Understanding the Contributions of Operations, Technology, and Design to Meeting Highway Capacity Needs

DETAILS

0 pages | 8.5 x 11 | PAPERBACK
ISBN 978-0-309-43512-3 | DOI 10.17226/22803

AUTHORS

Kittelson & Associates, Inc.

The Second
STRATEGIC HIGHWAY RESEARCH PROGRAM

SHRP 2 REPORT S2-C05-RW-1

Understanding the Contributions of Operations, Technology, and Design to Meeting Highway Capacity Needs

Kittelson & Associates, Inc.

in association with

Institute for Transportation Research and Education

Ruhr University Bochum

University of Utah

TRANSPORTATION RESEARCH BOARD

WASHINGTON, D.C.
2014
www.TRB.org

Subject Areas

Highways
Operations and Traffic Management
Planning and Forecasting

The Second Strategic Highway Research Program

America's highway system is critical to meeting the mobility and economic needs of local communities, regions, and the nation. Developments in research and technology—such as advanced materials, communications technology, new data collection technologies, and human factors science—offer a new opportunity to improve the safety and reliability of this important national resource. Breakthrough resolution of significant transportation problems, however, requires concentrated resources over a short time frame. Reflecting this need, the second Strategic Highway Research Program (SHRP 2) has an intense, large-scale focus, integrates multiple fields of research and technology, and is fundamentally different from the broad, mission-oriented, discipline-based research programs that have been the mainstay of the highway research industry for half a century.

The need for SHRP 2 was identified in *TRB Special Report 260: Strategic Highway Research: Saving Lives, Reducing Congestion, Improving Quality of Life,* published in 2001 and based on a study sponsored by Congress through the Transportation Equity Act for the 21st Century (TEA-21). SHRP 2, modeled after the first Strategic Highway Research Program, is a focused, time-constrained, management-driven program designed to complement existing highway research programs. SHRP 2 focuses on applied research in four areas: Safety, to prevent or reduce the severity of highway crashes by understanding driver behavior; Renewal, to address the aging infrastructure through rapid design and construction methods that cause minimal disruptions and produce lasting facilities; Reliability, to reduce congestion through incident reduction, management, response, and mitigation; and Capacity, to integrate mobility, economic, environmental, and community needs in the planning and designing of new transportation capacity.

SHRP 2 was authorized in August 2005 as part of the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU). The program is managed by the Transportation Research Board (TRB) on behalf of the National Research Council (NRC). SHRP 2 is conducted under a memorandum of understanding among the American Association of State Highway and Transportation Officials (AASHTO), the Federal Highway Administration (FHWA), and the National Academy of Sciences, parent organization of TRB and NRC. The program provides for competitive, merit-based selection of research contractors; independent research project oversight; and dissemination of research results.

SHRP 2 Report S2-C05-RW-1

ISBN: 978-0-309-12929-9

Copyright Information

Note: SHRP 2 report numbers convey the program, focus area, project number, and publication format. Report numbers ending in "w" are published as web documents only.

Notice

The project that is the subject of this report was a part of the second Strategic Highway Research Program, conducted by the Transportation Research Board with the approval of the Governing Board of the National Research Council.

The members of the technical committee selected to monitor this project and review this report were chosen for their special competencies and with regard for appropriate balance. The report was reviewed by the technical committee and accepted for publication according to procedures established and overseen by the Transportation Research Board and approved by the Governing Board of the National Research Council.

The opinions and conclusions expressed or implied in this report are those of the researchers who performed the research and are not necessarily those of the Transportation Research Board, the National Research Council, or the program sponsors.

The Transportation Research Board of the National Academies, the National Research Council, and the sponsors of the second Strategic Highway Research Program do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of the report.

SHRP 2 Reports

Available by subscription and through the TRB online bookstore:
www.TRB.org/bookstore

Contact the TRB Business Office:
202-334-3213

More information about SHRP 2:
www.TRB.org/SHRP2

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 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. C. D. (Dan) Mote, Jr., 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. Victor J. Dzau 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. C. D. (Dan) Mote, Jr., 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

SHRP 2 STAFF

ACKNOWLEDGMENTS

This work was sponsored by the Federal Highway Administration in cooperation with the American Association of State Highway and Transportation Officials. It was conducted in the second Strategic Highway Research Program (SHRP 2), which is administered by the Transportation Research Board of the National Academies. The project was managed by Stephen Andrle, Deputy Director for SHRP 2. The authors thank the other members of the research team, whose contributions and insights made this research project possible.

FOREWORD

Stephen J. Andrle, *SHRP 2 Deputy Director*

Adding transportation capacity improves safety, accessibility, and the economic health of a region but often comes with an economic, social, or environmental price. The public is aware of the price and must be convinced that new capacity is really needed before they will support expansion. This implies that operational improvements to the roadway may be warranted to improve efficiency some years before a major capacity expansion is proposed. This report addresses the question, What gains in sustainable flow can be obtained from operational improvements to roadways? The report uses the term *sustainable flow* rather than *capacity*, to avoid confusion with other definitions of capacity. The researchers used enhanced simulation methods and network diagnostics, including travel time and travel time reliability, to test various types of operational improvements. This report will be of interest to traffic engineers, freeway managers, roadway designers, and transportation planners.

This report builds on an emerging body of literature that suggests capacity is not a constant value but is a variable. Capacity (or sustainable flow) depends on driving behavior, the composition of the drivers at the moment, driver familiarity with the roadway, the mix of vehicle types, trip purposes, weather, signalization, and the presence of upstream or downstream bottlenecks. Therefore, if operational improvements are made to a roadway, the sustainable flow will be influenced by these variables. The research also suggests that the success of operational improvements in increasing sustainable flow is partially dependent on the configuration of the roadway network. The same effect is not necessarily achieved with every application. Finally, drivers learn and adapt to new roadway configurations. To properly analyze the effect of an operational improvement, this learning behavior should be recognized. The implication of these contextual factors is that simulation methods are needed to estimate the increase in sustainable throughput achieved from a package of operational improvements.

The researchers added capabilities to existing simulation models to reflect the stochastic nature of capacity for freeways and arterials and day-to-day learning in the route selection algorithms. They also used link, corridor, and network diagnostic features for evaluation of alternatives and to identify locations on the network where the benefits of implementing operational strategies appear high.

The researchers evaluated emerging technologies and network operations treatments, from which they selected 25 for testing, including actions such as ramp metering, interchange modification, queue management, narrow lanes, adaptive signals, and pretrip information. A simulation modeling approach was used to test strategies on networks and data from Fort Worth, Texas, and Portland, Oregon.

The results of applying the simulation techniques to the various strategies indicate the following:

- Multiple performance measures are needed to obtain a complete assessment.
- Measures should reflect link, corridor, and network characteristics.
- Driver's route choice must be considered to analyze network-level effects.
- A representative cross section of corridors and O-D pairs should be evaluated.
- The effectiveness of a particular operational improvement can only be evaluated in context.
- The reliability of travel times may improve even when the actual travel time remains unchanged.

The report describes the simulation modeling framework and the equations used for the enhancing simulation software.

CONTENTS

Executive Summary

Continuing growth in urban travel demand will inevitably require more physical capacity in the transportation system. However, because of limited financial resources, high construction costs, environmental considerations, long timelines, and an increasingly complex regulatory process, capacity-adding projects have become actions of last resort. It therefore behooves decision makers, planners, and engineers to evaluate operational improvement strategies that can—singly or in combination—eliminate or mitigate the need for a more traditional highway construction project.

Effectively evaluating the wide range of operational improvement strategies that are available is not a trivial matter, particularly when their performance is to be compared against the construction of new lanes. Traditional travel demand forecasting models are not effective for this kind of comparative analysis for several reasons:

- They assume that all drivers have perfect knowledge of the travel time on each of the travel paths available to them, an assumption that masks the effectiveness of operational improvement strategies aimed at improving driver awareness.
- They assume that the capacity of a freeway link or an arterial segment is a constant value, whereas an emerging body of research indicates that such capacity is better represented as a random variable (*1–3*). This limitation reduces the effectiveness of traditional tools for comparing alternatives because fluctuating capacity introduces variability that measurably affects vehicle assignments and network performance characteristics.
- They are not usually sensitive to the effects that upstream bottlenecks and blockages can have on downstream service rates. As an example, the models do not generally recognize that when the upstream queue of a separate turn lane extends into the adjoining through lane, this blockage prevents through traffic from reaching the downstream intersection for as long as the blockage exists, even when the downstream signal is green.
- They assume that all vehicle trips identified in the origin–destination (O-D) matrix will be completed by the end of the time period being analyzed, regardless of whether there is actually sufficient capacity to accommodate these trips within the specified time window. Thus, each vehicle trip is assigned to an entire travel path from origin to destination, even if some bottlenecks along that path operate with a volume/capacity ratio greater than 1.0.

Some modeling advancements are beginning to address these issues, but the advancements have not yet reached the point of practical and regular application, nor do they address all of the issues simultaneously. Ideally, the analysis methods should enable evaluation of improvement strategies that cut across the full spectrum of operations, technology, and design. They should also provide multiple performance measures that can be used to evaluate different strategies according to their impacts at the point, link, corridor, and network levels.

This report summarizes the results of a capacity project undertaken through the second Strategic Highway Research Program (SHRP 2) to advance the state of practice in this area. The objectives of this project were to (a) quantify the capacity benefits, individually and in combination, of operations, design, and technology improvements at the network level for both new and existing facilities; (b) provide information and tools to analyze operational improvements as an alternative to traditional construction; and (c) develop guidelines for sustained service rates (SSRs) to be used in planning networks for limited access highways and urban arterials (*2*).

Strategies Selected for Testing

Table ES.1 lists 25 operational strategies that were selected from an initial list of more than 100 as particularly effective in enhancing the performance characteristics of links, corridors, and networks. Some of the strategies are applicable only to freeways, some only to arterials, and some to both.

These strategies stood out from others because of the following characteristics:

- Their ability to reduce recurring congestion effects during peak periods;
- Their ability to be implemented rather quickly by agency decision makers, compared with major capital improvement projects;
- The feasibility of implementing them considering economic, social, political, and environmental factors;
- Their general capacity-enhancing effects; and
- The number, location, and characteristics of known successful applications.

Network Operations Modeling Approach

The effectiveness of each operational strategy listed in Table ES.1 was found to vary according to the context in which it is applied. Physical factors such as network structure as well as the existence and relative proximity of freeway/arterial alternatives have an important influence on a particular strategy's effectiveness, as do travel desire lines and overall demand levels. It is thus not possible to reliably estimate the effectiveness of a particular operational strategy in a particular network and demand setting from static, location-blind lookup tables. Instead, some form of a travel demand forecasting model is necessary.

Table ES.1. Non-Lane-Widening Strategies to Improve Capacity

Freeway	Arterial	Both
HOV lanes	Signal retiming	Narrow lanes
Ramp metering	Signal coordination	Reversible lanes
Ramp closures	Adaptive signals	Variable lanes
Congestion pricing	Queue management	Truck-only lanes
Pricing by distance	Raised medians	Truck restrictions
HOT lanes	Access points	Pretrip information
Weaving section improvements	Right and left turn channelization	In-vehicle information
Frontage road	Alternate left turn treatments	VMS/DMS
Interchange modifications		

Note: HOV = high-occupancy vehicle; HOT = high-occupancy toll; VMS = variable message sign; DMS = dynamic message sign.

Dynamic traffic assignment (DTA) models are especially advantageous for evaluating strategy effectiveness because they provide a realistic assignment of traffic in oversaturated networks:

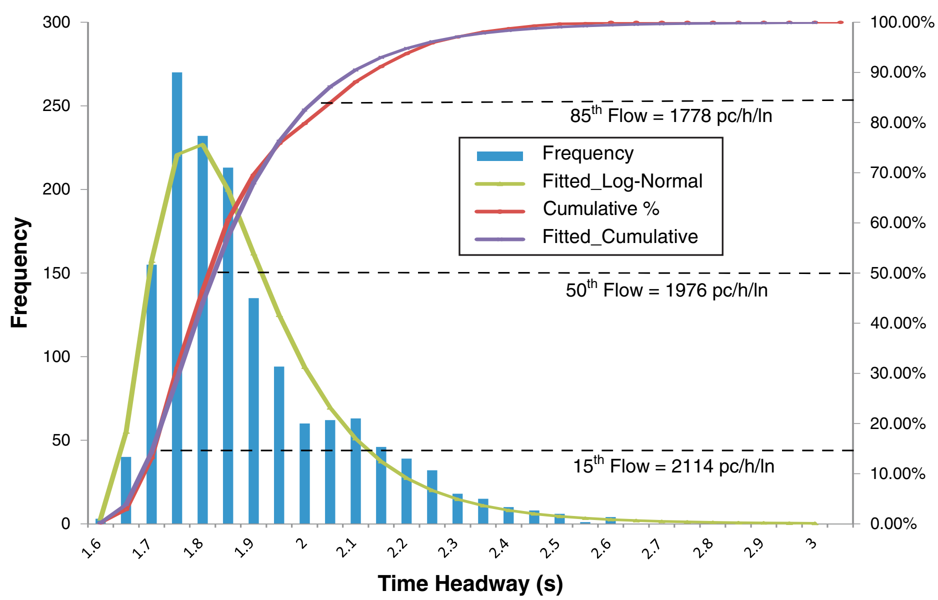
- They recognize that drivers have varying levels of knowledge about the travel time on each of the travel paths available to them.
- They recognize that the effects of congestion and queues can prevent drivers from reaching their destinations in a timely manner and therefore do not assume that all vehicle trips identified in the OD matrix will be completed by the end of the time period being analyzed.

The capabilities of DTA models overcome some but not all of the limitations associated with traditional travel demand models. A review was conducted of available DTA models, and none included all the modeling capabilities desired. Even so, several DTA models use a common and generally accessible graphical user interface (GUI) for input data; among these are DYNASMART-P (Dynamic Network Assignment Simulation Model for Advanced Roadway Telematics: Planning version) (*4*), DynusT (*5*), and DTALite (*6*). For this research project, the internal logic of DYNASMART-P was modified to incorporate several analytic enhancements (described in the following subsections), and the new version served as the test engine for the validation and demonstration activities. The modified version of DYNASMART-P developed in this project is not available for general use, but other DTA models can incorporate these enhancements and at least one open source model (DTALite) has already done so.

A summary of the modeling enhancements incorporated into the updated DTA model follows.

Stochastic Capacity for Freeway Bottlenecks

Traditional DTA models assume a constant capacity for freeway bottlenecks, which are generally located at merge points, lane drops, and weaving areas. However, empirical data reveal that breakdowns occur across a range of volumes, even at the same location. Therefore, a probabilistic approach was developed wherein random capacity values are generated at fixed time intervals (15 minutes) during queue-free time periods based on an empirically derived distribution of pre-breakdown headways as shown in Figure ES.1. This approach results in freeway bottleneck



Note: pc/h/ln = passenger cars per hour per lane.

Figure ES.1. Stochastic headway/capacity flow distribution for freeway bottlenecks.

activity that randomly varies from day to day. During simulation periods when these random freeway queues are present, the queue discharge rate at active bottlenecks is also represented with random variation by modeling it as a time-correlated stochastic process. The simulated freeway environment created by the joint action of the random queue-free capacity and the time-correlated queue discharge provides a realistic representation of day-to-day variation in recurring freeway network congestion. Details on stochastic freeway models are available in Jia, Williams, and Rouphail (*3*).

Stochastic Capacity for Arterials

The bottleneck points for signalized arterials are usually very easy to identify—they are most often located at intersections—and traditional DTA models once again assume a constant saturation flow rate during the green time either for the approaching links or for individual turn movements at these locations. However, it is well known that the saturation flow rate for individual links and turn movements varies significantly according to the behavior of individual drivers. At signalized intersections this is revealed by saturation flow rates that vary from cycle to cycle. Therefore, a stochastic approach was developed that allows the capacity of a signalized intersection to vary during each time interval (typically 15 minutes) according to an empirically observed distribution such as that illustrated in Figure ES.2.

Improved Arterial Bottleneck Representation

The approaches to signalized intersections along arterial roadways often include left and right turn pockets as a way of separating turn movements and increasing capacity. But when the queue length of through and/or turning vehicles extends beyond the length of the turn pocket, the result is a demand blockage that prevents upstream vehicles from taking advantage of the capacity available at the intersection (Figure ES.3). This is an important phenomenon to model in oversaturated networks because it directly affects the efficiency and productivity (or SSRs)

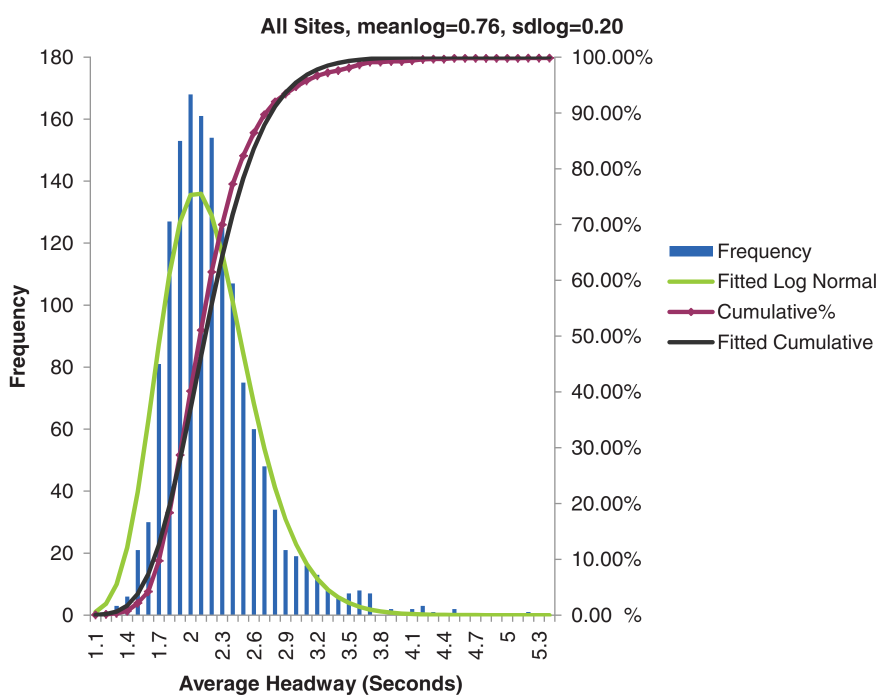


Figure ES.2. Stochastic headway/capacity flow distribution for arterials.

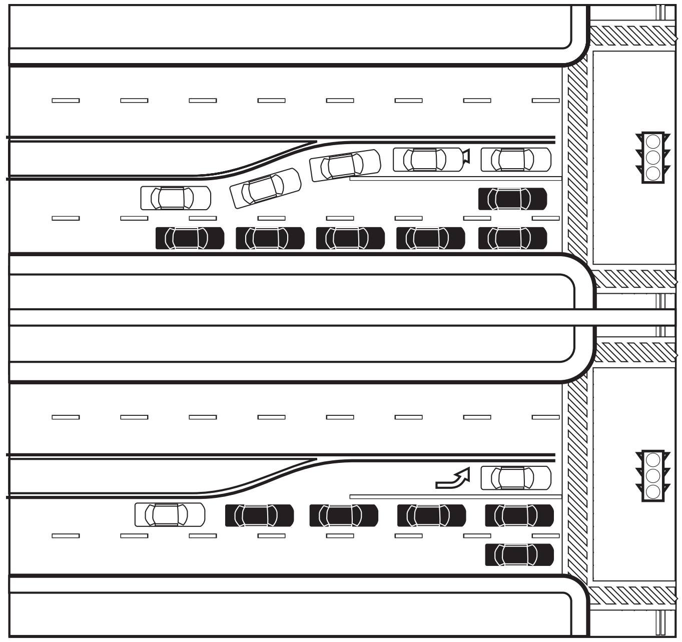


Figure ES.3. Blockage effects of short turn pockets.

of individual links and turn movements. Therefore, an enhanced model was developed that recognizes when queue lengths exceed available storage lengths at these locations and adjusts the downstream discharge rate accordingly.

Day-to-Day Learning

The consequence of stochastic freeway and arterial bottlenecks is that traffic flow and the presence or absence of breakdown conditions at a specific time on any particular link vary from day to day. This, combined with day-to-day variations in travel demand, means that real-world drivers take into account the conditions they have encountered across multiple days to make route choices based on the expected travel time of their available options. This more realistic representation of drivers' route selection processes was incorporated into the enhanced model by simulating multiple consecutive days and allowing a user-defined fraction of randomly selected drivers in each O-D pair the ability to remember their travel time experiences over the most recent days when making their route choices.

As illustrated in Figure ES.4, a simulation-based evaluation framework was used in this study to estimate the system performance for a multiday planning horizon under stochastic link capacity.

In addition to stochastic road capacity, day-to-day travel time variability is affected by the manner in which travelers obtain, process, and react to traveler information. To account for information uncertainty and cognitive limitations of individual travelers, the theory of "bounded rationality" (*4, 7*)—the concept that decision-making abilities are constrained by the information at hand and the available time to make a decision—is adapted in this study to describe route switching and departure time choice behavior.

One of the aims of this study was to enhance a mesoscopic dynamic traffic simulator by incorporating stochastic road capacity for both freeway and arterial links and by developing a new set of day-to-day learning and route updating models under stochastic travel time variations (introduced by variable capacity). The introduction of stochastic capacity at critical points in the network that suffer from queue and congestion more frequently, such as freeway bottlenecks and signalized intersections, enables reasonable and realistic modeling of travel time variability and sustainable flow rates.

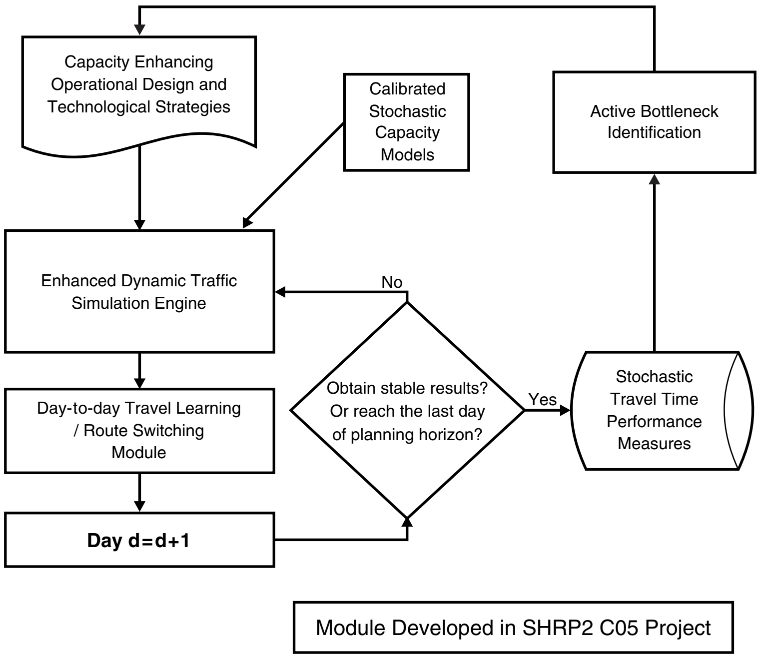
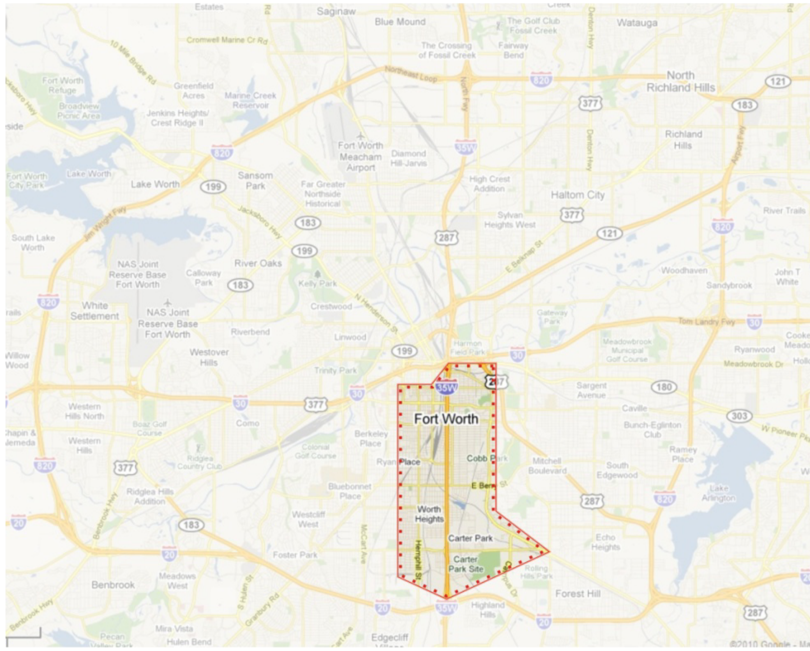


Figure ES.4. Module developed in SHRP 2 C05 project as a capacity-enhancing strategy evaluation framework.

Under stochastic capacity, the travel time experience on a single day can be dramatically affected by the underlying capacity on that particular day. (In this study, "day" is defined as any regular weekday; i.e., a day other than a weekend or a public holiday.) This study therefore developed a set of reliability-oriented system performance measures that consider multiple days' performance in order to systematically evaluate medium-term benefits of traveler information provision strategies.

In addition to the previously described model enhancements, the new network diagnostic features in the following list facilitate network evaluation as well as identification of points and corridors where the potential benefits of implementing one or more operational strategies appear to be high.

- *Active bottleneck identification.* Active bottlenecks are defined here as locations (on both freeways and arterials) where one or more actual breakdowns occur during the simulated days and time periods. They are represented with red circles on the network map. The diameter of each circle is proportional to the number of breakdowns observed during a particular analysis day. The analyst can thus easily identify the network points most susceptible to breakdown, revealing patterns and locations that can suggest specific operational strategies for congestion mitigation.
- *Movement-specific intersection delay.* The delay experienced by individual links and movements can be displayed both visually and in tabular format, allowing the analyst to quantify this important performance measure.
- *Stochastic link performance and breakdown probability.* Travel time variability is arguably at least as important to drivers as delay because variability dictates the amount of buffer time they must build into their travel schedule. This measure (defined as the difference between the 5th and 95th percentile travel times) is reported for each link, corridor, and/or O-D pair and provides further insight to the analyst on vulnerable links and corridors where one or more operational improvement strategies might be appropriate.



Source: © 2010 Google Maps.

Figure ES.5. Fort Worth area.

- *Queue growth and dissipation.* A sliding time bar can be used along with a visual representation of the network to observe the start, growth, and dissipation of queues. This allows the analyst to easily trace link breakdowns to their point of origin, again for purposes of identifying one or more operational improvement strategies that might be appropriate.

Baseline Models

Two separate networks were used to test the enhanced models and demonstrate both the usefulness and the usability of the new methodology. The first network was a very small subarea of the Dallas–Fort Worth, Texas, area (Figure ES.5). The small size of this network produced great efficiencies in testing and debugging the enhanced DTA models, and it was also a good platform for implementing and evaluating each of the 25 operational strategies presented in Table ES.1. The second network was a subarea of the Portland, Oregon, metropolitan area, encompassing approximately 210 traffic analysis zones, 860 nodes, 2,000 links, and more than 200,000 vehicle trips initiated during a 4-hour weekday time interval between 3 p.m. and 7 p.m. Both DTALite and DYNASMART-P were used in this network application; DTALite provided a regional baseline equilibrium as the starting point for the subarea analysis, which was conducted by using DYNASMART-P. This network clearly demonstrated the usefulness of the procedure in a real-world environment and featured the analytic elements of network diagnosis; identification and evaluation of treatment options; and interpretation of the results.

Strategy Testing

A straightforward method was developed to test the effectiveness of one or more operational strategies either as stand-alone projects or as alternatives to traditional new construction projects.

- First, the location of the operational strategy and/or new construction project to be tested is identified and a subarea or network that appropriately surrounds the location is established.

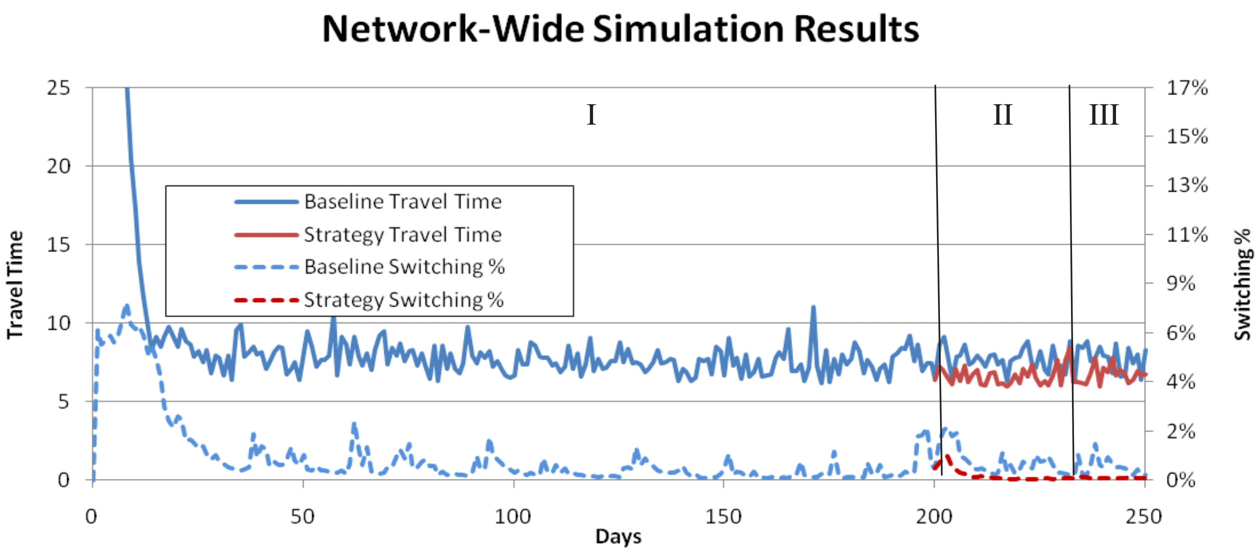


Figure ES.6. Overview of strategy testing plan under stochastic capacity conditions.

- Next, geometric, volume, and operational characteristics of each link in the subarea are identified and provided as inputs to the DTA model, including stochastic capacity distributions at the geometric or operational bottlenecks. Appropriate link, corridor, and/or network performance measures are also established for subsequent evaluation purposes.
- To effectively use the day-to-day learning process and generate results that can be usefully compared, the DTA model must be run under three separate regimes as shown in Figure ES.6. During the baseline stabilization period (Regime I), the DTA model is run for a period of simulated days to achieve equilibrium (i.e., without any of the operational strategies or new construction projects that are to be evaluated). The number of simulated days necessary to achieve equilibrium will vary according to the characteristics of the network and/or subarea being investigated. Figure ES.6 shows that for the Dallas–Fort Worth network the 200-day baseline stabilization period was longer than necessary. This was not a problem because the subarea was small and the runtime for each simulated day was very short. For larger networks, a baseline stabilization period of 50 days may be more appropriate.

After baseline stabilization has been achieved, the operational strategies and/or new construction projects to be evaluated are introduced into the network and the DTA model is run for an additional period of simulated days to allow driver adjustments and achieve stable conditions under the new scenario. This strategy stabilization period is illustrated as Regime II in Figure ES.6. A period of 30 simulated days is generally sufficient to achieve stabilization under Regime II conditions.

After conditions have stabilized in Regime II, the DTA model is run for an additional 20 simulated days (Regime III in Figure ES.6). The results of this period are compared with those of Regime I for the purpose of evaluating the effectiveness of the strategies being tested.

This methodology provides useful information about the effectiveness of the operational strategies and new construction projects being evaluated. As an example, consider the capacity addition scenarios that were tested for a southbound freeway corridor section in the Dallas–Fort Worth subarea network. Figure ES.7 illustrates the existing (baseline) condition and three lane addition projects (A, B, and C) that were contemplated and tested.

In addition to these lane addition projects, four operational strategy alternatives to the projects were evaluated:

- An advanced traveler information system (ATIS) strategy in which the fraction of drivers with access to pretrip information (e.g., via radio, television, or the Internet) increased from 1% to 10%;

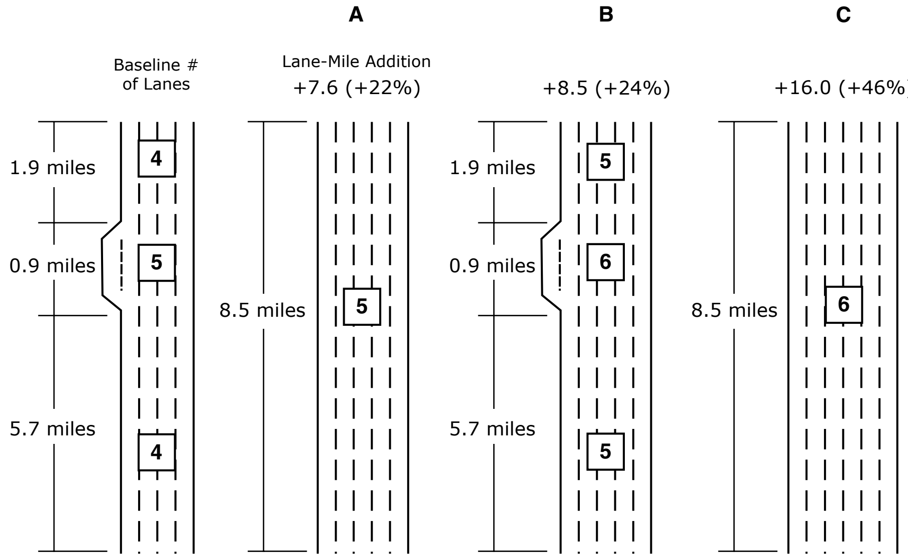


Figure ES.7. Capacity-enhancing scenarios for a southbound freeway corridor.

- Another ATIS strategy in which the fraction of drivers with access to en route information (e.g., through in-vehicle navigation systems) increased from 1% to 10%;
- An operational modification to the existing baseline condition in which the width of the freeway lanes and shoulder in a critical 3.1-mi (5-km) section of the southbound (SB) freeway corridor was narrowed so that a fifth lane could be introduced; and
- An operational modification to the existing baseline condition in which one northbound (NB) lane was reversed in the same 3.1-mi (5-km) section during the peak hour so that a fifth lane could be added in the southbound direction.

Figure ES.8 summarizes travel time results taken from a subarea test network for the southbound direction along an 8.5-mi freeway study section. The analysis was performed for a peak-period condition across a 20-day time horizon. The study section consists of three segments that have four, five, and four lanes, respectively. The gray bar represents performance for baseline conditions. The black bars represent the effects of individual non-lane-widening strategies. The white bars represent three lane-widening scenarios: (1) five lanes across all three segments; (2) one additional lane across all three segments; and (3) six lanes across all three segments.

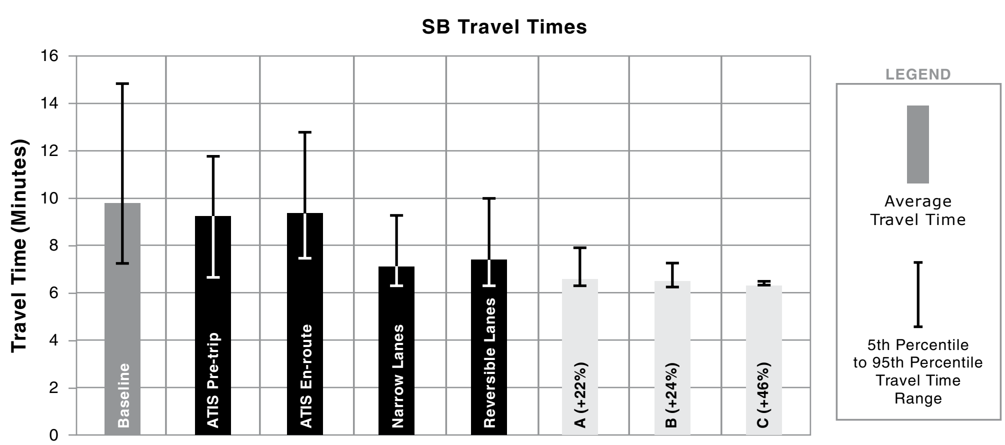


Figure ES.8. 20-day model results at corridor level.

Figure ES.8 demonstrates how trade-offs for improvement strategies can be examined in terms of their impact on average travel time (expressed as minutes of travel time) and travel time reliability (expressed as the range between the 5th and 95th percentile travel times). In some cases (e.g., the provision of pretrip information), travel time reliability associated with the tested option is significantly improved in relation to the base condition, even though the average travel time is largely unaffected. In other cases, such as the narrow-lanes strategy and each of the new construction projects, both travel time and travel time reliability are significantly improved by the tested option, although the narrow-lanes strategy may have negative safety impacts that were not considered in this analysis. Without the examination and assessment of reliability as a performance measure, a primary benefit of the strategies would go unrecognized, particularly for the non-lane-widening strategies.

These results were taken from a test network and should not be considered representative of outcomes that can be expected in other applications, because they are dependent on the particular characteristics of the network and the travel demand levels that are being modeled.

Model Portability Considerations

The enhanced DTA model described in this report can be applied in virtually any local network environment with only a few relatively straightforward adaptations:

- The stochastic capacity distribution functions for arterials and freeways should be modified to reflect local driving and car-following characteristics. This can be easily done by collecting and analyzing discharge headway distribution data at signalized intersections as well as pre-breakdown, breakdown, and post-breakdown speed-flow characteristics at freeway bottlenecks. In both cases, care should be taken to ensure that data are collected at locations not influenced by upstream or downstream intersections or bottlenecks.
- Before testing alternative operational strategies, the network structure and O-D patterns should be examined and calibrated under known existing conditions to ensure that the model adequately replicates them. This is identical to the calibration process used for many years with traditional travel demand forecasting models. Even so, it is likely that the network structure associated with a well-calibrated traditional travel demand forecasting model will need some modification before it can be used effectively by the enhanced DTA model. This is because the performance of the latter is more sensitive to certain network characteristics (e.g., the number, location, and length of centroid connectors; the type of intersection control; and the length of intersection turn lanes).
- The capacity adjustments that necessarily accompany some of the operational improvements strategies may need to be modified to better reflect local driving habits. For example, the capacity reduction that can be expected from the use of narrow lanes could differ by region or county.

Conclusions and Next Steps

This report presents an enhanced DTA model; new link, corridor, and network diagnostic tools; and an analytic methodology that can significantly improve the information available to decision makers and thus the robustness of their investment decisions. In today's environment where financial resources for new transportation construction projects are scarce and where environmental, regulatory, and policy constraints make such projects very difficult and time-consuming, it is incumbent upon both transportation professionals and decision makers to consider all viable options to new construction before making a final decision. The new analytic tools made available through SHRP 2 and reviewed in this report represent a significant new capability in this regard.

With regard to next steps, incorporating the ability to simulate the additional effects of events that cause nonrecurring congestion (e.g., crashes, other incidents, severe weather) will further enhance the usefulness and usability of these tools. In the meantime, they can still be effective in significantly improving investment decision making in the transportation industry.

References

1. Brilon, W., J. Geistefeldt, and M. Regler. Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity. *Proc., 16th International Symposium on Transportation and Traffic Theory,* College Park, Md., 2005, pp. 125–144.
2. Lorenz, M., and L. Elefteriadou. A Probabilistic Approach to Defining Freeway Capacity and Breakdown. *Transportation Research Circular E-C018: Proceedings of the Fourth International Symposium on Highway Capacity.* TRB, National Research Council, Washington, D.C., 2000, pp. 84–95.
3. Jia, A., B. M. Williams, and N. M. Rouphail. Identification and Calibration of Site-Specific Stochastic Freeway Breakdown and Queue Discharge. *Transportation Research Record: Journal of the Transportation Research Board, No. 2188,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 148–155.
4. Mahmassani, H. S. Dynamic Network Traffic Assignment and Simulation Methodology for Advanced System Management Application. *Networks and Spatial Economics,* Vol. 1, 2001, pp. 267–292.
5. Chiu, Y.-C., H. Zheng, J. A. Villalobos, W. Peacock, and R. Henk. Evaluating Regional Contra-Flow and Phased Evacuation Strategies for Central Texas Using a Large-Scale Dynamic Traffic Simulation and Assignment Approach. *Journal of Homeland Security and Emergency Management,* Vol. 5, No. 1, Article 24, 2008.
6. Zhou, X., and C. C. Lu. Traffic Flow Models in DTALite: White Paper. https://sites.google.com/site/dtalite/traffic-flow-model. Accessed July 27, 2010.
7. Simon, H. A. A Behavioral Model of Rational Choice. *Quarterly Journal of Economics,* Vol. 69, No. 1, 1955, pp. 99–118.

CHAPTER 1

Introduction

Project Background

The primary objectives of SHRP 2 Capacity Project C05 were threefold:

1. Quantify the capacity benefits, individually and in combination, of operations, design, and technology improvements at the network level for both new and existing facilities.
2. Provide transportation planners with the information and tools to analyze operational improvements as an alternative to traditional construction.
3. Develop guidelines for sustained service rates (SSRs) to be used in planning networks for limited access highways and urban arterials.

Taken together, these objectives support the development of methodologies to effectively determine the expected capacity gain from candidate operational improvements relative to the capacity gain from construction of an additional lane.

A variety of questions must be considered in efforts to address these objectives:

- Who is the audience?
- What is the range of operational strategies that should be considered?
- What performance metrics should be used?
- How can the operational effects of a particular strategy be fairly compared with the performance impacts of a traditional construction project?
- How can sustainable service flow rates be characterized and modeled?
- What tools can be used to implement the new methodologies?

Who Is the Audience?

The audience for this project is diverse in many respects. Groups that will benefit from a better understanding of the contributions of operations, technology, and design to meet highway capacity needs include the following:

- *Decision makers,* who will use the analysis results to make public investment decisions;
- *Traffic engineers and transportation planners,* who will use the methodologies developed in this project to plan and evaluate alternative operational improvement strategies;
- *Civil engineers,* who will use the methodologies to evaluate the adequacy of their highway design;
- *Researchers and educators,* who will use the methodologies to advance their research, improve their understanding of traffic phenomena, and train future transportation professionals; and
- *ITS designers,* who are interested in exploring the benefits of current and future ITS technologies.

What Improvement Strategies Should Be Considered?

The domains of operations, design, and technology all represent fertile ground for developing strategies that can effectively enhance corridor and/or network performance while also forestalling the need for new construction. The focus of this project was limited to strategies that have the potential to increase capacity or supply. Strategies to reduce or manage demand are worthy of consideration but were not explicitly incorporated into this project work.

What Performance Metrics Should Be Used?

The traffic performance measures that are particularly effective for a point analysis—for example, the computed volume/capacity ratio or level of service at an intersection—are often not especially meaningful in the context of a corridor, subarea, or network. Thus, it is desirable and helpful to consider multiple performance measures to assess capacity gains at both the local and network levels. A number of measures were considered and are supported by the analysis methodologies that were developed.

How Can the Capacity-Enhancing Abilities of the Improvement Strategies Be Characterized?

Figure 1.1 conceptually illustrates a way to use the analysis methodologies developed in this project to define the relationship between lane miles added onto a network in the case of a construction project and the effective lane miles added by a nonconstruction improvement strategy. As Figure 1.1 illustrates, a set of nonconstruction improvements that result in reducing the network travel time from, say, A to B, is equivalent to the addition of D minus C lane miles, increasing the network capacity by (D – C)/C%.

How Can Sustainable Service Flow Rates Be Characterized and Modeled?

Recent research indicates that highway capacity shows all properties of a random variable (*1–5*). The capacity of a highway facility is the result of driver behavior and therefore varies with driver population (i.e., by types of vehicles, motivation or trip purposes, experiences, familiarity with the freeway section or with the traffic operation at the specific time). In addition, highway capacity is a matter of systematic variability (e.g., due to accidents, incidents, weather, and work zones) that is the reason for most of the congestion delay on freeways. The amount of this delay increases with increasing demand. However, only in a limited part of the network does demand exceed the normal capacity of the infrastructure so that it becomes the main contributor to delay.

Modeling the stochastic nature of capacity on both freeway and arterial networks is an important advancement in improving the reliability of travel demand forecasts and operational analyses.

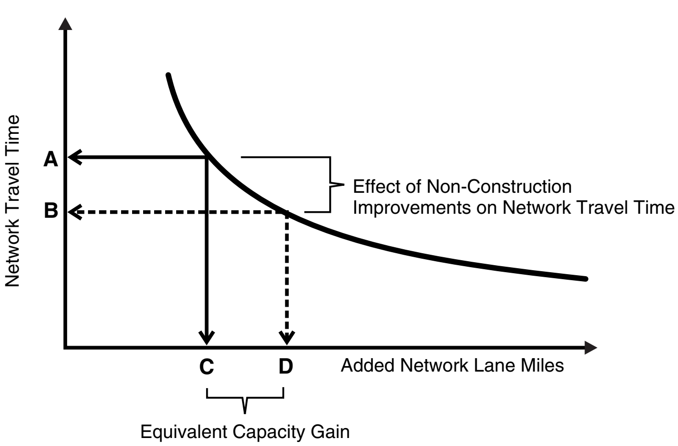


Figure 1.1. Equivalent capacity gain concept.

What Tools Can Be Used to Implement the New Methodologies?

The primary product of this project is new or enhanced analysis methodologies. Even so, their practical applicability in a real-world environment depends on the ability to implement them in useful and usable tools. Therefore, an important question at the start of this project was what assignment/simulation tools can be considered for this purpose? Several options were considered by using a decision support approach.

One simple theoretical approach to quantifying network-wide capacity would be to solve a max-flow min-cut optimization problem, but this method cannot account for the complex traffic-flow dynamics in a time-varying and complex traffic network.

Dynamic traffic assignment (DTA) modeling tools were targeted because of their unique ability to evaluate network performance under time-varying demand and supply conditions created by various operations-based, design-based, and technology based strategies. A range of network analysis tools are available, and all were initially considered for use in this project: DYNASMART-P (Dynamic Network Assignment Simulation Model for Advanced Roadway Telematics: Planning version), DynusT, DTALite, DYNAMIT, IDAS, Integration, VisSim, Paramics, SCRITS, and EMME/3.

For a number of pragmatic reasons unique to this project, DYNASMART-P was selected and used as the platform for developing and testing the improved analysis methodologies described in this report. However, there is no obvious reason preventing the incorporation of these methodologies into any of the other analysis tools identified, and in fact this has already happened with at least one of the tools (DTALite).

References

1. Brilon, W., J. Geistefeldt, and M. Regler. Randomness of Capacity—Idea and Application. *Proc., 5th International Symposium on Highway Capacity and Quality of Service,* Yokohama, Japan, Vol. 1, Japan Society of Traffic Engineers, Tokyo, 2006, pp. 147–157.
2. Brilon, W., J. Geistefeldt, and H. Zurlinden. Implementing the Concept of Reliability for Highway Capacity Analysis. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
3. Brilon, W., J. Geistefeldt, and M. Regler. Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity. *Proc., 16th International Symposium on Transportation and Traffic Theory,* College Park, Md., 2005, pp. 125–144.
4. Brilon, W., and M. Ponzlet. Variability of Speed-Flow Relationships on German Autobahns. *Transportation Research Record 1555,* TRB, National Research Council, Washington, D.C., 1996, pp. 91–98.
5. Elefteriadou, L., F. L. Hall, W. Brilon, R. P. Roess, and M. G. Romana. Revisiting the Definition and Measurement of Capacity. *Proc., 5th International Symposium on Highway Capacity and Quality of Service,* Yokohama, Japan, Vol. 2, Japan Society of Traffic Engineers, Tokyo, 2006, pp. 391–399.

CHAPTER 2

Analytic Enhancements to Enable Network Assessment of Strategies

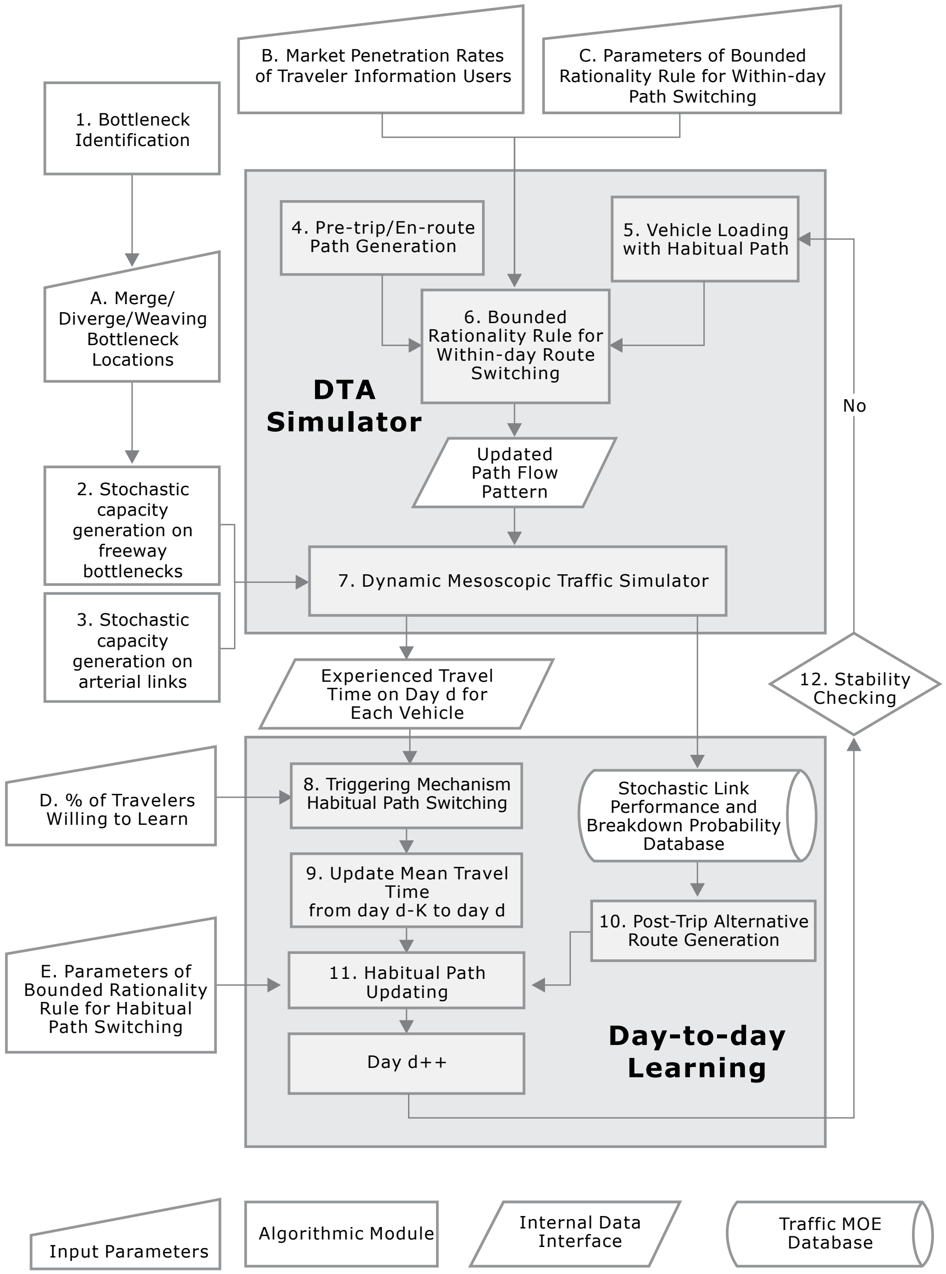
The four major methodological improvements described in this chapter dramatically increase the realism and sensitivity of existing dynamic traffic assignment tools. These improvements have been incorporated into two dynamic traffic assignment (DTA) models, namely, DYNASMART-P and DTALite, for purposes of providing transportation planners with the tools to analyze operational improvements as an alternative to traditional construction. Moreover, these improvements could be easily incorporated into other traffic simulation/analysis models as well.

Summary of Key Findings and Conclusions

- Speed and density thresholds, when used in conjunction with one another, are effective in identifying freeway breakdown conditions. However, local calibration is necessary for different study sites.
- Parameters for the average pre-breakdown headway probability distribution function vary by study site and require local calibration. The lognormal distribution appears to well characterize the pre-breakdown headway process.
- The queue discharge rate series are stochastic and strongly time correlated.
- *Highway Capacity Manual* (*HCM*) levels are more representative of the upper tail of the breakdown and queue discharge observations.
- Many commonly used network flow modeling platforms likely underrepresent the frequency, duration, and overall traffic operational effects of freeway system bottlenecks.
- Modeling results are substantially improved when the effects of demand blockages caused by short turn pockets on downstream saturation flow rates are taken into account.
- Modeling results are also improved by recognizing that the saturation flow rate on any signalized intersection approach is not a constant number but varies from cycle to cycle according to individual driver characteristics. Capturing this variability yields more realistic outcomes, especially when the modeling is performed over a number of simulated days.
- A stand-alone, time-dependent spreadsheet-based simulation tool (TPAST) developed as a supplemental product of this research effort can provide good insights into link-specific operational issues resulting from changes in the physical, traffic, and signal timing characteristics associated with the link, particularly those associated with short turn bays.
- Methodological enhancements are provided to seamlessly incorporate stochastic capacity models at freeway bottlenecks and signalized intersections in a mesoscopic traffic flow simulation environment.
- A systematic day-to-day learning and route choice framework has been developed enabling a realistic representation of drivers' route selection process under the travel time variability introduced by random capacity variations. A key related finding is that overall network travel time stabilizes after several iterations of the application of the learning process. The number of days over which this stabilization occurs is network dependent.
- Link, corridor, origin–destination (O-D), and network-level stochastic performance measures are identified and are consistent with traffic flow theory principles.

Overall Modeling Framework

The system evolution-modeling simulation framework for the network analysis tool (i.e., DTA simulator in this study) is illustrated in Figure 2.1. In the proposed modeling framework, static demand (i.e., the same number of vehicles with fixed departure times) is simulated over different days. Many of the proposed enhancements are imbedded in parts of the framework, as described in the following sections.



Note: DTA = dynamic traffic assignment; MOE = measure of effectiveness.

Figure 2.1. Comprehensive conceptual simulation framework.

The following three critical inputs (illustrated in Figure 2.1) should be prespecified by users:

- Time-dependent traffic demand;
- Bottleneck locations;
- Percentages of unequipped, pretrip, and en route users; and
- Parameters of the bounded rationality rule.

Figure 2.1 demonstrates the entire conceptual simulation framework implemented in DYNASMART-P. Two key components, stochastic capacity generation on freeway bottlenecks and signalized arterial and route choice mechanism, are not illustrated in detail; because of their significance, they are discussed in depth in the following sections of this chapter. There are four major enhancements:

1. Stochastic capacity for freeway bottlenecks and capacity allocation at freeway merge areas;
2. Stochastic capacity and turn pocket analysis on arterials;
3. Implementation of a day-to-day learning paradigm; and
4. New performance measures and implementation.

Enhancement 1: Stochastic Capacity for Freeway Bottlenecks

In this section, the stochastic nature of freeway breakdown and queue discharge is discussed through a comprehensive analysis of sensor data collected at bottleneck sites in the San Francisco Bay Area, California, and San Antonio, Texas. A new procedure is proposed to define the stochasticity of freeway breakdown and queue discharge based on time-indexed data of speed-flow profiles (*1*).

Stochastic Breakdown Characteristics

In the 2000 edition of the *Highway Capacity Manual,* freeway capacity is defined as "the maximum hourly rate at which vehicles reasonably can be expected to traverse a point or a uniform section of a roadway during a given time period under prevailing roadway, traffic, and control conditions" (*2*). In keeping with the definition of conventional freeway capacity, it is widely accepted by most traffic analysts that the facility will experience breakdown (i.e., a transition from an uncongested state to a congested state) only if the traffic demand exceeds a specified capacity value. Therefore, when freeway capacity is taken to be a constant value, breakdown is treated as a deterministic phenomenon. However, an emerging body of research (*3–8*) indicates that traffic flow rate during the time intervals preceding observed instances of freeway breakdown (called pre-breakdown flow rate in this study) is better represented as a random variable than a fixed value. Zurlinden (*9*) and Brilon (*10*) developed a methodology to derive roadway pre-breakdown distribution functions for the purpose of implementing this stochastic capacity concept. In the most recent study, Brilon (*11*) suggested the Weibull distribution for characterizing stochastic capacity based on traffic data from Germany. Based on the probabilistic nature of freeway capacity, Dong and Mahmassani (*12*) first illustrated the significant effects of the stochastic concept on the study of travel time reliability.

The conventional definition of fixed capacity for uninterrupted flow facilities has the practical drawback of underrepresenting the frequency, duration, and therefore traffic operational impact of freeway system bottlenecks. In order to quantify accurately the capacity benefits of operations, design, and technology improvements, it is necessary to develop realistic, implementable capacity models for freeway operations. The following sections provide details of the procedure followed to develop the stochastic pre-breakdown headway distribution for freeway bottlenecks implemented in DYNASMART-P.

Identifying Freeway Bottlenecks

As a basis for developing and validating stochastic freeway bottleneck models, the most common freeway bottleneck feature, on-ramp junctions, was selected for detailed study.

Data were assembled from the TransGuide system (*13*) in San Antonio and from PeMS data (*14*) archived for the San Francisco Bay Area (Caltrans District 4). Data for both locations are readily available from online databases.

The TransGuide database provides traffic volume, speed, and occupancy data gathered from the initial 26 miles of instrumented highways within the Texas Department of Transportation TransGuide project. The extracted data set for this study is the daily raw data (20-second intervals) from 01/01/2007 to 09/30/2008. The TransGuide data include a small number of missing observations which required removal of a small portion of data from the aggregated set. The Bay Area data used in this study are processed traffic data, which include volume, speed, and occupancy. The data cover the period from 01/01/2007 to 09/30/2008 and are aggregated at 5-minute intervals. Since the Bay Area data are already preprocessed and aggregated, the missing observations have been estimated in the data set.

Both TransGuide and PeMS databases provide detailed location information for each sensor on the freeway system. As discussed below, these data were important for selecting appropriate bottleneck locations for this study. The information from the TransGuide and PeMS databases is summarized in Appendix A.

The PeMS system identifies active bottleneck locations for the Bay Area; those that experience congestion for more than 10 days per month were selected for study. In contrast, the TransGuide database does not identify active bottlenecks in the freeway network. Therefore, geometric bottlenecks were identified through visual inspection by using Google Maps of the areas covered by TransGuide detectors. To control for possible confounding operational effects and to isolate ramp merge bottleneck effects, a systematic process was developed for study site selection. Suitable bottleneck sites met the following three criteria:

1. Sufficient distance between the on-ramp and the nearest downstream bottleneck. The longer the distance to a potential downstream bottleneck (e.g., on-ramp or off-ramp), the greater the likelihood that the data are not confounded by the presence of downstream queues regularly spilling back to the bottleneck location under consideration.
2. Suitably placed sensor data. An ideal detector is just downstream of the corresponding bottleneck.
3. Presence of traffic demand high enough to yield regular freeway breakdown. This criterion ensures an adequate sample size.

Because the candidate sites in the Bay Area are active bottlenecks, the third criterion was not a significant factor for the PeMS data. In the San Antonio area, however, traffic data were retrieved from the TransGuide system to evaluate the third criterion for sites that met the first and second criteria.

Table 2.1. Basic Information about the Study Sites

Site Number	Highway	Direction	Number of Lanes	Distance to Downstream Bottleneck (km)	Distance to Detector (km)
1	I-880 (BA)	S	4	3.0	0.1
2	I-680 (BA)	S	3	2.6	0.1
3	I-280 (BA)	N	4	NA	0.36
4	I-580 (BA)	W	4	NA	0.25
5	I-680 (BA)	N	4	6.1	0.12
6	I-35 (SAT)	N	3	3.6	0.18
7	I-35 (SAT)	S	3	2.2	0.38

Note: NA indicates that the distance to downstream bottleneck is very long (more than 10 km). BA = Bay Area; SAT = San Antonio.

Based on the three criteria, seven on-ramp bottlenecks (three with three travel lanes and four with four travel lanes) were selected as the study sample. Two of the sites are in the San Antonio area, and the remaining five are in the Bay Area. The basic information about each site is summarized in Table 2.1.

As mentioned above, the data in the Bay Area are aggregated into 5-minute intervals and the San Antonio data into 20-second intervals. Consistent with *HCM* 2000, both data sets were aggregated into 15-minute intervals prior to analysis.

Method for Implementing Stochastic Capacity

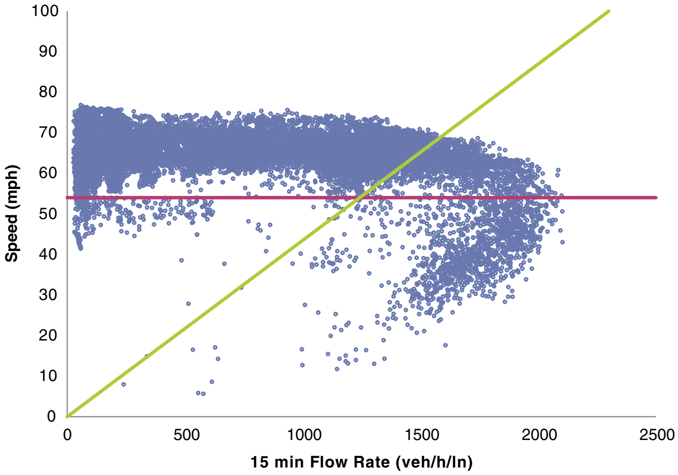
A critical step in developing pre-breakdown headway distributions is determining when breakdown occurred on the basis of a review of historical speed-flow data. The published studies specified a critical speed or a critical speed drop to define the freeway breakdown. However, while such a speed-based threshold can be well defined in reference to representative traffic flow characteristics, its value will vary by location. For example, extensive analysis of data from Los Angeles freeways suggests that free-flow speed is around 60 mph and that at breakdown the operating speed rapidly drops below 40 mph (*15*). Researchers in Los Angeles therefore specified a minimum speed differential of 20 mph and a speed threshold of 40 mph to identify a breakdown event by using 5-minute data. Using data obtained from Toronto freeways, on the other hand, Elefteriadou (*16*) defined that breakdown occurred "when the average speed of all lanes on the freeway dropped below 90 km/hr (56 mph) for a period of at least five minutes." These examples demonstrate that breakdown events are unique to local conditions and that use of a specific, universal speed threshold is not advisable. Freeway breakdown events should be defined on the basis of speed thresholds extracted from local speed-flow observations. Moreover, as discussed later in this chapter, it is problematic to use speed as the only criterion to define breakdown.

Breakdown determination is the critical starting point for both stochastic capacity and queue discharge studies. As mentioned earlier, although most previous studies used speed as the threshold, a single speed threshold was not considered appropriate for determining congested conditions based on the speed-flow relationships observed in field data. Figure 2.2 shows 21 months of 15-minute freeway data from I-880 in the Bay Area. The horizontal line superimposed on the plot indicates the speed boundary that was used to isolate congested conditions. As is apparent from the graph, a single speed threshold is not sufficient for determining congested conditions. Observed conditions exhibiting a flow rate lower than 1,000 vehicles per hour per lane but with speeds higher than 40 mph were considered to be reflective of anomalous free-flow conditions rather than congested conditions. The data pattern shown in Figure 2.2 is typical of the seven on-ramp sites, and the presence of low-flow observations below the critical speed threshold creates the need for a robust phase boundary for defining congested conditions.

A combination speed and density threshold (diagonal line in Figure 2.2) was applied to identify congested conditions, thereby avoiding the inclusion of anomalous low-speed data. Traffic states are considered to represent congested conditions only when

- The observed speed is below the critical speed.
- The observed density is greater than or equal to the boundary between levels of service C and D (LOS C/D).

As mentioned above, the critical speed and the density at the LOS C/D boundary [e.g., in *HCM* 2000, 26 passenger cars per mile per lane (pc/mi/lane) is the boundary that separates LOS C from LOS D] are locally calibrated for each



Note: veh/h/ln = vehicles per hour per lane.

Figure 2.2. I-880 speed-flow data.

of the specific study sites. The procedure for calculating these site-specific thresholds is described in the following paragraphs.

First, 15-minute flow rate values in the top 1 percentile tail are identified. The average of this sample of near maximum flows was generally equivalent to the traditional capacity defined in the HCM. The density for each 15-minute observation is then calculated as shown by Equation 2.1:

$$k = \frac{q}{\bar{\mu}} \tag{2.1}$$

where

k = density for each 15-minute observation (veh/mi/lane),
q = 15-minute flow rate for the top 1 percentile flows (veh/h/lane), and
$\bar{\mu}$ = space mean speed (mph)

The critical speed is then calculated as shown in Equation 2.2:

$$\bar{\mu}_{critical} = \frac{\sum q}{\sum k} \tag{2.2}$$

where q = 15-minute flow rates in the top one percentile flows and k = 15-minute density values corresponding to the top 1 percentile flows.

The equivalent density at capacity based on *HCM* definition is calculated by Equation 2.3:

$$k_{capacity} = \frac{1}{n}\sum k \tag{2.3}$$

where n = number of 15-minute observations in the top 1 percentile flow region.

Finally, the adjusted *HCM*-based critical density threshold (LOS C/D boundary) is calculated by Equation 2.4:

$$k_{critical} = \frac{26\left(k_{capacity}\right)}{45} \tag{2.4}$$

where 26 pc/mi/lane represents the maximum density per lane passenger car equivalent density for LOS C for basic freeway segments per *HCM* 2000 and 45 pc/mi/lane represents the corresponding per lane passenger car density at capacity.

Equation 2.4 provides adjusted values for the LOS C/D thresholds based on the observed density at capacity. In summary, traffic conditions observations are identified as representing congested flow when the observed 15-minute speed is less than the critical speed and the observed 15-minute density is greater than the critical density.

Each study site was analyzed independently. Speed and vehicle count data were summarized in 15-minute intervals across all lanes. The vehicle count data were then expressed as equivalent hourly flow rates per lane. The traffic parameters for each site are summarized in Table 2.2.

Breakdown from Free-Flow Conditions

In deterministic traffic models, such as that of the *HCM*, there are two basic traffic states in uninterrupted freeway operation: uncongested and congested flow. In defining stochastic capacity, the focus lies on the pre-breakdown state (i.e., the uncongested states just preceding the breakdown state). The breakdown state is the first in what may be a series of congested state observations. Once the breakdown states were identified, all the corresponding pre-breakdown states were selected from each data set. For purposes of implementation in the mesoscopic model, DYNASMART-P, the pre-breakdown flow rates were first converted into passenger car

Table 2.2. Calibrated Traffic Parameters for Each Study Site

Site Number	Highway	Average Top 1 Percentile Flow Rate (veh/h/lane)	Critical Speed (mph)	Density (C/D) (veh/mi/lane)
1	I-880 (BA)	2,052	56	21
2	I-680 (BA)	2,093	53	23
3	I-280 (BA)	2,183	53	24
4	I-580 (BA)	1,982	49	23
5	I-680 (BA)	2,127	54	23
6	I-35 (SAT)	1,992	47	23
7	I-35 (SAT)	2,172	63	20

equivalent flows and then aggregated into 15-minute pre-breakdown headways (i.e., 3,600/flow rate). Heavy vehicle count data were not available for the sites in this study. Therefore, the *HCM* 2000 default of 5% heavy vehicles and the passenger car equivalent for trucks and buses of 1.5 for a level general segment was used to convert the TransGuide and PeMS data to passenger car equivalent flow rates.

It is desirable to exclude pre-breakdown flow rate under nonrecurring conditions as much as possible. However, in the absence of incident logs, a statistical approach was applied to exclude outlying pre-breakdown flow rates. As shown in Equation 2.5, a pre-breakdown flow rate is identified as an outlier if:

$$q < Q_{0.25} - 1.5IQR \text{ or } q > Q_{0.75} + 1.5IQR \tag{2.5}$$

where

$Q_{0.75}$ = 75th percentile flow rate (pc/h/lane),
$Q_{0.25}$ = 25th percentile flow rate (pc/h/lane), and
$IQR = Q_{0.75} - Q_{0.25}$.

The speed-flow diagram in Figure 2.3 shows the pre-breakdown and outlier observations for one study site (I-880). Almost all flow rates below the *HCM* equivalent LOS C/D density boundary were identified as outliers. For the 15-minute aggregated traffic data, it is reasonable that the pre-breakdown flow rates would not occur at LOS C or better under the prevailing roadway, traffic, and control conditions.

Statistical tests were conducted to determine the probability distributions that reflect the stochastic characteristics of freeway pre-breakdown headway. The most common tests for goodness of fit are the Kolmogorov–Smirnov (K-S) test, chi-square test, and Anderson–Darling (A-D) test. All are used to decide whether a data sample belongs to a population with a specific distribution. A chi-square test could be applied to test any univariate distribution; however, the values of the chi-square statistic are quite sensitive to how the data are binned (*17*). The Anderson–Darling statistic is a measure of how far the data points lie from the fitted distribution. However, the A-D test is not a distribution-free test. The critical values for the A-D test must be calculated for each distribution, and they are only available for a very limited number of distributions (*18*). The K-S statistic also quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution. The K-S test is distribution-free in the sense that it makes no assumption about the underlying distribution of data (*19*). Another advantage of the K-S test is that it is an exact test, while the chi-square goodness-of-fit test depends on an adequate sample size for the approximations to be valid (*17*). Therefore, the K-S test was applied here to examine which distribution functions provide the best fit to the pre-breakdown headways.

The shifted lognormal, normal, exponential, Weibull, and gamma distributions were evaluated for fit with the field data. For each of these, a K-S statistic was calculated. As indicated in Table 2.3, the shifted lognormal distribution yields the lowest K-S statistic values across all study sites.

Figure 2.4 illustrates a sample pre-breakdown headway distribution for one study site on I-880 in the Bay Area. As shown in the figure, a single headway value is not appropriate for defining breakdown on freeways. The trend illustrated

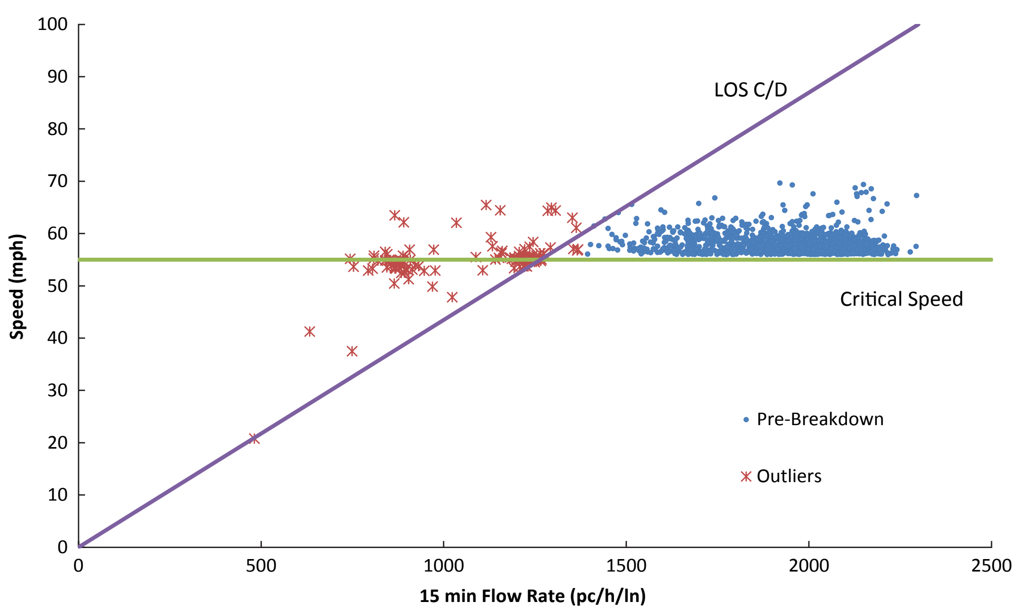


Figure 2.3. Pre-breakdown flows and outliers for I-880.

Table 2.3. Computed Kolmogorov–Smirnov (K-S) Statistic Values by Distribution

Tested Distribution	Site Number						
	1	2	3	4	5	6	7
Shifted lognormal	0.039	0.074	0.032	0.066	0.022	0.094	0.058
Normal	0.140	0.132	0.128	0.168	0.072	0.226	0.120
Exponential	0.239	0.333	0.226	0.227	0.296	0.189	0.133
Weibull	0.385	0.410	0.380	0.390	0.230	0.310	0.170
Gamma	0.128	0.122	0.119	0.156	0.063	0.209	0.106

also indicates that the slope continually decreases with increasing values of time headway (i.e., decreasing flow rate). This trend is consistent with findings from previous studies (*7, 11*) showing that the probability of breakdown increases with increasing flow rate. The figure also gives the corresponding 15th, 50th, and 85th percentile flow rates derived from the distribution. For example, if capacity is defined as a 15-minute flow rate that is sustainable 85% of the time, the corresponding capacity value could be 1,778 pc/h/lane.

Similar to the site-specific process for identifying breakdown observations, a local pre-breakdown headway distribution was estimated independently for each on-ramp site. The distribution parameters for the seven sites are summarized in Table 2.4.

Although the distribution parameters varied among study sites, it was found that the pre-breakdown headways of all seven sites are best modeled as a shifted lognormal random variable. The average shift at the seven study sites is 1.519 seconds (standard deviation of 0.053 seconds), which is equivalent to 2,370 pc/h/lane. This value is very close to value of predicted *HCM* 2000 capacity. The mean pre-breakdown headway at the seven study sites is 1.886 seconds (standard deviation of 0.094 seconds), which is equivalent to 1,909 pc/h/lane. The scale parameters vary among different sites and may depend on the specific characteristics of the analyzed freeway section, local differences in driver behavior, prevailing weather conditions, and other factors.

In order to implement the stochastic capacity in DYNASMART-P and then test the effects of stochastic freeway capacity on SSRs and network performance, a single shifted lognormal distribution was proposed by aggregating all the pre-breakdown headway observations in the above

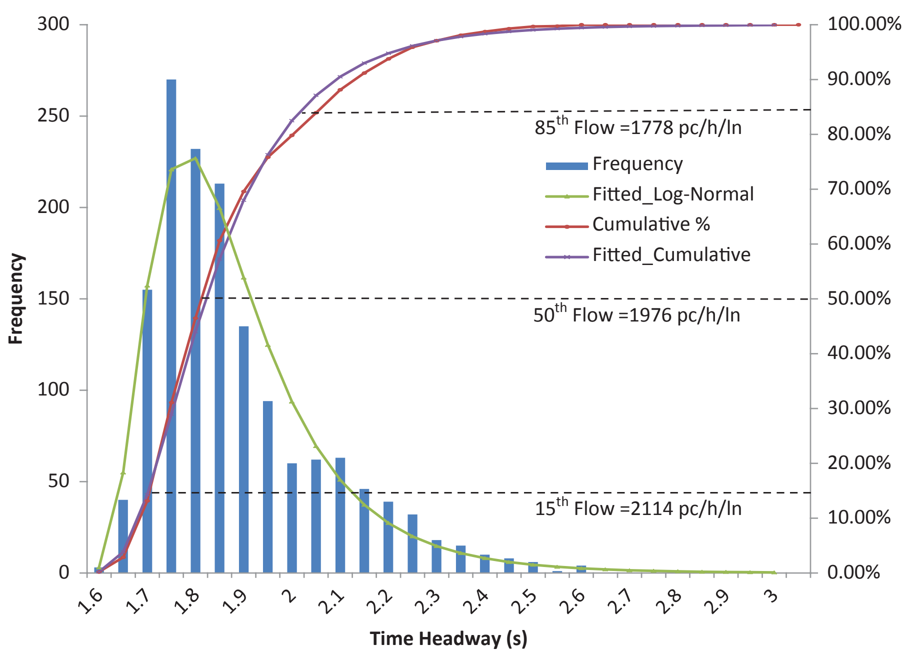


Figure 2.4. Pre-breakdown headway distribution for I-880.

Table 2.4. Summary of Pre-breakdown Headway Distribution Parameters

Site	Area	Lognormal Parameter			Maximum Pre-breakdown Flow (pc/h/lane)	Mean Pre-breakdown Flow (pc/h/lane)
		Shift (s)	μ	σ		
1	I-880 (BA)	1.536	–1.255	0.520	2,343	1,933
2	I-680 (BA)	1.486	–1.139	0.288	2,422	1,978
3	I-280 (BA)	1.626	–1.297	0.516	2,214	1,857
4	I-580 (BA)	1.489	–1.537	0.488	2,418	2,079
5	I-680 (BA)	1.527	–0.730	0.255	2,358	1,778
6	I-35 (SAT)	1.499	–1.147	0.689	2,371	1,893
7	I-35 (SAT)	1.470	–1.021	0.678	2,449	1,871

seven study sites. The corresponding parameters are shift = 1.5 seconds, $\mu = -0.97$, and $\sigma = 0.68$. It should be noted that if the point of interest is a single on-ramp bottleneck, as stated before, local calibration is recommended to develop the average pre-breakdown headway distribution model. For a relatively large network, however, a single set of parameters for the pre-breakdown model developed from the data available is efficient for model implementation and should be accurate enough for the network-level analysis.

Recovery from Breakdown Conditions

Similar to the conventional definition of capacity in the *HCM,* the queue discharge flow rate is also typically characterized in a deterministic manner. In other words, after a breakdown occurs, the queue will discharge at a constant flow rate. Based on field data, Lorenz and Elefteriadou (*7*) have clearly demonstrated that the queue discharge flow rate is also stochastic in nature. In a recent study, Dong and Mahmassani (*20*) suggested a linear relationship between queue discharge rate and the pre-breakdown flow rate. However, the studies for the queue discharge behaviors, especially the stochastic characteristics, are quite limited. In this section, queue discharge behaviors under the stochastic capacity scenario were discussed.

Considering the stochastic nature of freeway capacity, it is quite possible that the queue discharge flow rate is correlated with the stochastic pre-breakdown flow rate preceding the queue existence. In addition, the 15-minute field data have shown that the queue could be discharged over multiple time intervals. By examining the observational data, it was found that the queue discharge rate could be updated with stochastic time-correlated recursions. A simple first-order autoregressive model is proposed as follows in Equation 2.6:

$$C_t = \alpha C_{t-1} + \mu + \varepsilon_t \, (t \geq 1) \tag{2.6}$$

where

C_t = queue discharge rate at time interval t in pc/h/lane,
α = coefficient,
μ = intercept, and
$\varepsilon_t \sim N(0, \sigma^2)$ = random error.

When $t = 1$, $C0$ is the pre-breakdown flow rate. If we assume $\beta = 1 - \alpha$, then the first two terms of the model above can be rewritten as shown in Equation 2.7:

$$C_t = C_{t-1} + \beta(\mu_c - C_{t-1}) \, (t \geq 1) \tag{2.7}$$

where β = a linear parameter that models the strength of regression to the mean and μ_c = the average discharge rate in pc/h/lane.

For the DYNASMART-P model implementation, a random error term is added based on the error distribution of the fitted model. As shown in Equation 2.8, stochastic innovation term $\varepsilon_t \sim N(0, \sigma^2)$ is proposed. Therefore, the recursive model to update the queue discharge rate (per lane) is

$$C_t = C_{t-1} + \beta(\mu_c - C_{t-1}) + \varepsilon_t \, (t \geq 1) \tag{2.8}$$

The traffic data from the study site on I-880 in the Bay Area were used to fit the proposed queue discharge model. The fitting procedure began with the series only having two time intervals (i.e., pre-breakdown time interval and queue discharge interval). The duration of the queue discharge interval begins with one time interval and then is continuously extended by a time interval if queues are still present at the end of the time interval. Table A.3 in Appendix A shows fitted parameters for the various cumulative queue duration lengths. The results indicate that there is a statistically significant relationship between C_t and C_{t-1}. The relatively high R^2 value also suggests that the proposed model matches the data very well. At least 73% (the minimum R^2 value in Appendix A) of the variation in the response variable C_t, can be explained by the proposed model. Moreover, by using an average discharge rate of 1,850 pc/h/lane and β as 0.2 for all the queue duration lengths, a simple overall queue discharge model could be generated with few impacts on the goodness of fit of each subgroup. Based on the model fitting results, a stochastic innovation term $\varepsilon_t \sim N(0, \sigma^2)$ is also proposed with $\sigma = 100$ pc/h/lane. Figure 2.5 illustrates the proposed simplified recursive queue discharge model.

In summary, the queue discharge model is based on all breakdown flow observations with three primary characteristics:

- The queue discharge rate series are strongly time correlated.
- The queue discharge rates have a stochastic, random innovation component.
- The queue discharge rates converge to the mean discharge rate for breakdowns that are initiated with stochastic capacities that are in the tails of the pre-breakdown headway distribution.

Supplemental Freeway Capacity Enhancement: Capacity and Queue Allocation at Merge Points

When two traffic streams merge and the sum of their demand is greater than the capacity of the downstream roadway, traffic queues will be generated on the upstream links. How and where queues grow and dissipate at the merge point fully depends on how the available downstream capacity is allocated to the entering traffic streams. Following a comprehensive literature review, the research team developed a series of algorithms to allocate capacity at merge points. Basically, the proposed capacity allocation algorithms focused on two scenarios: freeway-to-freeway merge and on-ramp-to-freeway merge.

Freeway-to-Freeway Merge

The proposed allocation algorithm for freeway-to-freeway merge is relatively simple: the allocated capacities for the entering freeway streams are proportional to the incoming upstream flow rates. For example, if the capacity of the downstream freeway link is 3,600 passenger cars per hour (pcph) and the entering flow rates of the two upstream freeway links are 2,800 pcph (Freeway 1) and 1,400 pcph (Freeway 2), the allocated capacity for Freeway 1 is 2,400 pcph (i.e., 3,600 • 2/3) and for Freeway 2 is 1,200 pcph (i.e., 3,600 • 1/3). Therefore, the queue growth rates for Freeway 1 and Freeway 2 would be 400 pcph and 200 pcph, respectively, assuming no changes in demand.

On-Ramp-to-Freeway Merge

Compared with the freeway-to-freeway merge, the proposed allocation algorithm for on-ramp-to-freeway merge is relatively complex, because the freeway mainline flow has higher priority than that of the on-ramp. Therefore, the basic concept is to apply the capacity allocation according to the relative demand distribution between the two entering traffic streams. Based on the possible demand combinations, five demand regions have been defined. The capacity allocation process varies by region. As shown in Figure 2.6, the *x*-axis

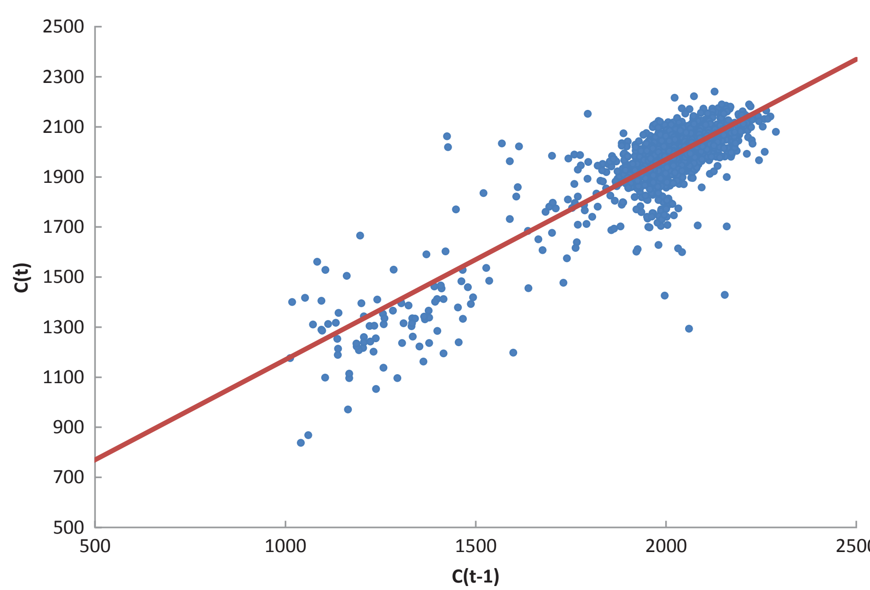


Figure 2.5. Simplified recursive queue discharge model.

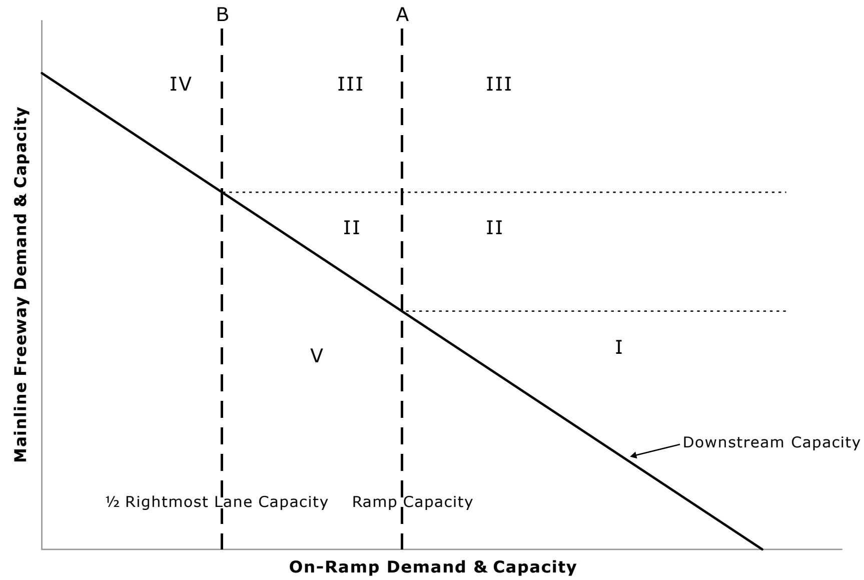


Figure 2.6. Demand regions for on-ramp-to-freeway merge.

represents the on-ramp demand and allocated capacity, while the y-axis represents the total freeway mainline demand and allocated capacity. The diagonal line represents the merge area—or downstream freeway—capacity (no served flows can be above that line); the two vertical lines represent the ramp roadway capacity (Line A), and 50% of the freeway rightmost lane capacity (Line B). All the dashed lines represent the region boundaries. Regions I through V depict the areas with various combinations of ramp and freeway mainline demands. It should be noted that Region V is below the diagonal line, which represents demand flows that are below the downstream capacity and therefore there is no effective queuing on either stream in this region and both ramp and mainline demand are fully served.

Regions I–IV are defined on the basis of relative demand distribution as follows. In all the following cases, demand exceeds that of Region V, that is

$D_F + D_R >$ Downstream Capacity

Region I: $D_F < C_{Ramp}$;
Region II: $C_{Ramp} < D_F < C - 0.5C_{RM}$;
Region III: $C - 0.5C_{RM} < D_F$ & $D_R > 0.5C_{RM}$;
Region IV: $C - 0.5C_{RM} < D_F$ & $D_R < 0.5C_{RM}$.

where

D_F = freeway demand (vph),
D_R = on-ramp demand (vph),
C = capacity at merge point (vph),
C_{Ramp} = on-ramp roadway capacity (vph), and
C_{RM} = freeway rightmost lane capacity (vph).

The following sections explain the basis for allocation of available downstream capacity in each region.

Region I

In this region, the freeway mainline demand can be fully served with the available downstream capacity. However, the on-ramp demand is greater than the on-ramp capacity, and therefore, the actual entering on-ramp flow rate at the merge point is at the on-ramp capacity value. In Figure 2.7, the point (D_F, D_R) represents a certain demand combination and the horizontal line demonstrates how the capacity at the merge point is allocated to both the freeway and on-ramp. C_F and C_R represent the allocated capacities for the freeway and on-ramp, respectively. In this region, no queue will be observed on either the freeway or the on-ramp links. Since the on-ramp demand is greater than the on-ramp capacity, queuing is expected to occur upstream of the on-ramp. The capacity at the merge point is also not fully used (Point C is below the downstream capacity line).

Region II

In Region II, the capacity of merge point is fully used and is allocated to the two incoming streams as shown in Figure 2.8.

Therefore, in this region, there is no queuing on the freeway mainline and queues occur exclusively on the on-ramp link. The rate of queuing at the merge point will be $C_{Ramp} - C_R$

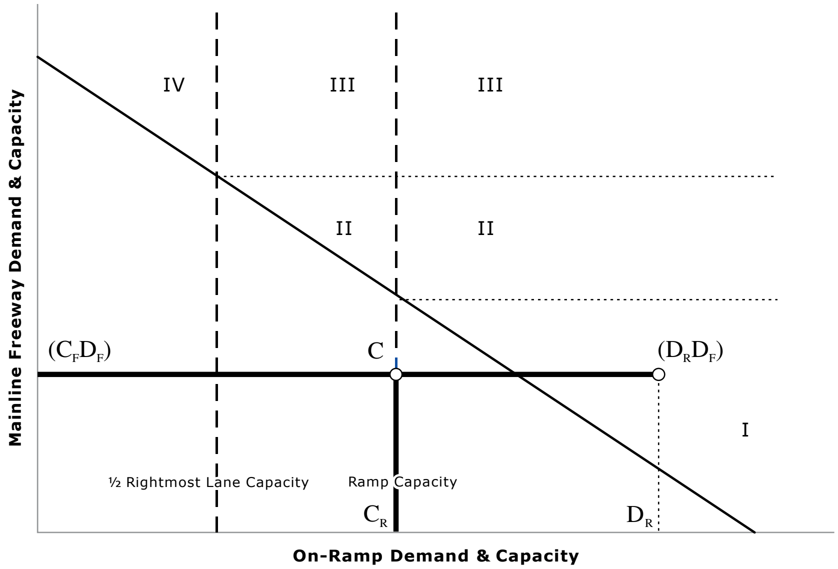


Figure 2.7. Capacity allocation for demand in Region I.

vehicles per hour and upstream of the on-ramp roadway at a rate of $D_R - C_{Ramp}$ vehicles per hour.

Region III

In Region III, the ramp demand exceeds one half the capacity of the freeway mainline, as shown in Figure 2.9. In this region, a single capacity allocation point ($C_R - C_F$) is recommended to allocate the available downstream capacity. Therefore, for all demand combinations in this region, the merge capacity allocated to the on-ramp traffic is exactly half of the freeway rightmost lane capacity (i.e., $0.5C_{RM}$), and freeway traffic will consume the remaining downstream capacity (i.e., $C - 0.5C_{RM}$). It should be noted that in Region III, queues will occur on both the freeway mainline and the on-ramp. The rate of queuing on the ramp is at a rate of $D_R - 0.5C_{RM}$ vehicles per hour, if $D_R < C_{Ramp}$, or at rate of $C_{Ramp} - 0.5C_{RM}$ vehicles per hour, if $D_R < C_{Ramp}$; and on the freeway mainline at the rate of $D_F - C_F$ vehicles per hour.

Region IV

In Region IV, the on-ramp demand is relatively low. The capacity at the merge point will be allocated as shown in Figure 2.10. Therefore, in this region, there is no queuing on the on-ramp and queues are exclusively allocated to the freeway mainline, and will occur at the rate of $D_F - C_F$ vehicles per hour.

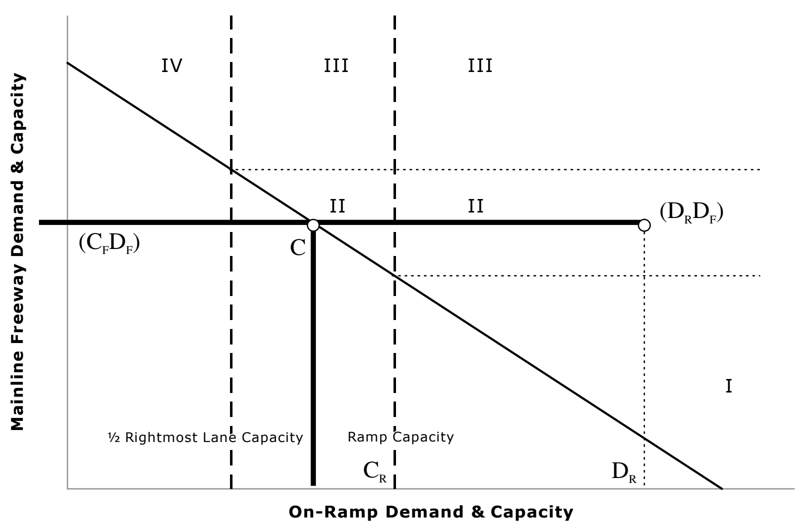


Figure 2.8. Capacity allocation for demand in Region II.

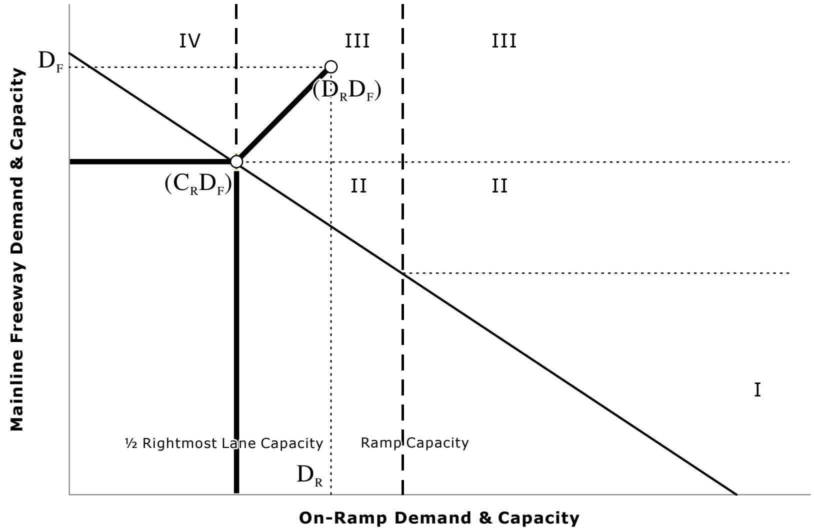


Figure 2.9. Capacity allocation for demand in Region III.

Enhancement 2: Stochastic Capacity and Turn Pocket Analysis on Arterials

Traffic flow along arterial street systems is affected by the operating characteristics of each individual approach along the arterial and system effects from upstream and downstream intersections (i.e., queue spillback or blockage). A significant body of knowledge is available in the *Highway Capacity Manual* (2000) for estimating capacity at individual approaches based on a number of factors including lane geometry, lane widths, signal timing, and turning movement demand. However, one of the traditional shortcomings of arterial analysis procedures is the lack of ability to model system effects, particularly in the context of a network and short of developing resource-intensive microsimulation models. The advancement of DTA models such as those applied in this research effort allows for the analysis of system-level effects that incorporate the unique factors of individual intersection approaches along with upstream and downstream conditions for each approach.

In order to improve the realism of operating conditions along arterials, two significant enhancements were made to the DTA models used in this research project to test the effects of non-lane-widening strategies:

1. *Stochastic capacity for arterials.* The bottleneck points for signalized arterials are most often located at intersections.

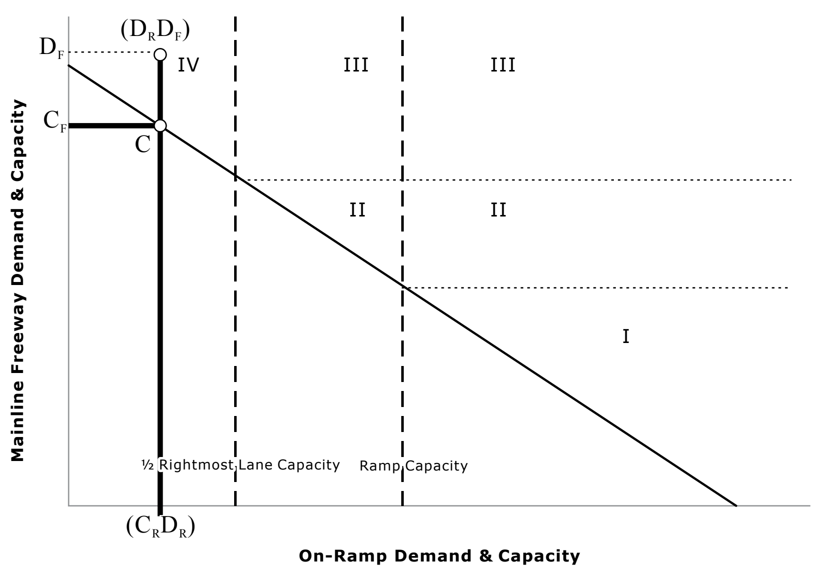


Figure 2.10. Capacity allocation for demand in Region IV.

Traditional DTA models assume a constant saturation flow rate during the green time for either the approaching links or for individual turn movements at these locations. However, it is well known that the saturation flow rate for individual links and turn movements varies significantly according to the behavior of individual drivers. At signalized intersections this is revealed by saturation flow rates that vary from cycle to cycle. Therefore, a stochastic approach was developed that allows the capacity of a signalized intersection to vary during each time interval (typically 15 minutes) according to an empirically observed distribution.

2. *Short turn pocket effects.* The approaches to signalized intersections along arterial roadways often include left and right turn pockets as a way of separating turn movements and increasing capacity. But when the queue length of through and/or turning vehicles extends beyond the length of the turn pocket, the result is a demand blockage that prevents upstream vehicles from taking advantage of the capacity that is available at the intersection. This is an important phenomenon to model in oversaturated networks because it directly affects the efficiency and productivity (or SSRs) of individual links and turn movements. Therefore, an enhanced model was developed that recognizes when queue lengths exceed available storage lengths at these locations and then adjusts the downstream discharge rate accordingly.

The following sections describe the enhancements made to the DTA models to incorporate the effects of stochastic variability of saturation flow rates and short turn pockets at signalized intersections.

Stochasticity of Saturation Flow Rates

Saturation flow rate is the maximum sustainable rate at which vehicles can discharge from a signalized intersection stop line. Saturation flow rates are expressed in terms of vehicles per hour of green time per lane (vphgpl) and are the inverse of saturation headways (defined as the average number of seconds of green time required to discharge a single vehicle from a single lane). Saturation flow rates are treated as a constant in most travel demand models, so although different values may be applied to left turn, through, and right turn movements, these same values are assumed to be constant across all signalized intersections and, for any particular intersection, throughout the time period being analyzed. Thus, for example, saturation flow rates of 1,600 to 1,800 vphgpl are commonly applied on urban streets.

However, traffic engineers have long known that saturation flow rates fluctuate over time, and even from cycle to cycle at the same intersection. Especially in congested environments, relatively small fluctuations can result in significantly worse intersection performance than will otherwise be predicted by standard travel demand forecasting models. This is because there is less excess capacity available to clear queues caused by temporary demand/supply imbalances when the lane group is operating near its capacity, and so disproportionately more stopped time and delay is incurred in order to clear the effects of these imbalances.

To overcome this deficiency and provide more realistic performance characteristics, a stochastic model was developed for predicting saturation headways at signalized intersections based on a mean value. The headway data used in this analysis were obtained from a recent research effort by the Florida Department of Transportation. As part of this project, a saturation headway database was developed, summarizing cycle-by-cycle headway observations for three different lane group types (through only, right only, and shared through/right) on 35 approaches across 19 different intersections. Extensive statistical investigations revealed that the lognormal probability distribution model provides the best fit to the empirical data that was collected. Figure 2.11 presents the headway density and cumulative probability plots for all the sites with lognormal distribution. The black lines represent the nonparametric fitted curves for probability density and cumulative probability plots. The blue dash lines represent the lognormal fitted curves.

The underlying variability did not change substantially even when the sites were grouped into different subsets. For example, Table 2.5 shows that data collected in small to medium-sized cities (population between 5,000 and 50,000) resulted in a somewhat higher mean and a somewhat lower variability than data that were collected in larger cities (population between 200,000 and 500,000). Despite this, the

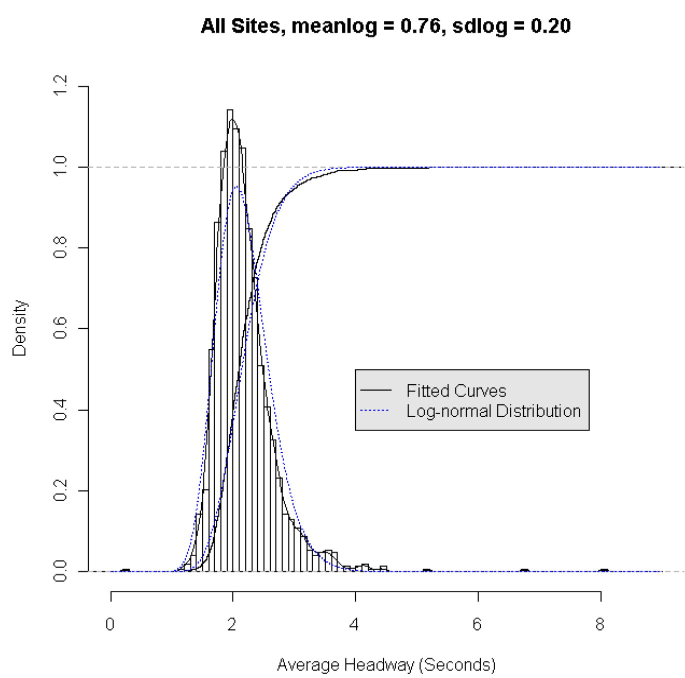


Figure 2.11. Headway density and cumulative probability plots for lognormal distribution.

Table 2.5. Lognormal Distribution Parameters: Through Lanes Only

Category	Mean of Lognormal Distribution	Standard Deviation of Lognormal Distribution
All sites	0.76	0.20
Small and medium cities	0.80	0.15
Large cities	0.75	0.20

standard deviation of the lognormal distribution for the two data sets remained in the same range as for the data set as a whole.

On this basis, it was concluded that the lognormal distribution model would be applied to the mean saturation headway predicted by the DTA model with a standard deviation of 0.15. The standard deviation parameter can and probably should be refined in the future, on the basis of more detailed and locally based studies of saturation headway data that capture a range of configuration types and roadway characteristics.

Short-Lane Effects at Signalized Intersections

A number of conditions upstream of the stop bar may affect the flow rate of vehicles, such as long queues that block vehicles from using one or more lanes with available space. Pocket blockage, for example, may occur when a continuous queue of through vehicles prevents left turning vehicles from accessing a short left turn lane (see Figure 2.12, bottom). This blockage may result in wasted green time while turning vehicles wait in queue before obtaining access to a turn pocket. Alternatively, long queues of turning vehicles can generate spillback, where queues extend beyond the back of the entrance to the turn pocket and impede the movement of vehicles on the adjacent through lane (see Figure 2.12, top).

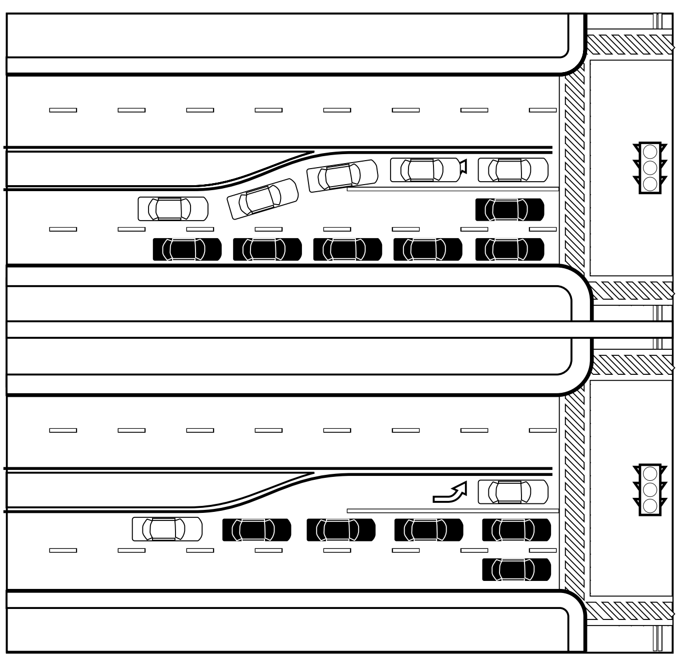


Figure 2.12. Blockage effects of short turn pockets.

The *HCM* ignores the effect of both spillback and blockage when calculating capacity, instead assuming that vehicles will be able to discharge from the intersection at all times when the light is green. With very long turn pockets, this is a fairly reasonable assumption, but when short turn pockets exist, ignoring the effects of spillback and blockage can have significant, compounding effects, and signals timed on the basis of accepted *HCM* practice may even exacerbate these problems.

The concept has been studied in a variety of contexts, but primarily in an effort to determine the probability of either spillback or blockage. By designing intersections with sufficiently long pockets with a low probability of either spillback or pocket blockage, the effects can largely be ignored. But conditions change, turning movement percentages shift over time, and geometric constraints can limit space available for turn pockets. It is therefore necessary for traffic engineers to be able to examine these effects in greater detail in order to effectively analyze mitigation options. Surprisingly, at the onset of this research, effectively the only option available to study these effects in detail was microscopic simulation, a time-intensive, expensive, and often complicated computer-based option. Ning Wu (*21*) developed a series of equations to predict discharge from a signalized intersection inclusive of these turn pocket effects based on simulated data, but their applicability is limited to intersections with a single approach lane. What the engineer and policy maker need is a simplified model capable of estimating the sustainable service rate of a signalized intersection, defined as the highest rate of flow that can be sustained over a peak demand period under prevailing conditions, inclusive of turn pocket spillback and blockage effects and the attendant lane changing behavior.

To accurately capture the propagation of queues induced by short left turn bays, this research extends the existing link-based mesoscopic simulation model by adding a gating mechanism at the entry point to the left turn pocket. Through a series of logical triggers, the gating mechanism allows for the vertical queuing of vehicles upstream of the turn pocket when arrivals exceed storage capacity.

Conceptually, a link with left turn bays can be partitioned into 3 parts: (a) left turn pocket with K bays and a length of L, (b) through pocket (adjacent to the left turn pocket) with M lanes and a length of L, and (c) upstream segment before the gate.

Figure 2.13 shows a representative approach link with two through lanes ($M = 2$) and a double left turn pocket ($K = 2$). The length of the minor pocket is denoted as H.

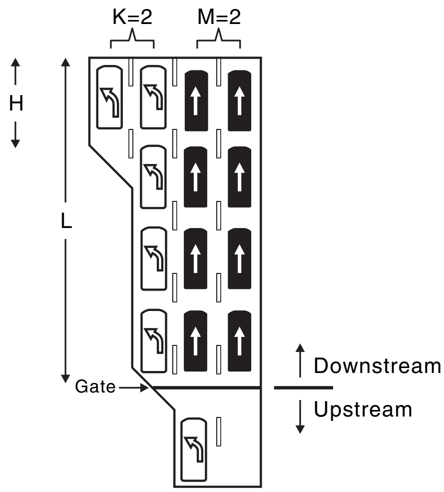


Figure 2.13. Illustration of a link with dual left turn pockets.

A clock-based simulation scheme is used in this study. The simulation time interval is denoted as ΔT, which should not be shorter than the shortest free-flow link travel time in the network (e.g., 6 seconds), so that a vehicle does not jump across two links during a simulation time interval. At each simulation time interval t, the position of vehicle i $x(i, t)$ is updated according to its speed $v(t)$. Similar to the modified Greenshields model used in DYNASMART-P, a minimal moving speed (e.g., 6 mph) is imposed to ensure that vehicles can move forward even at jam density. Without loss of generality, the position of the downstream end of a link is assumed to be 0, so a vehicle's position starts with the link length and moves decreasingly toward 0.

The simulation model uses a vertical queue or a point queue representation scheme, which leads to two important properties: (a) if the gating condition is not triggered, vehicles can always move to the end of the link and join the vertical queue, and (b) only vehicles in the vertical queue can be discharged to the downstream links. With the additional gating mechanism, if either a through or a left turn vehicle is blocked at the gate, then the vehicle cannot reach the end of the link (i.e., stop bar) and join the vertical queue. In this case, even if the green phase is displayed for the corresponding movement at a simulation time interval, a vehicle stopping at the gate is unable to be discharged, leading to wasted green time and a capacity loss due to blockage, thus giving rise to the concept of sustained flow rate or SSR.

For simplicity, the following discussion focuses only on left turn and through queues, as right turn vehicles typically have sufficient permissible time and storage space to be dissipated at all times. N^L and N^T represent the maximum numbers of vehicles that can be stored in the left turn and through pockets, respectively, at any given time. These two parameters can be viewed as the space capacity of each pocket, which is different from flow rate-based capacity (e.g., number of vehicles passing through a point during a certain given interval).

Consider the average vehicle length as AVL. As shown in Equation 2.9, the space capacity is determined by:

$$N^L = \left\lfloor \frac{L}{AVL} \right\rfloor + \left\lfloor \frac{H}{AVL} \right\rfloor, N^T = \left\lfloor \frac{L}{AVL} \right\rfloor \times M \qquad (2.9)$$

where

N^L = turn pocket storage (vehicles),
N^T = through lane storage (vehicles),
L = primary turn pocket length (feet),
H = secondary turn pocket length (feet),
AVL = average vehicle length (feet),
M = through lanes, and
$\lfloor\ \rfloor$ = rounding down to the nearest integer.

Accordingly, counters n^L and n^T are used to record the numbers of vehicles stored in the left turn pocket and adjacent through lanes, respectively.

There are four major events to be triggered, and the following binary flags are set to "false" by default:

f^L = true when the left turn pocket is full (i.e., $n^L = N^L$);
f^T = true when the through pocket is full (i.e., $n^T = N^T$);
b^L = true when a left turn vehicle blocks the gate; and
b^T = true when a through vehicle blocks the gate.

The conceptual discussion below aims to thoroughly describe the sequence of pocket full and blocking events and the interactions of left turn and through vehicles at the gate.

In Figure 2.14a, the left turn bay is occupied by two left turn vehicles, so f^L = true, while the second through vehicle can still travel through to the gate. In Figure 2.14b, the left turn blockage occurs when the third left turn vehicle arrives at the gate, and the vehicle has to stop there. Algorithmically b^L is triggered to true by an additional left turn vehicle after f^L = true. The fourth through vehicle in Figure 2.14c then cannot use the leftmost through lane to reach the downstream through pocket. Because only $M-1$ lane(s) are available for those following through vehicles, the following approximation formula is used in this study to determine the reduced through flow capacity at the gate:

$$C^T = \text{MFR} \times (M-1) \times \Delta T$$

where MFR is the maximum flow rate (i.e., the number of vehicles that can traverse a roadway segment) per lane per second, equivalent to the saturation flow rate.

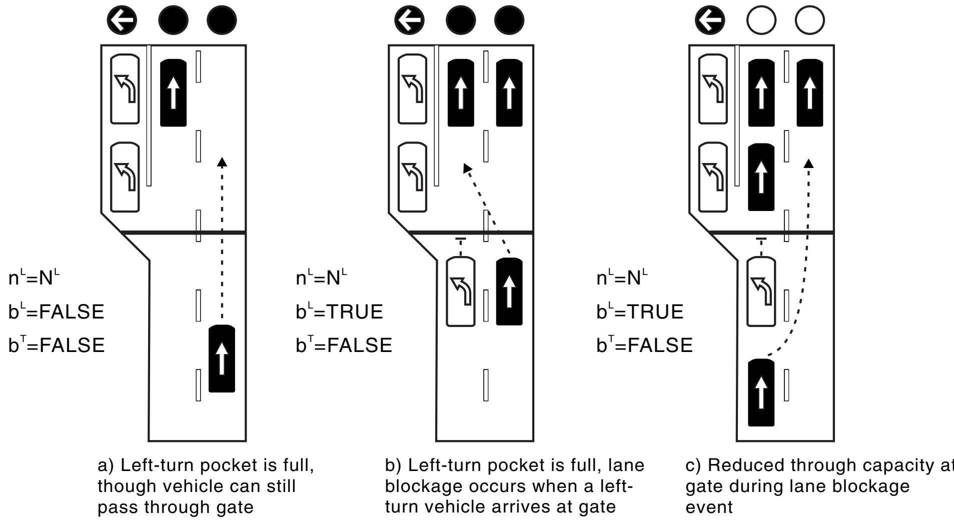


Figure 2.14. Illustration of left turn pocket blockage events.

In Figure 2.15a, four through vehicles occupy the downstream through pocket, so f^T = true, while the left turning vehicle can still enter the left turn bay. Figure 2.15b shows how the through blockage event is triggered when the fifth through vehicle arrives at the gate. That is, flag b^T is set to true by an incoming through vehicle when f^T = true. After the through blockage event occurs, no following left turn vehicles can pass through the gate, as illustrated in Figure 2.15c.

It should be noted that, in a real-world situation, if the fifth through vehicle queues in the rightmost lane instead, then blockage occurs later when another through vehicle enters the leftmost lane. This study adapts a simplistic deterministic assumption: the next through vehicle will block the leftmost lane.

Enhancement 3: Implementation of Day-to-Day Learning Paradigm

As had been stated previously, conventional traffic assignment methods assume static, deterministic road capacity. Therefore, the travel time of a path only depends on the flow pattern on that path. In other words, for a fixed networkwide path flow pattern, the corresponding path travel times do not change. However, real-world road capacities vary with time

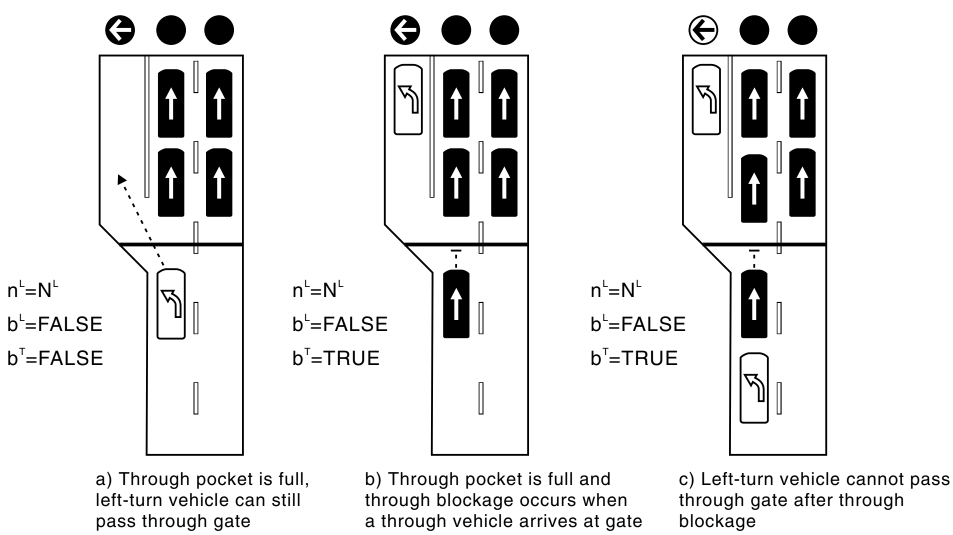


Figure 2.15. Illustration of through blockage events.

over a certain range, and a driver's traveling experience on a single day can be dramatically affected by the underlying realized capacity values on that particular day. In other words, travelers will experience different travel times on the same path over different days even for the same path flow pattern because of the inherent travel time variability introduced by stochastic capacity. As a result, conventional "within-day" or iterative route choice methods for reaching user equilibrium, such as the method of successive averaging (MSA), may not enable drivers to recognize and appropriately respond to the travel time variability/unreliability resulting from capacity fluctuation. A theoretically rigorous and practically useful traveler route choice model is crucially needed to adaptively capture the stochastic day-to-day travel time evolution process and also to maintain robustness under disruptions due to stochastic capacity reductions. To this end, a new route choice mechanism is proposed to simulate the drivers' route choice behavior under stochastic traffic process noise. By comparison, conversional stochastic assignment models focus on traveler perception errors under a deterministic traffic environment. The proposed mechanism includes two key components: a route choice learning module and a route choice decision module. In addition, different user classes, which receive and perceive different types of traffic information at different decision points along trips, are further investigated in this study.

Conceptual Overview

The day-to-day learning framework proposed by Hu and Mahmassani (*22*) and Jha, Madanat, and Peeta (*23*) provides a promising path for seamlessly integrating stochastic capacity models into the DTA simulator for large-scale networks. Generally speaking, the learning behavior in such a day-to-day learning framework is determined by each vehicle's historical traveling experiences, the traveler information obtained before and during the trip, as well as newly experienced travel times on the current day.

Conceptually, the model includes three components as shown in Equations 2.10 through 2.12:

Traffic flow assignment model: $f^{d+1} = A\left(f^d, T^d, w^d\right)$ (2.10)

Stochastic traffic system simulation process:
$t^d = S\left(f^d\right) + w^d$ (2.11)

Travel time perception model: $T^d = t^d + \varepsilon^d$ (2.12)

where

f^d = assigned route flow pattern on Day *d*, determined by traffic assignment model/function $A(\cdot)$;
t^d = true travel time on Day *d*, determined by dynamic assignment/simulation function $S(\cdot)$;
w^d = the system noise introduced by the stochastic capacity;
T^d = the observed travel time by a traveler; and
ε^d = the traveler perception error associated with perceived travel time in the network, introduced by the sampling error associated with personal experience and quality of information.

Most existing day-to-day learning models are implemented with stable road capacity, which assumes no system noise (i.e., $w^d = 0$), so the travel time is a deterministic vector for a given set of route flows, f^d in Equation 2.11. Accordingly, the focus in the previous research has been on how to reach the deterministic steady-state conditions, and how to construct realistic learning/updating models for the travel time perception error term ε^d related to Equation 2.12.

In this study, the research team enhances a dynamic traffic flow simulator, namely, DYNASMART-P, to describe a traffic simulation process with day-to-day varying system noise, w^d in Equation 2.11. Corresponding to the traffic flow assignment model, Equation 2.10, a day-to-day learning module is presented to describe adaptive traveler behavior across multiple days in a stochastic traffic evolution environment. The essential idea for the learning module is to enable certain users to use their historical traveling experiences to construct their estimates and make decisions under uncertain system travel times (introduced by the system noise). To simplify the route choice rules, the research team assumes $\varepsilon^d = 0$ in the following discussion. As a result, the proposed model does not involve the use of a Probit or Logit model to assign traffic flows and does not require a travel perception error updating process.

Day-to-Day Learning Simulation Algorithm

This study adapts a behaviorally sound route choice utility function, proposed and calibrated by Brownstone and Small (*24*) and Lam and Small (*25*), to consider the stochastic nature of traffic systems. As shown in Equation 2.13,

$$GT = T + \frac{VOR}{VOT} \times TSD + \frac{TOLL}{VOT} = T + \beta \times TSD + \frac{TOLL}{VOT} \quad (2.13)$$

where

GT = the generalized travel time,
T = the expected travel time for traveler,
TSD = perceived travel time variability,
β = reliability ratio [computed as the ratio of value of reliability (VOR) and value of time (VOT)], and
TOLL = road toll charge; it is assumed to be 0 in the following discussions as no toll-related strategies will be evaluated in this report.

It has been well recognized that travel time variability and reliability are important measures of service quality for travelers. In the above utility function, Equation 2.13, the travel time standard deviation (TSD) is used to measure system travel time variability associated with the underlying stochastic traffic process. This measurement contrasts with the perception error variance in a deterministic assignment model. For a single traveler v, the route choice decision is made by comparing the generalized travel time of habitual path, GT_v^h, and that of alternate path, GT_v^a, as shown in Equation 2.14:

$$GT_v^h > GT_v^a \tag{2.14}$$

where

v = traveler index,
h = index for habitual path, and
a = index for potential alternate path.

According to Equation 2.13, if the generalized travel time of the habitual path, GT_v^h, is greater than that of alternate path, GT_v^a, as shown in Equation 2.14, a driver should switch to the alternate path. The resulting decision rule could be derived as shown in Equation 2.15:

$$T_v^h - T_v^a > \beta\left(TSD_v^a - TSD_v^h\right) \tag{2.15}$$

In this study, T_v^h is equal to $\overline{T}_v^{d-K,d-1}$ as calculated in Equation 2.16, to take a traveler's multiday travel time experience into account. T_v^a is calculated by using the estimated travel time on the shortest path. It should be noted that the calculation varies for different user classes, which will be discussed in Chapter 3.

$$\overline{T}_v^{d-K,d-1} = \frac{T\left(P_v^{d-K}\right) + T\left(P_v^{d-K+1}\right) + \cdots + T\left(P_v^{d-1}\right)}{K} \tag{2.16}$$

where

d = day index,
K = number of days in the learning memory window,
$\overline{T}_v^{d-K,d-1}$ = traveling experience (i.e., average travel time) for traveler v from Day $d-K$ to Day $d-1$, on a particular path, and
$T\left(P_v^{d-1}\right)$ = travel time on path P_v^{d-1}, and P_v^{d-1} is the path traveled by vehicle v on Day $d-1$.

The right side of Equation 2.15 can be viewed as the minimum acceptable absolute tolerance needed for a route switch decision. This value arises from three components: the reliability ratio, β, the standard deviation of travel time on the habitual path, TSD_v^h, and the standard deviation of travel time on the alternate path TSD_v^a. The calibration study from Noland et al. (*26*) indicated a reliability ratio value of $\beta = 1.27$ based on survey data from more than 700 commuters in the Los Angeles region. The setting of parameter K depends on the signal-to-noise ratio in the traffic system. The more stable the travel time process, the smaller the parameter K can be and still yield a reliable mean travel time estimate. In general, K must be large enough to filter out the process noise from the stochastic traffic system.

The travel time variability measure, TSD_v^h, for the habitual path can be calculated from multiday travel times experienced by the traveler. The remaining challenge is how to estimate the standard deviation of travel time on the alternate path, TSD_v^a, where the traveler has little or no experience on this path. When there is no external pretrip or en route information available, TSD_v^a needs to be calculated from the traveler's prior experience. To the research team's knowledge, there is no widely accepted method to calibrate the standard deviation of perceived travel times on alternate paths for travelers without access to advanced traveler information systems and relying on prior knowledge only. In this research, the research team assumes that TSD_v^a is significantly larger than TSD_v^h due to the lack of precise information and the high level of uncertainty associated with the perceived alternate travel time. The calibration of the minimum acceptable absolute tolerance was beyond the scope of this study. Therefore, this research uses a simplified, single term model, $\beta(TSD_v^a - TSD_v^h)$, to represent the minimum acceptable absolute tolerance needed for a route switch decision. This simple model is intuitively sound, and using it eliminates the need for extensive calibration efforts.

In this study, a bounded rationality model, which states that drivers' decisions depend on their desired satisfaction level, is adapted to make the route choice comparison. The bounded rationality concept is employed because there has been growing attention [starting from the early work by Mahmassani and Herman (*27*)] to bounded rationality since Herbert Simon (*28*) pointed out that perfectly rational decisions are often not feasible given the limits of human cognition.

Based on the minimum acceptable absolute tolerance and the relative acceptable tolerance, a set of bounded rationality rules, shown in Equation 2.17, are used to describe users' route switching behavior. As opposed to the optimization theory in which users select the best option from all possible decisions, in the bounded rationality approach, users perform limited searches, accepting the first satisfactory decision.

$$\delta = \begin{cases} 1, \overline{T}_v^{d-K,d-1} - T_v^a > MAX\left[\alpha, \lambda T_v^h\right] \\ 0, \text{otherwise} \end{cases} \tag{2.17}$$

where

δ = 1, switch to an alternate path; 0, remain on the habitual/current path;
α = minimum acceptable absolute tolerance needed for a switch and $\alpha = \beta(TSD_v^a - TSD_v^h)$; and
λ = relative acceptable tolerance (i.e., relative improvement threshold).

Multiple User Classes and Conceptual Route Choice Simulation Framework

Three types of user classes are considered in this study: pretrip information users, en route information users, and unequipped users. The different user classes have access to different types of travel information to help them make their route choice decisions. Pretrip information is acquired by pretrip users before departure, through the Internet, TV, radio, and cell phones. En route information that describes the estimated time of arrival is provided to en route users during the trip by GPS navigation devices, radio channels, and variable message signs. The personal post-trip information acquired by unequipped users is typically based on a commuter's experienced travel time, in addition to potential external information sources from television and newspaper reports.

In this heterogeneous information environment, each user class has different ways to estimate travel time on the alternate path, and different decision-making locations and times. The pretrip users estimate on the basis of network real-time snapshot conditions just prior to their departure and make the route choice decision at the departure time. The en route users make route choice decisions each time they reach a node where alternate routes are available and estimate travel time based on the network real-time snapshot conditions. The unequipped users determine whether to change their habitual path on Day d at the end of Day $d-1$, when all trips complete and estimate the travel time on the shortest path based on average path travel times on Day $d-1$.

It should be noted that in reality, many people are creatures of habit and are unlikely to make route changes right away, if ever. Moreover, the information quality could vary for different user classes. Thus, the following assumptions are made about how different user classes receive information and how this information triggers route switching considerations:

- Pretrip and en route users are always willing to switch their routes.
- Only a certain percentage (p) of unequipped users have access to post-trip information and are willing to switch their routes.
- Pretrip/en route users receive the information with higher quality than unequipped users do.

The value of p also requires a site-specific calibration effort. For example, Haselkom, Spyridakis, and Meld (*29*) found that 20.06% of drivers were willing to switch their routes in the study area in Washington State, whereas Abdel-Aty et al. (*30*) found that to be true of only 15.50% of drivers in Los Angeles.

For different user classes, the implementation framework of the route choice models is shown in Figure 2.16.

Enhancement 4: New Performance Measures and Implementation Considerations

In this section, the stochastic performance measures, which could be generated from the enhanced DTA simulator, and the implementation considerations for the models described previously are discussed.

Performance Measures

Traditional transportation planning analysis has primarily focused on two performance measures: peak hour volume/capacity ratios by link and intersection level of service. The use of relatively few performance measures has traditionally been a result of limitations in traffic data and/or computational and modeling capabilities. However, in recent years these barriers have dissipated and many agencies have gained access to a wealth of data and modeling tools.

The advancement of mesoscopic and microscopic traffic models now enable more discrete analyses of traffic conditions and performance reporting in both space and time. Spatial analyses can be extended from a traditional node or link to an entire corridor, an O-D pair, or a whole network. Likewise, the time period for analyses can be extended from the traditional 15 minutes or one hour to multiple hours, a whole day, or time periods across multiple days. This last option allows transportation planners to examine not only the performance for a facility but also the reliability of the performance (represented by the variability that occurs across multiple days).

Ultimately, the type of performance measures that are appropriate to report and the scale or dimension at which they are reported are unique to the transportation network that is being analyzed, the mobility issue that is being addressed, and the treatment that is being tested.

Link-Level Performance Measures

For each link (i) in DYNASMART-P, the following measures of effectiveness (MOEs) are reported for every 15-minute interval over all simulation days:

- Link vehicle count (veh/15-minutes), v_i;
- Average travel time (minutes), t_i;
- Space mean speed (mph), s_i;
- Vehicle density (veh/mi/lane), d_i;
- Queue length, q_i (as an example, in DYNASMART-P, q_i is defined as the ratio of vehicle queue length to the link length);
- Freeway link breakdown indicator, b_i (for a freeway link, if at the end of every 15-minute interval, there is a queue, $b_i = 1$; otherwise, $b_i = 0$);

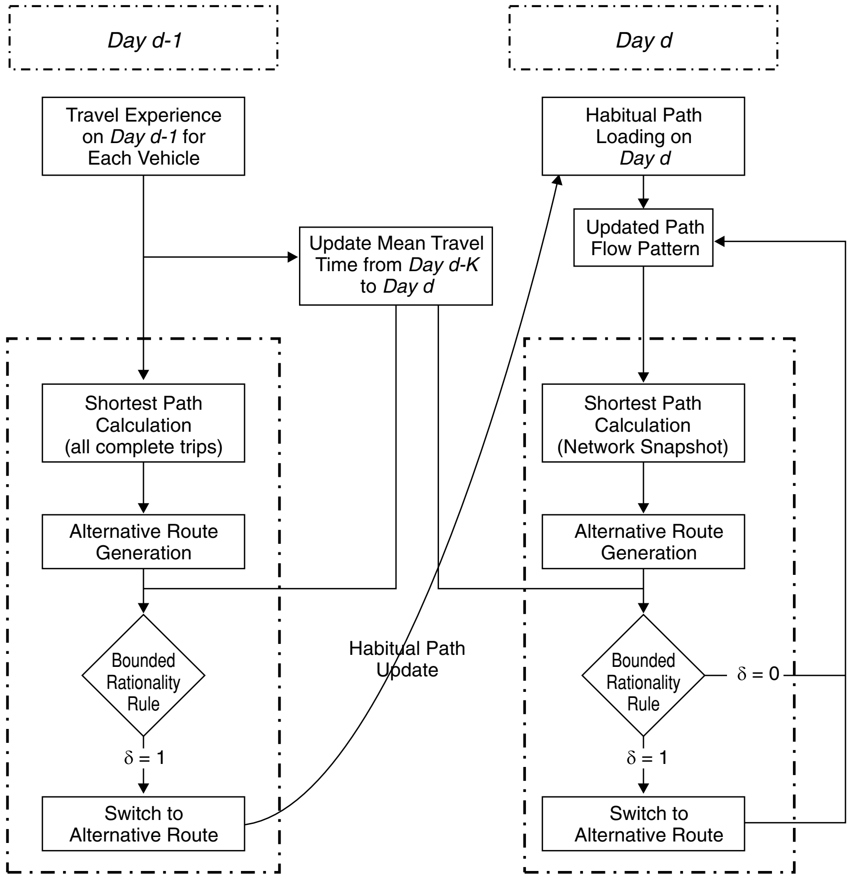


Figure 2.16. Implementation framework of route choice mechanism.

- Number of cycles with a queue at the start of the red phase (exclusively for signalized links); and
- Link capacity (veh/h/lane), c_i.

The link capacity is reported as an MOE because for certain links, capacities are not constant values. These links have stochastic capacities generated at the beginning of every 15-minute interval. Stochastic capacity for freeway segments and stochastic saturation flow rates for signalized intersection approach links are generated on the basis of a shifted lognormal distribution implemented in the simulator.

The users can summarize the link-level MOEs for any time scale within a single simulation day subject only to the 15-minute interval minimum mentioned above. As mentioned above, users can obtain descriptive statistics for these MOEs (e.g., mean and standard deviation) for the time scale of interest over multiple simulation days.

Corridor-Level Performance Measures

DYNASMART-P output cannot directly produce the MOEs for a roadway corridor of interest. However, several MOEs for the user-specified corridor can be generated by aggregating one or more link-based MOEs according to the following formulas. Since the corridor-level MOEs are aggregated on the basis of link-level MOEs, they should be reported for the same time scales as the link-level MOEs discussed above. These equations are not currently integrated within the mesoscopic simulator but are easily applied in postprocessing.

Density (veh/mi/lane) for a linear corridor encompassing multiple links (i) each with different length and number of lanes is shown in Equation 2.18:

$$D_c = \frac{\sum_i (d_i \times n_i \times l_i)}{\sum_i (n_i \times l_i)} \tag{2.18}$$

where

D_c = density for the user-specified linear corridor,
d_i = density for link (i) in the corridor,
l_i = link length (miles), and
n_i = link number of lanes.

Space mean speed (mph) for a linear corridor encompassing multiple links (i) is shown in Equation 2.19:

$$S_c = \frac{VMT}{VHT} = \frac{\sum_i (v_i \times l_i)}{\sum_i \left(v_i \times \frac{l_i}{s_i} \right)} \tag{2.19}$$

where

S_c = space mean speed for the user-specified corridor,
VMT = total distance traveled on the user-specified corridor,
VHT = total travel time on the user-specified corridor,
v_i = vehicle count for link i,
l_i = link length (miles), and
s_i = space mean speed for link i.

Queue length for a linear corridor encompassing multiple links (i) is shown in Equation 2.20:

$$Q_c = \frac{\sum_i (n_i \times q_i \times l_i)}{\sum_i (n_i \times l_i)} \tag{2.20}$$

where

Q_c = queue length on user-specified corridor as a fraction of corridor length,
n_i = number of lanes for link i,
q_i = queue length for link i, and
l_i = link length (miles).

Corridor travel time for a linear corridor encompassing multiple links (i) is shown in Equation 2.21:

$$T_c = \sum_i t_i \tag{2.21}$$

where t_i = travel time for link i.

Corridor breakdown count or cycle failure count for a linear corridor encompassing multiple links (i): The breakdown count or cycle failure count of the user-specified corridor is equal to the summation of breakdown counts or cycle failure count of each single link that is included in the corridor.

Origin-Destination Pair Performance Measures

For the origin–destination (O-D) pairs, average travel time (minutes/veh) is only MOE produced by DYNASMART-P, and it is reported for the entire simulation period. Before running the simulation, the user may define multiple critical O-D pairs, which could be an O-D pair with highest demand or an O-D pair users are interested in. For that user-defined critical O-D pair, average travel time is reported for multiple simulation days; while for other O-D pairs, this MOE is only reported on the last simulation day.

Network-Level Performance Measures

At the network level, DYNASMART-P generates average travel time (minutes/veh) for each of the vehicle/user subgroups listed below:

- Networkwide (all vehicles);
- Low-occupancy vehicle (LOV) group;
- High-occupancy vehicle (HOV) group;
- Unequipped vehicle group (vehicles that have no access to pretrip or en route information);
- Pretrip information vehicle group;
- En route information vehicle group; and
- Critical O-D vehicle group.

The average travel time for each subgroup on a single simulation day is reported for the entire simulation period. Users may also evaluate this MOE over multiple simulation days on the basis of their need. Table 2.6 summarizes the performance measures reported as part of the testing of select non-lane-widening strategies.

Model Implementation Considerations

In this section, the implementation considerations for the stochastic capacity models (both freeway bottlenecks and signalized intersections) and the short turn pocket model are discussed.

Stochastic Capacity Generation

As a significant component of the overall simulation framework, the implementation procedure of the stochastic capacity generation for both freeway bottleneck and signalized intersections is recommended as shown in Figure 2.17.

Table 2.6. Performance Measures by Spatial Type

MOE	Node/ Link	Corridor	O-D Pair	Network
Volume count	X			
Travel time	X	X (end to end only)		
Speed	X	X		
Density	X	X		
Queuing	X	X		
Freeway breakdown count	X	X		
Cycle breakdown count	X	X		
Capacity	X			
Average travel time			X	X

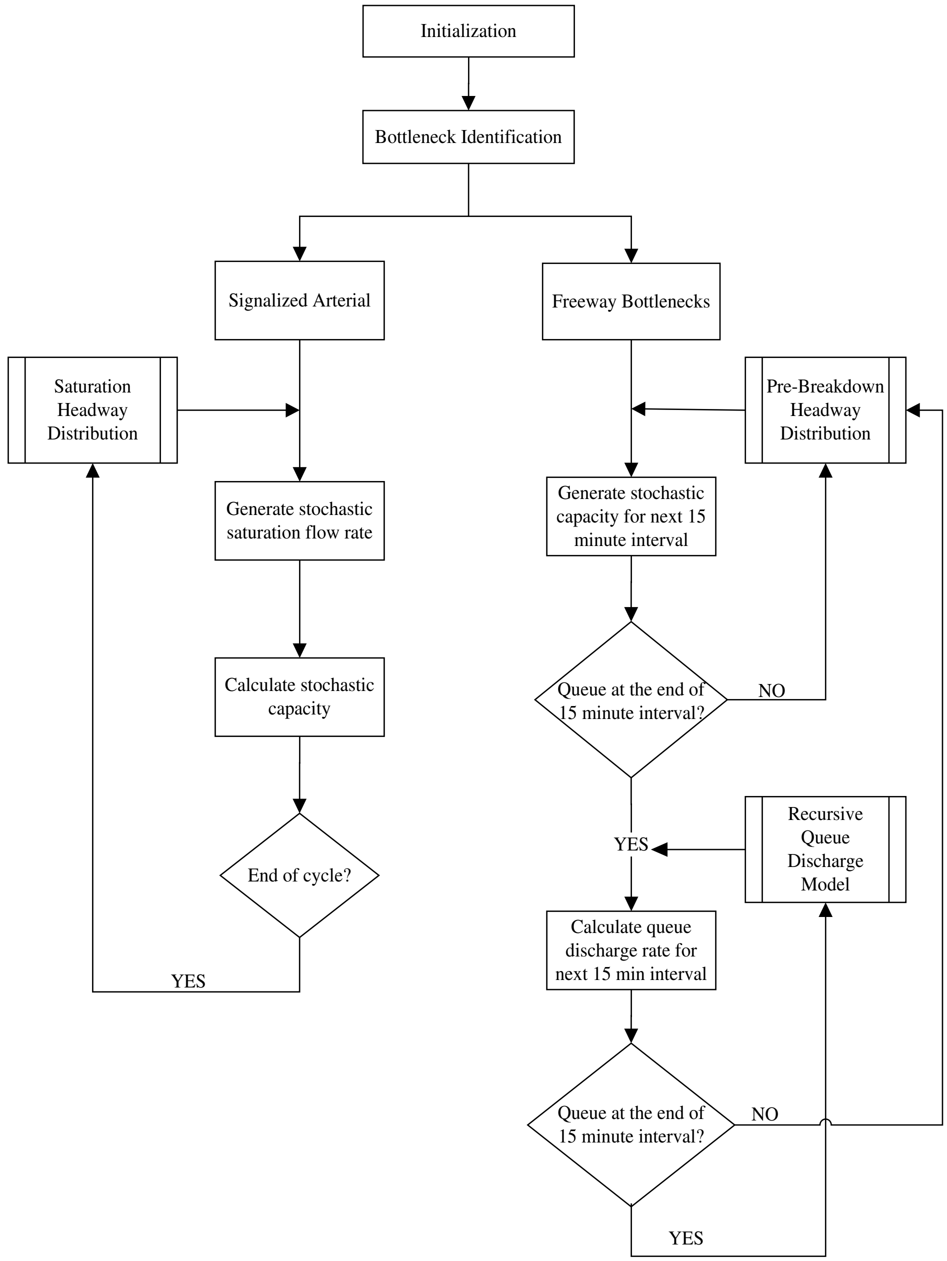


Figure 2.17. Implementation framework of stochastic capacity generation.

Short-Lane Effect Modeling

When the short-lane effects are modeled at signalized intersections, DYNASMART-P provides a promising alternative to microscopic simulation for analyzing short turn pocket effects. However, its mesoscopic platform does not model vehicle interactions or lane discipline, requiring some modifications in order to take into account the effects of blockage and spillback. The research team developed and described a "gating" mechanism, which serves as a flow regulator at the entrance to the turn pocket. Essentially, the gate operates by first checking for available queuing space downstream before allowing vehicles to pass. If the downstream lanes are full of vehicles, the gate prevents the vehicle from moving forward, which, in turn, limits flow for subsequent vehicles. In other words, the model bypasses the need to model individual vehicle interactions or lane discipline by simply reducing or preventing flow past the gate when a vehicle begins to generate blockage of a lane.

To capture the effects of short turn pockets on SSR in this mesoscopic model framework, four general requirements must be met. The model must (a) characterize and account for the effects of pocket spillback, (b) capture the wasted green time associated with turn pocket blockage, (c) appropriately measure differences in SSR due to phase order, and (d) account for dynamic lane assignment and the ability of through vehicles to utilize available lane space to bypass queues from pocket spillback. As discussed, although several macroscopic models are sensitive to one or more of these requirements, none accounts for the effects of pocket spill-back in a comprehensive manner.

References

1. Jia, A., B. M. Williams, and N. M. Rouphail. Identification and Calibration of Site Specific Stochastic Freeway Breakdown and Queue Discharge. *Transportation Research Record: Journal of the Transportation Research Board, No. 2188,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 148–155.
2. *Highway Capacity Manual.* Transportation Research Board, Washington, D.C., 2000.
3. Elefteriadou, L., R. P. Roess, and W. R. McShane. Probabilistic Nature of Breakdown at Freeway Merge Junctions. *Transportation Research Record 1484,* TRB, National Research Council, Washington, D.C., 1995, pp. 80–89.
4. Minderhoud, M. M., H. Botma, and P. H. L. Bovy. Assessment of Roadway Capacity Estimation Methods. *Transportation Research Record 1572,* TRB, National Research Council, Washington, D.C., 1995, pp. 59–67.
5. Persaud, B., S. Yagar, and R. Brownlee. Exploration of the Breakdown Phenomenon in Freeway Traffic. *Transportation Research Record 1634,* TRB, National Research Council, Washington, D.C., 1998, pp. 64–69.
6. Kuehne, R. D., and N. Anstett. Stochastic Methods for Analysis of Traffic Pattern Formation. *Proc., 14th International Symposium on Transportation and Traffic Theory,* Jerusalem, 1999.
7. Lorenz, M., and L. Elefteriadou. A Probabilistic Approach to Defining Freeway Capacity and Breakdown. *Transportation Research Circular E-C018: Proceedings of the Fourth International Symposium on Highway Capacity.* TRB, National Research Council, Washington, D.C., 2000, pp. 84–95.
8. Okamura, H., S. Watanabe, and T. Watanabe. An Empirical Study on the Capacity of Bottlenecks on the Basic Suburban Expressway Sections in Japan. *Transportation Research Circular E-C018: Proceedings of the Fourth International Symposium on Highway Capacity.* TRB, National Research Council, Washington, D.C., 2000, pp. 120–129.
9. Zurlinden, H. Ganzjahresanalyse des Verkehrsflusses auf Strassen (Whole-Year Analysis of Highway Traffic). Doctoral thesis, Institute for Transportation and Traffic Engineering, Ruhr University Bochum, No. 26, 2003.
10. Brilon, W., J. Geistefeldt, and M. Regler. Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity. *Proc., 16th International Symposium on Transportation and Traffic Theory,* College Park, Md., 2005, pp. 125–144.
11. Brilon, W., J. Geistefeldt, and H. Zurlinden. Implementing the Concept of Reliability for Highway Capacity Analysis. *Transportation Research Record: Journal of the Transportation Research Board, No. 2027,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 1–8.
12. Dong, J., and H. S. Mahmassani. Flow Breakdown and Travel Time Reliability. *Transportation Research Record: Journal of the Transportation Research Board, No. 2124,* Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 203–212.
13. TransGuide. www.transguide.dot.state.tx.us/. Accessed January 6, 2009.
14. PeMS. https://pems.eecs.berkeley.edu/?redirect=%2F%2F%3Fdnode%3DState. Accessed January 18, 2009.
15. Chen, C. A., A. Skabardonis, and P. Varaiya. Systematic Identification of Freeway Bottlenecks. *Transportation Research Record: Journal of the Transportation Research Board, No. 1867,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 46–52.
16. Elefteriadou, L., R. P. Roess, and W. R. McShane. Probabilistic Nature of Breakdown at Freeway Merge Junctions. *Transportation Research Record 1484,* TRB, National Research Council, Washington, D.C., 1995, pp. 80–89.
17. Snedecor, G. W., and W. G. Cochran. *Statistical Methods,* 8th ed. Iowa State University Press, 1989.
18. *NIST/SEMATECH e-Handbook of Statistical Methods,* 2005. www.itl.nist.gov/div898/handbook/.
19. Chakravarti, I. M., R. G. Laha, and J. Roy. *Handbook of Methods of Applied Statistics,* Vol. 1, John Wiley and Sons, 1967, pp. 392–394.
20. Dong, J., and H. S. Mahmassani. Flow Breakdown, Travel Reliability and Real-Time Information in Route Choice Behavior. *Proc., 18th International Symposium on Transportation and Traffic Theory,* Hong Kong, 2009, pp. 675–695.
21. Wu, N. Total Approach Capacity at Signalized Intersections with Shared and Short Lanes: Generalized Model Based on a Simulation Study. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2027,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 19–26.
22. Hu, T., and H. S. Mahmassani. Day-to-Day Evolution of Network Flows under Real-Time Information and Reactive Signal Control. *Transportation Research Part C: Emerging Technologies,* Vol. 5, No. 1, 1997, pp. 51–69.
23. Jha, M., S. Madanat, and S. Peeta. Perception Updating and Day-to-Day Travel Choice Dynamics in Traffic Networks with Information Provision. *Transportation Research Part C: Emerging Technologies,* Vol. 6, No. 3, 1998, pp. 189–212.
24. Brownstone, D., and K. A. Small. Valuing Time and Reliability: Assessing the Evidence from Road Pricing Demonstrations. *Transportation Research Part A: Policy and Practice,* Vol. 39, No. 4, 2005, pp. 279–293.
25. Lam, T., and K. A. Small. The Value of Time and Reliability: Measurement from a Value Pricing Experiment. *Transportation Research Part E: Logistics and Transportation Review,* Vol. 37, No. 2, 2001, pp. 231–251.
26. Noland, R. B., K. A. Small, P. M. Koskenoja, and X. Chu. Simulating Travel Reliability. *Regional Science and Urban Economics,* Vol. 28, No. 5, 1998, pp. 535–564.
27. Mahmassani, H. S., and R. Herman. Interactive Experiments for the Study of Trip Maker Behaviour Dynamics in Congested Commuting Systems. In *Developments in Dynamic and Activity-Based Approaches to Travel Analysis* (P. Jones, ed.), Gower, Aldershot, England, 1990, pp. 272–298.
28. Simon, H. A. A Behavioral Model of Rational Choice. *Quarterly Journal of Economics,* Vol. 69, 1955, pp. 99–118.
29. Haselkom, M., J. Spyridakis, and W. Meld. *Improving Motorist Information Systems: Towards a User-Based Motorist Information System for the Puget Sound Area.* Final Report. Washington State Transportation Center, University of Washington, Seattle, 1990.
30. Abdel-Aty, M. A., K. M. Vaughn, R. Kitamura, P. P. Jovanis, and F. L. Mannering. Models of Commuters' Information Use and Route Choice: Initial Results Based on Southern California Commuter Route Choice Survey. *Transportation Research Record 1453,* TRB, National Research Council, Washington, D.C., 1994, pp. 46–55.

CHAPTER 3

Operational Technologies and Treatments

This chapter identifies technologies and non-lane-widening treatments that improve network capacity and reduce the breakdown probability on arterial and freeway facilities and networks. The first section focuses on technologies, both emerging and visionary, that have the ability to positively impact overall network operations. The second section focuses on treatments, some of which incorporate the technologies described in the first section, and organizes the treatments on the basis of their applicability to facility type (arterial, freeway, or both) and their potential for increasing capacity and reducing the probability of breakdown.

The research team prioritized both technologies and treatments based on an extensive review of past and emerging research as well as discussions with experts in the field. From this prioritization, 25 unique treatments were identified for testing and application in the enhanced network model described in Chapter 2.

Key Findings and Conclusions

Results from an investigation of technologies that have the potential for widespread implementation in both the near term (less than 10 years) and long term (beyond 10 years) were identified and evaluated on the basis of their potential to improve capacity and throughput, ease of implementation, and timetable for implementation. Table 3.1 summarizes the top 10 most promising technologies.

Following a similar approach, a set of 25 non-lane-widening treatments were identified that offer agencies the ability to improve the operational efficiency of their roadway networks. These treatments improve network operations by increasing base capacity, reducing the probability of breakdown, and/or shifting demand to underutilized links on the network. Each of these treatments was considered for testing by using the enhanced operational model described in Chapter 2. Table 3.2 summarizes the 25 identified treatments.

Technologies Affecting Traffic Operations

Innovation in the transportation sector has provided metropolitan regions with an overwhelming array of possibilities to improve traffic operations and add capacity without constructing additional lanes. Through a scan of emerging practices and discussions with experts in the field, technologies were identified that have the potential for widespread implementation within the next 10 years, as well as promising technologies with slightly longer implementation timetables, to reflect the potential for technologies to build on one another with proper planning and foresight.

Communications Matrix

Two important elements to consider when evaluating technologies are the entities engaged in the transfer of data and the type of data being transferred. The key to maximizing traffic performance, in the form of increased or more reliable throughput, lies in finding ways to allow these entities to interact more effectively and efficiently. As an example, consider an individual vehicle as an entity in this framework. The potential for improvement is seemingly endless but becomes much clearer when broken down by the interacting agent. Vehicle-to-vehicle communication—where data concerning speed, congestion, incidents, road conditions, and other factors flow freely between vehicles—may certainly have a significant effect on network performance, but will not likely be commonplace within the next 10 years. Alternatively, localized control devices may serve as the interacting agent, obtaining data concerning speed, turning movements, volume, and the like through sensors or short-range communication to dynamically adjust signal timing, ramp metering, and lane availability. Such interactions may be reached in the next 10 years, but only with a highly organized, unified effort to place the necessary systems in the appropriate locations. As a third option, a

Table 3.1. Top 10 Promising Technologies with High Potential for Improving Capacity and Throughput

Potential for Immediate Widespread Implementation	Potential for Widespread Implementation in Major Metropolitan Areas Within 10 Years	Implementation Likely in 10 or More Years
Signal coordination	Plus lane	Automated vehicle and highway system
Electronic toll collection	Flow management	
Reversible-lane control	Capacity assignment decisions	
Network-level ramp metering	Route guidance	
Control coordination		

centralized network management system could also act as the recipient and processor of data sent from individual vehicles. This scenario would require a management system capable of effectively interpreting and distributing the appropriate data, as well as a long-distance communication channel such as a satellite-based system or high-capacity wireless network.

The Communications Matrix shown in Table 3.3 provides a means for organizing such emerging technologies based on the source of the data, the intended recipient, the type of data transferred, and the necessary communication channel. Clearly such a premise requires broad, yet distinct, unambiguous, and easily identifiable categories. Four types of sources/recipients seem to emerge:

- *Vehicle.* All data transmitted to or originating from vehicles or their passengers.
- *Infrastructure.* Devices that collect or transmit informational data about the vehicles (e.g., VMS, tag readers) or the infrastructure itself (e.g., ice and snow conditions).
- *Local control.* All local control devices such as signals and ramp meters.
- *Network control.* All control and management systems at the corridor or regional level.

When placed in a four-by-four format providing 16 individual directional categories, the interplay between these entities allows for a variety of ideas while still maintaining accessibility. The source of data, through either direct collection or data synthesis, is identified as the "From" category and is listed across the top. The recipient of that data is listed down the left side of the matrix, labeled "To," and each individual component is identified with a directional code. For example, vehicle-to-vehicle communication is denoted by a VV numbered code, network-to-vehicle by an NV, and so forth.

The matrix shown in Table 3.3 distinguishes among individual technologies that share similar objectives but have different control methods. For example, dynamically responsive traffic signals could receive information from local control or network control. A localized system bypasses the need for a central agency but necessitates software capable of communicating with nearby signals to accommodate real-time changes in volume, transit priority adjustments, emergency vehicles, and the like. A system based on network control has greater power for platoon formation over long distances, as well as the ability to integrate oversaturation controls but also requires the installation of sensor, software, and communication devices on all control devices within the region of interest.

Table 3.2. Selected Non-Lane-Widening Treatments to Improve Capacity

Freeway	Arterial	Both
HOV lanes	Signal retiming	Narrow lanes
Ramp metering	Signal coordination	Reversible lanes
Ramp closures	Adaptive signals	Variable lanes
Congestion pricing	Queue management	Truck-only lanes
Pricing by distance	Raised medians	Truck restrictions
HOT lanes	Access points	Pretrip information
Eliminate weaving sections	Right/left turn channelization	In-vehicle info
Frontage roads	Alternate left turn treatments	VMS/DMS
Interchange modifications		

Table 3.3. Communications Matrix

Technology Matrix								
To	From							
	Vehicle		Infrastructure		Local Control		Network Control	
Vehicle	VV1	Receive Nearby Vehicle Decisions	IV1	Infrastructure Status	LV1	In-Vehicle Display of Signals	NV1	Route Guidance
	VV2	Transmit Subject Vehicle Decisions	IV2	In-Car Speed Limit Adjustment Display	LV2	Reversible Lane Control	NV2	Network Highway Advisory Radio
	VV3	Collaborative All-Way Stop Control	IV3	Local Conditions via Advisory Radio	LV3	In-Vehicle Lane Assignment Display	NV3	511 Assistance
	VV4	Collaborative Gap Acceptance	IV4	Smart Work Zone Navigation	LV4		NV4	External Vehicle Speed Control
	VV5	Adaptive Cruise Control	IV5	Automated Parking Enforcement	LV5		NV5	
	VV6	Collaborative Driving System	IV6	Intelligent Vehicle and Highway System	LV6		NV6	
	VV7		IV7		LV7		NV7	
	VV8		IV8		LV8		NV8	
Infrastructure	VI1	Vehicle Dynamics Info	II1	Advance Incident Detection/Warning	LI1	Dynamic Advance Warning Signals	NI1	Capacity Assignment Decisions
	VI2	Weather/Road Conditions	II2	Congestion Detection/Warning	LI2		NI2	Weather Forecasts
	VI3	Electronic Toll Collection	II3		LI3		NI3	Dynamic Congestion Tolling
	VI4		II4		LI4		NI4	
	VI5		II5		LI5		NI5	
	VI6		II6		LI6		NI6	
	VI7		II7		LI7		NI7	
	VI8		II8		LI8		NI8	
Local Control	VL1	Approach Trajectory Input	IL1	Intersection Conditions	LL1	Signal Coordination	NL1	Control Coordination
	VL2	Turning Movement Options	IL2	Dynamic Speed Limits	LL2		NL2	Oversaturated Control
	VL3	Lane Use Options	IL3	Blocked Lane Information	LL3		NL3	Network-Level Ramp Metering
	VL4		IL4	Ramp Meter Override	LL4		NL4	Flow Management
	VL5		IL5	Sensor Status	LL5		NL5	Plus Lane
	VL6		IL6		LL6		NL6	Algorithms for Speed Adjustment
	VL7		IL7		LL7		NL7	
	VL8		IL8		LL8		NL8	
Network Control	VN1	Probe Information (AVI, AVL)	IN1	Volume Monitoring	LN1	Performance Capabilities	NN1	Regional Handoffs/Coordination
	VN2	Desired Paths	IN2	Congestion Monitoring	LN2	Local Control System Status	NN2	
	VN3	Desired Arrival Times, etc.	IN3	Speed Monitoring	LN3		NN3	
	VN4	Performance Limitations	IN4	Unmanned Aerial Vehicles	LN4		NN4	
	VN5		IN5		LN5		NN5	
	VN6		IN6		LN6		NN6	
	VN7		IN7		LN7		NN7	
	VN8		IN8		LN8		NN8	

As a result, two technologies with similar objectives actually have very different pathways for implementation, as well as very dissimilar methods of combining with other technologies.

The following provides a summary of the technologies listed in Table 3.3 and sorted by each of the four origin entities discussed below.

Vehicle

To Vehicle

These include all forms of direct communication between vehicles. Adaptive Cruise Control (VV5) is a basic form of this strategy in which vehicles automatically adjust their speeds based on distance and approach speed data from other vehicles. As most forms of VV communication require sensors and onboard display devices in both the sending and receiving vehicle, most of these ideas will not be widely available within the next 10 years. However, they are all likely to have dramatic effects on traffic safety and sustainable flow, such as VV4 Collaborative Gap Acceptance, which provides lane availability information among vehicles to ease merging, or information regarding the safety of turning in front of oncoming vehicles. Examples of these technologies are the ongoing VII program in the United States and the ubiquitous transportation (u-T) networks research program in South Korea. A schematic of the u-T system is depicted in Figure 3.1.

To Infrastructure

In the absence of a GPS-based system, in which vehicles communicate directly with a central management system, vehicle-to-infrastructure communications such as the one tested by Demers and List (*1*) provide second-hand information to other drivers concerning localized conditions. VI1 Vehicle Dynamics and VI2 Weather/Road Conditions allow vehicles to pass along information regarding speed, rapid braking, or pavement conditions to roadside systems such as VMS to inform digital warnings. VI3 Electronic Toll Collection is a more immediate representation of this category, which certainly has the potential to improve traffic operation, especially if progressed to the point of open-road tolling.

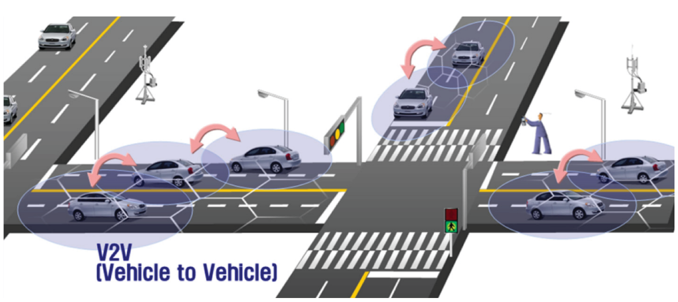


Figure 3.1. Vehicle-to-vehicle communication (u-T network, South Korea).

To Local Control

A basic form of this type of technology is outlined by Gradinescu et al. (*2*). Vehicles relay speed data to signals, thereby expanding the range for which signals can identify approaching vehicles as well as bypass the need for sensors. VL2 Turning Movement Options and VL3 Lane Use Options are both more advanced examples, in which vehicles relay their turning intentions or lane preferences to the signal for dynamic adjustments.

To Network Control

VN1 Probe Information and VN2 Desired Paths are examples of emerging technology with probe vehicles and GPS units aiding network management processes by providing both real-time travel information and route intentions for travel forecasts. As network management systems grow and develop, other useful information may emerge such as VN4 Performance Limitations, which relays speed, stopping ability, and traction information to the central system.

Infrastructure

To Vehicle

IV3 Advisory Radio provides a common example of this type of technology in use today. IV4 Smart Work Zone Navigation is beginning to emerge in the form of variable message signs, but throughput and safety may improve greatly as systems develop that allow work zone information to pass directly to vehicles through in-dash display, aiding speed management and merging. IV6 Intelligent Vehicle and Highway System is possibly the most obvious example of how technology will eventually completely overhaul transportation systems by allowing vehicles to coordinate with roadside sensors and computers to enter into an automatic mode, increase speed, and greatly reduce following distances.

To Infrastructure

Similar to vehicle-to-infrastructure technologies, these types of technologies simply have a slightly altered pathway. Sensors, rather than vehicles, collect volume, speed, or incident data and pass this information on to roadside digital signs.

To Local Control

This type of technology exists today in the form of detectors used at signalized intersections, but technological improvements will add new capabilities to sensor-to-control devices. IL1 Intersection Conditions and IL2 Dynamic Speed Limits (Figure 3.2), for example, would collect weather information and pavement conditions to inform signal timing adjustments

Figure 3.2. Variable speed limit (VSL) control in Stockholm, Sweden.

or safe speed limits displayed on digital overhead signs based on current weather conditions. IL4 Ramp Meter Override systems will also prove necessary within-ramp metering schemes to prevent arterial backups.

To Network Control

In addition to information collected directly from vehicles, pavement and roadside sensors will continue to play important roles in collecting speed and volume information for use at network management scale. While expensive, satellite-based systems (IN2 Congestion Monitoring) and IN4 Unmanned Aerial Vehicles may also prove valuable tools when used in conjunction with a centralized network management system.

Local Control

To Vehicle

Although these systems require some form of an onboard display system, suggesting a long path toward implementation, LV1 In-Vehicle Display of Signals may have significant safety implications, and LV3 In-Vehicle Lane Assignment Display could potentially have a large enough organizational effect on traffic to gain additional throughput. Reversible-lane controls, while already widespread, can lead to significant capacity improvements, especially when used in conjunction with real-time sensor, probe, or network management data.

To Infrastructure

Schultz (*3*) outlines the effects of Dynamic Advance Warning Signals, which provide early warnings to vehicles regarding current signal status via roadside or overhead displays. Such tools may become much more useful as signals advance from pre-timed to dynamically adjusted schemes, possibly leading to shortened or eliminated amber and all-red phases.

To Local Control

As previously discussed, LL1 Signal Coordination allows for local adjustments to cycle length, splits, and offsets through simple communication with nearby control devices, bypassing the need for network-level management.

To Network Control

To allow for network-level management of arterials, systems such as LN1 Performance Capabilities and LN2 Local Control System Status will need to be in place to relay the capabilities of the control devices, such as max greens, as well as the current status of cycle lengths and broken loops.

Network Control

To Vehicle

While NV3 511 Assistance currently serves as a major focus of information distribution regarding current roadway conditions, NV1 Route Guidance systems that utilize in-vehicle display of optimal routes will likely emerge as a more powerful tool. Demers and List (*1*) experimented with a system that used Wi-Fi, a pocket PC, and a synthesized voice to communicate route guidance to the driver. At this point it is unclear whether systems such as NV4 External Vehicle Speed Control, which limit a vehicle's speeds on the basis of its network location, will come into widespread use before an automated highway system eliminates the need for such an unpopular system.

To Infrastructure

The use of VMS to route traffic on the basis of current congestion (NI1 Capacity Assignment Decisions) or display area weather forecasts (NI2 Weather Forecasts) is emerging as a vital tool, especially during the long adjustment window needed for widespread adoption of in-vehicle display units (Figure 3.3).

Similarly, NI3 Dynamic Congestion Tolling provides a method for which network management systems can adjust tolls, creating greater incentive to follow network-suggested routes.

Figure 3.3. VMS queue-based routing in the Netherlands.

To Local Control

Network control over local control devices will likely emerge in a number of different ways, each intended to use real-time information processed at a central network agency to dynamically adjust traffic flow. NL3 Network-Level Ramp Metering and NL2 Oversaturated Control, where a network system completely overrides all signals in a region to flush an arterial network, are both powerful examples of this practice. NL5 Plus Lane is a more specific example of such a practice. Widely used in the Netherlands, as outlined by the International Technology Scanning Program (*4*), a Network Management System takes control over a freeway during congested conditions by opening the narrower left shoulder to traffic and lowering the speed limit across all lanes as depicted in Figure 3.4.

To Network Control

As network management systems expand their scope, there will likely be a need to break larger metropolitan networks into smaller regions with open sharing of information. Additionally, network-to-network communication may even be useful over greater distances to inform adjacent networks of approaching trucks or high volumes.

Relevance Assessment

The research team developed a method to rank and identify technologies that are believed to have the most immediate potential to provide improvements in traffic operations. Four categories of criteria were developed for the ranking:

- *Timetable.* A 1 through 3 grading scale was used on the basis of the following criteria with 1 representing the lowest score and 3 the highest: (a) technologies anticipated to exist in test-bed environments in the next 10 years, (b) technologies likely to see implementation in progressive metropolitan regions in the next 10 years, and (c) technologies likely to have widespread implementation in the next 10 years.
- *Ease of implementation.* A 1 through 3 grading scale was used on the basis of the following criteria, with 1 representing the lowest score and 3 the highest: (a) technologies with a large number of barriers, (b) technologies with a few barriers, and (c) technologies with relatively few barriers to implementation.
- *Capacity.* A 1 through 5 grading scale was used to stratify each technology on the basis of its expected impact on the ability of the system to process vehicles, with a 1 suggesting minimal impact, ranging up to a very significant improvement denoted by a 5.
- *Throughput.* A 1 through 5 grading scale was used to stratify each technology on the basis of the number of vehicles processed per unit time, with 1 suggesting minimal impact, ranging up to a very significant improvement denoted by a 5.

Figure 3.4. Implementation of plus lane (closed condition) in the Netherlands.

The research team applied individual knowledge, experience, and projections concerning each technology. While such a system is subjective in nature, it allows for a blending of ideas based on experienced judgment. This approach was further refined during a number of discussions in which each team member defended his or her rankings and assumptions. Table 3.4 presents the compiled rankings and averages. For the Timetable and Ease of Implementation categories, the averaged results were rounded to the nearest whole number, creating a final 1 through 3 ranking for each technology in each category. The Capacity and Throughput scores were also averaged, but the decimals were retained. The mean of these two averages is shown in the "Average Score" column. Table 3.4 is sorted first by Timetable, then by Ease of Implementation, and finally by Average Score.

The widely available and fairly easy-to-implement technologies rise to the top of Table 3.4, while those that are farthest off and most difficult to implement fall to the bottom. To balance near-term accessibility with significant long-term benefits, three cutoff scores were instituted (3.0, 3.5, and 4.0), increasing in value as the timetable for implementation increases. Those technologies that are available (3) and fairly easy to implement (2 or 3) needed to score higher than average (3 or higher) in order to achieve relevance for the purpose of this discussion.

Five of the 11 technologies in the top category topped this average score threshold:

- Signal Coordination (LL1);
- Electronic Toll Collection (VI3);
- Reversible Lane Control (LV2);
- Network-Level Ramp Metering (NL3); and
- Control Coordination (NL1).

(text continued on page 46)

Table 3.4. Ranked Technology Options

Ref	Technology	Details	Timetable (1 = test bed; 2 = selective; 3 = widespread)	Ease of Implementation (1 = low; 2 = medium; 3 = high)	Average Score
1	Signal Coordination	Local adjustments to cycle length, splits, offsets, without network-level control or guidance	3	3	3.21
2	Electronic Toll Collection	Toll collection from RFID or tag readers	3	3	3.14
3	Network Highway Advisory Radio	Networkwide radio broadcast of conditions/problem areas/detours	3	3	2.71
4	Volume Monitoring	Pavement sensors relay real-time highway volume information to network management system	3	3	2.57
5	Congestion Detection/ Warning	Use of detection devices and VMS to automatically display congestion warnings/ delay	3	3	2.57
6	Local Conditions via Advisory Radio	Localized radio broadcast weather/congestion/emergency conditions	3	3	2.07
7	Reversible Lane Control	Reversible lane control and indications.	3	2	4.14
8	Network-Level Ramp Metering	Network coordination of ramp meters for optimal highway flow	3	2	3.93
9	Control Coordination	Network-level adjustments for signal control	3	2	3.50
10	Local Control System Status	Ability to function, broken loops, current cycle length, splits, offsets, patterns in these same parameters over recent time, flash, emergency preemptions if they have occurred, and so forth	3	2	2.86
11	Ramp Meter Override	Queue detection allows for ramp meter override to prevent intersection backup	3	2	2.57
12	Plus Lane	Use of overhead signs to reduce speed limit and open left shoulder of a freeway to increase capacity during high volumes	2	2	4.07
13	Flow Management	Use of dynamic devices including overhead lights, in-pavement lighted lane dividers, or automatic movable barriers in order to maximize flow by opening or closing shoulder lanes or express lanes as needed for network optimization	2	2	3.86
14	Capacity Assignment Decisions	Use of VMS to route traffic based on network information (around incidents, alternate routes to avoid congestions and so forth).	2	2	3.57
15	Route Guidance	In-vehicle display of suggested route based on current/projected conditions	2	2	3.50
16	Oversaturated Control	Signal timing to flush arterial networks. Ramp meter closures. Signal timing plan changes to ameliorate oversaturation. Changes in lane use to provide surge capacity in specific directions on specific facilities.	2	2	3.43
17	Dynamic Congestion Tolling		2	2	3.14
18	Advance Incident Detection/ Warning	Use of detection devices and VMS to automatically display incident warnings	2	2	3.00
19	Blocked Lane Information	Infrastructure senses breakdowns, or degradations in lane condition and relays this information to local control	2	2	2.79

(continued on next page)

Table 3.4. Ranked Technology Options (continued)

Ref	Technology	Details	Timetable (1 = test bed; 2 = selective; 3 = widespread)	Ease of Implementation (1 = low; 2 = medium; 3 = high)	Average Score
20	511 Assistance	Interactive route guidance info via cell phone	2	2	2.64
21	Performance Capabilities	Performance capabilities of the local control: cycle length, max greens, storage lane lengths, where variable, spillback limitations, to the extent that they vary, sensed throughput capabilities, such as variations in saturation flow rates due to sun glare, local work zones, intense parking, and so forth.	2	2	2.64
22	Automated Parking Enforcement	Example from South Korea: Sensors and digital video collect tag info from illegally parked vehicles and automatically broadcast warning to owner over speakers/cell phone/Internet, and issue tickets by mail.	2	2	2.57
23	Probe Information (AVI, AVL)	Vehicle tracking data for network use	2	2	2.42
24	Dynamic Speed Limits	Speed limit adjustments made based on pavement conditions and weather	2	2	2.14
25	Speed Monitoring	Roadside sensors relay real-time highway speed information to network management system	2	2	2.14
26	Weather Forecasts	Network weather station provides localized forecasts to VMS	2	2	2.14
27	Sensor Status	Open channel communication between sensors and local control (faulty loop, stalled vehicle) to prevent poor control due to inaccurate information	2	2	2.00
28	Dynamic Advance Warning Signals	Overhead warning of signal status in advance, useful for high-speed intersections	2	2	1.71
29	Adaptive Cruise Control	Automatic cruise control which maintains a minimum following distance	2	1	3.14
30	Algorithms for Speed Adjustment	Network controlled algorithms for localized speed adjustment based on network conditions	2	1	2.50
31	Lane Use Options	Lane use options that the vehicles see, for example: need to be in the RH lane, need to be in the LH lane, indifferent about the through lanes, based on immediate and upcoming turning movements; also dimensional restrictions, such as need to be in the center lane, or restriction limitations, such as trucks have to be in the center lane.	1	2	2.93
32	Desired Arrival Times	Vehicle transmits desired arrival time at a specific location for feedback regarding best available route.	1	2	2.86
33	Smart Work Zone Navigation	Virtual cones, work zone navigation assistance—cone locations, trajectories, paths to follow, smoother merges	1	2	2.50
34	Intersection Conditions	Local control parameters adjust for ice, snow, sun glare, and so forth.	1	2	1.93
35	Weather/Road Conditions	Adjust infrastructure conditions (e.g., salt application) in response to vehicle behavior information	1	2	1.64
36	Intelligent Vehicle and Highway System	Driver inputs their destination and hands off vehicle control to an onboard computer when entering a freeway, which coordinates with other vehicles and roadside infrastructure. The intelligent system then optimizes lane usage and spacing with other intelligent cars, forming platoons and maximizing capacity.	1	1	4.21

(continued on next page)

Table 3.4. Ranked Technology Options (continued)

Ref	Technology	Details	Timetable (1 = test bed; 2 = selective; 3 = widespread)	Ease of Implementation (1 = low; 2 = medium; 3 = high)	Average Score
37	Collaborative Driving System	Automated driving system utilizing platoon formation and coordination with other vehicles	1	1	3.64
38	Collaborative Gap Acceptance	Yield or stop-controlled locations or lane changing on freeways or arterials, what are the gaps that can be collaboratively accepted or must be rejected, vehicles working together to accommodate needed lane changes; the idea is akin to forced merges, but a lot more polite.	1	1	3.00
39	Desired Paths	Vehicle transmits expected route to network management system.	1	1	2.93
40	In-Vehicle Lane Assignment Display	Local control provides vehicle lane assignment directly through in-vehicle display.	1	1	2.79
41	Congestion Monitoring	Satellite-based sensors relay imagery and traffic monitoring data to network management system.	1	1	2.71
42	Infrastructure Status	Intelligent speed adjustment and settings for braking, acceleration, and so forth.	1	1	2.43
43	Approach Trajectory Input	Vehicle relays approach speed and vehicle type (emergency, public transit, HOV, etc.) to signal.	1	1	2.36
44	Receive Nearby Vehicle Decisions	In-car display/broadcast of nearby vehicle movements	1	1	2.36
45	Collaborative All-Way Stop Control	Vehicle communication and preference relay at all-way stop.	1	1	2.29
46	Turning Movement Options	Vehicle relays turning movement intention to signal.	1	1	2.25
47	In-Car Speed Limit Adjustment Display	Vehicle receives variable speed limit information for in-dash display.	1	1	2.21
48	Vehicle Dynamics Info	Informs speed limit adjustments, heads-up display messages, and so forth.	1	1	2.07
49	External Vehicle Speed Control (EVSC)	Controls vehicle speed based on network speed limit maps and onboard GPS/EVSC system	1	1	2.00
50	Unmanned Aerial Vehicles	Unmanned aircraft for continuous traffic surveillance	1	1	1.93
51	Regional Handoffs/ Coordination	Air Traffic Control handoff-type thoughts	1	1	1.83
52	In-Vehicle Display of Signals	Allows for in-car display of signal/ramp meter status at time of arrival.	1	1	1.79
53	Performance Limitations	Acceleration, deceleration, stopping distance, adhesion (snow, ice), safe following distances as observed by the vehicle control system.	1	1	1.71
54	Transmit Subject Vehicle Decisions	Signaling/brake light info transmitted directly to nearby vehicles	1	1	1.64

(continued from page 42)

The next 17 technologies will take slightly longer to achieve widespread availability (2) and have a few, but not an overwhelming number of, barriers to implementation (2). Of these, 4 technologies score higher than 3.5:

- Plus Lane (NL5);
- Flow Management (NL4);
- Capacity Assignment Decisions (NI1); and
- Route Guidance (NV1).

The remaining technologies contain a 1 for either timetable or ease of implementation, suggesting that they are likely too distant or too difficult to implement for the purpose of this discussion. Intelligent Vehicle and Highway System (IV6), however, scores the highest of any technology considered.

Inventory of Network Operations Treatments

Treatments refer to the actions and applications that have the potential to improve sustainable service rates and reduce the probability of breakdown along freeway and arterial facilities and networks. Nearly 100 treatments were identified through a comprehensive literature search and discussions with technical experts in the field. In some cases the treatments incorporate technologies identified in the previous section.

The team ranked the effectiveness of each treatment with respect to the following three criteria:

- *Effect on peak-hour congestion.* The potential for the treatment to reduce peak-hour recurring congestion by increasing capacity and/or reducing the probability of breakdown.
- *Purview of decision makers.* Degree to which agency decision makers have the ability and institutional authority to implement the treatment.
- *Ease of implementation.* Ease of implementing the treatment considering economic, social, political, and environmental costs.

Based on the results of the ranking process, and following a consolidation of treatments into broader categories, a set of 10 broad categories emerged as effective, viable, and ready for implementation. Table 3.5 presents these categories and characterizes them according to their individual ability to increase capacity and decrease the probability of breakdown on freeways and arterials.

The following sections provide a brief description of each treatment as well as a description of the capacity-enhancing effects, known applications, and implementation needs.

Lane Treatments

Lane treatments result in an added or dedicated travel lane for a directional movement of traffic and/or a specific vehicle/user type in the current paved section of roadway. Their objective is to improve the vehicle-moving capacity for peak directional movements during congested periods of the day. Lane treatments can also be applied to increase the people-moving capacity of the facility and reduce travel demand in the case of bus-only or HOV lanes. Lane treatments apply to both freeway and arterial facilities and can be implemented on a static or dynamic basis. Lane treatments represent the most common class of treatments for recurring bottlenecks.

Common examples of lane treatments include

- Narrow lanes/use of shoulder lanes;
- Reversible lanes for arterials;
- HOV lanes on freeways;
- Variable lane controls at a signalized intersection; and
- On-street parking restrictions during peak periods.

Narrow lanes refer to adding a travel lane by re-striping a roadway with narrower lanes and/or converting part or all of the shoulder to a travel lane (also referred to as a "plus" lane, as described in Table 3.2).

Reversible lanes are used on arterial roadways, freeways, and bridges/tunnels to increase capacity for facilities that have directional peak traffic flows. Most reversible-lane applications on freeways are implemented by constructing a separated set of lanes in the center of the freeway with gate controls on both ends. For undivided facilities, a movable barrier can be applied to physically separate opposing directions of traffic flow.

Reversible lanes on arterials are typically implemented by using DOWNWARD GREEN ARROW and RED X lane-use control signs as described in the *Manual on Uniform Traffic Control Devices* (*5*). Implementation issues include driver awareness/education, enforcement, safety (potential for head-on collisions), maintenance and operations, and accommodation of left turn movements (*6*).

An *HOV lane* is reserved for the use of carpools, vanpools, and buses; motorcycles can usually use them as well. Most HOV lanes are applied to freeway facilities next to unrestricted general purpose lanes, but some are also used on arterial roadways. HOV lanes are intended to increase the person-moving capacity of a corridor by offering incentives for improvements in travel time and reliability. An inventory of existing and planned HOV facilities is provided through FHWA's Office of Operations website for HOV facilities.

Other lane treatments include converting a closely spaced on-off ramp sequence to a weaving section by extending the acceleration and deceleration lanes into a full auxiliary lane,

Table 3.5. Summary of Treatments for Achieving Improvements in Network Operations

Treatment	Freeway: Increases Capacity	Freeway: Decreases Probability of Breakdown	Arterial: Increases Capacity	Arterial: Decreases Probability of Breakdown
Operational Treatments				
Lane Treatments - Narrow lanes - Reversible lanes - HOV lanes - Variable lanes - On-street parking restrictions	•	•	•	•
Signal Timing - Signal retiming - Adaptive traffic control - Queue management - Transit/truck signal priority			•	•
Traffic Demand Metering - Ramp metering - Mainline metering - Ramp closures - Arterial demand metering		•		•
Congestion Pricing - Pre-set pricing - Dynamic pricing - Distance/vehicle class tolls - High-occupancy tolls - Central area pricing		•		•
Traveler Information - Pretrip information - In-vehicle information - Roadside messages - GPS navigation devices		•		•
Variable Speed Limits		•		•
Design Treatments				
Access Management - Raised medians - Access consolidation/relocation - Right turn channelization - Frontage roads			•	•
Geometric Design Treatments - Flyovers - Improving weaving sections - Alternate left turn treatments - Interchange modifications - Alignment changes	•	•	•	•
Truck-Related Treatments				
Truck/Heavy Vehicle Treatments - Truck-only lanes - Truck restrictions/prohibitions - Truck climbing lanes	•	•	•	•

adding lanes at off-ramps to mitigate the impact of a downstream signal on the surface street of an interchange, and adding temporary (median) lanes at weaving sections during peak periods.

Variable lanes at an intersection refer to the use of variable lane-use control signs that change the assignment of turning movements to accommodate variations in traffic flow. Conversely, many jurisdictions restrict left turn movements at intersections during peak periods through static or variable signing. While variable lanes are currently applied on a time-of-day basis, variable lanes could be applied dynamically with the proper technology and driver education/enforcement in place. The use of variable lanes to add turn lanes requires adequate turning radii, presence of a sufficient number of receiving lanes, variable-mode signal phasing, and advance warning signs.

On-street parking restrictions are often applied on urban roadways to provide additional through-capacity during peak commute periods and to preserve parking for local uses during off-peak periods. Enforcing the parking restrictions is a key challenge given that the potential capacity associated with the parking lane may not be achieved if one or more vehicles remain parked in violation of the restriction.

Signal Timing

Signal retiming is a process that seeks to optimize the controller's response to roadway user demand by implementing or modifying signal timing parameters (i.e., phase splits, cycle length, and offset), phasing sequences, and control strategies. Signal retiming can be carried out for an individual intersection, an arterial corridor, or an entire network. Effective signal retiming can increase capacity and reduce signal delay, which leads to lower travel times, improved reliability, and reduced driver frustration. Significant benefits can be achieved for intersections that have experienced changes in traffic flows and arrival patterns since the timing plans were last updated.

Adaptive traffic control systems (ATCS) use algorithms and system detectors to perform real-time optimization of traffic signals based on current or anticipated future traffic conditions. The adaptive software adjusts signal splits, offsets, phase lengths, and phase sequences to achieve a defined objective (e.g., minimize delay, reduce stops). There are five types of adaptive traffic control systems in use today. Two of the first ATCS systems developed are SCATS and SCOOT. SCATS (Sydney Coordinated Adaptive Traffic System) was developed in Australia in the early 1970s and SCOOT (Split Cycle Offset Optimization Technique) was developed in the United Kingdom a few years later. Recently, in the United States, other adaptive control systems have been implemented and/or are in testing, including RHODES (Real-Time Hierarchical Optimized Distributed and Effective System), OPAC (Optimization Policies for Adaptive Control), and FHWA's ACS-Lite.

Queue management is a signal timing technique for oversaturated arterials that seeks to minimize queue spillback within turn lanes or between links. A range of queue management techniques have been developed and applied around the world. Such techniques include zero offsets, reverse progression, gating, and metering (*7–9*), and diversion (*10*). These techniques are particularly critical at closely spaced intersections with limited queuing space, where queue spillback can block entries into critical intersections and cause a major reduction in arterial throughput (*11*).

Transit/truck signal priority gives special treatment to particular modes such as transit vehicles or trucks at signalized intersections. It does this by either extending the green phase or truncating the red phase for the approaching vehicle. It is unlike signal preemption in that it does not disrupt signal progression. The primary benefit of transit/truck signal priority is improved schedule reliability for transit vehicles and improved capacity/safety for the coincident traffic stream. Transit signal priority has also been shown to reduce travel time for transit vehicles. Transit signal priority does not have a significant effect on improving the vehicle-moving capacity of an arterial, although mainline movements typically benefit because green time increases when a priority call is placed. However, truck signal priority can have a significant effect on the capacity of the affected arterial movement.

Traffic Demand Metering

Traffic demand metering has useful applications for both freeways and, to a lesser extent, arterial networks. Demand-metering techniques are typically based on the goal of reducing the probability of breakdown of a freeway or major roadway by controlling the rate and location of additional new demand (e.g., from on-ramps and toll plazas). The metered traffic is allowed to enter the freeway or major road at a rate that is compatible with continuous or "sustained service flow" on the mainline. When appropriately applied, demand metering can increase the capacity of freeway and major road sections, and can contribute to the goal of this research, that is the maintenance of a sustained service rate (SSR). Metering may accrue other benefits, including those related to safety and the environment. Metering can also improve travel time reliability.

The three principal control methods for demand metering include local pre-timed, local-traffic responsive, and system-wide traffic responsive. These three approaches depend on the availability of local and system sensors as well as the ability of a local site (e.g., one on-ramp) to communicate with nearby on-ramps or an entire system of ramps.

One special application of ramp metering is the closure of one or more on-ramps during congested peak periods. In effect, this action does not increase demand on the downstream freeway section, but does require a high level of user information and signing to divert vehicles from the on-ramp to the adjacent arterial system.

The use of demand-metering techniques on a network of arterial streets is typically a more subtle form of freeway ramp metering. Traffic signal timing replaces ramp meters as the key element. Timing can be made more restrictive upstream of a bottleneck to reduce the rate of new demand that reaches the bottleneck. Thus, the delay and diversion activity is increased at the less congested upstream location while less demand is placed on the bottleneck.

The implementation of any demand-metering treatment requires a host of supporting actions, including traffic monitoring, communications, and control algorithms, typically based on a combination of historical and real-time information.

Congestion Pricing

Congestion pricing, also known as value pricing, implements a special type of toll to reduce traffic volume during particular times of congestion or in particular areas of congestion. Congestion pricing does not increase capacity but rather reduces the chance of breakdown during the most critical hours. The toll changes driving behavior by serving as an incentive for drivers to travel at different times, to find alternate routes, or to choose other methods of travel (e.g., mass transit, carpooling). It is meant to encourage drivers to be more conscientious and mindful of their driving habits.

There are several different ways to implement pricing. Preset pricing involves fluctuating tolls depending on the time of day, even if traffic flow is not congested. Conversely, if the pricing is a function of traffic flow, then the toll will fluctuate depending on the amount of congestion. There are different types of pricing: distance/vehicular classification types, open road types, and closed road types (*12*).

Distance/Vehicle Class tolls include tolls for driving a certain distance or for driving a truck. Open-road tolling involves tolling one way. Two examples of open-road tolling are high-occupancy-toll (HOT) lanes and express lanes. Both types are toll lanes adjacent to non-toll lanes on a roadway that allow drivers to have the opportunity to pay a toll to avoid congestion.

New tolling technology can help reduce the complexity and improve the accuracy of the congestion pricing as well as expedite the tolling process. Technology such as electronic tolling helps reduce the delay for paying the tolls.

While current experience and research indicate that congestion pricing has great potential, the ease of implementation and actual benefit will vary for each application. Political and public support is a key factor for implementation, as is the availability of alternate routes. Without alternate routes, an economic benefit will be achieved but capacity will not improve and significant driver resentment could result for the additional tolls.

A concept explored by FHWA to address equity concerns is called FAIR (Fast and Intertwined Regular) lanes (*13*). FAIR lane pricing creates tolled (express) lanes and non-tolled (general purpose) lanes on a freeway. When motorists use the general purpose lanes during rush hour, they would be compensated with credits that could be applied to the express lanes.

Traveler Information

Traveler information is information that can be provided to the driver that will allow him or her to make a well-informed decision regarding (a) what mode to take, (b) when to depart, and (c) the best route to travel. This information can be provided before and/or during the trip through the Internet, telephones, television/radio, roadside signs, and in-car displays and devices.

Many road agencies are implementing ATIS. These systems incorporate close-to-real-time information on roadways collected through cameras and traffic reports. The information gathered can be sent out through highway agency Internet sites or via private Internet sites.

Dynamic message signs (DMSs) and variable message signs (VMSs)—electronic signs that can be changed to provide current traveler information, as well as alternate route information—are being applied by highway agencies along highly traveled routes to provide updated information to travelers while they are en route.

Variable Speed Limits

Variable speed limits are applied to freeway sections, primarily in metropolitan areas with large traffic volumes and display traffic-actuated speed limits on variable message signs. The primary objective of variable speed limits is to increase road safety and homogenize traffic flow.

Variable speed limit systems usually combine speed limit signs with other traffic signs or text messages that are used to display incident or congestion warnings. Variable speed limit systems are often implemented in conjunction with automated speed enforcement.

In many applications, variable speed limits are embedded in traffic control systems that also adopt other measures such as lane control, ramp metering, or temporary hard shoulder running.

An efficient way to motivate drivers to display adequate speed behavior is to provide the traffic adaptive indication of expected travel times (in minutes) to characteristic points along the freeway (e.g., to well-known large intersections) by VMSs. This application contributes to a more patient behavior in case of congestion and a less nervous style of driving in flowing traffic.

Access Management

Access management is the "systematic control of the location, spacing, design, and operation of driveways, median openings, interchanges, and street connections to a roadway" (*14*). The intent of access management is to provide access to land development while still maintaining a safe and efficient transportation system.

The *Access Management Manual* identifies the following principles for maintaining land-use access and improving the safety and operations of the arterial roadway:

- Limit direct access to major roadways.
- Locate signals to favor through movements.
- Preserve the functional area of intersections and interchanges.
- Limit the number of conflict points.
- Separate conflict areas.
- Remove turning vehicles from through-traffic lanes.
- Use non-traversable medians to manage left turn movements.
- Provide a supporting street and circulation system.

Raised medians are applied to reduce turning movements and manage access to land uses along a corridor. Implementing a full barrier limits the number of interruptions in traffic flow. In addition to a full barrier, a limited access barrier can provide opportunities for drivers to make left turn movements where the agencies deem safe and appropriate.

There are also different techniques for improving the "margins" on outer edges of freeways or arterials to help improve capacity. First, lanes can be wider near the side of the road to provide more room for cars to maneuver. Another technique is to channelize right turn movements to minimize impedances to through movements. The use of frontage roads also creates separation between through and turning/local traffic that, in turn, can improve the capacity and sustainable service rate of the arterial roadway.

Geometric Design Improvements

Geometric design improvements refer to spot reconstruction or minor geometric widening that can be performed within the existing paved area. They are generally considered low- or moderate-cost improvements that are less significant than a major capital improvement project. They are often alternatives to facility widening projects.

Flyovers apply to interchange ramps and major through or turn movements at intersections. They are generally considered a spot treatment to address a high-volume movement as opposed to full reconstruction or lane widening of a facility. Flyover ramps can also be applied to at-grade intersections for high-volume left turn movements. A similar concept for urban areas is to depress the major through movement below the grade of the intersection. Flyovers or grade-separated movements add capacity by separating the traffic demand for the subject movement from conflicting flow.

Improving weaving sections primarily applies to freeways but can also apply to arterials. The improvement or elimination of weaving sections can be accomplished through changes in striping and lane assignment, use of medians to physically separate traffic flows, reconfiguration of ramps to add/remove movements, and realignment of ramps to increase weaving distance or remove the weaving movement. Weaving sections reduce speeds, capacity, and reliability (in addition to contributing to safety deficiencies). Improving weaving sections could potentially increase the roadway capacity to that of a basic freeway section or ramp merge/diverge.

Alternate left turn treatments for intersections refer to nonconventional intersections that convert left turn movements into other intersection movements in order to reduce the left turn signal phase. Examples include continuous flow intersections, jughandle intersections, superstreet intersections, and median U-turns. Alternate left turn treatments increase capacity and safety by eliminating one or more left turn phases and allowing more green time for remaining movements. Improved signal progression is generally achieved through a reduced signal phase.

Interchange modifications include changes to the interchange type, ramp configurations, and traffic control of the ramp terminals. An example of modifying an interchange type is converting a full cloverleaf interchange into a partial cloverleaf interchange to eliminate weaving sections. Other interchange modification techniques that could be made to increase freeway capacity include adding lanes on the entry or exit ramps, increasing the storage distance of on-ramps, changing the ramp alignment to reduce or increase travel speeds, and adding signalization or roundabout control at the ramp terminals. Ramp closures or restrictions to one or more vehicle types can be applied on a dynamic, temporary, or permanent basis to improve freeway performance.

Horizontal/vertical alignment changes apply to both freeways and arterials, and primarily older facilities that were designed and built before modern-day roadway design standards were put in place. Sharp horizontal or vertical curves affect the speed profile of vehicles and, anecdotally, can lead to sudden braking and increase the probability for breakdown.

Truck/Heavy Vehicle Restrictions

The management of truck traffic can have a positive impact on traffic operations for both freeways and arterial facilities. Traffic operations improvements relating to trucks and to capacity and sustained service rates may require additional factors to be considered such as allowable delivery schedules, weight restrictions on arterial streets, and vertical clearance limits. There are many possible truck operations techniques that relate to increasing capacity on freeways and major arterials. The most frequently used management techniques by state departments of transportation (DOTs) surveyed as part of NCHRP Synthesis 314 (*15*) include

- New or improved pavement;
- Truck climbing lanes;
- Lane restrictions for trucks;
- Restriction/prohibition of trucks on specific roads;
- Truck parking restrictions/prohibitions;
- Improved incident management;
- ITS strategies;
- Intelligent warning devices;
- Weigh-in-motion;
- Improved warning signing;
- Electronic screening; and
- Enhanced enforcement.

The major impacts on capacity created by removing trucks from mixed-use lanes are reduced headways, more consistent speeds, and possibly increased SSRs.

Selection of Operational Treatments for Testing Consideration

Following the comprehensive inventory and review of operational strategies and treatments identified in Table 3.5, a set of 25 strategies was selected for consideration in testing in the operational model developed as part of this project. The 25 strategies were selected on the basis of their ability to affect positive change in network operations. The treatments improve network operations by increasing base capacity, reducing the probability of breakdown, and/or shifting demand to underutilized links on the network.

The 25 strategies that were tested as part of this research project have been summarized in Table 3.2, organized by their application to freeway facilities, arterial facilities, or both.

References

1. Demers, A., and G. F. List. Probes as Path Seekers: A New Paradigm. *Transportation Research Record: Journal of the Transportation Research Board, No. 1944,* Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 107–114.
2. Gradinescu, V., C. Gorgorin, R. Diaconescu, V. Cristea, and L. Iftode. Adaptive Traffic Lights Using Car-to-Car Communication. Presented at IEEE 65th Vehicular Technology Conference, Dublin, 2007.
3. Schultz, G. G., R. Peterson, D. L. Eggett, and B. C. Giles. Effectiveness of Blank-Out Overhead Dynamic Advance Warning Signals at High-Speed Signalized Intersections. *Journal of Transportation Engineering,* Vol. 133, No. 10, 2007, pp. 564–571.
4. Mirshahi, M., J. Obenberger, C. A. Fuhs, C. E. Howard, R. A. Krammes, B. T. Kuhn, R. M. Mayhew, M. A. Moore, K. Sahebjam, C. J. Stone, and J. L. Yung. Active Traffic Management: The Next Step in Congestion Management. FHWA-PL-07-012. International Technology Scanning Program, FHWA, 2007.
5. Federal Highway Administration (FHWA). *Manual on Uniform Traffic Control Devices.* FHWA, U.S. Department of Transportation, 2003.
6. Rodegerdts, L. A., B. L. Nevers, B. Robinson, et al. *Signalized Intersections: Informational Guide.* Report No. FHWA-HRT-04-091. FHWA, U.S. Department of Transportation, 2004.
7. Messer, C. Extension and Application of Prosser-Dunne Model to Traffic Operation Analysis of Oversaturated, Closely Spaced Signalized Intersections. *Transportation Research Record 1646,* TRB, National Research Council, Washington, D.C., 1998, pp. 106–114.
8. Messer, C. Simulation Studies of Traffic Operations at Saturated, Closely Spaced Signalized Intersections. *Transportation Research Record 1646,* TRB, National Research Council, Washington, D.C., 1998, pp. 115–123.
9. Rouphail, N., and R. Akcelik. A Preliminary Model of Queue Interaction at Signalized Paired Intersections. *Proc., 16th ARRB Conference, Australian Road Research Board,* Vermont South, Victoria, Australia, Vol. 5, 1992, pp. 325–345.
10. Quinn, H. *A Review of Queue Management Strategies.* ITS University of Leeds. Deliverable No. 1 for Workpackage No. WP110. July 1992.
11. Prosser, N., and M. Dunne. A Procedure for Estimating Movement Capacities at Signalized Paired Intersections. Presented at Second International Symposium on Highway Capacity, Sydney, Australia, 1994.
12. Papacostas, C. S., and P. D. Prevedouros. *Transportation Engineering and Planning.* Prentice Hall, Inc., Englewood Cliffs, N.J., 1993, pp. 307–310.
13. Hecker, J. Reducing Congestion: Congestion Pricing Has Promise for Improving Use of Transportation Infrastructure. GAO-03-735T. U.S. General Accounting Office, 2003.
14. Transportation Research Board. *Access Management Manual.* Transportation Research Board of the National Academies, Washington, D.C., 2003.
15. Douglas, J. G. *NCHRP Synthesis of Highway Practice 314: Strategies for Managing Increasing Truck Traffic.* Transportation Research Board of the National Academies, Washington, D.C., 2003.

CHAPTER 4

Strategy Testing Results and Insights

This chapter presents the results of applying the enhanced DTA model features described in Chapter 2 to test the effectiveness of the selected operational improvement strategies shown in Table 3.2 and to confirm the applicability and usefulness of the developed model. The model was applied in two real-world test networks. The first is a small subarea network of the Fort Worth, Texas, region and the second is a larger subarea network within the Portland, Oregon, metropolitan area.

The chapter includes a description of the network, followed by an outline of the evaluation procedures used and the results from applying the operational, design, and vehicle technology strategies in each network. The first analysis conducted within the Fort Worth network focused on individual strategy applications and testing. This was in part due to the fact that all the enhancements described in Chapter 2 had to be incorporated, tested, and verified in the DTA tools and therefore could be better managed in a small network. The Portland network study entailed combinations of strategies and was intended to apply the lessons learned from the small network to a more realistic operational environment. This chapter also presents a comparison between a select set of strategies and various lane addition scenarios in order to demonstrate the concept of a strategy's equivalent physical capacity addition.

As is summarized in the next sections, it is clear—and not surprising—that the effectiveness of strategies is very much dependent on a number of contextual factors including the network congestion condition, the availability of unused capacity elsewhere on the network, the pattern of origin–destination demands (concentrated or dispersed) and the spatial and temporal extent of the strategy. Therefore there is no single "silver bullet" answer as to which strategy is most effective. What is important to note is that the tools developed in this research will enable planners and engineers to directly contrast the effectiveness of nonconstruction strategies such as ATIS or reversible lanes or HOT lanes versus traditional capacity additions on a level playing field, by not only incorporating both types in the tools but also accounting for the driver learning behavior in responding to strategies over time. Thus, in contrast to the *HCM* procedures, for example, which assume no demand elasticity to operational or technological changes, the enhanced tools provided here allow for some elasticity to be accounted for at least in terms of route choice over time. This consideration of the time element of response also enables the introduction and generation of reliability-based measures of effectiveness into the research findings as well and begins an integration process of capacity- and reliability-based research in the SHRP 2 program.

The following is a summary of key findings and conclusions from the Fort Worth network application, which is described more fully later in this chapter.

- The effectiveness of each strategy cannot be quantified in a simple lookup table. The effectiveness of any particular operational improvement strategy was found to be heavily dependent on the physical, traffic, and operating context in which it is applied. The results of the strategy applications described in this chapter are informative at a general level, but actual performance characteristics cannot be predicted for another application in a different network and/or a different context without using a tool such as the enhanced DTA model.
- The effectiveness of each strategy is related to the scale (link, corridor, and/or network) at which performance is being measured. The effectiveness of strategies that modify specific link characteristics (e.g., narrowing the lanes or introducing reversible lanes) is likely to be most pronounced at the link level and much less so at a networkwide level. The effectiveness of strategies that broadly affect all links (e.g., improved traveler information systems) is likely to be most pronounced at the networkwide level and much less so at the link level.
- The enhanced DTA model can be used to estimate the "equivalent lane addition" impact of one or more operational improvement strategies. This is an especially important capability for analysts and transportation investment

decision makers who must make difficult decisions about when and where to add new lanes of capacity to the existing transportation infrastructure.

- Important insights are gained from the enhanced DTA model on the travel time reliability effects of operational improvement strategies. Particularly in congested networks, it is significant that some operational improvement strategies can improve travel time reliability even if they do not materially affect the average travel time. Improving travel time reliability is an important and equally effective way of giving time back to drivers as a commensurate reduction in average travel time. It is also a way of improving the overall quality of life within a community. And yet, until now no practical method has been available to account for operational strategy impacts on travel time reliability, so that this important measure has often been overlooked.
- The usefulness and usability of the model will be significantly enhanced when the effects of incidents like crashes and severe weather can also be taken into account. Traditional operational models do not account for the effects of events and incidents such as these. Additionally, the overall effectiveness of many operational strategies, such as those evaluated under this research effort, is incomplete if only the recurring congestion effects are considered.

The following is a summary of the key findings and conclusions from the real-world Portland network application, which is more fully described later in this chapter.

- It is both feasible and practical to use the methodologies described in this report to assess alternative improvement scenarios within an urban subarea. The time and resource requirements associated with such an effort are well within the capabilities of most transportation agencies and metropolitan planning organizations.
- Travel time reliability is an important performance measure to consider at both the corridor and network levels. This is particularly true in congested networks where the primary benefit from strategically placed operational improvements is improved reliability, even when average travel times are not significantly affected.
- Multiple performance measures should be monitored when alternative improvement strategies are tested. Collectively, they should provide insights into capacity utilization, productivity, travel time, queuing, and reliability, which allow the user to obtain a significantly better understanding of overall impacts than would be the case if only one or two performance measures were used.
- Performance measures should also be monitored at multiple spatial scales when alternative improvement strategies are tested. Specifically, performance should be evaluated at the link, corridor, and network levels, and for critical O-D pairs, to gain a complete understanding of strategy impacts and any trade-offs that might take place.

Individual Strategy Testing: Fort Worth Network

Network Description

The study network used for strategy testing is located in Fort Worth, Texas. Figure 4.1 orients the study network location relative to the Fort Worth region located within the Dallas–Fort Worth–Arlington metropolitan area. The small map on the right side of Figure 4.1 shows the Fort Worth network as coded in the DTA tool DYNASMART-P. I-35W is located in the middle of the network and provides freeway access to downtown Fort Worth, just to the north. I-35W is connected to the adjacent arterial streets by parallel frontage roads, which serve as entry and exit points to the freeway facility. As shown in the map, DYNASMART-P network has a total of 13 traffic analysis zones. The land use of the study area is mainly residential area except Zones 1 and 2. Land use of Zones 1 and 2 is mostly industrial or institutional. Zones 3, 6, 9, 11, 12, and 13 include some industrial activity. Therefore, the O-D pair from 1 to 2, called the Critical O-D pair (or Primary O-D pair), has the largest numbers of trips.

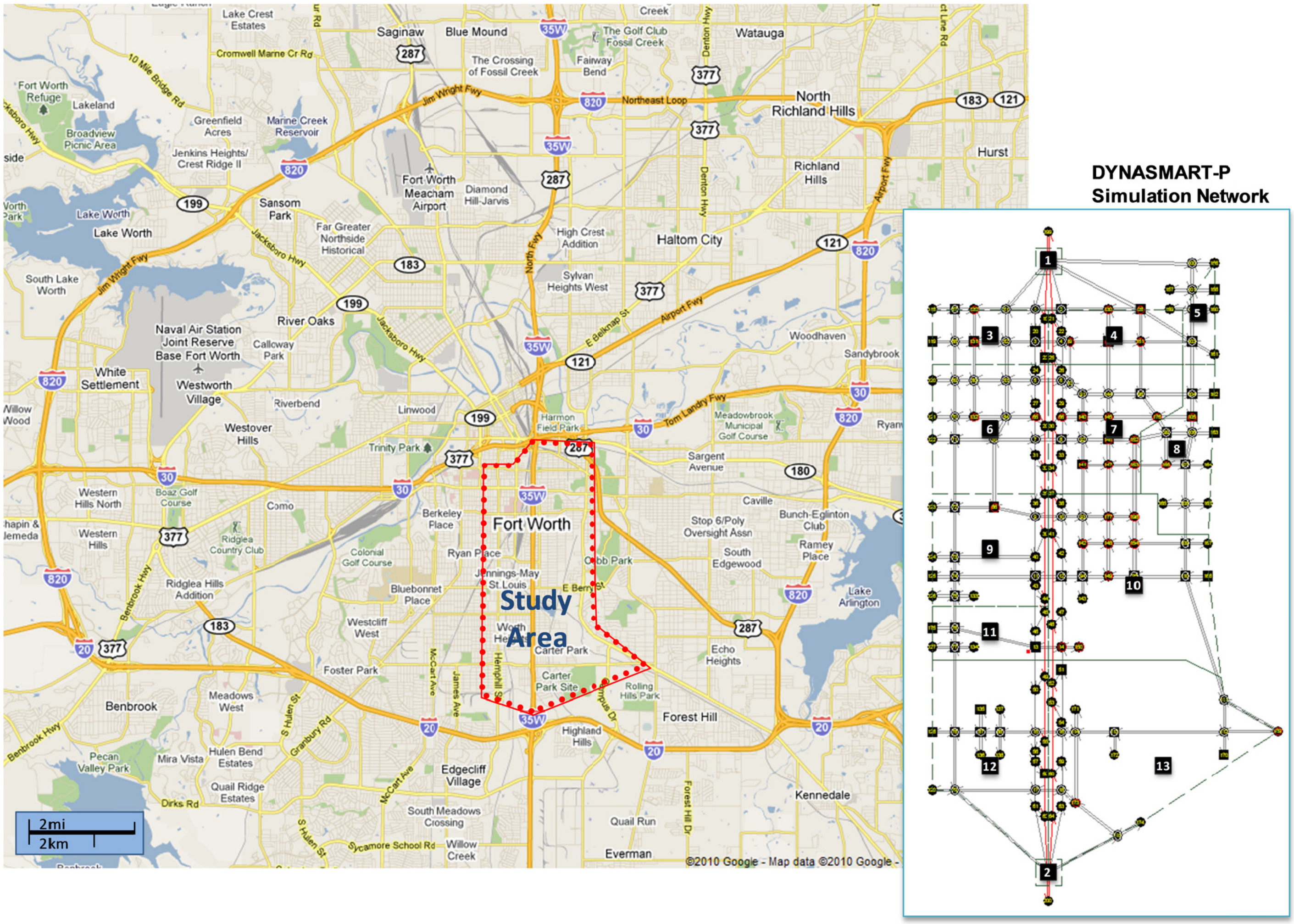
Roadway Attributes

The data set for this network was coded originally by the DYNASMART-P developers in the 1990s. The research team treated the network as an experimental framework and modified demand levels to generate a congested network that would be sensitive to various strategies of interest as well as roadway features to introduce features not originally included in the network, such as a double left turn pockets or six-lane arterial corridors. The following list provides a summary of modifications, with the number in parentheses indicating the total number of occurrences of each modification within the network.

- Modified demand profile (15-minute peaking);
- Modified overall LOV, HOV, and truck demand rates;
- Four-way stop to actuated signal control changes (2);
- Four-way stop to two-way stop changes (5);
- Single lane left turn pocket to dual left turn pocket (4); and
- Arterial link lane additions (8).

The principal road network attributes in this test network are

- Network data
  - Number of nodes: 180.
  - Number of links: 445.
  - Number of O-D demand zones: 13.



Map source: © 2010 Google Maps.

Figure 4.1. Fort Worth study area and DYNASMART-P simulation network.

- Node control type (number of intersections including on- and off-ramps)
  - No control (on- and off-ramps): 87.
  - Four-way stop: 24.
  - Two-way stop: 6.
  - Signalized (actuated control): 59.
- Traffic control data for signalized intersections
  - Two-phase control intersection: 10.
  - Three-phase control intersection: 18.
  - Four-phase control intersection: 35.
  - Max green: main 55 seconds, minor 25 seconds (or 20 seconds).
  - Min green: 10 seconds.
  - Amber: 5 seconds.

All links in the study network were defined as freeway or arterial links, characterized by a two-regime or single-regime modified Greenshields speed-flow model, respectively. Those models are explained in the DYNASMART-P user's manual (*1*). The settings for each of the two facility types are as follows:

- Freeway links
  - Maximum service flow rate: 2,200 pc/h/lane.
  - Saturation flow rate: 1,800 veh/h/lane.
  - Free-flow speed: 65 mph.
- Arterial links
  - Maximum service flow rate: 1,800 veh/h/lane.
  - Saturation flow rate: 1,800 veh/h/lane.
  - Speed limit: 40 mph.

Travel Demand Attributes

Time-Dependent Demand Profile

The actual analysis period is defined from 4:30 p.m. to 6:30 p.m., but in order to measure network statistics accurately over the full analysis period, vehicles were generated 30 minutes before the analysis time period (4:00 to 4:30 p.m.) as well as 30 minutes after the completion of the analysis period (6:30 to 7:00 p.m.). Over the entire simulation period, demand levels vary every 15 minutes, as shown below in Figure 4.2, with the height of each bar representing the ratio of 15-minute demand to the overall average demand. The first 30-minute period (4:00 to 4:30 p.m.) serves to load the network with vehicles before the start of the analysis period. The second period, or the peak analysis period (4:30 to 6:30 p.m.), has the highest demand levels and is the primary time period of interest. The third period (6:30 to 7:00 p.m.) is intended to simulate postpeak traffic and is referred to as the shoulder period. The fourth period (7:00 p.m. to network clear time) has zero demand and is included to allow sufficient time to collect statistics for all vehicles generated during the analysis period.

Demand Matrices by Vehicle Class

Demand for each of three vehicle types [i.e., low-occupancy vehicles (LOV), high-occupancy vehicles (HOV), and trucks] is expressed in twelve 15-minute O-D trip tables for the 3-hour demand period. The vehicle trip combinations used in the simulation network are LOV (85.0%), HOV (9.5%), and trucks (5.5%). LOV and HOV were assumed to have occupancy rates of 1 and 2.3 passengers per vehicle, respectively, while trucks use passenger car equivalent factors that vary based on the input roadway grade.

Demand Pattern for Alternate Baseline Network

Although demand levels were calibrated to generate a reasonably congested network, this network included only a small percentage of trucks. To test some truck-related strategies a modified high level of truck demand was simulated. Thus, rather than bias the results of the majority of the non-truck strategies tested, an alternate baseline was generated. The modified demand produces approximately 2,500 trucks on the major southbound O-D (from 1 to 2) and 550 trucks on the major arterial southbound O-D (from 3 to 12).

User Classes

User classes are defined in terms of access to travel information. The default user class, termed the Unequipped Class, has no access to real-time information and must base all route choice decisions on the learning methodology described in Chapter 2. However, two additional user classes, the Pretrip Information Class and the En Route Information Class, were developed to establish sensitivity to real-time information-based strategies. Each of these classes has access to a snapshot of current travel times along potential routes, either before departure, as with the Pretrip Information Class, or continuously along the route, as with the En Route Information Class. Within the baseline, 98% of drivers have no access to information (Unequipped), 1% can access pretrip information only (PT), and 1% can access continuous en route information (ER).

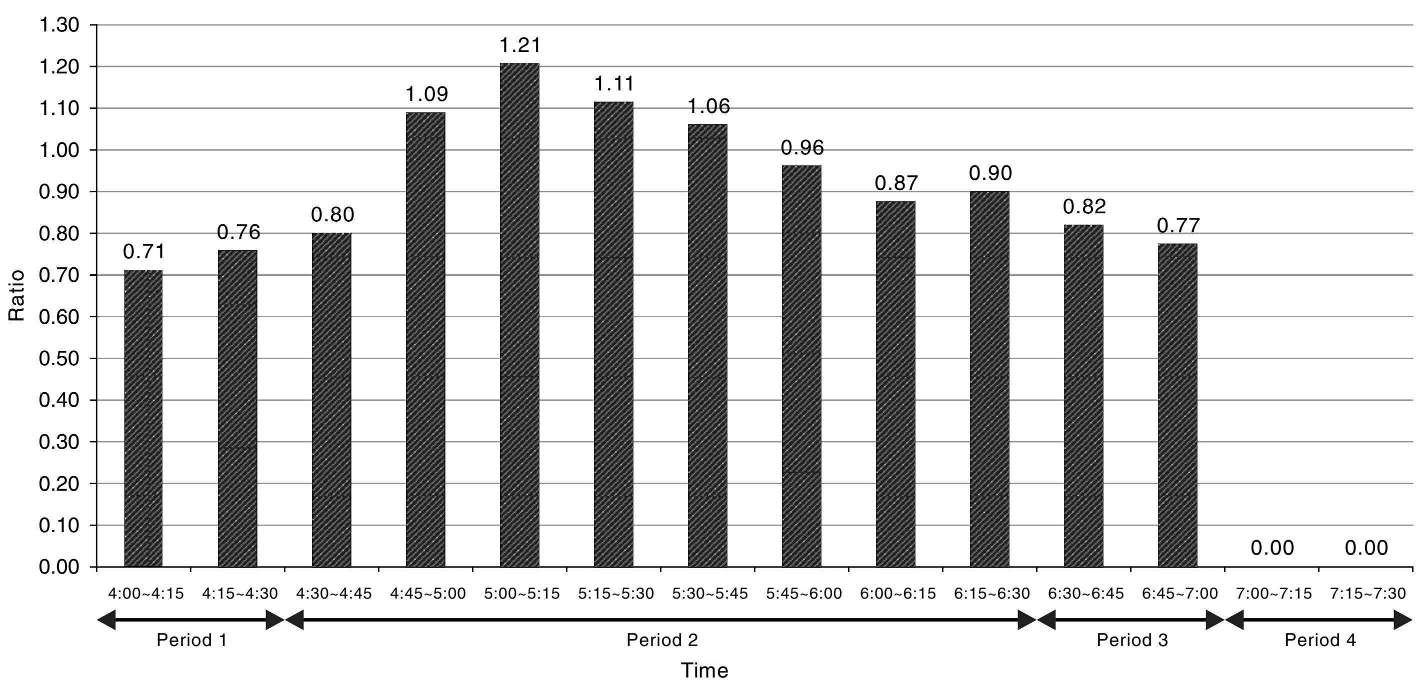


Figure 4.2. Baseline network demand profile.

The day-to-day learning procedure described in Chapter 2 was implemented within DYNASMART-P to model the ways in which drivers choose routes on a daily basis, as well as how they learn from previous travel experiences. The method establishes a minimum travel time improvement threshold, and drivers compare their default paths (originally based on minimum travel time) against travel times on alternate paths during preceding days. In order to simulate the gradual way in which drivers change their daily routine, only 15% of drivers in the Unequipped Class are permitted to have the option to switch their preferred path each day based on observations over the previous 5 days. All information users (PT and ER), on the other hand, were given the option to update their preferred paths on a daily basis. The "switching rate" refers to the actual total daily percentage of drivers who decide to take an alternate path based on this information, thus modifying their preferred path, and the variation of this value between 0% (minimum) and 17% (15% + 1% + 1% maximum) provides an indication of day-to-day network stability. Table 4.1 summarizes these thresholds.

Strategies Evaluated

A comprehensive review was undertaken of technologies affecting traffic operations. An inventory was also developed of operational, design, and technological strategies and tactics for achieving improvements in sustained service rates for freeways and/or arterial segments. Nearly 100 strategies and tactics were identified. A subsequent ranking process accounted for the effect of each strategy/tactic on road segment capacity; whether the particular strategy/tactic is in the purview of decision makers to implement; and barriers to actual implementation. On this basis, the approximately 100 initially identified strategies and tactics were distilled to the 25 presented in Table 3.2 and considered to be most promising for actual application.

Measure of Effectiveness

Each of the strategies tested represents an attempt to mitigate congestion and improve the productivity of the network, but not all could be expected to generate results significant at the network level. It was therefore necessary to compare results at a relative scale of interest. For example, with approximately 25% of the networkwide demand occurring on the southbound freeway facility, it was reasonable to examine networkwide results for many of the freeway-based strategies. However, many of the effects of the arterial strategies were only significant at the corridor level, requiring analysis at this scale. Some strategies dealt with individual intersections, making link-based analysis most appropriate. Therefore, although the baseline provided a standard set of comparison results, the scale of analysis was determined on a case-by-case basis. A variety of performance measures are available at the link, corridor, O-D, and networkwide levels from DYNASMART-P output. Additionally, freeway bottleneck summary information is provided as a potential diagnostic tool for practitioners.

Link-Level Performance Measures

For each link (i) in DYNASMART-P, the following MOEs are reported for every 15-minute interval over all simulation days:

- Link vehicle count (veh/15 minutes);
- Average travel time (minutes);
- Space mean speed (mph);
- Vehicle density (veh/mi/lane);
- Queue length (defined as the ratio of vehicle queue length to the link length);
- Freeway link breakdown indicator (for a freeway link, if at the end of every 15-minute interval, there is a queue, 1; otherwise, 0);
- Number of cycles with a queue at the start of the red phase (exclusively for signalized links); and
- Link capacity (veh/h/lane).

Corridor-Level Performance Measures

Several corridor MOEs such as density (veh/mi/lane), space mean speed (mph), queue length on corridor, travel time, and breakdown count or cycle failure count for a user-specified

Table 4.1. User Class Settings

	User Class	Unequipped	PT	ER
Day-to-day learning	Percentage of total vehicles	98%	1%	1%
	Daily learning rate (maximum switching %)	15%	100%	100%
	Daily learning improvement threshold (minutes)	5 minutes	5 minutes	5 minutes
	Maximum daily learning percentage	17%		
Daily path selection	Pretrip path improvement threshold (minutes)	—	2 minutes	—
	En route path improvement threshold (minutes)	—	—	2 minutes

corridor can be created by aggregating one or more link-based MOEs.

Network-Level Performance Measures

At the network level, DSP generates average travel time (minutes/veh) for each of the vehicle/user subgroups listed below:

- Networkwide (all vehicles);
- Low-occupancy vehicle (LOV) group;
- High-occupancy vehicle (HOV) group;
- Unequipped vehicle group (vehicles that have no access to pretrip or en route information);
- Pretrip information vehicle group;
- En Route information vehicle group; and
- Critical O-D vehicle group.

The average travel time for each subgroup on a single simulation day is also reported for the entire simulation period. Users may also evaluate this MOE over multiple simulation days based on their needs.

Simulation Procedure

In order to appropriately compare a baseline (unmodified network which represents current condition) to a strategy application case (modified network with enhancements), the user should follow these steps:

- Simulate the baseline network for 200 days by using the baseline O-D Demand Matrices.
- Export vehicle and path information.
- Simulate the baseline network for an additional 50 days by using the vehicle and path files exported from the original run.
- Simulate the modified network for an additional 50 days by using the vehicle and path files exported from the original run.
- Compare the baseline and the strategy results for the final 20 days of the simulation period.
- Subsequent applications suggest that this analysis time horizon is more than adequate even for larger networks, but the number of simulation days is a matter that can be individually judged at the time of each application.

Definition of Simulation Analysis Regimes

Figure 4.3 illustrates the defined simulation regimes that are defined to compare the results appropriately. The figure shows daily networkwide average travel time and daily route switching rate from the baseline and the strategy results.

Regime I: Baseline Stabilization Period

Before strategy testing can begin, the model must first allow drivers to learn the network and establish their preferred paths, just as anyone starting a new job may try several different routes initially in search of their preferred route to work. The daily switching rate, as noted, provides an indication of the overall stability of the network. The network was therefore allowed to run until relatively stable conditions were established, characterized by no observable general trend in the daily switching rate. Using 200 days provides a conservative estimate of this value, as shown for Regime I in Figure 4.3.

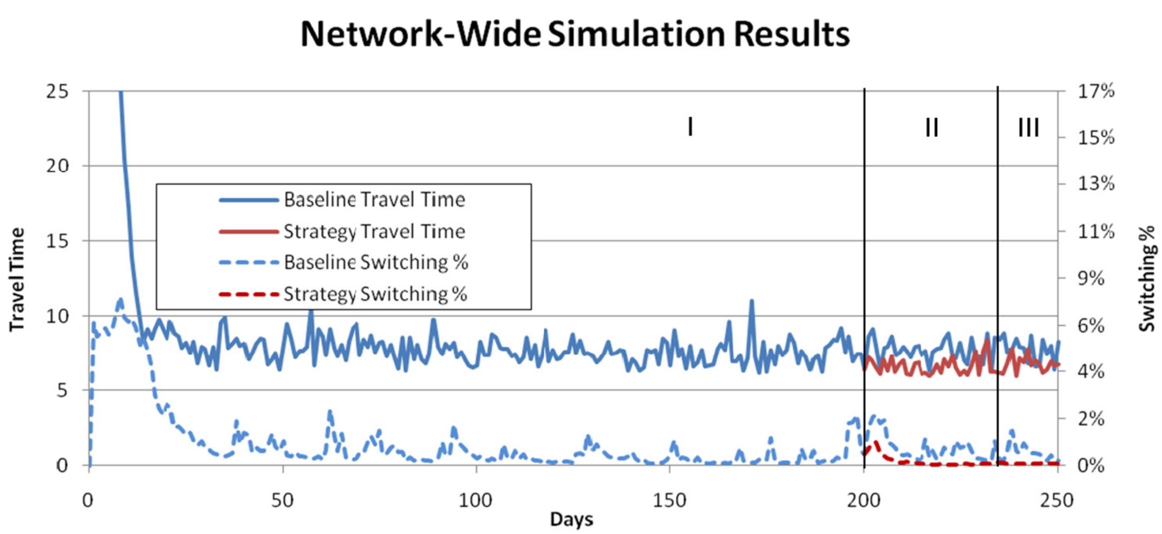


Figure 4.3. A 250-day simulation process.

Table 4.2. Baseline Networkwide Summary Statistics (Days 231–250)

Class	Overall	Unequipped	Pretrip	En Route	LOV	HOV	Critical O-D (1→2)
Average travel time (minutes)	7.49	7.50	6.94	7.00	7.49	7.53	9.58

Regime II: Strategy Stabilization Period

To test the effects of any type of network modification, the user must (a) allow drivers a sufficient number of learning days to adjust to the modified network and learn new paths as necessary before collecting statistics and (b) compare these results against a set of baseline runs generated by using the same sequence of random numbers. To do so, the user must first load the initial 200-day baseline results and change the settings from O-D-based vehicle generation to path-based vehicle generation to simulate the network by using the learned paths from the baseline stabilization period.

Regime III: Results Comparison Period

By simulating a modified network in parallel to the unmodified baseline network for 50 days, as discussed above, the final 20 days can serve as the stabilized comparison period that follows the same random number sequence, thus allowing for a correct protocol for comparison. As such, all strategy tests should follow the same procedure outlined above, comparing the final 20 days of simulation (Days 231 to 250) against the corresponding baseline simulation days.

Results and Key Insights

Baseline Results

The simulation results from the baseline, especially link volumes and speeds, can be used for the initial network calibration. Networkwide travel time can also be used for examining whether the developed network produces reasonable results. Networkwide average summary statistics for Days 231 to 250 of the baseline user class are shown in Table 4.2.

Link-based simulation performance measures from the baseline such as speed, density, and queue length can be used for diagnosing the network to find a potential starting point for strategy application. In addition to those, DYNASMART-P provides a bottleneck diagnosis function and statistics on the number of cycles with a residual queue at the start of the red phase on signalized links. Both can be reported in text output and in a visualized map as well. The results from the bottleneck diagnosis function are shown below. For each freeway, on-ramp, and off-ramp link, the following information is reported for each day:

- Total number of bottleneck-delayed vehicles (veh);
- Total bottleneck-caused delay (minutes); and
- Average bottleneck delay per vehicle (minutes).

Table 4.3 provides a 20-day average (Days 231 to 250) of the five bottlenecks that generate the largest amount of delay in the baseline network, with the location of each depicted in Figure 4.4. To improve these bottleneck sections, some strategies were implemented and evaluated.

Assessing the Effectiveness of a Freeway Reversible-Lane Strategy

In the course of the research, many of the 25 strategies presented in Table 3.2 were implemented in the Fort Worth network to assess their effectiveness in maintaining SSRs for freeways and/or arterial segments. In this section, the freeway

Table 4.3. 20-Day Average of Baseline Bottleneck Information (Days 231 to 250)

Bottleneck ID	Link Number	Type	Number of Vehicles Delayed	Total Delay (minutes)	Average Delay (minutes)	Frequency of Activation (% of 20 Days)
1	69	Freeway	3,907	14,063	1.90	100
2	89	Freeway	3,858	14,053	1.86	85
3	94	Freeway	4,457	10,605	1.77	100
4	86	Freeway	2,348	8,568	1.56	75
5	117	Freeway	1,618	5,585	1.04	65

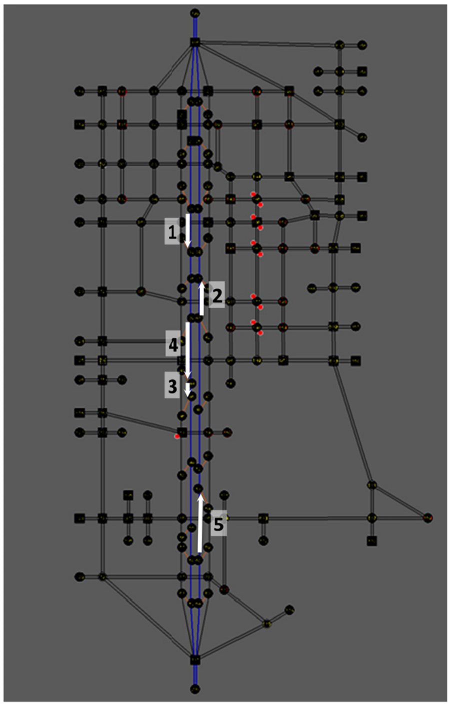


Figure 4.4. Visualized bottleneck locations.

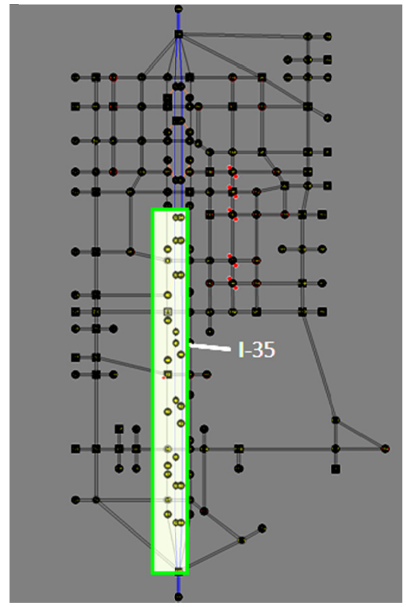


Figure 4.5. Reversible-lane implementation in Fort Worth network.

reversible-lane strategy was selected to demonstrate how to implement strategies in a DYNASMART-P network and what kind of results can be extracted from the simulation outputs. In order to measure the effects of the strategy, results from the baseline and strategy are compared.

Reversible lanes, or counterflow lanes, are lanes that allow traffic to flow in either direction through the use of dynamic message signs or movable barriers. A reversible lane is typically used to improve traffic flow during peak periods with highly directional flow. The team implemented this strategy to improve the apparent freeway bottleneck in Sections 1, 3, and 4, which are illustrated in Figure 4.4.

The I-35 West facility in the Fort Worth network has four lanes over the entire 3.1-mi corridor, as highlighted in Figure 4.5. It represents the peak direction of travel during the evening peak period. The corridor shown includes three on-ramps and five off-ramps. First, a lane was added to each of the nine links in the southbound direction so that reduction factors within DYNASMART-P could be used to simulate the opening and closing of lanes. Nine southbound links were modified to simulate a reversible-lane scenario with one northbound lane converted to a southbound lane during the peak hour. This equates to a 25% capacity reduction during the peak hour in the northbound direction, since four travel lanes are reduced to three by this strategy. The lane reduction in the off-peak northbound direction begins at Minute 45, continuing through Minute 135. The corresponding lane addition (lane opening) occurs between Minutes 60 and 120, where the reduction factor is removed to allow five full lanes of travel. Two 15-minute periods before and after the reversible-lane operation was to take effect were considered as the clearance time periods.

Network-Level Results Comparison

At the network level, although there was an average reduction in total breakdown count, bottleneck vehicle count, and vehicle delay caused by bottlenecks, overall average travel time actually increased by about 6%. Although networkwide travel time is highly variable due to the use of stochastic capacity, this travel time increase may suggest that the benefit to the southbound direction of an additional travel lane was outweighed by the substantial disbenefit caused to traffic in the northbound direction. The bulleted list that follows summarizes these results, with the number of standard deviations of improvement shown to the right. The vertical bar (|) represents the performance of the baseline. The star (*) represents the standard deviation of improvement. The number of stars represents the size of the improvement in terms of the number of standard deviations of the performance measure. If stars show up on the right side of the bar, it means the performance measure associated with the strategy shows an improvement. If they show up on the left side of the bar, it means the performance measure has degraded.

The average and 95th percentile values for several MOEs of interest are depicted in Table 4.4.

Table 4.4. Comparing Reversible-Lane Performance at the Network Level

	Baseline		Reversible-Lane Strategy		
MOE	Average	95th Percentile	Average	95th Percentile	Change in Average (%)
Travel time (minutes)	7.72	9.16	8.22	9.23	6.4
Breakdown count	24	29	18	23	–25.2
Bottleneck vehicle count	26,865	35,784	25,704	33,124	–4.3
Bottleneck delay (hours)	1,314	2,456	1,293	2,111	–1.6

Note: MOE = measure of effectiveness.

Network Performance:

- Travel time *
- Breakdown count **
- Bottleneck vehicle count
- Bottleneck delay

Corridor-Level Results Comparison

At the corridor level, the additional southbound capacity generated an increase in processing efficiency during the peak 15 minutes in the southbound direction and attracted a greater number of vehicles from adjacent routes to the freeway corridor during the peak period. All primary performance measures showed an improvement from this strategy, including reductions in travel time, density, queue length, breakdowns, and bottleneck delay.

Despite these gains, service degradation in the northbound direction was both significant and severe. Speed decreased significantly, corresponding to a significant increase in travel time and density. In short, the northbound direction did not have enough spare capacity to sacrifice a full lane to the southbound direction during the peak hour, leading to a significant decrease in quality of service with the reversible-lane in operation. The bulleted list below summarizes these results. Average and 95th percentile values are presented for each MOE of interest in Table 4.5.

Table 4.6, Table 4.7, Table 4.8, Figure 4.6, and Figure 4.7 provide illustrations of speed by link along the corridor. It should be noted that bottleneck vehicle counts and bottleneck delay are summarized for the entire simulation period, while other MOEs are presented only for the peak 15 minutes.

Southbound Direction:

- Speed *
- VMT *
- Travel time *
- Density *
- Queue length *
- Breakdown count *
- Bottleneck vehicle count *
- Bottleneck delay *

Northbound Direction:

- Speed ***
- VMT **
- Travel time *
- Density
- Queue length *
- Breakdown count
- Bottleneck vehicle count *
- Bottleneck delay *

Table 4.5. Reversible-Lane Peak 15-Minute Results (Freeway Corridor: Southbound)

	Baseline		Reversible-Lane Strategy		
MOE	Average	95th Percentile	Average	95th Percentile	Change in Average (%)
Speed (mph)	56.94	62.53	63.56	64.12	12
VMT (veh*mi/15 minutes)	5,153	5,787	5,809	6,109	13
Travel time (minutes)	5.17	10.61	3.70	3.72	–28
Density (pc/mi/lane)	31.48	59.08	19.62	20.80	–38
Queue length (mi)	0.15	0.36	0.04	0.09	–73
Breakdown count	0.95	2.00	0.15	1.00	–84

Table 4.6. Reversible-Lane Peak Daily Bottleneck Information (Freeway Corridor: Southbound)

MOE	Baseline		Reversible-Lane Strategy		
	Average	95th Percentile	Average	95th Percentile	Change in Average (%)
Bottleneck vehicle count	7,266	12,858	1,073	4,981	–85.2
Bottleneck delay (hours)	391	1,541	32	134	–91.7

Table 4.7. Reversible Lanes Peak 15-Minute Results (Freeway Corridor: Northbound)

MOE	Baseline		Reversible-Lane Strategy		
	Average	95th Percentile	Average	95th Percentile	Change in Average (%)
Speed (mph)	56.19	61.30	47.10	53.81	–16
VMT (veh*mi/15 minutes)	5,454	5,669	4,940	5,271	–9
Travel time (minutes)	6.08	10.60	7.98	12.86	31
Density (pc/mi/lane)	39.53	69.61	44.89	70.05	14
Queue length (mi)	0.33	0.57	0.17	0.32	–48
Breakdown count	1.10	3.00	1.15	2.00	5

Table 4.8. Reversible-Lane Peak 15-Minute Results (Freeway Corridor: Northbound)

MOE	Baseline		Reversible-Lane Strategy		
	Average	95th Percentile	Average	95th Percentile	Change in Average (%)
Bottleneck vehicle count	7,483	13,171	14,515	24,724	94.0
Bottleneck delay (hours)	440	1,251	905	1,661	105.6

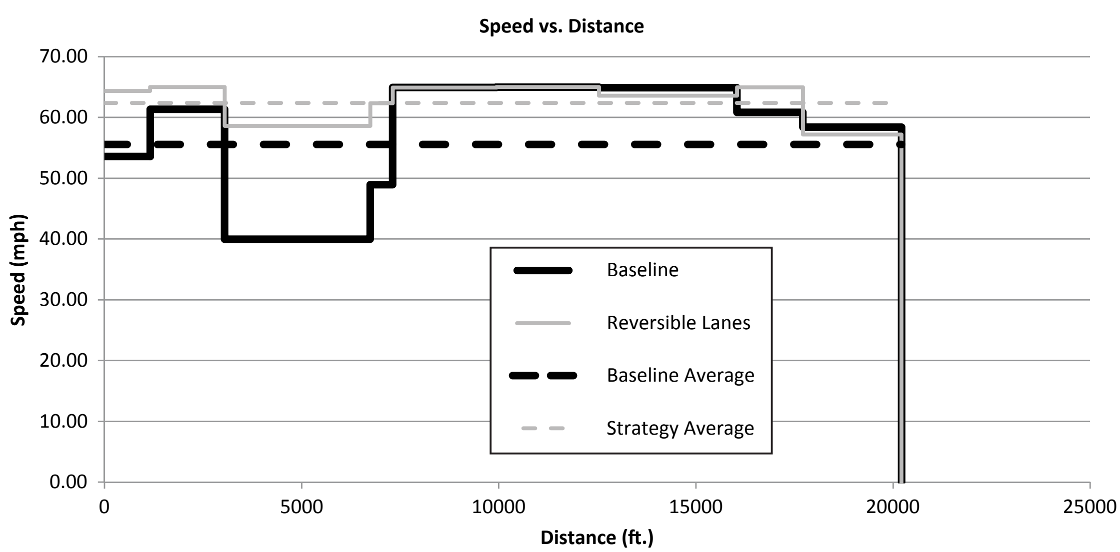


Figure 4.6. Southbound freeway corridor peak: 15-minute speed versus distance.

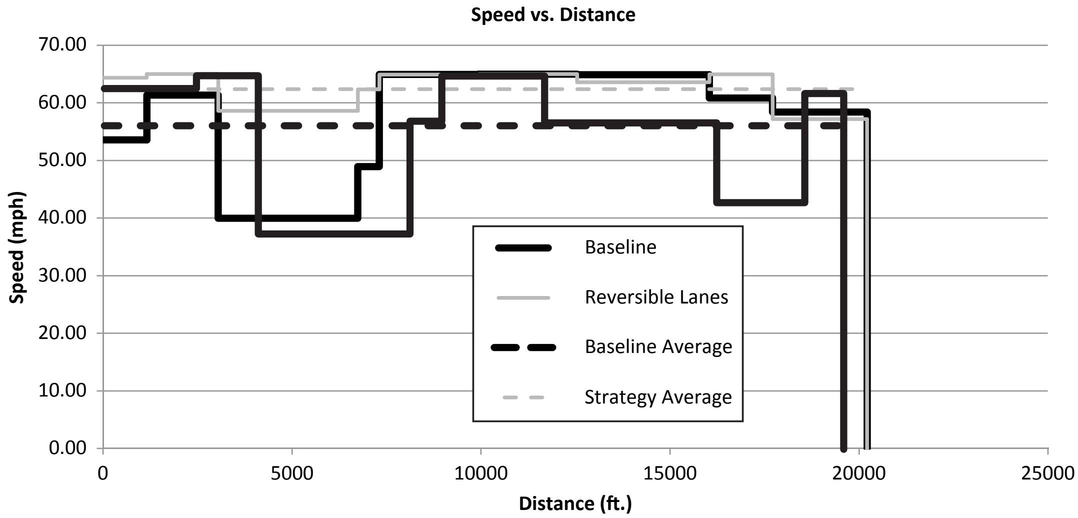


Figure 4.7. Northbound freeway corridor peak: 15-minute speed versus distance.

Thus, for this particular network, directional peaking along the freeway was not significant enough to warrant the removal of a lane from the off-peak direction to provide additional capacity to the peak direction. It should be noted, however, that the strategy did provided significant benefits to the peak direction, and for cases with lighter flow in the off-peak direction, or where the peak direction is highly critical for network performance, the strategy could have proved to be successful. Thus, the context in which the strategy is applied can have a significant effect on its overall effectiveness.

Strategies Involving Equivalent Physical Capacity Addition

Concept

In addition to comparing measures of effectiveness across multiple strategies, many practitioners would find it especially useful to be able to compare operational strategies to a typical lane addition scenario. In this manner, they would be able to determine the "equivalent capacity gain" of the candidate strategy. Figure 4.8 illustrates one possible way to do this through a hypothetical relationship between lane miles added to a network and performance, measured in terms of total network travel time. It may be useful to develop such a relationship from a variety of travel demand, *HCM,* or simulation models for a given network case study as lane mile additions represent the traditional way of increasing network capacity. It is important to note, however, the relationship shown in Figure 4.8 does not take into account the effect of latent demand on performance, nor is it necessarily continuous. Both of these issues are worthy of further exploration. Even so, the ability to implement the concept presented in Figure 4.8 would represent a significant step forward for the transportation profession and would greatly improve the quality and impact of information available to transportation investment decision makers. A set of nonconstruction improvements that reduces network travel times from A to B in Figure 4.8 is effectively equivalent to adding D minus C lane miles, or a construction-based capacity increase of (D – C)/C%. A method for undertaking the implementation of this concept is described in the following paragraphs.

Implementation

In order to illustrate the equivalent capacity gain concept for a few selected strategies, three physical capacity addition scenarios were applied to the primary freeway corridor

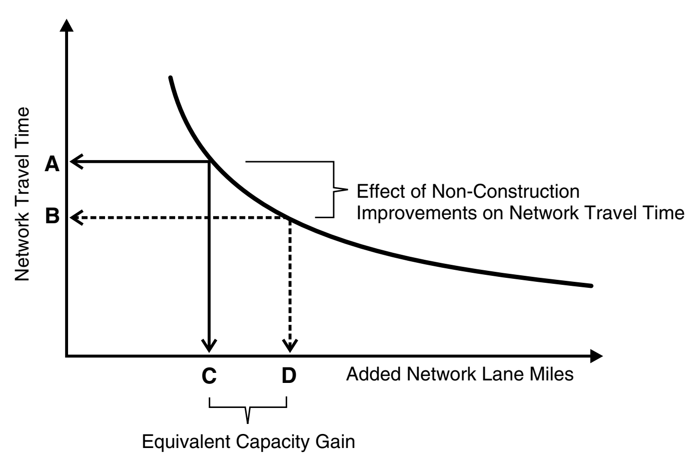


Figure 4.8. Equivalent capacity gain concept.

and tested in DYNASMART-P. As depicted in the map in Figure 4.9, lanes were added on I-35 Southbound, the primary path for O-D pair 1-2. The four illustrations in Figure 4.9 show the baseline condition and the locations of the lane additions for each of the three lane addition scenarios, A, B, and C. Scenario A involved adding one lane to the existing four lane segments, creating a continuous five-lane freeway corridor with no lane additions or reductions. In Scenario B, one complete lane was added throughout the entire corridor, maintaining the 0.9-mi segment with an additional lane. In Scenario C, the freeway was converted to a six-lane facility over the entire length of the corridor. The total lane miles added for Scenarios A, B, and C were 7.6, 8.5, and 16 lane miles, respectively. This is equivalent to 22%, 24%, and 46% lane mile increases for Scenarios A, B, and C, respectively.

Four operational improvement strategies were selected for comparison, involving three separate enhancement categories: pretrip information (technological), en route information (technological), additional narrow lanes in the existing cross section (design), and reversible lanes (operational) for improving identified bottleneck Sections, 1, 3, and 4 in Figure 4.4. The following sections present the comparison results associated with this particular application of the identified improvement strategies and are of course subject to the caveat that the learning model and strategy effectiveness models are representative of real-world behavior. Thus, while the general trends will hold, specific MOEs may be subject to some variation.

Day 1 Results

Examination of Day 1 results represents a traditional evaluation method for strategy testing with no learning or changes in demand. In other words, the examination of Day 1 results gives the analyst the opportunity to view the consequences of implementing the improvement strategies before drivers have a chance to respond either temporally or spatially, similar to what a highway capacity analysis procedure would yield.

For obvious reasons, O-D pair 1-2 was selected for the travel time analysis. As shown in Figure 4.10, the ATIS pretrip information and ATIS en route information strategies generated the same results as the baseline scenario, as both strategies rely on changes in route selection over time to have any meaningful effect. The remaining strategies exhibited significant improvement over the baseline scenario, with approximately a 5-minute travel time reduction for each. As capacity increased along the corridor for each strategy with no increase in demand, all vehicles were able to travel at free-flow speed, generating approximately the same results for each strategy.

Days 31 to 50 Results

A significant contribution of this research to the analysis of operational strategies was the addition of a driver learning algorithm which was implemented within the DYNASMART-P environment. Drivers are therefore able to respond to changes in the network and gradually learn the most efficient routes for a given O-D and departure time. To allow drivers time to respond to changes in the network, each scenario was

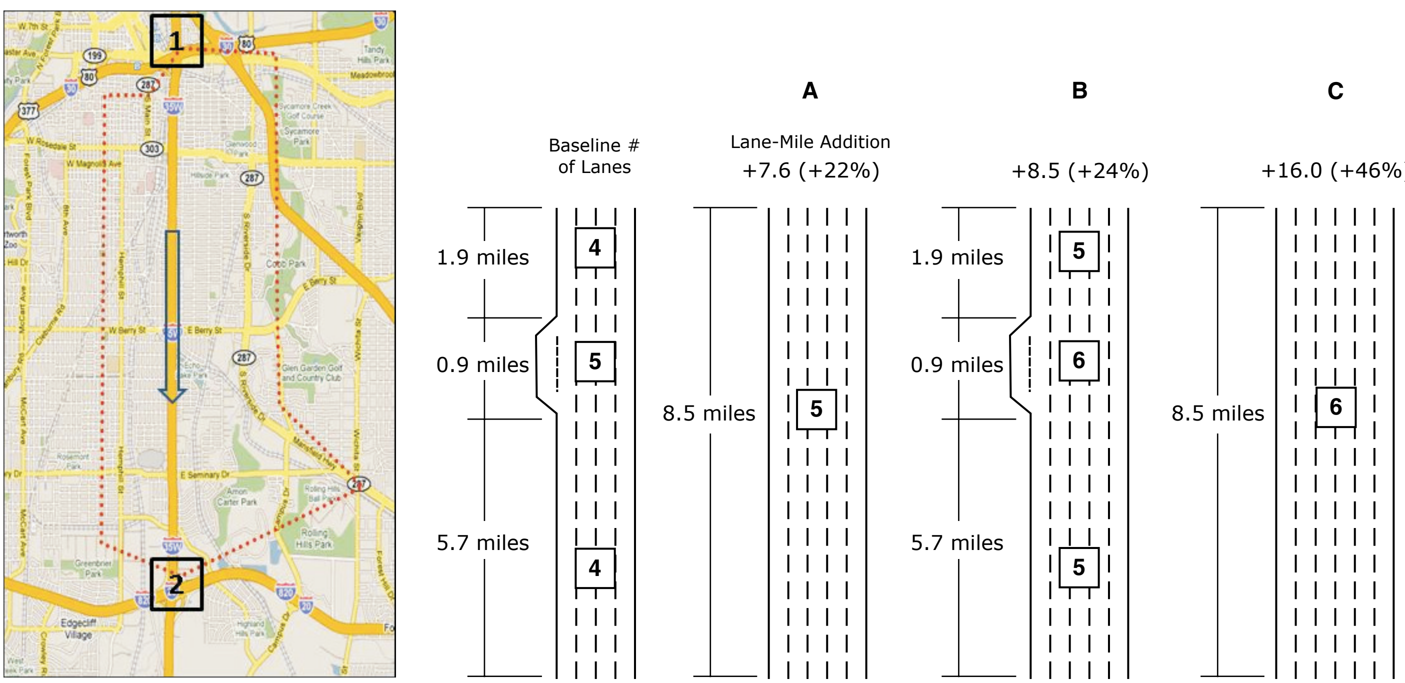


Figure 4.9. Illustrations of three physical lane addition scenarios.

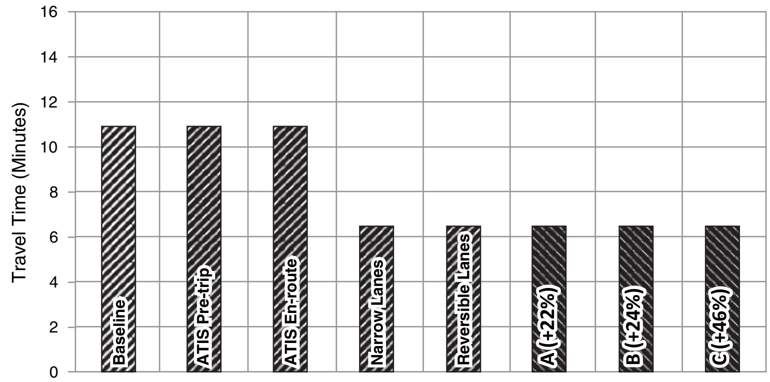


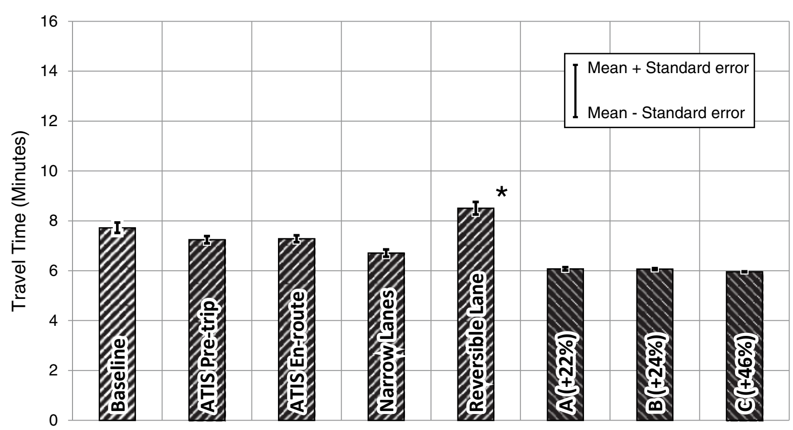
Figure 4.10. Mean travel time-Day 1-primary O-D 1-2.

simulated for a total of 50 days. The results depicted in the following exhibits represent average values from Days 31 to 50 to take into account supply stochasticity. Standard error bars are shown where appropriate. The following MOEs were selected for comparison:

- Networkwide
    - Mean travel time.
- Primary O-D (1–2)
    - Mean travel time;
    - 95th percentile travel time;
    - Travel time index = mean travel time/free-flow travel time; and
    - Buffer index = (95th percentile travel time – mean travel time)/mean travel time.

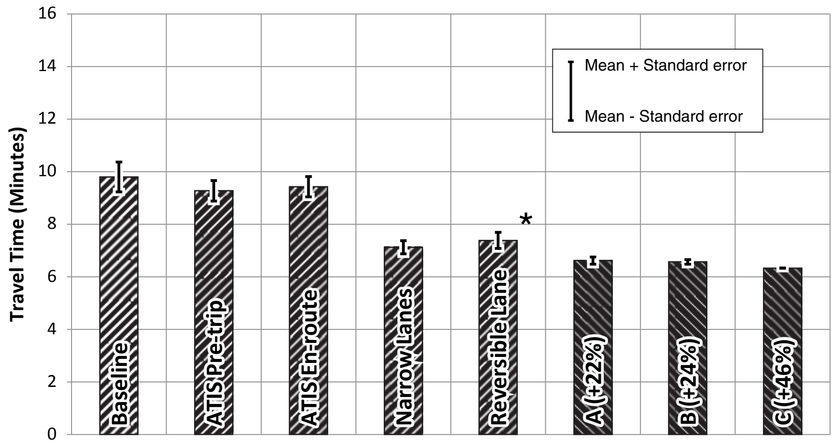
Figure 4.11 shows the networkwide mean travel times for all scenarios. Although both ATIS strategies reduced networkwide travel times, the narrow lanes strategy was the most effective among the nonconstruction alternatives. In all three cases however, each of the lane mile addition scenarios proved to be a more effective alternative. The reversible-lane strategy, on the other hand, increased networkwide travel times largely due to added congestion in the northbound direction caused by the peak-hour lane reversal. It is important to note that in addition to travel time savings, each of the three effective strategies also decreased the travel time variability on the primary O-D. This concept of reliability is of critical importance to the commuter, and an increase in reliability is a benefit largely ignored when only measuring average travel times. For each of the lane mile addition scenarios, reliability was increased to the point that variability in travel times all but disappeared entirely.

Figure 4.12 shows travel times along the primary O-D (1-2) for each of the evaluated scenarios. All strategies reduced O-D travel times relative to the baseline, including the reversible-lane strategy, as the primary O-D includes only travel times in the peak (southbound) direction. Overall, the narrow-lane strategy was the most effective in reducing travel times in the primary O-D, although the strategy was slightly less effective



Note: asterisk = accounts for added congestion effect on northbound freeway corridor due to lane reversal.

Figure 4.11. Average of Days 31 to 50 networkwide travel time.



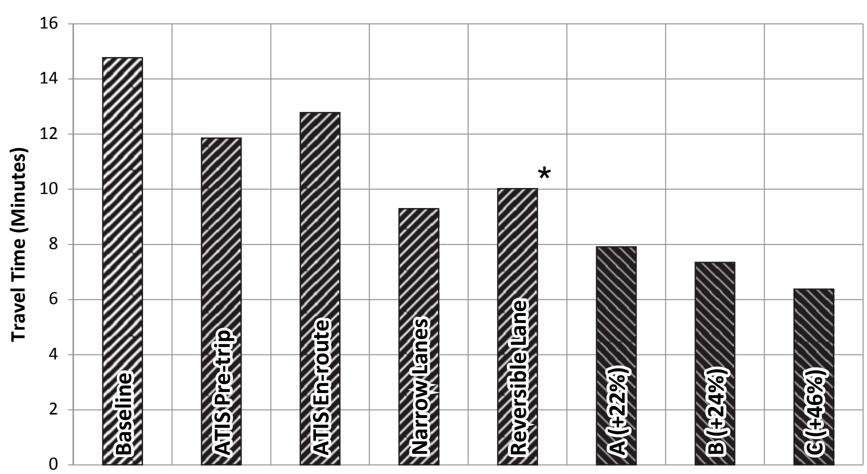
Note: asterisk = effects in peak direction with lane addition only.

Figure 4.12. Travel times for trips serving primary O-D 1-2.

than lane mile addition Scenario A. Error bars indicate that each of the strategies decreased the travel time variability for the primary O-D, indicating an overall increase in reliability. The 95th percentile travel times for the primary O-D, as shown in Figure 4.13, also demonstrate this trend.

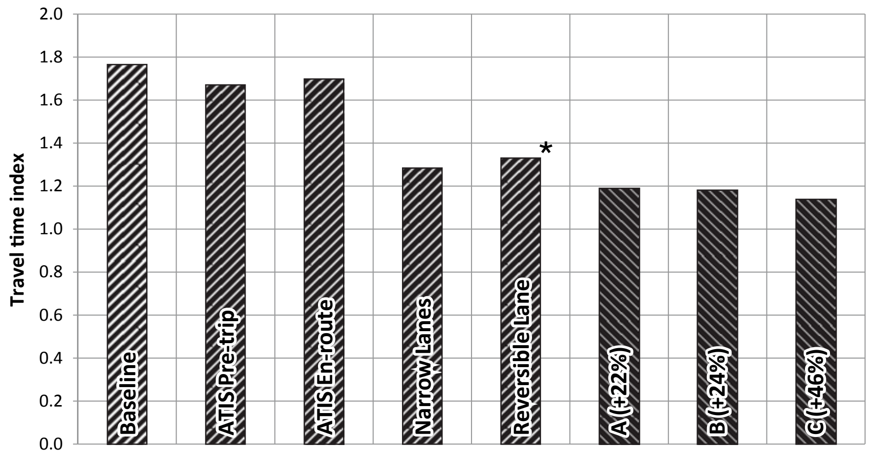
Figure 4.14 provides the travel time index for each of the evaluated scenarios on the primary O-D. The baseline scenario has a travel time index of around 1.7, which is to say that average travel times for the primary O-D were approximately 70% higher than under free-flow conditions. The travel time index decreased for each of the scenarios, and the narrow-lane strategy was able to decrease travel times to less than 30% greater than free-flow travel times during the peak hour. Although each of the lane mile addition scenarios generated lower travel time index values than the narrow-lane strategy, even the most expensive option (Scenario C) was only able to reduce the travel time index to slightly less than 1.2, highlighting how the method could be used in a cost-benefit analysis of potential strategies.

The buffer index shown in Figure 4.15 provides one method of evaluating travel time reliability for a given O-D or corridor of interest. Small buffer index values indicate low variability of travel times, or high reliability. As discussed previously, the buffer index clearly demonstrates that each of the scenarios led to an overall increase in reliability. Interestingly, however, the buffer index reveals that the ATIS pretrip strategy was the most effective of the four nonconstruction alternatives at increasing travel time reliability. Such an analysis allows the practitioner to evaluate strategies based on a number of potential objectives and perform a network-appropriate customized approach to congestion mitigation.



Note: asterisk = effects in peak direction with lane addition only.

Figure 4.13. 95th percentile travel times for trips serving primary O-D 1-2.



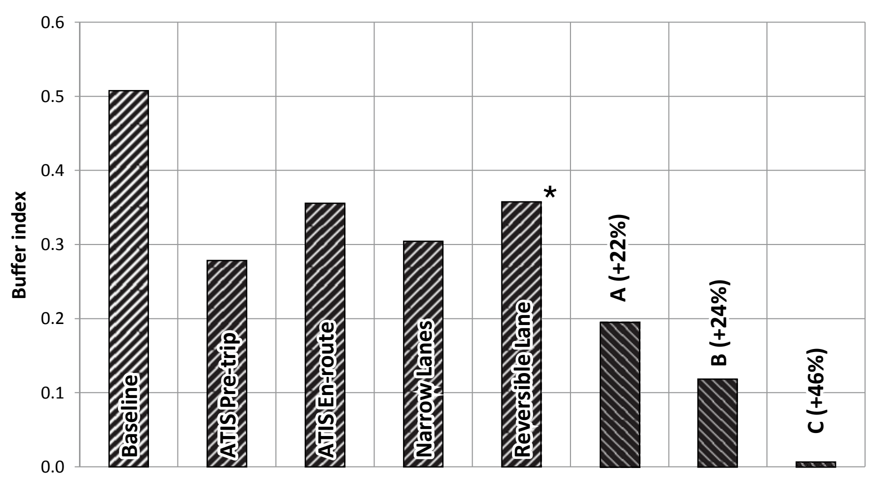
Note: asterisk = effects in peak direction with lane addition only.

Figure 4.14. Travel time index for trip serving primary O-D 1-2.

Limitations and Cautions

The enhanced DTA model developed through this effort and described in the preceding chapters for a test network provides a practical methodology for assessing the ability of various operational strategies, either singly or in combination with one another, to forestall or eliminate the need to construct additional lane miles of capacity within a transportation network. The methodology provides effectiveness assessments about both travel time and reliability at the link, corridor, and network levels. It is already integrated into two dynamic traffic assignment (DTA) modeling procedures (including the open source program DTALite as well as DYNASMART-P) and can be integrated into others as well.

Unfortunately, the usefulness of the methodology is constrained by the fact that the modeling environments in which it operates do not account for crashes, weather, and other events that can cause nonrecurring congestion. Nonrecurring congestion represents a significant part of the delay and frustration experienced by travelers and therefore has a substantial impact on both travel time and reliability. Yet, traditional operational models do not typically account for its effects. Additionally, the overall effectiveness of many operational strategies, such as those evaluated under this research effort, is misrepresented if only the recurring congestion effects are considered. As an example, the use of narrow lanes as an operational improvement strategy will typically result in more capacity, reduced travel time, and improved reliability in an environment where incident effects are not considered. But narrow lanes might also increase the potential for crashes, thereby reducing their effectiveness if nonrecurring congestion effects are taken into account. As another example, ramp metering does not add much in terms of capacity, but it



Note: asterisk = effects in peak direction with lane addition only.

Figure 4.15. Buffer index for trips serving primary O-D 1-2.

stabilizes flow, lowers the probability of breakdown during demand surges, and reduces the potential for crashes. In this case, the benefit of ramp metering will be underestimated unless nonrecurring congestion effects are also considered. Therefore, a comprehensive assessment of the overall effectiveness of these operational strategies needs to account for their effects on both recurring and nonrecurring congestion.

Much value will be gained from future efforts that extend the capability of the DTA methodology to include an assessment of nonrecurring congestion effects, incorporating this additional capability into at least one functional modeling environment.

Illustrative Application of Methods, Metrics, and Strategies

The overall net impact of a combination of operational improvement strategies deployed within a subarea or network will almost always be different from the sum of their individual effects, due to the influences each implemented strategy has on the others. The ability of a mesoscopic model to take these interactions into explicit account is one of the important advantages that come from implementing the methodological enhancements of this project in such an environment.

Combinations of various operational improvements strategies identified in this project were applied to a subarea of the Portland network in order to demonstrate the feasibility of this approach. In the process, useful insights were obtained that may have application in a broader context.

The subarea network selected for these purposes is identified in Figure 4.16. It was selected for this prototype application for a number of reasons:

- It is relatively large in size and therefore represents a good opportunity to test scaling issues associated with the method applications.
- It includes both north-south and east-west freeway segments, as well as multiple interchanges on both segments.
- It includes a full range of arterial streets that are generally organized around a grid pattern. Consequently, there are good opportunities within this particular subarea to conduct

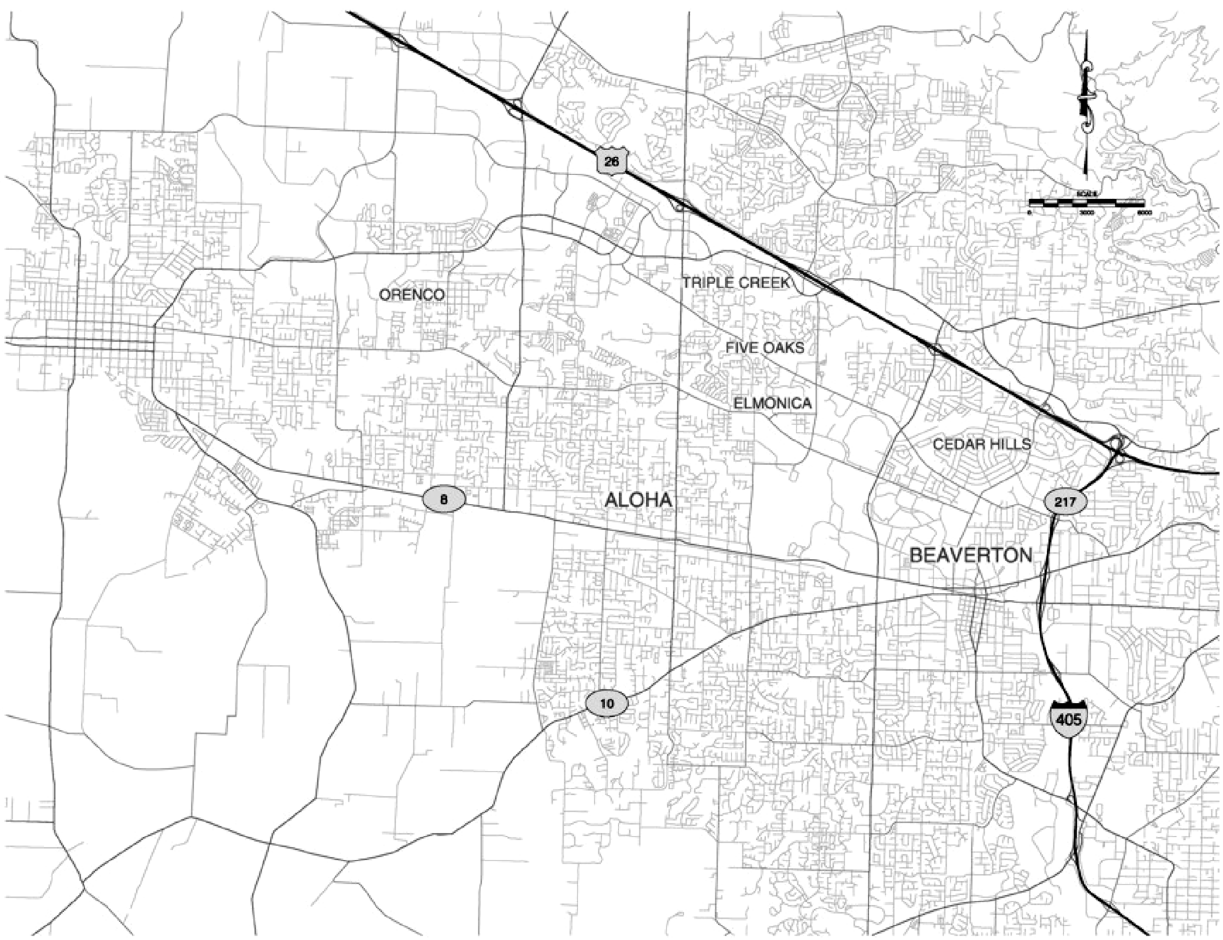


Figure 4.16. Portland network study area.

separate analyses of individual links, of separate corridors, and of the subarea network as a whole.

The area is characterized by significant congestion on both freeway and arterial segments during typical weekday evening peak hours. These conditions present opportunities to test solution alternatives ranging from new construction to a variety of operational improvement strategies.

The purpose of this application was to demonstrate, through the detail of an actual and real-world example, the manner in which each step in the procedure can be undertaken; the thought process that accompanies each step and informs decisions regarding next steps; and a representative range of findings/conclusions that can be anticipated as outcomes of the procedure. The application was developed specifically for the purposes of this report and was not used to inform actual investment decisions in the Portland metropolitan area.

There are references throughout this chapter to DTALite and DYNASMART-P, both of which are mesoscopic dynamic traffic assignment models. DTALite is a fully functional, open source dynamic traffic assignment model that can be downloaded without charge from http://sourceforge.net/projects/dtalite/. DTALite incorporates all the DTA model enhancements developed through this project. DYNASMART-P, Version 1.3.0, is available for a charge from the McTrans Center at http://mctrans.ce.ufl.edu/featured/dynasmart. Version 1.3.0 does not include any of the model enhancements developed through this project effort.

The two models were used together in the execution of this prototype application, but it should be emphasized that neither model is essential to the implementation and execution of the enhanced analytic and diagnostic methods described herein. These particular models were chosen for use in this particular application because of the research team's familiarity with them, and also to maintain consistency with the platforms used in the development and testing of the model enhancements and diagnostic tools described earlier. Other DTA models are available and could also be used in lieu of either DTALite or DYNASMART-P as the user prefers, providing the model enhancements and diagnostic tools developed within this project are appropriately integrated into them.

Description of the Subarea

The subarea that was selected for this prototype application, and illustrated in Figure 4.16, is located in the southwestern part of the Portland metropolitan area. It encompasses a fairly large area and includes facilities that have statewide, regional, and/or local significance.

- Highway 26, also known as the Sunset Highway, forms the northern boundary of the study area. It is the primary connector to an area known as the Silicon Forest for its high technology industry employment and surrounding residential population. Beyond the subarea's western boundary, Highway 26 continues westward for about 70 miles to the Oregon Coast and therefore is used extensively for freight movement as well as tourism and recreation. It is classified as a road of statewide importance because of the scope and scale of activities that rely on it.
- The eastern boundary of the subarea is defined by a portion of Highway 217, which connects between Interstate 5 (I-5) on the south and Highway 26 on the north. Therefore, in addition to serving travel demands within the corridor in which it is located, Highway 217 also represents an important connection for travelers from the south who are destined to points in the western part of the study area and vice versa.
- The southern boundary of the study area is defined by Farmington Road, which travels in a northeast-to-southwest direction and represents an effective boundary line between the western and southwestern parts of the Portland metropolitan area. Farmington Road transitions from an urban to a rural environment as it moves to the south and west, and in fact it is located outside the Urban Growth Boundary at the westernmost end of the subarea.
- The western boundary of the study area is defined by River Road, SE 10th Avenue, and NE Brookwood Parkway in such a way as to stay just to the east of the majority of Hillsboro's downtown core area and to connect between Farmington Road and Highway 26.

The interior of the subarea comprises a surface street system that includes a grid-like arterial road system for both east-west and north-south travel. Tualatin Valley Highway, also referred to as TV Highway or State Route 8, is the preferred route for east-west arterial travel and includes frequent bus service. It is also a fairly congested route, particularly during weekday peak hours.

Table 4.9 presents summary statistics about the Portland subarea network used in this application. It is a reasonably large

Table 4.9. Summary Statistics about the Portland Subarea Network

Network Characteristic	Entire Network	Subarea Network
Number of traffic analysis zones	2,013	208
Number of nodes	9,905	857
Number of links	22,748	1,999
Number of originating vehicle trips	1.2 million	212,000
Average travel time (evening peak period)	22 minutes	10.5 minutes

subarea with more than 200 traffic analysis zones (TAZs) and more than 200,000 originating vehicle trips during the 4-hour weekday time period (3 p.m. to 7 p.m.) that is analyzed.

This subarea is well suited to the purposes of this demonstration effort due to its size, the diversity and redundancy of various facility types, and the prevalence of congested conditions during typical weekday evening peak-hour conditions.

Building and Calibrating the Portland Subarea Network

An initial version of the Portland metropolitan area network was provided to the team in the same VISUM model format that Metro staff uses for their current travel demand forecasting activities. Metro staff also provided additional network detail from a variety of other sources, including signal control information that had been assembled some years earlier as part of their experimental work with the TRANSIMS model; detailed information about approach lane configuration; and the location and length of turn pockets. These data items were easily ported and transformed into a standard format that is used by multiple DTA programs (including DTALite, DYNASMART-P, and DynusT).

After the initial data sets were received and transformed, it was found to be necessary to conduct a variety of error-checking activities on the network. DTA models interact with the network differently from traditional travel demand models like VISUM, and so some network elements that do not cause problems in a VISUM analysis can still represent inconsistencies that need to be rectified for a DTA analysis. Example issues that were discovered and corrected through this process will help to clarify the kind of examination and corrective effort that is typically needed:

- Some centroid connectors from the VISUM model were found to be tied directly into real intersections. This will cause unrealistic operating conditions for the simulated intersection in a DTA model environment. To resolve this problem, the tie-in point for these centroid connectors was relocated to a mid-block location.
- Some TAZs had only one or two centroid connectors, causing too much volume to enter the network via a single street. More centroid connectors were added to resolve this problem, so the travel demand was dispersed more appropriately across adjacent road segments.
- The speed, length, and capacity of centroid connectors were sometimes defined within the VISUM model in ways that could significantly affect the overall travel time results for the network and/or corridors being investigated. The remedy to this problem was to adjust these parameters to be more representative of actual driving conditions on the local and collector street system inside each TAZ.
- The network structure at the subarea boundaries needed careful examination to ensure realistic performance characteristics during the DTA model runs. An example is shown in Figure 4.17. Here, the southernmost entry node for northbound traffic on Highway 217 initially connected to another mainline node on Highway 217, from which vehicles could either continue northbound on Highway 217 or exit onto an off-ramp (see Figure 4.17a). However, testing revealed that any backups from the off-ramp onto the mainline link would also block all northbound through traffic and cause unrealistically long delays to the through-traffic component of the entering traffic. This problem was resolved by reconfiguring the entry node so that it connected directly to two possible destinations: (a) the off-ramp and (b) the northbound through lanes on Highway 217. This is shown schematically in Figure 4.17b.

It is noteworthy that some inaccuracies (relative to true existing conditions) were purposely introduced into the network structure to ensure that this exercise is used only to

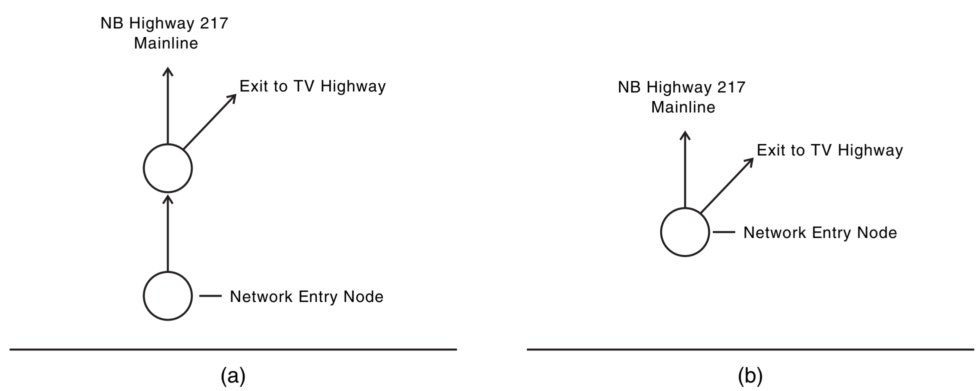


Figure 4.17. Network coding: (a) initial network coding and (b) modified network coding.

demonstrate the new methods and model capabilities. Thus, for example, the number of through lanes on some arterials was modified, as was the length of some left and right turn pockets. These inaccuracies do not affect the overall value of this demonstration exercise but do help assure that the results are not used to inform actual public investment decision making or policies.

The entire Portland area network was initially simulated by using DYNASMART-P and DynusT. This was done for a 4-hour time period (3 p.m. to 7 p.m.) across multiple days by using relatively high-end but still commonly available hardware. In particular, the hardware employed in this analysis included 64-bit machines with 16 GB of RAM and multiple processors allowing for higher-speed parallel processing. Unfortunately, the initial attempts to simulate the full Portland network by using either model were not successful on two levels.

The ratio of real time to simulated time was almost 1:1, which meant that the simulation of the 3 p.m. to 7 p.m. time period for a single day required about 4 hours of real time to complete. In the enhanced model environment where travelers adjust their path selection based on past experience, at least 35 to 50 simulated days are necessary for network travel times to stabilize around an equilibrium level. Thus, up to 200 hours (about 8 days) would have been needed to reach equilibrium for the Portland area network. This is clearly an impractical amount of time to allocate to such an activity, and particularly so in most day-to-day working environments.

Neither DYNASMART-P nor DynusT was able to successfully complete the required 35 to 50 days of simulated time, as both models crashed after only 1 to 2 days of simulated time. The reason is believed to be the large amount of data that must be carried forward in each iteration, with 1.2 million originating vehicle trips occurring on each simulated day.

An attempt was made to overcome these problems by aggregating the zones from 2,000 down to only 400 super-zones, and while this action had some beneficial effect, it still did not fully resolve either of the two issues identified above. On the other hand, it was found that both DYNASMART-P and DynusT were able to perform acceptably when modeling the smaller subarea network: a single simulated 4-hour day could be completed with DYNASMART-P in about 30 minutes by using the hardware environment described above, and in only 7 to 10 minutes when the computer processing environment was increased from two to eight cores working in parallel.

The DTALite model fared much better. Using the same hardware environment and without any aggregation of zones, a single 4-hour analysis period of simulated time for the entire Portland area network required only about 5 minutes to complete when using DTALite. At the time these investigations were ongoing, DTALite was still in some level of development and was not fully comparable to either DYNASMART-P or DynusT, particularly with respect to the manner in which signalized intersection control is modeled. These differences are only temporary and do not have much impact on the overall time requirement for the simulation. Even so, it was judged that the demonstration would be more robust and informative if consistency could be maintained with the modeling method used for the Dallas–Fort Worth network. Therefore, a two-step process was used to achieve this goal:

- DTALite was used to model the entire Portland metropolitan area network for a period of 50 simulated days.
- The results of the DTALite model were used to create an O-D matrix for the much smaller subarea network, and this became the basis for the DYNASMART-P modeling of the subarea that followed.

The following statistics provide a detailed description of the current road network attributes in the test network:

- Network data
  - Number of nodes: 858.
  - Number of links: 2,000.
  - Number of O-D demand zones: 208.
- Node control type (number of intersections including on- and off-ramps)
  - No control (on- and off-ramps): 689.
  - Four-way Stop: 0.
  - Two-way Stop: 0.
  - Signalized (actuated control): 169.
- Traffic control data for signalized intersections
  - Two-phase control intersection: 4.
  - Three-phase control intersection: 96.
  - Four-phase control intersection: 69.
  - Maximum green: 50 seconds.
  - Minimum green: 10 seconds.
  - Amber: 5 seconds.

In order to produce a realistic assessment of the entire four-hour simulation period, entering demand levels were adjusted every 15 minutes according to the Portland area demand profile shown in Figure 4.18. Here, the height of each bar represents the ratio of 15-minute demand to the original overall average.

User Information Classes

The day-to-day learning procedure described earlier in this report is used to model the ways in which drivers choose routes on a daily basis and how they learn from previous travel experiences. In this procedure, drivers have the opportunity to compare their experienced travel times against travel times on alternate paths and then make route changes for the coming days if they find they can save enough time to warrant the

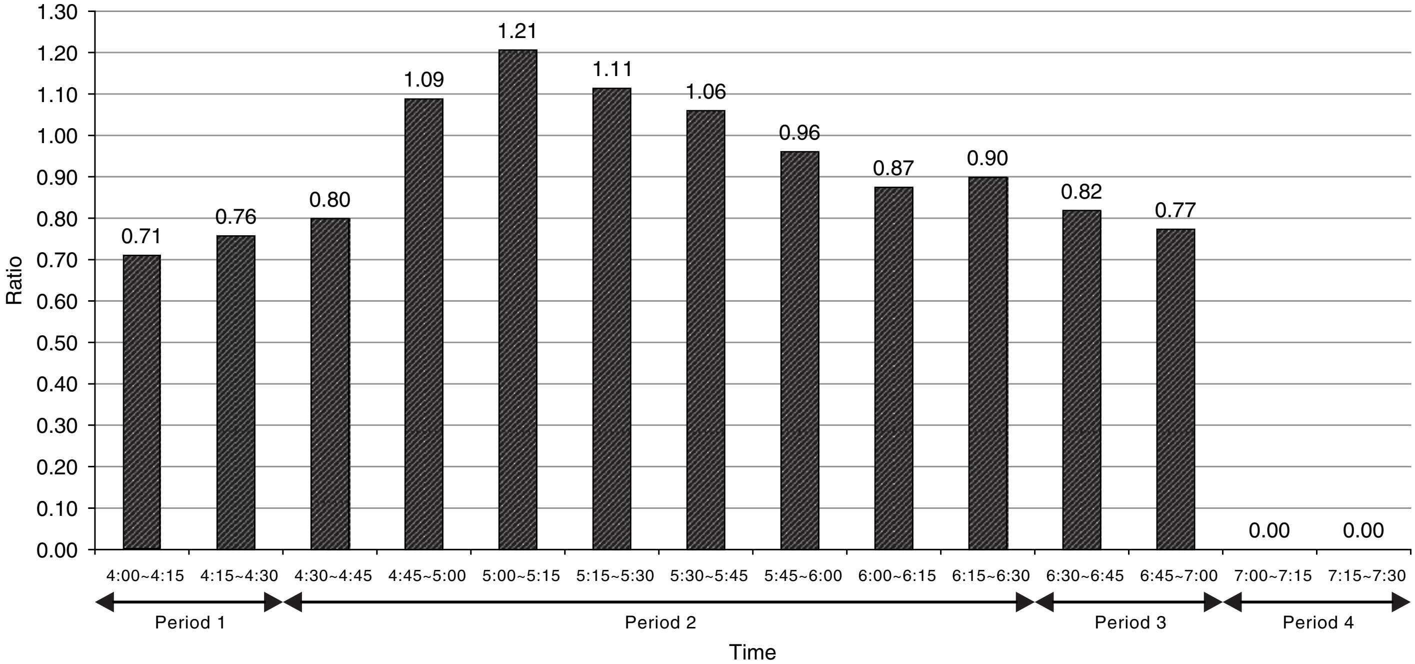


Figure 4.18. Baseline network demand profile.

switch. For this application, the percentage of drivers who could make a change in any given day was set to 15%, reflecting the reality that many people are creatures of habit and unlikely to make route changes right away, if ever. The amount of travel time drivers needed to save before they consider making a route change was set to 5 minutes. Neither of these assumptions is based on actual field data but come from the judgment of the project team members working in conjunction with the Portland Metro's planning and modeling staff.

Finally, the percentage of vehicles assigned to each of three available user classes was as follows: 98% of drivers were assumed to have no access to real-time information about network conditions (this user class is referred to as the Unequipped user class); 1% were assumed to have access to pretrip information only (this user class is referred to as the Pretrip or PT user class); and 1% were assumed to have access to continuous en route information (this user class is referred to as the Equipped or ER user class).

Simulation-Based Strategy Evaluation Procedure

A straightforward method was developed to test the effectiveness of one or more operational strategies either as standalone projects or as alternatives to traditional new construction projects.

1. First, the location of the operational strategy and/or new construction project that was to be tested was identified, and a subarea or network that appropriately surrounds the location was established.
2. Next, geometric, volume, and operational characteristics of each link within the subarea were identified and provided as inputs to the DTA model. Appropriate link, corridor, and/or network performance measures were established for subsequent evaluation purposes.
3. The DTA model was then run under three separate regimes in order to effectively use the day-to-day learning process and generate results that could be usefully compared, as shown in Figure 4.19. During the baseline stabilization period (Regime I), the DTA model was simulated for 50 days to achieve equilibrium under a baseline scenario (i.e., without any of the operational strategies or new construction projects that are to be evaluated).
4. After the baseline stabilization period was completed, the operational strategies and/or new construction projects to be evaluated were introduced into the network, and the DTA model was run for an additional 30 days of simulated time to allow driver adjustments and to achieve stable conditions under the new scenario. This is referred to as the strategy stabilization. Following immediately on this 30-day period was a simulation of an additional 20 days that formed the basis for the summary results output associated with the particular strategy being investigated.

Adherence to this methodology provided good insights into the effectiveness of the operational strategies and new construction projects being evaluated.

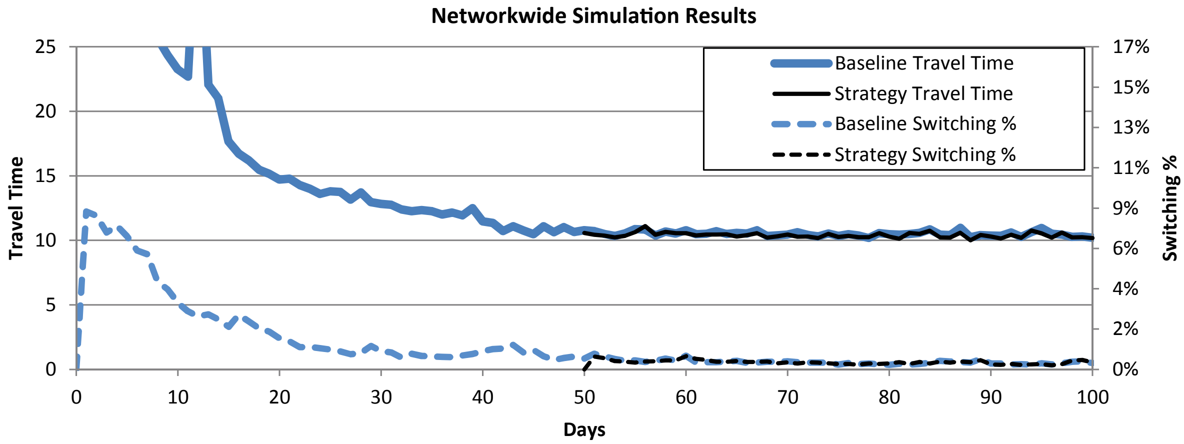


Figure 4.19. Overview of strategy testing plan under stochastic capacity.

Measures of Effectiveness

For the purposes of this demonstration project, a number of performance measures were monitored. Summary results for each performance measure were aggregated on a link, corridor, O-D-pair, and/or network basis, depending on the nature of the performance measure. The performance measures that were monitored in this application included the following:

- Peak hour (5 to 6 p.m.) link and corridor volume (vph);
- Total travel time (minutes) for links, corridors, and the entire network;
- Average travel time (minutes/veh) for links, corridors, and the entire network;
- Vehicle miles traveled (VMT) during the peak hour for each corridor and the network;
- Average speed (mph) for each corridor and the network;
- Density (veh/mi/lane) for each corridor; and
- Breakdown frequency for each corridor.

The results from the 20 simulated days were also used to compute an average value for each of these performance measures and values representing both the 5th and 95th percentile confidence levels for each performance measure.

Identification of Alternative Improvement Strategies

Figure 4.20 provides a visual overview of the subarea baseline conditions after completion of the 50-day Regime I period.

It is clear that Highway 26 (along the northern edge of the study area) and TV Highway (in the center of the study area) are both east-west facilities with significant amounts of congestion. This is also true in real life: both facilities are heavily used by westbound work-to-home commuters during the evening peak period, and both facilities also provide important regional connectivity services at the same time. On the basis of these facts, it was hypothesized that the following improvement strategies might be effective countermeasures and therefore worthy of additional investigation:

1. Expansion of TV Highway west of Murray Boulevard from the existing five-lane cross section to a new seven-lane cross section through new construction;
2. Addition of two new through lanes in the eastbound and westbound directions on TV Highway west of Murray

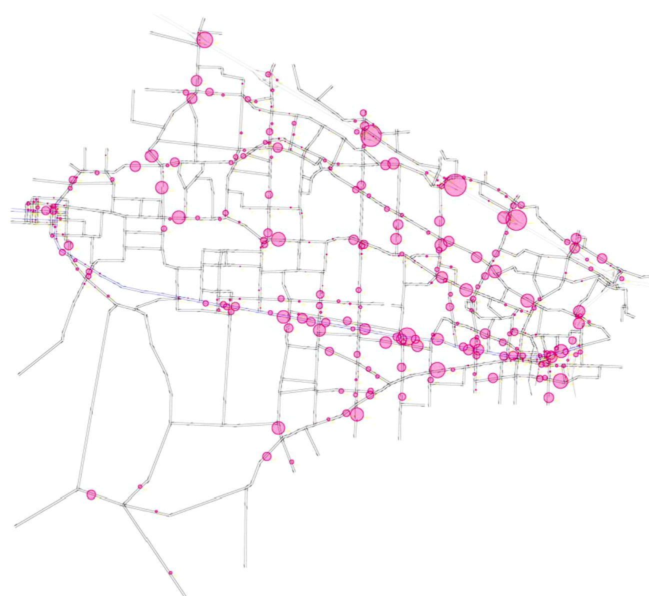


Figure 4.20. Overview of Portland subarea baseline conditions.

Boulevard, resulting in four through lanes in each direction of travel;
3. Conversion of TV Highway west of Murray Boulevard from the existing five-lane cross section to a new seven-lane cross section through use of narrower lanes and shoulders;
4. Improved signal timing as well as the addition of a new through lane in each direction on Baseline Road (between Murray Boulevard and the City of Hillsboro);
5. Improved signal timing on Baseline Road plus the introduction of congestion pricing on TV Highway between Highway 217 and Murray Boulevard, by adding a $1 surcharge to any trip using any part of TV Highway within these boundaries during the evening peak hour (5 to 6 p.m.);
6. Expansion of the availability of (and access to) pretrip traffic condition reports and information, such that 10% of all drivers take advantage of this opportunity (under baseline conditions, only 1% of all drivers are assumed to consult traffic condition reports/websites before departing on their trip);
7. Expansion in the use of en route information, from 1% under existing conditions to 10%.

Evaluation of Alternative Improvement Strategies

Each of the seven improvement strategies identified above, in addition to a “no change” baseline scenario, was analyzed. Each analysis began with traffic volumes and routing path conditions set as they existed at the end of the Regime I time period (see Figure 4.19). The improvement strategy was then introduced and an additional 50 days were simulated. Results data were then collected for the last 20 days of this simulation period.

Depending on the particular hardware being used, each 50-day simulation required between 6 and 13 hours to complete, using 64-bit machines with 16 to 18 GB RAM. More specifically, machines with two parallel processors required 12 to 13 hours to complete a 50-day simulation, whereas machines with four parallel processors required 6 to 7 hours to complete the same task.

Large output files are produced from each model run. As an example, approximately 5 million records relating to link performance characteristics were produced from each model run. These records were imported into a query (SQL) database and then post processed by using customized but simple routines in order to produce the summary results that follow.

Summary of Results

Figures 4.21 through 4.25 present summary results across all performance measures for a variety of corridors, for the subarea network as a whole, and for three separate O-D pairs. The indicated vertical lines depict the 95th percentile confidence interval for the mean value of the MOE and thus are a measure of the MOE reliability. Examination of these results reveals the following potentially important observations:

- Westbound travel time on TV Highway between Highway 217 and Hillsboro is most positively affected by the congestion pricing alternative. However, this same effect is not apparent for the section of TV Highway between Murray Boulevard and Hillsboro. Further, the congestion pricing strategy had a net adverse effect on travel times within the network as a whole.
- Network travel time performance benefited the most from increasing the percentage of drivers who make use of pretrip information, although this particular strategy did not significantly improve the travel performance of any of the three east-west corridors that were examined.
- The reliability of travel times on TV Highway between Highway 217 and Murray, as defined by the difference between the 5th and 95th percentile confidence levels, improved the most with the construction of an additional through lane in each direction on TV Highway; the provision of an additional lane and improved signal timing on Baseline Road; and increased usage of pretrip information. In other words, improvements in travel time reliability were as great with some low-cost strategies as was achieved with the construction of an additional lane, even when average travel time was not appreciably affected.
- All tested strategies resulted in a fairly significant reduction in average corridor density for each of the three corridors that were examined. However, this did not always translate into a corresponding reduction in average travel time.

These types of findings, which are an outcome of the enhanced network operational analysis procedures described in this report, provide a more detailed and complete assessment of the effectiveness of operational strategies at a network level. It is a significant step forward from the level of information produced by traditional transportation planning modeling and analysis tools. The types of findings described above are useful for not only transportation professionals but also decision makers who have responsibility for transportation investment decisions in the subarea boundaries.

The analysis reveals some fundamental take-away points for transportation professionals who undertake analyses of this type:

1. Multiple performance measures must be considered in order to obtain the most complete assessment of a particular operational improvement strategy. Levels of service and volume/capacity ratios continue to be important,

(text continues on page 78)

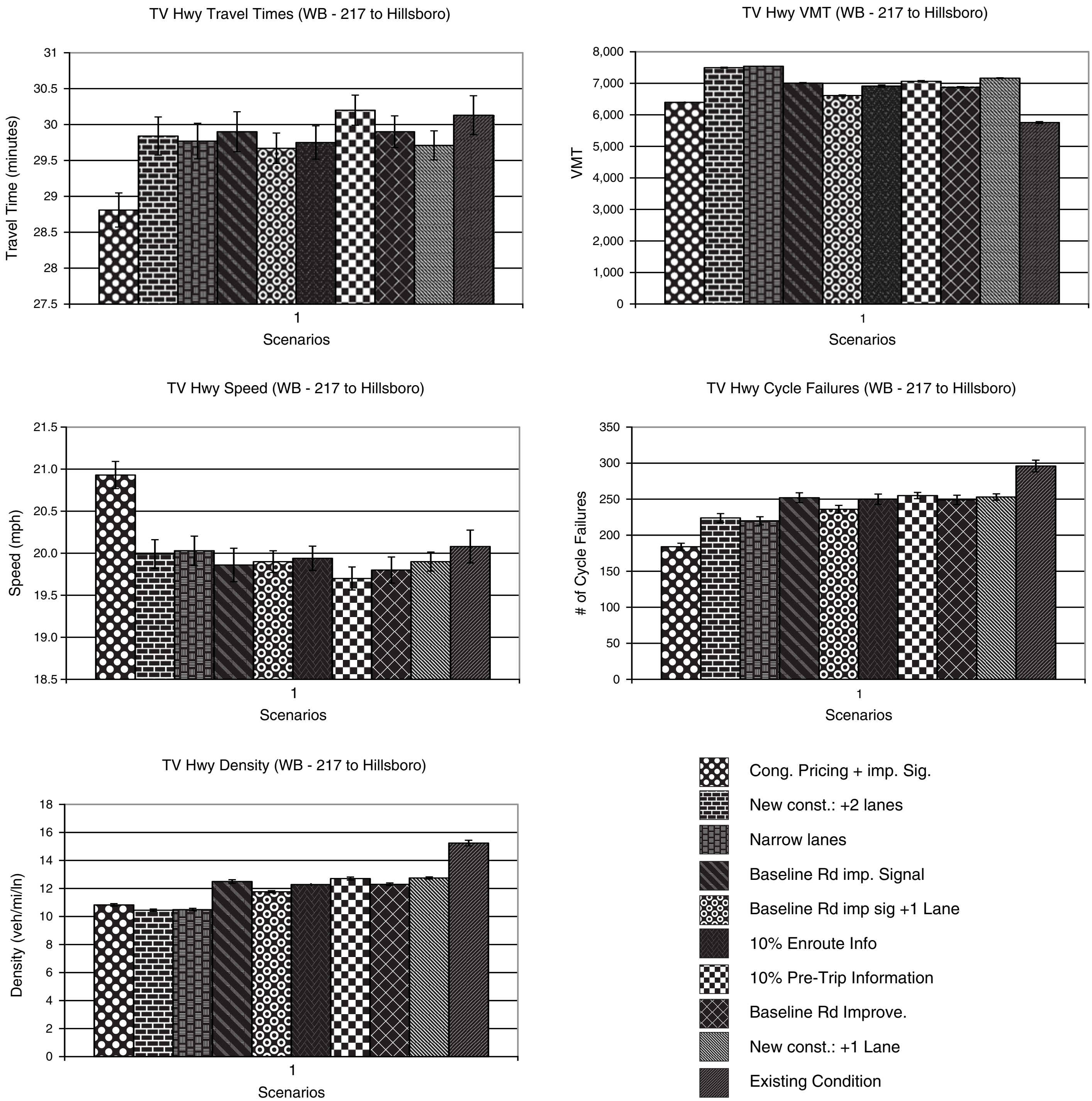


Figure 4.21. Performance characteristics of alternative improvement strategies on TV Highway (between Highway 217 and Hillsboro).

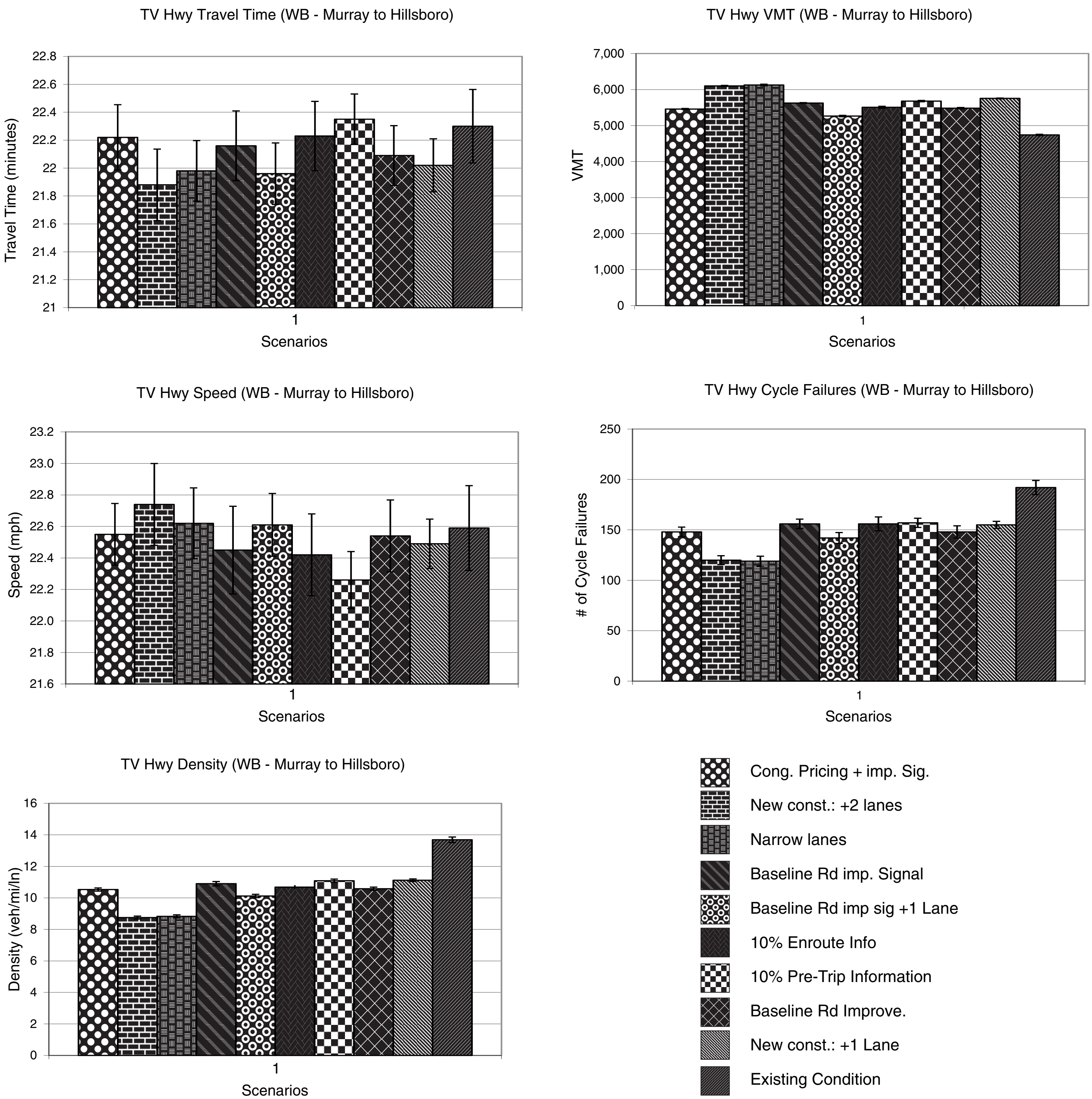


Figure 4.22. Performance characteristics of alternative improvement strategies on TV Highway (between Murray Boulevard and Hillsboro).

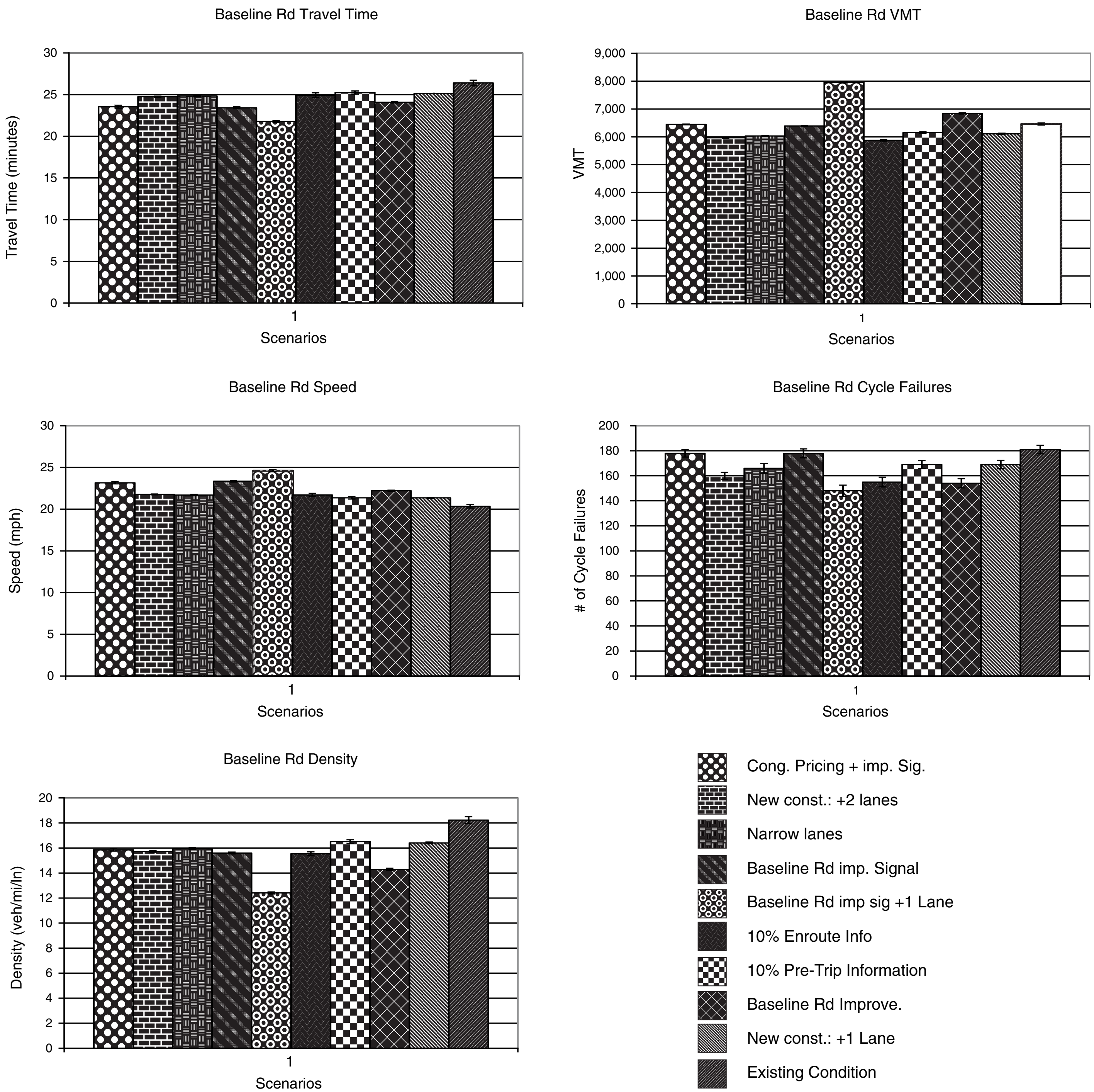


Figure 4.23. Performance characteristics of alternative improvement strategies on Baseline Road (between Murray Boulevard and Hillsboro).

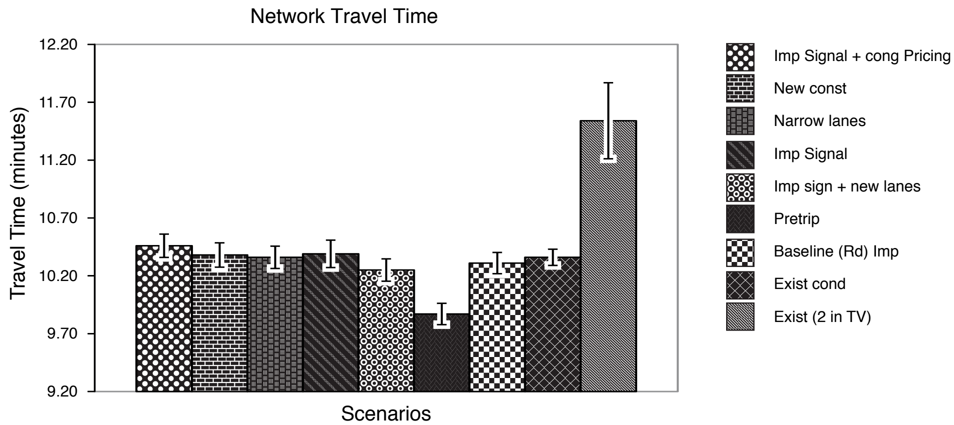


Figure 4.24. Networkwide travel time characteristics of alternative improvement strategies.

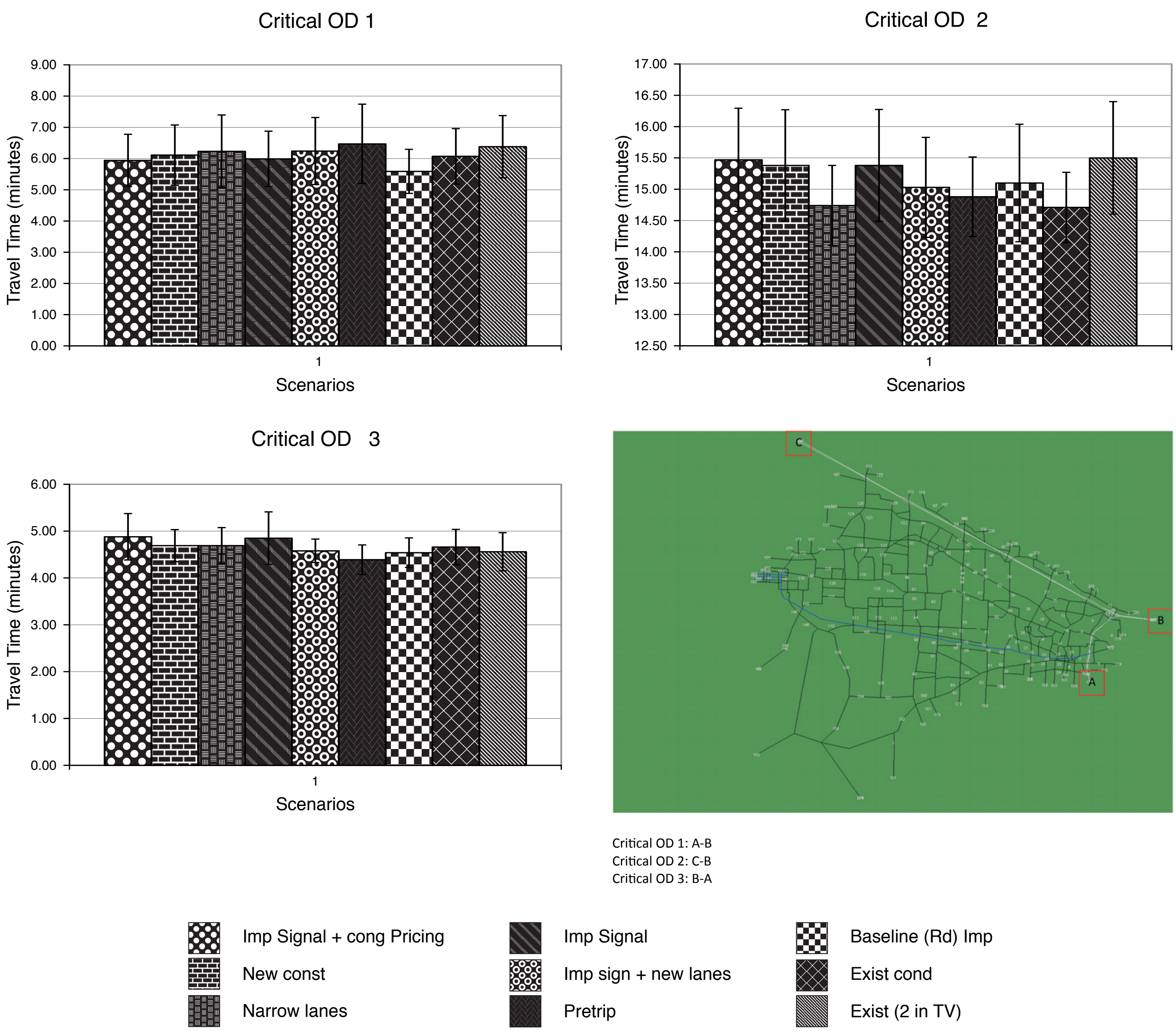


Figure 4.25. Performance characteristics for three origin–destination (O-D) pairs.

(continued from page 73)

of course, but these by themselves do not provide a complete or even adequate assessment of an operational improvement strategy, particularly in situations where a range of improvement strategies is being tested.

2. Performance measures must be selected to reflect link, corridor, and network characteristics. Alternative improvement strategies can have markedly differently effects at the link, corridor, and network levels, both individually and also relative to one another.
3. Evaluations of operational strategies at the network level must consider impacts on driver's route choice. As a result of the modeling enhancements made as part of this project, the Portland subarea exercise demonstrated the effects that operational strategies have on route choice. In some cases, operational strategies such as added lanes result in capacity increases that attracted vehicles to a particular route, whereas other strategies such as traveler information shifted demand from congested facilities to routes with available capacity. Understanding the relationship that capacity enhancements have on demand and vice versa is critical for congested networks.
4. A representative cross section of corridors and O-D pairs should be evaluated. Past analyses have typically been limited to the particular links and/or corridors within which the improvement strategy is implemented. However, the results of this demonstration application show that the performance characteristics of other corridors and O-D pairs are also likely to be affected and should therefore be taken into account.
5. The effectiveness of a particular operational improvement strategy can only be evaluated within the context of the network environment in which it will be implemented. It is not possible to accurately estimate the capacity-enhancing effects of a particular strategy through something as simple as a lookup table; so many other factors affect the effectiveness of an operational improvement strategy that it must be assessed in the context of the network within which it is situated.

Travel time reliability is an especially important performance measure to consider within a congested or oversaturated network. As demand outstrips supply and congestion levels increase, it becomes more and more difficult to show significant improvements in average travel time through the implementation of one or more operational improvement strategies. But quite often the reliability associated with the average travel time will improve with these operational improvements, even when average travel times remain virtually unchanged. This is a very important benefit because it has the effect of giving more discretionary time back to the driving public, just as an absolute reduction in average travel time would have done. It is also a benefit that has previously gone unnoticed, unmeasured, and/or unreported in analyses of this type.

Summary and Conclusions

The Portland subarea network represents a good venue for demonstrating the steps involved in applying the methodological enhancements developed in this project to a real-world environment. The demonstration described in this chapter has also documented the value and effectiveness of this new analysis tool in transforming the way alternatives analyses are conducted and, ultimately, the way important transportation investment decisions are made. Important insights have been achieved that will positively affect the nature, scope, and scale of performance measures used to judge different investment strategies.

Reference

1. Mahmassani, H., and H. Sbayti. *DYNASMART-P User's Guide Version 1.2.* Maryland Transportation Initiative, 2005.

CHAPTER 5

Next Steps

Accounting for the Effects of Nonrecurring Congestion

This project provides a practical methodology for assessing the ability of various operational strategies, either singly or in combination with one another, to forestall or eliminate the need to construct additional lane miles of capacity within a transportation network. The methodology provides effectiveness assessments about both travel time and reliability at the link, corridor, and network levels. It is already integrated into two dynamic traffic assignment modeling procedures (including the open source program DTALite as well as DYNASMART-P) and can be integrated into others as well.

Unfortunately, the usefulness of the methodology is constrained by the fact that the modeling environments in which it operates do not currently account for crashes, weather, and other events that can cause nonrecurring congestion. Nonrecurring congestion represents a significant part of the delay and frustration experienced by travelers and therefore has a substantial impact on both travel time and travel time reliability. Yet, traditional operational models do not account for its effects. Additionally, the overall effectiveness of many operational strategies, such as those evaluated under Capacity Project C05, is misrepresented if only the recurring congestion effects are considered. As an example, the use of narrow lanes as an operational improvement strategy will typically result in more capacity, reduced travel time, and improved reliability in an environment where incident effects are not considered. But narrow lanes might also increase the potential for crashes, thereby reducing their effectiveness if nonrecurring congestion effects are taken into account. As another example, ramp metering does not add much in terms of capacity, but it stabilizes flow, lowers the probability of breakdown during demand surges, and reduces the potential for crashes. In this case, the benefit of ramp metering will be underestimated unless nonrecurring congestion effects are also considered. Therefore, a comprehensive assessment of the overall effectiveness of these operational strategies needs to account for their effects on both recurring and nonrecurring congestion.

Therefore, the research team recommends additional work to (a) extend the capability of the DTA methodology to include an assessment of nonrecurring congestion effects and (b) incorporate this additional capability into at least one functional modeling environment.

Approach

Nonrecurring events that significantly affect the travel time and reliability of a transportation system can be defined across three dimensions:

1. The event type (planned versus unplanned);
2. The event's temporal scope (time of onset and duration); and
3. The event's spatial scope (location specific versus areawide).

Some existing DTA platforms, including both DTALite and DYNASMART-P, already allow users to specify the location, time, and duration of a location-specific planned event (e.g., a work zone) and then observe the effects this planned event has on network conditions and routing. However, unplanned events (such as a crash) and events that have areawide effects (such as the onset of a rainstorm) cannot be modeled.

A method for testing the effects of both planned and unplanned events should be developed in this further effort. With respect to modeling location-specific planned events, one approach might be to identify up to five strategies for work zones, on the basis of previous SHRP 2 work (particularly Reliability Project L03), and then test the effects of these work zone strategies by using DTALite and/or DYNASMART-P.

Additional enhancements are recommended to one or more DTA models to incorporate unplanned events such as crashes and weather events. For example, crash prediction models can be incorporated for freeway and arterial facilities based on methods such as those described in AASHTO's *Highway Safety*

Manual. This will result in a built-in stochastic model to predict crashes, similar to the stochastic model that is already in place for bottleneck capacity. Up to five strategies might be identified and tested that affect crash rates (either upward or downward) and/or the duration of the crash effects. These tests should be conducted on a real-world regional network so that reported results can be used to fairly measure the capacity and reliability effects at the corridor and network levels.

The results of SHRP 2 Capacity Project C01, as contained in the Transportation for Communities—Advancing Projects through Partnerships (TCAPP) website, provide the framework for an integrated planning and operations model. The results of SHRP 2 Capacity Project C05 and Reliability Project L05 will be referenced in TCAPP. This will effectively integrate planning- and operations-related findings from the SHRP 2 Capacity and Reliability program areas and serve as a significant step toward the implementation of the research.

Suggested Work Plan

Incorporating the ability to simulate the effects of nonrecurring events on network performance into DYNASMART-P (or any other tool, for that matter) is not a trivial matter, particularly in a stochastic and time-sensitive environment. A work plan that is recommended for consideration involves four basic steps:

1. Develop and articulate the specific strategy/method for representing the effects of nonrecurring congestion in one or more DTA models. There are several different techniques by which this can be accomplished. Examples include the introduction of link-specific disutility functions and use of a probabilistic-based simulation process. There are also other ongoing efforts in the broad field of nonrecurring congestion analysis that may serve as good springboards; an example of this might be some elements of the Surrogate Safety Assessment Model that is under investigation and development at FHWA. A by-invitation workshop might be convened with key SHRP 2 safety contractors, FHWA representatives, and DTA experts to obtain a critical review of the research team's proposed approach and also to receive additional input and guidance before starting the coding process. This work effort would reasonably require 3 to 4 months to complete.
2. Produce the necessary software code to implement the selected strategy/method within one or more existing DTA models. This is a straightforward process but nevertheless requires about 2 to 3 months after taking account of the need for testing and debugging.
3. Apply the enhanced DTA models to a real-world regional network under scenarios that involve selected operational strategies applied both singly and in combination. This effort is also straightforward and is expected to take about 2 to 3 months to complete.
4. Summarize the results of Step 3 and incorporate the resultant insights and findings into a final report. This final activity is expected to take about 2 months to complete.

APPENDIX A

Supplemental Materials for Uninterrupted Flow Facilities

Table A.1. Information Provided in TransGuide Database

Column	Units	Description
Date		Date of data as MM/DD/YYYY
Time		Time of data as HH24:MI:SS
Lane		Lane N; N ranges from 1 to the number of lanes at the location.
Roadway		Route number and direction
ID		Unique detector identifier
Speed	Mph	Format: Speed = **
Volume	Veh	Format: Volume = ***
Occ	%	Format: Occ = ***

Table A.2. Information Provided in Bay Area Database

Column	Units	Description
Timestamp		Date of data as MM/DD/YYYY HH24:MI:SS. Note that the time indicates the beginning of the summary period. For example, a time of 08:00:00 reports measurements from between 08:00:00 and 08:59:59.
Station		Unique station identifier. Use this value to crossreference with metadata files.
District		District number
Route		Route number
Direction of Travel		N | S | E | W
Station Type		CD = Coll/Dist FF = Fwy-Fwy HV = HOV FR = Off-Ramp OR = On-Ramp ML = Mainline
Station Length		Segment length covered by the station in miles/km.
Samples		Total number of samples received for all lanes.
% Observed	%	Percentage of individual lane points at this location that were observed (e.g., not imputed).
Total Flow	Veh/5-minutes	Sum of flows over the 5-minute period across all lanes. Note that the basic 5-minute rollup normalizes flow by the number of good samples received from the controller.
Avg Occupancy	%	Average occupancy across all lanes over the 5-minute period expressed as a decimal number between 0 and 1.
Avg Speed	mph	Flow-weighted average speed over the 5-minute period across all lanes. If flow is 0, mathematical average of 5-minute station speeds.
Lane N Samples		Number of good samples received for Lane N. N ranges from 1 to the number of lanes at the location.
Lane N Flow	veh/hour	Total flow for Lane N over the 5-minute period normalized by the number of good samples.
Lane N Avg Occ	%	Average occupancy for Lane N expressed as a decimal number between 0 and 1. N ranges from 1 to the number of lanes at the location.
Lane N Avg Speed	mph	Flow-weighted average of Lane N speeds. If flow is 0, mathematical average of 5-minute lane speeds. N ranges from 1 to the number of lanes.
Lane N Observed		1 indicates observed data, 0 indicates imputed.

Table A.3. Summary of Fitted Queue Discharge Model

Queue Duration Intervals (includes pre-breakdown)	Number of Instances (= queue duration)	Number of Instances (≤ queue duration)	Fitted Parameters and GOF				GOF with μ_c = 1,850 and β = 0.2	
			μ_c	β	R^2	RMSE	R^2	RMSE
2	94	94	1,934	0.1950	0.9116	97.57	0.9092	98.92
3	54	148	1,876	0.1589	0.8676	131.87	0.8655	132.90
4	39	187	1,872	0.1724	0.8658	121.76	0.8647	122.24
5	53	240	1,843	0.1791	0.8415	111.18	0.8410	111.37
6	56	296	1,838	0.1855	0.8228	103.94	0.8225	104.02
7	108	404	1,840	0.2035	0.7759	95.78	0.7757	95.81
8	53	457	1,848	0.2132	0.7441	98.48	0.7438	98.54
9	15	472	1,842	0.2115	0.7379	99.03	0.7375	99.11
10	6	478	1,842	0.2127	0.7306	100.19	0.7301	100.27
11	6	484	1,838	0.2117	0.7282	100.49	0.7277	100.59
12	5	489	1,832	0.2072	0.7318	102.35	0.7312	102.47
13	4	493	1,822	0.1986	0.7445	103.53	0.7437	103.68
14	1	494	1,817	0.1953	0.7467	103.73	0.7459	103.90
19	1	495	1,798	0.1771	0.7724	104.26	0.7707	104.64

Note: GOF = goodness of fit; μ_c = average discharge rate in pc/h/ln; β = linear parameter that models the strength of regression to the mean.

Related SHRP 2 Research

Framework for Collaborative Decision Making on Additions to Highway Capacity (C01)

Dynamic, Integrated Model System: Jacksonville-Area Application (C10A)

Transferability of Activity-Based Model Parameters (C10A)

Dynamic, Integrated Model System: Sacramento-Area Application (C10B)

Incorporating Reliability Performance Measures into Operations and Planning Modeling Tools (L04)

Identification and Evaluation of the Cost-Effectiveness of Highway Design Features to Reduce Nonrecurrent Congestion (L07)