TRANSPORTATION RESEARCH
CIRCULAR

Number E-C153 June 2011

Dynamic Traffic Assignment

A Primer

TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

TRANSPORTATION RESEARCH CIRCULAR E-C153

Dynamic Traffic Assignment

A Primer

YI-CHANG CHIU
JON BOTTOM
MICHAEL MAHUT
ALEX PAZ
RAMACHANDRAN BALAKRISHNA
TRAVIS WALLER
JIM HICKS

for the
Transportation Network Modeling Committee

June 2011

Transportation Research Board
500 Fifth Street, NW
Washington, DC 20001
www.TRB.org

TRANSPORTATION RESEARCH CIRCULAR E-C153
ISSN 0097-8515

The **Transportation Research Board** is one of six major divisions of the National Research Council, which serves as an independent adviser to the federal government and others on scientific and technical questions of national importance. The National Research Council is jointly administered by the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal.

The **Transportation Research Board** is distributing this circular to make the information contained herein available for use by individual practitioners in state and local transportation agencies, researchers in academic institutions, and other members of the transportation research community. The information in this circular was taken directly from the submission of the authors. This document is not a report of the National Research Council or of the National Academy of Sciences.

Transportation Research Board
500 Fifth Street, NW
Washington, DC 20001
www.TRB.org

Javy Awan, Production Editor; Jennifer Correro, Proofreader and Layout

Preface

Transportation planners and traffic engineers are faced nowadays with immense modeling challenges arising from several emerging policy, planning, and engineering developments. Hence, interest has grown in applying traffic analysis tools capable of analyzing travel activities and dynamic network performance for a corridor or region over peak hours or even extended daily hours.

Now, after decades of research and intensive market readiness developments, dynamic traffic assignment (DTA) models have become a viable modeling option. DTA models supplemental existing travel forecasting models and microscopic traffic simulation models. Travel forecasting models represent the static regional travel analysis capability, whereas microscopic traffic simulation models are superior for dynamic corridor-level travel analysis. DTA models fill in the gap by enabling dynamic traffic to be modeled at a range of scales from the corridor level to the regional with expanded and unique functional capabilities enabled by the DTA methodology.

The motivations for the TRB Network Modeling Committee (ADB30) in developing this primer were to provide neutral and factual information about DTA, to facilitate informed decision making by practitioners in planning or managing a DTA modeling activity, and to engage practitioners with educational material about modeling exercises and interpretations of results related to DTA.

The objectives of this primer therefore are to

- Explain the basic concepts of DTA and various DTA definitions and implementations,
- Highlight the types of transportation analysis applications for which DTA models could be found useful,
- Provide information about how to select a DTA model that best serves the intended application,
- Provide information regarding planning for and executing a DTA traffic analysis activity, and
- Describe the general DTA modeling procedure and modeling issues that may concern a model user.

This effort represents a first step in the committee's continuing commitment to facilitate development of practical analysis procedures in the DTA area and to improve communication among DTA researchers, developers, and user communities.

The authors of this document are Yi-Chang Chiu, University of Arizona; Jon Bottom, Steer Davies Gleave, Inc.; Michael Mahut, INRO Inc., Canada; Alex Paz, University of Nevada at Las Vegas; Ramachandran Balakrishna, Caliper Inc.; Travis Waller, University of Texas at Austin; and Jim Hicks, Parsons Brinkerhoff Inc. The authors' acknowledgments appear on page iv. In addition, two scholars made contributions to this publication during the writing and publication process: Steve Boyles, University of Wyoming, and Avinash Unnikrishnan, West Virginia University.

—Srinivas Peeta
Chair, Committee on Transportation Network Modeling

AUTHORS' ACKNOWLEDGMENTS

This publication is the result of a volunteer effort by individual members and friends of TRB's Committee on Transportation Network Modeling and is intended as a resource reference for transportation professionals who are interested in the general concept and modeling approaches of DTA. DTA is a general term encompassing a variety of problem definitions, formulations, and algorithmic solution procedures.

The authors strove to present DTA in a model–implementation independent manner; however, due to page and resource limitation, not all DTA model variations are included in this publication. This publication does not endorse any specific DTA definition or discuss how to select a DTA model that best serves the intended application or solution algorithm implementation, nor does it provide specific guidelines or best practice in applying DTA. Nonetheless, it is inevitable that certain discussions may exhibit judgmental expressions and opinions.

Opinions expressed in this publication represent those of the authors, not those of TRB.

The authors acknowledge the following individuals for providing valuable inputs to this document:

Technical Reviewers

- David Boyce, Northwestern University;
- Michael Florian, University of Montreal, INRO, Canada; and
- Dave Watling, University of Leeds, United Kingdom.

Practitioner Reviewers

- Asa Bergman, Portland Metro and Oregon State University;
- Kim Fisher, TRB;
- Sam Granato, Iowa Department of Transportation;
- Brian Gardner, FHWA;
- Dong Hu, City of Bellevue;
- Catharine Jensen, Michigan Department of Transportation;
- Leo (Lili) Luo, Maricopa Association of Governments;
- Vassilis Papayannoulis, Cambridge Systematics, Inc.;
- Eric Pihl, FHWA;
- Jeffry Shelton, Texas Transportation Institute;
- Aichong Sun, Pima Association of Governments;
- Chung Tran, FHWA; and
- Liang Zhou, Wilbur Smith Associates, Inc.

Contents

Why Dynamic Traffic Assignment?

FROM A TRANSPORTATION PLANNING PERSPECTIVE

Travel forecasting models are used in transportation planning to evaluate the impact of future changes in demographics, land use, or transportation facilities on the performance of a region's transportation system. Traveler behavior is introduced into these forecasting models through sequences of modeling steps. The traditional four-step process, for example, results in travel choices made by groups of homogeneous travelers in aggregate trip-based models. More advanced activity-based processes seek to represent travel choices made by individual travelers.

Cost and time of travel are key components of all travel models throughout the entire sequence of model steps. For example, a household's choice of the number of personal vehicles to own is often forecast subject to aggregate measures of the accessibility of the household. The less accessible a household is, the more likely it is to own automobiles. An accessibility measure is then some representation of the travel time and cost from the residence to work places or shopping places. Likewise, time and cost are clearly significant factors in other choices made, including residential, workplace, and discretionary activity locations, as well as factors in deciding which transportation services to use and which routes to follow when engaging in travel.

From a travel forecasting perspective, the time and cost of travel are critical factors. Those measures are also critical in quantifying impacts on a regional scale for the purpose of informing policy decisions. Travel time and cost measures determined using static network analysis (assignment) procedures use variables of interest that are time-invariant. It has become increasingly evident that these procedures are inadequate as explanations of influences on travel choices and as measures used to evaluate impacts when deciding how to develop policies for managing transportation systems, how to fund transportation system improvements, and how to measure environmental impacts related to systemwide travel.

Dynamic network analysis models seek to provide another, more detailed means to represent the interaction between travel choices, traffic flows, and time and cost measures in a temporally coherent manner (e.g., further improve upon the existing time-of-day static assignment approach). More specifically, dynamic traffic assignment (DTA) models aim to describe such time-varying network and demand interaction using a behaviorally sound approach. The DTA model analysis results can be used to evaluate many meaningful measures related to individual travel time and cost, as well as systemwide network measures for regional planning purposes.

FROM A TRAFFIC ENGINEERING PERSPECTIVE

Traffic engineers increasingly rely on traffic analysis tools to analyze and evaluate the current and future performance of transportation facilities for various modes of transport. There are a variety of analytical procedures and methodologies available that support different aspects of traffic and transportation analyses. Nowadays, most traffic analysts rely on either analytical–deterministic tools or microscopic simulation modeling to assess the performance of transportation systems of interest.

Publications such as the *Highway Capacity Manual* (HCM) (*1*) contain guidelines, concepts, and procedures for computing the quality of service and capacity of various highway facilities. HCM is an excellent resource for a sketch evaluation and high-level analysis of quality of service on planned or existing roadway infrastructure. This tool provides a set of analytical methods and practices including a logical methodology for assessing transportation facilities, but is limited in its ability to analyze effects such as oversaturation, queue spillback, dynamic routing, or peak spreading.

Other prevalent traffic analysis tools include microscopic traffic simulation models. Microscopic models simulate the movement of individual vehicles based upon car-following, lane-changing, and gap-acceptance theories. These models are often used to analyze various geometric design configurations, to evaluate and optimize localized individual intersections, and to analyze the interactions of multiple modes of transportation including cars, transit, rail, and pedestrians.

Newer microscopic models are route based, meaning vehicles select a route at departure and follow that route with or without further update along the journey during simulation. Most microscopic simulation models provide various ways by which a vehicle's route at departure or en route is selected or updated. Each approach is linked to a distinct route choice behavior and, while such flexibility can be of great convenience to the modeling work at hand, one need to realize the underlying route-choice behavior assumption associated with each method, as well as the impact of analysis outcomes depending upon which of the different available mechanisms is chosen.

For example, the one-shot (noniterative) assignment-simulation approach is commonly used in some microsimulators, in which vehicles departing at different times are given a route that is periodically updated in simulation based on instantaneous travel times—snapshot travel time measured at the time that the routes are generated without considering congestion during subsequent time periods. Such an assignment can be regarded as if travelers strictly follow some types of pretrip route guidance. Some microsimulation models allow the en route vehicles to update their routes based on the updated shortest route generated at a later time. This feature also implies a route choice behavior that strictly follows the en-route route guidance. While these two route choice behaviors exist in reality, it is important to realize that the majority of travelers may choose a route that leads to the minimal experienced travel time instead of minimal instantaneous travel time. The experienced travel time needs to be evaluated after the fact, by which point the traffic condition along the entire journey is revealed and experienced. In other words, choosing a minimal experienced travel time route at departure involves anticipation of future traffic condition along the journey. This anticipation is usually formed by learning from prior experience (e.g., try different routes). To account for this learning process, an iterative algorithmic process is needed. Such an iterative process reflects the learning and adjustment in route choice from one iteration to the next until the traveler cannot find a route with a shorter experienced travel time. More details are provided in the section on Instantaneous and experienced travel times.

The equilibrium-seeking DTA methods are based on iterative algorithmic procedures that are particularly aimed at describing such an individual route–departure time choice adjustment as well as at relating such changes to the network-level performance through simulation. These models apply iterative procedures involving the interplay of vehicular traffic loading and assignment algorithm to adjust the traveler route assignment in order for travelers departing at different times to select the respective minimal experienced travel time route. Many simulation-based DTA models adopt more computationally efficient traffic simulation logic (at the price of simplifying some simulation fidelity or detail) in order to be able to describe a corridor–regionwide traffic flow shift at a larger geographical scope (from a corridor up to a region) and over a longer time period (from peak hours to 24 h), compared with microscopic models. Most of these simulation methods are generally defined as mesoscopic simulation sharing common characteristics with microscopic models—individual vehicles are represented and vehicle dynamic states are simulated through simplified car-following or traffic flow theories without describing detailed intervehicle interactions (e.g., lane changing or gap acceptance).

STATIC VERSUS DYNAMIC MODELS

In a model defined on a relatively long time-of-day period, such as the peak period, the congestion properties of each link are described by a volume–delay function (VDF) or link time–performance function that expresses the average or steady-state travel time on a link as a function of the volume of traffic on the link. Such models are called static. The volume of traffic on the link is determined directly from the loading of the origin–departure matrix to links via routes. The travel times of each link on a route are added together to determine the route travel time. This approach has some limitations as far as the realism with which it represents the actual process (taking place on the road) that gives rise to congestion and increased travel time.

In a static model, inflow to a link is always equal to the outflow: the travel time simply increases as the inflow and outflow (volume) increases. The volume on a link may increase indefinitely and exceed the physical capacity (in vehicles per hour) of the link, as represented by a volume-to-capacity (V/C) ratio > 1. [Capacities used in static models generally do not correspond to maximum flow rates; it is typical to see capacity defined as the flow rate corresponding to level of service C or D. Often, it also incorporates the effect of downstream signals.] Since the link volume does not conform to the traffic flow limit that results from the physical characteristics of the roadway, the assigned link volume can be considered as demand—trips desired to traverse the link—instead of the actual flow. V/C > 1.0 means that the demand exceeds the capacity and subsequently congestion will occur. The drawback of using V/C is that it does not directly correlate with any physical measure describing congestion (e.g., speed, density, or queue). In dynamic models, as in reality, explicit modeling of traffic flow dynamics ensures direct linkage between travel time and congestion. If link outflow is lower than link inflow, link density (or concentration) will increase (congestion), and speed will decrease (fundamental speed–density relationship), and therefore link travel time will increase.

Outflow from a link may be reduced, and thus be potentially less than the inflow, for various reasons, such as

- Merging two lanes into one (e.g., at a freeway on-ramp) effectively reduces the capacity of each of the two merging lanes;
- Weaving (lane change maneuvers that cross over each other) also reduces link capacity;
- On arterial streets, traffic signals reduce the outflow capacity of links; and
- On both freeways and arterial streets, significant oversaturation for one exiting movement from a link can result in reduced flow rates on the other exiting movements, due to a local choke-off effect.

Traffic initially becomes congested (e.g., queuing occurs) at the end of a link because link inflow is greater than link outflow (put another way, a congested traffic state arises at the end of the link under these conditions). According to the basic tenets of traffic flow theory—upon which dynamic models are based—for a given value of outflow, there is a corresponding value of density and speed under congested conditions. This is best thought of in the case of a freeway, where the outflow is roughly constant, as opposed to a signalized road where outflow is constantly fluctuating. For purposes of this discussion, we assume that the outflow is in fact constant. The longer this condition (inflow > outflow) persists, the more vehicles accumulate on the link, and the portion of the link covered by the congested traffic grows in the upstream direction until it reaches the link entrance. At this point in time, the inflow is reduced. It is, in fact, equal to the outflow, and the link is in a steady-state condition, meaning that speed, density, and flow are essentially constant at all positions (in space) along the link. The speed and density on the link correspond to the flow (inflow and outflow, which are equal) in a well-defined mathematical way, called the fundamental diagram of traffic flow.

In a dynamic model, each link may be defined by its own fundamental diagram, if desired. This is sometimes thought of as the dynamic analogy to the static VDF, but this analogy is loose as the two mathematical relationships actually perform very different functions in the contexts of their respective models. In a static model, the VDF actually represents the congested condition, while in a dynamic model, the fundamental diagram describes how congestion at the exit node (reduced link outflow) is propagated upstream though the link, until it spills back onto the next upstream links.

This phenomenon brings forth the question of congestion spill-back, which is not represented in a static model. At the moment that the link inflow becomes equal to the outflow (as described above), the congestion then continues to spread upstream into whichever upstream links are feeding traffic into the congested link. The outflows of these links are thus reduced, and the process repeats as described above. This queue spillback process also describes how a long queue (congested traffic) can be represented over a sequence of links in a dynamic traffic model.

There is also the question of link FIFO (first-in, first-out). Static models, and even some dynamic models that are based on fluid mechanics, enforce the link FIFO rule. In a static model, this means that all vehicles traveling on the link experience the same travel time. In a dynamic model with FIFO, this means that all vehicles entering the link at a given point in time experience the same travel time. What this implies is that there is no overtaking between vehicles and, in particular, this means no overtaking between vehicles that exit the link by different turning movements. In reality, it is quite obvious that if there are two turning movements for exiting a link and if one is oversaturated and the other is not, then the vehicles in queue for the oversaturated movement can be overtaken by the other vehicles (assuming the link has more than one lane), and that the latter vehicles can have significantly lower travel times than the former. Models that move individual vehicles on discrete lanes of the roadway can model non-FIFO conditions realistically, and thus have no need of employing the FIFO assumption. Further, if the turn bay queue spills back to the through lane, the resulting capacity reduction also needs to be properly accounted for through appropriate traffic modeling.

Last, it is worth noting that, as there is no explicit representation of individual lanes in static models, there can be no distinction between the traffic conditions on different lanes of the same link. There is no way to represent the fact, for example, that the outside lane of a freeway is at a crawl due to an oversaturated off-ramp, while the other lanes are moving at a higher speed.

In summary, the limitations of static models due to their use of VDFs include

- Using VDFs, a link may have a V/C ratio greater than 1.0—the V/C ratio does not have intuitive traffic meaning;
- VDFs assume link FIFO, and therefore no overtaking;
- VDFs do not distinguish between different lanes on a roadway; and
- VDFs are based on a single value of link flow (or volume), implying that inflow is equal to outflow, and hence there is no accumulation of traffic on the link. As a result, there is no representation of the phenomenon of congestion spillback, i.e., where congested traffic spans a sequence of two or more links due to a downstream bottleneck.

Beyond the issues related directly to the use of the VDF and how travel time is determined in static models, other limitations include, for example, modeling of signal synchronization, modeling of lane-based effects, such as high-occupancy vehicle (HOV) or high-occupancy toll (HOT) lanes, as they require representing the special lane as a parallel link. Most intelligent transportation systems (ITS)–related applications, such as traveler information systems and advanced network control schemes (e.g., adaptive control and ramp metering), are beyond the modeling capabilities of static assignment models.

Notwithstanding the above critique, this document does not intentionally overlook the merit of static models. The widely recognized advantages of static models, including the ability to solve large-scale problems, to converge to precise equilibriums and to provide consistency of solutions (if a proper algorithm is used with a sufficient number of iterations) have been aiding policy–project decision making for agencies for decades. The critique is of benefit in demonstrating the contrast with dynamic models, but the contributions and merits of the static models should not be understated.

Dynamic Traffic Assignment in a Nutshell

In writing this primer, several practical considerations take priority to achieve broader communication. As a result, this chapter addresses only the core concepts necessary for the most straightforward and underlying DTA concepts as opposed to serving as a comprehensive synthesis of all DTA research. Numerous high-quality academic contributions cannot be given full attention, and compromise is required to meet state-of-practice needs. Perhaps the most critical position taken in this chapter is that of the default definition of DTA. Similar to what appears to be standard for current static assignment practice, in the absence of any prefix or suffix, the term dynamic traffic assignment is taken in this chapter to imply an equilibration based on experienced travel costs. While certain high-quality research may employ differing definitions and assumptions, the aforementioned definition appears both expedient and a usage that matches well with current understanding in the field based on static traffic assignment methods.

As a modeling process, the goal of traffic assignment is to determine the network traffic flows and conditions that result from the mutual interactions among the route choices that travelers make in traversing from their origins to their destinations, and the congestion that results from their travel over the network. In order to achieve this result, several assumptions need to be made, particularly regarding how traveler route choice behavior is modeled and how traffic flows and conditions are represented.

In practice, the common behavioral assumption is that travelers choose the available route having the least travel time between their origin and destination (O-D), reflecting the idea that travel is rarely a goal in and of itself, but instead involves some time, cost, or disutility that travelers would prefer to avoid. (More generally, travelers are sometimes assumed to choose the route having the lowest cost, or the minimum disutility, appropriately defined. Within limits, the particular impedance variable assumed to influence route choice does not affect the discussion here, so we will usually just say that travelers minimize their travel time with the understanding that the discussion could apply equally well to travel cost or other disutility measures.) Due to congestion effects, the travel time of a route between an O-D also depends on the choices made by other travelers, who are themselves also trying to choose the least travel time route between their own O-D. When every traveler succeeds in finding such a route, every used route has the minimum time or cost between O-D; moreover, for each O-D pair, every route used has the same travel time. This condition is known as user equilibrium.

Traffic assignment algorithms find these interactions to determine route and link volumes and travel times that satisfy this equilibrium condition through iterative procedures. At equilibrium, no traveler can find an O-D route that would lead to a reduction in travel time. If an equilibrium state is reached, it will persist as long as the network and travel demand do not change, because no travelers have any incentive to choose different routes.

The concept of equilibrium is an abstraction because of the simplifying assumptions that it entails. As noted, travelers are assumed to choose O-D routes that require minimum time (or disutility). Accordingly, travelers are assumed to know, and accurately perceive, travel times throughout the network, presumably through numerous trials of different routes to be able to identify their minimum travel time route given any congestion scenario. Finally, O-D flows and roadway characteristics are assumed to be fixed and known.

Despite these simplifying assumptions, the concept of equilibrium is meaningful for several reasons. First, it is believed to be a reasonable approximation of traveler choice, and leads to efficient solution methods and transferable conclusions, qualities that alternative simpler behavioral assumptions do not share. Second, the modeling time horizon is often long enough to assume that most travelers have discovered the shortest routes for their trips. Third, advances in advanced traveler information systems (ATIS) and other ITS technologies can make travelers much more aware of network conditions than was typical in the past. Finally, adoption of the equilibrium principle makes available methods from economics for evaluating the potential benefits (or disadvantages or costs) that accrue to travelers following a change in travel conditions due to implementation of certain transportation projects or policies. Thus, when comparative analysis is required, these features have made equilibrium-based traffic modeling the predominant approach used in practice to date.

The other major assumption in traffic assignment concerns the manner of representing traffic flow and conditions—that is, the way that travelers' route choices are related to networkwide congestion and travel times. Historically, traffic assignment methods focused on representing average or steady-state conditions over an analysis time period that was long compared to the time scale of traffic dynamics. In these problems, travel times and volumes on links and routes can be considered to be constant over the analysis period, meaning there is no need to account for their variations over time. Within this approach, the relationship between the average traffic level on a particular network facility and its average travel time is represented by a VDF, which typically is either a closed-form mathematical function or sometimes a user-specified piecewise linear curve.

This static representation has notable advantages in that mathematical properties of traffic assignment models (e.g., existence and uniqueness of equilibrium) can be obtained relatively easily. Further, the computation time needed to find an approximate equilibrium solution is acceptable even given the relatively limited computing power that was common decades ago. The static assignment approach has usually been considered suitable for long-range planning and, indeed, in the prime of its development the main application of static models was to the planning of large capacity expansion projects. However, a static approach, by definition, cannot reflect either variations over time in traffic flows and conditions or changes over time in characteristics of transportation system components. Thus, static assignment is ill suited to analyze either traffic congestion effects at a fine-grained temporal level, or many of the measures that can be taken to address congestion.

Started in the late 1970s, research into DTA, by representing time variations in traffic flows and conditions, has tried to reflect the reality that traffic networks are generally not in a steady state. To retain the advantages of an equilibrium approach, the notion of user equilibrium needed to be extended in two ways. The first extension generalizes the static model's perfect traveler information assumption and route choice criterion, recognizing that travel times on network links vary over time. Travelers are assumed to know or anticipate future travel conditions along the journey (through learning from the past trials) and, in choosing an O-D route, they are assumed to minimize the O-D travel time that they will actually experience; this will depend on when they arrive at the various links along a route and on the travel times that prevail on the links at those specific future times. (This is in contrast to a route evaluation approach that considers only the link travel-times that prevail at the instant of departure from the origin. More discussion about experienced versus instantaneous travel times is provided in the section Instantaneous and Experienced Travel Times.) Since travelers who depart from an origin to a destination at different times will experience different travel times, the second extension recognizes that, in a dynamic approach, the user equilibrium condition of equal travel times on used routes applies only to travelers who are assumed to depart at the same time between the same O-D pair. (An important generalization of DTA simultaneously determines travellers' choice of departure time and route. This model can directly analyze phenomena such as peak spreading in response to congestion dynamics or time-varying tolls. It is still the subject of active research, and will not be considered further here.)

These two extensions, although seemingly subtle, marked a fundamental departure from static assignment in traffic representation and algorithmic design. The first extension required the development of efficient ways of describing time-varying network traffic condition and finding shortest (least time) routes in networks where link travel times change over time. Describing time-varying network traffic condition requires careful treatment to ensure that, while network traffic is propagated forward in time, key traffic flow properties—in particular, flow conservation (e.g., total amount of link outflow at the present time cannot be greater than the total amount of inflow from the previous time step) and the fundamental relationship of traffic flow variables (e.g., flow equals the product of average speed and density)—are preserved temporally. This requirement motivates various network loading models including both analytical- and simulation-based approaches.

The need to calculate the shortest route that minimizes actual experienced travel time (in lieu of the shortest route that minimizes travel time based on some snapshot measurement of link travel time) motivated the development of time-dependent shortest path (TDSP) algorithm. (In transportation network modeling literature, the term "path" is mostly used as path is formally defined in general graph theory. However, throughout this document the term path is only used in the context which relates to literature in which path is used. In other general descriptions, the term route is used.) More discussions of the difference between the two types of shortest path algorithms are presented in the section on Instantaneous and Experienced Travel Times.

The second extension disaggregated the equilibrium condition, so that the equilibrium condition is to be established for each departure time (typically ranging from a few seconds to several minutes) rather than over the entire analysis period; this result is known as dynamic user equilibrium (DUE).

Finding a DUE solution (i.e., a set of time-varying link and route volumes and travel times that satisfy the DUE condition for a given network and time-varying O-D demand pattern) is a nontrivial exercise, because each traveler's best route choice (that is, least experienced travel time route) depends on congestion levels throughout the journey, which in turn depend on the route choices and progress through the network of other travelers who depart earlier, at the same time or later (Figure 1). This interdependence means that solutions must be found through an iterative process, starting from some initial set of route choices, and gradually improving them. This improvement process can continue indefinitely; in realistic-sized networks, finding an exact equilibrium is challenging. Rather, the goal of many current DTA models is to find an approximate equilibrium that is sufficiently converged to true equilibrium for the application at hand and that is obtainable in a reasonable amount of time.

As shown in Figure 2, the most common method of finding equilibrium in DTA is to apply the following three algorithmic components in sequence iteratively, until a defined stopping criterion is met:

- **Vehicle departing at different time are assigned with different routes.**
- **Vehicles departing at the same departure time between the same O-D pair but taking different routes should have the same experienced travel time.**
- **Experienced travel time cannot be realized at departure, but only at the end of the trip.**

FIGURE 1 Characteristics of a DTA solution.

- Network loading: Given a set of route choices, i.e., routes and route flows, what are the resulting route travel times?
- Path set update: Given the current route travel times, what are the new shortest routes (per O-D pair and departure-time interval)?
- Path assignment adjustment: Given the updated route sets, how vehicles (or flows) should be assigned to routes to better approximate dynamic user equilibrium.

Although sharing a similar overall model structure, most DTA models differ from one another in how these components are implemented. In the route evaluation step, the effect (in terms of time-varying link–route flows and travel times) resulting from vehicles following a given set of route choices is determined through a network loading process. There exists a variety of analytical and simulation-based network loading approaches: analytical models typically use exit functions to predict how traffic propagates in the network, while most simulation-based approaches use some type of mesoscopic simulation approach that represents changes in traffic flow at a resolution of 5 to 10 s.

The next step, path set update, involves analyzing the results of the network loading. Based on the congestion pattern and travel times identified in the network loading step, the routes with the lowest experienced travel time between every O-D pair, for each departure time

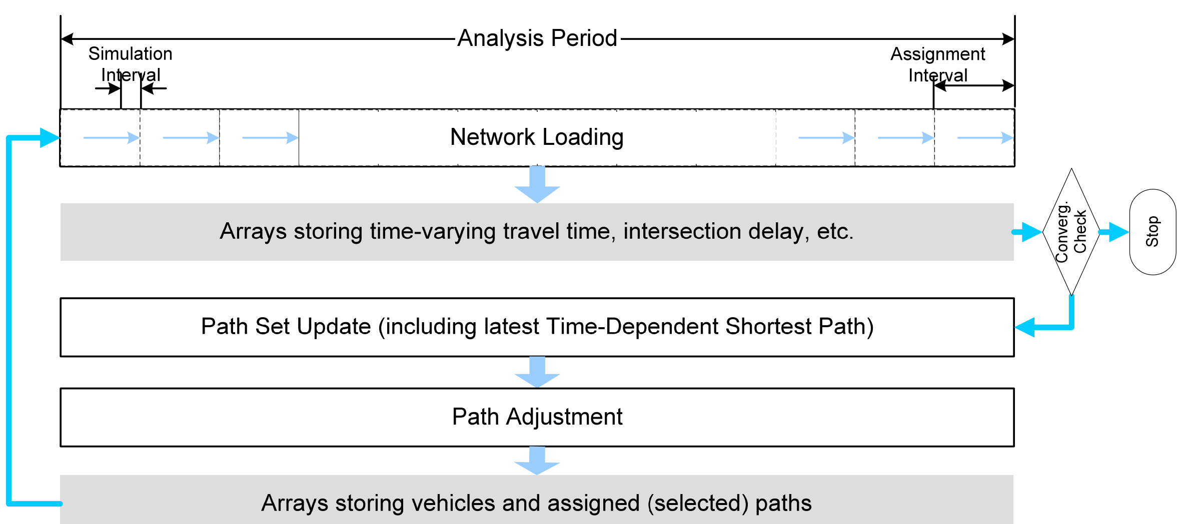


FIGURE 2 General DTA algorithmic procedure.

period (also called an assignment interval), are found by a TDSP algorithm. The newly found TDSP for a specific O-D pair and departure time period would be combined with all TDSPs found in previous iterations for the same O-D pair and departure time to form an updated path set.

Path assignment adjustment follows logically from path set update: if travelers shift their route choices towards the least experienced travel time routes (and away from longer routes), the assignment can be brought closer to equilibrium. Some care must be taken in this step since a major complication in finding an equilibrium solution is the interdependence between different travelers' route choices and travel times. If all travelers were to shift to the shortest routes found in the previous step, those routes would become highly congested and would no longer be shortest. Therefore, only some travelers' route choices should be adjusted, in order to avoid overcorrecting. Generally this step involves finding which routes in the set need to be increased with assignment flow–vehicles and which to be decreased, and by how much. Normally, the newly found TDSP along with several other good routes (with close to minimal travel time) are among those to be increased with flows. Underperforming routes (long travel time) are decreased with flow. It is also noteworthy that at this step, not all vehicles will select (or be assigned with) a new route. The adjustment made is only what is necessary in order to achieve equal travel among all routes in the current set.

After performing path assignment adjustment, the algorithm returns to the route evaluation step in order to determine the traffic pattern that would result from the new route choices (route flows). Thus, the three steps work in a sequential manner: the output of network loading provides the input for path set update; the output of path set update provides the input for path assignment adjustment; and the output of path assignment adjustment provides the input for network loading. These three steps are repeated until a stopping criterion is met. The algorithmic structure is illustrated in Figure 2.

The stopping criterion is typically computed at the end of the network loading step. Older DTA solution algorithms applied a solution approach called the method of successive averages (MSA). MSA impose a predetermined fixed amount of flow adjustment at each iteration, implying a slower convergence, but most recent algorithms employ the notion of relative gap as the stopping criterion. The following section offers further discussions of gaps.

DEFINING QUALITY OF DTA MODEL OUTPUTS

The quality of simulation-based DTA model outputs can be generally judged from three dimensions: convergence, sensitivity, and realism of traffic dynamics. These are individually discussed below.

Convergence

Almost all equilibrium-seeking DTA algorithms adjust the route assignment using an iterative solution procedure. As illustrated in Figure 2, at every step of the iteration, time-dependent link travel times from network loading are input to the TDSP routine to calculate minimum experienced time routes for every O-D pair and departure time period (assignment interval). The routes from the newly solved TDSP are combined with the existing route set and the flows between every O-D pair and departure time interval are then redistributed along the updated set.

The procedure is said to have converged, or reached an acceptable approximation to a user equilibrium solution, when there is no substantial incentive for a user to shift routes, i.e., a traveler won't improve his or her travel time by selecting another alternate route. This translates to no significant changes in flow pattern or experienced travel time after multiple iterations.

Thus, a DTA algorithm can test for convergence by calculating various metrics which measure the deviations in flow patterns or congestion indexes (such as experienced travel times) between successive iterations and checking to determine whether they are less than a prespecified tolerance level. The tolerance level is a measure of the amount of error (with respect to perfect equilibrium) permitted in the ultimate solution. That level is a measure of the amount of deviation of the ultimate solution from a true equilibrium solution. Even though ideally the tolerance would be very low, note that a lower tolerance leads to increased computational time. Therefore, a trade-off between convergence and computational time must be made when choosing the tolerance level.

A commonly used, but potentially problematic, convergence criterion in practice is the absolute change in link flows from one iteration to the next, which should be less than a prespecified tolerance level. A small change in link flows across iterations may indicate that most users are satisfied with their current route choice, but may also be an outcome imposed by the algorithm (e.g., MSA) and may, thus, have nothing to do with travel time. The outcome may simply indicate that the algorithm is stuck, not being able to find further improvement. In other words, MSA types of approaches guarantee that links flow changes continue to decrease over iterations, but this is an algorithmic construct unrelated to the equilibrium requirement of minimizing experienced travel time for all used routes.

Another more intuitive and sound convergence metric based on route times is termed the relative gap and is increasingly used as the convergence criterion. The relative gap is a rather common stopping criterion also used by static traffic assignment models. The typical definition of the total relative gap is

$$rel_{gap} = \frac{\sum_t \sum_{i\in I}\left(\sum_{k\in K_i} f_k^t \tau_k^t\right) - \sum_t \sum_{i\in I} d_i^t u_i^t}{\sum_t \sum_{i\in I} d_i^t u_i^t}$$

Where t is a superscript for an assignment interval or a departure time interval, i is a subscript for an origin-destination pair and k is a subscript for a route. Subscript i represents the set of origin destination pairs and K_i denotes the set of used routes connecting the origin destination pair i. f_k^t represents the flow on route k departing at assignment interval t, τ_k^t is the experienced travel time on used route k for assignment interval t. d_i^t denotes the total flow for origin-destination pair i at time interval t and u_i^t is the shortest route travel time for origin-destination pair i and departure time interval t. The numerator is the total gap, which measures how far the current the assignment solution is to the ideal shortest route time. Taking the total gap divided by the total shortest path times describes the ratio of the total gap to the total shortest path times. The intuition of the relative gap is that if all used routes have travel time very close to the shortest route travel time, then the numerator will be close to zero, and the relative gap value will be small. Since the travel time on all used routes will always be greater than or equal to the shortest route, the value of relative gap will never be negative. In most DTA applications, the solution is assumed to have converged to an equilibrium solution when the relative gap is less than a prespecified tolerance level.

Solution Sensitivity and Stability

Sensitivity and stability are two notions that relate to the way in which a problem's solution varies as a function of parameters that characterize the problem. These notions can be made precise, and the study of the sensitivity and stability properties of mathematically-formulated traffic assignment problems is already a well-established area of research for static models. Here we consider these notions in a more informal way, to describe the expected behavior of DTA problem solutions as the problem itself is changed. For example, given a DTA solution for a particular O-D pattern and network, what might be expected of the solution for a modified problem in which a link is added or changed? Rather than answering this question directly, we turn it around: if the solution to a modified problem exhibits unexpected features, this may indicate that a poor approximation to equilibrium has been computed for the original or the modified problem, or both.

To begin, recall that DTA models apply an iterative algorithm to compute an approximation to dynamic user equilibrium. From one iteration to the next, time-dependent route choices, link flows, and times change as the algorithm progresses towards an equilibrium solution. At some point, the convergence criteria (which may involve a gap function or be based directly on the changes in the computed solution between iterations) are satisfied and the algorithm terminates. However, if the algorithm were to continue for iteration, there would almost certainly be further changes in the computed solution. The magnitude of these changes depends on the strictness of the convergence criteria, and can be thought of as a measure of the precision of the solution.

Suppose now that a base network is modified to represent alternative projects under consideration, such as for example introduction of ramp metering, HOV–HOT lanes or new capacity. The model is solved with each modified network until the convergence criterion is satisfied. The output flows and times for each alternative are then compared, either directly or via any of a variety of aggregate measures of effectiveness. For this comparison to be valid it is imperative that the individual equilibrium solutions be computed to a precision that is greater than the differences between the solutions of the alternative problems; otherwise, any real differences between the alternatives will be lost in the imprecision of the calculated solutions. Failure to recognize this requirement can result in incorrect evaluation and ranking of alternative projects under consideration. (Strictly speaking, this can be considered more an issue of algorithm solution quality than of DTA problem sensitivity. Still, equilibrium solutions to individual problems must be computed to sufficient accuracy before comparisons between solutions to different individual problems can be validly carried out. In the context of static traffic assignment, this requirement is referred to as *link flow stability*.)

One should note that the appropriate solution convergence criterion can depend on the actual real world application. Among other things, it will depend on the differences between the alternative projects under consideration: the flows and times computed for projects with roughly similar impacts will need to be more precise than those for projects that are more strongly contrasted.

One way of characterizing problem sensitivity might be termed locality of impacts—a minor change in the network should generally not typically have large impacts on flows or conditions far from the location of the change. Existence of significant nonlocal impacts may be evidence of a poorly computed DTA solution. For example, a minor change, such as a speed limit change on a particular link, would not be expected to significantly affect flows and conditions far from the link in question; DTA model outputs that showed such effects should be closely examined. Of course, it is difficult to define minor changes and significant nonlocal impacts in an unambiguous and universally applicable way. However, when in doubt about an apparent nonlocal impact, it is best to verify the quality of the computed solution, for example by re-solving the DTA model with more stringent convergence criteria on the base and project networks.

The stability and sensitivity properties of DTA models are still areas of active research. Thus, when a DTA problem is perturbed in a minor way, it may be theoretically possible for the perturbed problem to have a very different DTA solution. (Daganzo showed that when the traffic flow model incorporates physical queues, the solution may exhibit chaotic behavior, but the extent and likelihood of such occurrence is largely affected by the DTA model implementation.) However, preventing this from happening for most practical situations is what the model developers should strive for and demonstrate. Practitioners should, therefore, treat the above guidelines as suggestions more than rules, but should remain alert to the possibility of poorly computed DTA solutions.

Realism of Traffic Dynamics

The primary outputs from static traffic assignment models are unique time invariant link flows and nonunique route flows. From the steady-state link flows, numerous other parameters, such as steady state speeds and V/C ratios, can be calculated. Note that link flows obtained from static models are viewed as average steady state conditions and may not adequately represent the impact of time dependent dynamics that are seen during peak periods. Another important issue with static link flows is consistency with measured real world traffic parameters such as vehicle counts, as model-predicted link volumes can be greater than link capacity.

In contrast, most simulation-based DTA models have the ability to provide trajectories of vehicles for every origin-destination pair and every departure time interval. From the knowledge of vehicle trajectories, detailed information characterizing the temporal and spatial dynamics of travel times and other congestion indices can be extracted. For example, it is easy to evaluate the average travel times, speeds, or densities on links for any given time period. Also, outputs obtained from DTA models are likely to be more consistent with real world traffic data compared with those from static assignment models. It is also relatively easy to obtain vehicle counts at a particular location within a specified time interval from DTA outputs by tagging in the model all vehicles whose trajectories pass through that particular location during the specified time interval. These features facilitate the validation of a DTA flow pattern using data from field observation.

Some of the important traffic flow characteristics that can be obtained from DTA outputs include traffic counts and speeds at specified detector locations; time-varying speed–travel time profiles on links, link sequences or routes; queue lengths; average density and flow; and time-varying density or volume profiles. One important feature of DTA model outputs is that obtaining time-varying speed profiles enables more accurate estimation and evaluation of air quality impacts or emissions from various projects.

Most prevalent DTA models apply mesoscopic traffic simulation for evaluating route choice decisions. Mesoscopic models can reasonably depict the aggregate or macroscopic properties of traffic flows (such as average speed, density and flow rates) without having to examine the interaction of individual vehicles in detail. Mesoscopic simulation provides greater computational efficiency that allows a much faster simulation (if compared with microscopic simulation on the same network size) or allows application to a much larger network (if compared with microscopic simulation for the same analysis period). Different mesoscopic models, however, incorporate different logic and may, in some cases, predict rather different traffic dynamics for a given network and demand situation. Commonly used metrics (i.e., measures of effectiveness) for evaluating microscopic traffic simulation models can generally be applied to mesoscopic models with the understanding that the coarser representation of traffic dynamics in mesoscopic models may not permit generation of the fine-grained statistics that could be produced by microscopic models. Both models should, nevertheless, conform to macroscopic traffic flow fundamental relationships.

STATIC AND DYNAMIC ASSIGNMENT IN A ONE-SHOT SIMULATION

Although this document is not aimed at standardizing or unifying the terminology usage for DTA, the descriptions follow widely accepted definitions of key terms from literature. For example, the term network loading refers to the representation of the movement of vehicles along specific routes from origin to destination, and the determination of the link and route volumes and travel times that result. In DTA models, network loading may be accomplished using analytical procedures or by simulating vehicles' movements along their routes as they carry out their journeys through the network, giving rise to the resulting link flows and travel times. Network loading takes place at each iteration of an assignment model, but does not by itself result in equilibrium. The term assignment encompasses all three algorithmic steps discussed in Figure 1, in which routes are updated and flows and times are adjusted to compute a modified set of route flows which are then evaluated by network loading or simulation, which allows the quality of the assignment (route flows) to be evaluated vis-à-vis the equilibrium property. In other words, the term assignment is closely related to the equilibrium searching mechanism (as noted in the preface).

One reason for the definition conventions adopted in this document is the difficulty which can arise with the broadest possible terminology. For instance, DTA-related terms can be used to describe rather different procedures and concepts that involve assigning a route to a vehicle. One example is the one-shot simulation approach that regularly updates and assigns routes to newly-generated or en route vehicles. The standard modeling mode of most microscopic simulation models can be connected to this approach. In this modeling mode, the term "assignment" refers to associating vehicles with a route. If the route set and flows are predefined and remain unchanged throughout simulation (see Figure 3), this has actually been referred to as static assignment in certain software documentation, but that use of assignment is distant from the concept explained in this primer.

A more advanced approach has shortest routes regularly updated based on prevailing traffic conditions (i.e., instantaneous link travel times) and has these routes assigned to newly generated vehicles at the start of the trip. This is sometimes referred to as dynamic assignment (see Figure 4) in certain software documentation, although this usage would not follow the terminology conventions adopted in this primer.

In a more flexible one-shot method, in addition to the above assignment approach for newly-generated vehicles, each vehicle (or a subset of vehicles called familiar travelers) reevaluates the current route at each decision node (or way-points), based on current (instantaneous) link travel times (refer to the sections on From a Traffic Engineering Perspective and Instantaneous and

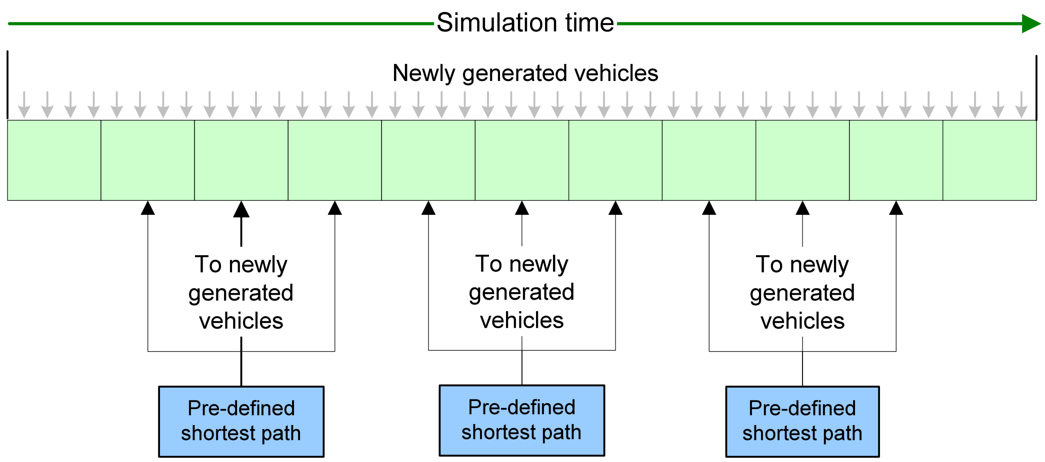


FIGURE 3 Static assignment in a one-shot simulation.

Experienced Travel Times for discussions on instantaneous versus experienced travel times). A decision node is one at which there is at least one feasible route to the destination on each of two or more of the outgoing links of the node. This approach allows the traveler to abandon the current route for a better one for the remaining trip, as a result of changes in link travel times since the last route choice was made (at an earlier decision node, or at the origin node). This method is sometimes referred to as one-shot dynamic assignment with feedback (see Figure 4) and, again, one needs to distinguish the difference between this method and the notion of assignment defined in this primer.

Although the term assignment is used in above one-shot, noniterative simulation approaches, it would be more precise to call this network loading with incremental route updating, since it does not attempt to achieve user equilibrium and does not reach consistency between the travel time used in route generation and the experienced route travel time.

In both above cases, travelers select the shortest routes, calculated based on instantaneous travel time. The implication is that their choices are based on some myopic decision rather than anticipating the traffic condition along the route so as to minimize the actual experienced travel

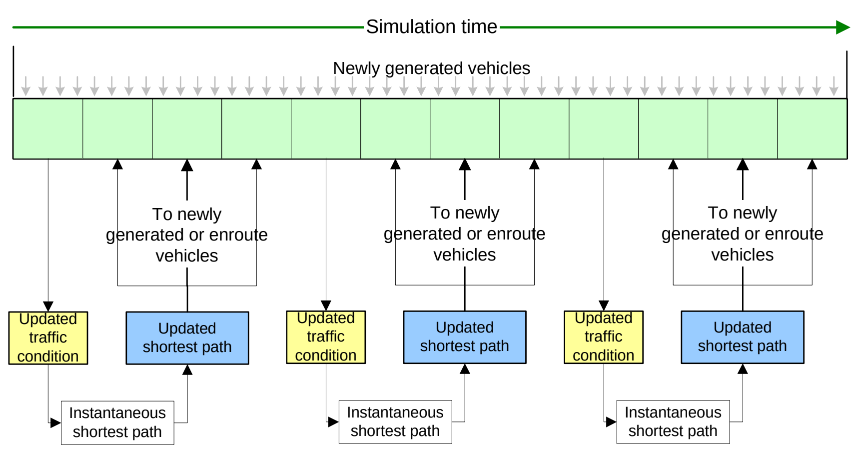


FIGURE 4 Dynamic assignment (with feedback) in a one-shot simulation.

time. While perhaps true for some travelers accustomed to making travel decisions based solely on traveler information sources, the majority of the traveling public may incline toward choosing from a finite number of routes that the traveler has learned and, thereby, to minimize his experienced travel time. In the simulated world, if all vehicles were to select myopic routes, the network congestion is likely to be overstated (i.e., the simulated traffic would be more congested than would be likely to occur in the real world, in which the shortest minimal experienced travel time route is sought by travelers), because travelers do not optimize their route choice based on the eventual outcome, but on short-term information.

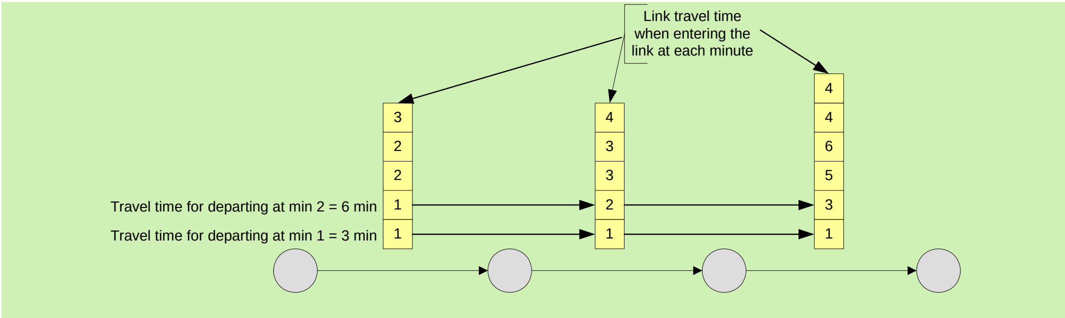
Another way to put this conclusion, as long as experienced travel time plays a significant role in the route choice criteria, which it does in most real-life situations, the iterative equilibrium solution provides the desired consistency between the route choice and the resulting finished). Any noniterative solution or non-TDSP path generation mechanism, be it based on random utility theory or otherwise, must necessarily have some degree of inconsistency between the time used for route choices and resulting experienced travel times, particularly when modeling large congested networks. (Some models have also been proposed in which random utility maximization is embedded within an equilibrium framework in order to provide the desired consistency on travel times.)

INSTANTANEOUS AND EXPERIENCED TRAVEL TIMES

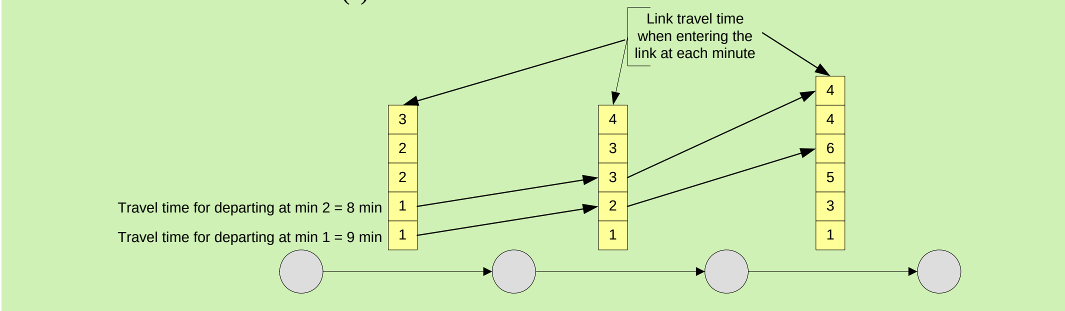
As can be seen from the previous discussions, experienced travel time plays a key role in establishing a dynamic equilibrium condition that is consistent with a traveler's route choice decision. The notion of experienced travel time departs from the notion of instantaneous travel time that is typically applied in the static assignment context as well as in the micro-simulation context for one-shot assignment-simulation modeling. What is the difference? In the example illustrated in Figure 5, there are four nodes and three links comprising a simple one-way network. The stack of values represents the different times to traverse a link when departing from the upstream node (and entering the link) at different times. Time-varying link travel time is common during peak hours due to congestion buildups.

As an example, the time needed to traverse Link 1 is 1 time unit when departing the upstream node at time 1, and 3 time units when departing the upstream node at time 5. Similarly, the travel time for Link 2 is 1 and 2 time units when departing the Link 2 upstream node at times 1 and 2, respectively. The instantaneous travel time for the entire route at each different departure time is calculated by summing up the link travel time corresponding to that same departure time for all links comprising the route. As an example, for vehicles departing at time 1, the travel time is 1 + 1 + 1 = 3 time units; for vehicle departing at time 2, the travel time is 1 + 2 + 3 = 6 time units.

The experienced travel time calculation accounts for the time needed for traversing one link, and looks up the downstream link travel time based on the time of entering that downstream link (assuming that traversing a node takes no time). Based on this approach, the travel time for the route when starting at Departure Time 1 should be 1 + 2 (vehicle entering Link 2 at Time 2 so the Link 2 travel time is 2 time units), plus 6 (vehicle arriving at Link 3 at Time 4, so the Link 3 travel time is 6 time units). The experienced travel time is 1 + 2 + 6 = 9 time units. Similarly for departure at Time 2, the instantaneous travel time is 6 time units, versus the experienced travel time = 8 time units.



(a) Instantaneous travel time calculation



(b) Experienced travel time calculation

FIGURE 5 Experienced travel time versus instantaneous travel time determination.

Clearly, these two methods produce different route travel times and, likewise, differing results for the shortest route(s). The questions to bear in mind are what is the conceptual and practical significance of such a difference? From the standpoint of evaluating behavior with respect to route travel time, what is the possible interpretation of both methods? From a traffic analysis standpoint, which one is a more likely response of a traveler to the traffic condition?

The shortest route obtained based on the instantaneous travel time calculation has the minimum travel time based on the snapshot of the link travel times prevailing at departure. However, because link travel times change dynamically (due to congestion), that route does not necessarily result in minimal experienced travel time because there is no provision in this procedure to reflect the anticipation of congestion that is to occur at a later time down the road (e.g., congestion caused by vehicles departing later but entering the same link, one which the vehicle being modeled is still traversing).

Assigning vehicles with an instantaneous travel time route is not necessarily incorrect, but its corresponding underpinning assumptions need to be understood. The route choice associated with instantaneous travel time may be interpreted as (*a*) travelers know what the shortest route is at departure through pretrip information (e.g., 511, news, or website) or en route in-vehicle information system (if the traveler is to take another route when en route); or (*b*) from day to day, travelers do not assess the route travel time from the experience standpoint, but rely instead on the traveler information.

In contrast, the shortest route obtained based on the experienced travel time calculation method will yield a time-dependent shortest route with minimal experienced travel time. This assumes that travelers are willing to seek routes that minimize their experienced travel time instead of the route that appears to be the best only at the departure.

Obviously, with the prevalence of traveler information systems, one could argue that both types of travelers may co-exist in the traveler population. Some models do have the capability to include multiple user classes and so represent travelers with different information accessibility and route choice behavior to be jointly modeled, but the ATIS market penetration is likely to remain relatively low.

The following example demonstrates how instantaneous travel time and the experienced travel time calculations generate different routes. This network consists of six nodes, seven links and one O-D pair. All vehicles depart at node 1 and head toward node 6. The time-varying link travel times are assumed to be provided by the network loading procedure as discussed previously. The time-varying link travel times are specified for each link with each number starting at the bottom of the stack representing the link travel time when entering the link at time 1, 2, …, etc.

There are three routes connecting nodes 1 and 6, including Routes 1-2-4-6, 1-2-5-6, and 1-3-5-6 as shown in Figure 6. Following the instantaneous travel time calculation, one would obtain travel times for Departure Time 1 for Route 1-2-4-6 to be 3 time units (adding the link travel time from the bottom cell of the stack for all links), for Route 1-2-5-6 to be 4 time units, and 1-3-5-6 to be 5 time units. Consequently, the Route 1-2-4-6 is the shortest instantaneous travel time route (see Figure 7).

Taking the same network and time-varying link travel times but applying the experienced travel time calculation approach, the experienced travel times for the three routes become 9, 5, and 4 time units, respectively. Consequently, the shortest experienced-travel time route is 1-3-5-6.

The same process can be applied to find the shortest routes for Departure Time 2. As shown in Figure 8, the instantaneous travel time and experienced travel time calculation approaches yield rather different travel time. The resulting shortest routes happen to be the same for Departure Time 2

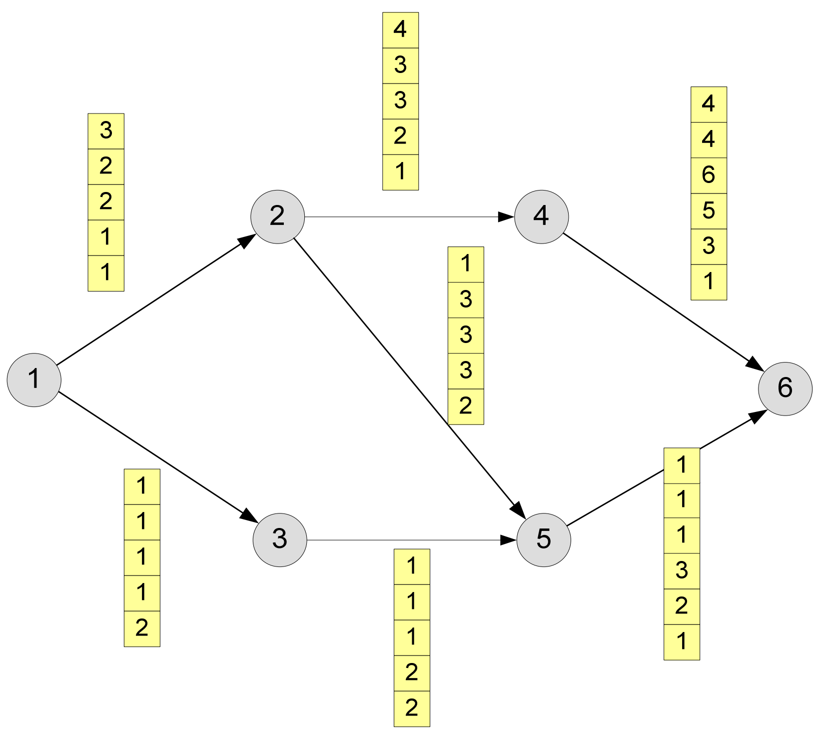


FIGURE 6 Sample network with time-varying link travel times.

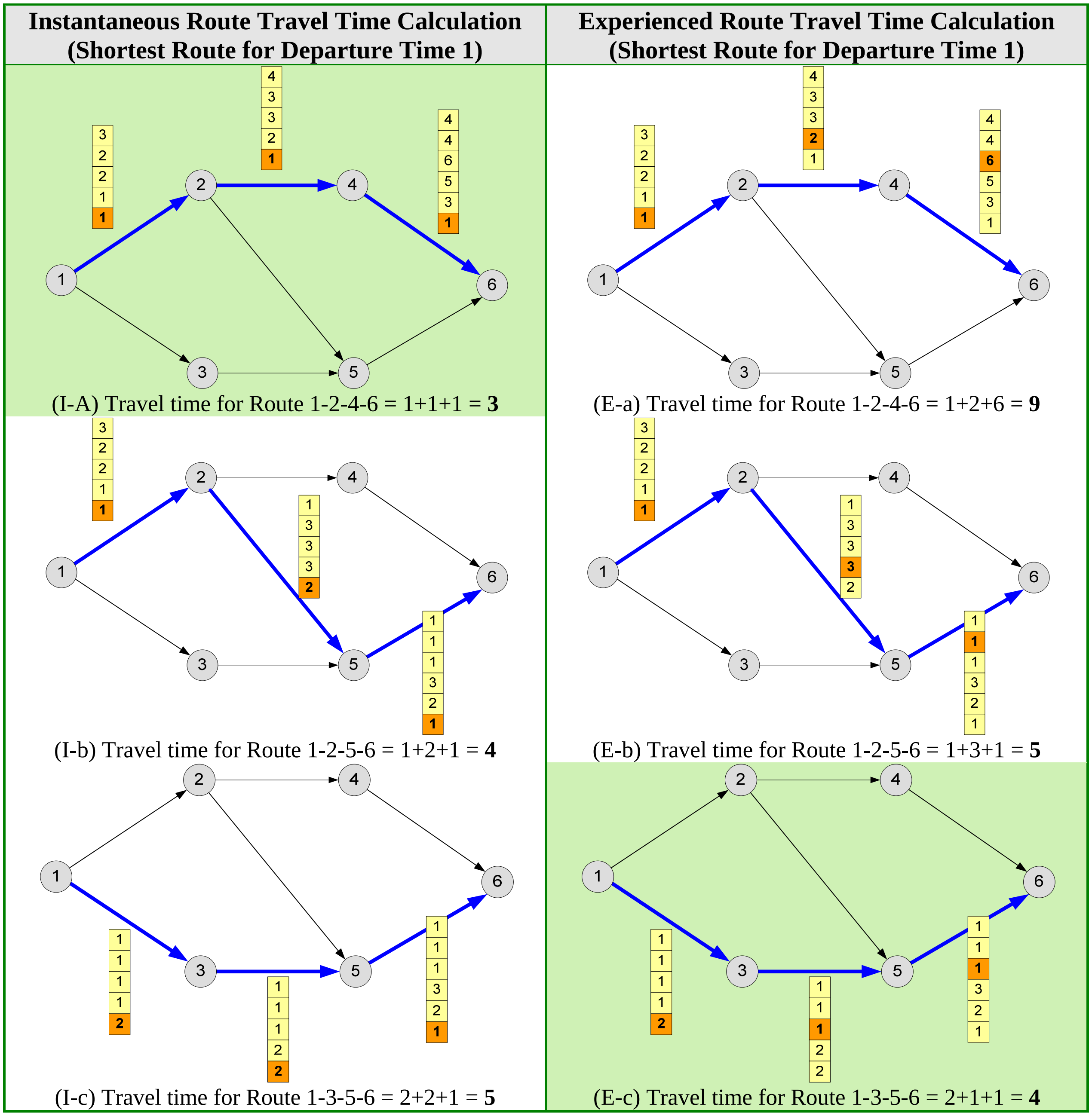


FIGURE 7 Different shortest routes obtained by instantaneous travel time and experienced travel time approaches (Departure Time 1).

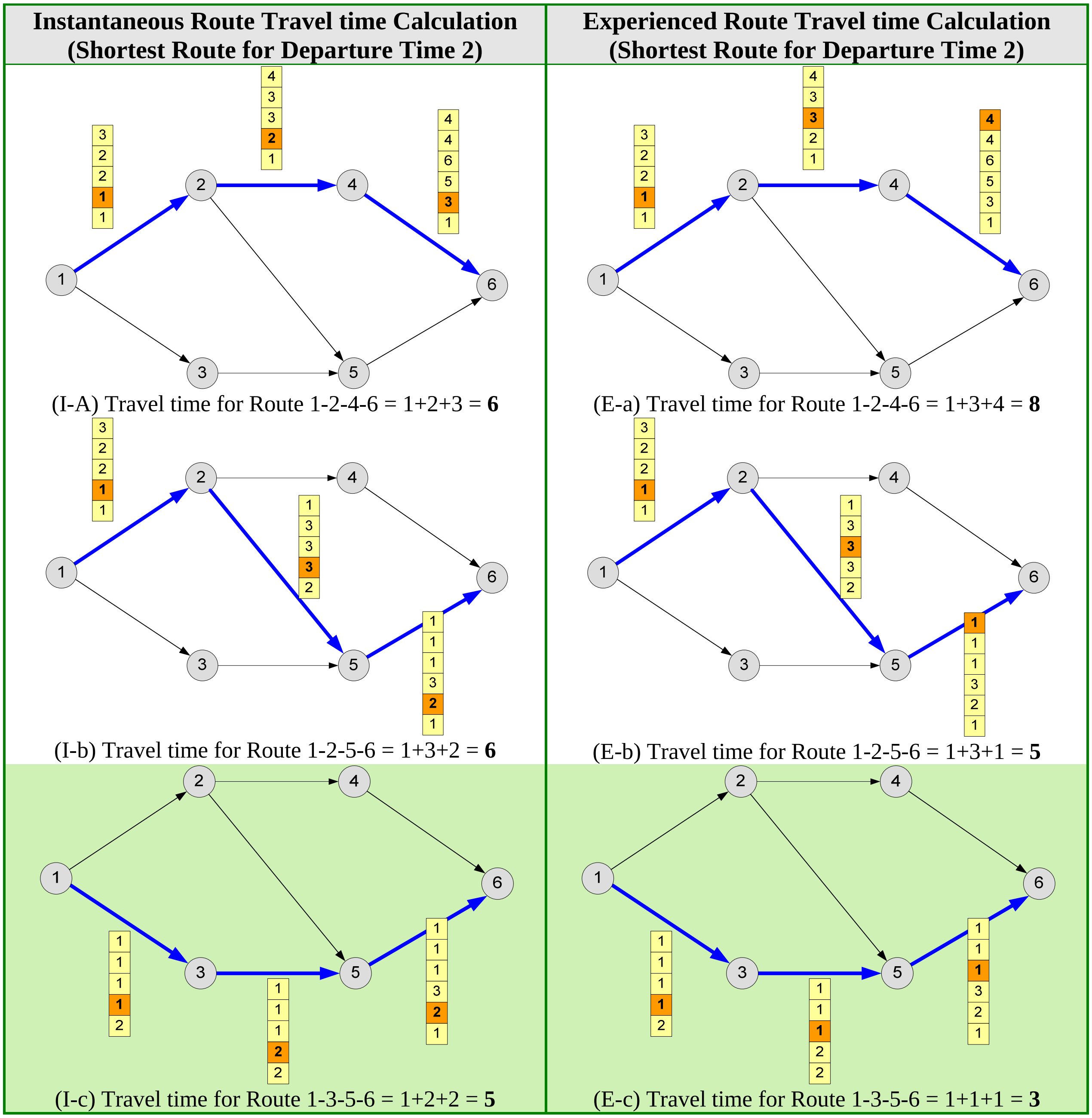


FIGURE 8 Different shortest routes obtained by instantaneous travel time and experienced travel time approaches (Departure Time 2).

but, in a real-world traffic network, shortest routes generated by the two approaches are likely to differ significantly.

DISEQUILIBRIUM VERSUS NONCONVERGENCE

Although the discussion here has emphasized equilibrium as a fundamental modeling principle, active research is ongoing in dynamic traffic network disequilibrium. While equilibrium assumes that travelers are well informed and make rational choices, disequilibrium research focuses on the learning process by which travelers' experiences in one period (typically a day) affect their expectations and decisions in subsequent periods (*2–4*). Disequilibrium analyses acknowledge that travelers do not have perfect information about network travel conditions, because of basic unfamiliarity with the network or random variability in conditions, and study the ways in which network conditions evolve from day to day as travelers continually adjust their behavior based on prior experiences. In disequilibrium analyses, it has been shown that, depending on the assumptions, the system can evolve in a variety of ways: in some cases, the learning process eventually converges to user equilibrium; in others, flow patterns and conditions repeat themselves over multiple days; and in others, the system exhibits chaotic behavior. Much less is known about the disequilibrium properties of dynamic networks, which remains an area of active research.

It is sometimes claimed that the unconverged results from an equilibrium model or the results from a one-shot, noniterative model represent a disequilibrium situation in the sense described here, which is not correct. The unconverged results simply correspond to outputs from an intermediate step in an algorithmic process designed to compute equilibrium, whereas the disequilibrium is the result of a detailed and careful representation of the evolution of network and demand conditions, and the development of travelers' learning processes over time.

Decision Making for Applying Dynamic Traffic Assignment Tools

The decision to employ DTA versus other modeling approaches for the application of interest needs to be based on careful consideration of several factors. General considerations include the type of analysis (project) of interest, the benefits provided by the DTA approach versus other comparable approaches, the limitations of DTA modeling tools, and the resources available (e.g., data, time, budget, personnel, hardware). This chapter discusses decision factors when considering applying DTA for the problem at hand.

WHAT APPLICATIONS FIND DTA MODELS ADVANTAGEOUS?

As previously discussed, DTA models offer dynamic network equilibrium modeling capability that is not available in static traffic assignment and most microscopic traffic simulation models. The primary application areas for DTA models can be identified as operational planning and real-time operational control of vehicular traffic systems. The former is more relevant to the target audience for this publication, whereas the latter application area is more relevant for traffic operations engineers who manage daily traffic.

Operational planning (or planning for operations) is aimed at making planning decisions for major operations, construction, or demand management actions that are likely to induce a temporal or spatial pattern shift of traffic among different roadway facilities at a corridor–networkwide level. Such types of projects include, but are not limited to (*a*) significant changes of roadway configuration (e.g., change downtown streets from one-way to two-way configuration), (*b*) freeway expansions, (*c*) construction of a city bypass, (*d*) adding or converting HOV–HOT lanes, (*e*) integrated freeway or highway corridor improvement–construction, and (*f*) travel demand management strategies such as peak spreading or congestion pricing.

In the above cases, the underlying characteristic motivating the use of DTA is that the change in demand or network conditions is significant enough to induce spatial or temporal traffic flow shifts, as a consequence of travelers wanting to use different routes or departure times in response to the demand or network changes. Hence, the new vehicle routes need to be reestimated by the DUE procedure. Because the model needs to be sensitive to congestion, DTA appears to be more suitable to addressing this problem than the static traffic assignment approach. Furthermore, because such traffic flow pattern shifts are likely to take place over a larger geographic area and over a longer time period, simulation-based DTA may be more cost-effective than microscopic traffic simulation in capturing the areawide traffic flow shift.

In recent years, increasing numbers of practitioners and researchers have applied or are applying DTA to support modeling efforts related to the following areas.

1. Interstate freeway corridor management (*5–8*);
2. Integrated corridor management (*9–11*);
3. Traffic management for short- or long-term network disruptions (*12–17*);
4. Managed lanes (*18*, *19*),
5. Downtown traffic management and street configurations (*20*, *21*);
6. Incident management (*22–24*);

7. ITS evaluation and information provision (*25–29*);
8. HOV–HOT lanes or congestion pricing (*30–32*);
9. Emergency management (33–41);
10. Air quality inventory analysis (*42*, *43*);
11. Campus transportation master plan assessment (*44*);
12. Multi-resolution (macro, DTA and micro) regional traffic models (*6*, *45*),
13. Integration with existing travel demand models (*46*) or activity-based models (*47–49*), and
14. Network reliability (*50*, *51*).

In the context of real-time operational control, DTA models are relevant for transportation engineers working on large-scale real-time traffic management or information provision problems. Real-time DTA models are appropriate to address these types of problems in a systematic manner because they provide capabilities to estimate future network conditions (flow patterns) that will result from a particular traffic management or information provision strategy. They are capable of updating the network states and developing new traffic management or information provision strategies based on real-time field data. Although there are advanced real-time DTA models, some important issues still need to be addressed to fully achieve effective deployment. For example, deployable models need to be computationally efficient to provide timely solutions (*52–54*).

WHAT TO EXPECT FROM DTA MODELS

Given the time-dependent nature of demand and network characteristics, DTA models are used primarily to estimate dynamic traffic flow pattern over the vehicular network. That is, DTA models load individual vehicles onto the network and solve for their routes so as to achieve systemwide or traveler class objectives. These objectives are based on the project characteristics. For example, planning studies typically require the estimation of the user equilibrium flow pattern which results when travelers cannot improve their travel times by unilaterally changing routes. Other studies may require the prespecification of the vehicular routes based on normal conditions or the real-time rerouting of vehicles. This characteristic is particularly important for studies involving ITS technologies, the evaluation of the effects of special and short-term events, or the provision of information. Hence, advanced DTA models should provide capabilities to handle different classes of travelers depending on the project characteristics.

DTA models provide a vast array of detailed outputs that describe time-dependent network states. They typically provide time-dependent system-level and link-level statistics. Examples of output files include system-level travel time, miles traveled, and stop times. There are also output files that include time-dependent link-level travel times, speeds, densities, queues, and stop times. In addition, DTA models provide a graphical user interface (GUI) to display these network characteristics and statistics graphically. Some GUIs provide capabilities to edit or build the project inputs and to handle more than one project or scenario simultaneously. Most DTA models output the trajectories followed by all the vehicles. This information can be used to develop any nonstandard statistic that the analyst may need.

CAUTIONS FOR USING DTA MODELS

In the context of modeling large-scale dynamic vehicular traffic networks, DTA models have reached sufficient maturity to provide meaningful results within acceptable solution times. However, several precautions are provided herein for consideration in applying DTA models.

First, DTA models are not the universal cure that can cost-effectively address all types of problems at hand. DTA models take more time and resources to construct and calibrate (as compared with static traffic assignment models) and represent traffic dynamics in a coarser granularity (as compared with microscopic traffic simulation models.) Practitioners are advised to match the choice of modeling approaches to the problem at hand. For long-term planning, for example, the available level of input data required for DTA may not be available, so that one may have to make many assumptions in order to construct such models. If the additional detail and precision in their output data compared to conventional network forecasting is not beneficial for the modeling question in mind, then it may not be worth the additional modeling effort. For a smaller bounded area in which a detailed representation of multiple modes (e.g., auto, transit, pedestrians) and facilities (roadways, parking, crosswalks, etc.) are required, microscopic models may be more appropriate and useful.

Most existing DTA models focus on route choice, and relatively few are implemented for departure time or arrival time choice. However, a wealth of academic literature can be found (*55*–*59*) that addresses the choice of departure or arrival time. Existing DTA models simulate transit (typically buses, and possibly rail-based transit), but dynamic transit assignment is generally not incorporated into existing models.

The existing simulation–assignment paradigm is vehicle-based instead of person-based, and travelers are generally considered homogeneous in many choice dimensions. In other words, the modeling of heterogeneity in individual sociodemographical attributes and choice preference are still a research topic and is not fully addressed by most existing equilibrium-seeking DTA models. Methodological concerns in this regard include possible deviation from the equilibrium condition (so as to negatively impact the stability and consistency of the model outputs) and the exacerbation of computational intractability. Limited recent research, however, indicates a promising direction for incorporating heterogeneity (*60*).

DTA models can, to a certain degree, represent the effects (capacity, delay, etc.) of most existing traffic signal control logics (pretime, actuated, stop signs, etc.) in their mesoscopic simulation logic. However, this representation is relatively simplistic. Most of the existing DTA models represent the major features of the different control types without including the exact logic and settings of the existing commercial controllers. If the application at hand requires detailed representation of signal timing and coordination or of intersection or ramp vehicle interactions, DTA models may not be as effective as microscopic simulation models. Furthermore, if the application requires signal timing optimization, this may imply the simultaneous optimization of routes and controls—generally referred to as the network design problem. Existing DTA models generally do not directly perform this analysis; additional customization is usually needed.

In some cases the traffic flow models within a DTA model may produce counterintuitive results that are difficult to explain because they are the consequence of interactions taking place over the entire network and across multiple time periods (spatiotemporal interactions). Counter-intuitive results should not be treated as incorrect out of hand: the motivation for building complex models is because the system in question is too complex to be evaluated intuitively.

What a planner may find challenging is to interpret the simulation results by relating flow, density, speed, and queue for a single location or a sequence of roadway segments. As an example, flow rate increases with increasing density and decreasing speed during the onset of congestion, but the flow rate will start to decrease once the speed is reduced below a certain threshold value. In other words, a single flow rate figure corresponds to two separate traffic states—free flow and congested situations. As such, speed or density is a better descriptor of congestion than flow rate. More discussions on this topic are provided in section on Calibration Methods.

The DTA GUIs are helpful for displaying the results, comparing projects and scenarios, and analyzing the network states. However, clearly illustrating why one project or scenario is better or worse than another is a question of results interpretation that cannot be automated or simplified. This difficulty is of course equally true for static assignment models, though the higher sensitivity of DTA models can bring this issue to a higher level. For example, adding capacity to the network may negatively affect some regions of the network or even the entire system. This occurrence is known as the Braess paradox and can easily be explained in the context of static traffic assignment but, because the effects may be spread out over the entire system and over several time periods, there may not be a straightforward way to show where and why the network is negatively affected.

At times the analyst may wish to convert quickly from an existing travel forecasting model to a DTA model. Most DTA model developers offer certain utilities for streamlining the initial network creation process. However, manual refinement of the model is always necessary to ensure data set quality. Most DTA models come with a GUI and simulation outputs are stored either in plain text files or databases. In many cases, post-processing tools may be made available by DTA model developers or have to be developed by the model user to extract desired statistics. The learning curve, required knowledge, time, and software development also need to be considered when planning DTA model development.

DTA models deal with large-scale dynamic networks where the network states are the result of many network and demand factors interacting over time and space. Calibration of the model is a critical step that requires knowledge of the DTA model and actual traffic conditions at the site of interest. A wealth of model calibration research can be found in the literature (*61–63*). However, most DTA developers provide guidelines and utilities to facilitate the calibration activities. Without loss of generality, practical frameworks and procedures for the calibration of DTA models are provided in section on Model Validation and Calibration.

Simulation-based DTA models generally do not strictly conform to mathematical properties such as uniqueness or existence of the equilibrium condition. This is due to the nonlinear, dynamic and potentially stochastic traffic conditions arising from complex interactions of human drivers, the control system and the roadway environment. A previous overview (*64*) describes situations where properties such as the convergence and uniqueness of the DTA solutions may not necessarily be the prerequisite from a practical standpoint due to the well-known ill-behaved nature of this complex problem. It is nonetheless argued that such properties are desirable and merit rigorous pursuit in the interest of stability and consistency of model outcomes. If DTA solutions are determined without convergence the results produced may be arbitrary or random network states. From a scenario comparison standpoint, this result is not desirable as many factors other than the scenario per se impact the final outcomes and an analyst has no way to tell whether the change of traffic condition in the compared scenario is strictly due to the scenario or is affected by other artifacts introduced by the solution algorithm. This issue is well known in static assignment modeling and holds true for DTA.

DECISION MAKING IN SELECTING DTA MODELS

It lies in the analyst's hand to judge, from a technical standpoint, whether DTA is the most suitable modeling tool for the problem at hand. However, a project or program manager may also need to assess other factors, such as schedule, budget, available data and computing resources. DTA-based analysis may not necessarily require more time or resources to perform compared with travel forecasting models or microscopic models. Since it is a relatively new emerging technology, however, it takes education for transportation agencies or consulting professionals to understand the pros and cons of DTA models as well as being able to match DTA models with suitable applications that include proper decision aids.

Several prior publications provide valuable decision-making guidelines for selecting traffic analysis tools, ranging from sketch planning tools to DTA to microscopic models (*65*, *66*). The general decision dimensions include geographic scope, facility types, travel modes, management strategies, travelers' responses, performance measures, and tool cost-effectiveness. Based on the original decision diagram as shown in Figure 9, the shaded areas in each decision dimension indicate the attributes or capabilities generally supported by DTA models. Column 7 in Figure 9 shows no shading as the decisions listed Column 7 are more related to subjective preference than model functionality.

Performing a DTA model exercise does not necessarily compete with resources planned for travel modeling or microscopic modeling. On the contrary, it leverages the investment of existing models as many DTA model developers provide conversion tools that allow the analyst to quickly convert the initial base model from existing travel models, saving time for initial DTA model creation. Some DTA models also offer subarea analysis capability, which allows a subnetwork to be extracted from the DTA network, along with the corresponding time-varying, gate-to-gate demand matrix or vehicle trajectories (routes), which can be converted to a microscopic simulation model for further detailed operational analysis and vehicle animation.

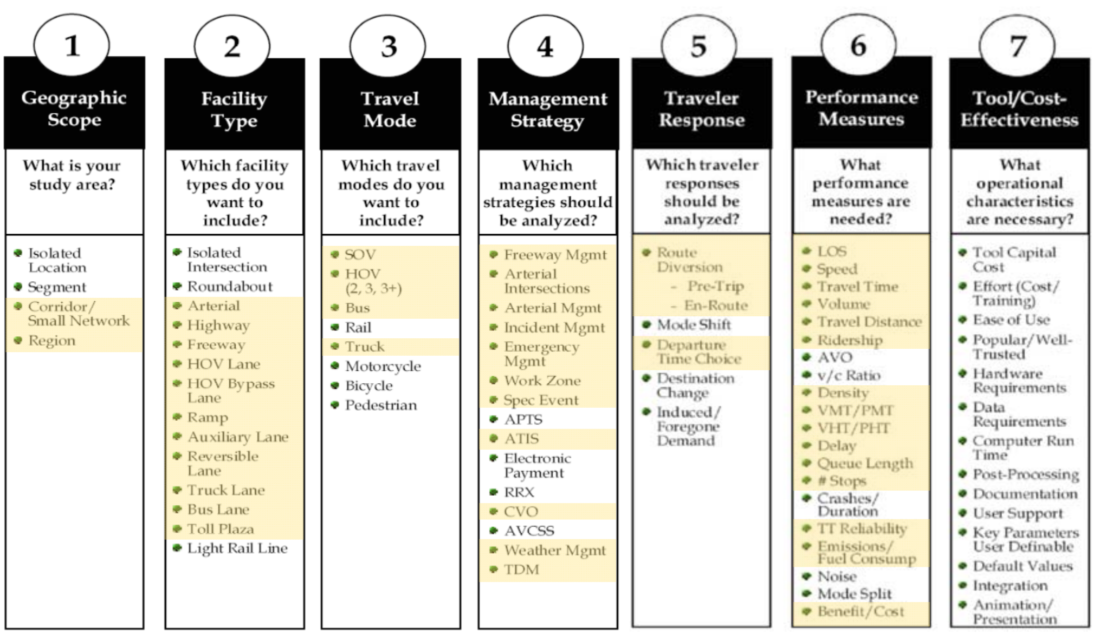


FIGURE 9 General decision-making process for selecting traffic analysis tools (*66*).

As computing power advances and software usability improves, the time and resources required for DTA modeling will continue to decrease. What is critical at the present moment for transportation agencies and transportation professionals is to appreciate the value and benefit of the DTA modeling capability and start planning for acquiring the know-how and building DTA models as part of the modeling infrastructure and knowledge base, so that one is better prepared to tackle the modeling challenges in the years to come with the complete range of available modeling tools.

PLANNING FOR DTA MODELING ACTIVITIES

The development of a DTA model generally requires a careful data analysis in order to have a successful base-year DTA model. Network data must be evaluated for consistency and accuracy, and typically enhanced to include some representational detail not required in static network models. Current traffic operations must be observed to count traffic or record speed profiles for model validation, to collect traffic control system information, and to observe travel times on routes across the region. Existing ITS elements also need to be adequately incorporated into the model.

Some analysts may not have access to resources to assemble the necessary information, either due to lack of particular expertise, lack of ability to hire additional help, or an insufficient mandate of responsibility for such information. Often data exist in other departments that could be used, but previous modeling practices (either traditional travel forecasting or micro-simulation) have not required the departments to work closely together and, hence, develop the necessary channels of communication. Lack of common referencing systems in the data or lack of consistent representational detail can also introduce significant issues.

Assuming that all the necessary data can be obtained and the DTA model set up, expertise is required to adjust traffic control information, demand data, and network attributes to reflect operating characteristics. The existing regional travel model can be used as the skeleton for the initial construction of a DTA model. It can provide the demand for the period of interest as well as the network characteristics. However, it does not provide the control settings as travel models often do not represent traffic signals with the use of actual control logic. Although the travel model provides the network characteristics, depending on the level of detail in the travel model, the model user may need to collect additional information and adjust the DTA model to ensure proper modeling for traffic simulation. For example, the presence of a left turn pocket and its characteristics, such as length, typically have no direct effect on the travel time calculated in a static model, but could be important to the success of a traffic simulation model.

Calibrating a DTA model is relatively time-consuming and requires extra care. The calibration process is typically an iterative approach that requires the experience to adjust capacity, demand, and behavior parameters based on evaluations of existing solutions. Network-side calibration generally involves setting parameters values for the models or logic governing traffic simulation, assigning saturation flow rate values or jam density values to links based on generic characteristics such as area type and facility type, targeted adjustments to typical links with capacity restrictions related to pedestrians, grade, narrow pavement, or complex geometry. Calibration of demand is needed for the simulation model to better represent field-observed traffic data. Such calibration includes matching field-observed counts and bottleneck in terms of temporal and spatial extent of congestion (which may be viewed from either speed or density, in lieu of counts.)

Many model users new to DTA models have found that actively engaging DTA model developers throughout the modeling process improves the development process and resulting model. Experience has found that such a hand-holding partnership significantly improves the modeling experience for the model user. This working relationship also provides valuable feedback for the developers' future model enhancement efforts.

General Modeling Process

The process of applying or deploying a DTA model generally encompasses the following basic steps, which are discussed briefly in the following sections.

DATA SET PREPARATION

DTA models typically have numerous inputs and parameters that need to be specified before model application. The exact nature of these inputs and parameters depends largely on the individual components making up the DTA model. At a high level, however, they can be grouped into demand-side and network-side quantities. The demand quantities typically include, at a minimum, time-dependent O-D matrices or trip tables and traveler behavior model inputs and parameters. The network quantities include capacities, link performance functions, traffic control information and strategy information such as incident impact parameters or ITS elements. It is easy to see that this set of inputs and parameters can be very large in real-world applications. For example, the number of nonzero flows in a set of time-dependent O-D matrices grows rapidly with the size of the network (the number of zones or centroids) and also with the chosen temporal modeling resolution (the length of the analysis time period and the number of intervals contained within). The network quantities are also numerous, in that they can be specified at the level of the individual link or segment. The modeler must select or specify these inputs and parameters appropriately as part of the deployment effort.

Demand Data

Time-dependent trip tables are common demand inputs to DTA models, although some may also accept individual trip activity records (e.g., trip tour or trip chain). The patterns can vary across origins, destinations and departure times. The most common method for capturing these variations is through a series of trip tables, each containing information about the trip departures within a relatively short time interval. The duration of this interval depends on the variability of real-world demand patterns as well as the desired modeling accuracy. Deployments and laboratory experiments have included intervals between 5 min and an hour. Fifteen-minute intervals are rather common.

The size of each trip table is determined by the number of zones defined by a model user, and can be large for realistic regions. While each matrix may be very large, the number of O-D pairs (cells in each matrix) with positive flows between them may be relatively small. The exact extent of this sparsity will depend on the geographic distribution of the zones and the density of the transportation links connecting these zones. The total number of O-D flows to be supplied to the DTA is thus a product of the number of nonzero cells of one matrix and the number of time intervals chosen for the modeling exercise. Typical O-D tables contain decimal-value entries. These values are eventually converted to integers during simulation as the basic entity in simulation is discrete (e.g., vehicle or person). This conversion step creates minor discrepancies between the number of vehicles generated between each O-D pair and the specified number of trips in the O-D tables.

If a DTA model is developed from an existing static database, it may be desirable to use a finer zone structure implying a larger number of zones (especially if considering integrating the effort with activity-based models). This choice is motivated by the higher fidelity of DTA models in general and the way that zones are connected to the network. In a DTA model, due to the higher realism of the representation of traffic flow, vehicles generally enter and exit the network using links that represent actual roads (rather than abstract centroid connectors), which helps to avoid creating false congestion points. Generally, DTA models can better portray the advantages of simulation by employing a higher resolution of the spatial distribution of demand and networks.

Most DTA models simulate multiple vehicle types or classes. Separate classes should be used for the purposes of controlling access to different network elements (e.g., prohibited turns and special-use lanes) and information sources (e.g., in-vehicle devices). Every class should have its own demand matrix (or matrices, if permitted). Typical classes might be single-occupancy vehicle, HOV, light truck, heavy truck, taxi, etc. Depending on the methodology used, multiple classes may also be needed in the context of modeling congestion-pricing schemes or toll roads (e.g., to capture sensitivities of different classes to out-of-pocket costs).

The solution of a dynamic model starts with an empty network, accumulates vehicles according to the input demand rates, and gradually empties out during low-demand intervals. Because it takes time to load the network, it is important that the demand period start earlier than the time window (observation period) over which the model is expected to produce meaningful results regarding the state of traffic. It is also useful to end the demand period beyond the desired study end period to allow all (or most) vehicles loaded during the actual study period to clear the network. Having the demand period longer than the study period will produce more realistic results, as the process of the network clearing (which will experience lower travel times due to lack of demand) will not significantly affect routes choices of travelers during the study period. As a general rule, the start and end buffer windows should be as long as it takes to get across the network (longest trip) under the prevailing traffic conditions. Note that the end buffer window should still contain realistic demand rates so that the vehicles already on the network experience reasonable travel times and make route choices accordingly. Using a zero-demand end buffer window can bias the speeds upwards and travel times downwards, resulting in inaccurate route choices. Furthermore, at the end of simulation, certain vehicles may still exist in the network and have not completed the trip. It is important to understand how these vehicles are dealt with in the model output statistics to avoid biased statistics (e.g., incorporating incomplete trips may result in shorter distance and travel time).

Time-varying demand data may be derived from several sources. The most convenient way is to utilize the existing trip tables associated with travel forecasting models. Most planning agencies have O-D tables for different periods in a day (e.g., a.m. peak, p.m. peak, and off-peak), with each table spanning several hours. If hourly factors are available for the time of interest, they can be used to derive a temporal profile in order to disaggregate the existing tables into finer time resolutions (e.g., hourly or 15-min tables). However, one should be warned that simply applying the hourly factors to a 24-h table to derive the hourly table is a flawed exercise, as the directionality of O-D trips are typically lost when trips are aggregated into the 24-h table. Factoring a 24-h table does not retrieve the critical O-D directionality information. The travel pattern would deviate a great deal from reality on the ground. Some planning agencies maintain trip tables representing a.m. peak, p.m. peak, or off-peak periods. The directionality is more likely to be preserved in these time-of-day tables than the 24-h table. If DTA is applied only for peak-hour analysis, then a corresponding time-of-day table can be a reasonable starting point. The temporal profile within the period of interest may still need to be specified by the model user, but the trip spatial directionality is generally maintained. For a 24-h simulation and assignment, one may consider stitching these time-of-day tables to form 24-h demand tables.

Ideally, these O-D matrices converted from a travel forecasting model should be adjusted or calibrated to match traffic data, as described in the section on Model Validation and Calibration.

Network Data

DTA models are generally more data-intensive than static models. For example, though both models work on a network of the study region, DTA requires a more detailed network including the number of lanes on each link, the presence of acceleration–deceleration lanes and turn bays and lane connectivity. Such data must be collated from various sources. The network representation, for example, could be based on existing models, geographic information system (GIS) files, online maps or aerial photographs. In the absence of a network already coded in the format required by the chosen DTA software, the model user must create such a network from scratch. Most existing DTA models provide GUI for this purpose. This step usually involves the definition of nodes and links. Nodes represent urban intersections and freeway diverge or merge points. Links represent the physical sections that connect these nodes. In some cases, links may be further divided into segments to capture linear variations in roadway cross-section geometry. GIS files can considerably simplify this step, as the centerlines and other geometry information from such sources can be expected to be reasonably accurate. Additional work may be involved in defining all allowed and prohibited lane movements at link and segment boundaries. Online maps and aerial photographs can be invaluable sources of data at this stage, especially for validating lane connections. Existing data sets will most probably be derived from static planning models. The network from such a data set must be upgraded to include at least the basic DTA requirements. Such an upgrade can be time-consuming, depending on the spatial extent and density of the network and the level of detail in the static network representation. The model user should further remember that static networks are often based on nodes connected by straight-line segments. A move to a more accurate geography will ensure better results and output visualization that will be true to the real-world transportation system.

In contrast to some static models, the geometry and flow characteristics of zone connectors have increased physical significance in dynamic models, and should therefore be modeled as real physical roadways. In particular, this requirement implies that connectors may not be incident to major intersections, as is sometimes the case in static models, but rather be moved to mid-block locations or distributed on the link in a manner that preferably corresponds to trip origins and destinations. Furthermore, spillbacks on origin zone connectors may introduce a discrepancy between the assumed (via the time-dependent O-D matrix) and actual number of vehicles entering the network at a given time, which may not be desirable. Some DTA models may allow vehicle loading to be distributed among a set of generation lines, which allows more spreading out of the vehicle loading.

At present, most DTA models generally do not perform multimodal assignment involving both private vehicles and public transit. Some DTA models represent transit vehicles, such as buses, with dwell times at stops as an exogenous input. This approach provides a framework for evaluating the impacts of transit vehicles on private vehicles in terms of congestion and travel time, and vice versa. The impacts of various network or traffic control modifications—such as reserved lanes and special signal phases—on the travel times of both private and public vehicles can thus be evaluated. In this context, transit line data represent another input to the model and are generally thought of as a subset of the network data.

Control Data

Coding signal timing and ramp meter control are critical from both the simulation and assignment standpoints. Properly coded signals realistically represent the delays at surface streets, permitting more realistic assignment results. DTA models can typically represent the operation of pretimed (fixed) traffic signals, potentially with the necessary parameters to represent signal synchronization. Various software systems may have different options available for representing actuated signals, though these options normally involve a set of parameters such as minimum and maximum green times per phase, which are used to approximate the operation of this type of controller in a simplified way.

Some DTA software approximates actuated signals using fixed-time plans by obtaining average green splits per movement from a signal timing–optimization software package. Some DTA software allows analysts to directly enter actual timing including actuated control so that actuated controls are more explicitly represented. Often, however, the detailed logic of modern traffic controllers is not modeled, but approximated.

Particular DTA software may also provide for the specification of standard uncontrolled intersections such as all-way stop controlled, two-way stop controlled, roundabouts, freeway merges and yield signs. Signal data usually reside at a municipal traffic department and are in one or more of a variety of formats from text files to GIS-mapped databases. These procedure options would dictate the amount of time needed to enter the signal data into the model. Some DTA models offer a default setting for actuated signal timing during initial conversion and allow the analyst to populate more detailed data at a later time. Experience shows that not all agencies have a complete inventory of unsignalized intersections. Field surveys may therefore be needed or assumptions may be made about all arterial intersections that are not inventoried as signalized intersections.

Another commonly raised question is about the signal timing setting for existing intersections or for those existing only in future planning years. One suggestion is to leave the existing time unchanged and apply actuated timing for future intersections, then run test DTA runs and identify locations where sustained congestion exists. Adjusting the timing for these intersections manually or integrating with other signal optimization models to optimize the signals may suffice for future-year scenarios. However, in the attempt to optimize signal timing, the safety considerations should also be adequately addressed. The *Manual on Uniform Traffic Control Devices* is also a valuable source for planning reasonable future traffic control consideration, for example, signal warrants from an existing stop control, e.g., whether the minimal past conditions that warranted consideration for the existing signal timing will still obtain in the future. This step is important to represent properly future traffic control settings and resulting traffic conditions.

Scenario Data

Scenario data usually relate to the application of interest. Properly specifying scenario data is crucial for using the model properly for the problem at hand. The activities involved are often referred to as the modeling techniques pertaining to different model packages. In addition to carefully reading the user manual, it is useful to closely interact with the model developers for specific modeling questions as the user manual may not exhaustively document all the modeling techniques and their application to various problems. Often, on the one hand, modeling a certain problem in a certain way is the judgment call of the model user. The needed judgment usually comes with experience and mastery of the software. On the other hand, it is not uncommon for model developers to claim that their models can perform a wide range of analyses for various applications as many modeling questions can be addressed through certain modeling techniques or workarounds. It is, however, paramount that the model user investigates in detail whether the said model technique is a theoretically sound approach or merely a workaround. In-depth investigation and examinations will reveal the advantage or disadvantage of various models for the problem of interest. Unfortunately, this is the burden the model user has to bear.

Once the O-D demand, network, traffic control, transit line, and scenario data are properly coded in the correct format, the next step is model calibration and validation. This involves identification of all other DTA model inputs and parameters such as link or segment traffic flow parameters, capacities or performance functions, O-D adjustments, route choice model parameters, historical (or perceived) network travel times. Different DTA models differ in how they respond to model parameter adjustments. Discussions in the section on Model Validation and Calibration (page 35) are offered for only general considerations; model-specific details need to be sought from developers.

CHARACTERIZING A DTA SOLUTION

A generic DTA model as illustrated in Figure 10 can be used to describe the process of calibration and validation as described in later sections. A DTA model accepts several inputs and

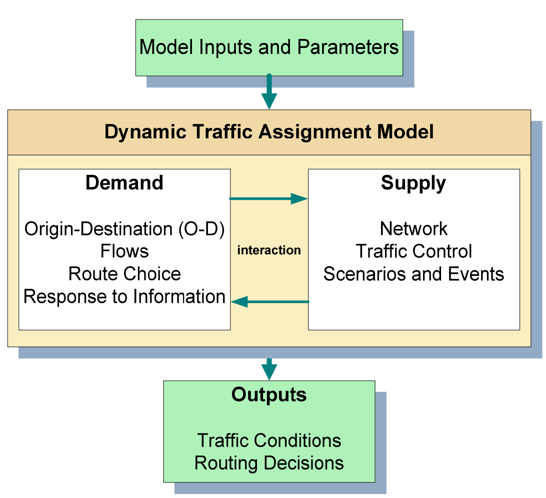


FIGURE 10 Structure of a generic DTA Model

parameters, as described in the section on Data Set Preparation (page 29). These inputs and parameters are used by the various demand and network model components, which interact to predict the spatial and temporal evolution of networkwide traffic conditions.

The outputs of large and complex DTA models depend heavily on the input values selected. It is therefore crucial, even mandatory, to ensure that DTA models are adequately calibrated so that their outputs compare favorably with real-world traffic observations made in the study region. If the model can replicate the current (or baseline) conditions well, it instills some confidence in the results generated for future scenarios that obviously lack real-world validation measurements.

Before looking at typical model outputs, such as link-based measures of flow, speed and density, there are several global measures that can be used to characterize the model results and put them in context. These are primarily convergence measures and certain networkwide measures that are particularly important when reviewing the initial DTA runs of a model.

The most valuable convergence measures are those that quantify how close the current solution is to equilibrium, such as the relative gap. This measure reflects the difference between the minimum (best) route cost and the average route costs, relative to (divided by) minimum cost, as a weighted average across all O-D pairs. This measure must be calculated for (and reported by) each departure time interval. A plot of relative gap by departure interval over all iterations should display a pattern of lines converging down to minimal, stable values. A typical trend is to see increasing values of relative gap with increasing departure time, partially due to the increasing congestion levels encountered by travelers leaving later rather than earlier.

Due to the added level of realism and detail in DTA models when compared with static models, DTA models typically have significantly higher values of relative gap in the final calibrated solution. There is a fairly wide range in terms of level of realism and detail across different DTA models in their representation of traffic dynamics, but the same rule applies: the higher the level of realism and detail, the higher the values of relative gap the model is likely to produce for a given network or scenario. Moreover, the value of relative gap inevitably increases with the average level of congestion in the network. As a result, it is difficult to recommend specific thresholds of relative gap for DTA models. More important is that the convergence measure, when observed over iterations, is relatively stable when the DTA run is terminated.

Certain network-wide measures can also provide a strong indication of the general quality of the model results if they are reported as time-varying outputs. These include the number of vehicles in the network and the number of vehicles waiting to enter the network. The latter measure reflects the fact that once congestion spills back to an origin zone, the entrance flow rate is restricted and may be lower than the demand flow rate for entering the network through this connector. This difference between demand and entrance rates causes vehicles to wait outside the network until they can be loaded. Time-varying, spatially averaged network speed is also a useful measure but tends to be very strongly correlated (inversely) with average network density and hence the number of vehicles in the network.

These networkwide measures cannot be compared with empirical measurements because such data are not available. Rather, it is the shape of the time-series plots and their relative values to each other that are most informative. For example, it should be taken as a warning sign if the graph of the number of vehicles on the network is continuously increasing up to the end of the forecasting period. This indicates that the network is not in a stable state, but is getting increasingly more congested. If this is considered reasonable under the circumstances (i.e., compared to reality), it would be advisable to increase the duration of the period to include more of the off-peak demand.

It may also be somewhat disconcerting to see the number of vehicles waiting to enter the network increase continuously until the end of the demand period. This trend could indicate that in certain zones departure volumes are high with less loading points, and could also indicate that traffic controls at or near those loading points may not be adequate to discharge generated vehicles. If vehicles take too long to be loaded, their actual loading time becomes much later than their generation time, which then means the O-D temporal pattern is distorted.

Observing how long it takes for the network to clear after the end of the period is also a very strong qualitative indicator with respect to the results: if it simply takes too long to clear, the results should not be considered acceptable.

MODEL VALIDATION AND CALIBRATION

The process of validation (verification) compares the model outputs to the observed traffic conditions such as traffic counts and speeds to assess the quality of model outputs. Model calibration involves the identification of a set of DTA model inputs and parameters that results in model outputs that are reasonably close to those field observations. Traffic data can come from various sources with varying detection technologies such as loop detectors, acoustic sensors or video-based detections. Automatic vehicle identification technologies commonly used in toll collection are also useful for O-D calibration in addition to point-to-point count and travel time data.

The measured traffic data can represent a wide range of quantities, some of which are

- Vehicle counts (by link or by lane) measured at detector locations;
- Average vehicle speeds at detector locations;
- Average link or segment density or detector occupancy;
- Queue lengths;
- Link or subroute travel times; and
- Intersection turning movement counts.

Since a DTA model's outputs are time-varying, the data used to calibrate and validate the DTA must also be dynamic. Thus, the various data listed above must be collected at a relatively fine temporal resolution (such as every 5 or 15 min). The data collection interval must also be compatible with the desired modeling time interval selected for the particular application. For example, if the project goal is to model the hourly variations in demand and network performance, then hourly data should be collected at a minimum. Theoretically, the finer time resolution that the data are in, the better the calibrated model represents the real-world situation. However, one needs to exercise several cautions as stated in what follows.

Data sets collected from different sources may be expected to show some degree of inconsistency between them. Inconsistency can occur for various reasons. Data from different days, weeks, months, or years may reflect different demand levels, trip patterns and network–infrastructure conditions such as work zones. Even on the same day, some sensors may introduce measurement errors due to malfunction or failure. Another issue is to understand the source of the data. Hourly traffic counts generated from daily counts using hourly multipliers are not as accurate as actual hourly counts from the field. It is thus worthwhile to do some consistency checking in addition to the usual data cleaning procedures. The ideal case is to be able to make direct comparisons based on overlap between two data sets if they cover some of the same links or turning movements. Basic statistical tests, such as linear regression analyses, can be used to quantify the goodness of fit between two overlapping subsets. Even a relatively small sample of overlapping data can give a general indication of the consistency between the data sets as a whole.

Due to the day-to-day variability of traffic conditions, care needs to be exercised when using traffic data for model validation and calibration. One commonly used approach is to take traffic data that is averaged over a number of days, and those days should be selected in such a way that they are representative of the O-D demands in the model or scenario. Note, however, that the presence of unrecorded major disruptions such as severe incidents or weather effects across days can cause significant bias if the data from these days are simply averaged. In some cases, traffic simply fluctuates to a large extent from day to day, and simply taking the average would smooth out the congestion (worse case) and thereby would leave out valuable information. Therefore, before averaging the data, one needs to consider categorizing data by weekday, weekend, holiday, etc., and may also need to reflect seasonal variations and weather conditions. The goal is to reduce the data variations within the same category. Another alternate approach is first to take the average and then select one actual data set that resembles the average. This approach would represent the average and retain the within-day variation.

The process of initially validating and subsequently calibrating a DTA model can be broken down into two sequential analysis stages: qualitative and quantitative. The qualitative analysis stage (also referred to as preliminary validation) is what typically starts after the very first model runs, when there may still be many errors in the input data to be found. In these situations, it may be of little practical value to begin comparing the model outputs to empirical data, especially if the model is not converging to a stable solution. Once the model has been improved to a certain extent, the quantitative analysis (also referred to as calibration) starts. Quantitative analysis is based on a direct comparison of model outputs and empirical data and investigating the outliers in order to further refine the model.

Qualitative Analysis (Preliminary Validation)

The section on Characterizing a DTA Solution discusses measures to be used for characterizing the results of a DTA run: this activity applies to all model runs, and should be considered at all stages of the calibration process. However, these measures are particularly important in the early stages of the calibration when they are most likely to indicate that the model results are unsatisfactory for one or more reasons.

Generally, errors in the coding or inputting of network and traffic signal data are found to produce outputs showing more congestion rather than less. Because of the phenomenon of congestion spillback, a queue grows in space and engulfs vehicles that do not directly contribute to the original cause of the queue (they will turn off the road before reaching the downstream bottleneck). In extreme cases, queues that are initially separate grow into one, causing congestion to grow even faster and spread out in many directions. This can cause an entire section of a network to be engulfed in heavy congestion. If deadlock (gridlock) sets in, the vehicles will literally be standing still.

While not all DTA models may display such behavior in their initial runs, it can happen in situations where the network is highly congested in reality. Under such circumstances, not much is gained in computing statistical measures of comparison between the outputs and empirical traffic data. There is a need to identify the initial bottlenecks behind the heavy congestion and to fix input errors that result in unrealistic capacity reductions. At the same time, care should be taken to retain realistic capacity values, as the bottlenecks may also be caused by unrealistic O-D demand inputs or route choice models. Indiscriminate capacity increases might remove or dilute bottlenecks that do exist in the real world and may also invalidate scenario analyses based on other assumptions about traffic demand levels. In some situations, it may help to advance the simulation start time to allow for a more gradual evolution of congestion patterns. A simple diagnosis strategy is to conduct a DTA run with correct network and control setting but with a uniformly reduced O-D demand level to create relatively free-flow conditions, and observe if unusual congestion still develops at certain locations. Tracking such congestion patterns and locations usually leads to the discovery of coding errors.

Another artifact of overcongested model results, and in particular if gridlock has set in, is that the assignment algorithm will often not converge to a stable solution. If the assignment is not stable, it is also of questionable value to be comparing the model outputs to empirical data, because the results may still have been varying significantly from iteration to iteration when the model run was stopped.

The purpose of the qualitative analysis stage is primarily to achieve model results that exhibit a stable solution, are free of gridlock and, if possible, in which the overall congestion pattern at least resembles the actual conditions on the street. In many cases, the initial DTA runs may already display these properties, and the calibration work can begin immediately with the quantitative analysis described below.

Quantitative Analysis

This stage of the calibration process is based on direct comparisons between model results and empirical observations. Various statistical measures may be used to quantify the goodness of fit between the DTA output and the observed data, but the actual process of improving the fit by adjusting input data is essentially the same as used in the qualitative analysis stage. It is based on an understanding of traffic phenomena and causes of congestion along with common sense and modeling judgment. The following section provides recommendations for a general process of deductive analysis to be used for improving or calibrating a model.

In its simplest form, the quantitative analysis work consists of investigating one outlier at a time in order to determine if it is not the result of an error in the input data. In this context, an error might simply reflect a need for a minor adjustment of some input value. The most important data to consider for calibration, especially in the initial stages of this work, are traffic counts.

In some cases, it may be possible to identify a single problem that, when corrected, simultaneously improves several traffic count outliers at once. This type of improvement happens when two or more outliers occur over a set of links (or turning movements) that can be joined together using other links (or movements) for which there are no observations, to form part of a route. Fixing the error causes these vehicles to change to an alternate route, and thus several outliers are improved simultaneously. In the ideal case, there are also observations on the alternate route, indicating an error with the opposite sign, which are also improved at the same time. In some cases there are simply pairs of outliers, one positive and one negative (on alternate subroutes) that are correlated in this way.

Understanding how changing a parameter like link or movement capacity, e.g., by modifying signal timing parameters—which directly affects link travel time—impacts route choices in the model is critical to identifying possible correlations of this kind in advance of making any adjustments to address them. The user-equilibrium property ensures that there will be a consistent relationship between experienced travel time and route choice. However, this relationship cannot be accomplished without having the model reaching proper convergence.

Once the inputs have been corrected or adjusted to the point where the outputs are deemed acceptable, the calibration process is complete. There is an expression about modeling that is good to keep in mind when calibrating: "All models are *wrong* (i.e., imperfect), but some are *useful*" (*67*). The challenge is to make the model good enough to be useful, namely, to make useful predictions.

Calibration Methods

This section presents a general methodology for investigating discrepancies between model outputs and field data. The underlying logic is applicable to the qualitative context as well as the quantitative context of the calibration process, as presented above. Since traffic counts are currently the most common type of field data, the term model volume(s) will be used in reference to model outputs that are comparable with empirical traffic counts. These outputs may be link-based or turning movement-based. Other empirical data such as speed are also important for understanding the source of discrepancy. Queue lengths, if available, may be used within the procedure as supporting information to aid in the interpretation of model volumes.

Before starting to interpret discrepancies between model volumes and traffic counts, it is imperative to understand a fundamental property of real traffic that is respected by dynamic models (but not by static models). As illustrated in Table 1, when the model volume is higher than the observed volume, one needs to understand whether this occurs under a free-flow or congested regime by checking the speed data. If the speeds are in the free-flow regime, it means that the model link density is higher than actual density (see Condition 1 in Table 1; density can be depicted from the reciprocal of the slope of the line connecting the traffic point and the origin, a smaller slope means a higher density). Higher model link density can be caused either by lower model downstream capacity or higher model upstream demand. If the speeds are in the congested regime, then the model density is lower than the observed (see Condition 4 in Table 1), caused by either higher model downstream capacity or lower model upstream demand.

In the opposite case in which the model volume is lower than the observed counts, two separate conditions need to be examined as in Conditions 2 and 3 in Table 1, depending on the prevailing speed regimes.

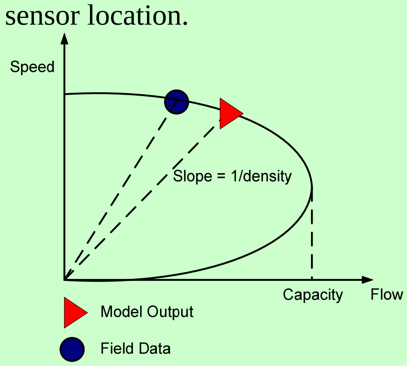
A discrepancy between model volumes and counts at a particular observation location (the plural here is used to refer to time-varying data) is essentially due to an imbalance in the model between capacity and demand. There are three basic factors that contribute to this imbalance: (*a*) local network capacity and control timing parameters; (*b*) local traffic demand because of the assignment process, in the form of route flows; and, (*c*) global demand as represented by the O-D matrix. Each of these three influences can be a possible source of error. Their respective contributions to a specific outlier can be determined by following a simple process of investigation and elimination.

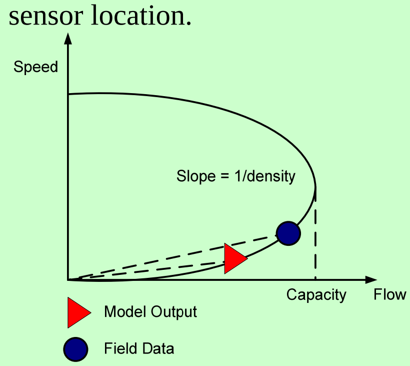
This process is discussed in detail later in this section but can be summarized as follows:

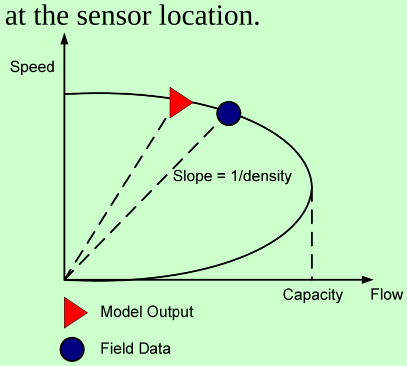
1. Compare model volume and observed volume. If model volume is higher than the observed volume, Conditions 1 and 4 in Table 1 apply. The next step is to check speed. If speed

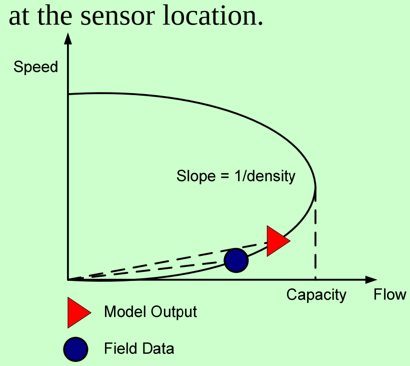
TABLE 1 Network and Demand Effects Contributing to Discrepancies Between Model Outputs and Field Data

<table>
<tr><th></th><th>Free-Flow Speed</th><th>Congested Speed</th></tr>
<tr><td>Model speed < observed speed</td><td>(1) Model volume > observed volume

Model downstream capacity is lower than actual, or model upstream demand is higher than actual, causing higher model density at the sensor location.

</td><td>(2) Model volume < observed volume

Model downstream capacity is lower than actual, or model upstream demand is higher than actual, causing higher model density at the sensor location.

</td></tr>
<tr><td>Model speed ≥ observed speed</td><td>(3) Model volume < observed volume

Model downstream capacity is higher than actual, or model upstream demand is lower than actual, causing lower model density at the sensor location.

</td><td>(4) Model volume > observed volume

Model downstream capacity is higher than actual, or model upstream demand is lower than actual, causing lower model density at the sensor location.

</td></tr>
</table>

is congested speed, then Condition 4 applies. The discrepancy can be attributed to either higher downstream capacity or lower upstream inflow demand. To further isolate the cause, the diagnostic actions could include checking the local capacity and signal timing parameters upstream and downstream (as may be applicable) of the observation location. A similar process also applies to other volume discrepancy situations.

2. If the local capacities are not (or are no longer) a major contributing factor, the next step is to check the assignment, which is responsible for the local demand.
3. If the local capacities are correct and the assignment is acceptable, the only remaining possibility is the global O-D demand: it may have to be revised.

The next three sections will further discuss these influencing factors in order: capacity, assignment and demand.

Capacity Effects

The capacity of a link or turning movement is determined by a number of factors. First, there will be a maximum capacity or ideal saturation flow rate, which is an exogenous input parameter that is based primarily on the number of lanes and a per-lane saturation flow rate. The traffic signal timing at the downstream end of a link allocates green time to each movement, and signal synchronization will affect how well this available capacity may be utilized. If there are shared lanes at the intersection—lanes that service more than one turning movement—the effective capacity of these two (or more) movements will be interdependent. Right-turn-on-red at an intersection is another source of variability, since its utilization will depend on available gaps in the conflicting movements and on the existence of non right-turning vehicles in the right-hand lane. Vehicle mix is yet another factor that impacts capacity. A large fraction of slow-moving trucks, for example, could result in a significant drop in capacity. Grade is another possible factor affecting the capability.

The HCM provides fairly detailed guidelines for the determination of the average capacity of various types of highway links. It is important to note that capacity in mesoscopic traffic simulation model may be used in a variety of different ways, and generally it is not perceived and used the way the microscopic models do. Microscopic models don’t directly apply capacity in simulation as capacity is the macroscopic outcome that results from the car-following interactions of individual vehicles. Since mesoscopic models are not driven by the car-following mechanism, but by macroscopic fundamental relations (e.g., speed-density or volume-density relationships), it is possible for capacity to be applied directly to influence the vehicle simulation, depending on specific implementation procedures in different models. If a mesoscopic model is driven by a speed-density relationship, then vehicle movements are affected by local density, and the capacity is implicitly determined by the specified speed-density relationship (e.g., flow rate is the product of speed and density, $q = k \cdot v$; for each speed-density relationship, there exists a maximum flow rate or capacity mathematically). For mesoscopic models that rely instead on the flow-density relationship, the link capacity is an exogenous given.

In both cases, the link capacity is explicitly or implicitly determined by the model’s input parameters. Even so, this fact should not be related directly to how the concept of capacity is applied in VDF in static models. Unlike static models, where capacity is the only link attribute used to estimate link travel time, in mesoscopic models, it is the speed-density or flow-density functions that are used to determine the changing prevailing speeds and positions of vehicles on each link within each simulation period. Average link travel time (by movements) is the outcome of simulation reached by taking the average of link travel time (taking the difference of link exit and entry time) for all vehicles making the same movement within each simulation period.

The existence of oversaturated movements will further affect the operational (or effective) capacity of a link or movement. If the downstream link of a movement is congested to the point that the inflow rate is restricted below the ideal capacity, this will further reduce the effective capacity of its incoming movements. If one turning movement is significantly oversaturated, the resulting queuing could impede vehicles destined for other turning movements exiting from the same link.

It is also useful to keep in mind that incidents (both those known and those unobserved) may have imposed capacity restrictions during the study period and may affect the field data. However, such an anomaly could be detected by comparing multiple data sets from different days at the same locations.

The above factors serve to explain why volume on a link or movement in a dynamic model will often be much lower than the simple theoretical capacity estimate (given, for example, by the product of the effective green, number of lanes, and per lane saturation flow rate for arterials). In general, a DTA model will account for many of the factors discussed above, and understanding the causes behind the model volumes is a key component of the calibration process.

Some common coding errors that lead to reduced capacity and hence excessive congestion are

- Incorrect lane allocations to movements,
- Lack of turn lanes,
- Insufficient green time,
- Incorrect signal phase definition, and
- Incorrect signal synchronization.

Poor synchronization is rarely the cause for extreme congestion: synchronization has its biggest positive impact when traffic flows are within the available capacity (offered by green time). In this situation, lack of synchronization will result in higher travel times, but may not cause a complete breakdown of traffic conditions. One exception to this rule is the existence of closely spaced signals, where lack of proper synchronization can lead to a significant drop in flow. It is worthwhile to identify such cases in advance and verify the synchronization parameters rather than depending on the model outputs to bring the problem to light.

The traffic flow functions, such as speed-density or speed-flow curves, underlying mesoscopic models as discussed earlier must be calibrated to reflect the ground truth at different types of locations (e.g., freeway mainlines, ramps, weaving sections and arterials by number of lanes). Typically, the two required traffic descriptors are plotted against each other and a curve fitted for representative sensor locations. Since sensors are generally not expected on every link or segment, grouping strategies are adopted to assign a traffic flow function to each link or segment.

The process of fitting a fundamental diagram curve through sensor data can be challenging. The data are usually scattered and often do not show a tight relationship. The data may have to be weighted to ensure that the most common conditions are captured by the fitted curve. There is also the possibility of overfitting local curves at the expense of networkwide performance.

Assignment Effects

The assignment of O-D demand to route flows—i.e., the aggregation of travelers' route choices—results in the traffic demand for any individual link or turning movement in the network. This can be thought of as the ideal or assigned demand. Due to the nature of dynamic models, in terms of realism and detail, extra attention is needed for establishing demand and capacity parameters at the level of individual turning movements.

Though the assigned demand for a particular turning movement may be greater than its capacity, this may not result in oversaturation of the movement and congestion on its upstream link. The reason is that there may be bottlenecks upstream of the movement in question which are restricting the flow rate at which the assigned demand can reach its upstream link. A simple rule for queuing on a link is that its inflow rate must exceed the outflow rate. The inflow rate on a link, which is the result of the assigned demand and the possible metering effect of upstream bottlenecks, is referred to as the local demand for that link. The inflow rate on a link can further be broken down into the local demands for each of its outgoing movements, and the same queuing rule applies: if the inflow (at the link entrance) for an exit movement is greater than the outflow of that movement, this will result in an accumulation of vehicles on the link and hence an increase in link travel time.

Evaluating the impact of the assignment on a particular traffic-count outlier essentially consists of qualitative analysis techniques, as there is generally no empirical information about travelers' routes through the network. However, knowledge of the network is often sufficient to make reasonable and educated judgments that can go a long way, particularly for using route analysis to find more coding errors in the capacity-side (network and traffic control) inputs. Route analysis generally consists of two basic approaches: select-link route analysis and O-D route analysis. As a general rule, the capacity-side data should be verified as much as possible (as discussed in the Capacity Effects section, above) before commencing route analysis.

Select-link analysis is commonly used in static assignment modeling. This tool generates all routes including a link or turning movement (or, if desired, a specified combination of links and movements). There are a few different ways that select-link analysis can be adapted to the dynamic context. Possibly the simplest and most intuitive approach to dynamic select-link analysis is what is called select-link simulation. The outputs in this case are typical time-varying, link-based measures such as flow (volume), but only the vehicles on the corresponding subset of routes are counted. Whatever the specifics of the select-link tool employed, the underlying information is the same: it is simply a question of how it is presented for analysis.

Select-link analysis is a particularly useful approach when there is too much local demand, and thus too much congestion, at a particular location in the network (link or turning movement). The question that needs to be asked is where should these vehicles be instead of here? This question can only be answered by knowing the origins, destinations and the routes used by the vehicles in question. In this situation, the select-link information can produce three basic outcomes: there are some routes that seem unreasonable (in particular, excessively long or circuitous), all the routes seem reasonable, or there are routes that ought to be used but are not.

If the routes seem reasonable, there is no particular reason to question their validity. This outcome can be followed up with a visual inspection of the complete route set for the O-D pairs identified by the select-link tool to put the select route in context, but this step rarely changes existing conclusions about whether a route is reasonable or not. Route choice model parameters may then be analyzed to verify whether they reflect realistic traveler behavior.

If the select-link routes seem reasonable and it is certain that the capacity-side data are correct, the problem is almost certainly due to excessive demand for one or more of the O-D pairs in question. Coming to this conclusion from select-link route information is not as unlikely as it may first appear. Depending on the location of the link or movement in question, the select-link analysis may in fact produce relatively few O-D pairs, and those O-D pairs may have relatively few reasonable alternative routes. If there is a route which is obviously unreasonable, or at least questionable, the next step is to visualize the full set of routes for the corresponding O-D pair. This process will likely indicate much more reasonable routes that are also being used, but which are probably carrying too little flow. In many cases, it will be immediately obvious that there is too much congestion on one of the reasonable routes, and it may turn out that result is due to an undiscovered capacity-side coding error. As a result of the objective of the assignment model—i.e., to minimize each traveler's travel time—the additional travel time due to the excessive congestion pushes the travelers to use the unconventional route(s). If no particular causes of excessive travel time on the more reasonable routes are identified, and if it is certain that the capacity-side data are correct, the only remaining option is that there is excessive O-D demand for one or more of the O-D pairs in question. Nevertheless, it is desirable first to eliminate unreasonable or unrealistic routes from the choice set to ensure that future scenario runs do not run into similar problems. While performing this task, care should be taken to include alternative routes that may become reasonable when regularly used links are disabled through incidents, work zones, etc. If using movement-based (intersection turning) counts for the analysis, and if the movement flow is shown to be near capacity in both models and actual data, the upstream congestion could still be rather different. This is a good example of why it is often said that speed or queue length data are essential in addition to traffic count data.

Demand Effects

As discussed in the Assignment Effects section above, it is possible in some situations to draw definitive conclusions about excessive demand for certain O-D pairs from a rigorous investigation of traffic-count outliers, preferably in conjunction with empirical data about link speeds or queue lengths. In these cases, how much to adjust a specific O-D demand value is largely a matter of common sense and modeling judgment.

Automated processes and algorithms for adjusting dynamic O-D matrices using empirical data are an active area of academic research. Some existing DTA models, in conjunction with optimization formulations and solvers, allow systematic adjustment of O-D matrices for multiple vehicle classes on large networks with thousands of zones. Nevertheless, manual O-D adjustment is also commonly used in calibrating a DTA model. It should also be noted that, unlike the well-defined physical properties of roads and traffic signals, there is often a significant degree of uncertainty and day-to-day variability in the O-D demand data.

Compared to static models, the additional detail and realism of DTA models mean that they are fundamentally more sensitive to input errors, and this applies equally well to O-D demand data. The added sensitivity of DTA is one of its primary motivations, but this trait brings with it higher requirements for precision in the required input data. Several factors can contribute to errors in the O-D matrices. Due to the limitations in how static models represent congestion, static models could be somewhat forgiving of an overestimated demand matrix. More severe congestion can result when such a matrix is used in a DTA model. If an automated adjustment procedure (algorithm) is used to adjust the demand data based on traffic counts, it should be noted that such a procedure can introduce errors into the matrix, especially if there is not enough count data available. If there are important links for which there is no count data, those links may end up carrying significantly too little or too much flow as a result.

One of the most common errors made after the first run of a DTA model is to observe only the volumes and not the levels of congestion and to assume that, due to the low flows on the links, the demand values must be too low when in fact the opposite is true. If the low flows are due to congestion, which can be easily verified, increasing the demand in the model will only increase the levels of congestion shown and result in even lower link flows. This kind of error is a strong motivation for using additional data to calibrate and validate demand (and DTA) models. Readers can refer back to discussions related to Table 1.

It should be noted that given the same total demand volume (number of cars), a time-sliced matrix which exhibits a time-varying demand profile will inevitably generate more congestion than a flat (static) demand. This is because it is the peak of the demand profile that will represent the peak loading conditions, which may result in congestion or even gridlock. The same number of cars spread out evenly in time will necessarily have the lowest possible peak-loading conditions.

For this reason, it may be advisable to begin the first runs of a new DTA model with a flat demand matrix (scaled down to reduce oversaturation), until there is confidence that all of the capacity-side coding errors have been corrected. It is usually not necessary to use peak-loading conditions on a network for the purpose of identifying coding errors as discussed above. If the congestion is particularly acute, it is advisable to reduce the total demand until the capacity-side coding errors are discovered and corrected.

Time-varying O-D inputs may also be estimated from traffic data such as counts, speeds and travel times. This topic continues to receive significant research attention, as there are typically far too many O-D flow variables to allow manual adjustments within a reasonable time frame. Those links with observed data will also constitute only a subset of all possible O-D pairs.

O-D calibration–adjustment is simple in concept. A starting set of O-D matrices is loaded into the DTA model. At convergence, its outputs are compared against the observed data (mostly counts). An objective function is evaluated to quantify this fit (usually the total count differences over all observed links), and its value is used in some systematic way to adjust the O-D flows automatically. The adjusted O-D matrices are loaded into the DTA model again and run to reach convergence. Hopefully, the next evaluation of the objective function will yield a small discrepancy. The process continues until the total count discrepancy reduces to a certain threshold. In other words, the estimation approaches are therefore iterative so that the last estimates are fed back into the DTA model and the process repeated until the observed, real-world traffic data are reasonably replicated.

While the above description sounds simple, there are several complicating factors. Even assuming that the DTA model is error-free (from the perspective of algorithms, network coding, etc.), the count data may be erroneous or inconsistent across space and time. A perfect fit between the model output and the data is therefore difficult to achieve, though effort must be expended to ensure that the data used for calibration are reliable. Simulation-based DTA models are also stochastic (the level of stochasticity is typically much lower than that in microscopic models), so that multiple runs with the same inputs can generate slightly different outputs. The nonanalytical and nonlinear nature of the measures of fit and the DTA itself further complicate any attempt to devise robust and efficient optimization algorithms. Finally, as has been discussed earlier, dynamic O-D estimation is a large-scale problem that grows rapidly with geography and temporal resolution. Most of the existing least-square formulations may not be able to handle a typical regional model with thousands of zones. Some linear transformation approximation techniques appear to be computationally efficient in handling a vast number of zones.

It is worth noting that O-D calibration should not be perceived as being independent of capacity calibration. After all, the output of the DTA model (for a given set of O-D flows) still depends on other DTA inputs such as route choice parameters, capacities and link or segment performance functions. As discussed earlier, the capacity effects and assignment effects need to be addressed prior to commencing O-D calibration, to ensure the presence of reasonable capacity and assignment quality during O-D calibration.

When trying to add a time-varying profile to a demand matrix, it must be recognized that this process has a certain amount of uncertainty, especially if the methodology is based directly on survey data (e.g., the times at which people report that they arrive to and leave from work), as opposed to traffic measurements. In particular, it may be necessary to apply some amount of time-smoothing to reduce the demand peak if it is suspected that it may be too high. A relatively small reduction of the peak 15-min demand, e.g., 10%, may be enough to make a significant difference in the resulting congestion levels. From a practical standpoint, it is best to start adding the demand profile gradually, starting with larger time slices, and then refining to smaller ones if required or justified.

Locations from which traffic data are collected are often a concern for a modeler. In most cases, certain traffic or speed data may exist through prior studies. Sometimes, these data may concentrate only on specific freeways or corridors. Some regions, however, may have relatively richer traffic count data for freeways from several hundreds of sensors spread out over the entire region over several years that can be easily retrieved (e.g., Twin Cities, Minnesota). At times, additional resources may be available to collect more data. The commonly asked questions are how many locations are considered adequate for DTA and where should they be? Compared to the calibration for a microscopic simulation model which usually encompasses a small area so that calibration could be carried out for most links or turning movements, it may not be practical to perform the same level of calibration details and coverage for a relatively larger DTA network. An intuitive approach is to select the minimal number of locations which combined carry traffic from as many O-D pairs as possible. In other words, if one is to add more data, the best locations may be those carrying traffic from O-D pairs not previously covered by existing locations. Determining these sensor locations can be nontrivial and time-consuming. While this is still an active research area, some DTA developers are trying to include an optimization–data analysis tool to assist model users in making a better decision in this regard.

After all, after demand adjustment, comparison between original and adjusted matrices should be made to ensure changes are correlated and reasonable.

After conducting careful examination and correction actions at network, assignment and O-D levels, the overall calibration quality is supposedly improved as each step contributes to better or more correct network coding and better matching with field data. In achieving the overall calibration outcome, it is recommended that one checks the fit between the model output and the observed data for each time interval in the study period, as this will provide important clues about specific remaining issues in the network. Typically, error statistics such as the root mean square error and root mean square normalized error are computed for each time interval, with each measurement providing one data point per statistic. Other error statistics such as the root mean square percent error have also been reported in the literature.

While the statistics are important in imparting rigor to the calibration process, graphical comparisons may serve as quick checks of accuracy. Plots of model outputs against observed data can help one easily to decide whether the current solution is acceptable. Contour plots of data such as speeds can also be useful in intuitively comparing the location, start and duration of congestion.

SCENARIO ANALYSIS

Analysts develop DTA models that reflect observed transportation network conditions so that comparable models of alternative scenarios can be developed and evaluated relative to the base model. Alternative scenarios might be developed to reflect the effect of different future demand and network scenarios. Future year scenarios usually include future predicted demand, but the network configuration may be depicted by various build or no-build scenarios. In addition to scenarios related to network capacity, scenarios that can also be evaluated in DTA models include those involving alternate network control (e.g., ramp metering or signal coordination), information strategies (e.g., placement of dynamic message signs, or setup of various pretrip or en route information mechanisms), value pricing, or integrated corridor management. A wealth of literature in the last decade demonstrated use of a DTA model can be applied to a wide spectrum of application; several such instances are also discussed in the section on What Applications Find DTA Models Advantageous.

The build and no-build scenarios and their comparative analyses are not that different from similar analyses done with static network models. However, extending based year demand and network specifications to future years may require additional considerations. In the base year, signal timing settings are usually determined during the network building process based on actual signal timing data. In future years, signal timing settings are usually unknown. It is not reasonable to assume that current traffic control settings will apply without change in the future, given that future flow patterns that emerge from future demand and network relationships may cause the current setting to perform poorly in future years.

To approximate traffic control settings for a future scenario, one could use the base-year settings to generate an initial assignment solution. The DTA model solution given these base-year control settings will provide a new traffic flow pattern. Traffic signal optimization software or HCM analyses can be applied with these flows, and new signal timing parameters determined. The DTA model can be solved again, and the procedure repeated until a reasonably stable flow pattern is developed. At times, if integrating with a signal optimization model is not possible, manual adjustment of signal timing at isolated congested intersections may also lead to reasonable timing setting and traffic flows.

Besides the issue of updating traffic control settings for future scenarios, another potential difficulty may exist when seeking to apply the calibrated base-year demand to future year. If demand matrices are adjusted or calibrated during the building of the base-year model, how should one use such calibrated demand matrices in the future year scenario? This is an open question. There is no rationale to support the idea that the adjustments done in the base-year should also be done in the future, or should be done proportionally relative to the likely total growth in demand, or should be applied in a spatially uniform way (although techniques of this kind are typically applied in practice). Future demand may also include a new land use development that does not exist in the base-year network; therefore, updates made to that particular zone in the base year are not relevant to the predicted future demand. However, some adjustment must be done, as the same inadequacies that can be measured in the base year may still be present in the future year demand, although unquantifiable. If demand adjustment has to be done, one may consider scaling up the similar adjustment to future demand, plus special considerations to specific zones with known future growth factors (e.g., a planned community due to scheduled deployment of military personnel over a certain period). Since this suggestion is given out of necessity and not necessarily based on any supportive theory, it is recommended that as much effort as possible be put into calibration of the network, traffic control, and corrections to the demand model as possible to minimize the extent to which demand adjustment for future year is required.

CONTINUAL SYSTEM MONITORING AND RECALIBRATION

A transportation network is constantly changing in various ways. In the short term, people are combining activities that help them achieve their daily goals. These goals and activities can change from day to day, causing variations in demand patterns. For example, the typical weekday commute to and from a work place can be perturbed by the need to visit a grocery store or drop off or pick up children from after-school programs. In the medium term, people may change their lifestyles, resulting in activity pattern modifications. In the long term, residential location choices and job changes can cause traffic patterns to shift. Land use changes may also cause activity patterns to reorganize in space and time as people try to work their usual activities into a changing landscape of shopping areas, recreational opportunities and employment centers.

The above demand fluctuations must be juxtaposed to changes in the physical network itself. Repair and maintenance schedules might warrant that some links be fully or partially closed temporarily. New transportation links might be added, while older ones may be decommissioned. There will undoubtedly be a learning process as travelers, both seasoned and otherwise, adjust to small and large changes.

Traffic conditions, therefore, may always be in a state of flux. Before reusing an existing model on a new project, the model user must remember to validate his or her model with new (current) data to ensure that its predictions are sufficiently accurate. When significant deviations are identified, recalibration and revalidation may be necessary to restore the desired accuracy level. This step will require a regular data collection effort supported by expertise in the area of rigorous model calibration and validation. Ideally, if a planning agency realizes and appreciates the value and benefit of a regional DTA model, continual efforts and resources may be planned and committed to regularly keep the DTA model up to date. This update can be put in place in conjunction with regular update of the travel demand model.

Conclusion

As a practitioner once said "DTA is now well heard but still not well understood." Increasingly used by transportation model developers, even in respect to various distinctly different model concepts and implementations, the term DTA has caused escalated confusion for practitioners in recent years. Such confusion hinders the willingness of a practitioner to apply DTA technology. Further, those who have decided to use one particular model without having sufficient basic understanding of DTA may be subject to future difficulty in interpreting model results and thus risk mismatching the appropriate dynamic model to specific application problems.

The main goals of this document is not to set the standard for DTA, but to present and depict the concept of DTA as defined by literature, to discuss general modeling issues and to present, with respect to adoption of DTA, decision-making considerations for both novice and experienced transportation modeling practitioners.

As a part-time effort of a group of enthusiastic authors consisting of researchers, DTA model developers, and experienced practitioners, this work is by no means exhaustive due to the numerous constraints of volunteer effort. Nonetheless, it represents the commitments of the TRB Committee on Transportation Network Modeling's long-lasting interest in helping advance the field of transportation planning and traffic analysis practices. It is hoped that this document will serve as the catalyst for more dialogues and interest among the DTA application, research, and model development communities. Finally, revised or additional expanded editions are certainly possible, but readers' constructive inputs and suggestions will play a crucial role in guiding future publication activities. An effective mechanism is currently planned for readers to interact with the Primer's authors to facilitate extended discussions and idea exchanges within or beyond the scope of this publication in the general context of DTA.

References

1. *Highway Capacity Manual 2000*. Transportation Research Board of the National Academies, Washington, D.C., 2000.
2. Friesz, T. L., R. N. Tobin, N. Mehta, and S. Ganjalizadeh. Day-to-Day Dynamic Network Disequilibrium and Idealized Driver Information Systems. *Operations Research*, Vol. 42, No. 6, 1994, pp. 1120–1136.
3. Friesz, T. L., D. Bernstein, and R. Stough. Dynamic Systems, Variational Inequalities and Control Theoretic Models for Predicting Urban Network Flows. *Transportation Science*, Vol. 30, No. 1, 1996, pp. 14–31.
4. Cantarella, G. E., and E. Cascetta. Dynamic Processes and Equilibrium in Transportation Networks: Towards a Unifying Theory. *Transportation Science*, Vol. 29, No. 4, 1995, pp. 305–329.
5. *Traffic Analysis Tools Videos exclusively from FHWA*. FHWA, U.S. Department of Transportation, 2008. www.developmentspace.net/trafficanalysistools/dyna_vid2.htm.
6. Shelton, J., and E. Nava. Evaluation of Alternatives for Zaragoza/I-10 Interchange—Final Report. Texas Transportation Institute, El Paso, 2007.
7. Shelton, J., and E. Nava. Analysis of Mitigation Strategies for I-10 Corridor Hot-Spots—Final Report. Texas Transportation Institute, El Paso, 2007.
8. Kettner, S. Examining Capacity Increasing Measures with the Dynameq Model: Equilibrium Traffic Assignment. Presented at 13th World Congress and Exhibition on Intelligent Transportation Systems, London, 2006.
9. *Integrated Corridor Management Systems Program Plan*. FHWA, U.S. Department of Transportation, 2008. www.fhwa.dot.gov/crt/roadmaps/icmprgmplan.cfm. Accessed October 24, 2008.
10. Zhou, X., H. S. Mahmassani, and K. Zhang. Dynamic Micro-Assignment Modeling Approach for Integrated Multimodal Urban Corridor Management. *Transportation Research Part C: Emerging Technologies*, Vol. 16, No. 2, 2008, pp. 167–186.
11. Mahmassani, H., S. Eisenman, and S. Hamdar. Evaluation of Operational Strategies for Integrated Corridor Management. Presented at 85th Annual Meeting of Transportation Research Board, Washington, D.C., 2006.
12. *Integrated Corridor Management (ICM) Quarterly Newsletter—Winter 2008*. Research and Innovative Technology Administration, 2008. www.its.dot.gov/icms/docs/knowledgebase/html/news_winter08.htm. Accessed December 1, 2008.
13. Wilson, S. Applications of Dynamic Traffic Assignment in Twin Cities Area. Presented at 2008 Innovation in Travel Model. Portland, Ore., 2008.
14. Hu, H., B. Williams, N. Rouphail, A. Khattak, and X. Zhou. Modeling the Role of Transportation Information in Mitigating Major Capacity Reductions in a Regional Network. Presented at 88th Annual Meeting of Transportation Research Board, Washington, D.C., 2009.
15. Tian, X., M. Mahut, M. Jha, and M. Florian. Dynameq Application to Evaluating the Impact of Freeway Reconstruction. Presented at 86th Annual Meeting of Transportation Research Board, Washington, D.C., 2007.
16. *R11: Strategic Approaches at the Corridor and Network Level to Minimize Disruption from the Renewal Process*. Strategic Highway Research Program 2, Transportation Research Board of the National Academies, Washington, D.C., 2009. http://144.171.11.40/cmsfeed/TRBNetProject Display.asp?ProjectID=2678.
17. *C05: Understanding the Contribution of Operations, Technology, and Design to Meeting Highway Capacity Needs.* Strategic Highway Research Program 2, Transportation Research Board of the National Academies, Washington, D.C., 2009. http://144.171.11.40/cmsfeed/ TRBNetProjectDisplay.asp?ProjectID=2164.

18. Shelton, J., Y.-C. Chiu, and B. Kuhn. Evaluation of Managed Ramp Strategies for Incident and Congestion Management. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.
19. Kuhn, B., K. Balke, N. Chaudhary, D. Jasek, G. Karkee, K. Obeng-Boampong, J. Shelton, and S. Venglar. Managed Lanes Strategies Feasible for Freeway Ramp Applications. Texas Transportation Institute, College Station, 2008.
20. Chiu, Y.-C., X. Zhou, and J. Hernandez. Evaluating Urban Downtown One-Way to Two-Way Street Conversion Using a Multiple Resolution Simulation and Assignment Approach. *ASCE Journal of Urban Planning and Development*, Vol. 133, No. 4, 2007, pp. 222–232.
21. Mahut, M., M. Florian, N. Tremblay, M. Campbell, D. Patman, and Z. K. McDaniel. Calibration and Application of A Simulation-Based Dynamic Traffic Assignment Model. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1876,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 101–111.
22. Palma, A. D., and F. Marchal. *Estimation of Non-Recurrent Congestion Using Dynamic Traffic Simulations.* 2004.
23. Sisiopiku, V., X. Li, K. Mouskos, C. Kamga, C. Barrett, and A. M. Abro. Dynamic Traffic Assignment Modeling for Incident Management. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1994,* Transportation Research Board of the National Academies, 2007, pp. 110–116.
24. Wirtz, J., J. Schofer, and D. Schulz. Managing Large Scale Transportation Disruptions: Using Simulation to Test Traffic Incident Management Strategies: Illustrating the Benefits of Pre-Planning. Northwestern University, Evanston, Ill., 2004, p. 21.
25. Bottom, J. *Consistent Anticipatory Route Guidance*. Department of Civil and Environmental Engineering. Massachusetts Institute of Technology, Cambridge, 2000.
26. Boxill, S., and L. Yu. An Evaluation of Traffic Simulation Models for Supporting ITS Development. In *Technical Report 167602-1*, Center for Transportation Training and Research, Texas Southern University, 2000, p. 114.
27. Chiu, Y.-C., and N. Huynh. Location Configuration Design for Dynamic Message Signs Under Stochastic Incident and ATIS Scenarios*. Transportation Research Part C: Emerging Technologies*, Vol. 15, No. 1, 2007, p. 33.
28. Paz, A. Behavior-Consistent Deployable Real-Time Traffic Routing Under Information Provision. School of Civil Engineering, Purdue University, West Lafayette, Ind., 2007.
29. Chien, S., K. Mouskos, and A. Ziliaskopoulos. Development of a Dynamic Simulation/Assignment Model for NJDOT's I80 ITS Priority Corridor. New Jersey Institute of Technology, Newark, 2005.
30. Abdelghany, A. F., K. F. Abdelghany, H. S. Mahmassani, and P. M. Murray. Dynamic Traffic Assignment in Design and Evaluation of High-Occupancy Toll Lanes. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1733,* TRB, National Research Council, 2007, pp. 39–48.
31. Vovsha, P., W. Davidson, and R. Donnelly. Making the State of the Art the State of the Practice: Advanced Modeling Techniques for Road Pricing. *Proc., Expert Forum on Road Pricing and Travel Demand Modeling,* Volpe National Transportation Systems Center, Washington, D.C., 2006.
32. *C04: Improving Our Understanding of How Highway Congestion and Pricing Affect Travel Demand.* Strategic Highway Research Program 2, Transportation Research Board of the National Academies, Washington, D.C., 2009. http://144.171.11.40/cmsfeed/TRBNetProjectDisplay.asp? ProjectID=2163.
33. Henk, R. H., A. J. Ballard, R. L. Robideau, W. G. Peacock, P. Maghelal, M. K. Lindell, C. S. Prater, L. Loftus-Otway, P. Siegesmund, R. Harrison, L. Lasdon, Y.-C. Chiu, H. Zheng, J. Perkins, C. Lewis, and S. Boxill. Disaster Preparedness in Texas: Final Report. Texas Transportation Institute, College Station, 2007, p. 360.
34. Tuydes, H., and A. Ziliaskopoulos. Tabu-Based Heuristic for Optimization of Network Evacuation Contraflow. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006.

35. Sbayti, H., and H. S. Mahmassani. Optimal Scheduling of Evacuation Operations. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1964*, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 238–246.
36. Balakrishna, R., Y. Wen, M. Ben-Akiva, and C. Antoniou. Simulation-Based Framework for Transportation Network Management for Emergencies. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2041*, Transportation Research Board of the National Academies, 2008, pp. 80–88.
37. Gao, Y., Y.-C. Chiu, S. Wang, and S. Küçükyavuz. Optimal Refueling Station Location and Supply Planning for Hurricane Evacuation. Presented at 89th Annual Meeting of Transportation Research Board, Washington, D.C., 2010.
38. Abdelgawad, H., B. Abdulhai, and M. Wahba. Multiobjective Optimization for Multimodal Evacuation. Presented at 89th Annual Meeting of Transportation Research Board, Washington, D.C., 2010.
39. Zheng, H., Y.-C. Chiu, P. Mirchandani, and M. Hickman. Modeling of Evacuation and Background Traffic for an Optimal Zone-Based Vehicle Evacuation Strategy. Presented at 89th Annual Meeting of the Transportation Research Board, Washington, D.C., 2010.
40. Lin, D.-Y., N. Eluru, S. T. Waller, and C. Bhat. Evacuation Planning Using the Integrated System of Activity-Based Modeling and Dynamic Traffic Assignment. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
41. Kwon, E., and S. Pitt. Evaluation of Emergency Evacuation Strategies for Downtown Event Traffic Using a Dynamic Network Model. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1922*, Transportation Research Board of the National Academies, 2005, pp. 149–155.
42. Bai, S., Y.-C. Chiu, and D. Niemeier. Using Dynamic Assignment to Improve Regional Mobile Emissions Estimation. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.
43. Bai, S., Y.-C. Chiu, and D. Niemeier. Comparative Analysis of Using Trip-Based Versus Link-Based Traffic Data for Regional Mobile Source Emissions Estimation. *Atmospheric Environment*, Vol. 41, No. 35, 2007, pp. 7512–7523.
44. Meritz, D. $8 Million Project Planned to Ease Access to I-10 near UTEP. *El Paso Times*, 2007.
45. Velan, S., and P. Volet. Linking Demand from a Regional Travel Demand Forecasting Models to a New DTA Model: The Montreal Experience. Presented at Innovations in Travel Modeling 2008, Portland, Ore.
46. Xyntarakis, M., E. Sall, J. Hicks, and B. Charlton. Regional Dynamic Traffic Assignment for Real-World Travel Demand Models: Trying It Out in San Francisco. Presented at 89th Annual Meeting of the Transportation Research Board, Washington, D.C., 2010.
47. Lin, D.-Y., N. Eluru, S. T. Waller, and C. R. Bhat. Integration of Activity-Based Modeling and Dynamic Traffic Assignment. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.
48. *C10 (A): Partnership to Develop an Integrated, Advanced Travel Demand Model and a Fine-Grained, Time-Sensitive Network,* Strategic Highway Research Program 2, Transportation Research Board of the National Academies, Washington, D.C., 2009. http://144.171.11.40/cmsfeed/TRB NetProjectDisplay.asp?ProjectID=2829.
49. *C10 (B): Partnership to Develop an Integrated, Advanced Travel Demand Model and a Fine-Grained, Time-Sensitive Network*, Strategic Highway Research Program 2, Transportation Research Board of the National Academies, Washington, D.C., 2009. http://144.171.11.40/cmsfeed/TRBNet ProjectDisplay.asp?ProjectID=2828.
50. *L04: Incorporating Reliability Performance Measures in Operations and Planning Modeling Tools.* Strategic Highway Research Program 2, Transportation Research Board of the National Academies, Washington, D.C., 2009. http://144.171.11.40/cmsfeed/TRBNetProject Display.asp?ProjectID=2193.

51. Dong, J., H. S. Mahmassani, and C.-C. Lu. How Reliable Is This Route? Predictive Travel Time and Reliability for Anticipatory Traveler Information Systems. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1980*, Transportation Research Board of the National Academies, 2006, pp. 117–125.
52. Hawas, Y., H. S. Mahmassani, Y.-C. Chiu, N. Huynh, and A. Abdelfatah. Development of DYNASMART-X Software for Real-Time Dynamic Traffic Assignment. Center for Transportation Research, Austin, Texas, 1997.
53. Ben-Akiva, M. Development of a Deployable Real-Time Dynamic Traffic Assignment System. Evaluation Report (Part A): Evaluation of Estimation and Prediction Capabilities. Massachusetts Institute of Technology, Cambridge, 2003, p. 144.
54. Wen, Y., R. Balakrishna, M. Ben-Akiva, and S. Smith. Online Deployment of Dynamic Traffic Assignment: Architecture and Run-Time Management. *IEE Proceedings Intelligent Transport Systems* (now IET Intelligent Transport Systems), Vol. 153, No. 1, 2006, pp. 76–84.
55. Mahmassani, H. S., and R. Herman. Dynamic User Equilibrium Departure Time and Route Choice on Idealized Traffic Arterials. *Transportation Science*, Vol. 18, No. 4, 1984, pp. 362–384.
56. Ziliaskopoulos, A. K., and L. Rao. A Simultaneous Route and Departure Time Choice Equilibrium Model on Dynamic Networks. *International Transactions in Operational Research*, Vol. 6, No. 1, 1999, pp. 21–37.
57. Huang, H. J., and W. H. K. Lam. Modeling and Solving the Dynamic User Equilibrium Route and Departure Time Choice Problem in Network with Queues. *Transportation Research Part B*, Vol. 36, 2002, pp. 253–273.
58. Ben-Akiva, M., and M. Bierlaire. Discrete Choice Models with Applications to Departure Time and Route Choice. In *Handbook of Transportation Science* (H. R. Kluwer, ed.), 2003.
59. Li, Y., A. Ziliaskopoulos, and S. Waller. Linear Programming Formulations for System Optimum Dynamic Traffic Assignment with Arrival Time-Based and Departure Time-Based Demands. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1667*, Transportation Research Board of the National Academies, Washington, D.C., 1999, p. 52–59.
60. Lu, C.-C., H. S. Mahmassani, and X. Zhou. A Bi-Criterion Dynamic User Equilibrium Traffic Assignment Model and Solution Algorithm for Evaluating Dynamic Road Pricing Strategies. *Transportation Research Part C: Emerging Technologies*, Vol. 16, No. 4, 2008, p. 371.
61. *Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software*. Turner-Fairbank Highway Research Center, McLean, Va., 2004, p. 146.
62. Lundgren, J. T., and A. Peterson. A Heuristic for the Bilevel Origin-Destination-Matrix Estimation Problem. *Transportation Research Part B: Methodological*, Vol. 42, No. 4, 2008, p. 339.
63. Balakrishna, R. Off-Line Calibration of Dynamic Traffic Assignment Models. Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, 2006, p. 210.
64. Peeta, S., and A. K. Ziliaskopoulos. Foundations of Dynamic Traffic Assignment: The Past, the Present and the Future. *Networks and Spatial Economics*, Vol. 1, Nos. 3–4, 2001, pp. 233–265.
65. *Traffic Analysis Toolbox Volume I: Traffic Analysis Tools Primer*. Turner-Fairbank Highway Research Center, McLean, Va., 2004, p. 34.
66. *Traffic Analysis Toolbox Volume II: Decision Support Methodology for Selecting Traffic Analysis Tools*. Turner-Fairbank Highway Research Center, McLean, Va., 2004, p. 108
67. Box, G. E. P., and N. R. Draper. *Empirical Model-Building and Response Surfaces*. Wiley and Sons, New York, 1987.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **www.TRB.org**

www.national-academies.org

TRANSPORTATION RESEARCH BOARD
500 Fifth Street, NW
Washington, DC 20001