic rules at each link and node (operational level) and is therefore limited by the capacity of traffic networks. An excess of demand results in congestion and delays experienced by the drivers. As a consequence, drivers may reconsider their way of travelling (evaluation level) and shift to other alternatives of travelling (1) during one trip e.g., by acquiring new information en-route and updating their tactics (2) on future journeys, by changing their travelling strategies or (3) on a typically longer time-scale, changing place of origin or destination, or not travelling at all. Modelling and predicting tactical changes is crucial for instance in the application and assessment of, e.g., Advanced Travellers Information Systems (ATIS) or Dynamic Traffic Management (DTM) measures like route guidance or real-time traffic control systems. Strategic changes are obtained also due to infrastructural changes (enhanced capacity of the existing, or development of new, infrastructures and means of transport) or because of drivers' learning and adaptation processes that suggest changing their current strategy of travelling. At a larger scale, the costs of travelling and the consequent level of accessibility and connectivity of each origin–destination pair may determine mobility changes in the overall distribution of activities and demand. This is commonly considered in planning problems that analyse the effects of pricing schemes, infrastructure investments, etc.

The role of DTA is, in essence, to provide a functional relationship between the demand for mobility and the network supply. To specify this functional relationship, two elements are fundamental: (1) how travellers' perceive, and respond to, the costs related to mobility, i.e. their decision-making process, and (2) how transport networks are capable of coping with the demand for mobility, and how costs for mobility are generated, i.e. the network loading process. Typically, transportation problems require an integration of effects measured and/or estimated at both aggregate and disaggregate levels. Travellers' choices are the results of individuals' perception of travel costs, which also vary from user to user. On the other hand, travel costs are specified only once all users are loaded onto the network, especially in congested scenarios. Both elements are characterized by dynamic and stochastic properties, as will be described in the next sections, and by many sub-elements that contribute to determine these
properties. Looking at Figure 1.2, four choice modules typically describe the individual decision-making process of road users before and during a trip. They can select the most convenient destination for their purpose, which mode of transport, at what time to leave the place of origin and by which route.

**Socio-demographic level**

- Population
- Network topology
- Acquired information
- Acquired experience

**Decisional level**

- Demand for mobility
- Destination choice
- Departure time choice
- Modal choice
- Route choice
- Trip chaining
- Scheduling
- Multimodal
- Dep. time adjustments

**Operational level**

- Network loading
- Route flows
- Accidents, special events
- Rerouting
- Traffic management systems
- Network Capacity
- Traffic information systems
- External factors

**Evaluation level**

- Congestion levels
- Travel costs (time, reliability, accessibility)

**Figure 1.2** Generation, decision-making and network loading processes.
Typically this four-stage process is proposed in this sequence (sequential process), given the increasing sensitivity of each decision level to changes in the system (Cascetta, 2001). However, this assumption is not only mathematically and computationally advantageous but also implies obvious simplifications of the real choice process. Conversely, many approaches propose all choice levels should be dealt with simultaneously (simultaneous approach) as will be described in detail in the next section. The (individual) choice process of drivers is later used to translate the demand for mobility, often assumed given or derived from exogenous parameters, into origin-destination flows. Note that in the following we adopt a terminology used in assignment problems for road networks. Most of the concepts described can, however, be used essentially in the same way for other modes of transport. The (disaggregate) choice process of drivers is then aggregated into route flows, which are loaded onto the network and converted into link flows through the Dynamic Network Loading (DNL) process.

Traffic flows are, however, bound to the finite capacity of the roads and intersections, and, if they exceed these limits, congestion sets in, causing delays to the drivers. Road capacity is determined primarily by the geometry of the road sections, by the imposed traffic rules (speed limits, traffic priorities at intersections etc.) and the adopted traffic management and control strategies. However, this value of capacity is rarely constant, as it also depends on individuals’ driving behaviour (distance-keeping, desired speeds, gap acceptance, etc.) and on external or incidental factors, such as weather conditions, accidents, special events, etc. Thus, road and intersection capacities have a stochastic character as much as the demand for mobility (Brilon et al., 2005). Different drivers’ tastes and types, information media, experience acquired, road conditions, and so forth, determine the strong unpredictability of travel times experienced by each traveller and justifies the complexity of computational algorithms able to model these effects.

Due to the vast research area around DTA we focus on three aspects, which serve as an introduction to the chapters in this book and to some aspects that have recently emerged in this area.

1. **Solution properties of DTA:** Classical DTA approaches aim to provide a unique, feasible state of the system, given a number of assumptions on drivers’ behaviour and network characteristics. However, in reality, traffic hardly shows the same pattern every day, and changes are observed because of variations in daily (or weekly) activities and schedules, because of the dynamic effects of congestion and due to the variable choices and driving behaviour of drivers. Novel approaches propose looking at a distribution of feasible solutions, or at transient processes from known traffic and assignment solution states that evolve
to possible future states. The traditional equilibrium approaches and these new approaches will be introduced and discussed in section 1.2.

2. Travel choice behaviour: Classical DTA problems deal mainly with the consistency between route flows and costs. However, by dealing explicitly with individuals' decision making (i.e., which route, departure time, mode, etc.) DTA can be framed within more general contexts. Moreover, more realistic travel choice processes can be introduced (e.g., bounded rationality, choice under risk and uncertainty, etc.). Section 1.3 discusses these approaches.

3. Dynamic Network Loading models: To be able to model drivers' choices realistically, including their experienced travel times and congestion levels, the way traffic propagation is modelled is of great importance. However, many DTA applications rely on simplified traffic models, for the sake of mathematical and computational tractability, or to guarantee properties of existence, uniqueness, convergence and stability of the solution. However, consistency with the real dynamic and stochastic behaviour of traffic is far from being achieved. Section 1.4 describes the traffic models currently available.

1.2 SOLUTION PROPERTIES OF DTA

Traditionally, DTA has been formulated as a fixed-point problem, i.e. determine a 'self-consistent' point prediction of traffic volumes and costs, based on economic principles (Hazelton and Watling, 2004). Therefore, research in DTA has been initiated mainly around the concept of equilibrium, focusing on its existence, uniqueness, convergence and stability. Seminal papers that discussed these properties are, among others, those of Smith (1979), Tobin and Friesz (1988), Zhang and Nagurney (1996), Zhu and Marcotte (2000) for the case of route choice, Ben-Akiva et al. (1984), Arnott et al. (1987), Newell (1988) for the departure time choice, Huang and Lam (2002) for the simultaneous route-departure time choice, and Wie (1990), Wie et al. (2002) for the case of elastic demand. In the remainder of this section we categorize the different DTA approaches available in literature.

1.2.1 Analytical DTA Models

Analytical DTA models are concerned primarily with the macroscopic properties of traffic, i.e. they provide direct functional relationships between input parameters (origin-destination demand, link capacities, levels of information, management strategies etc.) and traffic states (link/route flows, travel times etc.). However, this approach is often characterized by restrictive assumptions, hardly acceptable in realistic networks (e.g., monotonicity and...
separability of the cost functions). DTA has been formulated analytically initially as a Mathematical Program (MP-DTA) by Merchant and Nemhauser (1978a, 1978b) and later elaborated by Ho (1980), who formulated the problem using linear programming, while Carey (1986, 1987, 1992), Janson (1991a, 1991b), and Birge and Ho (1993) extensively studied the mathematical properties of this approach for System Optimum (SO), deterministic and stochastic User Equilibrium (UE) respectively. MP-DTA formulations are supported by well established and standard solution algorithms. However this approach is restricted to a discrete time domain and requires very restrictive assumptions for the cost functions to guarantee existence and uniqueness of the solution. New approaches are under development, which make use of more realistic dynamic relations between flows and road capacities. For example, Ziliaskopoulos (2000) proposed using the Cell Transmission Model (CTM, Daganzo, 1994) for modelling the propagation of traffic in a MP-DTA problem.

An alternative to MP-DTA formulations is the Optimal Control formulation (OC-DTA). This approach was first proposed by Friesz et al. (1989) both for the SO and the UE cases in a single destination system. OC-DTA formulations overcome two limitations of MP-DTA. They can formulate the problem using continuous parameters, and link performance functions can be treated as control variables and they do not require the specification of a functional form for the costs. These two properties allow the investigation of existence and uniqueness properties under more general assumptions.

A third class of analytical DTA models is represented by Variational Inequality formulations (VI-DTA), which have considerable advantages compared to MP-DTA and OC-DTA formulations, as they combine equilibrium and optimization conditions in one formulation. Properties of this approach have been extensively analysed both in the static (Dafermos, 1980) and in the dynamic context (Friesz et al., 1993; Nagurney, 1993). Chen (1999) showed that VI-DTA can be successfully formulated for modelling the four-stage travel choice process. Later Bliemer and Bovy (2003) extended the VI-DTA formulation to account also for multiple user classes. However, computational tractability still remains an issue and application of these systems is done at the expense of simplified traffic performance models, typically developed under deterministic assumptions (e.g., Chabini, 2001; Bliemer et al., 2004). These simplified theories allow, however, a direct functional relationship between time-varying flows and costs, and have interesting mathematical advantages (Carey, 2006). Therefore, although more realistic traffic performance models are being introduced to better capture the dynamic and stochastic character of traffic networks, and especially due to the fast development of (micro)simulation techniques, simplified analytical models are still widely considered essential in DTA theory, as they are often
capable of explaining trends and properties of some applications, following the principle of parsimony. For example, in this volume, Smith (Chapter 3) uses basic linear cost functions to show the effectiveness of different control strategies.

1.2.2 Simulation-Based DTA Models

On a parallel track, simulation-based DTA models (SB-DTA) are being developed to fill the gap between computational tractability and realism. As pointed out by Peeta and Ziliaskopoulos (2001) this class of models is less concerned with the mathematical properties and with the consistency between estimated flows and costs in the network, while it focuses on the consistency between estimated flows/costs with those measured in the field. For this reason the key element characterizing SB-DTA models is the way in which traffic propagation is simulated. Simulation has been proposed on all three modelling levels, i.e. macroscopic (e.g., METANET, Messmer, 2000; SATURN, Hall and Van Vliet, 2002), mesoscopic (e.g., Cellular Automata, Nagel and Schreckenberg, 1992; DynaMIT, Ben-Akiva et al., 1998; DYNASMART-P, Mahmassani et al., 2001; DYNAMEQ, Florian et al., 2006) and microscopic (e.g., VISSIM, PTV, 2003; AimsunNG, Barceló et al., 1994; INTEGRATION, Van Aerde and associates, 2001). SB-DTA models are particularly suited for real-time applications and this justifies their growing application in practical studies. In this volume the simulation-based approach is adopted by Pel et al. (Chapter 15), Chen et al. (Chapter 17), Cipriani et al. (Chapter 18) and Behrisch et al. (Chapter 19). The work of Flöterod (Chapter 14) also uses a simulation approach, but differs from the other studies since it uses simulation also for modelling the travel demand.

The choice of which traffic performance model to adopt is not the only determinant to guarantee existence, uniqueness, convergence and stability properties of DTA solutions. An important contribution in this sense is also given by the travel choice models adopted, and in particular the route swapping. An opportune choice of the route swapping criterion can in fact be as important as the specification of an efficient DNL for convergence speed and stability of the solution. In this volume Mounce and Carey (Chapter 6) and Rahamim et al. (Chapter 7) discuss this issue and analyse the convergence rate of different route swap models. A discussion of these route swap processes follows in Section 1.3.

1.2.3 Dynamic Process Models

The above studies are all based on the principle of equilibrium and focus on finding its properties, i.e. stability, uniqueness, convergence. Recently, many studies have questioned the validity of this principle, arguing that equilibrium in dynamic traffic networks is rarely observable and transient behaviour
might be a more important aspect to incorporate in DTA. There is in fact not yet empirical proof that traffic flows follow the rules and constraints imposed by the equilibrium theory. Recent research prefers to view equilibrium as an attractor, i.e. a point that could be achieved if all conditions are kept constant or at least they remain such that convergence to the same attractor is expected. The Doubly Dynamic DTA formulation (DD-DTA) or Dynamic Process models, originally introduced by Horowitz (1984), adopt this view. In this approach the evolution over time towards an attraction point is important, as it may determine its convergence and stability properties. Evolution in these systems is modelled as a Markov Chain process, defining a unique probability distribution on the link traffic volumes as time-dependent processes from a known initial state. The properties and practical application of DD-DTA models have been shown in various studies, e.g., Cascetta (1989), Cascetta and Cantarella (1989), Watling (1999), Cascetta et al. (1991) and Cantarella and Cascetta (1995), Hazelton and Polak (1997), Hazelton (2002), Watling and Hazelton (2003), Cantarella and Velonà (2003), Hazelton and Watling (2004), Lo and Bie (2006), Nakayama (2006), Balijepalli et al. (2006) and Bie and Lo (2008). DD-DTA allows one to model convergence to attractors and to analyse equilibrium stability around these attractors. In this way, day-to-day and within-day dynamics are analysed in a unique framework. In this volume the Dynamic Process models approach is used to analyse the uniqueness and stability of signal control strategies in day-to-day processes (Cantarella, Chapter 2). An extensive review of dynamic process models is found in Watling and Hazelton (2003), together with a thorough discussion of equilibrium and non-equilibrium approaches.

The question whether it is necessary to investigate non-equilibrium or transient traffic behaviour is currently one of the main topics of discussion in this area. Equilibrium analysis has obvious advantages in terms of computational and mathematical tractability. Equilibrium approaches are also found to be more interpretable and understandable for many applications. Nevertheless it should be discussed whether it always makes sense to look for an equilibrium point in all applications of DTA, especially in those where boundary conditions and perturbations prevent (unique) equilibrium conditions from ever occurring. Calculating a distribution of the possible state of the system depending on the observed current state may be a promising direction of research, but implies the difficult task of predicting the variability of traffic states, which requires insight into the nature and magnitude of perturbations and the parameters that caused these perturbations. Better understanding is needed on how to deal with recurrent, observable traffic variability (e.g. daily fluctuations of demand and supply) and with non-recurrent, unpredictable traffic variations (incidents, special events). Moreover better insight is needed on how drivers respond to both recurrent and non-recurrent congestion (e.g., by rerouting, or by day-to-day travel
alternative shifts), and how they perceive and value these uncertainties, as will be discussed in section 1.3.

1.2.4 Game-Theoretical Models

The equivalence between Wardrop and Nash equilibria (see Haurie and Marcotte, 1985), has opened new perspectives for the explicit modelling of day-to-day and within-day fluctuations, as well as transient effects such as learning and updating information, using techniques already developed and applied in the field of Game Theory (Friesz et al., 2006). Concepts from Game Theory have already been applied and integrated with DTA to, for example, elegantly formulate bi-level problems such as traffic control optimization (Chen and Ben-Akiva, 1998), or network design problems such as dynamic road pricing (Joksimovic et al., 2005), or to model multiple user classes (Ramadurai et al., 2008). Very recently, Yang et al. (2008) demonstrated game theoretical models to be a very powerful approach for solving DTA, to incorporate many different aspects, such as a mixture of competitive and cooperative drivers, information levels, different network management strategies etc. By using this approach, they could prove the existence of different equilibrium points for different types of road users. In this volume, Iryo (Chapter 4) finds that the existence or non-existence of (Nash) equilibria could be associated to the network topology and that existence and uniqueness can be guaranteed under specific network layouts.

1.3 TRAVEL CHOICE BEHAVIOUR

In this section we discuss the issues arising when more choice levels or more refined models are incorporated in DTA formulations.

1.3.1 Integration of Different Choice Levels

Traditionally, DTA theory dealt with the consistency between route flows and route costs. Within this view, assignment has been mainly associated with route choice, while all other choice levels have been assumed exogenously given in the problem. However, many applications require other choice levels to be dealt with simultaneously to route choice. Empirical and theoretical findings have shown, for instance, that route and departure time choice have comparable dynamisms, and switching decisions should be modelled in a joint process, especially to estimate the effects of e.g. congested networks, traffic information, dynamic pricing schemes, travel time variability (e.g., Mannering, 1989; Caplice and Mahmassani, 1992; Hatcher and Mahmassani, 1992; Khattak et al., 1995; Mahmassani and Liu, 1999; Szeto and Lo, 2004; Lim and Heydecker, 2005; Viti et al., 2005). Joint modelling of route-
departure time choice also allows a more correct allocation of the total daily demand (Heydecker and Addison, 1998), and to consider trip-chaining strategies (e.g., Lam and Huang, 2002; Polak and Heydecker, 2006).

DTA approaches that integrate different choice levels have been initially proposed in a route choice-like manner. That is, for example, the case with the hypernetwork or supernetwork approach (Sheffi and Daganzo, 1979), where different choice levels are modelled as alternative routes. Van der Zijpp and Lindveld (2001) proposed a simultaneous route–departure time choice model using a hypernetwork approach to model the peak spreading phenomenon in congested networks, while Benjamins et al. (2002) applied this methodology on multimodal transport networks, and recently Ramadurai and Ukkusuri (2009) proposed using supernetworks for modelling activity–travel behaviours. Despite its computational advantage, the hypernetwork approach implies many simplifications (separable cost functions, time discretization etc.). Moreover, by incorporating different choice levels the generation of route alternatives is often too large even for small networks (Bovy, 2006).

Special attention should be drawn to the integration between DTA and demand for mobility, i.e. the origin–destination flows. Despite the fact that a large part of demand estimation models can be categorized as non-DTA-based (gravitational-type models, estimation using proportional assignment, direct sampling methods from behavioural surveys, etc.), it is believed that dynamic demand estimation procedures should be seen as inverse DTA problems (Bierlaire, 2002). Traditionally, DTA-based OD estimation approaches in literature have followed a bi-level structure, whereby the equilibrium assignment is at the lower level while a minimization of the distance between estimated and measured flows is solved at the upper level. The problem is that the relationships between link flows/costs and route choices when both are allowed to vary dynamically are rather complicated. For this reason, in practice all approaches which incorporate these two dynamic factors follow a simultaneous or sequential, iterative approach, where the demand at each time period is preliminarily determined by fixing the travel costs and later these costs are updated to become consistent with route choices. The process becomes therefore a sequential correction towards a consistent solution where the measured time-varying traffic counts are justified by time-dependent route flows represented by time-varying OD tables. Examples of such models are those proposed by van der Zijpp (1996) and Ashok (1996). The main challenge in integrating demand estimation with DTA models remains the underdeterminedness of the problem, which implies that the reliability of the solution strongly depends on the validity of the adopted behavioural and traffic models, the complexity and size of the networks, the number of available measurements and on the adopted
optimization techniques. In this volume Frederix et al. (Chapter 10) and Cipriani et al. (Chapter 18) discuss these issues.

The need to incorporate travel behaviour explicitly in a DTA framework to capture the dynamics of traffic in a realistic way is increasingly acknowledged by the transportation community. Within-day dynamics are in fact closely related to departure time choices of individual drivers, as well as their activity–travel behaviours, which also affects day-to-day dynamics as many activities are scheduled on a weekly basis (Viti et al., 2010). In this volume, for instance, a simultaneous route–departure time model approach is proposed by Friesz et al. (Chapter 5) to integrate day-to-day and within-day dynamics, Adnan (Chapter 13) proposes incorporating activity scheduling in DTA to account for flexible working hours and trip chaining, for instance, while Ge and Stewart (Chapter 16) investigate the effects of different dynamic pricing schemes on peak period congestion.

1.3.2 Choice under Uncertainty

Following a trade-off between simplicity and realism and, as already said, to guarantee properties of existence, uniqueness, convergence and stability, popular choice models adopted in DTA are often simple, ranging from the deterministic all-or-nothing assignment to stochastic Logit-type models. Although this choice is certainly advantageous in terms of mathematical tractability, it is rather questionable in practical applications, especially in cases where the drivers' response is fundamental to the effectiveness of a management policy or when we want to predict the effects of an information system. If considering different choice levels is widely acknowledged as a fundamental step to capture the dynamic character of traffic, more sophisticated choice models are needed to relax the simplifying assumptions of fully rational and informed road travellers. When people face a decision that yields an uncertain result, the choice is sometimes not fully rational or optimal. Variability of costs, conflicting objectives, competing alternatives, heterogeneous risk attitudes etc., make decisions often unpredictable. Decisions are made with incomplete information about the real travel costs or under time constraints, or even depend on the traveller's emotional state at the time the decision is made. Some people rely more on personal past experiences than information and sometimes it is the other way around.

Empirical findings assessed the importance of risk perception and irrational behaviour under uncertain costs (e.g., Avineri and Prashker, 2003). In modelling travellers' utility, some authors have already attempted to include a component that represents a degree of uncertainty. Small (1982), Mirchandani and Sorosh (1987) and van Berkum and van der Mede (1992) proposed including the standard deviation of experienced travel times in the utility function as a measure of uncertainty. Luo and Lo (2003) proposed the
use of the budget time, which depends not only on the standard deviation of travel times but also on the probability of being late weighted by the individual risk attitude. Lam and Small (2001) showed, however, that the 90th percentile is a better measure for reliability than the standard deviation. Some other studies criticize the use of standard probabilistic methods, and suggest a more radical shift towards new utility concepts, inspired by psychological theories of choice under risk and uncertainty (e.g., Von Neumann and Morgenstern, 1944; Kahneman and Tversky, 1979). Katsikopoulos et al. (2002) and Viti et al. (2005) found with route choice experiments that drivers generally do not follow the linear expected utility principles when comparing certain with uncertain outcomes. Non-linear utility theories, which deal with the influence of risk attitude and perception of uncertainty in travellers’ decision-making, have been proposed recently (e.g. Avineri and Prashker, 2004; de Palma and Picard, 2005; Avineri, 2006). However, first steps have highlighted the substantial complexity of tuning such non-linear models (Avineri and Bovy, 2008).

1.3.3 Learning, Habit and Limited Rationality

As mentioned in the previous section, for the sake of simplicity, interpretability and mathematical tractability, the adopted route choice processes in DTA have initially been very simple, and swapping rules have been chosen more for their computational advantages than for their realism (e.g., the traditional Method of Successive Averages, MSA). The importance of capturing the effects of dynamic management and information systems suggests that modelling transient periods might be as important as assuring the existence, uniqueness and stability of an equilibrium.

For this reason, attention has been given recently to learning mechanisms governing day-to-day processes as factors influencing the dynamics and determinants for convergence to (one or another) equilibrium. Horowitz (1984) made one of the first contributions to this field by modelling the mean perceived travel cost by a weighted average of the realized costs experienced in past days. Information was integrated later in this framework by, e.g., van Berkum and van der Mede (1992), Polak and Heydecker (2006), Jha et al. (1998), Chen and Mahmassani (2004) and Chorus et al. (2008). The interaction between past and actual information and experience has been identified as a key issue for understanding day-to-day choice processes. Ben-Akiva et al. (1991) studied the users’ response to information extending their previous research in dynamic network modelling. Srinivasan and Mahmassani (2003) considered various information and socio-economic attributes and estimated their effects in a day-to-day context. The type of information (i.e. descriptive vs. prescriptive information), its correctness and completeness, the availability of feedback information and some generic information effects
like under- and overestimation errors all proved to have significant effects on route switching behaviour.

Apart from more sophisticated learning mechanisms and models for including uncertainty in decision-making processes, questions are to what extent road users perceive uncertainty and differences in utility between travel alternatives, and in particular when they can be considered to be satisfied with their choice and stop looking for a more convenient alternative. Stopping criteria in DTA models have been chosen mostly for the sake of computational advantages, while more psychological concepts may support theories of satisfaction, habit and bounded rationality. The concept of satisfaction was firstly introduced in travel demand forecasting problems by Daganzo (1979) and later elaborated and generalized by Cantarella (1997). Its simplicity has clearly determined its popularity in travel behaviour analysis, but despite its mathematical advantages its property of dependence of its value from the number of travel alternatives seems a strong drawback. For this reason theories of habit and bounded rationality have recently been introduced. For example, in modelling drivers’ route and departure time choices, Mahmassani and Jou (1998) proposed a model that explicitly includes bounded rationality and habit in route choice by specifying indifference bounds around the currently chosen route costs. They assumed that as long as the outcome of the best route choice falls within an indifference bound, the traveller is not inclined to change his/her route choice. Nakayama (2006) later studied the advantages of bounded rationality for the stability of traffic flows in equilibrium. New modelling techniques, based on empirical data, aiming at modelling the limited rationality of drivers are being introduced (e.g., fuzzy-logic, van Zuylen and Kikuchi, 2001; Dell’Orco and Kikuchi, 2005; Dell’Orco et al., 2008) and embedded in the assignment process (e.g., Henn, 2000; Liu et al., 2003). These approaches are the more appealing since they deal elegantly with uncertainties and can be driven by behavioural data (Kikuchi and Pursula, 1998).

1.4 DYNAMIC NETWORK LOADING MODELS

As mentioned previously, traffic performance and the way traffic flows are propagated onto the network is an important element in DTA problems as much as the travel choice models adopted. It is not surprising that considerable research effort has been dedicated to this element as well.

A schematic representation of the network supply is commonly defined by links and nodes. Each characteristic contributes to the traffic performance in the network. Link performance is primarily determined by the time a vehicle takes to traverse the link, i.e. its driving time, while at nodes it is more useful
to determine the delay or waiting time rather than the driving time. Traffic performance models can generally be subdivided into macroscopic, mesoscopic and microscopic, depending on the level of detail chosen for the input and output parameters. In this chapter we do not give an overview of time-independent models, adopted in static assignment and planning applications. These models have been proved to be inadequate in DTA applications since the seminal works of Merchant and Nemhauser (1978a, 1978b). We focus instead on time-dependent models and traffic flow propagation models. Moreover, we do not cover mesoscopic and microscopic approaches, as they are not (yet) used extensively in DTA problems – except in some simulation-based approach – due to their computational complexity. For a more extensive overview of such models we refer to Hoogendoorn and Bovy (2001), Szeto and Lo (2006) and Maerivoet (2006).

The difficulty in deriving instantaneous measures for each vehicle and the need for computationally efficient models in DTA frameworks motivate the more extensive use of models based on average conditions, i.e. a macroscopic approach. In these models traffic states are by descriptive parameters like average flows, densities and speeds. Traffic states are then estimated without consideration of each single vehicle trajectory but they are representative of all vehicles in the section, following well-known fundamental relationships between speed, flow and densities (e.g., the fundamental diagram proposed by Greenshields, 1935), and spatio-temporal properties to define the traffic state dynamics (e.g., the hydrodynamic theory of Lighthill and Whitham, 1955, and Richards, 1956, known as LWR theory). Based on the way traffic flow dynamics are modelled, macroscopic models adopted in DTA problems can be categorized in whole-link models, and traffic flow propagation models. In the following of this section we describe these two approaches.

1.4.1 Whole-Link Models

In whole-link models, the propagation of flows through a link is described by relationships between whole-link variables, e.g., travel times, number of vehicles, inflow and outflow rates etc. (Carey and McCartney, 2002). We can distinguish two categories of whole-link models: (1) those based on the estimated number of vehicles in a link to calculate the expected vehicles outflow rates (link exit functions, see e.g., Merchant and Nemhauser, 1978a, 1978b; Carey, 1986, 1987; Friesz et al., 1989; Wie et al., 1995; Carey and McCartney, 2002; Carey et al., 2003) and (2) those estimating the expected travel time per unit of distance as function of the density per unit of time (link performance functions, see e.g., Janson 1991a; Ran et al., 1997; Chen and Hsueh, 1998).

Whole-link models are characterized by serious drawbacks. They suffer from limited realism, are difficult to be specified and they typically violate
New Developments in Transport Planning

the First-In-First-Out (FIFO) condition (Carey, 1986, 1987), which is essential to guarantee the existence of DTA solutions. Moreover, in some applications the assumption of vehicles homogeneously distributed in the link may be a strong defect, especially in networks where the dynamics of queues are important. Carey and McCartney (2002) noted, however, that in cases of sufficiently short links and opportune time discretization then the whole-link model results can be considered acceptable. The FIFO violation limitation was overcome by Carey (2004a, 2004b).

Advantages of whole-link models are obviously their relative simplicity and monotonicity, which make them particularly suited for theoretical studies on the mathematical properties of existence convergence, uniqueness and stability. However, they have been found weak in the way they deal with queuing and especially in the propagation of queues on the network. It is for this reason that when the effects of queuing are analysed, especially in the departure time adjustment problem, the use of the so-called bottleneck models is often preferred. These models do not keep track of the physical extension of the queues (they are for this reason often referred to as vertical queues, or point-queue models), but they are interested only in the expected delay or waiting time. This approach has been adopted for instance by Tsubota and Kuwahara (Chapter 12) and by Ge and Stewart (Chapter 16).

To give physical properties to queues, spatial or horizontal queues have been proposed. These models often subdivide the link into two elements: a part where vehicles drive freely and one where vehicles are queued up. There are, however, two difficulties in using this approach: (a) the queue length may change while traversing the link, and (b) the outflow capacity may change while traversing the link (Bliemer, 2006). However, these parameters are known only when travel time is realized, i.e. only once a vehicle has exited the link. This implies that horizontal queuing approaches are particularly weak to model variable queues.

1.4.2 Traffic Flow Propagation Models

Frederix et al. (Chapter 10, this volume) show the significant errors caused by vertical and horizontal queues in DTA-based OD estimation problems, and the outperforming of traffic flow propagation models in capturing congestion dynamics and the effects of spillback. Traffic flow propagation models overcome these limitations of whole-link models, but at the expense of a certainly higher mathematical and algorithmic complexity. These models describe the physical propagation of flows on the links and across the nodes. The traditional approach is to consider traffic as a compressible fluid and to control the variation of flow, density and speed using the conservation law of fluid-dynamics. Lighthill and Whitham (1955) and Richards (1956) developed and applied a theory for vehicular traffic, initiating the so-called
first-order LWR models, or Kinematic Wave Theory (KWT) models, since
the main elements in these models are the forward and backward shockwaves
emerging when a change in one of the three fundamental parameters is
observed. Therefore the crucial difference between whole-link models and
KWT models is that homogeneity is not assumed within the link, but between
the kinematic waves. Some research criticized the assumption of homogeneity
within those regions delimited by the shockwaves. Higher-order models have
later been proposed to resolve this shortcoming, arguing that in this way
phenomena like stop-and-go waves, queue discharge and traffic hysteresis
could be better captured (e.g., Payne, 1971; Helbing, 1996; Zhang, 1998;
Messmer, 2000). These models have been criticized by Daganzo (1995),
highlighting several physical flaws. However, Papageourgiou (1998) pointed
out that second-order models outperform first-order models in estimating the
effects of ramp metering control.

Numerical solutions of the KWT models applied in DTA problems are the
Cell Transmission Model (CTM, Daganzo, 1994), which was first proposed
in a DTA framework by Ziliaskopoulos (2000) and Szeto (2003), and
recently extended by Sumalee et al. (2009) in a stochastic fashion to account
for traffic state uncertainties, and by Kalafatas and Peeta (2007, Chapter 11 in
this volume) in a graph-theoretic approach to better capture congestion
phenomena, and the Link Transmission Model (LTM, Yperman et al., 2006;
Gentile et al., 2007), which adopts the simplified KWT model based on
cumulative flows proposed by Newell (1993). LTM models outperform CTM
models in their computational efficiency, since links are not subdivided into
discrete cells, while keeping consistent track of kinematic waves. This way,
spillback effects in congested networks can be captured, and the dynamics of
queues are modelled in a more realistic way, as shown in dynamic OD
estimation applications by Frederix et al. (Chapter 10, this volume). Corthout
et al. (Chapter 9, this volume) propose the extension of LTM to account for
stochasticities caused by incidents.

It seems that the main challenge for the near future in flow propagation
modelling is also in the integration of link performance and the traffic
behaviour at nodes. Tampère et al. (forthcoming) clearly addresses this
problem, highlighting the difficulty of capturing the joint effects of spillback
and traffic priority at nodes. Gentile (Chapter 8, this volume) presents a more
general formulation of LTM which applies to any concave fundamental
diagram and node topology, but generalization is done at the expense of
modelling simplifications. If current DNL models seem to be adequately
developed for predicting traffic on motorway networks, the lack of valid node
models, together with the complexity of route selection criteria (both pre-trip
and en route) impedes a satisfactory application of DTA in the urban context.
1.5 CLOSURE

This chapter has given a (certainly not exhaustive) overview of the research elements characterizing DTA theory, and we hope that it may serve the reader as a useful introduction to the chapters presented in the remainder of this volume. We have tried to cast the following chapters in this state-of-the-art review, giving particular emphasis to the most interesting new emerging theories, while at the same time keeping a link with the classical vision of assignment processes.

REFERENCES


Brilon, W., J. Geistefeldt and M. Regler, (2005), ‘Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity’. In H.S. Mahmassani (ed.) 16th International Symposium on Transportation and Traffic Theory (ISTTT16), University of Maryland, College Park, Maryland, Elsevier.


New Developments in Transport Planning


Viti, F., E.A.I. Bogers and S.P. Hoogendoorn (2005), ‘Day-to-day Learning under Uncertainty and with Information Provision: Model and Data Analysis’, Presented at the 16th International Symposium of Transport and Traffic Theory (ISTTT16), University of Maryland, Maryland.


