Towards Disaggregate Dynamic Travel Forecasting Models

Moshe Ben-Akiva\textsuperscript{1}, Jon Bottom\textsuperscript{2}, Song Gao\textsuperscript{3}, Haris Koutsopoulos\textsuperscript{4}, Yang Wen\textsuperscript{1}

\textsuperscript{1}\textit{Massachusetts Institute of Technology}
77 Massachusetts Avenue, Room 1-181, Cambridge, MA 02139, USA (Tel: 1-617-253-5324, Fax: 1-617-253-0082, email: \{mba,wenyang\}@mit.edu)

\textsuperscript{2}\textit{CRA International}
200 Clarendon Street T-33, Boston, MA 02116, USA (Tel: 1-617-425-3392, Fax: 1-617-425-3132, email: jbottom@crai.com)

\textsuperscript{3}\textit{Caliper Corporation}
1172 Beacon Street, Newton, MA 02461, USA (Tel: 1-617-5274700, Fax: 1-617-527-5113, email: song@caliper.com)

\textsuperscript{4}\textit{Northeastern University}
437 Snell Engineering Center, Boston, MA 02115, USA (Tel: 1-617-373-4635, Fax: 1-617-373-4419, email: haris@coe.neu.edu)

Abstract

We argue that travel forecasting models should be dynamic and disaggregate in their representation of demand, supply and supply-demand interactions, and we propose a framework for such models. The proposed framework consists of disaggregate activity-based representation of travel choices of individual motorists on the demand side integrated with disaggregate dynamic modeling of network performance, through vehicle-based traffic simulation models, on the supply side. The demand model generates individual members of the population and assigns to them socioeconomic characteristics. The generated motorists maintain these characteristics when they are loaded on the network by the supply model. In an equilibrium setting, the framework lends itself to a fixed-point formulation to represent and resolve demand-supply interactions. The paper discusses some of the remaining development challenges and presents an example of an existing travel forecasting model system that incorporates many of the proposed elements.
Keywords: Disaggregate, dynamic, travel forecasting

1. Introduction

Transportation systems are highly complex, and for this reason planners and engineers often use computer models to analyze and design possible system improvements. Travel forecasting models are commonly used \([1][2]\) to predict the travel flows and conditions that result from the interaction between the demand for travel by transportation system users and the travel options that the transportation system supplies to its users. In general terms, travelers’ decisions (with respect to where, when and how to travel), as they relate to the service provided by different travel options, will be referred to as demand, and the response of the transportation system (in terms of travel time, cost, reliability and other service attributes) to a given level of demand will be referred to as supply. As Figure 1 illustrates, travel forecasting models represent supply and demand and incorporate a mechanism to determine the outcome of their interactions. These models simultaneously determine both the performance of the network and the response of its users, thus enabling planners and engineers to investigate how potential interventions will affect transportation system flows and service conditions.
The conventional framework

Early travel forecasting models, originally developed in the mid-1950s \[3\] and taking a standard form by the late 1960s, were designed to predict steady state flows and conditions over a representative analysis time interval such as a peak period; for this reason, they are called static models. Their original applications were to analyze the effects of large transportation system capacity additions, such as building new roads. Such projects can produce effects over a long period of time, and the totality of these effects must be considered in deciding whether a project is worthwhile, so the models are often used in these analyses to make long-term forecasts. Because of the inherent degree of imprecision in such forecasts, and because predictions of steady state flows and performance are sufficient to evaluate such large-scale interventions, static models are well suited to this task. These models are still very widely applied, although sometimes to problems for which they are not well suited.
We will refer to travel forecasting models based on these early principles as *conventional* models, recognizing that particular model system implementations may incorporate some differences with respect to the standard form. Nonetheless, the term conventional is appropriate, because the overwhelming majority of travel forecasting models in use today continue to follow the general design.

Conventional travel forecasting models are characterized by static, aggregate and deterministic supply and demand modules and relationships. More specifically:

- Travel demand is predicted based on the aggregate characteristics of geographic areas called zones and is defined in terms of trips made directly from one zone to another without intermediate stops. Conventional demand models predict the steady-state level of demand that would be maintained over an analysis period of one or more hours in the presence of steady state network conditions. A few user classes at most are considered, with each user class being characterized by distinct travel behavior (e.g. cars and trucks). The specific relationships used to predict travel demand as a function of zone characteristics and network supply are generally analytical (closed form mathematical formulas or tables) and deterministic.

- Network supply reflects the steady-state network conditions that result from a steady-state traffic flow during the analysis period. Because they are aggregate, conventional supply relationships do not consider individual vehicles; and because they are static, they do not consider the finite time required for traffic to move along a link or from one link to the next along a path. The specific relationships used to predict network conditions (e.g.
travel time or cost) as a function of flow are generally analytical and deterministic.

- Supply-demand interactions are assumed to result in a steady-state equilibrium situation, where travelers acting independently would have no incentive to change their decisions. The outcome of the supply-demand interactions, in terms of traffic flows and network conditions, reflects the aggregate, static and deterministic nature of supply and demand models, and the trip-oriented definition of travel. (Even the so-called stochastic user equilibrium is in fact a deterministic concept.)

Travel forecasting models can capture supply-demand interactions at various levels of time scale, and the choice of scale dictates which choices are endogenous to the forecasting model. This includes household choices, such as where to work, where to live, the number of automobiles to own, which are not themselves travel choices, but are closely related to them. (For example, residential and employment location choices determine the origins and destinations of home-based work trips.) Travel forecasting models with long-term supply-demand interactions make the above choices dependent on network performances generated from the supply side. Some travel choices are of shorter term, such as whether to travel and where to travel for non-work trips, and mode choice and departure time choice for both work and non-work trips. Travel forecasting models with supply-demand interactions at this level treat long-term choices as exogenous input, while making network performance an influencing factor of these choices.

Path choices are of even shorter term, and a model that captures the supply-demand interactions at this level is a traffic assignment model, which takes as input longer-term
choices (trips in a given time period by origin, destination and mode) and performs several distinct functions: it implements the supply relationships, represents travelers’ path choice behaviors, and resolves the interaction between supply and the path choice component of demand; its outputs are traffic levels and network conditions. Interactions between supply and demand components other than path choice may be resolved via “feedback” of the traffic assignment model outputs to the corresponding longer-term demand sub-models, iteratively invoking the longer-term demand sub-models and the traffic assignment model until a satisfactory indication of convergence to equilibrium is attained. (The various demand sub-models together with a traffic assignment model are often called the four-step model system. This is by far the most commonly applied travel forecasting model system framework in the world.)

Path choices can be further classified by time scale: those based on recurrent network conditions and those based on non-recurrent network condition, such as incidents, work zones, etc. In the former case, the result of supply and path choices is usually conveniently represented by equilibrium conditions, while in the latter case it seems implausible to talk about equilibrium.

**Modeling advances**

Conventional travel forecasting models continue to preserve these general features many decades after their original development. However, during this period research into the individual supply and demand components of travel forecasting models has resulted in considerable advances in the sophistication and capabilities of these components. The
evolution of both supply and demand models has proceeded in several directions [4]:

- From *static* to *dynamic* consideration of time: models are increasingly focused on capturing the temporal aspects of traveler behavior (choice of departure time, reaction to time-dependent conditions and information) and on predicting the variations in network flows and conditions at a relatively detailed time resolution;

- From *aggregate* to *disaggregate* representation of travel: models are increasingly focused on representing individual travelers and vehicles to capture the heterogeneity of their decisions and movements. This, in turn, has led to decreased reliance on analytical and closed form mathematical models, and increased use of simulation models, where traveler and vehicle behavior are implicitly represented through logic incorporated in the simulation model. Related to this, there has been increased recognition that deterministic modeling relationships may not be appropriate in all circumstances, with increased interest in accounting for model stochasticity.

During this period, supply-demand interaction models have also evolved, although to a lesser extent. The principal direction of progress has been from a static to a dynamic representation of time [5]. In the mid-1970s, research began on traffic assignment models able to represent the time variations in traffic flows and conditions that result from changing levels of travel demand, changes in network capacity (reduced, for example, by a vehicle breakdown or a crash), and the finite speed of vehicles moving over the network. These are known as dynamic traffic assignment (DTA) models, and research in this area is still very active.
DTA models have found application in operating and managing transportation networks, due in part to advances in technologies for collecting and disseminating traffic data in real time. It is now possible, for example, to provide travelers with real-time information messages to help them make better travel decisions, and to control elements of the transportation system, such as traffic signals or ramp meters, to provide improved system performance. Conventional traffic assignment models are of little help in providing these capabilities because the models represent steady-state conditions. The challenge now is to model and design transportation system interventions that affect the short-term dynamics of transportation system flows and conditions, and this is precisely what DTA models are designed to do.

However, much of the development in DTA modeling has taken place independently of and remained separate from the more general advances mentioned above in supply and demand modeling. Most notably, travel forecasting models still retain an aggregate representation of travel, and have not followed the evolution of demand and supply models towards increasingly disaggregate representations and relationships. Travel forecasting models are thus not fully benefiting from available improvements in supply and demand modeling – improvements that would enable them to address a wider range of transportation system interventions, and with greater modeling fidelity.

We attribute this situation in part to lack of a travel forecasting modeling framework able to accommodate both dynamic and disaggregate features of supply and demand models. In this paper, we propose a framework for such models.
In the sections that follow, we consider the evolution of the modeling of transportation demand, supply and their interactions, highlighting their advances compared to the conventional modeling approach. This discussion leads into a presentation of a framework for disaggregate dynamic travel forecasting models. Existing model systems are next examined. Final sections summarize some of the issues that must be confronted in moving disaggregate dynamic models into mainstream practice, and present the conclusions of this paper.

2. Evolution of demand and supply modeling

Demand models

Demand models represent user travel choices (e.g. to travel or not, at what time, to what destinations, by what modes, on what paths) as these are affected by network conditions. Conventional models predict total demand for travel in a given time period by a user class from an origin zone directly to a destination zone by a particular mode. In conventional models, the prediction of user flows from origin to destination by mode is carried out by a number of successive sub-models that account for the socio-economic characteristics of the origin and destination zones, and for the time or cost required to travel between them. Path choice is conventionally handled in the traffic assignment model.

Travel demand models have been the subject of continual research and development over many decades. The following paragraphs briefly describe major directions in the evolution of travel demand modeling, in the light of the trends identified above.
From static to dynamic. While travel demand models are not strictly speaking dynamic, there has nonetheless been a clear trend towards increasing incorporation of time dependency in model inputs and outputs.

On the one hand, models increasingly include explanatory variables that reflect time-varying network conditions such as dynamic tolls, unexpected congestion and real-time travel information. On the other hand, development of demand models that are able to represent a traveler’s choice of trip departure time is an area of significant research and development activity. This capability is essential in order to represent traveler response to time-varying transportation system attributes (such as congestion levels) and policies (such as congestion pricing) through effects such as peak spreading.

From aggregate to disaggregate. Conventional travel demand models are aggregate in the sense that they directly predict the total amount of travel by user class from one zone to another. Only a few distinct user classes are typically considered. These models are statistically estimated from data on zone-to-zone flows, zonal socio-economic characteristics, and the quality of transportation service offered between zones. However, when travelers within a zone are heterogeneous, this procedure can produced biased results. Moreover, much travel data is collected at the household or individual level, so a more efficient analysis, involving a better set of explanatory variables, can be carried out directly using disaggregate (i.e. individual traveler-level) data and model relationships. These factors led to the development of disaggregate travel demand models starting in the late 1960s and early 1970s [6][7][8][9]; research in this area is still very active. Since many aspects of individual traveler
behavior are discrete choice (e.g. choice of mode or path), discrete choice analysis methods have played a prominent role in this development\textsuperscript{[10][11][12]}. 

From analytical to simulation-based. Conventional demand models typically apply analytical relationships, based on closed form mathematical equations or table lookups, to predict total zone-to-zone travel demand. As modeling has shifted to a more disaggregate approach, with the improved availability of data and the rapid development of computer technology, simulation-based demand models have become more widely adopted. Simulation-based demand models predict the response of individual travelers based on individual-level characteristics; aggregate demand results from the cumulative response of all the individuals. In effect, each traveler becomes a distinct user class. A simulation model may include sub-models that involve mathematical relationships, but is not confined to analytical expressions of this type and may implement any logic. Simulation has proven to be a flexible, versatile, and comprehensive tool for demand modeling.

It should be noted that simulation models could require extensive amounts of highly detailed data: for example, data on the characteristics of every household or individual in an urban area. Travel surveys or population census data only provide some of the required data, for example aggregate distributions of population characteristics. Accordingly, analysts have developed new tools, called population generators or synthesizers\textsuperscript{[13][14][10][15][16]} to generate the disaggregate data necessary for micro-simulation or sample enumeration in forecasting. Population generation methods generate detailed household or traveler characteristics in a way that is consistent with known aggregate population or travel characteristics. For
instance, TRANSIMS includes a population synthesizer that generates a set of synthetic households and household members that match aggregate demographic data \(^{[15]}\), while the DynaMIT demand simulator disaggregates known origin-destination (O-D) matrices into the individual drivers who make the corresponding trips \(^{[17]}\); similar approaches have also been used earlier in the travel demand model system of the San Francisco area \(^{[10]}\) and the Dutch national model system \(^{[18]}\).

*From trip-based to tour-based and activity-based.* Conventional demand models consider the basic unit of analysis to be a trip made directly from one location (zone) to another. This definition does not recognize that travelers may make intermediate stops, resulting in more complex travel patterns called tours. In addition, the conventional definition of a trip ignores the fact that travel is often a by-product of travelers’ participation in other activities, so that understanding travel demand requires understanding the activities that trigger it \(^{[1][19]}\).

Tour-based and activity-based travel demand models have been developed to overcome these limitations. The former explicitly chain trips into tours. The latter derive trips and tours from other activities of individuals and households, taking account of logical, physical and other constraints, and so can represent the effects of a variety of transportation policies \(^{[20][21]}\). In either case, trips are no longer treated as point-to-point movements resulting from independent decisions, and behavioral fidelity is enhanced. Both types of models have proven amenable to improved modeling of departure time choice \(^{[22]}\). Further discussion of theoretical advances in and applications of activity-based demand modeling can be found in \(^{[23]}\). Recent model systems that incorporate activity-based demand models include
From deterministic to stochastic. Discrete choice models, which are the foundation of much of modern travel demand modeling, are based on an assumption of random utility maximization. The outputs of such a model are probabilities of choosing each of the available alternatives, and so the actual choice predicted in a model application is inherently random. Moreover, most simulation-based demand models involve stochasticity, either in an attempt to reflect the actual uncertainty of the real world, or as a computational device. When demand models are applied in the context of traffic assignment models, and in many other travel forecasting applications as well, the inherent stochasticity of their outputs is usually ignored, however.

Supply Models

Supply models perform two functions: they represent the attributes of transportation service that a link (or sometimes a node) offers based on the traffic using it; and they also represent the propagation of traffic through the network according to the departure time, mode and path that a demand model has predicted. The former function is referred to as a link (node) performance model, while the latter function is referred to as a network loading model. The two are closely inter-related: the total traffic on a link or node that is input to a performance model results from network loading, and a network loading model applies performance models to calculate the travel times that govern the movement of traffic through the network.

Conventional supply models seek to represent steady state traffic patterns; they are static, and traffic dynamics are not modeled at all. Traffic is represented as an aggregate continuous
flow, rather than as individual vehicles. Link performance functions are usually closed form or tabular relationships, referred to as speed-density functions, volume-delay curves or similar names. Travel times are used to assess path utility, but not to propagate flows. Indeed, because these models are static, flow propagation does not need to be explicitly represented, and network loading simply consists of computing link flows by accumulating the flows on the different paths that use the link, using the link-path incidence matrix. All relationships are deterministic.

We briefly note the directions that supply models have evolved in compared to the conventional approach.

From static to dynamic. In dynamic supply models, traffic moves through the network at a finite rate that is determined by the travel times from link (or node) performance models. Traffic enters the network at its origin and in a given time interval. (Most operational supply models treat time as a discrete rather than a continuous variable.) The first link’s performance function determines the amount of time that the traffic will require to traverse that link. The traffic then enters the next link in a correspondingly later time interval, and so on. A change in traffic flow that occurs in one time interval (resulting from increases or decreases in the level of traffic entering the network, or from capacity changes on network facilities due to incidents or operational controls) is reflected in downstream traffic and conditions in a later time interval. This is in contrast to static supply models, in which the assumption of steady-state flows and conditions makes it unnecessary to consider traffic’s finite speed of movement through the network in propagating flows.
In determining the rate of movement of traffic along a link or through a node, performance models in dynamic supply models explicitly or implicitly account for the nature and intensity of vehicle-vehicle and vehicle-infrastructure interactions. Thus the traffic speed on a facility in a given time interval directly or indirectly depends on its traffic level in that interval (and possibly in preceding ones). On the other hand, the traffic level on a facility in a given time interval depends on the speeds encountered on upstream facilities in preceding time intervals. Dynamic network loading models typically resolve this mutual dependency either analytically or through simulation. Again, this is in contrast to static supply models, in which the volume on a link is computed by adding up the volumes on the paths that go through it: network loading is a simple accounting operation.

As noted above, a static representation of traffic levels may be sufficient to design and evaluate large-scale transportation network interventions, such as major capacity additions. However, it is increasingly recognized that dynamic supply models are required for a more detailed understanding of the performance of network facilities, and for the development and implementation of network management and operations policies.

Examples of early dynamic supply models used direct generalizations of static link performance functions, such as delay functions \(^{[27][28]}\) and exit-flow functions \(^{[29][30]}\). These are suitable for theoretical analysis, but are not particularly realistic in modeling traffic dynamics \(^{[31]}\). Other dynamic models are adaptations from traffic engineering practice, such as point-queue models. They usually perform well in modeling isolated bottlenecks but do not readily extend to a network context. One serious drawback of the above models is their
inability to model shock waves, physical queues or spillbacks.

Dynamic supply models based on hydrodynamic theories of traffic flow treat traffic as a compressible liquid, and use partial differential equations to describe the relationships among traffic state variables – speed, density and flow – and the dynamic evolution of traffic (by solving the resulting differential equations for given boundary conditions). The first example of hydrodynamic supply models were due to Lighthill and Whitham \[32\] and to Richards \[33\], who established the so-called LWR theory. Payne \[34\] extended the LWR model by incorporating the impact of acceleration (second-order model).

Finally, dynamic, simulation-based, models that capture traffic dynamics as the result of interactions among vehicles have emerged, e.g. DynaMIT \[35\], DYNASMART \[36\], Nameq \[37\][38][39], METROPOLIS \[40][41\] and MITSIMLab \[42\].

*From aggregate to disaggregate.* An important distinction among the dynamic models discussed above is that some (e.g. models based on the hydrodynamic theories) use an aggregate representation of traffic (flow-based), while others utilize a disaggregate representation (e.g. MITSIMLab). Disaggregate supply models represent traffic at the individual vehicle level with vehicle-vehicle interactions captured at various levels. This detailed representation enables treatment of many important traffic phenomena that are difficult to capture in aggregate models that do not represent the heterogeneity of drivers, vehicles and their interactions. Examples include vehicle interactions at intersections (effects of left turning vehicles, intersection blocking), impacts of heavy trucks, capacity-reducing effects of lane changing, performance of weaving sections, and interaction
among multiple user classes (heterogeneity of traffic flows). In these models, waves, queues and spillbacks emerge as natural results of detailed vehicle-vehicle and vehicle-infrastructure interactions.

Disaggregate, dynamic, supply models are able to accurately represent traffic control and management devices and systems, such as signals, ramp metering, and controlled intersections. This ability is required for operational and intelligent transportation system applications. Delays at these control/management facilities can be explicitly modeled, rather than roughly estimated, and thus disaggregate model results provide a more trustworthy base for the evaluation of the operations/strategies.

Disaggregate supply models are consistent with detailed models of traveler behavior. As discussed above, the use of disaggregate choice models in travel demand forecasting has become common. If an aggregate supply model is used with a disaggregate demand model, some form of demand aggregation is needed and errors may be introduced. It is therefore advantageous to use a disaggregate supply model that matches seamlessly with the demand model, and thus provides an overall coherent modeling framework.

Disaggregate supply models are usually simulation-based. There are two categories of disaggregate supply models, depending on the representation of vehicle movements in the network. One is microscopic simulation models, where individual vehicle movement predictions result from the application of detailed car following, lane changing and other models. The other is mesoscopic simulation models, where aggregate traffic flow relationships are used to predict individual vehicle movements: for example, vehicles move at
a speed obtained from a speed-density curve. The first traffic simulation models appeared in the early 1950s and dealt with the modeling of signalized intersections\[43\][44][45]. Actually, in one of the most celebrated papers in traffic engineering, Webster reports on the use of simulation in his work \[46\]. Early models were facility specific (i.e. separate models for arterials, intersections, freeways). Furthermore, they suffered from inconsistencies, due to the lack of a travel behavior component to model route choice. Instead, at intersections, turning percentages were defined, and vehicles were allocated to a particular approach based on these percentages. Traffic simulation models quickly became very popular for modeling intersections, urban corridors, and freeway control and a special conference took place to review the developments and assess future needs \[47\].

Current traffic simulation models are a complex synthesis of a number of individual models and pay particular attention to capturing traffic dynamics and modeling large scale networks. Microscopic simulation provides the most realistic and detailed modeling, while mesoscopic simulation is an effective compromise between modeling realism and computational efficiency. Most mesoscopic simulation models have been developed as part of a more comprehensive DTA model \[35\][36][48][49], while most microscopic simulation models are used as stand-alone network loaders without being integrated into DTA models, largely because of their high computational and input data preparation requirements.

*From deterministic to stochastic.* As the reality and accuracy requirements of supply models increase, the uncertain aspects of traffic phenomena cannot be ignored. Disaggregate, simulation-based, supply models provide a better platform to deal with system uncertainty.
They are able to handle uncertain factors that are distributed over vehicles and drivers, such as performance characteristics and desired speeds; as well as uncertain factors that derive from the infrastructure, such as incidents.

**Supply-Demand Interaction**

In conventional travel forecasting models, supply-demand interactions are assumed to result in an equilibrium situation where individual travelers have no incentive to change their decisions.

Travelers are assumed to choose, from among the available options, the one that maximizes the utility that they would receive based on their expectations about the available travel options. The demand model represents the outcome of this decision-making process. Similarly, interactions between different vehicles and between vehicles and the transportation infrastructure determine the conditions that a traveler will experience when making a particular trip. Because of congestion, the conditions will generally deteriorate as more travelers use a particular element of the transportation network. The supply model represents the response of the network to traffic flowing over it.

The network response to travelers’ decisions may result in conditions that do not correspond to the expectations that the travelers had when planning their trip. Over multiple trips, travelers revise their expectations and decision-making rules based on what they have learned from prior experiences. The outcome of this learning process (e.g. trip-to-trip changes in travel choices and network conditions) can potentially be quite complex.
When the outcome of the supply-demand interaction is such that travelers have no reason to revise their expectations or decision rules after making a trip, the result is termed supply-demand equilibrium. At equilibrium, travelers’ expectations about conditions on which they base their choices, and the conditions that they experience when making their trips, are fully consistent; as a result, the predicted choices and conditions are stable over repeated trips.

Day-to-day dynamic analysis of transport systems has been introduced by Horowitz [50] and Ben-Akiva et al.\cite{51,52}, and further discussed by Cascetta [53,54] following a stochastic process approach. A general framework presenting both deterministic and stochastic dynamic processes and including equilibrium as a particular case is in [55]. A further analysis of deterministic process is in [56] and [57], whilst [58] and [59] provide a further insight about stochastic process models. Cantarella et al.\cite{60} describe a doubly-dynamic assignment model and apply it to a real case.

The concept of travel equilibrium is derived from fundamental principles of economic theory, and provides powerful approaches and methods for the prediction of travel flows and conditions and for the evaluation of the resulting economic user benefits. Despite its unquestionable value as an abstraction, however, the equilibrium concept must also be judged in terms of how accurate a basis it provides for the prediction of travel flows and conditions. As noted above, this depends on the time scale of interest. At very short time scales (e.g. in incident situations), travelers have little basis from which to form valid expectations, so equilibrium may be a questionable assumption in these situations. At medium- or
longer-term time scales, some travel-related decisions (e.g. path choice) may be near
equilibrium while others (e.g. location decisions) are still adjusting. Current research into
traveler response to information and day-to-day learning behavior, to name just two areas, is
providing new data that ultimately may allow a more empirically validated approach to travel
forecasting.

Conventional travel forecasting models consider only one or a few homogeneous user classes
(where members of each class have the same travel behaviors), and in this case the interaction
of independent traveler utility maximization with network performance characteristics implies
that overall network flows and conditions must verify specific network-level properties. For
example, when each traveler is assumed to choose a path that minimizes individual travel
time, the resulting deterministic user ("Wardrop") equilibrium has the network-level property
that the travel time on all used O-D paths is the same, and is less than or equal to the time on
any unused O-D path.

Computational methods for solving traffic assignment models are usually found by
identifying mathematical problems whose solutions also verify these network-level properties,
and then developing algorithms that numerically solve the “equivalent” mathematical problem.
Equivalent optimization, non-linear complementarity, variational inequality and fixed point
problems have been used for this purpose [61]. To establish the correspondence between
solutions of an assignment model and the equivalent mathematical problem, conditions (e.g.
continuity, separability, symmetry, monotonicity) on the involved supply and demand
functions must generally be imposed.
Most common methods for applying the four-step model system (trip generation, trip distribution, mode split, and traffic assignment) to determine an overall equilibrium between all demand model components and the supply model can only be considered heuristics without a rigorous mathematical basis. Therefore, in discussing the evolution of supply-demand interaction modeling, we focus primarily on traffic assignment models since these are the most rigorous. As noted in Section 1, traffic assignment models have not experienced development in all the directions that individual supply and demand models have. The principal aspects of the evolution of these models are described below.

*From static to dynamic.* The most substantial development in supply-demand interaction models, compared to the conventional approach, has been in the formulation and solution of dynamic traffic assignment (DTA) models. The analysis of these models is more challenging than of static models, but there have been considerable advances in these areas as a result of the efforts of theoretical and applied researchers over the years. To a large extent, the needs of the DTA research have dictated the approaches adopted to represent the assignment model demand and supply components. In other words, the independent advances in demand and supply modeling discussed above have not, for the most part, been incorporated in DTA model development.

*From aggregate to disaggregate.* Little progress has been made in methods for determining the outcome of supply-demand interactions in disaggregate models. The conventional assignment model solution approach may not be applicable in disaggregate models because the heterogeneity of travel behavior (in effect, the large number of different user classes)
tends to reduce or eliminate the network-level properties of equilibrium traffic and conditions that the approach relies on. Moreover, increasing detail and realism in supply and demand models, and recourse to simulation-based methods, will make it difficult or impossible to verify these models’ mathematical properties for the purpose of establishing an equivalent mathematical problem. A few algorithmic methods are commonly applied to compute supply-demand equilibrium in simulation-based models, but these are generally heuristics, without a rigorous mathematical justification for this application. Cantarella and Cascetta [55] describes a disaggregate stochastic process model for demand assignment and discuss the results of its application to a toy network.

*From deterministic to stochastic.* Even when individual supply or demand models are inherently stochastic, most approaches to compute the outcome of supply-demand interactions either ignore the stochasticity or focus only on the expected values of the random variables. This may sometimes be justified as being a “large population” approximation, although the justification becomes less tenable when such models are applied at a disaggregate level, to individual travelers or vehicles. A true stochastic equilibrium would be characterized by mutually consistent and stable probability distributions of network traffic and condition variables. Relatively little research has been done in this area. Fully stochastic static equilibrium is examined from a purely probabilistic and atemporal perspective in [62], while [53][54][55][58][59] address stochasticity in static and dynamic network flows and conditions using stochastic process models that incorporate day-to-day user learning behavior.
3. **A Framework for Disaggregate Dynamic Travel Forecasting Models**

As the brief overview in Section 2 indicates, travel demand and supply models have seen significant advances in recent years, and current disaggregate dynamic demand and supply models possess the ability to address a wide range of problems with a high level of modeling realism. However, supply-demand interaction models have not evolved to the same extent: most progress has been in the area of traffic assignment models. Nonetheless, although these have progressed from static to dynamic time representations, the resulting dynamic traffic assignment (DTA) models still do not fully incorporate available advances in disaggregate supply and demand modeling approaches.

A disaggregate dynamic travel forecasting model, consisting of disaggregate and dynamic demand and supply components together with appropriate methods for resolving their interactions, would combine in a consistent way the rich spectrum of travel behavior represented by disaggregate demand models with the high-accuracy depiction of traffic phenomena by disaggregate supply models, and provide improved modeling realism and breadth of applicability compared to aggregate dynamic models.

Disaggregate dynamic models bring together a number of advantages. They can accurately simulate the time-varying nature of congestion during peak periods and represent traveler responses to time-varying transportation system attributes and policies (such as congestion pricing), including departure time choice, pre-trip path choice, and en-route response to traffic information. Their tour- and activity-based demand model enhances behavioral fidelity by recognizing that travel is a by-product of travelers’ participation in other activities and that
point-to-point trips may be components of a more complex travel pattern. Their disaggregate demand representation enables efficient use of data at household or individual level and incorporation of better set of behaviorally-meaningful explanatory variables. Finally, their disaggregate supply representation allows treatment of many important causes of congestion that are impossible to capture in aggregate models and ensures an accurate representation of traffic control and management devices and systems, including a variety of Intelligent Transportation Systems (ITS) technologies. A disaggregate and dynamic framework also provides a better platform to deal with inherent system stochasticity. This forecasting framework has wide applicability, since it is sensitive to a large variety of policies ranging from long-term land-use to very short-term dynamic operational and ITS ones.

Figure 2 illustrates the proposed framework, emphasizing its relationship to the general supply-demand framework presented in Figure 1. Inputs include suitably detailed descriptions of the traveling population and of the traffic network. Depending on the application, it may be necessary to provide distributions of socioeconomic characteristics and information access (both pre-trip and en-route) over the population, possibly together with land use information. The traffic network description covers both the infrastructure (highways, arterials, local roads) and the traffic management, control and information systems. Outputs from such a model include disaggregate data on vehicle movement and individual travel choices. Aggregate network-level performance attributes and travel demand summaries can be computed from the disaggregate results.
Three main components use the above inputs to generate the desired outputs:

- disaggregate demand microsimulation
- disaggregate supply simulation
- supply/demand interactions resolution.

The *disaggregate demand simulator* includes a population generation process and a micro-simulator that applies travel choice models to the individual travelers generated by the population generator. The population generation process results in individual travelers who reflect the known distribution of population and/or travel characteristics. For trip-based models, the population can be generated from an aggregate dynamic origin-destination trip table, and the generated individuals then make mode, departure time, and/or path choices according to the appropriate demand models. For activity-based models, on the other hand, households and individuals are usually generated from demographic data. The choice
models typically incorporate time-varying explanatory variables, and predict departure time choice, among other travel decisions. Choices are based on the attributes of the various travel alternatives (e.g. paths) and the traveler’s own characteristics, and may represent a random selection based on the choice probabilities output by the demand model.

The *disaggregate supply simulator* represents the time-dependent movement of individual vehicles over the network in accordance with the travel choices predicted by the demand simulator. Vehicles are moved according to either macroscopic or microscopic traffic performance models. Stochasticity in vehicle movement is explicitly represented. The supply simulator produces both vehicle- and network-level measures of travel performance, such as vehicle space-time trajectories, link travel time and cost, and congestion delay.

An important feature of the proposed framework is the full and seamless integration between the demand and supply modules. Full integration means that the vehicles loaded onto the network correspond directly to the population generated in the demand module, in terms of their characteristics, and travel decisions regarding path and departure time choices. In addition, the vehicles loaded on the network maintain their socioeconomic characteristics and other attributes, so that en-route decisions and response to information is consistent with these characteristics, and their pre-trip behavior. Furthermore, the full integration between demand and supply facilitates the representation of the supply/demand interactions.

The *supply/demand interactions* module predicts equilibrium conditions based on the interactions of the disaggregate demand and supply models. The framework accommodates the interactions at all levels of time scale, from those involving long-term land use choices
(e.g. where to work, where to live), to medium-term travel choices such as mode and
departure times, to short-term path choices. (Note that if land use is included in the demand
model, then it should be removed from the input in Figure 2.)

With disaggregate representation of supply and demand, the computation of equilibrium
cannot be based on aggregate network-level properties of traffic and conditions. Fixed point
formulations offer a promising way to express and solve for equilibrium in this context.
These formulations can be expressed in terms of either behavior (demand) variables $b$ or
condition (supply) variables $c$, e.g.:

$$S \ast D(c) = c$$
$$D \ast S(b) = b$$

where $D()$ designates the demand function, $S()$ designates the supply function and $\ast$
designates functional composition. In the first case, for example, a set of network conditions
represent an equilibrium if, after travelers make decisions based on the conditions and the
network then reacts to the decisions that were made, the resulting network conditions are the
same as the initial network conditions, so that travelers have no reason to revise their
decisions. A similar interpretation applies to the second case.

Fixed point approaches have occasionally been applied in the past to analyze conventional
travel forecasting models, but have never been among the mainstream methods applied to
solve these models, in large part because computational methods that exploit the
network-level properties of equilibrium flows and conditions with homogeneous users tend to
be more efficient. Nonetheless, this formulation is interesting for disaggregate dynamic
travel forecasting because it captures the essential supply-demand consistency property of equilibrium, even at a disaggregate representation level. It also appears to impose the least stringent conditions on the involved functions (i.e. the supply and demand models). Moreover, as a general approach it is robust across a wide variety of equilibrium problem details such as the amount of behavioral heterogeneity and the extent of stochasticity in the demand and supply relationships.

These problem details dictate the appropriate approach for solving the equilibrium fixed point problem, but many general methods are available. For example, gradient-based methods may be appropriate for deterministic problems; stochastic approximation (or averaging) methods – including both conventional algorithms such as the method of successive averages (MSA) \cite{63,64,56,2} as well as accelerated methods such as the one due to Polyak \cite{65} – are appropriate for problems that allow stochasticity but represent its outcome in terms of expected values (such as the so-called Stochastic User Equilibrium model); and Markov Chain Monte Carlo methods such as Gibbs sampling are appropriate for computing the equilibrium probability distributions of traffic and condition variables in fully stochastic problems \cite{62}. With increased recognition of the value of this formulation for disaggregate dynamic problems, new research may lead to more efficient fixed point computational methods for these problems.

4. Examples

Elements of the disaggregate dynamic travel forecasting framework presented in Section 3 can be found in a number of model systems that exist or are under development, including
DynaMIT [66], DYNASMART [36], Dynameq [37][38][39], METROPOLIS [40][41], ILUTE [25], TRANSIMS [24], etc. We focus here on DynaMIT as an example of a model system that is built on the principles of this framework and that includes many of the requirements outlined in the previous sections. DynaMIT combines inter-related models of travel behavior (demand simulator), dynamic network performance (supply simulator), and demand/supply interactions. It captures the stochastic nature of transportation systems and can represent the functioning and effects of ITS technologies. DynaMIT is particularly useful for short-term planning applications involving work zones, special events, high occupancy vehicles and high occupancy/toll (HOV and HOT) facilities and congestion pricing.

Demand Simulator. The DynaMIT demand simulator is a microscopic, disaggregate travel behavior simulator and consists of modules for demand disaggregation, disaggregate travel choice prediction and, depending on the application, O-D matrix estimation. The role of the demand disaggregation is to generate the population of drivers from given O-D matrices. As each driver is generated, it is assigned a number of socioeconomic and trip characteristics. Based on these characteristics, individual travelers make decisions, using disaggregate choice models, about path, departure time, and possibly travel mode. The path choice model is a key component and determines the paths that vehicles follow over the network, incorporating pre-trip and/or en-route information. The departure time models capture an important behavioral response to congestion, incidents or dynamic pricing. In contrast to the framework presented in this paper, the demand simulator in DynaMIT currently does not simulate activity patterns, and is an example short-term demand model.
An important component of the demand simulator is the (aggregate) estimation of dynamic O-D flows from relevant measurements (e.g. traffic counts). The O-D matrix estimation problem is one of combining and reconciling information from diverse sources and with various error characteristics. This functionality is critical for short-term planning applications.

Supply Simulator. The DynaMIT supply simulator is a disaggregate, time-based, mesoscopic traffic simulator. Given a set of travelers and their departure times and paths as predicted by the demand simulator, and given the network characteristics and implemented control strategies, the supply simulator predicts the performance of the network in terms of time-dependent travel times, queue lengths, etc. The complexity of traffic dynamics in the network is captured through the integration of three classes of models: capacities associated with roadway features, incidents, and control strategies; queuing reflecting the effect of bottlenecks; and macroscopic speed-density relationships representing uninterrupted flow. The simulator is designed to operate at different levels of granularity, depending on the requirements of the application.

The demand and supply simulators in DynaMIT have also moved towards full integration. The O-D estimation process is a form of population generation. The demand simulator generates the population of drivers who are assigned socioeconomic characteristics, based on their origin, and make pre-trip route and departure time decisions. These drivers are then loaded on the network accordingly. Their characteristics are maintained throughout their trip and hence, if they receive information, or encounter a Variable Message Sign (VMS) they
respond based on these individual attributes and the corresponding models employed by the demand simulator.

**Demand-Supply Interactions.** DynaMIT utilizes an MSA-type algorithm \cite{35} for the solution of a fixed-point problem that captures the dynamic interactions between demand and supply. The formulation is in terms of time-dependent link cost variables, and allows for the effects of en-route path revision based on real time travel information. The solution algorithm simultaneously solves for time-dependent traffic levels, network conditions and travel information.

5. **Challenges**

A number of issues, both theoretical and practical, require further attention and research to make the disaggregate dynamic travel forecasting framework discussed in this paper useful and practical. In particular, some of the important challenges include:

- The algorithmic approaches used for resolving the demand-supply interactions need to be better understood in terms of their properties with respect to convergence and rate of convergence. Establishing these properties is not trivial, due to the simulation-based nature of the involved supply and demand models, and the stochastic nature of the modeling framework.

- Data availability and methods for system calibration are important requirements for the success and applicability of models based on this framework. Required data include both socioeconomic and network performance characteristics. In general, calibration of a disaggregate dynamic model system should consist of (disaggregate) calibration of
individual models, such as travel behavior models, using disaggregate data from detailed surveys and travel diaries; as well as (aggregate) calibration of the overall model system using aggregate data that is readily available and collected by common sensor technologies, such as counts and speeds from loop detectors. Emerging point-to-point data sources (such as automated vehicle identification or AVI sensors) will likely provide valuable information that can improve the effectiveness of the calibration task.

- Validation of the framework and its underlying assumptions is an important and challenging task, especially given the type of the available data. Validation of the behavioral assumptions underlying the equilibrium conditions sought by the supply/demand interactions can be particularly challenging.

- Solution of stochastic models results in joint probability distributions of the output variables rather than point values. Calculating such distributions can be a very significant computational challenge, and appropriate methods for interpreting and acting on such outputs are still not well understood.

- More generally, the computational performance of disaggregate dynamic model systems can be a major impediment to their acceptability. It is critical that these model systems be able to handle large scale urban networks with reasonable computational effort. Achieving this goal will require both basic research on efficient algorithms, and implementation strategies that take advantage of efficient data structures and utilize advanced simulation technologies.

As disaggregate dynamic travel forecasting model systems are implemented, deployed and applied, the experience obtained will help identify practical aspects related to their use and
contribute to making them operational and practical for a wide range of applications, such as operations planning, long-term planning, and emergency/disaster response.

6. Conclusion

Travel forecasting models have experienced substantial growth in their development and application. Some of the factors that have contributed to this include:

- Focus on better utilization of available capacity and hence more emphasis on operations management and planning
- Need for evaluation and modeling of complex time-dependent system interventions due to the emergence of ITS
- Increased availability of detailed, time-dependent data on network performance and use.

In order to provide the functionality needed to support the new application focus, state of the art DTA models are a synthesis of a number of advanced models representing the demand and supply characteristics of the transportation system, along with sophisticated algorithms to capture the complex interactions between supply and demand.

Although DTA models have moved forward from the earlier, flow-based supply and aggregate demand formulations of the traffic assignment problem, they can be further improved. In particular, this paper argues that disaggregate dynamic travel forecasting models are the most promising approach for satisfying all the requirements of current (and future) applications, including policy analysis. These models provide increased realism in capturing both travel behavior and network performance and so provide considerable flexibility in representing a
The main characteristics of disaggregate dynamic travel forecasting models are:

- Disaggregate activity-based representation of the choices of individual travelers, including departure time choice, pre-trip path choice, and en-route response to traffic information,
- Detailed, disaggregate, dynamic modeling of network performance, through vehicle-based traffic simulation models that capture the dynamic evolution and dissipation of queues and the formation of spillbacks, and that represent the operations of the traffic control system at an appropriate level of detail,
- Full integration between the demand and supply modules through the generation of a population of drivers in the demand module, who are loaded (along with their individual attributes) on the network by the supply simulator, and
- Algorithms that solve the resulting demand-supply interactions as fixed point problems.

Despite significant advances in the last few years that have resulted in a comprehensive and consistent modeling framework and in flexible, simulation-based model implementations, a number of challenges still remain and further work is needed to make these models operational.

**Acknowledgment**

The authors are indebted to Giulio Cantarella and André de Palma for their insightful and useful comments and suggestions.
References


[37] Florian, M., Mahut, M. and Tremblay N. A simulation based dynamic traffic assignment: the model, solution algorithm and applications. Presented at the International Symposium of Transport Simulation 2006 (ISTS06), September 4-6, 2006, Lausanne, Switzerland. ISTS06 Compendium of Papers CD-ROM.


[50] Horowitz, J.L. The stability of stochastic equilibrium in a two link transportation network.


