Transportation System Simulation Manual

November 2019

U.S. Department of Transportation
Federal Highway Administration

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
xx-xx		

4. Title and Subtitle

Transportation Systems Simulation Manual

5. Report Date

November 2019

6. Performing Organization Code

7. Authors

George List, Richard Dowling, Robert Bertini, David Hale, Shannon Warchol, Sean Qian

8. Performing Organization Report No.

9. Performing Organization Name and Address

Leidos
11251 Roger Bacon Drive
Reston, VA 20190

10. Work Unit No. (TRAIS)

11. Contract or Grant No.

DFTH6116D00030

12. Sponsoring Agency Name and Address

Federal Highway Administration
U.S. Department of Transportation
1200 New Jersey Avenue, SE
Washington, DC 20590

13. Type of Report and Period Covered

Final report, September 2016–November 2019

14. Sponsoring Agency Code

HRDO

15. Supplementary Notes

The government task manager is John Halkias.

16. Abstract

In the United States and in many parts of the world, the increasing rate of surface traffic congestion is outpacing the available roadway infrastructure in urban areas. Agencies and governments are pursuing intelligent transportation systems solutions, active transportation demand management strategies, connected vehicle technologies, and alternative intersection designs that significantly improve traffic flows, without requiring construction of additional lanes. Given the expenses and complexities associated with advanced traffic management, simulation analysis is becoming increasingly vital as a mechanism for evaluating potential solutions and strategies prior to implementation. However, despite the growing need for traffic simulation analysis and expertise, the transportation industry is struggling to adapt. Worldwide surveys demonstrate that engineers consistently lack the time, money, expertise, and guidance to properly calibrate their simulation models. A proliferation of simulation products has led to a lack of standardization, interoperability, and proper tool selection for disparate engineering objectives. Several countries have developed simulation guidelines, but these have not gained recognition in the United States. Numerous State departments of transportation have produced their own simulation protocols, but none of them have gained nationwide acceptance. Agencies and States are requesting the development of a national Transportation System Simulation Manual (TSSM), which can provide the necessary guidance to support 21st century traffic analyses. This document represents the first-ever complete draft of the TSSM.

17. Key Words	18. Distribution Statement
Simulation, calibration	No restrictions.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No of Pages	22. Price
Unclassified	Unclassified	456	N/A

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized.**

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2D	two dimensional
3D	three dimensional
AADT	annual average daily traffic
AAE	average absolute error
ACHD	Ada County Highway District
ADT	average daily traffic
AM	morning peak
ANOVA	analysis of variance
API	application programming interface
ATIS	advanced traveler information systems
ATM	active traffic management
ATR	automatic traffic recorder
BDAE	bounded average error
BMP	bitmap image file
BSM	basic safety message
CAD	computer-aided dispatch
CAD	computer-aided design
CATER	Center for Advanced Transportation Education and Research
CATT	Center for Advanced Transportation Technology
CAV	connected and/or autonomous vehicles
CCTV	closed-circuit television
CDF	cumulative distribution function
CFI	continuous flow intersection
CMF	crash modification factors
CMS	Congestion Management System
COMPASS	Community Planning Association of Southwest Idaho
CSV	comma-separated values
CV	connected vehicles
D/C	demand-to-capacity
DAF	demand adjustment factor
DARPA	Defense Advanced Research Projects Agency
DCM	data capture management
DDI	diverging diamond interchange
DEQL	denied entry queue length
DF	degrees of freedom
DLL	dynamic-link library
DLM	dynamic lane merge
DLT	displaced left-turn intersection
DMS	dynamic message sign
DOT	department of transportation
DOT	department of transportation
DSED	Data Services Engineering Division

DTA	dynamic traffic assignment
EPA	Environmental Protection Agency
EPA	Environmental Protection Agency
FCFS	first come, first served
FDF	frequency distribution function
FDOT	Florida Department of Transportation
FHWA	Federal Highway Administration
FIFO	first in, first out
ft	feet
ft/s	feet per second
GhG	greenhouse gas
GIF	graphic interchange format
GIS	geographic information systems
GPS	global positioning system
GUI	graphical user interface
HCM	Highway Capacity Manual
HGV	heavy goods vehicle
HOT	high occupancy toll
HOV	high occupancy vehicle
HTTP	hypertext transfer protocol
I–270 MD	Interstate 270 Maryland
I–35 KS	I–35 Kansas
I–66 VA	I–66 Virginia
ID	identification
ID	Idaho
IDT	Idaho Transportation Department
ITS	intelligent transportation systems
JPEG	Joint Photographic Experts Group
KC	Kansas City
KDOT	Kansas Department of Transportation
km/h	kilometers per hour
LHD	Latin hypercube sampling design
L-HGV	large heavy goods vehicle
LIFO	last in first out
LOS	level of service
LPR	license plate recognition
max	maximum
MD 121	Maryland Route 121
MD 187	Maryland Route 187
MDOT SHA	Maryland Department of Transportation State Highway Administration
ME	mean error
mi/ln	miles per lane
min	minimum
min/mi	minute per mile

MOVES	MOtor Vehicle Emission Simulator
MP	milepost
mph	miles per hour
MPO	metropolitan planning organization
NCDOT	North Carolina Department of Transportation
NDOT	Nevada Department of Transportation
NEMA	National Electrical Manufacturers Association
NHS	national highway system
Nox	nitrous oxides
NPMRDS	national performance management research data set
NV	Nevada
OBU	onboard units
O-D	origin-destination
ODOT	Oregon Department of Transportation
OR 211	Oregon Route 211
OR 216	Oregon Route 216
OR 35	Oregon 35
OR 99W	Oregon Route 99W
PBPD	performance based practical design
PDF	probability density function
PHD	person-hours of delay
PHT	person-hours traveled
PM	evening peak
PMT	person-miles traveled
PSRC	Puget Sound Regional Council
QAP	queue accumulation polygon
QAP	queue accumulation polygons
QL 1	surveyor location 1
QL 2	surveyor location 2
RDE	research data exchange
RITIS	Regional Integrated Transportation Information System
RMSE	root mean square error
RTC	Regional Transportation Commission
RTMS	remote traffic microwave sensor
RTOR	right turn on red
rv1	random value 1
rv2	random value 2
sec/veh	seconds per vehicle
SPUI	single-point urban interchange
SQM	simulation quality metrics
SSAM	Surrogate Safety Assessment Mode
SSM	safety surrogate measures
STV	smart travel van
SUV	sport utility vehicle
TDC	traffic data collector

TDM	travel demand model
TIFF	tagged image file format
TMC	government agency traffic management centers
TMC	traffic message channel
TOD	time of day
TRANSCOM	Transportation Operating Coordinating Committee
TRM	triangle regional model
TSMO	transportation system management and operations
TSSM	Transportation System Simulation Manual
U.S.	United States
U.S. 101	U.S. Route 101
U.S. 26	U.S. Route 26
UAV	unmanned aerial vehicles
USDOT	United States Department of Transportation
V/C	volume-to-capacity
VA	Virginia
VCV	verification, calibration, and validation
VDOT	Virginia Department of Transportation
veh/hr	vehicles per hour
veh/hr/ln	vehicles per hour per lane
veh/km/ln	vehicles per kilometer per lane
veh/mi	vehicles per mile
veh/mi/ln	vehicles per mile per lane
veh/p	vehicles per time interval
VHD	vehicle hours delay
VHT	vehicle hours traveled
VMS	variable message sign
VMT	vehicle miles traveled
WisDOT	Wisconsin Department of Transportation
WSDOT	Washington State Department of Transportation
XML	extensible markup language
Y+AR	yellow plus all red

LIST OF SYMBOLS

Δ	increment
©	copyright
™	trademark
%	percent
@	at
∑	summation of
√	square root of
λ	weighting factor
>	greater than
<	less than
∞	infinity
#	number
’	feet
&	and
$	dollars
e	error value
α	confidence level
μ	true mean value
Δ	difference
^	estimated value
±	plus or minus
®	registered trademark
+	plus
-	minus
≈	almost equal to

VOLUME I. BASIC CONCEPTS OF SIMULATION

Volume I of the Transportation System Simulation Manual (TSSM) introduces the basic concepts of simulation: what it is, how and when it is typically used, and what it involves. It is intended to be a source of information for readers who are not familiar with simulation-based analyses and need a quick review of the basics before they can tackle the more technical issues covered in Volume II. As such, Volume I is a highly compressed summary of what might be covered in a beginning graduate-level course on simulation analyses.

Since such a compressed summary cannot possibly educate readers on all the details of simulation analysis, this volume points to additional sources, outside the TSSM, that readers should consult to gain additional insights into the bases for simulation models and ideas about how they can and should be employed.

Chapter 1 defines simulation and describes the differences between simulation and analytic approaches to performance analysis. It describes the conditions when simulation is used and refers the reader to various (Federal Highway Administration) FHWA reports for deciding on the appropriate analysis approach.

Chapter 2 describes the various levels of aggregation or resolution at which simulation can be applied: macroscopic, mesoscopic, and microscopic. The similarities and differences among the different resolutions of simulation modeling are described. Multi-scale and multi-resolution modeling are described.

Chapter 3 provides an overview of basic simulation concepts.

Chapter 4 provides short briefings on a collection of issues typically encountered in simulation. Practitioners will find it worthwhile to review the topics listed in this chapter to see if their specific issues are covered and how to avoid or at least anticipate simulation issues as they arise.

CHAPTER 1. WHAT IS SIMULATION?

This chapter provides a definition of simulation. It illustrates how simulation is different from analytical analyses and describes how the two techniques can be used together. It illustrates when and why simulation is useful and the benefits and costs. It also provides examples of situations where simulation has been used, why it was chosen, how it was employed, and the reasons it was beneficial.

1.1 DISTINCTIONS BETWEEN SIMULATION AND ANALYTIC ANALYSES

Simulation is the "imitation of the operation of a real-world process or system over time" (Banks 1998). The key words in this definition are over time. Analytic analyses predict the average system performance over a selected time duration or for a specific point in time.

The key difference between the two analysis approaches is that simulation splits up time into a series of intervals, or steps, and has the system's operation and performance unfold across them, from one time step to the next. The length of the time steps can be all the same, in the case of a time-based simulation, or varied in length, in the case of an event-driven simulation. An example of the first would be the simulation of a vehicle moving along a highway where, in every time step, its status—location, speed, acceleration, and deceleration—is updated and the new forces to be applied are determined. An example of the latter is a pre-timed traffic signal's operation, where the status is the combination of light indications, and nothing changes until the end of each interval is reached. In a typical microscopic, time-based simulation, the length of each time step is often one-tenth of a second, so that vehicle dynamics and signal timing events can be modeled accurately. (In one-tenth of a second at 60 miles per hour (mph), a vehicle moves 8.8 feet; many signal timing parameters, such as gap times, are specified in one-tenth of a second.) In an event-based simulation model, the lengths of the time steps are determined by how long it is until the next vehicle arrives at a downstream stop line or a change in traffic signal status arises, or some other event.

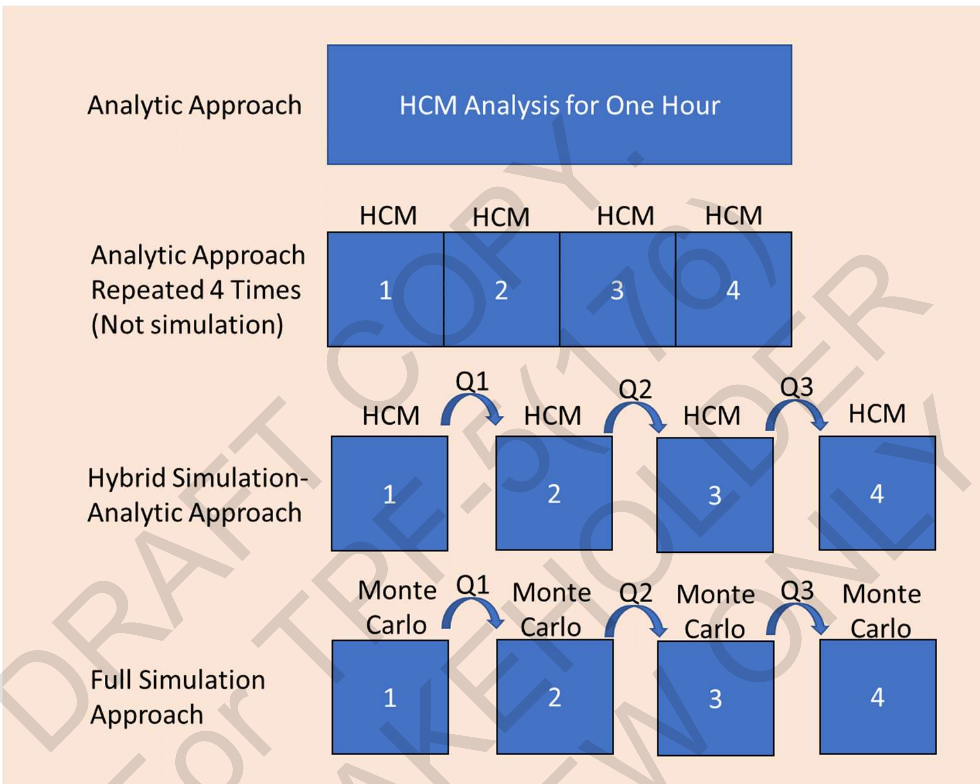
That said, the distinction between simulation and analytic approaches is blurry. For example, a simulation analysis often employs analytic equations to evaluate conditions within each small time period, and analytic analyses for several time periods may be linked together in a simulation-like manner. A couple examples may help distinguish between the two (follow along on Figure 1):

1. If the Highway Capacity Manual (HCM) is used to solve for the average delay at a signal for the peak hour, this is an analytical assessment.
2. However, if the HCM method is used in the following way, it resembles a simulation.
    a. Split the peak hour into four 15-minute subperiods and apply the HCM method to each one. So far, this is still an analytical approach, but one that is repeated four times instead of done just once.
    b. Then, use the HCM results from the first 15-minute period as inputs for the second 15-minute period, etcetera. Now the HCM is being used in a simulation

style. The carryover of queued vehicles from one 15-minute time period to the next is the reason this is true.

3. Finally, if a Monte Carlo-based[1] technique is used within the analytical subprocesses to capture the effects of uncertainties in the analytical relationships between inputs and performance outputs, then this is a simulation-based analysis.

The HCM method cited in Figure 1is an example of one of many analytic approaches that may employ simulation as part of the procedure.



Source: FHWA.

Figure 1. Illustration. The differences between simulation and analytic approaches.

Macroscopic, mesoscopic, or microscopic simulation methods may also be applied in a stochastic or deterministic manner. In the latter case, the use of random numbers and Monte Carlo-style value sampling is omitted. For example, the truck travel time estimation method employed in the Mixed Flow Model of the HCM employs a deterministic microscopic simulation model to compute travel times for trucks across undulating grades (Essentially the hybrid simulation analytic approach shown in Figure 1 but without employing an HCM macroscopic analysis methodology.) (Dowling et al. 2014).

---

[1] Monte Carlo techniques use repeated random sampling to obtain results.

Computer simulation uses an abstract representation of the system, captured in logical relationships, that describes the system in such a way that its behavior can be predicted. The simulation model uses virtual objects, called entities, to represent the way in which objects interact, through time and across space, in a manner that mimics the way in which the system behaves in real life.

The logic and rules that define how that interaction takes place are also referred to as simulation algorithms. Most of these algorithms are stochastic in nature, in that one or more of the algorithm parameters are subject to random variability. The inputs to simulation algorithms are the current state of the system, with the logic predicting changes to the system state in the next time step.

1.2 WHEN SHOULD SIMULATION BE USED?

The choice between analysis approaches (simulation or analytical) is covered in depth in two FHWA resources in the Traffic Analysis Toolbox series:

- Volume I: Traffic Analysis Tools Primer (Alexiadis et al. 2004).
- Volume II: Decision Support Methodology for Selecting Traffic Analysis Tools (Jeannotte et al. 2004).

The essential message of these two documents is that the analyst should use simulation whenever and wherever it seems the most cost-effective approach to achieving the desired level of accuracy in predicted traffic system performance. Often, a motivation is that the interactions between parts of the system are too complex to study analytically; or the cascading effects of varying demand across time cannot be addressed analytically. Note that more than cost needs to be considered. The data must be available to make the simulation-based analysis possible. The selected modeling approach must have the appropriate sensitivities to issues and options of concern to the analyst. For example, simulation may be the best (or only) choice when the complexity of the interactions within the system exceed the ability of available analytic approaches to obtain a solution.

> Simulation should be used when it is the most cost-effective approach to achieving the desired accuracy with the appropriate sensitivities.

Some examples of situations where one or the other (or both) of these approaches are appropriate are:

- *Analytical*: An isolated intersection where congestion does not exceed 1 hour, and downstream congestion does not affect the operation of the upstream signal; this can be evaluated using an analytic approach, like the HCM.
    - However, if the analyst wishes to test dynamic traffic responsive transportation system management and operations (TSMO) strategies to reduce fluctuations in delay due to demand surges, incidents, and weather, then a simulation analysis might be most cost-effective tool to use.
- *Simulation*: An intersection where downstream congestion, or left-turn pocket queues overflow into the through lanes involves complex relationships; this is an illustration of a situation where simulation is quite appropriate.

- *Analytical*: A freeway without mainline queuing during the peak hour; this may be most cost-effectively evaluated as isolated pieces (basic, merge, diverge, and weaving sections) using the HCM analytic methods for isolated freeway segments.
    - However, if dynamic TSMO strategies are to be tested, then a simulation analysis might be most cost-effective.
- *Analytical*: A freeway with some mainline queuing that does not last beyond the peak period and does not spill back to on-ramp intersections would require a system analysis using a hybrid simulation and analytical approach, such as the HCM's Freeway System Analysis method.
- *Simulation*: A freeway with off-ramp queues spilling back onto the freeway mainline involves complex lane-by-lane queueing relationships is most cost-effectively evaluated using simulation.

CHAPTER 2. WHAT ARE THE AVAILABLE SIMULATION OPTIONS?

This chapter explains available options for conducting simulation analyses. The main options for simulation relate to the degree of aggregation employed: microscopic, mesoscopic, and macroscopic. This chapter also talks about multi-resolution simulation and agent-based simulation. It provides examples of situations where and why these modeling options have been employed. It also illustrates the differences among the input requirements and the types of outputs provided.

A more extensive discussion of the options can be found in the first two volumes of the Federal Highway Administration's (FHWA) Traffic Analysis Toolbox series:

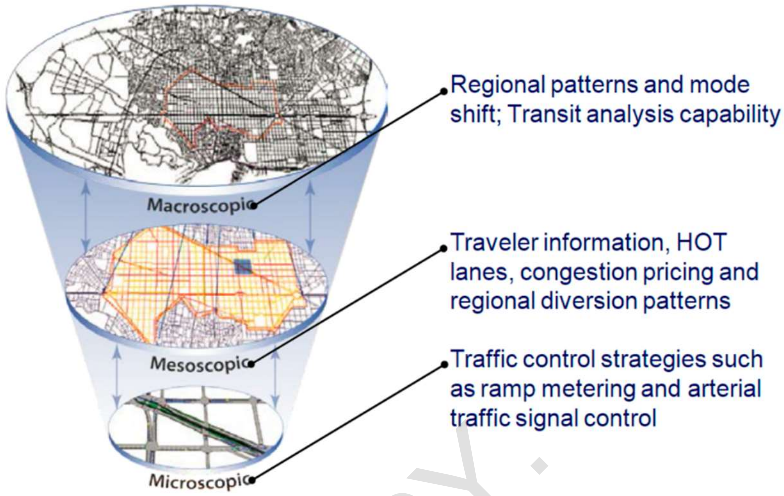
- Volume I: Traffic Analysis Tools Primer (2004).
- Volume II: Decision Support Methodology for Selecting Traffic Analysis Tools (2004).

Traditionally, simulation models for transportation applications are classified according to the level of detail with which they represent the traffic stream. The three principal types of simulation tools are:

- Microscopic models.
- Mesoscopic models.
- Macroscopic models.

The principal difference between the three resolutions is tied to the level of detail with which the network is represented; which directly affects computational efficiency and the size of the network and duration of time to which the model can be applied, as illustrated in

Figure 2.



Source: FHWA.

Figure 2. Diagram. The varying demand modeling and traffic analysis simulation resolutions.

A brief overview of each level of analysis is given in the following sections.

2.1 MICROSCOPIC TRAFFIC ANALYSIS MODELS

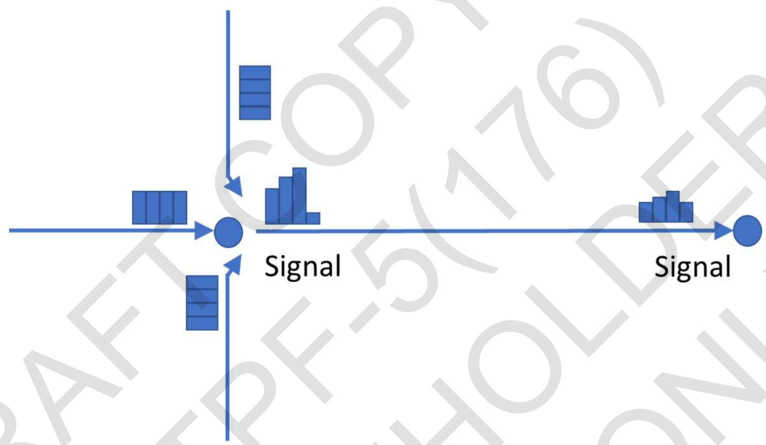
Microscopic models of traffic flow represent the movement of each individual vehicle in the traffic stream. Changes in the state of every vehicle are modeled. The microscopic algorithms predict how each vehicle will behave in response to other vehicles, roadway geometry, and traffic control objects in the network. The basic microscopic algorithms include modeling car-following behavior, acceleration and deceleration, lane changing behavior, gap acceptance behavior, routing behavior, and reaction to traffic control devices. The state of the simulated system or network emerges from the aggregate of these individual microscopic decisions, rather than being predicted directly.

The exquisite detail of microsimulation enables the modeling of a wide variety of phenomena, including changes in vehicle technology that may affect individual driving behavior (such as car following and lane changing). This is illustrated by the Traffic Flow Theory Monograph (FHWA 2015), Barcelo (2010), and Traffic Analysis Toolbox (FHWA 2015).

2.2 MESOSCOPIC TRAFFIC ANALYSIS MODELS

Mesoscopic models of traffic flow also focus on the movement of individual vehicles, but at a more aggregate level. Their individual behavior is not modeled explicitly. The simulation algorithms estimate changes in the state of a link or segment, as in its density, flow rate, or average speed. The prediction of the future state of a link or segment is based on macroscopic traffic stream models, which are a function of the density of vehicles on the link. These traffic stream models are applied to each vehicle on the link so that they move to the next time step, individually and collectively, in accordance with the findings from the macroscopic analysis. As such, mesoscopic models fall between macroscopic and microscopic models, providing a greater resolution than the former, and greater computational efficiency and reduced processing time relative to the latter.

An example of mesoscopic modeling of traffic operations is the manner in which one simulation software estimates the arrival pattern at the downstream signal for vehicles entering at the upstream entry to the link (University of Florida McTrans 2018; TRL Software - TRANSYT 2018). As shown in Figure 3, the second-by-second arrival rates of vehicles is converted into an arrival pattern per cycle. The arrival pattern gradually flattens as the traffic moves down the link using assumed platoon dispersion factors until the downstream signal arrival pattern is obtained. The signal indications determine the signal discharge pattern by turn movement, which becomes the arrival patterns for the entries to the downstream links.



Source: FHWA.

Figure 3. Diagram. TRANSYT-7F mesoscopic approach to simulating arrival patterns.

2.3 MACROSCOPIC TRAFFIC ANALYSIS MODELS

Macroscopic models focus on vehicle flow rates, or the movement of streams of vehicles across time and through space. The variables are typically flow rates, speeds, and densities defined across time and space. To illustrate, time might be divided into 1-second steps and space divided into one-tenth-mile segments. Macroscopic algorithms are governed by the paradigm of conservation theory or continuity theory (no vehicles are created or destroyed). The flow rate, speed, and density of the segment in time step k for segment n is determined by the values of those variables in time step $k-1$ for segment n as well as segments $n-1$ and $n+1$. If the vehicle density on a freeway was 90 vehicles per mile lane (veh/mi/ln) and it had three lanes, then in a one-tenth-mile segment, there would be 27 vehicles. In a microscopic simulation model, the movement of each vehicle would be represented, but in the macroscopic model, only three variables are involved. This means much larger size networks can be modeled because far fewer variables are involved. Macroscopic models can be useful for very large networks, or for sketch-planning applications that predict corridor or networkwide performance, including the effects of traffic management strategies. Macroscopic models are also appropriate in the development of traffic management and control systems and can be used to estimate and predict average traffic flow characteristics. It should be noted that, although the representation of traffic is typically macroscopic, the behavioral rules describe the dynamics of the flow may be microscopic (e.g., gas-kinetic models). There is limited guidance in current practice on the use of macroscopic simulation models (Barcelo 2010).

The Highway Capacity Manual (HCM)[2] Freeway Facility Analysis Method is an example of a macroscopic simulation model. Analysis results for individual segments and time slices are propagated to adjacent segments and time slices (Transportation Research Board 2016).

2.4 DYNAMIC TRAFFIC ASSIGNMENT

Dynamic traffic assignment *(*DTA) is a feature, often incorporated in simulation models, that allows the path choices (and potentially departure times) to be altered during the simulation. This feature helps with the analysis of incident response strategies. It is often included in mesoscopic models, but it can also be found in both microscopic and macroscopic models. Mesoscopic models do not have to employ DTA, and DTA can be applied at all three levels, including microscopic and macroscopic. Refer to Barcelo (2010) and Traffic Analysis Toolbox (FHWA 2015) for further guidance. The greater computational agility of mesoscopic simulation models enables the easier incorporation of DTA, compared to microscopic, and it is often used to represent the dynamic aspects of demand behavior such as rerouting and peak spreading due to congestion.

2.5 SIMILARITIES AND DIFFERENCES AMONG MODEL RESOLUTIONS

Table 1 summarizes key features, similarities, and differences among the three independent resolution levels. As shown in this figure, each level of simulation modeling has certain advantages over the others for specific modeling tasks. However, it is also true that there is a great deal of overlap in the capabilities of the different levels of resolution for simulation. The analyst can also combine models of different resolutions, using a multi-resolution approach to obtain the analytical results of microscopic simulation along with the computational speeds of mesoscopic or macroscopic simulation. Finally, any simulation model can be linked to a demand model to obtain forecasted demands.

[2] Sixth Edition of the Highway Capacity Manual.

Table 1. Similarities and differences of simulation model resolution types.

	Macroscopic	Mesoscopic	Microscopic
Appropriate size of network	Regionwide, corridor, facility	Same as macroscopic	Subarea, corridor, facility, segment, or intersection
Network resolution	Link/node level	Link/node level	Detailed lanes and connectors, lane by lane and turn lanes
Computational complexity	Requires comparatively fewer computations; therefore, can efficiently cover large networks	Intermediate between macroscopic and microscopic resolutions	Requires more computations; therefore, most efficient on smaller networks
Vehicle interaction models	Traffic stream models	Traffic stream models	Vehicle-to-vehicle interaction
Representation of individual vehicles	No	Varies	Yes
Functionality	Long and short-range planning	Areawide operations, traffic diversions	Detailed operations
Time scale	Typically, the finest level considered is 15 minutes; results can be aggregated to longer time periods	Varies	Typically, split-second by split-second analysis with results aggregated as desired
Level of detail	Directly outputs static, average system performance results; can also output link and intersection turn movement specific results; capacity is an input	Can produce both static and dynamic performance results	Very detailed (can be lane by lane); dynamic performance outputs must be aggregated to obtain macroscopic results; queue discharge capacity output only when and where queues are present

Table 1. Similarities and differences of simulation model resolution types. (continuation)

	Macroscopic	Mesoscopic	Microscopic
Number of scenarios modeled	Can be many	May be several	Often just a few
Demand modeling capabilities	Regional traffic operations network analysis or simulation models linked to demand models can best evaluate the trip generation, distribution, mode choice, and routing impacts of changes in highway operations	Subarea models can handle demand changes within the subarea being modeled but cannot address regional effects	Microscopic models generally have very limited demand modeling capabilities (generally only rerouting within the subarea); demand modeling capabilities can be added by linking microscopic model to demand model

2.6 MULTI-SCALE AND MULTI-RESOLUTION MODELING

Multi-scale or multi-resolution models attempt to obtain the advantages of modeling some phenomena that are best handled at the regional level (such as travel demand) while retaining the ability to microscopically evaluate traffic operations. The key is to obtain the improved accuracy without paying too high a price in terms of extra study resources, including added computer run times.

With advances in computational efficiency and increased development of commercial simulation tools, the lines between the three classic types begin to blur. Emerging technologies often feature multi-scale or hybrid simulations, which combine mesoscopic or macroscopic models for most of the network and microscopic models in the areas of interest (subarea analysis). Hybrid models have the benefits of providing the user with one platform or interface and letting the user scale the level of detail for a specific application. As such, they combine the high fidelity of microsimulation in areas of specific interest, with the ability to more accurately represent routing decisions of the surrounding areas and the overall network.

Multi-resolution modeling may link together a regional demand model (that models link traffic operations at the macroscopic level), a mesoscopic simulation model (that employs DTA to refine the demand forecasts coming out of the demand model), and a microscopic simulation model to more precisely model traffic operations at the intersection, segment, and lane level. For more information on multi-resolution modeling, see the FHWA report on Effective Integration of Analysis, Modeling, and Simulation Tools (Nevers et al. 2013).

CHAPTER 3. PRINCIPLES OF SIMULATION

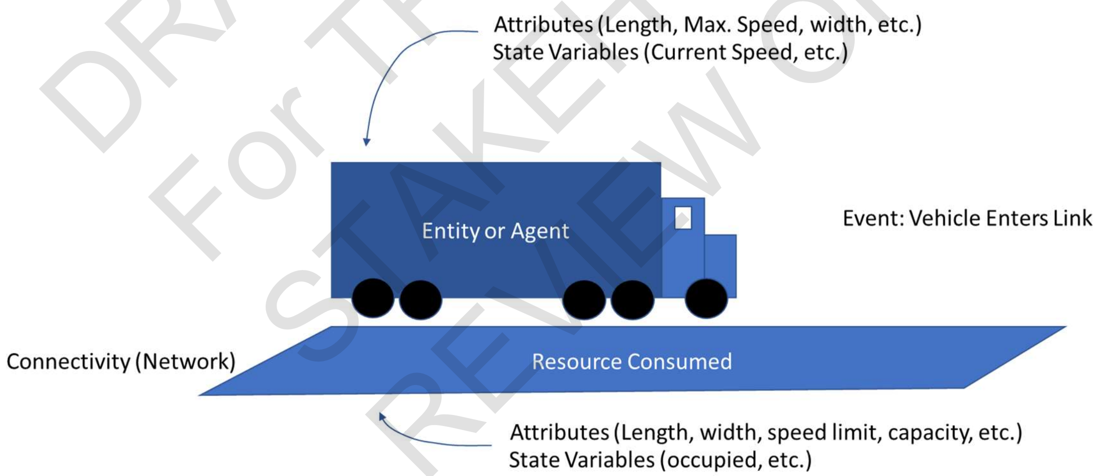
This chapter presents a brief introduction to the principles upon which microscopic, mesoscopic, and macroscopic simulation models are based.

3.1 BASIC SIMULATION CONCEPTS

This section describes the basic ideas upon which all simulation models are based. The material is largely based on two sources: part 5 of the Highway Capacity Manual (HCM) (Transportation Research Board 2000), which talked about simulation in a general sense, and a book (Kelton et al. 2010) that describes how to use a general purpose simulation software package called Arena.

Entities or Agents

Most simulation models are based on the use of entities or agents (which are like avatars in computer games). In a microscopic simulation model, the vehicles are entities, and there may be others. Entities can move around in the simulation, alter their statuses, affect the statuses of the system or other entities, and can be affected by other entities. They are typically the source of data used to generate performance measure output from the simulation. Entities are defined by the user and generated by the model. They move around for some time and then leave the system and/or are disposed at the end of the simulation. In all cases, these entities must be created either by the user or automatically by the software. In transportation simulations, entities can include vehicles, pedestrians, cyclists, transit vehicles, etc. (see Figure 4).



Source: FHWA.

Figure 4. Illustration. Basic simulation concepts.

Attributes

Attributes create the identities of the entities. An attribute is a characteristic. It can define the driving style, a weight, horsepower, communication capability (equipped or unequipped), or anything else the simulation model needs to know to properly represent the way an entity behaves. In a microscopic simulation model, the vehicles (entities) have car-following behavior parameters that affect how they behave. Typically, attributes are captured by parameters that have specific values. The values can be different from one entity to another. It is up to the user to decide what attributes the entities need to have, to name them, attach values to them, change them as suitable, and then use them when it is time to do so.

Network, Links, and Nodes

In transportation simulation models, the entities (e.g., the vehicles) typically move across a predefined network. In many instances, the network consists of links, nodes, and connectors. Links contain attributes about the number of lanes, lane width, grade, etc., used in various simulation algorithms to allow or control movement of entities. An intersection or junction of two links can be represented by a node (common in macroscopic and mesoscopic simulation), or by a series of connectors (more common in microsimulation) to allow movement from one link to another. The combination of links, nodes, and connectors in the simulation network represents the physical infrastructure upon which other control elements in the simulation are configured. In some multimodal simulations (e.g., pedestrian crowd modeling) the network may also include areas, zones, and obstacles that define where entities can and cannot move.

State Variables

A state variable is an item of information that indicates the status of some aspect of the system. In contrast to attributes, state variables are not necessarily tied to specific entities. Rather, they can relate to the system at large. Examples include the location and speed of the vehicles, the occupancy of a parking lot, the status of the signal timing at an intersection, or the status of an incident. The set of state variables identified (enumerated) is sufficient if it completely and unequivocally defines the status of the system at a specific point in time. By knowing the values of this set of state variables, nothing important about the status of the system is unknown.

Resources

Resources are objects the entities use. Entities either occupy or seize (take possession of) resources, and then release them as they move through the system. A resource might be a specific space on the highway, a position in queue, space within an intersection, or space in a storage facility. Typically, two entities cannot have or occupy the same resource at the same time. Throughout the simulation, entities seize resources when available, do what they need to do using the resource(s), and then release them when finished. If more than one resource exists, like the number of parking spaces in a parking lot, each one is called a unit of that resource.

Events

Events are the things that happen. An example would be a vehicle passing over a detector. Events occur at specific points in (simulated) time. The status of the system changes when an event occurs. An attribute may alter its value, a decision may be made, or data may be collected. There are many possibilities. More than one event can occur at the same time. Making sure that events take place in chronological order is critical to ensure the simulation takes place correctly. Otherwise, the simulation results are invalid.

Simulation Clock

A clock is used by the simulation model to keep track of time. The clock may mimic real time (e.g., start at 3 p.m. and run until 8 p.m.), or it may keep track of relative time (e.g., start at 0 minutes and run until 300 minutes). Most, if not all, simulation clocks advance time in small units, called time steps or intervals, like the ticking of the second hand on a wall clock. The time steps can be of the same duration, or to save simulation time, they can sometimes be of varying durations to speed up overall processing time

Processing Logic

The processing logic determines how the state of the system changes with time. Different types of processing logic exist, such as preset if-then rules (common in microscopic simulation), evolving rules that change over time (e.g., agent-based simulation), or predictive equations (common in macroscopic simulation).

Performance Measures

Performance measures (or measures of effectiveness) are the metrics by which the user monitors the performance of the system. Performance measures can be tied to different levels of detail: system, subarea, corridor, facility, link, and node (intersection). Examples include total delay, system average speed, queue lengths, total energy consumption, emissions, noise levels, and travel times. Other examples include volume-to-capacity (V/C) ratios for critical facilities, like bottleneck locations, that are often congested.

3.2 TIME, SPACE, AND STATE EVOLUTION

This section describes how simulation models are based on the evolution of a system's status quo across time and space. It does this in the context of macroscopic, mesoscopic, and microscopic models. This helps illustrate how these modeling paradigms are different. It talks about the variables used to represent the system, how they vary by model type, and how the values of those variables are updated as simulation time progresses. It also talks about the issue of granularity, the representations of time and space, how different levels of granularity affect the level of resolution in the system representation and how it affects the results that can be obtained.

Treatment of Time

The treatment of time is particularly important in simulation models. Since time does not progress continuously in simulation models as it does in the real world, rules are needed to determine how time will advance. Two options are the most common. In the first, time advances in small steps that are always the same size (e.g., 0.1 second). This is called time-based simulation. In the second, time advances in variable size steps from one event to the next. This is called event-based simulation. For further guidance, refer to Traffic Flow Theory Monograph (2015) and HCM (2000).

Generally, a time step-based simulation uses fixed duration time steps, and the simulation clock advances at these fixed steps (typically one second or fractions of a second). Most transportation simulations (both microscopic and mesoscopic) are time-based. Event-based simulations on the other hand are driven by specific (scheduled) events, and generally no computations are performed between events. For example, in the simulation model for a toll booth, the events might be: arrival at back of queue, arrival at first-in-queue, entry into the server, and release from the server. Nothing of any significance transpires in between these events. The simulation model uses random variables to determine when each vehicle will enter the system (join the back of queue) and how long it will spend being processed (from two separate random distributions). It uses this information to schedule the events. Because nothing happens in between these events, no computations are performed to enhance computational efficiency.

Using math symbols, time can be described as advancing from time t_k, the k^{th} value of time to time t_{k+1}, the $k + 1^{st}$ value of time. The increment of time, or time step, can either be of a constant size Δt (e.g., 0.1 second) in the case of time-based simulations or of variable size Δt_k (i.e., Δt is not constant, but varies with k) in the instance of event-based simulations. That is, a time-based simulation has a constant step size, such that $t_{k+1} = t_k + \Delta t$; and an event-based simulation has variable sizes for the time steps, $t_{k+1} = t_k + \Delta t_k$.

In either case, it is assumed that no decisions need to be made between t_k and t_{k+1}. That is, the state of the system at t_{k+1} can be completely determined by the status quo at t_k and the decisions that were made at that point in time. Put another way, there has to be certainty about how the state of the system is going to change between t_k and t_{k+1}. Nothing of significance can be overlooked or missed.

For example, assume a vehicle at position $\underline{d}_k$ at time t_k with a velocity $\underline{v}_k$ (speed and direction) and an acceleration $\underline{a}_k$. The underline indicates that these variables are vectors—they have both a direction and a magnitude—that describe movement in three dimensional space. These pieces of information make it possible with certainty to predict where the vehicle will be at time t_{k+1} and how fast it will be traveling, that is $\underline{d}_{k+1}$ and $\underline{v}_{k+1}$. At time t_{k+1}, a new acceleration value can be obtained $\underline{a}_{k+1}$ and the process can repeat. This holds true for all vehicles in the system.
Put differently, if the time is incremented correctly, nothing happens to the status of the system between t_k and t_{k+1} that is not accounted for by evaluating the system status at t_k and then moving the system forward in time by Δt or Δt_k. Or alternately, if the times in between t_k and t_{k+1} were examined, the state of the system that pertains at t_{k+1} would be the same.

Time-Based Simulation

Simulation models that are time-based have a constant value for the time increment, Δt. Microscopic traffic simulation models are almost always of this type. The motivation is the fact that vehicles are moving in time and space, and that representing their kinematics (position, speed, acceleration) is very important. The time step increment is typically set at Δt equal to 0.1 second. Such a time step is short enough that nothing about the vehicle dynamics is likely to be missed or misrepresented, and the signal timing will be consistent with what happens in the real world.

For example, at 70 mph, a vehicle is traveling approximately 10 feet in 0.1 second. That is less than a vehicle length and/or the length of a detector. It is not likely that the movement of the vehicle will be misrepresented or a detector actuation will be missed. Insofar as signal timing is concerned, many controllers allow users to enter control values to the nearest 0.1 second. So if the Δt is set to 0.1 second, the simulated behavior of the controller will closely match that of the real world.

Discrete Event Simulation

Discrete event simulation models use a variable Δt that is a function of the (variable) event-driven time step size, k, and such becomes Δt_k. Events become the significant decision-making points, not time. The model steps from one event to the next and skips over the time in between. Queueing simulation models are often set up this way. In these simulations, the only things that can happen are: a) a new entity joins the queue, b) an entity begins its processing by the server, or c) an entity finishes its processing by the server. Hence, the model can step from one event to the next and no detail of the simulation will have been missed. An event list keeps track of the times when events are scheduled to happen. This ensures that no events are missed. The model can have thousands of entities, servers, and queues; but the only thing the model must keep track of is: When does the next event occur?

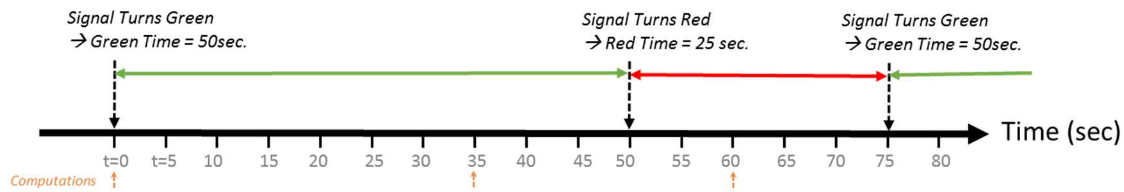
Source*:* FHWA.

A. Subfigure of pre-timed signal modeled as discrete event simulation.

Source: FHWA.

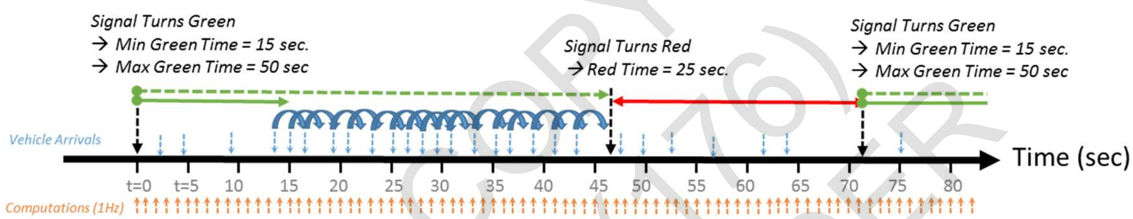
B. Subfigure of actuated signal modeled as time-based simulation.

Figure 5 presents a comparison of discrete event simulation and time-based simulation, using an example of a pre-timed traffic signal and a traffic-actuated signal, respectively.



Source*:* FHWA.

A. Subfigure of pre-timed signal modeled as discrete event simulation.



Source: FHWA.

B. Subfigure of actuated signal modeled as time-based simulation.

Figure 5. Illustration. Time-based and discrete event simulations.

Spatial Aggregation

Vehicle trip origins and destinations may be aggregated within a simulation model into groups of street links within a general area (such as a traffic analysis zone), or they may be assigned to specific street links. The link origins and destinations may be further disaggregated to specific parking lots and driveways. Generally, the finer the level of aggregation, the longer the processing times to simulate a given time period and the more precise the simulation model can be in predicting traffic on specific links.

The position of a vehicle within a link may be represented macroscopically, as being somewhere within a cell within a link (cellular automata), or microscopically, as being at a specific location, an exact number of feet downstream of the link's beginning end, and lane number.

Uncertainty Content

Simulation models are either deterministic or stochastic. If the model is deterministic, then no randomness exists. The event sequence is the same every time the model is run and the outcomes do not vary. If the model is stochastic, then randomness exists in one or more places. The event sequences can vary from one simulation run to another. As a result, the outcomes can be, and typically are, different (Traffic Flow Theory Monograph 2015).

Deterministic

A good example of a deterministic simulation model is one that predicts the travel times for vehicles traversing vertical and horizontal alignments. Forces are applied to make the vehicle accelerate or decelerate, so that it stops where it is supposed to, and it does not exceed the designated speed limits. The objective is to estimate travel times. A good example is the kinematics model used to estimate truck times for the mixed flow model in the HCM (2016). Another deterministic example simulates the operation of a pre-timed traffic signal. The number of cycles that occur during the simulation of a given amount of time is always the same. The amount of green cycles given to any given approach is also always the same, etc. All of the events, e.g., the durations of the green, yellow, and red intervals, always have the same durations. The event sequence is always the same. The entity interactions are defined by if-then relationships. Such simulation models have known inputs and they produce results that are always the same unless something about the model is changed. Oftentimes, these deterministic processes can be part of a larger stochastic simulation, such as a pre-timed signal in an otherwise stochastic simulation.

Stochastic

Stochastic models have randomness in one or more aspects of the logic. Queueing models are a good example: the headways between arrivals vary; the service time varies from one entity to the next, and as a result, the outcome varies. Random numbers are drawn from random number generators to determine what transpires. The average delay may be large for one run and much smaller for another.

In the context of a traffic simulation model, different vehicles may take different amounts of time to traverse a given section of freeway. Their decisions about lane changing may be probabilistic. With stochastic models, different values and decisions arise at each time step in each model run. Multiple runs are needed to obtain meaningful results. Simulation runs must be repeated until enough data are collected so that the predictions of system metrics are defensible. In the end, the distribution of the outputs is most important. Means and variances are important. The likelihood that a given outcome is going to be obtained is important. Section 5 discusses these concepts and implications for analysts in greater detail.

Queuing Models

Queueing is often a major focus of traffic simulation models. Vehicles queue upstream of freeway bottlenecks. They queue at signals while the light is red. Queues also arise in parking lots, upstream of incidents and work zones, and at toll booths.

A good reference on queueing models is Introduction to Transportation Analysis, Modeling, and Simulation (Möller 2015).

Some arterial network simulation models are effectively queuing models. Vehicles advance from one intersection approach to the next, and at each one they join the queue, advance to the stop bar, and are then released. Other instances where queues arise include toll booths, freeway ramps, parking lot entrances and exits, and transit stops.

Queuing System Parameters and Performance Measures

Every queue has three major components: the queue, server, and entities in the queue. Entities enter the system by joining the queue. They then wait to be serviced, receive their servicing, and depart (leave the system). Descriptive parameters include the distribution of the headways between arrivals, the distribution of the servicing times, the discipline used for queuing, and the number of servers that exist. A few of the more common measures of performance are as follows:

- Average headway: the average time between arrivals (based on the distribution of the headways).
- Average service time: the average amount of time required by the server to service an entity (based on the distribution of the service times).
- Average time in queue: the average amount of time that entities typically spend in queue waiting to be served.
- Average time in system: the average amount of time that entities typically spend in the system from joining the back of queue until being released by the server.
- Utilization rate: the percentage of time the server is busy.
- Throughput: the number of entities that leave the system per unit time.
- Mean queue length: the average length of the queue in vehicles or as a distance.

Note that while the list above describes *average* performance measures, the random nature of simulation and associated distribution of performance measures allows analysts to also compute varying percentiles from those distributions. For queuing systems, 85th percentile or 95th percentile queue lengths are commonly used as outputs that, in turn, drive the design and sizing of storage bays (e.g., length of a left-turn pocket at an intersection).

Queuing Disciplines

Queuing disciplines determine which entities will be processed after they join the queue.
First in, first out (FIFO), or first come, first served (FCFS), is a queuing discipline in which the entities (vehicles, passengers, etc.) that arrive first are serviced first. This is the typical queuing discipline in traffic. The vehicles waiting in queue at an intersection stop bar is a good example of this type of queue.

Queue jumping may be implemented in the field where bicycles or transit vehicles and other priority users may be allowed to bypass the queue of other vehicles. High occupancy vehicle (HOV) lanes at a bottleneck may be another example of queue jumping.

There are also several other queueing disciplines in the literature not commonly seen in traffic simulation: last in, first out (LIFO); random; round robin scheduling; priority queue; and shortest activity first.

For further reading on queueing theory, see the textbooks Traffic Flow Fundamentals (May 1990) or Introduction to Queuing Theory (Cooper 1981).

3.3 INPUT DATA

This section summarizes information on input data provided in Chapter 6. of volume II. Table 2 below summarizes data employed by the various types of models. Note that for completeness, the table includes both input data and data required for model calibration. Readers should consult chapter 6 for a fuller discussion of input data needs and sources.

Table 2. Summary of input and calibration data needs by simulation type.

Data	Description
Physical system inputs	Road and intersection geometry; vehicle fleet characteristics
Traffic control inputs	Speed limits; intersection controls
Demand inputs	Turn moves; major mid-block driveway volumes; and/or origin-destination table; vehicle mix (heavy vehicles, buses, etc.)
Calibration/validation data	Link speeds; travel times; link volumes; queue discharge flows; start and end times of queues; maximum queue length
Calibration parameters	Driver characteristics (aggressiveness); for macroscopic and mesoscopic simulation: capacity and free-flow speeds; for microscopic simulation: minimum headways and free-flow speeds

The basic data needs are similar for simulation, whether it is performed at the microscopic, mesoscopic, or macroscopic level. The primary difference is in the temporal and geographic resolution required. Macroscopic analysis can typically use hourly demand and signal timing data while microscopic analysis may require 5–15 minute detail. See Chapter 6. for more details.

3.4 PERFORMANCE MEASUREMENT

Performance measures are the outputs and metrics used to evaluate and quantify the performance of the system under study. However, each simulation software package computes performance measures in slightly different ways. It is thus important that users are aware of the definitions of various performance measures and their differences. This assures that they can understand the results provided by one package, and compare them to the results seen in other packages and in the field. A thorough understanding of performance measures is also critical for calibration and validation, and to assure that definitions between field-measured data and the simulation are compatible. For some performance measures, including delay and queue lengths, getting consistent measurements between field and simulation can be quite challenging (volatility of queues in the field; numerous parameters in the simulation model affecting queuing).

The most common performance metrics include:

- Travel times, travel rates, and their distributions.
- Delays and delay distributions.
- Spot speeds, space-based speeds, and their distributions.
- Queue lengths and their distributions.

A wide variety of metrics can be collected. The ones listed above are almost always included, but others include the number of stops, signal timing performance, lane utilization data, or even full trajectory output.[3] From the trajectory-level detail, simulation outputs can also readily be used to assess environmental impacts like air and noise pollution, or even to evaluate crash and incident rates (i.e., changes in these rates due to the alternatives considered). For simulation models with (dynamic) traffic assignment capabilities, shifts, increases or decreases in traffic demands are key outputs of interest.

Issue: Selection of Performance Metrics

In selecting performance measures, critical decisions include the need to:

- Determine what metrics will be monitored and reported.
- Determine how the system will be instrumented to collect data that can be used to create values for the metrics.
- Determine when the data should be collected and for what period of time.
- Determine what information will be derived from the observations obtained (e.g., only the average value(s) or the probability density function (PDF) or (CDF).
- Develop an understanding of how the values of the metrics might vary between the baseline conditions and the alternatives, and among the alternatives.

Issue: Statistical Assessment of Varying Performance Results

Performance measures can be extracted and reported through descriptive statistics (e.g., mean, median, variance), or can be reported as the underlying distributions through use of a PDF or CDF. The use of and emphasis on reporting distributions is growing in popularity because of recent emphasis on reliability assessment. For example, SHRP-2 project L02 (List et al. 2014) and SHRP-2 project L08 (Zeeger et al. 2014) stressed the importance of looking at the distribution of metrics, not just the means or similar single-point values.

Issue: Collecting Performance Data during Warm-Up and Cool-Down Periods

Another important consideration is the way the simulation runs are conducted. In other engineering disciplines, simulation analyses often focus on assessing the performance of the system across its duty cycle. A duty cycle captures the different operating conditions that the system encounters (or to which it is subjected) and the relative frequency (or duration of time) for which each condition arises. In a reliability assessment, the duty cycle might be an entire year

[3] A trajectory is a collection of location observations for a vehicle over time.

and the simulation model assesses the performance of the system in each operating condition that arises. In a peak period performance analysis, the duty cycle might be the temporal and spatial loads to which the system is subjected, from off-peak (midday or early morning) to peak loading, with demands varying by time and space, and back to off-peak (evening or midday). The analysis needs to capture the system's entire response to the load condition, from when free-flow pertains and there are no queues until the time when free-flow conditions once again pertain. The performance of the system is observed for the entire simulation, and the response of the system is completely recorded.
Traffic engineers typically refer to the before-peak condition as the warm-up time when the system is filling with traffic and the pre-peak-loading conditions are established. The system is then subjected to peak-load demands; and then a cool-down period follows where the peak-load conditions assuage and the system returns to post-peak-load conditions. For example, if the analysis of the morning (AM) peak period is of interest, the warm-up time would last from the start of simulation until steady state pre-peak conditions were established. A rule of thumb is that the warm-up time should be at least as long as twice as long as the longest travel time for any origin-destination (O-D) flow. Alternately, it should be long enough that all link flow rates, queue lengths and delays, and O-D travel times have stabilized. The cool-down period should last until all the vehicles that start trips during the peak period have reached their destination. Alternately put, all the link flow rates, queue lengths and delays, and O-D travel times have stabilized at post-peak values. In heavily congested urban areas with large networks, long simulation times may be needed to return to off-peak conditions; simulating the entire day (from early morning to late evening) might be required; and, so, sensible, carefully reasoned decisions might be needed to make the simulation analysis affordable.

In general, the plan should be to select (to the extent that modeling and data collection resources permit) a simulation period that starts in an uncongested condition and ends in a similar uncongested condition, recognizing and adjusting for less realistic performance results on the fringes of the simulation period as the model warms up from zero to the realistic demand values and cools down after the maximum demands are applied. The selection of warm-up and cool-down periods, and their treatment in the computation of performance results is addressed in more detail in Chapter 7. .

3.5 WORK TASKS, SCOPE, AND BUDGET FOR SIMULATION

Chapter 5 provides information on the general work tasks, scope, and budget for performing simulation analyses. This section previews that material. Simulation modeling software is designed to be applicable, through proper calibration, to a wide range of traffic conditions. Every simulation modeling effort must be calibrated against local real-world conditions (such as field observed speeds and vehicle volumes). The predictions of an uncalibrated model are generally misleading.

> The predictions of an uncalibrated simulation model are generally misleading.

As a rule, the development, coding, and calibration of a new simulation model is a significant undertaking. Some of the major considerations are:

- *Temporal and spatial bounds*. For example, should it be one hour in duration, two, or three? Should it encompass intersections (interchanges) upstream and downstream of the study area?
- *Vehicle composition*. Should it model different vehicle types? Should trucks be modeled separately?
- *Traffic flow patterns*. Should the demand inputs be based on O-D flow patterns or turning movement percentages?
- *Operating conditions*. How many operating conditions should be considered? What typical days should be used for calibration? Should the post-project completion operating conditions be simulated?

Although these are challenging questions, and the ensuing development of the model is both expensive and time consuming, once a model has been calibrated, the testing of project alternatives usually requires much less effort, especially when processing results from multiple model runs is automated in some way.

CHAPTER 4. CHALLENGES THAT CAN ARISE IN SIMULATION

This chapter highlights issues that can arise when using simulation. It is a precursor to volume II in that it shows where these issues can arise during the steps of model preparation and analysis. The appropriate chapter in volume II should be consulted for more details.

4.1 SCENARIO GENERATION AND SPECIFICATION

One challenge that can arise is the correct specification of the conditions to be examined. The simulations are incorrect if they cannot individually or collectively answer the questions of interest. If they do not adequately represent all the important conditions that might arise, analysis will be flawed and lead to inconclusive findings.

For results from a simulation analysis to really indicate how a system will perform, the runs should include scenarios that represent all conditions under which it will operate. Design engineers often refer to this as identifying the duty cycle of a device or system, or the set of conditions it will see and the relative frequency with which they occur. The Environmental Protection Agency (EPA) uses this idea to specify the driving cycles used to estimate the fuel economy of new vehicles.

One challenge can be proper temporal and spatial delineation of the scope of simulations. Most commonly, if the scope is too limited, in spatial or temporal extent, then the wrong results will be obtained; or, the modeling parameters will be tweaked inappropriately to produce the observed performance even though the causal factors are outside the bounds of the simulation, spatially or temporally.

Another challenge is to ensure that all the important operating conditions are included. In the SHRP-2 L02, List et al. (2014) introduced the idea of regimes to describe the operating conditions that may arise. For example, heavy demands in conjunction with bad weather would be a regime. Moderate demands with good weather is another. The regimes are defined as a combination of a demand level (light, moderate, heavy) and an environmental condition (nothing unusual, bad weather, an incident, a work zone, excessive demand, or combinations of these). Environmental does not mean air quality, etcetera, in the sense of environmental impact or concerns; but rather the combination of external factors that are influencing how the system is performing other than the demands. Once identified, a probability of occurrence for each regime is identified based on historical data for the time frame of interest (e.g., the AM peak period on work days when schools are in session). If a reliability assessment is of interest, as suggested by the updated volume III of the Traffic Analysis Toolbox, and the time interval used for analysis is 5 minutes, then there would be 105,120 intervals across the year. The challenge is to categorize these intervals based on their operating regime [demands and external influences]. For example, [low demands and nothing unusual] will be one of the regimes (admittedly, of limited interest). If the demands are low from 11 p.m. to 5 a.m. and external influences, such as bad weather and incidents, occur 10 percent of the time during those hours, then there will be 23,652 intervals of 5 minutes ($0.9 \times 365 \times 6 \times 12$) that fall into this regime of [low demands with no external influences]. This constitutes 22.5 percent of the 5-minute intervals; that is, it occurs 22.5 percent of the time. The simulation analysis must reflect that frequency of occurrence.

Other research efforts, like SHRP-2 L08 by Zegeer et al. (2014), L03 by Cambridge Systematics (2013), and L13 by Tao et al. (2011), use variants of this idea. The benefits of this regime-focused idea are twofold. First, the analyst can have a clear sense of the load conditions that should pertain during the analysis period of interest. Second, the analysis is not confounded by performance impacts that are caused by other external influences. The SHRP-2 L08 project suggests techniques for generating these scenarios based on historic data about traffic, weather, and non-recurring events. Another good source of information is volume XI of the Federal Highway Administration's (FHWA) Traffic Analysis Toolbox series, Weather and Traffic Analysis, Modeling and Simulation, from the traffic analysis toolbox by Park et al. (2010).

4.2 STATIC VERSUS DYNAMIC ANALYSES

Another issue that can arise is picking the wrong type of simulation to conduct. Selecting static when the conditions are dynamic or vice versa. The analyst needs to be careful not to use a model with static inputs to simulate dynamic conditions. A good example is an analysis of a peak hour where constant O-D inputs are assumed. Clearly, this is not likely to be correct. In most instances, the O-D patterns during the peak hour are time varying. Then, the analyst adjusts the model parameters so that the predicted performance matches that observed to the best of the model's ability. This is completely wrong. The analyst has forced the model to match the performance produced by dynamically changing demands by adjusting the model parameters to match that performance based on constant demands. The reason the performance arises is because of the dynamic demands, not static values. If the calibrated parameter values are applied to any other situation, it is highly unlikely that defensible results will be obtained. It is critical that the real, dynamic inputs be used if the analyst is to calibrate the model so that the observed performance is matched.

In general, the conditions to be examined can be either static or dynamic; and, they can be deterministic or stochastic. Any pairwise combination is possible. It is important that the analyst ascertain what combination of these conditions pertains in a given setting.

> If the conditions are static and deterministic, then nothing of substance changes during the simulation and all model inputs are fixed in value. The *k*+ time step is just like the next, as illustrated in

Source: FHWA.

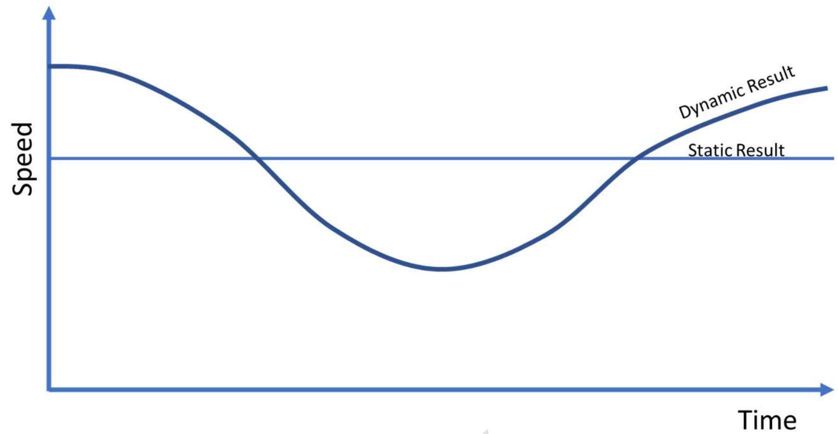
Figure 6. The simulation of a pre-timed signal is a good example. All the interval durations are fixed in length. A time-based simulation simply moves forward until the current interval ends and then the signal indications are changed.

If the conditions are static and stochastic, then randomness exists in some aspect of the system, but the inputs driving the performance are constant. A good example is the simulation of a toll booth's operation where the demands are constant, but the headways and service times vary. The queue length and delays will vary and CDFs of the delays, time-in-server, and time-in-system can be prepared, but no change in those distributions is occurring. Another example is the simulation of a traffic network late at night. Not that it is of great interest, but the O-D flows are not increasing or decreasing, and yet, because the vehicles are entering the network with varying headways, the travel times have distributions as do other aspects of the system's operation, such as the duration of the green times at signals, queue lengths, and delays.

If the conditions are dynamic but deterministic, the inputs are varying but there is no randomness in the simulation itself. This condition is uncommon, but an example would be a simulation of truck movements across an undulating grade where the headways are fixed but varying in duration, and the flow rate is increasing and then decreasing. The simulation model will predict truck decelerations and accelerations, maintaining safe headways, and as a result, travel times and delays. However, every simulation will be identical to every other one.

If the conditions are dynamic and stochastic, the inputs are varying and there is randomness in the simulation itself. This is the most common condition for simulation analyses. The performance of the network during a peak-load condition is of interest. The O-D patterns are changing across time, as well as signal control parameters, tolls, and other aspects of the system. Often, the question is: Where should capacity be added and/or how should the operating plan be changed to achieve better performance? Assembling the data for such a simulation is the most challenging of the four condition combinations; but, it is important to assemble the right data; because, if that is done, the model will be empowered to predict the observed performance. The inputs that caused the observed performance will be in use to drive the system's operation.

In the case of dynamic conditions, there is no single time step that is appropriate for performing an adequate analysis of a dynamic situation under all conditions. It varies according to the volatility of the conditions being evaluated. For example, shorter time periods may be needed if conditions (demand or performance) vary greatly within each 15-minute period.



Source: FHWA.

Figure 6. Illustration. Dynamic and static analysis results.

4.3 REAL-TIME VERSUS OFF-LINE SIMULATION

With increasing frequency, simulations are being run in real time. This is a challenge. The simulation model must be able to produce its prediction of performance in step with or faster than real time (for example, predicted traffic speeds for the next 15 minutes are produced in 1 minute of computer run time). Real-time simulation allows interactions between the simulation model and the control system, like signal controllers (which always run in real time), whose performance is of interest.

4.4 HARDWARE-IN-THE-LOOP, SOFTWARE-IN-THE-LOOP SIMULATION

An occasional challenge is to connect external hardware, or an external software module, to the simulation model. This is called hardware-in-the-loop or software-in-the loop simulation. The need might arise if the performance of a new traffic signal controller is of interest; or, the introduction of automated (software-controlled) vehicles into the traffic stream. If such situations arise, the analyst must couple the hardware for the device to the simulation model via a communications channel; or, the device software to the simulation model via a dynamic-link library (DLL). Historically, the interest in doing this has stemmed from the testing of new signal controllers. Each controller is unique, with special features; and, the generic signal control software provided with simulation packages is unlikely to contain these features. Hardware-in-the-loop is often limited in the number of controllers that can be run at any one time by practical constraints, such as available stacking space in the room, power supply, or simply the cost of purchasing or renting numerous units of the hardware.

Hardware-in-the-loop has the advantage of using the actual controllers with the actual installed software as would be found in the field. Software-in-the-loop runs the risk of slight to significant discrepancies between the version of the software emulated in the simulation model and the actual controller software installed in the field.

4.5 CHALLENGES IN ESTIMATING DEMANDS FROM COUNTS OR DEMAND MODELS

All simulation models need to have demand inputs, either via explicit specification of the O-D flows or entry flows and turning percentages. Estimating these travel demands is difficult. Numerically, if a network contains 100 locations where O-D demands can originate or terminate, then, excluding the 100 flows that are to and from the same location, 9,900 O-D pairs exist. So, at a minimum the flow rates for 9,900 O-D pairs must be provided. Of course, many of them may be zero. Moreover, if the originating flows change every 15 minutes, then the values for 39,600 originating O-D flow rates must be provided. However, it is unlikely that enough information is available to determine what these values are; although, the situation is becoming better as more probe data become available. If the O-D flows are estimated from the link flow rates. In that regard, if the network contains 300 one-way arcs (3 arcs on average departing each node, where a rectilinear grid would have 4), then there would be, at most, 1,200 observations of directional flow rates across the 1-hour time frame; and, even if the turning movement percentages are known for each 15 minutes, then no more than 3,600 flow rate observations are available. This means 3,600 observations are available to determine the value of 39,600 variables. No unique answer is possible. This is the Achilles' heel of both planning analyses and simulations.

Moreover, while it is possible to estimate demands from traffic counts under uncongested conditions, it is extremely difficult to do so when the network is congested. Then, only the vehicles that can make it through system capacity bottlenecks can be observed. There are also time delays between when the counted vehicles enter the network and when they exit. The counted vehicles at the exit ramp from a freeway may reflect demands that entered the freeway at least 5 minutes earlier. The exit ramp counts may also be reduced by vehicles stored in mainline queues on the freeway. The best that can be hoped for is to develop a reasonable estimate of the unobservable true demand and to recognize the limitations of that estimate. In addition, except through observations of probe vehicles, it is very difficult to ascertain the actual O-D flow patterns.

There are several methods to estimate traffic demands for simulation models, each with specific strengths and weaknesses, and include:

1. Estimating demand from turning movement counts.
2. Estimating demand from traffic counts.
3. Estimating demand from a demand model.

A brief summary of the issues is presented below. More details on estimating demand are given in chapter 6.

Estimating Demand from Turning Movement Counts

One way to estimate the O-D flows is implicit, through arc flows (directional link flows) and turning movement percentages. If the flow rate is 100 vehicles per hour (veh/hr) at an entry node and the downstream turning percentages are 10 percent lefts, 70 percent throughs, and 20 percent rights, then the O-D trip table implicitly has 10 percent of the flows that turn left at that intersection (to go elsewhere), 70 percent that go straight, and 20 percent that go to the right. These splits in the O-D flows occur in cascading fashion across the network until destination nodes are reached. (There is nothing to keep the flows from looping around sequences of nodes.) If these turning-percentage-based flows are traced out for all origins, an O-D trip table is produced. It is consistent with the turning movement percentages, but it may or may not be the actual O-D pattern to which the network is being subjected.
The user input percent turns are fixed. The user may change them from time period to time period within the simulation, but the model will not change those percentages based on any congestion that arises during the simulation.

This method results in significant resource savings for the user, in that an O-D table does not need to be generated by the user; however, the resulting O-D table is fixed. Traffic cannot be dynamically rerouted by the model to alternate routes based on congestion.

Another advantage of this approach is that the simulated traffic will more closely match the counted turning movements at each intersection.

Users must watch that the link counts are not constrained by upstream congestion or downstream queueing backing up into the subject link. Manual adjustments to the link counts may be necessary to better reflect true demands.

In addition, the use of turn probabilities may result in unrealistic circular travel routes for individual vehicles.

Estimating Demand from Traffic Counts

Link traffic counts with or without intersection turning counts can be used to synthesize the O-D table(s) for the network (using software available for demand modeling). Since there are always insufficient links and turning movements reflecting unique paths between each O-D pair, the estimation of the O-D table is an under-constrained problem. There are an infinite number of O-D tables available that will all produce the same observed counts, even if the analyst has counts for 100 percent of the links and turning movements in the network. (See the earlier discussion.) Consequently, O-D synthesis focuses on finding the best, or most likely, O-D table to produce the observed counts. The definition of best or most likely varies by O-D synthesis method. Generally, one starts with an O-D table developed by a demand model and tries to find a new O-D table closest to the original O-D table that will reproduce the counts.

The major advantage of this approach is that by using O-D table(s) as input(s), the analyst enables the simulation software to dynamically adjust the routing of vehicles through the network as congestion builds and declines throughout the simulation. A secondary advantage is that the final selected O-D table should reasonably reproduce the observed counts. The disadvantage is that capacity constraints in the network may cause some of the counts to be a poor basis for estimating the true O-D table or tables for the network.

The analyst must watch that the link counts are not constrained by upstream congestion or downstream queueing backing up into the counted link. Manual adjustments to the link counts and/or turn movements may be necessary (before O-D synthesis is applied) to better reflect true demands.

Estimating Demand from a Demand Model

It is also possible to use a demand model to estimate the O-D flows. This is often done, especially for future years. There are two difficulties often encountered when this approach is used:

1. The estimated O-D table for existing conditions (used for calibration purposes) may not match the day and times when travel time and link count calibration data was collected. This makes it difficult to calibrate the simulation model to obtain the observed travel times or link counts.
2. The forecasted O-D table for future conditions may greatly exceed the capacity of the transportation system to deliver those demands to the simulated network.

Both these difficulties are not hard to overcome with sufficient resources.

The existing O-D table can be adjusted to better match the link counts through the O-D synthesis process described earlier. The initial demand model O-D table (or tables) becomes the seed table to start the O-D synthesis process. As noted earlier, the analyst should carefully verify that the counts are not capacity constrained, and if so, the analyst must adjust the counts using professional judgement before they are used in the O-D synthesis process.

The future O-D table can be adjusted through a capacity constraint process to reflect the capacity constraints on the transportation network links feeding the simulation network study area. Again, an O-D synthesis approach is used. The forecasted O-D table is reduced until the demands on the external links feeding the simulation network are equal to their capacity. Note that this is done only for links feeding the simulation network, not the links exiting the simulation network.

The strength of this approach is that the O-D tables used in the simulation reflect the sophisticated data and methods used in the travel demand model to estimate demand.

4.6 PROBLEM-FOCUSED VERSUS RELIABILITY-FOCUSED SIMULATIONS

Simulation analyses are undertaken for a variety of purposes. The appropriate simulation approach depends on the purpose of the simulation.

If the purpose of developing and applying the simulation model is to solve a recurring congestion problem, such as weekday peak-period congestion that occurs in fair weather without crashes or incidents, then it is appropriate to simulate peak period, fair weather, nonincident conditions. If the purpose of the simulation model is to identify solutions, such as transportation system management and operations (TSMO) strategies, to address nonrecurring congestion problems that occur under incident or foul weather conditions, then the simulation should focus on a variety of demand, weather, and incident conditions. Multiple scenarios for each forecast year will need to be evaluated in the simulation model. Chapter 5. , volume II provides more details.

4.7 TRAJECTORY MANAGEMENT IN SIMULATION MODELS

This is a significant challenge. The ability of a specific software implementation of microsimulation to develop strategic trajectories[4] may be an issue of concern to the analyst. While almost all microsimulation software implementations employ car-following and lane-changing submodels to move vehicles within a link, these implementations vary in their abilities to accurately reflect how real-world drivers pre-position themselves for downstream turns or exit ramps.

In complex weaving situations this may be an issue. Various link-specific and vehicle-type specific manual overrides may be available in the software implementation to force or adjust the pre-positioning before a turn or an exit. The analyst might also write his or her own software DLLs[5] to supplement or override the default driver behavior models in the microsimulation software. Note that writing DLLs requires specialized software writing knowledge.

4.8 STOCHASTICITY IN SIMULATION

Stochasticity is often an integral element in simulation, most often in microscopic simulation. Instead of moving vehicles in lockstep down the link at the same speed with fixed spacings, the software introduces more realism by adding randomness in the speeds and spacings drivers select to move down the links. Stochasticity is also used to determine how quickly a driver may accelerate at a green light, how soon the driver changes lanes, how closely the driver follow other vehicles, and many other aspects of driving behavior for which analytical models and the necessary driver data are not available.

Stochasticity is introduced into simulation models through a random-number generator that is either part of the simulation software or provided by the computer's operating system. Random-number generators create reproducible sequences of numbers, typically between zero and one, that appear to be in random order. A seed value, selected by the user or the software, is used to start the random number sequence. When the same seed value is used, the random number sequence will always be the same. The analyst can (and should) introduce additional stochasticity into the simulation by using different seed numbers. Otherwise, even though the

[4] Strategic trajectories look several intersections ahead in plotting out the route.

[5] Dynamic link library.

sequence of numbers appears to be random, the simulation model will produce the exact same results since the number sequence is the same.

4.9 STATISTICAL ANALYSIS OF SIMULATION RESULTS

Simulation often generates a wealth of data. Analyzing that data is a challenge. Fortunately, statistical analysis tools are available to facilitate this activity. If stochastic simulation is employed (typical of microsimulation software implementations) then statistical analysis techniques will help the analyst appropriately characterize, qualify, and interpret the results. Most analysts are familiar with computing averages (a single point representing the midpoint of the simulation results). However, there is much additional insight to be gained by also looking at the distribution of the values: the standard deviation, the 85th percentile, the range, the 95th percentile, among others. Additional information can be generated by considering and computing confidence intervals for the results. Hypothesis testing may be used to quantify the likelihood of making an error in choosing one alternative over another based on the simulation results.

An introduction to statistical analysis (as well as discussion of advanced statistical techniques) is beyond the scope of this Transportation Systems Simulation Manual (TSSM). The analyst should consult the appropriate statistical analysis textbooks.

4.10 RECONCILING MACROSCOPIC AND MICROSCOPIC RESULTS

Sometimes the analyst may be presented with two sets of apparently divergent performance results. One set was produced by a macroscopic model. The other set was produced by a microsimulation model. The challenge is determining under what circumstances divergences in the results are due to shortfalls in the specific analysis method, errors in the specific application of the analysis method, and differences in how the analysis methods define and compute the performance measures themselves.

The problem is that macroscopic, mesoscopic, and microscopic use different fundamental approaches to computing their performance measures. Macroscopic analysis methods often extract their performance measures from queue accumulation polygons (QAP) and other macroscopic flow profiles. Microscopic analysis methods obtain performance measures from the vehicle trajectories. The two methods of computing performance measures therefore define each performance measure (such as queues or delay) differently, which will result in modestly different results even under otherwise identical conditions. The differences will vary from situation to situation and cannot be eliminated through a simple set of fixed correction factors.

Trajectory-Based Measures (Microscopic Simulation)

Figure 7 illustrates the basic concept of trajectory-based delay calculation on a signalized intersection approach.

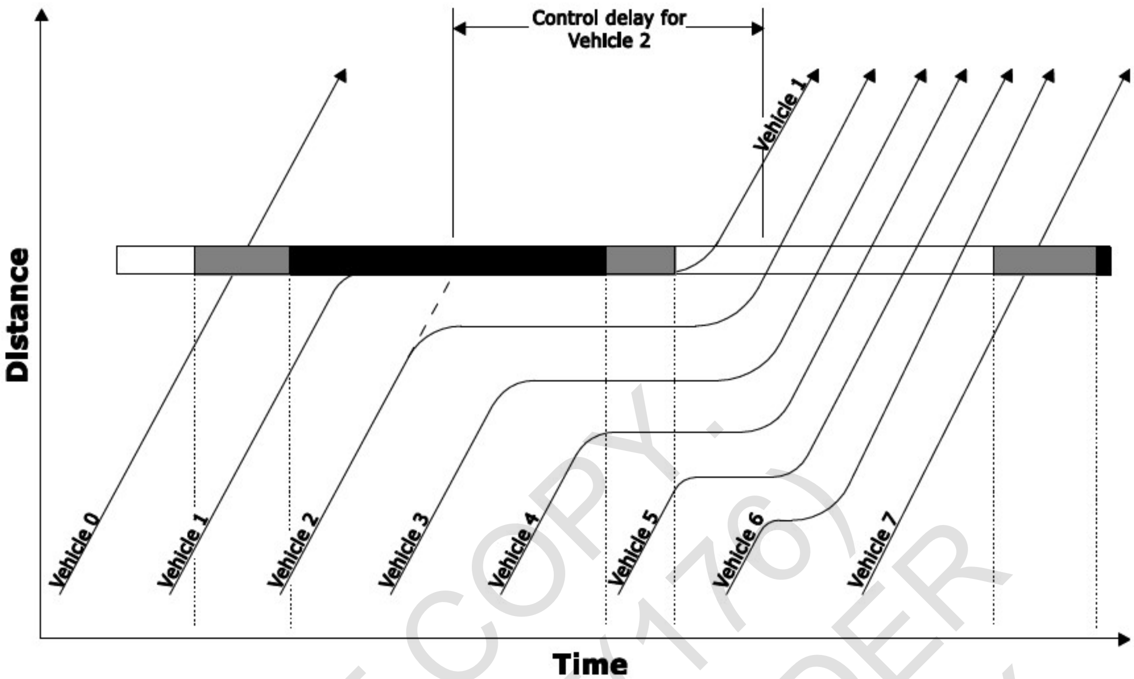


Figure 7. Chart. Trajectory-based computation for control delay.

Total delay is typically obtained by subtracting actual travel time from the (hypothetical) free-flow travel time for each vehicle. Note that free-flow travel time depends on the desired free-flow speed of each driver and will thus vary between vehicles. Stop delay may be assumed whenever vehicle speed falls below a certain threshold (e.g., 5 feet per second [ft/s] per the Highway Capacity Manual [HCM]). Travel time may be calculated as segment length divided by the average speed of each vehicle. To obtain density, the average number of vehicles on a segment can be divided by the segment length.

To determine queue-related outputs (e.g., average queue, maximum queue, 95th percentile queue), there must first be a rule or procedure for identifying a queued state. This procedure may consider the gap between a vehicle and its leader, vehicle speed, vehicle acceleration, and distance to the stop line. Refer to the HCM (2016, chap. 36) for more details on the recommended procedure for identifying a queued state.

Flow Profile-Based and Queue Accumulation Polygons-Based Measures (Macroscopic Analysis)

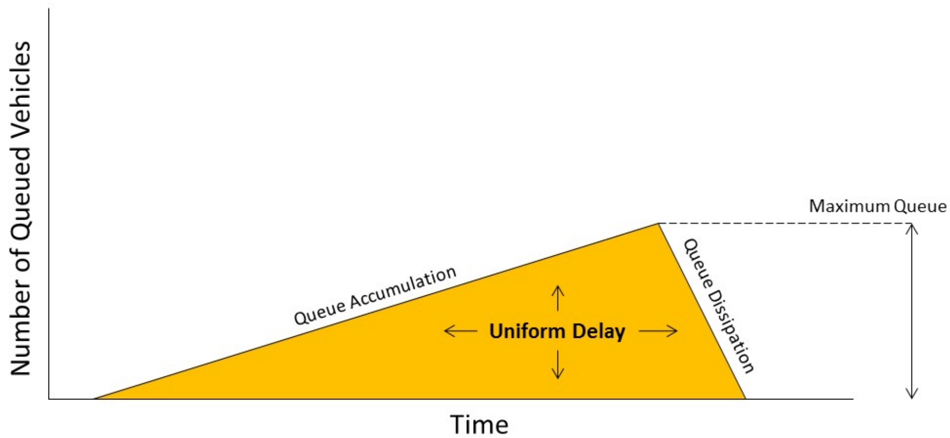
> In macroscopic analysis methods, the number of queued vehicles on a roadway segment accumulate when input flow rates exceed available capacities, or when traffic streams are temporarily stopped by a traffic signal. A sample QAP is illustrated in

Source: FHWA.

Figure 8. The queue will typically dissipate when a subsequent time period brings a lower input flow rate, or when the traffic signal turns green. After the queue has been fully eliminated, uniform delay can be computed as the two-dimensional area under the QAP. In order for delay to be computed accurately, the number of queued vehicles should be zero on both the left and right sides of the QAP. Indeed, this motivates a selection of time periods that avoids oversaturated conditions at both the very beginning and very end of a simulation. The maximum queue can also be obtained as the highest point of queue accumulation.

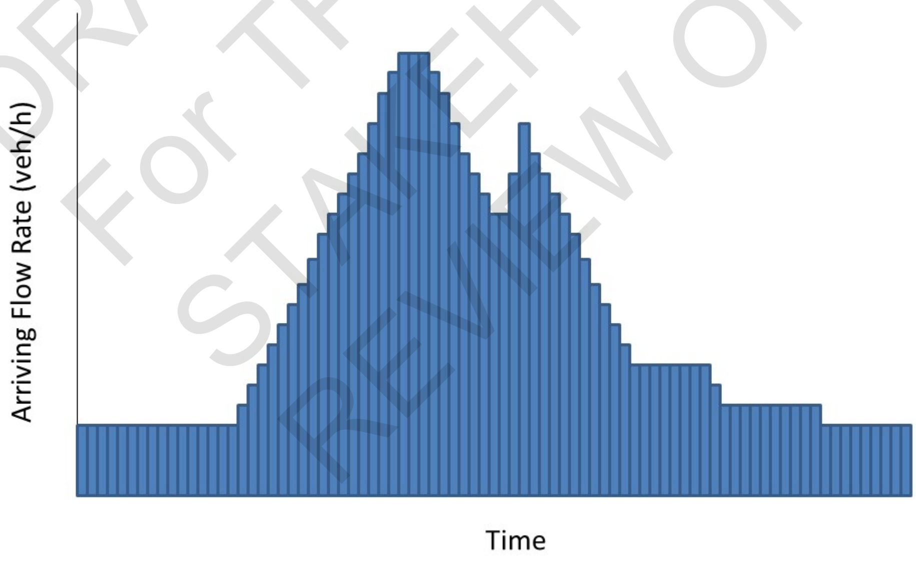
Source: FHWA.

Figure 9 illustrates an arrival flow profile, which can be used to determine a proportion of arrivals on green. Finally, if discharge flow and saturation flow profiles are constructed, two-dimensional areas under these profiles can be used to compute throughput and capacity, respectively.



Source: FHWA.

Figure 8. Illustration. Queue accumulation polygon used to compute multiple performance measures.



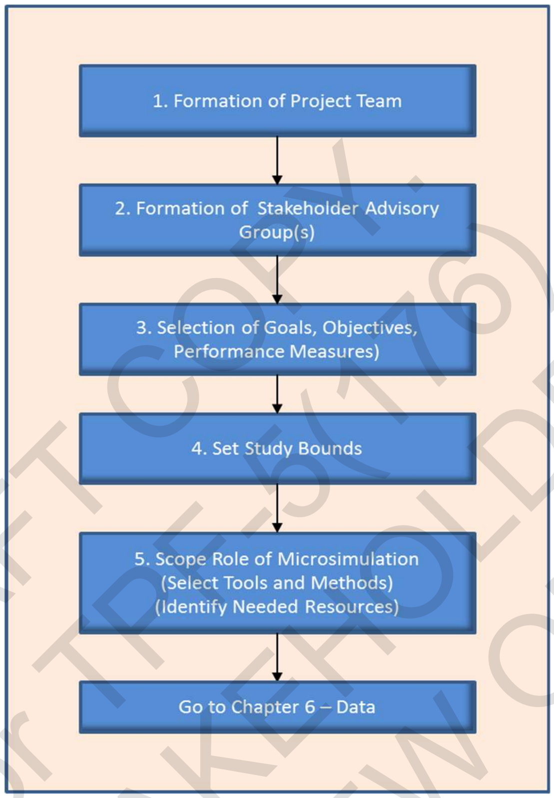
Source: FHWA.

Figure 9. Chart. Arrival flow profile used to compute proportion of vehicles arriving on green.

VOLUME II. THE PROCESS

CHAPTER 5. DEFINING AND SCOPING THE SIMULATION ANALYSIS PROBLEM

Defining the simulation problem to be studied and scoping the extent of the analysis are two important aspects of conducting a simulation analysis. The project team must set the scope and approach, and therefore the budget, and this has a profound impact on the ultimate success of the effort.[6] This phase of the effort has five steps, as shown in **Error! Reference source not found.**.



Source: FHWA.

Figure 10. Diagram. Microsimulation task flow (part 1).

The first two steps involve forming the project team and a stakeholder advisory group (if appropriate) to advise on scoping of the analysis. These steps are critical to ensuring the project team has correctly scoped the simulation analysis and that interested peers have been assembled to support scoping decisions. Since these steps are administrative, they are not addressed here in detail. However, it is important to realize that among the stakeholder group, there is a need for people who can critically guide and evaluate decisions about what to simulate (and, ultimately,

---

[6] The reader will find many similarities between this chapter and the scoping steps and themes in module 2 of the Federal Highway Administration (FHWA) guide, *Scoping and Conducting Data-Driven 21st Century Transportation System Analyses* (FHWA-HOP-16-072). The problem statement, project goals, identification of affected stakeholders, and bounding the problem are all critical elements of any analytical effort, whether using continuous monitoring data, microsimulation, or both.

who will evaluate results). Steps 3–5 deal with selecting performance measures, setting the study bounds, and scoping the simulation analysis portion of the project.

In defining and scoping the simulation analysis, the project team must ask and answer the following key questions:

1. What are the project objectives? Is it a planning, design, or transportation system management and operations (TSMO) project? What are the appropriate system- and project-level performance measures for measuring the abilities of design and operations alternatives to meet agency, stakeholder, and project objectives?
2. Given the expected range of design and operations alternatives and their likely effects on selected performance measures, what are the appropriate study bounds?
    a. Geometric bounds: exactly at the limits of the project, one intersection/interchange upstream, to the next upstream bottleneck.
    b. Temporal bounds: only the analysis period, 30 minutes before and after, the entire peak?
    c. Time frame: current year, design year, after-build, *n* years in the future, morning (AM) peak, evening (PM) peak, both peaks, other demand conditions?
    d. Load conditions: normal (nothing abnormal), adverse weather, incidents, both, abnormally high demand, during maintenance work?
3. Given the above, what is the appropriate role, schedule, and budget (scope) for the simulation effort in the overall transportation operations analysis? Put another way, where is simulation needed, and why? To what extent can other, less resource- and time-intensive tools be used? Can the scope of the simulation effort be reduced and still achieve the project goals?

The remaining sections of this chapter provide advice about how to answer these questions. They also provide examples of how they might be answered in specific settings. Each section addresses one of the three question areas.

While these key scoping questions are laid out sequentially, they are a group—a set of questions to be addressed. No doubt it will be helpful to consider them simultaneously. In that regard, it may be useful to employ a Delphi technique to iterate among the possible answers, bouncing the project team's thoughts off the stakeholders', and vice versa (or some other refinement process). It may also be useful to triage thoughts about scope and analysis options. For example, some questions might be addressed using less resource- and time-intensive tools, such as the Highway Capacity Manual (HCM) methods, to narrow down design and operation options to the most promising for more intensive simulation analysis.

5.1 SELECTING THE PERFORMANCE MEASURES

Performance measures play two roles in simulation analysis: they are needed to evaluate design and operations alternatives; and they are needed to evaluate the quality of the simulation model. An important question is: Does the model adequately replicate the system's operation so that the effects of future alternatives can be assessed?

Regarding the evaluation of alternatives, the selected performance measures will drive the choice of solutions. A successful analysis must therefore consider all stakeholder and agency objectives in its selection of performance measures to be produced by the simulation analysis. For example, the project team might want to focus on peak hour performance, choosing HCM-type metrics that focus on peak hour level of service (LOS). If this is done, then the simulation model, and the temporal and spatial scope, will be crafted so that LOS-related performance can be assessed. On the other hand, as is sometimes the case, safety is also of concern. However, most simulation models are weak at producing metrics that address safety-related concerns. It might be that supplemental, postprocessing tools are needed to develop safety-related metrics; or, that the spatial and temporal limits must ensure that areas (and times) of safety-related concerns are addressed. For example, accidents are often associated with increasing congestion, when queue lengths are growing, rather than when congestion is at steady state. Hence, the temporal and spatial bounds ought to capture the time leading up to the peak load and places where queues form.

> An excellent reference on tying agency goals and objectives to performance measures and the identification of project alternatives is NCHRP Report 785: Performance Based Analysis of Geometric Design of Highways and Streets (Ray 2014), available from NCHRP as a free download. This report focuses on geometric design; however, its recommendations are equally applicable to simulation analysis conducted to support performance based practical design (PBPD).

Table 3 provides an example of how a State department of transportation's (DOT) management goals may be translated to specific project performance measures. In this example, the transportation agency's goals are to improve the mobility, safety, environment, health, and economic welfare (economics) of its citizens in a cost-efficient (efficiency) manner. These broad agency goals are aligned with one or more specific project objectives. Each project objective is associated with one or two performance measures designed to measure achievement of that objective.

Note that additional tools (beyond simulation) may be necessary to address some of the performance measures shown in Table 3.

Being Holistic

It is advisable for the project team to involve internal and external stakeholders in the identification of stakeholder needs and appropriate performance measures for addressing those needs. This ensures that the project team will properly scope and budget the simulation analysis to address the needs of all critical stakeholders and decision makers who may influence the project direction.

Each stakeholder brings a unique perspective; and stakeholders outside the agency often have more specific information needs than those inside. For example, the local agency may be concerned about queues on individual local streets while the State agency is more concerned about overall system performance. The initially selected performance measures should be reviewed to verify that they also address all specific stakeholder needs or concerns. If there are gaps, then the project team should consider adding to the list of performance measures specific ones to address stakeholder concerns and decision-making needs. For this, the project team should make the stakeholders aware of the analysis impacts and simulation resources needed for such performance measures to see whether these considerations are cost effective.

Table 3. Example identification of candidate performance measures from agency objectives.

Agency goal	Project objectives	Example performance measures
Mobility	• Reduce delays. • Increase speed. • Improve reliability.	• Person-delay (PHD). • Average speed (PMT/PHT). • 95th percentile travel time (auto). • 95th percentile travel time (transit).
Safety	• Reduce crash rate. • Reduce queuing.	• Injury + fatality rate or SSMs. • No. of queue storage ratios > 1.00.
Environment	• Reduce pollutants. • Reduce GhG.	• NOx emissions. • GhG equivalent emissions.
Health	• Increase pedestrian and bicycle volumes.	• Bicycle and pedestrian volumes. • Pedestrian and bicycle crash rates.
Efficiency	• Maximize the benefit-cost ratio.	• Life cycle benefit-cost ratio.
Economy	• Increase access to jobs. • Reduce freight costs.	• Percent of jobs within xx minutes • 95th percentile travel time (truck).

Note: The selection of agency goals, project objectives, and performance measures is unique to each agency and project. This table is intended as an example of the variety of performance measures an agency might consider addressing its specific goals and objectives.
PMT = person-miles traveled. PHT = person-hours traveled. PHD = person-hours of delay. NOx = nitrous oxides. GhG = greenhouse gas. SSM = safety surrogate measures.

Objective-Based versus Tool-Driven Performance Measures

The selection of performance measures should be driven by the goals and objectives of the agency and stakeholders, not the capabilities of the simulation tool. For example, some of the performance measures listed in Table 3 cannot be produced by currently available traffic simulation software packages. Thus, the project team should consider how the scope of the analysis effort might be adjusted to supplement the simulation analysis as necessary with other tools and methods to produce the needed performance measures.

Safety versus Efficiency

Often, the trade-off in system design is between safety and efficiency (or productivity). Nominally, the primary objective is to ensure that trips can be made safely; and, the challenge is to allow them to be efficient as well, minimizing the resources required, such as time, capacity, land area, energy, cost, negative external impacts, etcetera. When this is the case, it is important to include metrics that address these aspects of the problem, as implicitly portrayed in Table 3.

Operational Performance Metrics

Metrics that focus on operational performance often measure total travel times, total delays, queue lengths, energy consumption, etc. An important thought here is that, for many of these metrics, the number of network locations that contribute to the value is quite small. For example, in the case of delay, it is only the places where bottlenecks occur. There may be delays in other places because traffic increases the travel rate (or decreases the speed), but, most of the delay accumulates at locations where the vehicles are standing still or moving very slowly. Hence, to improve system performance, the challenge is to improve the performance of these bottleneck locations (and, in the process, not allow hidden bottleneck locations to emerge). Hence, while aggregate metrics like total system delay may be interesting to examine, they do not provide a clue as to what needs to be done to improve performance. The delays at the bottleneck locations need to be examined. Hence, it might be as meaningful, or more helpful, to have a top 10 set of bottleneck locations and monitor the performance of those places and see what can be done to improve their performance, rather than endeavoring to tweak system-level options to achieve improvements.

Calibration Metrics

Calibration metrics are a separate thought from performance measures. It might be true that metrics useful for calibration are also useful for performance assessment; but, they also may not be. Calibration metrics make it possible for the study team to ensure the simulation model is producing credible results. As Chapter 8. explains, these calibration metrics focus on the behavior of bottlenecks and other locations that dramatically affect the ability of the model to match observed system behavior. Briefly, calibration metrics are likely to include the dynamics of queues at bottlenecks, travel times through sections of the network, flow rates on links, trajectories of vehicles passing through weaving sections, etcetera. These metrics need to be specified, and the field data to support them collected, but they may not be indicators of performance that will be presented to the public, or be of interest, broadly, to the stakeholders. Other calibration-related concerns include: Is the model loading all the existing and future demand onto the network (is traffic being denied entry)? Is the model adequately replicating the real world? Calibration performance measures likely to be useful are described in chapter 8. The selection of performance measures to determine if all demand has been loaded on the network is discussed in the following section, Setting Study Bounds.

Safety Performance Metrics

In some projects, there is an interest in improving safety, such as changing the geometry so that accidents are less likely. Many safety problems (except for driver distraction like texting) occur because the lead time for making complex trajectory management decisions is too short, the geometry is poor, and/or the signage is inadequate. However, simulation models do not directly output safety-focused statistics, like crash rates; and, inherently, they are programmed so that crashes do not occur. In addition, many simulation models do not yet couple the geometric data to vehicle behavior. Vehicles do not know to slow down because the turns are sharp; and, there is no connection to the signage.

Given the current capabilities of simulation models, the project team has two analysis scoping options to address these safety-related concerns:

- Estimate crash rates using tools based on the Highway Safety Manual (HCM). See AASHTO (2014). Manually adjust the geometric design or operational control to improve performance.
- Select safety performance measures that can be obtained from simulation models. These are called safety surrogate measures (SSM) or proxies for safety.

Examples of SSM are the probability distribution of headways at the bottleneck locations, the distribution of deceleration rates, the number of lane changes per lane mile, the percentage of lane changes for an exit that occur within *x* feet upstream of the exit, and the frequency of following vehicles being within 2 seconds of a lead vehicle. For whatever metric(s) are chosen, the project team should have at-the-ready references that indicate how and why these surrogate measures provide useful information about the likelihood that accidents will occur. More information about the selection of SSM and processing vehicle trajectories to estimate them from simulation output is provided in Chapter 9. and in Gettman and Head (2003).

Estimating Vehicular Pollutant Emissions and Noise

Sometimes, energy and pollution-related metrics are of interest. Some simulation software packages can directly estimate pollutant emissions and/or noise while others cannot. The project team should consult an emissions expert (or a noise expert, as appropriate) to ensure the emission rates model and noise model (if present) built into the simulation model will produce satisfactory estimates for the project stakeholders. The project team can also scope the analysis to include a manual link to an emission forecasting tool, like the MOtor Vehicle Emission Simulator (MOVES) model. See Environmental Protection Agency (2018). More information on linking simulation output to emission and noise models is provided in Chapter 9. .

5.2 SETTING THE STUDY BOUNDS

No analysis can be completed in a finite amount of time without setting bounds. This section covers the process of setting of geographical and temporal limits to the analysis, identifying the base and forecast years, and selecting the operational conditions (demand, incident, and weather scenarios) for evaluating the traffic operations effects of planning, design, TSMO, and other simulation alternatives.[7]

---

[7] The guidance in this chapter, by necessity, is high level, laying out general principles. For additional detail, readers should consult Traffic Analysis Toolbox, Volume III: Guidelines for Applying Traffic Microsimulation Software (Wunderlich 2016) and simulation software user's manuals.

Process for Setting Study Bounds

It is advisable for the project team to form and consult with a group of internal and external stakeholders and peers to confirm the setting of the following study bounds:

- Geographic limits.
- Temporal limits (peak period[s]).
- Base and forecast years.
- Operational conditions (demand level, weather, incidents, events, etc.).

While there are many possible processes for setting the study bounds, the following process describes one way to logically approach this task.

1. Identify the study bounds based on current recurring congestion conditions.
2. Test these bounds for continuing validity under future-project, project-alternative, and no-project conditions using various macroscopic tools, like the HCM.
3. Review the simulation outputs to verify the continuing validity of the study bounds under the various future scenarios, which may include nonrecurring congestion.

Step 1: Identify Initial Study Bounds

In this first stage, an initial set of study boundaries are set based on traffic counts and field observations of existing and potential future congestion issues. Traffic counts should be used to identify current peaking patterns. Field observations or other archived sources of performance data should be used to indicate times and locations of congestion.

Geographic and Temporal Boundaries

The temporal and geographic limits should be set so that they contribute to the accurate characterization of the system's performance. For example, if route diversions are anticipated, then the alternative routes for affected O-D pairs must be part of the model; else, external impacts are missed. The goal is to select study area limits and times of day for the analysis the project team expects will encompass a significant portion of the future congestion and expected impacts of the project under a variety of scenarios.

Picking an overly narrow geographic study area and temporal range saves on analysis resources and improves precision of the study results, but risks missing too much of the project impacts. Picking an overly broad geographic study area and temporal range imposes a greater burden on analysis resources, reduces precision of the results, and dilutes the project impacts by mixing them with a larger background of travel activity unrelated to the project.

The project team should take care not to develop a simulation model for too large a geographic area as this can lead to project budget over-runs, missed deadlines, and unsatisfactory results (often due to inadequate resources to collect data, and error check inputs and outputs).
The selection of study area geographic boundaries and temporal boundaries is ultimately a judgement call on the part of the project team, with input required from other specialty units who

will rely on the analysis output. The project team should consider other planned and programmed projects that may affect performance of the project under study, even if other projects are outside of the study boundary.

Note that when evaluating multiple improvement projects with overlapping impacts, the study area may need to be expanded to address the larger impact area.

Travel demand models and other macroscopic analysis tools can be used to assist the project team in setting the boundaries of the simulation analysis.

Forecast and Base Years

The appropriate base and forecast years for the mobility analysis should be selected to meet the requirements of Federal, State, and local stakeholders in the project. Evaluations of opening day and opening day, plus 20 years are common. Note that the further one forecasts into the future, the more complex the analysis and the less certain one can be of the results.

The base year for analysis should be selected based on stakeholder requirements and the availability of demand and facility performance data for developing and calibrating the simulation model.

Intermediate forecast years are often appropriate for large, staged construction or development projects. The year the facility opens is a good choice. See Chapter 10. . Intermediate forecast years may also be appropriate when funding is less certain for the later stages. Finally, intermediate forecast years may be appropriate when performing a performance based practical design (PBPD) analysis and when considering short-term improvement alternatives, such as TSMO strategies, that can be implemented much more rapidly than conventional capacity improvements.

It is often most cost effective for the project team to round up the desired forecast years to match those for which a statewide travel demand model or the local (metropolitan planning organization [MPO], city, or county) model have already prepared land use, demographic, and transportation network forecasts. The project team can then take advantage of land use data and transportation networks already coded for the area.

It is critical that the project team verify in advance of starting the analysis that the selected base and forecast years will meet all the requirements of the stakeholders for the project analysis.

Demand

Simulation modelers often confront the problem of having forecast years with demands significantly greater than the amount of traffic that can be physically delivered to the study area by the future transportation system. In such cases the project team must consider whether it is realistic to design future improvements within the study area to carry demands that cannot be delivered to it. This is a policy decision on the part of agency management regarding the future transportation system it wishes to build. If it is decided to design the highway improvements for external demands that are capacity constrained, dynamic traffic assignment (DTA) is one approach for generating constrained demands. Other approaches, such as gateway capacity constraint, are addressed in the latest edition of FHWA's Traffic Analysis Toolbox, Volume III: Guidelines for Applying Traffic Microsimulation Software (Wunderlich 2016).

Travel Time Reliability

The design of a facility to address travel time reliability effects is an agency policy decision. The analysis of reliability effects involves advanced simulation practices. These involve investigating system operation under differing demand, special event, incident, work zone and weather conditions. Particularly, TSMO strategies provide benefits under a variety of load conditions. Thus, evaluation of a range of conditions or scenarios is critical for estimating the benefits of TSMO strategies.

> Estimating the effects of design and operational improvements on reliability requires consideration of a range of operating scenarios, such as weather, incidents, and work zones.

Design and operational strategies will have different levels of mobility benefits depending on overall levels of congestion and specific causes of that congestion. Scenario analysis is used to understand how the mobility benefits of the design and operational alternatives vary under differing demand, weather, crash, and work zone conditions. The benefits of each alternative under each condition can then be better appreciated. The results can then be extended to an estimate of full-year benefits by applying probabilities of each scenario occurring over the course of 1 year.

If travel time reliability is to be simulated, the project team must identify the range of operational conditions appropriate for evaluating the reliability impacts of design and operational improvements for the project. This is best done by gathering a reasonable number of consecutive weekdays' worth (and/or weekends if appropriate for the study area and project) of traffic counts, weather reports, incident logs, special event calendars, and work zone logs for the study area. A full year of data, reflecting all conditions that arise during a complete duty cycle is most desirable.

To do these holistic assessments, the project team should consider employing a clustering analysis, such as the one described in the Traffic Analysis Toolbox: Volume III. Another source of guidance is the MathWorks article, Introduction to Cluster Analysis (MathWorks).

This section provides an overview of the procedures for how to use historic data to combine demand levels with weather and incidents for the purposes of testing design, TSMO, and operations alternatives under an appropriate range of conditions. More information and guidance can be found in Dallas Testbed Analysis Plan for AMS Testbed Development (Yelchuru et al. 2016).

A sufficient number of days of simultaneous historic demand, performance, incidents, and weather data are needed to identify the appropriate scenarios and appropriate weighting for each scenario. The identification of scenarios for evaluation follows these steps: [8]

1. **Identify Attributes** that best describe the load conditions.
2. **Assemble and process data** needed to describe the attributes quantitatively and make the clustering analysis possible.
3. **Normalize the attribute data** to common scale, such as 0–1, for the minimum to the maximum.
4. **Select attributes** that are most influential in distinguishing among the load conditions.
5. **Perform the cluster analysis** to group observed data into individual model scenarios. Repeat this step, and steps 1–4, as necessary to obtain a meaningful and manageable set of distinctly different load conditions.

A well-documented example of scenario analysis for microsimulation modeling can be found in the Dallas Testbed Analysis Plan, FHWA-JPO-16-373 (Yelchuru et al. 2016), freely available from FHWA to download.

The project team should take a strongly proactive role in the cluster analysis, using the statistical results to inform but not control the selection of load conditions to study. In the selection of these scenarios, the project team should take into strong consideration the likelihood that the design and operations alternatives will significantly affect traffic performance under each scenario. For example, if the proposed operations strategies will not operate on snow days, then there is little reason to include snow days in the simulation scenarios.

Once the clusters are identified, representative days need to be selected (from the data, not a hypothetical day). These days should be representative of the cluster. They should have attribute values that closely match average values for all days in the cluster. The simulation model should then be applied to and calibrated for the observed congestion on that day. The validation should then be conducted based on each of the days in each cluster.

The project team should take a strongly proactive role in the selection of scenarios for microsimulation analysis.

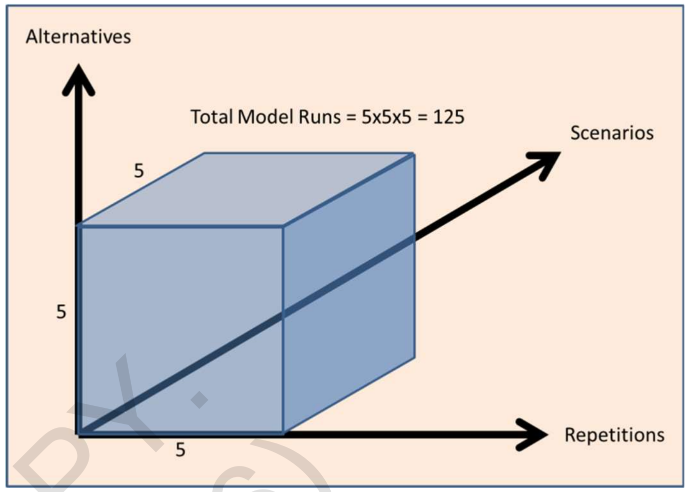
The project team should try to minimize the extent to which the parameters are distinctly different for each cluster. This is not to say they should not be changed; but rather that the changes should be logical and explainable, not just dial changes to achieve a good result. Examples of differences where cluster-specific values make sense are adverse weather, work zones, and incidents.

Adverse weather may affect car following, acceleration, and braking calibration parameters. Weather may affect the number of lanes open (snow removal). Work zones and incidents may affect the number of lanes open as well as car following and desired speed parameters due to distracted driving near the incident. Special events will affect demand levels, origin-destination patterns, and peaking.

[8] Using the procedures described in Chapter 2. of the Update to the Traffic Analysis Toolbox, Volume III, see Wunderlich, Vasudevan, and Wang (2016).

Guidance on calibrating a simulation model for different load conditions can be found in Chapter 8. .

For practical reasons, it is often necessary to seek a compromise between the extent of the cluster analysis and study budget. The clusters add a third dimension to the number of simulation model runs that must be completed. Increasing the number of clusters evaluated (even from one to two) multiplies the number of scenarios to be simulated. The Dallas Testbed Analysis Plan (Yelchuru 2016) provides an example of how the number of clusters can be managed through thoughtful consideration. They decided it was important to examine scenarios that combined two levels of demand with five external-factor conditions: no-external-influences (called dry) and four incident conditions (none, minor, medium, and severe). Among the 10 (2 × 5) combinations that could have been examined as a result, eight were analyzed using simulation.



Source: FHWA.

Figure 10. Diagram. Effects of scenarios on model runs.

The project team should include in the scope preparation of documentation of the cluster analysis approach used for determining analysis scenarios.

Step 2: Review of Study Boundaries for Future Conditions

Before coding the simulation model, the project team should employ a macroscopic tool like the HCM to do a quick validity check on the selected study boundaries under the various future scenarios and project alternatives. These checks should be repeated using the forecasted future volumes. The recommended macroscopic performance measures to use for evaluating study boundaries under future conditions are:

- Volume-to-capacity (V/C) ratios at key bottlenecks in the network.
    - While the peak hour V/C ratios can exceed 1.00, the study period should be long enough so the bottleneck V/C ratios over the entire planned study period (including the AM and PM peak periods, if applicable) for the simulation are less than 1.00 for all bottlenecks in the network. The entire system should be undersaturated at the beginning and end of the simulation (see Chapter 7. and Chapter 8.).
- Queue lengths for key bottlenecks in the network.
    - The peak hour queue lengths for bottlenecks immediately downstream of the entry links for the network should be estimated to ensure the median queue lengths are not expected to extend into the entry links.

Step 3: Review the Simulation Results to Assure Validity of the Study Bounds

Once the bounds have been set for the simulation model, the project team should review the simulation results to verify the validity of the spatial and temporal study boundaries. For example, the simulation analysis may reveal that the initially selected time duration is insufficient to clear congestion in the forecast year under the no-project alternative or that the spatial boundaries are inadequate to capture the back-of-queue locations on the study boundaries. The project team may want to expand the peak period, extend the geographic coverage,[9] or both. Alternately, if this is too challenging, they should take steps to address the issue post-simulation in computing performance results. See Chapter 9. for guidance on dealing with congestion remaining at the end of the simulation analysis period and for queues extending beyond the study limits (vehicles denied entry to the network during the simulation).

> It is desirable that the selected model bounds cover most of the forecasted recurring congestion under the variety of future alternatives under consideration. However, it may not be feasible to achieve this level of confidence in heavily congested networks or for nonrecurring congestion scenarios involving severe incidents. In such cases, it is generally not cost effective to continue expanding the simulation model limits. The project team should instead consider applying the output post-processing methods described in Chapter 9. to partially correct for this issue.

Assuming the simulation focuses on a peak-load condition, which is common, where the demands increase from an off-peak, low- or moderate-demand condition to peak load and then return to a post-peak, low- or moderate-demand condition, the recommended simulation model performance measures for assessing the validity of the study bounds are:

- Vehicles denied entry.
    - This is a running tally of vehicles generated by the model within each time interval (say each 15 minutes) that could not be placed on the network because of excessive congestion on an entry link (or load link). Some simulation models delete these vehicles from the reported results; there is no obvious record of the fact that their desire to traverse the network was not accommodated.
    - Vehicles denied entry should be as near zero as possible during the entire course of the simulation. The project team should revisit assumptions about the demand levels, the peaking of demand within the study period, and the geometry of the entry/load links if the vehicles denied entry are frequently significantly greater than zero. (For example, the entry links should perhaps be made long enough to ensure that sufficient storage is available to accommodate all demand, even if the lengths of those entry links are much longer than the lengths of the links in the real world.) Put in a negative manner, the inability of the external links to deliver

[9] Any significant additions to the geographic coverage of the model, such as adding more intersections, may require revisiting the initial model calibration. Lengthening an entry link may not require revisiting the initial model calibration.

the estimated demands to the network should not be a reason why the simulation model succeeds at providing good performance.

- Vehicle flow rates on exit links.
    - This performance measure can be used to validate the selected duration of the simulation model warm-up period (the elapsed simulation time until performance statistics are collected).
    - It can also be used to validate the duration of the cool-down period; and, implicitly, the duration of the simulation overall.
    - The project team should select the most important exit links to monitor. These are links with high volumes and/or ones that require a long time to return to post-peak flow rates. They also must be locations where the project team has a field count to compare the simulation results against.
    - The warm-up period (and the start of collection of performance statistics by the simulation model) should be long enough to reach consistency between the model's predictions of the exiting flow rates and those observed in the field. For example, if there was only one O-D pair, then the warm-up period must be long enough for that flow to fill the system, including queues at intermediate bottlenecks. The potential end of the warm-up period begins when the simulation model shows vehicles arriving at the destination (when the flow rate into the destination becomes non-zero.) At a juncture slightly later, the flow reaches a point where the exiting flow rate becomes consistent with the entering flow rate given the dynamics of the entering flow, the dynamics of the bottlenecks, and the time required to traverse the network (on the longest path used). A useful rule of thumb is that the warm-up period should be twice the duration of the longest O-D path travel time.
    - The cool-down period duration is a similar thought. More to the point, the duration of the simulation should be such that the exiting flow rates all return to post-peak conditions, consistent with field observations. This ensures all delays experienced by all vehicles that wanted to use the system have been captured. No delays have been missed, and consistent with the vehicles denied entry thoughts, no entering vehicles have been discarded by the simulation model because the duration of the simulation was too short to allow them to be serviced.
- Vehicles in the system.
    - This is a running tally of the vehicles on the network at any given time.
    - It is an instantaneous value that pertains at the end of (or the beginning of) a specific time interval.
    - While it is not a metric for which field observations are, today, available (it may be in the future as the density of probes increases), it is a very useful metric to examine to ensure that the temporal duration of the simulation is adequate.
    - Field-based values of the number of vehicles in the system can be estimated by computing densities and then multiplying by the lengths of the links. For example, if on a link, at the end of a 15-minute interval, the observed flow rate is 1,200 vehicles per hour (veh/hr) and the average speed is 30 miles per hour (mph), then the density of vehicles on the link is about 20 vehicles per mile (veh/mi) and the number of vehicles on the link would be 40 (20×2). Extending this thought to all

the links in the system produces the number of vehicles in the system at a given point in time.

- It is very unlikely that the number of vehicles in system ever reaches a steady state value because the O-D demands are constantly changing (even if a single set of O-D flow rates is used). However, it should stabilize, statistically, at a peak-load value, and then drop to a post-peak level (again in a statistical sense) at the end of the simulation.
- If, for example, the flow rates on the links in the system after the peak match observed flow rates for the post-peak condition, then the number of vehicles in the system (the flow rates multiplied by the lengths of the links and then by the duration of the observation interval) should also level off to the post-peak conditions.

5.3 SCOPING THE ROLE OF SIMULATION ANALYSIS

Properly scoping the role of simulation analysis is critical to the project's success. The project team must balance the needs of the agency and stakeholders for accurate information against the resources available and scale back the effort or seek additional resources when a mismatch is identified. That said, there are many alternative tools, like the HCM procedures (Transportation Research Board 2016), that can be used to obtain answers sufficient for reducing the range of alternatives or the geographical bounds of the study area to be evaluated.[10] The more costly simulation analysis methodology can then be focused on more promising alternatives at truly critical locations.

> Scoping balances decision-making needs against analysis resources. It identifies the most cost-effective tool to use at each stage of the analysis.

Identifying the Role of Simulation

Simulation is a very powerful tool, but it is also expensive and labor intensive. It should be used wisely.

The project team should start scoping the transportation analysis by recognizing that simulation is one of many special tools available to the project team for conducting the analysis. There will be stages of the analysis (especially early stages when study boundaries are being defined) where other tools like sketch planning, regional demand modeling, and HCM analysis may be more cost effective. Later, when alternatives are narrowed down, simulation becomes the cost-effective, and sometimes necessary, tool for evaluating the nuances among alternatives.

> Generally, the precision and detail of each stage of the analysis should be commensurate with the objectives of that stage of the analysis and proportionate to the complexity of the mobility and environmental issues being evaluated at that stage.

[10] When using alternative tools, the project team should be aware of the specific limitations of those tools. The TSSM cannot cover all the limitations of all the potential alternative tools.

For more information on the available analysis tools and their strengths and weaknesses for transportation analyses, see Traffic Analysis Toolbox Volume I: Traffic Analysis Tools Primer (Alexiadis et al. 2004) and Traffic Analysis Toolbox Volume II: Decision Support Methodology for Selecting Traffic Analysis Tools (Jeannotte et al. 2004).

Some additional thoughts are:

- Sketch-planning methods require the fewest resources and provide the fewest operational details. They are best for regional or statewide analyses where only a single areawide result is needed.
- State, regional, county, or city travel demand models require more detail and provide system and link level performance results. Their forté is demand forecasting. They are less precise for operations analysis and generally are not directly sensitive to the traffic operations improvements of TSMO strategies.
- Analytical models of traffic operations, such as HCM-based tools, provide very useful but limited assessments of the impacts from changes. They are not very sensitive to real-time control (traffic adaptive) TSMO strategies. They do not forecast demands, and generally work best in the analysis of individual freeways, streets, and intersections.
- Microsimulation models, when implemented in suitably capable software, provide the greatest detail in operational results and enable the project team to program in most any real-time adaptive control strategy. To the extent that queuing can be contained within the modeled geographic area and time interval, microsimulation models provide superior estimates of oversaturated conditions, especially when demand is greater than capacity. They can model systems of facilities and provide performance results in exquisite temporal and spatial detail.
- Mesoscopic simulation models, particularly those that include DTA capabilities, can provide an intermediate level of traffic operations analysis detail between travel demand models and microsimulation models. Mesoscopic simulation models provide more flexibility for the temporal shifting of demands than microscopic simulation models but less traffic operations analysis detail. Mesoscopic simulation models provide superior traffic operations analysis detail to travel demand models but with less ability to model trip generation, distribution, and mode shifts of travel demand.

While it is natural to seek the greatest precision in one's analysis, one must consider the objectives of the mobility analysis, context, resource and time constraints, and built-in limitations of the simulation tool.

Transitioning between Macroscopic and Microscopic Analysis

At earlier stages of the analysis, when design and operations alternatives are numerous, it may be practical to use other more static and analytical tools to filter out the less promising alternatives. These high-level analyses may also confirm or refine the initial decisions to bound the study. Intersections may be added to or dropped from the study area. The study hours may be expanded or contracted.

In the later stages of the analysis, with fewer alternatives to evaluate and with a firmer understanding of the appropriate temporal and geographic limits of the study, it becomes cost effective to employ simulation.

Interfacing Demand and Simulation Models

It is often necessary to tie a simulation model to a demand model to predict the effects of improvement alternatives. Here, a demand model means a model used by a metropolitan planning organization, or other similar entity, to understand the broad-brush, urban network-scale impacts of making capacity investments or operational changes. See Meyer and Miller (2002) for a thorough, although dated, discussion of these models.

> Combined with mesoscopic simulation, DTA provides a connection between demand models and microsimulation models.

A mesoscopic model that incorporates DTA can be used to create an interface between the planning-level model and a microscopic simulation model. For example, demand models tend to overpredict the O-D flow rates that can be accommodated by a network during a peak period. In the case of 20-year forecasts this can result in hourly forecasts for which it is difficult to obtain meaningful simulation results. The problem is that the planning model is assigning more traffic to the links than the capacity can accommodate (V/C ratios are greater than one) and the link travel times do not adequately capture the impacts of the traffic flows. A mesoscopic simulation model combined with DTA can reinterpret the planning model outputs and generate network loading trends more consistent with the capacity constraints. This meso-DTA model can also provide feedback to the planning model about more realistic link travel times so that the demand model produces more realistic demand forecasts.[11]

Software (Tool) Selection

Once the project team has selected a general approach to the analysis, there may be a question about what software (tool) to use. Sometimes, only one choice is available, as mandated by agency policy. However, it is always important to carefully consider the analysis capabilities of the tools available even if it is only one. There are many commercial or public domain computer software tools available to do simulation analyses. While this Transportation System Simulation Manual (TSSM) provides only general advice on methods and tool selection, the selection of the appropriate tools and methods for analysis is very important and the project team must be aware of the software tool's capabilities and limitations. The project team should also consider the availability of that tool to parties who will be reviewing the analysis, as well as the accuracy and availability of raw data attributes/data sources that feed into the selected tool.

[11] For additional information on DTA, Dynamic Traffic Assignment: A Primer. (Chui et al. 2011). Retrieved from onlinepubs.trb.org/onlinepubs/circulars/ec153.pdf.

Prototype Simulation Analysis Scope

The following bullets provide a prototypical traffic operations analysis scope in which simulation is one of the methods used:

1. Project scope.
    a. Identify study objectives considering stakeholder needs.
    b. Select performance measures.
    c. Select analysis approaches for each stage of analysis.
    d. Set model calibration targets.
    e. Estimate staff time.
2. Data collection.
    a. Define data collection methods (e.g., field collection, agency archives, crowd sourced).
    b. Gather data, base maps, inventory.
    c. Field observations.
    d. Prepare data collection report.
    e. Deliver database.
3. Base model development.
    a. Input data.
    b. Develop quality assurance.
    c. Error Checking: Review inputs, Review animation.
4. Calibration.
    a. Develop operational scenarios.
    b. Compare model performance to field data for each scenario.
    c. Adjust model parameters as needed for each scenario to calibrate to targets.
    d. Prepare scenario development and model calibration report.
5. Alternatives analysis.
    a. Forecast demands, capacity constrain demand using DTA.
    b. Base case analysis.
    c. Project alternatives analysis.
    d. Review results for reasonableness.
    e. Prepare alternatives analysis report.
6. Final Report.
    a. Key results.
    b. Visualization of results.
    c. Technical documentation.
    d. Delivery of model files and user's guide to explain use of files.

Source: FHWA.

Figure 11. Illustration. Prototype simulation analysis scope.

CHAPTER 6. DATA

6.1 INTRODUCTION AND DATA LANDSCAPE

Data collection is one of the most critical steps in a simulation-based study. As models are calibrated to the data collected, the true validity of any model is dependent on the quality of data it is fed. See Wunderlich et al. (2016), Virginia Department of Transportation (2015), and Transportation Research Board (2016) for good discussions on this topic. As indicated by the papers in Antoniou et al. (2014), availability of real-world data can improve reliability of the models' predictions. Therefore, it is critical to identify key data required for the study very early in the project. To properly conduct traffic operations and/or safety analyses, the guide from Florida Department of Transportation (2014) indicates the transportation analysts need to make informed decisions on what data to collect, when to collect it, where to collect it, how long to collect it, and how to manage the collected data.

As an indicator of the importance of data considerations in a simulation environment, a recent project in Europe (known as MULTITUDE) identified "absence of appropriate data" as one of 12 key issues of importance (Antoniou et al. 2014). The MULTITUDE team discussed the need for an optimized data resource library (or libraries) to aid in the improvement of simulation modeling, particularly in the calibration and sensitivity analysis procedures. The need for greater data quality and quantity was identified as a major conclusion of the MULTITUDE project.

Regarding the importance of data in the calibration and validation of simulation analyses, the MULTITUDE team stated:

> *Although there might be rare occasions where to do a study without any data, we strongly advise against this. If there are no data available, get them. If there are still no data available, increase your efforts to get them. Even if you learn from these data that the parameters in your model (as put into it by the manufacturer of the model) fit the data you have obtained quite well without any calibration, your study is now much stronger and more trustful than it was without data.*

The objective of this chapter is to establish a framework and provide guidance for obtaining and applying a wide range of data to the simulation enterprise.

6.2 TRAFFIC DATA MEASUREMENT

Generally, in the United States (U.S), traffic data can be derived from different types of sources:

- Government agency traffic management centers (TMC).
    - Point measures from sensors.
    - Segment measures from probe vehicles.
    - Incident locations/logs from closed-circuit television (CCTV) and TMC staff.
- Fleets or private organizations.
    - Segment measures from probe vehicles.

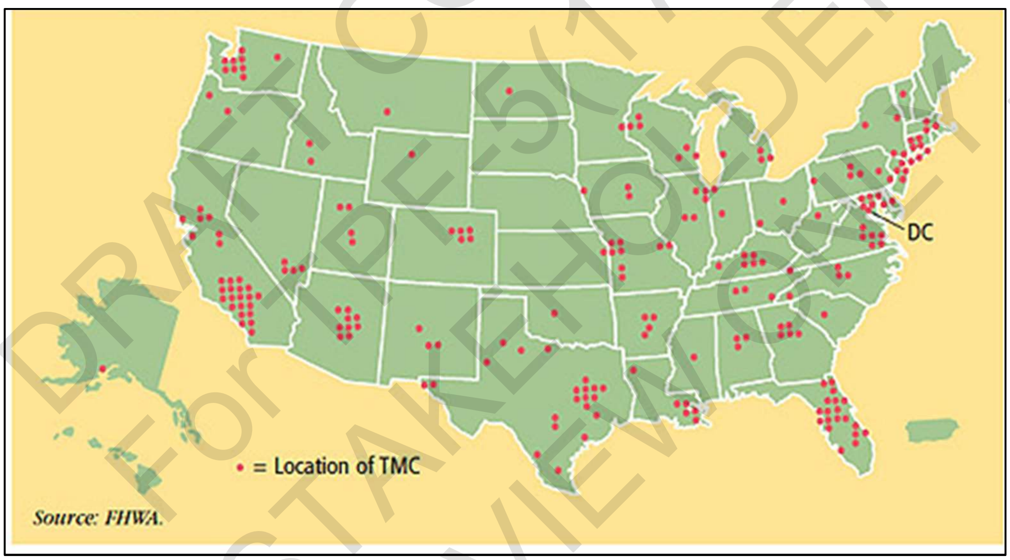
- Community crowdsourced.
    - Matched anonymous position data from cellular users.
    - Incident locations from communities of users.

Government Agency Traffic Management Centers

Most U.S. metropolitan areas have traffic management centers. As of 2011 there were 266 located across the U.S. (see

Source: FHWA.

Figure 12). The scope of TMC covers a range of functions, including freeway management (usually performed by States), arterial management (usually performed by cities), and emergency and incident response. TMC are operated according to local needs and may not be staffed 24 hours per day/7 days per week. Sensors and CCTV cameras usually feed information into the TMC and to the public via web-based services. Incident management may include surveillance, verification, dispatch, response, and resolution via the most appropriate agency (e.g., department of transportation [DOT] maintenance, law enforcement, and emergency medical response).



Source: FHWA.

Figure 12. Map. Locations of traffic management centers in the United States in 2010.

Fleets or Private Organizations

Travel times are also collected by fleet management systems. As of 2013 there were more than 1.5 million equipped vehicles owned by more than 10,000 fleet operators. Other fleets or private entities may provide travel time data that can be used by traffic data providers.

Community Crowdsourced

Several of the traffic data providers use matched anonymous position data from cellular users and incident locations from communities of users (e.g., Waze). Speeds are linked to roadway segments and usually converted to a color on the map (green-yellow-red spectrum). These are collected passively, as opposed to active crowdsourcing where personnel are present to actively place markers to aid data collection. For high-flow facilities, such as freeways and arterials, it is likely that the speeds/travel times are frequently refreshed. For low-flow links, however, it is not always clear how historical data are merged with real-time data. From a user's perspective, it is not possible to determine how recent the data is. Also, anecdotally and based on some prior research, the traffic data providers do quite a bit of smoothing, incorporating historical data that reduces the ability to detect abrupt changes to traffic conditions.

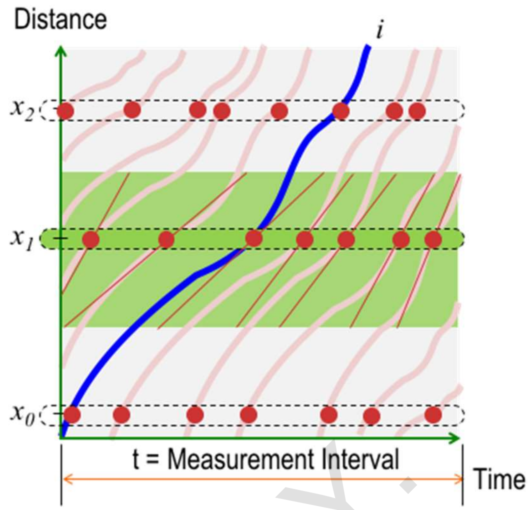
6.3 FUNDAMENTALS OF TRAFFIC MEASUREMENTS

Usually point-based sensor data (typically count, occupancy, and speed) are collected along roadway segments and made available free of charge to traffic data providers in simple-to-use formats (e.g., extensible markup language [XML]). Some regions use probe vehicles from tolling systems or license plate recognition to provide segment-based travel times that are displayed via variable message sign (VMS) and made available via XML. Detectors are technically able to record actual passage times, speeds, and lengths of individual vehicles. But traditionally these metrics are aggregated.

Source: FHWA.

Figure 13 is a time-space diagram (time on the *x*-axis and distance on the *y*-axis) that can be used to illustrate how point measures are used. A set of sample vehicle trajectories are shown. For example, at location x_1, a total of 7 vehicles are counted during the time interval shown, such that the flow during the interval t is $7/t$. Flow (also sometimes referred to as volume or throughput) is always measured at a point. Speed can also be directly measured with pairs of loop detectors, microwave devices, and other sensors.

Figure 14 also aims to illustrate how speeds are measured and extrapolated upstream and downstream to characterize a segment of roadway. The green shaded area illustrates the influence area of the detector at location x_1. Note that the actual trajectories in this situation are approximated as straight lines, assuming constant speeds over the influence area. In normal practice, point measures are usually aggregated over time intervals of 20 seconds, 30 seconds, 1 minute, 5 minutes, or even 15 minutes. Speeds measured from successive sensors can be applied to successive influence areas; the areas' travel times can be summed to obtain a corridor travel time.



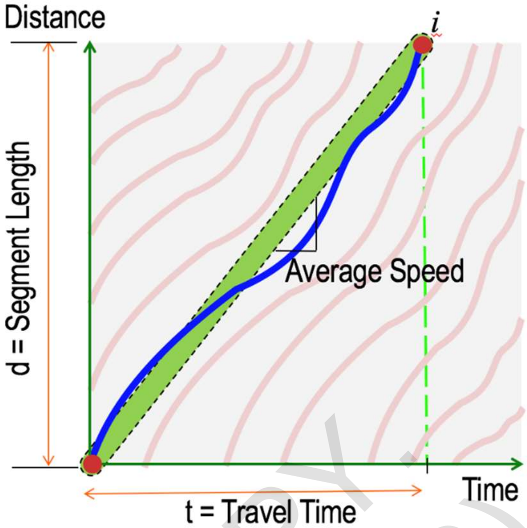
Source: FHWA.

Figure 13. Chart. Point measures of traffic conditions extrapolated over a highway segment.

In some situations, speed can be approximated by identifying a vehicle at two sequential positions along a highway.

Source: FHWA.

Figure 14 illustrates the method for estimating a vehicle's speed over a segment of highway, where its position is recorded at the two locations shown (separated by distance *d*) and two times (represented by travel time *t*). The vehicle's average speed over a distance, or space mean speed, is calculated as *d*/*t* (space mean speed).



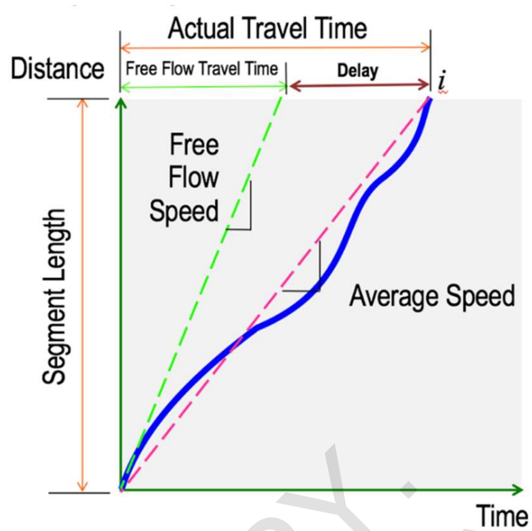
Source: FHWA.

Figure 14. Chart. Vehicle position matching speed estimation method.

In transportation, delay is often used as an important performance measure.

Source: FHWA.

Figure 15 shows that delay is defined as the difference between free flow travel time and actual (or average) travel time.



Source: FHWA.

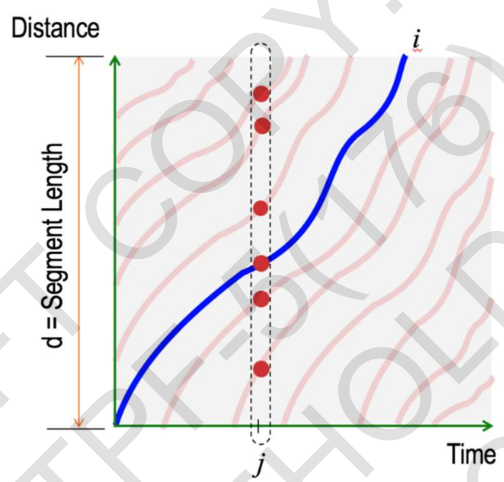
Figure 15. Chart. Definition of delay.

In addition to the commonly used floating-car method (used for calculating both speeds and travel times), actual speed can be measured via sensors or probe vehicles. The preferred method of determining free-flow speed is observing traffic speeds during periods of low flow (e.g., non-congested, off-peak daytime hours), though some practitioners use a posted speed limit (or posted speed limit plus 5 miles per hour [mph]) as an assumed free-flow speed.

For completeness, it is worth mentioning that some fundamental traffic related measurements are defined spatially but in practice are difficult to obtain.

Source: FHWA.

Figure 16 illustrates that density is defined at time *j* over segment length *d*. In this case 6 vehicles are observed over distance *d* at time *j*, resulting in density 6/*d* vehicles/distance (typically reported as vehicles per mile [veh/mi]). Density is often used as a performance measure but is difficult to measure in practice. Even though density is not a commonly available data source right now, the introduction of data collection by UAVs is poised to bring about new possibilities in this regard.



Source: FHWA.

Figure 16. Chart. Spatial measurements.

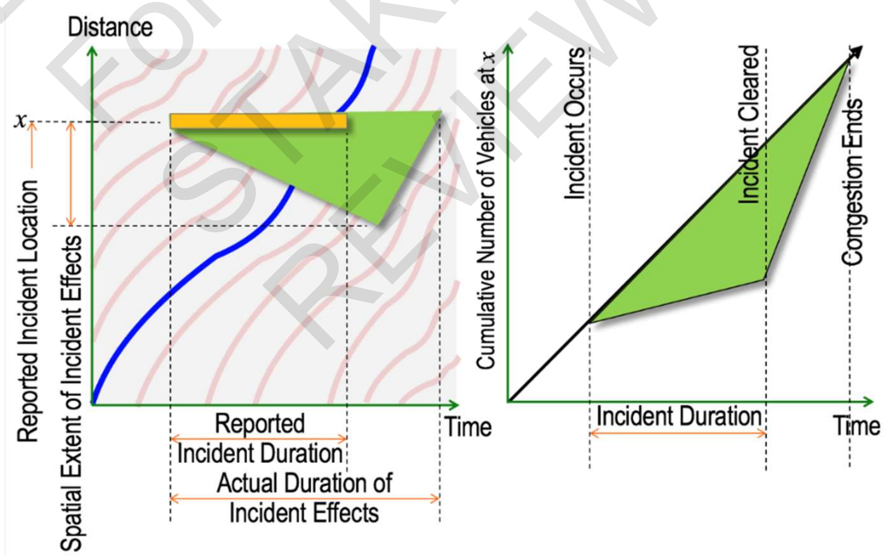
6.4 INCIDENT-RELATED TRAFFIC MEASUREMENTS

Incidents are defined as crashes, breakdowns, and other random events that occur on the roadway. Incidents (from both recurring and nonrecurring sources) contribute to a significant share of delays on highways, lead to blockages and road closures, increase drivers' exposure to hazardous conditions, and cause secondary crashes. Incidents also divert maintenance resources and reduce productivity.

Incidents are logged when TMC are staffed (typically during peak periods) via a computer-aided dispatch (CAD) system, which is sometimes linked with the region's emergency dispatch system (9-1-1 center). The logs include the incident start time and location (a point on the roadway, usually by milepost [MP] and direction; for example, northbound Interstate 95 at MP 23.45). When the incident is cleared, usually the same incident that was initiated is given an end time. In TMC, the spatial extent of an incident is rarely logged. Also, the end time of the queue or congestion triggered by the incident are generally not recorded. Sometimes, incidents are erroneously left open, or multiple logs for the same incident are entered. Incidents that are just congestion or a queue are not usually logged. When TMC are not staffed, or are understaffed, not all incidents will be captured.

Source: FHWA.

Figure 17 illustrates that the incident is logged at a specific point and its duration does not typically account for the spatial and temporal realities. State DOT often use the same dispatch system to indicate planned construction or maintenance activities, so there are often open incidents that cover long time periods that may not correctly indicate the activity's time, location, and duration. This is distinct from situations in other countries where an incident is typically defined by its spatial and temporal extents.



Source: FHWA.

Figure 17. Chart. Incident location and duration measurement in the United States.

6.5 DATA TYPES AND SOURCES

The precise data requirements for developing and calibrating a transportation analysis model vary based on the software tool used; however, they all require the following basic types of input:

- Geometric data.
- Traffic count data.
- Traffic signal operations data.
- Calibration data.
- Safety data.

6.6 GEOMETRIC DATA

Geometric data include any data required to characterize a network. See VDOT (2015) and FDOT (2014) for examples. There are many different types of geometric parameters but not all parameters are required for each analysis type. Further data may be required based on the selected analysis tool. The most recent version of the available aerial imagery must be used to field verify and validate the accuracy of data. Table 4 below lists typical geometric data required. Additional geometric data may be required based on the specific analyses.

Table 4. Typical geometric data required.

Geometric data	
Number of lanes	Interchange configuration
Location of preemption devices	Ramp length and radii
Approach grade	Acceleration/deceleration lane lengths
Lane widths	Distances to adjacent interchanges
Storage bay lengths	Distance to upstream warning signs
Taper lengths	Driveway spacing
Intersection approach widths	Median data
Shoulder widths	Distance to constricting infrastructure
Lane designations	Vertical and horizontal sight distances
Presence of pedestrians or bike lanes	Location of traffic control devices
Length of passing lane(s), if present	Location of stop bars
Length of no-passing zone(s)	Turn lane radii

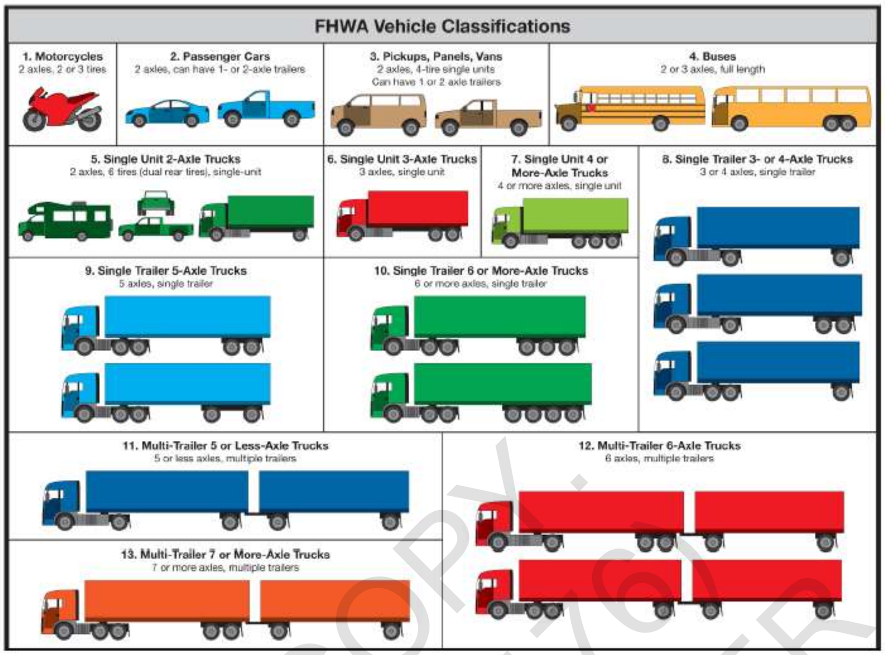
Roundabout approach widths	Splitter island locations
Ped/bike crossing distances	Transit stop locations
Bike lane/path cross slope	Transit bay location
Roadside shoulder slope	Transit reserved lanes

6.7 TRAFFIC COUNT DATA

Accurate counts lay the foundation for a well-executed traffic analysis. Traffic count data include all data necessary for documenting existing or future traffic conditions, including passenger cars, trucks, bicycles, and pedestrians. The Federal Highway Administration (FHWA) vehicle classification levels are shown in

Source: FHWA.

Figure 18 below. Since traffic counts and aerial imagery are constantly changing, care should be taken to ensure that the latest versions are used for the analysis. VDOT (2015) indicates that where possible, aerial imagery should be field verified to validate its accuracy. Even if the roadway has not experienced major geometric and/or traffic control changes since the data collection period, the appropriate age of the collected data should not exceed 2 years at the time of the analysis. In areas with seasonal variations in traffic, it may be worthwhile to apply seasonal factors. Seasonal factors (if needed) are available with most DOTs.



Source: FHWA.

Figure 18. Infographic. FHWA vehicle classification levels.

The various traffic count data parameters include:

Table 5. Traffic count data parameters.

Traffic count data	
Peak period turning movement counts	Speed data
Automated traffic recorder counts	Demand Profile
Short-term counts	Toll plaza and gate lane data
Parking maneuvers	Payment choices
Transit service data	Time of day restrictions
Vehicle classification data	Driveway spacing

6.8 TRAFFIC SIGNAL OPERATIONS DATA

For analyses that include either traffic signals or pedestrian signals, the guide prepared by VDOT (2015) suggests data on traffic signal operations are typically collected to analyze existing conditions on-site. The various traffic signal operations parameters that could be included in traffic analyses are:

Table 6. Traffic signal operations data parameters.

Traffic signal operations data	
Green times (minimum green/minimum initial and maximum green)	Vehicle extension and gap time
All-red time/red clearance and yellow time/yellow clearance	Recall mode
Cycle lengths	Time of day clocks
Offsets	Pedestrian crossing times (WALK and DON'T WALK)
Type of controller (NEMA, fixed time, etc.)	Transit priorities
Sequencing and phasing diagrams	Preemption timings
Actuation type	Ramp metering data (processing splits, capacity criteria, etc.)

NEMA = National Electrical Manufacturers Association.

6.9 CALIBRATION DATA

Calibration data are required to accurately develop a traffic simulation model. Without the calibration data, the analyst is unsure whether the model will accurately predict the performance on-site (VDOT 2015). Calibration data assist the analyst in adjusting the model to reflect local conditions. Calibration data for operational analyses shall be no more than 2 years old, provided the roadway has not experienced any major geometric and/or traffic control changes since the last data collection. It is considered good practice to calibrate a dynamic traffic model with both travel time/speed and count/flows, because as in reality, there are two conditions with the same flow and different speed. To collect queue data for calibration purposes, queue lengths should be collected at 5-minute intervals during data collection. This should be done by observing and documenting the length of queue for all movements. In oversaturated conditions, the entire extent of the queue should be observed and documented, even if the queue extends past an adjacent intersection.

The various calibration parameters used for traffic analyses include but are not limited to:

Table 7. Calibration data parameters.

Calibration data	
Peak period traffic demand	Driver behavior data
Pedestrian and bicycle travel speeds	Toll lane and gate processing time by payment choice
Mainline speed data	Travel times
Ramp speed data	Queuing data
Throughput data	

6.10 SAFETY DATA

Although safety is not a direct component in simulation models, traffic analysts often collect safety related data for critical assessment of transportation facilities. Crash databases are maintained by State DOTs and include information such as crash location, type, and severity. Safety data parameters collected for most traffic analyses include:

Table 8. Safety data parameters.

Safety data	
Historical crash data	Crash modification factors (CMF)
Field collected data	Safety calibration factors
Roadway alignment data	Crash and severity distributions

6.11 DATA SOURCES

Collecting data for traffic analyses may involve contacting multiple agencies to obtain data that they maintain. It is worthwhile to document what data was recently collected for other analyses so there is minimal duplication, if any. It may be possible to use existing data that was collected by other organizations. See Wunderlich et al. (2017) and VDOT (2015) for examples. Some sources that may be contacted for data include:

- City or municipal departments.
- County roadway departments.
- Regional councils of governments.
- Planning agencies.
- State departments of transportation.

Wunderlich et al. (2017) suggests that physical geometry can be obtained from rectified aerial photography and base mapping files prepared as part of the design effort for projects. Origin-destination (O-D) trips can be obtained from the local metropolitan planning organization's (MPO) regional travel demand model. Note, however, that these are generally 24-hour estimates, so they must be adjusted and refined to produce peak period estimates for use in simulation models. Other sources of O-D data include time-lapse aerial photography, feed from license plate readers, toll-tag data, data from analytics companies, and Bluetooth® detector network data. Vehicle characteristics data can be obtained from the State DOT or air quality management agency. Nationwide fleet data can be obtained from car manufacturers, the Environmental Protection Agency (EPA) and FHWA.

Variable message signs (VMS) in the study area can be obtained from geographic information systems (GIS) files, aerial photographs, and construction drawings. Data on transit are available with the local and regional transit operators. It includes schedules and stop locations, along with other data on dwell times at stops. Wunderlich et al. (2017) and Antoniou et al. (2015) suggest that event data can be received from public agencies, such as TMC logs. Crash databases should be verified since data may not always be recent and may not be for the specific study area.

Traditionally, vehicle counts are determined through traffic surveys or traffic detectors. In today's world of innovative technology, several different types of data sets are available to provide more accurate information such as probe data, connected vehicle data, crowdsourced data, and other big data sources. Big data has the potential to provide vast amounts of demand, vehicle, speed, flow, event, and behavioral data with minimal human intervention. Antoniou et al. (2015) suggests that although their market penetration is still not high enough to justify replacing traditional data sources, emerging data sources are alternative means for data verification or adjustment. The modern-day analyst can incorporate these emerging data sets into various traffic simulation analyses. It is important to note that these sources can be directly used to assess travel times, but they are just a sample for O-D matrices, i.e., they require an adjustment process with link counts to expand them and produce a demand that can be used as input for the simulation. Some details on these emerging data sources are provided in the following subsections.

First-Generation Probe Data

Probe data are obtained by wireless communications with global positioning system (GPS) enabled vehicles or mobile devices moving in the network and processed to characterize current and historic trends of traffic congestion. See Antoniou et al. (2015). These providers leverage location and speed data from many participants, and in coordination with historical data, present a comprehensive profile of travel time and traffic congestion. The data collected are processed in real time and meshed with historical records to provide a continually evolving model of traffic information, traffic forecasts, travel times, and traffic counts to businesses and individuals.

In addition to probe data providers, probe data technology solutions involve targeted collection of vehicle position and location data on specific routes, passively detecting and re-identifying vehicles moving in the transportation system through several methods. Traditional license plate recognition (LPR) is a technology used to count vehicles and estimate arterial travel time. Using cameras or CCTV, this approach applies optical character recognition to read vehicle registration plates. This is used by the police, as well as electronic toll collection agencies on pay-per-use roads. LPR can catalog the movements of traffic or individuals to generate path travel times and develop O-D matrices. Similarly, toll-tag technologies and Bluetooth® reader technologies can be used to match vehicles in one location that appear later in other parts of the network. These probe data are a readily available resource to the modern-day data analyst and have some key features of interest to fill gaps or support travel time analyses.

National Performance Management Research Data Set

The national performance management research data set (NPMRDS) is an archived speed and travel time data set obtained and sponsored by FHWA that covers the national highway system (NHS) and roadways near 26 key border crossings with Canada and Mexico. See CATT (Center for Advanced Transportation Technology) Lab (2017). Three sets of NPRDMS data are available that include speed and travel times at 5-minute intervals over 400,000 road segments: (i) passenger vehicles, (ii) trucks, and (iii) trucks and passenger vehicles combined. Unlike other probe-based speed and travel time sources, NPMRDS is updated monthly, not in real time.

Combining Probe Vehicle Trajectory Data with Segment Speed Maps

As an example of how probe vehicle data can be combined with segment speed maps, consider a sample that used a 4-mile section of Oregon Route 99W (OR 99W) from Durham Road to I–5. In this example, approximately 12 hours of probe vehicle data were collected over 5 days (66 runs that covered weekday peak, weekday off-peak, and weekend midday). GPS data were recorded at 1-second intervals. One-minute INRIX data were available for the TMC segments coinciding with the corridor, including:

- Time.
- TMC code.
- Speed.
- Average speed.
- Reference speed.
- Travel time (in minutes).
- Confidence score.
- C-value.

Source: FHWA.

Figure 19 shows a time-space diagram that contains a color plot of the INRIX speeds as a backdrop (time is on the *x*-axis and distance on the *y*-axis; speed is denoted by color, where red is slow and green is fast). The probe trajectories from the same day and time are plotted as well, with colors indicating their speed.

Source: FHWA.

Figure 19. Chart. Southbound Oregon Route 99W INRIX speeds with probes.

While comparisons are not the point here, B. Subfigure comparison of southbound Oregon Route 99W probe and INRIX times.

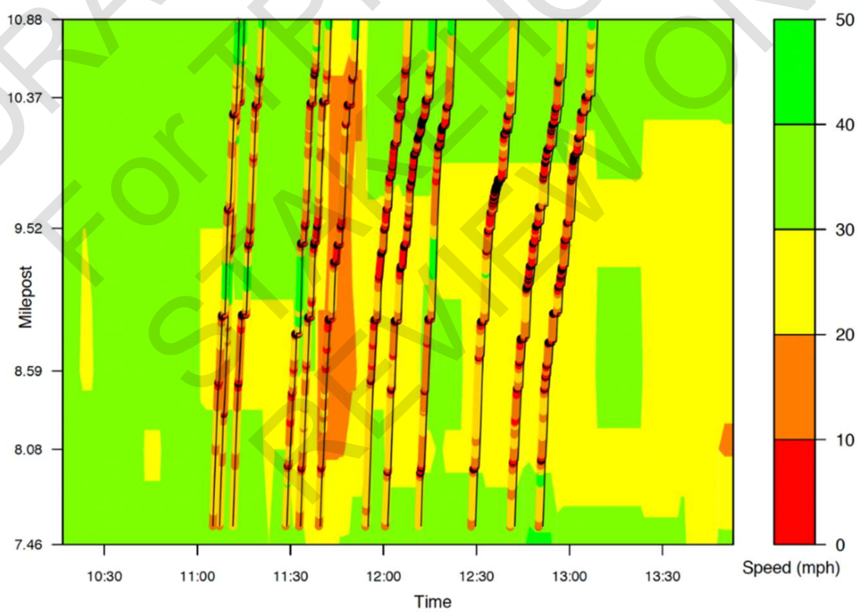
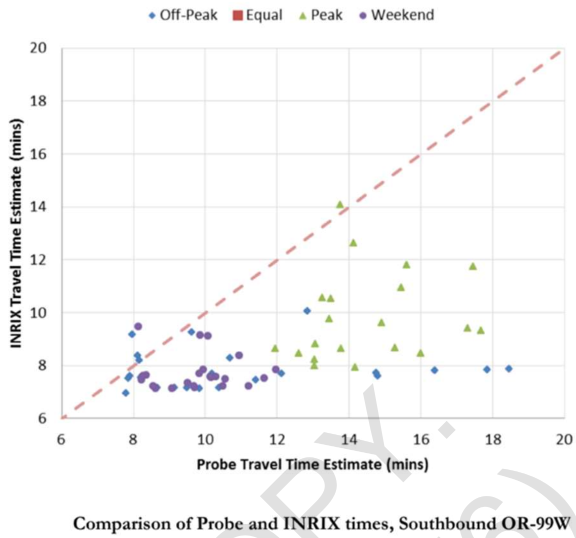


Figure 20 shows some travel time comparisons between the probe vehicles and the INRIX data on the subject arterial highway corridor.

Source: FHWA.

A. Subfigure comparison of northbound Oregon Route 99W probe and INRIX times.



Source: FHWA.

B. Subfigure comparison of southbound Oregon Route 99W probe and INRIX times.

Figure 20. Charts. Sample INRIX comparison to probe vehicle travel times.

Description of Big Data Sources

In this subsection, some sources of big data are described.

Traffic Data Providers

Traffic data providers such as INRIX, TomTom, Google® Maps, and HERE monitor more than 300,000 miles of U.S. and Canadian highways, providing real-time traffic information for more than 60 countries across North America and Europe. Data are collected from more than 300 million real-time anonymous mobile phones, connected cars, trucks, delivery vans, and other fleet vehicles equipped with GPS locator devices. INRIX also reports incidents and unique local variables while offering developers real-time traffic and routing information using application programming interface (API) access (*5*).

GPS Data from Large Vehicle Fleets

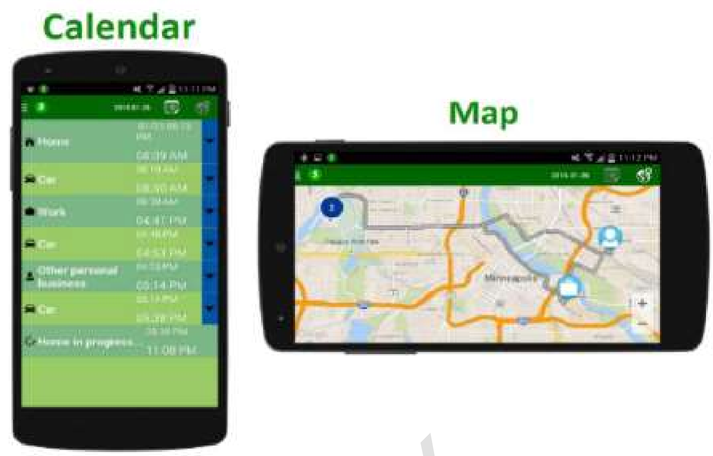
There are large vehicle fleets with individual GPS-equipped vehicles regularly transmitting location data to a central database. An example of such a system is the New York City Taxi and Limousine Commission, which collects more than 40 million records every year and integrates them into a single database (5). Empty trips are not included in this data set. Data are recorded per trip, including fields such as trip start and end times, trip duration, O-D information, and travel cost. Xie et al. (2013) show that similar taxi GPS data can provide rich information to calibrate traffic safety models. The taxis, even those without traffic surveillance units, can serve as probe vehicles. Therefore, they can provide more data sources (i.e., routing and travel time) for both analytical traffic as well as simulation models.

Cellular Network Data

Cell phone data have the potential to provide real-time information about mobility on a large scale at low cost. Cell phone location is usually a good proxy for the user's location. Thus, cell phone data can be used to provide traffic and travel demand data. Despite being available in large numbers, cell phone data still have many issues that need to be addressed before they are used for calibrating simulation models. Antoniou et al. (2015) suggest the primary issues include data coverage, sample bias, data availability and resolution, data suppression, and geographic level of detail. Since cell phone data lack the various demographic characteristics required for travel demand models, they may not be sufficient for independent use. However, they could be used in conjunction with baseline data, using travel demand surveys to establish a basic pattern and fine-tuning using travel demand estimates from the cell phone data.

Crowdsourced Data

Travel times can be collected using crowdsourcing data from online services that provide real time or historical traffic data. The web mapping APIs deliver several hypertext transfer protocol (HTTP) web services such as static map, directions, distance matrix, elevation, geocoding, and places. While web mapping applications provide very efficient methods to visualize large amounts of data sets, real-time traffic data are also available to users. The popular map service Google® Maps has more than 1 billion monthly active users in more than 200 countries who generate map data. The City of Chicago created dotMaps, an interactive web mapping service built on Google® Maps and Google® Cloud Platform that helps the city reduce unnecessary roadwork and ease traffic congestion. This initiative has improved quality of life, allowed for more efficient work management, and saved $24 million of taxpayer money by reducing duplicative work. The crowdsourced methodology can be an attractive approach to transportation agencies as an alternative low-cost data collection and surveillance method. An example is shown in Figure **21**.



© Fan et al. 2015.

Figure 21. Screenshot. Daily detail views for displaying predicted daily activities and trips from cell phone.

Event Data

Several agencies collect event data related to incidents and crashes as well as other related events. Events such as major construction, sporting events, concerts, or inclement weather may cause delays in the transportation network. For instance, Antoniou et al. (2015) suggests that Transportation Operating Coordinating Committee (TRANSCOM), an agency that coordinates activities of all transportation agencies in the New York and New Jersey regions, collects volume, speed, and travel time data through electronic readers during such events to enable calibration of simulation models for similar types of events in future.

Unmanned Aerial Vehicle Imaging Data

Unmanned Aerial Vehicles (UAV) are semi-autonomous or fully autonomous aircraft that can carry cameras, sensors, communication equipment, or other payloads. Prominent since the 1950s, Srinivasan et al. (2004) suggests that UAVs have been used since 2010 by Defense Advanced Research Projects Agency (DARPA) to increase applications in the military. Coifman et al. (2004) indicates that in transportation, UAVs have found applications in transportation operations and planning that allow more rapid assessments during incident response, monitoring freeway conditions, coordinating among a network of traffic signals, gathering traveler information, guiding emergency vehicles, tracking vehicle movements through intersections, and estimating O-D flows. Hansen (2016) says that UAVs have found applications in monitoring congestion as well as analyzing lane-change maneuvers in Denmark.

Connected Vehicle Data

The United States Department of Transportation (USDOT) initiated the connected vehicle research program (United States Department of Transportation 2018) to explore the potentially transformative capabilities of wireless technologies to make surface transportation safer, smarter, and greener and to enhance livability for Americans. © Fan et al. 2015.

Figure 22 provides a graphical illustration of a fully connected vehicle environment and the elements of vehicle data involved. For example, connected vehicle safety applications depend on a basic safety message (BSM), which provides basic vehicle information, such as vehicle size, position, speed, heading acceleration, and brake system status. Vehicles equipped with connected vehicle onboard units (OBU) will broadcast BSMs. Data providers have begun to combine vehicle sensor data with other sources to provide new data feeds augmented with connected vehicle data. For example, in-vehicle temperature sensor and traction control data can be combined with traditional atmospheric weather information to give drivers advanced warnings about dangerous weather-related road conditions, keeping them safer on their route.

© Fan et al. 2015.

Figure 22. Illustration. A fully connected vehicle environment.

6.12 DATA REQUIREMENTS BY ANALYSIS TYPE

Data requirements for building and calibrating traffic simulation models vary based on the type of analysis being considered. Table 9–Table 15 provide a complete overview of the various geometric, traffic count, and calibration data requirements per analysis type.

Table 9. Geometric data requirements for analysis types–signalized/unsignalized intersections, roundabouts, non-traditional intersections/interchanges, ped/bike lanes.

GEOMETRIC PARAMETERS	ANALYSIS TYPES			
	Signalized intersections, unsignalized intersections, and arterials	Roundabouts	Non-traditional intersections/ interchanges (SPUI, CFI, DLT)	Pedestrians and bicycles (on- and off-street)
Aerial imagery	✓	✓	✓	✓
Number of lanes	✓	✓	✓	✓
Location of preemption devices	✓	-	✓	-
Approach grade	✓	✓	✓	-
Lane widths	✓	✓	✓	✓
Storage bay lengths	✓	✓	✓	-
Taper lengths	✓	-	✓	-
Intersection approach widths	✓	✓	✓	✓
Shoulder widths	✓	-	✓	✓
Lane designations	✓	✓	✓	-
Presence of pedestrians or bike lanes	✓	✓	✓	✓
Length of passing lane(s), if present	-	-	-	-
Length of no-passing zone(s)	-	-	-	-

Source: Adapted from Virginia Department of Transportation (2015).
SPUI = single-point urban interchange. CFI = continuous flow intersection. DLT = displaced left-turn intersection.
- = data not required.

Table 9. Geometric data requirements for analysis types–signalized/unsignalized intersections, roundabouts, non-traditional intersections/interchanges, ped/bike lanes. (continuation)

GEOMETRIC PARAMETERS	ANALYSIS TYPES

	Signalized intersections, unsignalized intersections, and arterials	Roundabouts	Non-traditional intersections/ interchanges (SPUI, CFI, DLT)	Pedestrians and bicycles (on- and off-street)
Roundabout approach widths	-	✓	-	-
Ped/bike crossing distances	✓	✓	✓	✓
Roadside shoulder slope	-	-	-	✓
Interchange configuration	-	-	✓	-
Ramp length and radii	-	-	✓	-
Acceleration/deceleration lane lengths	-	-	✓	-
Distances to adjacent interchanges	-	✓	✓	-
Distance to upstream warning signs		-	✓	-
Driveway spacing	✓	-	✓	-
Median data	✓	✓	✓	✓
Distance to constricting Infrastructure	✓	✓	✓	✓
Payment choices	-	-	-	-
Time of day restrictions	✓	-	-	-

Source: Adapted from Virginia Department of Transportation (2015).
SPUI = single-point urban interchange. CFI = continuous flow intersection. DLT = displaced left-turn intersection.
- = data not required.

Table 10. Traffic count data requirements for analysis types–signalized/unsignalized intersections, roundabouts, non-traditional intersections/interchanges, ped/bike lanes.

	ANALYSIS TYPES

TRAFFIC DATA PARAMETERS	Signalized intersections, unsignalized intersections, and arterials	Signalized intersection preemption and transit priority	Roundabouts	Non-traditional intersections/ interchanges (SPUI, CFI, DLT)	Pedestrians and bicycles (on- and off-Street)
Peak hour turning movement counts	✓	✓	✓	✓	✓
Automated traffic recorder counts	✓	✓	-	✓	✓
Annual average daily traffic (AADT)	-	-	-	-	-
Parking maneuvers	✓	✓	✓	✓	-
Transit service data	✓	✓	✓	✓	-
Vehicle classification data	✓	✓	✓	✓	-
Speed data	✓	✓	✓	✓	-
Toll plaza and gate lane data	-	-	-	-	-

Source: Adapted from Virginia Department of Transportation (2015).
SPUI = single-point urban interchange. CFI = continuous flow intersection. DLT = displaced left-turn intersection.
- = data not required.

Table 11. Calibration data requirements for analysis types–signalized/unsignalized intersections, roundabouts, non-traditional intersections/interchanges, ped/bike lanes.

	ANALYSIS TYPES

CALIBRATION DATA PARAMETERS	Signalized intersections, unsignalized intersections, and arterials	Signalized intersection preemption and transit priority	Roundabouts	Non-traditional intersections/ interchanges (SPUI, CFI, DLT)	Pedestrians and bicycles (on- and off-Street)
Peak hour/period traffic demand	✓	✓	✓	✓	-
Pedestrian and bicycle travel speeds	-	-	-	-	✓
Mainline speed data	✓	✓	✓	✓	-
Ramp speed data	-	-	-	✓	-
Toll lane and gate processing time by payment choice	-	-	-	-	-
Travel times	✓	✓	-	✓	-
Queuing data	✓	✓	✓	✓	-
Existing crash data	-	-	-	-	-

Source: Adapted from Virginia Department of Transportation (2015).
SPUI = single-point urban interchange. CFI = continuous flow intersection. DLT = displaced left-turn intersection.
- = data not required.

Table 12. Geometric data requirements for analysis types–freeways/interchanges, two-lane highways, multilane highways.

GEOMETRIC PARAMETERS	ANALYSIS TYPES

	Freeways/ interchanges (merge, diverge, weave, collector-distributor)	Two-lane highways	Multilane highways
Aerial imagery	✓	✓	✓
Number of lanes	✓	-	✓
Location of preemption devices	-	-	-
Approach grade	✓	✓	✓
Lane widths	✓	✓	✓
Storage bay lengths	-	-	✓
Taper lengths	-	-	✓
Intersection approach widths	-	-	-
Shoulder widths	✓	✓	✓
Lane designations	✓	-	✓
Presence of pedestrians or bike lanes	-	✓	✓
Length of passing lane(s), if present	-	✓	-
Length of no passing zone(s)	-	✓	-
Roundabout approach widths	-	-	-
Ped/Bike crossing distances	-	-	-
Roadside shoulder slope	-	-	-
Interchange configuration	✓	-	-
Ramp length and radii	✓	-	-
Acceleration/deceleration lane lengths	✓	-	-
Distances to adjacent interchanges	✓	-	-
Distance to upstream warning signs	✓	-	-
Driveway spacing	-	-	✓
Median data	-	-	✓
Distance to constricting infrastructure	✓	✓	✓
Payment choices	-	-	-
Time of day restrictions	-	-	-

Source: Adapted from Virginia Department of Transportation (2015).
- = data not required.

Table 13. Traffic count data requirements for analysis types–freeways/interchanges, two-lane highways, multilane highways.

TRAFFIC DATA PARAMETERS	ANALYSIS TYPES

	Freeways/ interchanges (merge, diverge, weave, collector-distributor)	Two-lane highways	Multilane highways
Peak hour turning movement counts	-	-	✓
Automated traffic recorder counts	✓	✓	-
Annual average daily traffic (AADT)	-	-	-
Parking maneuvers	-	-	✓
Transit service data	-	-	✓
Vehicle classification data	✓	✓	✓
Speed data	✓	✓	✓
Toll plaza and gate lane data	-	-	-

Source: Adapted from Virginia Department of Transportation (2015).
- = data not required.

Table 14. Calibration data requirements for analysis types–freeways/interchanges, two-lane highways, multilane highways.

CALIBRATION DATA PARAMETERS	ANALYSIS TYPES		
	Freeways/ interchanges (merge, diverge, weave, collector-distributor)	Two-lane highways	Multilane highways
Peak hour/period traffic demand	✓	✓	✓
Pedestrian and bicycle travel speeds	-	-	-
Mainline speed data	✓	✓	✓
Ramp speed data	✓	-	-
Toll lane and gate processing time by payment choice	-	-	-
Travel times	✓	✓	✓
Queuing data	✓	-	✓
Existing crash data	-	-	-

Source: Adapted from Virginia Department of Transportation (2015).
- = data not required.

Table 15. Geometric data requirements for analysis types–work zone traffic, toll plazas, gated lanes, managed lanes, ramp metering, safety analyses.

GEOMETRIC PARAMETERS	ANALYSIS TYPES

	Work zone traffic (freeway or arterial)	Toll plazas and gates	Managed lanes or ramp metering	Safety analyses
Aerial imagery	✓	✓	✓	✓
Number of lanes	✓	✓	✓	✓
Location of preemption devices	✓	-	-	-
Approach grade	✓	-	-	✓
Lane widths	✓	✓	✓	✓
Storage bay lengths	✓	-	-	-
Taper lengths	✓	-	-	-
Intersection approach widths	✓	-	-	-
Shoulder widths	✓	-	-	✓
Lane designations	✓	✓	✓	✓
Presence of pedestrians or bike lanes	✓	-	-	✓
Length of passing lane(s), if present	✓	-	-	-
Length of no passing zone(s)	✓	-	-	-
Roundabout approach widths	✓	-	-	-
Ped/bike crossing distances	✓	-	-	✓
Roadside shoulder slope	✓	-	-	✓
Interchange configuration	✓	-	-	✓
Ramp length and radii	✓	✓	-	✓
Acceleration/deceleration lane lengths	✓	✓	-	✓
Distances to adjacent interchanges	✓	-	-	✓
Distance to upstream warning signs	✓	✓	✓	-
Driveway spacing	✓	-	-	✓
Median data	✓	-	-	✓
Distance to constricting infrastructure	✓	✓	✓	✓
Payment choices	✓	✓	-	-
Time of day restrictions	✓	-	✓	-

Source: Adapted from Virginia Department of Transportation (2015).
- = data not required.

6.13 EXAMPLE DATA COLLECTION INITIATIVES

This section provides examples of data collection initiatives across (1) a freeway corridor, (2) a freeway interchange, and (3) a highway corridor. The objective of these examples is to provide the analyst with a firsthand account of the various steps involved in collecting data to analyze/simulate traffic operations across such facilities.

Table 16. Traffic count data requirements for analysis types–work zone traffic, toll plazas, gated lanes, managed lanes, ramp metering, safety analyses.

TRAFFIC DATA PARAMETERS	ANALYSIS TYPES			
	Work zone traffic (freeway or arterial)	Toll plazas and gates	Managed lanes or ramp metering	Safety analyses
Peak hour turning movement counts	✓	-	-	-
Automated traffic recorder counts	✓	-	✓	-
Annual average daily traffic (AADT)	-	-	-	✓
Parking maneuvers	✓	-	-	-
Transit service data	✓	-	-	-
Vehicle classification data	✓	✓	✓	-
Speed data	✓	✓	✓	-
Toll plaza and gate lane data	✓	✓	-	-

Source: Adapted from Virginia DOT (2015).
- = data not required.

Table 17. Calibration data requirements for analysis types–work zone traffic, toll plazas, gated lanes, managed lanes, ramp metering, safety analyses.

	ANALYSIS TYPES

CALIBRATION DATA PARAMETERS	Work zone traffic (freeway or arterial)	Toll plazas and gates	Managed lanes or ramp metering	Safety analyses
Peak hour/period traffic demand	✓	✓	✓	-
Pedestrian and bicycle travel speeds	-	-	-	-
Mainline speed data	✓	✓	-	-
Ramp speed data	✓	-	-	-
Toll lane and gate processing time by payment choice	✓	✓	-	-
Travel times	✓	✓	✓	-
Queuing data	✓	✓	-	-
Existing crash data	-	-	-	✓

Source: Adapted from Virginia Department of Transportation (2015).
\- = data not required.

Sample Freeway Corridor Data Inventory

For the analysis of a freeway corridor, in many locations, high-resolution sensor data are available. The first recommended step is to create a functional diagram that clearly shows the locations of geometric features (on- and off-ramps, lane drops, etc., and sensor locations). See

Source: FHWA.

Figure 23. Diagram. Oregon Route 217 functional diagram.

(Oregon Route 217 [OR 217] in Portland, Oregon, as a case study).

For this corridor, inductive loop and radar detector data are available at 20-second resolution (minimum) as well as 5-minute, 15-minute, and 1-hour aggregations. The data are lane by lane count, occupancy, and speed. Dual loop detectors are used so speed is measured (not estimated), and as of 2016, vehicle lengths are also available. When averaged across lanes, the speed is weighted by volume. A sample output of 5-minute data for 1 detector is shown in Source: FHWA.

Figure 24.

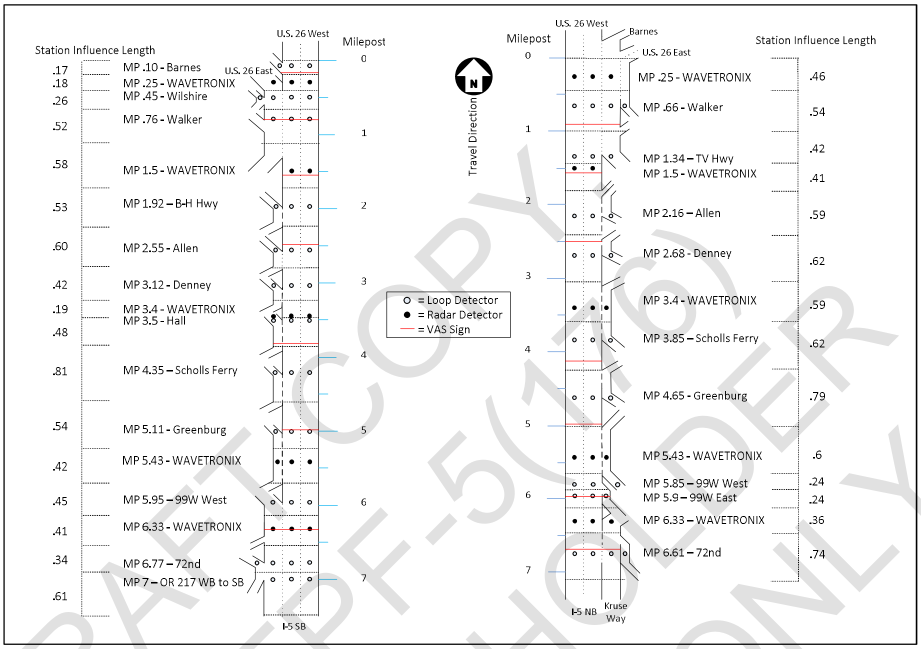
For one station, the data can be plotted and analyzed in a nearly unlimited number of ways. As an example,

Source: FHWA.

Figure **25** shows time series plots of speed and volume over 1 weekday using 1 month of detector data for midweek days (Tuesday–Thursday). In this location, it's clear that there are peaks in the morning and afternoon. Extrapolating point measurements such as county segment is possible and most often uses the midpoint method. As shown in

Source: FHWA.

Figure **26**, it is possible to assume that a speed value measured at one detector represents speed conditions half the distance to the next downstream detector and half the distance to the next upstream detector for a specific resolution interval.



Source: FHWA.

Figure 23. Diagram. Oregon Route 217 functional diagram.

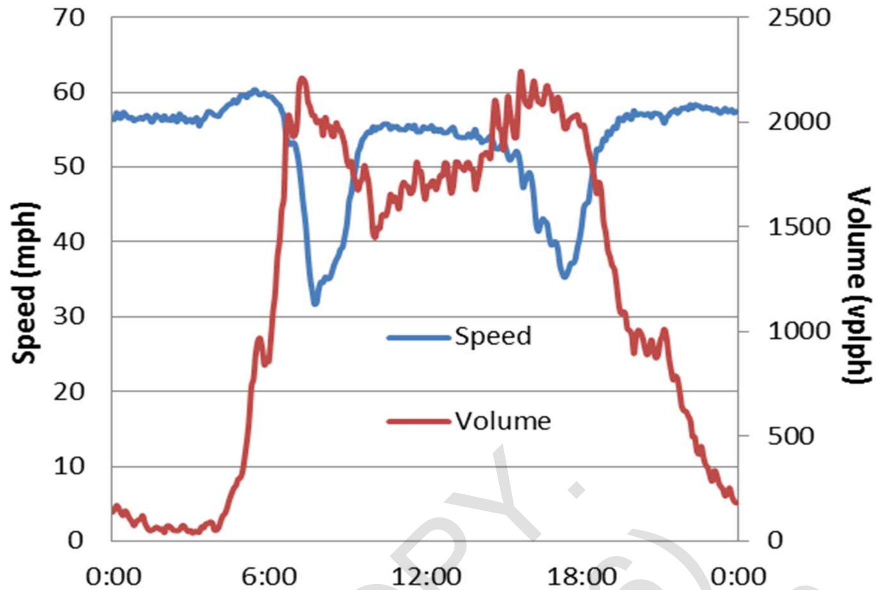
starttime	speed	volume
10/18/17 15:00	66.97	1260
10/18/17 15:05	64.56	900
10/18/17 15:10	64.21	672
10/18/17 15:15	65.05	876
10/18/17 15:20	66.61	1344
10/18/17 15:25	64.93	972
10/18/17 15:30	64.06	1272
10/18/17 15:35	64.63	936
10/18/17 15:40	61.11	1584
10/18/17 15:45	61.34	1164
10/18/17 15:50	59.39	1296
10/18/17 15:55	54.47	1056
10/18/17 16:00	61.15	1236
10/18/17 16:05	59.31	744
10/18/17 16:10	54.99	1632
10/18/17 16:15	22.1	588
10/18/17 16:20	26.69	1500
10/18/17 16:25	21.76	588
10/18/17 16:30	44.62	1200
10/18/17 16:35	53.85	1356
10/18/17 16:40	50.71	1320
10/18/17 16:45	33.71	924
10/18/17 16:50	28.56	1356
10/18/17 16:55	27.97	900

Source: FHWA.

Figure 24. Illustration. Sample 5-minute data.

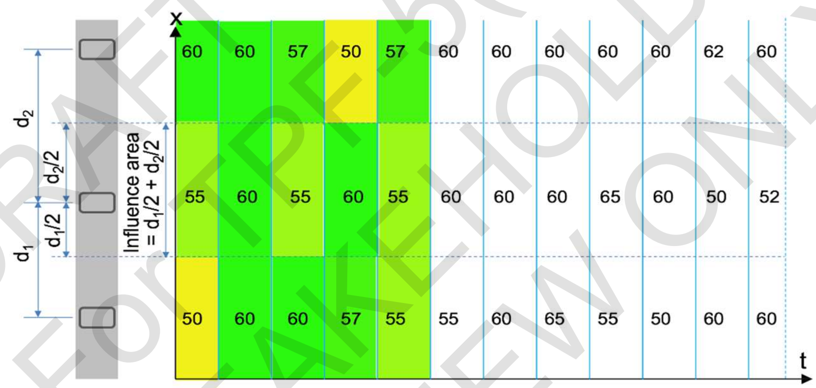
Source: FHWA.

Figure 27 shows a plot of measured speed along the OR 217 freeway corridor having applied the midpoint method described above. From this perspective, it is possible to see the peak periods and the bottleneck location—very useful inputs for a simulation activity.



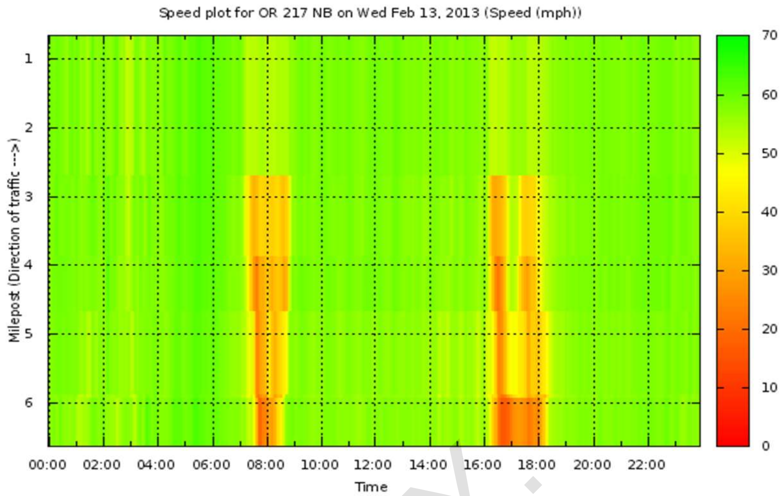
Source: FHWA.

Figure 25. Chart. Daily speed count data.



Source: FHWA.

Figure 26. Chart. Midpoint method illustration.



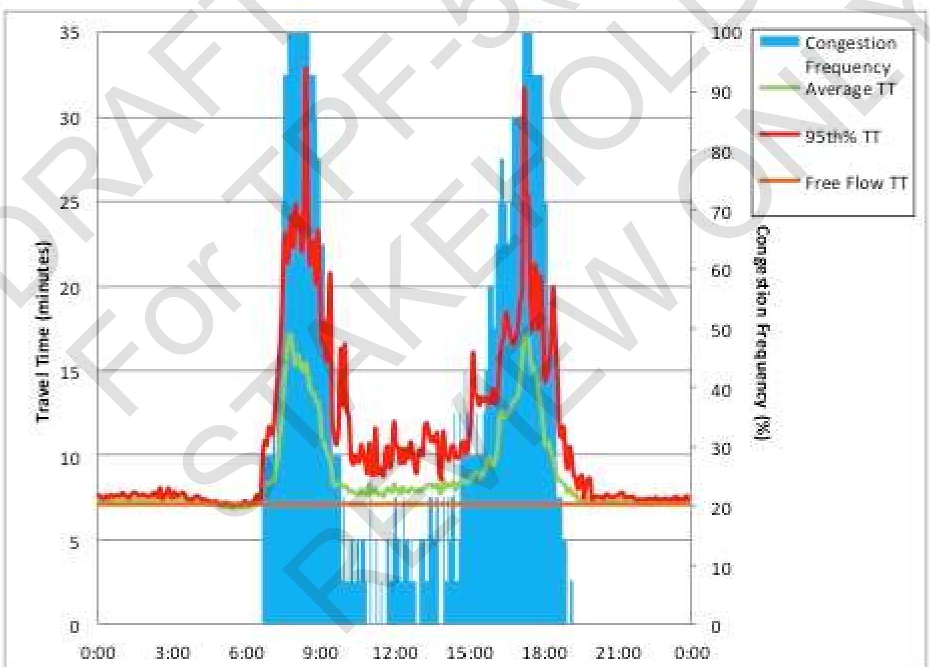
Source: FHWA.

Figure 27. Chart. Speed plotted over time and space.

Speeds can also be converted to travel time by taking the segment length and dividing by speed. A corridor travel time can be constructed by summing successive segment travel times.

Source: FHWA.

Figure 28 shows how a collection of corridor travel times (in this case using 5-minute travel times from midweek days over 1 month) can be plotted. This figure shows the mean travel time as well as the 95th percentile travel time from this specific distribution. Also shown is the estimated free-flow travel time and a bar chart showing the percent of observations during which congestion was observed for each time interval.

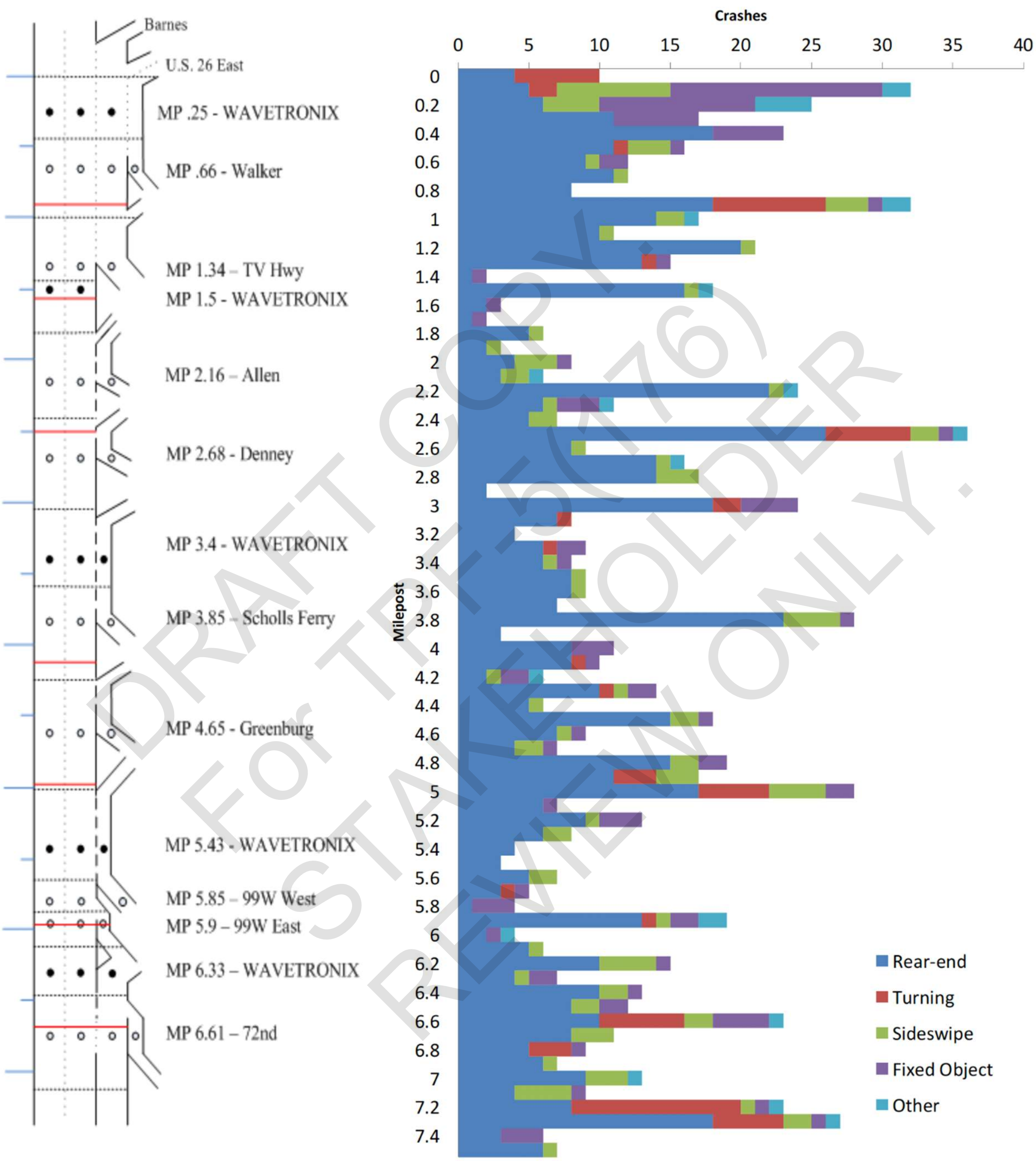


Source: FHWA.

Figure 28. Chart. Oregon Route 217 corridor travel time reliability.

Source: FHWA.

Figure 29 illustrates one method of organizing some of the crash data typically available from State DOTs for freeway corridors. This figure shows bar charts at each 0.1 MP indicating different crash types: rear end, turning, sideswipe, fixed object and other. The source of these data is the Oregon Department of Transportation (ODOT) Reported Crashes Database.

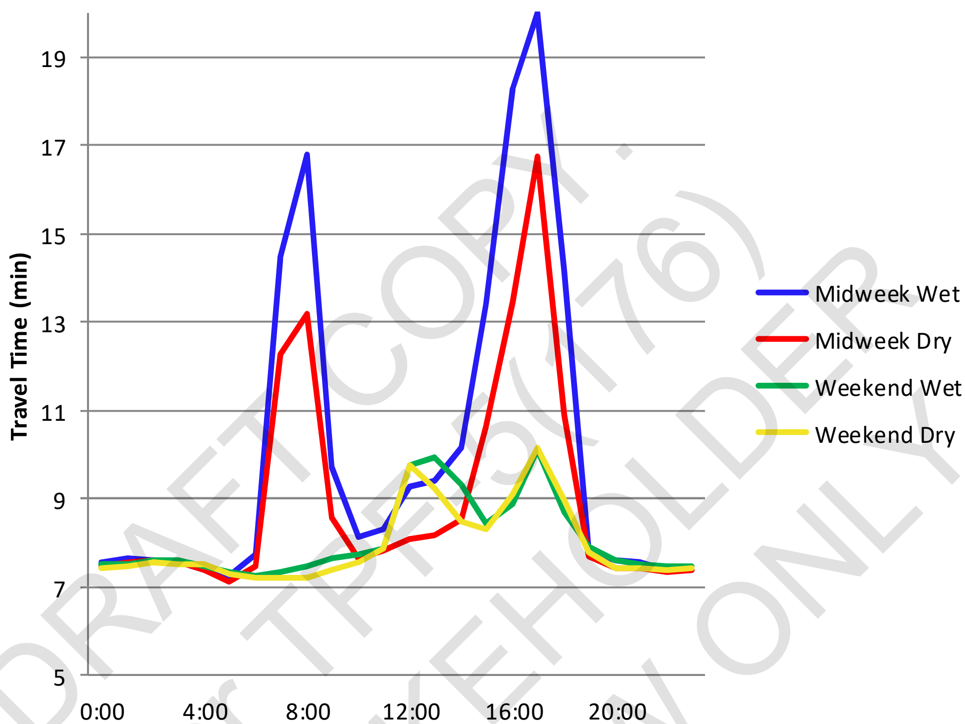


Source: FHWA.

Figure 29. Illustration. Freeway corridor crash analysis.

Weather and rainfall data may also be available for preparation of freeway data for simulation. As an example, corridor travel times along OR 217 are shown in Source: FHWA.

Figure 30**Error! Reference source not found.** for four scenarios (using a 3-month period): midweek days with precipitation (wet), midweek days without precipitation (dry) and weekend days with and without precipitation. As shown in the figure, travel times on midweek days were noticeably longer on wet days.



Source: FHWA.

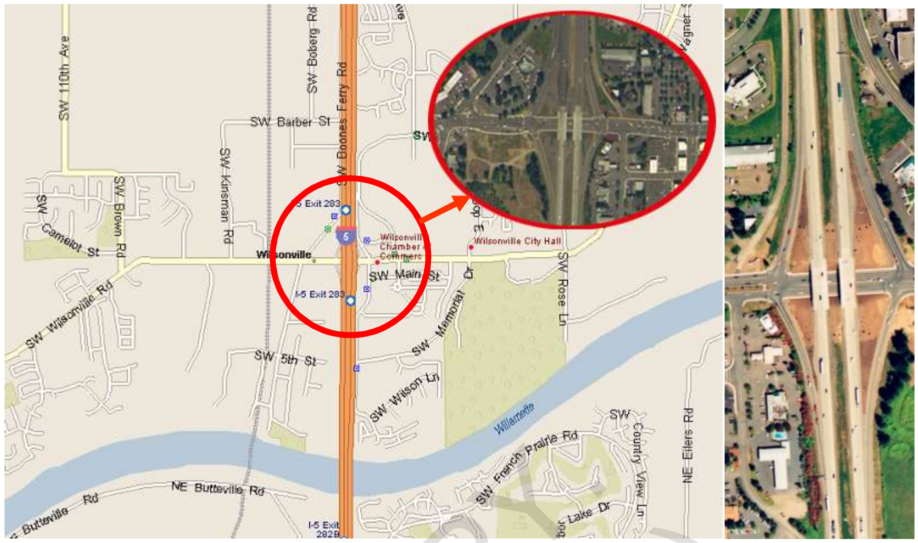
Figure 30. Chart. Weather impacts on travel times.

Sample Freeway Interchange Data Collection

In order to prepare simulating traffic operations at a simple freeway diamond interchange, the first step is to obtain basic aerial mapping of the site. In this example, the I–5/Wilsonville Road interchange was studied in Wilsonville, Oregon. The east-west arterial includes four consecutive signalized intersections (including freeway ramps) and two non-signalized intersections (see Source: FHWA.

Figure 31Source: FHWA.

Figure).



Source: FHWA.

Figure 31. Map. Interchange map and aerial views.

At the time of the analysis, there were concerns about heavy traffic volumes entering the City of Wilsonville and a nearby retail shopping area. Data collection included (A. Subfigure of turning movements.

Source: FHWA.

B. Subfigure of site reconnaissance photos.

Figure 32Source: FHWA.

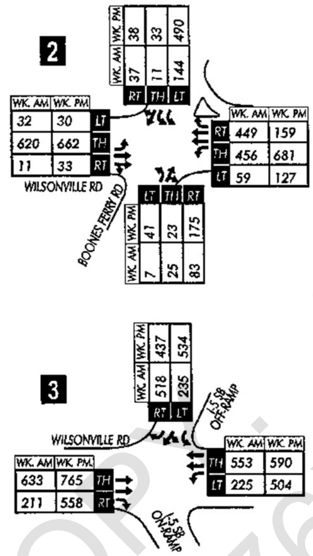
A. Subfigure of turning movements.

Source: FHWA.

B. Subfigure of site reconnaissance photos.

Figure):

- Aerial photographs from ODOT.
- Computer-aided design (CAD) drawings from the City of Wilsonville.
- Turning movement counts from DKS Associates.
- Field data: Evening (PM) peak period signal cycle times (10 data sets) and sight distances, lane widths, island sizes, and pavement markings.



Source: FHWA.

A. Subfigure of turning movements.

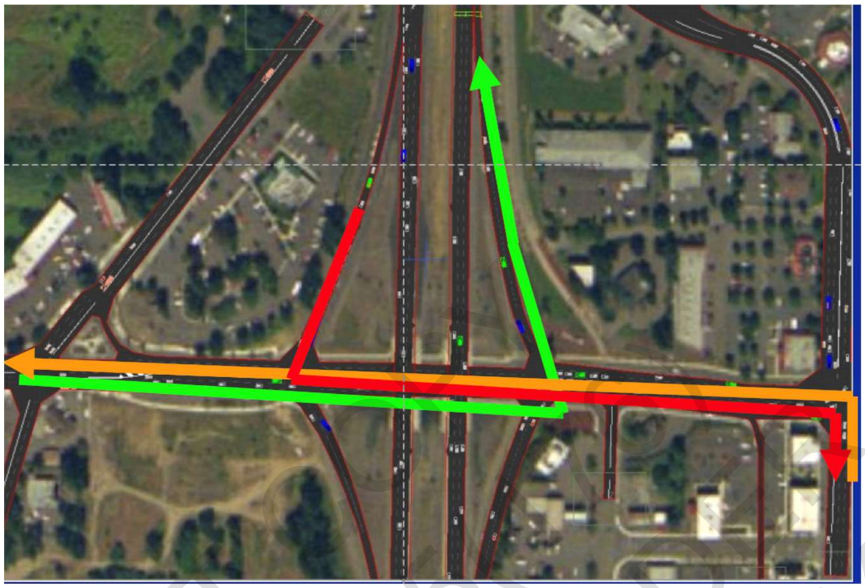
Source: FHWA.

B. Subfigure of site reconnaissance photos.

Figure 32. Illustrations. Turning movements and site reconnaissance.

Travel time runs were also conducted for three routes shown in Source: FHWA.

Figure 33:



Source: FHWA.

Figure 33. Map. Travel time runs.

Sample Highway Corridor Data Inventory

Outside of urban areas, data sources may be more dispersed spatially and may have lower temporal resolution. As one example, U.S. Route 26 (U.S. 26) and Oregon 35 (OR 35) are the two highways that provide access to recreational areas in Mt. Hood National Forest east of Portland, Oregon. Both highways are primarily two lanes and have posted speed limits of 55 mph (see © Oregon Department of Transportation.

Figure 34**Error! Reference source not found.**).

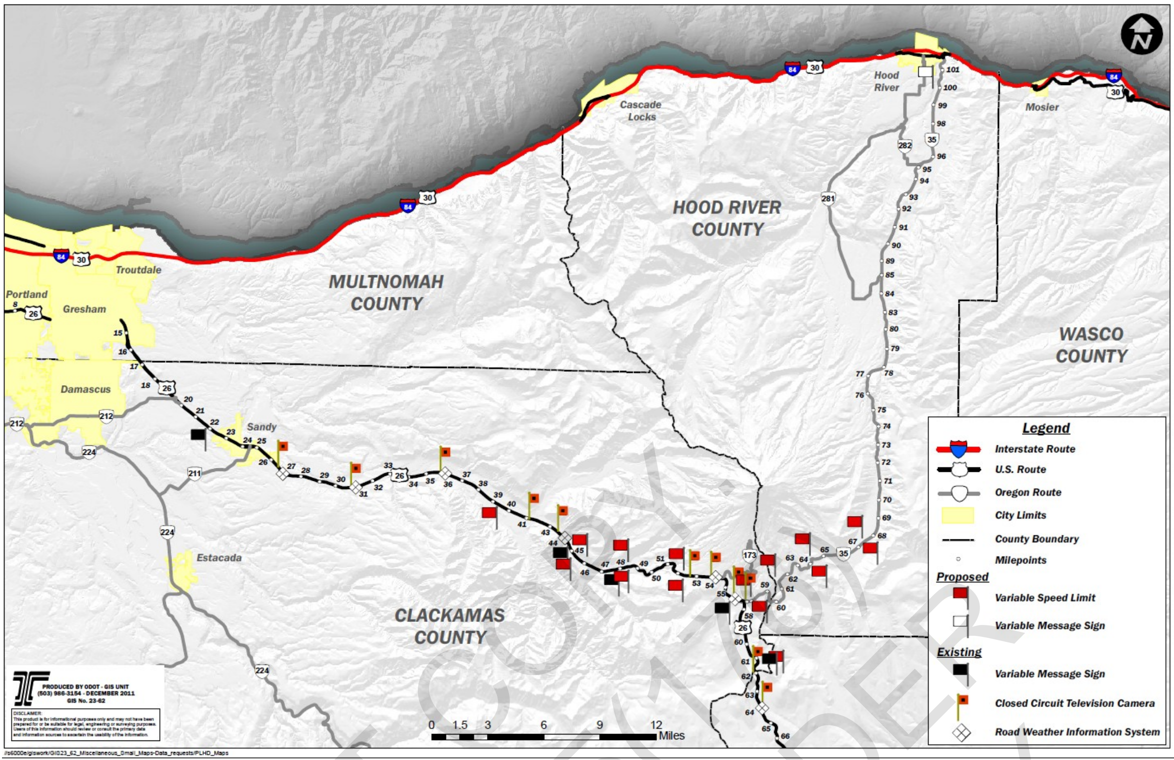


Figure 34. Map. U.S. Route 26 and Oregon Route 35.

Despite its more rural setting, the Mt. Hood highway has CCTV cameras and permanent Automatic Traffic Recorder (ATR) stations throughout the corridor that provide count and vehicle classification data for planning purposes. © Oregon Department of Transportation.

Figure 35 is a map of all ATR stations throughout the state.

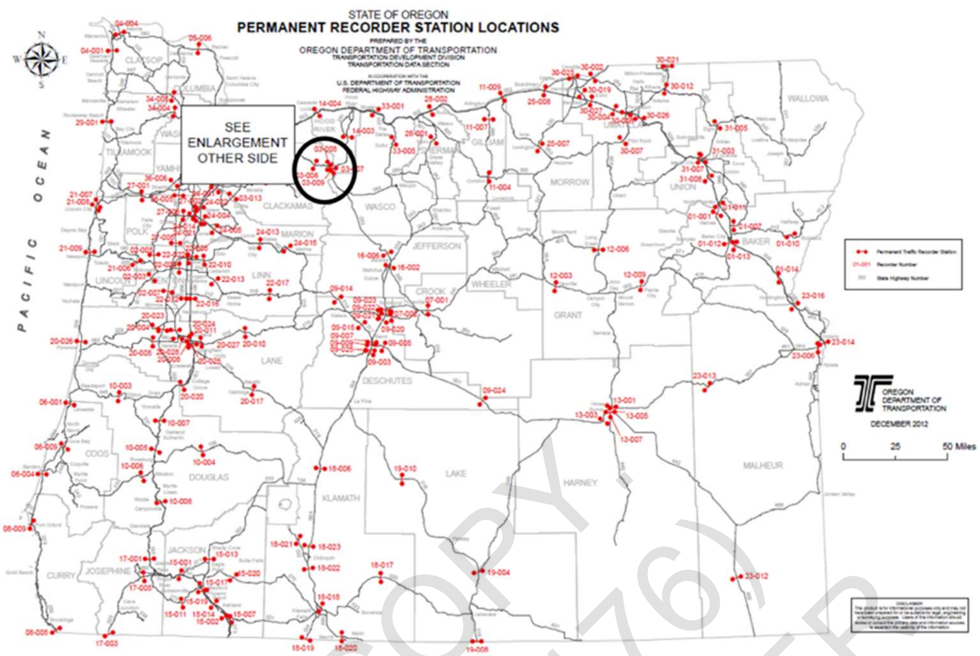


Figure 35. Map. Oregon permanent automatic traffic recorder stations.

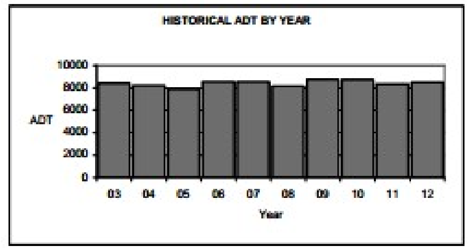
Figure 36 is a sample of the data that can be obtained from the ATR system. The figure summarizes average daily traffic (ADT) estimates by year and month, as well as a vehicle classification breakdown, for one ATR station near Rhododendron, Oregon.

Location: US26; MP 46.38; MT. HOOD HIGHWAY NO. 26; 0.30 mile east of Camp Creek Rd (USFS 28)

Site Name: Rhododendron (03-006)
Installed: August, 1995

HISTORICAL TRAFFIC DATA

		Percent of ADT				
Year	ADT	Max Day	Max Hour	10TH Hour	20TH Hour	30TH Hour
2003	8416	222	23.8	21.3	19.5	18.8
2004	8211	205	22.4	19.5	18.8	18.2
2005	7906	216	23.0	20.3	19.6	18.9
2006	8535	231	22.4	19.1	18.5	18.0
2007	8542	198	21.2	18.9	18.1	17.7
2008	8162	233	22.9	20.4	19.2	18.8
2009	8736	197	22.3	19.6	18.4	17.8
2010	8714	207	21.6	19.8	18.9	18.5
2011	8330	214	24.7	20.0	18.6	18.1
2012	8480	227	24.0	21.0	20.2	19.4



2012 TRAFFIC DATA

	Average Weekday Traffic	Percent of ADT	Average Daily Traffic	Percent of ADT
January	6280	74	9101	107
February	5935	70	8640	102
March	6043	71	8166	96
April	4963	59	7106	84
May	5905	70	7473	88
June	6977	82	8337	98
July	9021	106	11229	132
August	9139	108	11656	137
September	7152	84	8708	103
October	5198	61	6190	73
November	4781	56	5676	67
December	7726	91	9477	112

Classification Breakdown	Percent of ADT
Motorcyles	1.16
Passenger cars	70.48
Light Trucks	15.73
Buses	0.59
Single unit trucks (2 axles)	1.85
Single unit trucks (3 axles)	0.42
Single unit trucks (4 or more axles)	0.01
Single trailer trucks (4 or less axles)	1.87
Single trailer trucks (5 axles)	3.95
Single trailer trucks (6 or more axles)	1.50
Multi trailer trucks (5 or less axles)	0.22
Multi trailer trucks (6 axles)	0.19
Multi trailer trucks (7 or more axles)	2.03

Figure 36. Screenshot. Automatic traffic recorder system output.

In rural areas, probe vehicle data from NPMRDS, now provided by INRIX (formerly provided by Nokia/HERE), can be a resource for obtaining speed data. This can complement volume data that may be available from ATRs or other counting programs. For NPMRDS and other probe data, corridors are broken into unique traffic message channel (TMC) codes (see sample from another highway in © Oregon Department of Transportation.

Figure 37), and travel speeds and times are estimated for every minute or every 5 minutes, depending on the level of aggregation. U.S. 26 is broken into 12 eastbound and 12 westbound TMCs across 51 miles between Oregon Route 211 (OR 211) and Oregon Route 216 (OR 216). Oregon 35 is broken into 8 TMCs northbound and 7 TMCs southbound between I–84 and U.S. 26.

tmc	road	direction	intersection	state	county	zip	start_latitude	start_longitude	end_latitude	end_longitude	miles	road_order
114+04411	OR-217	NORTHBOUND	72ND AVE/EXIT 7	OR	WASHINGTON	97223	45.421573	-122.745525	45.423394	-122.7476364	0.163615	1
114P04411	OR-217	NORTHBOUND	72ND AVE/EXIT 7	OR	WASHINGTON	97223	45.423394	-122.7476364	45.428313	-122.754869	0.493205	2
114+04412	OR-217	NORTHBOUND	OR-99W/PACIFIC HWY/EXIT 6	OR	WASHINGTON	97223	45.428313	-122.754869	45.430527	-122.757172	0.189651	3
114P04412	OR-217	NORTHBOUND	OR-99W/PACIFIC HWY/EXIT 6	OR	WASHINGTON	97223	45.430527	-122.757172	45.437806	-122.765532	0.649114	4
114+04413	OR-217	NORTHBOUND	GREENBURG RD/EXIT 5	OR	WASHINGTON	97223	45.437806	-122.765532	45.441681	-122.77306	0.453808	5
114P04413	OR-217	NORTHBOUND	GREENBURG RD/EXIT 5	OR	WASHINGTON	97223	45.441681	-122.77306	45.445374	-122.780801	0.454741	6
114+04414	OR-217	NORTHBOUND	OR-210/SCHOLLS FERRY RD/EXIT 4	OR	WASHINGTON	97223	45.445374	-122.780801	45.447658	-122.783118	0.198848	7
114P04414	OR-217	NORTHBOUND	OR-210/SCHOLLS FERRY RD/EXIT 4	OR	WASHINGTON	97223	45.447658	-122.783118	45.458191	-122.784649	0.736732	8
114+04415	OR-217	NORTHBOUND	HALL BLVD/EXIT 4A	OR	WASHINGTON	97223	45.456458	-122.78449	45.457541	-122.784568	0.074941	9
114P04415	OR-217	NORTHBOUND	HALL BLVD/EXIT 4A	OR	WASHINGTON	97223	45.457541	-122.784568	45.463056	-122.786003	0.387816	10

Figure 37. Screenshot. Sample traffic message channel codes.

Figure 38 is a sample of the data obtainable from INRIX's Massive Raw Data Downloader application. For each TMC and minute, a speed, historic average speed, reference speed, and travel time are given. The confidence score is a number indicating the source of the estimated speeds and travel times. A value of 30 indicates the data are entirely from real-time sources, a 10 indicates the data are entirely from historical sources, and a 20 indicates the data are from a mix of real-time and historical sources. The C value is an estimate, on a scale from 0-100, of how reliable that row of data are.

tmc_code	measurement_tstamp	speed	average_s	reference	travel_tim	confidenc	cvalue
114-05968	6/16/2013 22:09	55	54	54	9.29	30	100
114+05275	6/16/2013 22:09	56	56	56	9.12	10	0
114-05275	6/16/2013 22:09	56	51	51	2.33	30	100
114+05969	6/16/2013 22:09	54	54	54	2.42	10	0
114P05275	6/16/2013 22:09	53	53	53	0.6	10	0
114N0527	6/16/2013 22:09	52	52	52	0.61	30	100
114-05970	6/16/2013 22:09	55	55	55	12.4	10	0
114+05273	6/16/2013 22:09	56	56	56	12.12	10	0
114-05274	6/16/2013 22:09	52	52	52	6.48	10	0
114+05970	6/16/2013 22:09	53	53	53	6.36	10	0
114-05969	6/16/2013 22:09	44	47	47	15.7	30	100
114+05274	6/16/2013 22:09	52	52	52	13.29	10	0
114-05273	6/16/2013 22:09	28	28	28	1.1	30	100
114N0527	6/16/2013 22:09	30	30	30	0.04	10	0
114P05273	6/16/2013 22:09	30	30	30	0.1	10	0
114+05277	6/16/2013 22:10	59	59	59	12.58	10	0
114-05967	6/16/2013 22:10	64	60	60	1.7	30	69
114+05968	6/16/2013 22:10	59	59	59	1.85	10	0

Figure 38. Screenshot. INRIX data sample.

Figure 39 shows average hourly speeds by season, using minute-level speed data from 1 year for both westbound and eastbound.

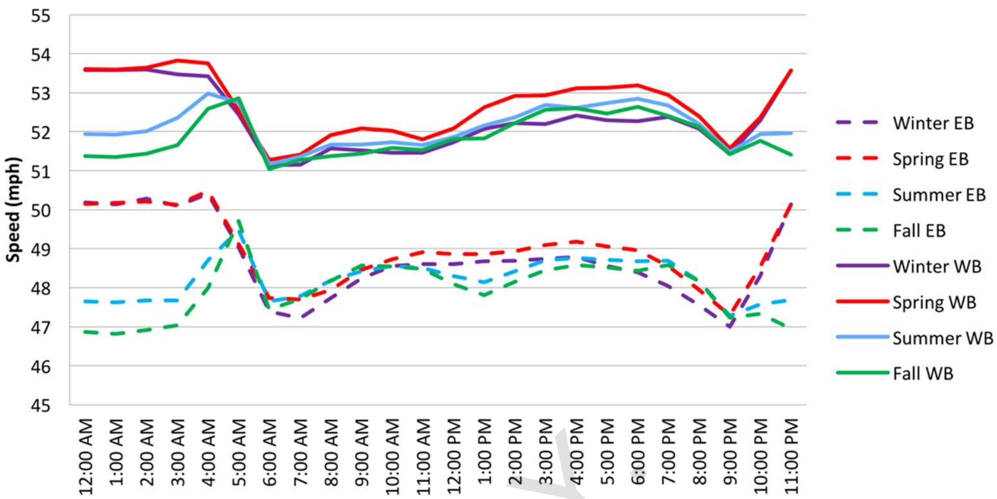
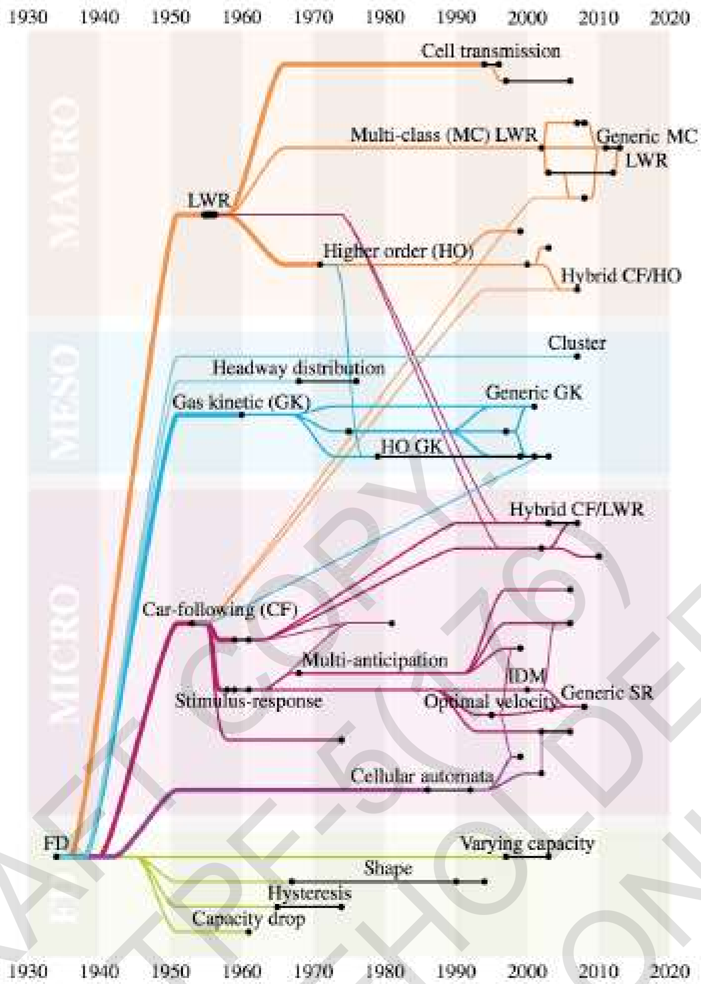


Figure 39. Chart. Sample hourly INRIX speeds by season.

As indicated, the speeds do vary. Note that the early-morning and late-night hour averages are largely compromised of purely historical data (confidence scores of 10), so they are unlikely to be very reliable. In all four seasons, U.S. 26 eastbound (uphill) is clearly much slower than U.S. 26 westbound (downhill). In both directions, average speeds tend to be highest in spring, with little difference among average speeds for the other three seasons. This indicates that speeds don't drop as much during winter as might be expected.

6.14 DATA REQUIREMENTS BY MODEL TYPE

The modeling of traffic flow can be performed along a spectrum of three scales—microscopic, mesoscopic, and macroscopic—from the most detailed to the least detailed, in that order. The following sections look at data requirements for these model categories in further detail.



© van Wageningen-Kessels et al. (2015).

Figure 40. Diagram. Genealogy of traffic flow models.

Macroscopic Models

Macroscopic models describe entities and their activities and interactions at a low level of detail. For example, the traffic stream may be represented in some aggregate manner, such as a statistical histogram or by scalar values of flow rate, density, and speed. Lane change maneuvers would probably not be represented at all; Ben-Akiva (1996) suggests that the model may assert that the traffic stream is properly allocated to lanes or employ an approximation to this end. Dynamic macroscopic models, such as the LWR model by Lighthill and Whitham (1955) and Richards (1956), describe the evolution of traffic over time and space using a set of differential equations. The solution to these equations can be obtained analytically or using simulation. Simulation is normally used when temporal and spatial interactions of traffic flows in road networks need to be evaluated. In Daganzo's (1994, 1995) cell transmission model, the LWR model is discretized into cells. The road is represented by many small sections (cells). The simulation model keeps track of the number of vehicles in each cell, and at every time step calculates the number of vehicles that cross the boundaries between adjacent cells. This flow from one cell to the other depends on how many vehicles can be sent by the upstream cell and how many can be received by the downstream cell. The number of vehicles that can be sent is a function of the density in the upstream cell and the number that can be received depends on the density in the receiving cell. Daganzo's (1999) lagged cell transmission model is a refinement of this scheme, where the number of vehicles that a cell can receive (from the adjacent upstream cell) depends on the density some time earlier in the cell.

Mesoscopic Models

Mesoscopic models generally represent most entities at a high level of detail, but describes their activities and interactions at a much lower level of detail than a microscopic model. For example, the lane-change maneuver could be represented for individual vehicles as an instantaneous event with the decision based on, say, relative densities rather than detailed vehicle interactions as in the case of the microscopic model developed by Lieberman and Rathi (1997). Per Burghout (2004), these models can take varying forms.

One mesoscopic paradigm is that of individual vehicles grouped into cells which control their behavior. The cells traverse the link and vehicles can enter and leave cells when needed, but not overtake cells ahead of them. The speed of the vehicles is determined by the cell, not the individual driver's decisions (Ben-Akiva 1996). Alternatively, another approach is used in some models (Jayakrishnan et al. 1994; Gawron 1998; Mahut 2001), where the roadway is modeled in two parts: a queuing and a running part. The lanes can be modeled individually. Although the vehicles are represented individually and maintain their individual speeds, their behavior is not modeled in detail. The vehicles traverse the running part of the roadway with a speed that is determined using a macroscopic speed-density function, and at the downstream end a queue-server transfers the vehicles to connecting roads. The main application area of mesoscopic models is where the detail of microscopic models might be desirable; but is infeasible due to a large network or limited resources available to be spent on the coding and debugging of the network.

Microscopic Models

Microscopic models, describe traffic at the level of individual vehicles and their interaction with each other and the roadway infrastructure, as suggested by Burghout (2004). Normally this behavior is governed by a set of rules which determine when a vehicle accelerates or decelerates, when it changes lanes, and how and when a vehicle chooses and changes its route to its destination. Therefore, models that govern the vehicle's behavior can often be divided into a car-following model, a lane-change model, and a route-choice model. The car-following model describes the breaking and acceleration patterns that result from the interaction of the driver with the vehicle in front as well as other factors (such as speed limits and roadway geometry).
The lane-changing model describes the decisions pertaining to change of lanes, based on the driver's preferences and the situation in both the current lane and other lanes (speed of the vehicle in front, presence of sufficiently large gap in adjacent lane, etc.), as described by Burghout (2004). The route-choice model describes how drivers determine which path to take from their starting location (origin) to their destination, and how they react to traffic and route information along the way. Relative to macroscopic and mesoscopic models, microscopic

models are costly to develop, execute, and maintain, as indicated by Lieberman and Rathi (1997).

6.15 SOME CHALLENGES WITH DATA

Collecting all necessary data, verifying their quality, and documenting any assumptions are key to justifying the results to stakeholders and the public. Some challenges analysts face with data are discussed in the subsequent sections.

Data Comprehensiveness and Reliability

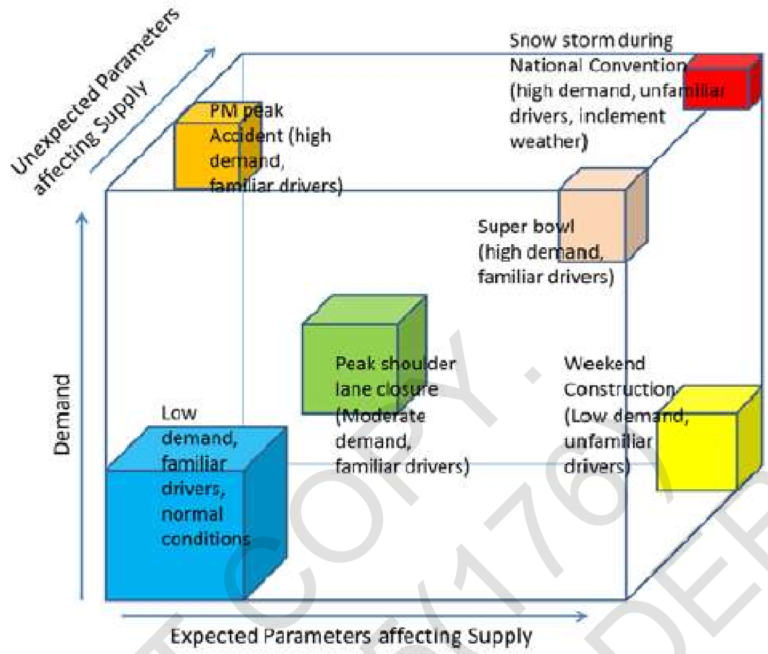
Traffic counts should be taken at key locations in the study area, including freeway segments, major intersections, interchanges, and major on- and off-ramps. Ideally, data collection should be conducted simultaneously at all important observation points. If this is impractical, it should be done during similar times of day (and for the same length of time) at all locations, under similar weather conditions and traffic demands. Automated data sources are used for collecting long-term data needed for developing and calibrating most transportation models. However, many automated data sources lack robustness or reliability to effectively aggregate different data sets. Therefore, a thorough review of data quality is recommended at a very early stage to address any deficiencies (Wunderlich et al. 2017; VDOT 2015; FDOT 2014).

Data Accuracy

Accuracy is the measure of the degree of agreement between data values and a source assumed to be correct. Wunderlich et al. (2017) and FDOT (2014) suggest it is also a measure of the error. It is important to have accurate, internally consistent, and recent data. If information on forecasted traffic conditions is not available, it could be useful to study data from other regions that may have similar bottlenecks and traffic patterns to those envisioned for the study area. This is important as the models are not only expected to represent observed conditions but also to present a clear picture of the future based on assumed future characteristics.

Data Fusion and Aggregation

A data analyst must ensure that each data set is of the same time frame before combining data sets. Wunderlich et al. (2017) suggests that sometimes, due to delays in transferring data from roadway sensors to the database or because of the collection frequency, an analyst may have to adjust the time frame so that the combined data set makes sense. Like the temporal issues, an analyst may also need to align data sets spatially. Per Traffic Analysis Toolbox Volume XIII: Integrated Corridor Management (ICM) Analysis, Modeling, and Simulation Guide (2012), it must be ensured that data from multiple sources also be for concurrent periods to neutralize seasonal and other travel pattern variances that can affect data (Wunderlich 2002). Cross validation is a way to ensure the proper temporal and spatial integration of different data sets. (Wunderlich et al. 2017). When integrating data sets, a data analyst must ensure data sets are consistent with each other. For example, volume and speed along a roadway should match ($Q=D\times V$) at any given time; if volume and speed data are affected by an incident, the effects on both should occur at the same time and location. Another step is checking to ensure traffic movements make sense; for example, left-turn vehicles will almost always come from the most left two lanes.



Source: FHWA.

Figure 41. Diagram. Traffic conditions that require data for calibration.

When data comes from multiple sources, issues of storage, licensing, and ownership are also critical. Having a common metadata framework across all systems and using common controlled vocabularies are keys to ensure consistency and reliability of metadata applied to the information and data assets. For example, the USDOT Data Capture Management (DCM) Program (USDOT 2018) developed a Research Data Exchange (RDE) platform to share archived and real-time data from multiple sources and modes to better support the needs of Intelligent Transportation Systems (ITS) researchers and developers while reducing costs and encouraging innovation. The USDOT published metadata guidelines for the RDE to be adopted by public- and private-sector data providers to increase usability of their data. Creation of metadata should be included in plans for the procurement of any data collection effort; otherwise, there is a risk that data will be misinterpreted or abandoned as too arcane to support future analyses. A history of detector numbering should be included in the metadata so that the analyst can link data sets from different years.

6.16 DATA QUALITY CONTROL

There are many considerations relating to data quality control applicable to the simulation environment. In general, it is useful to consider six dimensions to the data quality measurement challenge:

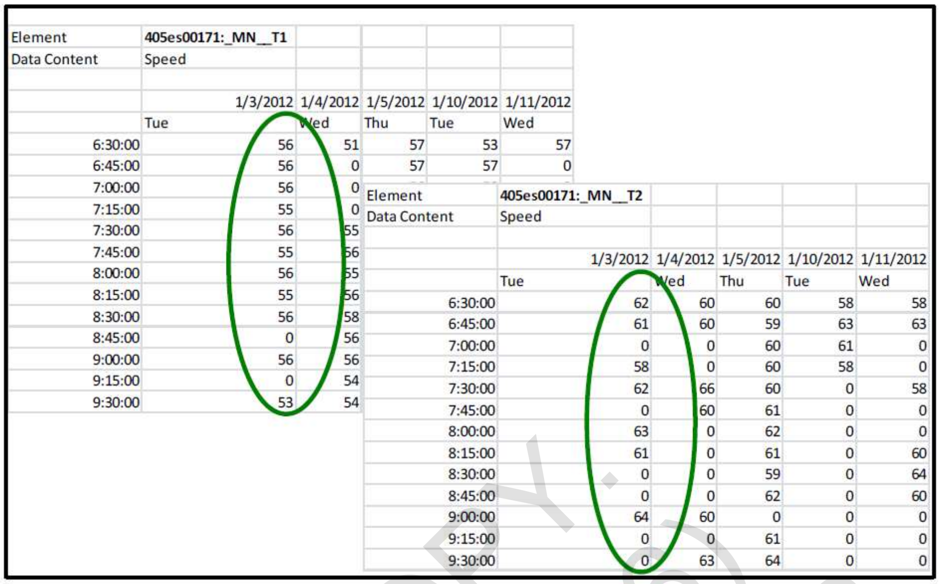
- **Accuracy** – How closely does the collected data match actual conditions?
- **Confidence** – Are the data trustworthy?

- **Delay** – How quickly are the collected data available for use in advanced traveler information systems (ATIS) applications?
- **Availability** – How much of the data designed to be collected are made available?
- **Breadth of coverage** – Over what roadways, or portions of roadways, are data being collected?
- **Depth of coverage (density):** How close together or far apart are the traffic sensors?

Wunderlich, Alexiadis, and Wang, in a study (2017) on Scoping and Conducting Data-Driven 21st Century Transportation System Analyses, state that "When conducting quality control of data or integrated data, an analyst must make sure to avoid open-ended quality control procedures and focus on those errors that are most likely to impact the model." Due to the presence of a multitude of factors, the analyst tries to control some of the factors since it isn't possible to control everything. In some cases, as Wunderlich et al. (2017) and VDOT (2015) suggest, the analyst needs to preserve outlier data to capture the variability in traffic conditions. Therefore, the analyst must have a clear understanding of why any outliers appeared and whether it is likely that similar values will continue to appear. It is important to preserve outliers not attributable to sensor and processing errors so that the full range of conditions can be visualized. Most quality control procedures of today are automated. However, some temporal and spatial consistencies problems in the data are impossible to detect through a basic quality control procedure. Wunderlich et al. (2017) suggest these inconsistencies (such as widely inconsistent input/output counts for adjacent traffic count sensors) can be problematic, even if the data have passed multiple elemental-level checks. If the data are used for model calibration, eliminating temporal and spatial inconsistencies must be an important focus of quality control.

> Wunderlich et al. (2017) give the following guidance on missing data when conducting quality control: When faced with missing data during quality control, the analyst must decide whether to discard the data point or impute missing data. Imputing data can be practical and realistic if the imputed data does not alter the overall trend observed with the data set. Sometimes, the imputation leads to illogical relationships (e.g., unequal directional count data at a tunnel entrance and exit). When this happens, the modeled system cannot be calibrated. Another error could occur if data is imputed to a data set; but isn't properly marked. If all the analysts aren't aware which data has been imputed, the overall data set might not make sense, possibly due to illogical or incongruent relationships. Smoothing data (e.g., averaging values of certain time intervals) to minimize the impact of missing data is a common way of imputation adopted by analysts. The analyst must avoid averaging data too much, or they will lose information about the variations in traffic patterns and their impacts. Source: FHWA.

Figure 42 demonstrates one instance where improperly addressed missing data could lead to inaccurate results. On January 3 (shown by the green circle), Lane 2 (right table) has more missing values than Lane 1 (left table). If the analyst chooses to average the speeds across both lanes without applying quality control, the calculated speed for the roadway segment will be inaccurate, causing issues for later calibration.



Element	405es00171:_MN__T1				
Data Content	Speed				
	1/3/2012	1/4/2012	1/5/2012	1/10/2012	1/11/2012
	Tue	Wed	Thu	Tue	Wed
6:30:00	56	51	57	53	57
6:45:00	56	0	57	57	0
7:00:00	56	0			
7:15:00	55	0			
7:30:00	56	55			
7:45:00	55	56			
8:00:00	56	55			
8:15:00	55	56			
8:30:00	56	58			
8:45:00	0	56			
9:00:00	56	56			
9:15:00	0	54			
9:30:00	53	54			

Element	405es00171:_MN__T2				
Data Content	Speed				
	1/3/2012	1/4/2012	1/5/2012	1/10/2012	1/11/2012
	Tue	Wed	Thu	Tue	Wed
6:30:00	62	60	60	58	58
6:45:00	61	60	59	63	63
7:00:00	0	0	60	61	0
7:15:00	58	0	60	58	0
7:30:00	62	66	60	0	58
7:45:00	0	60	61	0	0
8:00:00	63	0	62	0	0
8:15:00	61	0	61	0	60
8:30:00	0	0	59	0	64
8:45:00	0	0	62	0	60
9:00:00	64	60	0	0	0
9:15:00	0	0	61	0	0
9:30:00	0	63	64	0	0

Source: FHWA.

Figure 42. Screenshot. Speed data on two lanes.

This chapter concludes with a boxed-out section that briefly introduces some considerations that are helpful for analysts conducting reliability analysis.

Reliability Considerations

In the context of travel time, Xie et al. (2013) suggests that reliability can be defined as the probability that a certain trip (from a given origin to a given destination) can be made successfully within a specified interval of time. Small et al. (1999) introduced reliability as the uncertainty about arriving at one's destination at a predicted time. FDOT (2014) has defined reliability as the percent of travel that takes no longer than the expected travel time plus a certain acceptable additional time. Tu et al. (2008) says travel time reliability can capture the variability experienced by individual travelers. Small et al. (1999) argue it is an indicator of the operational consistency of a facility over an extended period. FDOT (2000) suggests reliability of travel time can also be used as a measure of quality of service. The standard deviation, median, coefficient of variation, buffer time, 95th percentile of travel time, buffer index, planning time index, misery index, skew, width, and congestion frequency are some of the other metrics used to measure travel time reliability.

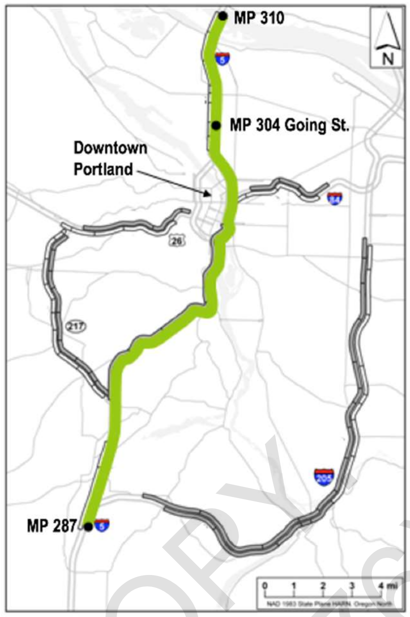
Existing Travel Time and Reliability Measures

Travel time related performance measures are usually constructed from measurements of volume (measured at a point), speed, and/or travel time for a specific corridor or O-D pair. For example, the travel time index is usually computed as the mean travel time divided by an assumed free-flow travel time to traverse a specific distance. In addition, measures that incorporate travel time reliability have been proposed and demonstrated under different circumstances. Commonly used reliability measures include the following:

- 90th or 95th percentile travel time.
- Standard deviation of travel time.
- Coefficient of variation: computed as standard deviation of travel time divided by mean travel time.
- Buffer index: computed as difference between 95th percentile travel time and mean travel time, divided by mean travel time.
- Planning time index: computed as 95th percentile travel time divided by assumed free-flow travel time.
- Misery index: computed as difference between mean travel time for worst 20 percent of trips and overall mean travel time, divided by mean travel time.
- Skew of travel time distribution: computed as the difference between 90th and 50th percentile travel time, divided by the difference between 50th and 10th percentile travel time.
- Width of travel time distribution: computed as the difference between 90th and 10th percentile travel time, divided by the 50th percentile travel time.
- Congestion frequency: percent of time that mean speed drops below a specific speed.
- Lateness and earliness indices: can be based on the log-normal distribution.

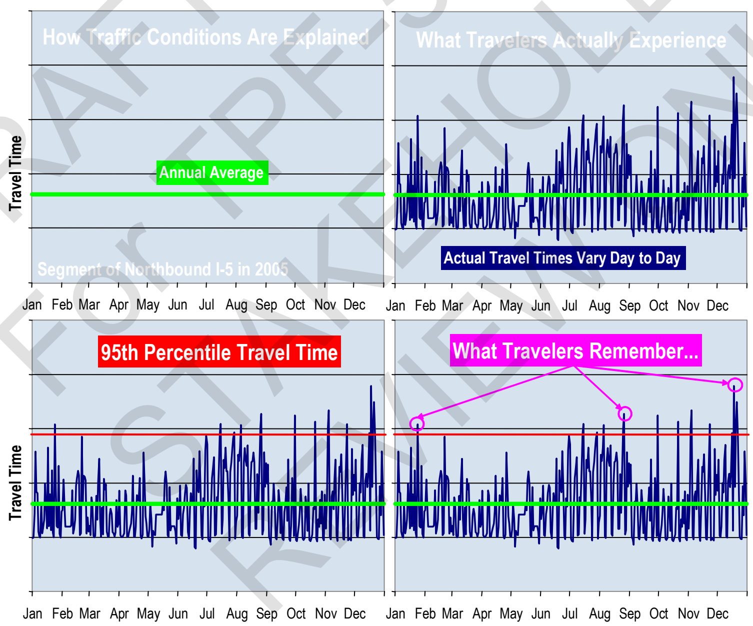
Segment Level Analysis

To illustrate some of these points, **Error! Reference source not found.** (Nam et al. 2005; Zhang et al. 2003) illustrates the reliability concept using real loop detector data from a 23-mile Portland, Oregon, freeway corridor (**Error! Reference source not found.**). Traffic conditions are usually explained as a point value, but users experience day-to-day variations. Even when the 95th percentile concept is understood, travelers remember the very worst experiences.



Source: PORTAL.

Figure 43. Map. Portland region.

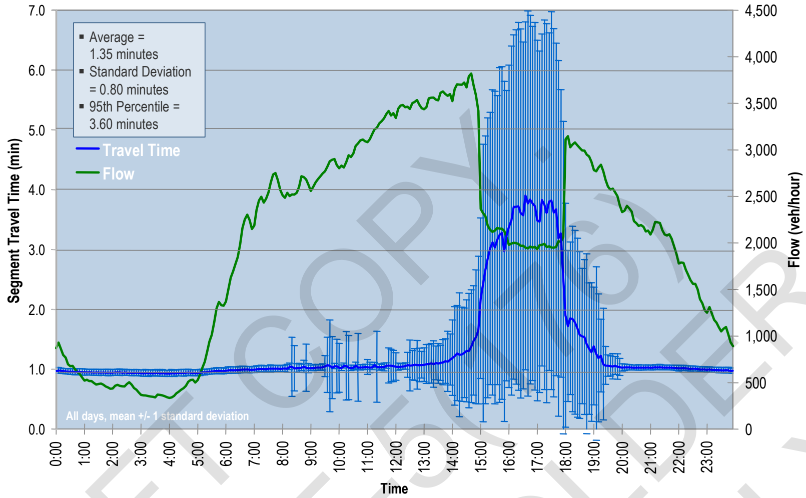


Source: FHWA.

Figure 44. Chart. Basic reliability introduction.

Loop detectors in this corridor are associated with segments of differing lengths. © Lyman and Bertini (2008).

Figure 45 shows 5-minute resolution travel time and flow data for a 0.75-mile segment for 1 year (>1.5 million data points). Travel time through the segment varied throughout the day (solid line), and its variability over the year increased during the PM peak (vertical error bars show ±1 standard deviation). The traffic flow also dropped during the most congested period.



© Lyman and Bertini (2008).

Figure 45. Chart. Going St. segment travel time for 1 year.

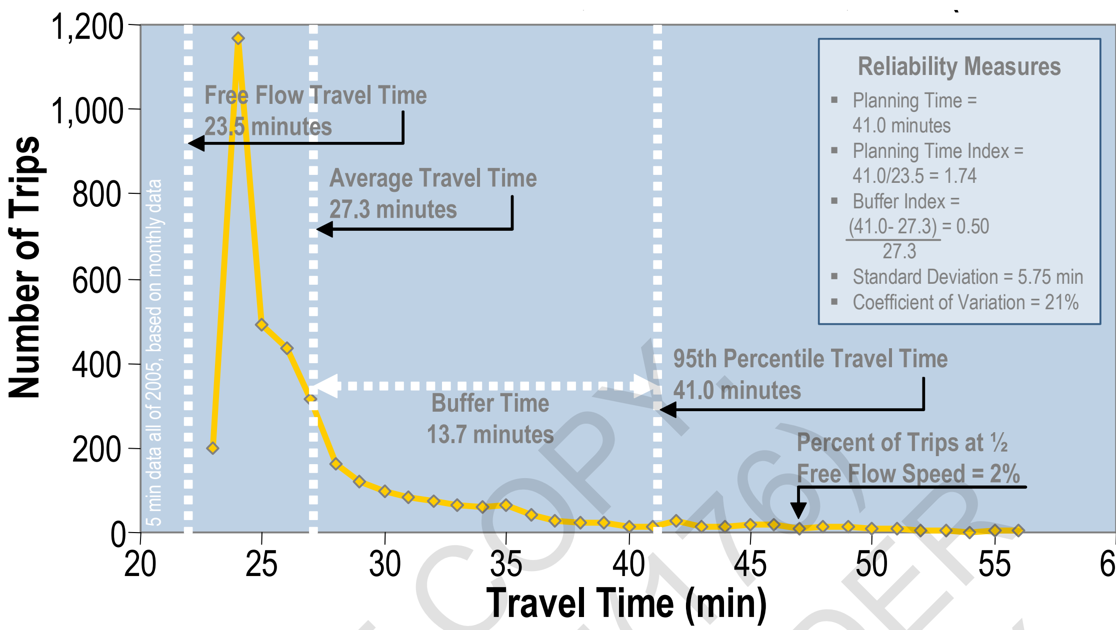
Corridor Level Analysis

Using all measured corridor travel times (5-minute resolution) for one year, Source: © Lyman and Bertini (2008).

Figure **46** shows a sorted travel time distribution. Free-flow travel time was 23.5 minutes, mean travel time was 27.3 minutes, standard deviation was 5.75 minutes, coefficient of variation was 21 percent, and 95th percentile travel time was 41.0 minutes. The figure also shows the buffer time (13.7 minutes), planning time index (1.74) and buffer index (0.50). For a traveler who wants to traverse the corridor and be on time 95 percent of the time (i.e., late once per month), a total travel time of 41.0 minutes should be reserved for this trip.

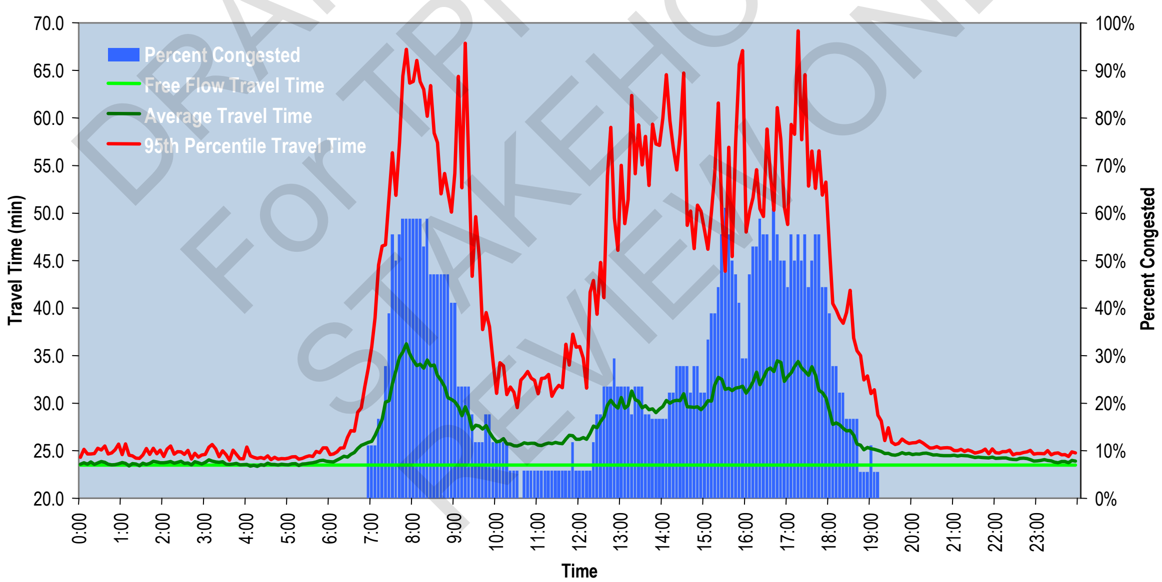
It's also possible to view how travel times vary over the day. Using 1 month's corridor travel time data (5-minute resolution), Source: © Lyman and Bertini (2008).

Figure **46** shows the free-flow travel time, mean travel time, 95th percentile travel time and the percent of monthly readings that were congested (threshold was defined at 1.3 times the free-flow travel time). The free-flow trip time is about 23 minutes, and an average 8 a.m. trip would take about 35 minutes. The 95th percentile travel time at 8 a.m. is about 67 minutes, revealing a required buffer time of 32 minutes at that time of the morning. As the 95th percentile curve drops closer to the mean travel time curve, the buffer time requirement is reduced.



Source: © Lyman and Bertini (2008).

Figure 46. Chart. Corridor travel time distribution for 1 year.



Source: © Lyman and Bertini (2008).

Figure 47. Chart. Corridor travel time and its variation for 1 month.

Reliability Considered over Time and Space

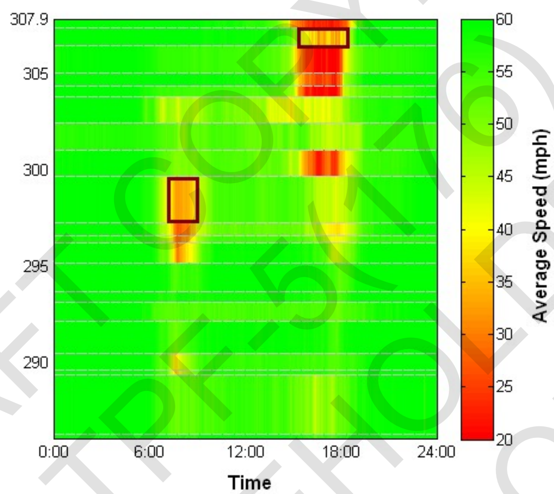
Reliability issues can be visualized across time (*x*-axis) and space (*y*-axis). Saberi and Bertini (2010).

B. Subfigure of congestion frequency.

Figure -A shows average speed for the same corridor in Portland using 5-minute data from weekdays during 2 months. Saberi and Bertini (2010).

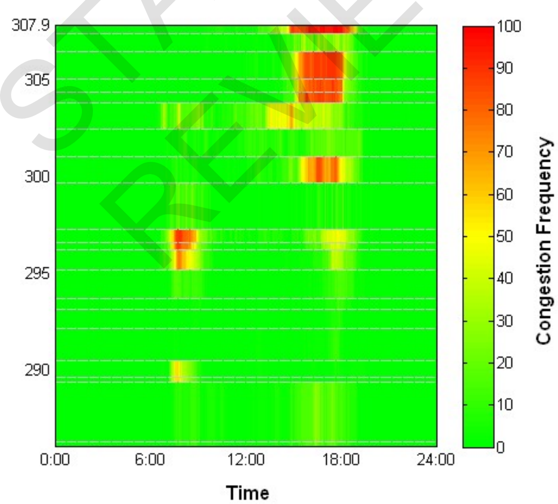
B. Subfigure of congestion frequency.

Figure -B shows congestion frequency using a threshold of 40 mph.



© Saberi and Bertini (2010).

A. Subfigure of average speed.

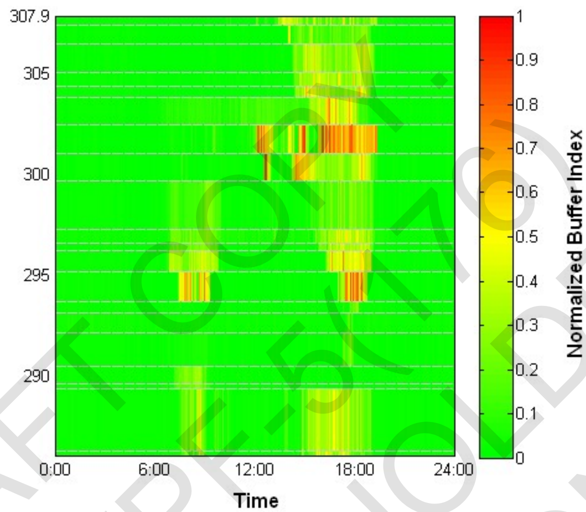


© Saberi and Bertini (2010).

B. Subfigure of congestion frequency.

Figure 48. Charts. Heat maps used to visualize reliability issues across time and space.

This can be extended to reliability measures, as shown in

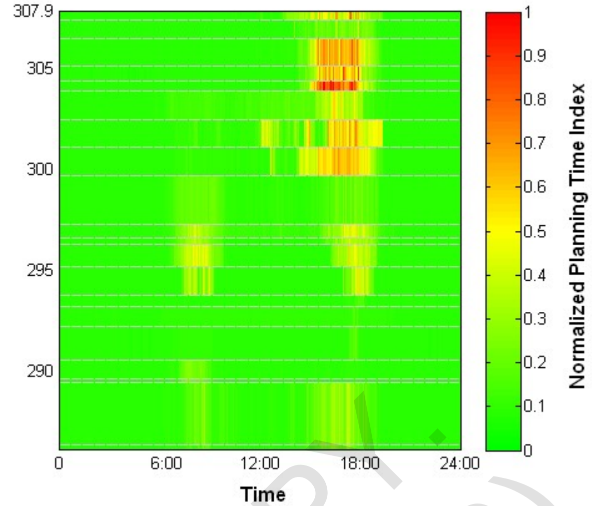


© Saberi and Bertini (2010).

D. Subfigure of coefficient of variation.

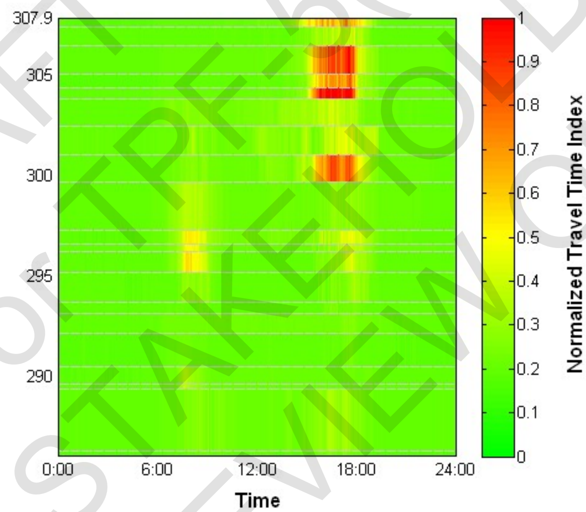
Figure 49 (buffer index, planning time index, travel time index, and coefficient of variation). In these examples, normalized values are used such that each measure is divided by its own statistical range (maximum–minimum). These figures indicate that each index would show approximately the same trends along the corridor and throughout the day. The travel time index (without reliability considerations) seems to heighten the trends more than planning time index.
© Saberi and Bertini (2010).

A. Subfigure of buffer index.



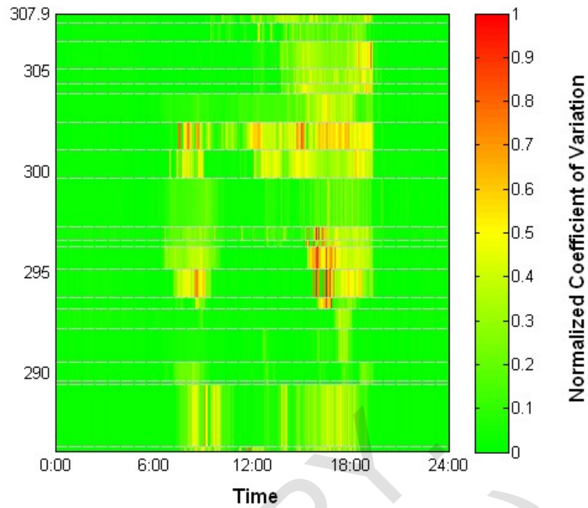
© Saberi and Bertini (2010).

B. Subfigure of planning time.



© Saberi and Bertini (2010).

C. Subfigure of travel time index.



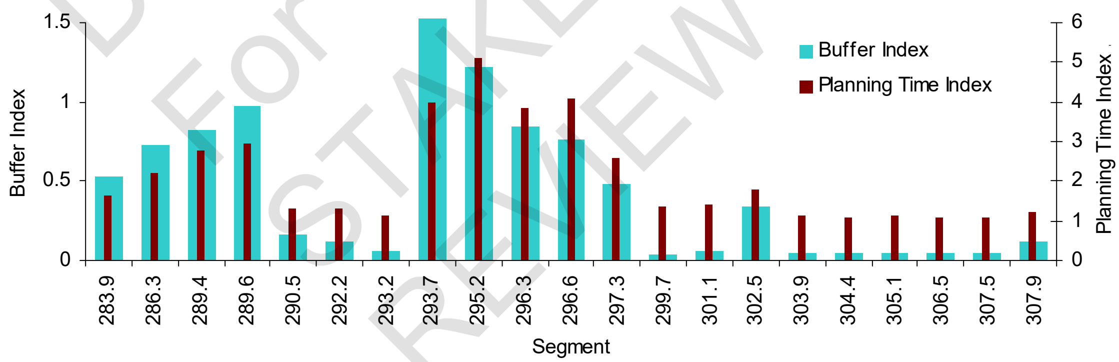
D. Subfigure of coefficient of variation.

Figure 49. Charts. Heat maps illustrating normalized corridor reliability.

Reliability Considered over Segments

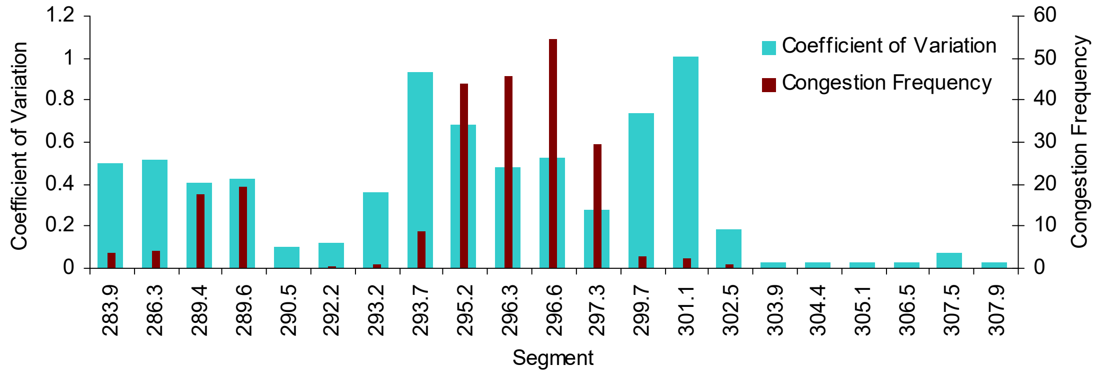
Figure 50 shows one final way of viewing reliability across corridor segments. Here weekday data from a 2-month period were used to compare AM peak reliability measures.

Travel time reliability, no matter which measure we use, varies substantially across freeway segments. Note that freeway segments are correlated, which influences the variability of corridor travel time and should not be ignored. Different reliability measures present different portraits of the reliability aspects of a freeway corridor. The buffer index and the coefficient of variation have a high consistency compared to other measures. The planning time index and the congestion frequency seem to follow similar trends. Segment travel time reliability measures can be used in regional transportation planning and operations to evaluate and prioritize roadway segments. Travel time reliability measures, both at the corridor level and segment level, can be used to highlight corridors or segments that are candidates for operational improvements. It is important to have access to several days of real data at each location for calibration purposes. A model should never be calibrated to reproduce a single day of data, because that day may not be representative of average conditions. It should also not be calibrated just to reproduce the average but also the variability observed in real data.



© Saberi and Bertini (2010).

A. Subfigure of buffer index and planning time.



B. Subfigure of coefficient of variation and congestion frequency.

Figure 50. Charts. Morning peak corridor segment buffer index, planning time index, coefficient of variation, and congestion frequency.

CHAPTER 7. MODEL DEVELOPMENT

7.1 INTRODUCTION

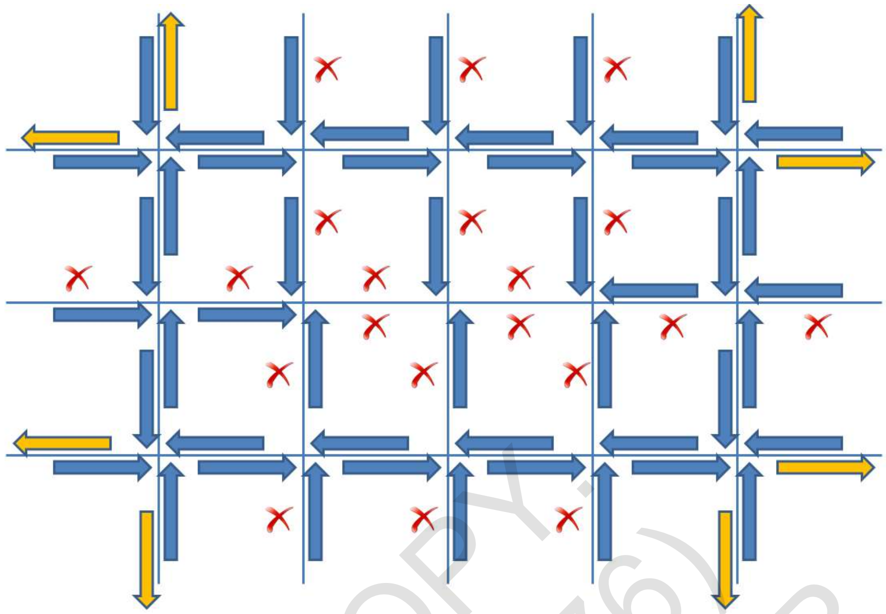
This chapter describes the process by which the project team should create the data files used as inputs for the simulation model.

7.2 INITIALIZATION, FILL TIME, AND SPATIOTEMPORAL LIMITS

> Many traffic simulation tools offer an input entry for initialization time, warm-up time, fill time, or some similar name. If the input flow rates are constant, the initialization time should allow the simulation model to reach a state in which the number of vehicles entering and exiting the network is roughly equal. This facilitates a more accurate simulation analysis.[12] An intersection, corridor, or network model should have an approximately equal number of vehicles entering or exiting the virtual system, prior to collection of vital performance measures. Failure to achieve this balance may mean vehicles have not had enough time to fill up the system, or indicate coding errors (e.g., missing links or nodes). In other cases, it may mean that conditions were oversaturated during the system initialization. In many cases, the accuracy of the simulation results may be compromised. Source: FHWA.

Figure 51 illustrates a traffic network where vehicles have not yet had enough time to fill and exit the network at all links and nodes. In such a case, congestion-related performance measures will be overly optimistic. A red X indicates links not yet reached by a significant number of vehicles.

[12] It is fine to not do this and have the simulation analysis encompass the entire simulation duration; but, the analyst then must realize that the output results include the transient associated with filling the network with traffic.



Source: FHWA.

Figure 51. Diagram. Traffic network during the initialization time.

When the number of vehicles exiting the network is not consistent with the inputs, the analyst should review the animated graphics to assess the underlying reason. In the case where vehicles have not yet had enough time to fill and exit the network, the initialization time should be increased. In the case where oversaturated conditions are preventing vehicles from exiting the system, the accuracy of the results can be increased by starting the simulation during an earlier time interval, in which congestion has not yet formed. The rule of thumb is that simulation results are best when conditions are undersaturated *both* during the first interval (following initialization) and during the final interval. It is fine for conditions to be oversaturated during any other interval.

Chapter 5. previously cited the benefits of well-defined spatial and temporal limits. When determining such limits, it is helpful to reflect on key questions the analysis is endeavoring to answer. For example, when testing control strategies to mitigate the three most congested areas in a city, it may be preferable to perform independent simulations of those specific corridors or junctions in isolation, as opposed to simulating the entire city. Moreover, those three bottlenecks may not occur during the same time of day, such that different temporal limits should be applied to each location. However, to simulate the best evacuation procedures and routes for the city, a single (large-scale) simulation spanning numerous time intervals may be necessary. Finally, simulations developed solely for public presentation may involve customized spatial and temporal limits; but when developing such data sets, engineers should consider the possibility of whether the data set may someday be reused for a more results-oriented analysis.

Beyond the key questions the analysis is intended to answer, another consideration involves proper measurement of congestion-related traffic network performance. When conditions are undersaturated throughout the typical day, conventional wisdom indicates that a peak 15-minute analysis is often adequate when assessing existing conditions. However, when analyzing oversaturated conditions, performance measures cannot be accurate unless congestion forms and dissipates within the chosen time intervals and physical network structure. In other words, the beginning of the first interval (and end of the final time interval) should exhibit undersaturated conditions at all links and nodes. This can be visualized by the congested area of a two-dimensional heat map being fully encapsulated within the analysis box, as shown in

Figure 52. This is sometimes referred to as shouldering (i.e., adequate modeling and analysis of time intervals at the shoulders of the peak periods).

In some cases, practical considerations prevent the use of ideal spatial and temporal limits. For example, even though oversaturated conditions persist for 2 hours, available resources or data only support a 15-minute simulation. In another example, although oversaturated conditions extend across 10 miles, resources and/or data only support simulation of a 2-mile radius. The scenario of inadequate temporal limits is discussed first.

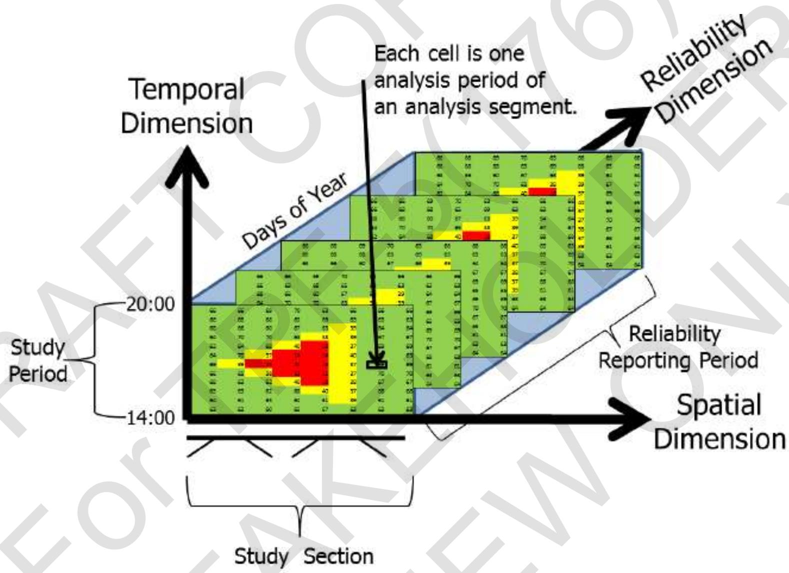
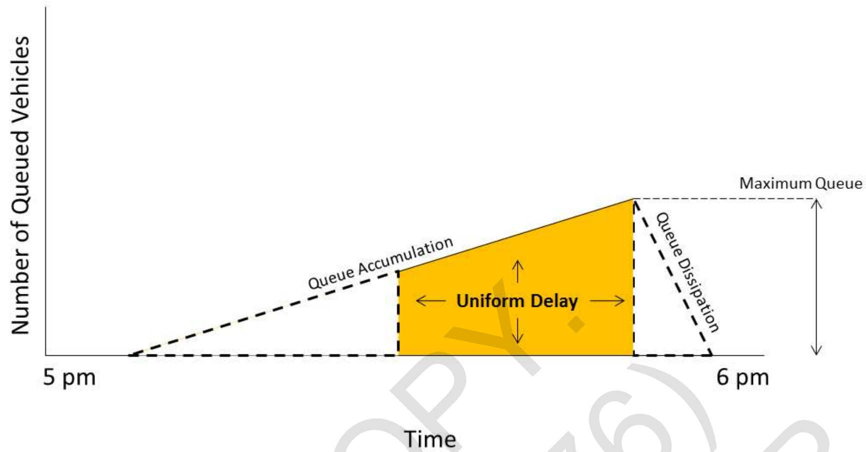


Figure 52. Diagram. Spatiotemporal heat maps in three dimensions.

The impact of temporal limits can be conveyed through the queue accumulation polygon (QAP). Source: FHWA.

Figure 53 illustrates a QAP for a hypothetical roadway segment. At approximately 5:10 p.m., the QAP starts to develop a steadily increasing queue length, which may have been invoked by any number of factors (e.g., demand exceeding capacity on this segment, queue blockage from a downstream segment, queue spillover from a turn pocket, or the presence of a midblock shopping center). The QAP depicted in Source: FHWA.

Figure 53 represents real-world conditions that the analyst wishes to capture in a simulation model. However, if the analyst chooses restrictive temporal limits (e.g., 5:30–5:45 p.m.), this will obscure key portions of this operation.



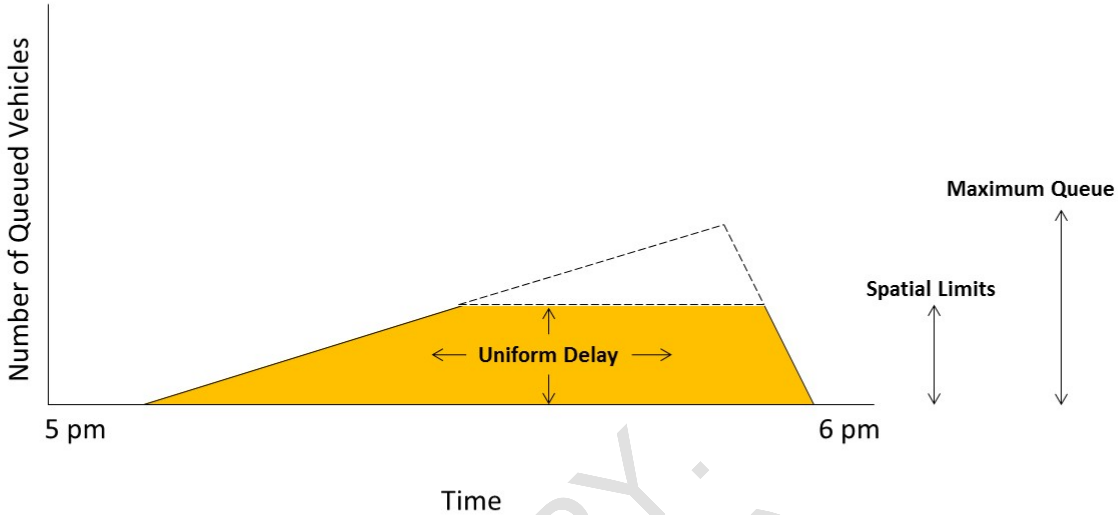
Source: FHWA.

Figure 53. Illustration. Queue accumulation polygon (QAP) with restrictive temporal limits.

Many of the core mobility-related performance measures available through simulation are closely related to uniform vehicle delay, which may be computed as the two-dimensional area under the QAP. As such, the ideal temporal limits for analyzing this roadway segment would at least include the 5–6 p.m. time window to capture the full QAP shape. Moreover, in a Monte Carlo analysis or scenario analysis, some realizations may produce QAP shapes that extend beyond the 5–6 p.m. window. Thus, the inclusion of some extra buffer time within the temporal limits may be helpful. If the simulation were to be terminated early (e.g., at 5:45 p.m.), output performance measures may be highly inaccurate, because delays experienced by vehicles entering between 5:30 and 5:45 p.m. cannot be fully measured without simulating 5:45–6 p.m. However, in cases where moving vehicle animation is a higher priority than the numeric output statistics, an early end to the simulation will not compromise the accuracy of such animation. Conversely, if the simulation were to begin late (e.g., at 5:30 p.m.), queue lengths at the start of the run (following the initialization fill time) might not be realistic. The analyst could try to compensate by calculating a special initialization time, which would accurately reproduce the initial set of queue lengths. The risk of this approach is that the initial queue lengths would presumably be more accurate if the prior time interval(s) had been fully simulated.

The scenario of inadequate spatial limits is discussed next. Source: FHWA.

Figure 54 illustrates another representative QAP for a hypothetical roadway segment. In this case, the maximum queue length on this roadway segment is quite long, but the spatial limits are only able to capture approximately half of that distance. As a result, numeric output statistics and moving vehicle animations will only reflect approximately half of the actual queue length.



Source: FHWA.

Figure 54. Illustration. Queue accumulation polygon (QAP) with restrictive spatial limits.

The analyst should use judgment to assess the consequences of restrictive spatial limits. On the one hand, decision makers reviewing the animated graphics would presumably get the point that oversaturation extends beyond this point of the network for certain periods of time. For certain public presentations or problem assessments, this may be perfectly adequate. In other words, there may be little benefit from investing the additional resources that would be required to capture the full extent of congestion. Moreover, roadway segments that experience queue spillback beyond the entry nodes may not be the most important segments in the analysis. They may be low-priority locations whose numeric output statistics will not factor into the decision-making process. Some tools even offer special performance measures, such as unmet demand, unserved vehicles, and latent delay, to estimate the impact of the unseen vehicles.

In other cases, the restrictive spatial limits may prove unacceptable. Queues may spill back beyond the model boundaries. Demand may be lost because vehicles cannot enter the network. Delays may be underestimated because the QAP is artificially capped at an arbitrary level (see Source: FHWA.

Figure 54). In addition, signal timing optimizations accounting for delays on all approaches may not assign adequate amounts of green time to approaches whose spatial limits have been artificially truncated.

Given these factors that can make spatial limits acceptable and unacceptable, the analyst must make informed decisions to extend the spatial limits in certain critical areas while allowing restrictive spatial limits in other, less critical network locations.

7.3 NETWORK DEVELOPMENT

Today's simulation tools offer several basic methods of data entry and traffic network development. These network development methods present several corresponding advantages and disadvantages related to:

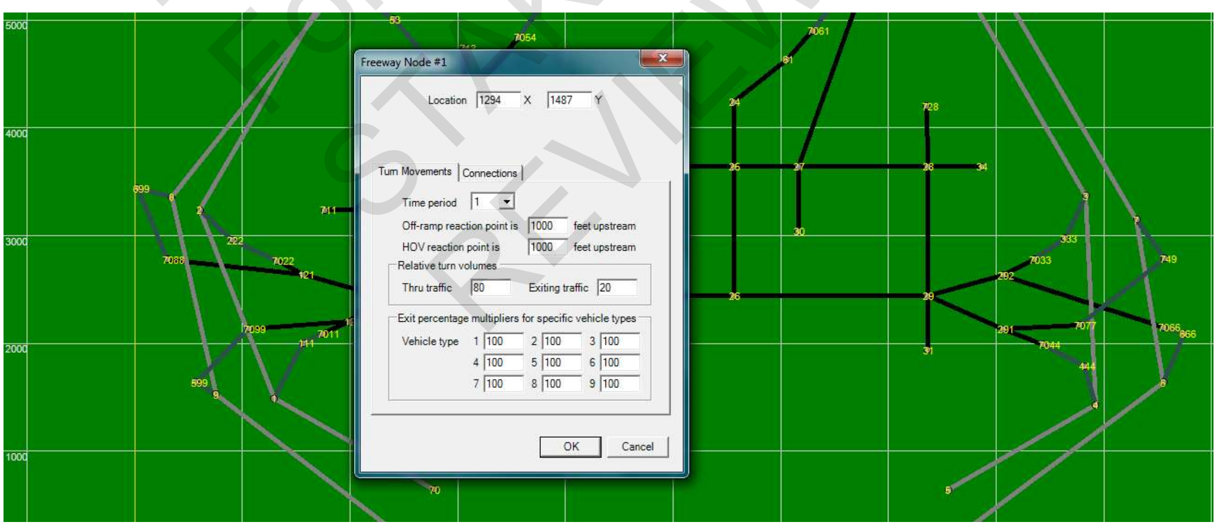
- Level of effort (required number of person-hours).
- Required expertise.
- Monetary cost of software tools.
- Precision of the simulated network.
- Assumptions and approximations within the simulated network.

The choice of network development method should reflect the analysis objectives and resources. The best network development method for one traffic modeling project may not be an efficient or effective choice for subsequent projects. As is often the case, engineering judgment is paramount. This chapter may provide some insights to facilitate better judgment.

Conventional Windows-Based Data Entry

Following the widespread adoption of Microsoft® Windows in the 1990s, many traffic simulation tools adopted a sketch-based data entry approach. In this approach, simulation tool users would begin by essentially drawing a series of two-dimensional links and nodes. This approach was intuitive, because the link-node diagram could resemble the area being modeled. Users could then click on specific links or nodes to view or edit their properties. For example, users could click on a freeway node having two downstream (i.e., mainline and off-ramp) links, and could then view or edit the off-ramp exiting flow rate at that location. This concept is illustrated in Source: FHWA.

Figure 55. By clicking on links and nodes to edit their properties, users could specify location-specific data such as the number of lanes on each segment, segment lengths, auxiliary lane lengths, demand volumes, signal timings, free-flow speeds, and so on. Global data, such as the number of time intervals, initialization fill time, and random number seeds, could be entered in some other area of the software unrelated to the links and nodes.



Source: FHWA.

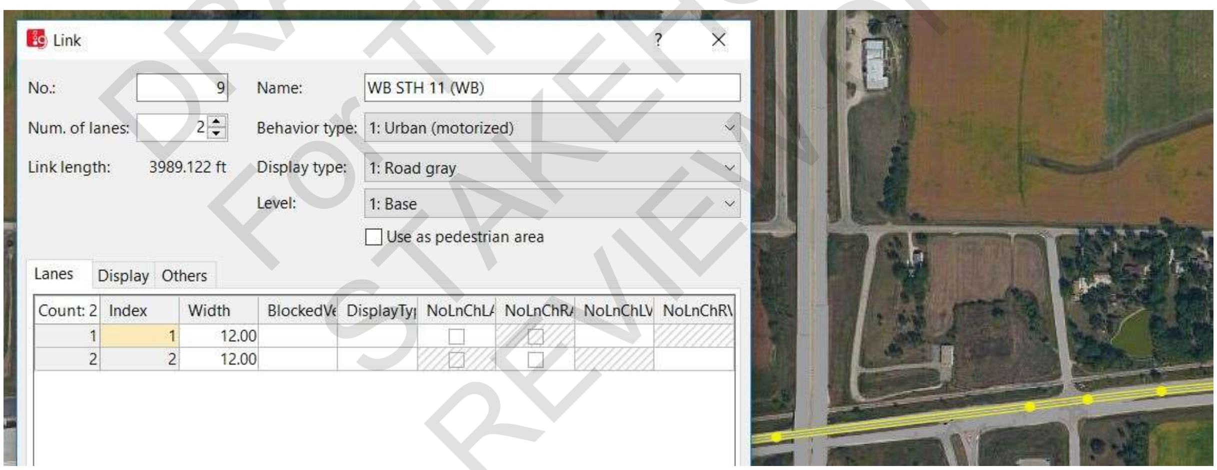
Figure 55. Screenshot. Sketch-based data entry for traffic simulation.

Data Entry over a Bitmap Background

Shortly after the sketch-based data entry approach was developed, traffic simulation vendors followed up with a helpful supporting feature which would facilitate more intuitive data entry. They implemented the ability to superimpose the link-node or link-connector diagram on top of an aerial photo or map image. This bitmap background could be implemented and/or presented in several electronic image formats such as bitmap image file (BMP), Joint Photographic Experts Group (JPEG), tagged image file format (TIFF), graphic interchange format (GIF), and others. Before the aerial photo can be used, there is typically a scaling step in which the engineer must indicate how the image size (e.g., measured in inches) relates to real-world distances (e.g., measured in feet), and how the image is oriented with respect to the simulator's *x-y* coordinate system. In addition to making the data entry process more intuitive (e.g., © Wisconsin Department of Transportation.

Figure 56), the bitmap background also improved visualization of the traffic animation (e.g., © Wisconsin Department of Transportation.

Figure 57). The bitmap background advancement provided the ability to contextualize the traffic network for better understanding by engineers, decision makers, and the public. Despite their usefulness, bitmap backgrounds are often omitted for smaller projects and/or research studies.



© Wisconsin Department of Transportation.

Figure 56. Screenshot. Typical background photo for sketch-based traffic simulation network editing.

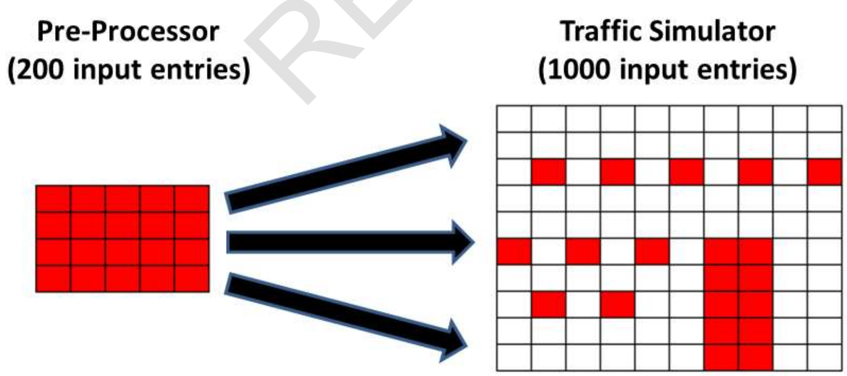
© Wisconsin Department of Transportation.

Figure 57. Screenshot. Traffic animation superimposed above a bitmap background.

Pre-Processors

Simulation products usually offer at least one native sketch-based network editor, which can edit the full array of available input data. However, a separate set of pre-processor tools are designed to generate traffic networks more quickly and easily. This is because the number of input data entries is much smaller in the pre-processor. For example, a pre-processor might generate an error-free simulation network after the analyst enters only 20 percent of the required input data (see Source: FHWA.

Figure 58). For the other 80 percent of the data, default values are typically generated. Therefore, pre-processors can be a valuable mechanism for saving time.



Source: FHWA.

Figure 58. Illustration. Data entry efficiency provided by a pre-processor.

Pre-processors provide access to the most important and fundamental input data (e.g., demand volumes, number of lanes, signal timings). Some pre-processors require the drawing of a two-dimensional link-node diagram, whereas others are capable of automatically generating a default link-node diagram (usually a linear facility with multiple junctions). Pre-processors often do not provide access to optional input data that are used less often. These optional data may pertain to calibration, complex roadway geometry, advanced traffic signal timing parameters, and advanced traffic management strategies. As such, for major simulation projects requiring advanced analysis, use of a pre-processor may not be sufficient. In such cases, the analyst may need to perform all their data entry in the native editor (which may be a time-consuming process); or switch over to the native editor after initially generating a baseline network in the pre-processor. © McTrans Center.

Figure 59 provides an example of a pre-processor's data entry screen, which automatically generates a linear freeway facility having multiple junctions. This data entry process would be much faster than the traditional method of drawing a link-node diagram, and then clicking on all links and nodes to edit their properties.

	From	To	Input Segment Type	Length (ft)	Terrain	Edit Segment	Adjusted Demand (veh/h)	% Trucks	% RVs	Number of Lanes	Free-Flow Speed (mi/h)	Adj. On-Ramp Demand (veh/h)	On-Ramp % Trucks	On-Ramp % RVs	On-Ramp Lanes	On-Ramp FFS (mi/h)
1	a	b	Basic Segment	5280	Level	Edit	4505	5.00	0.00	3	60					
2	b	c	On-Ramp	1500	Level	Edit	4955	5.00	0.00	3	60	450	5.00	0.00	1	40
3	c	d	Basic Segment	2280	Level	Edit	4955	5.00	0.00	3	60					
4	d	e	Off-Ramp	1500	Level	Edit	4955	5.00	0.00	3	60					
5	e	f	Basic Segment	5280	Level	Edit	4685	5.00	0.00	3	60					
6	f	g	Weaving	2640	Level	Edit	5225	5.00	0.00	4	60	540	5.00	0.00	1	40
7	g	h	Basic Segment	5280	Level	Edit	4865	5.00	0.00	3	60					
8	h	i	On-Ramp	1140	Level	Edit	5315	5.00	0.00	3	60	450	5.00	0.00	1	40
9	i	j	Ramp Overlap	360	Level	Edit	5315	5.00	0.00	3	60					
10	j	k	Off-Ramp	1140	Level	Edit	5315	5.00	0.00	3	60					
11	k	l	Basic Segment	5280	Level	Edit	5045	5.00	0.00	3	60					

© McTrans Center.

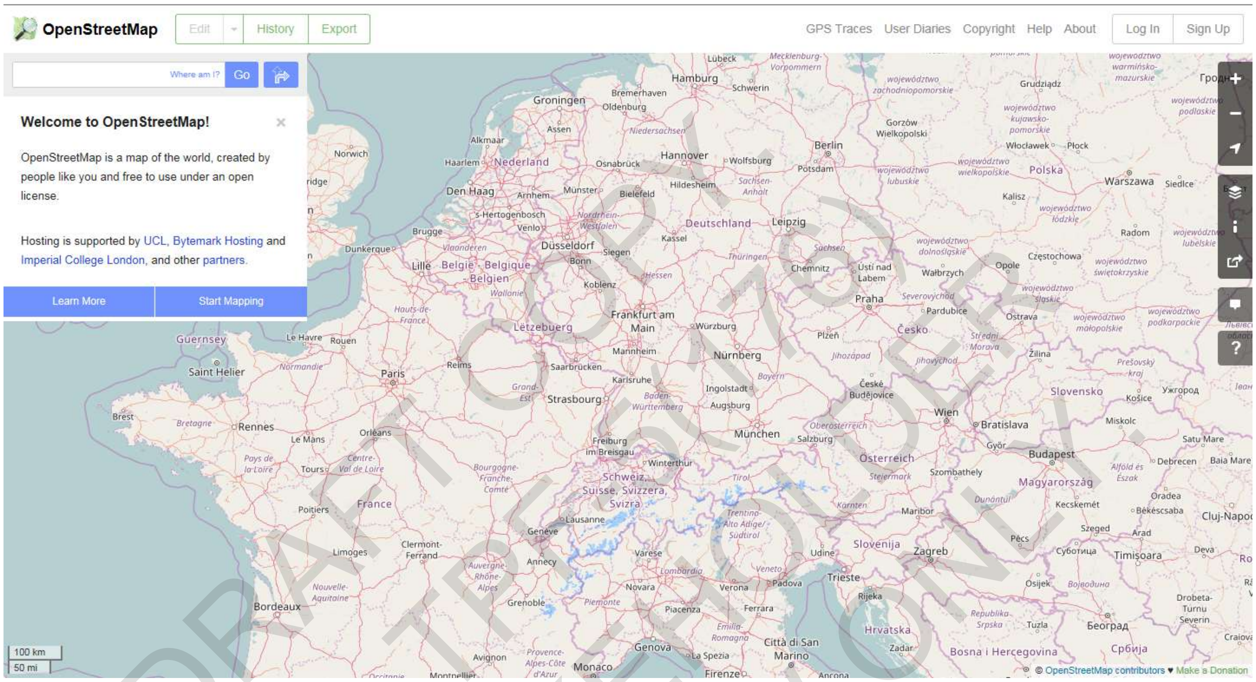
Figure 59. Screenshot. Sample expedited data entry in a pre-processor.

Importing Detailed Traffic Data

In addition to pre-processors and native sketch-based editors, a third network creation method involves importing detailed traffic data, usually from spatially accurate geographic information systems (GIS). This method allows the efficient importing of large quantities of input data, saving time and money. This method may be just as fast (or faster) than pre-processors. Some of today's simulation tools can import link-node geometries, lane configurations, posted speed limits, demand volumes, turn movement volumes, and signal timings. When the process works perfectly, complex traffic networks may be simulated within minutes, without any data entry. However, the process rarely works perfectly. In most cases, imported data must be edited and corrected in the simulator's native sketch-based editor. © 2019 OpenStreetMap™ contributors.

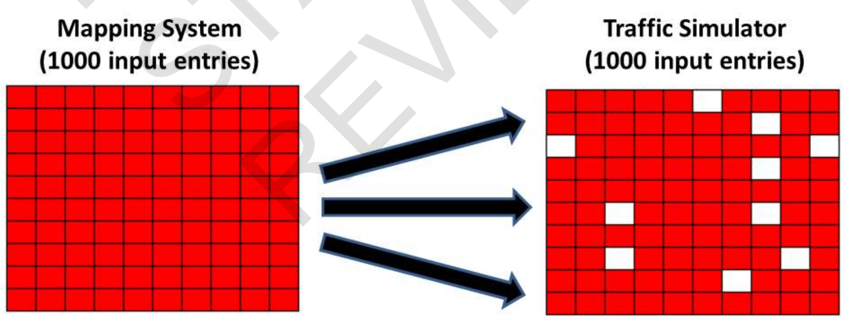
Figure 60 and Source: FHWA.

Figure 61 are intended to illustrate that although GIS may contain just as much data as the traffic simulator, some highly specialized and esoteric data items (e.g., advanced signal controller settings, surveillance detector locations, and vehicle fleet composition) are probably not available from GIS. Thus, the data entry process will involve filling in missing (specialized) data and fixing any data that were not properly imported. The efficiency of creating a network this way will depend on what percentage of data are properly imported, the analyst's proficiency with the software tools, the size or complexity of the traffic network, and other factors.



© 2019 OpenStreetMap™ contributors.

Figure 60. Map. Traffic data available through OpenStreetMap™.



Source: FHWA.

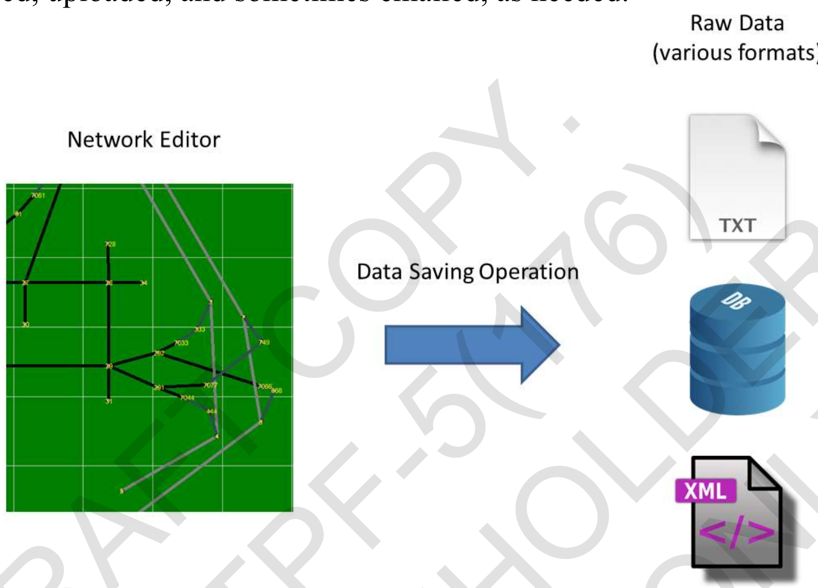
Figure 61. Illustration. Data entry efficiency provided by a GIS import process.

Raw Data Entry

When traffic simulation data must be saved for future use, each network editor automatically archives the information into one or more data sets. These data sets come in many raw formats, including extensible markup language (XML), comma-separated values (CSV), binary, simple text, and others as illustrated in Source: FHWA.

Figure 62 and Source: FHWA.

Figure 63. These data sets may then be reopened later for further analysis. The data sets may also be shared, copied, uploaded, and sometimes emailed, as needed.



Source: FHWA.

Figure 62. Illustration. Raw data storage formats.


```
<PhaseSplits ID="NBT">41.0</PhaseSplits>
<PhaseSplits ID="SBL">0.0</PhaseSplits>
<PhaseSplits ID="SBT">41.0</PhaseSplits>
<Recall ID="EBL">Off</Recall>
<Recall ID="EBT">Min</Recall>
<Recall ID="WBL">Off</Recall>
<Recall ID="WBT">Min</Recall>
<Recall ID="NBL">Off</Recall>
<Recall ID="NBT">Off</Recall>
<Recall ID="SBL">Off</Recall>
<Recall ID="SBT">Min</Recall>
<DualEntry ID="EBL">No</DualEntry>
<DualEntry ID="EBT">Yes</DualEntry>
<DualEntry ID="WBL">No</DualEntry>
<DualEntry ID="WBT">Yes</DualEntry>
```


Source: FHWA.

Figure 63. Illustration. Example of raw data stored in extensible markup language (XML) format.

In some cases, there is little benefit to directly editing the raw data without assistance from a graphical network editor. However, in other cases, it is more efficient to directly edit the raw data, and the ability to do so may become important to a project. For example, if it became necessary to systematically view and/or edit a specific segment characteristic across all segments, it might be much faster to accomplish this inside the raw data set, instead of repeatedly clicking on hundreds of links to view and edit their properties. Some simulators also offer a scripting tool, which can systematically edit a certain characteristic across all nodes or segments.

7.4 GEOMETRIC DATA

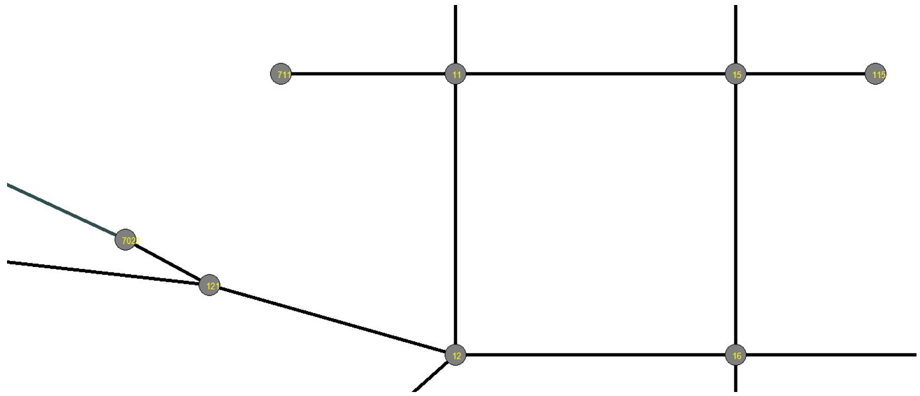
Different simulation tools offer fundamentally different ways of specifying traffic network geometry. These methods provide various advantages and disadvantages in terms of modeling fidelity, granularity, and speed of data entry. Accurate specification of geometric data is a prerequisite to any worthwhile simulation effort. Geometric data are typically entered very early in the network creation process. As such, critical information (such as segment lengths, segment curvature, intersection locations, number of lanes, auxiliary[13] lane lengths, and merge/diverge junction locations) can be verified prior to the specification of more detailed data (e.g., volume demands, intersection control, posted speed limits, signal timings, load conditions).

Links

> A link, generally, is a roadway segment that connects two features, junctions, or intersections. A link may also connect a roadway feature to a junction, or a junction to an intersection, or an intersection to a feature. A roadway feature may be a horizontal bend in the roadway, a location that changes the posted speed limit, a lane drop location, or a change in roadway elevation. A junction may be a freeway merge location, a freeway diverge location, or an uncontrolled intersection. The term intersection is often applied to sign-controlled or signal-controlled locations on surface arterials. In some countries and in some traffic simulation tools, the term junction may be used to describe intersections. The terms link and segment are somewhat interchangeable. Source: FHWA.

Figure 64 illustrates a typical link-node diagram.

[13] Auxiliary lanes may be left-turn and right-turn pockets at intersections, or acceleration and deceleration lanes on freeway ramps. Acceleration and deceleration lanes are also present on some surface arterials.



Source: FHWA.

Figure 64. Illustration. A typical link-node diagram.

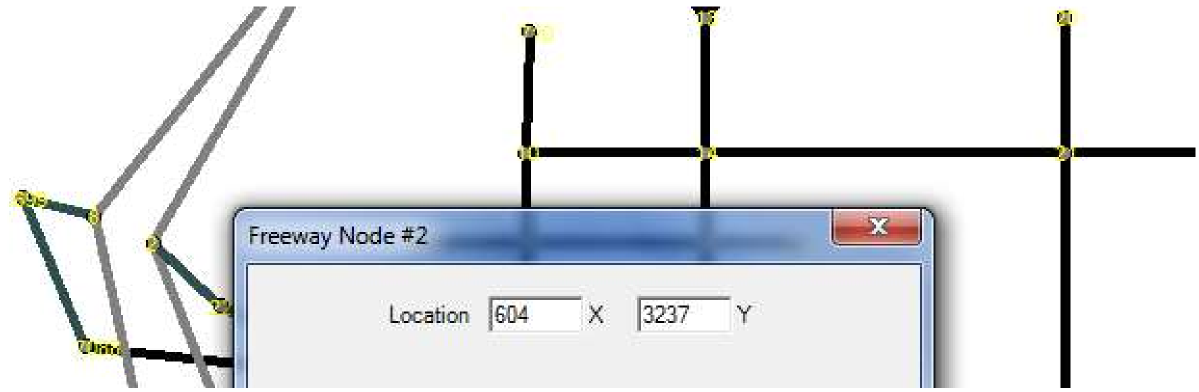
While the link length is one of the most important link characteristics, links may contain other characteristics such as curvature data, number of lanes, existence of auxiliary lanes, one-way versus two-way travel, auxiliary lane lengths, and link type. Regarding link type, some tools require specifying whether any given link is a surface link, freeway link, or ramp link. Other tools offer different link types for different area types (e.g., rural, urban, suburban). In most graphical network editors, links may be drawn or stretched using the mouse. Drawing or stretching a link may automatically update the associated link lengths and curvature, but this tool-specific functionality should be confirmed by the end user.

Nodes

A node, generally, is an intersection, a junction, or a feature point. While intersections and junctions are usually obvious choices for node locations, some node locations may require more engineering judgment. As stated previously, a roadway feature may be a horizontal bend in the roadway, a location that changes the posted speed limit, a lane drop location, a change in roadway elevation, or another geometric or system change that could affect traffic flow. The use of nodes to define such locations may be strategic on the part of the engineer. For example, if nodes are not placed at such locations, it may in some cases compromise the accuracy of the model. On the other hand, if nodes are placed in too many locations, this may increase the complexity of the model, increase the probability of data entry errors, and inflate the effort level required for network development. Eliminating unnecessary nodes can allow faster simulations. Some simulation tools can model lane drops, warning sign locations, and other load condition changes within the link itself, without needing a new node. Other tools offer the ability to define feature points that contain some node characteristics but are generally less significant than nodes.

> Many editors offer specification of node coordinates in terms of longitude (x), latitude (y), and sometimes altitude (z). Nodes can often be added, dragged, dropped, and edited. In some editors, dragging a node from one location to another may automatically update the associated node coordinates and connecting link lengths. Node coordinates are sometimes proportional to real-world units of measurement such as feet or meters. Source: FHWA.

Figure 65 illustrates the node coordinate concept.



Source: FHWA.

Figure 65. Illustration. Node coordinates.

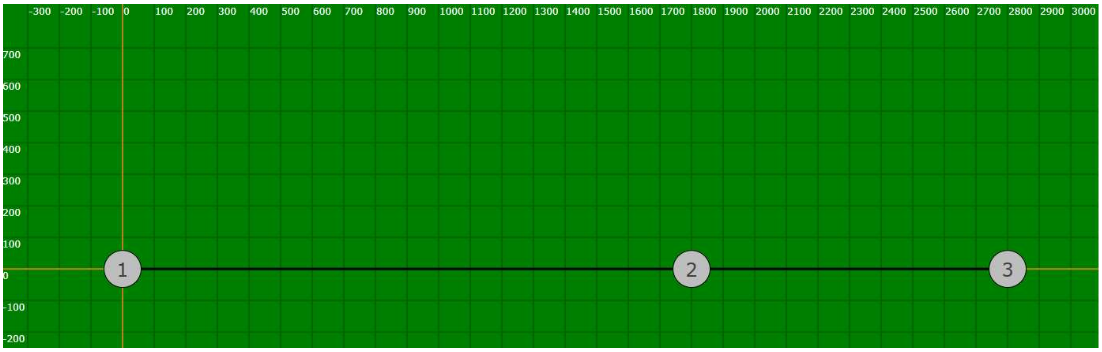
The *z*-coordinates may only be relevant when the simulator can model vertical grade effects on vehicle acceleration, or when the simulator supports three-dimensional (3D) animation. The *x*- and *y*-node coordinates should generally reflect the segment length between them, although segment lengths that exceed the (Pythagorean) distance between node coordinates may appropriately indicate a certain amount of curvature. Otherwise, inconsistencies between node coordinates and their associated link lengths may produce anomalies in the moving vehicle animation, and in the worst case may compromise the accuracy of simulation results.
Beyond the node coordinates, other node characteristics may include control type (e.g., uncontrolled, stop-controlled, yield-controlled, roundabout, pre-timed signal, actuated signal, ramp meter) and node type. Node types may include freeway nodes, surface nodes, and interface nodes that connect freeways and surface streets. However, not all simulation tools require specification of a node type.

Link-Node versus Link-Link Connections

Although links and nodes are common elements among the architecture of traffic simulation tools, some simulation tools do not offer or recognize nodes. Instead, they only recognize direct link-to-link connections. This paradigm presents significant advantages and disadvantages that should be considered during the tool selection process. Source: FHWA.

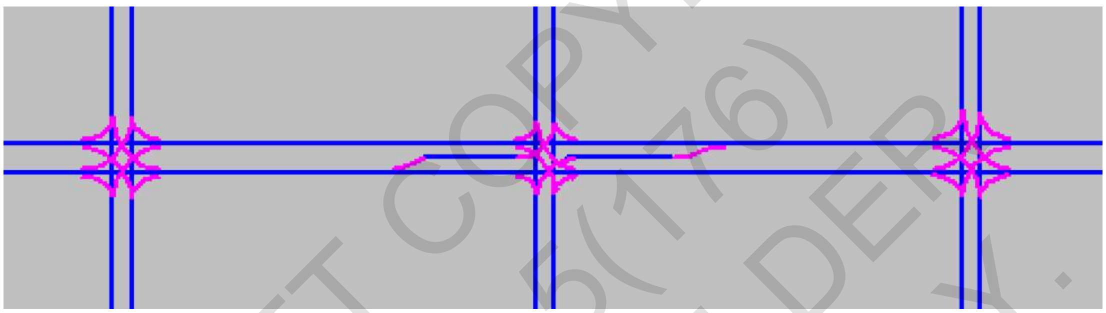
Figure 66 and © Wisconsin Department of Transportation.

Figure 67 provide a visual contrast between the two basic types of traffic network architecture.



Source: FHWA.

Figure 66. Illustration. Link-node architecture.



© Wisconsin Department of Transportation.

Figure 67. Illustration. Link-to-link architecture.

Link-to-link connection architectures are generally more flexible than link-node architectures, in terms of their ability to simulate advanced traffic network geometries and control strategies. One example of this is lane-by-lane modeling, in which individual lanes can be defined as unique links. This allows individual lanes to have unique demand volumes, free-flow speeds, saturation flow rates, and sign or signal control. Another example is contraflow lanes. Link-node architectures tend to require all lanes on a link to serve traffic flow in only one direction, but link-to-link architectures support spatially overlapping links that can serve traffic flow in any direction. Link-to-link architectures can seamlessly represent unconventional geometries, like the displaced left-turn intersection (also known as the continuous flow intersection), signalized intersections with more than five approaches, and different control types (e.g., stop sign, yield sign, signal, ramp meter) applied to adjacent lanes.

Although advantageous in many complex modeling situations, link-to-link networks tend to require more data entry time, resources, and experience. By contrast, link-node architectures may be easier to use. They may reduce the possibility of data entry error and facilitate the coding of large networks in shorter time. Due to their inherent simplicity, link-node networks may be easier to calibrate and may allow faster running times on the computer.

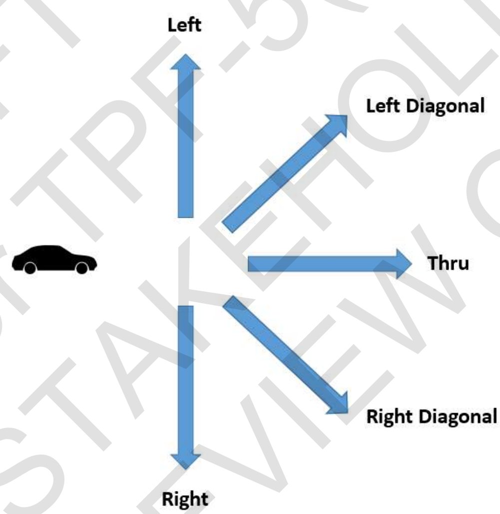
Orthogonal and Cardinal Directions

Traffic simulation tools offer different recognition and treatment of orthogonal directions (e.g., left, through, right, left diagonal, right diagonal) and cardinal directions (e.g., north, south, east, west, northeast, northwest, southeast, southwest). When the orthogonal and cardinal directions are fully supported and recognized by a given tool, this can facilitate the data entry process, and can make the output reports more intuitive.

However, the need to define directions may also be limiting. Some link-connector architectures are direction agnostic, which gives them great flexibility to simulate unusual geometries. For example, any upstream link can be allowed to feed any downstream link. This may allow lane-by-lane input data and output data. It could allow simulation of an unlimited number of intersection approaches. Ultimately, the fine-grained technical details associated with how a tool handles orthogonal and cardinal directions can determine the proportion of real-world conditions that can be simulated, ease of data entry, user-friendliness of output reports, interoperability with other tools, and many other factors.

Source: FHWA.

Figure 68 illustrates an example of orthogonal and diagonal turn movement directions. The treatment and recognition of diagonal movements varies widely among the tools; some tools do not differentiate between left turns and right turns.



Source: FHWA.

Figure 68. Illustration. Orthogonal turn movement directions.

Source: FHWA.

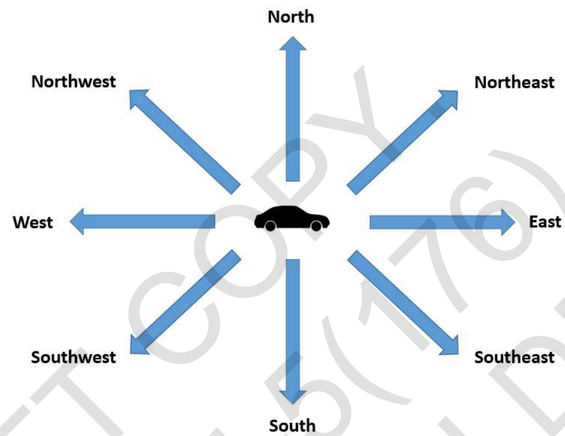
Figure 69 and Source: FHWA.

Figure **70** illustrate cardinal directions.

Source: FHWA.

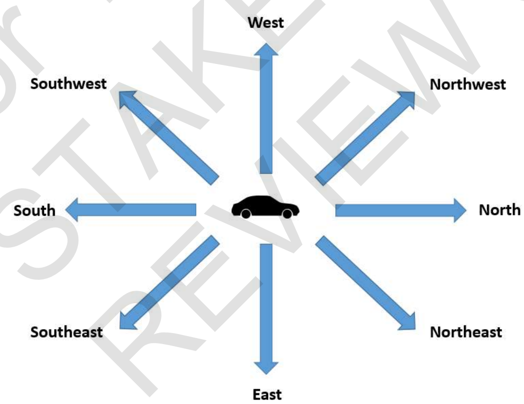
Figure 69 shows the most common and conventional treatment, in which northbound vehicles are assumed to move from the bottom of the screen (or page) toward the top. However, some simulation tools allow the user to select from many customized sets of cardinal directions. For example, in Source: FHWA.

Figure **70**, northbound vehicles are defined as moving from left to right. In most cases, cardinal directions exist primarily for human understanding and interpretation. In other words, they would not impact numeric simulation results as the orthogonal directions potentially could.



Source: FHWA.

Figure 69. Illustration. Conventional cardinal directions.



Source: FHWA.

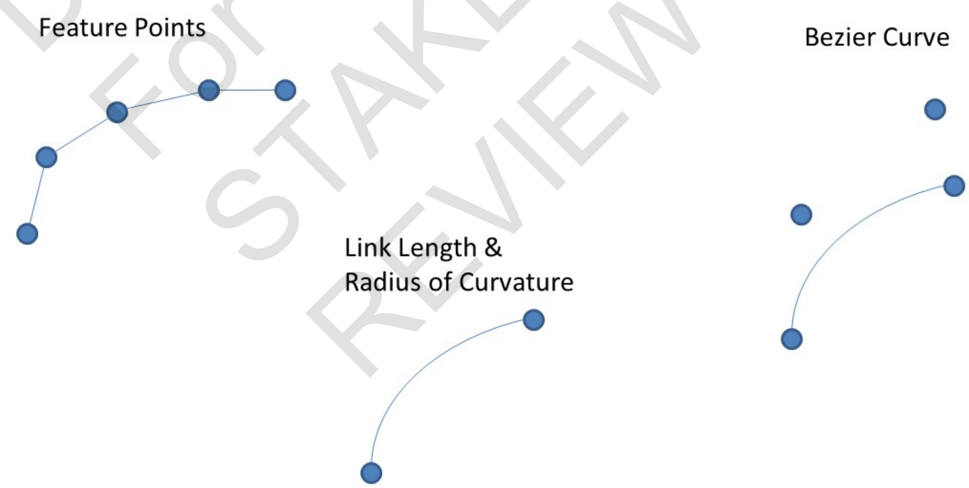
Figure 70. Illustration. Customized cardinal directions.

Link Lengths and Horizontal Curvature

As stated earlier, the terms link and segment are somewhat interchangeable in traffic simulation modeling. Collectively, segment lengths are a critical parameter that governs traffic simulation outcomes. Older traffic analysis tools were only able to model isolated intersections, linear signalized arterials, linear freeway facilities, and two-dimensional (2D) grid systems with 90-degree approaches. Today's traffic simulation tools can model highly complex and unusual roadway geometries, including all types of roadway curvature. This includes both horizontal and vertical curvature. This subsection addresses horizontal curvature, while the next subsection addresses vertical curvature (i.e., super-elevation and grade).

When discussing horizontal link curvature, it is helpful to differentiate between data entry methods and simulation modeling methods. The actual simulation models, regardless of whether they are at the microscopic, mesoscopic, or macroscopic levels of detail, may not explicitly or automatically adjust vehicle speeds, accelerations, emissions, or other characteristics in response to horizontal curvature in the network. Instead, the end user may be expected to calibrate input parameters (such as free-flow speed, speed limit, acceleration/deceleration, car-following behavior, and lane-changing behavior) on segments having significant horizontal curvature. In some advanced simulation tools, the end user can specify advanced vehicle dynamics parameters that would come into effect on segments having significant horizontal curvature.

Data entry methods for expressing horizontal curvature are thus used to produce more realistic looking data entry maps (which can help to prevent data entry errors), and more realistic looking traffic animations. Some tools offer many feature points along each segment. Although each section of roadway is completely straight between feature points, a larger number of feature points will still produce a curved appearance. Some tools assume link curvature if the link length exceeds the (Pythagorean) computed distance between end points, although the degree of curvature can be influenced by a user-specified radius of curvature. Finally, many tools use the Bezier curve method of graphical curve drawing. In this case, the user can literally draw the shape of the curve, and the software need only store the control point coordinates that are needed to draw the curve.



Source: FHWA.

Figure 71. Illustration. Common data entry methods for horizontal curvature.

Super-Elevation and Grade

When discussing vertical link curvature and grade, it is again helpful to differentiate between data entry methods and simulation modeling methods. Some simulation models automatically and explicitly adjust vehicle accelerations, emissions, and saturation flow rates when those vehicles encounter vertical grades in the network. Otherwise, a level link and a steep link would produce identical outputs if they have the same input length. In some advanced simulation tools, the end user can specify advanced vehicle dynamics parameters that would come into effect on segments having significant vertical grades.

For proper modeling of vertical grades, some illustrative concepts and examples may be helpful. First, regarding intersection operations, the parameters saturation flow rate and queue discharge headway are highly correlated. Vertical grade tends to significantly impact both, in that climbing a slope tends to increase headways (decreasing flow rates) whereas driving downhill tends to decrease headways (increasing flow rates). Nominally, one could say that a 2-second headway between vehicles would produce a saturation flow rate of 1,800 vehicles per lane per hour (veh/ln/hr), or that any per-lane saturation flow rate could be computed as 3,600 divided by the discharge headway. In practice, an accurate estimate may not be that simple. The Highway Capacity Manual (HCM) suggests recording the time of passage of the fourth and 10th vehicles over several cycles. Ideally, discharge headways and saturation flow rates should be obtained from local field measurements, but reasonable estimates are available through HCM methods and/or State-specific guidelines. On freeways, vertical grade would have a similar effect (e.g., lower capacities on up-grades). Regarding acceleration, deceleration, and emissions, vertical grades may particularly affect heavy vehicles. Even for passenger cars, some vehicle-specific calibration may be warranted if the effects are not modeled automatically by the simulation tools in question.

Beyond the simulation modeling methods, super-elevations and vertical grades may be entered at the node and/or link levels. For rapidly changing vertical grades, splitting a longer link into two links may allow for improved modeling and visualization. Some simulation tools that support or allow 3D graphics (
Figure **72**) allow *z*-coordinates to be entered in addition to the typical *x*- and *y*-coordinates, for specific nodes and/or feature points.

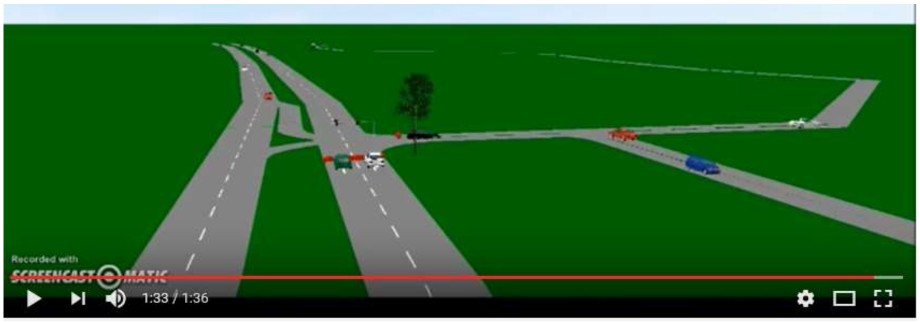


Figure 72. Screenshot. Example of 3D traffic animation.[14]

Lane-Specific Data

Different simulation tools offer varying levels of support for lane-specific data and modeling. Some are not capable of differentiating between lane-specific inputs (e.g., demands, speed limits, saturation flow rates) or outputs (e.g., queue lengths, throughput, average speed), while others require explicit data entry for each lane. Another set of simulation tools may assume equivalent input data for each lane but provide optional data entry features to specify asymmetrical lane characteristics when applicable. Some examples of lane-specific data are shown below.

Traffic volume demands. Discussion of lane-specific demand data is included later in this chapter in section 7.5.

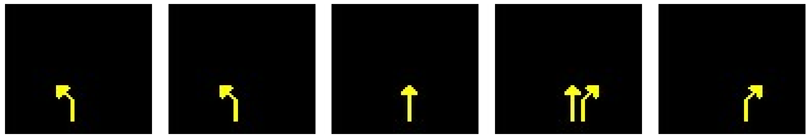
Saturation flow rates (or queue discharge headways). It is relatively rare for adjacent lanes to exhibit significantly different saturation flow and queue discharge characteristics beyond turn type (e.g., right-turning vehicles exhibit relatively large turning headways), for which queue discharge effects may be handled automatically by the simulation tool. However, special data entry may be needed in some cases. For example, the simulation user may need to indicate cases where right-shoulder parking and/or pedestrian influence causes significant saturation flow reductions in the right-most lane.

> **Channelization** (i.e., allowed turn movements from each lane). Some simulation tools may assume that turn movements are allowed from the left-most or right-most lane at an intersection or at a ramp junction. However, in many cases, the user will need to specify which turn movements are allowed from each individual lane, as shown in Source: FHWA.

Figure 73.

---

[14] ncstatefan888, VISSIM simulation (with 3D), March 25, 2015. https://www.youtube.com/watch?v=Cs0-u00nhNk. https://www.youtube.com/watch?v=Cs0-u00nhNk.

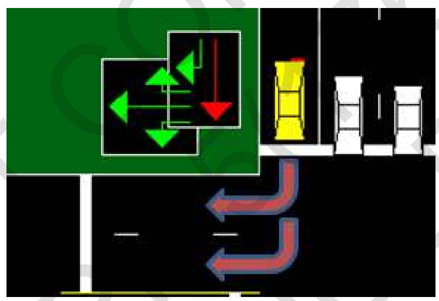


Source: FHWA.

Figure 73. Diagram. Lane channelization for a five-lane approach.

Lane alignments. Lane alignment data are primarily applicable to microsimulation, as shown in Source: FHWA.

Figure 74. If the user does not specify lane alignment data, the simulation tool may make incorrect assumptions about the downstream lanes into which vehicles from an upstream lane may enter. For example, an analyst might not align a southbound off-ramp lane (where vehicles turn left onto a surface arterial) with the left-most eastbound arterial lane, because this lane leads back to the freeway; by default, the simulation software may not make this distinction.



Source: FHWA.

Figure 74. Illustration. Lane alignments for a right-turning vehicle.

Speed limits** or **free-flow speeds. Some, but not all, simulation tools offer the ability to define unique speed limits in each lane, as shown in © Texas Transportation Institute.

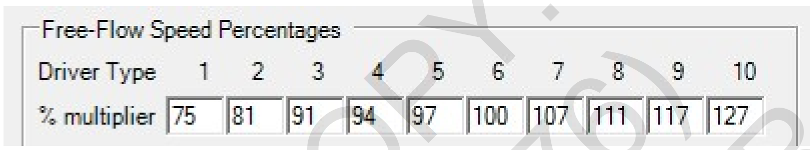
Figure 75.

© Texas Transportation Institute.

Figure 75. Photo. Dynamic lane-specific speed limits on I–5.

Other simulation tools require entry of free-flow speeds instead of speed limits. The simulation user may wish to review their tool's definitions or relationship between free-flow speeds and speed limits. For example, some tools may assume a specific ratio of free-flow speed to speed limit, unless that ratio is calibrated by the user. Finally, to reflect varying levels of aggressiveness among drivers, microsimulation tools often assume a distribution of speeds, which can be calibrated. One example of such a distribution is shown in Source: FHWA.

Figure 76.



Source: FHWA.

Figure 76. Screenshot. Different free-flow speeds applied to different driver types.

Allowed vehicle types. In some simulation tools, the user can specify a percentage of vehicle types using individual lanes. Practical applications of this may include:

- Heavy vehicles restricted from exiting at downtown area off-ramps.
- Heavy vehicles restricted from using the left-most freeway lane(s).
- Heavy vehicles biased towards using the right-most freeway lane(s).
- High occupancy vehicle (HOV) lanes.
- Toll plazas.
- Bus-only lanes.

Signal or sign control. In some tools, the sign or signal control may be specified for each individual lane. For example, during a traffic signal cycle, when a through lane is given a red ball indication, the adjacent turning lane may be given a green arrow at the same time. Similarly, the adjacent lanes of a toll plaza may require fundamentally different settings for signal and/or sign control.

7.5 TRAFFIC DATA

Demands versus Volumes

Traffic demand volumes are naturally a critical input parameter for traffic simulations. Traffic demands influence congestion levels in both the real and virtual worlds. These demand volumes are sometimes entered by the user at entry nodes, intersections, off-ramps, and mid-block source nodes. Many simulation tools also offer the ability to specify origin-destination (O-D) volumes. Although the terms demand and volume are often (and sometimes inappropriately) used interchangeably, it is critical the user enters demands instead of volumes. This is because volumes observed in the field are only reflective of vehicles discharged, whereas demands indicate how many vehicles are attempting to use the system. For undersaturated time intervals, demands and volumes are assumed equal. By properly entering demands instead of volumes for the oversaturated time intervals, excess demand vehicles will be able to enter the system at the appropriate time. When volumes are entered instead of demands, simulated networks are less likely to be oversaturated, and simulation results will tend to be overly optimistic. In fact, graphical user interfaces for some simulation tools say “volume” on data entry fields where demands should be specified.

Figure 77 illustrates an example of the common pitfall of entering volumes instead of demands. The entry node in question leads into a two-lane freeway mainline, having a capacity of approximately 2,200 vehicles per hour per lane. This means the effective 15-minute capacity is approximately 1,100 vehicles total. Figure 77 shows that a total of 5,800 vehicle drivers would like to enter the system between 4:45 and 6:00, but only a maximum of 5,400 could do so. If the user entered field-observed volumes instead of demands, the excess 400 vehicles would never be included in the computed performance measures and would never be simulated in the 6:00–7:00 time period. The result would be an overly optimistic simulation.

The pitfall of entering volumes instead of demands has affected many engineering analyses, such that simulated networks were not producing failing levels of service as observed in the real world. This also ties into the concept of trying to choose proper spatial and temporal limits for an analysis, as discussed in 0Specifically, when the number of simulated time periods is insufficient to capture the full duration of oversaturated conditions, overly optimistic simulation results may result. Regarding units of measurement, demands are often entered in either vehicles per hour (veh/h) or vehicles per time interval (veh/p). In some simulation tools, demands may be entered in veh/h even for time intervals that do not last a full hour. In this case, they reflect the hourly flow rate that exists during that time interval.

Time Period	Demand	Volume
4:45-5:00	1000 veh	1000 veh
5:00-5:15	1200 veh	1100 veh
5:15-5:30	1400 veh	1100 veh
5:30-5:45	1200 veh	1100 veh
5:45-6:00	1000 veh	1100 veh
4:45-6:00	**5800 veh**	**5400 veh**

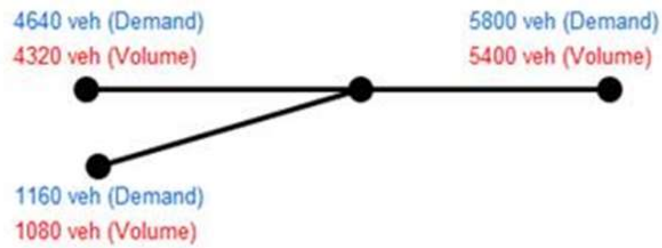


Figure 77. Illustration. Entry of demands or volumes at an entry node.

Origin-Destination Demands and Matrices

While the term demands generally refers to the number of vehicles whose drivers desire to use a specific element in the transportation network, such as a road segment, an intersection or a stretch of ramp, O-D demands denotes the number of vehicles departing from a designated traffic analysis zone (represented by an origin node) at a specific time of day and heading for another traffic analysis zone (represented by a destination node). An illustration is shown in Figure 78. It is assumed that all trips are generated from and attracted to a single node located at the centroid of a traffic analysis zone. Those nodes are defined as the origin nodes and destination nodes for the respective traffic analysis zone. The green arrows in Figure 78 indicate the direction of O-D demands between an origin node and destination node pair.

The set of origins and destinations are predetermined by modelers. They usually represent a collection of infrastructure components with ingress and egress traffic, such as aggregated residential zones and parking garages. O-D demand is often measured in vehicles per hour or the total vehicles during a time interval. Static O-D demand refers to the number of vehicles among all pairs of origins and destinations that depart from origins during a predetermined analysis period, such as 7–8 a.m. Thus, static O-D demand can be represented by a matrix. O-D demand can also be time dependent. Dynamic or time-varying O-D demand is the number of vehicles among all pairs of origins and destinations that depart from origins during a time-of-day interval, such as 7–7:15 a.m., 7:15–7:30 a.m., etc. Dynamic O-D demand can be represented by a series of demand matrices, where each matrix represents O-D demand for a time-of-day interval. It is important to note that the time dimension of O-D demand is generally defined based on departure time from origin nodes, rather than arrival time to destination nodes, or any other components along the routes of trips. Given the time-varying O-D demands matrix, the demand and volume of an infrastructure component (e.g., a road segment or an intersection) at a time-of-day interval can be estimated through traffic simulations based on route choices derived from that matrix.

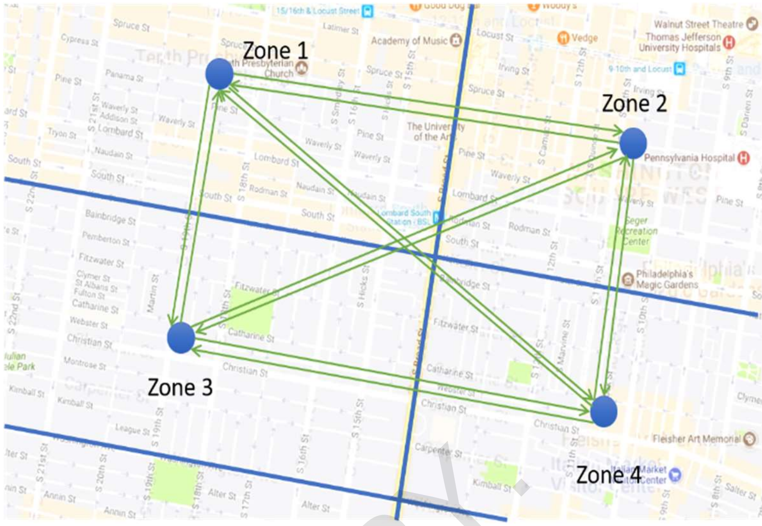


Figure 78. Illustration. Origin-destination (O-D) demand among traffic analysis zones.

In © 2019 Google® Maps. Overlay provided by Dr. Sean Qian, Carnegie Mellon University.

Figure 79 locations of origins and destinations in the Philadelphia region are plotted as an example. Each origin and destination node is assigned with a unique identification (ID).

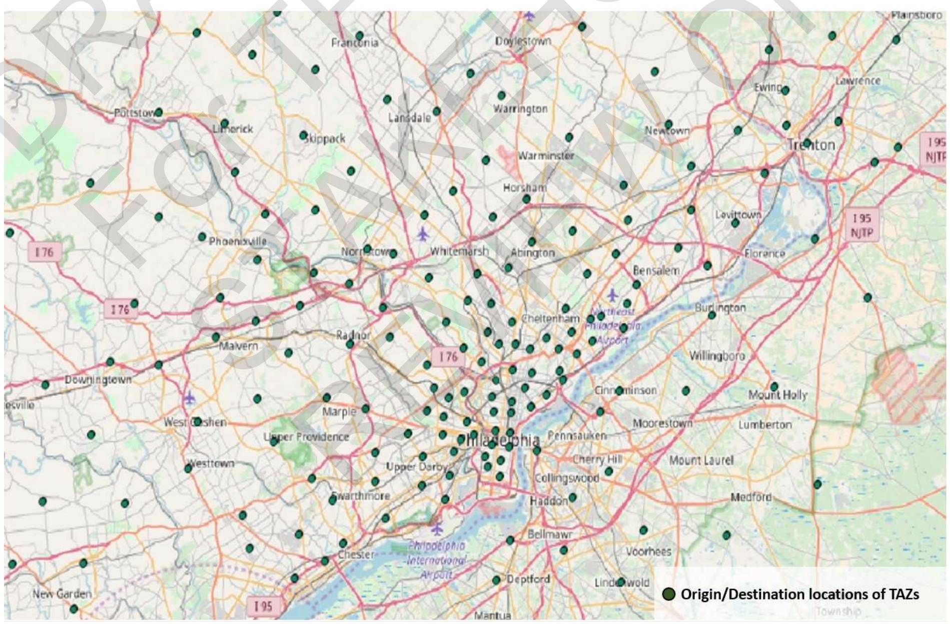
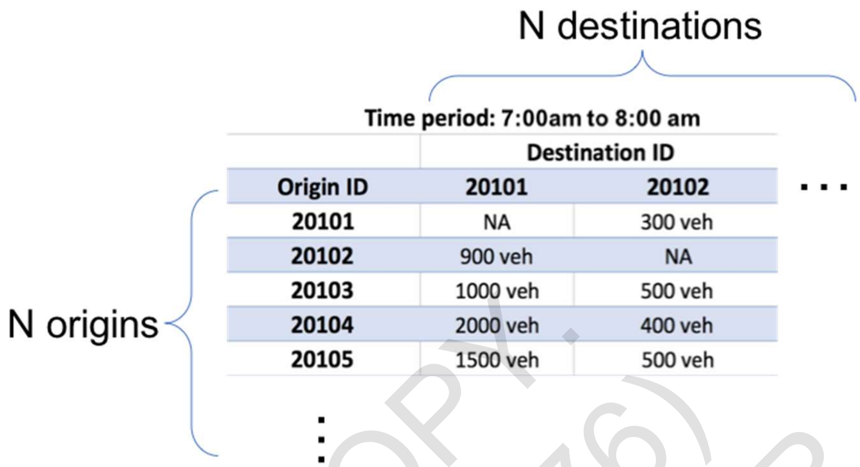


Figure 79. Map. Locations of origins and destinations in the Philadelphia region.

An example O-D demand matrix for the morning (AM) peak hour (7–8:00 a.m.) is shown in © Dr. Sean Qian, Carnegie Mellon University.

Figure 80.



	Destination ID	
Origin ID	20101	20102
20101	NA	300 veh
20102	900 veh	NA
20103	1000 veh	500 veh
20104	2000 veh	400 veh
20105	1500 veh	500 veh

© Dr. Sean Qian, Carnegie Mellon University.

Figure 80. Illustration. Origin-destination (O-D) demand matrix for morning (AM) peak hour in the Philadelphia region.

Intersection Turn Movement Demands

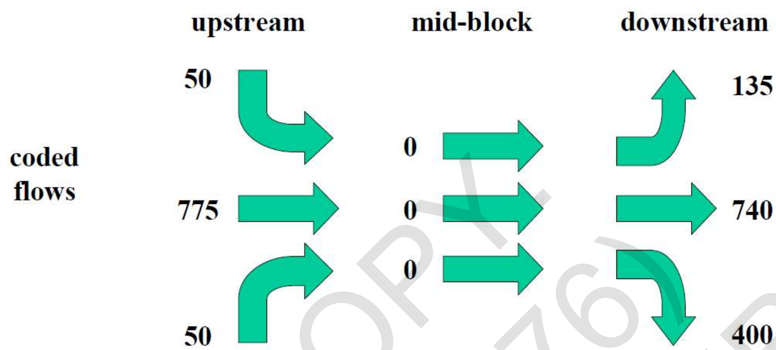
The most basic data entry for intersection turn movement demands involves entering hourly demands for three basic turn movements (left, thru, right) at an intersection with four approaches, for a total of 12 hourly demands. Beyond this, additional complexities could involve specifying demands across multiple time intervals, intersections not having four approaches, approaches at 45-degree angles, imbalanced demands at nearby intersections, local O-D data, specifying demands in vehicles per period (flows) instead of per hour (flow rates), and specifying turn movement percentages (instead of flows or flow rates).

Intersection turn movement demands are specified differently for each simulation tool. The data entry format may depend on whether the underlying tool utilizes a link-node or a link-connector architecture. Tools that implement a link-node architecture are intuitive and efficient for capturing standard conditions, but sometimes have trouble adapting to complex conditions. For example, at complex intersections, the differences between through movements, turn movements, and diagonal movements may not be clear, and may need to be defined by the user. Link-connector tools may be more adaptable to complex geometries, because the user simply specifies feeding link ID numbers (and their flow rates).

Regarding imbalanced demands, vehicles in a simulation model are not supposed to disappear. However, available volume data often do not produce consistent or logical flows in the upstream and downstream directions, regardless of whether those volumes were obtained in the field or from other

> models. Different tools may handle this dilemma in different ways. © TRANSYT-7F Users Guide.

Figure 81 illustrates the unadjusted volumes. Some tools assume the downstream volumes are accurate, and automatically adjust upstream volumes or flow profiles to become consistent. Other tools assume the opposite, and automatically adjust downstream flows. To avoid unexpected results, it is preferable to balance these demands before they are entered.



© TRANSYT-7F Users Guide.

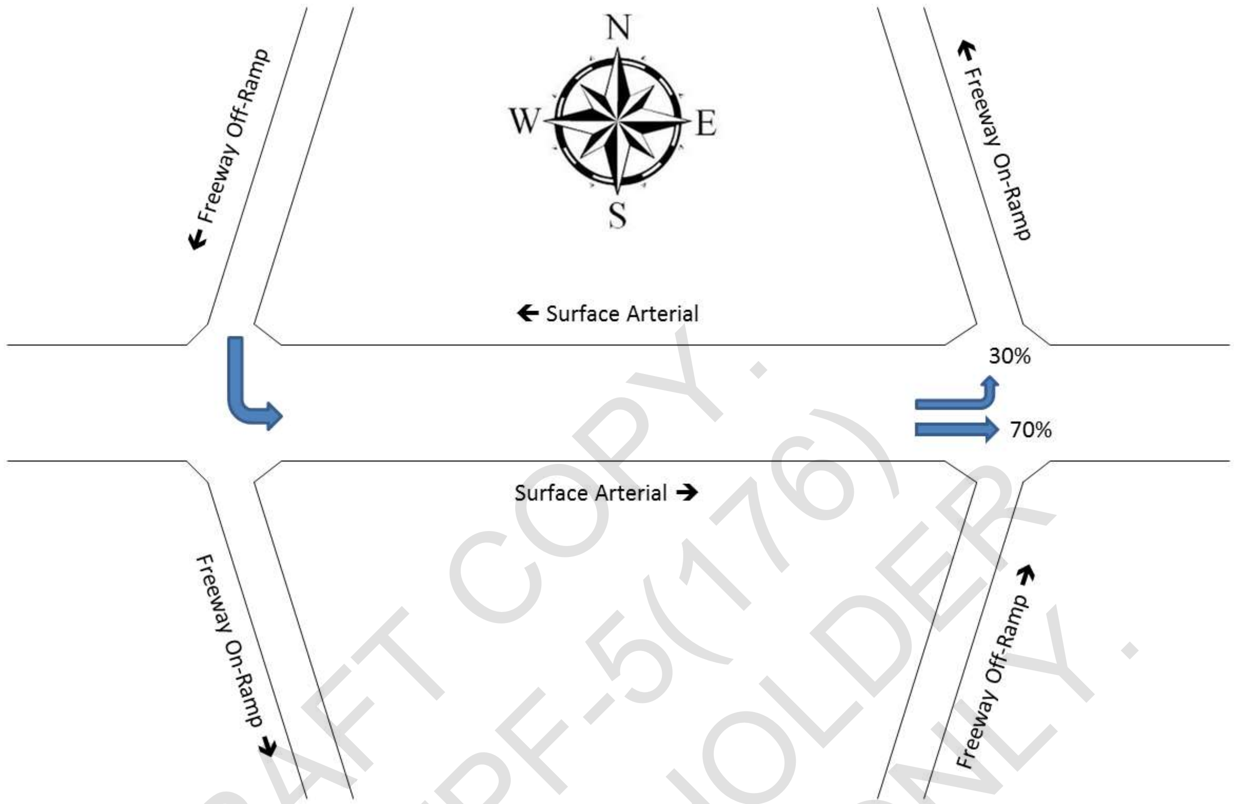
Figure 81. Diagram. The dilemma of unbalanced flows.

Because vehicles in a microsimulation model should not be created or destroyed simply due to unbalanced demands, some tools simply convert the downstream demands to turn movement percentages. For example, nearly identical results would be seen if the downstream demands in **Error! Reference source not found.** were entered as 11 percent left, 58 percent thru, and 31 percent right. However, these default turn movement percentages may be overridden by O-D demands entered at the local level (e.g., two, three, or four closely-spaced intersections) or the networkwide level. O-D demands at the local level are sometimes called conditional turn movements, and these are discussed in the next subsection.

Conditional Turn Movements

> O-D demands at the local level are sometimes called conditional turn movements. These optional input specifications, which are primarily applicable to microscopic traffic simulation, can be used to make the simulation more realistic. Source: FHWA.

Figure 82 illustrates a typical condition that could be appropriately addressed by adding conditional turn movements. At the intersection to the right, 30 percent of all eastbound vehicles are making a left turn to enter the freeway. However, the vast majority of upstream drivers who just exited the freeway do not wish to immediately reenter the freeway. Therefore, conditional turn movements can be used to indicate that, for example, only 1 percent of all upstream, southbound left-turning vehicles wish to make an eastbound left-turn maneuver at the downstream intersection. The simulation tool must then ensure that more than 30 percent of eastbound through vehicles at the upstream intersection make a subsequent eastbound left turn at the downstream intersection, so the overall 30/70 turn movement split is preserved at the downstream eastbound approach. Alternatively, the user may specify exact conditional turn movement percentages for all upstream and downstream turn movements, but this strategy is more time consuming and prone to data entry errors.



Source: FHWA.

Figure 82. Diagram. Example condition to be modeled by conditional turn movements.

Thus, common applications of conditional turn movements include:

- Preventing vehicles that just exited the freeway from immediately reentering the freeway.
- Preventing vehicles that just entered the freeway from immediately exiting the freeway.
- Preventing vehicles that just turned on to the surface arterial major street (from a side street) from immediately turning off at the next downstream minor street.
- Lane utilization and pre-positioning (see next subsection).

Lane Utilization and Pre-Positioning

Lane utilization expresses the proportion of vehicles using each lane. On multilane freeways and surface arterials, simulation tools usually assume equal or slightly unequal lane utilization unless the user indicates otherwise, or unless vehicles unequally distribute themselves to achieve user-specified downstream turn movement percentages. In macroscopic or mesoscopic simulation tools, unequal lane utilizations may be calibrated using inputs such as the lane utilization adjustment factor or single highest lane volume, and pre-positioning may not be relevant.

Regarding microscopic simulation, the method for achieving unequal lane utilizations may depend on whether the tool in question follows a link-node or a link-connector architecture. In microscopic simulation tools with a link-connector architecture, each individual lane could be classified as its own link, which would explicitly allow the user to specify a unique demand volume for each lane. Microscopic simulation tools with a link-node architecture may allow explicit specification of the proportion of vehicles in each lane or the proportion of vehicles for each turn movement. If the tool is limited to only specifying the proportion of vehicles for each turn movement, it may become necessary to define a duplicate or redundant turn movement (e.g., a 45-degree left-diagonal movement that feeds the same downstream node as a left-turn movement), thus allowing a percentage of vehicles to be restricted to only lanes channelized for that duplicate or redundant turn movement. In the Figure 83 example shown below, eastbound vehicles at the right-most intersection could be proportionally assigned to a left-turn lane and a left-diagonal lane, even though both lanes ultimately make the same left turn.

Figure 83. Illustration. Example of lane utilization and pre-positioning.

Pre-positioning behaviors could require additional data entry related to conditional turn movements and lane channelization. For example, suppose eastbound vehicles in Figure 83 have a tendency to move into the two left-most lanes far in advance of the freeway on-ramp on the far right. The user would then use conditional turn movement input data to indicate that nearly 100 percent of left-turn or left-most vehicles at the far-left intersection would proceed to remain left-turn or left-most vehicles at the middle intersection, and that nearly 100 percent of left-turn or left-most vehicles at the middle intersection would proceed to make a left turn to enter the freeway at the right-most intersection. In conjunction with this, it is likely that nearly 100 percent of right-turn or right-most vehicles would be conditioned to move straight through at the downstream intersection(s). Finally, at locations where vehicles exiting the freeway rarely reenter the freeway immediately, the left-most lanes on the southbound approach at the far-left intersection could be exclusively aligned with the right-most lanes on the eastbound surface arterial.

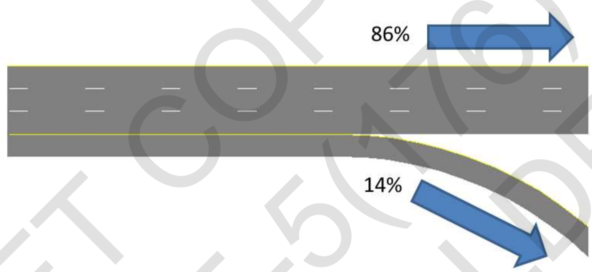
Off-Ramp Exit Demands

Users of simulation must typically specify either the proportion (e.g., as shown in Source: FHWA.

Figure **84**) or the explicit number of vehicles that will exit at each off-ramp. In simulations involving traffic assignment, the proportion of vehicles exiting at each off-ramp may be determined by the traffic assignment procedure.

Complications may arise if:

- Different vehicle types require different off-ramp exiting percentages.
- Freeway congestion prevents vehicles from reaching the off-ramp, such that simulated off-ramp flows differ from user-entered off-ramp flows. (Simulated off-ramp flows differ from user-entered off-ramp flows for unknown reasons.)
- An imbalance exists between upstream and downstream flows.



Source: FHWA.

Figure 84. Diagram. Off-Ramp Exit Volume Percentage.

In cases where different vehicle types require different off-ramp exiting percentages, some simulation tools allow the user to specify unique off-ramp exiting percentages for each vehicle type. Perhaps heavy vehicles rarely use off-ramps in a specific large city center. Passenger cars and sport utility vehicles (SUV) may be prohibited from exiting at way stations. In cases where simulated off-ramp flows differ from user-entered off-ramp flows, a diagnostic or calibration effort may be needed. Perhaps drivers tend to behave more aggressively at such off-ramps in the real world; this is something that can be achieved and calibrated through various input parameters. Or, perhaps the simulated congestion preventing vehicles from reaching the off-ramp is unrealistic, and calibration is needed to reduce the unrealistic congestion. In other cases, origin-destination flows may need to be defined or revised, to override the default assignment of off-ramp vehicles. Or, discrepancies between upstream and downstream flows may need to be reconciled (e.g., © Wisconsin Department of Transportation.

Figure 85); this reconciliation could achieve proper flow balancing, conservation of flow, and simulated off-ramp volumes.

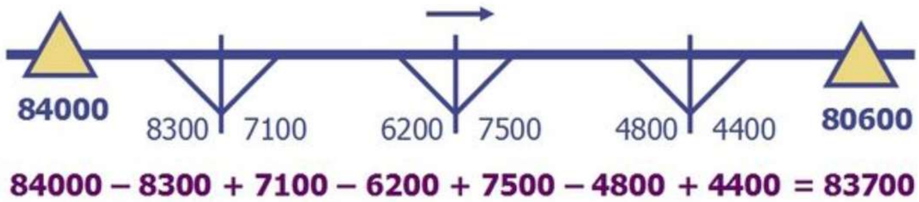


Figure 85. Diagram. Imbalanced upstream and downstream volume counts.

Vehicle Fleet Composition

A traffic simulation's accuracy may be aided by realistic specification of vehicle fleet composition and various vehicle type characteristics. Vehicle type characteristics may include length, occupancy, moving headway, queued headway, acceleration/deceleration capability, fuel consumption, emissions, free-flow speed, and managed lane transponders. Vehicle type classifications may include passenger car, truck, bus, carpool, emergency vehicle, and advanced technology vehicle. Clearly, some of these classifications may overlap, such as a certain percentage of passenger cars that are carpools or a certain percentage of trucks that are emergency vehicles.

Some simulations may assume uniform vehicle fleet compositions throughout the network. Other simulations may support different vehicle fleet compositions within subnetworks (e.g., 64-foot-long cargo trucks may rarely use a certain congested signalized arterial). The user may be allowed to specify unique vehicle type characteristics (e.g., percentage of trucks, carpools, and HOV vehicles) at each entry node surrounding the traffic network. Some example vehicle fleet settings are shown in Source: FHWA.

Figure 86.

Surface	25%	45%	25%	5%	0%
Freeway	10%	45%	35%	5%	5%
Headway	Normal	Normal	Normal	Large	Large
Length	14 feet	18 feet	22 feet	35 feet	64 feet

Source: FHWA.

Figure 86. Illustration. Example of vehicle fleet composition and settings.

The number of queued vehicles that may fit between a pair of traffic signals (or between closely spaced interchanges) will depend on the vehicle lengths. Thus, the calibration of vehicle lengths may affect the degree of queue spillback within a congested network. Heavy vehicles may accelerate more slowly on roadway segments with a steep incline and may generate increased pollutant emissions relative to the passenger cars. At toll plazas, certain vehicle types may be restricted to certain lanes.

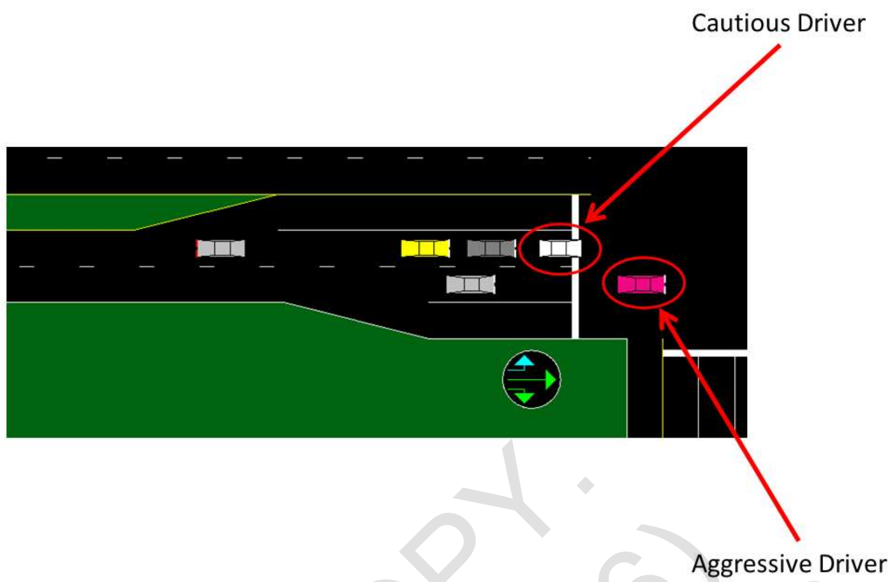
When buses are simulated, it may be necessary to define bus routes, bus stop locations, and time headways between buses. Bus station characteristics may include average dwell times, dwell time distributions, and bypass percentages.

Some simulation tools allow the user to specify unique O-D flow rates and/or percentages for each vehicle type. This amount of data entry may be more time consuming, but may allow for more accurate simulation outputs in traffic networks where different vehicle types are known to choose significantly different paths through the network.

Driver Characteristics

> The different characteristics of human drivers are on full display during the animations of most microscopic simulations, and to a lesser extent within the mathematical underpinnings of mesoscopic and macroscopic simulations. In microsimulation, one example of a driver characteristic is the desired free-flow speed. Some drivers have higher desired speeds than others. Some simulation tools allow the user to calibrate the distribution of desired speeds. In mesoscopic and macroscopic simulations, the discrepancy in desired speeds produces a certain degree of platoon dispersion, which may also be subject to calibration. Another example of a driver characteristic could be the average start-up lost time, which occurs when a signal turns from red to green. Source: FHWA.

Figure 87 illustrates an example where the aggressive driver has a short start-up lost time (e.g., 1.4 seconds), whereas the adjacent cautious driver has a longer start-up lost time (e.g., 2.8 seconds).



Source: FHWA.

Figure 87. Illustration. Aggressive driver and cautious driver reacting to a green light.

In some microscopic simulation tools, the level of driver aggressiveness may be expressed as a driver type. Some simulation tools allow the user to calibrate the percentage of aggressive and cautious drivers in the traffic network, as well as the numeric settings associated with each driver type. The following is a sample list of driver-based characteristics:

- Free-flow speed.
- Start-up lost time.
- Desired headway between vehicles.
- Time required to make a lane change.
- Distance required to make a lane change.
- Acceptable gaps for permissive left turns and right turns on red.
- Willingness to cooperate with other drivers (e.g., when blocking an intersection).

Headway and Saturation Flow Rate

On both freeways and surface arterials, the efficiency of traffic flow is highly sensitive to headways between vehicles. Drivers maintain a minimum headway from the immediate downstream vehicle while motionless in a standing queue (i.e., jam spatial headway), and typically a larger minimum headway while moving (i.e., optimal spatial headway). Larger headways tend to occur around heavy vehicles, as shown in **Error! Reference source not found.**. Microscopic simulation tools may handle such headways by automatically applying a default headway multiplier (e.g., 120 percent) to heavy vehicles. Macroscopic and mesoscopic simulation tools may handle such headways by automatically adjusting link capacities and saturation flow rates as a function of heavy vehicle percentages. Regardless of the modeling methodology, vehicle headways and saturation flow rates should be calibrated to local conditions wherever and whenever possible to ensure a realistic simulation.

Source: FHWA.

Figure 88. Photo. Example of larger headways for heavy vehicles.

One example of vehicle headway calibration would involve closely spaced interchanges on a freeway. During the peak commuting periods, drivers may follow more closely and aggressively in such areas. Thus, a microscopic simulation tool may allow link-specific and time interval-specific headway multipliers. Saturation flow rates are associated with signalized and un-signalized intersection approaches. For example, a saturation flow rate of 1,895 veh/ln/hr might result from an average of 1.9 seconds of headway space between discharging vehicles. Depending on the tool, the engineer may be expected to enter saturation flow rates or queue discharge headways as inputs to the simulation model. Ideal (theoretical) saturation flow rates may be reduced by friction factors such as narrow lane widths, uphill grades, heavy vehicles, conflicting pedestrians, and buses, as described in the HCM.

7.6 CONTROL DATA

Pre-Timed Traffic Signal Control

Most simulation tools at the microscopic, mesoscopic, and macroscopic levels provide data entry and modeling support for pre-timed traffic signals, which are also known as fixed-time traffic signals. Pre-timed signals provide a guaranteed, fixed-cycle length for as long as the timing plan remains in effect. Individual phase durations also remain fixed from cycle to cycle. Many pre-timed signals are timed by time of day (TOD). This means, for

example, that most green time in the morning could be given to movements going in the direction of downtown, whereas most green time in the evening could be given to movements leaving the direction of downtown. Therefore, some simulation tools provide support for entering multiple pre-timed signal timing plans, and the user can specify exactly when they take effect. There may also be a transition period between different timing plans, and this period may be handled differently by different simulation tools.
Figure **98** shows one example of a data entry form for pre-timed signals.

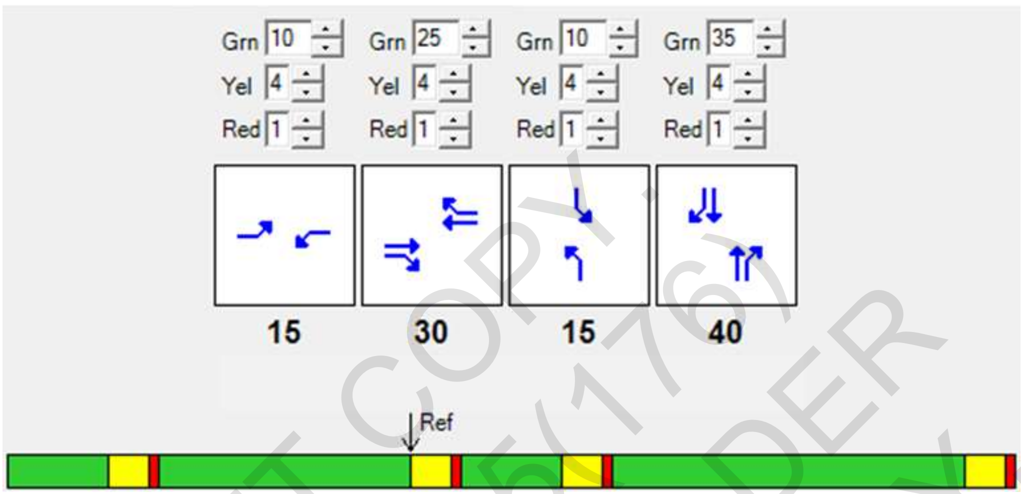


Figure 89. Screenshot. Sample data entry form for a pre-timed traffic signal.

Some additional considerations for pre-timed signals are as follows:

- When modeling such intersections, the simulation user may consider related parameters such as the offset, reference interval, cycle length, start-up lost time, extension of effective green time, and amber response. However, not all simulation tools support all of these settings and parameters.
- Some simulation tools support the copying and pasting of a signal timing plan from one intersection to another.
- For the simulation tools that support signal timing optimization, there may be additional input parameters associated with optimization (e.g., minimum and maximum cycle length, minimum phase time).
- In some simulation tools, users may define an actuated traffic signal, and then specify max recall on each phase (to emulate pre-timed operation).
- In some simulation tools, the user may define pedestrian demands and/or pedestrian signal timings (e.g., walk, flashing don't walk).

Actuated Traffic Signal Control

For macroscopic and mesoscopic simulation tools, the modeling and data entry for actuated traffic signals and pre-timed signals may be somewhat similar. The user may need to indicate whether the actuated signal is coordinated or uncoordinated. Generally, if the signal is uncoordinated, each signal phase is an actuated phase whose duration may fluctuate between a minimum green and maximum green value. If the signal is coordinated, the user may need to specify the background cycle length, the offset (or yield point), and the major street direction (i.e., north-south major street versus east-west major street). For microscopic simulation, the modeling and data entry for traffic actuated signals may be substantially different. The phasing sequence is typically specified in an 8-phase dual ring format (e.g., Figure 90). The number of input parameters is generally much larger, considering the controller parameters (e.g., Figure 91), detector settings (see next page), and sometimes coordination settings (e.g., Figure 92) that must be entered for each actuated signal. Thus, the additional input data needed for microsimulation of actuated signals requires more time and expertise from the end user. Data entry errors and phase duration errors are more likely for actuated signals than for pre-timed signals. Pre-processors may simplify the data entry process, but once the model is up and running, it is preferable to replace the software default values with collected, real-world parameter values. Ultimately, the degree of realism provided by microsimulation of actuated signals is often very beneficial to engineering studies.

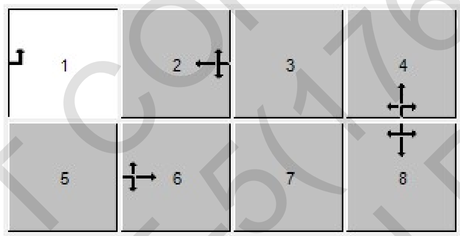


Figure 90. Screenshot. Sample data entry form for traffic actuated phasing sequence.

	Mingreen	Maxgrn	Yellow	All Red	Walk	Ped Clr	Split %	Split	Passage	Max Ext
1	3	10	3.0	2.0	0	0	16.7	15	2.0	0
2	3	35	3.0	2.0	0	0	33.3	30	2.0	0
3	0	0	0.0	0.0	0	0	0.0	0	0.0	0
4	10	30	3.0	2.0	0	0	50.0	45	3.0	0
5	0	0	0.0	0.0	0	0	0.0	0	0.0	0
6	3	35	3.0	2.0	0	0	50.0	45	2.0	0
7	0	0	0.0	0.0	0	0	0.0	0	0.0	0
8	10	30	3.0	2.0	0	0	50.0	45	2.0	0

Figure 91. Screenshot. Sample data entry form for traffic actuated controller parameters.

	Period 1	Period 2	Period 3
Start Time (sec)	0	25	0
End Time (sec)	30	40	0
Service Phase 1	☐	☑	☐
Service Phase 3	☐	☐	☐
Service Phase 4	☑	☐	☐
Service Phase 5	☐	☐	☐
Service Phase 7	☐	☐	☐
Service Phase 8	☑	☐	☐

Figure 92. Screenshot. Sample data entry form for traffic actuated controller permissive periods.

Actuated Traffic Signal Detectors

As alluded to earlier, the ability of microsimulation tools to explicitly emulate 1) individual vehicles rolling over specific detectors in the roadway and 2) immediate traffic signal timing responses to those actuations provides a crucial amount of modeling realism. Having said this, macroscopic and mesoscopic simulation tools can often produce reasonable estimates of actuated signal phase durations. The definition of traffic signal detectors for actuated signals can play a significant role in the amount of modeling realism that is possible. In traffic simulation, the terms detector and sensor are often used interchangeably.

Basic actuated control usually involves presence detectors deployed near the stop bars of a signal-controlled intersection. It is common for such detectors to have detector lengths of 30 feet or 40 feet to ensure that a vehicle near the stop bar is detected. For coordinated semi-actuated signals, the major-street through lanes may not have detectors and may not properly function if detectors are placed there. Pulse or passage detectors tend to be smaller in length (e.g., 6 feet) and may be used for more advanced control functions.

Detector setback indicates the distance between the front of the detector and the stop line. The setback distance may be zero for detectors deployed at the stop line. In one sample real-world application, detectors are placed near the upstream end of a left-turn pocket, and the exclusive left-turn phase is only called if the turn pocket queue stretches back to the sensor. In such a case, the detector setback might be nearly equal to the turn pocket length. In addition, the agency might specify a non-zero delay time, requiring the sensor to be occupied for the specified consecutive number of seconds before the associated phase is called. Thus, with a non-zero delay time, a vehicle passing over the detector at typical speeds would not call the phase, but a left-turn queue backing up to the detector would call the phase after a few seconds of detector occupancy.

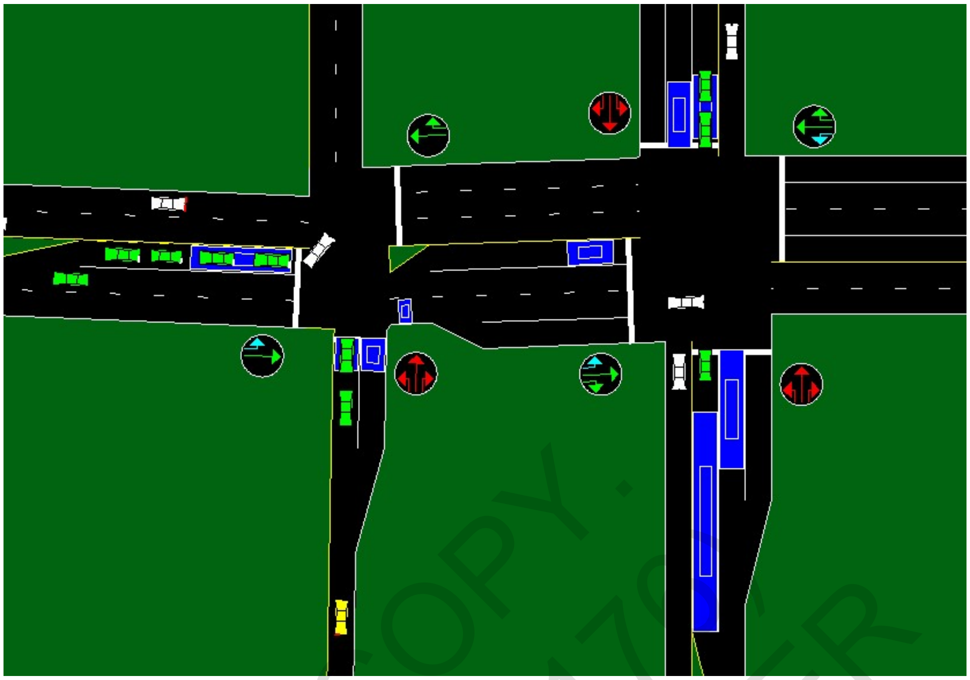


Figure 93. Screenshot. Examples of varying detector lengths, setbacks, and lane numbers.

Ramp Meters

Many simulation tools support the modeling and analysis of ramp meters. Ramp meters are traffic signals typically installed near the on-ramp merge point. These signals meter incoming on-ramp to mitigate congestion problems on the mainline freeway, as illustrated in Figure 94.

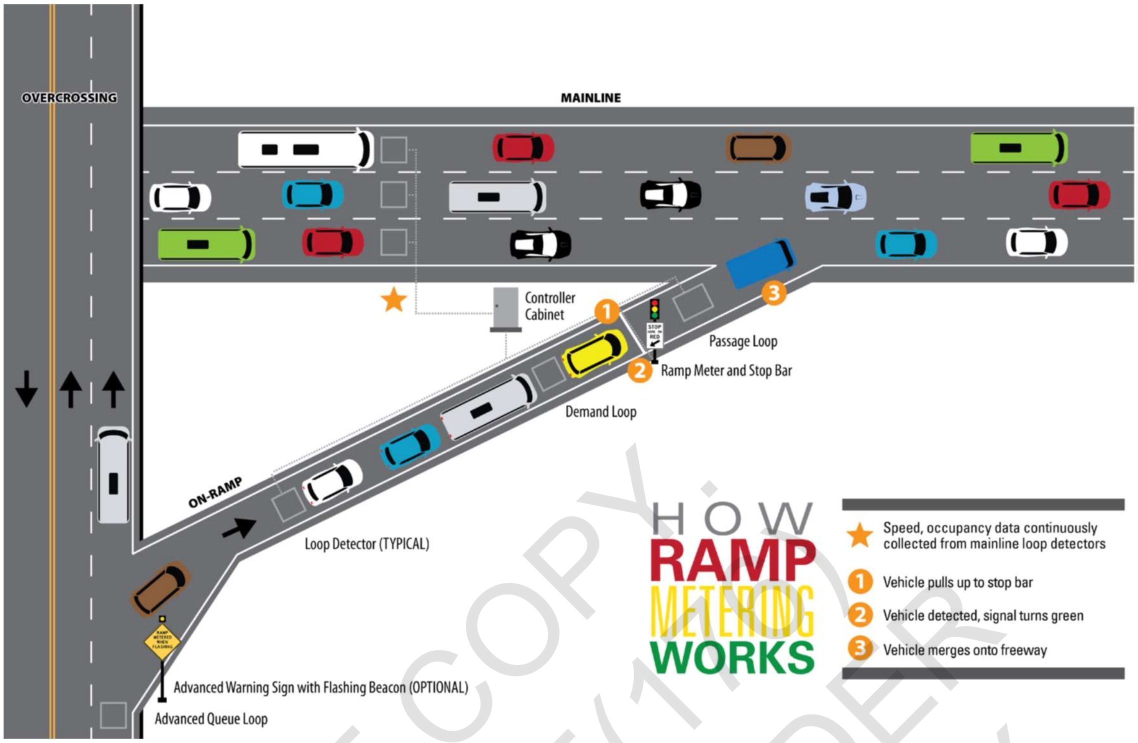


Figure 94. Diagram. How ramp metering works.[15]

Fixed-time meters may require data entry for the number of vehicles per green that can enter the mainline or for the duration of green and red intervals. Traffic responsive meters require some sort of detection such that ramp meter timings can adapt to freeway congestion levels. Ramp meter detectors are discussed in the next subsection. Traffic simulation users are often allowed to choose among available ramp metering algorithms and algorithm parameters. Some simulation products offer run-time extensions or application programming interfaces (APIs) that can customize the operation of various ramp metering algorithms.

Ramp meters are intended to either delay the onset of freeway bottlenecks or to prevent them from forming altogether. At known freeway bottleneck locations, ramp metering effects can be visualized by the Rice Experiment.[16,17,18,19] When rice is poured too quickly into the funnel, a

---

[15] United States Department of Transportation, "About Ramp Metering," accessed October 10, 2019, https://ops.fhwa.dot.gov/freewaymgmt/ramp_metering/about.htm.

[16] United States Department of Transportation, "Recurring Traffic Bottlenecks: A Primer," accessed October 10, 2019, https://ops.fhwa.dot.gov/publications/fhwahop18013/index.htm.

[17] Wisconsin Department of Transportation, "Doug MacDonald - Rice and Traffic Congestion," April 20, 2007, YouTube video, 6:19, https://www.youtube.com/watch?v=8G7ViTTuwno.

[18] The Seattle Times, "Rice is nice when trying to visualize highway traffic," December 29, 2006, retrieved from https://www.seattletimes.com/seattle-news/rice-is-nice-when-trying-to-visualize-highway-traffic/.

[19] Washington State Department of Transportation, "The PRICE Is Right!," retrieved from http://www.wsdot.wa.gov/NR/rdonlyres/4D0BFFF3-E59B-45EE-ABB4-1F00F762C5DE/0/MacDonaldRiceHandout.pdf.

bottleneck quickly forms, which prevents the rice from being discharged in a timely manner. When rice is poured too slowly, this also prevents the rice from being discharged in a timely manner. Only by pouring the rice at an optimized medium rate can throughput and speed be maximized.

Ramp Meter Detectors

Ramp meter detectors function similarly to actuated signal detectors (i.e., timings respond in real-time to detected vehicles). Simulation users can sometimes choose among different detector types, as described in the next subsection. Traditional deployments use feed-forward logic in an open loop control system. Downstream conditions are not monitored and cannot provide feedback. Downstream problems may not be detected until congestion reaches the detector upstream of the ramp. New generation ramp meters have facilitywide detectors, with timings coordinated in a similar fashion to the signals along surface arterials. Figure 95 illustrates feedback logic in a closed loop control system. Traffic conditions are measured at downstream bottlenecks to determine critical occupancy and how much traffic can enter upstream.

Downstream flow conditions are known in real time, providing feedback to determine ramp flow (q_r) and optimal occupancy (q_{capN}). When applied on a holistic, coordinated, systemwide basis, all ramps continuously communicate with each other to resolve complex traffic flow situations.

© WSP, 2017 IBTTA TRB Symposium.

Figure 95. Diagram. Coordinated ramp metering detection.

Freeway Surveillance Detectors

In a freeway simulation, typical performance measures include average density, speed, and number of vehicles discharged. These measures are usually provided regardless of whether the simulation is microscopic, mesoscopic, or macroscopic. Freeway surveillance detectors and data stations are used (typically in microscopic simulation) to obtain extra measures that would not otherwise be available.

Figure 96 illustrates a freeway detector in the middle lane of a three-lane mainline. Typical input parameters associated with the data station or surveillance detector include lane number, *x*-*y*-*z* coordinate location, distance to the upstream or downstream node, detector length, and detector type. Sample detector types include single loop, double loop, and Doppler radar.

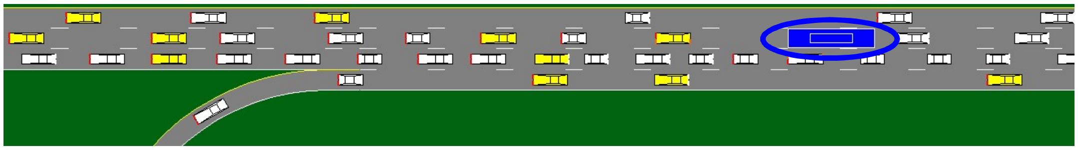


Figure 96. Freeway surveillance detector (circled).

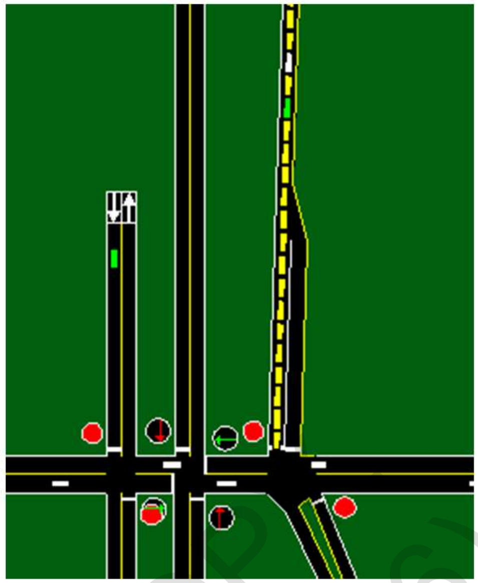
Queue length is one example of a performance measure that would be easily obtained at intersections along surface arterials but would be difficult to define or measure on the freeway without detectors or data stations. Although queuing is common on congested freeways, it is difficult to predict where (spatially) a queue will begin. It is also impractical to install detectors in all locations where a queue might begin. Instead, the data stations and/or surveillance detectors can offer performance measures that provide evidence of where a queue exists. These performance measures include percent of traffic at or below a given headway, percent of traffic at or below a given speed, lane-specific flow rate, mean occupancy rate, and mean headway.

Sign Controlled Intersections

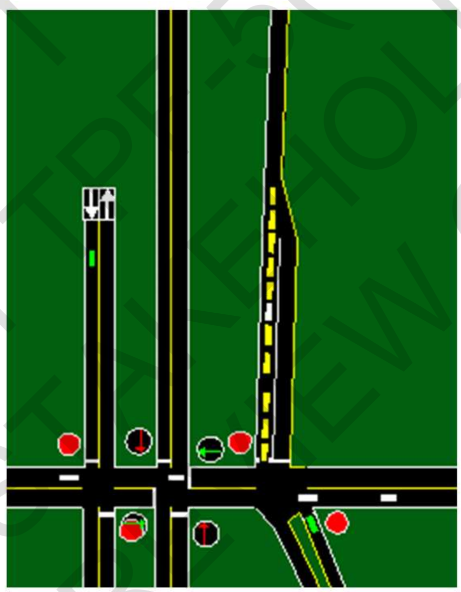
Defining a stop-sign controlled intersection approach or a yield-sign controlled intersection approach is generally easy in a simulation tool. However, calibration is often needed to achieve realistic queue lengths and delay times. The input parameters to be calibrated will depend on whether the tool is a microscopic, mesoscopic, or macroscopic simulation tool. For microscopic simulation analysis, calibration parameters may include gap acceptance times, follow-up times, and vehicle entry headway distributions. For mesoscopic and macroscopic simulation analysis, calibration parameters may include gap acceptance times, follow-up times, and maximum flow rates.

B. Subfigure of calibrated sign-controlled intersection.

Figure 97 illustrates the difference between a calibrated and an uncalibrated sign-controlled intersection in a microsimulation tool.



A. Subfigure of uncalibrated sign-controlled intersection.



B. Subfigure of calibrated sign-controlled intersection.

Figure 97. Diagrams. Comparison of calibrated and uncalibrated sign-controlled intersections.

Figure 98 shows an example of default gap acceptance values for various levels of driver aggressiveness. Macroscopic calibration parameters and outcomes are discussed next.

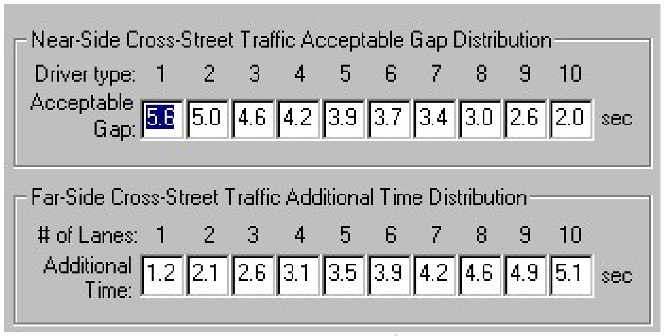
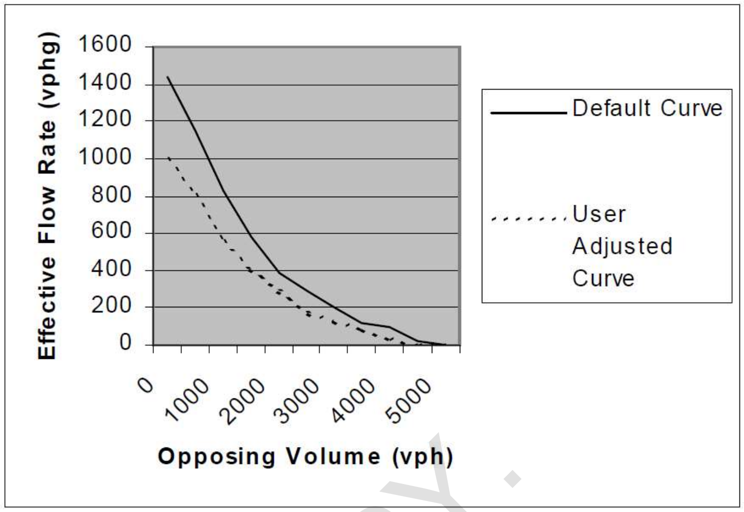


Figure 98. Screenshot. Sign-controlled intersection gap acceptance times subject to calibration.

For macroscopic or mesoscopic simulation tools, the maximum flow rate (*MFR*) equation shown below is an example of a gap acceptance model that can be calibrated for sign-controlled intersection operations in local areas. The variable A_i indicates a movement's flow rate when no opposing vehicles are present. Variable $v_o(t)$ indicates the opposing flow rate at time t. Variables B and C are model parameters automatically defaulted by the macroscopic or mesoscopic tool to simulate a stop-controlled approach, a yield-controlled approach, or a permissive left turn at a signalized intersection. Figure 99 provides an example of what might happen to effective flow rates at an intersection when the *MFR* model coefficients are calibrated.

$$MFR(t)_i = A_i \bullet \exp\left(-B_i \bullet v_o(t)^{C_i}\right) \tag{7.1}$$



© TRANSYT-7F Users Guide.

Figure 99. Diagram. Flow versus opposing volume for macro and mesoscopic models.

Roundabout Intersections

Data entry for roundabout intersections like the one shown in Figure 100 may be extensive in micro-simulation tools that treat each approach to the roundabout as its own intersection.

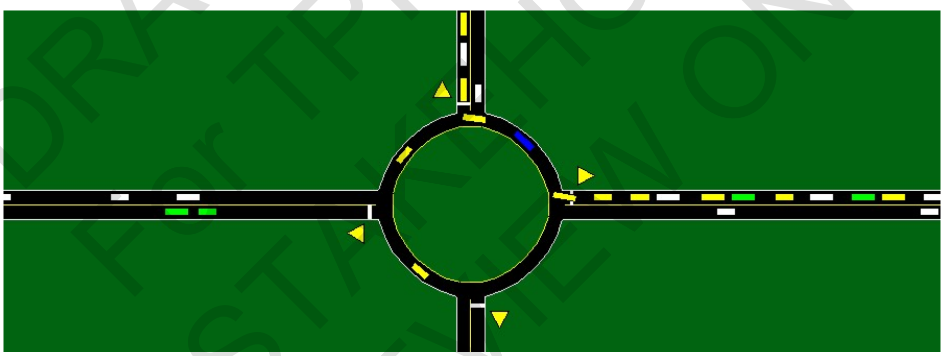


Figure 100. Illustration. Microsimulation of a roundabout intersection.

As a result of this modeling design, O-D volume data (or at least conditional turn movement flows) are needed to achieve realistic traffic patterns. In such cases, it may be beneficial to start with a template roundabout intersection coded in the past, and simply customize the parameters to local conditions.

Freeway Warning Signs and Toll Plazas

Nearly all simulation of urban freeways will require simulation of off-ramps. In microsimulation tools, off-ramp reaction distance or warning sign distance may be entered for each off-ramp, so drivers will know when to move into their goal lane for exiting the freeway. This is analogous to exit-only signs, but in some cases a calibrated warning sign distance will reflect driver behavior more so than physical sign location. In other words, the simulation should reflect the real-world location where drivers typically change lanes, even if that does not correspond with the location of the physical warning signs.

Figure 101 illustrates an off-ramp warning sign location.

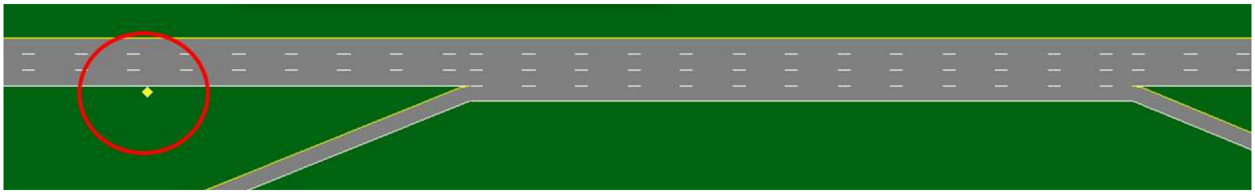


Figure 101. Illustration. Off-ramp warning sign.

Many freeway simulations will also require modeling of toll plazas. Often, there will be simplifying assumptions, because toll plaza modeling can potentially be very complex. Some tools offer toll plazas as an explicit object, making the setup much easier. Certain vehicle types may be prohibited from using certain lanes. Free-flow speeds, posted speed limits, and tollbooth service times may differ on a lane-by-lane basis. Lane utilization may vary significantly. Some parameters and conditions will probably vary across time intervals. In simulations involving traffic assignment, these settings may affect freeway demands and capacities. One example of simplifying assumptions is shown in

Figure 102, where toll booths are modeled using yield signs and stop signs. This operation can be calibrated to some extent through start-up lost times and follow-up times.

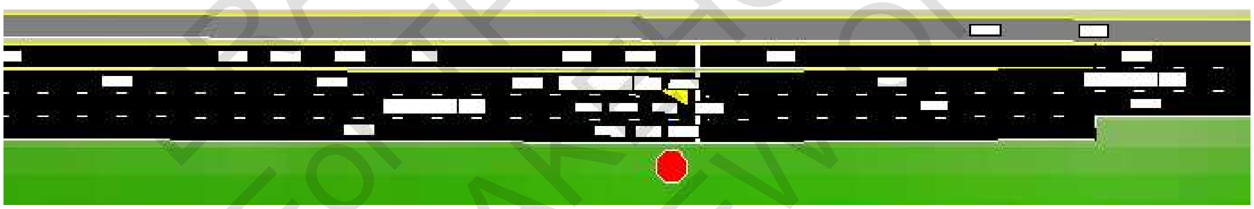


Figure 102. Illustration. Implicit microsimulation of a toll plaza.

By contrast,

Figure 103 illustrates a more explicit form of toll plaza modeling. In this simulation, the user may indicate lane-specific toll collection methods, service times, and vehicle utilization for each time interval. Additional calibration settings may include upstream reaction distance, lane change sensitivity, and setback distance for the second vehicle in queue.

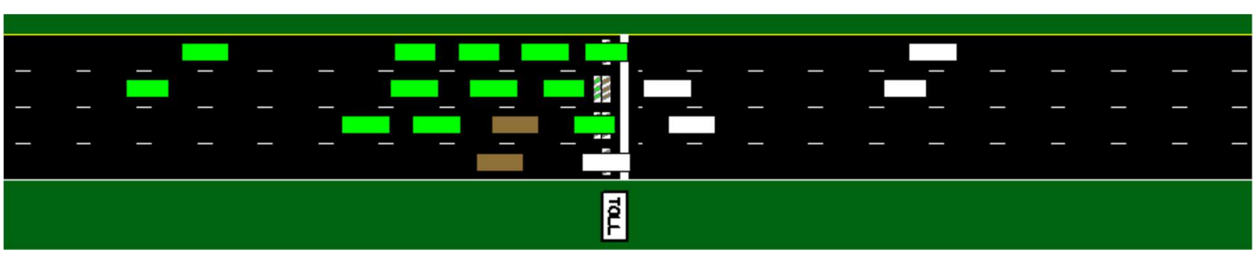


Figure 103. Illustration. Explicit microsimulation of a toll plaza.

Managed Lanes

According to FHWA, managed lanes are defined as highway facilities or a set of lanes where operational strategies are proactively implemented and managed in response to changing conditions. On HOV and high occupancy toll (HOT) lanes, drivers usually expect an improved level of service (LOS) and more efficient flow. Simple examples of managed lane microsimulation are shown in
Figure 104 and
Figure 105.

Figure 104. Illustration. Two-dimensional freeway with a left-side managed lane.

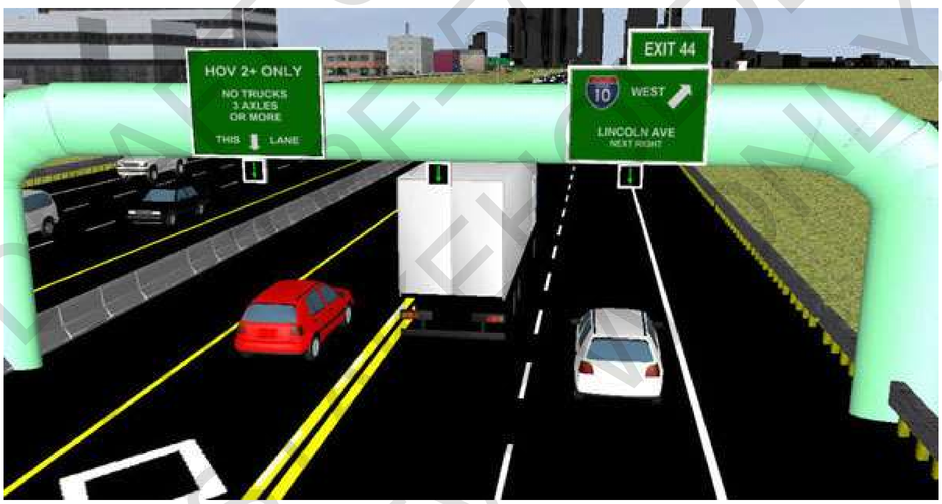


Figure 105. Illustration. Three-dimensional freeway with a managed lane.

Although managed lanes can be simulated at all resolutions (macro, meso, micro), explicit simulation of vehicle movements in and out of managed lane access points can only be handled by microsimulation, regardless of whether a managed lane physically provides exclusive or nonexclusive access. Similar to freeway off-ramps, HOV/HOT lanes having exclusive access may have reaction distance as a calibration parameter within microsimulation tools. Because macroscopic and mesoscopic simulation tools do not simulate lane changing in as much detail, they may provide simple capacity adjustments and speed adjustments for managed lanes. In traffic assignment simulations, these capacity adjustments may affect managed lane demand volumes.

Input data related to vehicle types will have a significant impact on microsimulation outcomes. Vehicle occupancies, percentage of HOV vehicles using the HOV lane, percentage of vehicles having transponders, and toll-paying vehicle types are all possible inputs. Other managed lane input data unrelated to vehicle types may include pricing algorithm parameters and the user-perceived value of time.

7.7 LOAD CONDITION DATA

Demand Variability

Traditional simulation projects have focused on analyzing typical peak hour traffic conditions. The danger of this approach is that when decisions are made based on most likely (i.e., 50th percentile) outcomes, these decisions may fail badly when the inevitable 75th–95th percentile conditions materialize. Some traditional projects have focused on analyzing the 30th highest hour demand volume scenario. While this is probably a step in the right direction, the analysis of only a single scenario still leaves engineers largely in the dark about how to manage risk and reliability, because they cannot visualize the full range of possible outcomes. Different combinations of demand, weather, incidents, work zones, special events, and driver behaviors ultimately produce a wide range of outcomes. These outcomes will only be captured by simulation if a variety of input scenarios are considered. Regarding demands,
Figure 106 and
Figure 107 illustrate the hourly, daily, and seasonal variability that should be considered.

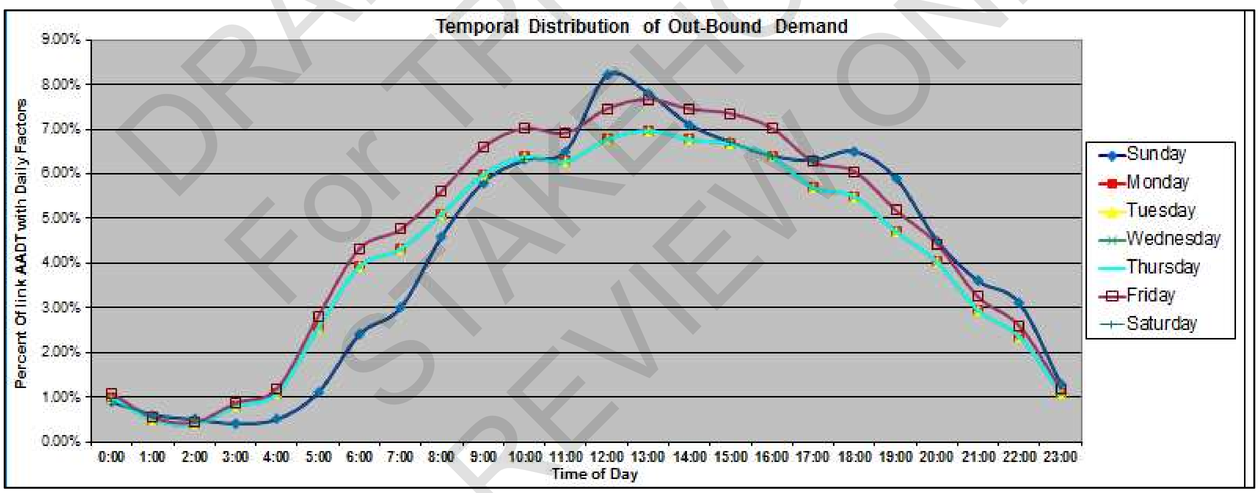


Figure 106. Chart. Hourly and daily demand distribution.

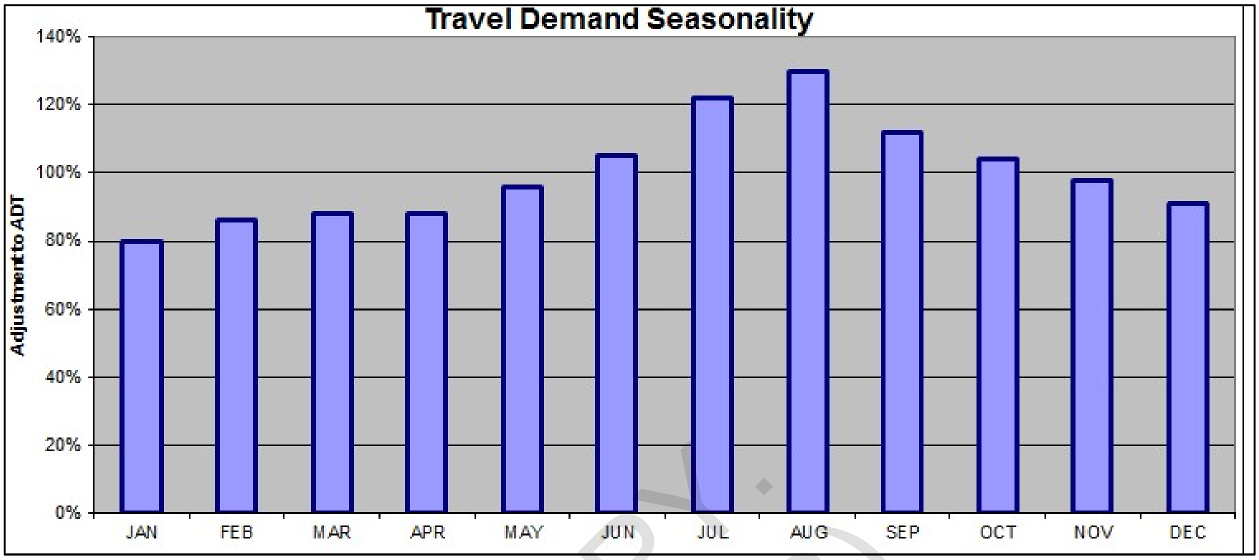


Figure 107. Chart. Seasonal demand distribution.

Incidents

Traffic incidents may be simulated implicitly or explicitly, depending on which tool is being used and what type of analysis is desired. Most microsimulation tools contain car-following models and lane-changing models that expressly prohibit unsafe vehicle maneuvers, such that collisions will never occur. However, incidents can still be simulated in such tools by specifying a lane blockage at a specific location, with a specific spatial length, and with a specific duration. Some tools may offer a range of possible incident locations, affected lengths, and affected durations. Monte Carlo simulation methods and random number seeds may affect such outcomes.
Figure 108 illustrates an example of a specific incident location and affected length of roadway.

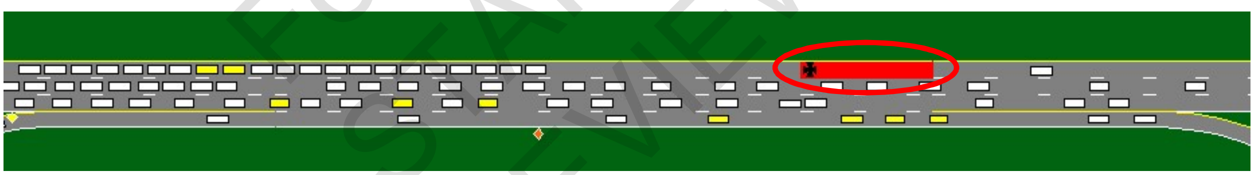


Figure 108. Illustration. Microsimulation of a traffic incident (circled).

In some tools, the impact of incidents on surrounding driver behavior can be calibrated. For example, there may be an incident warning sign location, which affects lane-changing behavior. In addition, there may be link-specific and/or lane-specific settings related to speed reduction (e.g., temporarily reduced speed limits), rubbernecking, and capacity reduction.

According to research that includes the SHRP2 L03 report (Cambridge Systematics et al. 2013), traffic incidents may also have a significant impact on traffic flow patterns, including the origins and destinations of vehicles. As described earlier in Chapter 5. (cluster analysis) and Chapter 8. , alternative scenario data sets and/or O-D matrices may be prepared for time intervals affected by incidents. This would be appropriate for macroscopic, mesoscopic, and microscopic simulation tools. The SHRP2 L04 (Mahmassani et al. 2013) report further describes methods for incorporating incidents into simulation projects. The simulated scenarios and/or weighting factors applied to scenarios affected by incidents, must reflect the fact that incidents affect only a relatively small number of minutes per year on any given roadway segment. Thus, a robust simulation study may consider incident frequency and the effects of incidents on speeds, capacities, demands, and travel time reliability.

According to the FHWA site on Analysis, Modeling, and Simulation for Traffic Incident Management Applications,[20] "Microsimulation tools are superior to other tools for evaluating the sensitivity of operations to small changes; such as individual incidents, or a small number of incidents throughout a large network. Mesoscopic tools often employ multiresolution demand and network modeling with dynamic traffic assignment (DTA) and selected subarea simulation, to predict how system performance and traffic demand will vary in response to an incident." Moreover, simulation models may now be used to support real-time incident management decision-making, including playbooks of pre-vetted congestion mitigation strategies. In the future, as computer speeds increase, it is conceivable that microscopic simulation will be more frequently used for DTA and/or real-time applications.

Weather and Visibility

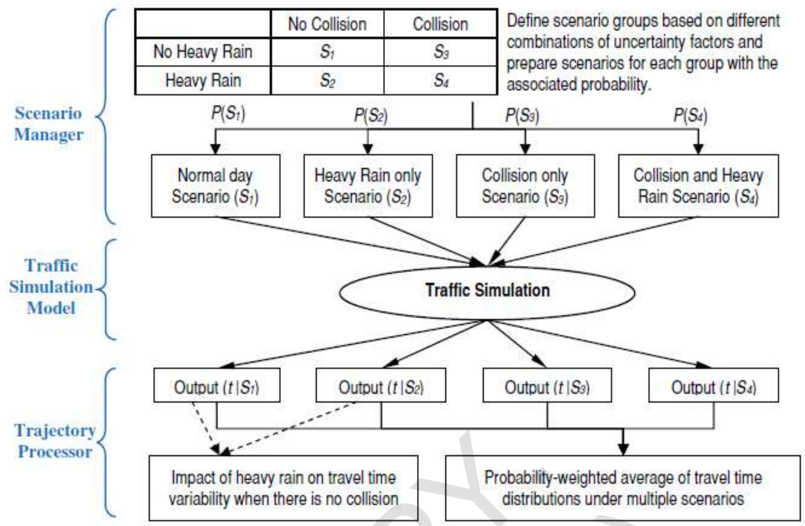
Like incidents, severe weather may also have a significant impact on traffic flow patterns, including the origins and destinations of vehicles. Indeed, when severe weather occurs, a substantial portion of drivers may choose to telework, use different routes, or select different destinations (e.g., a closer movie theater). As described in Chapter 5. (cluster analysis) and Chapter 5. , alternative scenario data sets and/or O-D matrices may be prepared for time intervals affected by severe weather. This would be appropriate for macroscopic, mesoscopic, and microscopic simulation tools.

The SHRP2 L04 report (2014) describes methods for incorporating severe weather into simulation projects.[21] The simulated scenarios and/or weighting factors applied to scenarios affected by weather must reflect the fact that severe weather affects only a relatively small number of minutes per year on any given roadway segment. Figure 109 illustrates these concepts.

---

[20] United States Department of Transportation, "Analysis, Modeling, and Simulation for Traffic Incident Management Applications," accessed October 10, 2019, https://ops.fhwa.dot.gov/publications/fhwahop12045/2_synth.htm.

[21] Stogios, Y., H. Mahmassani, and P. Vovsha (2013), *Incorporating Reliability Performance Measures in Operations and Planning Modeling Tools*, SHRP 2 Report S2-L04-RR-1, http://www.trb.org/Main/Blurbs/170716.aspx.



© SHRP2 L04.

Figure 109. Diagram. Simulation of different incident and weather scenarios.

Although the HCM primarily describes analytical methods not requiring or involving simulation, new travel time reliability procedures and chapters from the HCM, 6th edition, provide significant guidance on the impacts of incidents and weather on traffic speeds and capacities. Some of this same guidance could be applied during the data entry process for simulation scenarios affected by weather and incidents. Factors such as fog, which affect visibility, may also influence traffic in some areas. Thus, a robust simulation study may consider severe weather frequency and the effects of weather on speeds, capacities, and demands.

Work Zones and Special Events

In all major simulation resolutions (macro, meso, micro), the data entry requirements for work zones and special events may be like those for incidents. In fact, it may not be uncommon for engineers to model work zones as incidents, or vice-versa. The fundamental inputs may again include warning sign locations, free-flow speed adjustments (e.g., temporarily reduced speed limits), and capacity reduction.

Prior to creating the network, the engineer or agency will have defined the problem (Chapter 5.) by deciding if the simulation analysis will focus exclusively on time intervals affected by work zones and special events, or whether the analysis will also consider a proportion of time intervals unaffected by work zones and special events. As with incidents, a key consideration for work zones and special events is the impact on travel demand patterns. When drivers become aware of a work zone or a special event, the probability of route changes and departure time changes is high. An accurate simulation may not be possible unless these demand impacts are reflected by the input data. If the overall analysis must consider a mixture of time intervals with and without work zones and special events, the scenario-driven L04 approach illustrated earlier in Figure 109 is a viable option for capturing different demand patterns. When they occur in the proximity of intersections, planned work zones and special events are likely to warrant alternate traffic signal timing plans. A scenario-driven approach can also capture these alternate signal timing plans.

Traffic assignment simulations can determine driver responses to work zones and special events, but some tools allow the user to specify what percentage of drivers will divert to other routes. Figure 110 illustrates the results from a macroscopic simulation analysis of driver behavior (i.e., cancel trip, mode shift, hour time shift, takes detours) in response to a work zone.

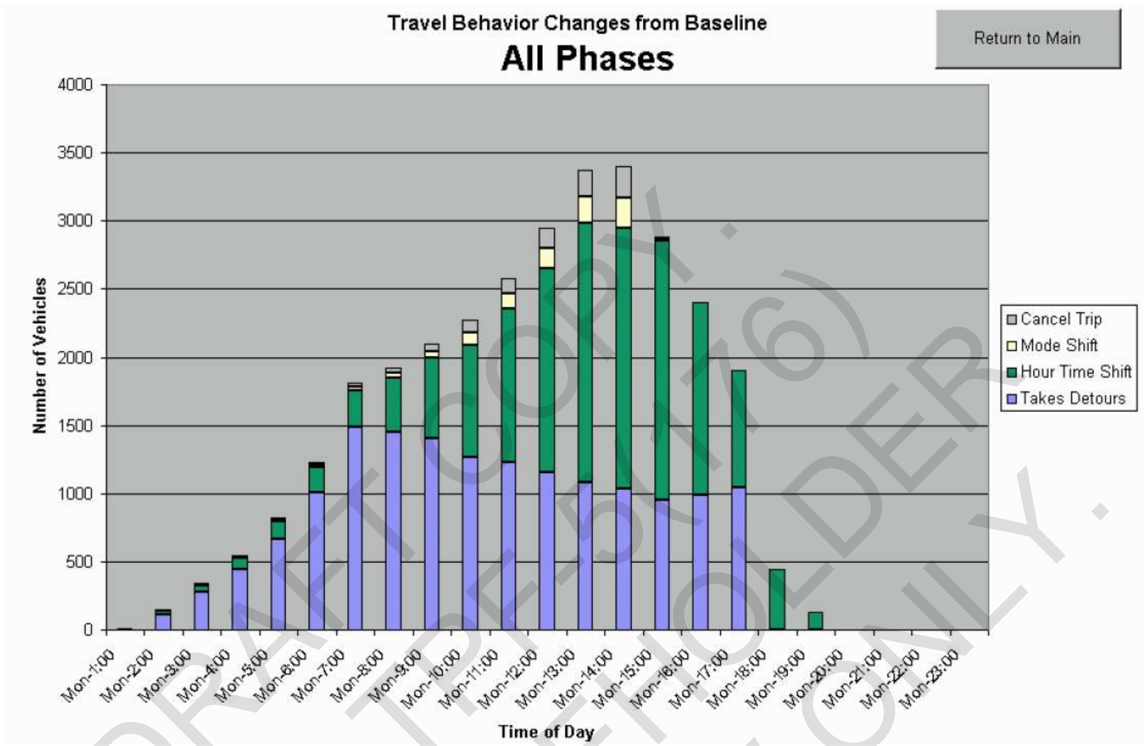


Figure 110. Chart. Work zone impacts on travel behavior.

7.8 TYPICAL CALIBRATION DATA

Chapter 8. provides details on the verification, calibration, and validation (VCV) process; however, the following provides a summary of some of the calibration elements to consider during the initial creation of the model.

Freeway System Calibration

Although there are different calibration approaches and philosophies, a small set of key input and output parameters must frequently be considered. For urban freeway segments and junctions such as those shown in
Figure 111, the following output parameters are commonly considered:

- ***Number of trips***: Throughput, which is a term that represents the number of vehicles discharged from any given segment or node, is perhaps the most fundamental simulated output parameter that must reasonably match the ground-truth. The number of discharged vehicles should be reasonably accurate at most links and nodes, if not all.
- ***Lane utilization***: The proportion of vehicles in each lane should be reasonably accurate.
- ***Average speed***: The speed of travel is quite sensitive to breakdown conditions. Inaccurate average speed results may mean that the simulated freeway capacity is not consistent with the real-world freeway capacity.
- ***Density***: Density is simply another way of expressing throughput divided by speed, and has historically been used by the HCM to determine freeway LOS.
- ***Travel time distribution***: A probability density function (PDF) of travel times allows computation of many important performance measures related to travel time reliability, which is an important consideration beyond simple average results.

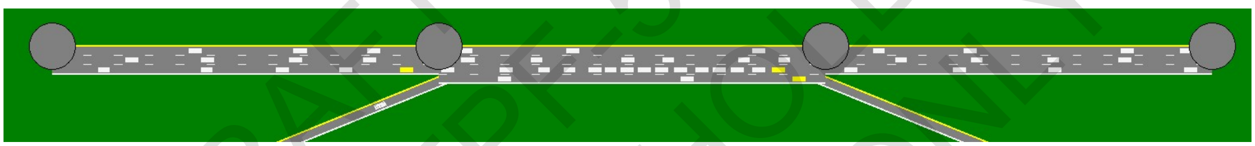


Figure 111. Illustration. Microsimulation of an unstable freeway weaving section.

To achieve accurate outcomes for the key measures listed above, calibration of the following input parameters is often considered:

- ***Car-following headway***: Urban freeway capacity often cannot be directly specified in microsimulation models. Car-following headway may be the most fundamental input parameter to influence freeway capacity. Highly designed freeways may achieve capacities of 2,500 veh/ln/hr, whereas lower-quality freeways may be limited to 1,900 veh/ln/hr. When a simulation tool is used, out of the box, with no calibration, the default capacity may produce very inaccurate results in some regions. Moreover, car-following headway must often be more aggressive in tight weaving areas.
- ***Warning sign distance***: This affects how far in advance vehicles try to move into the exiting lane for a desired off-ramp.
- ***Lane change settings***: The time and distance required to make a lane change may sometimes be set to more aggressive values if the situation dictates.
- ***Off-ramp heavy vehicle exit fractions***: Despite a significant number of heavy vehicles on some corridors, these vehicles might not exit at downtown areas in the real world.

Arterial System Calibration

Although there may be many different calibration approaches and philosophies, a small set of key input and output parameters must frequently be considered. For signalized corridors such as the one shown in
Figure 112, the following output parameters are commonly considered:

- ***Number of trips***: Throughput, which is a term that represents the number of vehicles discharged from any given segment or node, is perhaps the most fundamental simulated output parameter that must reasonably match the ground-truth. The number of discharged vehicles should be reasonably accurate at most links and nodes, if not all.
- ***Lane utilization***: The proportion of vehicles in each lane should be reasonably accurate.
- ***Control delay***: Control delay is the amount of delay caused by signal or sign control at an intersection. It has historically been used by the HCM to determine intersection LOS.
- ***Proportion of vehicles arriving on green***: Even when control delays are accurate, the quality of platoon progression between intersections should also be reasonably accurate to indicate a robust model.
- ***Traffic signal phase times***: Given the prevalence of actuated and adaptive traffic signals, the accuracy of phase times has been notoriously problematic for traffic modelers. Any verification or validation process should include some sort of audit of the phase times being simulated and how well these match real-world phase times.

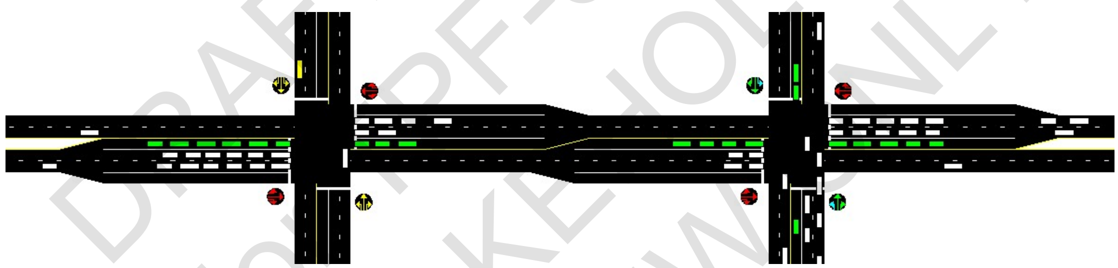


Figure 112. Illustration. Microsimulation of a signalized corridor.

To achieve accurate outcomes for the key measures listed above, calibration of the following input parameters is often considered:

- ***Adaptive and/or actuated signal parameters***: Accuracy of phase times is often problematic because the typical input data (e.g., maximum green, force off, permissive period, detector layout) cannot be used to determine phase times prior to simulation.
- ***Saturation flow rate and/or queue discharge headway***: Along with the traffic signal phase times, queue discharge headway is another fundamental input parameter that influences arterial capacity. Depending on a wide range of factors described in the HCM, signalized movement capacities may range between 1,300 and 2,300 veh/ln/hr. When a simulation tool is used, out of the box, with no calibration, the default queue discharge headway may produce highly inaccurate results in some areas.

- ***Extension of effective green time and/or amber response time***: During congested conditions, aggressive drivers may use the full yellow plus all red (Y+AR) clearance intervals between phases.
- ***Acceptable gaps***: For permissive left turns, right turn on red (RTOR), and stop sign operations, acceptable gap values within the software have a significant impact on simulated capacities.

Combined System Calibration

When calibrating a relatively large network (which may have numerous merge/diverge junctions and/or intersections), there may be a mixture of locations where results are acceptably accurate and locations where results have poor accuracy. In addition, many of the junctions and/or intersections having inaccurate results may be adjacent to one another. This implies the possibility that incorrect input data and/or model parameters at one specific location may be having a ripple effect through a series of adjacent links and nodes. In such cases, a thoughtful approach to the calibration process could save valuable time.

When simulating full networks as opposed to simple corridors or isolated junctions, it is helpful to differentiate between internal and external links. In a link-node diagram, external links are the segments having no upstream nodes. In
Figure 113 below, the northbound approach to node 1 is one example of an external link. By contrast, the southbound approach from node 2 to node 1 is an internal link. The red X indicates locations having inaccurate simulation outputs.

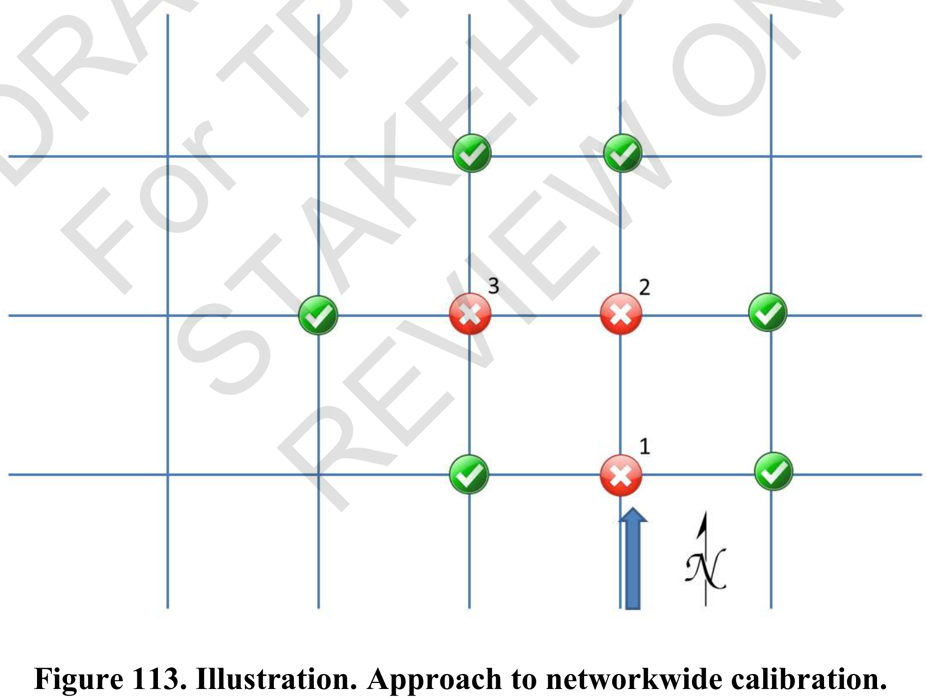


Figure 113. Illustration. Approach to networkwide calibration.

External links are easier to calibrate because their expected operation and results are much more predictable. In the

Figure 113 example where simulation output results are inaccurate at nodes 1, 2, and 3, it may save time to start by examining the only external approach to a node whose results are inaccurate (i.e., the northbound approach to node 1). If incorrect input data and/or model parameters can be identified here, the resulting correction may ripple through to rectify the accuracy of all downstream results. If any incorrect results remain after external link problems have been ruled out, a second phase of calibration may begin. This second phase may prove more challenging because internal link operations are affected by complexities such as queue spillback, platoon dispersion, shockwaves, and mid-block disturbances.

CHAPTER 8. VERIFICATION, CALIBRATION, AND VALIDATION

8.1 INTRODUCTION

This chapter addresses the tasks of verification, calibration, and validation (VCV) as they pertain to simulation models of transportation systems. A diagram of the steps involved is shown in Figure 104:

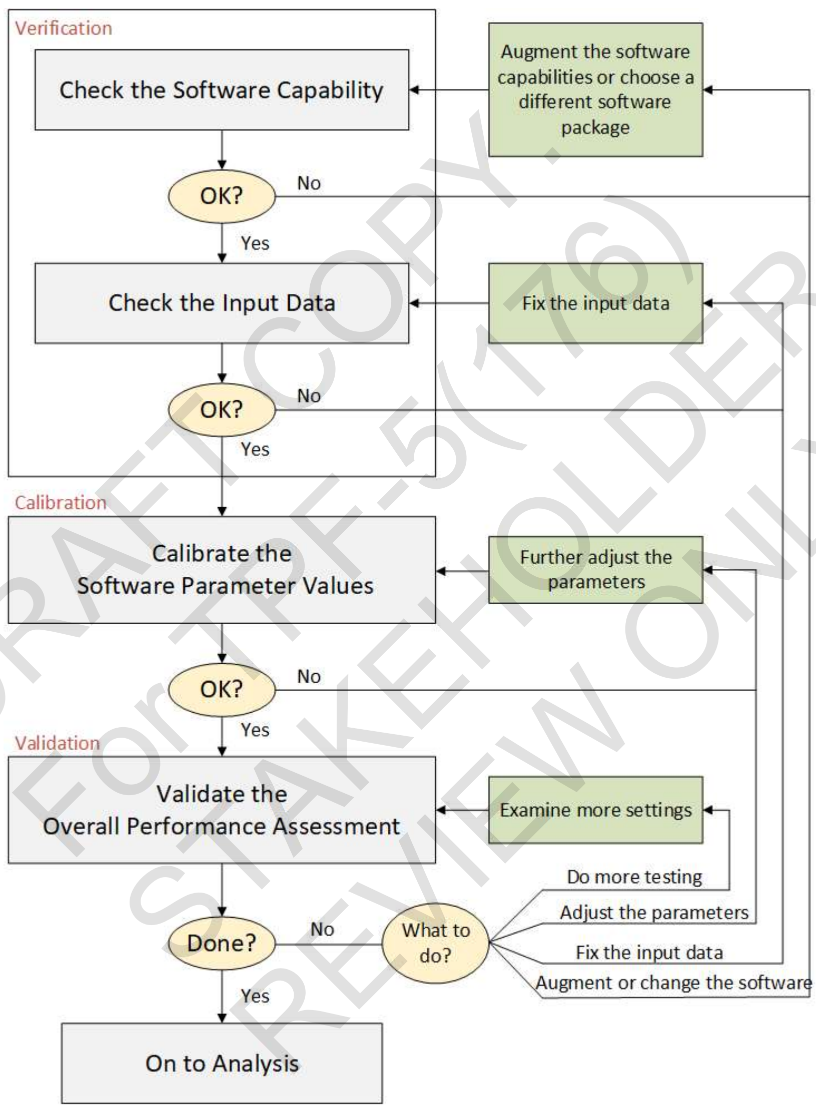


Figure 114. Diagram. The verification, calibration, and validation (VCV) process.

It is useful to remember that the simulation effort is intended to find new designs (or redesigns) for a system so that it performs acceptably for existing and/or future load conditions. The load conditions might exist "now" if current performance is unacceptable (congestion levels are too high); or they might pertain to a future design horizon when traffic growth produces demands which exceed capacity. Historically, the design changes have focused on the supply side, like altering the geometric design and operational controls; but today, they are beginning to focus on the demand side as well, such as the use of connected and/or autonomous vehicles (CAV) to boost throughput through more careful path and trajectory management.

The remainder of the chapter is organized as follows: section 8.2 discusses basic ideas and concepts, section 8.3 presents possible VCV metrics, section 8.4 discusses verification, section 8.5 focuses on calibration, and section 8.6 deals with validation. Section 8.7 revisits these three topics by presenting step-by-step procedures that can be followed to ensure that the best possible model results are obtained. Finally, section 8.8 summarizes the chapter's contents.

8.2 BASIC IDEAS AND CONCEPTS

In a sense, the word calibration has two meanings in simulation-related activities. One is the process of setting up the simulator so that it can predict the behavior of existing system(s) accurately. That is the focus of this chapter. The other, which might instead be called fine-tuning is the process of adjusting the design of the system being created so that it meets expectations.

A discussion about the design of an auto helps put these thoughts into perspective. Most people think about auto design as a physical process. First, a prototype is created, often called a mule. Then, the prototype is checked (verified) to ensure it has no missing elements, its subsystems are correctly connected, and they interact in the manner intended. Then, the prototype is calibrated. Its physical design, control systems, computer hardware and software are adjusted, based on testing, so that its performance is acceptable for all test load conditions.[22] Most of this calibration is done on test tracks, but some occurs on local highways and streets. When it does, other than for magazine test drives, it is done using driving circuits with specific features of interest, not present on the test tracks.[23] The load conditions test the performance limits of the car; they answer the question: When will it fail? The design is changed to improve performance (e.g., spring and dashpot rates, the connection points for the springs and dashpots, throttle controls, brake system settings, steering system parameters). When this calibration is complete, the design is validated through more extensive testing under field conditions on normal roads, maybe using normal drivers (not test drivers), and under normal (and abnormal) driving conditions. This final step helps ensure nothing has been missed in the calibration and verification steps, no situations are likely to arise whose evaluation has been missed, and the auto is unlikely to be asked to do things for which it is unprepared. This also helps ensure that customer (normal driver) perceptions of performance will be consistent with the design intentions and expectations.

Often, today, this physical design work is preceded by computer simulation. As with transportation systems, design ideas are checked to see if they might be acceptable. The first step is to calibrate the simulation software (the way that word is used in this chapter) to ensure it can

---

[22] Historically, as with the transportation systems, the design emphasis has been on the physical design of the car and its driver interface subsystems. Increasingly, today, the human side is becoming important as there are subsystems that aid or take over the driving task.

[23] These field driving circuits are used because they contain features that are important, but not present on the test track. There are such circuits in and around Detroit, for example.

predict the observed behavior of existing cars. This means adjusting parameters that pertain to the components of the auto, the building blocks upon which the model is based, to ensure that they are represented correctly: the power train, suspension system, body flexure and stiffness, tires/pavement, etc. An important element, here, is the virtual driver used to control the virtual car's movement. It is like the virtual drivers that control path choice and trajectory management in transportation system simulators. Once this calibration effort is complete, a computer-based model of the new car is created. This virtual prototype is verified for completeness by checking the data files. Its design is then fine-tuned (calibrated) by simulating the car's performance on virtual test roads. (These roads and their environmental conditions are also described using input data files.) Once this is complete, the car's design is then further validated by simulating its operation under additional load conditions; this may include the use of a driving simulator and real drivers.

In the case of transportation systems, however, a full-scale prototype is never built. The system is never tested outside of the context in which it will be used. There is no factory. There is no mule. The system is built where it will be used, it is tested under real-world load conditions, it only sees the load conditions that arise, and it may never see load conditions that push it to its performance limits.

This means simulation is very important. It provides a way for the new transportation system design to be tested and refined. It can play a major role in helping the analyst determine how to expand capacities, add new facilities (links, interchanges, intersections, etc.), reconfigure existing facilities, change the operating plan, introduce new use policies, etc.

Like the car example, simulation of the transportation system involves similar steps. A calibration effort must take place to ensure that the simulator can predict the performance of the existing system under load conditions that have been observed. Once that is complete, data files can be created that describe the proposed system and the load conditions to which it will be subjected. The new system's design can then be fine-tuned (calibrated) so that performance targets are met. Once acceptable performance is achieved, the final design can be tested more broadly (validated) for additional load conditions.

This idea of a load condition is very important. For the car, it is the combination of a trajectory (on the test track or road) and a context (such as inclement weather). For the transportation system, the idea is similar. It is a set of traffic demands (spatiotemporal flows) to be imposed linked with a context such as everything is normal, adverse weather, maintenance work, incidents, etc.

To be useful for calibration, these load conditions must be observed. As Figure 115. Illustration. Components of a load condition.shows, each load condition data set must have five elements: (1) a description of the physical system,[24] (2) the operational controls that pertain to the setting being examined,[25] (3) the applicable traffic demands,[26] (4) a description of any other attributes of the

[24] Most importantly, the facilities and vehicles.
[25] How the system's operation is managed as in the signal timings and tolls.
[26] The temporal-spatial matrices that describe the origin-to-destination (O-D) flows.

load condition examined (e.g., weather, maintenance activity, incidents, special traffic flow conditions, special operational controls, etc., and (5) the observed system outputs (performance).[27] The load condition is the traffic patterns in combination with the external environment. The system to which it is applied is the physical network, the vehicles, and the operating plan.

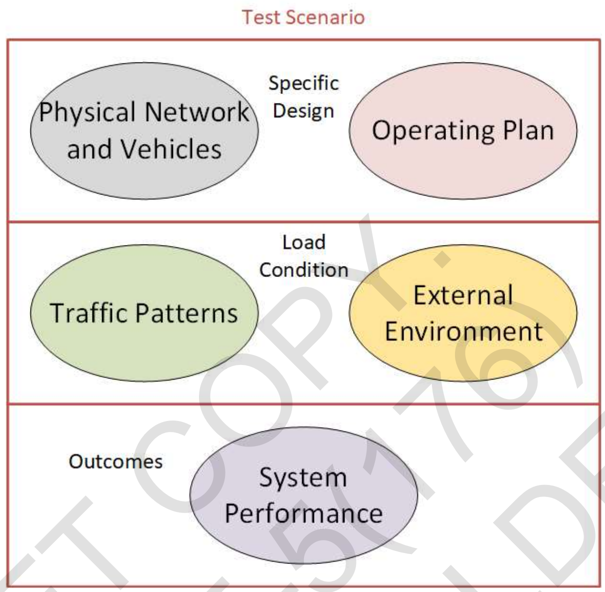


Figure 115. Illustration. Components of a load condition.

In the real world, the same (or nearly the same) load condition is likely to occur repeatedly. For example, each year, there will likely be many times when the traffic pattern is the morning (AM) peak on a weekday when schools are in session and the external environment is normal.

There also tends to be a repetition cycle for the load conditions, typically 1 year. Based on this load cycle, the frequency with which the load conditions occur can be ascertained, and the load conditions can be clustered into groups. This is equivalent to the clustering ideas presented in the update to Volume III of the Traffic Analysis Toolbox. See Wunderlich, et al. (2018). In SHRP 2 L-02 (List et al. 2014) , the research team found that for three interstate routes in San Diego, the breakdown of load conditions shown in Figure 116 was useful.

In the L-02 case (2014), the categorization of the traffic pattern was categorical: uncongested, low, moderate, or high. The external environment categories were: normal conditions, higher-than-normal demand, weather, special events and incidents. The *n* values in the table indicate the number of 5-minute intervals in 2011 during which each specific load condition arose. For example, in the case of the I–5 route, the normal, uncongested load condition occurred in 55,533 of the 72,000 recorded 5-minute intervals. This is 77 percent of the 5-minute intervals. Incidents were underway during high traffic conditions during 466 of those 5-minute intervals, or 0.6 percent. (The SV values are the semivariances for the observed 5-minute intervals based on an

[27] Weather, maintenance, and other external activities that constrain or influence the system's operation.

acceptable travel rate of 1.33 minute per mile [min/mi], the equivalent of 45 miles per hour [mph]. The SV values provide an indication of the unreliability of the travel times for that condition, where higher values mean worse reliability.)

Route	Cond	Normal		Demand		Weather		Special Events		Incidents		Σ(SV*n)	Facility
		SV	n	SV	n	SV	n	SV	n	SV	n	(000)	Total
I-5	Uncong	7	55533	60	1250	46	797	111	135	172	285	573	5278
	High	205	12783	1415	472	2563	175	1399	104	1769	466	4705	
CA-15	Uncong	15	24491	47	147	68	229	29	77	139	55	400	9465
	Low	27	15931	118	102	106	193	0	0	97	25	457	
	Mod	46	14863	127	13	151	271	0	0	93	103	740	
	High	241	13918	2415	665	3751	162	3113	168	3032	587	7868	
CA - 163	Uncong	11	32823	13	1019	61	277	21	29	54	102	386	9561
	Mod	56	20950	169	519	399	333	601	344	684	354	1841	
	High	261	12764	1789	1028	1924	254	1424	243	1385	961	7333	

Figure 116. Screenshot. Load conditions for three freeway routes in San Diego.

Similarly, for purposes of the VCV effort, the observed load conditions should be grouped into clusters for which the load condition is the same or similar. The number of clusters to use is addressed in Wunderlich et al. (2018), as well as how to identify them. If only one load condition is of interest, such as the evening (PM) peak on normal weekdays when all other conditions are normal, then maybe only one cluster is needed. However, if the PM peak is also of interest on days when there is inclement weather, then two or more clusters might be necessary (e.g., heavy rain might be different from snow).

When a holistic analysis of the entire duty cycle (whole year) is of interest, then several clusters are likely to exist. Figure 117 shows the clusters presented by Wunderlich et al. (2018) for a hypothetical network. Bottleneck duration and travel time determine the clusters. The duty cycle is a year.

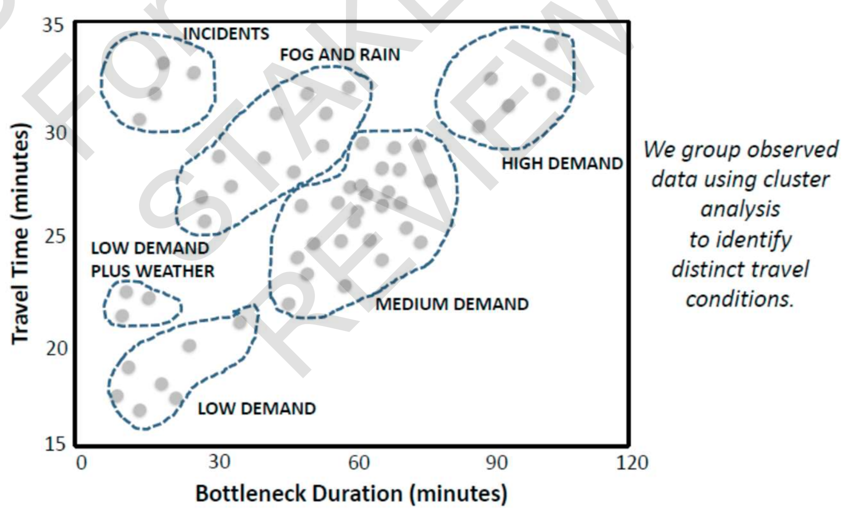


Figure 117. Illustration. Hypothetical clustering for 1 year's data.

Also, as Wunderlich et al. (2018) suggest, a representative day from each cluster should be selected for calibration, not a synthetic day created by combining data sets. This representative day should be as close to the mean values as possible for the inputs and outputs of greatest interest. It should also have the best, cleanest data, especially in terms of the correspondence between demands and performance. The remaining load condition observations can be used later for validation. Figure 118 shows the hypothetical selection of representative days from the clusters of days shown in Figure 117.

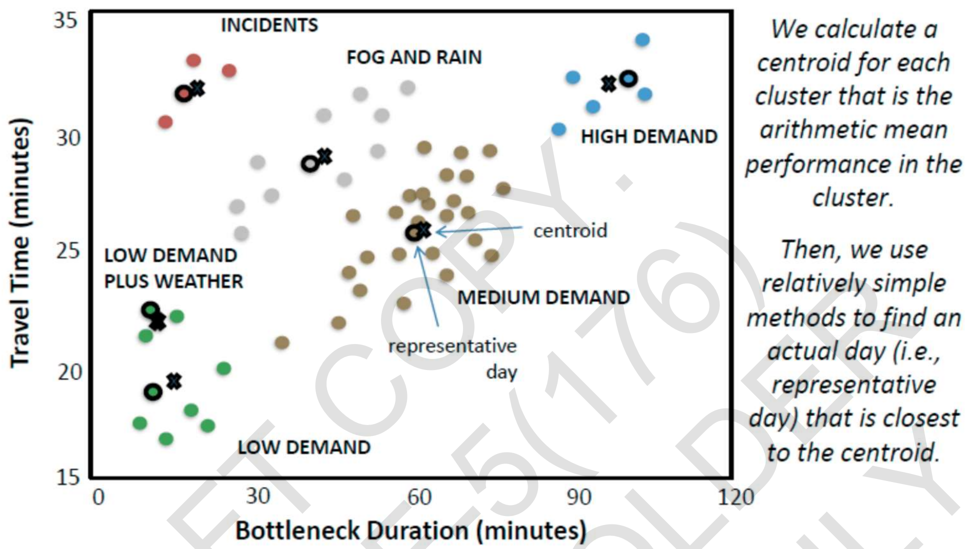


Figure 118. Illustration. Selection of representative days.

This means not just one set of parameters exist; but, rather, sets of parameter values for each cluster of load conditions. For example, one set of values might pertain to the PM peak for adverse weather conditions, like snow. Another might pertain to the morning (AM) peak under normal conditions. In the snowy condition, the drivers may be more cautious. They may want larger gaps and be more hesitant about lane changing. The flow rates may be lower and the trips may be more spread out temporally. It is also possible that the signal timing plans may use larger gap times, different offsets, and different minimum (min) and maximum (max) greens. The analyst should expect to find that multiple sets of parameter values will be found.

It is helpful to categorize the model parameters. A scheme using four categories is shown in Figure 119. The user behavior category includes parameters that affect path choice (in the context of dynamic traffic assignment) and trajectory management (on and across specific facilities, such as car following and lane changing). This is the category that analysts typically think of when the word calibration is mentioned. The demand category includes the origin-destination (O-D) flow rates and their variations with time. (The importance of these parameters is often overlooked.) For multimodal systems (e.g., where transit is involved), this demand category can include multiple O-D data sets. (The user behavior category can also include parameters indicating how transit users make path and departure time choices.)

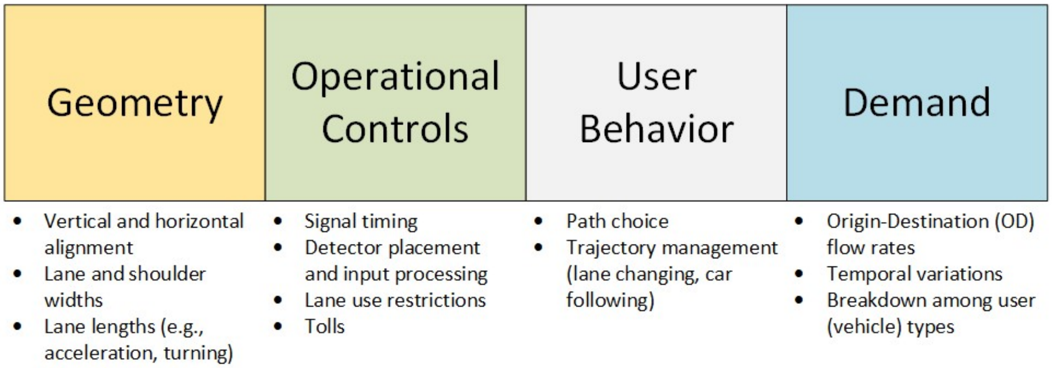


Figure 119. Illustration. Parameter value categories for simulation models.

The geometry category includes parameters that describe the physical design of the system. Examples include the number of lanes on specific facilities and the lengths of acceleration and deceleration lanes. The operational controls include parameters for how the system operates; how it controls and directs the flow of users (drivers, vehicles). An example is signal timing. The parameters in these two categories have historically been thought of as being part of the design process, not calibration.

In the context of today's evolving transportation world, it is important to realize that the parameters in the user behavior category are becoming part of the design set as well. They are not just involved in calibration. As CAVs become involved in trajectory management (e.g., car following and lane changing) and path choice (in the context of dynamic traffic assignment), the parameters for these control systems also need to be adjusted.

When the simulator is being calibrated so that its predictions match existing, observed behavior, it is likely that focus should be on user behavior and demand parameters. The system's physical design is already set and operational controls specified. The analyst should *not* expect to adjust these values to get the simulator to match observed behavior. It *is* important to check that they are correctly specified; for example, the way in which the detectors collect and process data, or the way the signals and other control devices use those inputs. Conversely, if the analyst senses that parameters should be adjusted, or if changing them improves the match with observed performance, then it is likely that something else about the model is misspecified, like the demands.

Once the simulator matches existing behavior, it is likely that the user behavior parameters should not be changed.[28] An exception would be where the vehicle technology is changing, as with driver-assist features; another would be where connected vehicles are becoming part of the vehicle mix.

The parameter values most likely to be changed in new designs are those in the categories of geometry and operational controls. This includes the amount of capacity provided, geometric

---

[28] Of course, different parameter values may pertain to each load condition.

configuration, and operational controls, including tolls. To the extent that demand-side changes such as CAVs are also likely to change, parameters for those subsystems should be adjusted as well.

Finally, it is useful to think about the models used in the VCV process as having five components, as shown in
Figure 120. They are: 1) the logic statements (the if-then relationships and equations that determine how the model works), 2) the software (code) that implements that logic, 3) the parameter values used by the logic (and by the software) to determine how the simulation will progress (e.g., the inputs to the path and trajectory management models that determine how the vehicles behave), 4) the data that describe the settings of interest (network geometry, operational controls, demand, load conditions), and 5) the output data that show how the system does perform in response to the inputs. Verification pertains to checking the logic, the software, and the input data; calibration involves checking the parameter values based on the output data for data sets deemed useful for calibration purposes; and validation pertains to checking the output predictions based on the output data given the input data for the validation data sets.[29]

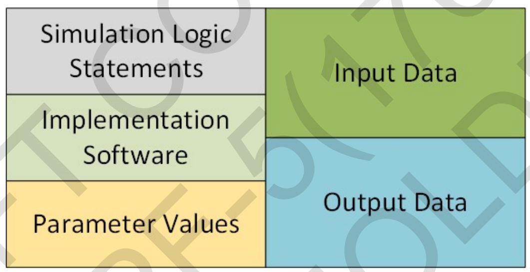


Figure 120. Illustration. Simulation model elements.

8.3 A SIMPLE EXAMPLE

A hypothetical signalized network study helps illustrate these ideas. Assume the network has spillback problems because of traffic levels and close intersection spacing. Assume the signals are actuated but not coordinated and the traffic engineer thinks that coordination will help improve performance.

The analyst might proceed as follows. First, a set of base conditions would be selected, ones for which system performance has been observed. These would be used for calibration or validation. If more than one load condition were of interest, clusters would be created. For each cluster, a representative day would be selected for calibration. That day would typify the cluster conditions and have the best, cleanest data, especially in terms of the correspondence between demands and performance. Each representative day's input data files would be verified for correctness and

[29] That is, to check the box for verification, the software must be able to model the conditions of interest, and the input data must be clean. For calibration, the parameter values, in combination with the test inputs, must be able to produce the test outputs. For validation, the outputs must arise from the inputs, based on the calibrated parameter values. In all cases, the inputs and outputs must be congruous for the conditions examined. That is, the inputs specified *will produce* the corresponding outputs *if* the model is correctly calibrated.

completeness. The model would then be calibrated for each representative day by finding sets of user behavior and demand parameters that allow the observed behavior to be matched. The user behavior parameters would be addressed first. The trajectory management (car following, lane changing) parameters would focus on matching saturation flow rates and lane-changing behavior on critical approaches. The path choice parameters would be set so that route diversions can be matched in response to things like congestion and incidents. Next, the demand values would be adjusted (within reason), especially the temporal and spatial O-D patterns, so that the observed spatiotemporal delays and queue dynamics were matched. Once this was done, the calibration results would be validated by using the other days in each cluster set. The user behavior parameters for each cluster would be used. The demands would be adjusted (again, within reason).

If problems were identified with the results, the process would repeat. The model might have difficulty matching the results for the data sets in the validation group.

One thing the analyst might do is to make the user behavior parameters location specific. For example, they might need to be different for specific signal approaches because of sight distances, vertical and horizontal geometry, or other factors. They also might need to vary with the flow rates. Drivers might be more aggressive about both path and trajectory management decisions for high levels of congestion. In making these adjustments, the user behavior would be adjusted first, remaining consistent with traffic engineering principles; then, if necessary, the demand values would also be adjusted.

For the future conditions, the geometry, operational control, and demand values would be specified. (In this case, the demand values would be unconstrained inputs. They do not have to be matched to observed performance.) For each load condition, the appropriate set of user behavior parameter values would be chosen based on the cluster to which the load condition belongs. (Once the user behavior parameter values are set for a given cluster, they should not be changed during the design analyses unless something about the mix of vehicle types changes.) Once these steps were complete, the geometry and operational controls would be adjusted so that desired performance is achieved.

In all of this, it is important to remember that the trajectory management and path choice algorithms in all simulators are robotic (without any complaint intended). The logic they use produces the same response to the same inputs every time they arise. There is no random variation, no inattention to task, no inability to monitor and sense the location and speeds of all nearby vehicles. This means that, in calibration, the analyst is calibrating these robotic drivers so they produce user behavior that is consistent with that which produced the observed system performance.

As this example implicitly suggests, it is critical that the performance of the basic elements be correct, individually and in combination. If not, the simulation analysis is doomed. The analyst should first check the model's predictions of the performance of individual freeway sections, arterial sections, interchanges, intersections, weaving sections, merge and diverge sections, etc., *not* just the whole system. The analyst should check how these components perform in isolation *and* in combination; in the latter case, as with closely spaced intersections, freeway bottlenecks immediately downstream of interchanges, complex weaving sections, etc.[30] If these elemental and subsystem level models can predict behavior correctly, then, by combining them, they should also be able to produce defensible predictions. If they cannot produce defensible results, individually or as small subsystems, then the parameter values need to be fine-tuned more; otherwise, something about the system description is awry (or some fundamental aspect of the system's behavior has been overlooked and is not being modeled).

Moreover, if these elements, when combined, can predict the observed behavior of the base case systems, it is likely that they will be able to produce defensible predictions of performance for future, altered systems, which is the reason the simulation analyses are conducted. If they cannot do this, then those future predictions will not be defensible. It is *only* if those building blocks have been calibrated to produce defensible results that predictions of future configurations for future load conditions has a reasonable chance of producing defensible results. (Moreover, as indicated previously, much of this emphasis needs to be on the demand side, ensuring the models for predicting path and trajectory management decisions are appropriately calibrated for, and in the models of the future, correctly selected for the site and load conditions of interest.)

8.4 PERFORMANCE ASSESSMENT AND METRICS

Picking the best metrics to use for the VCV tasks is critical. These simulation quality metrics (SQM) must indicate whether the model's predictions of performance are consistent with observed performance or not.[31] They must also be useful to all the audiences that review the study findings. Some audience members may be interested in metrics that are different from those the analyst would otherwise select. Examples might be air and noise pollution measures. The best metrics to use for a given study may depend on the purpose and the type of model used (e.g., micro, meso, macro, etc.).

It is useful to think about the metrics as being in upper and lower levels, related to both supply and demand. The upper level SQMs should relate to overall system performance. On the demand-side, an illustration is aggregate travel delay (demand); on the supply side, it could be the number of links whose travel rates (min/mi) are higher than desired values. In both cases, average values of the metrics are useful; but, their distributions of values are better (e.g., the distribution of travel times). It may be that the average value is OK, but the higher percentile values are unacceptable.

The lower level SQMs should focus on component and subsystem performance. On the demand side, an example is the set of travel time distributions for individual O-D travel times/rates (demand) or bundles of O-Ds. On the supply side, it might be the distribution of discharge rates

[30] An important reason for doing this is that the system may contain hidden bottlenecks, ones that emerge once other problems are resolved. If this component-level assessment is done, then, when those bottlenecks emerge, the model will be able to represent their behavior correctly, or at least in a defensible manner, since no observations exist. Otherwise, if this is not done, the model may create erroneous predictions of performance.

[31] It is possible the SQMs may need to be modified as the VCV process unfolds. That is fine. The main challenge, at the outset, is to identify a set of SQMs likely to provide the quality assessment desired.

delays or queue lengths for specific freeway bottlenecks, or saturation flow rates, queue lengths, and approach delays for signalized intersections.

The probability density function (PDF) or cumulative density function (CDF) for these metrics are better to use than average values or single percentiles.[32] The distributions help show the range of performance that arises, within a given load condition or across a set of them. They capture the effects of stochasticity. They can show when the capacity investments or operational changes are producing significant changes in performance. This guidance is consistent with the admonishments of studies that have focused on reliability assessment such as SHRP-2 L02 (2014), L04 (Mahmassani 2014), L08 (Zeeger et al. 2014), C05 (Kittelson & Associates 2010), and L13 (Tao 2013). While it is not reasonable to focus on the distributions of travel times or rates for all the O-D pairs for a given network setting, it is reasonable to focus on the distributions for the top 10 O-D pairs (in terms of volumes) or a select set of O-D pairs that form a representative sample of the types of trips being served by the network.

To elaborate more about the demand (traffic) side metrics, these SQMs should measure the quality of the service provided. They should focus on performance that is meaningful to the users. Travel times (and travel rates) are good choices. The travel times can be for specific origin-destination (O-D) pairs or point-to-point paths through the network. The variability of these travel times is also important, as in the nature and shape of the PDFs[33] for travel times and travel rates, and delays for specific paths and specific O-D pairs, and for the system overall.[34]

While it is possible to monitor the travel times of all O-D pairs, this would be a large effort. Instead, it might be better to focus on a select set of O-D pairs, say the top 10 (20, 100, etc.); perhaps, the most significant ones or a set of representative pairs.[35] This limited set, if carefully chosen, can keep the workload manageable, and the process better focused. In fact, a well-chosen set can help identify significant problem spots in the network. For example, the set might include an O-D pair for every congested location; and/or it could have one or more O-D pair for every bottleneck location.

For supply-oriented metrics (i.e., for the operator or agency), an appropriate focus might be on the behavior of individual elements and subsystems. Of interest might be the response of these elements and subsystems to demand, as in queues, delays, spillover effects, and facility interactions. For a freeway corridor, the analyst might want to ensure that individual bottleneck locations function correctly: that the queue dynamics are correct for traffic inputs up to capacity.

---

[32] CDFs are companions to PDFs. They describe the percentage of observations that are equal to or less than a given value. If an approach delay of 30 seconds per vehicle (sec/veh) is at the 90th percentile, then 90 percent of the delays are that value or less.

[33] A PDF indicates the percentage distribution of values for a variable whose values vary. For example, if a delay is 20 seconds 80 percent of the time and 10 seconds 20 percent of the time, then the PDF would indicate that this was the case.

[34] It is always possible the average performance is acceptable, but the worst (higher percentile) performance is not.

[35] Another option is to identify all the bottleneck (problematic) locations in the network and ensure that a few O-D pairs are monitored for each one. It is also possible that intermediate, hypothetical O-D pairs are useful to monitor, such as the traffic going from node A to node B (regardless of the traffic's ultimate origin and destination), because that traffic passes through the bottleneck(s) of concern.

Then, if there are spillover effects, the analyst should also check the system response is correct for those interactions. That is, the individual facilities should be calibrated so they would produce correct outputs (given first principles of traffic engineering) up to capacity; and then, when they are combined, as they are in the field, the model should produce correct results based on field observations of the operations.[36]

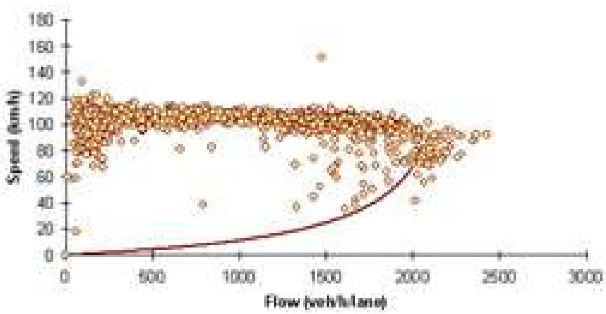
It is also important to ensure that other portions of the system produce results consistent either with first principles or field observations. One example is weaving sections. The delays and speeds should be defensible, and so should the lane changes. Another example is the traffic trends upstream of major exit ramps; the exiting traffic should be trending toward the ramp in a manner that is consistent with first principles and field observations, whether those observations are formal or informal. If there are hidden bottlenecks that might emerge if current bottlenecks are remedied, then the performance of those locations should be checked, through parametric analyses (independent of what the traffic flows presently are or are forecasted to be), to ensure their behavior is consistent with first principles if they were to become bottlenecks.

Put another way, the analyst should be assured that these facilities, separately, or in small clusters (where system effects arise), are being modeled correctly: do the predicted outputs match up with those that would be expected if the traffic flows were such that performance is consistent with that which is observed. This means acceptable facility-specific travel times (and travel rates), queue lengths (predominantly, on a per-lane basis), bottleneck discharge rates, segment capacities (by direction), saturation flow rates at intersections, volume-to-capacity (V/C) ratios, and demand-to-capacity (D/C) ratios. These SQMs should indicate whether the operational controls are suitable, and the designs are acceptable. If the simulation model can assess safety performance as well, as in producing surrogate safety measures (SSM) then metrics like the frequency of near-collisions are useful. SQMs for pollutant outputs, noise impacts, and other side impacts of the system's operation would also be of interest. Additional information on simulating pollutant outputs and noise impacts provided in Chapter 5. .

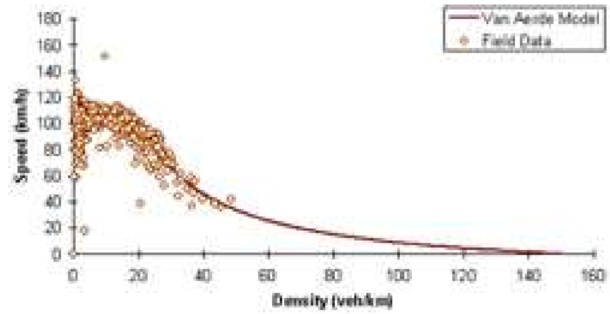
Ensuring the links perform in accordance with the fundamental diagrams of traffic flow is a useful way to view this disaggregate assessment.
Figure 121 provides illustrations of freeway relationship between speed, flow, density, and headway.[37]

Simulation models should be able to reproduce these relationships with appropriate values (e.g., the free-flow speeds or travel rates, the maximum flow rate or capacity, and the decline in speed with density) appropriate for the freeway and arterial facilities being modeled.

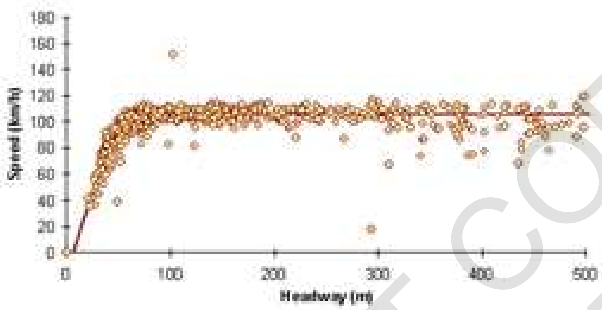
[36] Pursuing these ideas may create the need for hypothetical datasets that focus on testing an element's performance, or that of a subsystem, for a range of loading (input) conditions.
[37] The relationships illustrated can be found at https://ops.fhwa.dot.gov/publications/weatherempirical/sect3.htm.



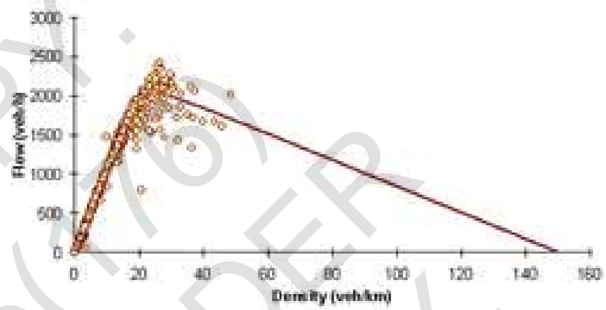
A. Subfigure of speed and flow.



B. Subfigure of speed and density.



C. Subfigure of speed and headway.



D. Subfigure of flow and density.

Note: free speed: 106 kilometers per hour (km/h). Jam density: 150 vehicles per kilometer per lane (veh/km/ln). Speed at capacity: 85 km/h. Capacity: 2,031 veh/h/ln. Wave speed: -17.0 km/h.

Figure 121. Charts. Fundamental relationships among speed, flow, and density.

A top 10 set of metrics is useful here, as it was on the demand side. This set should focus on fundamental relationships that pertain to the links and nodes. For example, they should focus on queues at bottleneck locations, delays at those same locations, corridor travel times, or O-D travel times.[38] The choices should be predicated on the likelihood that the metric will always be in the top-10 list. The distributions of these metrics, far more than their average values, will be very helpful when making design decisions. At signalized intersections, the distribution of queue length is a logical choice for an SQM, with an associated thought that a specific percentile value may be of greatest interest. From a design standpoint, the 80th percentile value might be useful. This would mean the queue length would exceed that value only 20 percent of the time (one day a week). The higher level assessment metric might be the sum of the squares of the deviations between simulated and observed 80th percentile values, at a 5-minute level of resolution.

In the context of the dynamics associated with peak period operations, such as the PM peak, good facility-level SQMs include the temporal patterns of queues, delays, arc flow rates, and travel times. A good system-level SQM is the length of time it takes the system to return to uncongested operation. This metric is easy to observe in the field, and if the simulation model predicts an erroneous value, something clearly needs to be recalibrated. It is even better to

[38] Top 10 may also be replaced for a top 20, top 100, or some other basis.

examine the PDF for this time and gain a sense of how short and how long the simulation model predicts this time might be and how the time is distributed.

Clearly, the overall performance of the system should be monitored, such as the average speed (travel rate), the average travel time, the aggregate delay, and the average delay. However, such aggregate metrics rarely provide clear insights about where the problems are or how to fix them. It is better to calibrate the predictions of the most problematic locations in the network first and then focus on overall performance.

The type of data that can be used to develop these SQMs depends on the type of simulation model being used (micro, meso, macro, etc.) Microscopic models can yield both aggregate measures and detailed data about individual vehicle trajectories. For meso and macroscopic models, queue lengths, flow rates, densities, speeds, travel times, delays, and shock wave measures are available.[39]

For microscopic simulation models, using trajectory-level data is becoming increasingly popular. These data show how individual vehicles are traversing the network. Such data can facilitate an understanding of the simulation outputs, be used to puzzle out unexpected results, and provide unbiased comparisons between results from multiple tools. Figure 122 illustrates the basic concept of trajectory-based delay calculation on a signalized intersection approach.

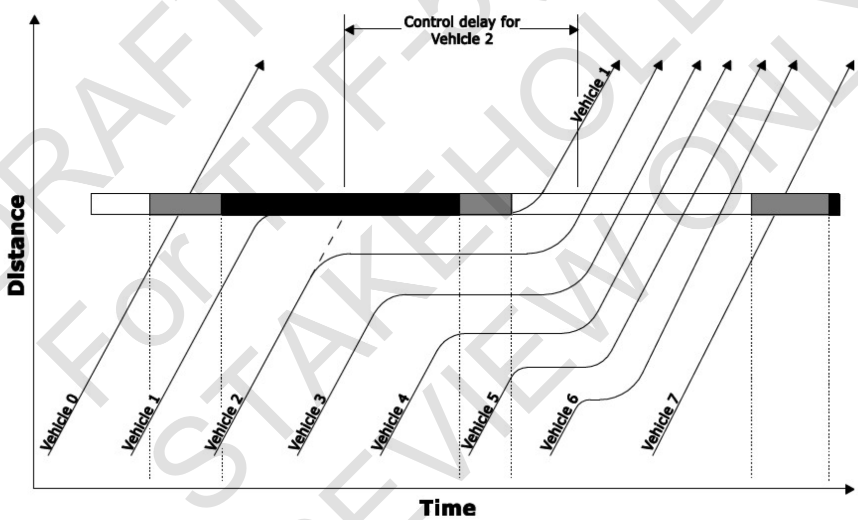


Figure 122. Chart. Trajectory-based computation of control delay.

Total delay is typically obtained by subtracting the (hypothetical) free-flow travel time from the actual travel time for each vehicle. Note that free-flow travel time depends on the desired free-flow speed of each driver, and thus, will vary between vehicles. Stop delay accrues whenever a

[39] It is very important to remember that for microscopic simulation models, capacity is an output (a manifestation of the parameter values chosen), while for meso- and macroscopic models, capacity is an input.

vehicle's speed falls below a certain threshold, such as 3–5 mph. Travel times may be calculated as segment length divided by the average speed of each vehicle. To obtain density, the average number of vehicles on a segment can be divided by the segment length.

These data can also be used to compute aggregate assessments of delays, queue lengths, flow rates, and speeds (both spot speeds and space-mean speeds). The trajectory data are also very useful for calibrating specific submodels (e.g., car following, lane changing, saturation flow rate discharge).

To determine queue-related outputs (e.g., average queue, maximum queue, 95th percentile queue), there must first be a rule or procedure for identifying a queued state. This procedure may consider the gap between a vehicle and its leader, vehicle speed, vehicle acceleration, and distance to the stop line. Refer to NCHRP Report 385 (Roess 1998) and/or HCM (2016, chap. 36) for more details on the recommended procedure for identifying a queued state.

Aggregate data can also be used. This is performance information prepared by the simulation software package based on internal equations and measurement techniques. The minor problem with such information is that it may be difficult to determine how the measurements were taken and how the metrics were computed. But if the analyst understands the answers to both those questions, there is certainly nothing wrong with using this information. It eliminates the time and effort involved in post-processing disaggregate data. And it can be used in fine-tuning the calibration.

For macroscopic models, the facility-level testing focuses on calibrating the speed-density-flow relationships for specific facility types (and specific locations). The queue dynamics upstream of bottleneck locations (on freeways or at signals) are a primary concern. Vehicles on a roadway segment can accumulate when input flow rates exceed available capacities, or when traffic streams are temporarily stopped by a traffic control device. A sample queue accumulation polygon (QAP) is illustrated in
Figure 123.

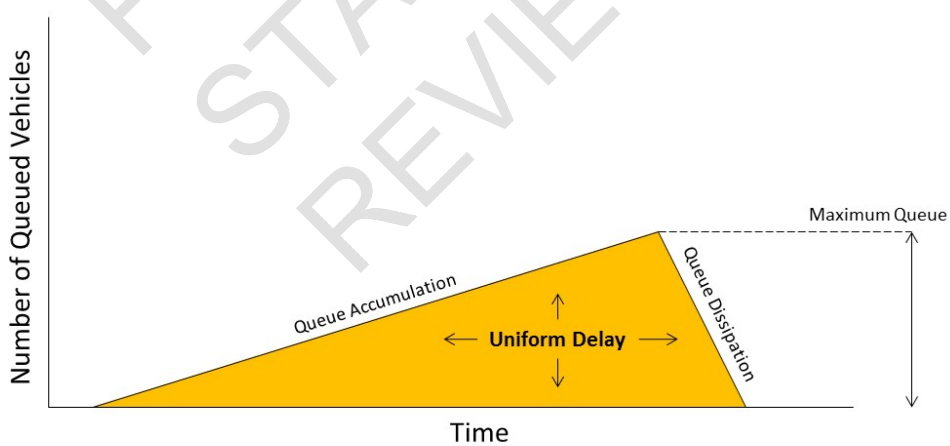


Figure 123. Queue accumulation polygon (QAP).

The queue will typically dissipate when a subsequent time period brings a lower input flow rate, or when the traffic signal turns green. After the queue has been fully eliminated, uniform delay can be computed as the two-dimensional (2D) area under the QAP. For delay to be computed accurately, the number of queued vehicles should be zero on both the left and right sides of the QAP. Indeed, this motivates a selection of time periods that avoids oversaturated conditions at both the very beginning and the very end of a simulation.

The maximum queue can also be obtained as the highest point of queue accumulation.

Figure 124 illustrates an arrival flow profile, which can be used to determine a proportion of arrivals on green. Finally, if discharge flow and saturation flow profiles are constructed, the 2D areas under these profiles can be used to compute throughput and capacity, respectively.

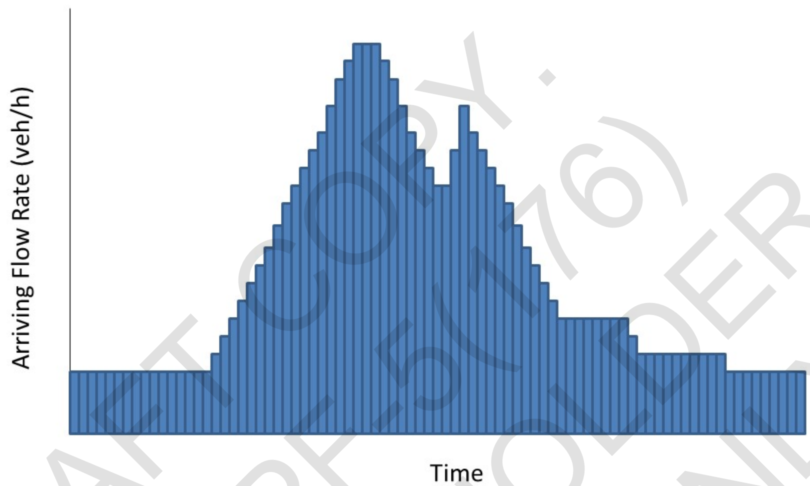


Figure 124. Chart. Arrival flow profile.

8.5 VERIFICATION

Verification is the first step in the VCV process. As Figure 104 previously showed, this step involves two substeps: 1a) ensuring the simulation software can be used to analyze the settings of interest and 1b) checking the input data accurately represent the system. These substeps, especially the second, may be performed more than once during the VCV process. It may also be that problems identified later in the VCV process require revisiting the software's capabilities.

Because the technological frontier for autos and trucks is evolving rapidly, the first substep is important. The analyst needs to be sure the software can adequately simulate the load conditions foreseen for the future conditions being used for performance assessment. For example, the future conditions might include connected and automated vehicles (CAV). The software must be able to simulate their behavior. Otherwise, a workaround analysis strategy must be developed, another software package must be employed, or the software must be augmented so these vehicles can be modeled (which researchers often do). Ideally, this check was performed when the project was scoped (see Chapter 5.), but it needs to be done.

Completing the first substep should be straightforward. The user manuals for the software should be reviewed to make sure the capabilities are understood. If there are questions, contacting the technical support staff for the software is an option. Speaking with colleagues who have also used the software is another option. Checking the Web for documents that provide assessments of the software's capability is another possibility. Contacting a researcher who has used the software (and perhaps augmented it) might also have value. The main objective is to minimize the likelihood the software is being used to study one or more settings for which it is not prepared or intended.

Substep 1b), which is more traditional, focuses on checking that there are no misspecifications system. On the demand side, this is a check of the input volumes (flow rates) and/or the O-D flow patterns. On the supply side, it is checks of the physical network, its links and nodes, its geometric configuration, the speed limits, speed restrictions, vertical and geometric alignments; operational controls like signal timings, ramp-metering controls, time-of-day use restrictions, procedures for changing the direction of use (e.g., for high occupancy toll [HOT] facilities), and tolls (where applicable). This verification step needs to be completed for every data set used in the analysis. The data sets used for either calibration or validation must be checked.

Congruence is an extremely important concept here. For a data set to be congruent, the inputs and outputs must match up.[40] That is, the outputs must be caused by the inputs. While chapter 6 describes how to use care in preparing the data sets, the VCV process might identify incongruence or inconsistency in the data. The prime example is traffic inputs obtained at a different time than that for the measured outputs (performance), and possibly from a different source. On a surface arterial, the turning movement counts might have been obtained on different days; and at times different from when the delays and queue lengths were measured. This will present a significant calibration challenge. It is also possible that the traffic flows were obtained from a planning model while the performance was observed in the field. It is unreasonable to presume that the average demands portrayed by the planning model will necessarily produce the performance values observed on a given day. This will also lead to calibration trouble. Subtle changes in the values of the inputs may be needed to achieve a successful validation.

As with facility design work, it is valuable to have a second set of eyes review the data. Fresh reviews will often spot incongruences or missing elements. Some agencies have peer-consulting firms review each other's work. In a similar vein, having the team in one office of a consulting firm review the work of teams in other offices is a possibility. The objective is to minimize the likelihood the input data have missing items or incongruent information.

In summary, this verification activity can be described as follows:

1a) Ensure the simulation software can address all issues that arise in the scenarios to be examined.[41] Examples would be CAVs, HOT lanes, and tolling strategies. To do this, the analyst should check the software user manual, talk with colleagues, consult the Web, and

[40] The demands might be too high or too low, and the timing of those demands might be too early or too late.
[41] Hopefully, this was done in the scoping effort. See chapter 5. But before verification, calibration, and validation proceeds further, the ability of the software to address the settings of interest should be checked.

seek other sources of information. If all the issues can be addressed, then continue. Otherwise see if the user manual indicates how the issues can be addressed, see if technical support from the vendor can provide guidance, and/or get a colleague to suggest a solution strategy. If issues remain unresolved, then devise a workaround strategy and determine the implications of using it, or select a different software platform. Another option is to shift to a different type of simulation (e.g., mesoscopic or macroscopic). A third possibility is to create or find software that augments the simulation software's capabilities. This latter option is not likely to be feasible in most cases, but it is a possibility.

1b) Check that the input data sets for all the scenarios are clean, complete, and congruent (inputs and outputs match up). Check the physical network's description. For example, check that there are no dead ends. Make sure link connectors are in appropriate places. Make sure right and left-turn lanes are correctly connected to the upstream arcs. Check the operational controls. See if the signal controls at all intersections work appropriately. Make sure traffic stops when the signals are red. Make sure the signal timing is sensible and consistent with the field. Check the traffic demands. Make sure no O-D flows are incorrect (an order of magnitude too high or too low because zeros were added or missed). Make sure the changes in O-D flows from one time period to the next make sense. Do tests to see if traffic can get from each origin to all destinations. See if the paths chosen make sense. Check that the field observations of performance are congruent with the input data. At freeway bottlenecks, make sure the upstream demands (based on the O-D data and the paths that will be used) are able to produce the queues observed (not too high or too low).

8.6 CALIBRATION

Calibration is the second step in the VCV process. This is where the model's parameter values are adjusted so that the model produces outputs which closely match observed performance. It can also be the step where plausible load conditions are explored to see if the model can produce reasonable outputs for them as well.[42] Wunderlich et al. (2017) has stated that the objective of calibration is "to perform a systematic calibration of the error-free base model to reproduce observed throughput and other performance measures in load conditions of interest."

Calibration is likely to be a task of great interest to most analysts. It is challenging to describe clearly because people who work in this area have strong opinions about how calibration should be done; and those opinions vary. There is agreement, however that through calibration, the model's parameter values are adjusted so that the model's predictions match the observed outputs (to the maximum degree possible).

Wunderlich et al. (2018) stress the following points about calibration:

- Calibrate selectively, only for key performance measures.

[42] This latter comment is particularly important because the calibration effort is endeavoring to prepare the model to predict the performance of systems that do not yet exist. Hence, the model needs to be tested for a wide variety of conditions, not just one or a few situations for which field data are presently available. It is important that the model's predictive capabilities are robust.

- Use performance measures based on good observed data.
- Use the same model variant for each travel condition.
- Use a representative day rather than a synthetic day that combines multiple days.
- Focus on bottleneck dynamics as well time variant performance measures.

Calibration can be viewed as having four substeps. Substep 2a) involves assembling the data sets that will be used for calibration. An important aspect of the task is to ensure that the inputs and outputs in each of these data sets are consistent or congruent. This means the model should be able to produce the observed outputs based on the inputs if the model is correctly calibrated. Substep 2b) focuses on ensuring that individual system elements are correctly calibrated. That is, if they were to operate in isolation, they should produce outputs that are consistent with fundamental traffic engineering ideas. A good example would be that every basic freeway section in the model should have a capacity (maximum throughput) that is consistent with what the analyst would expect for each facility. The same thought pertains to freeway weaving sections, signalized intersections and un-signalized intersections. If this cannot be demonstrated, the integrated model will not be able to produce defensible system outputs. Substep 2c) checks the calibration of subsystems within the model (an entire arterial, an entire freeway section, short sections of freeway) to ensure the facility interactions are correctly captured. Finally, substep 2d) ensures the system, taken in its entirety, produces defensible outputs. (If the first three substeps are completed successfully, substep 2d) should be perfunctory.)

Once data cleanliness and congruence are assured by completing substep 2a), then step 2b) of the calibration effort can commence. This involves fine-tuning the model parameter values so the output data set values are produced based on the inputs. In a sense, the process is simple, like a regression analysis: adjust the parameter values, given the inputs, so that the simulation outputs match the observed outputs to the maximum extent possible. However, given the number of parameters that can be varied, from element-specific values to universal constants, the complexity is large. The challenges are at least twofold: 1) to not get lost and 2) to not set a value for parameter A that completely confounds the rest of the model and leads to erroneous values for all other parameters. The strategy presented here is to work from the bottom-up as opposed to top-down. Make sure individual system components (elements) have parameter values that allow them to predict correct outputs based on congruent inputs (as though the components were operating individually); then repeat this process for collections of system components (elements) that interact, like the signals along an arterial; and finally, check the performance of the overall system, like its ability (time required) to recover from a peak-load condition.

A simple model of a single toll booth is a useful example. Assume the system has an approach lane, a toll booth, and an exit lane. Assume the objective is to calibrate the service time distribution for the toll booth. Further, assume prior research suggests the service time can be represented as a uniform distribution from a minimum service time up to a maximum. (The research also suggests that more complex distributions can be used, but a uniform distribution is adequate.) Assume the analyst has a calibration data set that includes an arrival rate (700 vehicles per hour [veh/hr]), a distribution of delays (see
Figure 125), and an estimate of the minimum service time (2 seconds).

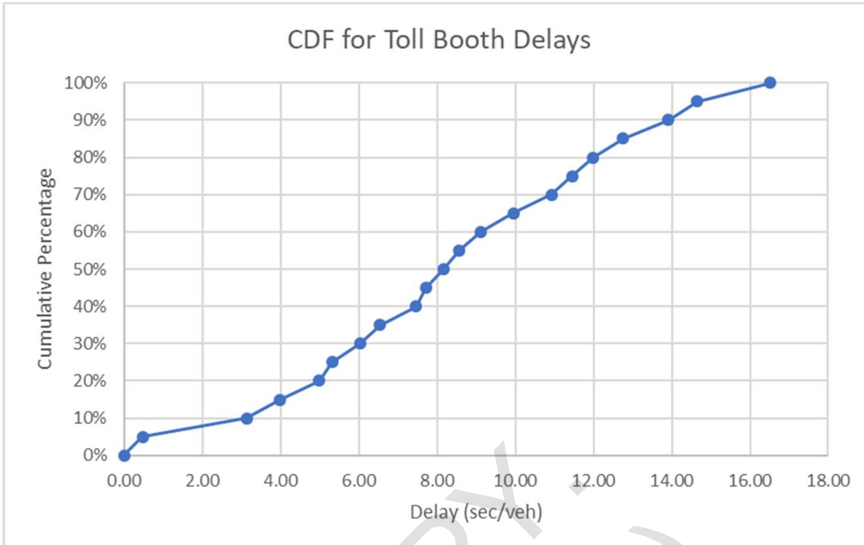


Figure 125. Chart. Hypothetical cumulative density function (CDF) for observed delays at a toll booth.

To do the calibration, the analyst sets up a simulation model in Microsoft® Excel. A screenshot of the simulator is shown in
Figure 126. A set of 100 vehicle headways is synthesized using the 700 veh/hr arrival rate; and an assumed shifted negative exponential distribution for the headways where the minimum headway is 1.5 seconds and the average is 5.14 seconds (the inverse of 700 veh/hr). (This procedure is akin to the ones used by microscopic traffic simulators.) Column B shows the vehicle number. Column C shows the random number sequence used to generate the headways (created and the pasted as numbers so the values do not change). Column D shows the headway based on random value 1 (rv1). Column E shows the arrival time in system based on summing the headway values. Column F shows the time when each vehicle enters the toll booth. It is either a) when it arrives or b) when the previous vehicle exits. Column G contains the set of random values (pasted as numbers) used to generate the toll booth service times. Column H shows the resulting service time based on the minimum service time shown in cell C4 and the range value shown in cell C5. (For example, if random value 2 [rv2] were 0.5, then the service time would be 3.55=2.0+0.5×3.11). Column I shows the time in system (departure minus arrival) and column J shows the delay (time in system minus the service time).

B	C	D	E	F	G	H	I	J	K
	700	= Flow Rate (veh/hr)							
	1.5	= Min Headway							
	5.14	= Avg Headway (sec)							
	2.00	= Min Service Time (sec)							
	3.11	= Range of Service Time (sec)							
# Veh	rv1	tHdwy	tArr	tEnt	rv2	tSvc	tExit	tSys	Delay
1	0.17	4.34	4.34	4.34	0.34	3.070756	7.41	3.07	0.00
2	0.37	3.05	7.40	7.41	0.15	2.465807	9.88	2.48	0.02
3	0.39	3.00	10.40	10.40	0.03	2.103299	12.50	2.10	0.00
4	0.94	1.60	12.00	12.50	0.98	5.062452	17.56	5.56	0.50
5	0.84	1.78	13.78	17.56	0.03	2.101304	19.66	5.88	3.78
6	0.75	1.97	15.75	19.66	0.22	2.6987	22.36	6.61	3.91
7	0.02	7.46	23.21	23.21	0.63	3.963775	27.17	3.96	0.00
8	0.54	2.48	25.69	27.17	0.79	4.457719	31.63	5.94	1.48
9	0.17	4.31	30.00	31.63	0.91	4.841864	36.47	6.47	1.63
10	0.03	7.17	37.17	37.17	0.58	3.810491	40.98	3.81	0.00

Figure 126. Screenshot. Hypothetical calibration worksheet.

The solver in Excel is then used to estimate the width of the uniform distribution. (The value shown in cell C5.) The solver minimizes the sum square deviations of the CDF values at every 5th percentile; comparing the values provided by the 100 synthesized vehicles (this a small number for such a purpose) and the observed distribution (as shown in Figure 123). The value shown in cell C5 is what results from assuming the arrival rate of 700 vehicles per hour (veh/hr) is correct.

Of course, in performing this analysis, the analyst must be careful to ensure that all delays are captured. In this case, the system starts empty and it should return to empty before data collection stops.

Figure 127 shows that this was true.

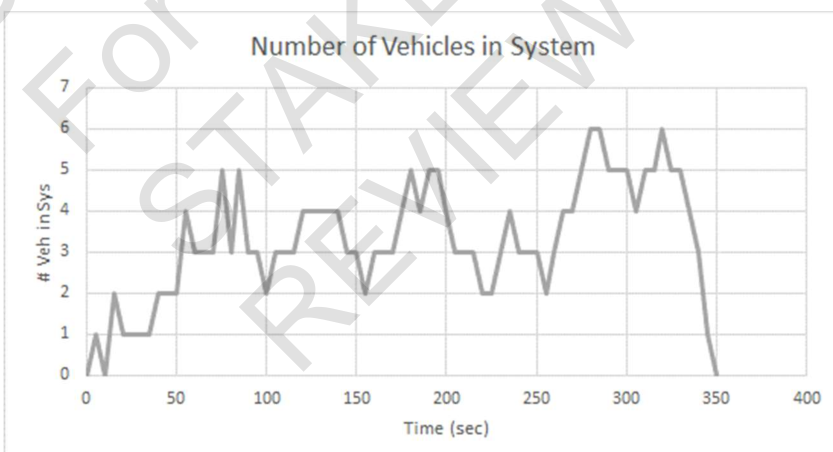


Figure 127. Chart. Temporal trend of the number of vehicles in system for the calibration analysis.

The twist, here, is that the 700 veh/hr is not correct. Unbeknown to the analyst, the actual flow rate was 600 veh/hr when the delays were recorded. The arrival rate at the toll booth is typically

700 veh/hr; so, it was reasonable to assume that value; but, on the day the data were collected that was not the flow rate.

If the correct arrival rate of 600 veh/hr had been used, the width of the uniform distribution would have been found to be 4.0 seconds, not 3.11 seconds. Of course, unless the analyst is aware of this error, there is nothing wrong with the calibration result. The 3.11 seconds is correct based on the 700 veh/hr employed.

However, if a simulation is preformed of different arrival rates based on these 2 values (the erroneous 3.11 seconds and the actual 4.0 seconds), the results are very telling. See Figure 128. The left-hand side shows the numerical values of average delay for flow rates from 100 to 1,000 veh/hr. The right-hand side plots the average delays. (Notice, importantly, that the average delay for the 600 veh/hr flow rate and the 600 veh/hr calibration of 8.23 seconds matches the average delay for the 700 veh/hr flow rate and the 700 veh/hr calibration of 8.25 seconds. This should be the case, because these were the calibration cases!)

	Average Delay	
Flow	@600	@700
100	0.23	0.14
200	0.52	0.35
300	0.97	0.59
400	1.77	0.99
500	3.68	1.74
600	8.23	3.38
700	26.21	8.25
800	41.67	21.85
900	53.75	33.93
1000	63.42	43.60

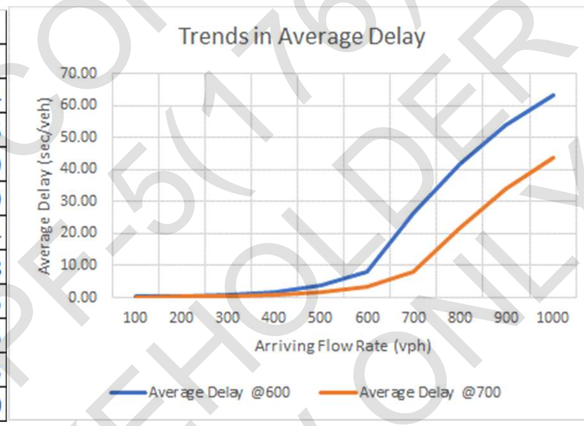


Figure 128. Screenshot. The calibration workbook sheet.

The punchline is this: If an incorrect arrival rate is used in the calibration (700 veh/hr), the simulator will then predict incorrect average delays for all the flow rates, including the condition under which the calibration was performed (26.21 seconds versus 8.25!). In this case, the average delays are significantly underestimated, by more than 30 percent at 1000 veh/hr. As an aside, even if there was noise in the observations, if at least two flow rates were used for the calibration, the likelihood these erroneous results would be created would be less, unless there is a systemic bias. This is the point about calibrating individual system elements, in this case a toll booth, so that their performance across a spectrum of load conditions is correct. This also reinforces the idea that it is useful to check the performance of system elements based on first principles and general traffic engineering-based expectations.

This element-focused testing might be followed by additional subsystem and system-focused testing-intended data sets that ensure the cascading effects of system overload conditions are correctly handled. This is equivalent to the limit-condition testing to which the car is subjected. This idea is not stressed here; but, maybe, practice will move in that direction.[43]

One thing that can be helpful in ensuring reasonable calibration is to use episodic, not steady state, conditions. The simulator must be able to deal with inputs that vary. Transportation systems do not typically have constant inputs. The traffic demands are overlapping, temporal transients. Moreover, the control systems, like adaptive control, respond differently depending on the spatiotemporal demand patterns. Condition-responsive tolls (affect path and departure time choice) are another example.

As
Figure 129 illustrates, the traffic demands tend to rise and then fall. (In the case of the toll booth example, they started at zero and returned to zero.) In Figure 129, the black line indicates the input flows of vehicles entering the system, wishing to be accommodated in accomplishing their trips. The red line is the flow of vehicles exiting the system, reaching their destinations. The temporal separation between the two shows the trend in the number of vehicles in the system, accomplishing their trips. If the system introduces no delays, the exiting graph will look just like the entering graph, simply displaced in time. If delays occur, the exiting graph will extend over a longer time and have an output profile that covers a larger span of time than the input; its duration from initial rise to final settling will be longer.

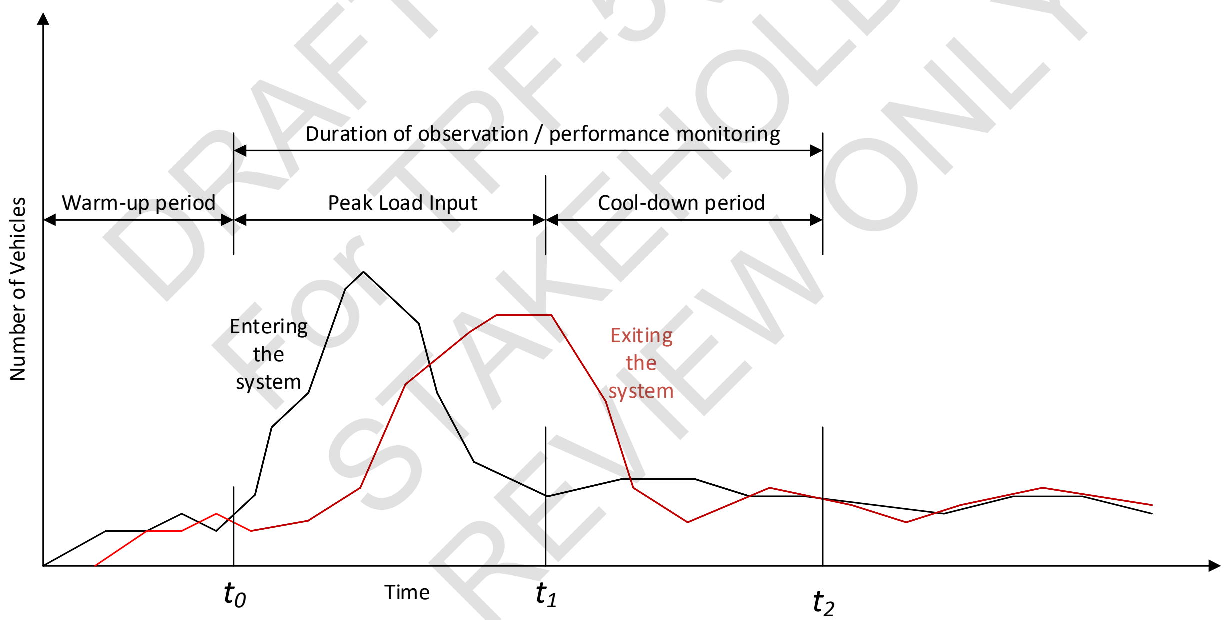
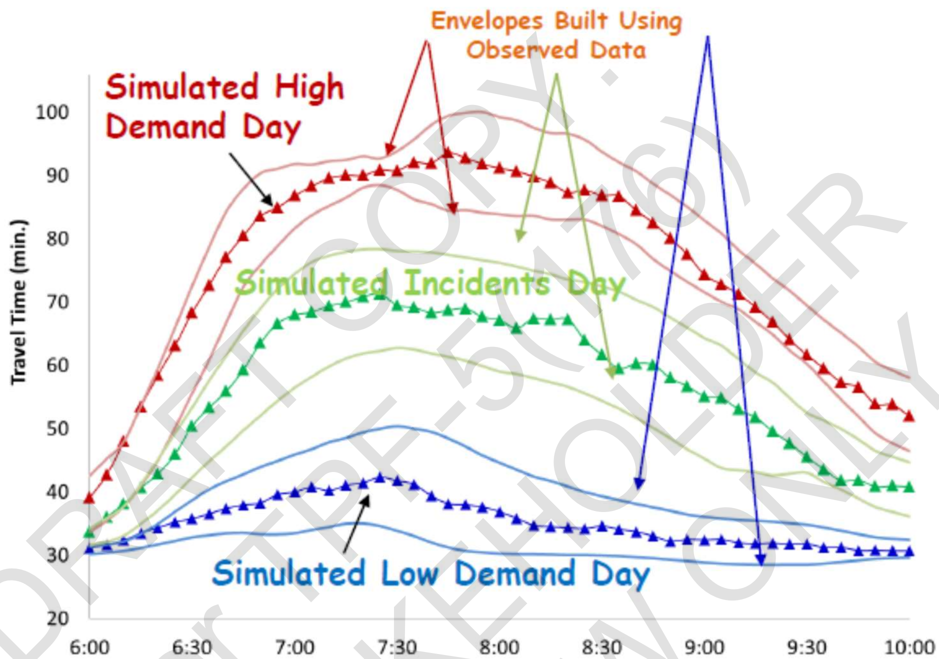


Figure 129. Chart. Simulation of transportation system episodic behavior.

[43] There is no excuse for creating a simulation model whose individual elements produce results inconsistent with traffic flow theory and traffic engineering principles. Bottlenecks, if examined in isolation, must produce queue dynamics consistent with their capacity and geometry. Signalized intersections, if individually studied, must produce queue dynamics and delays that are consistent with their geometry, saturation flow rates, and signal timings.

The calibration test data sets should encompass these entire transients. They should begin before the start of the rise in demand so that the system's accommodation of all vehicles is captured, and they should extend to, and potentially beyond, the time when all the peak period vehicles have been processed through the system (reached their destination) or when the system has returned to a nominal, uncongested condition.

Wunderlich et al. (2018) suggest selecting one or metrics to match during calibration and then develop acceptance envelopes for the simulator's estimates of those metrics.
Figure 130 shows envelopes for three clusters: 1) high demand day, 2) incident day, and 3) low demand day.



© Wunderlich et al. (2018).

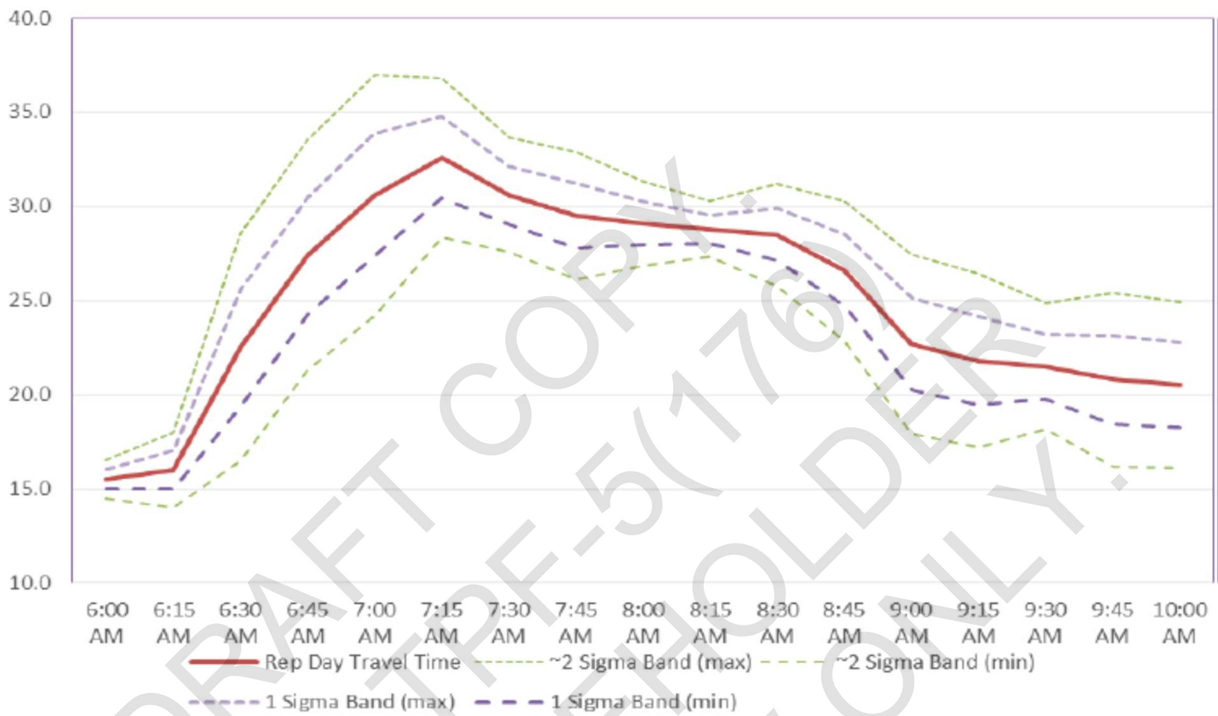
Figure 130. Chart. Acceptable performance envelopes for three clusters.

The envelopes, shown as solid lines, indicate the temporal patterns of the performance metric (travel time) across the facility at the 5th and 95th percentiles. (Using other metrics and percentiles is possible.) The focus of these percentile values is based on the representative day and the standard deviations in each time interval for the observed days in each cluster (including the representative day). The trend lines shown with triangles are the outputs from the simulator's predictions of the travel time based on the input data for the representative day.

For a single load condition, Figure 131 shows the 1- and 2-sigma bounds for the performance metric under study. Wunderlich et al. (2018) suggests using four criteria to determine if calibration has been achieved (for a given load condition):

1) 95 percent (19 of every 20 values) for the performance metric(s) of interest from the simulation of the representative day should lie inside the 2-sigma envelope.

2) 67 percent (2 of every 3) values for the performance metric(s) of interest from the simulation of the representative day should lie inside the 1-sigma envelope.
3) The average absolute error (AAE) from the simulated results for the representative day over all time intervals should be less than or equal to a bounded average error (BDAE) threshold.
4) The AAE from the simulated results for the representative day over all time intervals should be less than or equal to one-third of the BDAE threshold.



© Wunderlich et al. (2018).

Figure 131. Chart. One- and two-sigma bands for the temporal trend of a performance metric.

The equations involved are:

$$BDAE\ Threshold = \frac{\sum_{i \neq r} \sum_{t} \frac{|c_r(t) - c_i(t)|}{N_T}}{N_{cluster} - 1} \quad (8.1)$$

where $c_r(t)$ is the observed value on the representative day during time interval t, $c_i(t)$ is the observed value for day i (a different day) during time interval t, N_T is the number of time intervals and $N_{cluster}$ is the number of observed days in the cluster.

The AAE is given by:

$$AAE = \frac{\sum_{t} \left| c_r(t) - \tilde{c}_r(t) \right|}{N_T} \tag{8.2}$$

where $c_r(t)$ and N_T are defined as before and $\tilde{c}_r(t)$ is the simulated value of the observed value on the representative day during time interval.

Examples of criteria (1) and (2) seem useful to present.
Figure 132 shows the temporal trends in a performance metric for a given load condition. The dotted lines show the two-sigma bands. The green diamonds show the location of the observed trend for the performance metric for the representative day. The orange triangles show the locus of the simulated results. Seventeen data points are shown. Only one, at about 8 a.m., lies outside the two-sigma bands. Wunderlich et al. (2018) seems to indicate that one outlier (out of an entire set is allowed if the total number of intervals is less than 20, but not more than 1). That means this simulation result passes the calibration test for this criterion.

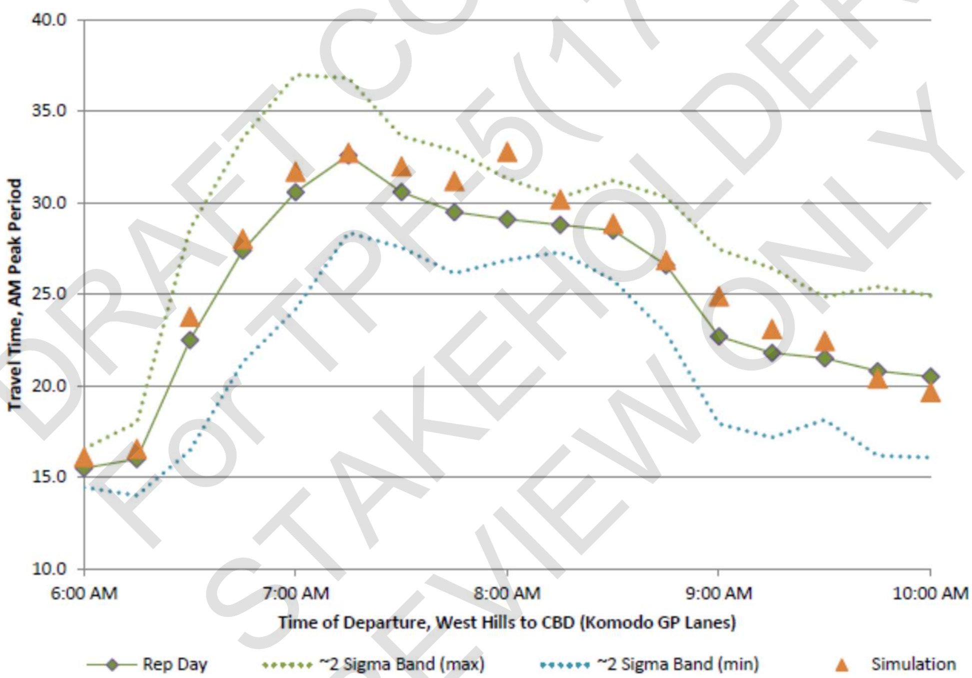


Figure 132. Chart. An illustration of the criterion 1 test.

Figure 133 shows the results for criterion 2. The dotted lines show the locus of the one-sigma bands. The green diamond and orange triangle plots are as before. By eye, it appears that two of the simulated values lie outside the one-sigma bounds. This means 15 out of 17 lie within the 1-sigma bounds, which is significantly more than 66 percent, so this criterion for acceptability has also been satisfied.

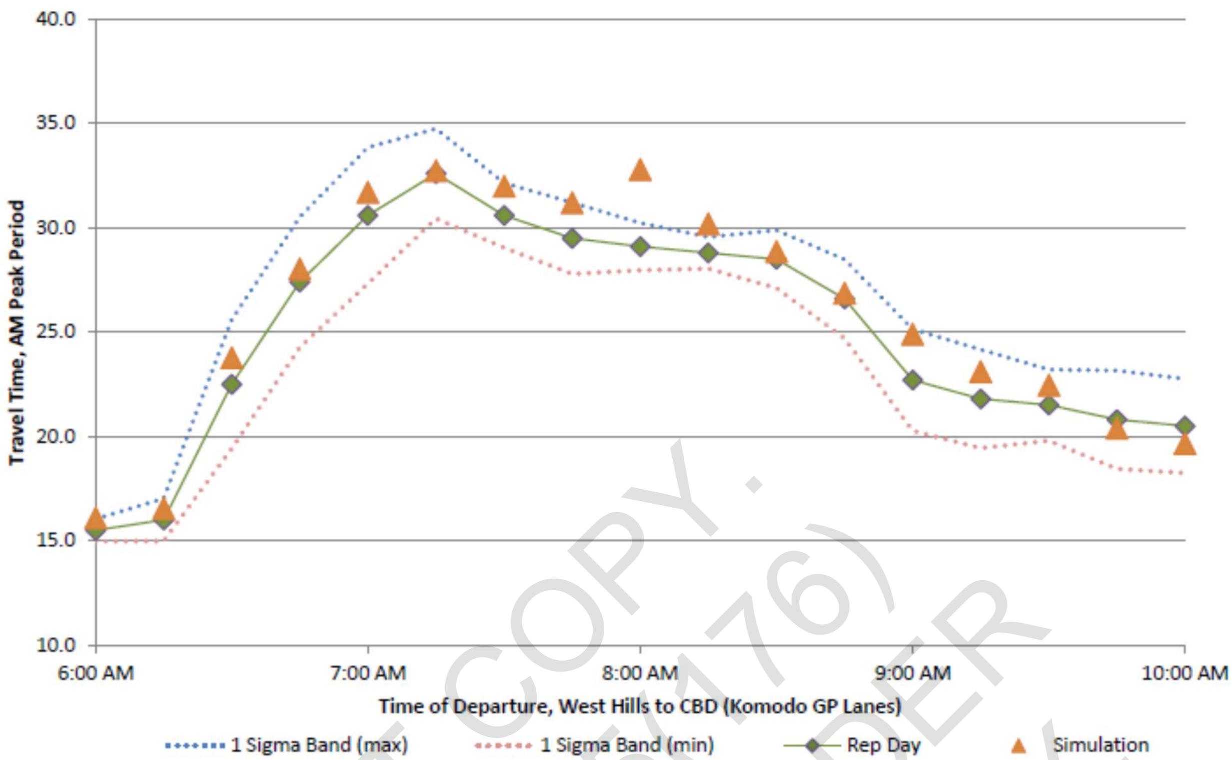


Figure 133. Chart. An illustration of the criterion 2 test.

Wunderlich et al. (2018) presents numerical illustrations of the analysis for criteria 3 and 4. Those evaluations are not duplicated here.

Compromise is to be expected during the calibration effort. Because of this, the analyst should rank the performance metrics so that the ones which are most important are studied most intensely. A strategy should exist for striking a balance between metrics that are in conflict. It may be that robustness is more important than error minimization. The model must provide good, defensible estimates of performance for all scenarios of interest, not just one. This helps ensure that design decisions are based on good predictions.[44]

---

[44] The notion of all scenarios needs to be applied in a thoughtful way. A model matching performance metrics well for all settings may be difficult to achieve. Even if it is possible to create such a model, it might be costly to do so. However, simulation models are often used to study tough problems where tipping points exist. System performance may change dramatically if alterations are made. Examples would be adding an additional through lane, narrowing the lanes in a work zone, creating an elevated highway section, adding a second lane on a flyover ramp, removing slip lanes, keeping a second left turn lane in a diverging diamond interchange (DDI) interchange, replacing a signal with a roundabout, or creating a new bridge crossing. These examples illustrate the point that the simulation model needs to be tested, calibrated, and validated in such a manner that the analyst is sure that the model will produce defensible results (reasonable numerical results) when it is used to examine these with/without questions. The dollar value of such decisions can run well into the millions or hundreds of millions. Hence, the analyst needs to be confident the model will do a good job of producing defensible results. The numerical meaning of good job and defensible results is a tough decision the project team and agency need to make. The main point is the model should not go awry in its predictions because it has not been adequately tested.

Figure 134 illustrates this idea of achieving an acceptable compromise.

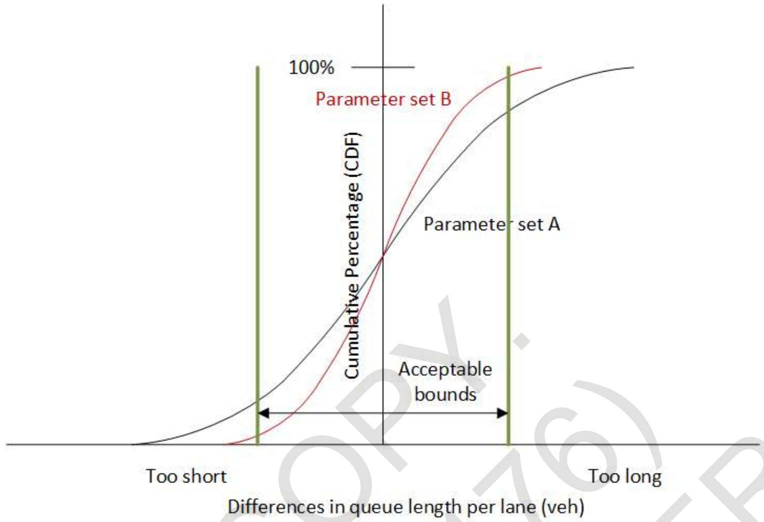


Figure 134. Chart. Getting the best calibration results.

To illustrate, assume that queue lengths (per lane) are of interest since they affect design decisions. Further assume the 95th percentile queue length is the performance measure[45] and the simulation model can provide this value, within a run or across a set of runs. Then, the analyst prepares a plot like Figure 135 that shows the difference between the model-estimated and field-measured 95th percentile queue lengths on the *x*-axis and the cumulative percentage on the *y*-axis for *all* significant queue lengths in the system being modeled. Differences that are negative indicate the model is underestimating the queue length; differences that are positive represent overestimates.

Quite appropriately, the analyst might use the sum of the squares of the deviations as the overall performance metric. However, examining the differences at individual percentiles is also important. Notice in Figure 134 that the mean queue lengths are the same (among all the locations where queues form), but the distributions are different. The differences based on parameter set A has a wider range of values than those for parameter set B; and, for set B, enough of the differences fall within the acceptance threshold that this alternate parameter set can be accepted. The distributions are providing information that the averages do not capture.

Calibration can, in principle, be done simultaneously or sequentially, provided things are set up appropriately in each case. Examples of both can be found in the literature. In the simultaneous case, a gradient-like search is conducted, with the objective of minimizing the differences between the observed and simulated (estimated) output (performance) values. The simultaneous approach has the advantage that the selection of parameter values is done all at the same time

---

[45] For example, the 95th percentile queue length (per lane) at freeway bottlenecks, on freeway ramps, or at strategically important intersections, because these queue lengths affect major design decisions.

and the objective function, that captures the integrated effects of all the parameter value choices, ensures that the best overall fit is achieved.

Sequential calibration is suggested here because it is somewhat more practical, it is a little easier to understand and describe to a layperson, it breaks down the process into steps, it reduces the scope of parameters that are calibrated at each stage, and it appeals to the cause-and-effect logic that underpins engineering-based thinking. Simultaneous calibration can be used for final fine-tuning, where the interactions among the model parameters are reviewed.

The sequential process is inherently iterative. It involves successively and thoughtfully refining the values of the model parameters so that the observed field performance is matched satisfactorily. Stated in an it-won't-happen sense, it is unreasonable to think the best set of values will be obtained in a single pass. Even if a solver is used, it employs a search process that involves iterations to identify the best set of values. In addition, it is best to adjust the parameter values slowly and carefully so that the best combination of matches is obtained for all the metrics of interest.

Sequential calibration can be done either top down or bottom up. In the top-down case, parameter values are set modelwide first (like driver behavior parameter values) and then problem spots are identified where further fine-tuning is required. This seems like a good idea, but a shortcoming is that the global parameter values may be offsetting underestimates of performance in some places with overestimates in others; and, neither one may be bad enough to be visible as a problem.

A bottom-up approach is suggested here because of this. The idea is that simulation parameters that are local in scope should be calibrated first, even if they pertain to many facilities and/or locations. For example, the parameters that describe driver behavior can be set for the normal freeway conditions so that every freeway link produces defensible results in terms of speed, flow, and density. Then, those values can be customized for the links where special geometric conditions, etc., pertain. (This is an illustration of the test track ideas for the car example presented earlier.) A range of demand flow rates should be examined, not just one or a few, including flow rates up to and including capacity, to ensure that the model's predictions of behavior align with observed and anticipated behavior. Doing this helps ensure that misspecified parameters do not confound the calibration results by creating erroneous predictions. The parameter values that are more global in nature, such as the ones that pertain to path choice, can be calibrated later.

For every setting that needs a specific set of parameter values, the analyst needs congruent input and output (performance) data. For example, on a freeway, the analyst must know the upstream flow rate as well as the discharge rate to be able to predict the average delay and the length of the queue. Similarly, for a signalized intersection, the turning movement demands must be known as well as the saturation flow rates and the seconds of green for each observation (e.g., seconds of green in a 5-minute interval). Otherwise, it will be impossible to calibrate the parameters. The causal conditions (input data) must be inconsistent with the observed performance.[46]

[46] In fact, it may be that many calibration dilemmas are caused by this lack of congruence.

Opinions vary about what parameters to include in the calibration effort. Most calibration analyses include the parameters used to predict car following, lane changing, and path choice. Some include the O-D flows as well, especially for mesoscopic and macroscopic models. It is suggested here to include the O-D flows for microscopic models as well.[47,48] Chu et al. (2004) suggests four steps for the calibration effort: (1) driving behavior model parameters, (2) route choice model parameters, (3) the traffic demands,[49] and (4) overall model fine-tuning. Toledo et al. (2003) suggests a similar sequence.[50] Using sequence is urged here as well. Effectively, the parameters focused on link-level phenomena are calibrated first, like those dealing with car following, lane changing, queue discharge, etc.; then route choice, like the parameters in probabilistic path choice models; then the traffic demands, like the spatial and temporal variations in the O-D flows, to ensure that performance metrics such as delays, queue lengths, and travel times can be predicted based on the input flows; and then an overall fine-tuning. This helps ensure that lower level issues are resolved before system-level problems are addressed.[51] The model will then be prepared to predict appropriate cause-and-effect relationships at the facility level (e.g., at individual freeway bottlenecks or signalized intersections). And system-scale issues will not be caused by lower scale miscalculations.[52]

Another option is to adjust the parameters one at a time, but this is not recommended. It only partly explores the cause-and-effect relationships. The interactions between parameters are missed. Calibration procedures that examine groups of parameters simultaneously, where they all influence a specific aspect of the model's operation, is recommended. Global sensitivity analysis is a strategy presently being investigated by researchers. But it has not yet reached the point where it is ready for practitioner use.

Fine-tuning the traffic demands is an activity that is sometimes, but not always, included in the VCV effort. The recommendation is that it should be included. The traffic flows have a profound effect on the performance metrics produced. See the footnote below about achieving congruence between the inputs and outputs. Hence, it is important to calibrate the traffic demands so that they are consistent with the performance metric values being used for assessment. Since this seems to be an uncommon thought, the process of fine-tuning the traffic demand values is reviewed in the appendix. In the future, it is hoped that simulation software products will have procedures that can automatically fine-tune the traffic demands. But, for the present, it is at least important to understand that it should be done and how it might be done. The appendix provides

---

[47] Incongruent O-D flows may make it impossible to predict the queues and delays that are observed.
[48] Toledo, T., H. Koutsopolous, A. Davol, M. Ben-Akiva, W. Burghout, I. Andreasson, T. Johansson, and C. Lundin, "Calibration and Validation of Microscopic Traffic Simulation Tools: Stockholm Case Study," *Transportation Research Record 1831*, pp. 65-75, 2003.
[49] Specifically, the temporal and spatial nature of the O-D flows.
[50] Chu, L, H. Liu, J-S Oh, and W. Recker, *A Calibration Procedure for Microscopic Traffic Simulation*, UCI-ITS-WP-04-2, University of Irvine, 2004.
[51] This ensures the cascading effects of facility-to-facility interactions do not confound the calibration effort.
[52] A possibility is that the system-level differences are due to incongruity between the traffic demands and the observed performance metric values. That is, the O-D flows being used as input cannot produce the performance values observed.

a mathematical description of one very flexible way to do the fine-tuning. Others exist. The analyst should select one that works well and fits within the project scope and budget.

A final observation is that, for many networks, the significant numerical contributions to the performance metrics only come from a few locations. It may be, for example, that 80 percent of the delays and queues only arise on 20 percent or less of the links in the network. This means, in calibration, that there are specific problem spots where the performance must be predicted correctly for a successful calibration to be achieved.[53]

A very useful suggestion about how to approach the calibration and validation task is provided by Toledo et al. (2003) and presented in Figure 135. First, the parameters for the individual submodels in the simulator (e.g., driving behavior and route choice models) are calibrated using disaggregate data. Link-level data include vehicle trajectories. Next, more aggregate data (e.g., time headways, speeds, flows) are used to tune more general parameters.

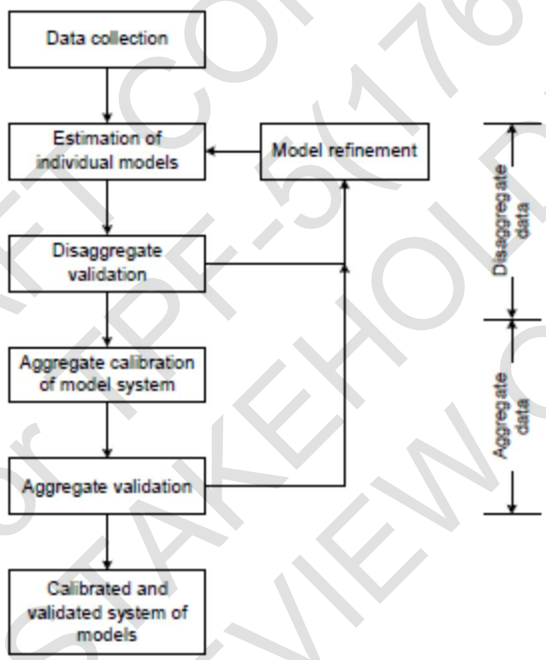


Figure 135. Diagram. Calibration procedure suggested by Toledo et al.

[53] It is possible that these problem spots may vary with the scenario examined. For example, they may be in different spots for the AM and PM peak condition. This means the analyst needs to focus on different locations during the calibration effort depending on the load condition being examined. But it also still means the number of problem locations that need to be studied intensely is limited.

To assess the quality of the calibration, an overall quality measure such as the root mean square error (*RMSE*) is suggested. The *RSME* is calculated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(m_n^{sim} - m_n^{obs}\right)^2} \qquad (8.3)$$

A companion quality measure is the mean error (*ME*). It indicates whether a bias exists (positive or negative) in the estimates overall. That equation is:

$$ME = \frac{1}{N}\sum_{n=1}^{N}\left(m_n^{sim} - m_n^{obs}\right) \qquad (8.4)$$

A possible metric on which to predicate these assessments is the average queue length (per lane) at bottleneck locations, measured every 5 or 15 minutes.[54] It is an observable value over short periods of time (at 5 minutes this average might be based on values from only 3–5 cycles) and relatively stable.[55]

Equation 8.1 can be embellished to combine two or more metrics, say queue length and travel time. One option is to compute individual $RSME_k$ values for each metric k and then use weights λ_k to combine those values:

$$RMSE = \sum_k \lambda_k RMSE_k \text{ where } RMSE_k = \sqrt{\frac{1}{N_k}\sum_{n=1}^{N_k}\left(m_{n,k}^{sim} - m_{n,k}^{obs}\right)^2} \qquad (8.5)$$

Another option is to combine the weighted metrics before taking the square root:

$$RMSE = \sqrt{\sum_k \lambda_k \frac{1}{N_k}\sum_{n=1}^{N_k}\left(m_{n,k}^{sim} - m_{n,k}^{obs}\right)^2} \qquad (8.6)$$

The same two enhancements can be applied to the *ME*.

A slightly different approach, consistent with Figure 135, juxtaposes the percentile values of the field and simulated performance metrics. Clearly, because of the episodic nature of the network dynamics, these CDFs might want to be compared across short epochs (every one-half hour) of

---

[54] This means the subscript *n* is referring to both a location and a point in time. That is, the value of *N* is computed by multiplying the number of observation locations by the number of time slices for which observations exist.
[55] There is a subtlety here which is important. When the PDFs and CDFs are examined, the focus is on the entire distribution of these values for a single simulation run or across a set of runs. In Equation 8.1, on the other hand, a direct comparison is being made between the average value recorded in the field for a given movement within a given time slice and the one predicted by the simulation model for a specific point in time. The likelihood that these time-varying, traffic-driven values will match at such a high level of granularity is low. If they do, the temporal dynamics of the simulation model are matching the field conditions extremely well.

the simulation. In addition, comparing them for the entire duration of the simulation has value as well.

Given these overarching evaluation metrics, and consistent with the observation that calibration is likely to be iterative, the analyst should expect that calibration will progress through successive refinement of the parameter values until satisfactory results are obtained for all the data sets (settings) examined. Figure 114 showed the iterative procedure involved in calibrating a given model. Adjustments are made to the parameter values, the differences between predicted and observed values for the performance metrics are reviewed, the goodness-of-match quality measure is computed (e.g., *RMSE*) and this result is compared with the best value computed so far. If the new result is better, it is kept along with the set of parameters that produced the result. If not, the previous set of parameter values is kept and a new parameter is selected for testing. The process continues until the goodness-of-match metric cannot be reduced further; or, until a satisfactory level of match has been achieved. (Or, until the money for calibration runs out.) Diminishing returns will occur, as would be expected.

Calibration in stages is recommended. This helps to untangle the interactions among the parameters and ensure that reasonable parameter values are identified.

The first stage should focus on the performance predictions for the links in the network (i.e., the network segments lying between junctions and/or bottleneck locations). Many research efforts have focused on doing this for freeway and arterial segments. The important objective is to ensure the model can correctly predict the behavior of the segment under a range of load conditions. This is equivalent to saying the simulation model must be able to correctly predict the relationships among speed (or travel rates), flow, and density that pertain to each link.[56] For microscopic models, since capacity is a derived value and not an input, it means the model must be able to correctly predict the maximum possible throughput as well (and the likelihood that various flow rates can be achieved without breakdowns occurring). To do this, a range of flow rates should be examined, not necessarily just the flow rates observed in the field that pertain to the conditions for which performance data are available. The model's ability to be robust in its predictive capabilities must be tested. Consistent with the earlier observation that most of the poor network performance arises at only a few problem spots, the analyst probably does not need to focus on all the links in the network. Default, or reasonable, parameter values may be adequate for most links. Instead, his or her effort should be concentrated on the links near these problematic locations.[57] Moreover, the most important flow rates to study are those near (or beyond) capacity. What is important is the performance of the link when the flow rate is high, delays are changing significantly, and queues are forming and dissipating. It is the response of the link to these high load conditions that is important. Among the model parameters that should be adjusted are those related to driver behavior (car following, lane changing, etc.). Links should

---

[56] Several problems can arise here if the parameters are not correctly adjusted. In microscopic simulation models, the capacity value may not be correct. Another is that spontaneous bottlenecks may arise that do not occur in the field. A third is the lane-changing behavior may be inconsistent with that observed in the field. For freeways, the latter is particularly true of segments upstream of major off-ramps or diverge locations. For arterials, it is upstream of major left turns.

[57] Segments upstream affect the dynamics of the queues created. Segments downstream affect the ability of those queues to dissipate.

also be checked to see if site-specific parameter values are needed to explain the speed, flow, and density relationships observed. This stage can end when all the problem spot links have been examined and it seems unlikely that further improvements in link performance predictions can be achieved by further refining the link-related parameter values.

The second stage should focus on the behavior of the conflict points where congestion regularly occurs and queues arise. These are the places that generate the significant delays, create large travel times, and cause unreliability. Again, much research has focused on identifying parameter values that enable these junction facilities to be simulated so their performance matches observed field data. The analyst should optimize the parameter values that predict the performance of these facilities. Examples are: 1) bottleneck locations in both the freeway and arterial networks caused by lane drops, significant grades, limited visibility, sun glare, or other causes, 2) interchanges in the freeway network where the weaving movements, merges, and diverges create conflicts, and 3) critical intersections in the arterial network. The objective is to ensure the model can provide defensible predictions of the performance of these facilities for the range of load conditions that might arise.[58] Queue dynamics and delays should be the major focus; also consider the likelihood that queues will form depending on the flow rate being accommodated. For freeway facilities, the parameter values to check are those that influence trajectory management, such as look-ahead distances and lane-changing behavior. For the freeway interchanges, it also means also checking the parameter values that influence performance on the ramps (if those locations were not checked during the first stage). For intersections, it means checking the parameters that control saturation flow rates as well as trajectory management in anticipation of turning movements (e.g., getting in the appropriate lane or lanes to make left and right turns).

The third stage should focus on checking the performance of subsystems, such as critical portions of the freeway and arterial networks (and in some cases, combinations of the two, such as freeways that have frontage roads). Here, ensuring the model correctly predicts the system-level effects is the focus. The system-level effects compound and amplify the queuing and delay caused by the critical links and junction points.[59] Much research has focused on calibrating simulation models so they will correctly predict the performance of linear freeways and arterials. The analyst should check the queue dynamics are correctly predicted for bottleneck locations, especially when system-level effects and interactions occur. For arterials, the thoughts are similar. The analyst should check for system-level effects, such as queue spillback and the influence of signal coordination on network travel time distributions. Making sure the subsystems perform to expectations will help eliminate the possibility that parameters have been misspecified and that overall system-level performance predictions will not be correct.[60]

[58] Without the influence of system-level effects so that the isolated performance of the junctions can be checked.
[59] If the predictions are wrong for the links and junction points when separately examined, these wrong predictions will be amplified and exaggerated by the system-level interactions. Hence, it is important that the facility-specific performance is checked first, in isolation, so the system-level interactions are correctly increasing the impact the facility-specific performance has on the overall system performance.
[60] Similarly, testing the behavior of a freeway segment for hypothetical demands will help identify whether the car-following and lane-changing parameters need to be adjusted or not.

The fourth and last stage should focus on overall system-level performance. Here, the emphasis should be on: 1) path choice and traffic loading and 2) system-level interactions among the subsystems. The analyst should check that queues arise in the right locations, and for the right traffic flow conditions. System-level effects that are known to exist should be checked. Travel time distributions should be checked for critical O-D pairs, using probe data to provide field observations. The queue dynamics at critical locations should be checked to see that queues arise when they should, they reach lengths that are consistent with field observations, and that they dissipate at rates that are consistent with the field observations. Where dynamic traffic assignment is part of the simulation framework, path choice decisions should be checked. The flow rates on critical links should be checked to ensure they are consistent with field observations.

A summary of the suggested calibration process is as follows:

1) Check the operation of individual system segments (links between junctions). Make sure their performance can be correctly predicted. The relationships among speed, flow, and density are critically important. The capacity is a part of this assessment. These checks should be conducted without the influence of system effects. That is, the assessment should pretend these links are operating in isolation. The links most important to check are the ones near bottleneck locations where queues and delays arise. Links both upstream and downstream of these locations should be checked. The ones upstream influence how fast queues grow. The ones downstream influence how fast they dissipate.
2) Check the operation of critical junctions in the network. For freeways, this means bottleneck locations such as lane drops, as well as critical off-ramps, weaving sections, and diverges. The most critical junctions to check are ones that produce the most significant queues and delays. For freeways, make sure the predictions of queue dynamics are appropriate. Make sure the queue grows at a defensible rate and dissipates at a defensible rate. Where capacity drops occur in conjunction with bottleneck formations, ensure those capacity drops are captured. For intersections, ensure the saturation flow rates are appropriate for all individual lanes and lane groups. As with the first set of checks, leave out the system-level effects. Check the performance of the junction for different traffic demands.
3) Check the operation of critical subsystems. Make sure the system-level interactions for these subsystems are correctly captured. Ensure the queue dynamics are reasonable. Check the distributions of travel times for key O-D pairs.
4) Check the predictions for the overall system. Use several traffic flow (O-D) patterns to test the model's ability to predict how the network will respond to the spectrum of load conditions. For several O-D pairs, check the decisions made by the path choice model if one is used. Check that paths chosen are reasonable. For O-D pairs that have paths with similar travel times, check that all competitive paths are used and the percentage shares of usage are reasonable. Check the resulting arc utilization percentages are reasonable.[61] This is also an opportunity to double-check the completeness and cleanliness of the input data sets. Unreasonable path choice decisions may be caused by errors in the network

[61] These are derived from the path use percentages. If multiple paths traverse the same segment, their path use percentages combine to indicate the segment utilization percentage.

data. Check that there are queues where queues would be expected. Make sure the queue lengths are reasonable. Ensure that all traffic gets processed. End the simulation once the system is empty or has returned to post-peak conditions. Fine-tune the parameter values if necessary. If the queue lengths and delays do not seem reasonable, it could be that the traffic demands are incongruous with the performance expected. Check the arc flow rates upstream of the congested locations and ensure they are reasonable.[62] They must be able to produce the expected queue dynamics. Adjust the traffic demand patterns (temporal and spatial O-D flows) before making substantial changes to other model parameter values.[63] Also, make sure the path choice patterns are reasonable for the main O-D pairs that should be involved in causing the anticipated congestion.

5) Complete the network-level calibration runs. Check the worst (most deviant) predictions. Ensure they lie within acceptable bounds. For the worst differences, go back and check the O-D flows can produce the expected performance. Also check the network description is correct and the operational controls have been correctly specified. At or near capacity, little changes can have a major impact on predicted performance.
6) Record and summarize the results.

8.7 VALIDATION

Validation is the process of ensuring the model provides an accurate representation of the way in which the real-world system behaves. It applies the calibrated model to additional settings and checks that the model can defensibly predict (indicate) how the system will perform for those conditions. Validation is a robustness check on the calibration results. Like calibration, it assumes that the simulation software can be used to analyze the situations of interest and that the setting-based input files are clean and correct.

Validation involves testing the calibrated model for other conditions not used in the calibration. The idea is like the difference between testing a car on a test track (calibration) and then driving it on real roads (validation). If the validation results are unacceptable, then further calibration is needed. The validation scenarios or settings should be important situations for which the system should perform well. Additional field observations of inputs and outputs are needed to do this checking, beyond those used in calibration. The validation checks should focus on robustness. The model should produce defensible results, for conditions not used in the calibration. If this is true, the VCV process is complete and analyses can commence. Otherwise, further VCV work should be performed. It could be that the software needs to be augmented with capabilities it does not have, the input data need to be further adjusted, or the parameter values need to be modified.

Validation has two substeps. Substep 3a) involves creating or assembling the data sets to use for validation. Following the guidance of Wunderlich et al. (2018), this should include the other observation days within a given cluster. The analyst must again ensure the validation test inputs and outputs are consistent or congruent. Substep 3b) involves applying the calibrated model to

[62] If there is no queue, the upstream demand may be too small. If there is a queue but none would be expected, the upstream demand may be too large.
[63] That is, the inputs are capable of producing the performance values observed.

the validation test data set inputs and checking to see if the outputs are correctly predicted. If not, the calibration step needs to be revisited. Most likely, additional limit condition situations need to be examined to ensure that the calibrated model can correctly predict system performance for those situations. There must be something about the conditions in the validation data sets that is pushing the model in directions for which it has not been correctly calibrated. Once the calibration has been refined, the validation data sets should be examined again to see if the predictions have been improved and the calibration is now acceptable.

It may be useful to use validation data sets, including outputs, for other operating conditions of interest from a robustness perspective. This might pertain to new, proposed physical configurations (1) and new set of operational controls (2). For the latter, of course, no performance data would exist, but performance data might exist for a different system that already has a similar physical configuration coupled with similar operational controls. And checking the predictions of the calibrated model for that alternate system might, in fact, be valuable. Checking the predictions for the future do-nothing alternative might even be useful, during calibration, to see what predictions would be created. That condition is comprised of the existing physical system (1) and its controls (2) combined with the predicted traffic flows and some default environmental condition. Calibration (and/or validation) must ensure that the simulation model can predict defensible results for all the combinations of conditions (1) through (4) that are of interest.

The validation step is extremely important. If the calibrated model is only checked against the calibration data sets, the validation assessment is circular. In this case, the same data used to calibrate the model are also used to validate its predictions. A much better plan is to use separate data sets to see if the calibrated model can produce defensible results for conditions that were not used for calibration. If it can, then it has been calibrated. It has been trained to provide useful predictions. If not, it should be calibrated further so that it has a predictive capability that is more robust.

It may be valuable to experiment with the classification of the data sets as being used for calibration and validation.
Figure 136 illustrates this thought.[64] It may be that one classification scheme results in an easy calibration and a robust validation, while another makes it more challenging. For example, if the calibration data set is only the morning (AM) peak and normal weather conditions, the calibrated model may not be prepared to deal with the evening (PM) peak, adverse weather conditions, new geometric designs, other traffic demands (e.g., holiday conditions), and other new features found in the future conditions. If these scenarios are only in the validation data set, the calibrated model may have difficulty producing defensible results.

[64] The complete set of field data sets can be subdivided into those used for calibration and validation in a wide variety of ways. It pays to be very thoughtful about how they are divided so that the best possible calibration results are obtained. A device manufacturer might very well deem the calibration data sets are ones obtained in the lab or in very controlled conditions where the outputs can be readily predicted based on the inputs. The validation data sets, on the other hand, might come from field applications where the inputs are not as well controlled.

Clearly, the data sets used for calibration should not be used for validation and vice versa. Having said this, the calibration data sets must be chosen carefully. They must span the conditions the system will see, so the model calibration has been checked for all the settings to which it will be applied.

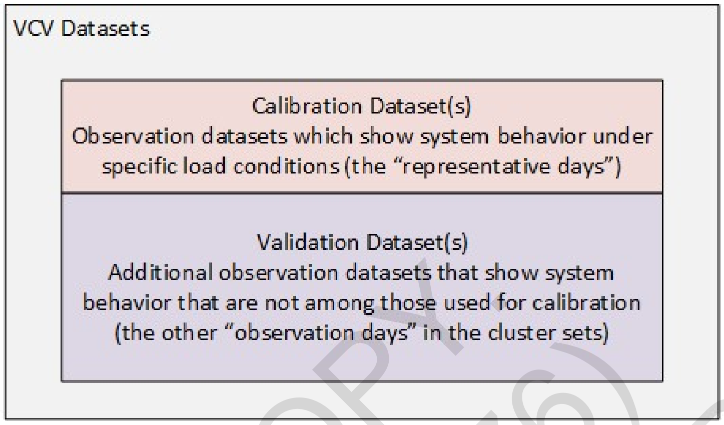


Figure 136. Illustration. Calibration and validation data sets.

The validation data sets, then, are used to: 1) double-check the model’s parameter values work well for all settings and 2) the model is robust in its predictive capabilities.[65]

The other load conditions referenced in
Figure 136 might also refer to different scenarios, such as PM peak volumes for a model focused on the AM peak. It could also refer to nonrepresentative days within a cluster. These are observational other than the representative day from a cluster analysis (Chapter 5.). It might also mean off-peak hours. If the analyst is trying to accurately replicate an AM commuter peak on a section of freeway that sees congestion every nonholiday weekday, even without any incidents, then responses to incidents might be considered; or performance on holiday weekdays; or performance when public schools are not in session; or off-peak hours where people don't drive as aggressively; or the PM peak, where the primary flow is in the opposite direction. In some instances, the load condition might be sporting events, or variations in anticipated traffic increases.

Another thought is that data from other networks may be useful in completing the VCV effort. This is illustrated in
Figure 137.

[65] Picture that manufacturers build standard intersections, arterials, freeway interchanges, etc. Transport agencies buy and install them—as might be the case for prefabricated bridges. The manufacturer of these facilities tests them as best it can, as auto manufacturers do, to ensure the facilities perform as advertised for various traffic load conditions. This is equivalent to the calibration step. Then, the performance of those facilities is observed in real-world conditions, and that performance validates the fine-tuning (i.e., calibration) that was done by the manufacturer. The validation is successful if the facility performs as advertised and it meets customer expectations.

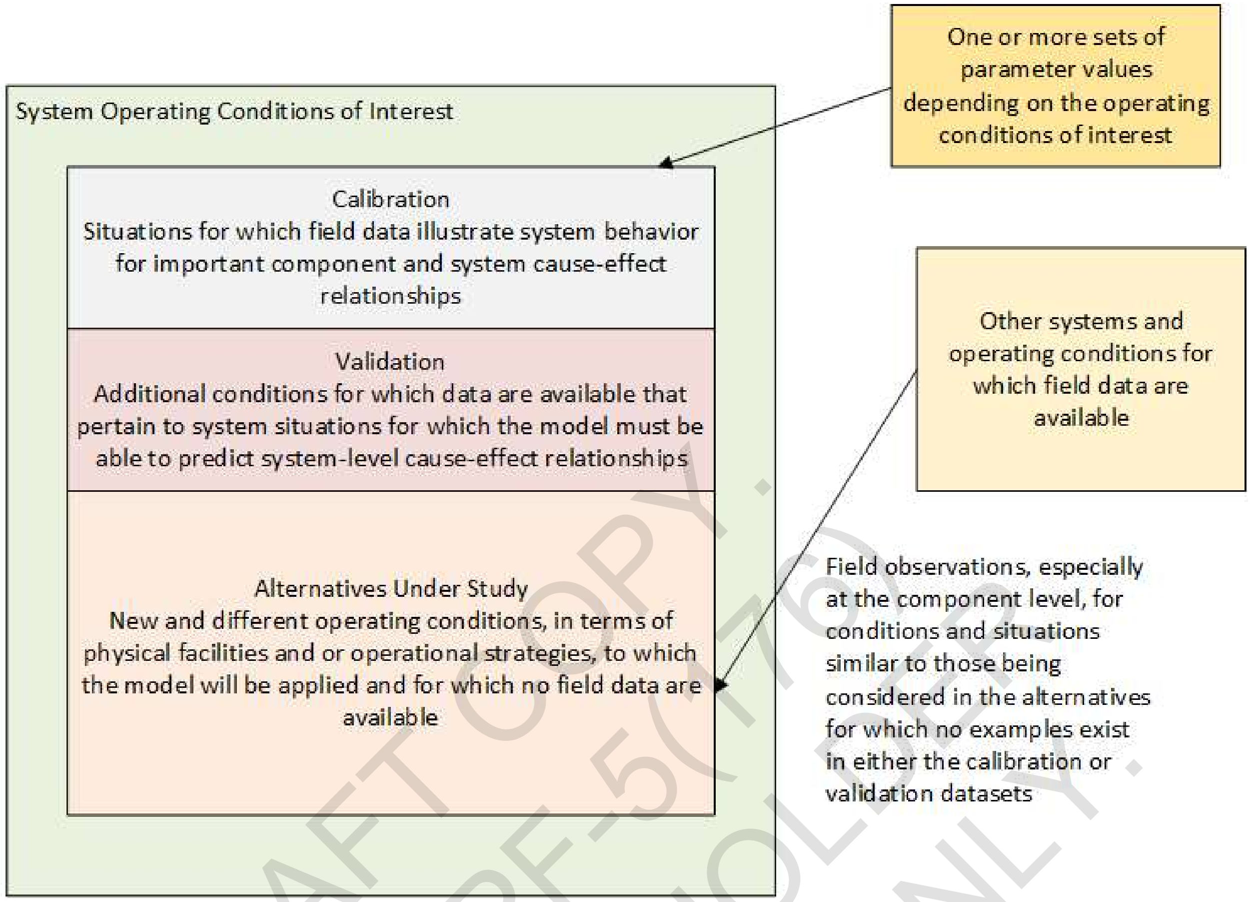


Figure 137. Illustration. Preparing the model for the alternatives of interest.

This would be true if there are future scenarios that involve physical facilities or technologies that do not exist in the present system. Two illustrations are a new type of interchange and the use of CAVs. Two more are the introduction of a centralized traffic control system and the use of HOT lanes. If new elements are part of the future system scenarios, then, as illustrated by Figure 137, the analyst should obtain data (inputs and outputs) from other systems where these facilities and/or vehicles do exist to test the model's ability to simulate their operation and performance.[66]

Excerpted data sets are fine. The entire systems do not have to be simulated. But enough of the system should be excerpted so that the model's ability to predict the performance of these systems can be assessed.[67]

---

[66] This underscores the value of creating an archival database of simulation analyses that have been conducted in the past, especially the field data for observed performance and behavior.

[67] Wisconsin Department of Transportation (WisDOT) has used data from other networks, though specifically to capture the effects of contraction in the duration of the peak (which is something that additional capacity could produce). Data from another corridor were used to see how much peak contraction might occur.

An illustration of this last point seems useful. It also illustrates the use of simulation to test the performance of a new technology. A recent research project explored the potential value of using connected vehicles (CV) to help eliminate spontaneous bottlenecks in a tunnel.[68] Field data showed that flow breakdown frequently occurred on the sag curve in the tunnel when the inbound flow rates were high. That is, there would be a spontaneous bottleneck inside the tunnel. Drivers were apparently more cautious and myopic in their driving behavior.[69]

It was thought that, in the future, CV technology could help reduce the likelihood these spontaneous bottlenecks would arise. The CV-based control could smooth the traffic flow by seeing further ahead, monitoring the flow rates, speeds, and densities, and using a speed-control feedback mechanism to reduce the variance in flow dynamics. It might be able to eliminate the breakdowns, and/or reduce the inbound flow rate so the queues are located upstream of the tunnel entrances. This is a good example of using simulation to explore the impacts of a technological change.

Two microscopic simulation models were created to explore these ideas. One was a code-augmented version of VISSIM. The other was a single-lane, car-following model completely contained within MATLAB®. The code augmentation in VISSIM allowed the researchers to emulate a CV-based control system that adjusted target speeds based on data collected about speeds and headways within the traffic stream. The MATLAB® code did the same, but based on a standalone, self-contained microscopic simulation representation of the system.

The simulation models showed that if the tunnel were not present, no breakdowns would occur. Flow rates up to and including the capacity of the facility (2,400 vehicles per hour per lane [veh/hr/ln]) could be accommodated. The impacts of the sag curve and reduced visibility in the tunnel would be removed.

The input demand employed was time-variant and trapezoidal in shape as shown in Figure 138. The flow rate started at 0, rose to values up to and including 2,400 veh/hr/lane across 10-minutes, stayed at those rates for 30 minutes, and then dropped back to 0 across another 10 minutes. The simulation continued until all vehicles had exited the system, which was typically at 60 minutes or less.

---

[68] Cetin, M. E. Beheshtitabar, R. Nezafat, G. List, and E. Willliams, Car-Following Behavior in a Tunnel and Impacts of Increasing Downstream Observability on Traffic Flow, Proceedings of the 97th Annual Meeting of the Transportation Research Board, on-line, Washington, DC, January 7-11, 2018.

[69] This means the car-following model (and parameters) applicable to the sag curve in the tunnel is different from that which pertains to the freeway sections upstream and downstream of the tunnel.

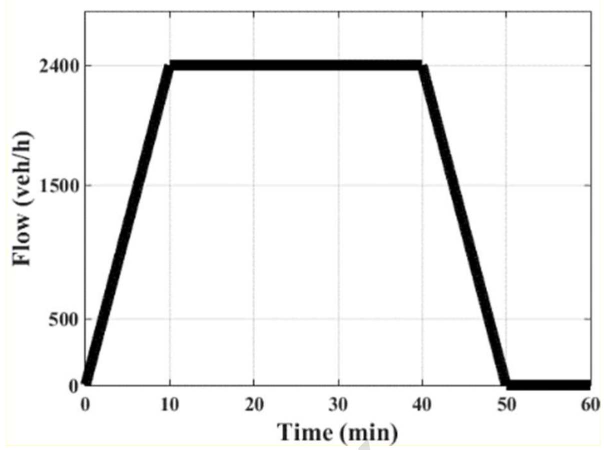
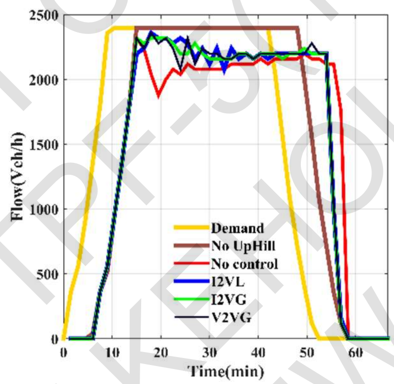
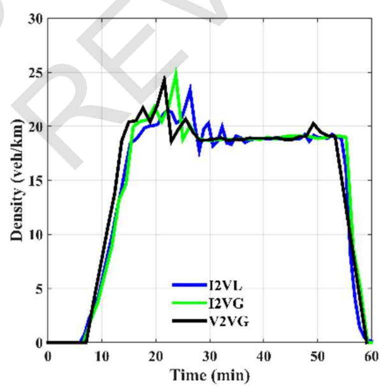


Figure 138. Chart. An illustrative, episodic travel demand input.

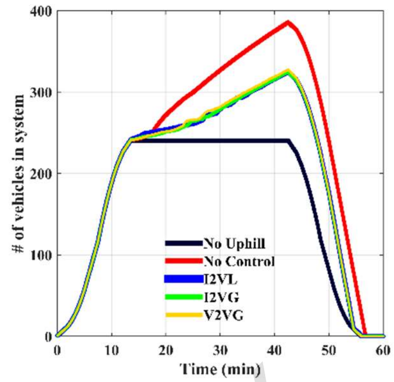
Figure 139 shows the findings. Subfigure A shows the input and output flow rates for various load conditions. Subfigure B shows the densities measured by various data collection strategies. And subfigure C shows the resulting number of vehicles in the system.



Subfigure A. Flows



Subfigure B. Densities



Subfigure C. Vehicles in system

Figure 139. Charts. Illustrative, episodic system performance.

The bottom line to the analysis was the CV technology did produce a clear benefit. In subfigure C, the black line shows the trend in the number of vehicles in the system when no sag curve is present. The red line shows the result when the sag curve exists, but no CV-based control is imposed. And the green, blue, and yellow lines show the impacts of employing various CV-based technologies. In the case of the red line, the number of vehicles rises to just shy of 400, because of the queue created by the bottleneck, and then drops back to zero once the bottleneck clears and there is no more input traffic. In the case of the CV-based controls, there is a clear impact on the number of vehicles in the system. The peak value drops from nearly 400 to just a little more than 300. These differences have a dramatic effect on the delays and the travel times.

Subfigure B shows that different CV control strategies (not discussed here) have an impact on the ability to monitor the density of traffic in the tunnel, which affects the quality of the control strategy.[70] But, as the graphs in subfigure C indicate, all three of the CV control options have similar impacts on system performance.

As with calibration, the validation data sets must span the range of conditions for which the calibrated model will be used. The data sets can be facility-level as well as system-level. The facility-level data sets can be very helpful in checking the calibration of lower level parameter values. They also allow the model to be exposed to settings that may exist in the future scenarios but are not included in the present system (e.g., a new interchange type, CAVs, HOT lanes).

Metrics different from those employed in calibration should be used for validation. In validation, the objective is to see if wrong or erroneous predictions are provided. It is not to fine-tune the parameter values. (If the validation data sets are used to tune the parameter values, then they effectively become part of the calibration data sets and new validation data sets become needed.) The metrics need to focus on how far the model's predictions are from the observed values. And

---

[70] These are described in the project report.

deviations should be expected. But the deviations should not be large. A useful metric would be a count of the number of instances where the model's predictions are outside specific bounds, like being more than a standard deviation away. If, for example, for all the validation scenarios, the model can predict the observed values with less than a 20 percent error, that may be acceptable. Or perhaps the count needs to be the number than are more than 30 percent different. Another possibility is a PDF that shows the absolute percentage deviations among all the metrics monitored. If more than 20 percent of these predicted metric values are greater than a standard deviation away from the observed value, then perhaps the calibration is not yet satisfactory. This idea is presented in
Figure 140. For the hypothetical situation portrayed, parameter set A does not have a high enough percentage of the SQM metrics that have percent differences within the acceptable bound. Parameter set B, on the other hand, does meet the criterion of having a sufficient percentage of the SQM metrics with absolute percent differences less than the acceptable bound.

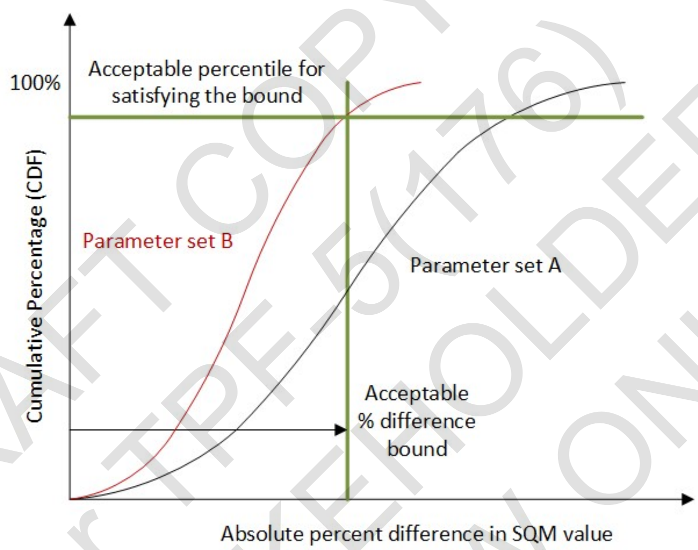


Figure 140. Chart. Distribution of absolute differences in simulation quality metrics (SQM) values.

Clearly, this is a robustness-focused metric. It checks the model does not produce predictions of performance too significantly different from the observed values. A step-by-step representation of this validation process is as follows:

1) Check among the validation data sets to see if there are any isolated facility tests that would be useful to conduct before focusing on the system-level scenario data sets. Examples of this might be new types of facility that do not appear in the calibration data sets, new operational controls, or new O-D pairs (e.g., a new stadium or shopping center). For this list of new conditions, conduct limited scope assessments focused on the issues that might arise due to the introduction of these new system elements. Check the model's predictions are reasonable for these new settings. If not, go back to the calibration task and make further adjustments. Or replace some of the calibration data sets with ones that are part of the validation set so these untested conditions can be used in the calibration effort.

2) Begin doing network-level simulation runs with the validation data sets. Fine-tune the parameter values if necessary (but based on the calibration data sets, not the calibration and validation data sets). Be careful not to adjust them too much. If the queue lengths and delays do not seem reasonable, see if the traffic demands are incongruous with the arc-level traffic flows that would produce the observed behavior. One thing to check is that arc flow rates upstream of the congested locations are reasonable.[71] They must be able to produce the queue dynamics expected. Adjust the traffic demand patterns (temporal and spatial O-D flows) before making substantial changes to other model parameter values.[72] Also, make sure the path choice patterns are reasonable for the main O-D pairs that should be involved in causing the anticipated congestion.

Complete the network-level calibration runs. Check the worst (most deviant) predictions. Ensure they lie within acceptable bounds. For the worst differences, go back and check the O-D flows can produce the expected performance. Also check the network description is correct and the operational controls have been correctly specified. At or near capacity, little changes can have a major impact on predicted performance.

8.8 STEP-BY-STEP PROCESS

This section represents step-by-step guidance previously provided in the individual sections about verification, calibration, and validation.

Verification

1) Ensure the simulation software can address all issues that arise in the scenarios to be examined.[73] Examples would be CAVs, HOT lanes, and tolling strategies.
   a. Check the software user manual, talk with colleagues, consult the Web, and seek other sources of information.
   b. If all issues can be addressed, then continue.
   c. Otherwise, see if the user manual indicates how the issues can be addressed, see if technical support from the vendor can provide guidance, and/or get a colleague to suggest a solution strategy.
   d. If issues remain unresolved, devise a workaround strategy and determine the implications of using it or select a different software platform. Another option is to shift to a different type of simulation (e.g., mesoscopic or macroscopic). A third possibility is to create or find software that augments the simulation software's capabilities. This latter option is not likely to be feasible in most cases, but it is a possibility.
2) Check the data sets for all the calibration and validation observations are clean, complete, and congruent (inputs and outputs match up).

[71] If there is no queue, the upstream demand may be too small. If there is a queue but none would be expected, the upstream demand may be too large.
[72] That is, the inputs can produce the performance values observed.
[73] Hopefully, this was done in the scoping effort. See chapter 5. But before the verification, calibration, and validation process proceeds further, the ability of the software to address the settings of interest should be checked.

a. Check the physical network's description. For example, check there are no dead ends. Make sure link connectors are in appropriate places. Make sure right- and left-turn lanes are correctly connected to the upstream arcs.
b. Check the operational controls. See if the signal controls at all intersections work appropriately. Make sure traffic stops when signals are red. Make sure signal timing is sensible and consistent with the field.
c. Check traffic demands. Make sure no O-D flows are incorrect (an order of magnitude too high or too low because zeros were added or missed). Make sure the changes in O-D flows from one time period to the next make sense. Do tests to see if traffic can get from each origin to all destinations. See if the paths chosen make sense.
d. Check that the field observations of performance are congruent with input data. At freeway bottlenecks, make sure upstream demands (based on O-D data and paths that will be used) are able to produce the observed queues (not too high or too low).

Calibration

1) Check the operation of individual system segments (links between junctions). Make sure their performance can be predicted correctly. The relationships among speed, flow, and density are critically important. A capacity check should be part of this assessment. These checks should be conducted without the influence of system effects. That is, the assessment should pretend these links are operating in isolation. The most important links to check are those near bottleneck locations where queues and delays arise. Links both upstream and downstream of these locations should be checked. The ones upstream influence how fast queues grow. The ones downstream influence how fast they dissipate.
2) Check the operation of critical junctions in the network. For freeways, this means bottleneck locations such as lane drops, as well as critical off-ramps, weaving sections, and diverges. The most critical ones to check are those that produce the most significant queues and delays. For freeways, make sure predictions of queue dynamics are appropriate. Make sure the queue grows at a defensible rate and dissipates at a defensible rate. Where capacity drops occur in conjunction with bottleneck formations, ensure those capacity drops are captured. For intersections, ensure the saturation flow rates are appropriate for all individual lanes and lane groups. As with the first set of checks, leave out the system-level effects. Check the performance of the junction for different traffic demands.
3) Check the operation of critical subsystems. Make sure the system-level interactions for these subsystems are correctly captured. Ensure the queue dynamics are reasonable. Check the distributions of travel times for key O-D pairs.
4) Check the predictions for the overall system. Use several traffic flow (O-D) patterns to test the model's ability to predict how the network will respond to the spectrum of load conditions. For several O-D pairs, check the decisions made by the path choice model if one is used. Check that the paths chosen are reasonable. For O-D pairs that have paths with similar travel times, check that all competitive paths are used and the percentage shares of usage are reasonable. Check that resulting arc utilization percentages are

reasonable.[74] This is also an opportunity to double-check the completeness and cleanliness of the input data sets. Unreasonable path choice decisions may be caused by errors in the network data. Check there are queues where queues would be expected. Make sure the queue lengths are reasonable. Ensure all the traffic gets processed. End the simulation once the system is empty or the flows have returned to post-peak values. Fine-tune the parameter values if necessary. If the queue lengths and delays do not seem reasonable, it could be that traffic demands are incongruous with the performance expected. Check the arc flow rates upstream of the congested locations and ensure they are reasonable.[75] They must be able to produce the expected queue dynamics. Adjust the traffic demand patterns (temporal and spatial O-D flows) before making substantial changes to other model parameter values.[76] Also, make sure the path choice patterns are reasonable for the main O-D pairs that should be involved in causing the anticipated congestion.

5) Complete the network-level calibration runs. Follow the ideas suggested by Wunderlich et al. (2018). Check the worst (most deviant) predictions. Ensure they lie within acceptable bounds. For the worst differences, go back and check if the O-D flows might be producing these deviations. Adjust as necessary. Also check the network description is correct and operational controls have been correctly specified. At or near capacity, little changes can have a major impact on the predicted performance.
6) Record and summarize the results.

Validation

1) Check among the validation data sets to see if there are any isolated facility tests that would be useful to conduct before focusing on the system-level scenario data sets. Examples might be new types of facility that do not appear in the calibration data sets, new operational controls, or new O-D pairs (e.g., a new stadium or shopping center). For this list of new conditions, conduct limited scope assessments focused on the issues that might arise due to the introduction of these new system elements. Check the model's predictions are reasonable for these new settings. If not, go back to the calibration task and make further adjustments. Or replace some of the calibration data sets with ones that are part of the validation set so these untested conditions can be used in the calibration effort.
2) Begin doing network-level simulation runs with the validation data sets. Fine-tune the parameter values if necessary (using the calibration data sets). Be careful not to adjust them too much. If the queue lengths and delays do not seem reasonable, see if the traffic demands are incongruous with arc-level traffic flows that would produce the observed behavior. One thing to check is that arc flow rates upstream of the congested locations are

[74] These are derived from the path use percentages. If multiple paths traverse the same segment, their path use percentages combine to indicate the segment utilization percentage.
[75] If there is no queue, the upstream demand may be too small. If there is a queue but none would be expected, the upstream demand may be too large.
[76] That is, the inputs are capable (should produce) the performance values observed.

reasonable.[77] They must be able to produce the expected queue dynamics. Adjust the traffic demand patterns (temporal and spatial O-D flows) before making substantial changes to other model parameter values.[78] Also, make sure the path choice patterns are reasonable for the main O-D pairs that should be involved in causing the anticipated congestion.

3) Complete the network-level calibration runs. Check the worst (most deviant) predictions. Ensure they lie within acceptable bounds. For the worst differences, go back and check the O-D flows can produce the expected performance. Also check the network description is correct and operational controls have been correctly specified. At or near capacity, little changes can have major impacts on the predicted performance.

8.9 SUMMARY

This chapter has addressed the tasks of VCV as they pertain to transportation system simulation models. The VCV process is perceived as having four integrative, sequential, and iterative steps: 1) check the simulation software package can analyze the settings of interest, 2) ensure the input data correctly describe those settings, 3) adjust the simulation parameters so they have the best values for producing outputs consistent with field observations, and 4) check the model produces overall results that are defensible for the setting(s) of interest. Verification pertains to the first two of these; calibration, the third; and validation, the fourth. A diagram depicting the VCV process was shown in Figure 114. Several observations were made about the VCV process that are important to recognize if the effort is to be successful:

1) The overall calibration models have five components: 1) logic (if-then relationships and equations) that describes how the system is assumed to function, 2) software that implements that logic, 3) parameter values that affect the numeric results obtained from applying the logic, 4) data files that describe the load conditions of interest, and 5) additional data files that describe the performance observed.
2) Verification pertains to the first four of these components.
3) Calibration and validation pertain to the third, fourth, and fifth components.
4) The load conditions are combinations of traffic demands and an environmental condition.
5) The system comprises physical elements and the operational controls.
6) The load conditions are episodic.
7) During calibration and validation, multiple tests at different levels need to be performed.
8) The system-level data sets that indicate system performance should be divided into two sets. One set should be used for calibration, the other set for validation. Following Wunderlich et al. (2018) the representative days in each of the clusters should be used for calibration purposes. The remaining observation days in each cluster can be used for validation.
9) Data from other networks may be needed if there are future scenarios that involve physical facilities or technologies that do not exist in the present system.

[77] If there is no queue, the upstream demand may be too small. If there is a queue but none would be expected, the upstream demand may be too large.
[78] That is, the inputs are able to produce the performance values observed.

10) The VCV process involves compromise. It is highly unlikely the field observations of performance will be matched exactly. Maybe, if the software logic is perfect and the input data are congruous with the performance observations, the model might be able to match the field observations.

SQMs must be chosen to determine whether the simulation model’s predictions of performance are acceptable or not. These SQMs are two-tiered. Lower tier SQMs measure the performance of specific aspects of the simulation, such as behavior of freeway and arterial segments, bottlenecks and intersections. Upper tier SQMs combine these metrics to provide overall assessments of performance (such as the sum of the squared percent errors). Possible metrics for microscopic, mesoscopic, and macroscopic simulation models are reviewed in 8.4.

Verification is described as having two parts. The first part focuses on checking the ability of the software to simulate the system designs of interest, like HOT lanes, CAVs, new control strategies, and new types of facilities. The second part aims to check the completeness and correctness of input data. Even though this review is theoretically done in preparing the input data sets, this double-check is intended to identify items that might be missing or incongruences. Details of the verification step are described in 8.5.

Calibration is the process of adjusting the model’s parameter values so the simulation model can produce outputs that closely match the performance observed in the field (for load conditions that have been observed) or are plausible (for load conditions that have not been observed, either in the present or in the future). It is inherently iterative, and it is unreasonable to expect the calibration effort will result in a parameter set that allows the field observations to be matched exactly. Details of the calibration step are discussed in 8.6.

Some of the calibration parameters are local in scope and some global. Those that affect local behavior should be calibrated based on detailed analyses of localized performance, such as flow on freeway segments and queue dynamics at the stopbar. Those that are global should be calibrated based on system-level behavior, such as path travel times and overall levels of congestion.

More than one set of parameter values may be needed, even for a given model setting. A single set may not be sufficient. Driver behavior, for example, is affected by geometric design, especially in tunnels, in work zones, and where the geometric design is substandard.

Chu et al. (2004) suggest four steps for the calibration effort: (1) driving behavior model parameters, (2) route choice model parameters, (3) the traffic demands,[79] and (4) overall model fine-tuning.[80] Toledo et al. (2003) suggests a similar sequence.[81] Using sequence is urged here as

---

[79] Specifically, the temporal and spatial nature of the O-D flows.

[80] Toledo, T., H. Koutsopolous, A. Davol, M. Ben-Akiva, W. Burghout, I. Andreasson, T. Johansson, and C. Lundin (2003). Calibration and Validation of Microscopic Traffic Simulation Tools: Stockholm Case Study, Transportation Research Record 1831, pp. 65-75.

[81] Chu, L, H. Liu, J-S Oh, and W. Recker (2004). A Calibration Procedure for Microscopic Traffic Simulation, UCI-ITS-WP-04-2, University of Irvine.

well. Effectively, the parameters focused on facility-level phenomena are calibrated first, like those dealing with car following, lane changing, queue discharge, etc.; then route choice, like the parameters in probabilistic path choice models; then the traffic demands, i.e., the spatial and temporal variations in the O-D flows, to ensure that performance metrics such as delays, queue lengths, and travel times can be predicted based on the input flows; and then an overall fine-tuning.

Validation focuses on checking the calibrated model. It takes the calibrated model and applies it to additional settings to see if the model can defensibly predict (indicate) how the system will perform for those conditions. It is effectively a robustness check on the calibration results. Section 5 provides a detailed discussion about the issues involved in validation.

Section 6 presents a step-by-step process that is suggested for the VCV tasks. A 19-step process is described. And for some steps, substeps are described. The objective is to present a clear process that analysts can follow to ensure the simulation model has been thoroughly tested before it is used to conduct the assessments.

Undoubtedly, over time, the best way to perform the VCV tasks will evolve. New techniques will be developed. New VCV procedures will be developed. Hopefully, in time, the VCV tasks will become more automated so that consistency can be established among VCV efforts that accompany projects. The challenge, as is apparent today, is to find a step-by-step procedure that ensures the best parameter values are identified, and that the VCV tasks proceed based on the most useful information that can be provided.

CHAPTER 9. DATA OUTPUT ANALYSIS AND SIMULATION

9.1 RUNNING SIMULATION MODELS

Once verification, validation, and calibration (VCV) are complete, alternatives can be created and the analyses can be conducted. A data set is created that describes the entire set of files the software package uses to represent the system and record the results of the analyses. This chapter describes that process.

Note: Throughout this chapter, data set is used to describe the entire set of files a software package may generate to represent a model.

9.2 MANAGING DATA SETS

For each alternative within a project, there will be multiple data sets (e.g., each time period over which the analysis is being performed is likely to have a unique simulation data set). In the text that follows, the words data set and model are used interchangeably. They both refer to an electronic input data file or files that describe the project in a given alternative and for a specific scenario. While the alternatives may consider unique geometric elements, there should be consistency between the alternatives of the geometric elements at the edges of each model. To ensure those details are identical across all of the simulation models, base models should be created for every operating regime examined. Once those models are verified, validated, and calibrated, numerical adjustments (e.g., volume, signal timing) can be made to study the alternatives of interest.

Naming Structure

For a project with numerous alternatives, the number of model data sets can become onerous to manage. Creating a condensed but readable naming structure is critical. The naming structure should include a segment for each input which varies from model to model. For instance, if there are three geometric alternatives, each with two time period analyses, the naming structure might be *ProjectRef_Alternative#_TimePeriod_Version*, with one data set named *50476_Alt1_AM_V1*.

Dates can be used instead of version numbers, although that may create a problem if multiple models are created in a single day.

It is best practice to create a file name that contains no spaces. Use hyphens, underscores, or other special characters to tie the name delimiters together. This is very important if a programing script will be used to analyze the resulting data.

File names of excessive length should be avoided especially if the file is stored within multiple subfolders on the computer. The file path may have a character limit imposed by the computer operating system.

Example:

Figure 141 is a set of files that had two inputs that varied: the volume scenario and the signal timing. The file naming structure is *ProjectRef_Volume_SignalTiming*.

It is important to note that while the naming structure does not need to be independently understandable to a bystander (the reader may not understand what *HLOff* or *3CritHoldOr* means), all internal team members should understand the structure. It should also be noted the file naming conventions below do not use the version control tracking recommended above.

Name	Date modified	Size
300_HLOff_2CritRampO.inpx	10/7/2015 2:35 PM	374 KB
300_HLOff_3CritHoldOr.inpx	10/5/2015 2:47 PM	375 KB
300_HLOff_4CritCCSM.inpx	1/5/2016 7:36 PM	375 KB
300_HLOn_2CritMainO.inpx	9/24/2015 7:36 PM	374 KB
300_HLOn_3CritHoldOr.inpx	10/5/2015 2:49 PM	375 KB
300_HLOn_4CritCCSM.inpx	1/5/2016 7:29 PM	375 KB
300_HROff_2CritMainO.inpx	9/24/2015 7:45 PM	374 KB
300_HROff_3CritHoldO.inpx	9/24/2015 7:49 PM	375 KB
300_HROff_4CritCCSM.inpx	1/5/2016 7:35 PM	375 KB
300_HThru_2CritMainO.inpx	9/24/2015 7:24 PM	374 KB
300_HThru_3CritHoldO.inpx	9/24/2015 8:55 PM	375 KB
300_HThru_4CritCCSM.inpx	1/5/2016 7:26 PM	375 KB
300_HTLOff_2CritMainO.inpx	11/17/2015 2:16 PM	375 KB
300_HTLOff_3CritHoldO.inpx	11/17/2015 3:59 PM	375 KB
300_HTLOff_4CritCCSM.inpx	1/5/2016 7:22 PM	375 KB
300_HTLOff_no.inpx	10/29/2015 12:51 PM	374 KB
300_HTLROff_2CritMainO.inpx	11/17/2015 2:16 PM	375 KB
300_HTLROff_3CritHoldO.inpx	11/17/2015 3:56 PM	375 KB
300_HTLROff_4CritCCSM.inpx	1/5/2016 7:21 PM	376 KB
300_HTROff_2CritMainO.inpx	11/17/2015 2:02 PM	375 KB

Figure 141. Screenshot. Example naming structure for simulation input files.

Archiving Strategy

Each data set will be created at a different stage in the project and many versions of the same basic data set will exist. The desire to model alternatives may also increase the number of data sets.

It is recommended that a spreadsheet or other electronic document be used to record: 1) the current state of each data set and 2) data sets which have been abandoned, including the reason.

Example:

Figure 142 shows a spreadsheet listing each data set. A color coding scheme has been used to assist in tracking the status of each model. A comment column provides special notes about the models.

LEGEND	
	Skip for Now
	Need to Run
	Run & data Pulled (Steps 1-7)
	Data in Charts besides Green (step 8)
	Python run and Green Results in (Step 9)

MODEL	COMMENTS
HLOffMeter	Error upon running
HROffMeter	Error upon running
HThruMeter	
HThruSuper	
HThruMUTR	
HThruMUT	
HThruHlfSu	
HROffSuper	
HROffMUTR	
HROffMUT	
HROffHlfSu	
HLOffSuper	
HLOffMUTR	
HLOffMUT	
HLOffHlfSu	
HLOnSuper	
HLOnMUTR	
HLOnMUT	

Figure 142. Screenshot. Example status tracking system for simulation models.

9.3 PREPARING FOR DATA OUTPUT

In addition to the inputs of a simulation model, information must be provided about what outputs to create (see 0). In a sense, the model data files describe the system's geometrics, operating plan, demands, and external conditions (which may affect the calibrated parameter values employed).
It is best to establish the output specs of the base model before creating additional models for the alternatives. Many types of data can be collected from a simulation model. Travel times are a common example, and those outputs are dependent on the placement of a sensor. For a travel time measurement, variability can be introduced if the sensor in one model is placed in a slightly different location than another model.

In some scenarios, the development of an alternative model may result in the movement of a data collection sensor. Additionally, it may be necessary to compare two pieces of data from the same model against one another. For example, it may be of interest to know how a change in traffic signal timing helps or hinders two different origin-destination (O-D) pairs. Because of this, a consistent reference system should be used for placement of sensors.

If exact placement of the sensors is possible and a consistent reference system is established for each type of data being collected, configuring the model for data collection may occur after alternative models have been created.

Example: The model makes a change to the traffic signal timing in a downtown grid and the analyst is interested in seeing how the delay for traffic moving eastbound through downtown compares to traffic moving northbound.

In this scenario, the starting position and ending position of eastbound and northbound traffic are in two different locations. Therefore, the project team must establish a uniform method for determining the area of roadway over which the delay is measured. The project team may decide to begin measuring delay 10 feet downstream of the vehicle input point (assuming the model has been built to contain all queues) with the ending point of the delay measurement 15 feet downstream of final intersection in the O-D path.

Setting the Naming Convention

Like the naming of simulation data sets, the output data must also have an understandable naming convention. The storage of the data will depend on the method of data extraction.

For data that is collected through a copy and paste method, two options exist:

1. All of the data collected from a simulation model can be placed in the same workbook with a separate spreadsheet for data type. The workbook name should have the same format and information as the simulation model from which it came, and each tab of the workbook should denote the type of data contained within.
2. Each type of data collected can have a unique workbook with a separate spreadsheet for each simulation model. The workbook name should denote the type of data contained within, and each tab should have the same format and information as the simulation model from which it came.

How the data are prepared for analysis will help inform which option is preferred.

Example:
Figure 143 shows a workbook organized under the first option. The simulation data set name matches the workbook name and each sheet is labeled by the type of data it contains. Figure 144 shows a workbook organized under the second option. The data type is the workbook name with a sheet for each data set.

For data sets collected through internal or external file writing systems, it is likely multiple output files will be created for each simulation model, one for each type of data collected. The file name should have the same format and information as the simulation model from which it came, with an addition at the end of the file name denoting the type of data contained within.

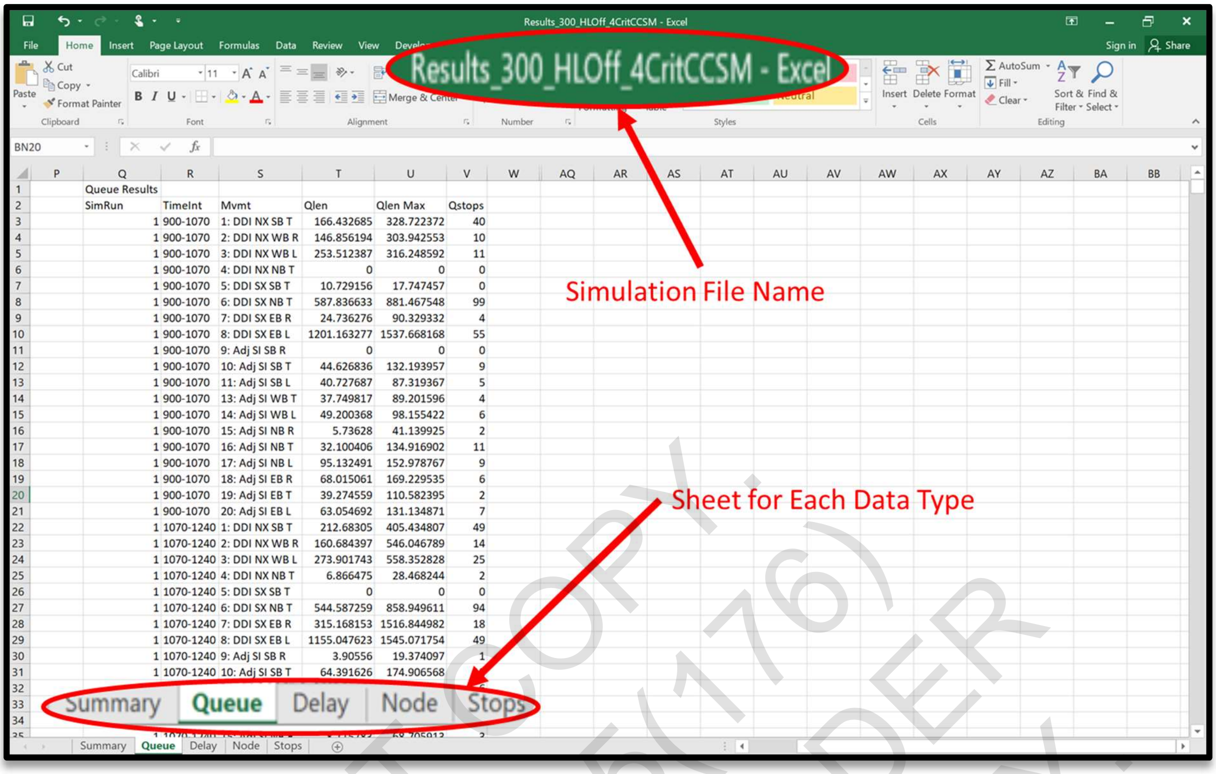


Figure 143. Screenshot. Example data storage with workbook by model and sheet by data type.

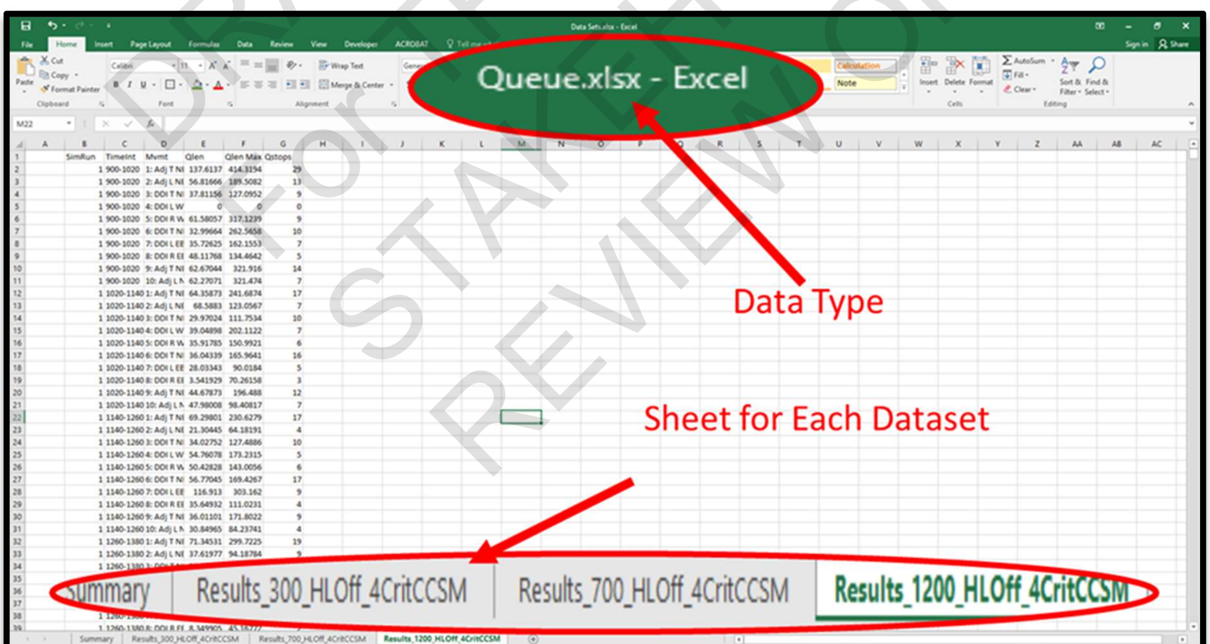


Figure 144. Screenshot. Example data storage with workbook by data type and sheet by model.

Determining the Optimal Data Output Format

An initial set of runs on a test data set should be conducted before additional models are created. This ensures the desired output data are being collected and aggregated in a reasonable manner. Additionally, it allows the project team to set up a system for storing and analyzing the data collected from all the simulation files.

Once the data output format is determined, the analyst should prepare the method of data analysis. This will most likely be either a commercial statistical software program, a spreadsheet program with formulas enabled, or a computer script developed in-house. A mock data set can be used to test the data analysis method. Likely, there will need to be iteration to ensure the data output format from the simulation matches the data input needs of the analysis method.

9.4 SPECIFYING THE SIMULATION PERIOD

The simulation period is the duration of time for which the simulation is performed. For example, if the evening (PM) peak is of interest, and it extends from 3:30–6:30 p.m., then, it might make sense to have the simulation start at 3 p.m. to allow a warm-up period and have it end at 7:30 p.m. or 8 p.m. after all traffic associated with the peak condition has left the network (all peak period O-D flows have reached the destinations). Consistent with this thought, the discussion in this chapter assumes a warm-up period is used, along with an analysis period, and a cool-down period. It is possible to include a different perspective in which the warm-up and cool-down periods are included in the analysis. However, the project team must understand how this is different and how to deal with the outputs produced. (In such cases, the network performance is assessed holistically, starting from the point when traffic starts to enter the network until the end of the simulation; and the analyst looks to see how the system responds to and deals with the demands to which it has been subjected. However, most project teams do not do this.

A very useful way to think about the duration of these time periods is in terms of trends in the probability density functions (PDF) and cumulative density functions (CDF) for metrics of interest across moving time horizons. Picture the approach to a signalized intersection. Consider the length of the queue. It varies from one cycle to the next; and, for a set of n cycles, the distribution of queue lengths can be determined. Assume at the end of cycle k this distribution is developed; and it is developed again at the end of cycle $k+1$, using overlapping data. (It does not make any sense to develop the distributions other than on a cycle-by-cycle basis, since a new data point becomes available only each cycle.) If the distribution of the queue lengths at the end of cycle $k+1$ is not statistically different from the one at the end of time step k, and the same holds true for a significant number of cycles, like n, since at that point all of the original observations will have been replaced, then the queue length distribution is likely to be in steady state. The same idea would pertain to O-D travel times, where the distribution can be refreshed with every vehicle arrival, or for queues on freeway bottlenecks, where a difference in time, say 1 minute, can be the interval with which the distribution is refreshed.

Warm-Up Period

The warm-up period allows the system to reach a steady state pre-peak-load status quo before the performance assessment starts. In the sense the system demands are dynamic in both spatial (O-D) and temporal dimensions, the warm-up period should end before the demands begin to increase. Doing this makes it possible to assess, holistically, how the system performs when subjected to the dynamic loads.

One way to specify the warm-up period is to treat it as modeling the system's behavior before the peak load starts. As indicated immediately before, the link flow rates, O-D flow rates, queue lengths, and delays should stabilize statistically before the warm-up period ends. If the pre-peak O-D flow rates are held constant, since they are not being studied anyhow, then the network should reach a condition in which the average flow rates on the links and the average outputs from the O-D flows are not changing. For example, if the analysis is concerned with a 2-mile freeway corridor where the average speed is 45 miles per hour (mph), then it will take traffic from the upstream entry point 3 minutes to reach the downstream end. Some researchers argue the warm-up period should be twice this length; or, more generally, twice the longest O-D travel time through the network (Antoniou et al. 2014; Dowling 2004). If this guidance is followed, then, in this instance, the warm-up time should be at least 6 minutes long. As discussed in the preceding section, the validity of using this warm-up time period duration can be checked by examining the PDFs for the end-to-end travel times to ensure they are not changing from 1 minute to the next. The situation for an arterial is a bit more complicated. Now, the queue length distributions at the signals should stabilize before the warm-up period ends. That is, statistically, if a PDF of the queue lengths is created on a moving-horizon basis for, say, once every cycle for 10 cycles (so that the number of observations is not too small), and the cycle length is 120 seconds, then the first complete PDF can be created after 20 minutes of simulation time, and stability in the distribution of the queue lengths might be observable five or so cycles later. If this were true, the warm-up period should be 30 minutes long. The analyst should look at the PDF distributions for the queue lengths to see if more time is required. Once stability is reached, the warm-up time period can end, and the peak-period simulation can commence.

Data Collection Period

The data collection period should begin immediately following the warm-up period. The project team should strive to capture all of the impacts associated with the peak-load condition, *including* the way in which the system deals with the increasing demands. As is probably obvious, adding capacity to the network or changing the operating plan has the impact of decreasing the severity of the impacts. It pushes off the point in time when congestion reaches its maximum, it keeps queues from forming so quickly and so early, and it may prevent unacceptable congestion conditions from arising.

Hence, the project team should be interested in seeing how the revised system deals with the dynamics (the transient) associated with the entire duration of the peak demands, for *all* the peak period travelers, from the onset of the peak condition until the end.

This means the data collection period should not end until the network's performance has returned to post-peak, steady state conditions (again, statistically speaking, as explained earlier.) Figure 5 provides an example. Assume that the purpose of the simulation analysis is to explore the impacts of incidents on freeway operation during peak-load conditions. Assume that, normally, there are no queues. The freeway is busy, but breakdowns do not occur. However, if an incident takes place, queues form, delays occur, and system performance suffers.

Assume further that during the peak-load conditions, the density on the freeway is five vehicles across the length of this section of freeway, as shown in Figure 145. Then, if the system starts empty, the warm-up period can end when, for 10 successive 30-second snapshots, the PDF for the number of vehicles on the link is not changing significantly and the average is stabilized at five. Assume this is the condition depicted in (A) vehicles proceed with no queue. Data collection can start. A little later, the crash occurs. A queue begins to form. The queue reaches a maximum length of six vehicles by the time the incident is cleared. The queue begins to diminish in size. The front of queue moves upstream. Eventually, the clearing wave catches up with the building wave, the queue disappears, and the freeway returns to the condition shown in (A) vehicles proceed with no queue. At this point, the analysis period can stop. It might be wise to continue data collection for a short while to ensure there are no aftershocks, but the data collection can stop. The system has returned to the nominal, steady state condition that would have existed if no incident had occurred. (Data collection *cannot* stop when the incident is cleared because there are still vehicles in queue. It must continue until the queue has disappeared and all vehicles affected by the incident have left the system.)

Cool-Down Period

The cool-down period allows the system to return to post-peak, steady state conditions. (Steady state in a statistical sense, as described earlier in terms of PDFs.) Another way to describe the cool-down period is that it allows for *all* impacts from the peak-load demands to dissipate. At the end of the cool-down period, there should be no evidence anywhere in the system that the peak-load condition occurred.

The cool-down period can start when the peak-load condition ends. This treats the diminishing demands as being outside the peak-load condition. The assertion is that cool-down starts when the demands start dropping. That is fine; but, it should not end, as indicated before, until the system has reached steady state in the post-peak condition. The input flows should not terminate at the end of the peak period, they should continue and decline to post-peak levels. This keeps the simulation model from perceiving the system is empty, after the peak, except for the dissipation of the peak-period traffic. Instead, volumes representative of those following the analysis period should be used.

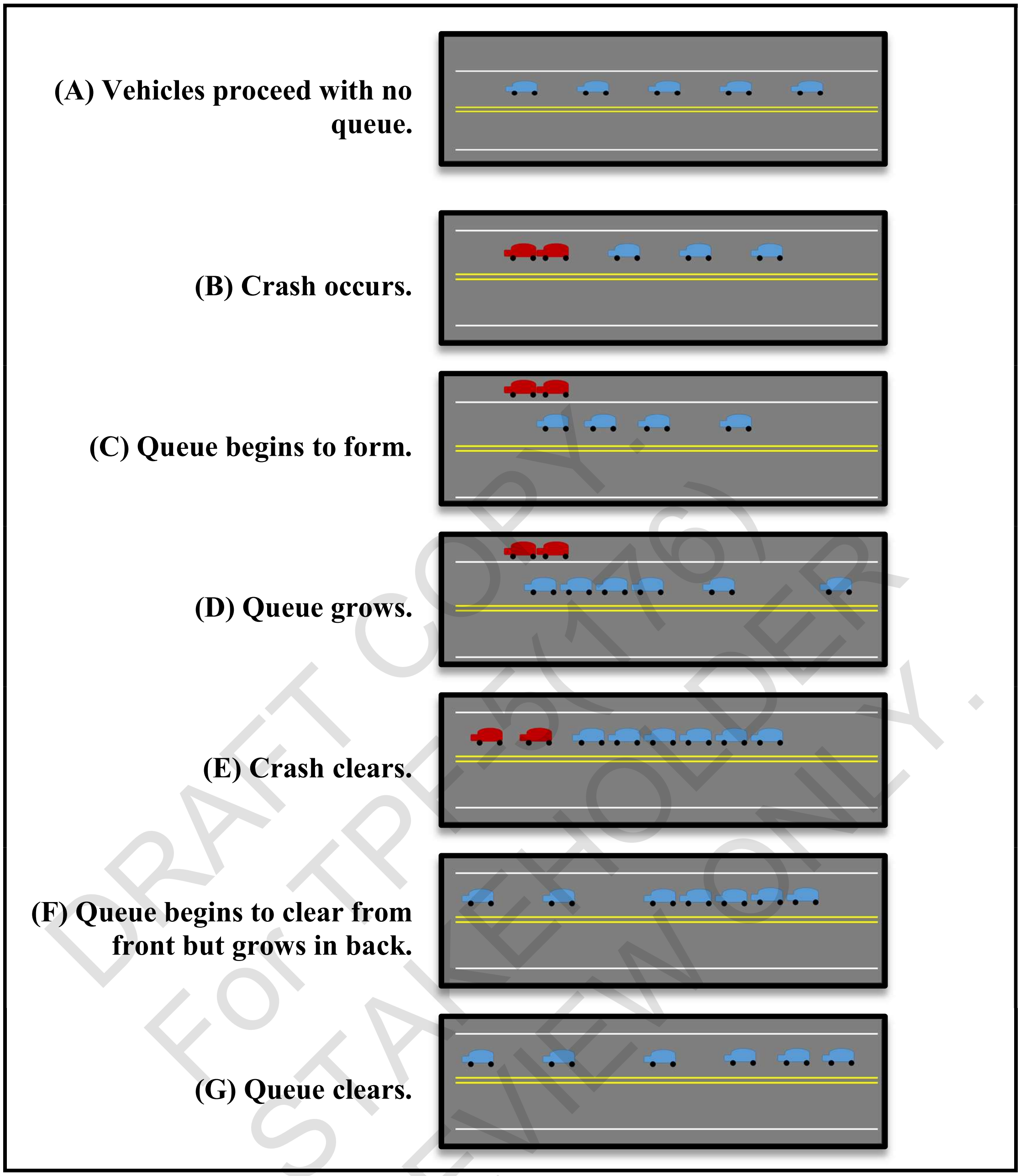


Figure 145. Illustration. Seven stages of queue formation and clearance.

9.5 DATA EXTRACTION METHODS

There are many methods for extracting data from a simulation model, ranging from copying and pasting output into a spreadsheet to writing a script that extracts the data. Each method has advantages and disadvantages, as listed in Table 18.

Table 18. Advantages and disadvantages of data extraction methods.

Copy/paste	+ Initial aggregation completed. + No coding knowledge required. - Limited control over which parameters are available. - Method of aggregation must be investigated. - Higher potential for errors during transfer.
Internal file writing system	+ Raw data may be available. + Limited coding knowledge required. - Limited control over which parameters are available.
External file writing system	+ Most flexible control over parameters that can be measured. + Raw data likely available. - Extensive programing knowledge needed. - More complex data management required.

The simulation software package may have an internal data processing function that summarizes often-used metrics and reports the information in a table. To present a manageable amount of data or to match the measurement interval of the input data, the information is typically aggregated at some level. Before using this data, it is critical to understand how the software has aggregated the data. Care should also be taken to ensure all necessary data are properly cut and pasted from the model to the data storage system.

The software package may also provide some data as separate files written during or immediately following the simulation execution. These are typically comma-separated values (CSV) files that can then be imported into a spreadsheet or analyzed with a computer script. Data may be aggregated or raw.

If the software program provides an application programming interface (API), a computer script can be written to pull data from the system. This option provides the greatest flexibility in selecting parameters to measure. APIs also allow for flexibility in the structure of data storage, which is an additional consideration that must be made by the project team.

9.6 SOFTWARE-BASED DATA PROCESSING

Some software tools aggregate data in time or space, or both, before presenting them to the user. When reviewing or using software-based data, it is critical to determine if and how the data were aggregated.

Determining Whether the Data Are Aggregated

It is critical to understand whether a data set provided by the software is aggregated or raw, and if aggregated, how it was aggregated. Data which may initially appear to be raw (e.g., travel time through an O-D path in a microsimulation) may be an average value of all measurements during a time period. Data about an O-D trip may be aggregated over a series of road segments. To determine the extent to which data have been aggregated, the software user manual should be consulted or a small test simulation should be conducted with outputs compared to hand calculations.

Weighted versus Unweighted Aggregation

A summary measure (e.g., mean, 95th percentile) can be weighted or unweighted. If each data point contributing to the summary is of equal importance, the measure is said to be unweighted. The measure should be weighted if one or more of the data points is of greater importance or has a greater weight. To find the weighted measure, the sum of the product of all data points and their respective weights is divided by the sum of the weights.

$$Weighted\ Measure = \frac{\sum_{i=1}^{n} data\ point_i \times weight_i}{\sum_{i=1}^{n} weight_i} \quad (9.1)$$

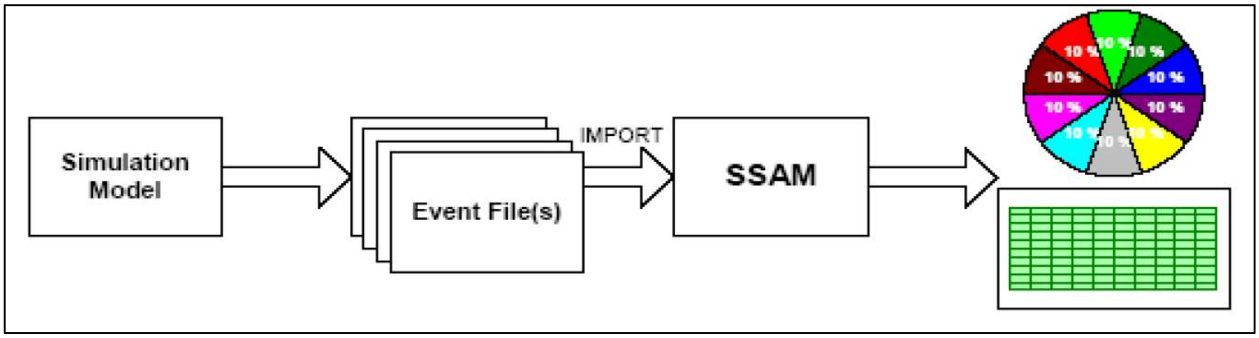
Example: Figure 146 provides travel times and volumes between four zones. The travel time between zones two and four is much higher than all other zone pairs. Because the volumes for those pairs are also large compared to other volumes, the weighted mean considers those two travel times more heavily than the travel time from zone two to three, for instance, which has a volume of only 70 vehicles per hour (veh/hr).

Origin	Destination	Volume	Travel Time
		vph	*min*
1	2	100	2.2
1	3	125	4.5
1	4	40	3.4
2	1	110	2.1
2	3	70	1.3
2	4	150	6.7
3	1	100	4.6
3	2	90	1.5
3	4	25	2.6
4	1	30	2.9
4	2	130	8.8
4	3	75	2.6
Unweighted Mean			3.6
Weighted Mean			4.14

Figure 146. Screenshot. Weighted versus unweighted mean travel time calculated from raw data.

Surrogate Safety Assessment Model

When considering the performance measures to be considered, the objective of the modeling activity may drive a desire to understand safety impacts. While safety is difficult to measure directly, surrogate safety measures can be used. For microscopic simulations, vehicle conflicts—the occurrence of two or more vehicles that will collide if an evasive maneuver is not made—can serve as a surrogate. The Federal Highway Administration (FHWA) provides a tool, the Surrogate Safety Assessment Model (SSAM) (2018), which analyzes outputs from many microscopic simulation software packages. Specifically, the tool requires the input of trajectory files and outputs a file of identified conflicts as shown in Figure 147 (FHWA 2018).



© Gettman and Head (2003).

Figure 147. Diagram. Surrogate Safety Assessment Model (SSAM) operational concept.

9.7 SOURCES OF OUTPUT VARIATION

A mathematical formula, $a + 2b$, is deterministic. If you use the same values as inputs for *a* and *b*, the output will always be the same. Deterministic models have one source of output variation: variation in the inputs or picking a new number for *a* or *b* or both will result in a new output.

Stochastic models incorporate random number seeds to create fluctuations in the modeling conditions to manipulate items such as the individual headway and order of vehicle, pedestrian, bicycle, etc., creation in the model. Two identical models with identical inputs but nonidentical random number seeds will not produce the same outputs. The variation in the two outputs is due to the fluctuations in modeling conditions created by the random number.

Transportation simulation models are frequently stochastic. Even for the same demand, vehicles queued at a signal may arrive in a slightly different pattern: four through vehicles, a left-turning vehicle, and then a right-turning vehicle or a right-turning vehicle, two through vehicles, a left-turning vehicle, and then two through vehicles. These differences will result in a variation in delay for the intersection. In macroscopic models, stochasticity may be created by considering the distribution of demand and a distribution of capacity for a single regime or between regimes (e.g., as affected by weather or road work).

Input variation accounts for differences between alternatives while random number seed variation accounts for differences within successive runs of an alternative.

Example: a stochastic model is created to determine the impact of pedestrians exiting a downtown opera house on the surrounding traffic signal operations. The east exit has a flowrate of six pedestrians per second. Model 1 has 20 percent of the pedestrians traveling south and the remaining 80 percent traveling north. Model 2 has 30 percent of pedestrians traveling south and the remaining 70 percent traveling north. Each model is executed four times. Table 19 shows the average delay for pedestrians traveling north. The difference in the delay within model 1 is due to random number seed variations. The difference in the delay between model 1 and model 2 is due to input variation.

Table 19. Average delay for pedestrians under various inputs and random seeds.

Model	Run	Seed #	Delay (min)
1 (20 % southbound)	1	15	1.1
	2	23	1.2
	3	84	1.0
	4	13	0.8
2 (30 % southbound)	1	15	1.5
	2	23	1.7
	3	84	2.0
	4	13	1.6

9.8 DATA AGGREGATION

Once the output data are created, the analyst should spend time determining whether and how to combine the data before analyzing.

Data Aggregation Size

Care should be given in considering the physical and temporal limits of synthesis for a performance measure. If the desired performance measure is volume-to-capacity (V/C) ratio for links throughout the network, the analyst must decide if each link will be individually presented or as a combined value. Additionally, over what time period should the V/C ratio be aggregated? These decisions depend on what question the analyst is trying to answer.

Traffic performance indicators may be aggregated and reported across a variety of time period durations. The choice of performance reporting aggregation may affect the quality of understanding and decision-making that flow out of the reporting process. Engineering judgment should be applied in selecting aggregation levels that convey the proper information.
Figure 148 illustrates travel times aggregated in bins that are 24 minutes wide.

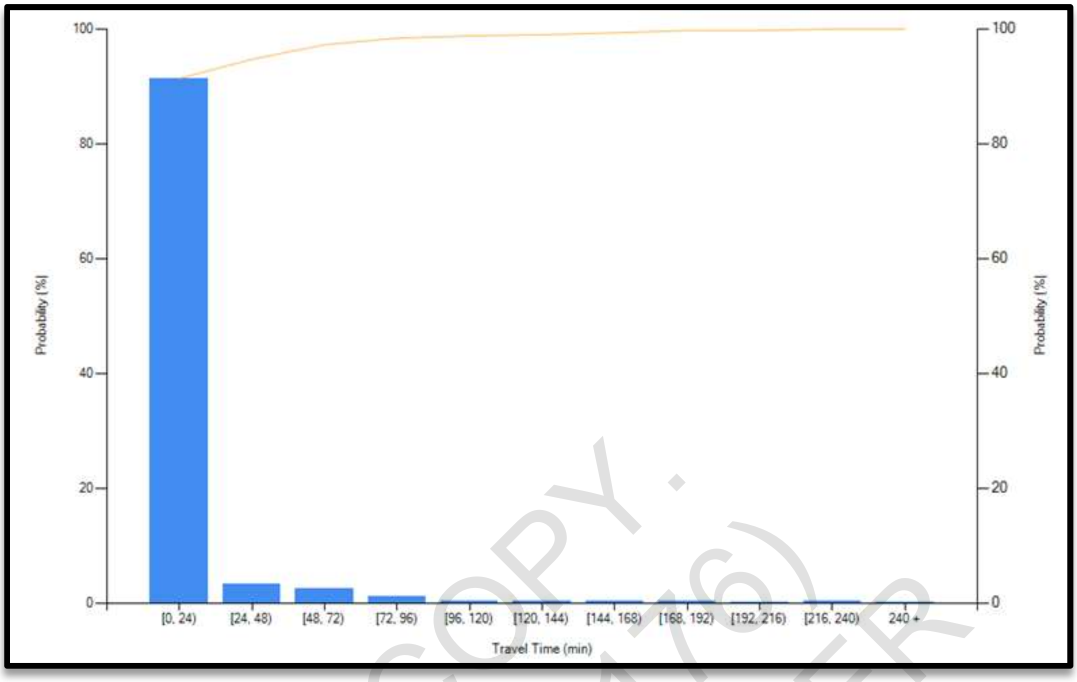


Figure 148. Screenshot. Probability density function (PDF) with large bins generated using HCM-CALC software.

However, more than 90 percent of the reported travel times are fewer than 24 minutes. Therefore, unless all travel times fewer than 24 minutes are considered ideal performance and the analyst does not need to distinguish between them, the aggregation interval should be reduced so that the fuller set of positive and negative outcomes can be viewed. By contrast, Figure 148 illustrates travel times aggregated in bins that are 1 minute wide. Although the attempt to capture more information may have been well-intentioned, the resulting excessive amount of output may be more difficult to interpret than a smaller amount of output.

Aggregation can also occur over space. The travel times over a set of links can be combined to provide the travel time for a specific route. Consideration should be given for the time periods over which the travel times were collected.

Example: An analyst has the average travel time of five consecutive freeway segments for each 1-minute interval over a 30-minute peak period. To determine the total travel time over all five segments, the analyst will need to determine how to stitch the travel times together. One option (shown in blue in
Figure 149) would be to take the travel time for each segment at a specific time in the peak period (t = 5 minutes) and add the five values together giving a travel time of 15 minutes.

A second option (shown in red in
Figure 149) would be to select a starting time in the peak period (t = 5 minutes), determine the travel time on the first segment at that time (2 minutes), and pull the travel time for the second segment at a time equal to the starting time (5 minutes) plus the travel time in the first segment

(2 minutes). This represents the travel time a user would experience as the users moves through the system.

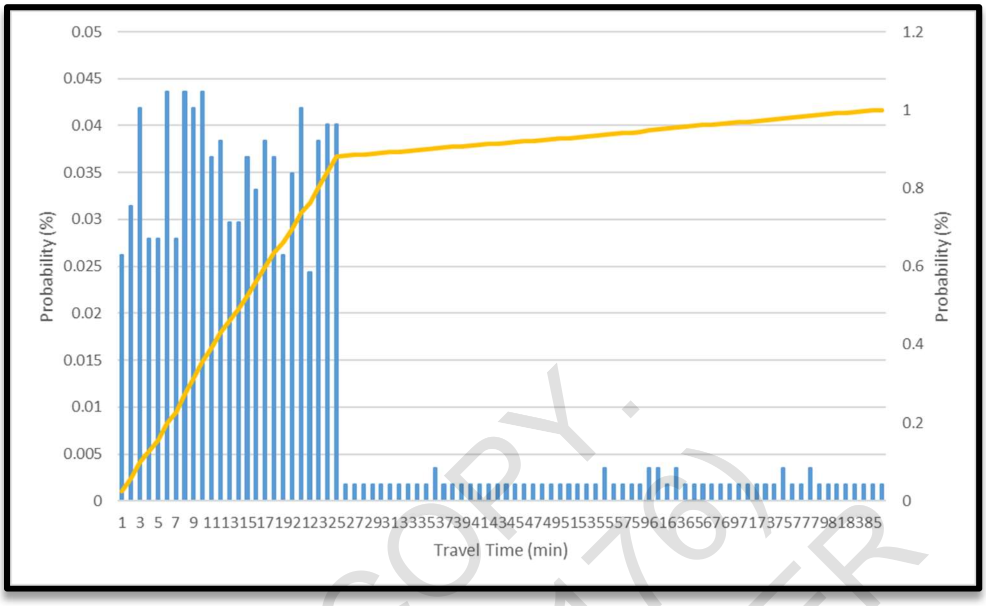


Figure 149. Screenshot. Probability density function (PDF) with small bins generated using HCM-CALC software.

As can be seen in
Figure 150, the travel time for the first option is 15 minutes while the travel time for the second option is 14 minutes.

		Time																														
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Segment	1	1	2	4	2	5	2	4	2	2	3	1	5	2	1	1	3	2	5	3	1	5	4	5	3	3	5	4	5	2	1	4
	2	5	2	2	5	1	1	1	4	3	1	4	1	2	3	2	2	5	2	5	2	1	1	3	1	4	4	1	5	3	1	1
	3	4	3	2	4	5	5	5	2	1	2	4	1	3	2	3	3	5	2	2	5	5	5	2	4	5	5	4	1	5	1	3
	4	3	2	3	1	5	3	3	3	1	1	3	2	3	5	1	2	4	2	5	2	1	2	2	1	5	3	3	1	1	4	3
	5	5	4	1	1	5	4	1	5	3	2	3	5	5	2	3	4	5	1	4	5	5	5	2	1	4	5	4	4	4	2	4

Figure 150. Screenshot. Travel times for five segments in 1-minute bins.

Spatial aggregation can be thought of as a question of facility-based results or trip/link/node-based results. Furthermore, in microsimulation, the facility may be thought of as a single interchange within the system. Providing a value for interchange delay (a facility-based measure) may mask the poor performance of a single movement through the interchange (trip based).

In weighing the decision to present facility vs trip/link/node-based results, the analyst should consider the purpose of the simulation. Is the trip/link/node the focus of the simulation? If not, is the trip/link/node result significant enough to rise to the level of alarming? Asking these questions should assist the analyst in determining how to present the data. It should be noted that the analyst may choose to present the facility-based result alongside one or more trip/link/node-based results.

Visual Inspection of the Data

Before conducting statistical tests, all data should be visually inspected. Visual inspection assists in identifying trends that may not be captured using a planned statistical method. This inspection can be modeled using scatter plots, trajectories, histograms, time-space diagrams, or other methods.

While the statistical methods can be used to measure differences between alternatives at a specific level of precision, the visual representation of the data can be used to tell the story of *how* the alternatives differ.

Example: linear regression can be used to determine the relationship between two or more variables. However, the absence of a strong linear relationship does not indicate there is an absence of any relationship. The plot in
Figure 150 shows there is a poor linear relationship between the *x* and *y* variables, as given by the horizontal dotted line, but the variables are clearly related. Modeling the data using a linear regression without first inspecting the data would lead to the incorrect conclusion that there is no strong relationship between the variables.

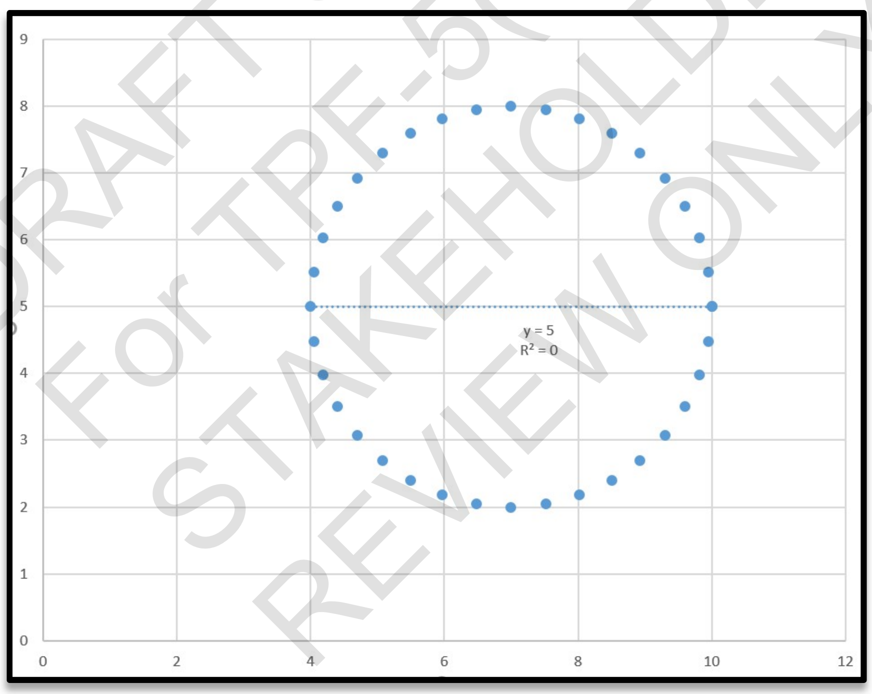


Figure 151. Chart. Example chart demonstrating poor use of linear trend line for nonlinear data.

Representing Distributions

When measuring performance in a transportation model, the measure of interest is often captured multiple times. In a macroscopic model, a V/C ratio may be captured over many links in a model. In a microscopic model, the occurrence of cycle failure at a specific approach may be captured over multiple cycles within each simulation. This results in a collection of measures which can be described as a distribution.

When synthesizing data, it can be helpful to create a distribution to understand the performance variability. When considering reliability, for example, the results from two simulation models could have identical means but very different distributions. The first may be horizontally compact with small tails—that indicates a uniformity of the travel times. The second may be bimodal with one peak occurring on the left side of the graph and a second peak on the right side of the graph. This indicates some trips had a very small travel time while other trips had a very long travel time. When considering reliability, both distributions may have the same mean travel time, but the first is much more reliable than the second.

The PDF visualizes the relative likelihood of a specific value occurring (e.g., what percent of the time was the V/C ratio > 0.83?; what percent of the time did cycle failure occur?). The value on the *y*-axis describes the likelihood while the value on the *x*-axis is the numerical or categorical value measured (e.g., the range of all V/C measured; the categorical values of yes or no).

A cumulative density function visualizes the likelihood of any value up to and including the value of interest occurring (e.g., what percent of the time was the V/C ≤ 0.83?). Cumulative density functions (CDF) generally carry no meaning for categorical performance measures.

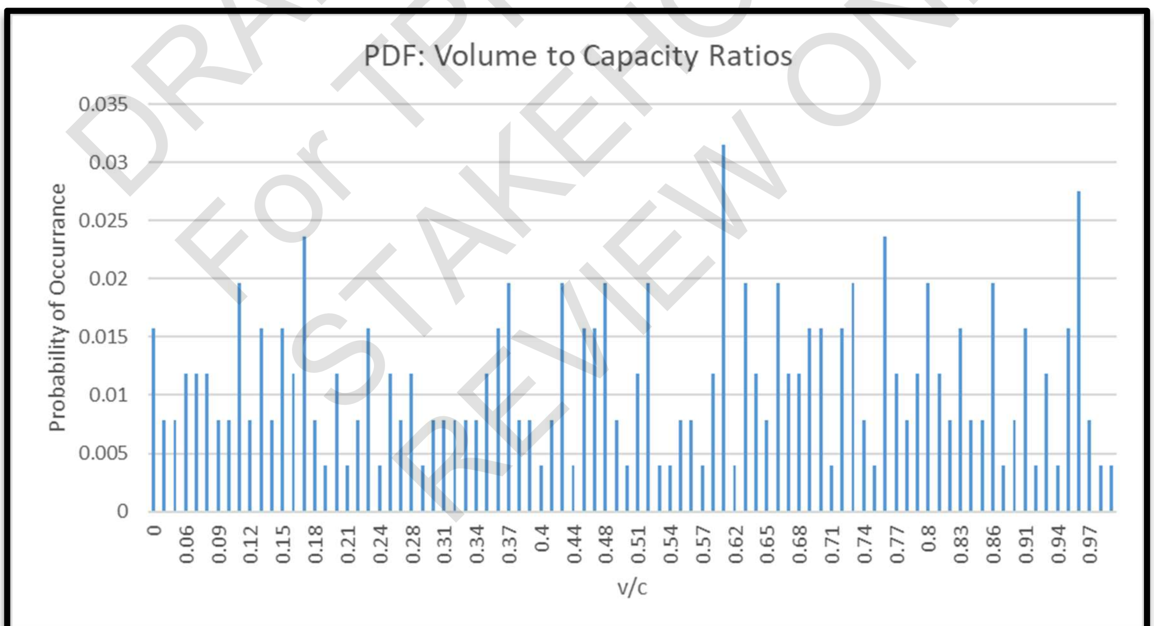


Figure 152. Chart. Probability of a range of volume-to-capacity ratios occurring.

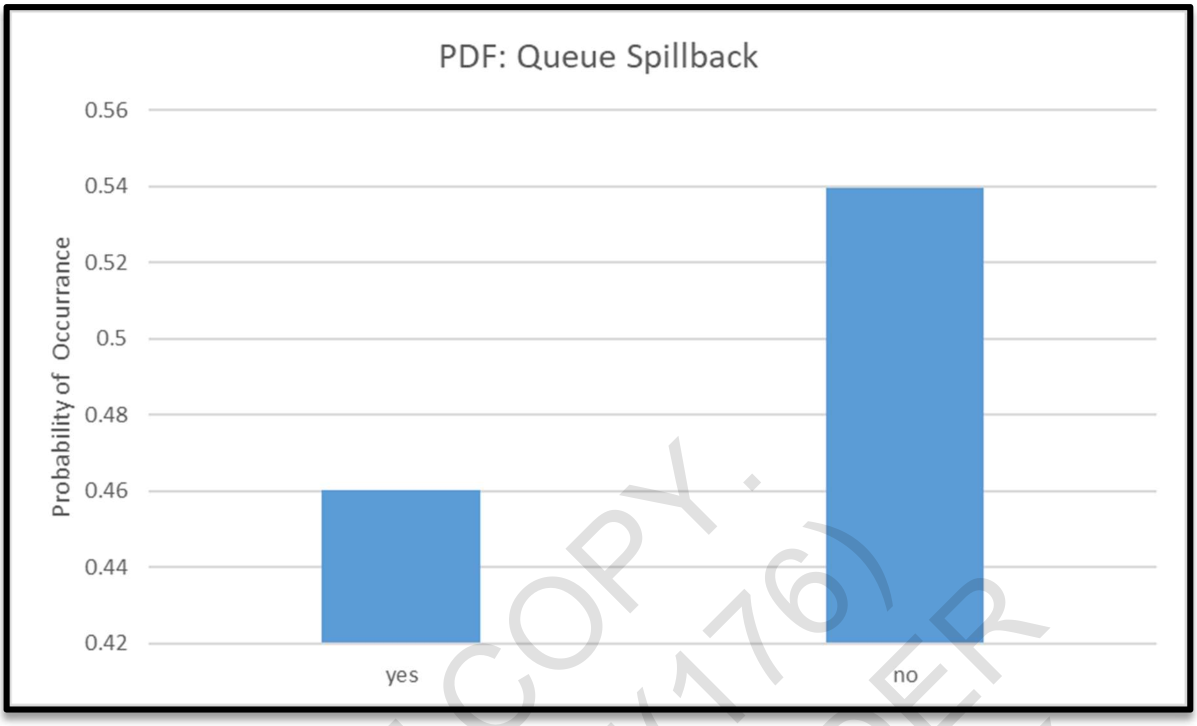


Figure 153. Chart. Probability of queue spillback occurring.

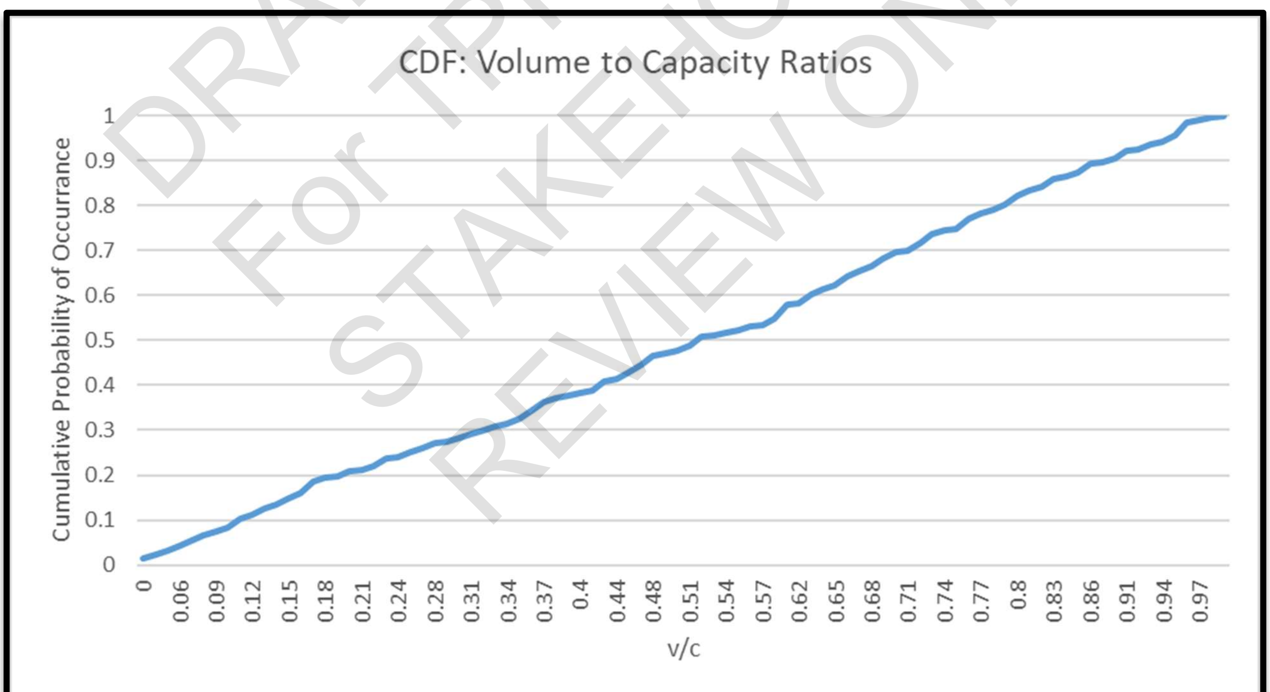


Figure 154. Chart. Cumulative probability of different volume-to-capacity ratios occurring.

It is from cumulative distributions with numerical, continuous performance measures that a specific percentile likelihood can be extracted. The most frequently used likelihood is the median, or 50th percentile; however, other values can be used. It is helpful to think of the percentiles as a measure of how frequently an event will occur. For instance, when considering the value that represents the 80th percentile, a larger value can be expected to occur once in every 5 days, minutes, cycles, etc. Table 20 provides commonly used values.

Table 20. Relationship between percentile values and likelihood of a larger value occurring.

Percentile	Likelihood of a Larger Value Occurring
50%	1 in 2
80%	1 in 5
90%	1 in 10
95%	1 in 20

Temporal Trends

The creation of a cumulative distribution and PDF orders data by magnitude. Another ordering that can provide significant insight is temporal. Assuming the model has metrics which are collected in regular intervals (e.g., 5 minutes, 15 minutes, every cycle length), these values can be charted temporally to determine trends. Inspection of temporal trends can provide a critical understanding of how the model accommodates intense volumes seen in the peak periods.

For microscopic arterial models or any type of mesoscopic and macroscopic models where queue lengths are selected as a performance measure, a trend which shows the queue length (either average or maximum back of queue) continuing to increase as the peak period ends reinforces the need for a cool-down period. Increasing queues as the cool-down period ends indicates the length of time selected or cool down may be insufficient. Similar conclusions can be drawn from temporal trends in freeway models. If speed is a performance measure, the analyst should study the trends to determine if, at the end of the cool-down period, speeds have returned to near free-flow levels, as shown in
Figure 155.

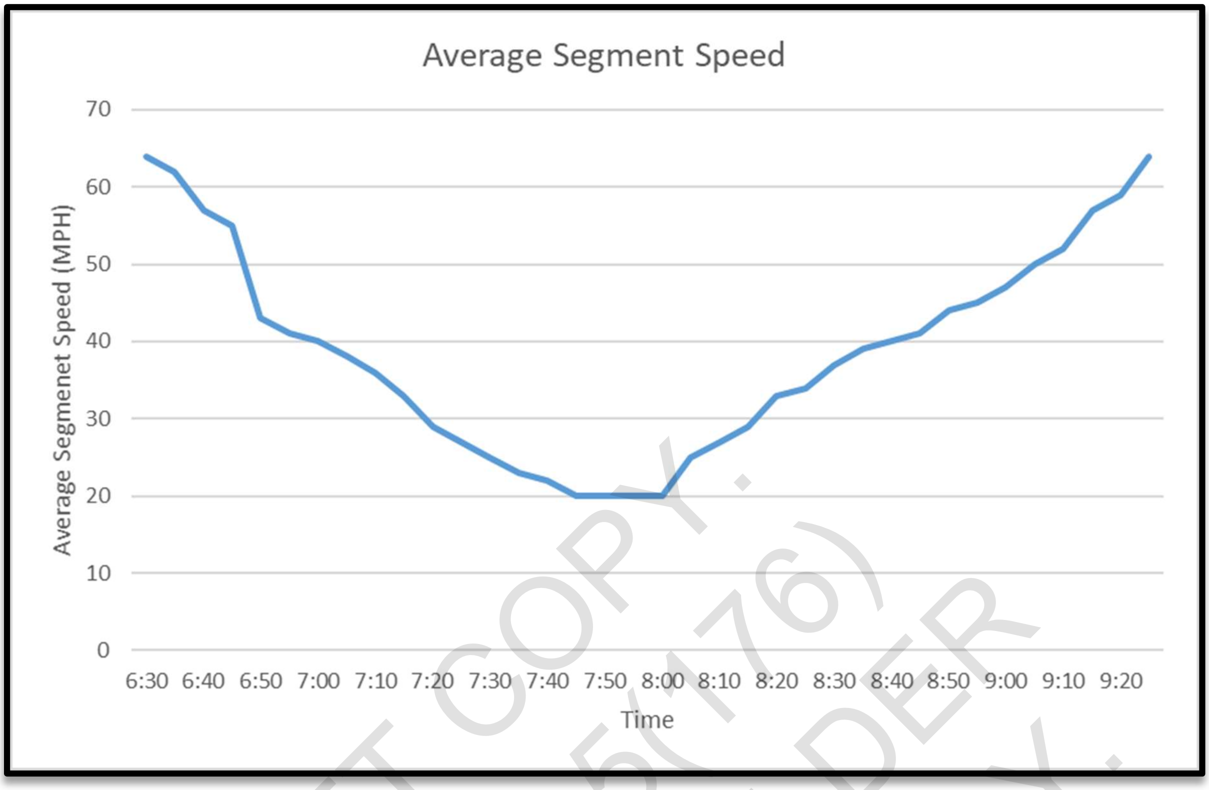


Figure 155. Chart. Temporal trend of average speed along a freeway segment.

9.9 DATA ANALYSIS

This section provides an overview of data analysis as that activity pertains to the analysis of outputs from simulation models.

Definitions

Common terms used in statistical analysis are defined in this section.

Independent/Dependent Variables

Independent variables are those variables that are being intentionally and systematically varied or controlled for (held constant) through the input parameters. Dependent variables are those values that are being tested and measured. Generally, the performance measure is a dependent variable.

Categorical/Numerical Values

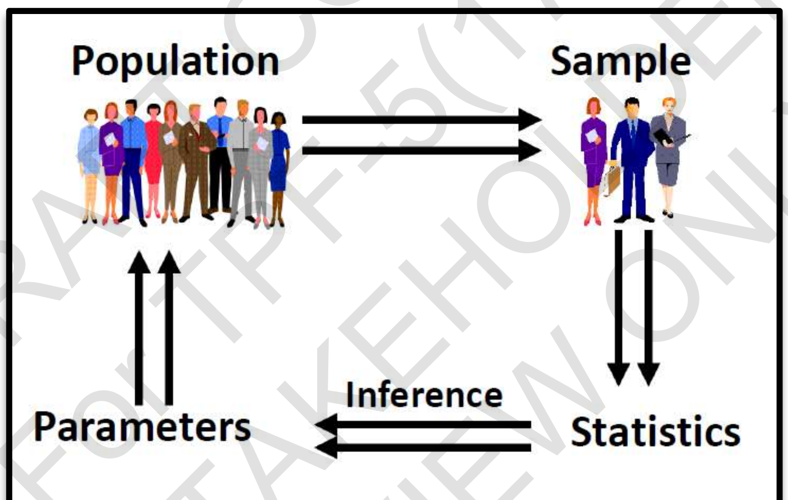
Categorical or qualitative variables are recorded by groups or categories (e.g., binary report of if the capacity was exceeded, types of gas released by a combustion engine). Graphically, categorical values are often represented by pie charts or bar charts. Numerically, categorical values are summarized using proportion or percent in each category.

Numerical or quantitative variables are recorded using numeric values for which arithmetic operations, such as adding and/or averaging, make sense (e.g., travel time, volume). Graphically, numerical values are often represented with histograms, box plots, distribution charts, etc. Numerically, numerical values are summarized using means, standard deviations, cumulative percentages, etc.

Statistic, Sample, Parameter, and Population

Simulation is used to model a subset of possible scenarios likely to occur in the built world. This subset of scenarios is known as a sample of the larger population of scenarios which occur in the built world. From the sample, we generate a summary of a variable. This summary is known as the statistic. The statistic is meant to inform the understanding of the summary of the variable from the larger population. The summary of the variable of the larger population is known as the parameter. The relationship between these four terms are shown in © McGowan (2014).

Figure 156.



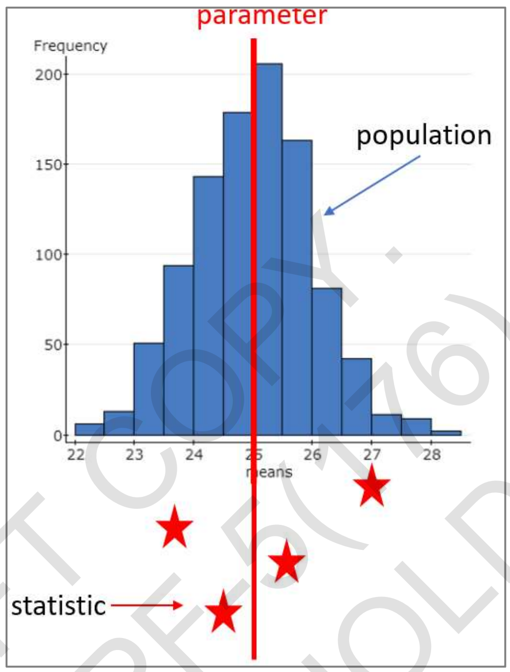
© McGowan (2014).

Figure 156. Illustration. Relationship between population, sample, statistics, and parameters confidence interval.

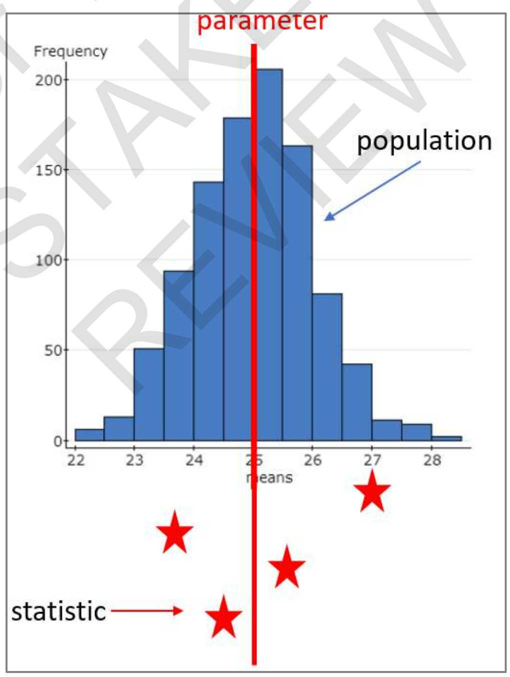
Often, a sample statistic is reported with a confidence interval. Because a statistic is developed from only a sample of the population, the statistic likely is not precisely equal to the population parameter. Therefore, the confidence interval consists of a range of values that are good estimates of the population parameter. The size of the range depends on the variability of the sample, the sample size, and the confidence level.

In
Figure 155-A below, we have a population with a known parameter. In
Figure 155, we ran four experiments and gathered four sample statistics. In

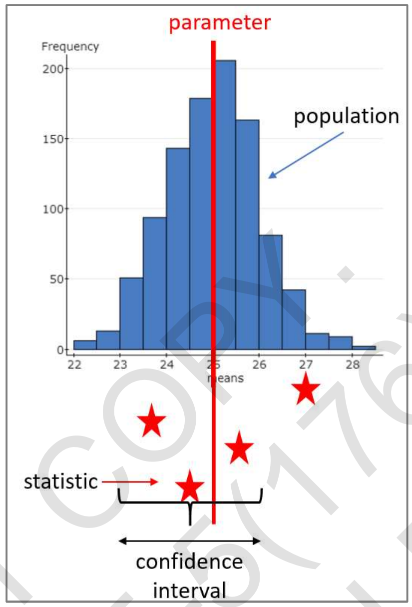
Figure 155, we developed a confidence interval around the statistic. In this case, the confidence interval correctly captures the parameter.
In Figure 157, we apply a confidence interval to all other statistics and find it often captures the parameter. In this figure, even though the top most statistic is generated from a sample of the population, the confidence interval does not capture the true population parameter.



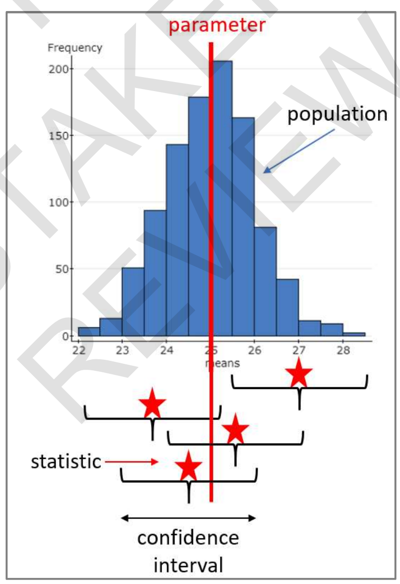
Subfigure A. Population with a known parameter.



Subfigure B. Sample statistics.



Subfigure C. Confidence interval.



Subfigure D. Application of the confidence interval to all other statistics.

Figure 157. Charts. How a confidence interval may or may not capture a population parameter.

Confidence Level

The confidence level is the percent of all possible samples in which the performance measure statistic's confidence interval captures the true population parameter.

Consider an experiment in which an infinite number of scenarios are simulated and the performance measure is recorded with the confidence interval. While the confidence interval consists of a range of values that are good estimates of the population parameter, the true population parameter may not be in the confidence interval. Some proportion of those confidence intervals *will* contain the true population parameter. That proportion is the confidence level.

If a sample statistic and confidence interval is reported with a confidence level of 95 percent, it *can* be understood that 95 percent of the confidence intervals generated from every possible sample will contain the population parameter; however, it *does not* mean:

- The specific confidence interval presented has a 95 percent likelihood of containing the population parameter.
- 95 percent of the data collected from the sample lie within the interval.
- There is a 95 percent probability the sample statistics from a future experiment will fall within the confidence interval.

Alpha

Alpha, also known as the significance level, is the likelihood that the true population parameter is outside of the confidence interval. Therefore, the sum of alpha and the confidence level must equal 1.

Alpha is also the probability of making a type I error, that is, the likelihood that a hypothesis test rejects the null hypothesis even though the null hypothesis is true. If the hypothesis test is determining if two alternatives are different based on a statistic, a type I error would occur if the test concluded one alternative was better than the other when they are the same.

Beta

Beta is the probability of making a type II error, that is, the likelihood that a hypothesis tests accepts the null hypothesis even though the null hypothesis is false. If the hypothesis test is determining if two alternatives are different based on a statistic, a type II error would occur if the test concluded the alternatives are the same when they are distinct from one another.

Power

Power describes how likely a statistical test is to find a difference if one exists. It is the probability of not making a type II error. Therefore, the sum of beta and the power must equal 1.

Stochastic versus Deterministic

A mathematical formula, such as $a + 2b$, is deterministic. If you use the same values as inputs for a and b, the output will always be the same. The output is fully determined by the input.

Stochastic models contain inherent randomness, generally through the use of a random number seed. Two identical models with identical inputs will not produce the same outputs. The variation in the output is due to the fluctuations in modeling conditions created by the random number seed.

Random Seed

The random seed is a number used to begin the generation of other pseudorandom numbers. These newly generated numbers are selected from a probability distribution. In stochastic simulation models, select inputs are generated through the pseudorandom number generation process. It should be noted that the random seed does not need to be randomly generated. The title of random comes from the pseudorandom numbers the seed creates.

The random seed should be changed for each run completed within a given model. However, when comparison between the models is expected, the same set of random seeds should be used for each model tested.

Significance—Practical versus Statistical

Statistical significance indicates the observed effect (or difference between two effects) is due to a difference in the populations, as opposed to being due to having picked an extreme sample of the population.

Practical significance indicates that the observed effect (or difference between two effects) is meaningful in the real-world application. For example, given a specific confidence level, sample size, and variance, there may be statistical difference between the travel time under the no-build and the preferred alternative for a project; however, if that travel time difference is 4 seconds per vehicle (sec/veh) over a 45-minute trip, that difference is likely not practically significant.

Determining the Number of Replications Needed

When an engineer has access to outputs generated by many realizations or runs, corresponding to an equally large number of changes to the random seed(s), combined results computed from these outputs become more accurate. The amount of performance variability also becomes clearer. Conversely, when the number of realizations is too low, the range of the confidence interval for a given confidence level can be very large. This makes it difficult to determine if one model is statistically different from another model.

Example: While performing sensitivity analysis to assess the impact of various input parameters on congested freeway speeds, an engineer notices that reducing the off-ramp reaction distance from 1,500 to 500 feet (ft) improved the efficiency of traffic flow. Table 21 illustrates these results, which were based on only one realization. Normally, giving drivers less time to react would produce inefficient last-minute lane changes, which would degrade performance. Some consultants and researchers have made the mistake of submitting and publishing incorrect conclusions based on only one realization.

By contrast, Table 22 illustrates the same set of results averaged across 10 realizations. Here, reducing off-ramp reaction distance to 500 ft reduces average vehicle speeds by a significant percentage. Not only are these results more intuitively sensible, but they are also statistically more likely to represent the true population speed. Although this example focused on average results, note that a robust simulation analysis should consider the variance and stability of outputs.

Table 21. Misleading results based on only one realization.

	Average Speed (mph)
Base case	26.0
Anticipatory lane change 500 feet	27.7
Off ramp reaction point 500 feet	27.0
Off ramp exit fraction 24%	45.8
Car following sensitivity + 0.1	26.4
5% trucks	21.6

Source: McTrans.

Table 22. Improved results based on 10 realizations.

	Average Speed (mph)
Base case	28.4
Anticipatory lane change 500 feet	28.0
Off ramp reaction point 500 feet	24.1
Off ramp exit fraction 24%	45.4
Car following sensitivity + 0.1	25.2
5% trucks	22.9

Source: McTrans.

Among other simulation guidance documents, Guidelines for Applying Traffic Microsimulation Modeling Software (Dowling 2004) details the process for selecting the number of replications. The number of replications must be calculated for each performance measure separately. In order to accurately calculate the number of replications needed, the exact performance measures must be known.

Example: consider two O-D pairs—one with a high volume and a second with a low volume. If the desired performance measure is travel time along the O-D path, more replications will be needed to achieve the desired level of confidence for the second O-D pair than would be needed if the analyst were concerned with only the first O-D pair. The number of observations is the total number of vehicles traveling over the O-D pair during all replications. As the first pair has a higher volume than the second, fewer replications are needed to achieve the same number of observations.

If the model is developed to provide only the average travel time over the analysis period for each O-D pair, then each replication provides only one observation. To achieve the desired confidence level, many more replications would be needed than if individual travel times were gathered.

Selecting Alpha and the Confidence Level

Typically, a confidence level of 95 percent or higher is accepted in the transportation field, however a different value may be selected if the analyst desires. A higher value will require more replications of the model while a lower value reduces the probability that the future sample statistic and confidence interval will capture the true population parameter.

Selecting the Confidence Interval

The size of the confidence interval is at the discretion of the analyst. For models where it is necessary to distinguish between small variations (the models are similar), a smaller confidence interval should be used. For a given confidence level, the smaller the confidence interval, the more repetitions are required.

Reporting the selected confidence interval and confidence level is important. For example, if a model or a set of measurements indicates an average travel time of 7 minutes through a specific freeway corridor, a confidence interval could indicate the true average travel time lies between 6.5 and 7.5 minutes with 90 percent confidence. However, depending on the observed variance, there might only be 50 percent confidence that the true average travel time lies between 6.5 and 7.5 minutes in a different model. Unless confidence intervals and levels are included in the performance reporting, the average travel time would be reported as 7 minutes in both cases, and a decision maker would have no insight into how trustworthy these results were. Important decisions could potentially be made based on unreliable information.

9.10 SUMMARY STATISTICS

When assessing stochastic systems, each replication of simulation can present a unique value for a given performance measure. Providing a summary of that performance measure over all replications commonly occurs in two forms: a sample distribution, or a mean and standard deviation.

Summary statistics (including mean, mode, range, standard deviation, maximum value, minimum value, etc.) are used to provide numerical detail about the sample distribution.

Care should be taken when describing the summary statistic. If the data point measured directly from each simulation replication is an averaged value (e.g., average travel time), then presenting the mean over all replications would be the average of the average value (e.g., average of the average travel time).

Calculating a Confidence Interval

While all the summary statistics assist in understanding the distribution, the most commonly used summary statistic is a mean. By providing a confidence interval for the mean, a second piece of information is provided that describes the deviation of the distribution. If the population distribution is bell shaped or the number of observations is large, the analyst can use common statistical methods to develop a confidence interval for the mean.

Four values must be known to calculate the confidence interval: the mean, confidence level, standard deviation of the sample distribution, and number of observations. The confidence interval is given by:

$$\bar{y} \pm t \frac{s}{\sqrt{n}} \tag{9.2}$$

Where:

$\bar{y}$ = mean value over all replications
s = standard deviation of the sample distribution
n = number of observations

and *t* is found using a two-tails T-distribution table (Statistics How To 2018).

Using a Two-Tailed T-Distribution Table

To use the table, two values must be known: the number of degrees of freedom (DF) and the confidence level. When finding the confidence interval, the number of DF is one less than the number of observations. (Note: the number of DF may be different when conducting hypothesis testing). Using a two-tails T-distribution table, the *t* value is presented at the intersection of (1 - confidence level) and the number of DF.

Example: For a confidence level of 95 percent (1 – 0.95 = 0.05) and 29 DF, the *t* value is 2.045 as shown in Table 23.

Table 23. Two-tails T-distribution showing the intersection of a confidence level of 95 percent and 29 degrees of freedom (DF).

Two Tails T Distribution Table

DF	A = 0.2	0.10	0.05	0.02	0.01	0.002	0.001
∞	t_a = 1.282	1.645	**1.960**	2.326	2.576	3.091	3.291
1	3.078	6.314	**12.706**	31.821	63.656	318.289	636.578
2	1.886	2.920	**4.303**	6.965	9.925	22.328	31.600
3	1.638	2.353	**3.182**	4.541	5.841	10.214	12.924
4	1.533	2.132	**2.776**	3.747	4.604	7.173	8.610
5	1.476	2.015	**2.571**	3.365	4.032	5.894	6.869
6	1.440	1.943	**2.447**	3.143	3.707	5.208	5.959
7	1.415	1.895	**2.365**	2.998	3.499	4.785	5.408
8	1.397	1.860	**2.306**	2.896	3.355	4.501	5.041
9	1.383	1.833	**2.262**	2.821	3.250	4.297	4.781
10	1.372	1.812	**2.228**	2.764	3.169	4.144	4.587
11	1.363	1.796	**2.201**	2.718	3.106	4.025	4.437
12	1.356	1.782	**2.179**	2.681	3.055	3.930	4.318
13	1.350	1.771	**2.160**	2.650	3.012	3.852	4.221
14	1.345	1.761	**2.145**	2.624	2.977	3.787	4.140
15	1.341	1.753	**2.131**	2.602	2.947	3.733	4.073
16	1.337	1.746	**2.120**	2.583	2.921	3.686	4.015
17	1.333	1.740	**2.110**	2.567	2.898	3.646	3.965
18	1.330	1.734	**2.101**	2.552	2.878	3.610	3.922
19	1.328	1.729	**2.093**	2.539	2.861	3.579	3.883
20	1.325	1.725	**2.086**	2.528	2.845	3.552	3.850
21	1.323	1.721	**2.080**	2.518	2.831	3.527	3.819
22	1.321	1.717	**2.074**	2.508	2.819	3.505	3.792
23	1.319	1.714	**2.069**	2.500	2.807	3.485	3.768
24	1.318	1.711	**2.064**	2.492	2.797	3.467	3.745
25	1.316	1.708	**2.060**	2.485	2.787	3.450	3.725
26	1.315	1.706	**2.056**	2.479	2.779	3.435	3.707
27	1.314	1.703	**2.052**	2.473	2.771	3.421	3.689
28	1.313	1.701	**2.048**	2.467	2.763	3.408	3.674
29	**1.311**	**1.699**	**2.045**	**2.462**	**2.756**	**3.396**	**3.660**
30	1.310	1.697	**2.042**	2.457	2.750	3.385	3.646

Hypothesis Testing

Hypothesis testing is used to determine if a statement is true. In hypothesis testing, knowledge of statistics is used to create a frame of reference about what a distribution would look like if the statement were true. Then, using the collected data, a distribution is generated and a test is performed to determine if the distribution from the collected data matches the hypothetical distribution.

In hypothesis testing there are four steps:

1. Establish the null and alternative hypotheses.
2. Find the test statistic.
3. Specify the null distribution and find the p-value.
4. Interpret the p-value and make a conclusion.

Hypothesis testing allows the analyst to make a claim about the difference (or sameness) of two or more distributions.

For any hypothesis testing to be valid, the sample must be random (typically guaranteed through the random number seed) and the population must be normally distributed. If it cannot be shown that the population is normally distributed, a large sample size (greater than 30 observations) should be used.

Paired T-Test

Frequently, simulation may be used to determine the difference between two alternatives or the base condition and one alternative. The t-test for a difference assumes there is no difference between the two options (for the performance measure being tested) and tests whether that is false at a given confidence level. This assumption is called the null hypothesis.

The result of the t-test for a difference is either that the null hypothesis is rejected (there is a statistical difference between the two options) or it is not rejected (there is not enough evidence to say a statistical difference exists between the two options). If the null hypothesis is not rejected, the analyst must either accept the finding or increase the power. This is done by either increasing the number of replications of each sample or relaxing the confidence level. As the confidence level is often set based on project goals, it is more likely the analyst will choose to generate more replications.

In order for the paired t-test to be valid, the samples must be independent as well as random, the population variances must be equal, and the populations must be normally distributed (or the sample size must be larger than 30). While there are statistical tests that quantitatively determine if the variances are equal, it can be qualitatively verified by ensuring the two sample distributions have a similar width and shape.

The mathematical formulas for the paired t-test can be found in any statistics reference book as well as in the Guidelines for Applying Traffic Microsimulation Modeling Software (FHWA

2015). It should be noted that unlike for the summary statistic in a paired t-test the number of DF value is the total number of observations across both tests minus 2.

Analysis of Variance

When considering the difference among more than two samples, an analysis of variance (ANOVA) should be used first. This test determines if a statistical difference exists between any of the alternatives. ANOVA requires the same conditions be met as the paired t-test. That is, the samples must be independent as well as random, the population variances must be equal, and the populations must be normally distributed (or the sample size must be larger than 30). While there are statistical tests that quantitatively determine if the variances are equal, it can be qualitatively verified by ensuring the two sample distributions have a similar width and shape.

Similar to the paired t-test, the ANOVA starts with an assumption—the null hypothesis—that there is no difference between any of the alternatives. If the test statistics show the hypothesis to be rejected, there is a difference between at least one pair of alternatives. If the test statistic is not rejected, the analyst can either accept the finding, or increase the power (lower the confidence level, or increase the number of replications).

The mathematical formulas for ANOVA can be found in any statistics reference book as well as in Guidelines for Applying Traffic Microsimulation Modeling Software (FHWA 2015).

If the null hypothesis is rejected, the analyst should move on to determining which pairs of alternatives have differences. The paired t-test can be used for this test; however the net confidence level is reduced with each pair tested even if the value used in the formula remains the same. If there are three alternatives, there are three possible pairs to test: A to B, A to C, and B to C. Even if the paired test is completed using a confidence level of 95 percent, the fact that it is completed three times reduces the net confidence level of all three tests to 0.95^3, or 85.7 percent.

To avoid the reduction in the confidence level, the analyst can use Tukey's Honestly Significant Difference test. The mathematical formulas for Tukey's Honestly Significant Difference can be found in any statistics reference book as well as in Guidelines for Applying Traffic Microsimulation Modeling Software (FHWA 2015).

9.11 INTERPRETING THE RESULTS

Absolute Performance versus Relative Performance

When considering reporting data, the validation should be recalled. A perfect validation may not have been achieved. In those cases, data should be reported as a relative performance (i.e., in comparison to the base model) as opposed to an absolute performance. Even with models with strong validation, presentation of data in using relative versus absolute performance may improve stakeholder understanding.

When relative performance is used, care should be taken for reporting performance measures that are nonlinear. If there are key segments or nodes where performance measures are critical to decision-making, those nodes or segments should be a focus during validation to provide additional confidence in the results.

Link-Based versus Lane-Based Results

Lane-specific simulation output data and output reports may provide additional insights.

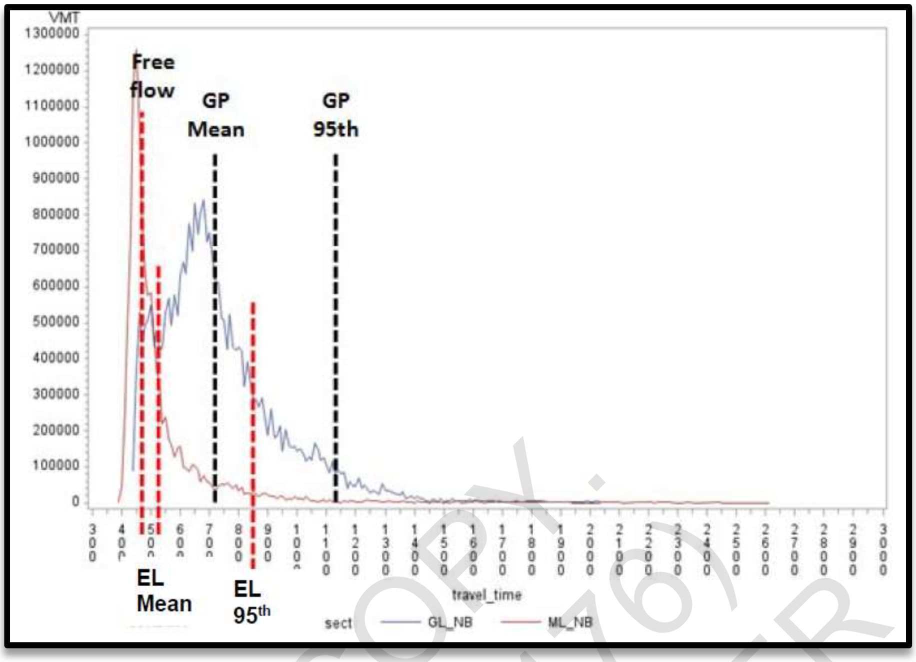
Example: high-occupancy-toll (HOT) vs. general purpose (GP) lanes: LA 2014 report (Cambridge Systematics 2014) compared the performance of HOT lanes versus the performance of GP lanes. The report recommended consideration of the 95th percentile travel time index (also known as the planning time index), and the percentage of vehicle miles traveled (VMT) at a speed greater than 45 (mph). The summary report illustrated in Table 24 shows a clear contrast in reliability between the GP lanes and HOT lanes. These results are based on the travel time graphs shown in Figure 158 which provide a direct comparison between GP and HOT lane distributions.

Table 24. Lane-specific performance reporting.

	TTI	Average TTI (Minute)	P95TTI	P95_TT (Minute)	Percentage of VMT with Speed >45 mph
GP Lanes					
5-minute	1.52	12.1	2.33	18.7	
15-minute	1.52	12.2	2.33	18.7	24.72%
Hourly	1.52	12.1	2.29	18.4	
Weekday	1.14	8.8	1.77	14.2	
95 Express					
5-minute	1.15	8.9	1.73	13.8	
15-minute	1.15	8.9	1.73	13.9	97.04%
Hourly	1.15	8.9	1.63	13.1	
Weekday	1.05	7.6	1.27	10.1	
Composite					
15-minute	1.41	11.2	2.26	18.1	41.41%

TTI = Travel Time Index; and VMT = Vehicle Miles Traveled.

Source: adapted from Florida Department of Transportation (2014).



Source: adapted from Florida Department of Transportation (2014).

Figure 158. Chart. Travel times of high-occupancy toll (HOT) lanes and general purpose (GP) lanes.

9.12 DATA PRESENTATION

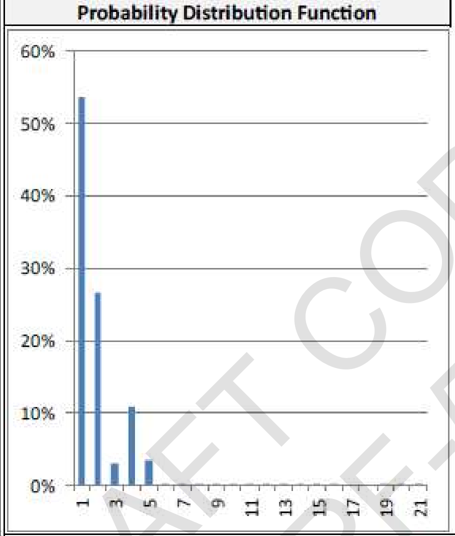
In a robust simulation analysis, no single performance measure is perfect. Univariate measures may be misleading. It is helpful to consider several relevant numeric and visual performance measures, as shown in Figure 159. Additional examples of combinations of text and visual presentation are included in the Virginia Department of Transportation (VDOT) Traffic Operations and Safety Analysis Manual (VDOT 2015). Moving vehicle animations, which may provide additional insights, should be reviewed in addition to any numeric statistics and static graphics.

Reliability Analysis Summary Report for Sample Facility

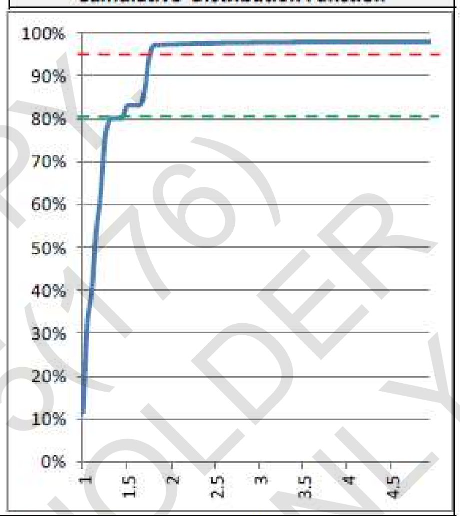
Facility Description			
Number of Segments	4	Number of Weather Scenarios	80
Facility Length (miles)	1.0	Number of Incident Scenarios	192
Number of Total Scenarios	334	Numb. of Incident + Weather Scen.	54

Facility Reliability Performance Measures			
Mean TTI	1.25	Standard Deviation	5.10
PTI	1.77	Semi-Standard Deviation	5.58
80th percentile TTI	1.32	Percent of VMT at TTI > 1.33	51.73%
Misery index	3.30	Percent of VMT at TTI > 2	25.97%

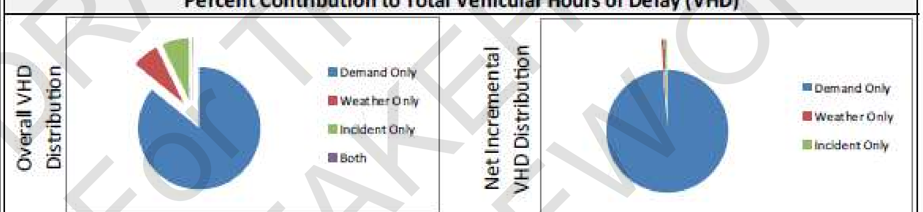
Probability Distribution Function



Cumulative Distribution Function



Percent Contribution to Total Vehicular Hours of Delay (VHD)



Analysis Details for Reliability Reporting Period

Recurring Congestion (Demand Only)		Non-Recurring Congestion Only	
Maximum d/c Ratio	1.2	Maximum d/c Ratio	3.48
Maximum TTI	1.8	Maximum TTI	13.2
% Time with Queuing	2.9284%	% Time with Queues	0.0004%
% Time without Queues	97.07%	% Time without Queues	100.00%
Percent of VMT in Recurring-Only	2.6%	Percent of VMT in All Periods	97.4%
Total % Unserved VMT in RRP	100.00%	Total % Unserved VMT in RRP	96.28%

Figure 159. Infographic. One-page summary report of multiple performance measures. Textual presentation.

As with any presentation, the text and images should be created with the audience in mind. While the analyst may have spent significant time developing, running, and analyzing the model, the audience will likely need to be brought up to speed on the relevant details of the project.

Details about the original objective for the study as well as the physical and/or temporal context can assist in orienting the audience to the correct frame of reference.

When presenting numerical data in textual form, it may be beneficial to present as a percent change in the baseline as opposed to a raw number. This orients the audience to considering relative differences between alternatives (or between the base case and the alternative).

While specific forms and charts may be required by the client, all formal reports should include the following:

- Original scoping limits, alternatives analyzed, and performance measure selection.
- Field data collection methodology and results.
- Model development process and assumptions.
- Verification, calibration, and validation methods and results.
- Data analysis results.
- Assumptions and unknowns.
- Recommendation(s).

9.13 VISUALIZATION OPTIONS BY DATA TYPE

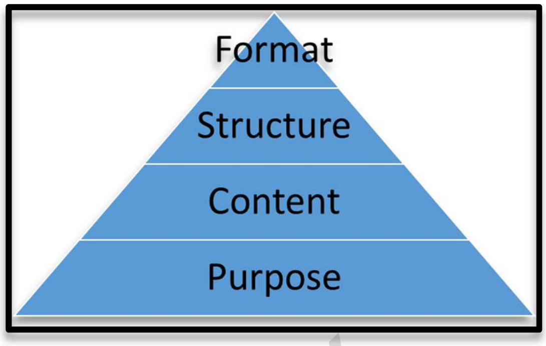
The University of Reading Statistical Services Centre (2000) suggests text can only be used to convey three or four numbers. Once that is exceeded, a visualization is necessary.

When developing visualizations, a strict hierarchy of priorities should be observed as shown in Figure 160. The purpose should be considered first, followed by content development, then structure, and finally formatting.

A clear purpose for the presentation should be identified which assists in developing communication goals. There is likely a specific question the analyst is trying to answer for the viewer.

The next step is to review the content. Ensure the data help answer the specific question and purpose of the presentation. The amount of data is important as well; generally, anything fewer than 20 data points is likely best served in a table. The United Nations Economic Commission for Europe further discusses the decision between using a table or a chart as well as laying out keys for proper presentation (United Nations Economic Commission for Europe 2009, 12).

The structure of the visualization should suit the data. It should establish a strong structure and highlight key facts while reducing complexity. Strive to focus on the most important attributes.



© Polley (2016).

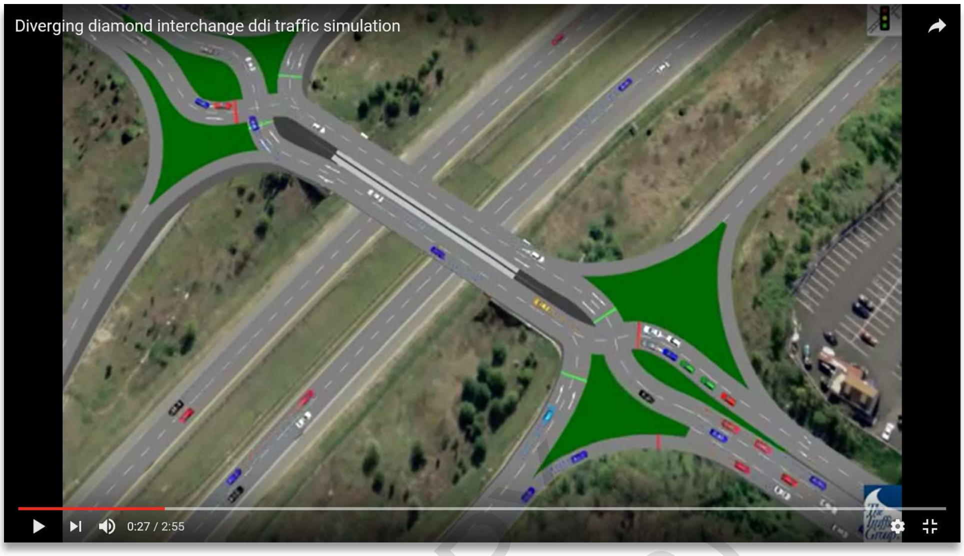
Figure 160. Diagram. Hierarchy of visualization priorities.

Finally, the format of the data should be accessible to all viewers. Selection of color, shape, and location should consider the cultural notions the viewers possess (e.g., in American culture, green tends to indicate good performance while red is associated with poor performance).

United Nations Economic Commission for Europe (2009, 11) provides a checklist for developing good data visualizations.

Simulation Tool Generated Graphics

Moving vehicle animations, such as the one shown in Figure 161, are critical for engineers and decision makers in communicating traffic operational performance and concepts; however, fewer users are familiar with the helpful static graphics features that many simulation products offer.



© adapted from The Traffic Group (2011).

Figure 161. Screenshot. Moving vehicle animation of a diverging diamond interchange.

These static graphics can sometimes produce insights that would not have been possible through observation of dynamic animations. These include flow profile diagrams, time-space diagrams, probability density functions, pie charts, and color-coded maps that reveal network performance, such as the one shown in

Figure 162.

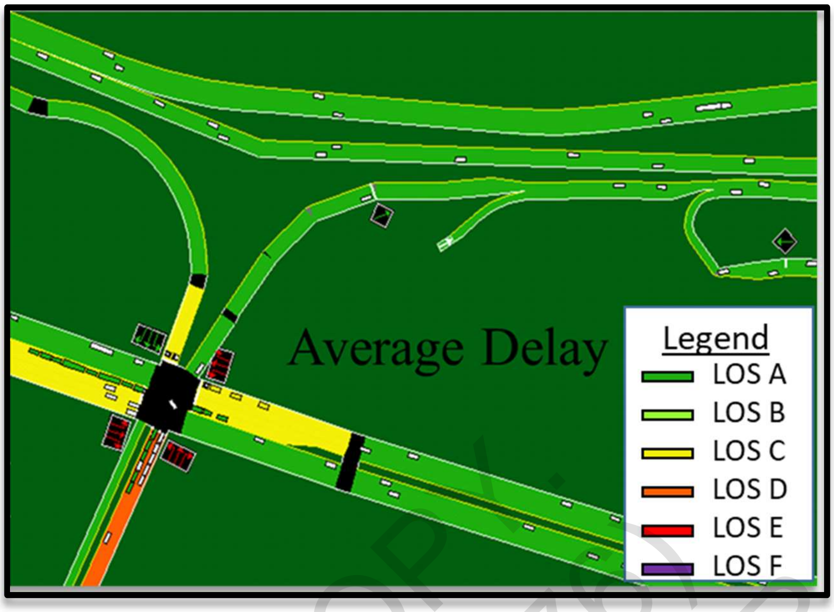


Figure 162. Map. Color-coded segments based on performance measures.

One website that can assist in selecting appropriate data visualization options is RAWGraphs (http://raw.densitydesign.org/). The tool allows users to upload data for immediate development of multiple visualizations. Alternatively, users can work with existing data sets to explore what types of data (categorical vs. numerical) are most appropriate for each type of presentation style.

Simulation Visualization

In presenting to nontechnical audiences, it may be helpful to provide a simulation (in two or three dimensions) of the alternatives or preferred alternatives. This can help the audience understand atypical operations of the system. A presentation angle from the view of the user (e.g., driver, bicyclist, etc.) may be more helpful than a plan view. Additionally, the presence of existing buildings and landmarks in the simulation can assist in orienting the audience to the area.

If the goal of the presentation is to assist the audience in understanding how to perform a specific movement or traverse through the system, it may be appropriate to modify the system to induce specific behavior (e.g., increasing the volume at a roundabout to demonstrate an inbound vehicle yielding). However, care should be taken to communicate the intention of the visualization.

If the goal of the visualization is to demonstrate the performance of the system (relative to another model or to current day operations), a model that has been properly verified, calibrated, and validated should be used. This file can be selected from one of the analysis models. Guidelines for Applying Traffic Microsimulation Modeling Software (FHWA 2015, 65) discusses options for how to select a representative and/or worst case model.

It should be noted that some agencies consider simulation visualization—or animation—to be separate from the simulation process. While the same tools may be used for animation and simulation, some agencies have separate procedures and documents for each use of the tool.

Corridor Visualization

Heat maps, such as the three shown in Figure 163, can be especially effective at communicating the change in travel time, segment speed, average speed, and reliability along corridors over time. By modeling the corridor along the *x*-axis and time on the *y*-axis, variations in color allow the user to see how changes occur in both time and space.

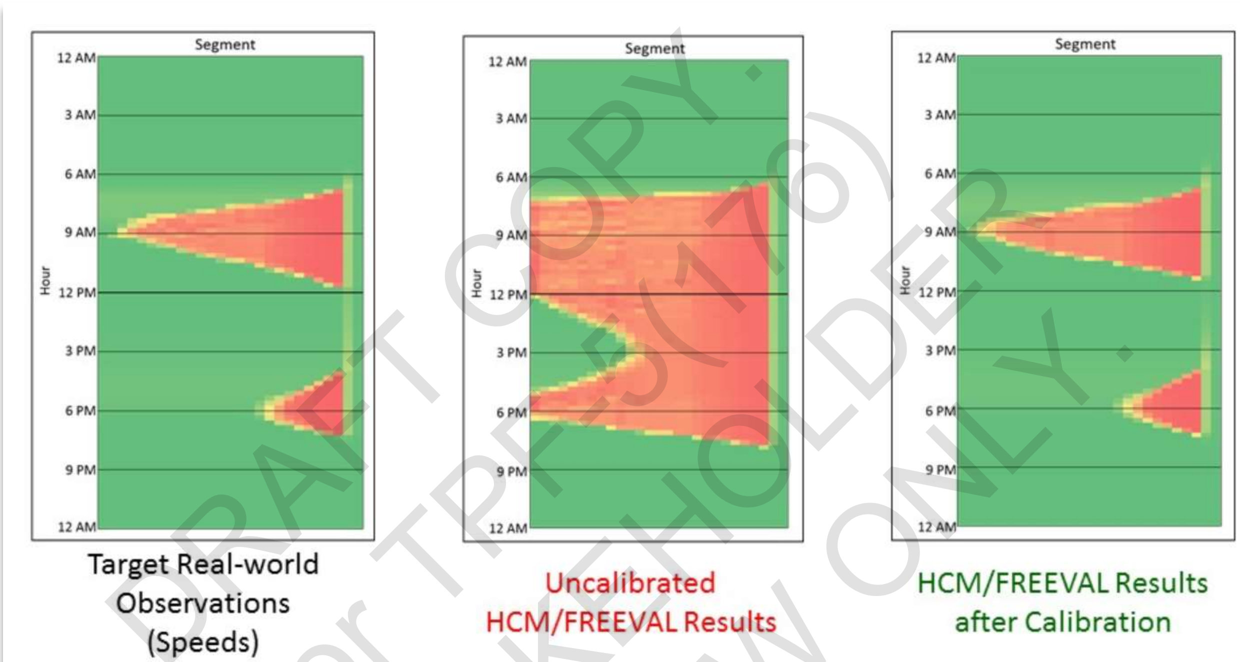


Figure 163. Charts. Example heat maps showing speed over time and across segments.

Tables versus Charts

Detailed discussion about using tables and the more common charts (scatter, bar, pie) can be found through the University of Leicester (2018).

Creating Accessible Visualizations

When creating a visualization, the analyst should consider how to convey the material to audience members with limited or no eyesight.

Color Selection

Many tools exist to assist in understanding how those with partial or full color blindness see the color spectrum as well as tools to assist in selecting accessible color palettes.

The Color Laboratory simulation controls allow users to see a color palette under nine different color visualization capabilities. Additionally, users can use the web page color filter to view any website under four types of color blindness.

Color Brewer 2.0 allows users to describe the type of data they have; and the system will provide options for colorblind, photocopy, and print friendly palettes (accessed).

Adobe Color CC provides appealing color palettes for use in various mediums.

Contextual Descriptions

In considering an audience focused presentation, it may be more appropriate to present distances in context as opposed to numerically. Specifically, for community engagement when the audience does not have the same frame of reference for distances as many practitioners in the transportation field do, referencing buildings and landmarks can be more accessible.

For example, when discussing queue lengths, instead of providing numerical reductions in distances—there is a 430-foot reduction in queue—locational placement should be used: the queue currently backs up to Cobb Elementary School but under alternative 1, vehicles are expected to queue only to Wendy's.

CHAPTER 10. POST PROJECTION VALIDATION

Simulation modeling efforts only have value if they help to inform decision-making about capital investments and operational changes. They have to provide predictions that are useful and defensible. Post-project validation can help in two ways: 1) checking the calibrated model was able to defensibly predict system's post-project performance and 2) improving the quality of the simulation analyses performed by the agency and the project team.

Ideally, simulation model predictions should be checked when the design year or planning horizon is reached. These are the conditions for which the improved facility was designed. But, typically, those times are far in the future. Moreover, the planned conditions may never materialize in exactly the way they are portrayed in the modeling work. The land use may evolve in unanticipated ways, vehicle technology may evolve, and the control technology may change (e.g., tolling may be introduced, and/or mileage taxes). In addition, the simulation modeling software may itself no longer be operable on the computer platforms then available. It is also possible the software may have changed enough that the predictions from future releases may be different from those obtained when the analysis was done. Finally, it is unlikely the project team which did the simulation modeling, and were familiar with the analysis, will be available.

Hence, it seems useful to focus on the conditions that exist once the project is finished. It does represent an opportunity to check the ability of the calibrated and validated model to predict future conditions; and it can serve as a basis for determining whether there are changes to the model and/or analytical process that will help improve the quality of the predictions obtained. It involves waiting only until the project is complete. The land use, vehicle technology, and other conditions are not likely to have changed dramatically. Moreover, the software used to do the assessment is likely to still be useable.

10.1 THE CONTEXT

Post-project validation offers the opportunity to improve the quality of the simulation models and analysis efforts without having to wait for design years to materialize. The idea is that post-project validation can become part of a continuous quality improvement program. To do this, the principal questions are: how should the assessment be done? What should be its scope? What data are needed? What metrics should be used? What outcomes should be expected?

There is a scoping question that arises. Should the prediction of the post-project load conditions be in or out of scope? That is, the assessment of future conditions involves two steps: 1) preparing input data files that describe the post-project conditions as they are, in case anything has changed since the original post-project analyses were performed, and 2) preparing fresh outputs that predict the system's performance for those conditions. If the prediction of the future load conditions (especially the traffic demands) is improved, then the performance predictions will more closely align with those that will be observed. And if the simulation effort is improved, then the predictions will be better for whatever load conditions are forecast and assessed.

This chapter focuses on the second step, preparing fresh outputs and checking the simulation model's predictions. Clearly, the first step is extraordinarily important. And if the predicted load conditions are different from those that materialize, the simulation model must be re-run for the conditions that do materialize for the validation effort to proceed. (This is effectively the same as implicitly including enhancements to the forecasting step and assuming that it can exactly predict the load conditions that do arise. The importance of this observation cannot be stressed enough. If, for example, the forecasting step predicted that facility use would increase by 5 percent, but it increases, instead, by 15 percent, then the forecasting process needs to be improved, the original simulation results cannot be used, and new simulation runs are needed. Similarly, if the temporal-spatial demands were assumed to persist for 30 minutes, but they exist for 1 hour, then, again, the original results cannot be used (checked) and outputs from new simulation runs are needed.

Clearly, insights may arise from this effort about changes to simulation model parameter values that would be useful, or better ways to represent the system's operation, better error-checking procedures, or ways to cluster situations into scenarios that are more useful, etc. But, the purpose is *not* to make these changes *ex post facto* so that the predictions are better. Rather, it is to make decisions about how these practices can be changed in the future. Put another way, this is *not* a calibration effort. It is an analysis of the model's ability to predict future performance, for previously unobservable load conditions.

The rest of this chapter describes the overall process of doing the post-project validation, presents a step-by-step process for conducting the validation, and gives ideas for how the findings from the effort can be used to improve the simulation modeling work. These ideas are discussed throughout in the context of all modeling scales from microscopic to macroscopic, as well as multi-scale models.

10.2 EVIDENCE OF THE POTENTIAL VALUE

One can ask whether such an effort would have value; or if agencies are doing this post-project validation already. In a recent survey, this question was posed to a collection of state agency representatives involved in doing or overseeing simulation modeling efforts. Their answers were as follows:

- One respondent indicated that post-project validation is typically done for signal timing improvements but not for geometric improvements (either arterial or freeway). Such analyses may be conducted if there are citizen complaints. This having been said, crash and safety post-project validations are conducted.

- The most supportive response was that, for very high-profile projects, follow-up is demanded by the stakeholders. However, the extent of validation depends on financial constraints. In high-profile projects, the performance of the network may be monitored throughout the duration of the project, and after completion. The agency sometimes uses consultants to collect this information.
- Another respondent indicated that while post-project validations are not performed, they would be beneficial. In one study, 50 percent of traffic forecasts were correct. But for those that were correct, they were right for the wrong reason(s). A consultant was hired to do a post-project evaluation of the travel demand modeling. Moreover, although the agency does not require post-project validations, some regions do it for their own edification. One reason it is difficult to mandate such studies is that even though they can be valuable for improving the work on future projects, it is hard to justify the funding.
- One agency recognizes that post-construction validations should be conducted, but they do not do them for every project. High-visibility projects, or projects that have opening-day problems, are reviewed. Having a build year, selected alternative model (the opening-day model) is helpful when issues arise on opening day. The agency can quickly reference the model to see which assumption(s) were incorrect and produced the dichotomy between the predicted and observed performance.
- Another agency indicated they are endeavoring to do such studies, either in-house or with consultants. But a challenge is collecting the field data. The sensors are often not active. The agency does ask consultants to provide opening-day results. And in that regard, traffic diversion is often an issue. Changes in traffic routing often result in differences of 5–10 percent in the flow rates; sometimes the differences can be as much as 15–20 percent.
- One comment was that such efforts can be risky because the entire analysis might be wrong due to unforeseen events. However, the models and data sets are kept for at least 3 years. This helps with monitoring traffic volumes, in addition to keeping track of the models employed.
- Another indicated that post-project analyses are sometimes performed, but not often.
- A few said these validations are never done.
- Several respondents indicated they wish they could do such analyses, but they do not have resources to make it happen.
- Another respondent shared a related thought that the agency is considering policies that would make it possible for them to save models and data sets.
- One respondent indicated that such an activity could have potential liability implications.
- Another respondent indicated that such studies would be valuable, but the funding is tied to the project, and once the project is built, the funding goes away; no budget is available for doing post-project analyses.
- One agency indicated that it does do before-and-after studies, but that effort rarely includes an examination of the simulation model to ensure that its predictions were consistent with those in the field.
- Finally, one other agency indicated that usually there is no post-project validation. Consultants might save data sets on a project-by-project basis, but there is no formal

process for doing so. However, the agency might develop a database for storing data sets in the future.

As these responses indicate, the value in conducting the post-project validations is typically perceived. The impediments seem to be funding and contracting arrangements, although risk and liability implications were mentioned. The related observation is that several agencies are considering creating databases that archive the models employed and the results obtained.

10.3 SUGGESTED PROCESS

The post-project validation effort should commence with data collection and conclude with comparative analyses and post-validation debriefing. The effort should be a validation, rather than a calibration, exercise, in that the model's parameter values should not be changed. They were set during the pre-project calibration effort, and they were asserted to be the right values to use for the analysis. Rather, the objective in this instance should be to compare the model's predictions of system performance against those observed in the field. The assumed parameter values can and should certainly be checked. And notes should be made about how they could be changed in the future. But for purposes of the validation, they should not be altered. It is also possible the model logic should be checked. A good example is the routing (rerouting) logic, where it has been employed. It might be assuming driver decision-making behavior that does not materialize in the field. It might also be useful to check the logic pertaining to gap acceptance, lane-changing, car-following, and other traffic-flow phenomena.

The steps involved in the process are as follows:

1) *Collect data regarding the post-project load conditions and performance*. This includes information about traffic flows, queues, delays, travel times, and capacities. The reason this needs to be done is so the project team can ensure the load conditions (and performance data) are consistent between the field and simulation model. As mentioned previously, it is quite possible the traffic flows are different from those assumed. And if this is true, the simulation model's predictions will be inherently different than those observed.
2) *Re-run the simulation model, if necessary*. If the field data suggest the post-project load conditions are different from those assumed in the pre-project analyses, the simulation model needs to be re-run to produce new results.[82]
3) *Compare performance*. Compare the performance predicted by the simulation model with that observed for the post-project condition(s). This step has the same attributes as the validation efforts portrayed in Chapter 8. .
4) *Identify modeling enhancements*. This step focuses on examining the findings from step three, and determining what enhancements to the modeling effort should be made to improve the quality of results. It might be the model parameter values need to be adjusted. Or the process by which the model is calibrated needs to be altered. But weaknesses in the simulation model logic (the software) might also be identified. The project team can take actions with regard to the first two. But changes related to the third need to be passed back to the software developer. And, until changes are made in that logic, the analyst will need to identify ways to work around the software limitations. A good example of something that might arise is issues related to lane choice management for vehicles that are about to make left and right turns or diverge at freeway off ramps.

---

[82] This could be possible if the planning model used to estimate the post-project traffic volumes failed to appropriately account for the impacts of the project. This is not a weakness or shortcoming in the simulation model, *per se*. Rather, it is a weakness in the planning model used to estimate what those flows would be. And clearly, if the flows that materialize are different from those the planning model predicted, the simulation model needs to be re-run before its predictions of post-project performance are compared with the performance observed.

5) *Indentify ways to improve the modeling process*. If the validation effort indicates the model parameter values (ranges) should be updated (e.g., because the performance capabilities of the vehicles can be better represented), the guidance documents kept by the project team (agency) should be updated to reflect the insights gleaned from the assessment. If it indicates the calibration procedure should be enhanced, the guidance documents can be updated to reflect the new insights.
6) *Document the findings*. A record of the results of the analysis should be created as well as the actions taken to make changes, and the rationale.

Following this procedure will result in an increasingly better process to follow in conducting the simulation analyses. And the parameter values employed in the simulations, and the guidance for determining what they should be, will be enhanced.

10.4 DETAILED GUIDANCE

Some guidance, like that provided in Chapter 8. , pertains to this effort. That guidance is presented here.

As with the verification, calibration, validation (VCV) effort described in chapter 8, the post-project validation involves examining load conditions to which the updated system is subjected. Each of those conditions involves combinations of: (1) the updated physical system,[83] (2) the updated operational controls,[84] (3) the new traffic demands,[85] and (4) the extant environmental condition.[86] This idea was illustrated in Figure 115. Illustration. Components of a load condition.. For example, if the objective of the project is to resolve a capacity issue, the project team should focus on how delays and queues have been reduced or eliminated during the peak-load conditions. The post-project observations should focus at least on normal weather conditions in the absence of incidents since that is the nominal condition for which the system should perform best. The project team might also want to test other abnormal load conditions such as adverse weather to assess the system's robustness. Certainly, if funding allows, post-project performance data should be collected for all load conditions for which improved performance was promised or anticipated.

As with the VCV effort, the post-project validation needs to simulate the episodic demands to which the system is subjected. Field data need to be collected so the episodic trends can be observed. These trends need to be compared with the simulation model's predictions.

As with the VCV effort, simulation quality metrics (SQM) should be chosen. These should indicate whether the simulation model's predictions of the post-project performance are acceptable or not.[87] They should again be two-tiered. Lower level SQMs should measure system

[83] Most importantly, the facilities and vehicles.
[84] How the system's operation is managed.
[85] The temporal-spatial matrices that describe the origin-to-destination (O-D) flows.
[86] Weather, maintenance, and other external activities that constrain or influence the system's operation.
[87] It is possible that the SQMs may need to be modified as the VCV process unfolds. That is fine. The main challenge, at the outset, is to identify a set of SQMs likely to provide the quality assessment desired.

performance for specific aspects of the model's predictions, such as travel rates and times, capacities, queue lengths, travel times, bottleneck discharge rates, and saturation flow rates at intersections. Upper level SQMs should focus on overall assessment metrics based on the lower level data (such as the sum of the squared percent errors). The best SQMs for a given analysis depend on the type of model used (micro, meso, macro, etc.) and the settings being examined.

Nominally, these SQMs should be consistent with, or identical to, those used in the VCV effort. On the demand side (traffic), the SQMs should measure the quality of the service provided. Travel times (and travel rates) are important, as are delays. The reliability of the travel times is also important, as in the probability density function (PDFs)[88] for travel times, travel rates, and delays for specific paths, for specific O-D pairs, and for the overall system.[89] Monitoring all the O-D pairs is certainly possible, but focusing on the top 10 (or top 20, top 100, etc.) O-D pairs is more practical and more useful.[90] Focusing on a limited number of O-D pairs helps identify significant problem spots in the network. The O-D pairs selected might have the largest flows. They might also be O-D pairs that pass through heavily congested locations in the network. It would be good to have at least one O-D pair for every bottleneck location.

On the supply side, the SQMs should ensure the performance of individual facilities is consistent with the field observations. This means focusing on travel times (and travel rates), queue lengths (predominantly on a per-lane basis), bottleneck discharge rates, segment capacities (segments having a direction), saturation flow rates at intersections, volume-to-capacity (V/C) ratios, and demand-to-capacity (D/C) ratios. If the simulation model can assess safety performance as well, then metrics like the frequency of near-collisions would be useful. A top-10 list of metrics on which to focus is useful here as well. These might be the fundamental diagrams for critical links in the freeway and arterial networks, distributions of queues at bottleneck locations, the delays at these same locations, distributions of corridor travel times, or O-D travel times.[91] The choice should be predicated on the likelihood the metric being monitored will always be in the top-10 list. The distributions are very helpful when making design decisions. At signalized intersections, the distribution of queue length is a logical choice for an SQM, with an associated thought that a specific percentile value may be of greatest interest. From a design standpoint, the 80th percentile value might be useful. This would mean the queue length would exceed that value only 20 percent of the time (one day a week). The higher-level assessment metric might be the sum of the squares of the deviations between simulated and observed 80th percentile values, at a 5-minute level of resolution.

In the context of the dynamics associated with peak period operations, such as the evening (PM) peak, good facility-level SQMs are the temporal patterns of queues, delays, arc flow rates, and

---

[88] A PDF indicates the percentage distribution of values for a variable whose values vary. For example, if a delay is 20 seconds 80 percent of the time and 10 seconds 20 percent of the time, then the PDF would indicate this was the case.
[89] It is always possible the average performance is acceptable, but the worst (higher percentile) performance is not.
[90] Another option is to identify all the bottleneck (problematic) locations in the network and ensure that a few O-D pairs are monitored for each one. It is also possible that intermediate, hypothetical O-D pairs are useful to monitor, such as the traffic going from node A to node B (regardless of the traffic's ultimate origin and destination) because that traffic passes through the bottleneck(s) of concern.
[91] Top 10 may also be replaced for a top 20, top 100, or some other basis.

travel times. A good system-level SQM is the length of time it takes the system to return to uncongested operation. This metric is easy to observe in the field, and if the simulation model predicts an erroneous value, something clearly needs to be recalibrated. It is even better to examine the PDF for this time and gain a sense of how short and how long the simulation model predicts this time might be and how the time is distributed.

Clearly, the overall performance of the system must be monitored, such as the average speed (travel rate), the average travel time, the aggregate delay, and the average delay. However, such aggregate metrics rarely provide clear insights about where the problems are or how to fix them. It is better to identify the most problematic locations in the network and focus on assessing the model's performance for those spots, individually, and combined.

The type of data that can be used to develop these SQMs depends on the type of simulation model being used (micro, meso, macro, etc.). Detailed data about individual vehicle trajectories are available for the microscopic models as well as aggregate measures provided as output by the simulation software. For mesoscopic and macroscopic models, queue lengths, flow rates, densities, speeds, travel times, delays, and shock wave measures are available.[92]

10.5 A BRIEF EXAMPLE

An example of doing this analysis is based on Route 7, locally known as Alternate Route 7, a short section of freeway just north of Albany, New York, that connects I–87 to I–787. The configuration of the facility is illustrated in

Figure 164,

[92] It is very important to remember that for microscopic simulation models, capacity is an output (a manifestation of the parameter values chosen) while for mesoscopic and macroscopic models, capacity is an input.

Figure 165, and

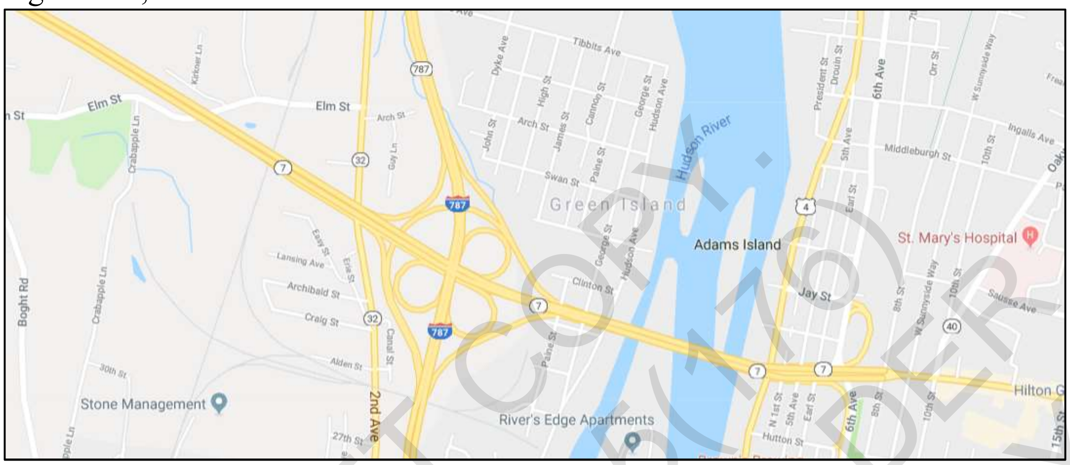


Figure 166.

Figure 164 shows the overall facility with its two main interchanges at either end.

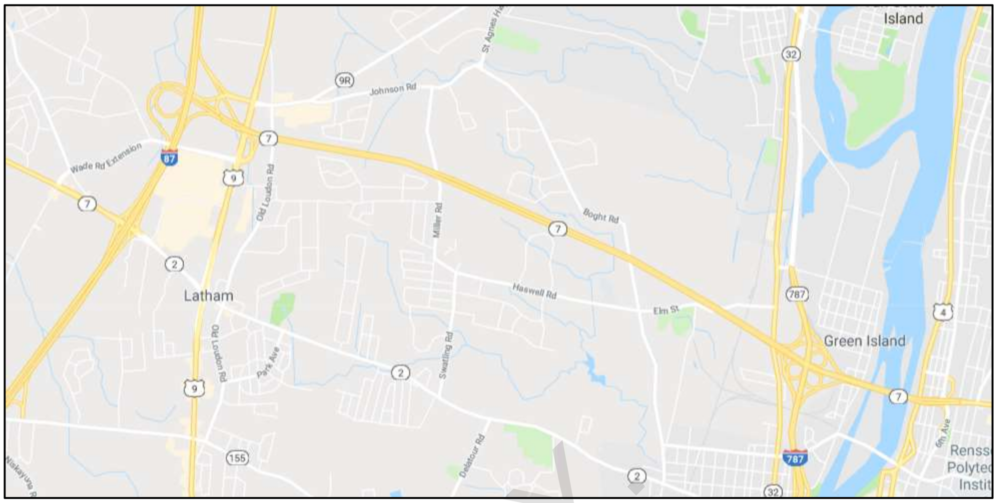


Figure 164. Map. Alternate Route 7, north of Albany, New York.

Figure 165 shows the configuration of the interchange at the western end of the facility.

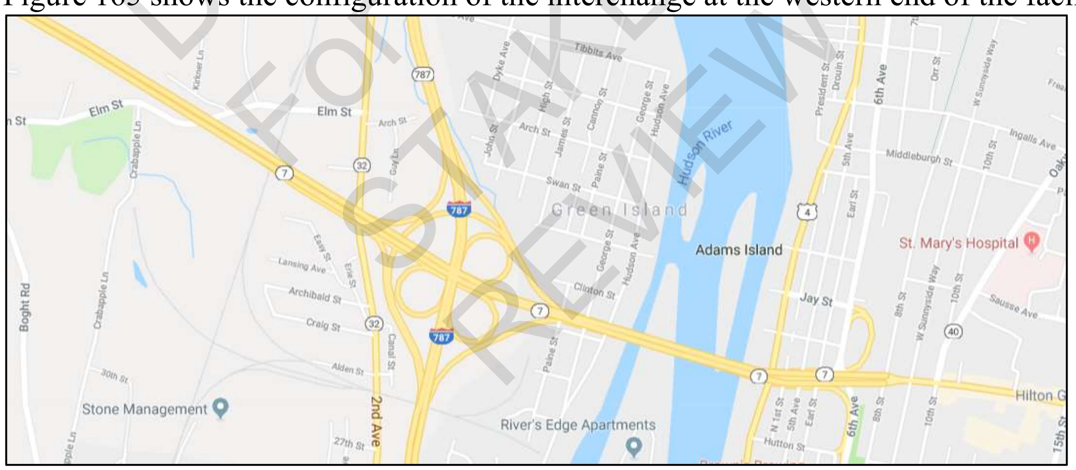


Figure 166 shows the configuration at the eastern end.

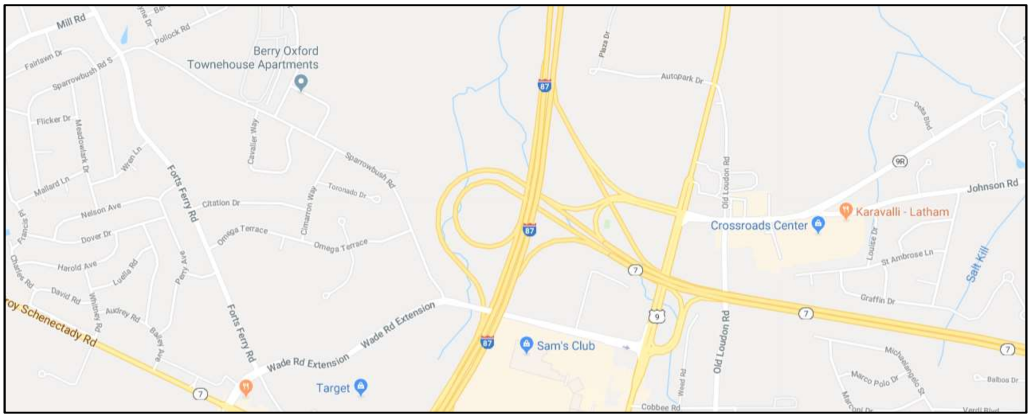


Figure 165. Map. Interchange configuration at the western end.

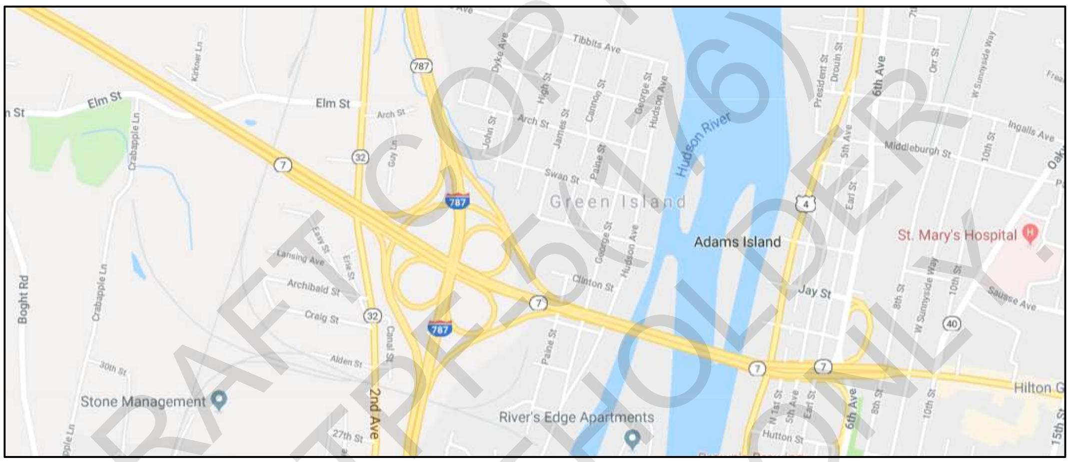


Figure 166. Map. Interchange configuration at the eastern end.

Clearly, the expectation was the large flows on the western end would be from Route 7 west toward I–87 south, and on the eastern end from Route 7 west to I–787 south. But neither of these conditions ever materialized. On the western end, the heavy flows are from I–87 south to Route 7 east, and on the eastern end they are from I–787 north to Route 7 west. These are daily commuter flows to and from downtown Albany.

This means the traffic forecasts for the design year were predicated on very different assumptions about traffic flow patterns than those that have materialized. (In their defense, the anticipation was the I–787 would continue eastward to Vermont at the interchange with Route 7. But construction of the freeway east of the I–787/Route 7 interchange never materialized.)

It also means the post-project traffic flow patterns would have been very different from those assumed for the design year. And since the facility transitioned from no flow to capacity during the peak periods within 2 years after construction, it is likely that even the post-project predictions of traffic volumes would have been very different from those that arose.

And because the heavy flows were imposed on the limited capacity of loop ramps and very short right-hand ramps, significant queues arose and large delays occurred. Even today, during the PM peak, the queue on Route 7 westbound extends about half the length of the distance between the interchanges. And the same is sometimes true in the morning (AM) peak for the right-hand ramp at the eastern interchange.

Hence, this would be a situation where the simulation model (if one was used) employed to study the performance of this facility might have predicted post-project performance that would have been very different from the conditions that did arise; and it would be important to see if the calibrated car-following, lane-changing model parameters would actually predict the queue dynamics observed on the ramps during the peak periods. Much could be learned about the quality of the simulation efforts being conducted based on this very simple network.

10.6 SUMMARY

This chapter has described the way in which project teams can use post-project validations to improve the quality of their simulation models and analysis efforts. The use of simulation to inform decision-making about capital investments and operational changes only has value if its predictions are defensible. So, a process of continuing quality improvement needs to exist. The principal questions to be answered are: Does the model produce credible predictions? Does it indicate how the system will actually perform once changes are made? Can it capture the nuances in performance observed in the field?

Nominally, it would be best to check the quality of the predictions for the design year or the planning horizon conditions, but typically those situations are far in the future; it is not reasonable to expect the simulation model or its results will be held ready for reexamination for such a long period of time. Nor is it likely that the actual planned conditions will materialize. Land use will evolve in unanticipated ways, vehicle technology will evolve, the operational technology will change (e.g., tolling), and the simulation modeling software used to do the analysis may no longer be available or it may not work on the computer platforms available. It is also unlikely the people who did the simulation modeling and are familiar with issues that arose, will be available to do the assessment. And a feedback loop that involves an approximate 20-year lag is not particularly helpful.

Hence, this chapter suggests that examining the model's predictions of system performance after the alterations are complete would serve the same purpose and be far more useful. Predictions of system performance could be created and then checked against the performance observed once the changes are implemented; after the facility is built or modified; or the operational changes are put in place

VOLUME III. CASE STUDIES

This volume contains case studies that illustrate how the guidance provided in volumes I and II should be and has been applied in real-world analyses. In that this first edition of the TSSM may be presenting ideas that have not been used previously in practice, the examples presented here, presently, do not necessarily follow the manual's guidance in all respects. Moreover, analysts are always empowered to deviate from suggested practice when their assessment suggests that is the best path to follow.

Consistent with the current use of simulation models, especially microscopic ones, examples are provided for three contexts: surface arterials, freeways, and combined arterial-freeway networks. Chapter 12. deals with combined freeway-arterial examples; Chapter 12. , freeway examples; and Chapter 13. , arterial examples. Also, since, as indicated, the examples predate the existence of the Transportation System Simulation Manual (TSSM), rather than using case studies that embrace all the ideas TSSM presents, which would likely have yielded none, each chapter cherry picks work across several case studies to highlight ways in which the practices suggested by TSSM have been followed. (Over time, these case studies will be replaced by others that more fully embody the practice suggestions TSSM puts forth.)

Consistent with the structure of volume II, the case study chapters highlight major issues deemed to be important in chapters 5–9. An example of arterial topics stressed, by chapter, is shown in Figure 167.

Boise	VDOT	Moana	
			Chapter 5: Defining the Problem
Yes	Mostly	Mostly	• Define scope, goals, approach, budget, objectives, performance measures
Yes	Yes	Mostly	• Study boundaries (temporal and spatial)
Yes	No	Yes	• Role of microsimulation (should just be "simulation")
Partial	No	Partial	• Formation of project team and stakeholders
Partial	No	Partial	• Identify needed resources
Base	Base	Base	• Define forecast and base years
No	No	No	• Range of operational conditions to consider
No	No	No	• Screening of initial alternatives through non-simulation methods
Partial	No	Partial	• Explicit linking of performance measures to agency goals or project objectives
			Chapter 6: Data
Yes	Yes	Yes	• Detailed description of the origins of data used in the analysis
No	No	No	• Description of speed measurement method
No	Yes	No	• Collection of vehicle type classifications
No	No	No	• Pedestrian volume counts
Yes	Yes	Yes	• Obtaining signal timings from the field or the DOT
Yes	Yes	Minimal	• Collection of calibration data (e.g., queuing, travel times, throughput)
No	No	No	• Collection of crash data and roadway alignments
Aerial	Aerial	Aerial	• Use of probe data, NPMRDS, vehicle trajectories, aerial imagery
No	No	No	• Application of data processing principles (e.g., aggregation, fusion, imputation)
No	Minimal	No	• Statistical analysis of field data sample size
			Chapter 7: Creating the Network
No	No	No	• Intelligent selection of initialization fill time
No	No	No	• Analyzing the consequences of inadequate study boundaries
No	No	No	• Informed selection of overall data entry method
Yes	No	No	• Use of pre-processors and/or OpenStreetMap
Yes	No	Yes	• Entry of demands instead of throughput volumes
Yes	No	Yes	• Analysis of O-D demands
No	No	No	• Application of a specific volume balancing method
No	No	No	• Application of conditional turn movements
No	Yes	No	• Application of data entry methods to achieve pre-positioning and custom lane utiliza
No	No	Yes	• Entry of detector (sensor) data
No	No	No	• Entry of freeway warning sign distances
No	No	No	• Entry of operating condition data
Yes	Yes	Partial	• Selection of key input parameters for calibration
Yes	Yes	Partial	• Identification of key network locations for calibration

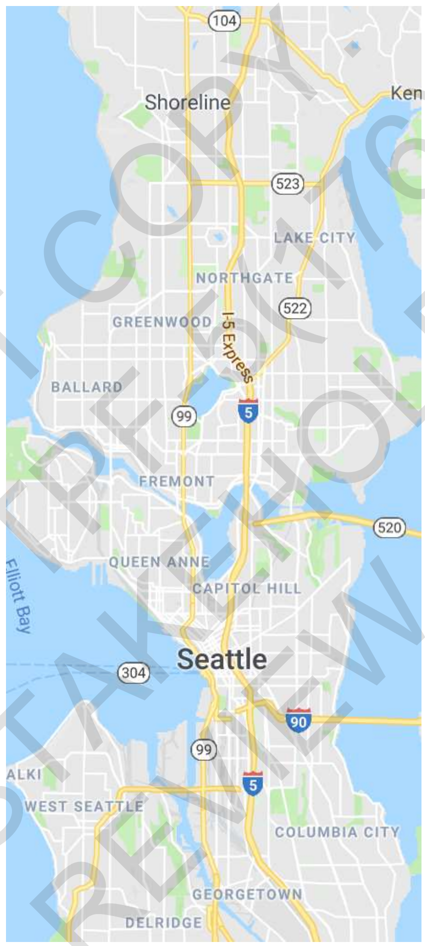
Boise	VDOT	Moana	
			Chapter 8: VC&V
Yes	Yes	No	• Analysis of WHY the results are incorrect before deciding on next step
Yes	Yes	No	• Verifying that the software can model conditions of interest
Yes	Yes	No	• Verifying that the input data are clean, congruent, and accurately repres
Yes	Minimal	No	• Validating the model's predictions for a broad range of conditions
Yes	Yes	No	• Selection of both upper (e.g., sum of squared percent errors) and lower
Yes	No	No	o Rank the SQMs in terms of importance
Yes	Yes	Yes	• Focusing on a limited number of critical OD pairs
No	Yes	No	• Evaluation of PDFs or CDFs that describe distributions of the performa
No	Minimal	No	• Ensuring that links perform in accordance with the fundamental diagram
No	No	No	• Analyzing reasonableness of vehicle trajectories
No	No	No	• Using vehicle trajectories to calibrate certain parameters (e.g., car follov
No	Minimal	No	• Calibrating the speed-density-flow relationships for macroscopic model
No	No	Yes	• Having "a second set of eyes" review the data
Yes	No	No	• The calibration effort needs to deal with episodic demands
Yes	No	No	• Selection of a preferred calibration sequence (e.g., driving behavior, rou
No	Yes	No	• Identification of key problem spots where performance must be predict
No	Yes	No	• Validate with separate datasets to see if the calibrated model can produ
Yes	Yes	No	• Review the best-case and worst-case scenario outcomes
			Chapter 9: Data Processing, Analysis, and Presentation
No	No	No	• Using a naming structure for the scenario datasets
No	Yes	No	• Automated post-processing analysis (e.g., commercial statistical softwa
No	No	No	• Weighted aggregation of performance measures
No	Yes	Yes	• Analysis of stochastic model outcomes under different random number
No	No	No	o Determining the number of replications needed
No	No	No	• Intelligent choice of bin size for histograms and probability density functi
No	Yes	No	• Use of probability density functions and cumulative density functions
No	Yes	No	• Use of confidence intervals
Yes	Yes	No	• Use of statistical hypothesis testing
No	Yes	No	• Analysis of lane-specific results
Yes	Yes	Yes	• Use of multivariate outputs and accompanying visualizations
No	No	Yes	• Presentation of percent changes instead of raw numbers
Yes	Yes	Partial	• Final documentation of assumptions, unknowns, and recommendations
Yes	Yes	Minimal	• Presentation of static graphics and moving vehicle animation

Figure 167. Chart. Arterial topics stressed in the case studies, by chapter.

CHAPTER 11. COMBINED FREEWAY-ARTERIAL STUDIES

This chapter discusses three traffic simulation case studies for combined freeway-arterial networks. Also, it cherry picks work across several case studies to highlight ways in which the practices suggested by the Transportation System Simulation Manual (TSSM) have been followed. (Over time, these case studies will be replaced by others that more fully embody the practice suggestions TSSM puts forth.)

The first case study was sponsored by Washington State Department of Transportation (WSDOT). It focused on I–5 in Seattle. The approximate study area is shown in Figure 168.



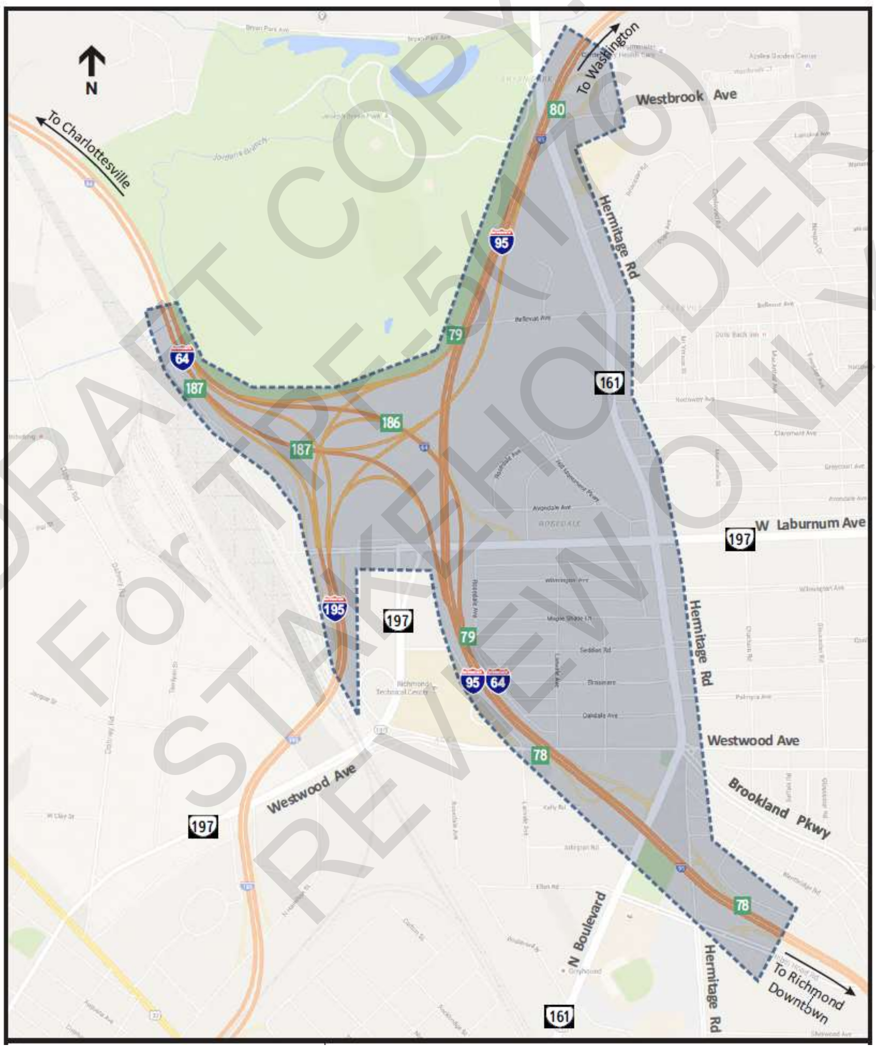
© Google.

Figure 168. Map. Approximate dynamic traffic assignment (DTA) study area.

The report for the study, entitled I–5 DTA Model Development North Segment: Corson Street to SR 104, was prepared for WSDOT in 2014. The report provides the following introduction:

The Washington State Department of Transportation (WSDOT) requested support to develop and maintain an integrated mesoscopic Dynamic Traffic Assignment (DTA) model for evaluating location-specific and corridor-wide transportation strategies on a selected "pilot" segment of the I-5 corridor. The intent of this work was to establish a preliminary DTA model network based on Dynameq (INRO) software that focuses only on I-5 from Corson to King/Snohomish Co. Line MP 161.23 TO 177.82. The project also assumes the use of a readily available travel demand model (TDM) from the Puget Sound Regional Council (PSRC) to support development of a stand-alone pilot-segment-only DTA model for the above I-5 segment.

The second case study focuses on the Bryan Park Interchange between the I–95/I–64/I–195. The interchange is shown in Figure 169.



© Virginia Department of Transportation.

Figure 169. Map. Bryan Park Interchange study area.

The project is described in the report, Bryan Park Interchange Dynamic Lane Merge Assessment (Schneider Electric and RK&K 2015), prepared for VDOT. The report provides the following introduction:

> *The Virginia Department of Transportation (VDOT) has identified the need to improve traffic mobility and relieve recurring congestion at the I-95/I-64/I-195 Bryan Park Interchange. Within the project area, three (3) interstates (I-95, I-64 and I-195) converge, and given the proximity to Bryan Park (which bounds the north end of the interchange), this convergence is known locally as the Bryan Park interchange (shown in Figure 1). The Bryan Park Interchange serves as the primary convergence point for the traffic originating from northern Richmond and Henrico County. The I-95/I-64 overlap section provides direct access to Downton Richmond while the I-195 segment provides access to western part of the Downtown area. The confluence of these uses, plus the heavy merging and through traffic volumes on I-95/I-64 at the interchange, leads to heavy congestion in this interchange area.*

The third case study is a study of I–40 in Raleigh, North Carolina. The network-level model is shown in Figure 170.

Figure 170. Illustration. Mesoscopic base network.

The red lines are freeway facilities with a total of 1,165 links and 649 on- and off-ramps. These include I–40, I–85, I–440, and I–540 as the major interstate freeways in the network. The blue lines are highway facilities (principal arterials, 2,644 links) and black lines represent arterial roads (minor arterials and collectors, 15,792 links). The mesoscopic network comprises 2,389 traffic analysis zones, 9,527 nodes, and 20,250 links.

The project is described in the 2014 report, Work Zone Traffic Analysis & Impact Assessment (Schroeder et al. 2014). While the analysis included both a mesoscopic and macroscopic model, this case study will focus on the mesoscopic model. The report provides the following introduction:

> *The objective of this research is to quantify and illustrate network-level impacts from the I-5311/I-5338 project work zone. The work zone is scheduled to be deployed on Interstate Highway 40 from US1 interchange to I40/440 split, and continues on Interstate highway 440 from I40/440 split up to US64/264 interchange. The work zone section extends from Exit 293 on I-40 through Exit 301 at I-40 and Exit 14 on I-440, with a total length of around 11 miles. Exhibit 1 shows the work zone extent on a map of the area.*
>
> *The [mesoscopic] network used in this project was obtained from NCDOT project HWY-2009-05 (Williams et al., 2001). The network was converted from the TRM (Triangle Regional Model), which was developed in 2010 for the planning year 2015.*

11.1 DEFINING AND SCOPING THE SIMULATION ANALYSIS PROBLEM

0describes the need to define and document the scope, goals, approach, budget, objectives, and key performance measures for any traffic simulation study. For the most part, all three case studies accomplish this.

Fundamental Project Parameters

The I–5 report (WSDOT 2014) documents the general objectives and scope in the introduction. The Bryan Park report (2015) details the purpose of the report as an operational assessment of various modifications including implementation of dynamic lane merge (DLM):

> *The operational needs intended to be addressed by the proposed DLM modifications include existing and future congestion within the Bryan Park Interchange area. The proposed DLM modification will utilize the I-95 northbound excess lane capacity dynamically, and relieve congestion along the I-64/I-195 ramp by providing an additional merge lane towards I-95 northbound. Furthermore, the Exit 80 (Hermitage Road) off-ramp will be permanently closed to eliminate potential unsafe weaving maneuvers downstream of the I-64/I-195 merge area. Exit ramp traffic will be rerouted within the study area (discussed in the further sections of this report). Additionally, the intersection of Hermitage Road at N Boulevard/Brookland Parkway/Westwood Avenue is modified as a roundabout to accommodate rerouted traffic and improve existing traffic operations at the intersection.*

The I–40 case study's (2014) primary objective was to understand the travel time impacts on the corridor under construction as well as the alternative routes under various work zone configurations.

Spatiotemporal Boundaries

5.2 emphasizes the importance of defining an appropriate set of spatial and temporal study boundaries, which can capture all relevant congestion.

The I–5 report (2014) includes extensive details of the freeway spatial extents and, to a lesser degree, the arterial extents stating:

> *The study area of the DTA model includes an approximately 16-mile section of the I-5 corridor between Corson (MP 161.23) and the King/Snohomish County line (MP 177.82). This includes all freeway interchanges along this section as well as signalized intersections immediately adjacent to these ramps. For some areas through downtown Seattle, additional arterial segments and signalized intersections were represented to reflect travel patterns to/from local streets, especially in highly congested zones between Denny Way and James Street. These external intersections were included to evaluate the extent of the queues, if any, both on I-5 and the local arterials. The study area does not include alternate parallel routes to I-5…The DTA model will be developed for two weekday periods: AM peak period from 06:00 to 09:00, and the PM peak period from 15:00 to 18:00.*

The Bryan Park report (2015) provides the following description of the study zone:

> *The study area includes mainline I-95/I-64 Overlap from the Exit 78 (Hermitage Road) interchange to the south to just north of the Exit 80 (Westbrook Avenue) interchange. Hermitage Road was analyzed from the I-95 northbound terminal to Westbrook Avenue, including the signalized intersections at Laburnum Avenue and Brookland Parkway/Westwood Avenue. Laburnum Avenue was analyzed from the I-195 off-ramp to the intersection of Hermitage Road.*

The I–40 case study (2014) spatial limits are included in Figure 170. While a previous model was used for the starting point of this model, extensions were made to expand the model to the east for inclusion of additional potential alternative routes drivers were expected to select. The temporal limits were a 4-hour morning (AM) peak period (6–10 a.m.) and a 4-hour evening (PM) peak period (3:30–7:30 p.m.).

Role of Simulation

In the I–5 case study (2014), the role of simulation was discussed during the introduction. The report discusses the benefits of a mesoscopic model stating:

> *A DTA modeling platform for the I-5 corridor was deemed suitable for the purposes of investigating corridor-level performance and eventual route and pathway diversion (as*

the model network expands) due to its blending of traffic assignment capabilities with the intersection/link operational analysis characteristics of traffic simulation tools thereby bridging the "gap" between the more commonly used macroscopic and microscopic paradigms.

The Bryan Park report (2015) indicated that analytical methods from the Highway Capacity Manual (HCM) were used for analysis of merges, diverges, and weaves along some segments of the study area; however, it was determined that microscopic simulation would "best capture the impacts of DLM and upstream/downstream congestion in the study area."

The I–40 report (2014) provided a strong justification for the use of mesoscopic modeling:

Mesoscopic simulation models provide a more computationally-efficient approach to analyzing large transportation networks [than microscopic modeling], by using macroscopic traffic stream relationships (e.g., link-based speed/flow/density) to model behavior of individual vehicles. Mesoscopic models have previously been used for very large network analyses, for example for predicting traffic patterns from a downtown evacuation event (Kwon et al, 2005), as well as evaluation of Active Traffic Management Strategies in the Triangle region in North Carolina (Williams et al, 2011). Mesoscopic models implement DTA algorithms that can update routing decisions of vehicles based on capacity reductions (from work zones), congestion patterns, and even impacts of deploying various technologies for enhanced traveler information. For this research, mesoscopic tools are therefore ideally suited for modeling a large and complex network efficiently, while allowing for evaluation of diverse traffic management strategies.

Formation of Project Team and Stakeholders

While each case study details the customer of the report (i.e., WSDOT, VDOT, NCDOT), few details were provided about other stakeholders or how the project teams were formed.

Identify Needed Resources

The I–5 report (2014) used Dynameq software, travel demand model, hourly counts for roadways and ramps, intersection turning movement counts, speed data, travel times, Google® Maps Directions application programming interface (API), signal timing and phasing data, and ramp metering rates as modeling resources. The Bryan Park report cited VISSIM software, Highway Capacity Software (HCS), traffic volumes, a previously created microscopic model, and previously created Synchro files as necessary resources.

The I–40 report (2014) cited the use of DynusT, DTALite, origin-destination (O-D) matrix, point-based traffic volumes, point-based speed estimates, route-based travel time and speed estimates, and a list of planned construction projects that were included in the model from which they started but needed to be removed. The model from which the team started had a base year of 2015 while the base year for the case study was 2011; therefore, there were some significant alterations to the alternative routes which would not yet have been implemented.

Define Forecast and Base Years

In I–5 model was created to be used as a model for future alternatives analysis. As such, the base year of 2015 was modeled as was a future baseline year of 2025.

The Bryan Park model base year was 2011 with a forecast year of 2022 selected for the DLM alternatives analysis.

The I–40 model used the same base year and forecast year (2011). It is implied the construction had already started and recommendations for the work zone would be almost immediately implemented.

Range of Operational Conditions to Consider

The I–5 report (2014) does not indicate the creation of future year models for specific operating conditions. However, as the model is to serve as a base for future alternatives analyses, those analyses could include more detailed travel time reliability testing including varying traffic demands, weather, incidents, work zones, and special events. The neither the Bryan Park nor I–40 reports indicate operational conditions were considered.

Screening of Initial Alternatives through Non-Simulation Methods

As the I–5 model was not created to directly consider alternatives, there was no discussion of screening of such alternatives. The Bryan Park report (2015) indicates the entire corridor was modeled in HCS 2010; however, the impacts of DLM were not able to be captured by analytical software, so microsimulation was necessary to compare alternatives.

Part way through the I–40 evaluation (2014), the team switched from DynusT to DTALite. Initial results from the DynusT model showed a few of the alternatives were infeasible and therefore unlikely to be constructed; therefore, they were removed from the list of alternatives tested in DTALite. While DynusT is not a non-simulation method of screening alternatives, the process of reviewing results and determining feasibility still applies.

Explicit Linking of Performance Measures to Agency Goals or Project Objectives

As the I-5 model was not created to directly consider alternatives, there was no discussion of performance measures for evaluating alternatives.

The Bryan Park report (2015) states the performance measures of density and speed were selected for the freeway analysis to "verify that the proposed modifications will provide more efficient and safer operations along the freeway."

The I–40 report (2014) stated the goal of the research was to "estimate how severe queuing impacts on the freeway [were] expected to be and more importantly, how drivers [were] expected to adjust their travel patterns towards alternate routes." As such, a key performance measure used was percent of route diversions. Table 25 shows the estimated diversion needed to achieve the optimum diversion travel times and volumes. Queue length was evaluated using a macroscopic tool and is covered in the Freeway Case Study (Schroeder 2014).

Table 25. Partial table showing travel times and volumes for no-diversion versus optimum-diversion models and estimated percent of diversion necessary to achieve those values.

Route	Work Zone	Direction	Pre-Construction Travel Time (min)	No-Diversion Travel Time (min)	Optimum-Diversion Travel Time (min)	Pre-Construction Volume (vph)	No-Diversion Volume (vph)	Optimum Diversion Volume (vph)	Estimated Diversion
Route A (I-40 Milepost 293 through 300)	WZ 1	EB	8.0	8.0	7.9	3333	3254	3173	-5%
		WB	8.4	39.7	12.7	4541	1875	2078	-54%
	WZ 5	EB	8.0	8.0	8.0	3333	3314	3167	-5%
		WB	8.4	22.1	10.7	4541	3060	3037	-33%

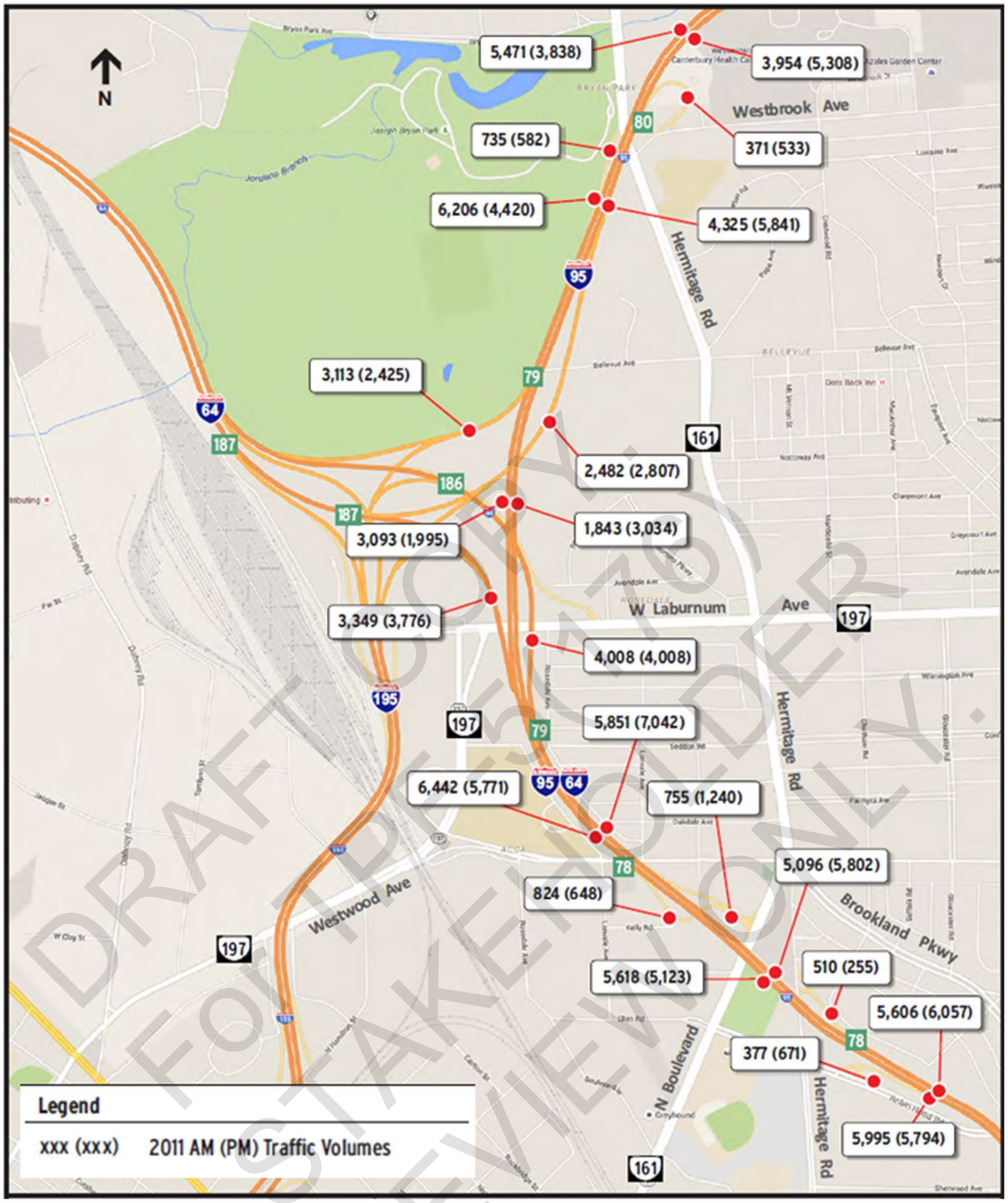
Source: North Carolina Department of Transportation.

11.2 DATA

Origins of Data Used in the Analysis

The I–5 report (2014) categorized data used for model development and calibration into four groups: volumes, speeds, travel times, and intersection control. WSDOT provided hourly volumes for roadways as well as ramps, while turning movement counts were collected both from WSDOT and the City of Seattle. Tabular and contour maps were provided by WSDOT detailing average speeds by time and location. Google® Maps Directions API was used to estimate most of the travel times for calibration for most origin-destination paths that were not exclusively contained on I–5. WSDOT was able to provide point-to-point travel times for the I–5 corridor. The City of Seattle, City of Shoreline, and WSDOT provided signal timing and phasing data.

The Bryan Park model was built off a previous study, so much of the traffic volume and network data was sourced from the previous model. Figure 171 shows the peak hour traffic volumes along the freeway facilities. Additional volume data for four intersections were collected through field observation and from VDOT.



© Virginia Department of Transportation.

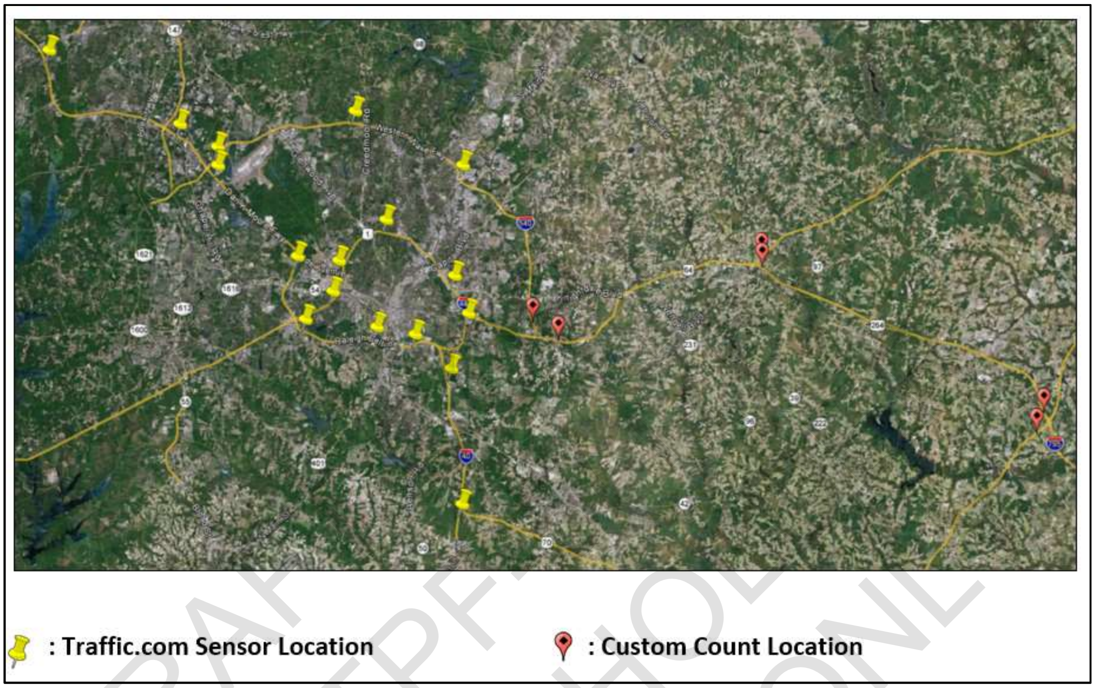
Figure 171. Map. 2011 Peak hour volumes along freeway facility.

The I–40 report needed the following data:

- O-D matrix obtained from the TRM.
- Point-based traffic volumes and speeds obtained from traffic.com (https://wego.here.com/traffic) side-fire radar stations as well as field data collection by NCDOT at six key locations and arterial traffic counts on likely diversion routes.

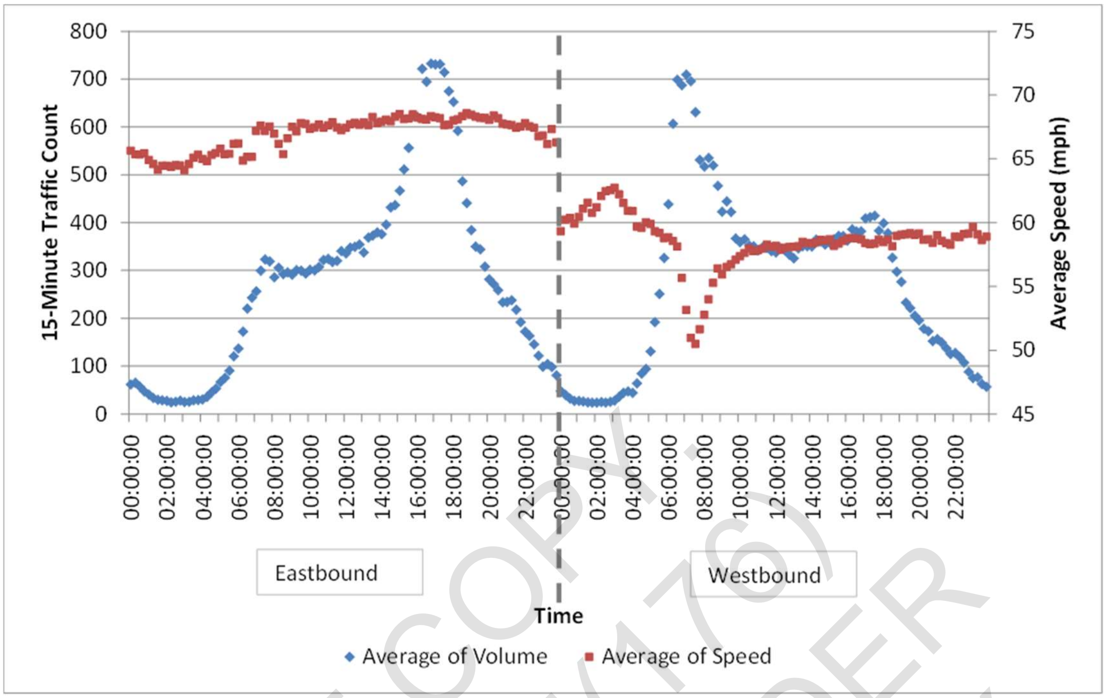
- Route-based travel times generated from INRIX.com probe-based travel time data along segments.

Figure 172 shows the location of the traffic.com sensors as well as the custom count locations requested form NCDOT. Figure 173 shows a speed-volume provide for one traffic.com sensor on I–40 aggregated over 15-minute periods.



© North Carolina Department of Transportation.

Figure 172. Map. Point-based data collection sites.



© North Carolina Department of Transportation.

Figure 173. Chart. Speed-volume profile for a sensor on I–40.

Description of Speed Measurement Method

The I–5 report (2014) indicates time mean speeds were provided by WSDOT, as gathered at specific point locations. Time-mean speeds are measured directly at a single point in contrast to space-mean speeds which require collecting travel times over a short segment and deriving the speeds by dividing the collected travel time by the distance of the segment. While travel times were also collected for this work, the report implies these were used for calibration of O-D travel times as opposed to segment speeds.

The Bryan Park model did not discuss the origins of the speed data.

The I–40 report used both time-mean speeds from side-fire radar as well as segment-based travel times and space-mean speeds from INRIX.com.

Collection of Vehicle Type Classifications

- While the I–5 report (2014) details the classes of vehicles modeled, it does not indicate the underlying data that led to the inclusion of classes. Neither the Bryan Park (2015) nor I–40 report (2014) discusses which vehicle types were modeled or how that data was collected.

Pedestrian Volume Counts

Pedestrian counts were not discussed in any of the reports.

Obtaining Signal Timings from the Field or the Department of Transportation

As discussed in the Origins of Data Used in the Analysis section, signal timing for the I–5 model was obtained both from the Department of Transportation (DOT) and cities along the corridor. Additional intersection details were imported from Synchro.

The I–5 model also included ramp metering. While the entrance ramps used a dynamic rate setting algorithm, the DTA model required a fixed-time rate. Therefore, the modeling team used the on-ramp flow rate as a first approximation of the metering rate, as described in the report:

> *For example, a single lane signalized on-ramp with a peak hour flow of 600 vehicles per hour was initially estimated to meter vehicles at a rate of 6 seconds per vehicle. Field observations during weekday AM and PM peak periods generally validated and confirmed these rates. However, some discrepancies were noted for selected on-ramps and adjustments in the estimated metering rates were made prior to inclusion in the existing conditions model.*

The Bryan Park report (2015) does not discuss the origins of signal timing data for the four intersections that were added to the model. Five other intersections were included in the microscopic model from which the Bryan Park model was derived. The signal timing plans for those intersections were likely already developed in the original model.

The I–40 report (2014) did not discuss signal timings.

Collection of Calibration Data

In the I–5 model discussion on model development methodology, it was noted that observed traffic counts and speed data along I–5 were used for calibration although specific details regarding field collection were not included.

As the Bryan Park model was largely taken from a previously calibrated model, there are few details regarding additional calibration conducted on the updated model. Traffic patterns were collected from Google® Maps and used as a visual calibration of freeway level of service for the base year.

In the I–40 report (2014), both the traffic.com point data and INRIX.com segment data were available in 15-minute aggregates for multiple years. The team decided to use Tuesdays, Wednesdays, and Thursdays, from the 2nd week of the month for all of 2011 as well as January, 2012, for a total of 39 observations.

Collection of Crash Data and Roadway Alignments

None of the reports mentioned safety-oriented data, such as crashes. Roadways alignments were also not mentioned.

Use of Probe Data, National Performance Management Research Data Set, Vehicle Trajectories, Aerial Imagery

While not explicitly mentioned, the I–5 model likely used probe data secondhand through the Google® Maps Directions API. The API was used to collect O-D travel times along the corridor including travel along surface streets. Google® uses historical and real-time probe data—among other data sources—to generate travel time estimates.

The Bryan Park report (2015) also used Google® Maps to gather typical traffic data.
The I–40 report used traffic.com side fire radar data as well as INRIX.com probe-based data.

Application of Data Processing Principles (Aggregation, Fusion, Imputation)

The I–5 report (2014) contained extensive detail regarding the demand matrix preparation for the DTA model in the Demand Matrix Preparation section. The description of the overview is provided below and additional details for each step can be found in the full report.

> *The current year demand for the AM and PM peak periods along the I-5 corridor were based on an Emme static model which was originally developed in 2014 to study the I-90 corridor in the Seattle area. This static model – henceforth referred to as the "4K Model" – contains 3,865 zone centroids, eleven classes, and includes the entire I-5 corridor study area.*
>
> *The demand calibration for the I-5 corridor DTA was originally intended to be based on the Puget Sound Regional Council (PSRC) Emme static model, which was developed in 2015 and provided to INRO by WSP, however, upon closer inspection it was revealed that this model lacked the detail required for a proper DTA conversion and the 2014 I-90 4k Model was used instead.*
>
> *The process of calibrating the current year demand involves the following procedures:*
>
> - *The conversion of the older 4K Model macro-based Standard traffic assignments to SOLA traffic assignment specifications compatible with Emme 4.3.2.*
> - *Defining the nodes, centroids, and links within the 4K Model which correspond to the I-5 study subarea.*
> - *Performing minor tweaks and corrections to the 4K Model to reflect count data and to ensure both intersections and links are a reflection of real-world geometry.*
> - *Performing a subarea extraction (including external gate traversal analysis) of the I-5 study area.*
> - *Conversion of the 3-hour subarea network to a 1-hour network.*

- *Processing and inputs of count data, in 30-minute increments, to the I-5 subarea network.*
- *Performing a demand adjustment using the weighted real-world count data, resulting in an adjusted subarea demand matrix for each of the eleven traffic classes in 30-minute increments (six 30-minute time slices for each peak period, 12 slices in total).*

Figure 174 shows a plot of the traffic counts versus modeled volume following the demand adjustment where a dot along the line y=x indicates the observed volume was equal to the modeled volume.

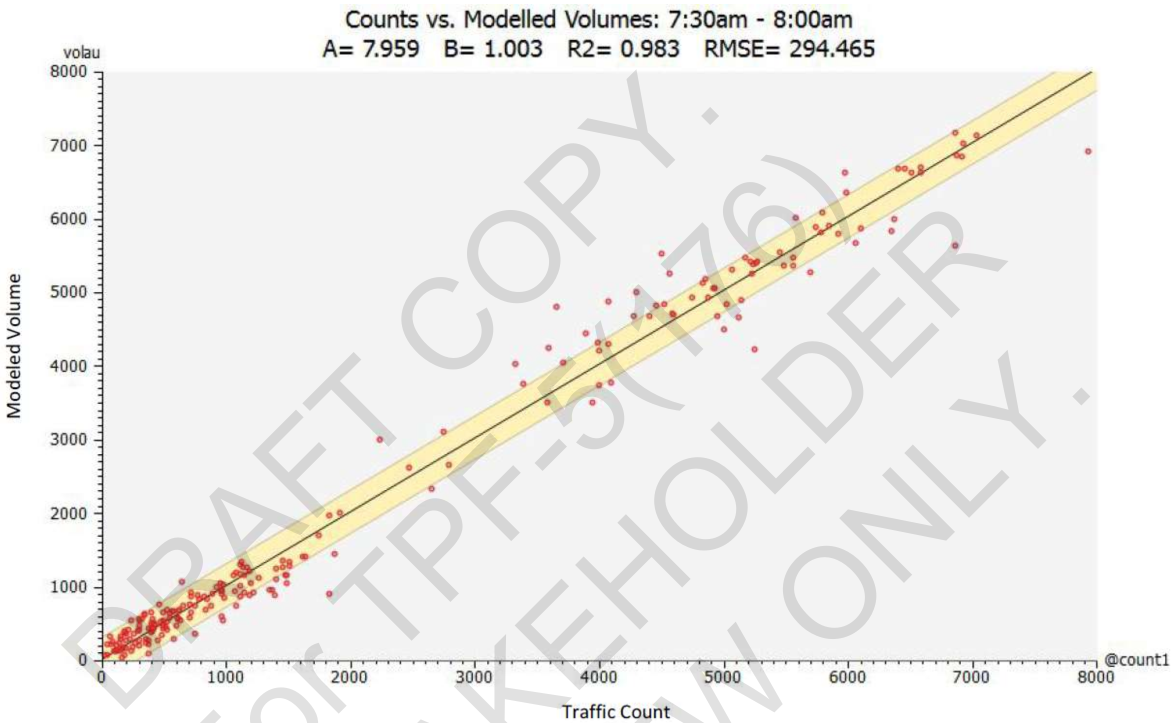


Figure 174. Chart. Estimated and observed traffic volume comparison.

In the I–40 report, the INRIX.com segment-based data was aggregated to provide route-based travel times. In stitching segment data together to generate route data, consideration should be taken into the method applied; that is, a decision needs to be made about what time period is used for each segment. Consider a scenario in which a route has 10 segments and data is aggregated at the 5-minute interval. If the travel time over the first two segments is 4.5 minutes using data from 7 to 7:05 a.m., it may be prudent to then use the travel time from 7:05 to 7:10 a.m. for travel time over segment 3.

Statistical Analysis of Field Data Sample Size

Neither the I–5 (2014) nor Bryan Park report (2015) detailed the analysis of field data sample sizes.

The I–40 report (2014) used 39 samples of route travel times (3 days per week, 1 week per month for 13 months). The team then calculated the average travel time as well as the standard error, as shown in Figure 175. The team does not discuss how the standard error was used, but it could have been used in determining the reasonableness of travel times when validating the model.

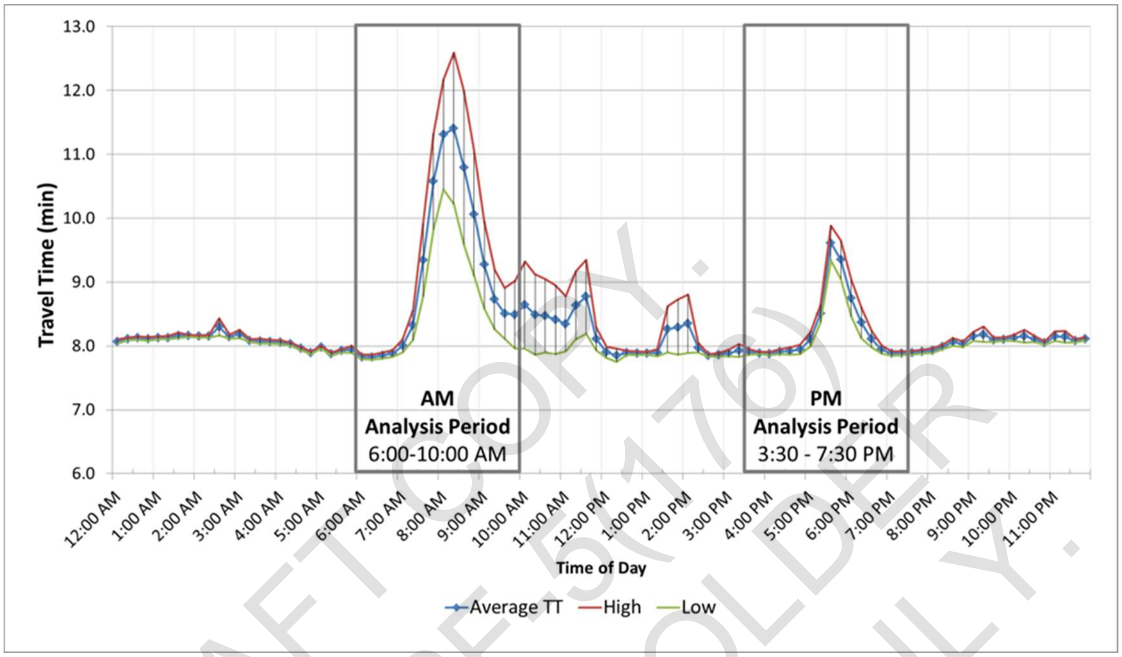


Figure 175. Chart. Average travel time (blue) with the standard error interval.

11.3 CREATING THE NETWORK

Intelligent Selection of Initialization Fill Time

The I–5 report (2014) details the model had a 1-hour fill time as well as a 1-hour cool-down time. While the report does not detail why 1 hour was selected, it does provide details about the volumes used during the fill time and some rationale for why those volumes were selected. Additional details regarding why the fill time and cool-down time of 1 hour was selected would reinforce to reviewers the conditions of the specific scenario modeled were considered in the decision.

The Bryan Park model does indicate that a fill time was used; however, it does not provide details about the length of time used. The report notes that VDOT's Traffic Operations Analysis Tool Guidebook (Chiu 2011) was used as a reference during modeling, and that document includes the following note regarding the fill time:

> *The seeding time should be long enough to distribute traffic throughout the entire network. Typically, seeding times should be approximately equal to either the actual peak*

hour travel time or twice the off-peak travel time, when traversing from one end of the network to the other.

The I–40 (2014) report does not address this item directly; however, it does mention in the discussion of validating the DTALite model that although a 4-hour peak hour was modeled, the primary performance measures were captured for the second hour of the period.

Analyzing the Consequences of Inadequate Study Boundaries

In a traffic simulation, when the spatial limits of the virtual network fail to fully capture the extent of congestion, this can negatively affect certain aspects of an analysis. However, other aspects of the analysis may be unaffected. Engineering judgment can be used to assess the consequences of inadequate study boundaries.

The I–5 report (2014) does discuss challenges faced by modeling some low volume O-D pairs including unrealistic queuing. These pairs were identified as long and improbable because the more direct route, although available in the real world, was not included in the DTA model. The report details the impact these unlikely O-D pairs may create in the model and the decision to create a pruning algorithm to remove pairs in which the shortest free-flow path was 2.2 times as long as the straight-line distance between the origin and destination. The report does not discuss why the study boundaries did not include the more direct routes.

The Bryan Park report (2015) discusses the exclusion of a bottleneck along the edge of the spatial boundaries. This bottleneck would meter inbound traffic to the interchange resulting in a volume lower than the true demand. This challenge was identified before modeling began, and the bottleneck was excluded from all simulations.

The I–40 report (2014) did not discuss the impact of the study boundary selected.

Informed Selection of Overall Data Entry Method

The I–5 network was substantially taken from a larger DTA model with some manual addition along the southern and northern boundaries of the study area. Additionally, intersection data were imported from Synchro with ramp metering coded manually. Similarly, the Bryan Park model was taken from a previous microsimulation model with four intersections coded manually. The I–40 project DynusT model was also adapted from a previous model with additional roadways added to the network.

While the use of imported data has numerous advantages, chief among them time savings, the reports do not touch on possible disadvantages. This may include differences in local parameters between the original model and the added section.

Use of Pre-Processors and/or OpenStreetMap™

Some pre-processor tools can generate traffic network data sets that can be imported and processed by a simulation tool. In other cases, a simulation tool may have the ability to import data from third-party sources and/or formats. Some simulation tools can import detailed data from geographic information systems (GIS) systems such as OpenStreetMap™.

The I–5 report (2014) detailed the importing of Synchro intersection data into the DTA model. Demand adjustment were completed in Emme before populating demand matrices in the model. The Bryan Park (2015) and I–40 reports (2014) did not detail the use of pre-processors.

Entry of Demands Instead of Throughput Volumes

Although the terms demand and volume are often (and sometimes inappropriately) used interchangeably, it is critical the user enters demands instead of volumes. This is because volumes observed in the field are only reflective of vehicles discharged, whereas demands indicate how many vehicles are attempting to use the system. This does not preclude the later use of throughput volumes for calibration and validation purposes.

The I–5 report (2014) uses the term volume when detailing the data collected and demand when detailing the traffic inputs for the model. The report is silent on if, or how, these data are different.

The Bryan Park report (2015) uses the term volume when detailing the traffic inputs for the model. It does, however, consider the difference between volume and demand when considering the study boundaries. South of the Bryan Park Interchange, there is a bottleneck for northbound traffic on I–95. This section of the model was excluded to allow for the full, unmetered demand to reach the interchange within the model.

The I–40 report (2014) did not explicitly discuss the use of demands versus volumes. Spot field data volumes were collected and used for calibration and validation purposes, but the team also acknowledged the presence of bottlenecks internal to the model where the volume would be lower than the demand.

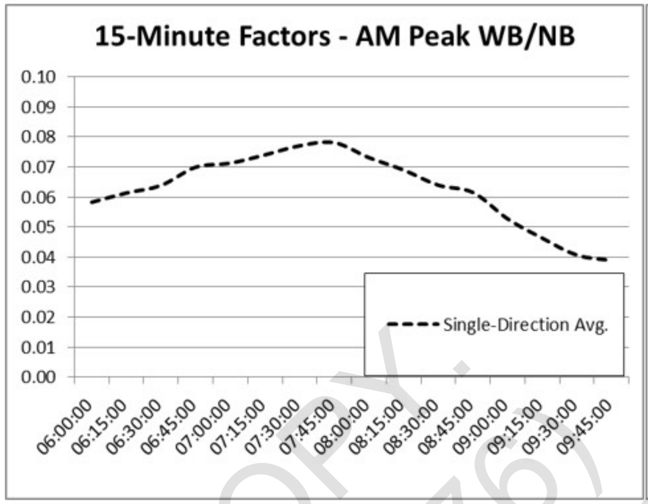
Analysis of Origin-Destination Demands

The I–5 model used a previously developed regional demand model with adjustments in Emme. Some O-D pairs were pruned from the model as discussed in the section, Analyzing the Consequences of Inadequate Study Boundaries. The Bryan Park report (2015) does not discuss O-D pairs.

The I–40 (2014) report gathered an O-D matrix from the TRM (triangle referring the larger metropolitan area of which Raleigh is part, along with the cities of Durham and Chapel Hill). This 4-hour peak period O-D matrix was split into four separate hourly O-D matrices using field measured volumes. The report describes the process of generating the matrices:

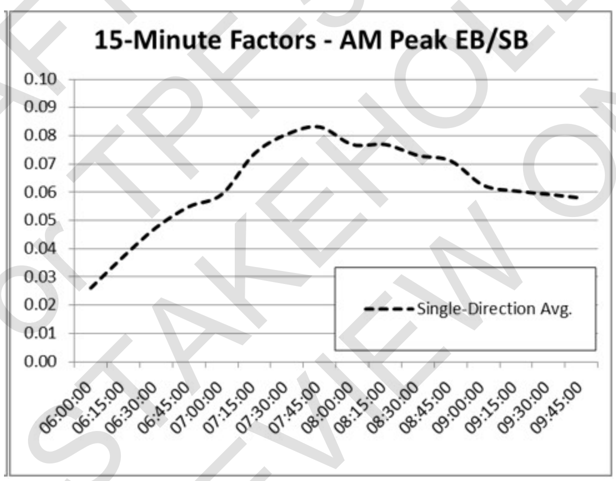
> *For each sensor in each direction in each peak period, the hourly factors were determined by dividing the hourly volume by the sum of 4-hour volume. This yielded four hourly factors for each peak period per direction per sensor. The network wide hourly*

factors for each direction were computed by averaging all hourly factors in that direction. These hourly factors are shown in Figure 12.9.



© North Carolina Department of Transportation.

A. Subfigure of 15-minute factors for westbound and northbound traffic.



© North Carolina Department of Transportation.

B. Subfigure of 15-minute factors for eastbound and southbound traffic.

Figure 176. Charts. Average directional hourly factors for the morning peak.

As shown in © North *Carolina Department of Transportation.*

A. Subfigure of 15-minute factors for westbound and northbound traffic.

© North Carolina Department of Transportation.

B. Subfigure of 15-minute factors for eastbound and southbound traffic.

Figure 176, hourly factors do not match between different directions due to differences in peaking patterns. As such, to recreate a more realistic peaking pattern, one needs to use different hourly factor for the different directions. Although this is possible in DynusT, it requires one to distinguish all origin-destination pairs feeding each direction, which is unfeasible. Therefore, the team used an average of the directional hourly factors as the networkwide hourly factors, see Figure 177.

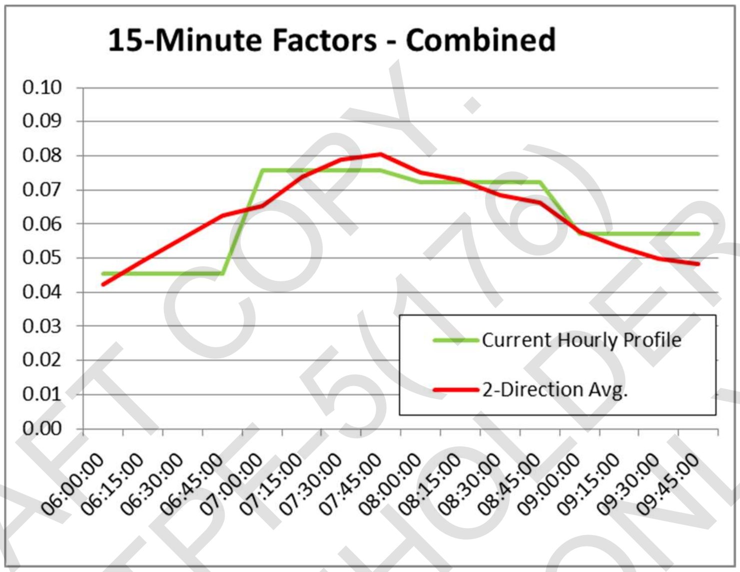


Figure 177. Chart. Networkwide hourly factor for morning peak.

Application of a Specific Volume Balancing Method

For various reasons, upstream and downstream traffic volume counts obtained from the field are not always consistent with one another illustrates an example of this phenomenon. Severely inconsistent counts are likely to produce inaccurate simulation results. Volume balancing methods are available to mitigate this problem. When such methods are used, their use should be documented for the reviewer's benefit and understanding. The Bryan Park report (2015) does not discuss volume balancing.

The I–5 report (2014) details the use of the multiclass demand adjustment tool to assist in adjusting demand. Details are provided as to which counts were considered most important and therefore should be preserved as much as possible. To assist in an understanding of how the counts differed from the modeled volume, the report included a plot of the estimated versus observed traffic, shown in Figure 174.

Application of Data Entry Methods to Achieve Pre-Positioning and Custom Lane Utilization

Lane utilization expresses the proportion of vehicles using each lane. On multilane freeways and surface arterials, simulation tools usually assume equal or slightly unequal lane utilization unless the user indicates otherwise, or unless vehicles unequally distribute themselves to achieve user-specified downstream turn movement percentages. In macroscopic or mesoscopic simulation tools, unequal lane utilizations may be calibrated using inputs such as the lane utilization adjustment factor or single highest lane volume, and pre-positioning may not be relevant. Regarding microscopic simulation, the data entry method for achieving unequal lane utilizations may depend on whether the tool in question follows a link-node or a link-connector architecture. Any such data entry methods should be documented for the reviewer's benefit and understanding.

No discussion of lane utilization factors or pre-positioning was made in the reports.

Entry of Detector (Sensor) Data

While the I–5 report (2014) does indicate detectors are used along the freeway in the real world for ramp meter purposes, such detectors were not replicated. The DTA model demanded a fixed-time ramp meter rate and therefore the modeling of such detectors was unnecessary. The Bryan Park report (2015) did not indicate the inclusion of detectors. The I–40 model used field-collected sensor data for calibration, but the report did not discuss the replication of such detectors in the model to collect performance data.

Entry of Freeway Warning Sign Distances

Freeway warning signs are used in microsimulation to indicate to exiting vehicles when to move toward the target lane upstream of the diverge point. TSSM Chapter 7. indicates these can be located at the real world exit sign location, or in the presence of data indicating driver pre-positioning locations, at the location where drivers move into the target lane. The I–5 (2014) and I–40 (2014) reports do not detail use of freeway warning signs as they are not microscopic simulations. Although the Bryan Park model was microscopic simulation, there was no indication freeway warning signs were used in the model.

Entry of Operating Condition Data

Traditional simulation projects have focused on analyzing typical peak hour traffic conditions. The danger of this approach is that when decisions are made based on most likely (i.e., 50th percentile) outcomes, these decisions may fail badly when the inevitable 75th–95th percentile conditions materialize. With this in mind, some traditional projects have focused on analyzing the 30th-highest-hour demand volume scenario. While this is probably a step in the right direction, the analysis of only a single scenario still leaves engineers largely in the dark about how to manage risk and reliability because they cannot visualize the full range of possible outcomes. Different combinations of demand, weather, incidents, work zones, special events, and driver behaviors ultimately produce a wide range of outcomes. These outcomes will only be captured by simulation if a variety of input scenarios are considered.

None of the reports addressed this topic.

Selection of Key Input Parameters for Calibration

As the purpose of the I–5 model was to create a baseline for future alternatives analyses, the report included significant details regarding the calibration of the model. The parameters included driver response time as well as gap acceptance behavior. A cost function was used to determine route choice behavior with the parameters of the function included as inputs to the calibration. The report (2014) also mentions the use of “targeted, time-dependent matrix adjustments.”

The Bryan Park report (2015) indicates the original model from which the Bryan Park model was derived was a calibrated model; however, no mention of additional calibration methods were detailed.

The I–40 report (2014) focused on calibration using volumes, travel times, free-flow speeds, and O-D matrices. The use of speed flow models is discussed in the Ensuring that Links Perform in Accordance with the Fundamental Diagrams of Traffic Flow section.

Identification of Key Network Locations for Calibration

For some calibration measures (e.g., link volumes, travel times), the I–5 report (2014) provides details over the entire model. For speed, three segments were selected for calibration including 17 miles of the I–5 southbound main line, 17 miles of the I–5 northbound main line, and 7.4 miles of the I–5 southbound express lane.

The I–40 team recognized the challenge of calibrating and validating a model stretching hundreds of square miles stating, “While the team went through an extensive calibration and validation effort, not all sensor stations could be validated against the field-measured sensor data. For a network of this magnitude a complete validation is very difficult to achieve, and consequently the team focused on links within the work zone, as well as key diversion routes” (Schroeder et al. 2014, 16).

11.4 VERIFICATION, CALIBRATION, AND VALIDATION

The Bryan Park model was largely based on a previously calibrated model. Few additional details were provided regarding verification, calibration, and validation (VCV) of the final Bryan Park model and as such discussion of the model will not be included in this section of the case study.

Analysis of Why Results Are Incorrect before Deciding Next Step

The pruning of O-D pairs from the I–5 model were previously discussed in the sections Analyzing the Consequences of Inadequate Study Boundaries and Analysis of O-D Demands. Although the report does not explicitly state the challenges resulting from these low-volume O-D pairs were discovered while calibrating the model, the report (2014) does discuss why the pairs had potential to create issues within the model:

> *The adjusted demand matrices were evaluated to determine if some minor O-D pairs should be removed due to feasibility considerations. The model includes several connections between the northbound and southbound directions which could be used to generate long, improbable routes. Although such routes would not cause significant issues in the behavior of the static model, the low capacities of these paths (through the local street network off the freeway) would likely result in significant queuing due to the hard capacity constraints enforced by the traffic simulation model.*
>
> *In reality, these O-D pairs would have much more direct routes available to them which would be the primary routes in the static model, but these routes are not physically included in the DTA model. Thus, the demands for these O-D pairs in the DTA model represents relatively small, residual percentages of the total demands for these O-D pairs, but for the reasons mentioned above, overuse of the corresponding low-capacity paths would add unnecessary congestion and VMT to the DTA model.*
>
> *Moreover, excessive congestion on paths using both the NB and SB directions of the I-5 would lead to unrealistically high interactions and thus interdependencies between the two directions in the DTA model, which in turn would have a negative impact on model stability and on its predictive/forecasting value.*

The I–40 (2014) report discussed two specific sources of errors encountered when conducting validation, the first related to differences in peak periods in the far southeastern portion of the model versus the northwestern portion. Recall the model started with one 4-hour O-D matrix that was then converted into four 1-hour O-D matrices using field data:

> *The team concluded the calibration and validation effort with greater confidence in the AM peak period results, with the PM period showing a slight overestimation of volumes and travel times. The challenges in calibration are attributed to the large size of the network, as well as highly-variable commuting patterns across the region. A close investigation of sensor data, revealed that westbound/northbound volumes from Johnston County generally peaked about one hour earlier than the eastbound/southbound movements from Orange and Durham County. With the model requiring generalize O/D volume factors over the entire network, the team used an average volume profile across sensor stations as the best feasible estimation across the network.*

Sometimes, unexpected results occur despite having validated the base model as the I–40 team encountered in the DynusT model with respect to density. This error occurred only in the alternatives model, after the base model had been validated. Figure 178 shows densities in one segment over 1,400 passenger cars per mile per lane (mi/ln), far in excess of the typical values of around 170–220 passenger cars per mi/ln. The team worked to determine why this occurred (queues were vertically stacking) and why it was not apparent in the base model (the extreme volume-to-capacity (V/C) ratio in the alternative was not present in the base model). This challenge highlights the need to review both episodic and best- and worst-case scenarios when validating the model.

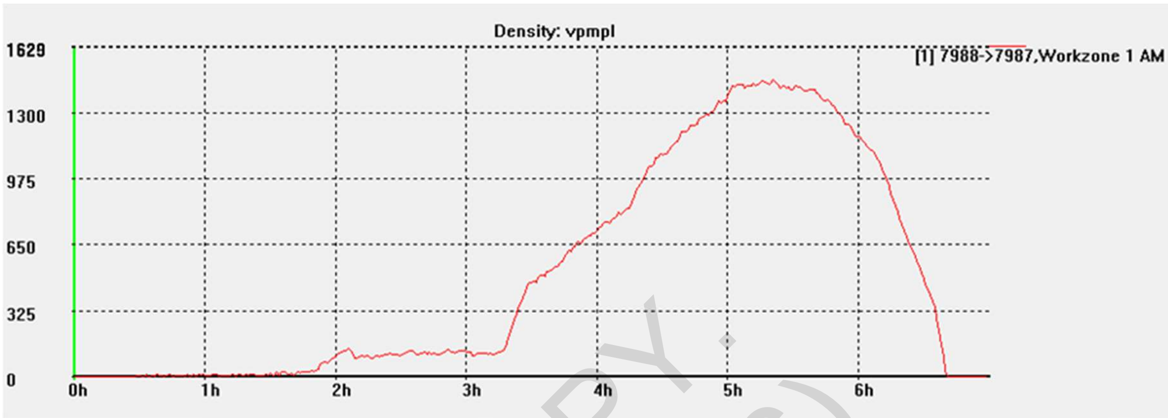


Figure 178. Chart. High segment density observed in the DynusT tool.

Selection of Both Upper and Lower Level Simulation Quality Metrics

In calibrating the model, the I–5 (2014) report indicated the use of relative gap ("the measure of the degree to which the drivers in the model have achieved their ideal path choices in response to recurring congestion patterns"), total travel time, vehicles traveling in the network, vehicles waiting to enter the network, speed, and link volumes as lower level simulation quality metrics (SQM). For upper level SQMs, the report selected goodness-of-fit regression statistics (R-squared values and slope).

For validation, the I–5 report (2014) indicated the use of relative gap, total travel time, vehicle exiting the network, vehicle miles traveled (VMT), vehicle hours traveled (VHT), vehicles waiting to enter the network, and speed. Upper level SQMs used for validation were not indicated.

The I–40 report (2014) noted the use of point-based volume data, point speed data, and segment-based travel times for validation of the DynusT and DTALite models. The results were presented as an average percent error.

Focusing on a Limited Number of Critical Origin-Destination Pairs and/or Key Problem Spots

The selection of key problem spots requires knowledge of the modeled area, the purpose of the study, and the model components.

With respect to combined freeway-arterial models, knowledge of the modeled area may include an understanding of where bottlenecks or hidden bottlenecks are in the network. Do these bottlenecks appear in the model? Does an alternative that relieve the active bottleneck activate a hidden bottleneck?

Thinking about the purpose of the study will reveal key areas that are of interest to the stakeholders. Is there an O-D pair the alternatives are attempting to address? Is there an interchange segment the alternatives are attempting to address?

Finally, freeway-arterial combined networks have a key model component of entrance and exit ramps. These represent a transition from interrupted to uninterrupted flow. Careful study of queue spillback onto the freeway, weaving segments, and intermittent merging onto the freeway due to platoon releases from a signal upstream of the entrance ramp are all components that should be of focus.

The I–5 (2014) report generally focused on global metrics, although there were three key segments over which the speed profiles were compared to field data:

- I–5 southbound main line (17 miles).
- I–5 northbound main line (17 miles).
- I–5 southbound express lanes (7.4 miles).

Speed contour plots were created for these segments and compared to the field collected speed contour plots. Figure 179 shows the plots of the field collected data (left) and model data (right) of the northbound main line from MP 161 to MP 178.

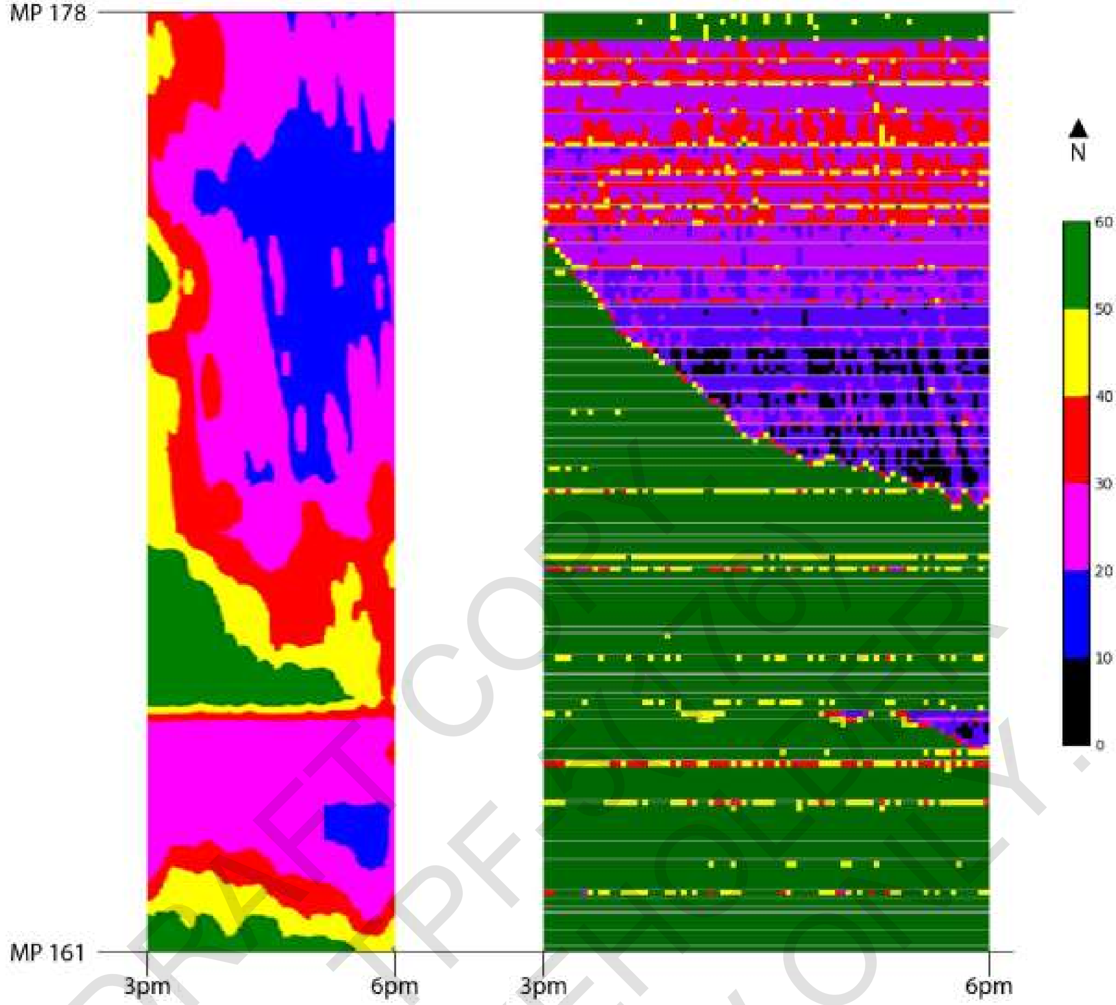
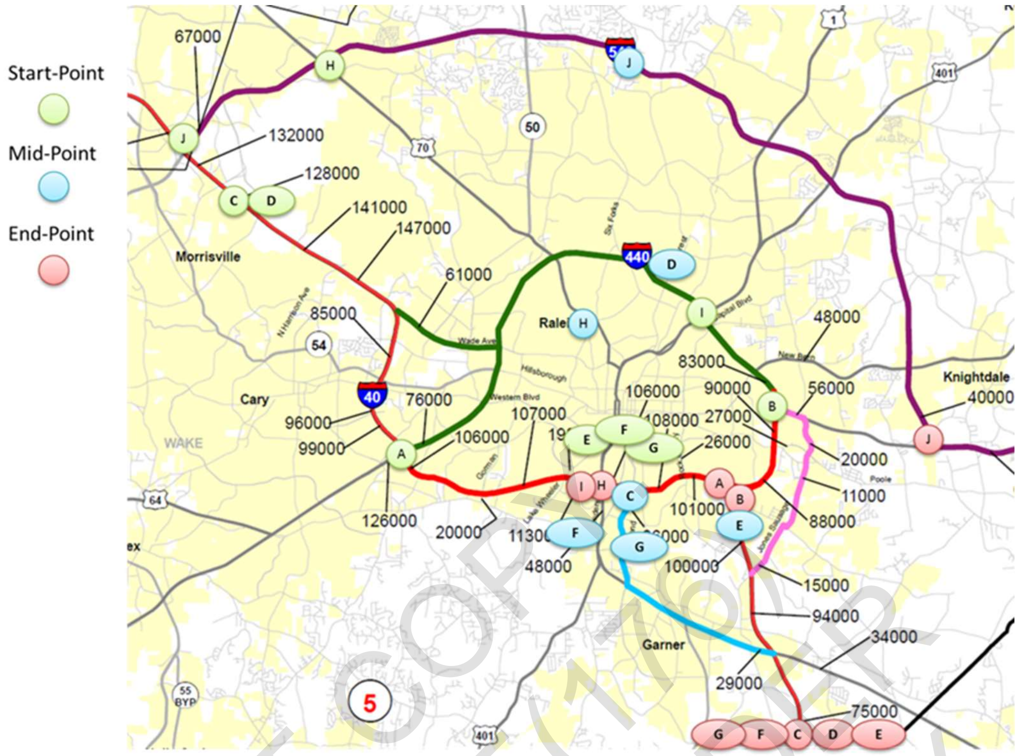


Figure 179. Illustration. Field data (left) and model data (right) of the northbound I–5 main line in the afternoon (PM) peak.

The I–40 report (2014) was driven by a need to understand work zone impacts. As such, the team focused validation efforts on the freeway first, followed by arterials. The work zone links and key diversion routes were of high priority. The team highlighted the key validation routes as shown in Figure 180.



North Carolina Department of Transportation.

Figure 180. Map. Key validation routes.

Evaluation of Probability Density Functions or Cumulative Density Functions That Describe Distributions of the Performance Metrics

None of the reports addressed probability density functions (PDF) or cumulative density functions (CDF).

Ensuring that Links Perform in Accordance with the Fundamental Diagrams of Traffic Flow

In addition to replicating field-measured performance, a well-calibrated simulation should obey the fundamental laws of traffic.
Figure 181 shows the typical shapes for speed vs. density, speed vs. flow, and flow vs. density plots.

The I–40 report (2014) used probe data to collect volume and speed data on the freeway which allowed for the creation of speed vs. flow plots, as shown in Figure 182. The report also details the adjustment of the speed limit in the DynusT model and developed custom speed-density models for weaving segments in the DTALite model. Specific focus was put on ensuring the traffic flow models were appropriate in the work zone.

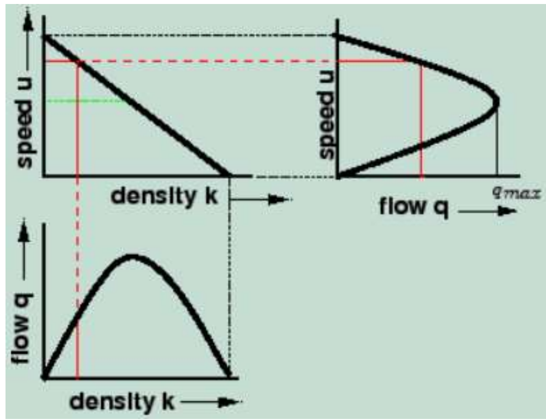
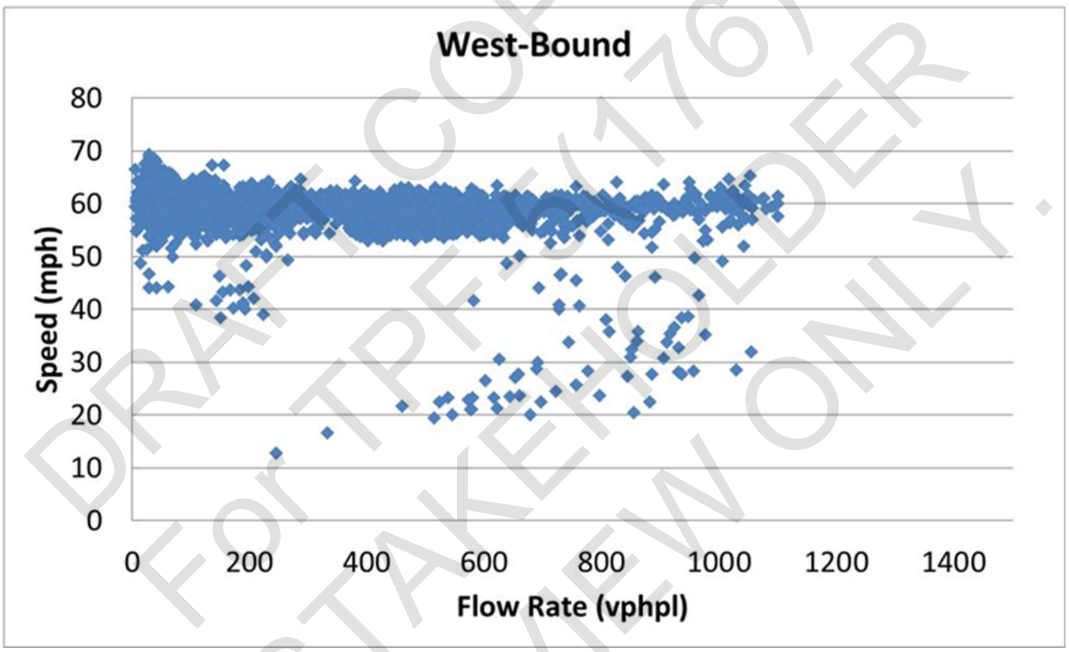


Figure 181. Diagram. Fundamental diagrams of traffic flow.[93]



© North Carolina Department of Transportation.

Figure 182. Chart. Speed-flow curve for one sensor on I–40.

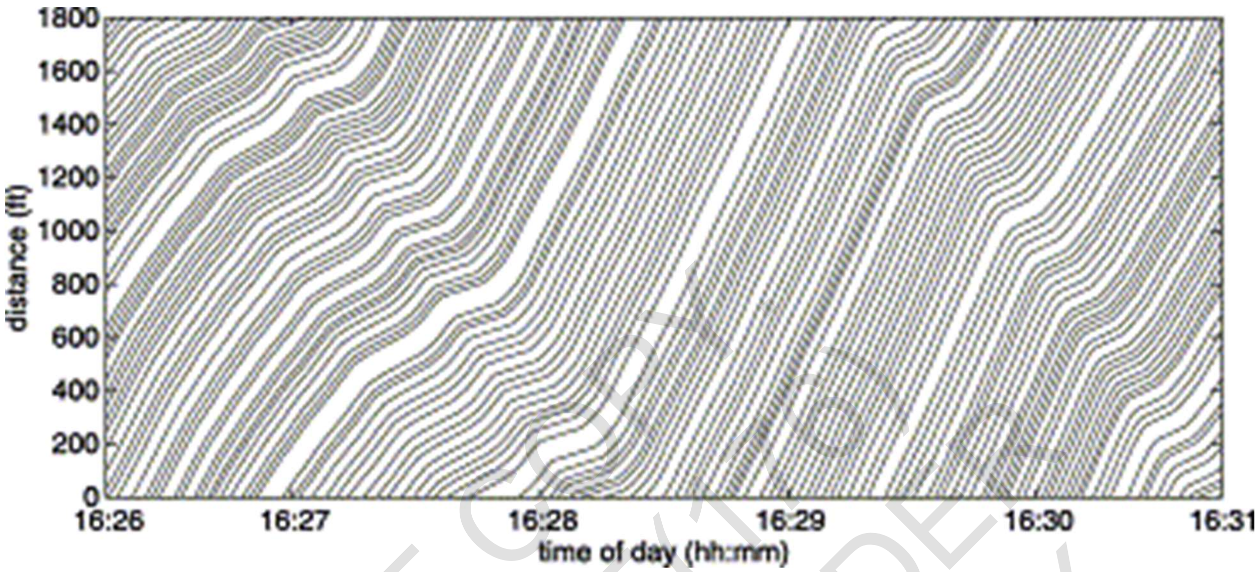
Analyzing Reasonableness of Vehicle Trajectories

Simulated vehicle trajectories, such as the example shown in © Benjamin Coifman.

Figure 183, can reveal unrealistic accelerations, decelerations, shockwaves, cruise speeds, lane-change behaviors, and so on. In some cases, microscopic simulation models producing

[93] "Fundamental Relations of Traffic Flow" (web page), SlidePlayer slide 21, accessed October 20, 2019, https://slideplayer.com/slide/12982685/.

seemingly accurate aggregate performance measures have been shown to be inaccurate at the trajectory level. Although the three combined freeway-arterial case study reports did not provide trajectory analysis in the context of their calibration efforts, analysis of the trajectories is certainly encouraged for important simulation projects.



© Benjamin Coifman.

Figure 183. Chart. A vehicle trajectory plot.[94]

Having a Second Set of Eyes to Review the Data

It is not uncommon for simulation files to be reviewed by customer-stakeholders or their representative. Often, the project report is drafted in part to guide the reviewer through the work. Therefore, the original author may not return to the project report and note that a review was conducted. Although none of the reports detailed an external review, making such a note may provide non-customer stakeholders additional confidence in the results.

Dealing with Episodic Demands

In recent years, transportation professionals have begun to realize the significance of annual performance and travel time reliability. This requires explicit analysis of operating conditions including varying traffic demands, weather, incidents, work zones, and special events. The TSSM authors recommend detailed analysis of these variabilities whenever possible; and when these conditions are analyzed, the methods of doing so should be well documented. Today's traffic modeling experts should become well-versed in the simulation-based travel time reliability analysis methods described in the SHRP2 L04 report (Mahmassani 2014), as well as

[94] Coifman, Benjamin. "Estimating travel times and vehicle trajectories on freeways using dual loop detectors." Transportation Research Part A: Policy and Practice 36.4 (2002): 351-364.

calibration methods described in Traffic Analysis Toolbox Volume III (Wunderlich, Vasudevan, and Wang 2016).

Selection of a Preferred Calibration Sequence

According to TSSM Chapter 8. , input parameters can be calibrated simultaneously or sequentially, with sequentially the recommended approach. One example of a recommended sequence was: (1) driving behavior model parameters, (2) route choice model parameters, (3) traffic demands, and (4) overall model fine-tuning.

The I–5 case study report (2014) did not explicitly state if the input parameters were calibrated simultaneously or sequentially.

The I–40 report (2014) provided a flow chart for the calibration method of the DynusT model in Figure 184.

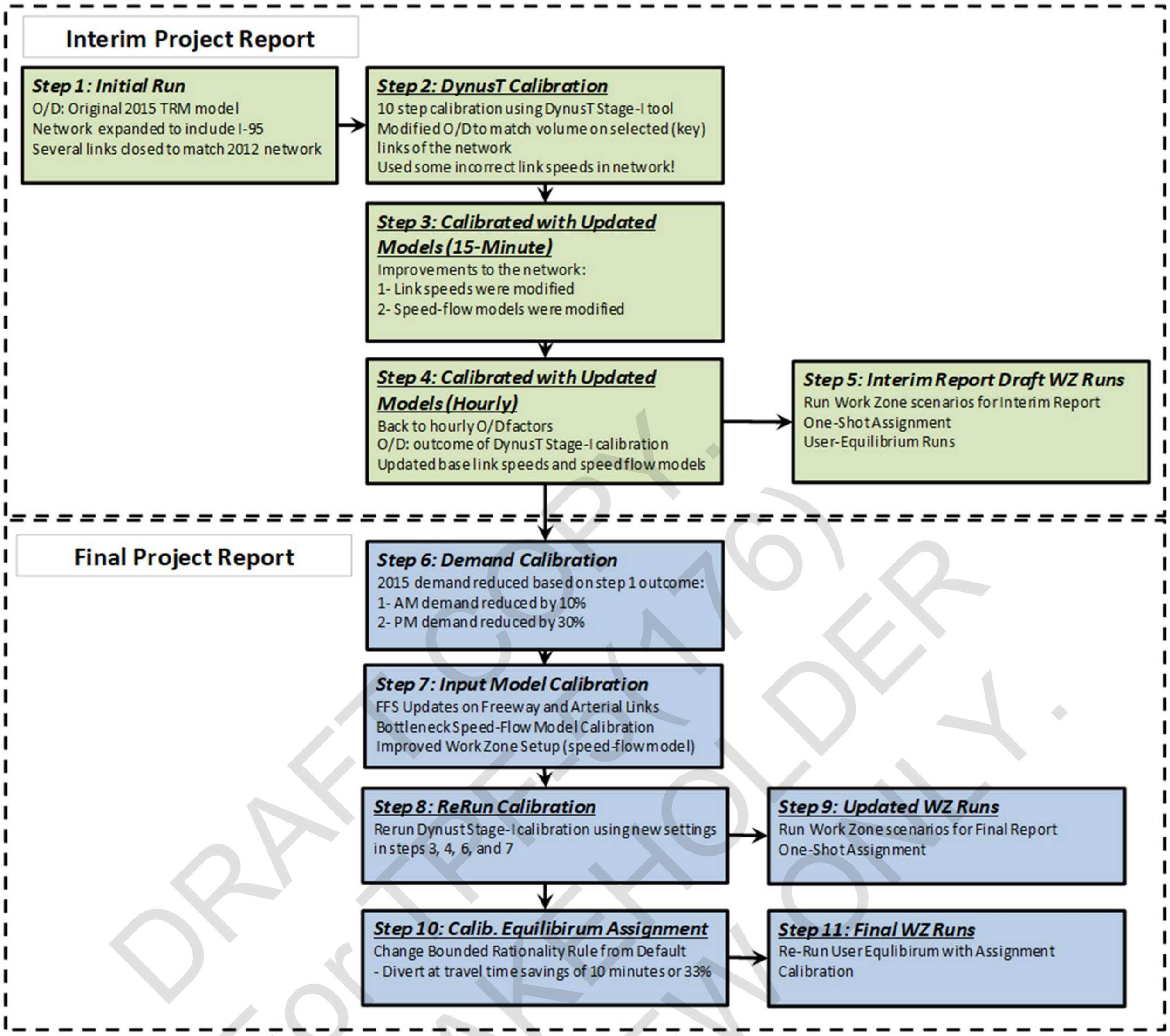


Figure 184. Diagram. Calibration flow chart for DynusT model.

Validate with Separate Data Sets

The I–5 model link volumes were compared to empirical traffic counts at 193 locations, but the report (2014) does not state if those traffic counts were used in developing the demand matrices. In comparing the model to the empirical data, the report considered goodness-of-fit metrics R-squared and slope as shown in Figure 185.

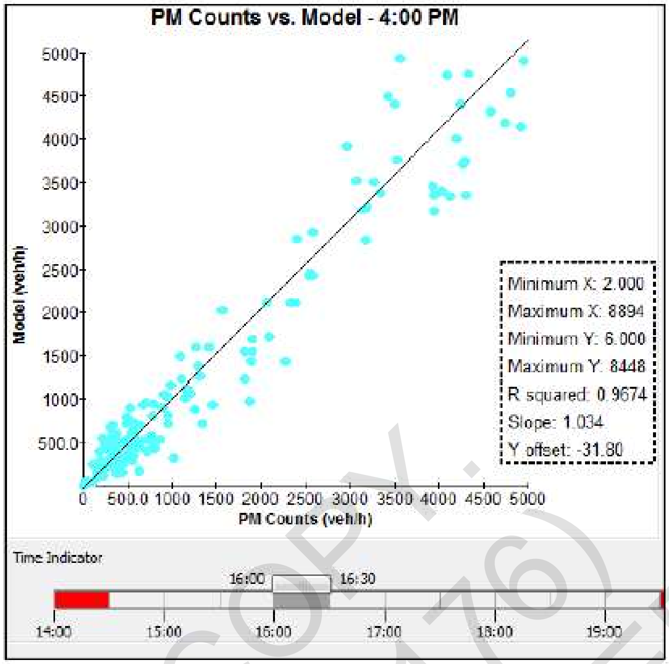


Figure 185. Screenshot. Empirical traffic counts versus model volumes from 4 to 4:30 p.m.

Review Best-Case and Worst-Case Scenario Outcomes

As previously discussed in the subsection on episodic demands, a comprehensive review of varying operating conditions (e.g., weather, incidents, work zones) and episodic demands is strongly encouraged. Such a review would naturally lead to discovery of best-case and worst-case scenario outcomes that could potentially materialize throughout the year.

None of the three combined freeway-arterial combined case studies discussed subjecting the model to best- or worst-case scenarios during calibration and validation. However, the I–40 report (2014) did consider worst-case scenarios while modeling alternatives. The mesoscopic models were run with one-shot path assignment where no diversion from the path assignment was allowed. This represented the worst-case scenario in which no vehicles would divert to alternative routes during construction. The team also ran user-equilibrium models, in which the simulation was repeated 20 times. After each simulation, some portion of vehicles would change to a route with a lower travel time, after which the simulation was re-run. Equilibrium is reached when additional runs would not result in further significant reduction in travel time.

11.5 DATA PROCESSING, ANALYSIS, AND PRESENTATION

Using a Naming Structure for the Scenario Data Sets

TSSM **Error! Reference source not found.**recognizes that scenario analysis is one of the most valuable capabilities of traffic simulation models. For a project with numerous alternatives, the number of input data sets can become onerous to manage. Creating a condensed but readable naming structure is critical. Chapter 9 recommends a naming structure that reflects an input that would vary from model to model. For example, given three geometric alternatives and two time periods, the naming structure might look like: *ProjectRef_Alternative#_TimePeriod_Version*, with one data set named: *50476_Alt1_AM_V1*.

Although highly recommended, the data set naming convention is a low-level detail that is often not documented in the final project reports. Nonetheless, for the benefit of future engineers who may review or build on the initial work, an appendix to the simulation report could be used to document the chosen data set naming convention.

Automated Post-Processing Analysis

Traffic simulation models are notorious for producing exhaustive amounts of numerical output data, especially microsimulation models requiring 10 or more Monte Carlo realizations per scenario. Post-processor tools are frequently needed to analyze the summation, mean, median, maximum, percentile, and variance of these results across numerous links, nodes, sensors, and data stations. In some cases, these may be third-party post-processing tools distributed or sold separately from the native simulation tool. In other cases, there may be a post-processing module included within the simulation package, or the simulator may seamlessly provide post-processed statistics without any mention of a separate tool/module.

Simulation project reports do not always disclose names of the tools used to obtain their post-processed statistics. Engineers have post-processed and visualized their outputs through spreadsheet software (e.g., Microsoft® Excel) and statistics software (e.g., SAS, R). Specialized post-processing tools have been customized for specific simulation products.

The I–5 report (2014) clearly used an automated post-processing tool to develop various graphs. In Figure 186, the right side of the image shows the graphical user interface (GUI) that can be used to manipulate display of the data. This includes adjusting the number of assignment intervals or iteration numbers depicted.

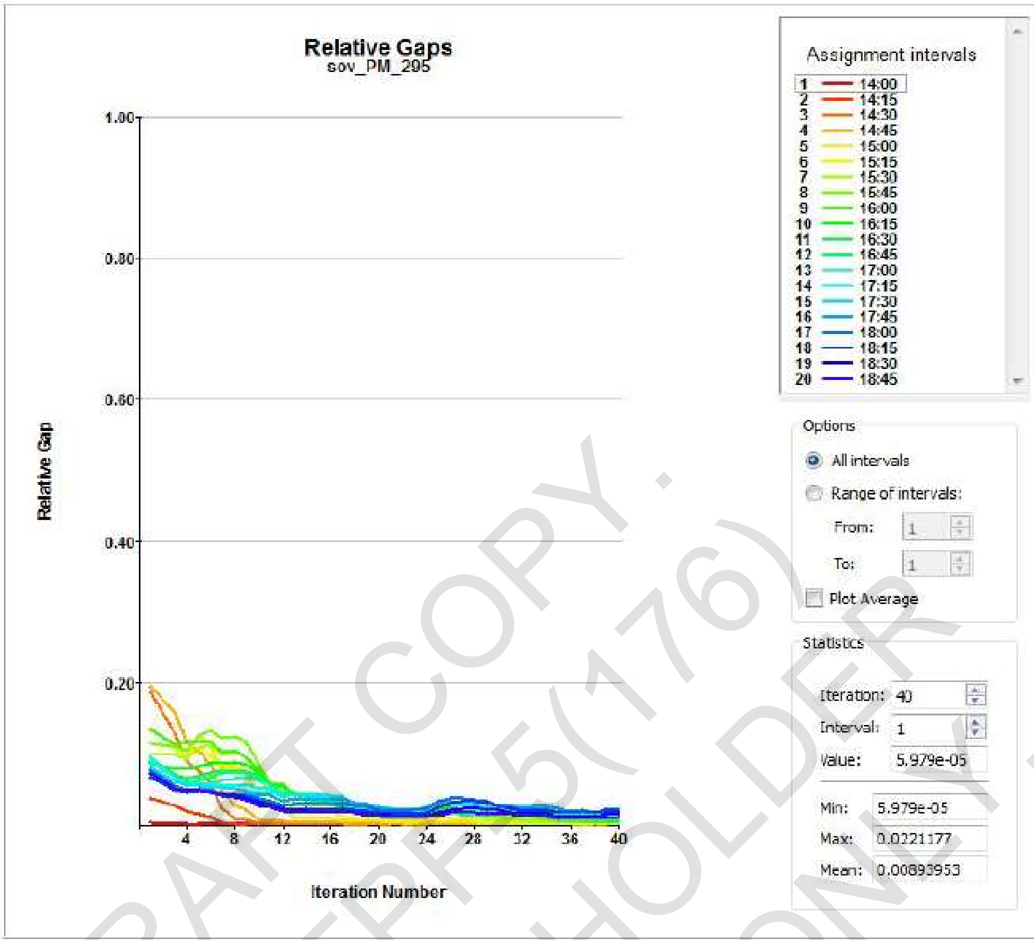


Figure 186. Illustration. Chart (left) generated using inputs from the graphical user interface (GUI) (right).

The Bryan Park microscopic model used the maximum back of queue as a performance measure in the alternatives analysis. Such collection of the maximum back of queue is a post-processing performed on the queue collection data, although some tools have this as an imbedded feature. Table 26 provides the maximum back of queue at each intersection approach as well as the average delay by intersection and turning movement.

Table 26. Future year no build analysis delay and maximum back of queue results.

2022 No Build AM - Delay (sec/vehicle)																		
Intersection	Overall LOS	Overall Delay	Northbound				Southbound				Eastbound				Westbound			
			L	T	R	Max Q**	L	T	R	Max Q**	L	T	R	Max Q**	L	T	R	Max Q**
Hermitage Rd at Westbrook Ave	A	6	-	3	1	60	2	2	-	80	-	-	-	0	46	-	2	175
I-95 NB Off-Ramp at Westbrook Ave	A	4	-	-	-	0	14	6	-	0	-	0	-	0	-	0	-	0
Hermitage Rd at Laburnum Ave	B	19	41	30	10	190	34	31	19	305	28	12	6	205	18	15	11	410
Hermitage Rd at Brookland Pkwy*	C	27	57	17	20	125	38	24	15	165	51	54	48	145	28	27	22	260
Laburnum Ave & I-64 EB Off Ramp	A	7	-	-	-	0	15	-	5	180	-	0	0	0	1	0	-	0
Laburnum Ave & I-195 NB Off Ramp	A	2	10	11	2	45	-	-	-	0	-	0	-	0	-	0	-	0
Laburnum Ave & I-64 WB On Ramp	A	1	-	-	-	0	-	-	-	0	4	0	-	0	-	1	2	80
N Blvd & I-95 Ramps	A	7	9	3	-	270	-	3	4	70	42	-	4	160	-	-	-	0
Hermitage & I-95 NB Off Ramp	A	5	-	0	-	0	-	1	-	0	-	-	-	0	8	-	5	85
2022 No Build PM - Delay (sec/vehicle)																		
Intersection	Overall LOS	Overall Delay	Northbound				Southbound				Eastbound				Westbound			
			L	T	R	Max Q**	L	T	R	Max Q**	L	T	R	Max Q**	L	T	R	Max Q**
Hermitage Rd at Westbrook Ave	A	5	-	3	1	120	4	2	-	50	-	-	-	0	53	-	3	160
I-95 NB Off-Ramp at Westbrook Ave	A	5	-	-	-	0	14	7	-	0	-	0	-	0	-	0	-	0
Hermitage Rd at Laburnum Ave	C	20	45	36	25	320	41	34	14	235	20	16	11	450	28	11	6	210
Hermitage Rd at Brookland Pkwy*	C	23	47	19	21	240	42	18	9	115	29	32	29	245	32	32	25	215
Laburnum Ave & I-64 EB Off Ramp	B	10	-	-	-	0	20	-	0	315	-	0	0	0	1	1	-	20
Laburnum Ave & I-195 NB Off Ramp	A	1	9	0	3	45	-	-	-	0	-	0	-	0	-	0	-	0
Laburnum Ave & I-64 WB On Ramp	A	1	-	-	-	0	-	-	-	0	3	0	-	50	-	0	2	10
N Blvd & I-95 Ramps	B	14	14	9	-	415	-	3	12	80	76	-	4	400	-	-	-	0
Hermitage & I-95 NB Off Ramp	A	3	-	1	-	0	-	0	-	0	-	-	-	0	5	-	4	5

*Note: Intersection of Hermitage Rd and Brookland Parkway is currently a 5-legged intersection with average approach delay of 6 seconds (AM) and 7 seconds (PM) along right-out Westwood Avenue (fifth-leg) approach. However, the overall intersection delay (23 Secs – both AM and PM) includes average delay of all five (5) approaches.

**Max Q represents maximum approach queue length in feet

Source: Virginia Department of Transportation.

The I–40 report (2014) used ArcGIS to illustrate networkwide impacts, as shown in Figure 187.

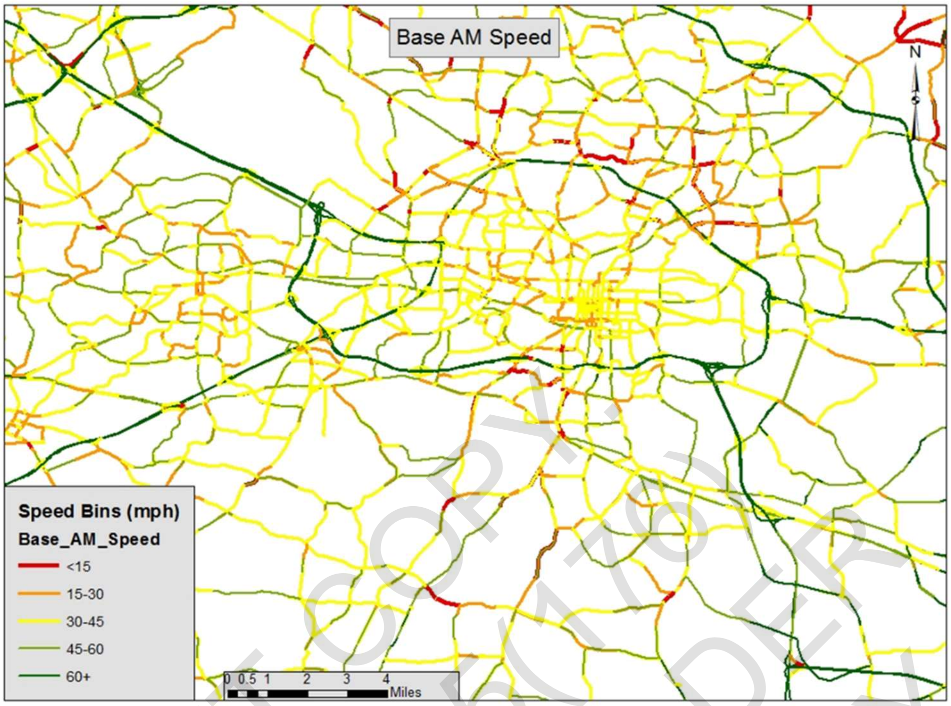


Figure 187. Map. Networkwide morning (AM) speeds for base model.

Weighted Aggregation of Performance Measures

Applying weighting factors to the performance measures at various network locations can help to convey the right information. Performance measures can be weighted by demand, throughput, time, distance, and other metrics. This can help to show the bottom-line traffic performance was most heavily influenced by:

- Links or nodes serving the greatest number of trips (i.e., throughput-weighted average).
- Conditions prevalent for the longest time period(s) (i.e., time-weighted average).
- Links or segments covering the greatest distance (i.e., distance-weighted average).

The I–5 report (2014) does not include weighted performance measures. The Bryan Park report (2015) includes overall intersection delay, shown in Table 26

, which is a volume-weighted aggregation of the individual movement delays. While the I–40 report (2014) details aggregated values, it does not note if they were weighted.

Analysis of Stochastic Model Outcomes under Different Random Number Seeds

Random number seeds are most often associated with microsimulation, although macroscopic and mesoscopic simulation tools can certainly incorporate random processes when the developer so chooses. According to conventional wisdom and several State DOT simulation guidelines, the number of Monte Carlo realizations should be at least 10 and should reflect important aspects of the study (e.g., congestion levels, network size). For example, an unstable freeway weaving section could warrant up to 100 realizations to obtain sufficient confidence in the reported results.

The Bryan Park report (2015) details the use of 10 realizations in accordance with VDOT's Traffic Operations Analysis Tool Guidebook (Chiu 2011).

Use of Probability Density Functions and Cumulative Density Functions

The I–5 (2014) and I–40 (2014) reports did not include PDFs or CDFs.

In a discussion of the DLM, the Bryan Park report (2015) provided a graph of the duration of the DLM in each of the 10 simulations. While this graph, shown in Figure 188, is neither a PDF nor a CDF, the data could be manipulated to create such a function.

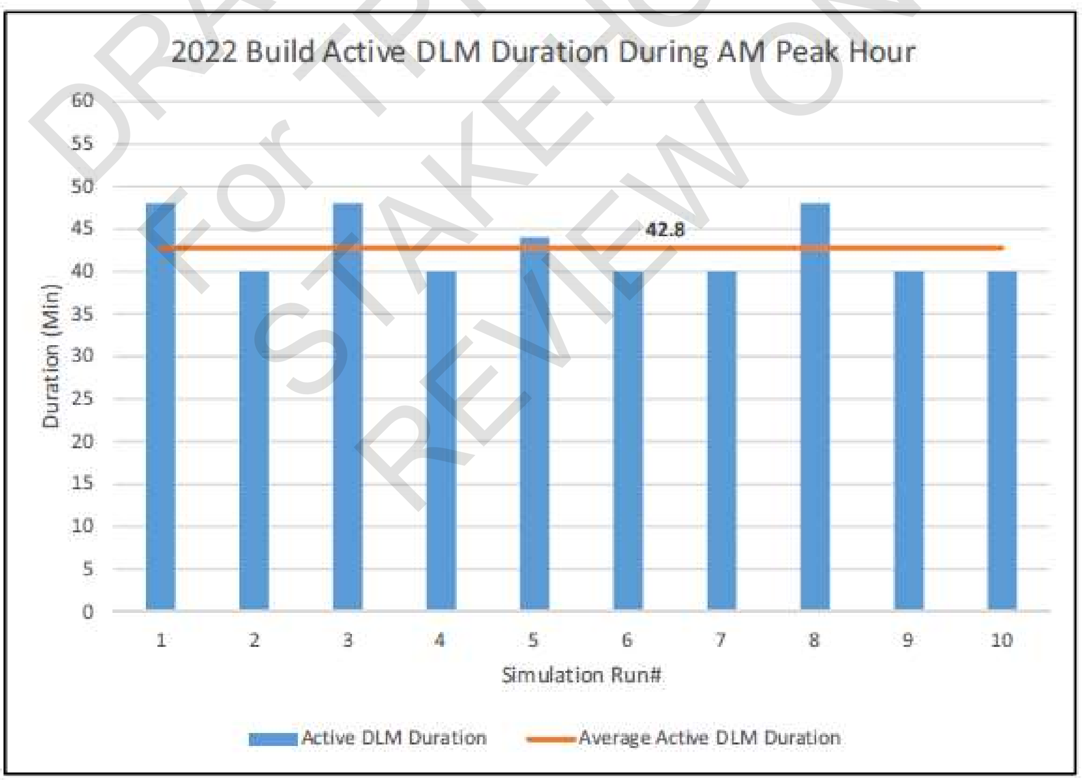


Figure 188. Chart. Duration of active dynamic lane merge (DLM) during morning (AM) peak hour by simulation.

Use of Confidence Intervals

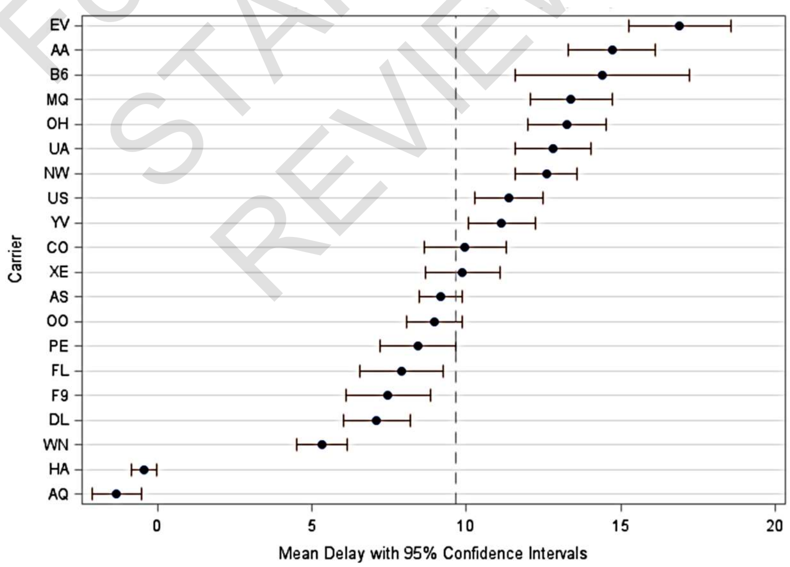
Performance measures typically generated from simulation tools may be unreliable unless the engineer provides a confidence interval, which indicates how likely it will be to observe a mean value within a specific range (e.g., the mean travel time has a 0.95 probability to fall between 1.45 and 1.57 minutes per mile).

Confidence intervals have often been used to clarify microsimulation outputs, but they can also be used to assess the sample size of traffic data measurements used to obtain the simulation inputs. Ideally, the analyst would like to keep confidence high while still having a narrow enough interval to sense where the true value may lie. It is also important to note that increasing the sample size results in a decrease in the width of a confidence interval, since the amount of error used to calculate the confidence interval is decreased.

> Confidence intervals can also be used to suggest whether two statistics are significantly different. If two confidence intervals overlap, this suggest the 2 means probably will not be significantly different at that respective confidence level. In © SAS Institute.

Figure 189, it is possible to see which carriers will most likely differ by comparing where the confidence intervals do not overlap. Apart from being easy to compare visually, it also helps to show how much variability is associated with each mean, by the examining the width of the intervals.

None of the reports provided confidence intervals.



© SAS Institute.

Figure 189. Chart. Daily airline delays.[95]

Use of Statistical Hypothesis Testing

Although goodness of fit cannot necessarily be classified as a hypothesis, the I–5 report (2014) used the R-squared statistic as a high-level SQM. The authors indicated the R-squared values for the regression of empirical traffic counts versus modeled volumes were all between 0.95 and 0.97.

Neither the Bryan Park (2015) nor I–40 (2014) report included details of statistical testing.

Analysis of Lane-Specific Results

None of the reports detailed lane-specific results.

Use of Multivariate Outputs and Accompanying Visualizations

No single performance measure is perfect. Univariate measures may be misleading. It is helpful to consider several measures and distributions. Performance reporting can include combinations of numeric performance measures, visualizations, and qualitative discussions. In this case, the analyst's understanding would have been definitively reduced if they had not had access to the visualizations. In other cases, the numeric measures and/or qualitative discussions are needed to highlight important subtleties within a complex visualization.

The I–5 report (2014) provided visualizations for many of the measures used in calibration and validation. In instances where goodness-of-fit-measures were reported, accompanying visualizations were provided to serve as witness to the statistics.

The I–40 report (2014) included Figure 180, which was a visualization of the speed throughout the network during the AM peak of the base scenario. They further provided tabular data regarding travel time and volumes along various routes within the network relative to the base scenario, as seen in Table 27
. Due to the complexity of the tabular data, the report follows with a written description of the important conclusions to be drawn:

> *Maintaining two lanes open on the peak direction in the AM peak hour significantly increases the travel time on all routes with or without diversion compared to the baseline*

---

[95] SAS Institute, Ranking with Confidence: Part 1, 2011, retrieved from https://blogs.sas.com/content/iml/2011/03/25/ranking-with-confidence-part-1.html#prettyPhoto/0/.

scenario. As expected, allowing drivers to change their routes diverts a considerable proportion of traffic volume from the key routes and results in travel times that are considerably lower than those observed in the no-diversion scenario.

Maintaining three open lanes on the peak direction in the AM peak significantly increases travel time when drivers do not change their routes. However, the extent of this increase in travel time is considerably lower than that of the scenario with two open lanes, as expected. When drivers are allowed to change their routes, maintaining three lanes open only slightly increases the travel time.

Table 27. Morning peak hour work zone impacts for selected routes.

Route	Work Zone	Direction	Pre-Construction Travel Time (min)	No-Diversion Travel Time (min)	Optimum-Diversion Travel Time (min)	Pre-Construction Volume (vph)	No-Diversion Volume (vph)	Optimum Diversion Volume (vph)	Estimated Diversion
Route A (I-40 Milepost 293 through 300)	WZ 1	EB	8.0	8.0	7.9	3333	3254	3173	-5%
		WB	8.4	39.7	12.7	4541	1875	2078	-54%
	WZ 5	EB	8.0	8.0	8.0	3333	3314	3167	-5%
		WB	8.4	22.1	10.7	4541	3060	3037	-33%
Route B (I-440 Milepost 14 through 16)	WZ 1	EB	2.0	2.0	2.0	5045	2490	5623	11%
		WB	1.7	33.2	2.0	2755	803	1627	-41%
	WZ 5	EB	2.0	2.0	2.0	5045	3709	5377	7%
		WB	1.7	3.9	2.0	2755	2117	1796	-35%
Route C (I-40 Milepost 284 through 300 via I-40)	WZ 1	EB	25.9	25.9	25.9	2996	2873	2854	-5%
		WB	27.5	349.2	33.3	4426	2112	3086	-30%
	WZ 5	EB	25.9	26.0	26.0	2996	2968	2850	-5%
		WB	27.5	126.5	29.7	4426	3248	3668	-17%
Route D (I-40 Milepost 284 through 300 via I-440)	WZ 1	EB	29.3	139.6	29.5	2938	2684	2745	-7%
		WB	30.1	319.0	30.3	4146	2437	4288	3%
	WZ 5	EB	29.3	36.4	29.5	2938	2872	2745	-7%
		WB	30.1	98.9	30.0	4146	3374	4228	2%

Source: North Carolina Department of Transportation.

Presentation of Percent Changes Instead of Raw Numbers

During TSSM stakeholder meetings, stakeholders pointed out that simulation reports are easier to understand when they provide percent changes in the numeric performance measures, in addition to (or instead of) the raw numbers.

The I–40 report (2014) uses percent changes in several ways including communicating how many vehicles divert to alternate routes in each scenario, the reduction in travel time, distance, and stop time, as well as the unserved vehicles at the conclusion of the model run. Some of these are exhibited in Table 28. Further, the authors note that due to the challenges of calibrating and validating such a large network and the assumptions which were made, readers "should therefore focus on the *relative impacts* of the work zone when compared to the (partially calibrated) baseline" (2014).

Table 28. Networkwide work zone impacts: absolute and relative to baseline.

Scenario		Average TT		Average Travel Distance		Average Stop Time		# of Unserved Vehicles	
		Min.	% Diff	Min.	% Diff	Min.	% Diff	Veh.	% Diff
Baseline		30.14	n/a	11.92	n/a	4.01	n/a	40168	n/a
WZ 1	No-Diversion	36.65	21.6%	11.65	-2.3%	10.07	151.1%	64088	59.5%
	With-Diversion	**31.58**	**4.8%**	**11.72**	**-1.7%**	**4.39**	**9.5%**	**44871**	**11.7%**
WZ 2	No-Diversion	35.12	16.5%	11.73	-1.6%	9.64	140.4%	57435	43.0%
	With-Diversion	**31.12**	**3.3%**	**11.73**	**-1.6%**	**4.07**	**1.5%**	**44815**	**11.6%**
WZ 3	No-Diversion	34.44	14.3%	11.75	-1.4%	9.17	128.7%	53325	32.8%
	With-Diversion	**31.17**	**3.4%**	**11.72**	**-1.7%**	**4.12**	**2.7%**	**44672**	**11.2%**
WZ 5	No-Diversion	33.06	9.7%	11.85	-0.6%	9.08	126.4%	46548	15.9%
	With-Diversion	**31.14**	**3.3%**	**11.73**	**-1.6%**	**4.08**	**1.7%**	**44606**	**11.0%**
WZ 7	No-Diversion	34.65	15.0%	11.83	-0.8%	9.43	135.2%	46868	16.7%
	With-Diversion	**31.26**	**3.7%**	**11.7**	**-1.8%**	**4.12**	**2.7%**	**44729**	**11.4%**

Source: North Carolina Department of Transportation.

Final Documentation of Assumptions, Unknowns, and Recommendations

The I–5 report (2014) was developed for the intention of future use by other modelers. As such, care was provided in detailing assumptions about data consistency in calibration as well as intentionally excluded inputs for the future year scenario. In addition to this warning, the report detailed how to address the potential need for such considerations in future models:

> *It is important to note that the supply side (traffic signals, ramp meters, etc…) model inputs were not optimized for the future demand in these model runs. As a result, the network may not be providing the best possible capacities, travel times or overall performance for the given demand scenarios.*

> *This has direct impact on both of the causes mentioned above for increased numbers of vehicles waiting outside of the network. With regards to congestion occurring on the I-5, some of the most significant bottlenecks are due to congested signalized off-ramps which then spill back onto the freeway and gradually choke off capacity for the main line traffic. The choke effects in turn result in lower congestion on the freeway itself and an increase in vehicles waiting outside of the network. Secondly, several locations can be observed where an increase in the number of vehicles waiting at certain zones is due to a signalized bottleneck just inside the network which cannot serve the increased demand. In these cases, congestion spills back directly up the entrance connectors from these bottlenecks without any downstream congestion being created deeper inside the network as a result of the increased demand volume.*
>
> *To compensate for these effects, the ramp meter flow rates (capacities) were increased in the future models by roughly 50%. This provided some improvements in overall throughput and reduction in delays outside the network.*
>
> *Expanding the models in the future to incorporate a larger study area including alternate routes would be recommended, as this would allow the model to adapt to the limited capacities and re-distribute itself in a more optimal way, resulting in a better match between demand and supply and less delay outside of the modeled network.*

The Bryan Park report (2015) details key assumptions broken out by category including:

- Study area.
- Opening year.
- Peak period for analysis.
- Existing traffic data.
- Future traffic forecasts.

Additionally, the report concludes with qualitative summaries of the findings regarding alternatives tested and a corresponding recommendation:

> *The proposed DLM alternative resulted in a minor overall improvement of operations at the Bryan Park Interchange along I-95. The DLM improvement along I-64/I-195 on-ramp had no negative impact on upstream traffic operations along northbound I-95. However, traffic operations improved slightly at the I-64/I-195 on-ramp merge area and downstream along I-95 northbound during AM and PM peak hours. During the critical PM peak hour, speeds would be expected to increase by 4 MPH to 6 MPH with density reduction of 4-5 passenger cars per mile per lane. The improvement in operations can be attributed to favorable (also safer) merging conditions for I-64/I-195 merging traffic while maintaining satisfactory freeway operations along northbound I-95. The analysis of multiple simulation runs indicated that active DLM control is expected for majority (43 minutes) of the AM peak hour and at least half of the time (34 minutes) during the PM peak hour… Based on the operational analysis, and anticipated operations/maintenance costs associated with the proposed improvements, the improved*

operations may not provide appreciable benefits to outweigh the financial implications of installation.

The I–40 report (2014) provided a section on model limitations which included a discussion of assumptions:

> *This research has provided valuable insights onto the expected impacts of the I–5311/I–5338 work zone on Interstates I-40 and I-440. Those will be useful in weighing trade-offs and relative operational performance of different work zone staging strategies.*
>
> *However, it should be emphasized that the current approach assumes no travel demand changes in response to the presence of the work zone. Thus, the results assume no trip reduction in the form of car-pooling, transit, telecommuting, peak-spreading, discretionary trip cancellation or other demand-reducing strategies. In reality, some demand reduction, spatial, temporal and/or modal shifts can be expected in response to a work zone of this magnitude. Findings from this analysis are important for public outreach as it relates to this project, which may also include the exploration of suggested (or mandatory) detours routes.*
>
> *One other important consideration of all results put forth in this report is that the contractor scenarios in the design-build construction contract had not been finalized at the time of this writing. Thus, the actual impacts of the work zone can be expected to differ from our estimates, depending on how close the final contractor phases and stages are to the modeled scenarios.*
>
> *Despite these caveats, the project has carried out an in-depth and comprehensive comparison and application of three software tools for evaluating corridor and network impacts of a significant freeway work zone. All three models were calibrated with a considerable amount of field volume data and local work zone capacity estimates (from prior research). The base models were validated with empirical speed and travel time data for various key routes in the Triangle region.*

The concluding chapter of the report also included overall recommendations, impacts to key diversion routes, and the value of diverting traffic. These key conclusions were also provided in summary in the executive report.

Presentation of Static Graphics and Moving Vehicle Animation

Moving vehicle animation is typically associated with microsimulation, but static graphics are provided for all forms of simulation.

The I–5 report (2014) detailed a mesoscopic model and included static graphics including Figure 174, Figure 178, Figure 185, and Figure 186.

The Bryan Park report (2015) detailed a microscopic model with static graphics including Figure 169, Figure 171, and Figure 188.

The I–40 report (2014) captured in this case study detailed two mesoscopic models with static graphics including Figure 170, Figure 172, Figure 173, Figure 175, Figure 76, Figure 177, Figure 178, Figure 180, Figure 182, Figure 184, and Figure 186.

CHAPTER 12. FREEWAY CASE STUDIES

This chapter discusses three traffic simulation case studies for freeway networks. The first case study is described in the final report, Work Zone Traffic Analysis & Impact Assessment (Schroeder et al. 2014), prepared for the North Carolina Department of Transportation (NCDOT). The work zone and surrounding traffic network are illustrated in Figure 190.

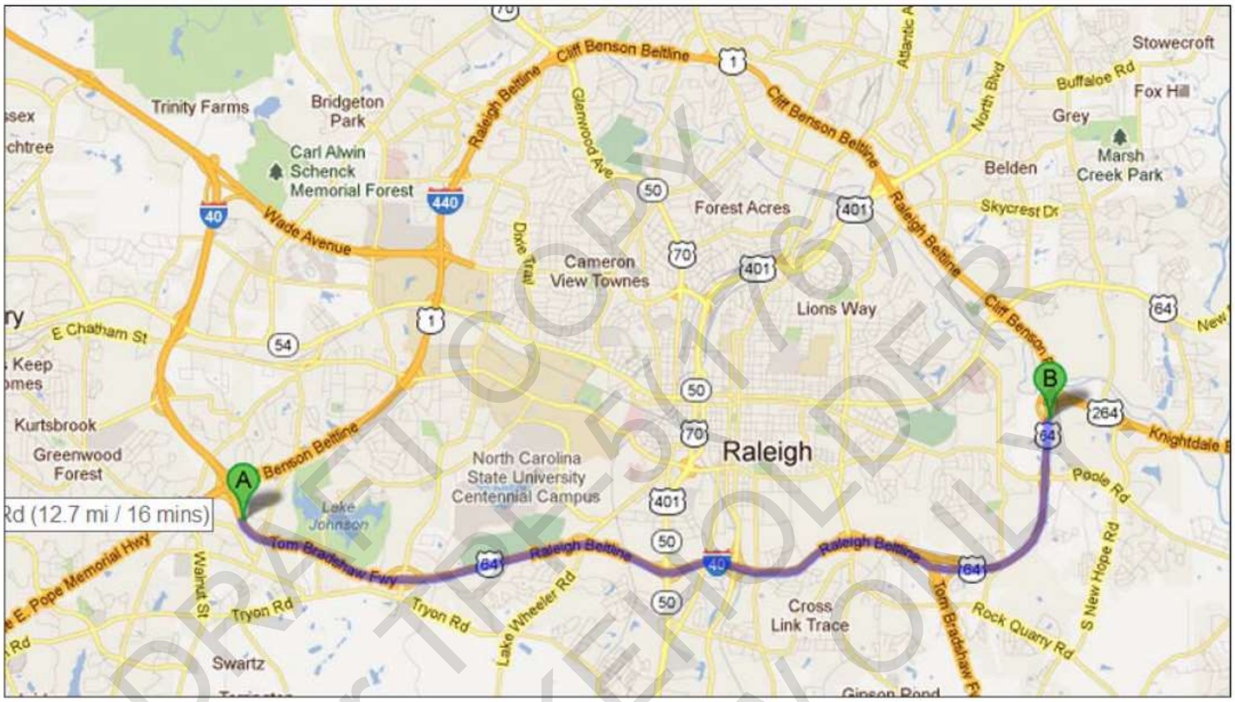


Figure 190. Map. I–40 work zone and study area.

The report provides the following introduction: "This report documents the results and project summary of NCDOT research project 2012-36: Work Zone Traffic Analysis & Impact Assessment. The project is tasked with assessing the estimated impact of proposed NCDOT TIP Project I-5311/I-5338, which is a pavement rehabilitation project on interstates I-40 and I-440 from Exit 293 to I-40 Exit 301 and I-440 Exit 14. The project aims to predict the network-wide impacts of this work zone during construction. The primary focus in the project and in this report is the development and calibration of a networkwide mesoscopic simulation model of the Triangle region using the DynusT and DTALite software tools, as well as a macroscopic evaluation in the FREEVAL tool. The geometric extents of the model cover the entire triangle region, with expansion to the east of the triangle to include additional sections of US264, I-40, and I-95. The model has been calibrated using field-estimated spot volume and speed data, as well as key route travel times obtained from INRIX. A variety of work zone scenarios were modeled to test the relative impacts of different lane closure configurations on route and network performance. The analysis further differentiates between no-diversion and with-diversion, where the latter relies on twenty iterations of a dynamic traffic assignment utility, which diverts traffic to minimize overall network travel time” (2014).

The second case study simulation (Whitman, Requardt & Associates LLP 2015) was originally developed for the Maryland Department of Transportation State Highway Administration (MDOT SHA), and then utilized further in a study entitled Narrowing Freeway Lanes and Shoulders to Create Additional Travel Lanes (FHWA, forthcoming). The section of roadway selected for analysis is in Montgomery County, Maryland, and spans approximately 5.5 miles in the freeway’s southbound direction between Shady Grove Road and I–270 West Spur to I–495. The corridor is illustrated in
Figure 191.

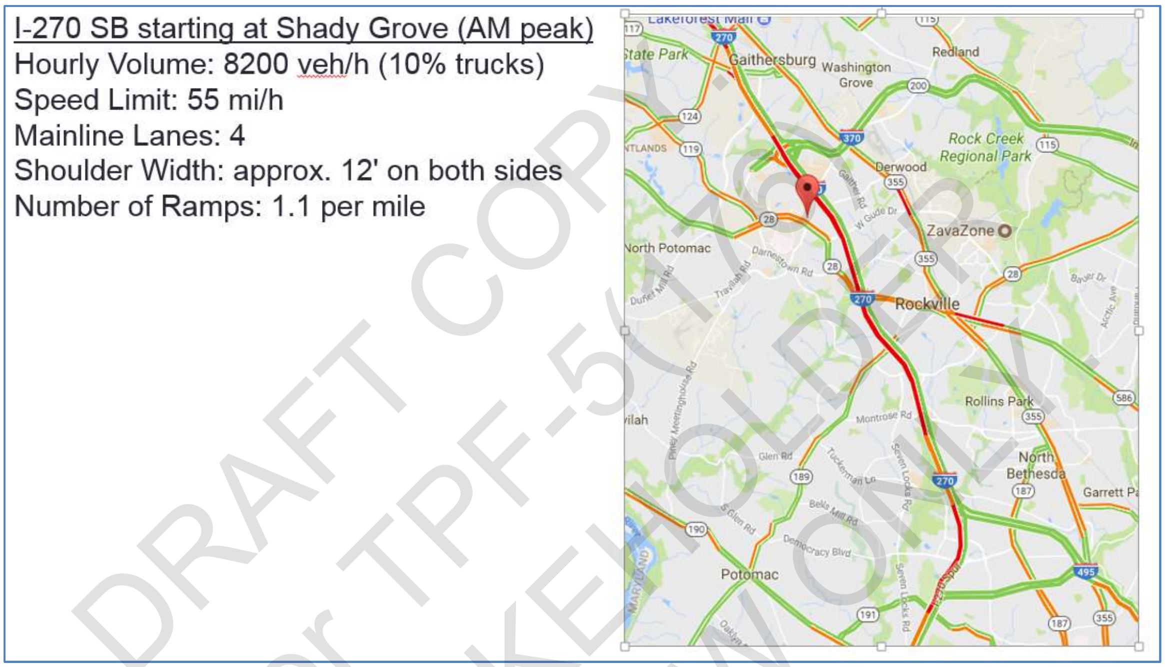


Figure 191. Map. Congested section of I–270 in Maryland.

The facility spans two collector-distributor lanes separated out from the four general purpose (GP) lanes by a jersey barrier and has shoulders widths varying anywhere from 2 feet to 10 feet. The GP lane closest to the median operates as a high occupancy vehicle (HOV) lane during the morning rush, from 6:30 to 9:30 a.m., Monday through Friday, and carries over 125,000 vehicles per day in the I–270 southbound direction alone. The collector-distributor lane ends at the southern end of this section of the I–270 corridor, just south of Montrose Parkway. In this 5.5 mile section of I–270, there are three slip ramps from the collector-distributor to the GP lanes and two slip ramps from the GP lanes to the collector-distributor lanes. Within the collector-distributor lanes, there are five merge locations, two diverge locations, and one weaving condition, excluding those generated by the slip ramps. The posted speed limit for both the collector-distributor lane is 55 miles per hour (mph). Noteworthy for the I–270 corridor in this section of roadway is the lack of intelligent transportation systems (ITS). There are no overhead gantries with lane assignment, variable speed signs, and only one dynamic message sign located at the southern end of the corridor. Bridge structures along the entire stretch of I–270 are a significant constraint to potential widening, as bridge piers interfere with the potential of shifting lanes or widening without significant structural costs. Additionally, this section of the I–270 corridor experiences significant delays due to capacity constraints just south of this corridor along the I–207 spur and further at the American Legion Bridge along I–495 going into Virginia. The I–270 corridor ranks in the top locations around the State for recurring and nonrecurring congestion, with a travel time index between 1.87 and 2.49 between Shady Grove Road and Montrose Parkway during the morning (AM) rush.

The third set of case study simulation models are described in the report, Alternative Designs to Alleviate Freeway Bottlenecks at Merge/Diverge and Weaving Areas (FHWA 2018). The report focused on two simulation models: the I–35 corridor in Kansas, and the I–66 corridor in Virginia. The facility is shown in
Figure 192.

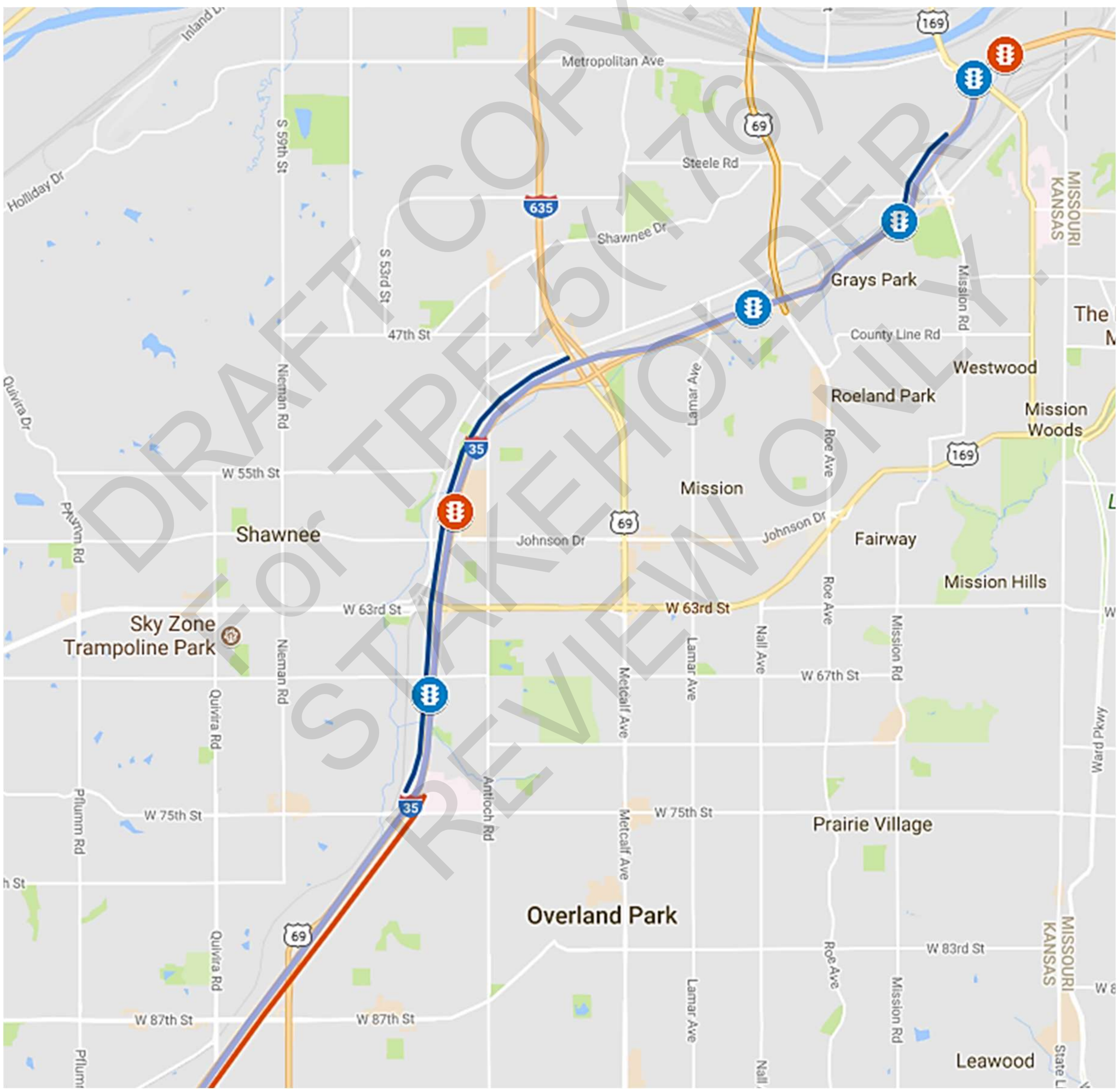


Figure 192. Map. I–35 Kansas City, Kansas, Facility.

Figure 193 illustrates the layout of the road. The existing HOV lane has open access to the adjacent lanes with two HOV-only entrances/exits.

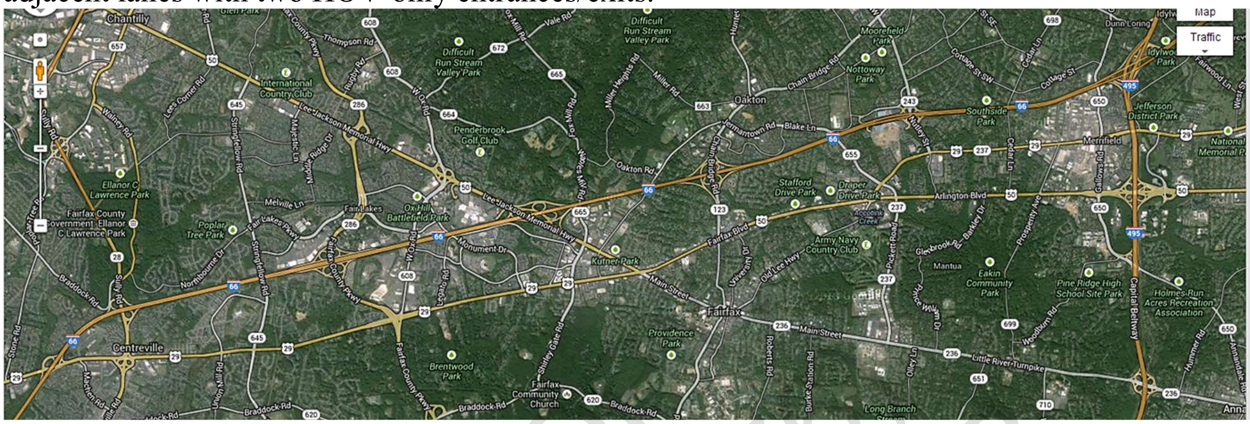


Figure 193. Map. Illustration of the I–66 freeway test bed.

I–66 outside the beltway:

- Start: mile marker 64→intersection with I–495.
- End: mile marker 51 → intersection with U.S. 29.
- Length: 13 miles.
- Number of interchanges: six.
- Average distance between interchanges: 2.3 miles.
- Number of lanes: four lanes/direction.
    - Normal lanes: three (in some sections, the right-most lane turns into hard shoulder).
    - One HOV lane.

The I–35 facility is approximately 10 miles long, and consists of several interchanges. The bottlenecks at the southbound direction are located at Southwest Boulevard (merge), Shawnee Mission Parkway (weave), and 67th Street (merge). The last two bottlenecks interact with each other and it appears that queue from 67th Street propagates upstream to Shawnee Mission Parkway, and even further upstream (spillback reaches Metcalf Avenue interchange). At the Southbound direction the Kansas Department of Transportation (KDOT) is operating ramp metering at 7th Street, Southwest Boulevard, 18th Street, and 67th Street. The I–66 Virginia (I–66 VA) simulation model has been used extensively for various research projects. The Virginia Department of Transportation (VDOT) has multiple active traffic management (ATM) deployments along the corridor. Currently, VDOT is upgrading I–66 outside the Beltway to include three regular lanes in each direction and two express lanes in each direction, from I–495 (the Capital Beltway) to Gainesville. The existing VISSIM simulation network is calibrated to traffic conditions prior to the upgrade. The simulation network is a 13-mile stretch of I–66 outside the Washington Beltway (I–495). This is a major east-west commuter corridor near Washington, DC, with four lanes in each direction. This stretch of the freeway experiences recurring congestion westbound in the evening (PM) peak every weekday.

12.1 DEFINING AND SCOPING THE SIMULATION ANALYSIS PROBLEM

Fundamental Project Parameters

Transportation System Simulation Manual (TSSM) 0describes the need to define and document the scope, goals, approach, budget, objectives, and key performance measures for any traffic simulation study. In the North Carolina study (2014), the project report described a desire to predict corridor and network-wide impacts of the I−5311/I−5338 project work zone during construction, including routes along the work zone corridor. In the Maryland study (2015), the project report described a desire to accurately depict I–270's existing conditions, as a vital step to effectively evaluate future traffic operations and alternatives along the corridor. Finally, the Alternative Designs report (2018) contained multiple sections (i.e., abstract, executive summary, introduction) to describe the objective of evaluating untested mobility solutions via microsimulation. Key performance measures for all projects will be described later in this case studies document.

Spatiotemporal Boundaries

TSSM 0emphasizes the importance of defining an appropriate set of spatial and temporal study boundaries, which can capture all relevant congestion. In the North Carolina study (2014), the spatial study boundaries were given in Figure 170 earlier. The temporal study boundaries were the morning (AM) peak (6–10 a.m.) and evening (PM) peak (3:30–7:30 p.m.). In the Maryland study (2015), the spatial scope was described previously as 5.5 miles in the freeway's southbound direction between Shady Grove Road and the I–270 West Spur to I–495. The temporal study boundaries were the AM peak (7–9 a.m.) and PM peak (4–6 p.m.). Finally, in the Alternative Designs report (2018), spatial scopes were described in the previous section of this document. For the I–35 Kansas (I–35 KS) analysis, temporal study boundaries were described as the afternoon peak (3:15–6:15 p.m.). For the I–66 VA analysis, temporal study boundaries were described as the PM peak (3:00–6:30 p.m.).

Role of Simulation

Macroscopic and mesoscopic simulation tools are commonly used to evaluate or predict regional impacts. For the North Carolina study (2014), the engineers chose to develop and calibrate a networkwide mesoscopic simulation model of the triangle region, as well as a macroscopic simulation of the work zone corridor. The macroscopic analysis tool FREEVAL was used to explore the estimated impacts on the work zone corridor. While FREEVAL is not able to predict diversion rates and networkwide performance, it has been proven to be a useful tool for assessing work zone impacts based on previous research conducted for NCDOT. FREEVAL is a faithful representation of the freeway facilities methodology in the Highway Capacity Manual (HCM), and as such, a benchmarked analysis tool across the United States (U.S.).

In the Maryland study (2015), the goal was to evaluate future traffic operations and alternatives along the I–270 corridor. Microscopic simulation via VISSIM was the chosen approach. Although mesoscopic simulation is less frequently selected for evaluating smaller facilities (such as a 5.5-mile stretch of I–270), the macroscopic FREEVAL approach would have also been viable here. The selection of microscopic versus macroscopic simulation for such analyses may depend on many factors including desired output data, available input data, staff expertise, and the need for moving vehicle animation. A multi-resolution approach (e.g., both macro and micro) is often valuable when resources permit.

Finally, in the Alternative Designs study (2018), the objective was to assess the mobility impacts of various innovative solutions (e.g., split merge design for acceleration lanes, managed lanes on the right, mainline metering, coordinated ramp metering, dynamic traffic calming). As such, microsimulation was chosen for this study. Microsimulation is typically preferable to mesoscopic or macroscopic simulation for calibrating and assessing the detailed lane-by-lane operations, which require car-following and lane-changing models.

Formation of Project Team and Stakeholders

The North Carolina project report listed eight co-authors from North Carolina State University's Institute for Transportation Research and Education, who were presumably part of the project team. The main customer was NCDOT. The report's front matter contained an acknowledgments section, naming nearly 2 dozen members from the project's Steering and Implementation Committee.

The Maryland project report was actually an internal memo to the Data Services Engineering Division (DSED) Travel Forecasting and Analysis. If the report's target audience had been external, a more explicit declaration of stakeholders and/or customers would have been useful.

Regarding the Alternative Designs project, a stakeholder group was formed early on in the project. This group consisted of subject matter experts from the public and private sectors. This group contributed their opinions and recommendations on project deliverables, including the final report. However, their names and organizations were not published in the final report. The customer was the Federal Highway Administration (FHWA) Office of Operations Research and Development. Several FHWA subject matter experts served on the project's review panel.

Identify Needed Resources

The freeway case studies in this document cited the following resources and data:

- DynusT and DTALite software tools.
- FREEVAL tool.
- FHWA Traffic Analysis Toolbox.
- HCM.
- Google® Earth.
- Triangle regional model (TRM) built in the TransCAD software.
- Point-based traffic volumes at key locations in the network.
- Point-based speed estimates at key locations.
- Route-based travel time and speed estimates along critical network routes.
- Sensor-based speed and volume data from traffic.com side-fire radar stations across the triangle region to support volume calibration and speed validation.

- Probe-based travel time data from INRIX to support validation of modeled route travel time to field observations.
- Custom point volume estimates requested from NCDOT at key locations outside of the traffic.com sensor coverage in the triangle.
- Sensor data for adjusting volume demand levels.
- Aerial imagery.
- Online mapping tools.
- Signal timings and travel times provided by the Department of Transportation (DOT).
- Heavy vehicle percentages.
- Posted speed limits.
- Speed and volume data collected by remote traffic microwave sensor (RTMS) trailers along major mainline segments.
- On- and off-ramp volume data collected by GoPro® cameras.
- Origin-destination (O-D) matrices estimated using the QueensOD software.
- Field data at freeway and on-ramp sensors through KCScout (i.e., the Kansas City ITS portal, http://www.kcscout.com/).

Define Forecast and Base Years

Forecast years are needed to determine whether a given traffic management strategy will remain the best option in the foreseeable future. They can also be used to assess how badly a status quo traffic network configuration may fail under increased future demands. However, the case studies from this document were primarily focused on existing conditions as opposed to future conditions. In the North Carolina study (2014), the TRM was developed in 2010 for the planning year 2015. In the Maryland study (2015), the goal was to develop an I–270 microsimulation model that would be calibrated to existing conditions in the year 2015. In the Alternative Designs study (2018), the microsimulation models were calibrated to 2015–2016 conditions.

Range of Operational Conditions to Consider

In recent years, transportation professionals have begun to realize the significance of annual performance and travel time reliability. This requires explicit analysis of operating conditions including varying traffic demands, weather, incidents, work zones, and special events. In the North Carolina macroscopic simulations, incident scenarios were analyzed as shown in

Table 29. This analysis eventually found the magnitude of the incident impacts is severely higher in work zone scenarios with only two lanes open instead of three lanes open.

Table 29. Incident scenario details.

Route	Run #	WZ Scenario/ Time Period	Incident Location	Lanes Open in Incident	Incident Duration	Incident Segment Number *	Incident CAF
C	I1	WZ2/PM	I-40 EB @ mm 293	2→1	2 hours (4-6 PM)	18	0.44
	I1b	WZ6/PM	I-40 EB @ mm 293	3→2	2 hours (4-6 PM)	18	0.47
	I2	WZ2/PM	I-40 EB @ mm295	2→1	2 hours (4-6 PM)	23	0.44
	I2b	WZ6/PM	I-40 EB @ mm 295	3→2	2 hours (4-6 PM)	23	0.47
	I3	WZ2/PM	I-40 EB @ mm 299	2→1	2 hours (4-6 PM)	35	0.44
	I3b	WZ6/PM	I-40 EB @ mm 299	3→2	2 hours (4-6 PM)	35	0.47
	I4	WZ3/AM	I-40 WB @ mm 297	2→1	2 hours (4-6 PM)	44	0.44
	I4b	WZ5/AM	I-40 WB @ mm 297	3→2	2 hours (6-8 AM)	44	0.47
	I5	WZ3/AM	I-40 WB @ mm 301	2→1	2 hours (6-8 AM)	21	0.44
	I5b	WZ5/AM	I-40 WB @ mm301	3→2	2 hours (6-8 AM)	21	0.47
D	I6	WZ2/PM	I-440 EB @ mm 10	3→2	1 hour (5-6 PM)	41	0.74
	I7	WZ1/AM	I-440 WB @ mm 5	3→2	2 hours (6-8 AM)	48	0.74

The Maryland I–270 microsimulation study (2015), designed to create a test bed for future analysis of treatment alternatives, focused on calibrating to existing conditions without consideration of the operating conditions. Similarly, in the Alternative Designs study (2018), the I–35 KS and I–66 VA microsimulation models were calibrated to existing conditions without consideration of the operating conditions. However, for a more robust calibration that accounts for operating conditions more explicitly, the cluster analysis procedures from the Traffic Analysis Toolbox Volume III (Wunderlich, Vasudevan, and Wang 2016) could be considered.

Screening of Initial Alternatives through Non-Simulation Methods

When a traffic analysis must consider a large number of traffic control alternatives, roadway geometric design alternatives, and/or demand management alternatives, it might not be practical to analyze each alternative through simulation. In such cases, non-simulation-based tools (e.g., sketch-planning tools) may help to weed out some portion of the ineffective alternatives being considered.

Another alternative is that faster simulation tools requiring less input data could be used to screen alternatives for the more data-intensive simulations. In the North Carolina study (2014), the macroscopic simulation did not weed out any alternatives, but did help to reinforce the finding that three-lane work zones would cause much less congestion than two-lane work zones.

Yet another alternative is commonly used in microsimulation studies, as exemplified by the Alternative Designs project. In this project, a significant amount of sensitivity analysis was conducted using relatively small toy networks, which do not contain real-world data. Such sensitivity analyses using toy networks are able to quickly examine a large number of input combinations to screen out the ineffective alternatives. Subsequently, the much larger and slower real-world simulations can be used to confirm and finalize the results.

Explicit Linking of Performance Measures to Agency Goals or Project Objectives

The objective of the North Carolina study (2014) was to predict work zone impacts on freeway operations. Table 30 illustrates the following chosen performance measures:

- Travel time.
- Maximum queue length.
- Maximum denied entry queue length (DEQL).
- Maximum demand-to-capacity (D/C) ratio.

Table 30. Key performance measures from macroscopic simulation.

Route	Direction/ Time Period	Scenario	Travel Time (Minutes)	Max Queue Length (Miles)	Max DEQL (Miles)	Max d/c Ratio
C	EB / PM	Base	27.4	3.3	0.0	1.16
		WZ2	320	16.2	25.1	3.4
		WZ2b	214.2	16.2	20.9	2.8
		WZ4	122.7	11.6	30.7	2.4
		WZ6	77.8	10.4	0.0	2.3
		WZ7*	106.5	10.3	19.2	2.4
	WB / AM	Base	26.8	0.5	0.0	1.2
		WZ1	115.9	13.5	1.38	3.8
		WZ1b	58.9	8.8	0.0	2.6
		WZ3	46.9	10.9	0.0	2.5
		WZ5	47.3	10.9	0.0	1.6
		WZ7*	36.2	10.9	0.0	2.5
D	EB / PM	Base	31.7	1.0	0.0	1.2
		WZ1	68.0	8.9	0.0	3.7
		WZ1b	40.3	3.0	0.0	2.5
		WZ3	39.5	2.6	0.0	2.5
	WB / AM	Base	47.4	7.2	0.0	1.7
		WZ2	92.9	16.8	0.0	3.0
		WZ2b	67.9	14.5	0.0	2.6
		WZ4	55.6	14.0	0.0	2.0

* This scenario is designed for the EB direction, but the analysis in the WB direction considers onlooker delay impacts

Source: Institute for Transportation Research and Education.

The objective of the Maryland project was simply to develop a calibrated I–270 microsimulation model, which could be used to evaluate various congestion mitigation strategies in the future. To accomplish this, the following performance measures were chosen:

- Speed comparison tables.
- Travel time graphs with +/- 10 percent bands.
- Travel time graphs with 95 percent confidence intervals.
- Traffic volume comparison tables.
- Travel time comparison tables.
- HCM 2010 corridor level of service (LOS).

Finally, in the Alternative Freeway Designs study (2018), the objective was to assess the mobility impacts of various innovative solutions (e.g., split merge design for acceleration lanes, managed lanes on the right, mainline metering, coordinated ramp metering, dynamic traffic calming). To illustrate these impacts, the researchers selected the following performance measures:

- Networkwide outputs.
- Segment-specific outputs (on both ramps and mainlines).
    - Throughput.

- Travel times.
- Speeds (sometimes using heat maps).
- Delays.
- Congestion duration.
- Latent delays and queues.
- Number of lane changes.

12.2 DATA

Origins of Data Used in the Analysis

The North Carolina case study report (2014) cited the following four origins of data used in the analysis:

- Sensor-based speed and volume data from traffic.com side-fire radar stations across the triangle region to support volume calibration and speed validation of model results.
- Probe-based travel time data from INRIX.com to support validation of modeled route travel time to field observations.
- Custom point volume estimates requested from NCDOT at key locations outside of the traffic.com sensor coverage in the triangle.
- Arterial traffic counts on key non-freeway routes in the triangle, which are likely to serve as key diversion routes to the proposed work zone.

The Maryland case study report (2015) cited the following origins of data used in the analysis:

Peak Hour Traffic Volumes

DSED Travel Forecasting and Analysis developed 2015 peak hour traffic volumes for the study area. The AM and PM peak hours were determined to be 7 a.m. to 8 a.m. and 5 p.m. to 6 p.m. The traffic demand was balanced throughout the entire network for both AM and PM peaks.

Signal Timings

Signal timing data was provided by DSED for signalized intersections within the corridor to ensure that the VISSIM models included accurate existing signal timings. The VISSIM model .RBC files were verified and/or modified to match the signal data.

Travel Times and Speeds

Travel time and Regional Integrated Transportation Information System (RITIS) data were provided by DSED for segments along the I–270 general purpose, HOV, and collector distributor lanes.

- The vehicular travel time data was collected between 7 a.m. and 9 a.m. and between 4 p.m. and 6 p.m.; runs were conducted in the peak direction along the I–270 GP lanes from Maryland Route 121 (MD 121) to the I–270 Spur and along the entire length of the I–270 collector-distributor lanes. These data were then used to validate the existing VISSIM models to reflect current travel behaviors.
- RITIS data were also used to ensure the models were producing reasonable travel speeds within the study area. These data were used to validate roadway segments not included in the travel time data: I–270 between I-70 and MD 121 and for all segments south and west of the I–270/I–270 Spur split.

The Alternative Freeway Designs project report (2018) cited the following origins of data used in the analysis.

Interstate 35 Kansas

Traffic volume and speed data at the mainline, on-ramp and off-ramp were obtained from the Kansas City (KC) Scout Portal (http://www.kcscout.net/KcDataPortal). The VISSIM model provided by KDOT was used to obtain the percentage of passenger cars, heavy goods vehicles (HGV), and large heavy goods vehicles (L-HGV). For the desired speed, the mainline detectors' speed readings obtained from KC Scout were reviewed, and a mean speed of 64 mph was identified. The desired speed distribution follows an S-shaped curve.

Interstate 66 Virginia

Field data were collected to build time-dependent travel demand O-D matrices. Traffic counts and speeds were collected on each individual lane, including HOV lanes, for the mainline, on-ramps, and off-ramps. A combination of RTMS radar detectors, video cameras, GoPro©

cameras, and manual counts were used to obtain speeds, volumes, and occupancies. Their deployment is shown in
Figure 194.

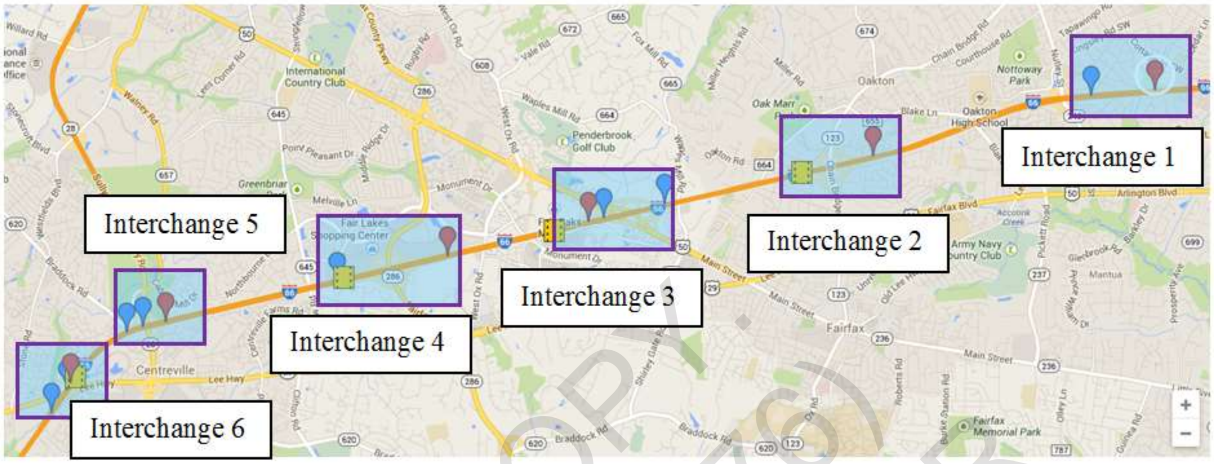


Figure 194. Map. Locations of remote traffic microwave sensor trailers and GoPro© cameras.

Description of Speed Measurement Method

The North Carolina study (2014) obtained point-based speed estimates from sensors at key locations. This included sensor-based speed data from traffic.com side-fire radar stations across the triangle region. For each key analysis route, the team extracted 15-minute average speed data on Tuesdays, Wednesdays, and Thursdays of the 2nd week of each month in 2011, as well as in January 2012. These analysis periods are consistent with the point data collected. Route data were obtained from the INRIX.com online data repository.

Point data were collected at 23 locations in the study area. Seventeen locations were covered by traffic.com sensors. Traffic.com uses side-fire radar technology to continuously record speed and volume data, which are archived online in different aggregation levels. At the remaining six locations with no sensors, a request for counts was submitted to NCDOT and traffic counts were collected for a period of 2 days at each location. At these locations, only 15-minute traffic counts were collected; speed data were not available. All point data collection locations are shown in Figure 195.

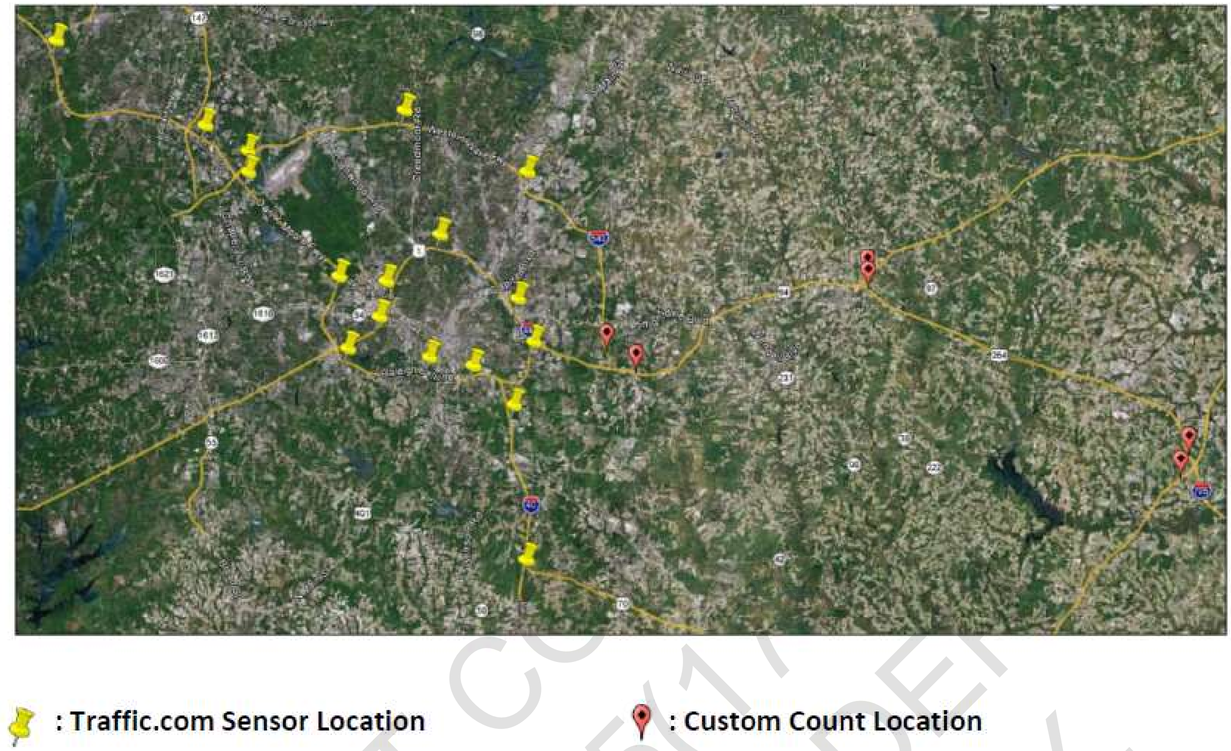


Figure 195. Map. Point data collection sites.

Traffic.com sensors collect speed and volume data aggregated over 5 minutes, 15 minutes, 1 hour, or 24 hour time intervals. In this study, a 15-minute aggregation level was used. Data were collected on Tuesdays, Wednesdays, and Thursdays of the 2nd week of each month in 2011, as well as in January 2012. Therefore, at each sensor location, 13×3=39 observations were available for speed and volume that were used to estimate the average speed and volume for each time interval. Speed-flow curves and speed and volume profiles were obtained for all traffic.com sensors.

In the Maryland study (2015), RITIS probe data were used to ensure the models were producing reasonable travel speeds within the study area. These data were used to validate roadway segments that were not included in the travel time data: I–270 between I–70 and MD 121 and for all segments south and west of the I–270/I–270 Spur split.

In the Alternative Designs project, I–35 KS speeds were obtained from a combination of RTMS radar detectors, video cameras, GoPro© cameras, and manual counts. I–66 VA speed and volume data were collected by six RTMS trailers along major mainline segments.

Collection of Vehicle Type Classifications

For the Maryland study (2015) and the I–35 KS simulation model, vehicle type classifications were obtained from previously developed microsimulation networks. However, the origins of these classifications were not given. Vehicle type classifications were not discussed in the North Carolina case study report (2014), or in the I–66 VA report (2018).

Collection of Calibration Data

For the freeway case studies described in this document, most of the data collection described earlier in this section were for calibration. Tables and figures containing these data will be presented later in this document, in the calibration and analysis sections.

Use of Probe Data, National Performance Management Research Data Set, Vehicle Trajectories, Aerial Imagery

In the North Carolina study (2014), probe data were used in the calibration of speeds and travel times. Tables and figures showing the results of this calibration based on probe data will be presented later in this document, in the calibration and analysis sections. In their simulation models, the North Carolina study team defined segmentation of key routes from aerial imagery.

In the Maryland study (2015), vehicle trajectories and probe data were both used in the calibration of speeds. Tables and figures showing the results of their calibration will be presented later in this document, in the calibration and analysis sections. Aerial imagery was used in their simulation network development process.

In the I–35 KS study (2018), speeds were calibrated using data from the KC Scout portal, but it is unclear whether those speeds were obtained from probe data.

Application of Data Processing Principles (Aggregation, Fusion, Imputation)

The North Carolina project used a combination of point-based sensor speed and volume data, and segment-based travel time estimates to compare the predictions of the mesoscopic simulation models to empirical data. Two types of field data were collected to be used in the calibration process of the mesoscopic dynamic traffic assignment model: (a) point data and (b) link data. Point data were collected at certain locations of the network and typically included 15-minute average speed and 15-minute volume. The team used two main sources to gather point data: 1) traffic.com sensors and 2) custom data collection. Link data mainly includes 15-minute average travel time and average speed along a link. Link data were downloaded from INRIX.com that covers all freeways and most of the arterial streets of interest.

Some additional traffic count data were needed outside the NCDOT sensor coverage in the triangle region, which were requested from and obtained by NCDOT for this project. The custom data collection included the following key locations in the study area, needed for calibration of the mesoscopic baseline model:

1. I–540 north of U.S. 64/U.S. 264 interchange.
2. U.S. 64/U.S. 264 east of I–540 interchange.
3. U.S. 64 north of U.S. 64/U.S. 264 split.
4. U.S. 264 east of U.S. 64/U.S. 264 split.

5. U.S. 264 northwest of I‒95 interchange.
6. I‒95 southwest of U.S. 264 interchange.

At these six locations, 15-minute traffic count data were collected for a period of 2 days. For each 15-minute time interval, the average of the two readings was used.

A close investigation of sensor data revealed that westbound/northbound volumes from Johnston County generally peaked about 1 hour earlier than the eastbound/southbound movements from Orange and Durham County. With the model requiring generalized O-D volume factors over the entire network, the team used an average volume profile across sensor stations as the best feasible estimation across the network.

In addition to the point data, 10 key calibration routes (A–J) were identified for data collection. These ten routes were defined to include the work zone, key alternate routes (I‒440, I‒540, etc.), and key urban arterials that could be used to carry some of work zone traffic. The calibration routes will be used to compare results of the base mesoscopic simulation model to real-world data. Calibration routes were selected to provide a representative sample of key routes in the triangle region. All calibration and analysis routes are shown in
Figure 196 and Table 31.

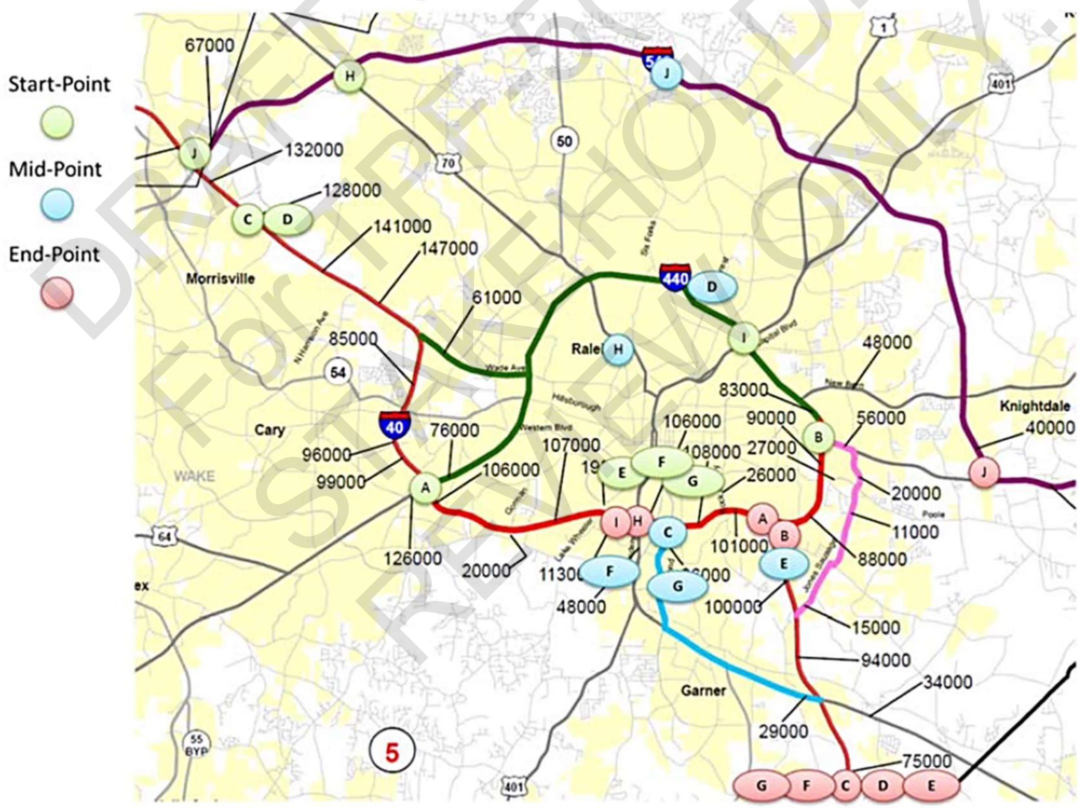


Figure 196. Map. Key route definitions.

Table 31. Description of routes.

#	Route Description	Start Milepost	Start Facility	End Milepost	End Facility
A	I40 WZ Section	293	I40-I440	301	I40-I440
B	I440 WZ Section	14	64/264 SPLIT	16	I40-I440
C	Triangle Route I-40	284	I-40	312	
D	Triangle Route I-440	284	I-40	312	
E	South Approach - I-40	Downtown	Saunders	312	I-40
F	South Approach - Saunders	Downtown	Saunders	312	I-40
G	South Approach - Hammond	Downtown	Saunders	312	I-40
H	US 70 Arterial	I540		I40	
I	US401 Arterial	I440		I40	
J	I540 Detour	1	I40	26	64/264

Source: Institute for Transportation Research and Education.

For each route, the team extracted 15-minute travel time and average speed data on Tuesdays, Wednesdays, and Thursdays of the 2nd week of each month in 2011, as well as in January 2012. These analysis periods are consistent with the collected point data. Route data were obtained from the INRIX.com online data repository. Each route includes several links, and the team summed up link travel times to yield route travel time. Like point data, for most of the routes, travel time for each 15-minute time interval is the average of 39 travel time observations throughout 2011 and January 2012. For each route, the team estimated average travel time, average travel time plus standard error, and average travel time minus standard error profiles.

In the Maryland study (2015), traffic demand was balanced throughout the entire network for both AM and PM peaks. Appendix A of the I–270 Modeling Calibration Methodologies memorandum (Whitman, Requardt & Associates LLP 2015) presents some of these details.

In the I–66 VA model development and calibration process, mainline volume data were obtained from RTMS trailers, but ramp volume data could not be obtained this way. Therefore, the ramp volumes were collected by GoPro© cameras. O-D matrices were then estimated using the QueensOD software.

Statistical Analysis of Field Data Sample Size

Many traffic engineers understand the value of statistical analysis of simulation outputs, but statistical analysis of collected field data may be less recognized. Field data are needed for both simulation inputs and ground truth calibration values. A statistical analysis of the collected field data can help to ensure the simulation input and output data both reflect reality. However, as with many simulation reports, the freeway case studies described in this document only provided statistical analysis of the output data.

12.3 CREATING THE NETWORK

Intelligent Selection of Initialization Fill Time

TSSM Chapter 7. states that many traffic simulation tools offer some sort of input entry for initialization time, warm-up time, fill time, or some similar name. The typical objective of the initialization time is for the simulation model to reach a state in which the number of vehicles entering and exiting the network is roughly equal. This would normally facilitate a more accurate simulation analysis. An intersection, corridor, or network model should have an approximately equal number of vehicles entering or exiting the virtual system, prior to collection of vital performance measures. Failure to achieve this balance may mean vehicles have not had enough time to fill up the system, or may indicate coding errors (e.g., missing links or nodes). If vehicles have not had enough time to fill the system, congestion-related performance measures could be overly optimistic. In other cases, it may mean that conditions were oversaturated during the system initialization. In many cases, accuracy of the simulation results may be compromised.

The North Carolina case study (2014) involved both macroscopic and mesoscopic simulation models. While fill times tend to be essential for microscopic simulation models, macroscopic and mesoscopic models vary in terms of their use of fill times. While the ability to edit and/or customize fill times as a function of traffic network size and/or congestion is nearly always available within microscopic simulation tools, this input parameter may not be provided for all macroscopic and mesoscopic simulation tools. In this case, the North Carolina case study report (2014) made no mention of fill times, initialization times, or warm-up periods.

By contrast, the other freeway case studies involve microsimulation, and the report from each of these studies provides information on the fill time. The Maryland I–270 microsimulation report (2015) gave the following description:

> *The AM model includes an initialization (seed) time of 5,400 seconds (1 hour 30 minutes), followed by 3,600 seconds (1 hour) of actual simulation time during which data was collected by the VISSIM model; the actual simulation time is the equivalent to the established AM peak hour. The seeding period was not only based on the total travel time from one end of the study area to the other, but also on the amount of congestion established prior to the AM peak hour. This initialization time was necessary to populate the network and to produce the appropriate congestion prior to VISSIM recording data for analysis. The PM model includes an initialization (seed) time of 1,800 seconds (30 minutes), followed by 3,600 seconds (1 hour) of actual simulation time during which data was collected.*

Similarly, the I–35 KS case study (2018) gave the following narratives for the sensitivity analysis and calibration efforts, respectively:

> *The simulation analysis is divided into three time periods. The first 900-second time period is assigned as the warm-up period, and thus no traffic evaluation measurements are conducted in this time period. Traffic evaluation measurements start at the beginning of the second 900-seconds period.*

For the model calibration, a day without incidents throughout the corridor or adverse weather was selected. The selected calibration day was April 22, 2016. The afternoon peak period starting from 3:15 PM to 6:15 PM was considered as the simulation period, and 25 minutes were added in the beginning of the simulation period for initialization.

Of these discussions, only the I–270 discussion provided a rationale for their chosen amount of fill time. Ideally, a simulation report should provide a discussion of this rationale whenever possible.

Analyzing the Consequences of Inadequate Study Boundaries

In a traffic simulation, when the spatial limits of the virtual network fail to fully capture the extent of congestion, this can negatively affect certain aspects of an analysis. However, other aspects of the analysis may be unaffected. Engineering judgment can be used to assess the consequences of inadequate study boundaries. For the freeway case studies described in this document, each published report provided a discussion of the chosen spatiotemporal boundaries. Documentation of these boundaries was included near the beginning of this arterial case study document, in 0.

The North Carolina project report (2014) provided the following discussion of study boundary adequacy:

Figure 197 *depicts the maximum DEQL and the maximum mainline queue length of work zone scenario 2. As expected, the queue length drops as the number of vehicles remaining on the mainline freeway is reduced. The exhibit makes evident that a 30% diversion is necessary to contain the queue within the extents of the modeled facility (DEQL equal zero), but that significant queuing on the facility remains even for this diversion percentage.*

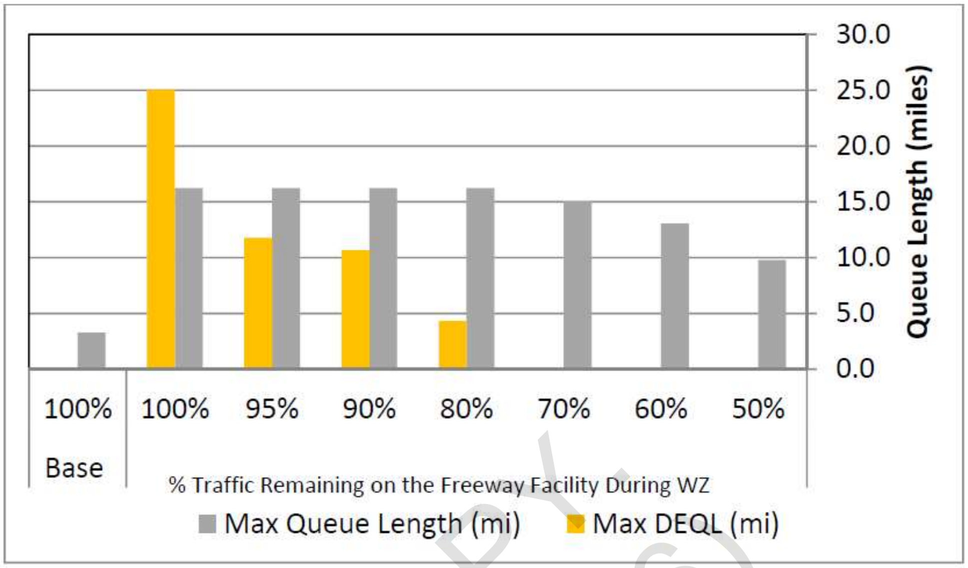


Figure 197. Chart. Work zone denied entry queue length versus mainline queue length.

The Maryland project report provided the following comment on study boundary adequacy:

> *For the PM model, speed reduction zones were coded at outbound links of the network to replicate lower speeds caused by friction created outside of the network. These reduction zones are located on I-270 northbound north of I-70, I-270 southbound south of MD 187, and I-270 Spur westbound south of Clara Barton Parkway.*

The I–35 KS project report (2018) cited a need for the calibration day to start in uncongested conditions and end in uncongested conditions, to verify the calibrated model is functioning properly. The spatial limits of the network play a role in determining the temporal limits, because some roadway segments experience longer lasting congestion than others. Although Figure 198 shows the vertical time periods contain green (i.e., uncongested) periods at the top and bottom, the horizontal spatial segments do not contain green periods on the far right. This implies that a more accurate model may have been obtained by extending the physical network boundaries beyond the far right segment (i.e., beyond the 67th Street Merge).

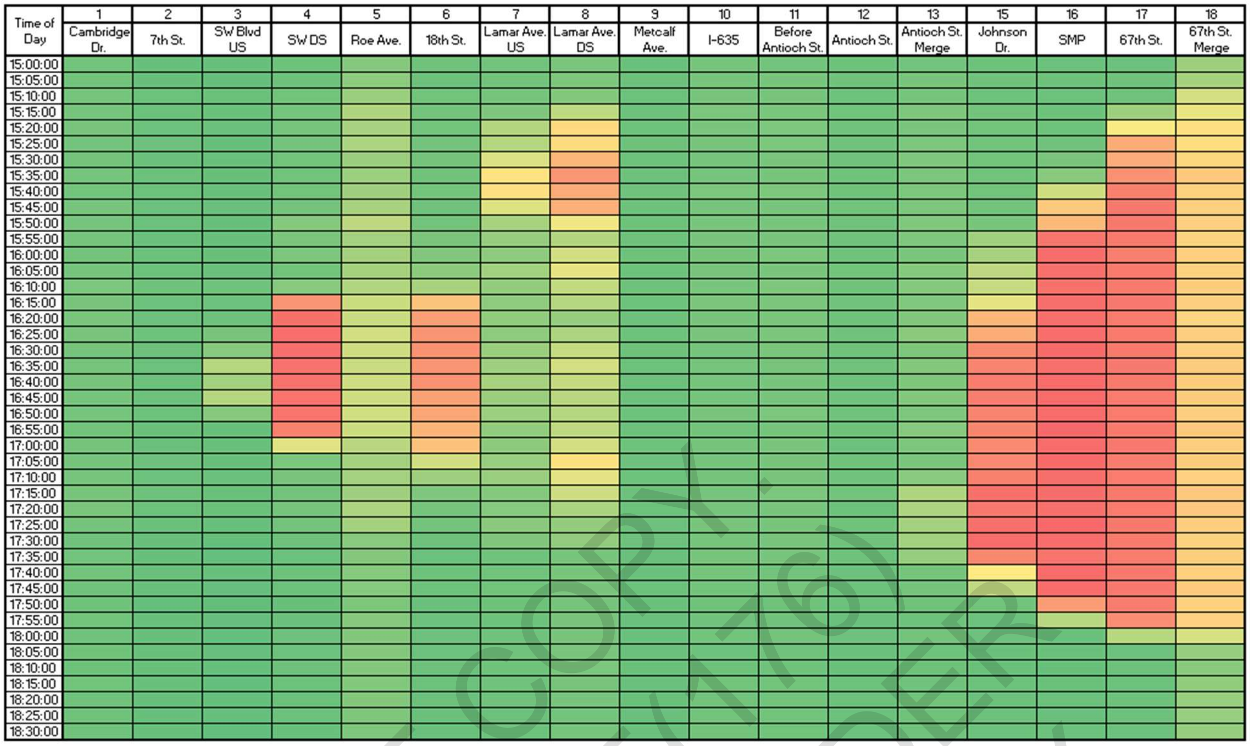


Figure 198. Heat Map. KC Scout speed profile.

Informed Selection of Overall Data Entry Method

TSSM Chapter 7. alluded to the advantages and disadvantages of various traffic network creation methods. These advantages and disadvantages are often considered during the traffic analysis tool selection process, but should also be documented for future reviewers. The North Carolina report gave the following narrative:

> *The network was converted from the TRM (Triangle Regional Model), which was developed in 2010 for the planning year 2015.*
> Figure *199 illustrates the original DynusT network. Red lines represent freeway facilities with a total of 1,165 links and 649 on and off ramps. These include I-40, I-85, I-440, and I-540 as the major interstate freeways in the network. Blue lines represent highway facilities (principal arterials, 2,644 links) and black lines represent arterial roads (minor arterials and collectors, 15,792 links). The DynusT network comprises 2389 traffic analysis zones, 9527 nodes, and 20,250 links.*
>
> *The team developed the [macroscopic simulation model] FREEVAL segmentation for key routes through the work zone from aerial imagery. The volume input was obtained directly from the calibrated DynusT baseline files for the AM and PM peak period.*

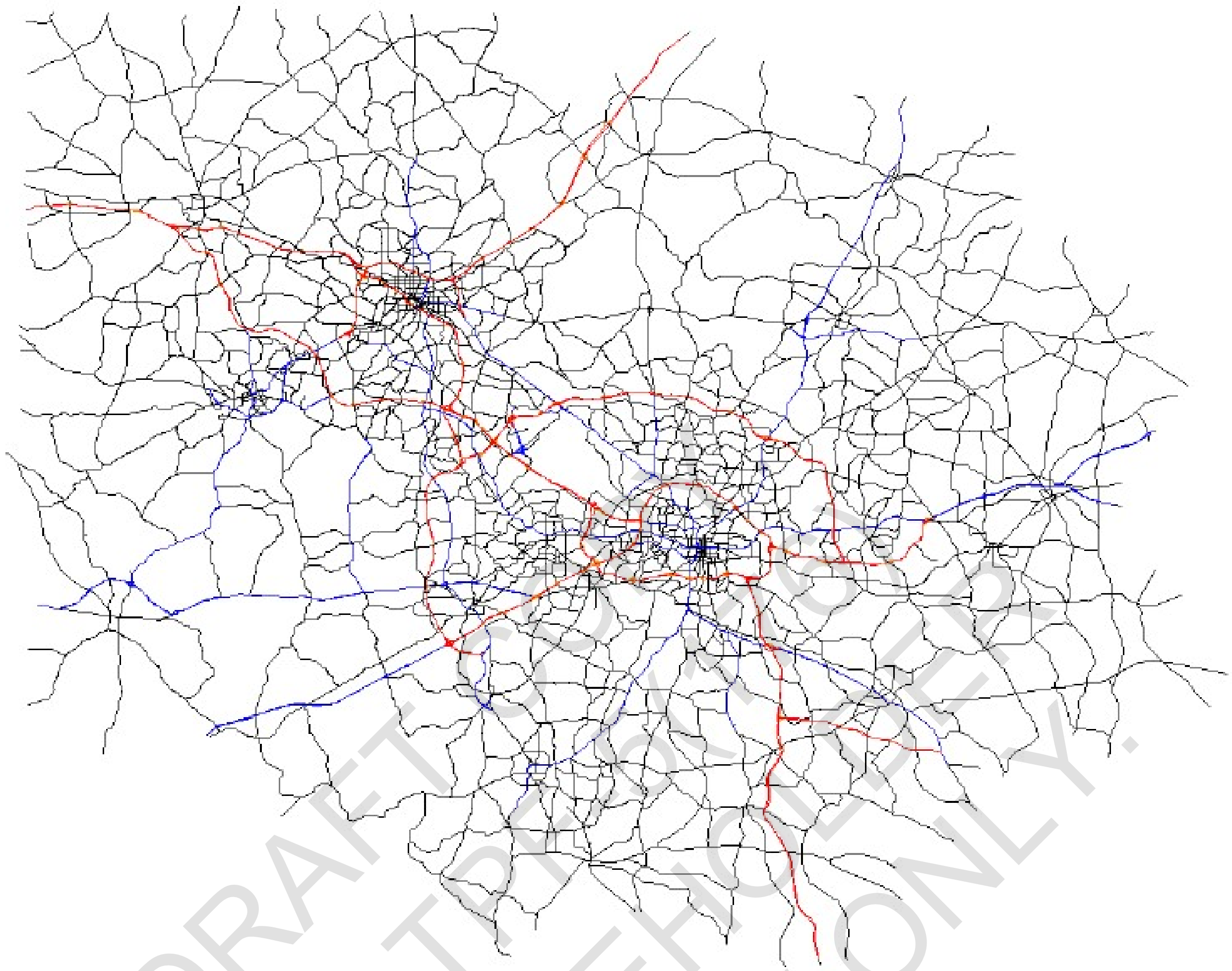


Figure 199. Map. Base network for mesoscopic simulation (North Carolina study).

Although the Maryland case study report provided a thorough discussion of the data they used, it did not comment on the data entry method they used. The I–35 KS network data set was obtained from the KDOT, and as such the original data entry method was unknown. The I–66 VA network data set was similarly obtained from prior studies. The QueensOD tool was used to re-estimate origin-destination flows as mentioned earlier.

Use of Pre-Processors and/or OpenStreetMap™

Some pre-processor tools can generate traffic network data sets that can be imported and processed by a simulation tool. In other cases, a simulation tool may have the ability to import data from third-party sources and/or formats. Some simulation tools can import detailed data from geographic information systems (GIS) systems such as OpenStreetMap™.

The North Carolina report (2014) mentioned their decision to import data from the TRM into their mesoscopic simulation model. This represents one example of a simulation tool having the ability to import data from third-party sources and/or formats. For their parallel macroscopic simulation effort, although they mentioned use of volumes from the mesoscopic model, as well as aerial imagery to obtain segmentation, there was no documentation to suggest pre-processing or automated importing of this data. The I–66 VA report (2018) mentioned use of the QueensOD method and/or tool (Van Aerde et al. 1998) to estimate O-D demands. This does not represent importing data from a third-party source, but rather a repurposing and reorganizing of traffic demand data already available to the analyst. The Maryland case study (2015) report did not mention use of third-party tools or data sources in the manner described by this sub-section.

Entry of Demands Instead of Throughput Volumes

Although the terms demand and volume are often (and sometimes inappropriately) used interchangeably, it is critical the user enters demands instead of volumes. This is because volumes observed in the field are only reflective of vehicles discharged, whereas demands indicate how many vehicles are attempting to use the system. This does not preclude the later use of throughput volumes for calibration and validation purposes. Due to growing traffic congestion and improving traffic analysis tools in the past few decades, more transportation professionals and academics are recognizing the importance of distinguishing between demands and volumes.

One estimation of hourly demand comes from the annual average daily traffic (AADT) figure. This accounts for all trips served throughout the day, even if some of the time periods were oversaturated. The North Carolina study (2014) used this method:

> *The I-40 and I-440 corridors carry approximate Average Annual Daily Traffic (AADT) volumes up to 113,000 and 90,000 vehicles per day (vpd), respectively. These demand levels already produce recurring peak-hour congestion near the proposed work zone.*

For analyses of future conditions, future demand estimates may be derived from growth factors applied to existing AADTs. In the Maryland study (2015), the year 2015 demand values were forecasted estimated by DSED Travel Forecasting and Analysis, (presumably a few years prior to 2015).

In the I–66 VA report (2018), the authors stated that field data were collected to build time-dependent travel demand origin-destination matrices.

Analysis of Origin-Destination Demands

Naturally, the most accurate traffic simulations provide realistic origin-destination travel paths. The North Carolina report (2014) had the following discussion:

> *The Origin-Destination (O/D) Matrix for the DynusT model was obtained from the Triangle Regional Model, which is a four-hour peak period model. As a result, the team needed to estimate hourly factors from field sensors, to split the four-hour demand into four separate hourly O/D matrices with the appropriate peaking characteristics.*

The Maryland case study report (2015) provided the following discussion of routing, which is analogous to the O-D concept:

The static routing decisions were coded in VISSIM such that the beginning of each route is as far upstream of the first decision point as possible; this method allows vehicles to make a routing decision as soon as possible, preventing unnecessary friction and congestion. In instances where routing decisions were close together, route combinations were applied to ensure realistic lane changing behavior.

The I–66 VA report (2018) provided the following O-D analysis information:

The field-collected data in this study could identify how many vehicles traveled between some, but not all, O-D pairs. To fill in the gaps, O-D matrices were estimated using the QueensOD software. For example, Table 32 lists an estimated O-D matrix for the 3:00-3:15 pm time period.

Figure 200 *compares estimated flows with observed flows of the same period. Results indicate an excellent correlation between estimated and field-measured O-D trips.*

Table 32. Estimated origin-destination (O-D) matrix for the 3–3:15 p.m. time period.

	2	3	4	5	6	7	8	9	10
1	457	97.2	375.8	149	1.4	228.1	756.1	417.9	3599.7
2		287.2	414.8		14.2		31.3		
3			861.4		72.2		0.1		
4					242		71.7	39.1	41.7
6							1164.5	8.9	401.8
8								64.2	1223.8
9									925.6

Source: FHWA.

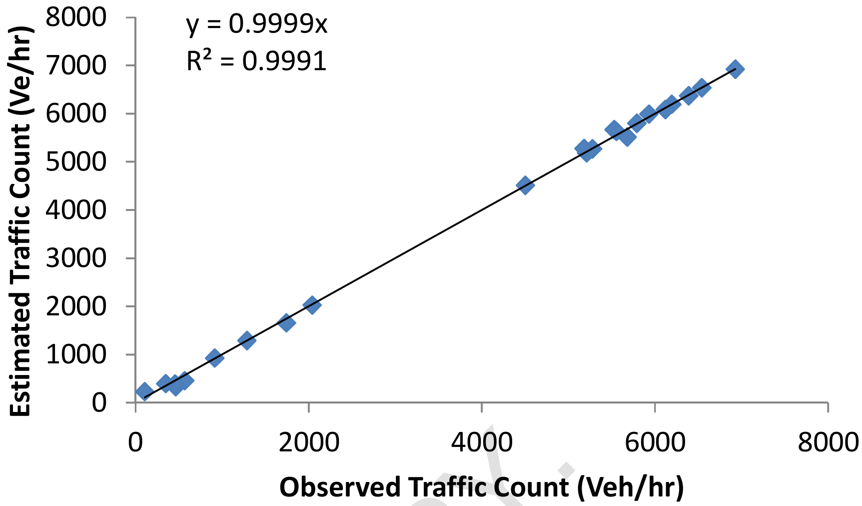


Figure 200. Graph. Comparison of estimated flow and observed flow: 3–3:15 p.m.

Application of a Specific Volume Balancing Method

For various reasons, upstream and downstream traffic volume counts obtained from the field are not always consistent with one another. Figure 201 illustrates an example of this phenomenon. Severely inconsistent counts are likely to produce inaccurate simulation results. Volume balancing methods are available to mitigate this problem. When such methods are used, their use should be documented for the reviewer's benefit and understanding.

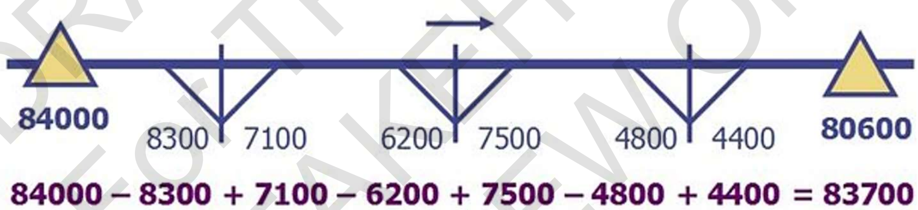


Figure 201. Diagram. Unbalanced upstream and downstream volumes.

As shown in

Figure **201**, the FREEVAL macroscopic simulation model focuses on the evaluation of a single facility (pipe), but with consideration of all merge, diverge, and weaving segments. The North Carolina study involving FREEVAL did not specifically mention flow balancing, but the HCM procedures implemented within FREEVAL automatically perform some of this balancing.

The Maryland case study report (2015) said that traffic demand was balanced throughout the entire network for both AM and PM peaks. Their demand balancing outcomes are given in Appendix A of the I–270 Modeling Calibration Methodologies Memorandum (2015).

Figure **202** illustrates a portion of these outcomes. Although these outcomes were given, the chosen methodology that led to these outcomes was not described in the I–270 report (2015). When possible, a simulation report should document the flow balancing methodology that was used.

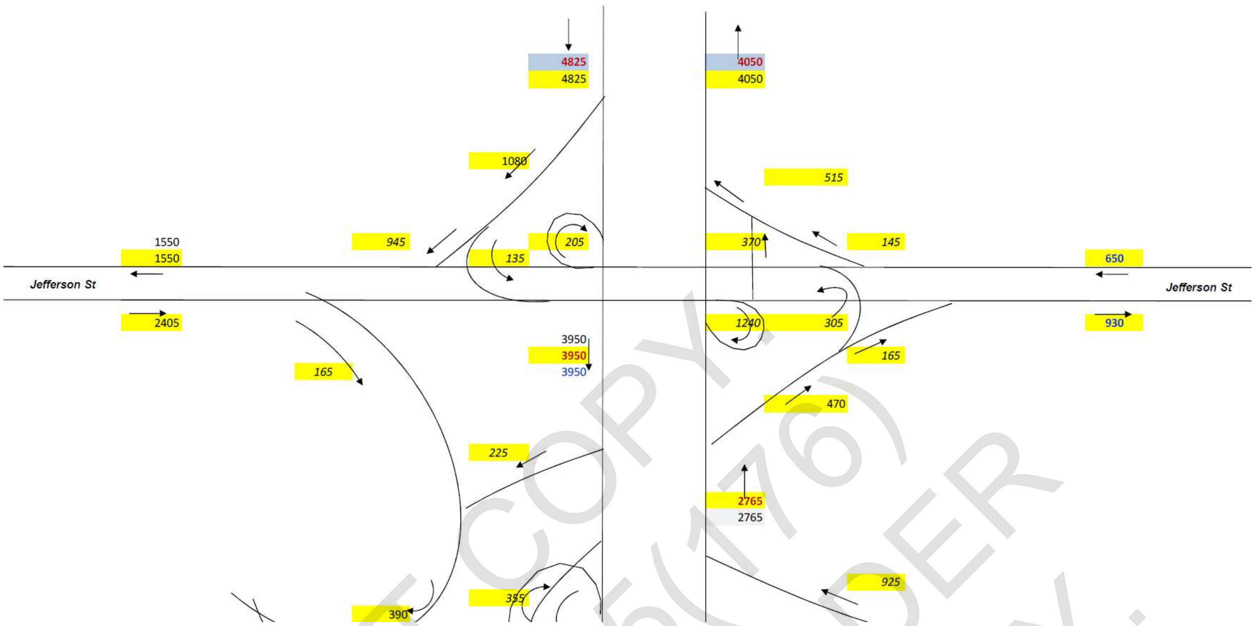


Figure 202. Illustration. Flow balancing outcomes in the I–270 simulation study.

Entry of Detector (Sensor) Data

Detectors are used on both freeways and arterials. Passage and presence detectors are used to control the traffic signals and/or ramp meter signals. Surveillance detectors can be used to collect additional traffic performance measures. Although the North Carolina case study report (2014) mentioned use of field sensors to obtain simulation input data, it made no mention of any surveillance detectors or ramp meter detectors in the simulated network. The I–35 KS case study report (2018) provided the following discussion of sensors used within the microsimulation model:

> *We developed a dynamic signal control algorithm that could be applied in an integrated ramp and mainline metering control strategy. The proposed dynamic signal control algorithm includes three main modules. In the first module, the traffic flow and speed are updated by the deployed downstream traffic sensors at current time. This information is updated at every time point, and thus the proposed dynamic signal control algorithm runs iteratively at the same discrete time points. Given that, the second module activates the dynamic signal controller provided that a certain traffic speed drop is detected at all lanes of the merge point. This module aims to stop the mainline traffic from entering the downstream queue until it is about to dissipate. Based on the information provided by the traffic sensors, the third module predicts the downstream traffic queue dissipation time with a reliable and efficient real-time algorithm (Ghiasi, Ma, et al. 2017), which is then*

used by the second module to let the upstream traffic smoothly merge into the downstream queue. For more detailed information about the developed dynamic signal control algorithm, please refer to (Ghiasi, Hale, et al. 2018).

Six sets of loop detectors are deployed in the simulation model in this area

Figure 203*, where each set contains several loop detectors embedded next to one another to cover all lanes at a specific longitudinal location. Thus, each loop detector set represents a traffic sensor in our analyses. These traffic sensors are evenly distributed along the merge section with the distance of 69 meters from one to another.*

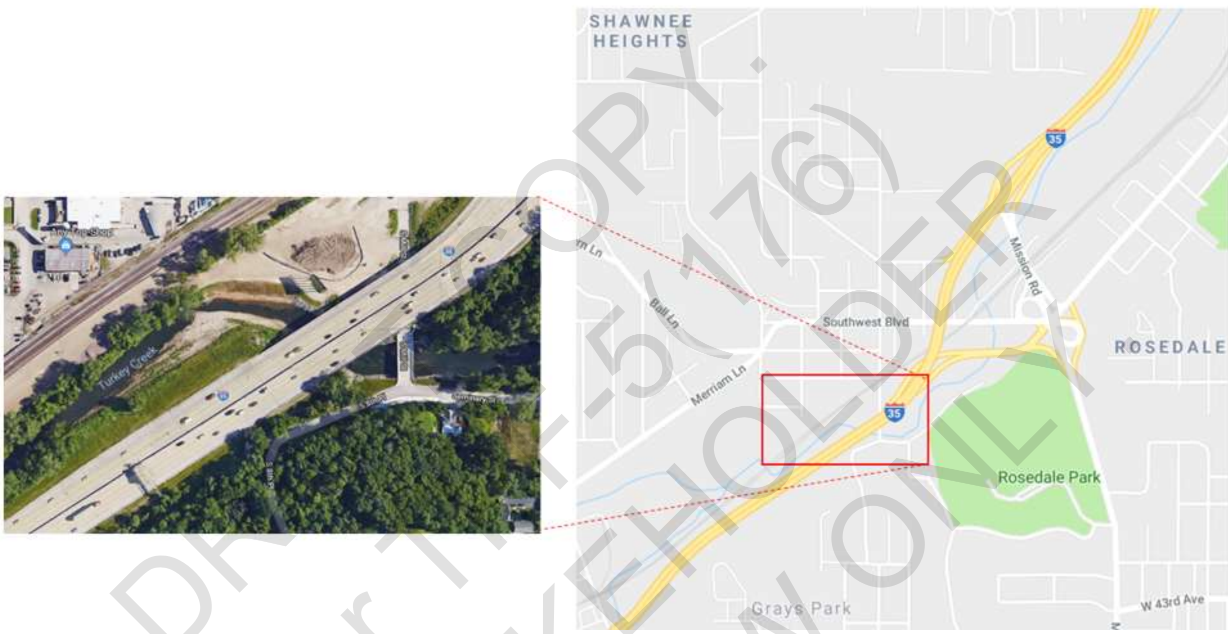


Figure 203. Map. Case study map for dynamic signal control: I–35 Kansas City, Kansas.

Entry of Freeway Warning Sign Distances

In traffic simulations of freeways, warning sign distances and/or reaction times may be specified for each off-ramp. The freeway case study reports described in this document did not discuss how warning sign distances or reaction times were simulated in the respective studies. Ideally, this information should be included in the simulation reports.

Entry of Operating Condition Data

Traditional simulation projects have focused on analyzing typical peak hour traffic conditions. The danger of this approach is that when decisions are made based on most likely (i.e., 50th percentile) outcomes, these decisions may fail badly when the inevitable 75th–95th percentile conditions materialize. With this in mind, some traditional projects have focused on analyzing the 30th-highest-hour demand volume scenario. While this is probably a step in the right direction, the analysis of only a single scenario still leaves engineers largely in the dark about how to manage risk and reliability because they cannot visualize the full range of possible outcomes. Different combinations of demand, weather, incidents, work zones, special events, and driver behaviors ultimately produce a wide range of outcomes. These outcomes will only be captured by simulation if a variety of input scenarios are considered.

The North Carolina study (2014) provided modeling of incidents and incident impacts. Incident scenarios were analyzed as shown in the previous **Error! Reference source not found.**. This analysis eventually found that the magnitude of the incident impacts is severely higher in work zone scenarios with only two lanes open instead of three lanes open. Additional information on the data entry process was given as follows:

> *The Capacity Adjustment Factors (CAF) for incidents in work zones are calculated using HCM 2010 suggested values in Exhibit 10-17. It is assumed that the capacity reduction due to incidents and work zones are independent, therefore the two adjustment factors were simply multiplied. For example, the suggested CAF for a freeway segment, which has two lanes and where the number of lanes is reduced to one due to an incident is 0.35. Please note that this adjustment factor accounts for two lanes and to use this value in FREEVAL computational engine, we must convert it to a lane by lane CAF, which is 0.70. On the other hand the suggested work zone CAF is 0.63 for such segment. The combinatorial CAF is calculated by multiplying the two values, which is 0.70*0.63 = 0.44. Route D incidents (I6 and I7) occur outside the work zone construction site and its CAF is 0.74, which just accounts for a capacity reduction due to an incident. Figure 13.15 depicts the incident locations on the map of the work area.*

Figure 204. Map. Incident locations in the North Carolina simulations.

Selection of Key Input Parameters for Calibration

In the North Carolina simulation study (2014), the researchers selected demand volumes, O-D flows, and link free-flow speeds as key inputs for calibration. They also calibrated the fundamental relationship between speeds and flows within work zone areas as follows:

> *With the primary objective of this project being the evaluation of work zone impacts, the team worked carefully to develop and calibrate customized work zone traffic flow models. These models are based on speed-flow model theory in the Highway Capacity Manual and calibrated to match work zone capacity estimates in North Carolina from prior research.*

The Maryland case study report (2015) listed the following key inputs for calibration:

- Traffic volume inputs and routing decisions.
- Traffic signal timings.
- Addition of turning speed reduction zones.
- Driver and link behavior types.
- Lane change distances.

In the I–35 KS case study report (2018), the researchers said that the model was calibrated by changing driver behavior parameters in the links and connectors. The I–66 VA report (2018) showed numerous outputs matching well between the simulation and field measurements, but did not explicitly say which inputs were calibrated. The lesson here is that the selection of inputs to calibrate is important, and should be documented along with the final calibration outcomes.

Identification of Key Network Locations for Calibration

For the North Carolina study (2014), the key point locations and routes for calibration were shown earlier in Figure 172 and Figure 173, respectively. For the Maryland study (2015), the calibration effort apparently considered all mainline links, although a special tabulation was given for the I–270 Spur section (see Table 33).

Table 33. Mainline link calibration results for I–270.

AM Peak Speed Comparison - General Purpose Lanes							
I-270 Northbound	**Field (mph)**	**VISSIM (mph)**	**Difference (mph)**	**I-270 Southbound**	**Field (mph)**	**VISSIM (mph)**	**Difference (mph)**
From I-495 interchange				From I-70			
to MD 187	60.7	60.5	-0.2	to MD 85	62.1	61.7	-0.4
to I-270 Split	60.5	56.7	-3.8	to MD 80	39.8	46.5	6.7
to Montrose Rd	64.9	63.0	-1.9	to MD 109	33.6	34.3	0.7
to MD 189	65.6	63.3	-2.3	to MD 121	46.3	47.7	1.4
to MD 28	66.1	62.9	-3.2	to MD 27	27.3	32.9	5.6
to Shady Grove Rd	65.5	63.0	-2.5	to MD 118	21.4	16.5	-4.9
to I-370	67.0	64.1	-2.9	to Middlebrook Rd	22.3	20.3	-2.0
to MD 117	66.6	63.8	-2.8	to MD 124	17.9	16.5	-1.4
to MD 124	66.8	63.9	-2.9	to MD 117	33.1	28.8	-4.3
to Middlebrook Rd	65.6	63.6	-2.0	to I-370	22.9	34.2	11.3
to MD 118	64.0	62.3	-1.7	to Shady Grove Rd	13.0	27.0	14.0
to MD 27	64.9	63.6	-1.3	to MD 28	12.3	13.5	1.2
to MD 121	65.0	63.7	-1.3	to MD 189	27.2	13.5	-13.7
to MD 109	67.6	62.6	-5.0	to Montrose Rd	26.3	12.3	-14.0
to MD 80	66.6	61.9	-4.7	to I-270 Split	38.8	25.4	-13.4
to MD 85	66.6	61.2	-5.4	to MD 187	54.1	52.3	-1.8
to I-70	63.1	62.7	-0.4	to I-495 interchange	55.9	51.7	-4.2
I-270 Total Speed (mph)	**65.5**	**62.4**	**-3.1**	**I-270 Total Speed (mph)**	**28.4**	**28.3**	**-0.1**
I-270 Total Percent Difference			**-4.7%**	**I-270 Total Percent Difference**			**-0.4%**
I-270 Spur Northbound				**I-270 Spur Southbound**			
From Cabin John Pkwy				From I-70			
to MD 190	62.6	60.3	-2.3	to I-270 Split	27.4	26.6	-0.8
to I-495	63.7	61.2	-2.5	to Democracy Blvd	30.5	29.8	-0.7
to Democracy Blvd	63.2	56.6	-6.6	to I-495	35.7	25.8	-9.9
to I-270 Split	64.4	62.9	-1.5	to MD 190	58.4	48.9	-9.5
to I-70	65.9	62.7	-3.2	to Cabin John Pkwy	58.9	58.6	-0.3
I-270 Spur Total Speed (mph)	**65.8**	**62.3**	**-3.5**	**I-270 Spur Total Speed (mph)**	**28.5**	**28.0**	**-0.5**
I-270 Spur Total Percent Difference			**-5.3%**	**I-270 Spur Total Percent Difference**			**-1.8%**

Source: Whitman, Requardt & Associates LLP.

Similarly, the I–35 KS calibration effort apparently considered all mainline links, as shown in the earlier Figure 192. In the I–66 VA study (2018), RTMS trailers for calibration were placed at key interchange locations, as shown in Figure 205.

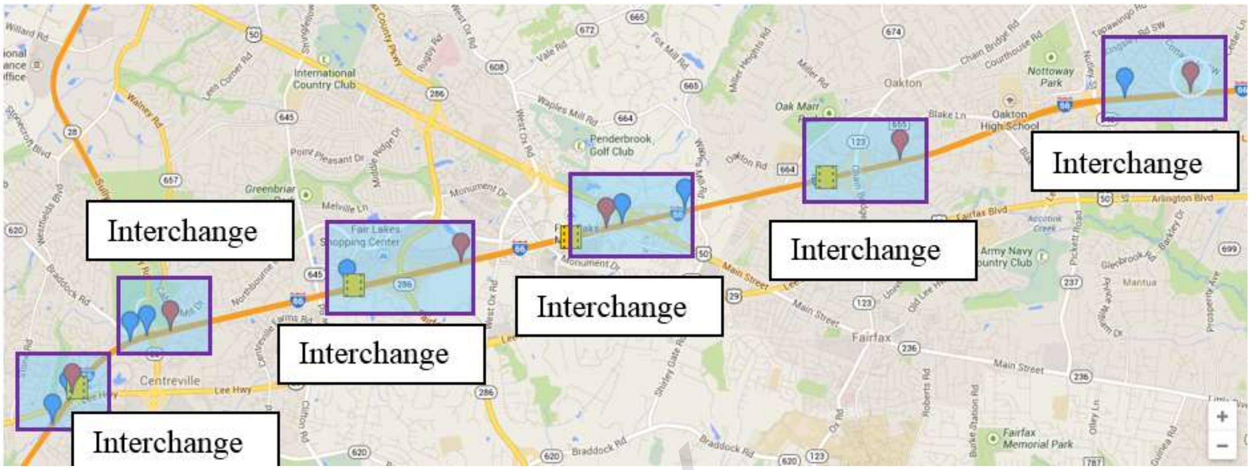


Figure 205. Map. Key interchange locations along I–66 VA.

12.4 VERIFICATION, CALIBRATION, AND VALIDATION

Analysis of Why Results are Incorrect before Deciding Next Step

As stated previously, the North Carolina study (2014) involved a mesoscopic dynamic traffic assignment (DTA) model for examining regional impacts, and a macroscopic simulation model for examining localized work zone impacts. When analyzing the initial DTA results, the analysts found very high work zone diversion rates, and dispersion of traffic across very minor links in the network. Because this was viewed as unrealistic, the team decided to change the bounded rationality rule, such that drivers would only divert if they could save 10 minutes or 33 percent of travel time.

For the macroscopic model, the team took a different approach because the model did not have traffic assignment capability. So instead of calibrating the parameters associated with traffic diversion rates, the team instead performed a sensitivity analysis of the demand adjustment factor (DAF), which would reflect a proportion of diverted vehicles as follows:

- 0 percent diversion, DAF = 1.0.
- 5 percent diversion, DAF = 0.95.
- 10 percent diversion, DAF = 0.90.
- 20 percent diversion, DAF = 0.80.
- 30 percent diversion, DAF = 0.70.
- 40 percent diversion, DAF = 0.60.
- 50 percent diversion, DAF = 0.50.

Results are shown in Table 34 and described in the report as follows:

> *As expected, by shifting diverting traffic from the subject facility, freeway traffic operations improved significantly, although still not to acceptable levels for the very severe work zone scenarios. The WZ2 travel times of 320 minutes without diversion, is reduced to 253.5 minutes with 20% diversion and 117 minutes with 50% diversion. The*

latter diversion is likely overly optimistic, and 20% is considered a reasonable assumption. With 20% diversion, the queue further still spills back beyond the bounds of the facility, and at least 30% diversion is necessary to contain the queue to 15 miles and within the modeled facility.

The DAF was thus calibrated to a value of 0.80 via a sensitivity analysis, which indicated that the other DAF options produced overly optimistic and pessimistic outcomes.

Table 34. Sensitivity analysis to diversion rates on work zone traffic impact.

Scenario	DAF	Travel Time (Minutes)	Max DEQL (Miles)	Max Queue Length (Miles)	Max d/c Ratio
Base (No WZ)	1.0	27.4	0.0	0.0	1.2
WZ2	0.50	117.2	0.0	9.8	1.7
	0.60	171.3	0.0	13.1	2.0
	0.70	216.3	0.0	15.0	2.4
	0.80	253.4	4.3	16.2	2.7
	0.90	279.2	16.4	16.2	3.0
	0.95	303.7	11.8	16.2	3.2
	1.00	320.0	25.1	16.2	3.4
WZ2b	0.50	50.8	5.7	8.0	1.4
	0.60	89.2	5.7	13.5	1.7
	0.70	138.2	5.7	15.8	2.0
	0.80	176.8	8.3	16.2	2.2
	0.90	200.1	16.0	16.2	2.5
	0.95	209.2	18.7	16.2	2.7
	1.00	214.2	20.9	16.2	2.8
WZ4	0.50	35.4	10.7	3.0	1.2
	0.60	52.8	10.7	5.9	1.5
	0.70	75.9	10.7	10.9	1.7
	0.80	105.6	10.7	11.3	2.0
	0.90	116.7	19.0	11.6	2.2
	0.95	115.8	21.0	11.6	2.3
	1.00	122.7	30.7	11.6	2.4
WZ6	0.50	29.0	0.0	0.1	1.1
	0.60	34.0	0.0	2.7	1.3
	0.70	47.6	0.0	6.0	1.5
	0.80	55.1	0.0	6.9	1.8
	0.90	62.7	0.0	8.1	2.0
	0.95	72.4	0.0	9.8	2.1
	1.00	77.8	0.0	10.4	2.3
WZ7	0.50	33.8	0.0	1.0	1.2
	0.60	43.2	0.0	1.7	1.5
	0.70	55.6	0.0	4.6	1.7
	0.80	73.5	8.8	9.1	1.9
	0.90	93.2	8.8	10.3	2.2
	0.95	100.9	14.5	10.3	2.3
	1.00	106.5	19.2	10.3	2.4

Source: Institute for Transportation Research and Education.

In the Maryland study (2015), the following calibration issues and solutions were identified:

- Modified lane-changing distances to ensure smooth yet realistic traffic flow in both peak and off-peak directions.
- Modified driver behavior parameters and link behavior types; driver and link behavior types from the provided files were maintained where possible.
- For the PM model, speed reduction zones were coded at outbound links of the network to replicate lower speeds caused by friction created outside of the network. These reduction zones are located on I–270 northbound north of I–70, I–270 southbound south of Maryland Route 187 (MD 187), and I–270 Spur westbound south of Clara Barton Parkway.
- Adjusted collector-distributor ramp volume demand based on updated counts provided by DSED or in relation to higher speeds as reflected by the travel time data primarily on the collector-distributor road during the PM peak hour.

The I–35 KS report said, "Through the simulation animation unrealistic driver behaviors were identified, and were resolved with the help of calibration" (2018). Ideally, a discussion of how and why driver behaviors were unrealistic should be included in the report.

Selection of Both Upper and Lower Level Simulation Quality Metrics

It was stated previously the North Carolina simulation models were calibrated using field-estimated spot volume and speed data, as well as key route travel times obtained from INRIX. These are the lower level simulation quality metrics (SQMs). The upper-level SQMs were simple percentages:

> *The calibrated DynusT model achieved an average percent volume error within 5% across all sensor stations in the network, with the exception of the first AM and last PM hours. Both of those periods showed higher volume estimates in the model than the field data. Point-based speed estimates of the DynusT model yielded an overall error between 5% and 9% on freeway stations (no arterial speed data was available) and an error of 6 to 11% along sensor stations in the work zone. In all cases, the speed prediction error was positive, indicating roughly 10% higher speeds (less congestion) in the model compared to the field data.*
>
> *For DTALite the AM peak period network was calibrated to an average error within 2% of sensor volumes for the last three hours of the peak period, and 11% of sensor speeds. The average error was 21% for the first hour of the AM peak period. This higher error was deemed acceptable, as the primary performance statistics were extracted for the second hour of analysis, and the first hour was generally uncongested due to lower volumes. In the PM peak, all volumes on average were within 4% of field-measured sensor volumes in the first three hours of the peak period, and within 9% of speeds in the peak hour. The average error was 21% in the last hour of the PM peak period.*
>
> *The team further validated route travel times for several freeway and arterial routes. The route travel time error was less than 10-15% for most freeway routes, including the work zone. For arterial streets, the validation error was higher with some errors in the 20%-*

50% range. The arterial error is attributed to the form of the mesoscopic simulation model, which has traditionally been used for freeway analyses, including past research for NCDOT.

The Maryland case study report (2015) provided the following discussion of low-level (e.g., travel time) and high-level (e.g., confidence intervals) simulation quality metrics (SQM):

The travel time data along I-270 showed high variability between travel times, specifically in the AM peak hour. The travel times between 7:00 AM to 8:00 AM were significantly higher than the travel times between 8:00 AM to 9:00 AM; therefore only the travel time runs during the AM peak hour (7:00 AM to 8:00 AM) were used for validation even though these also showed high variability between runs. For example, in the AM peak southbound direction, travel times ranged from 193 to 362 seconds between I-370 and Shady Grove Road. This high variability results in large 95% confidence intervals and is shown on the travel time calibration graphs. The PM travel time runs were consistent throughout the entire PM peak period.

The travel time data along I-270 also showed lower speeds between MD 121 and the I-270 Spur when compared to the RITIS data. The travel time data was chosen for validation to conservatively capture the amount of delay observed along the corridor. The RITIS data was used for all other roadway segment validation, where travel time data was not collected.

The validation targets for the I-270 model were:

- *Overall Corridor Speed and Travel Time – within +/- 10%.*
- *Travel Time – within +/- 10% confidence bands, 95% confidence intervals, and +/- 10 MPH.*
- *Traffic Counts – within +/- 10%.*

These targets were chosen based on the complexity and length of the network and the variability of the field-measured travel time runs along certain travel time segments. Priority was given to corridor travel speeds since it will be an important measure of effectiveness when quantifying and evaluating improvement alternatives. Traffic count locations at the end of the corridor (southernmost in the AM and northernmost in the PM) were used as the primary locations in determining throughput for the corridor. Other count locations along the corridor were used as a secondary validation measure against the travel time data due to the dynamic conditions along the corridor.

The I–66 VA study (2018) used speed and throughput (volume) for low-level SQMs. Time series plots and scatterplots were both used to visually demonstrate a reasonable level of agreement between simulated and field-measured performance. However, it also used the R-squared correlation coefficient as an upper-level SQM, to demonstrate agreement between the simulated and observed O-D flows. This is illustrated in
Figure **206**.

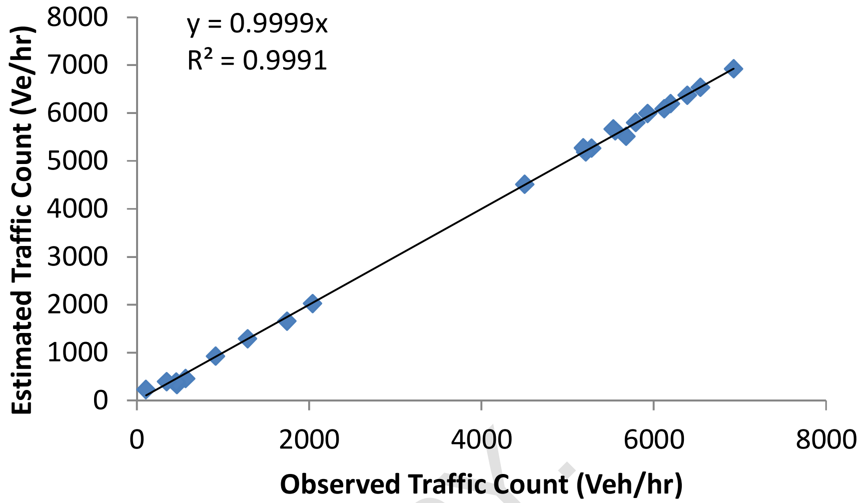


Figure 206. Chart. Comparison of estimated flow and observed flow.

The I–35 KS study (2018) similarly used speed and throughput for low-level SQMs. Heat maps and scatterplots were used to visually demonstrate a reasonable level of agreement between simulated and field-measured performance. Simple percentages were applied as the upper-level SQM: "The speed distribution from the simulation has to roughly match the speed distribution of real life data with an admissible error of 10%. At the locations where the calibration was successful, the speed difference between field and simulated data is within 10%."

Focusing on a Limited Number of Critical Origin-Destination Pairs and/or Key Problem Spots

The North Carolina calibrations focused on critical locations (shown previously in **Error! Reference source not found.**) as follows:

- *Point-based traffic volumes* at key locations in the network.
- *Point-based speed* estimates at key locations.
- *Route-based travel time and speed* estimates along critical network routes.

The Maryland study (2015) noted highly variable southbound travel times (ranging from 193 to 362 seconds), between I–370 and Shady Grove Road, in the AM peak direction. In the I–66 VA study (2018), RTMS trailers for calibration were placed at key interchange locations, as shown previously in
Figure 205.

Although three out of four case studies noted the general calibration results for O-D pairs, none of the freeway case studies here focused on a limited number of critical origin-destination pairs.

Evaluation of Probability Density Functions or Cumulative Density Functions That Describe Distributions of the Performance Metrics

Although all the freeway case study reports provided extensive graphs and tables to demonstrate their calibration results, none of them made use of probability density functions (PDF) or cumulative density functions (CDF). The Maryland case study report (2015) provides graphs that are somewhat analogous to the CDF, in the form of a cumulative travel time plots (e.g., Figure 207).

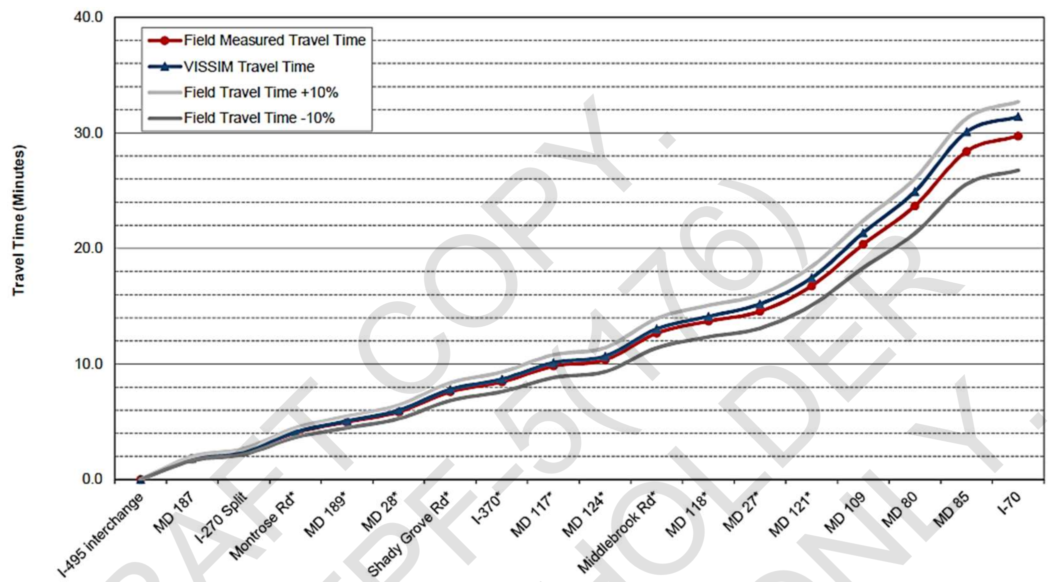


Figure 207. Chart. Cumulative travel time plot.

Ensuring That Links Perform in Accordance with the Fundamental Diagrams of Traffic Flow

In addition to replicating field-measured performance, a well-calibrated simulation should obey the fundamental laws of traffic.

Figure **208** shows the typical shapes for speed vs. density, speed vs. flow, and flow vs. density plots.

In the North Carolina study (2014), speed-flow curves and speed and volume profiles were obtained for all traffic.com sensors.
Figure 209 presents the speed-flow relationship for one of the westbound sensors.

The I–66 VA study (2018) produced speed-flow curves for each of the six RTMS locations. These are compound curves comparing the data points from simulation to data points obtained from the field.
Figure 210 illustrates the compound speed-flow curve for the first interchange. Similarly, Figure 211 illustrates a compound speed-flow curve for one of the I–35 KS locations.

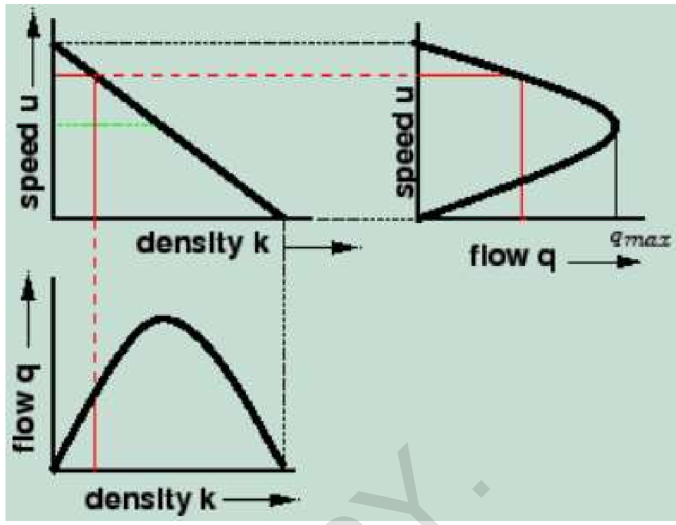


Figure 208. Diagram. Fundamental diagrams of traffic flow.[96]

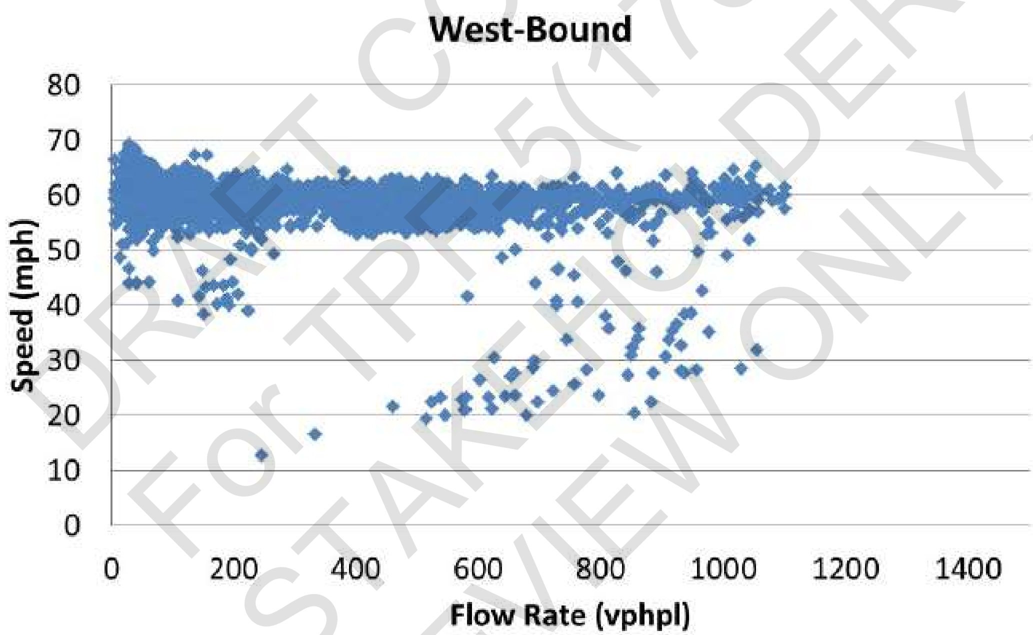
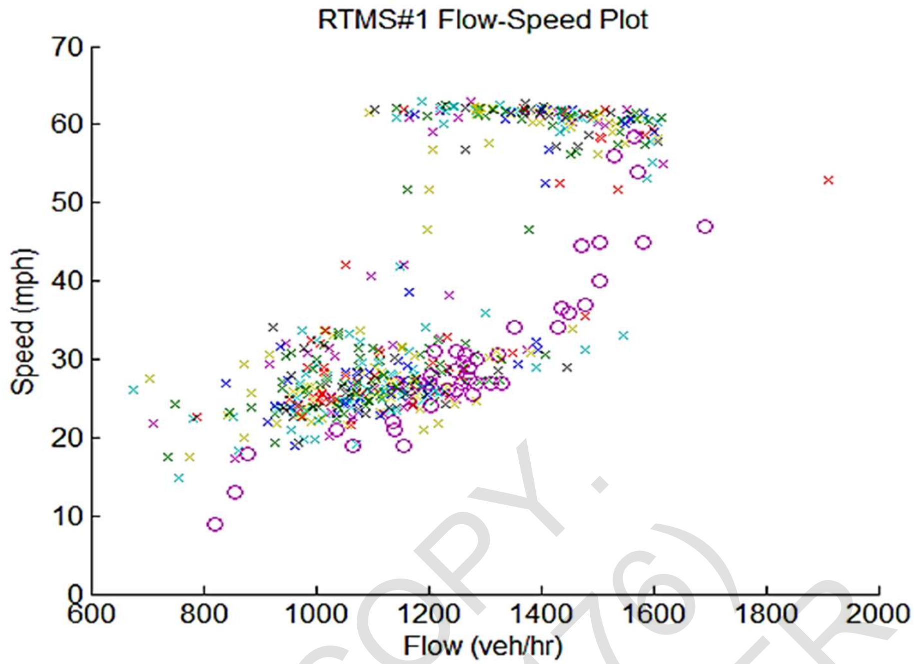


Figure 209. Chart. Speed-flow curve for sensor 040100 on I-40 westbound.

[96] "Fundamental Relations of Traffic Flow" (web page), SlidePlayer slide 21, accessed October 20, 2019, https://slideplayer.com/slide/12982685/.



(Circle: Field; x-mark: Simulation)

Figure 210. Chart. I–66 VA compound speed-flow curve depicting calibration results.

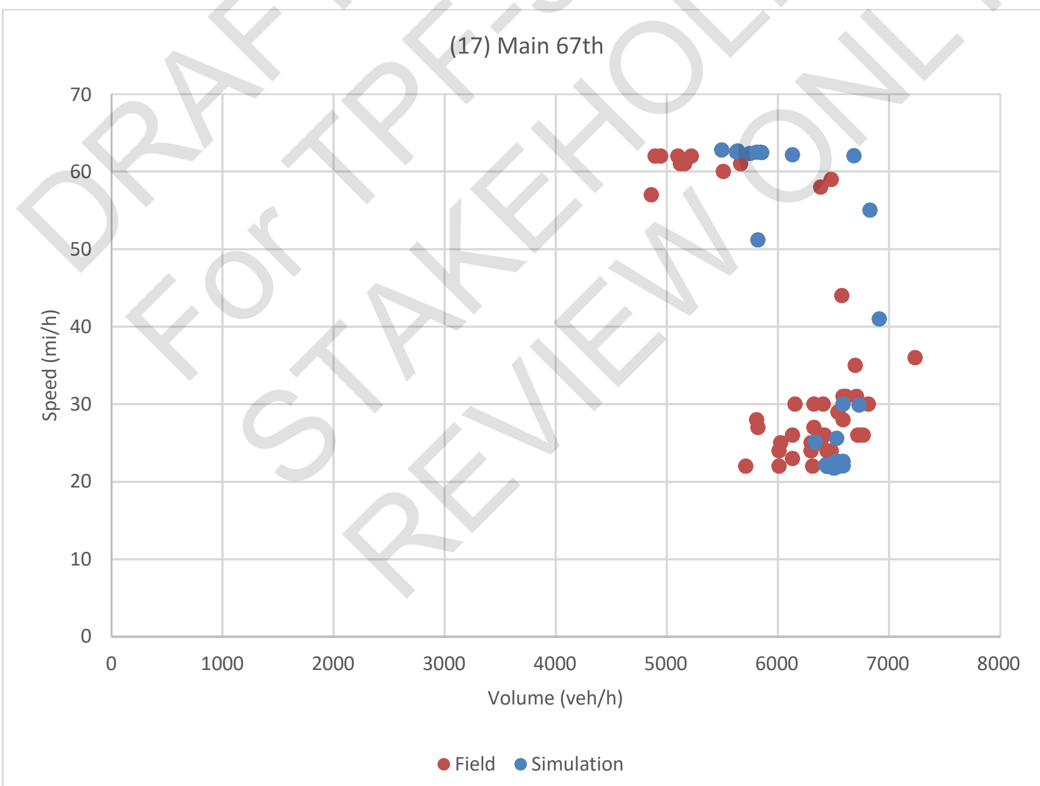


Figure 211. Chart. I–35 KS compound speed-flow curve depicting calibration results.

Analyzing Reasonableness of Vehicle Trajectories

Simulated vehicle trajectories can reveal unrealistic accelerations, decelerations, shockwaves, cruise speeds, lane-change behaviors, and so on. In some cases, microscopic simulation models producing seemingly accurate aggregate performance measures have been shown to be inaccurate at the trajectory level. The Maryland case study report (2015) provided probe vehicle trajectories and travel times along all key routes.
Figure 212 illustrates these trajectories and travel times along just one of the key routes.

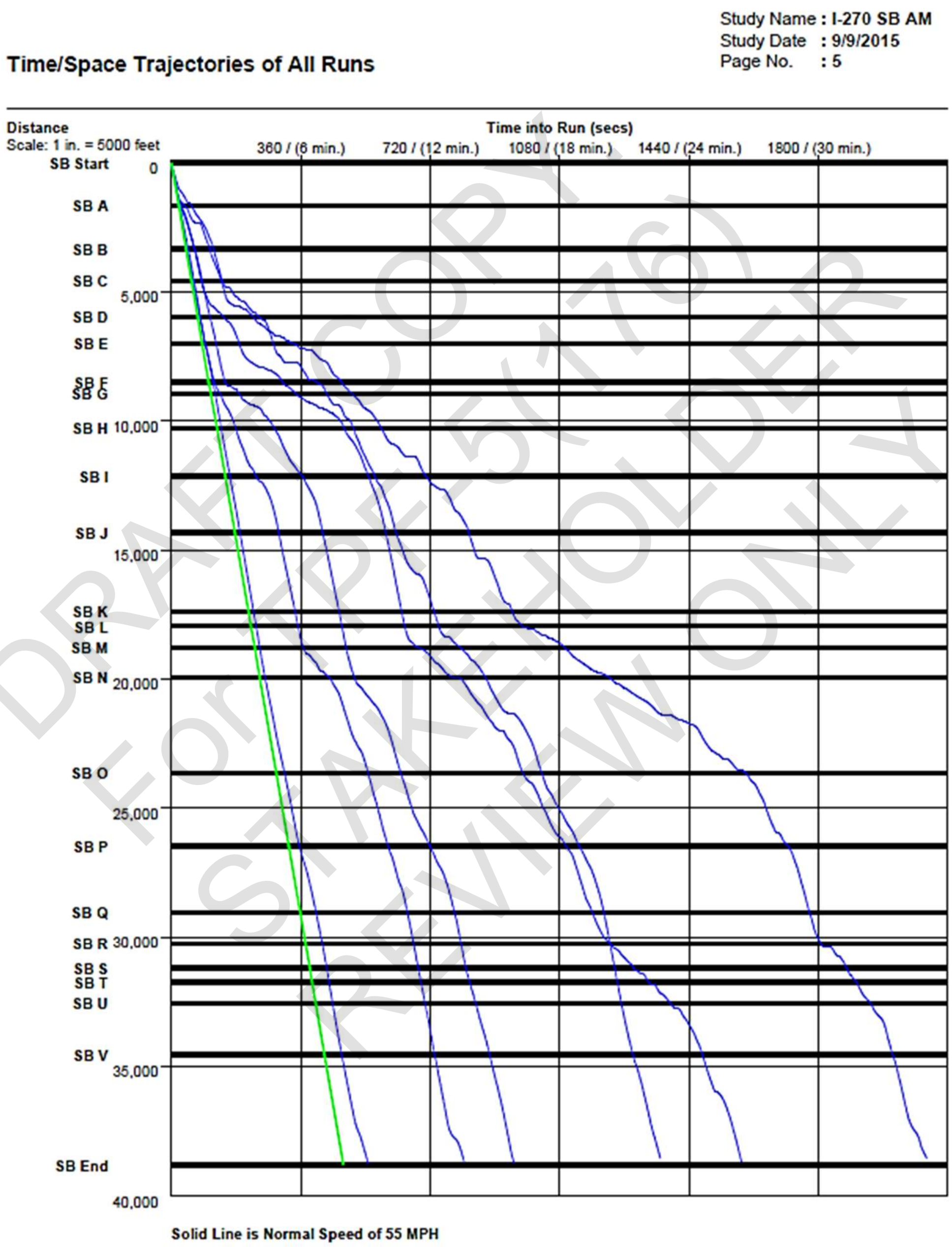


Figure 212. Chart. Trajectories and travel times along I–270 in Maryland.

Having a Second Set of Eyes to Review the Data

Stakeholders can provide a powerful mechanism for achieving quality control. Some simulation reports can help to build confidence in the model results by documenting their stakeholder meetings and interactions. The previous section, Analysis of Why Results are Incorrect before Deciding Next Step, talked about two dozen members from the Steering and Implementation Committee (North Carolina study), DSED Travel Forecasting and Analysis (Maryland study), and stakeholder group for the Alternative Designs project. These stakeholders provided critical reviews and suggestions for their respective studies.

Dealing with Episodic Demands

In recent years, transportation professionals have begun to realize the significance of annual performance and travel time reliability. This requires explicit analysis of operating conditions including varying traffic demands, weather, incidents, work zones, and special events. The TSSM authors recommend detailed analysis of these variabilities whenever possible; and when these conditions are analyzed, the methods of doing so should be well documented. Today's traffic modeling experts should become well-versed in the simulation-based travel time reliability analysis methods described in the SHRP2 L04 report (Mahmassani 2014), as well as the methods described in Traffic Analysis Toolbox: Volume III (FHWA year).

The North Carolina analysts recognized the importance of episodic demands when updating the demand profiles for both AM and PM periods. O-D matrices were obtained from the TRM, which is a 4-hour peak period model. As a result, the team needed to estimate hourly factors from field sensors, to split the 4-hour demand into 4 separate hourly O/D matrices with the appropriate peaking characteristics. Traffic performance is highly sensitive to these peaking characteristics. Although demands exceed volumes during oversaturated time periods, Figure 213illustrates a volume profile at one of the sensor locations. Figure 214 illustrates a similar volume profile at one of the I–66 RTMS locations. Heat maps from the I–35 KS case study report (2018) illustrate the peaking characteristics along that freeway. Those heat maps will be presented in the next case study section on Data Processing, Analysis, and Presentation.

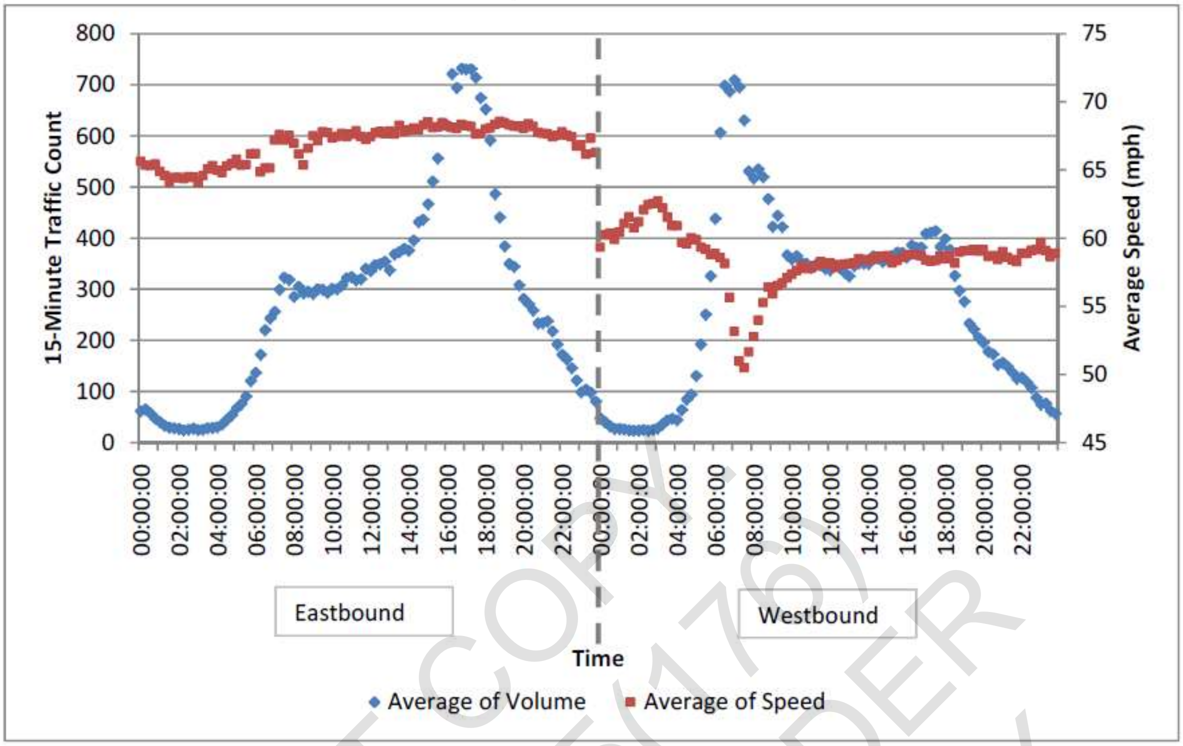


Figure 213. Chart. Speed-volume profile for sensor 040100 on I–40.

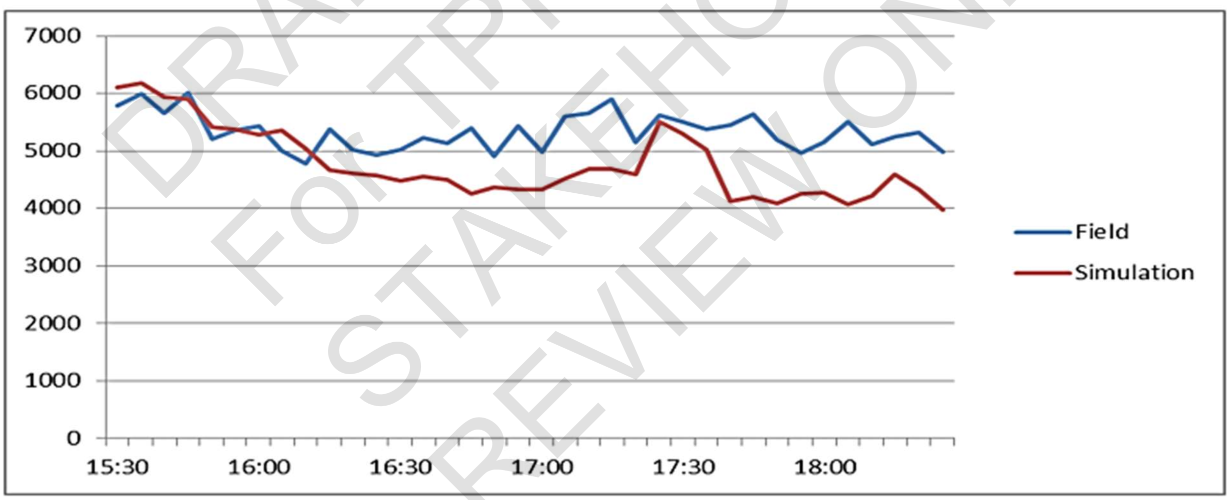


Figure 214. Chart. Volume profile for Interchange #1 on I−66.

Selection of a Preferred Calibration Sequence

According to TSSM Chapter 8. , input parameters can be calibrated simultaneously or sequentially, with sequentially the recommended approach. One example of a recommended sequence was: (1) driving behavior model parameters, (2) route choice model parameters, (3) the traffic demands, and (4) overall model fine-tuning. Although most calibration efforts follow some sort of sequence, the chosen sequence is often not fully documented as it should be. The best documentation of calibration sequence among the freeway case studies was given in the North Carolina report (2014). For this study, the calibration process is presented in Figure 215 below. Overall, the process included eleven calibration steps.

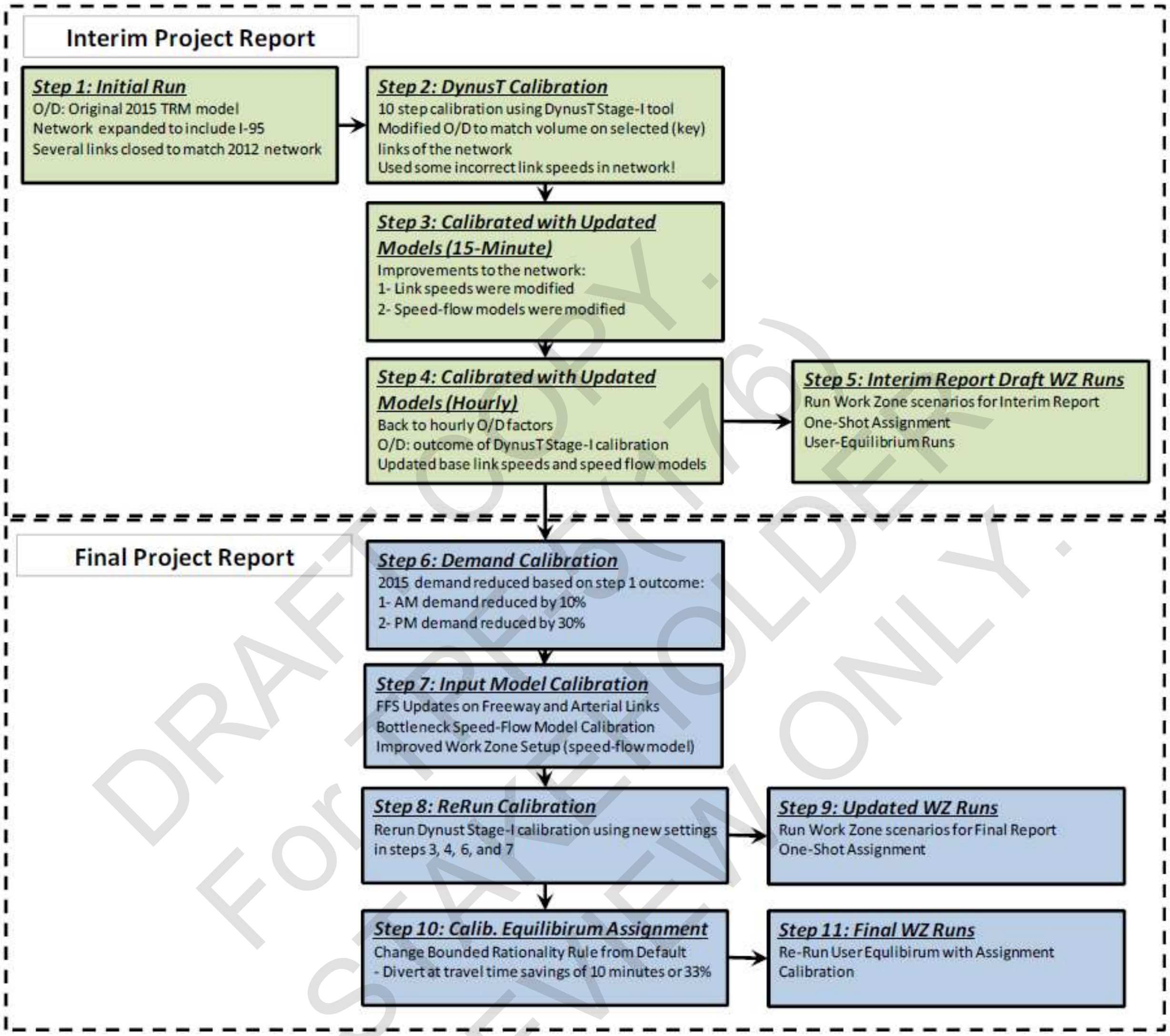


Figure 215. Diagram. Calibration flow chart.

Validate with Separate Data Sets

TSSM Chapter 8. says, "The validation step is extremely important. If the calibrated model is only checked against the calibration data sets, the validation assessment is circular. In this case, the same data used to calibrate the model are also used to validate its predictions. A much better plan is to use separate data sets to see if the calibrated model can produce defensible results for conditions that were not used for calibration. If it can, then it has been calibrated. It has been trained to provide useful predictions. If not, it should be calibrated further so that it has a predictive capability that is more robust."

The North Carolina case study reports said the base models were validated with empirical speed and travel time data for various key routes in the triangle region. In the Maryland study (2015), only the AM travel times were used for validation purposes, because these travel times were significantly larger than the PM travel times. The I–66 and I–35 calibration reports did not specifically mention validation. The North Carolina case study report (2014) and Maryland case study report (2015) did not specifically mention validation by using separate data sets.

Review Best-Case and Worst-Case Scenario Outcomes

The North Carolina study (2014) extensively examined worst-case scenario outcomes:

> *The one-shot modeling approach assumes that the origin-destination (O/D) matrix and the path assignment from the calibrated base network are unchanged when the work zone is put in place. The one-shot results therefore assume no diversion due to the work zone, with all traffic maintaining its optimum path from the base network. The one-shot results conceptually represent the worst-case conditions for the I-40 work zone facility, while representing low expected impacts on diversion routes and the surrounding network.*
>
> *For all scenarios, the analysis distinguishes between a no-diversion and with-diversion scenario. The no-diversion case corresponds to a "one-shot" assignment, where all vehicles remain on their baseline paths despite the presence of the work zone. Conceptually, this corresponds to a worst-case analysis of that work zone scenario where drivers are not aware of the presence of the work zone. The with-diversion results were obtained from running mesoscopic dynamic traffic assignment (DTA) modules to (near) user-equilibrium, by repeating the network assignment twenty times. After each assignment, each simulated vehicle in the network reconsiders its routes based on its experience in the previous run. After 20 iterations, the model therefore results in a modified traffic assignment, where each vehicle minimizes its origin-to-destination travel time, under the constraints imposed by the work zone. Conceptually, this corresponds to a user equilibrium solution, under the assumption of fixed demand.*
>
> *Figure* 216 *depicts that the magnitude of the incident impacts is severely higher in work zone scenarios with two lanes open. The incident scenario 3 is the worst scenario among all the incident scenarios. The analysis shows another advantage of maintaining three lanes open over just keeping two lanes open. There is no work zone scenario which maintains three lanes open during construction in route D.*

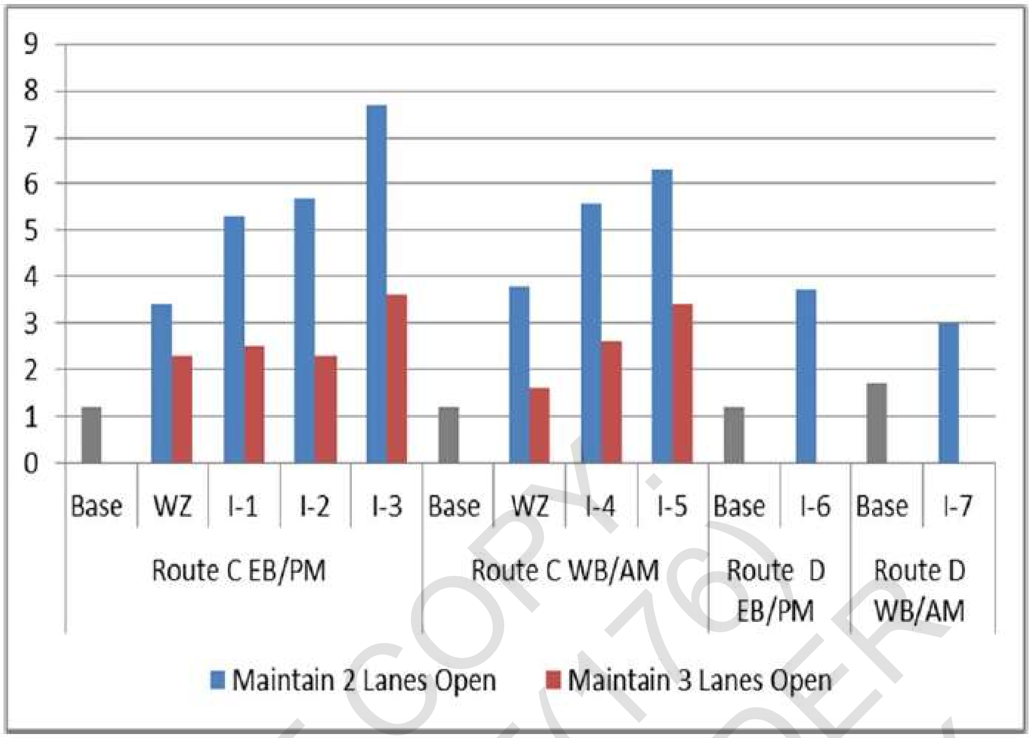


Figure 216. Chart. Capacity ratios for incident scenarios compared to base case and work zone scenarios.

For the no-diversion work zone scenarios, the average network travel time increased from 9.7% to 21.6%. This is a significant increase in average travel time in the network; however, it should be noted that no-diversion scenarios represent worst-case scenarios that are highly unlikely to occur in real-world conditions. However, they could also be viewed as a second baseline by which to gauge the ability of the existing diversion routes to carry the diverted traffic load. Scenarios with fewer open lanes (scenarios WZ-1 and WZ-2) yielded longer average delay than those with more open lanes (scenarios WZ-3, WZ-5, and WZ-75).

The increase in network-wide average travel time ranged from 4.8% to 16.4% for the various no-diversion scenarios. Again these scenarios represent the worst-case conditions and are very unlikely to happen in the real-world. In the diversion scenarios, the average travel time increased only by 2.8% to 3.6%, considerably lower than the no-diversion scenarios travel times. In the PM peak period, all with-diversion work zone scenarios resulted in approximately similar average network travel time, average travel distance, average stop time, and number of unserved vehicles.

In the I–66 VA study (2018), initial calibration was performed to narrow down parameter set candidates by using Latin hypercube sampling design (LHD) approach. A total of 500 scenarios created by LHD were evaluated by VISSIM with five replications for each scenario to choose the best candidate scenario.

12.5 DATA PROCESSING, ANALYSIS, AND PRESENTATION

Using a Naming Structure for the Scenario Data Sets

TSSM **Error! Reference source not found.**recognizes that scenario analysis is one of the most valuable capabilities of traffic simulation models. For a project with numerous alternatives, the number of input data sets can become onerous to manage. Creating a condensed but readable naming structure is critical. **Error! Reference source not found.**recommends a naming structure that reflects an input that would vary from model to model. For example, given three geometric alternatives and two time periods, the naming structure might look like *ProjectRef_Alternative#_TimePeriod_Version*, with one data set named *50476_Alt1_AM_V1*.

Although highly recommended, the data set naming convention is a low-level detail that is often not documented in the final project reports. Nonetheless, for the benefit of future engineers who may review or build on the initial work, an appendix to the simulation report could be used to document the chosen data set naming convention.

Automated Post-Processing Analysis

Traffic simulation models are notorious for producing exhaustive amounts of numerical output data, especially microsimulation models requiring 10 or more Monte Carlo realizations per scenario. Post-processor tools are frequently needed to analyze the summation, mean, median, maximum, percentile, and variance of these results across numerous links, nodes, sensors, and data stations. In some cases, these may be third-party post-processing tools distributed or sold separately from the native simulation tool. In other cases, there may be a post-processing module included within the simulation package, or the simulator may seamlessly provide post-processed statistics without any mention of a separate tool/module.

Simulation project reports do not always disclose names of the tools used to obtain their post-processed statistics. Engineers have post-processed and visualized their outputs through spreadsheet software (e.g., Microsoft® Excel) and statistics software (e.g., SAS, R). Specialized post-processing tools have been customized for specific simulation products.

Although the North Carolina case study report (2014) provided ample graphics and figures to visualize outcomes, it is not clear that post-processing tools were needed to extract outputs from the macroscopic and mesoscopic simulation tools that were used. Although the Maryland case study report (2015) was based on microsimulation, no discussion of random number seeds was included, so it is not clear that post-processing tools were used. However, a mild amount of post-processing was performed to produce the LOS analysis shown in Table 35. The bottom of the table shows the density range thresholds used to obtain LOS for each freeway segment. Thus, the post-processing of dozens of link densities (to obtain dozens of LOS outcomes) may have been performed manually, or may have been performed by an automated process.

Regarding the I–35 KS study, generation of a heat map based on simulated speeds was accomplished by one or more automated post-processing applications. Each cell of the heat map was based on an average speed, which was averaged over several random number seed realizations. This simulated speed profile (see Figure 217) was ultimately compared to the KC Scout speed profile (shown previously in Figure 198), to assess the degree of calibration accuracy.

Table 35. Morning peak period corridor level of service.

AM									
I-270 Northbound	Field Density (pc/mi/ln)	Field LOS	VISSIM Density (pc/mi/ln)	VISSIM LOS	I-270 Southbound	Field Density (pc/mi/ln)	Field LOS	VISSIM Density (pc/mi/ln)	VISSIM LOS
From I-495 interchange					From I-70				
to MD 187	24.7	C	24.6	C	to MD 85	26.4	D	26.3	D
to I-270 Split	17.1	B	17.9	B	to MD 80*	40.6	E	34.7	D
to Montrose Rd	18.1	C	18.5	C	to MD 109*	57.9	F	56.4	F
to MD 189	16.7	B	17.4	B	to MD 121	46.2	F	44.6	E
to MD 28	16.5	B	17.4	B	to MD 27*	56.6	F	47.6	F
to Shady Grove Rd	12.9	B	13.4	B	to MD 118*	72.4	F	94.4	F
to I-370*	10.7	A	11.0	B	to Middlebrook Rd	74.3	F	82.6	F
to MD 117	11.8	B	12.3	B	to MD 124	94.0	F	101.9	F
to MD 124	12.3	B	12.6	B	to MD 117	52.4	F	60.2	F
to Middlebrook Rd	15.5	B	15.9	B	to I-370*	123.2	F	82.6	F
to MD 118	13.0	B	13.6	B	to Shady Grove Rd*	105.6	F	50.8	F
to MD 27	14.5	B	14.8	B	to MD 28**	97.1	F	87.9	F
to MD 121	12.9	B	13.2	B	to MD 189*	48.7	F	97.3	F
to MD 109	17.9	B	19.4	C	to Montrose Rd*	50.4	F	106.6	F
to MD 80	17.7	B	19.1	C	to I-270 Split	42.1	E	63.8	F
to MD 85	19.9	C	21.7	C	to MD 187	24.2	C	24.6	C
to I-70	22.6	C	22.6	C	to I-495 interchange	27.6	D	29.7	D

HCM 2010 Freeway LOS (pc/mi/ln)	
<11	A
> 11-18	B
> 19-26	C
> 26-35	D
> 35-45	E
>45	F

Source: Whitman, Requardt & Associates LLP.

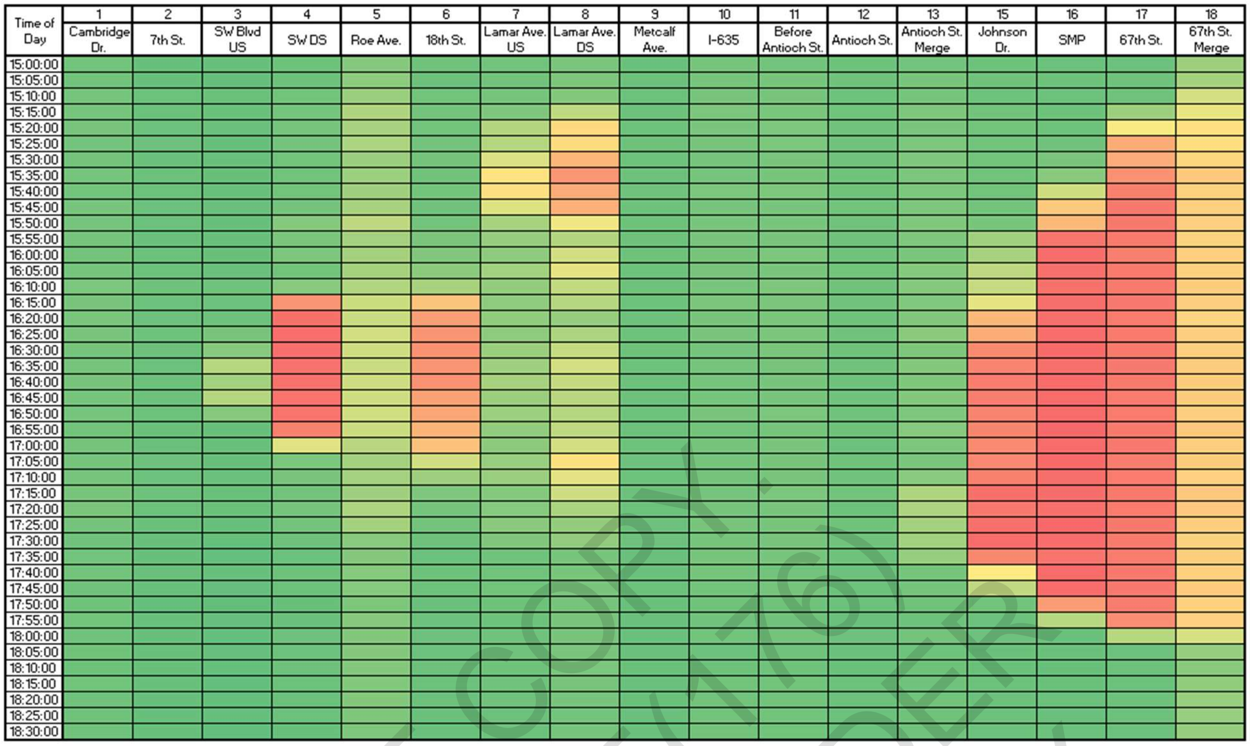


Figure 217. Illustration. Calibrated microsimulation model speed profile.

Weighted Aggregation of Performance Measures

Applying weighting factors to the performance measures at various network locations can help to convey the right information. Performance measures can be weighted by demand, throughput, time, distance, and other metrics. This can help to show that the bottom-line traffic performance was most heavily influenced by:

- Links or nodes serving the greatest number of trips (i.e., throughput-weighted average).
- Conditions prevalent for the longest time period(s) (i.e., time-weighted average).
- Links or segments covering the greatest distance (i.e., distance-weighted average).

Intersection LOS is one example performance measure typically obtained by weighted averages, in this case a throughput-weighted average frequently referred to as a volume-weighted average. The I–66 VA case study report (2018) provided a table (Table 36) showing weighted total throughputs of two designs for selected scenarios.

Table 36. T-test results: total delay and throughputs on general purpose and HOV lanes.

	HOV (%)	V/C ratio	On-ramp volume	Throughputs: Estimate mean of right-left	Throughputs: P-value	Delay: Estimate mean of right-left	Delay: P-value
Total	0.25	0.8	1,000	171	**4.59e-11**	-151	**7.49e-11**
			1,500	149	**7.29e-10**	-101	**3.67e-09**
		1.1	1,000	165	**2.05e-11**	-77	**7.67e-13**
			1,500	144	**5.60e-14**	-102	**1.19e-13**

Analysis of Stochastic Model Outcomes under Different Random Number Seeds

Random number seeds are most often associated with microsimulation, although macroscopic and mesoscopic simulation tools can certainly incorporate random processes when the developer so chooses. According to conventional wisdom and several State departments of transportation (DOT) simulation guidelines, the number of Monte Carlo realizations should be at least 10 and should reflect important aspects of the study (e.g., congestion levels, network size). For example, an unstable freeway weaving section could warrant up to a hundred realizations to obtain sufficient confidence in the reported results.

Unsurprisingly, the North Carolina case study report (2014) contained no discussion of stochasticity or random number seeds in the macroscopic and mesoscopic simulation tools that were used. However, the Maryland case study report (2015) also contained no such discussion, despite employing microsimulation. The I–66 VA report (2018) authors described the following calibration approach based on five random number seed replications per scenario:

> *Initial calibration was performed to narrow down parameter set candidates by using a Latin Hypercube Sampling Design (LHD) approach. A total of 500 scenarios created by LHD were evaluated by VISSIM with 5 replications for each scenario to choose the best candidate scenario. The selected candidate was fine-tuned to obtain the final simulation model.*

When comparing geometric alternatives, the I–66 VA researchers performed 10 runs with 10 different random number seeds to account for stochasticity. The I–35 KS report authors also accounted for stochasticity in their calibration and validation approach, although the number of replications was not specified:

> *The calibration was conducted by changing the values of multiple parameters throughout the software. The calibration day needs to start in uncongested conditions and end in uncongested conditions to verify that the calibrated model is functioning properly. The vehicle routes decision were modeled. To compare the simulation model to the calibration day detectors were established in the VISSIM model in the same locations they exist in real life. After that, initial simulation running using 1 random seed was conducted to compare calibration day speed profile to the simulation model speed profile. Through the simulation animation unrealistic driver behaviors were identified and were resolved with the help of the calibration. Also, to achieve a speed profile like the calibration day, different sets of driver*

behavior parameters in links were developed and tested. The process of running one seed and adjusting the parameters was executed until a similar speed profile was achieved. Then the simulation model was run multiple times with different seeds and the average speed profile of those runs should be close to the calibration day speed profile.

When the I–35 KS network was later used to test the mainline metering scenario, 10 runs were used to examine stochasticity, as shown in

A. Subfigure of throughput.

B. **Subfigure of average speed.**

C. Subfigure of **average delay.**

D. Subfigure of carbon monoxide emissions.

Figure **223** later in this section.

Use of Probability Density Functions and Cumulative Density Functions

From the North Carolina study (2014),
Figure 218 and
Figure 219 provide visualizations that are similar but not identical to the typical probability and cumulative density functions.
Figure 218 illustrates density over time on one specific freeway segment. The report points out that density appropriately maxes out at approximately 170 vehicles per mile per lane (veh/mi/ln), which reflects queue spillback to the upstream links.

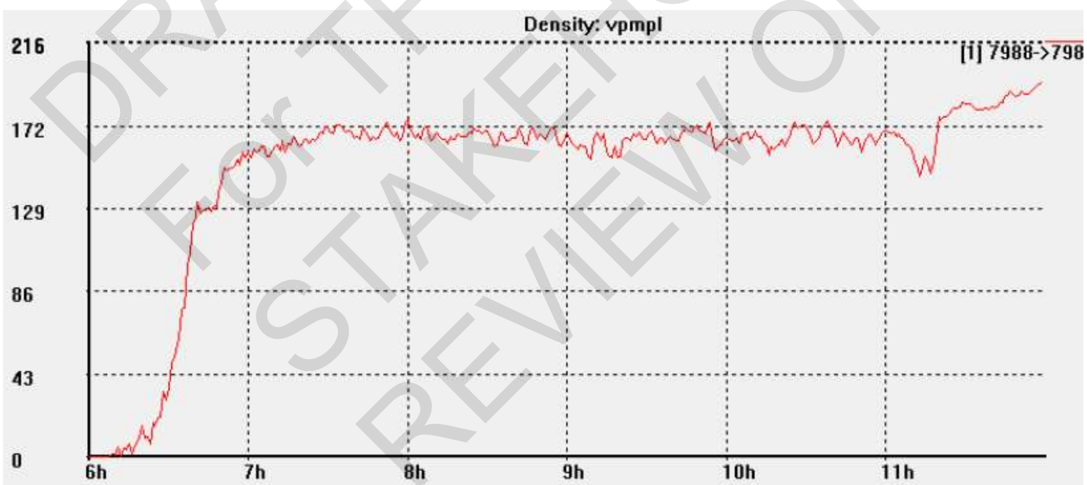


Figure 218. Chart. Link-specific density over time (North Carolina mesoscopic model).

Figure 219 illustrates traffic volume demand over time. The time series graph indicates that traffic peaks at approximately 8 a.m.

Figure 219. Chart. Demand volume over time (North Carolina mesoscopic model).

The Maryland case study report (2015) provides graphs that are somewhat analogous to the CDF, in the form of a cumulative travel time plots (see Figure 132).

Use of Confidence Intervals

In the I–66 VA study (2018), simulation results showed reductions in average delay. This improvement was statistically significant at the 95 percent confidence level, as illustrated in Figure 220 below. Average delay decreases from 643 to 608 seconds, representing a reduction of 35 seconds of delay for each individual traveler on average, and a percent improvement of 5.5 percent.

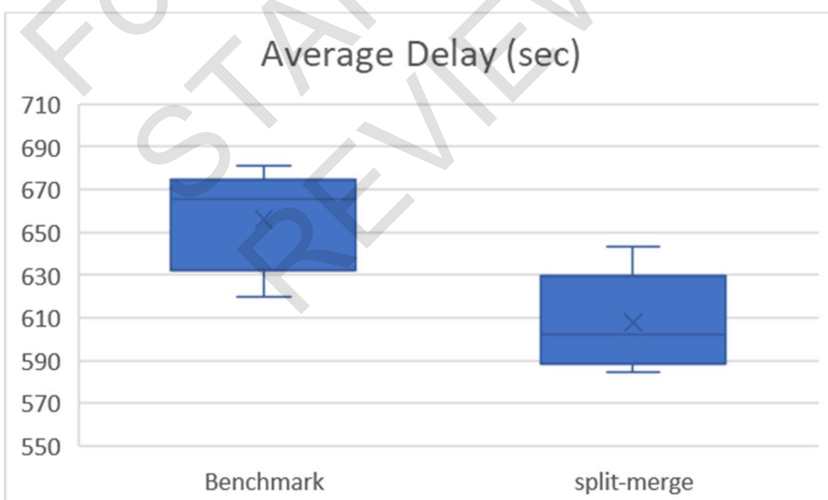


Figure 220. Chart. Delay reductions caused by the split merge design.

The Maryland case study report (2015) featured travel time validation targets for the I–270 model within +/- 10 percent confidence bands, 95 percent confidence intervals, and +/- 10 mi/h. Figures A13–A18 in Attachment A of the Maryland report (2015) show peak direction field measured travel times, VISSIM travel times, and 95 percent confidence intervals for the travel time segments. The 95 percent confidence intervals are based on the standard deviation calculated from the individual travel time runs. The graphs depict travel time and delay variability across the corridor with wide confidence bounds, specifically for the AM peak hour. The graphs show that almost all the VISSIM travel times fall within the calculated confidence intervals. One such graph is illustrated in
Figure 221 below.

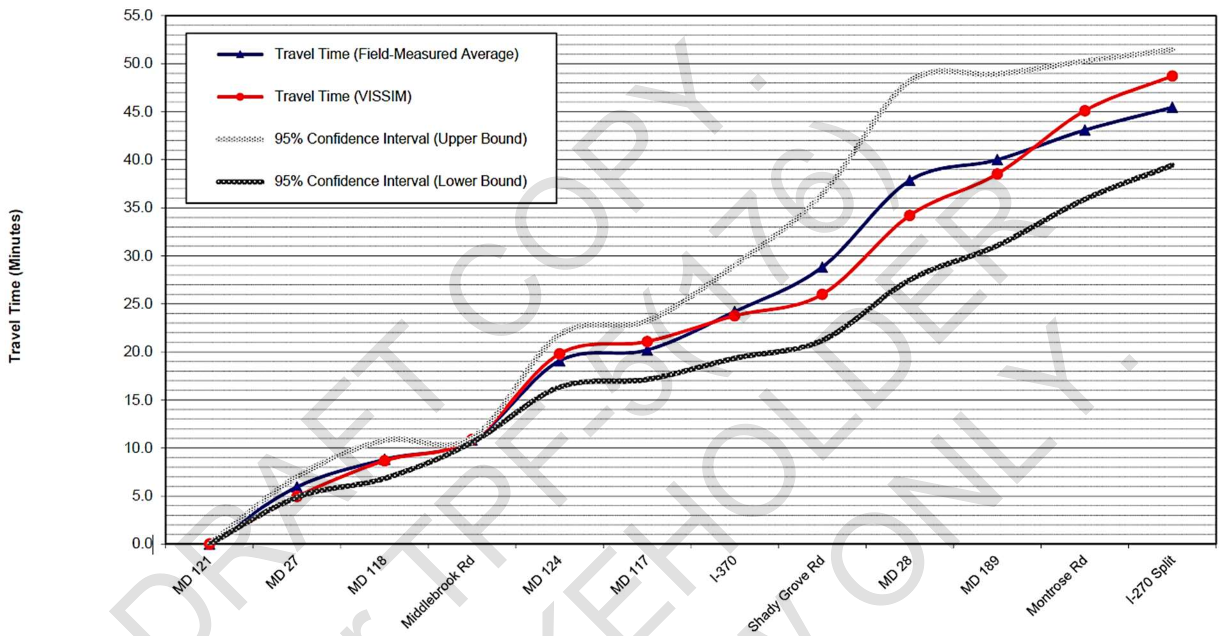


Figure 221. Chart. Cumulative travel time for 2015 weekday morning peak, southbound.

Use of Statistical Hypothesis Testing

The I–66 VA researchers used the following statistical testing to evaluate hypothetical impacts of relocating the managed lanes:

> *ANOVA Tests and Analysis of Merge Area. To further test and quantify the effects of each parameter on throughputs, ANOVA tests and Paired T-tests are applied to general purpose (GP) lane throughputs and HOV throughputs separately. Then, a regression analysis is applied separately. Table 37 shows the results of ANOVA tests in Throughputs and Delays. It suggests that all factors significantly affect the throughput values, except for the number of GP lanes. Like the throughput findings, all factors significantly affect the delay, except for the number of lanes is insignificant to HOV delay.*

Paired T-Tests. To test whether the two designs are different in each scenario, paired t-tests are performed. The null hypothesis is that the mean difference of the two designs is zero, while the alternative hypothesis is that the mean difference of the two designs is not zero. Estimations of mean difference and P values are provided. Results of representative scenarios are listed. There is a significant difference between the throughputs of GP vehicles as shown in Table 38 and Table 39.

Table 37. Throughput and delay analysis of variance table.

	Response	General purpose P-value	High occupancy vehicle P-value	Total P-value
Throughputs	HOV (right)	< 2.2e-16	0.003615	1.434e-07
	Number of lanes	0.1881	0.005253	< 2.2e-16
	V/C ratio	< 2.2e-16	< 2.2e-16	< 2.2e-16
	On-ramp volume	1.789e-09	1.092e-07	2.820e-10
	HOV percentage	3.578e-05	< 2.2e-16	< 2.2e-16
Average delay	HOV (right)	< 2.2e-16	< 2.2e-16	< 2.2e-16
	Number of lanes	< 2.2e-16	0.1932	< 2.2e-16
	V/C ratio	< 2.2e-16	< 2.2e-16	< 2.2e-16
	On-ramp volume	< 2.2e-16	< 2.2e-16	< 2.2e-16
	HOV percentage	0.00837	< 2.2e-16	9.211e-07

Table 38. T-test results: average delay and throughputs on general purpose and high occupancy vehicle lanes.

					General purpose throughputs		High occupancy vehicle throughputs	
	Number of lanes	HOV (%)	V/C ratio	On-ramp volume	Estimate mean of right-left	P-value	Estimate mean of right-left	P-value
Throughputs	4	0.25	0.8	1000	152	**4.02e-13**	0.1	0.88
				1500	130	**2.56e-06**	10.6	0.11
			1.1	1000	147	**2.18e-13**	-0.9	0.49
				1500	124	**7.73e-07**	19.1	0.29
Average delay	4	0.25	0.8	1000	-161	**1.34e-14**	19	**2.75e-07**
				1500	-131	**7.93e-09**	41	**2.41e-15**
			1.1	1000	-306	**5.90e-14**	25	**3.89e-10**
				1500	-223	**6.21e-06**	50	**5.20e-10**

Table 39. T-test results: total delay and throughputs on general purpose and high occupancy vehicle lanes.

				Throughputs		Delay	
	HOV (%)	V/C ratio	On-ramp volume	Estimate mean of right-left	P-value	Estimate mean of right-left	P-value
Total	0.25	0.8	1000	171	**4.59e-11**	-151	**7.49e-11**
			1500	149	**7.29e-10**	-101	**3.67e-09**
		1.1	1000	165	**2.05e-11**	-77	**7.67e-13**
			1500	144	**5.60e-14**	-102	**1.19e-13**

Contrarily, there is no significant difference between the HOV throughputs. The estimations of GP throughput differences between the design of the HOV lane on the right and the HOV lane in the normal design.

Table 38 also indicates that in terms of GP delay, the two designs are different. For the GP delays, the column "estimate mean of Right – Left" are all negative, indicating the GP delay is lower in the alternative design comparing with the normal design. In terms of HOV delays, the two designs are also significantly different. The column "estimate mean of Right – Left" are all positive, indicating the HOV delay is larger in the alternative design than the normal design. Table 39 shows the difference between the weighted total throughputs of the two designs for selected scenarios. Generally, when the V/C ratio is 0.55, the tests show there are no significant differences between the two. When the V/C ratio is higher the test results are significantly different. The total throughputs of the alternative design are higher than the normal design where the improvements range from 140 vehicles per lane per hour (veh/ln/hr) to 170 veh/ln/hr. The total average vehicle delay decreases in the alternative design, ranging from 70 seconds per vehicle and 150 seconds per vehicle. In total, the alternative design outperforms the normal design for the merge area for the synthetic simulation network and under the given system parameters.

Although the North Carolina study (2014) did not perform hypothesis testing, the report authors did include a helpful one-page section entitled Hypothesized Impacts. In this section, they developed some hypotheses for the expected impacts of various lane closure scenarios.

Analysis of Lane-Specific Results

The Interstate 270 Maryland (I–270 MD) network was used to evaluate various traffic management scenarios. Table 40 shows that speed improvements were particularly high on the right-most lanes in scenario 10.

Table 40. Lane-specific impacts of various traffic management scenarios (I–270 MD).

Average speed				
Scenario	**1**	**2**	**8**	**10**
Lane 1 (right)	20.4	19.2	37.2	63.3
Lane 2	17.2	16.1	35.0	62.6
Lane 3	18.6	17.4	35.7	61.7
Lane 4	23.7	21.7	40.9	63.5
Lane 5	34.5	29.2	46.7	65.4
Lane 6	42.4	37.9	52.9	66.6
Lane 7 (left)	N/A	45.0	57.4	68.7
Average speed change				
Scenario		**2**	**8**	**10**
Lane 1 (right)		-1.20	16.83	42.9
Lane 2		-1.12	17.80	45.42
Lane 3		-1.19	17.13	43.06
Lane 4		-1.93	17.27	39.87
Lane 5		-5.26	12.21	30.89
Lane 6 (left)		-4.50	10.55	24.26

N/A = data not required.

Use of Multivariate Outputs and Accompanying Visualizations

No single performance measure is perfect. Univariate measures may be misleading. It is helpful to consider several measures and distributions. Performance reporting can include combinations of numeric performance measures, visualizations, and qualitative discussions. In this case, the analyst's understanding would have been definitively reduced if they had not had access to the visualizations. In other cases, the numeric measures and/or qualitative discussions are needed to highlight important subtleties within a complex visualization.

Table 41 below contains multivariate performance measures from the North Carolina study (2014), while
Figure 222 provides accompanying bar charts that focus on the travel time results.

Table 41. Networkwide work zone impacts, morning peak.

Scenario		Average TT		Average Travel Distance		Average Stop Time		# of Unserved Vehicles	
		Min.	% Diff	Min.	% Diff	Min.	% Diff	Veh.	% Diff
Baseline		30.14	n/a	11.92	n/a	4.01	n/a	40168	n/a
WZ 1	No-Diversion	36.65	21.6%	11.65	-2.3%	10.07	151.1%	64088	59.5%
	With-Diversion	**31.58**	**4.8%**	**11.72**	**-1.7%**	**4.39**	**9.5%**	**44871**	**11.7%**
WZ 2	No-Diversion	35.12	16.5%	11.73	-1.6%	9.64	140.4%	57435	43.0%
	With-Diversion	**31.12**	**3.3%**	**11.73**	**-1.6%**	**4.07**	**1.5%**	**44815**	**11.6%**
WZ 3	No-Diversion	34.44	14.3%	11.75	-1.4%	9.17	128.7%	53325	32.8%
	With-Diversion	**31.17**	**3.4%**	**11.72**	**-1.7%**	**4.12**	**2.7%**	**44672**	**11.2%**
WZ 5	No-Diversion	33.06	9.7%	11.85	-0.6%	9.08	126.4%	46548	15.9%
	With-Diversion	**31.14**	**3.3%**	**11.73**	**-1.6%**	**4.08**	**1.7%**	**44606**	**11.0%**
WZ 7	No-Diversion	34.65	15.0%	11.83	-0.8%	9.43	135.2%	46868	16.7%
	With-Diversion	**31.26**	**3.7%**	**11.7**	**-1.8%**	**4.12**	**2.7%**	**44729**	**11.4%**

Source: Center for Advanced Transportation Education and Research.

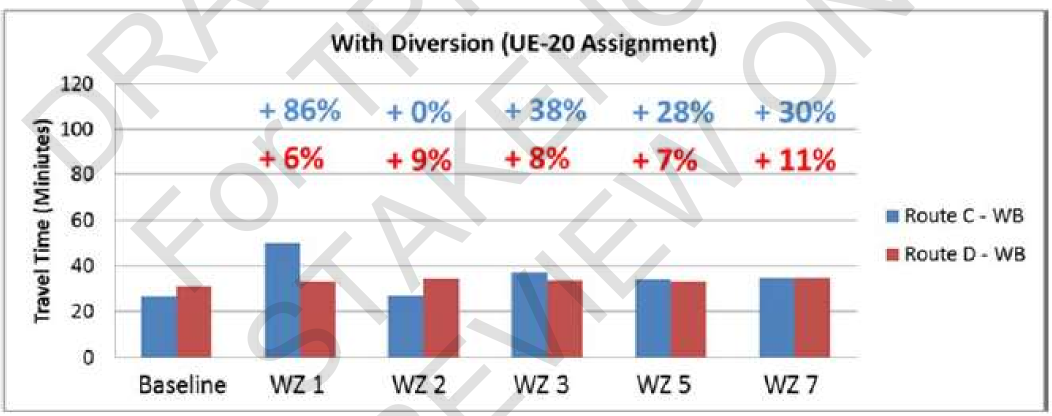


Figure 222. Chart. Travel time results for routes C and D, morning (AM) peak.

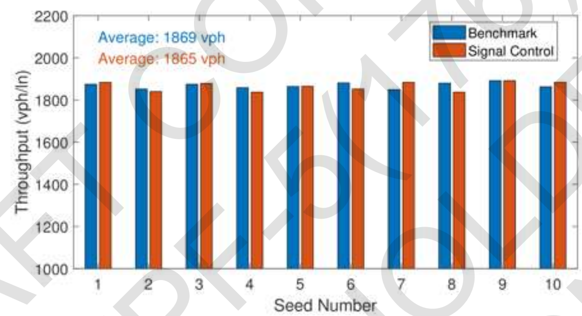
A. Subfigure of throughput.

B. **Subfigure of average speed.**

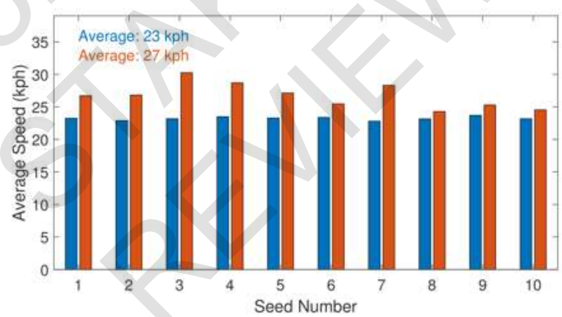
C. Subfigure of **average delay.**

D. Subfigure of carbon monoxide emissions.

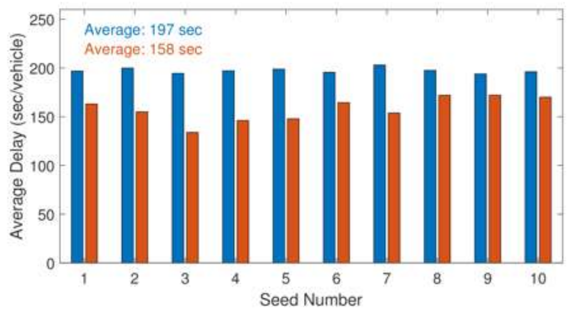
Figure **223** below illustrates four sets of results (i.e., throughput, average speed, average delay, CO emissions) generated by the I–35 KS network.



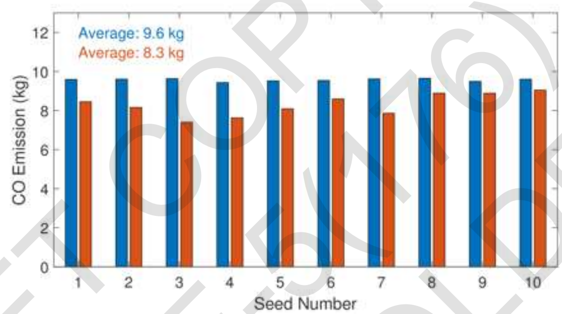
A. Subfigure of throughput.



B. Subfigure of average speed.



C. Subfigure of average delay.



D. Subfigure of carbon monoxide emissions.

Figure 223. Charts. Multivariate outputs generated by the I–35 KS network.

Presentation of Percent Changes Instead of Raw Numbers

During TSSM stakeholder meetings, stakeholders pointed out that simulation reports are easier to understand when they provide percent changes in the numeric performance measures, in addition to (or instead of) the raw numbers. The North Carolina case study report (2014) contained the following narrative:

> *[Mesoscopic simulation] results showed optimum diversion rates of 56% and 33% for the two-lane open and three-lane open options, respectively, in the AM peak under user equilibrium conditions. With these assumed rates of traffic volume reduction through the work zone, the network model estimates I-40 westbound travel times to increase through the work zone from 8.6 to 15.1 min for the two-lane pattern, and to 12.7 minutes for the three-lane pattern. In the more constrained two-lane open pattern, the model estimates volume increases in excess of 500 vehicles per hour (vehicles per hour) on Wade Avenue, US64, US1, and NC55, with smaller increases for other routes in the triangle. As a result, the model estimates travel time increases over 30% for I-440 eastbound (+32%), US70 northbound (+92%), Hammond Rd. northbound (+115%), and Rock Quarry Road westbound (+43%) in the AM Peak for the two-lane. For the three-lane open case, network-wide impacts are*

mitigated to some extent because fewer drivers select alternative routes. The model still estimates travel time increases over 30% for US70 northbound (+44%), Hammond Road northbound (+54%), and Rock Quarry Road westbound (+32%), and other significant impacts on I-440 EB (+10%), NC55 NB (+17%) and Hammond Road northbound (+22%).

By contrast, the segment-specific I–270 MD results were given in terms of raw numbers, as shown in Table 42 below.

Table 42. Vehicle speed summary for I–270 MD, afternoon peak.

PM Peak Speed Comparison - Local Lanes							
I-270 Local Northbound	**Field (mph)**	**VISSIM (mph)**	**Difference (mph)**	**I-270 Local Southbound**	**Field (mph)**	**VISSIM (mph)**	**Difference (mph)**
From C-D start				From C-D start			
to Montrose Rd	40.7	56.7	16.0	to Shady Grove	59.0	58.1	-0.9
to MD 189	33.8	35.9	2.1	to MD 28	56.8	53.3	-3.5
to MD 28	38.0	41.5	3.5	to MD 189	52.5	50.5	-2.0
to Shady Grove	18.5	17.9	-0.6	to Montrose	55.1	51.4	-3.7
to I-370	34.5	33.3	-1.2	to I-270 mainline	53.5	53.5	0.0
to MD 117	42.9	50.9	8.0				
to MD 124	23.9	12.7	-11.2				
to I-270 mainline	16.2	13.8	-2.4				
I-270 Local Total Speed (mph)	**27.8**	**25.4**	**-2.4**	**I-270 Local Total Speed (mph)**	**55.4**	**53.4**	**-2.0**
I-270 Local Total Percent Difference			**-8.6%**	**I-270 Local Total Percent Difference**			**-3.6%**

Source: Park, Won, and Perfater.

Final Documentation of Assumptions, Unknowns, and Recommendations

The North Carolina case study report (2014) provided the following discussion of limitations, assumptions, and recommendations:

This report provides NCDOT with results obtained from all tasks. It presents an estimate on the expected congestion impacts caused by the aforementioned work zone. Due to the extremely large scale of the modeled, caution should be taken in the interpretation of the absolute impacts of the work zone. While the team went through a rigorous calibration and validation effort, some errors in the calibration results remained. A partial calibration success is common in the application of mesoscopic models, as not all variability of traffic patterns in a large region-wide network can be predicted by simulation algorithms. The interpretation of the results in this report should therefore focus on the relative impacts of the work zone when compared to the (partially calibrated) baseline. As all work zone scenarios use the same baseline, this type of relative comparison is the most appropriate analysis approach.

In interpreting the model results, several key assumptions were made. First overall traffic demand was kept unchanged before and during the work zone, without any consideration *of trip reductions, additional carpooling, or telecommuting. The model further assumed that travelers only change their route between the same origin and destination, but not their departure time. In addition, no modal shift towards transit was modeled. All these travel strategies may result in further trip reduction and thus mitigate the impacts of the work zone on traffic congestion in the triangle.*

While it may seem intuitive that maintaining three lanes open to traffic (as opposed to two) in the work zone results in better operational performance, the magnitude of the difference between these two scenarios was striking and non-linear. For example, FREEVAL results suggest that a 30-40% diversion is necessary in the AM Peak hour to keep average travel speeds for a route through the work zone over 20 mph with a two-lane work zone pattern. With a three-lane pattern, FREEVAL estimates that only a 10-20% diversion rate will keep average travel speed above 40mph. For the PM Peak, the FREEVAL analysis suggests that a 40-50% diversion is necessary to keep the average speed over 10mph with the two-lane work zone pattern. With the three-lane pattern, it is estimated that a 40mph average travel speed can be maintained with approximately a 30-40% diversion rate. The results obtained from DTALite mirror these findings. They suggest that a diversion of 56% and 62% would be required for the AM and PM Peak period with the two-lane pattern, which may be an unrealistically high diversion goal. For the three-lane pattern, the model produces 33% and 36% diversion rates, which is still high, but likely more readily achievable with an intensive pre-trip and en-route traveler information campaign. Another benefit of the three-lane pattern is that it would lessen the severe impacts on alternate routes, because of the lower required diversion estimates.

Presentation of Static Graphics and Moving Vehicle Animation

Moving vehicle animation is typically associated with microsimulation, but static graphics are provided for all forms of simulation. One example of the static graphics associated with macroscopic or mesoscopic simulation is shown below in
Figure 224, where links are color-coded according to average vehicle speeds.

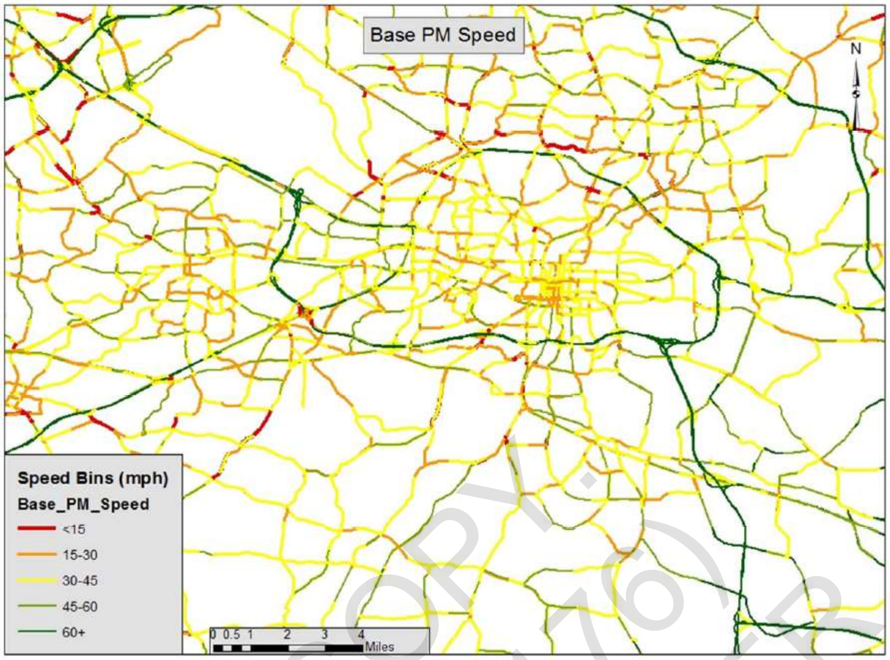


Figure 224. Map. Networkwide speed, afternoon peak.

Regarding the moving vehicle animation associated with microsimulation, the I–66 VA researchers documented an animation reasonableness check as part of their model verification and validation process:

The simulation calibration and validation have confirmed the reasonableness of simulation volumes at key areas (selected locations before and after merging areas) and travel time of key segments (especially those include the merging segments). This section focuses on checking the reasonableness of simulation behavior at key locations of the network, including car following behavior at basic segments, merging behavior at on-ramps and merging behavior at lane drops. Figure 225,
Figure 226, and

> *Figure* 227 *demonstrate simulation behavior of the above-mentioned three key areas. Animation checks at multiple key locations do not reveal abnormal behavior, and thus confirm the reasonableness of the simulation.*

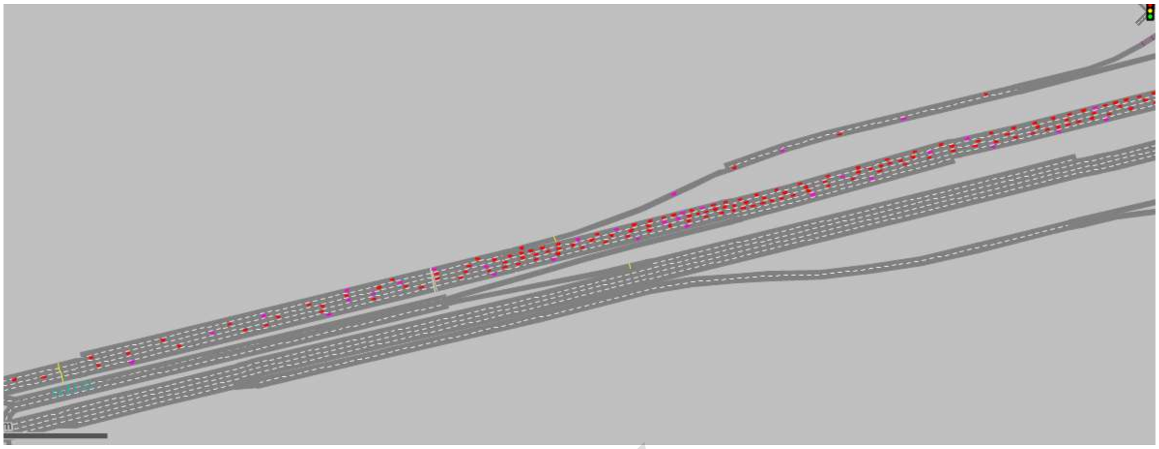


Figure 225. Illustration. Merging area simulation example 1 (bird's-eye view).

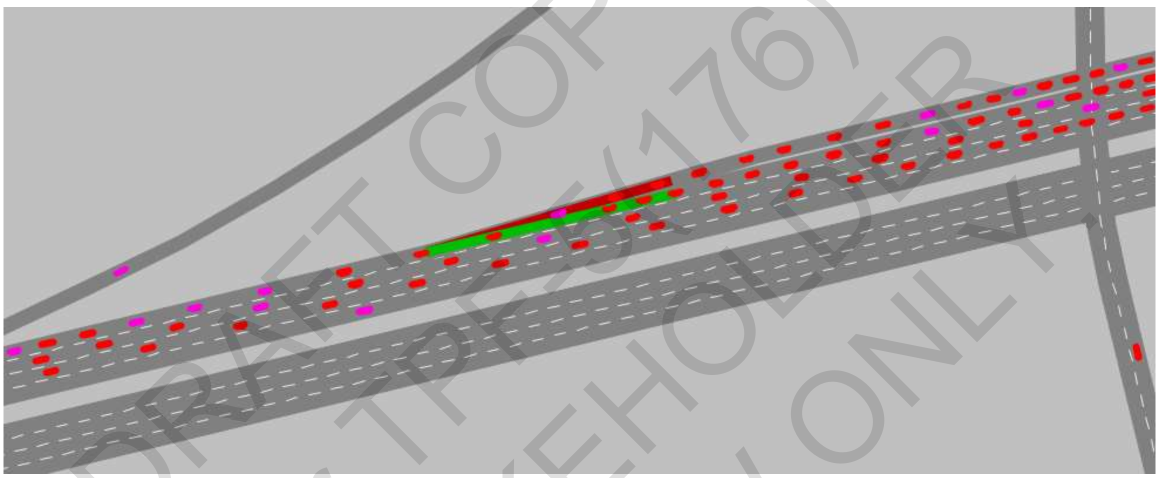


Figure 226. Illustration. Merging area simulation example 2 (zoomed-in view).

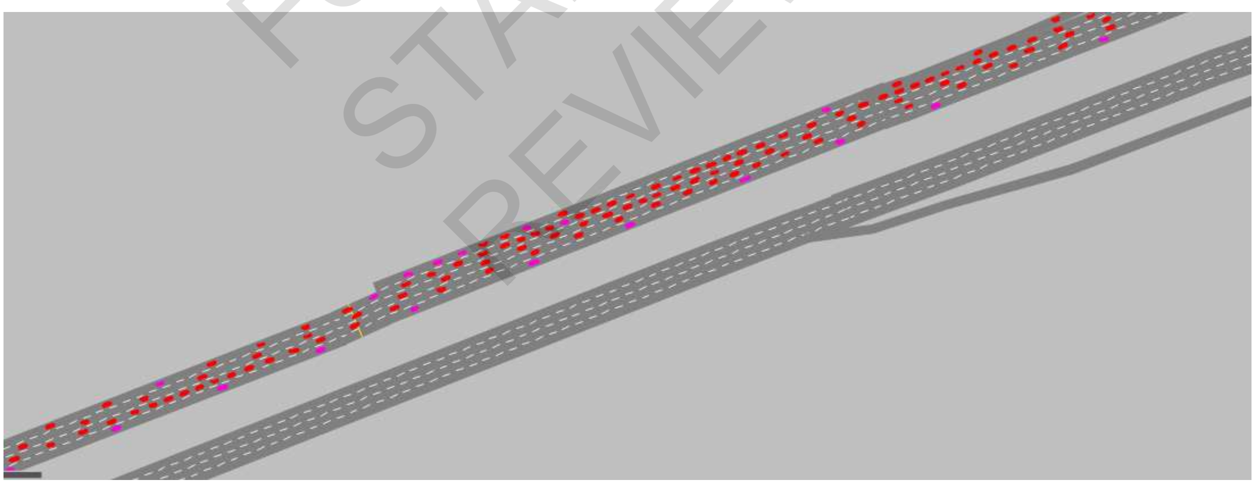


Figure 227. Illustration. Lane drop simulation (zoomed-in view).

For the I–270 MD animation reasonableness check, researchers compared simulated congestion levels against the congestion levels shown by probe data:

To check the reasonableness of the simulation animation and consistency with the field measurements, the animation results are compared with the field measured speed data downloaded from Regional Integrated Transportation Information System (RITIS) database.[97] The time period between October 1st 2017 and November 15th 2017 was selected for the speed data as in this period, traffic is not much affected by occasional trips and bad weather conditions. Moreover, only Tuesdays, Wednesdays, and Thursdays are selected for the data selection since traffic volumes are less likely to be affected by leisurely trips on these days.

Figure 228 **and**

Figure **229** *show the average speed heat map during 6:00-10:00 AM and 4:00-8:00 PM for the selected days, respectively. In these heat maps, the horizontal and vertical axes represent the location and time, respectively. As shown in the figures, three locations are already linked to the heat maps. With this, we can measure the scale of the location axes, and thus, link any other locations along the simulated corridor to the heat maps. We focus on the peak-hours that are bordered by the green block in the figures.*

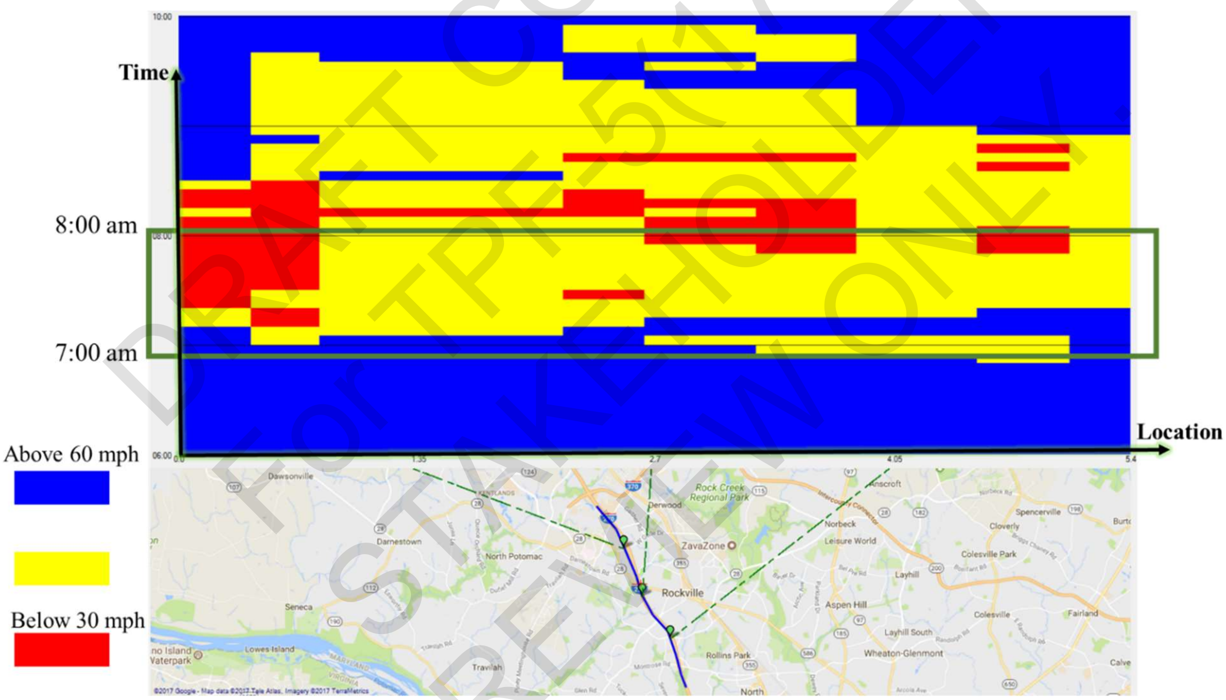


Figure 228. Illustration. Field-measured speed heat map during 6–10 a.m.

[97] "Probe Data Analytics Suite," RITIS, available at http://vpp.ritis.org.

Figure 229. Illustration. Field-measured speed heat map during 4–8 p.m.

We split the peak hour period into six 10-minute time intervals and then compare the simulation animation results with the field measured speed heat maps. In this report, we select four locations for each AM and PM peak hours as benchmarks.
Figure 230 and
Figure 231 zoom into the freeway geometry associated with the base data set to the peak-hour periods and tag the four selected benchmark locations. For comparison purposes, we classify vehicle colors in the simulation based on their speeds.

Table 43 *shows a snapshot of the vehicle color codes in the simulation animation. In this figure, the units are miles per hour. The remaining figures show examples of the comparison results. The comparison results indicate that the simulation animations match the field measured speed heat map for the most space-time points. We found a few inconsistencies between the animation and the field measured speeds (e.g., Animation check at 7:40 a.m.). However, most of these inconsistencies are related to microscopic traffic oscillations, and do not necessarily indicate serious inconsistencies with the macroscopic field speed measurements.*

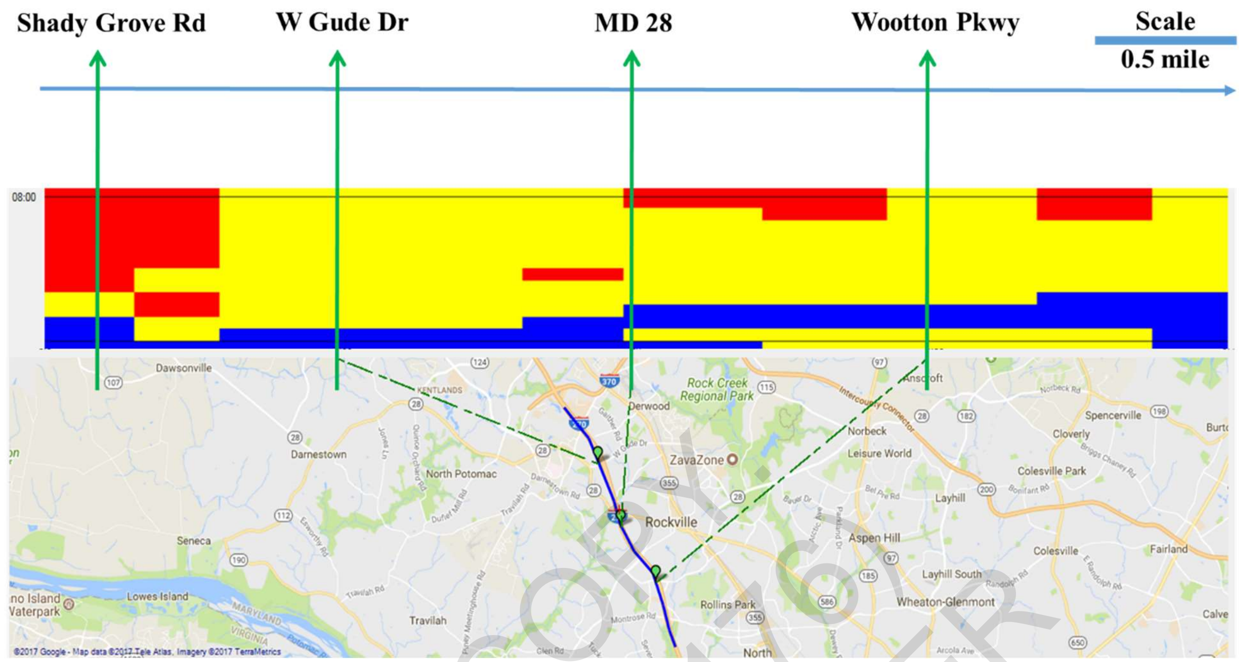


Figure 230. Illustration. Field-measured speed heat map for AM peak hour with location benchmarks.

In addition to these comparisons, we checked the simulation animation carefully to identify the existence of any stuck vehicle. No stuck vehicle was found in the I–270 freeway. We found only one minor problem on an on-ramp, as shown in

Figure 239. However, since this issue is located on an on-ramp and not on the mainline, it is not expected to significantly impact the simulation validity and results.

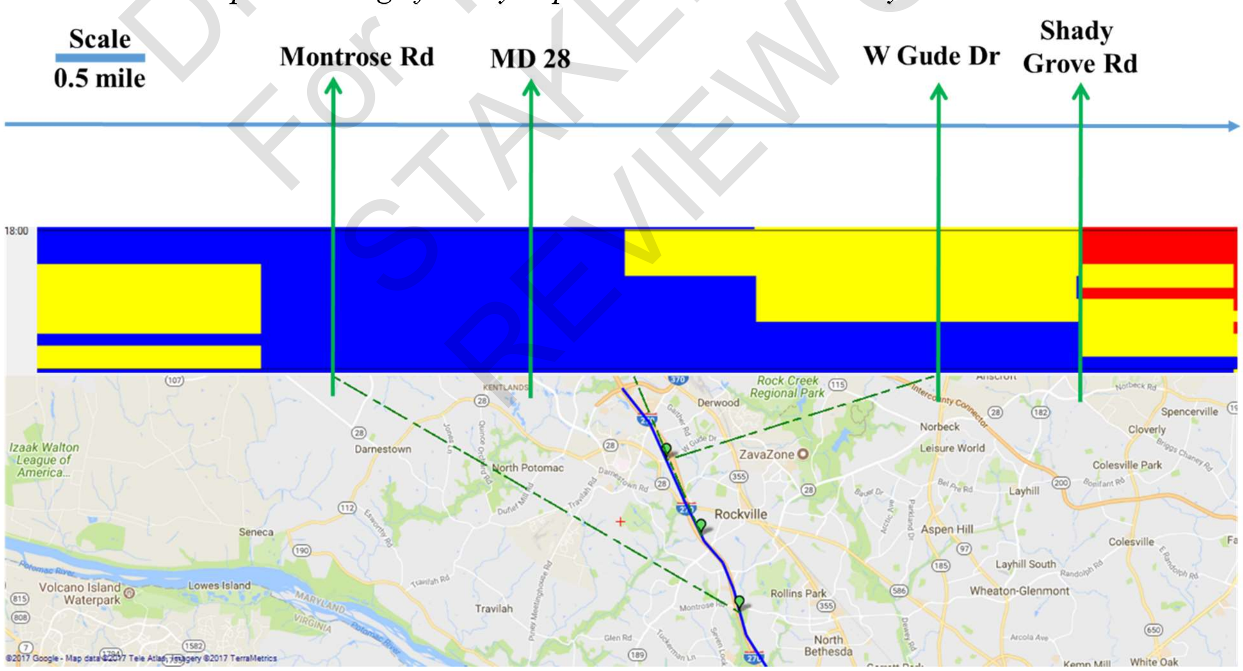


Figure 231. Illustration. Field-measured speed heat map for evening (PM) peak hour with location benchmarks.

Table 43. Simulation animation color codes.

Count: 11	LowerBound	UpperBound	Color
1	MIN	6.214	(255, 255, 128, 25)
2	6.214	12.427	(255, 255, 0, 0)
3	12.427	18.641	(255, 255, 128, 0)
4	18.641	24.855	(255, 255, 198, 0)
5	24.855	31.069	(255, 255, 255, 0)
6	31.069	37.282	(255, 198, 255, 0)
7	37.282	49.710	(255, 128, 255, 0)
8	49.710	62.137	(255, 0, 255, 0)
9	62.137	74.565	(255, 0, 187, 0)
10	74.565	124.274	(255, 0, 128, 0)
11	124.274	MAX	(255, 255, 255, 25)

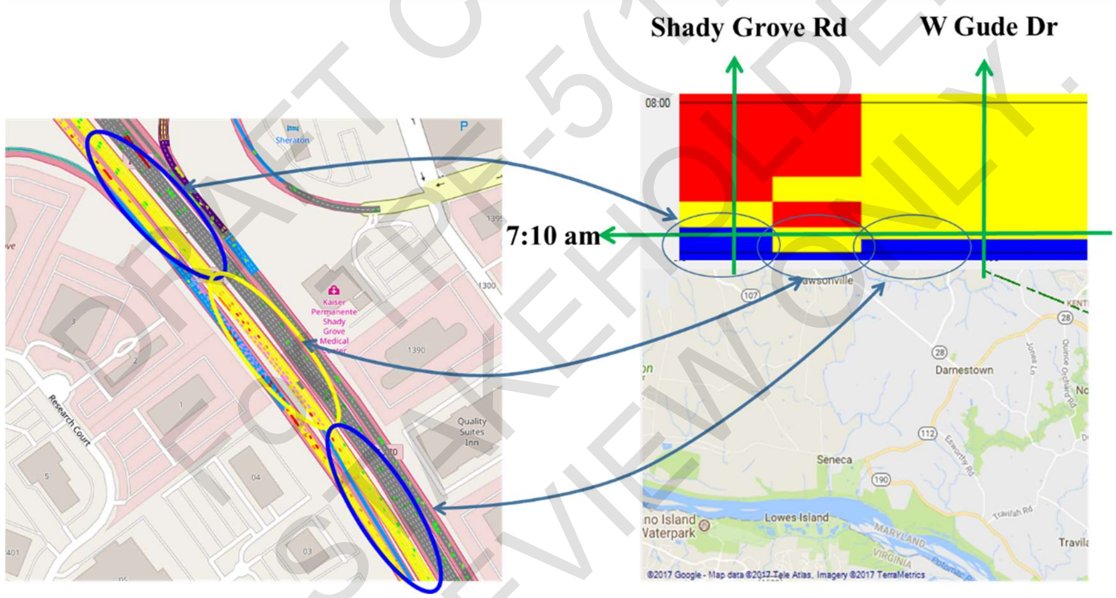


Figure 232. Illustration. Animation check at 7:10 a.m.

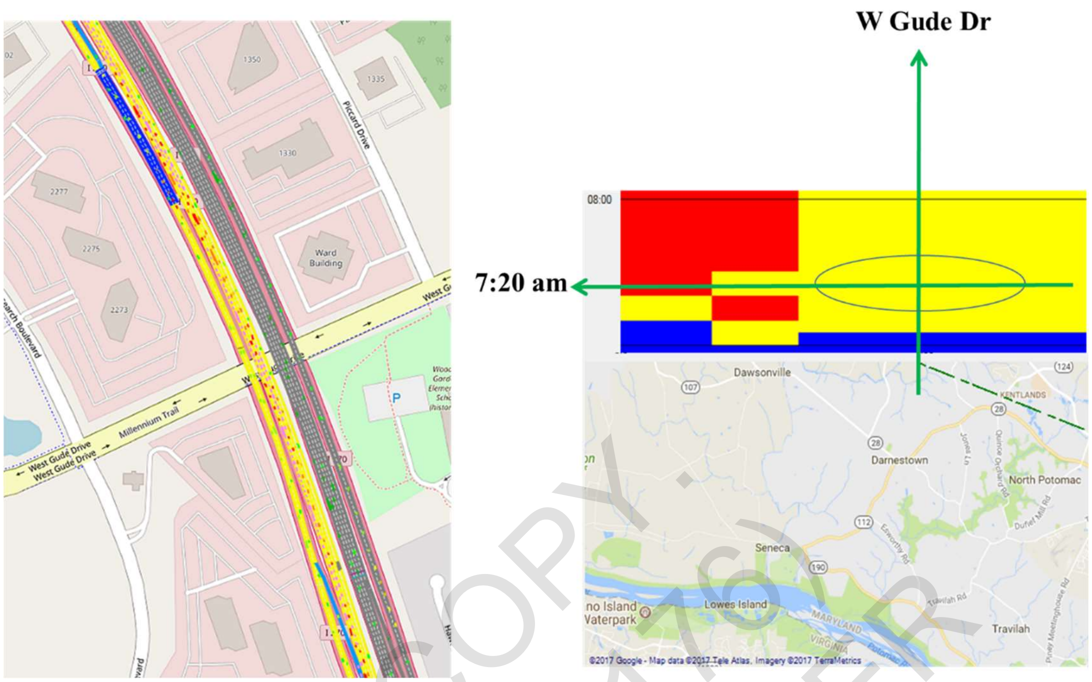


Figure 233. Illustration. Animation check at 7:20 a.m.

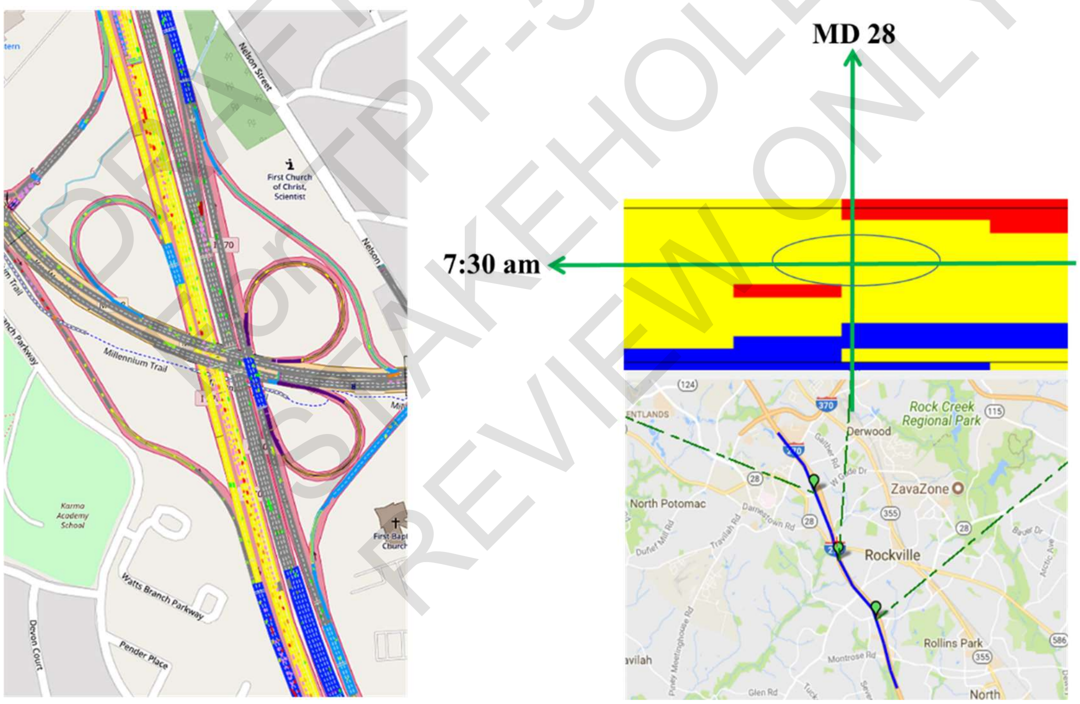


Figure 234. Illustration. Animation check at 7:30 a.m.

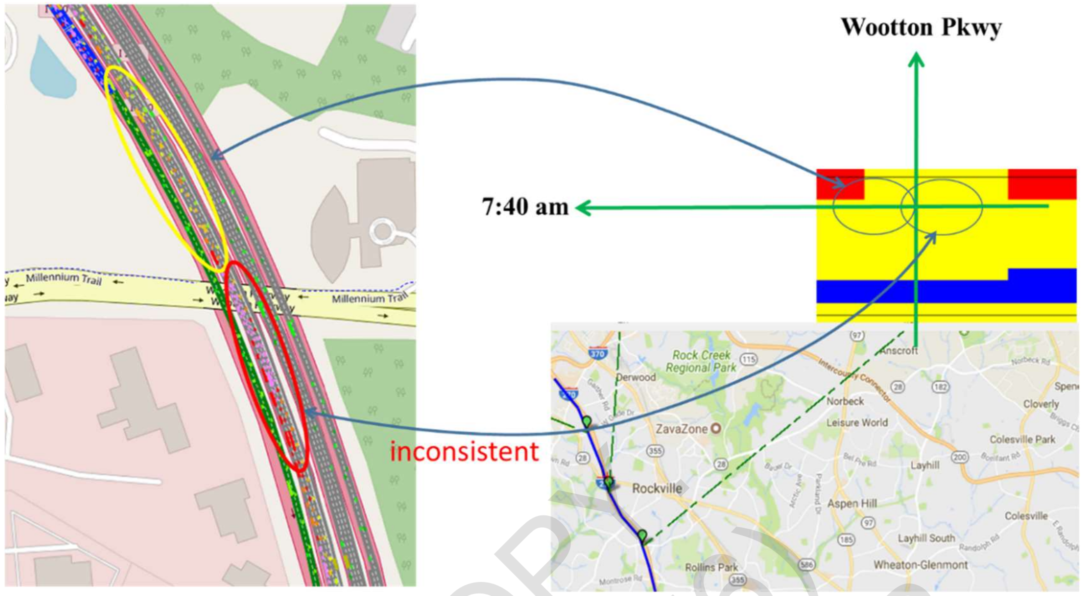


Figure 235. Illustration. Animation check at 7:40 a.m.

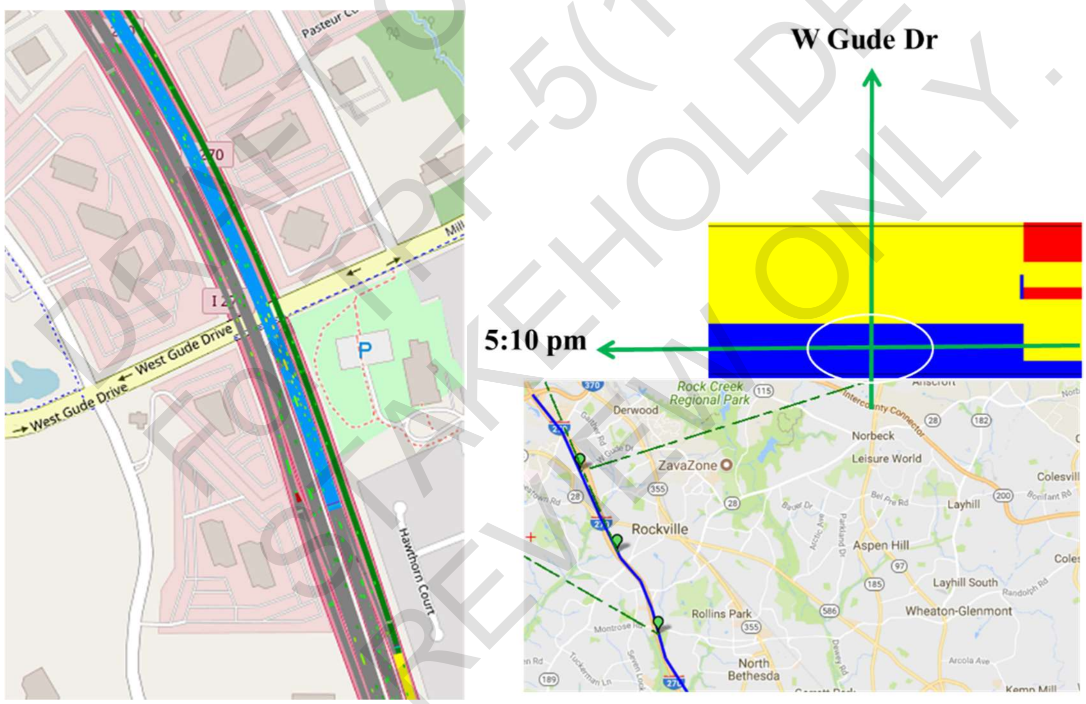


Figure 236. Illustration. Animation check at 5:10 p.m.

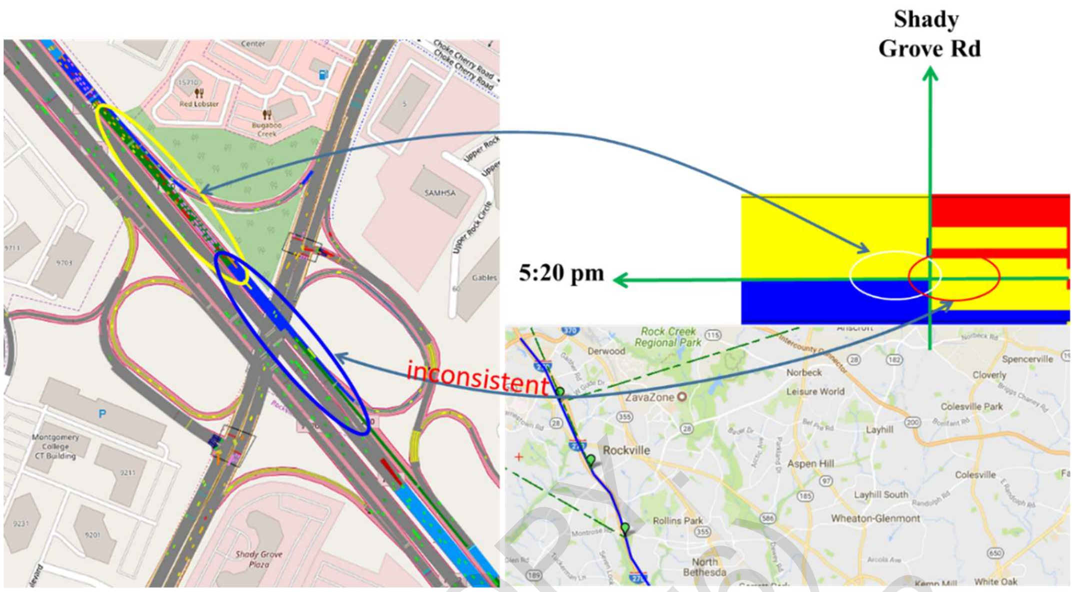


Figure 237. Illustration. Animation check at 5:20 p.m.

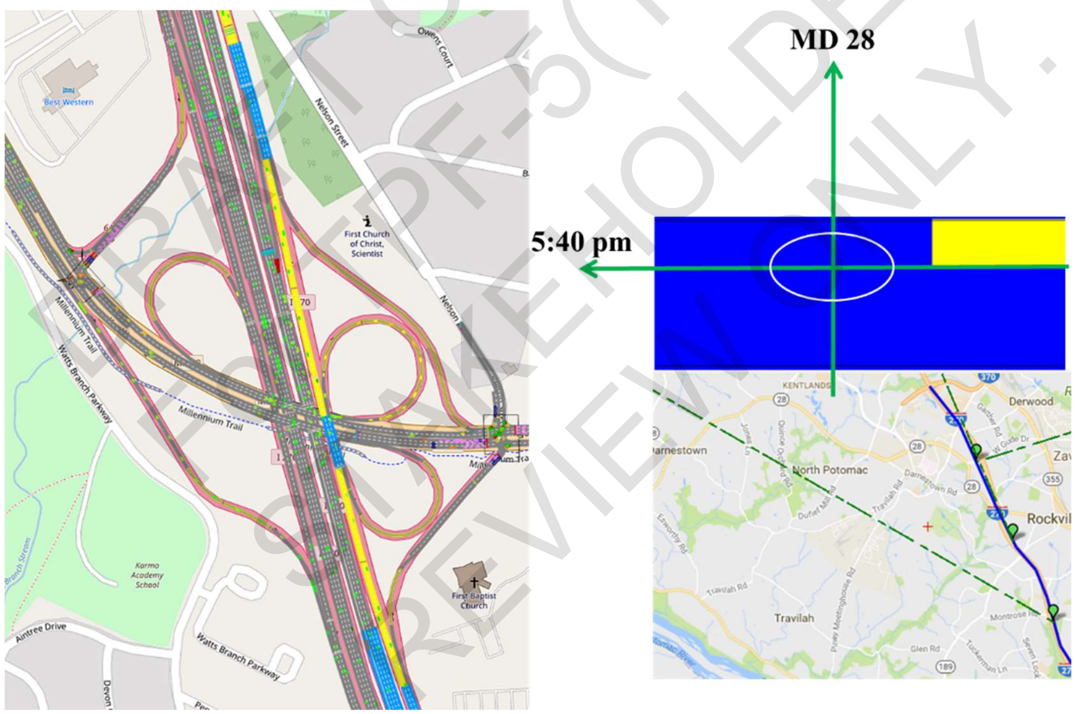


Figure 238. Illustration. Animation check at 5:40 p.m.

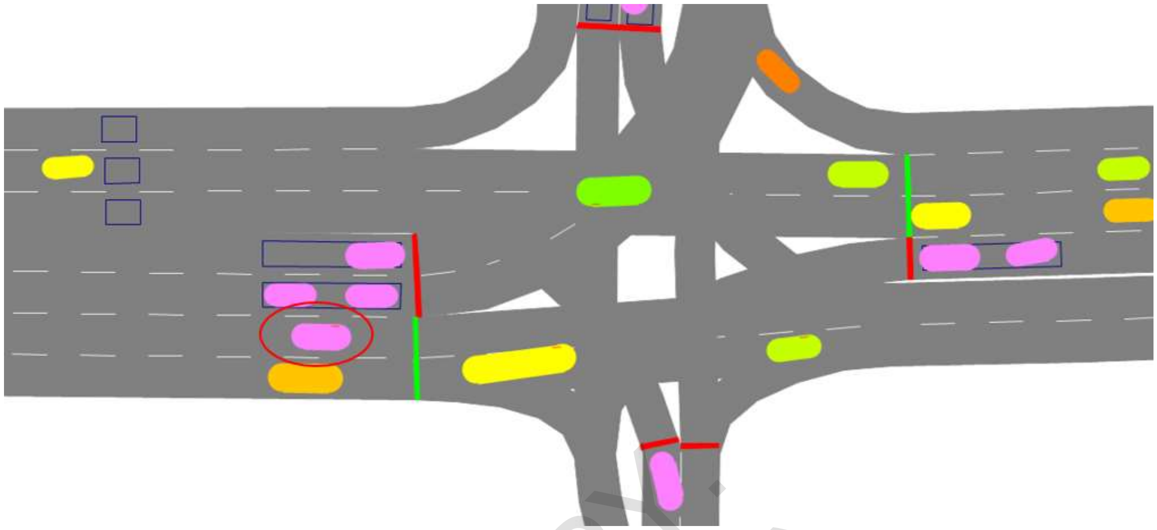


Figure 239. Illustration. A stuck vehicle on an on-ramp.

CHAPTER 13. ARTERIAL CASE STUDIES

This document discusses three traffic simulation case studies for surface arterial networks. The first case study is described in Traffic Analysis of Construction Phasing Options, Broadway Bridge, Boise (Six Mile Engineering 2013). The report provides the following introduction:

"The Idaho Transportation Department (ITD) has programmed a project to replace and widen the existing four-lane bridge on Broadway Avenue (US 20/26) over the Boise River. Construction funding for the project will be available in fiscal year 2015. The purpose of this study is to analyze the traffic impacts of two proposed bridge construction phasing alternatives and recommend a preferred alternative. The two proposed alternatives are: Partial Closure – Maintains two travel lanes on the bridge (one in each direction) for an anticipated construction duration of approximately 18 months. Full Closure – Closes all four travel lanes on the bridge for an anticipated construction duration of approximately 9 months. Both alternatives will result in traffic detouring from their normal travel route on the bridge. As a result, this study also identifies the potential detour routes, the anticipated traffic impacts to key intersections on those routes, and potential improvements to the intersections or routes to help alleviate traffic impacts during construction."

The bridge and surrounding traffic signals are illustrated in
Figure 240.

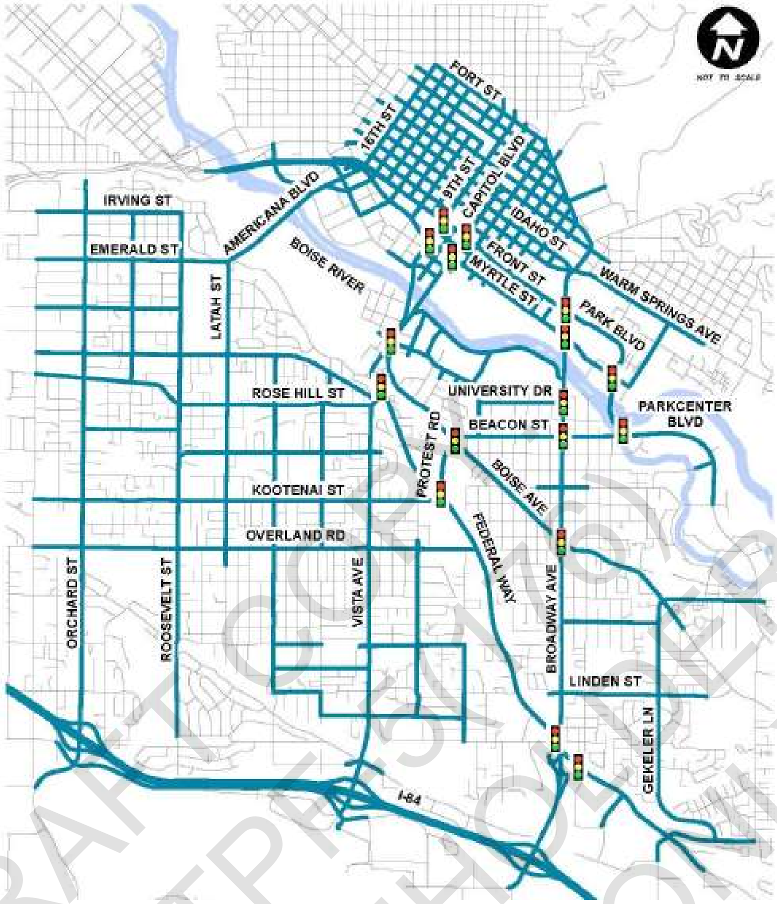


Figure 240. Map. Boise model area network and key intersections.

The second case study was developed in the context of a handbook, entitled Microscopic Simulation Model Calibration and Validation Handbook (Park 2006), prepared for the Virginia Department of Transportation. The handbook provides the following introduction:

"This test site is an urban network with 4 signalized intersections along Route 29 in Charlottesville, Virginia. Route 29 is one of the major routes in central Virginia because it serves most of the traffic volume from central Virginia to the Washington, DC, area. The test site is located at Emmet St. between Hydraulic Rd. and Barracks Rd., and includes a total of four intersections. Among them, two intersections at Emmet St. and Hydraulic Rd. and Emmet St. and Barracks Rd. work as metering intersections to consider boundary effects and only the number of vehicles that enter the network was considered. Another point that has to be considered is the existence of on- and off-ramps to Route 250 that connects Waynesboro to Richmond, Virginia. Because of this characteristic, the geometry of this network is more complex than other networks."

The four signalized intersections are indicated in the Figure 241 aerial photo.

Figure 241. Map. Four signalized intersections along Emmet Street.

The third case study is described in Proposed Signal Control at the Moana Lane/U.S. 395 Diverging Diamond Interchange (Center for Advanced Transportation Education and Research 2012). The report provides the following introduction:

"The Moana Lane/U.S. 395 interchange located in Reno, NV, is currently a standard diamond interchange and is controlled using two signal controllers. The Nevada Department of Transportation (NDOT) is planning to reconstruct this interchange into a Diverging Diamond Interchange (DDI). The Center for Advanced Transportation Education and Research (CATER) at the University of Nevada, Reno (UNR) was retained by NDOT to develop and evaluate feasible signal phasing schemes to operate the DDI. The main goal of this study is to evaluate the proposed phasing schemes using the hardware-in-the-loop simulation technology so that the schemes evaluated in simulation are readily implementable in the field once the interchange is reconstructed."

The original diamond interchange shown in
Figure 242. The proposed diverging diamond interchange (DDI) geometry is shown in Figure 243.

Figure 242. Map. Original Moana Lane/U.S. 395 standard diamond interchange.

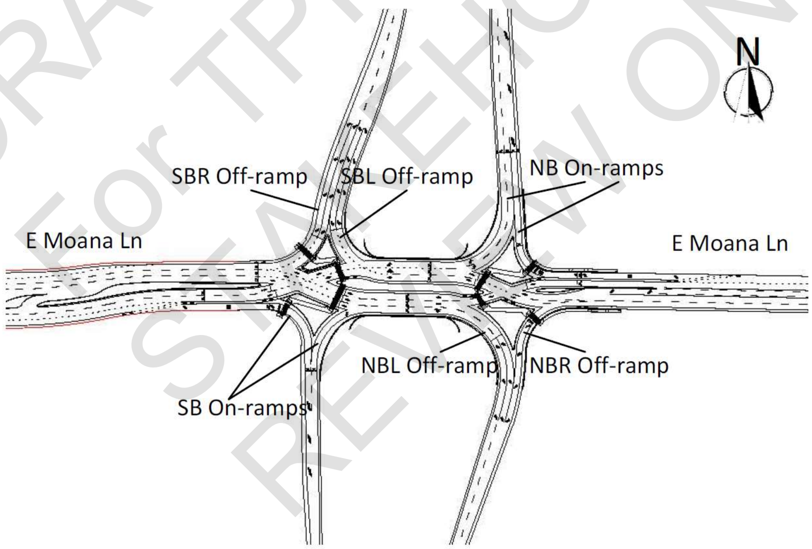


Figure 243. Proposed configuration of the Moana Lane/U.S. 395 diverging diamond interchange.

13.1 DEFINING AND SCOPING THE SIMULATION ANALYSIS PROBLEM

Fundamental Project Parameters

Transportation System Simulation Manual (TSSM) 0describes the need to define and document the scope, goals, approach, budget, objectives, and key performance measures for any traffic simulation study. For the most part, all three case studies accomplish this. For the Boise case study (2013) and Reno case study (2012), objectives and scope were largely given in the prior introductions from this document. For the Charlottesville case study (2006), the study objective was to demonstrate the proposed calibration and validation methods for VDOT, but this was implicitly understood due to the nature of the handbook. The scope of a simulation study can take on many components (e.g., spatiotemporal boundaries, base year(s), forecast year(s), alternatives under consideration).

Spatiotemporal Boundaries

0emphasizes the importance of defining an appropriate set of spatial and temporal study boundaries, which can capture all relevant congestion. The Boise report (2013) provided the following description:

“The DTA model study area (subarea network) measures approximately 3 miles by 4 miles. Its limits extend to Idaho Street to the north, Parkcenter Boulevard to the east, I-84 to the south, and Orchard Street to the west. Included in the area are 85 signalized intersections, 45 traffic analysis zones and three Boise River crossings, in addition to the Broadway Bridge. Of the 85 signalized intersections, 17 were identified as being potentially sensitive to impacts from the detoured traffic and were selected for detailed traffic analysis to quantify those affects. The evaluation period was limited to the weekday PM peak hour from 5:00 to 6:00.”

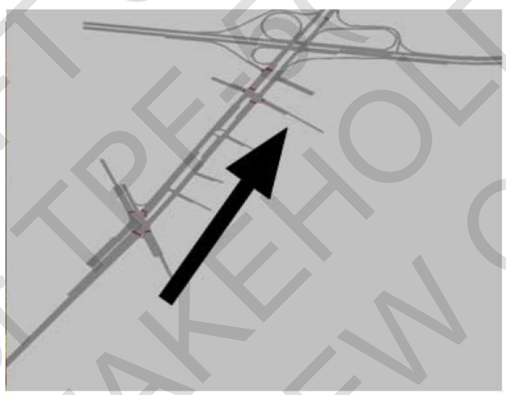
Although the VDOT report (2006) did not specify which time periods were evaluated for the Charlottesville network, it did describe efforts to extend the spatial network coverage as follows:

“During the preliminary site visit as well as data collection, a unique phenomenon was observed by the research team. As mentioned, Route 250 intersects with Route 29 within this network and Route 250 shares its roadway with the Route 29 bypass. Because of its high speed and capacity, a large percentage of vehicles on Route 29 south try to diverge to Route 250 / Route 29 bypass through a one-lane ramp and it forms quite a long queue on the right most-side lane that requires additional attention when building a network” (2006, 81).

B. Subfigure of **dynamic traffic assignment (DTA) model of the study area.**

Figure **244** shows the real field condition.

A. Subfigure of study area physical environment.



B. Subfigure of dynamic traffic assignment (DTA) model of the study area.

Figure 244. Maps. The need to extend the physical network to capture queue spillback.

The Reno case study (2012) focused on morning (AM) and evening (PM) peak hours, but the spatial boundaries apparently did not extend beyond the interchange. When conducting a simulation study for a localized facility improvement such as a diverging diamond interchange (DDI) conversion, TSSM authors recommend consideration of hidden bottlenecks. Anecdotally, some DDI constructions have worsened congestion on the wider traffic network, simply because elimination of the interchange congestion caused a bottleneck to move to a nearby hidden bottleneck. Simulation analysts can examine the nearby road network, to see if any hidden bottlenecks might exist.

Role of Simulation

The VDOT handbook (2006) did not address the role of simulation in carrying out the Charlottesville traffic study, presumably because the purpose of the handbook was to provide guidance on calibration and validation of simulation models. In the Reno case study (2012), simulation was chosen to evaluate signal timing plans directly from the signal controllers, using hardware-in-the-loop technology.

Unlike the other two case studies, analysts for the Boise case study had to choose among a wide variety of possible traffic models. Ultimately they chose Dynameq, a simulation-based dynamic traffic assignment model:

"Bridge construction phasing alternatives were initially analyzed using traffic forecasts from the Community Planning Association of Southwest Idaho (COMPASS) regional travel demand model. Different study area sizes and bridge closure scenarios were tested. The results from the analysis were unexpected because the network delays for several of the Partial Closure scenarios were less than the network delays for the existing conditions. It was recognized that the COMPASS regional travel demand model is a static model and cannot capture dynamic travel behavior or incorporate the effects of intersection delay and queuing that were expected to impact traffic distributions and network delays. To more accurately model traveler behavior and predict traffic patterns for the two construction phasing alternatives, the Federal Highway Administration (FHWA) recommended utilizing Dynamic Traffic Assignment (DTA) modeling. INRO Solutions (INRO) was selected through a proposal process as the firm to conduct the DTA modeling for this study."

Formation of Project Team and Stakeholders

Although the three case study publications were clear about who the customers were (i.e., IDT, VDOT, NDOT), they did not provide much information on the project team members or non-customer stakeholders.

Identify Needed Resources

The Boise case study report (2013) cited the Dynameq software, Synchro software, Google® Maps software, Emme software, Community Planning Association of Southwest Idaho (COMPASS) regional travel demand base model, travel demand data, traffic count data, geometric network data, and knowledge of local conditions as necessary modeling resources. The Charlottesville case study (2006) chapter cited VISSIM software, CORSIM software, spreadsheet software, traffic counts, heavy vehicle percentage, geometrical characteristics, detector locations, signal timing plan, speed limits, travel time data, maximum queue length data, video recording devices, electronic data collection devices, data collection sheets, probability density functions, genetic algorithm, contour plots, and analysis of variance (ANOVA) statistical tests as necessary modeling resources. The Reno case study (2012) cited detailed signal timing

plans from the Department of Transportation (DOT) and hardware-in-the-loop system as key resources. Existing signal timings from NDOT are illustrated in
Figure 245.

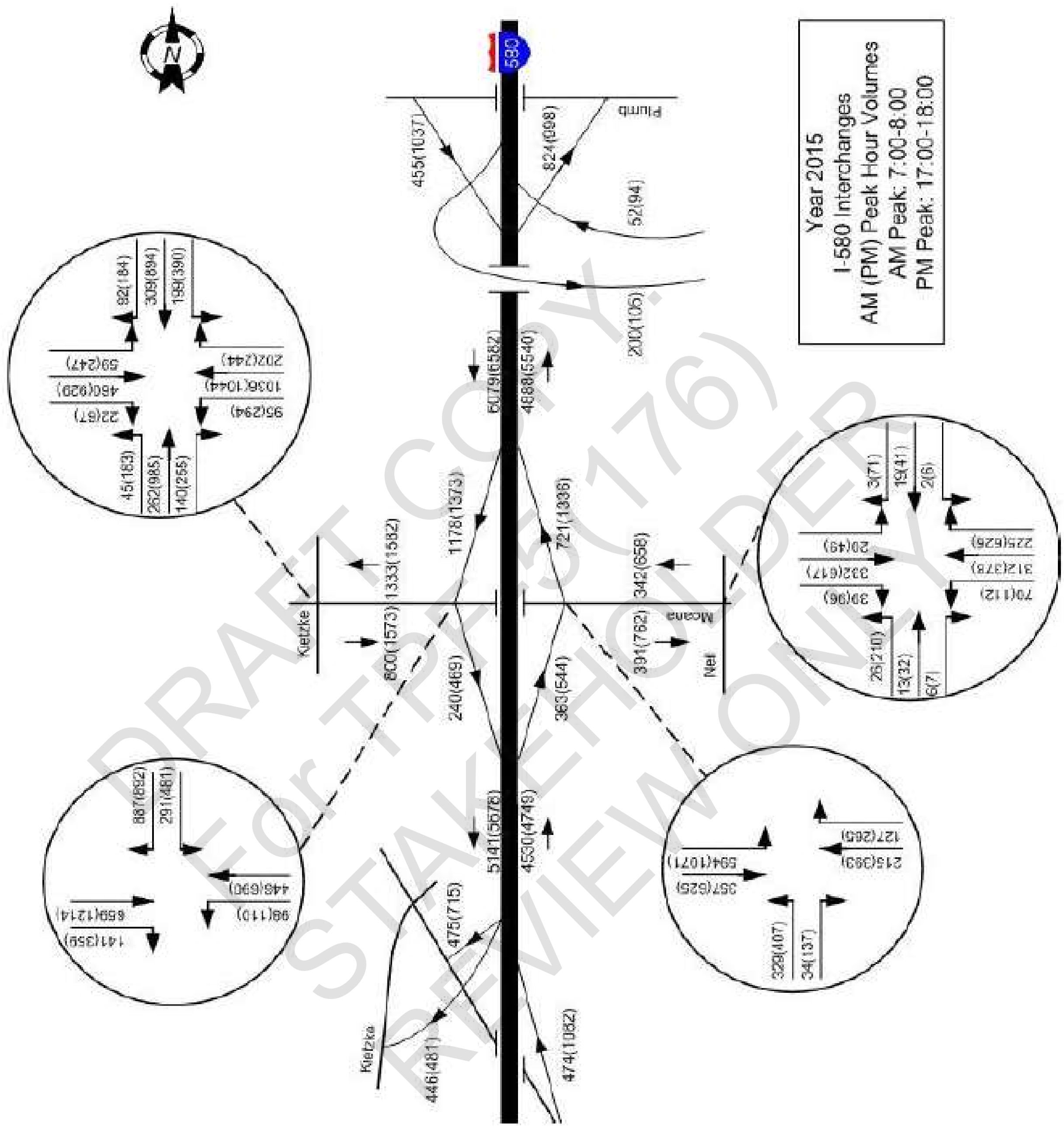


Figure 245. Diagram. Moana interchange existing signal timings.

Define Forecast and Base Years

The three case studies described in this document were all concerned with improving short-term traffic flows. In each case the base years were explicitly stated, but there were no forecast years.

Range of Operational Conditions to Consider

In recent years, transportation professionals have begun to realize the significance of annual performance and travel time reliability. This requires explicit analysis of operating conditions including varying traffic demands, weather, incidents, work zones, and special events. The three case study reports summarized in this document approached the traffic study using a more traditional approach, in which the operating conditions were not significantly discussed. However, TSSM authors recommend detailed analysis of the operating conditions whenever possible; and when these conditions are analyzed, the methods of doing so should be well documented. Today's traffic modeling experts should become well-versed in the simulation-based travel time reliability analysis methods described in the SHRP2 L04 report (Mahmassani et al. 2014), as well as the calibration methods described in the Federal Highway Administration's (FHWA) Traffic Analysis Toolbox Volume III (Wunderlich, Vasudevan, and Wang 2016).

Screening of Initial Alternatives through Non-Simulation Methods

When a traffic analysis must consider many traffic control alternatives, roadway geometric design alternatives, and/or demand management alternatives, it might not be practical to analyze each alternative through simulation. In such cases, non-simulation-based tools (e.g., sketch-planning tools) may help to weed out some portion of the ineffective alternatives that were being considered. The three case study reports summarized in this document did not need to consider many what-if scenarios, and as such did not require use of screening tools. The Boise case study (2013) included isolated intersection analysis through Synchro's Highway Capacity Manual (HCM) calculation method, which can be considered a non-simulation method, but this analysis was not done for screening purposes.

Explicit Linking of Performance Measures to Agency Goals or Project Objectives

In the Boise case study (2013), performance measures were clearly chosen to gauge areawide impacts of the bridge closure alternatives (i.e., partial closure vs. full closure):

- Percent of traffic rerouting away from the Broadway Bridge.
- Networkwide average vehicle hours delay (VHD).
- Networkwide vehicle hours traveled (VHT).
- Networkwide vehicle miles traveled (VMT).
- Major-street directional travel times.

Table 44. Travel time results: Broadway/Front to/from Broadway/Boise.

Direction	Origin	Destination	Travel Time Results (minutes)			Travel Time Difference					
			Base (via Broadway Bridge)	Partial Closure (via Broadway Bridge)	Full Closure (via Parkcenter Bridge)	Partial Closure vs. Base		Full Closure vs. Base		Full Closure vs. Partial Closure	
						min	%	min	%	Min	%
NB	Broadway/Boise	Broadway/Front	3.9	4.8	7.3	0.9	23%	3.4	87%	2.5	52%
SB	Broadway/Front	Broadway/Boise	4.2	4.5	7.7	0.3	7%	3.5	83%	3.2	71%

Source: Park, Won, and Perfater.

In the Charlottesville case study (2006), performance measures were chosen to demonstrate effectiveness of the recommended calibration and validation procedure. To accomplish this, travel time data for two different roadway sections were selected for use in the calibration procedure. Similarly, maximum queue length data at two different locations were selected for use in the validation procedure. In the case of the maximum queue length data, the queue length distances reported by the simulation tools differed from the number of queued vehicles collected in the field. Therefore, the simulation outputs were converted into numbers of queued vehicles.

In the Reno case study (2012), performance measures were chosen to quantify signal timing plan benefits for the proposed DDI. The chosen performance measures were average delay (seconds per vehicle [sec/veh]) and maximum queue (feet [ft]), which are two of the most common measures for quantifying signal timing plan benefits.
Figure 246 illustrates some of the projected short-term benefits.

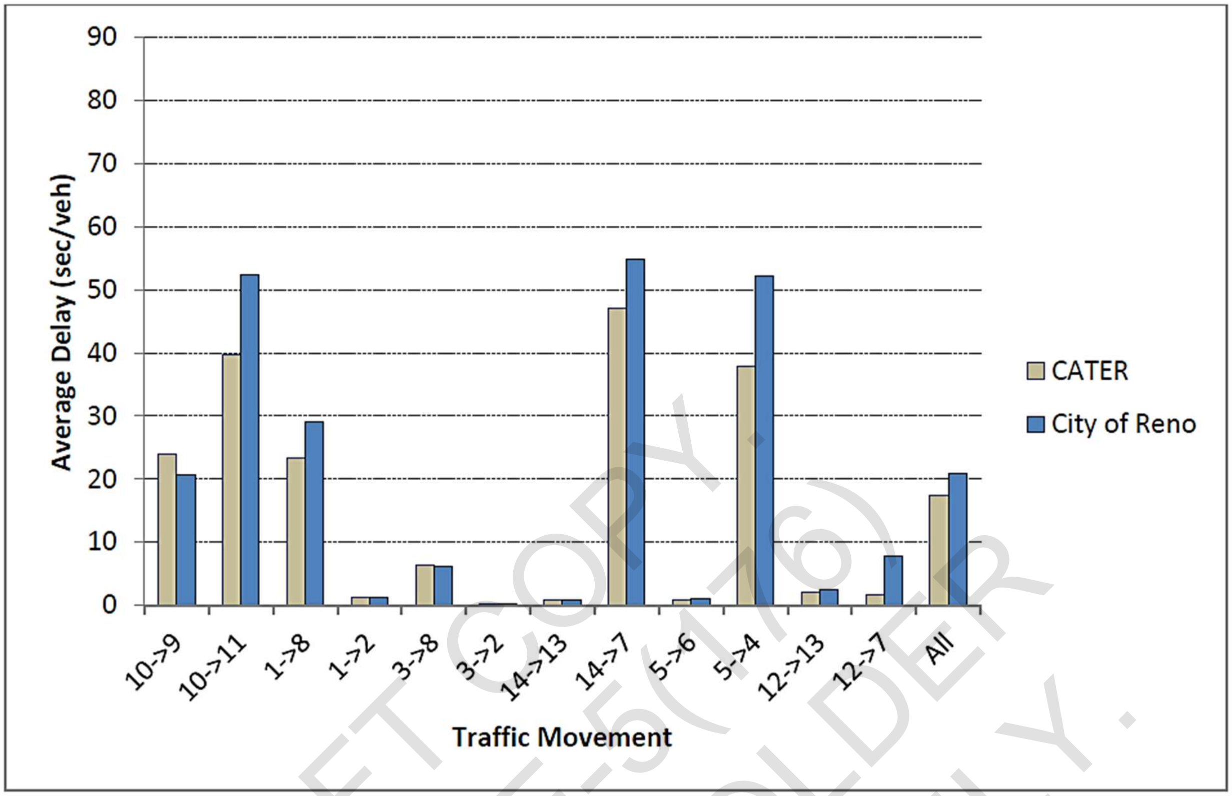


Figure 246. Chart. Delay reductions projected by diverging diamond interchange signal timing in Reno.

13.2 DATA

Origins of Data Used in the Analysis

The Boise case study provided the following discussion of data origins:

- Travel demand data was obtained from the COMPASS base (approved) 2012 model.
    - Origin-Destination (O-D) demand matrices.
    - Centroids and connectors.
- Geometry and volume data was obtained from Ada County Highway District's (ACHD) countywide Synchro model.
    - Network links and nodes, with corresponding attributes (speed, number of lanes, turn lanes, storage lengths, etc.).
    - Signalized intersection turning movement count data.
    - Signal timing data.

The DTA model was initially verified using information gathered from Google® Maps. The model was later validated utilizing travel time data collected by COMPASS on major roadways for the PM peak period.

Stop control data was gathered from Google® Maps Street View and coded manually into the network, including priority relationships. Demand matrices were imported from COMPASS demand model. Two hourly matrices were imported for the PM peak period, 4–5 p.m. and 5–6 p.m. The total number of trips for this period was 116,047.

A crucial element before model calibration and validation is to ensure that the empirical data, namely traffic counts and travel times, are valid and consistent. Traffic turn counts for the PM peak hour 5–6 p.m. were imported from Synchro for a total of 143 intersections. These counts were collected in a wide range of years from 1998 to 2013, as shown in Figure 247. To maintain consistency, it was decided to use only counts from 2010 and up. These counts were checked for significant gaps between the counts at two adjacent intersections, i.e., that the sum of turn counts that enter a network link is approximately equal to the sum of turn counts that leave it. It was decided to not use several counts that created a gap of more than 10 percent. The total number of turn counts that were found to be usable for model calibration was 697 at 89 intersections, which provided good coverage of the network.

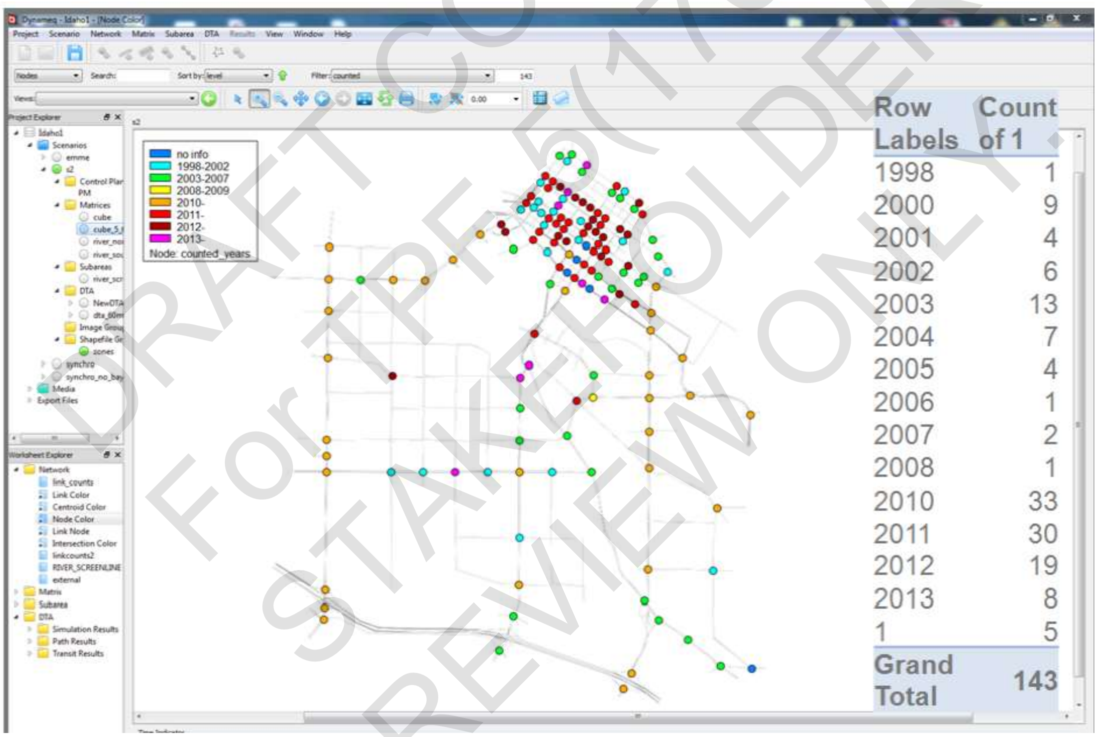


Figure 247. Screenshot. Synchro turn counts by year.

The DTA model was validated with travel time data for the PM peak hour 5–6 p.m., collected in floating car surveys. A description of the method can be found in a Treasure Valley Annual Congestion Management System (CMS) report by COMPASS.[98] Travel times for 18 routes were available for network validation.

[98] http://www.compassidaho.org/documents/prodserv/reports/2012CMSReport.pdf.

The Charlottesville case study (2006) provided the following discussion of data origins:

> The geometric attributes pertaining to the test network were obtained from SYNCHRO and CORSIM files, developed by VDOT. However, as these networks were developed in 2001, the changes in geometry such as lengths of left- and right-turn lanes or locations of bus-stops were updated using aerial photos and site visits. The current signal timing plans of Emmet St. (Rte. 29) at Angus Rd., Morton Rd. and Barracks Rd, which is in the jurisdiction of the City of Charlottesville, was provided by the traffic engineer in the City of Charlottesville. The signal timing plan of Rte. 29 at Hydraulic Rd. was obtained from VDOT, who manages the intersection. The traffic-related data were collected directly from the field using both manual counting and video recording. The data collection was conducted on a normal weekday, Tuesday through Friday, July 11, 2001, 2:45 p.m. and 4:15 p.m. A group of 17 people performed simultaneous manual counts along the test site as shown in
> *Figure* 248.

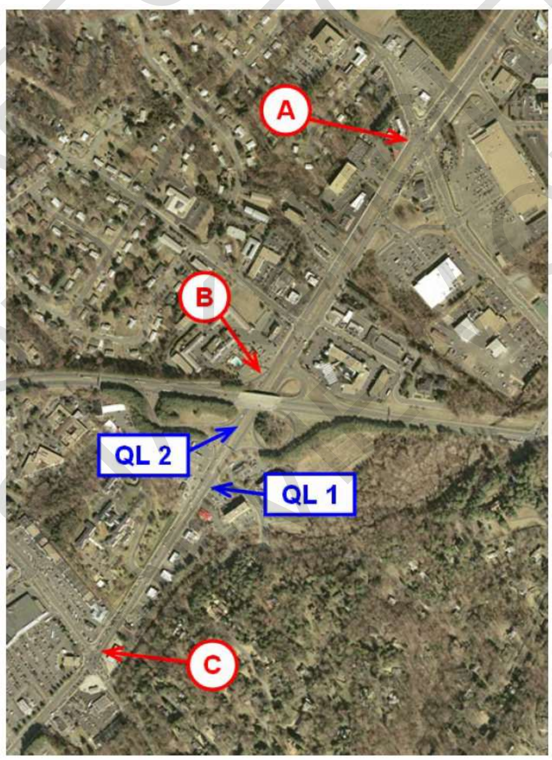


Figure 248. Map. Data collection location map in Charlottesville.

Traffic volume and heavy vehicle percentage were measured from every intersection and ramp by using an electronic data collection device (traffic data collector [TDC] TDC-12) as well as a data collection sheet. Due to the large number of vehicles entering Route 29 South at Hydraulic

Road (northern entrance to the network), traffic counts and other required data were collected by using Autoscope® in a smart travel van (STV). Four video cameras with synchronized clocks were positioned on the entry point and exit point for each section (two for location A and one for location B and C) and recorded a license plate number. Another two surveyors were located at two approaches at one intersection (surveyor location 1 [QL 1], surveyor location 2 [QL 2]) to collect the maximum queue length data by counting the number of stopped vehicles at the end of each red phase.

As mentioned previously, traffic volume data of each intersection were collected using an electronic data collection device TDC-12. After collecting all data collection devices from surveyors, they were connected to the computer and the traffic volume data in the devices were automatically transferred to computer in Microsoft® Excel format. Also, aerial photos were obtained from the City of Charlottesville to be used as a background image for the network building process in all different simulation models.

Videotapes on different days recorded from three different locations were reviewed and the license plate number of each vehicle was manually recorded and matched to extract the travel time of two sections. Each travel time was determined by subtracting the time when a vehicle passed the subject entry point from the time when the vehicle passed the subject exit point.

License plate numbers and recorded times were manually matched. Maximum queue length data were manually collected by writing down the maximum queue length at the end of every signal cycle. Data sheets were manually reviewed and typed in an Excel worksheet.

Regarding the Reno case study (2012), existing demand volumes and signal timings were obtained from NDOT, as illustrated previously in
Figure 245.

Description of Speed Measurement Method

Time-mean speed and space-mean speed have different definitions and measurement methods. TSSM Chapter 6. discusses how time-mean speed can be directly measured with pairs of loop detectors, microwave devices, and other sensors. Conversely, space-mean speed can be approximated by identifying a vehicle at two sequential positions along a highway. A vehicle's speed over a segment of highway is measured where its position is recorded at the two locations separated by distance *d*, and two times represented by travel time *t*, so its average speed is calculated as *d/t*. The three case studies in this document appear to be focused more on travel times than speeds. Although posted speed limits were obtained, none of them discuss field-measured speeds or calibrated speeds.

Collection of Vehicle Type Classifications

Discussions of vehicle type classification were not found in the Boise case study (2013) or Reno case study (2012). In the Charlottesville case study (2006), heavy vehicle percentages were obtained in the field as described in the previous section, Origins of Data Used in the Analysis.

Pedestrian Volume Counts

Pedestrian counts were not discussed or documented in any of the three case study reports described in this document. However, two of the case study reports addressed the topic of pedestrian signal timings.

Obtaining Signal Timings from the Field or the Department of Transportation

For the Boise study (2013), signal timings were obtained from the countywide Synchro model. In the Charlottesville case study (2006), signal timings were obtained from VDOT. In the Reno case study (2012), signal timings were obtained from NDOT, as illustrated in Figure 245.

Collection of Calibration Data

In the Boise study (2013), field data collection of volumes and travel times for calibration was described in the previous section, Origins of Data Used in the Analysis. Calibration was not explicitly discussed in the Reno case study report (2012), although volumes and signal timings were obtained from NDOT. The Charlottesville case study report (2006) described the following collection of calibration data:

The travel times of the following two sections were used as a performance measure for the calibration process:

- Section 1: From Hydraulic Road (A) to Barracks Road (C).
- Section 2: From Hydraulic Road (A) to Ramp 1 (B).

It should be noted the travel time of section 1 depicts the travel times of vehicles using the southbound left-most lane on section 1, and that of section 2 is the travel times of vehicles

traveling in the southbound right-most lane on the section. The travel times collected from the two sections are presented in Table 45 and

A. Subfigure of field travel time—section 1.

B. Subfigure of field travel time—section 2.

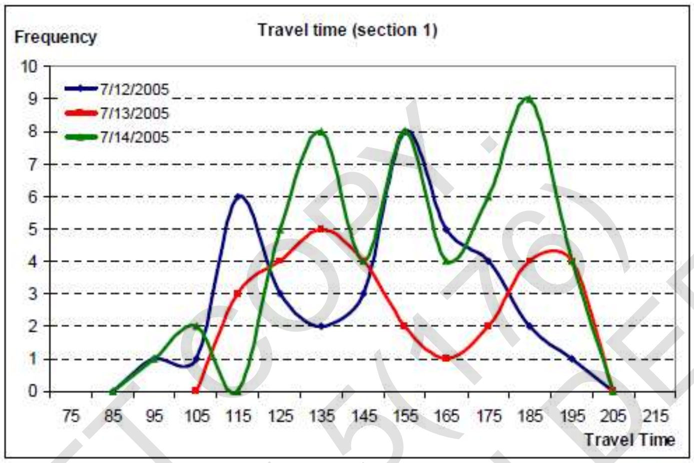
Figure **249**.

Table 45. Field-measured travel times from Charlottesville.

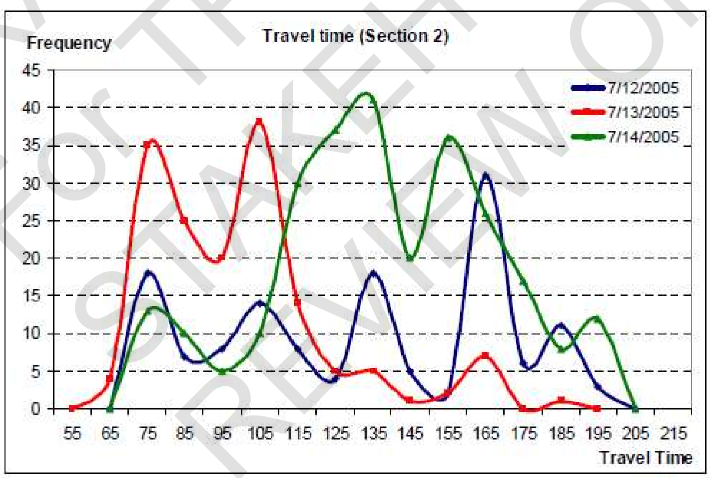
Section	Date	Number of Vehicles	Travel Time (sec)		
			Mean	Standard Deviation	Average
Section 1	7/12/05	36	147.9	26.0	150.3
	7/13/05	29	152.7	28.0	
	7/14/05	51	156.9	27.2	156.9
Section 2	7/12/05	135	132.9	37.5	116.3
	7/13/05	157	99.7	24.4	
	7/14/05	264	138.2	29.8	138.2

Source: Park, Won, and Perfater.

Maximum queue lengths were measured at the northbound approach of Morton Road. (QL 1) and the southbound approach of its overlap intersection (QL 2). The queue lengths were determined by counting the number of vehicles in a queue at the end of the red time for each cycle during the data collection period. The longest queue length was designated as the maximum queue length here. The maximum queue length in the two locations was used for validation. Table 46 shows the summary of surveyed queue lengths at the two locations.



A. Subfigure of field travel time—section 1.



B. Subfigure of field travel time—section 2.

Figure 249. Charts. Field-measured travel time variability from Charlottesville.

Table 46. Field-measured queue lengths from Charlottesville.

Location	Date	Queue Length, vehicles		
		Average	Standard Deviation	Max
Location 1	7/12/05	13	7	24
	7/13/05	10	5	24
	7/14/05	13	7	26
Location 2	7/12/05	9	4	15
	7/13/05	7	4	16
	7/14/05	11	5	23

Source: Center for Advanced Transportation Education and Research.

Collection of Crash Data and Roadway Alignments

The traffic simulation case studies for surface arterials were mobility-oriented studies; they did not mention safety-oriented data, such as crashes. Note that modern mobility studies cannot be considered robust without some analysis of travel time reliability, which would certainly be affected by crash data. Roadway alignment was also not mentioned.

Use of Probe Data, National Performance Management Research Data Set, Vehicle Trajectories, Aerial Imagery

Use of probe data and/or national performance management research data set (NPMRDS) data in traffic studies has exponentially increased in recent years. This document's case studies were conducted in 2013, 2006, and 2012, and did not take advantage of the probe data resource. Use of vehicle trajectory data has exponentially increased in the research world; but, practitioner use is lagging, and these case studies do not mention the trajectory resource. Few tools like the trajectory processor have been made available to practitioners. Aerial imagery is much more common and can be viewed towards the front of this document for all three case studies.

Application of Data Processing Principles (Aggregation, Fusion, Imputation)

The Boise report (2013) contained extensive data analysis via tables and graphs but did not explicitly describe any data processing principles. The Charlottesville report (2006) had a very helpful subsection entitled Data Reduction, whose contents are repeated here:

> *Traffic volume data of each intersection were collected using an electronic data collection device (TDC-12 Traffic Data Collector). After collecting all data collection devices from surveyors, they were connected to the computer and the traffic volume data in the devices were automatically transferred to computer in Excel format. Also, aerial*

photos were obtained from the City of Charlottesville to be used as a background image for the network building process in all different simulation models. Videotapes on different days recorded from three different locations were reviewed and the license plate number of each vehicle was manually recorded and matched to extract the travel time of two sections. Each travel time was determined by subtracting the time when a vehicle passed the subject entry point from the time when the vehicle passed the subject exit point. License plate numbers and times recorded were manually matched. Maximum queue length data were collected manually by writing down the maximum queue length at the end of every signal cycle. Data sheets were manually reviewed and typed in an Excel worksheet.

In the Reno case study (2012), O-D volumes were provided by NDOT for the AM and PM peak periods. The AM counts are shown in Table 47. Interestingly, the research team chose to estimate midday peak period O-D volumes by interpolating between AM and PM peak period counts. Interpolation can be considered one example of a data processing principle, whose use should be documented in a simulation study report.

Table 47. 2015 morning peak hour traffic demands in Reno (vehicles per hour).

O-D	1	2	3	4	Total
1	0	141	66	593	800
2	329	0	34	0	363
3	117	98	0	127	342
4	887	0	291	0	1178
Total	1333	239	391	720	2683

Source: Center for Advanced Transportation Education and Research.

Statistical Analysis of Field Data Sample Size

In its appendix on model calibration and validation, the Boise study report (2013) provided extensive, nonstatistical analysis of field data. Statistical analyses were performed when assessing correlation of field data to simulation data, but sample sizes were not addressed.

In the Charlottesville report (2006), a partial statistical analysis of field-collected travel time sample sizes was previously given in Table 45. The analysis could have been extended by developing confidence intervals to assess adequacy of the sample size of collected travel times.

The Reno case study report (2012) did not address field data sample sizes.

13.3 CREATING THE NETWORK

Intelligent Selection of Initialization Fill Time

TSSM Chapter 7. states that many traffic simulation tools offer some sort of input entry for initialization time, warm-up time, fill time, or some similar name. The typical objective of the initialization time is for the simulation model to reach a state in which the number of vehicles entering and exiting the network is roughly equal. This would normally facilitate a more accurate simulation analysis. An intersection, corridor, or network model should have an approximately equal number of vehicles entering or exiting the virtual system, prior to collection of vital performance measures. Failure to achieve this balance may mean vehicles have not had enough time to fill up the system, or may indicate coding errors (e.g., missing links or nodes). If vehicles have not had enough time to fill the system, congestion-related performance measures could be overly optimistic. In other cases, it may mean that conditions were oversaturated during the system initialization. In many cases, accuracy of the simulation results may be compromised.

The Reno case study report (2012) stated that in their simulation study, the warm-up time was 300 seconds, and the simulation time for measuring traffic performance was 3,600 seconds. If this information was missing from the report, it could raise questions or uncertainties in the minds of reviewers. And although the basic warm-up time and simulation time reported above meets that need, it would be even better to document the rationale behind these values to demonstrate they were intelligently selected.

Analyzing the Consequences of Inadequate Study Boundaries

In a traffic simulation, when the spatial limits of the virtual network fail to fully capture the extent of congestion, this can negatively affect certain aspects of an analysis. However, other aspects of the analysis may be unaffected. Engineering judgment can be used to assess the consequences of inadequate study boundaries. For the Boise (2013), Charlottesville (2006), and Reno (2012) arterial case studies, discussions of the chosen spatiotemporal boundaries were provided. Documentation of these boundaries was included near the beginning of this arterial case study document, in 0.

In the Boise report (2013), there was no evidence to indicate the chosen study boundaries were inadequate. In fact, their virtual network surrounding the reconstructed bridge was fairly large, as shown in the earlier

Figure 240**. By contrast, the Charlottesville analysts clearly considered and documented the impacts of their chosen boundaries: "During the preliminary site visit as well as data collection, a unique phenomenon was observed by the research team. As mentioned, Route 250 intersects with Route 29 within this network and Route 250 shares its roadway with the Route 29 bypass. Because of its high speed and capacity, a large percentage of vehicles on Route 29 south try to diverge to Route 250 / Route 29 bypass through a one-lane ramp and it forms quite a long queue on the right most-side lane that requires additional attention when building a network." (2006, 81)**

B. Subfigure of **dynamic traffic assignment (DTA) model of the study area.**

Figure **244** shows the extent of queue lengths observed in the field. Finally, the Reno case study report (2012) implied that spatial boundaries did not extend beyond the interchange. In a DDI study such as this, some analysis and documentation of possible nearby hidden bottlenecks is advisable, as discussed toward the beginning of this document.

Informed Selection of Overall Data Entry Method

Chapter 7. alluded to the advantages and disadvantages of various traffic network creation methods. These advantages and disadvantages are often considered during the traffic analysis tool selection process but should also be documented for future reviewers. The Boise report (2013) mentioned the decision to import data from the COMPASS regional demand model and Synchro traffic operations model. The report did not touch on advantages and disadvantages, although importing large quantities of data would presumably save a substantial amount of resources related to the data entry process. Regarding the Charlottesville case study (2006) and Reno case study (2012), although the origins of their data were discussed earlier in this document, their respective reports did not address their chosen methods of data entry.

Use of Pre-Processors and/or OpenStreetMap™

Some pre-processor tools can generate traffic network data sets that can be imported and processed by a simulation tool. In other cases, a simulation tool may have the ability to import data from third-party sources and/or formats. Some simulation tools can import detailed data from geographic information systems (GIS) such as OpenStreetMap™.

The Boise report (2013) mentioned the decision to import data from the COMPASS regional demand model and Synchro traffic operations model. This represents one example of a simulation tool with the ability to import data from third-party sources and/or formats. O-D adjustments were then executed in Emme to improve the fit of initial demand to the count data. The Charlottesville report (2006) mentioned use of the QueensOD method and/or tool (Van Aerde et al. 2003) to estimate (O-D) demands. This does not represent importing data from third-party sources, but rather a repurposing and reorganizing of traffic demand data already available to the analyst. The Reno case study report (2012) did not mention use of third-party tools or data sources in the manner described by this subsection.

Entry of Demands Instead of Throughput Volumes

Although the terms demand and volume are often (and sometimes inappropriately) used interchangeably, it is critical the user enters demands instead of volumes. This is because volumes observed in the field are only reflective of vehicles discharged, whereas demands indicate how many vehicles are attempting to use the system. This does not preclude the later use of throughput volumes for calibration and validation purposes.

Due to growing traffic congestion and improving traffic analysis tools in the past few decades, more transportation professionals and academics are recognizing the importance of distinguishing between demands and volumes. The Charlottesville report (2006) is relatively old; it uses the term volume exclusively. By contrast, the newer Boise report (2013) and Reno report (2012) both use demand in addition to volume and throughput.

Analysis of Origin-Destination Demands

Naturally, the most accurate traffic simulations provide realistic O-D travel paths. The Boise report (2013) discussed use of O-D demands from the COMPASS regional demand model, plus the subsequent O-D adjustments in Emme. The Charlottesville report (2006) discussed use of the QueensOD method and/or tool. The Reno case study (2012) described O-D demands provided by NDOT, as previously shown in Table 47.

Application of a Specific Volume Balancing Method

For various reasons, upstream and downstream traffic volume counts obtained from the field are not always consistent with one another.

Figure **250** illustrates an example of this phenomenon.

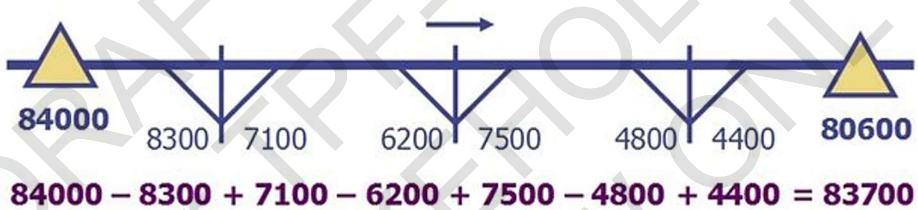


Figure 250. Diagram. Unbalanced upstream and downstream volumes.

Severely inconsistent counts are likely to produce inaccurate simulation results. Volume balancing methods are available to mitigate this problem. When such methods are used, they should be documented for the reviewer's benefit and understanding. Although the three arterial case studies in this document all performed calibration of simulated volumes to field-measured volumes, their reports did not mention use of a specific volume balancing method.

Application of Conditional Turn Movements

O-D demands at the local level are sometimes called conditional turn movements. These optional input specifications, which are primarily applicable to microscopic traffic simulation, can be used to make the simulation more realistic. The Boise case study (2013) used networkwide O-D data, but involved mesoscopic simulation. The Charlottesville case study (2006) and Reno case study (2012) involved microscopic simulation and O-D data, but made no mention of conditional turn movements.

Application of Data Entry Methods to Achieve Pre-Positioning and Custom Lane Utilization

Lane utilization expresses the proportion of vehicles using each lane. On multilane freeways and surface arterials, simulation tools usually assume equal or slightly unequal lane utilization unless the user indicates otherwise, or unless vehicles unequally distribute themselves to achieve user-specified downstream turn movement percentages. In macroscopic or mesoscopic simulation tools, unequal lane utilizations may be calibrated using inputs such as the lane utilization adjustment factor or single highest lane volume, and pre-positioning may not be relevant. Regarding microscopic simulation, the data entry method for achieving unequal lane utilizations may depend on whether the tool in question follows a link-node or a link-connector architecture. Any such data entry methods should be documented for the reviewer's benefit and understanding. In the Charlottesville case study (2006), the initial simulation network was built without any pre-positioning or custom lane utilizations. This led to the following outcomes described in the VDOT handbook:

> *Initially, the network was built without any special considerations on the long queue on the right-most side lane on Rte. 29 south. As a result, the vehicles that needed to diverge to the ramp toward Rte. 250 west did not generate a long queue but blocked the whole section because some of the vehicles changed their lane right in front of the diverging point. As a result, everywhere in the network was congested, which is not realistic at all. So, our research team assumed that the vehicles select their lane before they enter the network. To create that kind of condition, vehicle entry points for two different destination groups were separated with different lane groups:*
>
> - *Vehicle group that makes a right turn up to on-ramp toward Rte. 250 west.*
> - *Other vehicles.*

The concept of separate entries by lane is shown in Figure 251. The Boise case study report (2013) and Reno case study report (2012) did not broach the subject of this subsection.

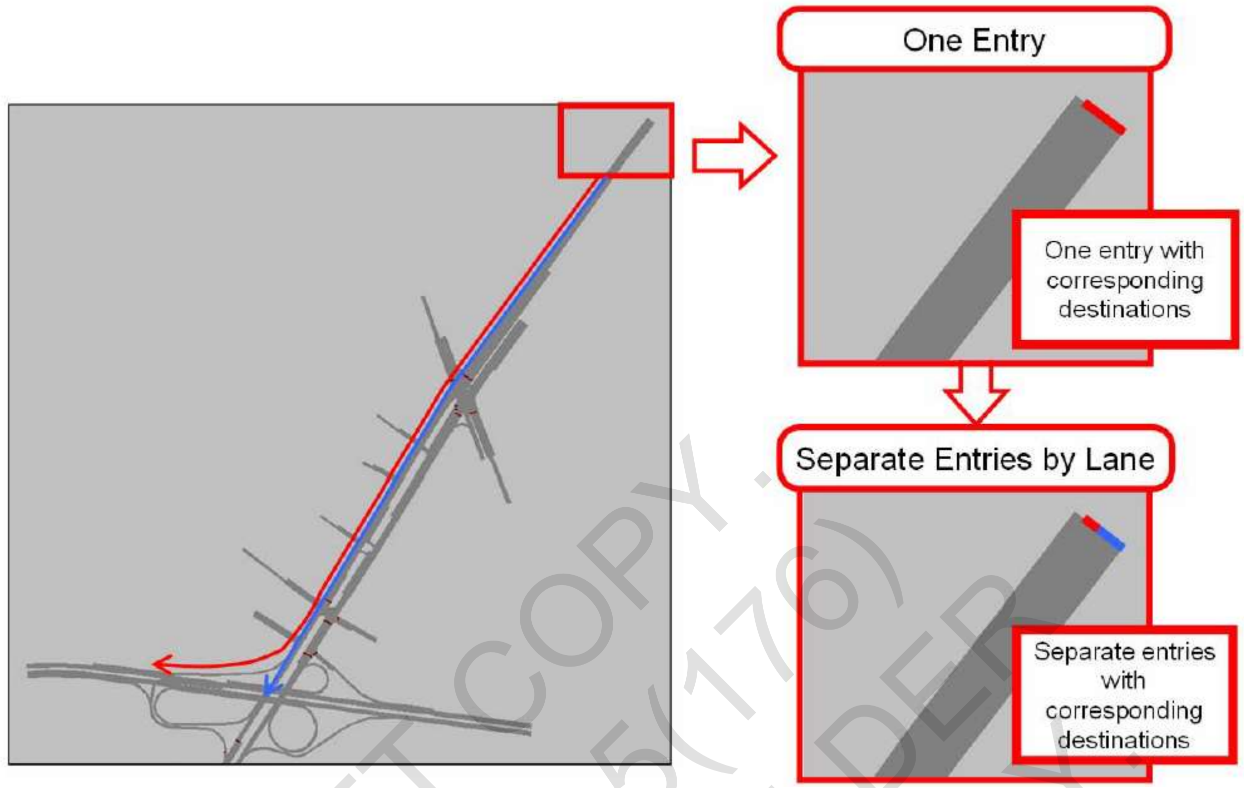


Figure 251. Illustration. Use of pre-positioning in the Charlottesville case study.

Entry of Detector (Sensor) Data

Detectors are used on both freeways and arterials. Passage and presence detectors are used to control the traffic signals and/or ramp meter signals. Surveillance detectors can be used to collect additional traffic performance measures. The arterial case study reports did not address this topic.

Entry of Freeway Warning Sign Distances

Freeway warning signs were not applicable to the arterial case studies.

Entry of Operating Condition Data

Traditional simulation projects have focused on analyzing typical peak hour traffic conditions. The danger of this approach is that when decisions are made based on most likely (i.e., 50th percentile) outcomes, these decisions may fail badly when the inevitable 75th–95th percentile conditions materialize. Consequently, some traditional projects have focused on analyzing the 30th-highest-hour demand volume scenario. While this is probably a step in the right direction, the analysis of only a single scenario still leaves engineers largely in the dark about how to manage risk and reliability because they cannot visualize the full range of possible outcomes. Different combinations of demand, weather, incidents, work zones, special events, and driver behaviors ultimately produce a wide range of outcomes. These outcomes will only be captured by simulation if a variety of input scenarios are considered. The arterial case study reports in this document did not address this topic.

Selection of Key Input Parameters for Calibration

According to the Boise case study report (2013), the simulated traffic flow from the DTA model versus the existing traffic counts showed an R^2 statistic of 0.91, indicating satisfactory calibration. The DTA model traffic flows at the Boise River bridge crossing were also compared with the existing traffic counts, and showed a good fit with an R^2 statistic of 0.97. The DTA model was validated utilizing travel time data collected by COMPASS on major roadways for the PM peak period. Modeled travel times were compared against observed travel times. As with model calibration, the goal is to achieve a minimum goodness of fit. The modeled travel times versus observed travel times show an R^2 statistic is 0.91, which exceeds the minimum for DTA model validation.

The Charlottesville case study (2006) was contained within a calibration guidance handbook. For this reason, 28 input parameters were chosen to demonstrate the calibration process. Two additional input calibration parameters were set to local default values, based on the analyst's familiarity with local conditions.

The Reno case study (2012) primarily focused on matching the peak period O-D counts and signal timings provided by NDOT.

Identification of Key Network Locations for Calibration

In the Boise case study (2013), input data were imported from the travel demand model and traffic analysis model as described earlier. The approach to calibration involved a thorough analysis of simulation results, which would identify unusual results in certain network locations. This process will be described further in the upcoming section, Verification, Calibration, and Validation (VCV).
In the Charlottesville case study (2006), the locations chosen for calibration were classified as Section 1: From Hydraulic Road to Barracks Road, and Section 2: From Hydraulic Road to Ramp 1. However, the rationale behind this choice of locations was not given.

The topic of calibration was not explicitly discussed in the Reno case study report (2012).

13.4 VERIFICATION, CALIBRATION, AND VALIDATION

The inclusion of three arterial case studies was intended to demonstrate the different ways in which real-world simulation analyses can illustrate TSSM principles. Another benefit was the ability to provide at least one tangible example for a higher percentage of TSSM principles, since few (if any) simulation reports exhibit 100 percent of the 65 principles selected for discussion in this document. This is particularly true of the case studies for TSSM Chapter 8. because the Reno DDI report (2012) did not focus very much on VCV principles beyond: a) focusing on a limited number of critical O-D pairs and b) having a second set of eyes review the data. Instead, the Reno study (2012) obtained volume and timing data from the DOT, and then focused on the relative benefits of various signal phasing plans. As such, most subsections within this VCV section of the arterial case studies document will not mention the Reno case study.

Analysis of Why Results are Incorrect before Deciding Next Step

The Boise case study report (2013) included a helpful 12-page section that provided many of the details behind their VCV effort. Although most pages within this section describe successful comparisons, which found that the simulation results had excellent correlation with field-measured data, some of the initial comparisons found discrepancies that needed to be addressed. In each of these cases, the report authors provided their analyses of why the results were incorrect, along with a subsequent explanation of how the discrepancy was rectified:

A path analysis for the O-D pair [53,178] shown in

> *Figure 252 reveals that almost half of the vehicles (46%) are taking a longer path to avoid the left turn at the node marked in blue (Capitol/Front St.).*

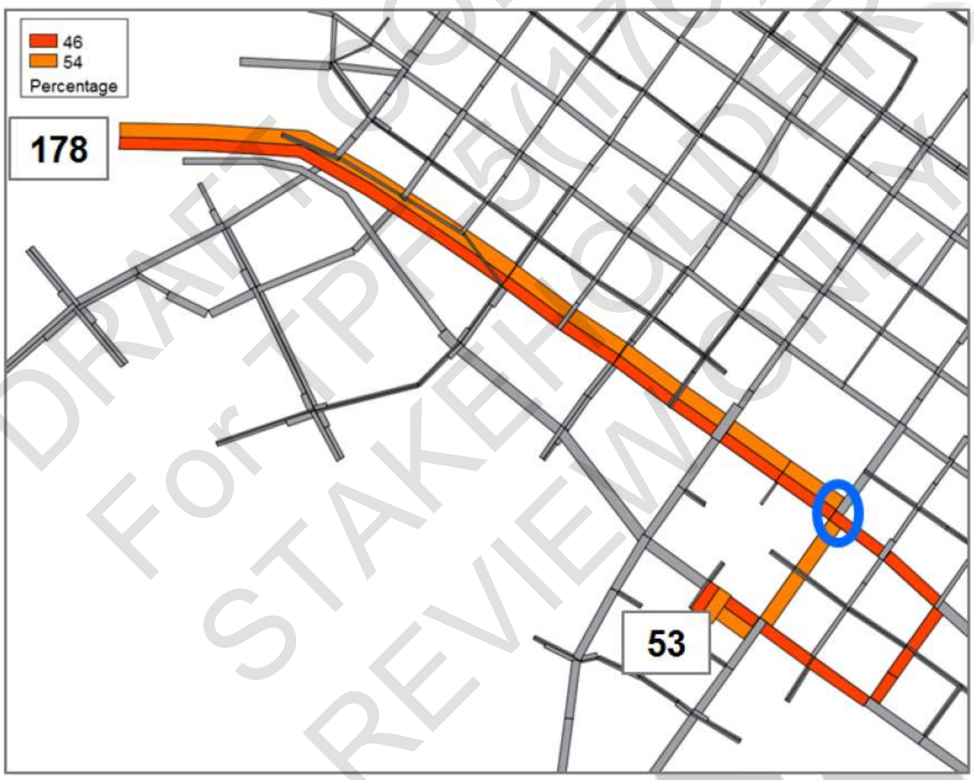


Figure 252. Map. Path analysis for origin-destination (O-D) pair.

> *This indicated that there might be a network coding issue at that node. The theoretical capacity of that turn, considering the saturation flow and effective green ratio, was coded in Synchro as 577 veh/hr, whereas the observed count was almost double at 930 veh/hr. Note in Figure 253 that the modeled flow of that turn was 577 veh/hr exactly. Fixing the coding error eliminated the problem, allowing the flow of the turn to equal the count.*

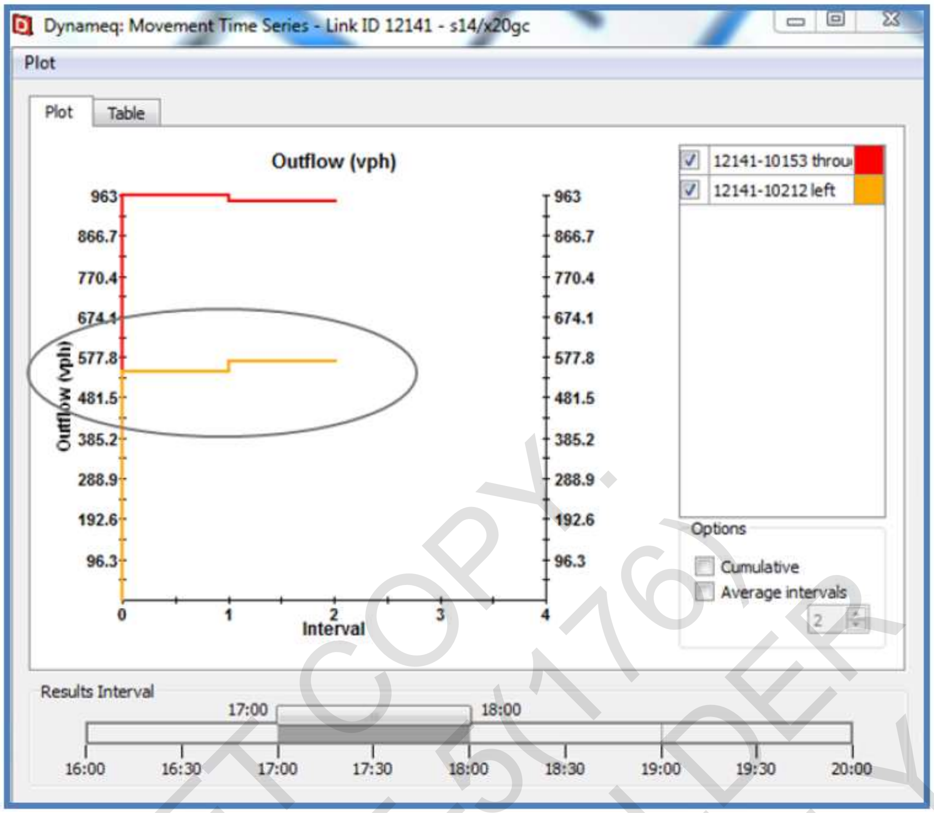


Figure 253. Screenshot. Movement time series for link 12141.

An example of a demand issue was found on 6th Street between Hays Street and W State St. The difference of 440 veh/hr in the counts as shown in the right side of

Figure 254 cannot be explained by the demand of the centroids that connect to the network between the counts locations, as shown in the left side of the figure. Since the counts were found to be consistent with the counts of adjacent intersections, the discrepancy can be attributed to a lack of demand. In this case, the demand of these centroids was increased.

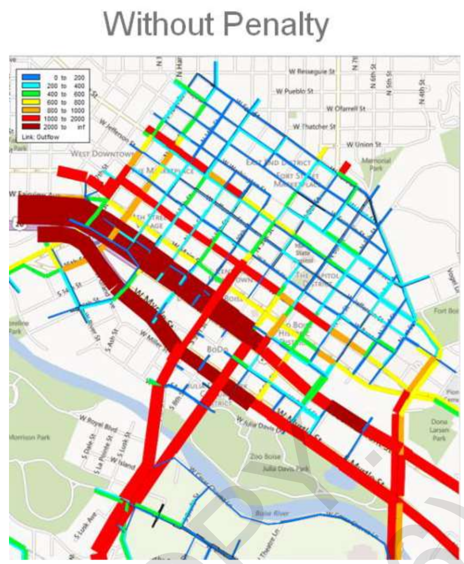
Figure 254. Map. Inconsistent counts between adjacent intersections.

During the calibration it was found that local streets attract too many trips as they bypass congestion on collectors and arterials. To reduce the attractiveness of these local streets, a penalty was added to the cost of using them; see

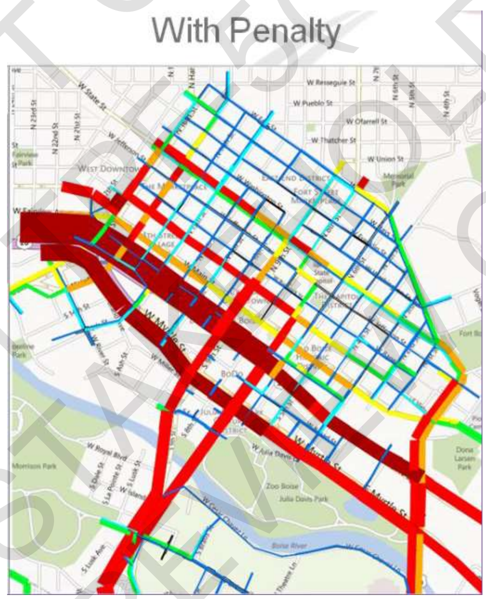
A. Subfigure of traffic assignment without penalty.

B. Subfigure of traffic assignment with penalty.

Figure **255**.



A. Subfigure of traffic assignment without penalty.

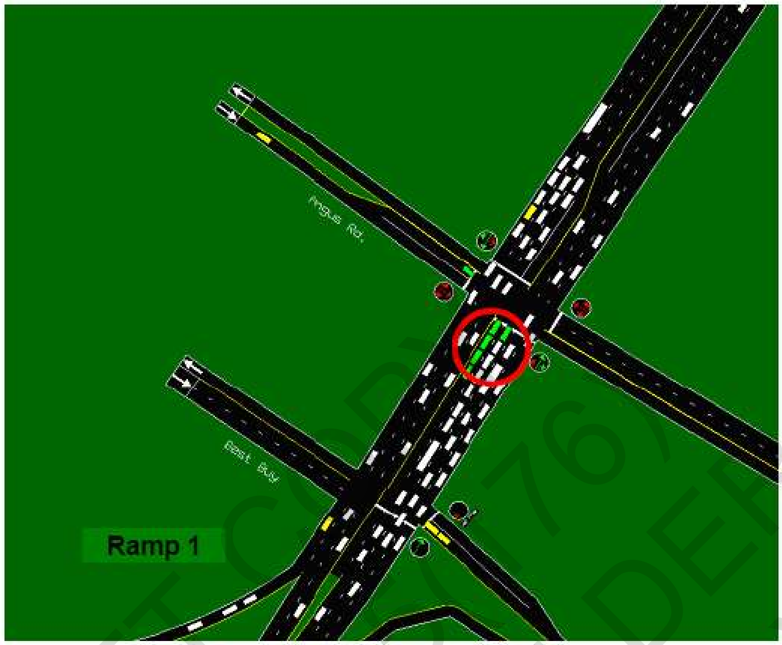


B. Subfigure of traffic assignment with penalty.

Figure 255. Maps. Traffic assignment results with and without penalty.

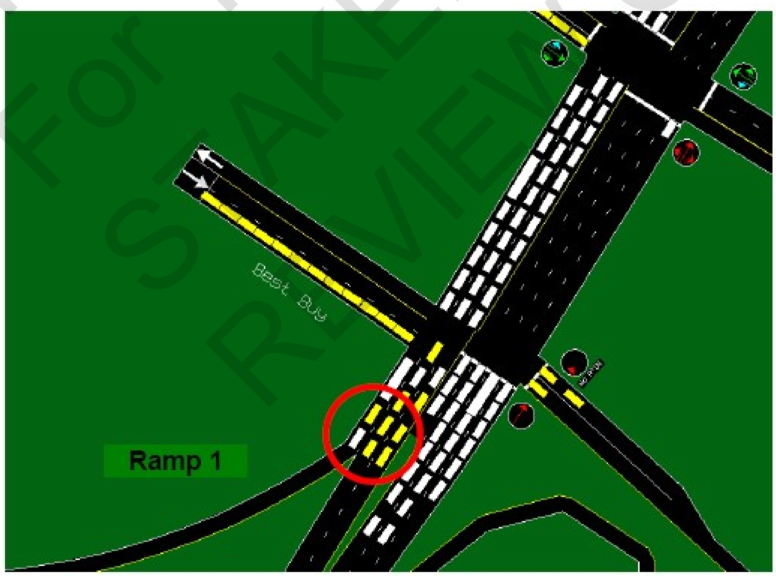
The Charlottesville case study report (2006) also provided some analysis of why the initial results were incorrect, along with subsequent explanations of how the discrepancies were rectified:

If the spacing of the route decision points (e.g., intersections, ramps, etc.) is too close, some vehicles cannot change their lane to the destination lanes due to those narrow spacings. When this kind of vehicle appears on the network, it occasionally blocks the lane and hinders the traffic flow.

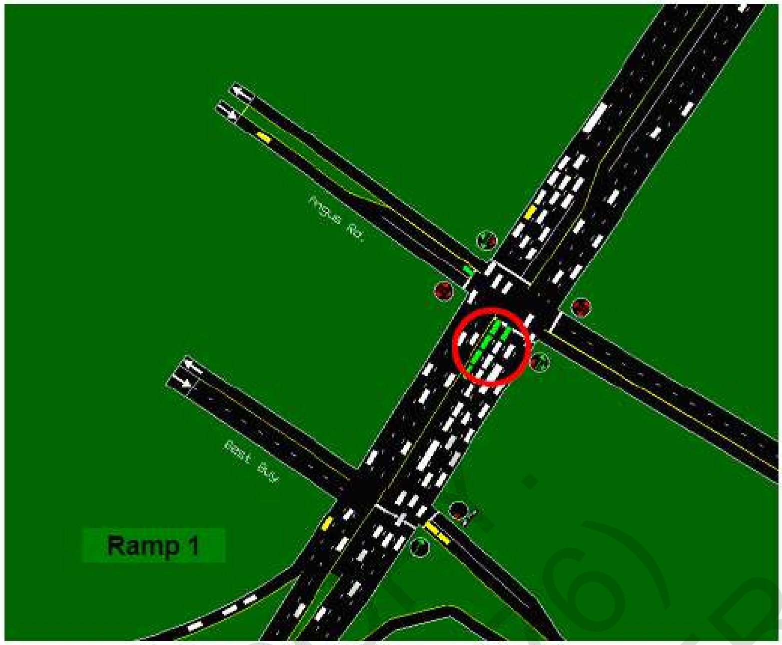


B. Subfigure of lane change difficulty at an intersection downstream from ramp 1.

Figure 256 *shows an example of the appearance of blocking vehicles in the network.*



A. Subfigure of lane change difficulty at ramp 1.



B. Subfigure of lane change difficulty at an intersection downstream from ramp 1.

Figure 256. Illustrations. Examples of lane change difficulty at closely-spaced intersections.

The unique traffic pattern between Hydraulic Rd. and the ramp to Rte. 250 that causes the blocking vehicle appearance can be addressed with the following two methods:

- *Use of Origin-Destination (O-D) assignment in an interchange.*
- *Use of turning percentage.*

The two methods were implemented alternatively, and then the latter was finally selected.

A lot of traffic from Ramp 7 made left-turns at Angus Rd. An unrealistic turning movement scene was [identified]. Therefore, by assigning links located in the area as an interchange, [the simulation] could replicate the actual field traffic pattern. In addition, vehicles turning right were using the right lane whereas the vehicles that wanted to make left-turns at the intersection of Emmet St. and Angus Rd. were using the left lane according to a road sign installed in the middle of the ramp. In practice, the vehicles sitting on the left lane could not freely enter Emmet St. due to high conflict traffic volumes on Emmet St., so they usually waited at the stop bar until the signal head turned to green indication. However, in the simulation, vehicles in the left lane aggressively entered Emmet St. as illustrated in
Figure 257 *(refer to green vehicles at the circle). To avoid these abnormal turning vehicles in the simulation, right-turn-on-red was placed on the approach from Ramp 7.*

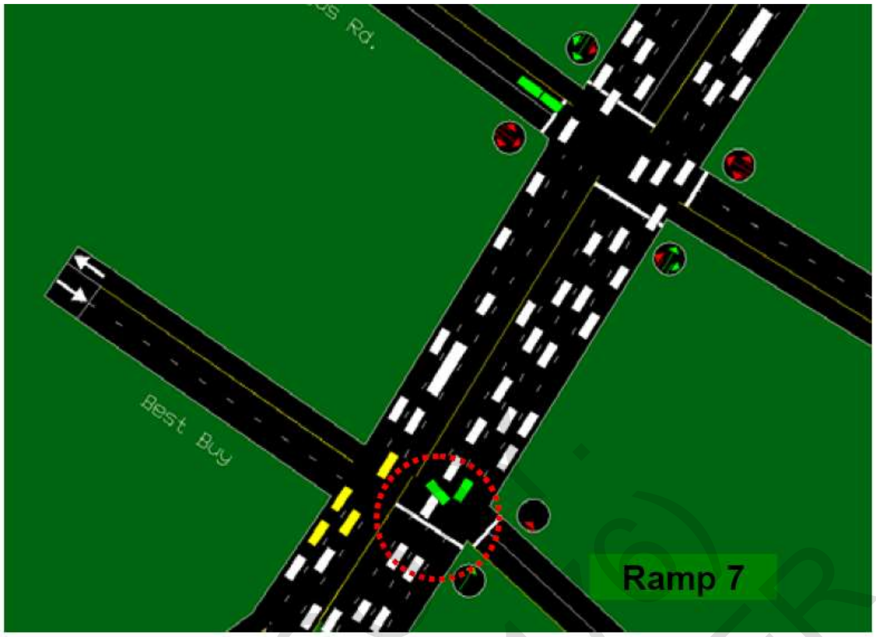


Figure 257. Illustration. Unrealistic turning movements at ramp 7.

Selection of Both Upper and Lower Level Simulation Quality Metrics

The Boise case study report authors appear to have selected one or more specific simulation quality metrics (SQM) for each stage of VCV:

Verification

- Upstream vs. downstream flows.
- Total networkwide trips.

Calibration

- O-D paths.
- Turn movement counts.
- Centroid demands.
- Relative gaps (difference between average and shortest travel time, normalized by shortest travel time, for each O-D pair).
- Networkwide measures (VHT, VHD) over time.

Link flows and speeds (see

- Figure 258).

Validation

- Travel time input data.
- Travel time outputs.

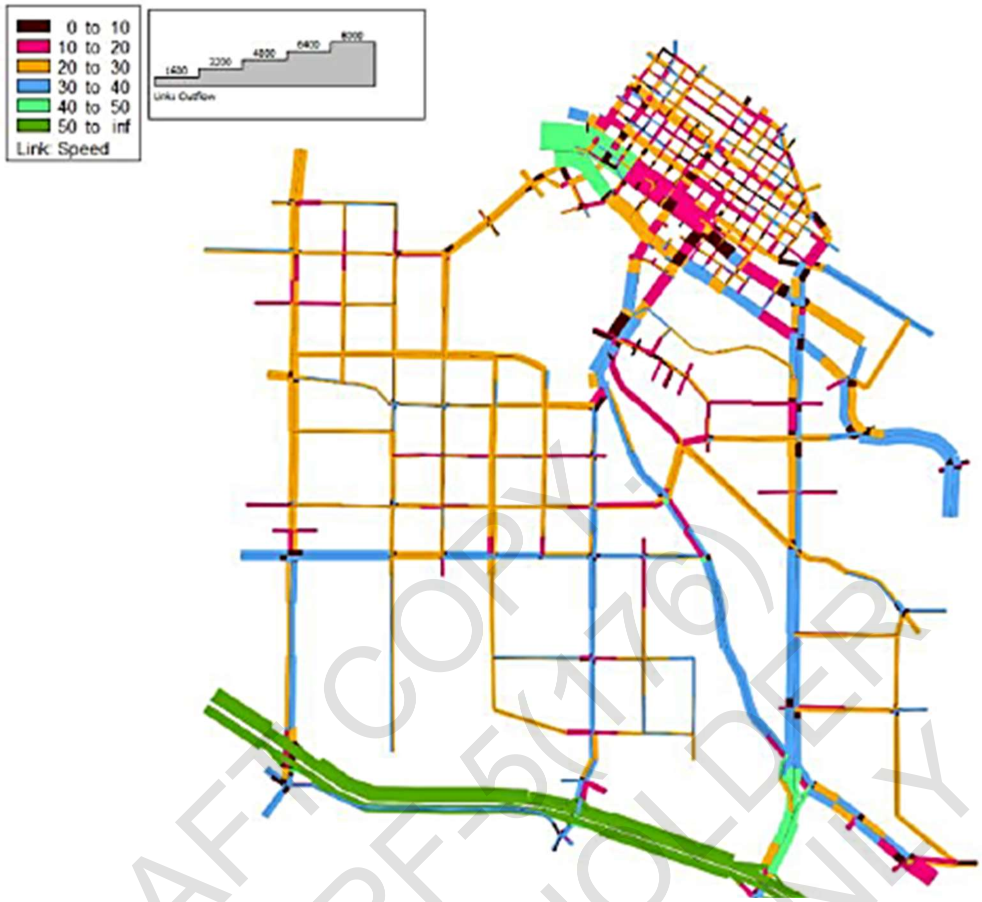


Figure 258. Illustration. Average link speeds in the Boise network (5–6 p.m.).

Chapter 8. also cites the utility of ranking SQMs in terms of importance. The Boise case study report authors place a special emphasis on the traffic counts: “A crucial element before model calibration and validation is to ensure that the empirical data, namely traffic counts and travel times, are valid and consistent. The purpose of calibrating the DTA model is to adjust the subarea model data provided by COMPASS to replicate the existing traffic counts as closely as possible” (2013, 5).

The SQMs discussed above are considered low level. The Boise case study report authors selected the R-squared (R^2) statistic as their high-level SQM: “The goal is to exceed a minimum goodness of fit, represented by the R^2 statistic, which measures how well the model approximates real data” (2013, 5). One example of this is given in Figure 259. The authors state that the R^2 value of 0.91 in Figure 259 is well within the acceptable range for DTA model validation.

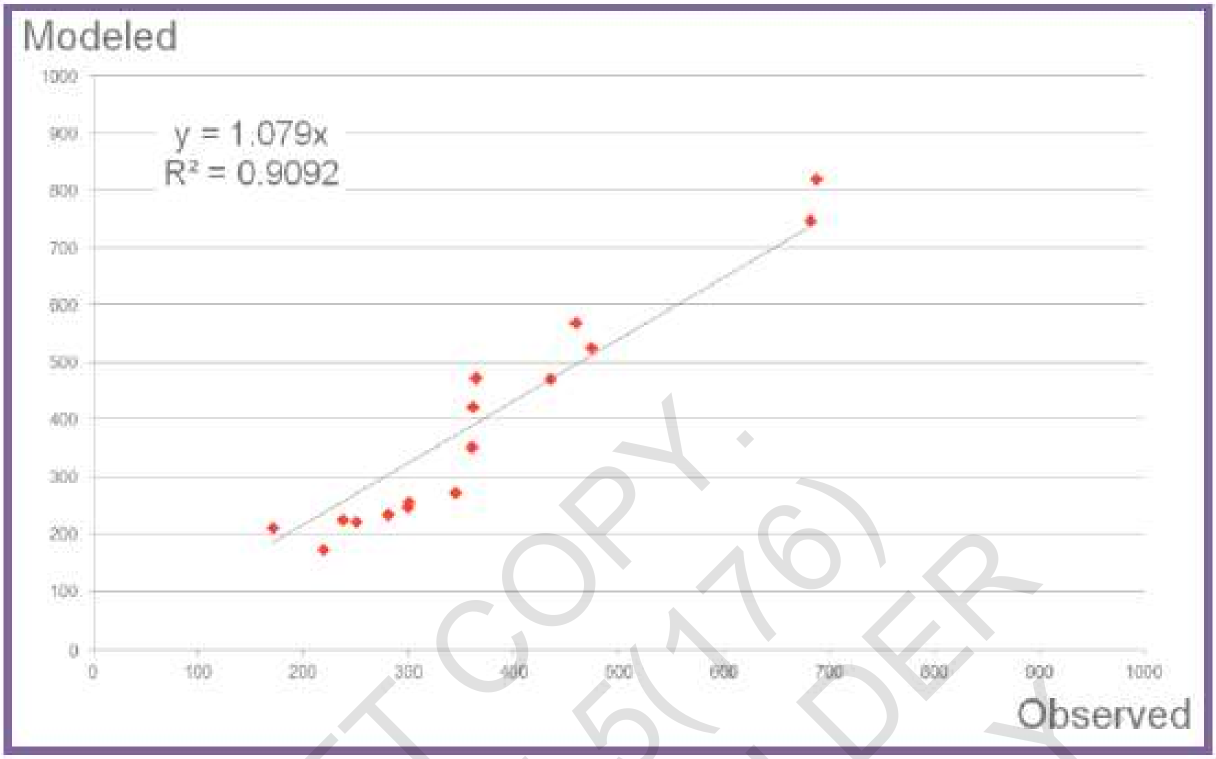


Figure 259. Chart. Modeled versus observed travel times for 16 routes in Boise.

For the Charlottesville case study (2006), the authors chose travel times and queue lengths as their baseline lower level SQMs, as illustrated in the Calibration Data of chapter 6. The equation below was used to compute their high-level SQM:

$$\frac{\sum(\text{Travel Time}_{\textit{Field}} - \text{Travel Time}_{\textit{Simulation (i)}})}{\text{Travel Time}_{\textit{Field}}} \tag{13.1}$$

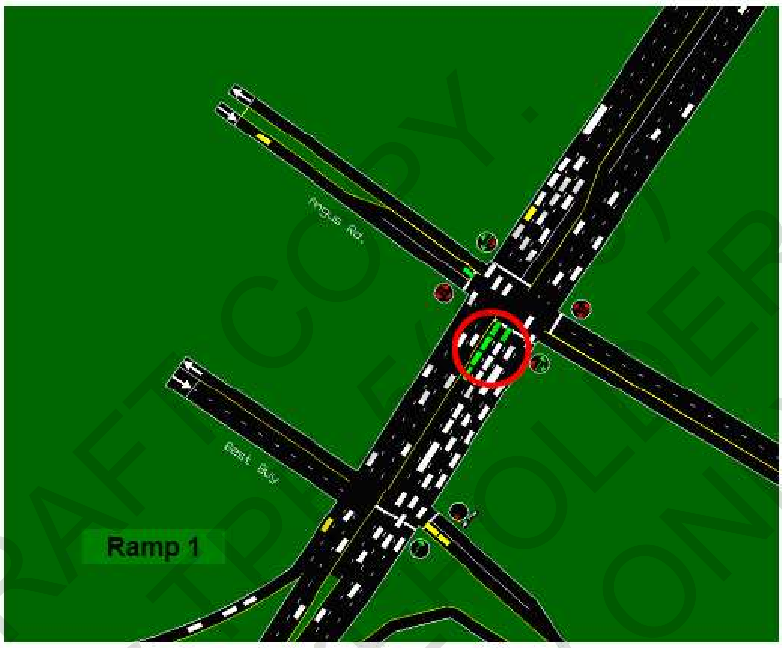
Focusing on a Limited Number of Critical Origin-Destination Pairs and/or Key Problem Spots.

The Boise case study authors performed a special analysis of route 14:

> *Although the overall fit of modeled travel times to observed travel times is within the acceptable range for model validation, a further analysis was done to understand why the modeled times are, on average, higher by 8% than the observed times as indicated by the slope of 1.08. Since the model objective was to assist construction decision for Broadway BR, it was decided to examine route 14 from Broadway & Warm Springs to Broadway & I-84 South ramp. The observed time on this route is 460 seconds while the modeled time was 568 seconds. The biggest discrepancy was on the stretch from Beacon St. to Boise*

Ave. where the observed time was 63.3 seconds and the modeled time was 108 seconds. According to the CMS report, the ideal (free flow) time for this section is 62.19 sec, which means that the observed delay on this section is only 1.1 sec. Given that the there are two signalized intersections at Highland St. and Boise Ave., which are not fully synchronized, the average delay should be higher. A possible explanation for the small observed delay could be that the floating vehicle did not experience delay at the signalized intersections and hence does not represent the average vehicle.

Since the Charlottesville case study (2006) focused on a route with only four signalized intersections, the authors focused on traffic performance at specific points (i.e.,



B. Subfigure of lane change difficulty at an intersection downstream from ramp 1.

Figure 256 and
Figure 257) and sections (i.e., Table 45 and Table 46). Similarly, the Reno case study (2012) focused on a small number of signalized intersections. Calibrated O-D flows were previously shown in Table 47.

Evaluation of Probability Density Functions or Cumulative Density Functions That Describe Distributions of the Performance Metrics

Although the Boise case study report (2013) contained time series graphs that resemble probability density functions (PDF) and cumulative density functions (CDF), it did not contain actual PDFs or CDFs. The Charlottesville case study report (2006) provided several graphs analogous to PDFs, albeit with frequency instead of probability on the *y*-axis. These are easily converted to CDFs if desired. The first set of frequency distribution

functions (FDF), which described field-measured travel time distributions, were previously presented in

A. Subfigure of field travel time—section 1.

B. Subfigure of field travel time—section 2.

Figure **249**. Additional sets of FDFs describing simulated travel time distributions are shown below in

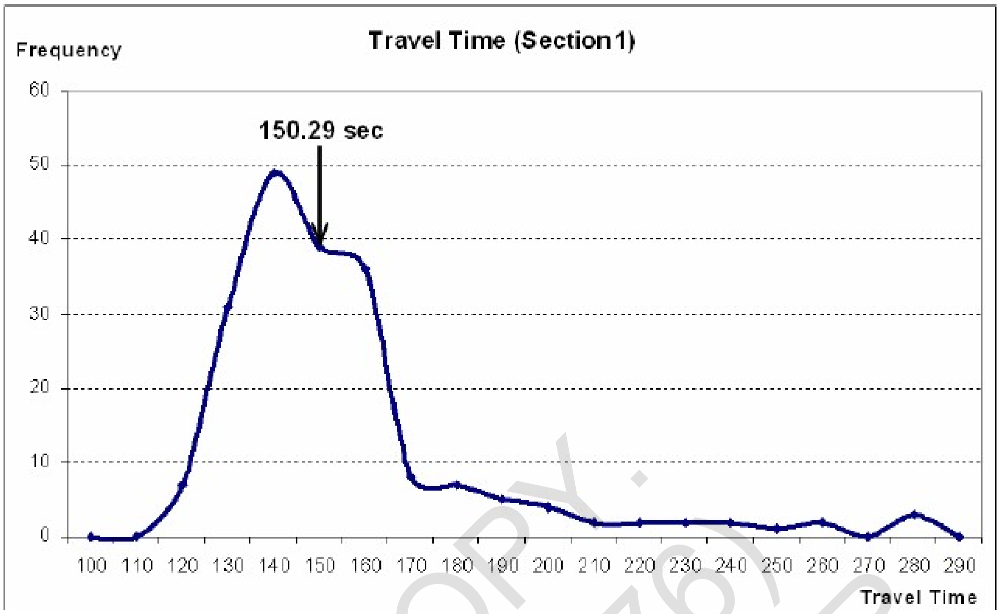
B. Subfigure of travel time of section 2.

Figure 260 **and**

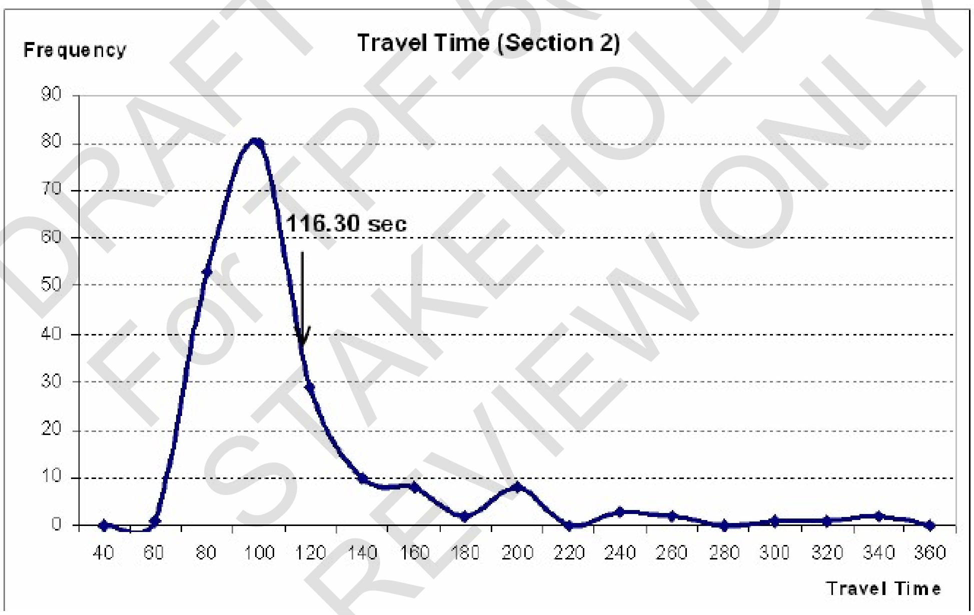
A**. Subfigure of travel time of section 1.**

B**. Subfigure of travel time of section 2.**

Figure **261**. The visually significant differences between these two sets of FDFs demonstrates the potential impact of calibration.

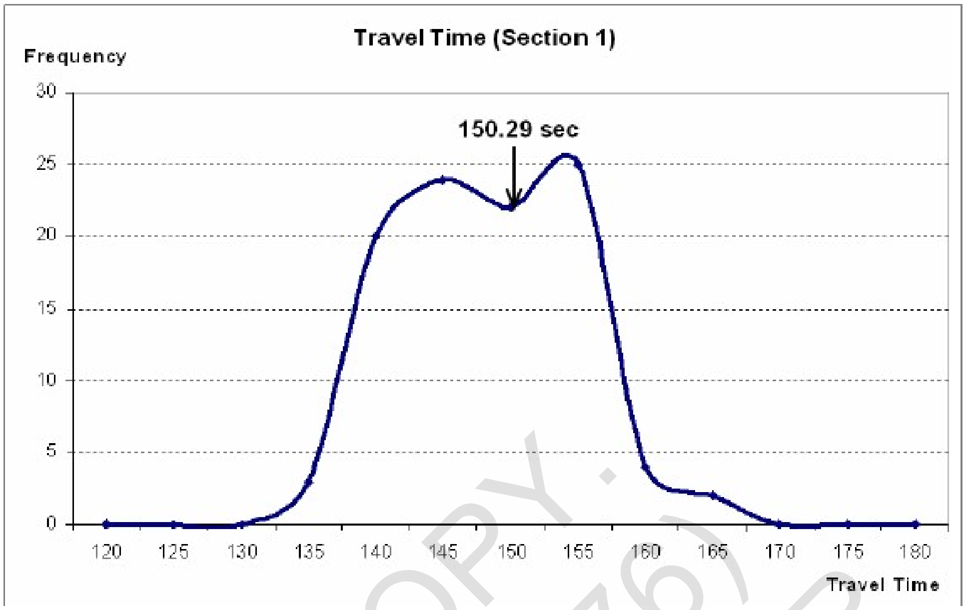


A. Subfigure of travel time of section 1.

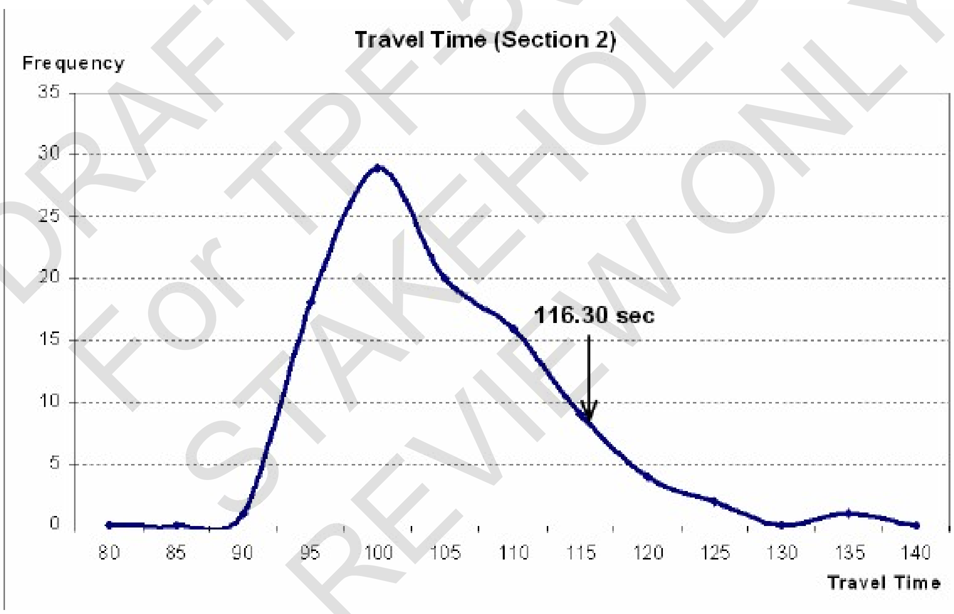


B. Subfigure of travel time of section 2.

Figure 260. Charts. Simulated travel time variability from Charlottesville (uncalibrated).



A. Subfigure of travel time of section 1.



B. Subfigure of travel time of section 2.

Figure 261. Charts. Simulated travel time variability from Charlottesville (calibrated).

Ensuring That Links Perform in Accordance with the Fundamental Diagrams of Traffic Flow

In addition to replicating field-measured performance, a well-calibrated simulation should obey the fundamental laws of traffic.
Figure 262 shows the typical shapes for speed vs. density, speed vs. flow, and flow vs. density plots. Figure 263 contains a real-world example of a speed vs. flow plot. Although the three arterial case study reports did not provide such diagrams in the context of their calibration efforts, generation of the fundamental diagrams is certainly encouraged for important simulation projects.

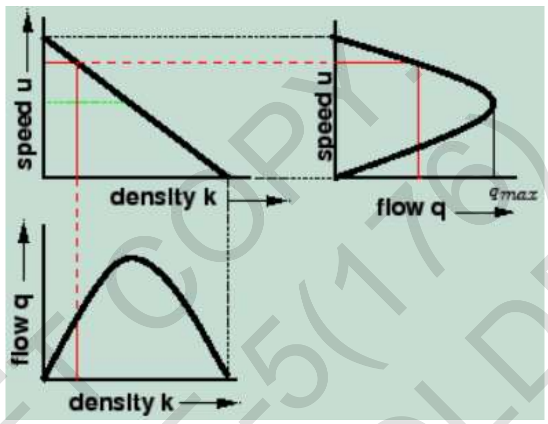
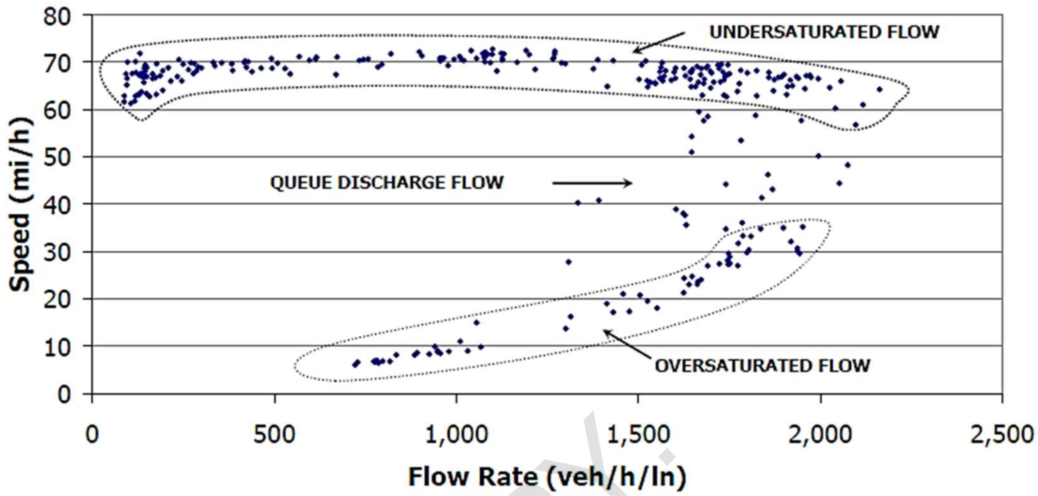


Figure 262. Diagram. Fundamental diagrams of traffic flow.[99]

[99] "Fundamental Relations of Traffic Flow" (web page), SlidePlayer slide 21, accessed October 20, 2019, https://slideplayer.com/slide/12982685/.



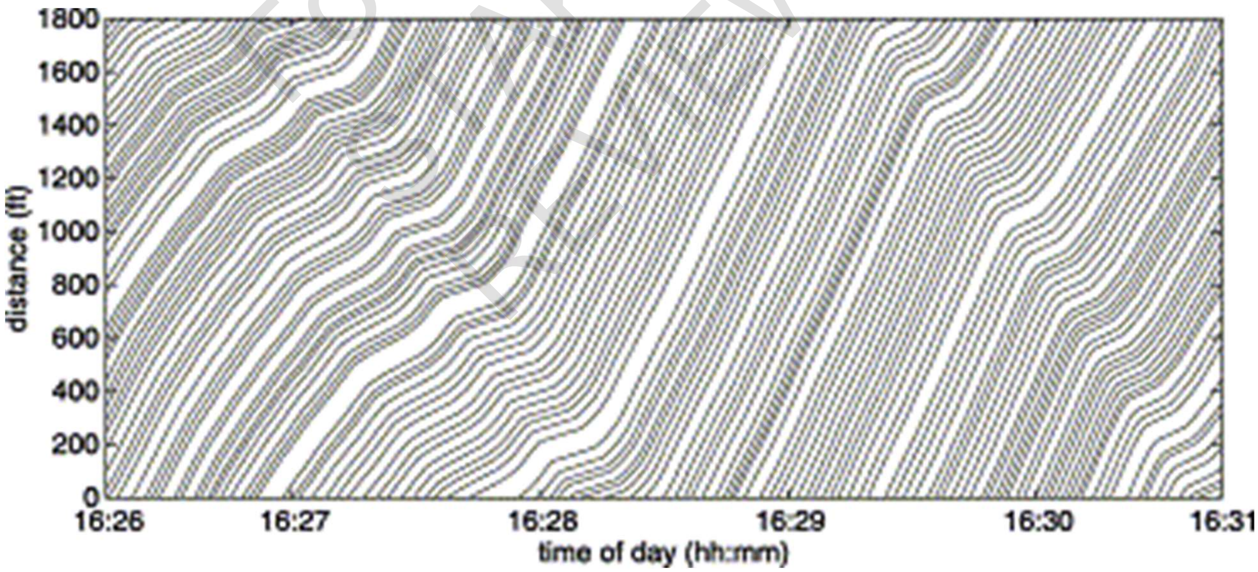
© HCM 6th edition (2016).

Figure 263. Chart. Speed versus flow on I–405, Los Angeles.

Analyzing Reasonableness of Vehicle Trajectories

Simulated vehicle trajectories can reveal unrealistic accelerations, decelerations, shockwaves, cruise speeds, and lane-change behaviors, as shown in © Benjamin Coifman.

Figure 264. In some cases, microscopic simulation models producing seemingly accurate aggregate performance measures have been shown to be inaccurate at the trajectory level. Although the three arterial case study reports did not provide trajectory analysis in the context of their calibration efforts, analysis of the trajectories is certainly encouraged for important simulation projects.



© Benjamin Coifman.

Figure 264. Chart. A vehicle trajectory plot.[100]

Having a Second Set of Eyes to Review the Data

Stakeholders can provide a powerful mechanism for achieving quality control. Some simulation reports can help build confidence in the model results by documenting their stakeholder meetings and interactions. For example, the Reno case study (2012) describes how researchers from the Center for Advanced Transportation Education and Research (CATER) initially met with staff from both the City of Reno and Washoe County's Regional Transportation Commission (RTC).

Dealing with Episodic Demands

In recent years, transportation professionals have begun to realize the significance of annual performance and travel time reliability. This requires explicit analysis of operating conditions including varying traffic demands, weather, incidents, work zones, and special events. The three case study reports summarized in this document approached the traffic study using a more traditional approach, in which the operating conditions and episodic demands were not significantly discussed. However, TSSM authors recommend detailed analysis of these variabilities whenever possible; and when these conditions are analyzed, the methods of doing so should be well documented. Today's traffic modeling experts should become well-versed in the simulation-based travel time reliability analysis methods described in the SHRP2 L04 report (2014), as well as the calibration methods described in FHWA's Traffic Analysis Toolbox Volume III (Wunderlich, Vasudevan, and Wang 2016).

Selection of a Preferred Calibration Sequence

According to TSSM Chapter 8. , input parameters can be calibrated simultaneously or sequentially, with sequentially the recommended approach. One example of a recommended sequence was: (1) driving behavior model parameters, (2) route choice model parameters, (3) the traffic demands, and (4) overall model fine-tuning. The Boise case study report authors developed and provided a helpful analysis flowchart, as shown in Figure **265** below. Beyond this overarching view of the analysis, substeps (and their sequence) within the calibration and validation processes are described qualitatively within the report. The Charlottesville case study calibration approach was more of a simultaneous optimization of all input parameters, although a validation step was included thereafter.

[100] Coifman, Benjamin. "Estimating travel times and vehicle trajectories on freeways using dual loop detectors." Transportation Research Part A: Policy and Practice 36.4 (2002): 351-364.

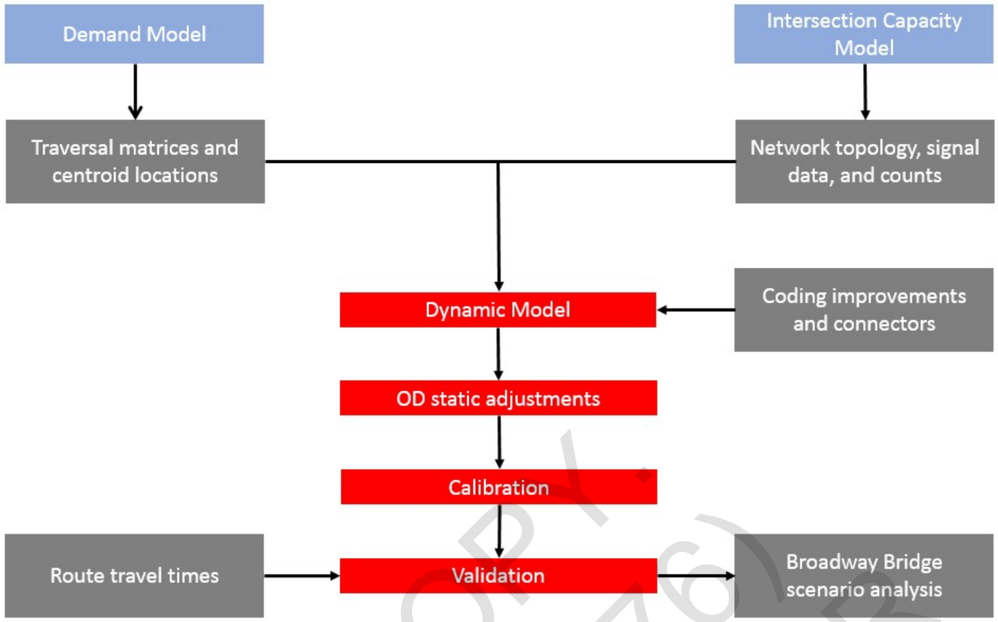


Figure 265. Diagram. Analysis sequence for the Boise case study.

Validate with Separate Data Sets

The Boise DTA model was validated with travel time data for the PM peak hour of 5–6 p.m., collected in floating car surveys.[101] By contrast, the Charlottesville microsimulation model was validated by applying a new and separate data set that was not used in the calibration process:

> *Even though the traffic pattern on that day (Friday) was quite different from the other days (Wednesday and Thursday), the simulated travel time was like the actual travel time. Especially, during the day, there were very long queues in the right-most lane from Hydraulic Rd. to the first entrance ramp to Rte. 250 and thus the travel time of Section 2 shows a relatively bigger discrepancy than that of Section 1. Also, the sizes of maximum queues are not very well matched.*

Table 48 shows the result of the validation procedure.

Table 48. Charlottesville case study validation results.

[101] http://www.compassidaho.org/documents/prodserv/reports/2012CMSReport.pdf.

MOEs	Standard Deviation	Actual Values
Travel time of Section 1, seconds	117.7 (27.8)	138.2
Travel time of Section 2, seconds	158.8 (14.7)	157.0
Maximum queue length at SB approach at Morton Rd. vehicles	17 (1.9)	23
Maximum queue length at NB approach at Morton Rd. vehicles	22 (9.6)	26

Source: Park, Won, and Perfater.

Review Best-Case and Worst-Case Scenario Outcomes

As previously discussed in the subsection on episodic demands, a comprehensive review of varying operating conditions (e.g., weather, incidents, work zones) and episodic demands is strongly encouraged in this day and age. Such a review would naturally lead to discovery of best-case and worst-case scenario outcomes that could potentially materialize throughout the year. Although the arterial case study reports highlighted in this document did not provide these comprehensive reviews, they did obtain cursory glimpses of multiple outcomes.

For example, the Boise case study examined two bridge closure options (i.e., full closure and partial closure), and introduced a traffic assignment penalty to "reduce the attractiveness of local streets" (2013, 10). The impact of this traffic assignment penalty was visualized previously in

A. Subfigure of traffic assignment without penalty.

B. Subfigure of traffic assignment with penalty.

Figure **255**. Next, the Charlottesville analysts conducted five Monte Carlo microsimulations (i.e., with varying random numbers seeds) for each combination of calibration (input) parameter values. Execution of multiple runs and model realizations is nearly always appropriate for traffic microsimulation, and allows recognition of different possible outcomes. The only limitation here is that the Charlottesville analysis only obtained these multiple realizations for one set of demands and operating conditions. Ideally, each combination of demands and operating conditions (also known as a scenario) should have been examined through multiple runs and model realizations.

13.5 DATA PROCESSING, ANALYSIS, AND PRESENTATION

Using a Naming Structure for the Scenario Data Sets

TSSM **Error! Reference source not found.**recognizes that scenario analysis is one of the most valuable capabilities of traffic simulation models. For a project with numerous alternatives, the number of input data sets can become onerous to manage. Creating a condensed but readable naming structure is critical. **Error! Reference source not found.**recommends a naming structure that reflects an input that would vary from model to model. For example, given three geometric alternatives and two time periods, the naming structure might look like: *ProjectRef_Alternative#_TimePeriod_Version*, with one data set named *50476_Alt1_AM_V1*.

Although highly recommended, the data set naming convention is a low-level detail often not documented in the final project reports. Nonetheless, for the benefit of future engineers who may review or build on the initial work, an appendix to the simulation report could be used to document the chosen data set naming convention.

Automated Post-Processing Analysis

Traffic simulation models are notorious for producing exhaustive amounts of numerical output data, especially microsimulation models requiring 10 or more Monte Carlo realizations per scenario. Post-processor tools are frequently needed to analyze the summation, mean, median, maximum, percentile, and variance of these results across numerous links, nodes, sensors, and data stations. In some cases, these may be third-party post-processing tools distributed or sold separately from the native simulation tool. In other cases, there may be a post-processing module included within the simulation package, or the simulator may seamlessly provide post-processed statistics without any mention of a separate tool/module.

Simulation project reports do not always disclose names of the tools used to obtain their post-processed statistics. Engineers have post-processed and visualized their outputs through spreadsheet software (e.g., Microsoft® Excel) and statistics software (e.g., SAS, R). Specialized post-processing tools have been customized for specific simulation products.

Of the arterial case study reports discussed in this document, all three clearly obtained results via some form of automated post-processing. For example, the Boise report (2013) provided tables (e.g., Table 49) and graphs (e.g., Figure 266) reporting the cumulative performance of more than 2,000 links in the network.

Table 49. Networkwide traffic performance (Boise case study).

MOE	Network Results			Percent Difference		
	Base	Partial Closure	Full Closure	Partial Closure vs. Base	Full Closure vs. Base	Full Closure vs. Partial Closure
VHD (hr)	2,882	2,884	3,163	0%	10%	10%
VHT (hr)	7,888	7,920	8,231	0%	4%	4%
VMT (mile)	174,963	175,412	176,736	0%	1%	1%

Source: Six Mile Engineering.

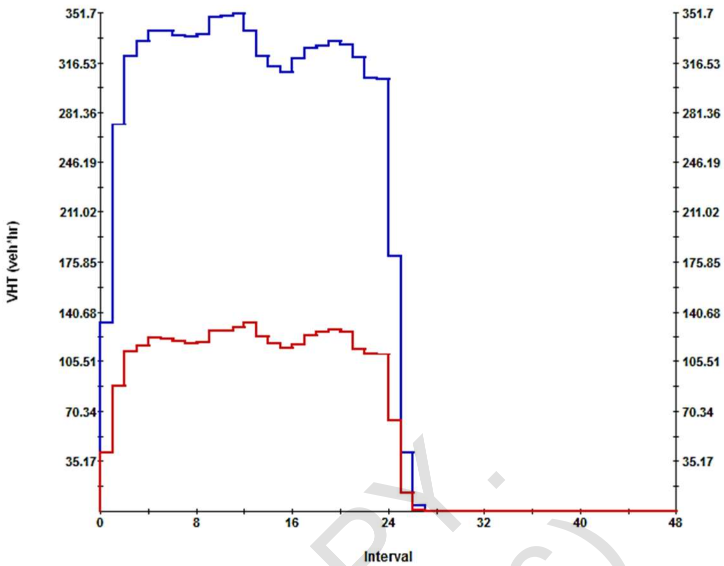


Figure 266. Chart. Networkwide vehicle miles traveled (VMT) over time (Boise case study).

Next, the Charlottesville report (2006) provided extensive results through spreadsheet software, some of which are shown in

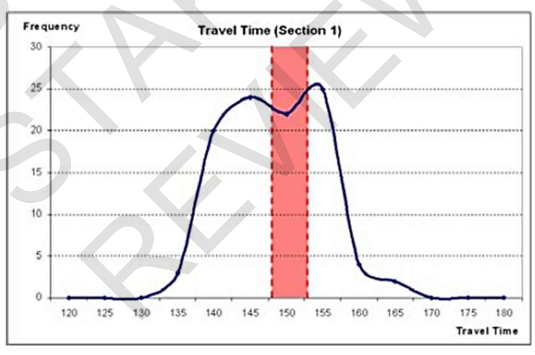
A. Subfigure of travel time (section 1).

B. Subfigure of travel time (section 2).

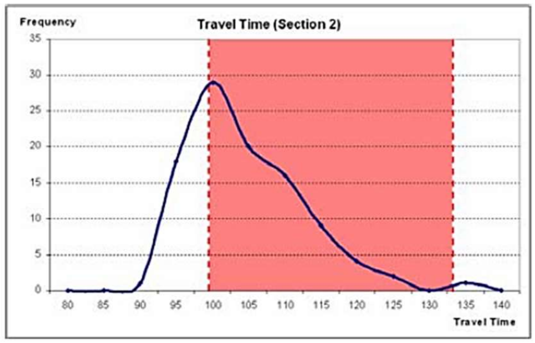
C. Subfigure of travel time (section 1) vs. traffic count.

D. Subfigure of travel time (section 2) vs. traffic count.

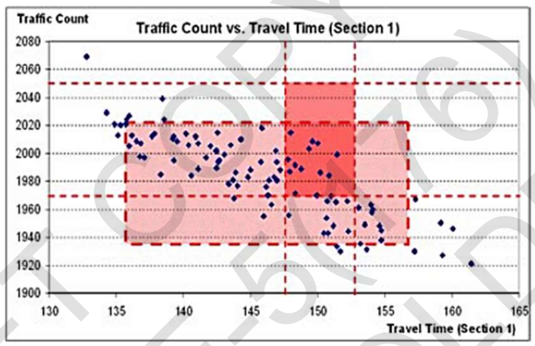
Figure 267. Statistical calculations (e.g., ANOVA, p-tests) were also obtained by the Charlottesville analysts through software not identified.



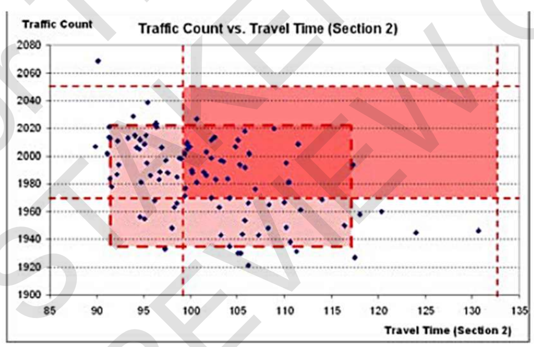
A. Subfigure of travel time (section 1).



B. Subfigure of travel time (section 2).



C. Subfigure of travel time (section 1) vs. traffic count.



D. Subfigure of travel time (section 2) vs. traffic count.

Figure 267. Charts. Calibration of flows and travel times (Charlottesville case study).

Finally, the Reno case study report (2012) provided six spreadsheet-generated histograms to illustrate the distribution of average delay and maximum queue for the AM peak, midday peak, and PM peak periods. One of these six histograms was previously illustrated in Figure 246.

Weighted Aggregation of Performance Measures

Applying weighting factors to the performance measures at various network locations can help to convey the right information. Performance measures can be weighted by demand, throughput, time, distance, and other metrics. This can help to show the bottom-line traffic performance was most heavily influenced by:

- Links or nodes serving the greatest number of trips (i.e., throughput-weighted average).
- Conditions prevalent for the longest time period(s) (i.e., time-weighted average).
- Links or segments covering the greatest distance (i.e., distance-weighted average).

Intersection level of service (LOS) is one example performance measure typically obtained by weighted averages, in this case a throughput-weighted average frequently referred to as a volume-weighted average. The Boise case study report (2013) provided maps of intersection LOS for their base scenario, partial bridge closure scenario, and full bridge closure scenario. See an example in
Figure 268. Ultimately the partial bridge closure scenario produced the greatest predicted benefits, and these maps helped to visually compare intersection LOS among the three scenarios. These maps also illustrated the vehicle throughputs at each intersection turn movement, such that those movements with the greatest throughputs would have the most influence on LOS.

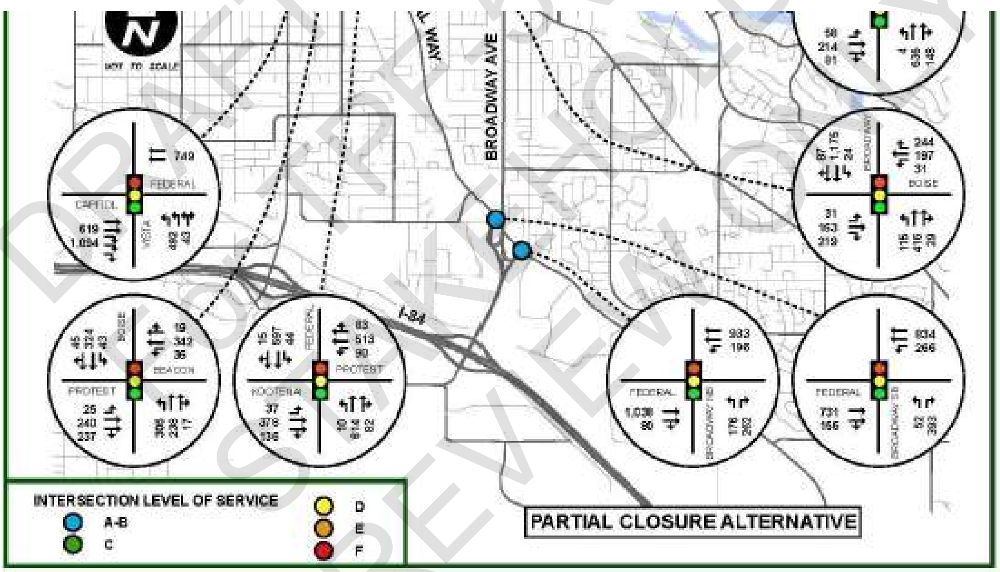


Figure 268. Illustration. Intersection levels of service within the Boise network (partial bridge closure alternative).

Analysis of Stochastic Model Outcomes under Different Random Number Seeds

Random number seeds are most often associated with microsimulation, although macroscopic and mesoscopic simulation tools can certainly incorporate random processes when the developer so chooses. According to conventional wisdom and several State DOT simulation guidelines, the number of Monte Carlo realizations should be at least 10 and should reflect important aspects of the study (e.g., congestion levels, network size). For example, an unstable freeway weaving section could warrant up to 100 realizations to obtain sufficient confidence in the reported results. The Charlottesville report (2006) described a rigorous, automated calibration procedure that considered 200 combinations of input parameter values. As such, they did not have enough time for additional realizations, and simply obtained the minimum 10 realizations per input combination (i.e., a total of 200×10=2,000 microsimulations). However, the Reno case study analysts decided to use only five realizations per scenario, even though there were only three scenarios (AM, midday, PM).

Use of Probability Density Functions and Cumulative Density Functions

The Charlottesville case study report (2006) provided several graphs analogous to PDFs, albeit with frequency instead of probability on the *y*-axis. These are easily converted to CDFs if and when desired. Earlier in this chapter, Figure 260and Figure 261visually demonstrated the potential impact of calibration via before-and-after frequency density functions. Another alternative to the PDF is the histogram, in which the performance measures are divided into discrete bins. The histogram shown in
Figure 269 is from the Charlottesville case study report (2006).

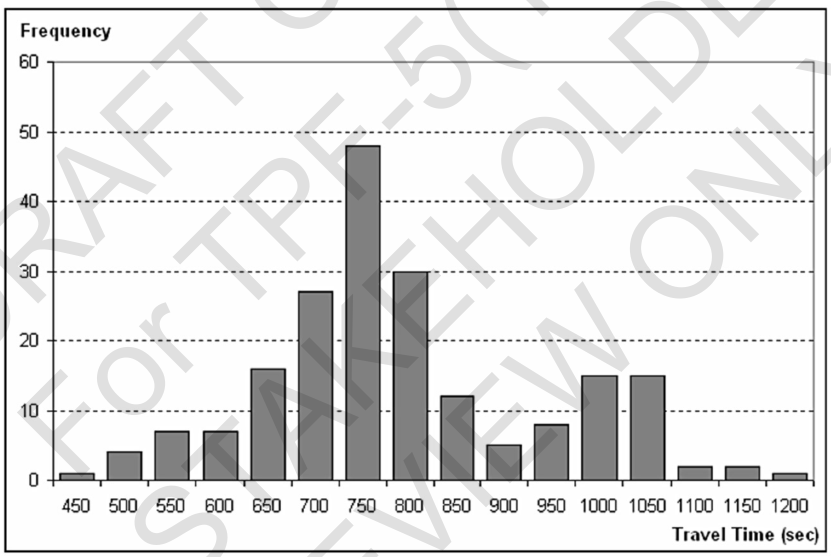


Figure 269. Chart. A sample travel time distribution generated by 100 microsimulation runs.

Use of Confidence Intervals

Performance measures typically generated from simulation tools may be unreliable unless the engineer provides a confidence interval, which indicates how likely it will be to observe a mean value within a specific range (e.g., the mean travel time has a 0.95 probability to fall between 1.45 and 1.57 minutes per mile). As previously shown in

A. Subfigure of travel time (section 1).

B. Subfigure of travel time (section 2).

C. Subfigure of travel time (section 1) vs. traffic count.

D. Subfigure of travel time (section 2) vs. traffic count.

Figure 267, analysts used confidence intervals to assess the effectiveness of their calibration:

The traffic count data were added to the consideration and presented as X-Y plots and compared with the field-collected data ranges. As shown in [

A. Subfigure of travel time (section 1).

B. Subfigure of travel time (section 2).

C. Subfigure of travel time (section 1) vs. traffic count.

D. Subfigure of travel time (section 2) vs. traffic count.

Figure 267] (c) and (d), 90% confidence interval region of simulation output data were overlapping with the field-collected performance measure data ranges. Therefore, it could be concluded that the enhanced procedure is not necessary.

Confidence intervals have often been used to clarify microsimulation outputs, but can also be used to assess the sample size of traffic data measurements used to obtain the simulation inputs. Ideally, the analyst would like to keep confidence high, while still having a narrow enough interval to sense where the true value may lie. It is also important to note that increasing the sample size results in a decrease in the width of a confidence interval, since the amount of error used to calculate the confidence interval is decreased.

> Confidence intervals can also be used to suggest whether two statistics are significantly different. If two confidence intervals overlap, this suggest the two means will probability not be significantly different at that respective confidence level. In © SAS Institute.

Figure 270, it is possible to see which carriers will most likely differ by comparing where the confidence intervals do not overlap.

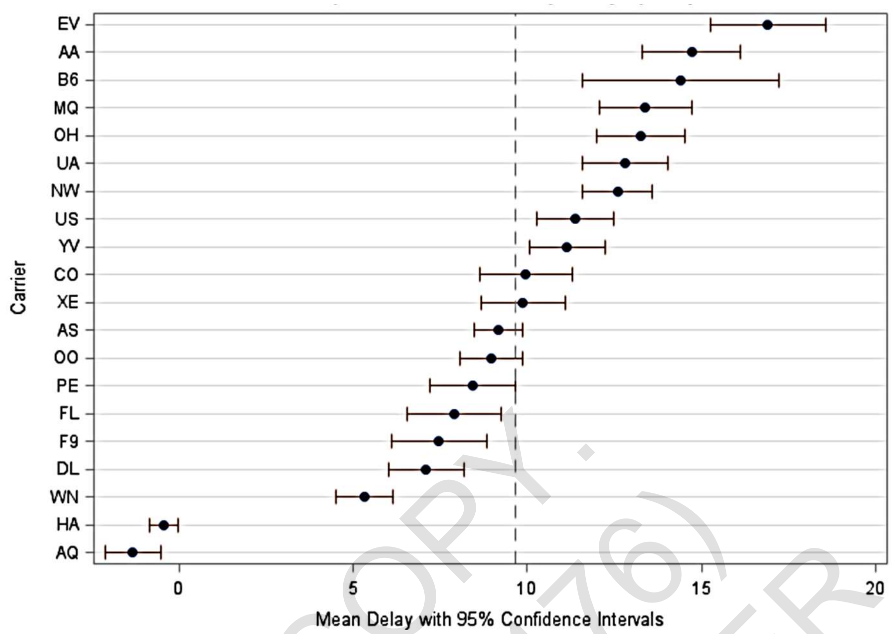


Figure 270. Chart. Daily airline delays.[102]

Apart from being easy to compare visually, it also helps to show how much variability is associated with each mean, by the examining the width of the intervals.

Use of Statistical Hypothesis Testing

Although goodness of fit cannot necessarily be classified as a hypothesis, the Boise case study report authors selected the R-squared (R^2) statistic as their high-level SQM, as stated earlier in the document: “The goal is to exceed a minimum goodness of fit, represented by the R^2 statistic, which measures how well the model approximates real data” (2013, 5). One example of this is given in
Figure 259 below. The authors state the R^2 value of 0.91 in
Figure 259 is well within the acceptable range for DTA model validation.

The Charlottesville analysts used statistical analysis (ANOVA) to identify key calibration parameters, as shown in Table 50.

[102] SAS Institute, Ranking with Confidence: Part 1, 2011, retrieved from https://blogs.sas.com/content/iml/2011/03/25/ranking-with-confidence-part-1.html#prettyPhoto/0/.

Table 50. Hypothesis testing on which factors have a significant impact on travel time.

Calibration Parameter	Section 1	Section 2
Mean value of start-up lost time	**0.00**	**0.00**
Mean queue discharge headway	**0.00**	**0.00**
Desired free-flow speed	0.07	0.30
Duration of a lane-change maneuver	**0.00**	**0.00**
Mean time for a driver to react to a sudden deceleration of the lead vehicle	**0.00**	**0.00**
Minimum deceleration for lane change	0.24	0.22
Difference in max/min acceptable deceleration for a mandatory lane change	0.50	0.55
Difference in max/min acceptable deceleration for a discretionary lane change	0.90	0.91
Deceleration rate of lead vehicle	0.77	0.86
Deceleration rate of following vehicle	**0.02**	**0.04**
Driver type factor used to compute driver aggressiveness	0.97	0.79
Urgency threshold	0.41	0.51
Safety factor X 10	0.85	0.96
Percentage of drivers who cooperate with a lane changer	0.94	0.80
Headway below which all drivers will attempt to change lanes	0.15	0.31
Headway above which no drivers will attempt to change lanes	0.56	0.68
Mean longitudinal distance over which drivers decide to perform one lane change	0.37	0.12
Left turn jumper probability	0.43	0.52
Left turn speed	0.89	0.69
Right turn speed	0.51	0.13
Left-turn lagging within 2 seconds	0.34	0.31
Left-turn lagging for 2~4 seconds	0.22	0.17
Acceptable deceleration for 10 driver types	0.61	0.80
Gap distribution for left turns	0.48	0.85
Gap distribution for right turns	0.79	0.91
Distribution of free flow speed by driver types	0.14	0.23
Start-up lost time distribution	0.07	0.15
Discharge headway distribution	0.05	0.08

* Note: level of significance 0.05 was used.

Source: Park, Won, and Perfater.

According to the results of the statistical analysis, the key parameters were as follows:

- Mean value of start-up lost time.
- Mean queue discharge headway.
- Desired free-flow speed.
- Duration of a lane-change maneuver.
- Mean time for a driver to react to a sudden deceleration of the lead vehicle.
- Start-up lost time distribution.

Analysis of Lane-Specific Results

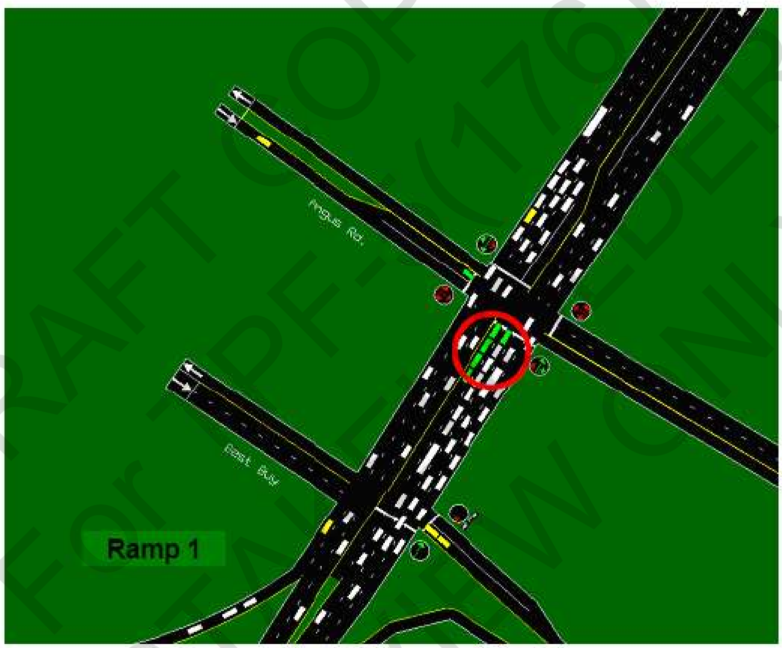
The Charlottesville case study report (2006) mentioned that one of the signalized intersection approaches exhibited heavy queuing in the right-most lane. However, prior to

the calibration effort, lane utilization was roughly equal on all signalized intersection approaches. These insights are valuable, and it is beneficial to include such information in the final project reports. It would also be beneficial to include any graphics or numeric outputs to highlight locations with significantly unequal lane utilization.

B. Subfigure of **dynamic traffic assignment (DTA) model of the study area.**

Figure **244**,

Figure 251,



B. Subfigure of lane change difficulty at an intersection downstream from ramp 1.

Figure 256, and
Figure 257 are graphical examples of this.

Use of Multivariate Outputs and Accompanying Visualizations

No single performance measure is perfect. Univariate measures may be misleading. It is helpful to consider several measures and distributions. Performance reporting can include combinations of numeric performance measures, visualizations, and qualitative discussions. In this case, the analyst's understanding would have been reduced if they had not had access to the visualizations.

In other cases, the numeric measures and/or qualitative discussions are needed to highlight subtleties within a complex visualization.

From the Boise case study (2013), some numeric performance measures were previously presented in Table 44and Table 49. Figure 258, which presented link-specific speeds on a color-coded map, was one example of an accompanying visualization. To complete the preferred trifecta of numeric results, graphical results, and qualitative discussions, the following qualitative discussion was provided:

> *The Partial Closure alternative is the recommended bridge construction phasing alternative based on the following findings from this study:*
>
> - *The Partial Closure alternative results in less delay and travel experienced by motorists during the weekday PM peak hour period than the Full Closure alternative. Although the estimated construction duration for the Partial Closure alternative is longer – 18 months compared to 9 months with the Full Closure alternative – the estimated user delay costs are estimated to be approximately $700,000 less than for the Full Closure alternative.*
> - *The Partial Closure alternative is expected to have less impact than the Full Closure alternative at to two critical intersections that are currently near capacity or overcapacity: Capitol Boulevard/Front Street and 9th Street/Front Street. These intersections are part of a closely spaced signalized network on three major corridors that are critical to PM peak travel in downtown Boise.*
> - *With the Partial Closure alternative, less than 30 percent of Broadway Bridge traffic is projected to re-route, which indicates that the river crossing is expected to remain the fastest route for a majority of traffic despite the lane and speed restrictions. The increase in travel times across the Broadway Bridge for the Partial Closure alternative are relatively small – between 20 and 50 seconds – compared with an additional 3.5 minutes for motorists re-routing over the Parkcenter Bridge with the Full Closure alternative.*

From the Charlottesville case study (2006), some numeric performance measures are presented in Table 51.

B. Subfigure of travel time of section 2.

Figure 260 **and**

A**. Subfigure of travel time of section 1.**

B**. Subfigure of travel time of section 2.**

Figure **261**, which graphically presented the route-specific travel distributions, were examples of accompanying visualizations.

Table 51. Simulated versus field-measured results (Charlottesville case study).

MOEs	Simulation Results	Actual Values
Travel time of Section 1, seconds	145.7	150.3
Travel time of Section 2, seconds	101.8	116.3
Maximum queue length at SB approach at Morton Rd, vehicle	16	16
Maximum queue length at NB approach at Morton Rd, vehicle	15	24

Source: Park, Won, and Perfater.

Regarding qualitative discussions, the following narrative was given alongside the previous

A. Subfigure of travel time (section 1).

B. Subfigure of travel time (section 2).

C. Subfigure of travel time (section 1) vs. traffic count.

D. Subfigure of travel time (section 2) vs. traffic count.

Figure 267:

After completing the calibration and validation procedure using single performance measure, the necessity of the procedure using multiple performance measures was evaluated by verifying simulation outputs that have been generated with the calibrated model using single performance measure. In other words, if the calibrated model satisfies the criteria provided in the enhanced procedure, then it is not necessary to conduct the calibration procedure using multiple performance measures. This is because the enhanced procedure would produce similar results.

As shown in [

A. Subfigure of travel time (section 1).

B. Subfigure of travel time (section 2).

C. Subfigure of travel time (section 1) vs. traffic count.

D. Subfigure of travel time (section 2) vs. traffic count.

Figure 267] (a) and (b), the distributions of travel time outputs with the calibrated model could include the field-measured travel time ranges. Based on two histograms, it could be concluded that the model is well-calibrated. The traffic count data were added to the consideration and presented as X-Y plots and compared with the field-collected data ranges. As shown in

A. Subfigure of travel time (section 1).

B. Subfigure of travel time (section 2).

C. Subfigure of travel time (section 1) vs. traffic count.

D. Subfigure of travel time (section 2) vs. traffic count.

Figure 267 (c) and (d), 90% confidence interval region of simulation output data were overlapping with the field-collected performance measure data ranges. Therefore, it could be concluded that the enhanced procedure is not necessary.

From the Reno case study (2012), vehicle delays and queue lengths were the primary performance measures. Some numeric measures are presented in Table 52 below. Figure 246, which graphically presented the delays in a histogram, was used as an accompanying visualization. Regarding qualitative discussions, the following brief statement was given at the report's end:

The simulation results indicate that both phasing schemes produced acceptable performance measures, though the enhanced phasing scheme proposed by UNR CATER showed better performances in most cases.

Table 52. Queue lengths in units of vehicles (Reno case study).

Peak Hours	AM		MD		PM	
Movement	CATER	City of Reno	CATER	City of Reno	CATER	City of Reno
10->9	287.2	288.2	346.8	333.3	427.1	345.1
10->11	175.3	206.4	213.4	248.1	267.8	479.4
1->8	229.9	250.7	387.2	406.0	594.8	634.5
1->2	30.8	41.7	117.1	138.1	325.4	365.0
3->8	69.0	62.0	66.4	61.8	84.3	74.0
3->2	37.7	13.5	40.4	9.4	67.0	50.9
14->13	0.0	0.0	0.0	0.0	0.0	0.0
14->7	279.7	274.8	414.1	447.2	526.3	632.1
5->6	0.0	0.0	0.0	0.0	0.0	0.0
5->4	174.8	222.1	216.8	247.6	245.5	308.9
12->13	7.8	0.0	17.7	8.3	59.2	100.3
12->7	95.6	101.8	112.7	164.6	238.9	279.6

Source: Center for Advanced Transportation Education and Research.

Presentation of Percent Changes Instead of Raw Numbers

During TSSM stakeholder meetings, stakeholders pointed out that simulation reports are easier to understand when they provide percent changes in the numeric performance measures, in addition to (or instead of) the raw numbers. One example of this was given in the Reno case study report (2012), and is shown in Table 53:

Table 53. Delay reduction caused by proposed diverging diamond interchange phasing scheme.

Peak Hour	AM	MD	PM
Movement	Change (%)	Change (%)	Change (%)
10->9	11%	16%	22%
10->11	-27%	-24%	-47%
1->8	-25%	-20%	-16%
1->2	5%	0%	-21%
3->8	14%	6%	-15%
3->2	-7%	-6%	-67%
14->13	-4%	0%	-59%
14->7	-8%	-14%	-25%
5->6	-8%	-10%	-12%
5->4	-33%	-27%	-38%
12->13	-8%	-14%	-18%
12->7	-75%	-80%	-86%
All	-17%	-16%	-28%

Source: Center for Advanced Transportation Education and Research.

Final Documentation of Assumptions, Unknowns, and Recommendations

The Boise case study report (2013) provided the following subsection:

> *Study Limitations and Assumptions*
>
> *The study effort was conducted with the following limitations and assumptions:*
>
> - *The signalized intersection analysis was conducted using Synchro 7, which follows the 2000 Highway Capacity Manual (HCM) methodology. The newest version, Synchro 8, incorporates the latest signalized intersection methodology from the 2010 HCM but was not used for this study because the software reports unexpected results when evaluating combination through and left-turn lanes. That lane configuration is present at several study intersections, including Broadway Avenue/Myrtle Street and 9th Street/Front Street. As confirmed by literature available from the Synchro 8 developer, the 2010 HCM computational methods have the potential to produce erroneous results (long delays) in these cases and a correction is currently pending.*
> - *With the 2010 HCM methodology, a signalized intersection is assigned a LOS F if the V/C ratio is equal to or greater than 1.0. With the 2000 HCM methodology, several study area intersections exceed a V/C ratio of 1.0 and report a LOS E. If re-evaluated with the 2010 HCM, these intersections will likely report a LOS F; however, the results were not adjusted to maintain consistency with the analysis methodology.*
> - *This study estimates the user delay costs for the full and partial closure alternatives. Additional impacts and costs of the construction phasing options are being considered by others in determining the preferred construction alternative.*
> - *The evaluation period was limited to the weekday PM peak hour from 5:00 to 6:00.*

The final recommendation from the Boise case study was to pursue the partial closure alternative as discussed earlier.

The Charlottesville case study report (2006) commented on study limitations and assumptions in more of a piecemeal fashion, as opposed to having a dedicated subsection like the Boise report. Examples of these statements included:

- *This case study assumed 100% of vehicles knowing the next two turn movements.*
- *Our research team assumed that the vehicles select their lane before they enter the network. To create that kind of condition, vehicle entry points for two different destination groups were separated with different lane groups:*
    - *Vehicle group that makes a right turn up to on-ramp toward Rte. 250 west.*
    - *Other vehicles.*

As discussed earlier, the final recommendation from the Charlottesville case study was that a univariate approach to calibration and validation was statistically sufficient.

Presentation of Static Graphics and Moving Vehicle Animation

Moving vehicle animation is typically associated with microsimulation, but static graphics are provided for all forms of simulation. The Boise case study (2013) used mesoscopic simulation, and thus the report provided several static graphics images.

Figure 253**,**

A. Subfigure of traffic assignment without penalty.

B. Subfigure of traffic assignment with penalty.

Figure **255**,

Figure 258 and

Figure 266 **are examples of this. The Charlottesville case study report (2006), which was based on microsimulation, provided both static graphics (e.g.,**

A. Subfigure of field travel time—section 1.

B. Subfigure of field travel time—section 2.

Figure **249**,
Figure 251,

B. Subfigure of travel time of section 2.

Figure 260**,**

A. **Subfigure of travel time of section 1.**

B. **Subfigure of travel time of section 2.**

Figure **261** and

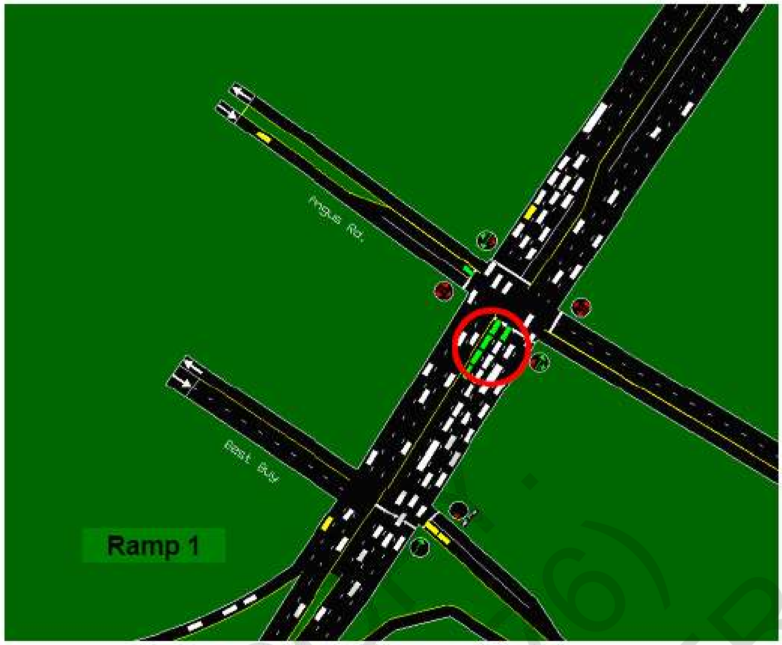
A. Subfigure of travel time (section 1).

B. Subfigure of travel time (section 2).

C. Subfigure of travel time (section 1) vs. traffic count.

D. Subfigure of travel time (section 2) vs. traffic count.

Figure 267) and snapshots of moving vehicle animation (e.g.,



B. Subfigure of lane change difficulty at an intersection downstream from ramp 1.

Figure 256 and
Figure 257). The Reno case study report (2012), which was based on microsimulation, provided histograms to present the vehicle delays and queue lengths.

APPENDIX A. SAMPLES

A FEW WORDS ABOUT SAMPLE SIZE

When thinking about data collection and manipulation to support transportation system simulation, regardless of facility type or simulation platform, it's important to consider the issue of sample size. The following is a list of key questions to remember:

1. **Is the sample size large enough for the accuracy desired?**

Imagine a set of measured speed data is being collected. To determine whether the sample size is adequate or not, the following general formula can be used for large populations (this formula is obtained from the formula for the confidence interval defined in question three below):

$$n = \left(s \frac{z_{\alpha/2}}{\varepsilon} \right)^2 \tag{A1}$$

Where:

n = minimum number of measured speeds required.
s = sample standard deviation.
$z_{\alpha/2}$ = constant corresponding to the desired confidence level.
ε = permitted error in the average speed estimate, miles per hour (mph).

For confidence levels of 90 percent, 95 percent, and 99 percent, use $z_{\alpha/2}$ values of 1.64, 1.96, and 2.58, respectively. Typically, a 95 percent confidence level is adequate for transportation applications.

2. **Given that a sample is of a certain size, how accurate is it?**

Solve the equation in part 1 for ε for the chosen confidence level.

3. **For a given level of confidence, what is the interval or range of points within which the true population mean could fall?**

The following general formula can be used for large populations to calculate a confidence interval for the true population mean based upon the sample mean, sample standard deviation, and sample size.

(A2)

Where:

- μ = true population mean.
- $\bar{x}$ = sample mean.
- s = sample standard deviation.
- n = sample size.
- $z_{\alpha/2}$ = constant corresponding to the desired confidence level.

Note: refer to **Error! Reference source not found.**for more details about confidence intervals and levels.

4. **For samples from two different populations, are the population means significantly different?**

To determine whether two sample distributions (representing two population distributions) are statistically significantly different from each other, the following formula can be applied for large populations:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \text{(A3)}$$

Where all variables are defined as before, with subscripts 1 and 2 referring to samples 1 and 2, respectively.

Compare the absolute value of th propriate z-value for the desired confidence level (i.e., 1.64, 1.96, ıe is greater than the z-value corresponding to the chosen conf in means is considered statistically significant.

$$\bar{x} - \frac{z_{\alpha/2}s}{\sqrt{n}} < \mu < \bar{x} + \frac{z_{\alpha/2}s}{\sqrt{n}}$$

Note that the variable s (the sample standard deviation) is used in the above equations instead of σ (the true population standard deviation) because the true population standard deviation is usually not known.

Additionally, the use of the z-variable is only valid for situations in which the sample size is large enough (usually a sample of greater than 30 is considered large). In cases where n is small (usually less than 30), the probability distribution is likely to be more spread out than a standard normal distribution (which the z variable represents). In this case, a more appropriate variable is t, representing the t distribution. This function still requires the population to be normally distributed but can treat smaller sample sizes more accurately. Use of the t distribution requires another parameter, the number of degrees of freedom of the distribution.

Note on the Number of Simulation Runs Required

Due to the stochastic nature of simulation packages, each model run with identical operational conditions but with different random seed numbers will produce different results. It is prudent, therefore, to execute a small number of simulation runs (say, 10) and with the statistical estimators of this trial, compute according to commonly accepted statistical principles, the number of simulation runs required to meet a stated objective (e.g., 80 percent, 90 percent, or 95 percent level of confidence). Based on probability and statistics, the following equation can be used to compute the required number of simulation runs Practitioners typically employ 30 simulation runs although a higher number of runs are required to meet a higher level of confidence.

$$n_r \geq \frac{(z_{\alpha/2})^2 \, s^2}{\varepsilon^2} \qquad \text{(A4)}$$

Where:

s^2 = variance (generally based on a relatively small number of initial simulation runs).

$z^2_{\alpha/2}$ = threshold value for the confidence interval.

n_r = number of runs required.

ε = maximum error of the estimate (user defined).

Example

For a specific freeway interchange, the following delay data for the northbound left movement at the northbound ramp intersection was determined from 10 runs of a microsimulation package. Each value represents seconds of delay.

45.2	41.9	42.2	41.9	59.9	46.8	62.1	58	46.9	39.2

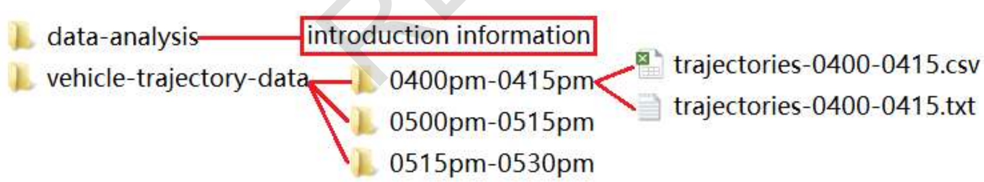
Mean of these data = 48.4 seconds.
s^2 of these data = 70.525 seconds^2.
s = 8.398 seconds.
α = 0.10 (corresponds with 90 percent confidence).
$\alpha / 2$ = 0.05 (corresponds with 90 percent confidence).
Z = 1.645 from statistical table (*1*).
ε = 1.5 seconds (based on reasonable error of delay estimate).

$$n_r \geq \frac{s^2 Z^2{}_{\alpha/2}}{\varepsilon^2} = \frac{(70.525)1.645^2}{1.5^2} \approx 85\ runs \quad \text{(A5)}$$

Traffic Data Measurements with Next Generation Simulation (NGSIM) Data

This section presents illustrative examples of traffic data measurements with the Next Generation Simulation (NGSIM) data (FHWA 2006). The data source used by the NGSIM program is a network of synchronized digital video cameras. A customized software application developed for the NGSIM program, NGVIDEO, transcribed the vehicle trajectory data from the video. The NGSIM program collected detailed vehicle trajectory data of southbound U.S. Route 101 (U.S. 101) and Lankershim Boulevard in Los Angeles, CA; along eastbound I–80 in Emeryville, CA; and along Peachtree Street in Atlanta, Georgia. The precise location of each vehicle within the study area was captured every one-tenth of a second. The following steps show how to use the NGSIM data to extract traffic characteristics such as average speed or delay.

The first step is to access data. The NGSIM data is accessible at https://www.its.dot.gov/data/. The NGSIM results can be found with the website's internal search function and exported as a comma-separated values (CSV) file. The released files are as shown in
Figure 271. The study area of the I–80 data set was approximately 500 meters (1,640 feet) in length and consisted of six freeway lanes, including a high-occupancy vehicle (HOV) lane. One on-ramp was located within the study area. A total of 45 minutes of data are available in the full data set, segmented into three 15-minute periods: 4–4:15 p.m., 5–5:15 p.m., and 5:15–5:30 p.m. These periods represent the buildup of congestion, or the transition between uncongested and congested conditions, and full congestion during the peak period.



Source: FHWA.

Figure 271. Illustration. Downloaded I–80 data set.

As shown below in Source: FHWA.

Figure 272, NGSIM data sets present detailed and accurate information about vehicle trajectories.

Vehicle_ID	Frame_ID	Total_Frame	Global_Time	Local_X	Local_Y	Global_X	Global_Y	v_Length	v_Width	v_Class
1	12	884	1.1134E+12	16.884	48.213	6042842.12	2133117.66	14.3	6.4	2
1	13	884	1.1134E+12	16.938	49.463	6042842.01	2133118.91	14.3	6.4	2
1	14	884	1.1134E+12	16.991	50.712	6042841.91	2133120.16	14.3	6.4	2

v_Class	v_Velocity	v_Acceleration	Lane_ID	Preceding Vehicle_ID	Following Vehicle_ID	Space_Headway	Time_Headway
2	12.5	0	2	0	0	0	0
2	12.5	0	2	0	0	0	0
2	12.5	0	2	0	0	0	0

Source: FHWA.

Figure 272. Chart. Next Generation Simulation data structure.

For any vehicle without lane-changing movements, the travel distance d traversed between the first data-collecting time point (at $t = 1$) and the current time point t can be estimated as the difference between Local_Y values: $d = Local\_Y(t) – Local\_Y(1)$. The Local_Y data are in units of feet. Thus, vehicles' trajectories can be plotted in a space-time diagram. In Figure 273, the solid blue line presents an example of trajectory for *Vehicle_ID = 1* during time interval $t = [1,600]$, which is 4 p.m.–4:01 p.m. In the NGSIM data set, each vehicle's speed at each time point is approximated by two sequential location data points using the equation $v = d/0.1s$. These data are available in the v_Velocity column. The unit of speed in NGSIM data is feet per seconds (ft/s). According to the data set, vehicles' speed range from around 3 ft/s to 33 ft/s during this time interval. If the free-flow speed is $75\ mph \approx 103.5\ ft/s$ (by assuming it is a bit higher than posted speed limit of 65 mph, then the delay in this case can be calculated as: *delay = actual travel time – free-flow travel time = 600 × 0.1 – 7.88 = 50.12 seconds*, as indicated in Figure 273.

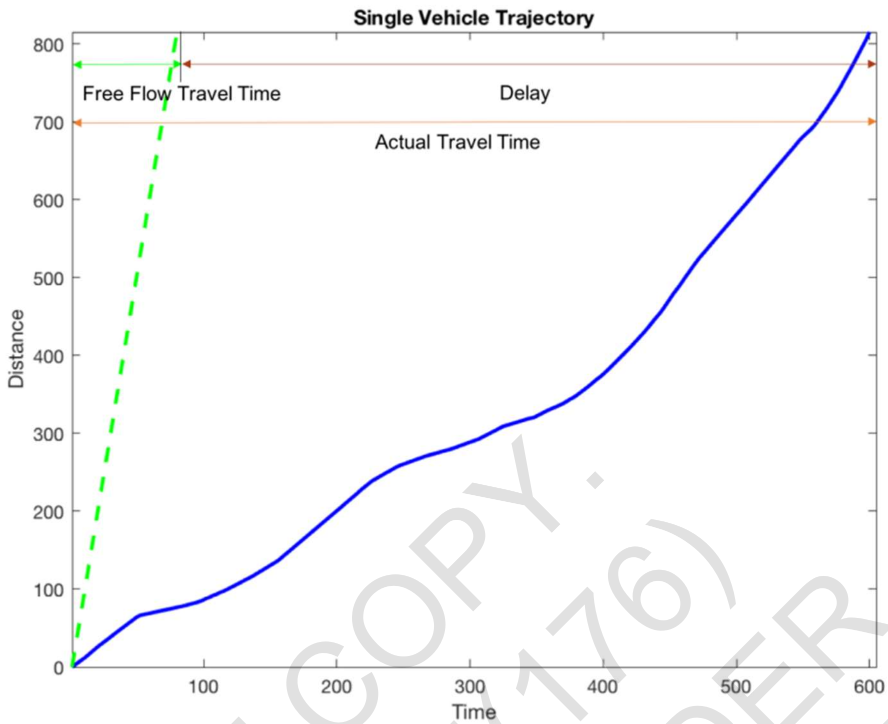


Figure 273. Chart. Single vehicle trajectory and delay.

With the NGSIM data, fundamental traffic related measurements can also be obtained. As illustrated before, density is defined by the number of consecutive vehicles observed over distance *d* at time *j*. For multiple vehicles, we can visualize their trajectories one by one in the same figure. In this example, we consider density for a single lane. Data cleaning and pre-processing may be needed to remove noise and errors in the database.

Figure 274 displays consecutive vehicle trajectories in lane 2 over segment *d = [0,1200]* over time interval *t = [600,700]*. At time point *j = 640*, the number of vehicles along segment [0,1200] is *n = 16*. Therefore, the density in $\frac{n}{d}=\frac{16}{1200}\ veh/ft\ /lane \approx 70\ vpmpl$ this case is equal to which corresponds to the severe traffic congestion during rush hour suffered by this segment.

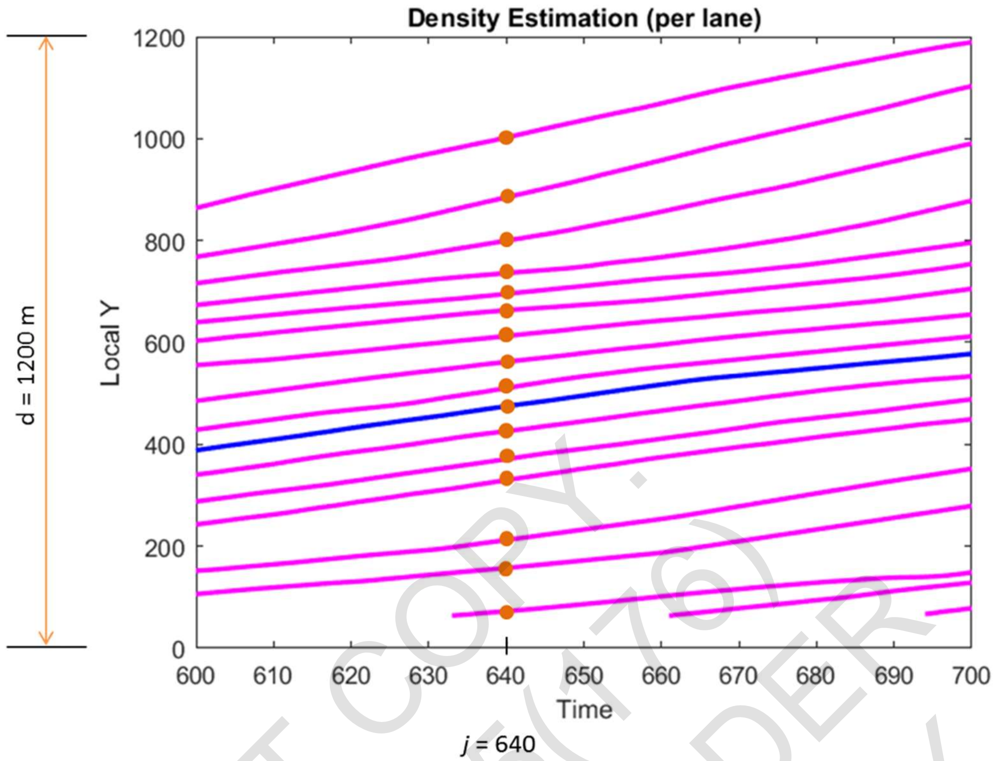


Figure 274. Chart. Spatial measurements example: density.

NGSIM data are among the most easily accessible and widely applied traffic data. As this example illustrates, applying and measuring traffic data are critical to extract various kinds of traffic characteristics for traffic simulation.

APPENDIX B. ORIGIN-DESTINATION FLOW ESTIMATION

In a qualitative sense, the process of adjusting the traffic flow inputs seeks to solve the following problem:

> *Minimize*: the mismatch between the estimated and observed arc flows[103] by time slice.
> *By*: adjusting the traffic demands (*O-D* flows by OD-pair and time slice).
> *While*: not deviating too far from the observed values of the arc flows (primarily) and the O-D flows (secondarily).

The structure of this problem is described graphically in Figure 275. The inputs are the observed arc flow rates and the observed O-D flows. The arc flow rates are the main input. They come from field observations. The observed O-D flows are the second main input. But they are typically obtained from a planning model and represent a set of average values. Probe data may also be available, and this data source will become more prevalent in the future. These flow data are more limited in scope and less certain.

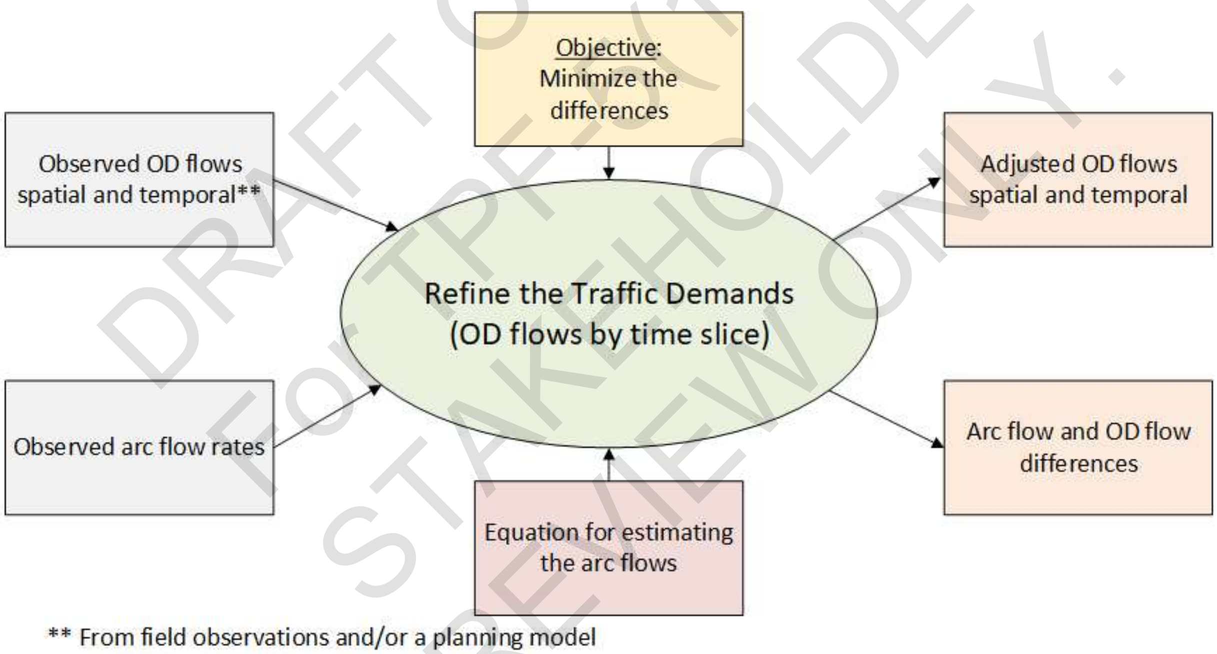


Figure 275. Diagram. The origin-destination (O-D) flow refinement process.

An equation (described later) allows the analyst to estimate the arc flows based on the estimated O-D flows. Finally, the objective is to minimize the differences between the observed values of both the arc flows and the O-D flows while not deviating too substantially from either.

---

[103] Link flows by direction.

The math involved is simple; but, it can look somewhat complex because of the subscripts that must be used. The variables being sought are adjusted values of the O-D flows $\hat{T}_{odk}$ for all *O*'s, *D*'s, and time slices *k*. A percentage difference variable Δ_{odk} captures the contrast between the estimated values $\hat{T}_{odk}$ and the observed values T_{odk}. The value of Δ_{odk} comes from the following simple equality relationship:

$$\hat{T}_{odk} + T_{odk}\Delta_{odk} = T_{odk}$$

(B1)

The estimated *O-D* flow rates $\hat{T}_{odk}$ are then used to compute estimated arc flow rates:

$$\hat{f}_{ak} = \sum_{odn} \alpha_{aodkn} * \hat{T}_{odn}$$

(B2)

The α_{aodkn} coefficients indicate the percentage of O-D flow commencing in time slice *n* that is crossing arc *a* in time slice *k*. The values of *n* and *k* are likely to be different because it takes multiple time slices to travel from *O* to *D*.[104]

Similar to the set of variables Δ_{odk}, a second set of difference variables, Δ_{ak}, captures the percentage differences between the estimated arc flow rates $\hat{f}_{ak}$ and the observed values f_{ak}:

$$\hat{f}_{ak} + f_{ak}\Delta_{ak} = f_{ak}$$

(B3)

These two sets of percentage difference variables are the focus of the objective function. It aims to minimize the sum of the squares of these values:[105]

$$Minimize \quad z = \lambda_1 \sum_{odk} w_{od}\Delta^2_{odk} + \lambda_2 \sum_{ak} w_a \Delta^2_{ak}$$

(B4)

The weights λ_1 and λ_2 indicate the relative importance of matching the O-D flows versus the arc flows, respectively. The weights inside the sums, w_{od} and w_a, indicate the relative importance of matching specific percentage differences for individual *O-D* pairs and arcs.[106] The focus on squared percentage differences implies that being either above or below the observed values are

---

[104] If α_{aodkn} happens to be 1.0, that means that all the *O-D* flow originating in time slice *n* crosses segment *a* in time slice *k*. A value less than 1.0 indicates that some portion of that *O-D* flow crosses the segment in time slice *k*.
[105] Minimizing the sum of the squares of the percent differences is also possible, consistent with the idea employed in equation 8.1.
[106] One option for these weights is to use the squares of the observed values.

equally bad. The use of the sum of the squares also means the refined values must be found using quadratic programming. Two embellishments can improve the search process and make it more efficient. The first is to minimize the sum of the absolute values of the percentage differences instead of the squared values. It decreases the emphasis on the large deviations, but makes the overall model linear:

$$Minimize \quad z = \lambda_1 \sum_{odk} w_{od} \left|\Delta_{odk}\right| + \lambda_2 \sum_{ak} w_a \left|\Delta_{ak}\right| \tag{B5}$$

The second embellishment, not elaborately described here, uses an enhanced absolute difference assessment to penalize larger deviations more heavily than small ones. It approximates the quadratic programming formulation. The essence of this idea is shown in Figure 276. The v_n is the computed value of the observed value, b_n is the observed value, and the deviations e_n^-, e_n^+, d_n^-, and d_n^+ are based on the differences between v_n and b_n.

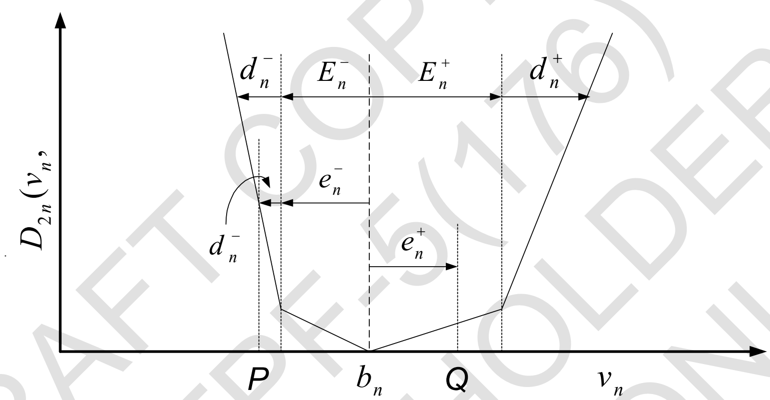


Figure 276. Chart. A varying sensitivity to differences from the observed values.

The sum of the small and large deviations is minimized. The large deviations are costlier than the small ones. Instead of a single absolute value percentage difference, Δ_{odk}, a composite percentage difference is computed that involves e_n^-, e_n^+, d_n^-, and d_n^+. For example, in the case of the arc flows:

$$\tilde{f}_{ak} + f_{ak}\Delta_{ak} = f_{ak} \quad where \quad f_{ak}\Delta_{ak} = d_{ak}^- + e_{ak}^- + e_{ak}^+ + d_{ak}^+ \tag{B6}$$

Only d_{ak}^- and e_{ak}^- can be non-zero or d_{ak}^+ and e_{ak}^+ can be non-zero for any difference between $\tilde{f}_{ak}$ and f_{ak} as long as standard math programming non-negativity constraints are imposed, both pairs cannot be non-zero at the same time. Also, if the weights associated with the *e* variables are less than those for the *d* variables, the *e* variables will be non-zero before the *d* variables are made non-zero.

REFERENCES

Adobe. (n.d.). Adobe Color CC. Accessed January 2, 2019. https://color.adobe.com/create/color-wheel/.

Alexiadis, V., D. Sallman and A. Armstrong. (2012). Traffic Analysis Toolbox Volume XIII: Integrated Corridor Management Analysis, Modeling, and Simulation Guide. FHWA-JPO-12-074.

Alexiadis, V., K. Jeannotte, and A. Chandra. (2004). Traffic Analysis Toolbox Volume I: Traffic Analysis Tools Primer. Washington, DC: FHWA-HRT-04-038. Retrieved from https://ops.fhwa.dot.gov/trafficanalysistools/tat_vol1/vol1_primer.pdf.

American Association of State Highway and Transportation Officials (AASHTO). (2014). Highway Safety Manual. Retrieved from http://www.highwaysafetymanual.org/Pages/default.aspx.

Antoniou C., J. Barcelò, M. Brackstone, H. Celikoglu, B. Ciuffo, V. Punzo, P. Sykes, T. Toledo, P. Vortisch, P. Wagner, V. Punzo, V., and M. Brackstone. (2014). Traffic Simulation: Case for Guidelines. Luxembourg: European Union. DOI:10.2788/11382.

Antoniou, C., G. List, and R. Bertini. (2015) Looking Back and Looking ahead: Celebrating 50 Years of Traffic Flow Theory, a Workshop. Transportation Research Circular Number E-C195. Retrieved from http://onlinepubs.trb.org/onlinepubs/circulars/ec195.pdf.

Banks, J. (1998). Handbook of Simulation. New York: John Wiley & Sons, Inc.

Barcelo, J. (2010). Fundamentals of Traffic Simulation. New York: Springer.

Ben-Akiva, M. (1996). Development of a Deployable Real-Time Dynamic Traffic Assignment System, Task D Interim Report: Analytical Developments for DTA System. MIT ITS Program, Cambridge, MA.

Brinckerhoff, P. (2007). Triangle Regional Model 2006 Documentation. Triangle Regional Model Service Bureau, Raleigh, NC.

Burghout, W. (2004). Hybrid Microscopic-Mesoscopic Traffic Simulation. Doctoral Dissertation, Royal Institute of Technology, Stockholm, Sweden. Retrieved from http://www.diva-portal.org/smash/get/diva2:14700/FULLTEXT01.pdf.

Cambridge Systematics Inc. (2014). Express Lanes Reliability Measures—Task Work Order 19 Summary Report. Prepared for the Florida Department of Transportation.

Cambridge Systematics Inc., Texas A&M Transportation Institute, University of Washington, Dowling Associates, Street Smarts, H. Levinson, and H. Rakha. (2013). Analytical Procedures for Determining the Impacts of Reliability Mitigation Strategies. SHRP 2 Report S2-L03-RR-1. Washington, DC: Transportation Research Board of National Academies.

CATT Lab. (2017). NPMRDS FAQs. Retrieved from https://npmrds.ritis.org/analytics/help/#npmrds.

Center for Advanced Transportation Education and Research. (2012). Proposed Signal Control at the Moana Lane U.S. 395 Diverging Diamond Interchange. Carson City: Nevada Department of Transportation.

Cetin, M. E. Beheshtitabar, R. Nezafat, G. List, and E. Willliams. (2018). Car-Following Behavior in a Tunnel and Impacts of Increasing Downstream Observability on Traffic Flow, Proceedings of the 97th Annual Meeting of the Transportation Research Board, on-line, January 7-11, Washington, DC.

Chen, C., A. Skabardonis, and P. Varaiya. (2003). Travel-Time Reliability as a Measure of Service. Transportation Research Record: Journal of the Transportation Research Board, No. 1855, pp. 74-79.

Chiu, Y.-C., J. Bottom, M. Mahut, A. Paz, R. Balakrishna, T. Waller, and J. Hicks. (2011). Dynamic Traffic Assignment: A Primer. Electronic Circular 153. Washington, DC: Transportation Research Board. Retrieved from http://onlinepubs.trb.org/onlinepubs/circulars/ec153.pdf.

Chiu, Y. C. (2011). Traffic Operations Analysis Tool Guidebook. Virginia Department of Transportation.

Chu, L, H. Liu, J-S Oh, and W. Recker. (2004). A Calibration Procedure for Microscopic Traffic Simulation. UCI-ITS-WP-04-2, University of Irvine.

Coifman, B., M. McCord, R. Mishalani, and K. Redmill. (2004). Surface Transportation Surveillance from Unmanned Aerial Vehicles. Proceedings of the 83rd Annual Meeting of the Transportation Research Board, Washington, DC. Retrieved from http://www2.ece.ohio-state.edu/~coifman/documents/UAV_paper.pdf.

Color Brewer (n.d.). Accessed January 2, 2019. http://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3.

Color Lab (2002). Color Lab. Retrieved from http://colorlab.wickline.org/colorblind/colorlab/.

Committee on Traffic Flow Theory (AHB45). (2001). Traffic Flow Theory, A State of the Art Report. Revised 2001. Miami: University of South Florida. Retrieved from http://tft.eng.usf.edu/docs/revised_monograph_2001.pdf.

Community Planning Association of Southwest Idaho (COMPASS). (2012). Treasure Valley Annual Congestion Management System Report. Report No. 09-2012. Retrieved from http://www.compassidaho.org/documents/prodserv/reports/2012CMSReport.pdf.

Community Planning Association of Southwest Idaho (COMPASS). (2018). Retrieved from http://www.compassidaho.org/.

Cooper, R. B. (1981). Introduction to Queueing Theory, 2nd Edition. New York: Elsevier North Holland.

Daganzo, C. (1994). The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. Transportation Research B 28:4.

Daganzo, C. (1995). The Cell Transmission Model, Part II: Network Traffic. Transportation Research B 29B (2): 79-93.

Daganzo, C. (1999). The Lagged Cell Transmission Model. International Symposium on Transportation and Traffic Theory, Pergamon, pp. 269-287.

Dowling, R., A. Skabardonis, and V. Alexiadis. (2016). Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software. Washington, DC: FHWA-HOP-16-070. Retrieved from http://www.dot.state.mn.us/trafficeng/modeling/resources/TAT%20Vol%20III%20E2.pdf

Dowling, R., List, G., Yang, B., Witzke, E., & Flannery, A. (2014). NCFRP Report 31, Incorporating Truck Analysis into the Highway Capacity Manual. Washington, DC: Transportation Research Board.

Dowling, R., Skabardonis, A., & Alexiadis, V. (2004). Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Software. Washington, DC: Federal Highway Administration, FHWA-HRT-04-040.

Dynameq. (2010). Version 1.4. 6, INRO. Retrieved from https://www.psrc.org/trip-based-travel-model-4k.

DynusT. (n.d.). Simulation-Based Dynamic Traffic Assignment Mesoscopic Model Software. Accessed January 2, 2019. http://dynust.net.

Elefteriadou, L. (2014). An Introduction to Traffic Flow Theory. New York: Springer.

Emam, E., and H. Al-Deek. (2006). Using Real-Life Dual-Loop Detector Data to Develop New Methodology for Estimating Freeway Travel Time Reliability. Transportation Research Record: Journal of the Transportation Research Board, No. 1959, pp. 140-150.

Environmental Protection Agency. (2018). MOVES and Other Mobile Source Emissions Models. Retrieved from https://www.epa.gov/moves.

Fan, Y., J. Wolfson, G. Adomavicius, K. Vardhan Das, Y. Khandelwal, and J. Kang. (2015). SmarTrAC: A Smartphone Solution for Context-Aware Travel and Activity Capturing. Center for Transportation Studies, University of Minnesota. Retrieved from http://hdl.handle.net/11299/173005.

Federal Highway Administration. (2006). Next Generation SIMulation Fact Sheet. FHWA-HRT-06-135. Retrieved from https://www.fhwa.dot.gov/publications/research/operations/its/06135/index.cfm.

Federal Highway Administration. (2013). Travel Time Reliability: Making It There on Time, All the Time. Retrieved from http://www.ops.fhwa.dot.gov/publications/tt_reliability/TTR_Report.htm.

Federal Highway Administration. (2015). Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software. Washington, DC: FHWA.

Federal Highway Administration. (2015). Traffic Flow Theory Monograph. Revised Monograph on Traffic Flow Theory. Retrieved from http://www.fhwa.dot.gov/publications/research/operations/tft/.

Federal Highway Administration. (2016). Narrowing Freeway Lanes and Shoulders to Create Additional Travel Lanes. Retrieved from https://ops.fhwa.dot.gov/publications/fhwahop16060/fhwahop16060.pdf.

Federal Highway Administration. (2017). Traffic Analysis Tools Program. http://ops.fhwa.dot.gov/trafficanalysistools/.

Federal Highway Administration. (2018). Alternative Designs to Alleviate Freeway Bottlenecks at Merge/Diverge and Weaving Areas. Retrieved from https://www.fhwa.dot.gov/publications/research/operations/18019/18019.pdf.

Federal Highway Administration. (forthcoming). Narrowing Freeway Lanes and Shoulders to Create Additional Travel Lanes. Washington, DC.

Federal Highway Administration. (2018). Surrogate Safety Assessment Model Overview. Retrieved from https://highways.dot.gov/safety/ssam/surrogate-safety-assessment-model-overview.

Florida Department of Transportation. (2000). The Florida Reliability Method in Florida's Mobility Performance Measures Program.

Florida Department of Transportation. (2014). Traffic Analysis Handbook. Systems Planning Office. Retrieved from https://fdotwww.blob.core.windows.net/sitefinity/docs/default-source/planning/systems/programs/traffanalysis/traffic-analysis-handbook_march-2014.pdf?sfvrsn=1edf5030_2.

Fundamental Relations of Traffic Flow. (n.d.). SlidePlayer web page, slide 21. Accessed October 20, 2019. https://slideplayer.com/slide/12982685/.

Gawron, C. (1998). Simulation-Based Traffic Assignment: Computing User Equilibria in Large Street Networks. PhD Thesis, University of Cologne, Cologne.

Gettman, D., and L. Head. (2003). Surrogate Safety Measures from Traffic Simulation Models, Final Report. FHWA-RD-03-050. Retrieved from https://www.fhwa.dot.gov/publications/research/safety/03050/03050.pdf.

Ghiasi, A., Hale, D., Bared, J., Kondyli, A., & Ma, J. (2018). A Dynamic Signal Control Approach for Integrated Ramp and Mainline Metering. From the 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 2892-2898. IEEE.

Ghiasi, A., Ma, J., Zhou, F., & Li, X. (2017). Speed Harmonization Algorithm Using Connected Autonomous Vehicles. Proceedings of the 96th Annual Meeting of the Transportation Research Board, Washington, DC.

Google® Earth™. (2018). Retrieved from https://www.google.com/earth/.

Google® Maps™. (2017). Chicago Department of Transportation. Retrieved from https://enterprise.google.com/maps/customers/chicago-department-of-transportation.html.

Google Maps™. (2018). Retrieved from https://www.google.com/maps.

Google® Maps™ APIs. (2017). Building the Next Generation of Location Experiences. Retrieved from https://developers.google.com/maps/.

Hansen, R. (2016). Traffic Monitoring Using UAV Technology. Retrieved from http://www.lidarmag.com/PDF/LIDARMagazine_Hansen_UASTrafficMonitoring_Vol6No5.pdf.

HERE Technologies. (n.d.). Traffic.com. Accessed October 2019. https://wego.here.com/traffic.

Hogg, R., and Tanis, E. (1993). Probability and Statistical Inference, 4th Edition. New York: Macmillan.

Hu, P., Tian, Z. Z., Xu, H., & Andalibian, R. (2014). An Advanced Signal Phasing Scheme for Diverging Diamond Interchanges. Proceedings of the 93rd Annual Meeting of the Transportation Research Board, Washington, DC.

INRIX. (2017). Retrieved from http://www.inrix.com.

INRIX. (2018). Retrieved from http://www.inrix.com.

Institute for Transportation Research and Education. (2018). FREEVAL Heat Maps. Retrieved from https://itre.ncsu.edu/.

Isukapati, I., B. Yang, and G. List. (2014). Use Cases for Travel Time Reliability Monitoring. Working paper produced as part of Establishing Monitoring Programs for Mobility and Travel Time Reliability (L02), Strategic Highway Research Program 2, Transportation Research Board, National Academies.

Jayakrishnan, R., H. Mahmassani , and T. Hu. (1994). An Evaluation Tool for Advanced Traffic Information and Management Systems in Urban Networks. Transportation Research C (3):129-147.

Jeannotte, K., Chandra, A., Alexiadis, V., & Skabardonis, A. (2004). Traffic Analysis Toolbox Volume II: Decision Support Methodology for Selecting Traffic Analysis Tools, FHWA-HRT-04-039. Washington, DC: Federal Highway Administration. Retrieved from https://ops.fhwa.dot.gov/trafficanalysistools/tat_vol2/Vol2_Methodology.pdf.

Kansas City ITS Portal. (n.d.). Accessed January 2, 2019. http://www.kcscout.com/.

Kittelson & Associates Inc. (2010). Understanding the Contributions of Operations, Technology, and Design to Meeting Highway Capacity Needs: Final Report. SHRP 2 Project C05. Washington, DC: Transportation Research Board of the National Academies.

Lee, S., and C. Messer. (2002). Assessment of Three Traffic Simulation Models for Diamond Interchange Analysis. Proceedings of the 81st Annual Meeting of the Transportation Research Board, Washington, DC.

Lieberman, E., and A. Rathi. (1997). Traffic Simulation. Traffic Flow Theory. Retrieved from https://www.fhwa.dot.gov/publications/research/operations/tft/chap10.pdf.

Lighthill, M., and G. Whitham. (1955). On Kinematic Waves II: A Theory of Traffic Flow on Long Crowded Roads. Proceedings of the Royal Society of London, Series A, 229.

List, G., M. Mashayekhi, R. Bertini, and B. Nevers. (2016). Development of a Transportation System Simulation Manual (TSSM) Framework: From Theory to Practice. Transportation Research Record (TRR), Journal of the Transportation Research Board.

List, G. et al. (2014). SHRP 2 Report S2-L02-RR-1, Establishing Monitoring Programs for Travel Time Reliability. Washington, DC: Transportation Research Board.

List, G., Williams, B., Rouphail, N., Hranac, R., Rodegerdts, L., Pincus, K, Schofer, J. K. (2014). SHRP 2 Report S2-L02-RR-2, Guide to Establishing Monitoring Programs for Travel Time Reliability. Washington, DC: Transportation Research Board.

Mahut, M. (2001). A Multi-Lane Extension of the Space-Time Queue Model of Traffic Dynamics. TRISTAN IV, Azores Islands.

Mahmassani, H., et al. (2013). SHRP 2 Report 52-L04-RR-1: Incorporating Reliability Performance Measures in Operations and Planning Modeling Tools. Washington, DC: Transportation Research Board of the National Academies.

Mahmassani, H. S., et al. (2014). SHRP 2 Report S2-L04-RR-1 Incorporating Reliability Performance Measures into Operations and Planning Modeling Tools. Washington, DC: Transportation Research Board.

MathWorks. (n.d.). Accessed January 2, 2019. https://www.mathworks.com/help/stats/introduction-to-cluster-analysis.html?requestedDomain=www.mathworks.com.

May, A. D. (1990). Traffic Flow Fundamentals. New Jersey: Prentice-Hall.

McGowan, H. (2014). Statistics 511-001: Experimental Statistics for Biological Sciences. Course notes.

Meyer, M., and M. Miller. (2002). Urban Transportation Planning, 2nd Edition. New York: McGraw-Hill.

Möller, D. P. (2015). Introduction to Transportation Analysis, Modeling and Simulation: Computational Foundations and Multimodal Applications. New York: Springer.

Monsere, C. (2012). Understanding and Communicating Multimodal Transportation Data. Pacific Crest, ISBN: 978-1-60263-440-4.

Nam, D., D. Park, and A. Khamkongkhun. (2005). Estimation of Value of Travel Time Reliability. Journal of Advanced Transportation, Vol. 39, No. 1, pp. 39-61.

Nevers, B., Nguyen, K., Quayle, S., Zhou, X., & Taylor, J. (2013). The Effective Integration of Analysis, Modeling, and Simulation Tools. Washington, DC: Federal Highway Administration, FHWA-HRT-13-036. Retrieved from https://www.fhwa.dot.gov/publications/research/operations/13036/13036.pdf.

Ni, D. (2016). Traffic Flow Theory - Characteristics, Experimental Methods, and Numerical Techniques. Waltham: Elsevier.

Park, B., Jones, T., & Griffin, S. (2010). Weather and Traffic Analysis, Modeling, and Simulation, Volume XI, Traffic Analysis Toolbox, FHWA-JPO-11-019. Washington, DC: Federal Highway Administration.

Park, B., Won, J., & Perfater, M. A. (2006). Microscopic Simulation Model Calibration and Validation Handbook. FHWA/VTRC 07-CR6. Virginia Transportation Research Council.

Polley, T. (2016). Introduction to Data Visualization. Workshop notes. Raleigh, NC.

PTV, A. (2011). VISSIM 5.40 User Manual. Karlsruhe, Germany.

R Core Team. (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.R-project.org/.

RAWGraphs. (n.d.). Accessed January 2, 2019. http://raw.densitydesign.org/.

Ray, B. (2014). NCHRP 785, Performance-Based Analysis of Geometric Design of Highways and Streets. Washington, DC: National Cooperative Highway Research Program. Retrieved from https://www.nap.edu/catalog/22285/performance-based-analysis -of-geometric-design-of-highways-and-streets.

Reading Statistical Services Centre, University of Reading (2000). Informative Presentation of Tables, Graphs, and Statistics. Retrieved from www.reading.ac.uk/ssc/resources/Docs/PresentationOfTablesGraphsAndStatistics.pdf.

Regional Integrated Transportation Information System. (n.d.). Probe Data Analytics Suite (web page). Retrieved from http://vpp.ritis.org/.

Richards, P. (1956). Shock Waves on the Highway. Operations Research 4: 42-51.

Roess, R. P., and J. M. Ulerio. (1997). NCHRP Report 385, Comparison of the 1994 Highway Capacity Manual's Ramp Analysis Procedures and the FRESIM Model. Washington, DC: TRB, National Research Council.

Rouphail, N. and B. Schroeder. (2010). User's Guide to FREEVAL. Raleigh, North Carolina.

Saberi, M., and R.L. Bertini. (2010). Beyond Corridor Reliability Measures: Analysis of Freeway Travel Time Reliability at the Segment Level for Hot Spot Identification. Proceedings of the 89th Annual Meeting of the Transportation Research Board.

SAS Institute. (1990). SAS Language: Reference: Version 6. Sas Inst.

SAS Institute. (2011). Ranking with Confidence: Part 1. Retrieved from https://blogs.sas.com/content/iml/2011/03/25/ranking-with-confidence-part-1.html#prettyPhoto/0/.

Schneider Electric and RK&K. (2015). Bryan Park Interchange Dynamic Lane Merge Assessment. Contract Number 40903, Task Order 17. Virginia Department of Transportation, Richmond, VA.

Schroeder, B., Rouphail, N., Williams, B., Hajbabaie, A., Jia, A., Sajjadi, S., & Narron, B. (2014). Work Zone Traffic Analysis & Impact Assessment. Institute for Transportation Research and Education. Raleigh, North Carolina.

Six Mile Engineering. (2013). Traffic Analysis of Construction Phasing Options. Boise: Idaho Transportation Department District 3.

Small, K., R. Noland, X. Chu, and D. Lewis. (1999). Valuation of Travel-Time Savings and Prediction in Congested Conditions for Highway User-Cost Estimation. Transportation Research Board. University of California, Irvine.

Srinivasan, S., H. Latchman, J. Shea, T. Wong, and J. McNair. (2004). Airborne Traffic Surveillance Systems – Video Surveillance of Highway Traffic. ACM 2nd International Workshop on Video Surveillance & Sensor Networks. DOI: https://dl.acm.org/citation.cfm?doid=1026799.1026821.

Statistics How To. (2018). T-Distribution Table (One Tail and Two-Tails). Retrieved from http://www.statisticshowto.com/tables/t-distribution-table/#two.

Tung, R., Z. Wang, and Chiu, Y. C. (2010). Step Model Framework: A Deployment Case Study in Seattle Model. Proceedings of the 3rd Conference on Innovations in Travel Modeling.

Stogios, Y., H. Mahmassani, and P. Vovsha (2013). Incorporating Reliability Performance Measures in Operations and Planning Modeling Tools. SHRP 2 Report S2-L04-RR-1. http://www.trb.org/Main/Blurbs/170716.aspx.

Tao, Z., Spotts, J., & Hess, E. (2011). S2-L13-RW-1, Requirements and Feasibility of a System for Archiving and Disseminating Data from SHRP 2 Reliability and Related Studies. Washington, DC: Transportation Research Board.

The Seattle Times. (2006). Rice Is Nice When Trying to Visualize Highway Traffic. Retrieved from https://www.seattletimes.com/seattle-news/rice-is-nice-when-trying-to-visualize-highway-traffic/.

The Traffic Group. (2011). Diamond Interchange DDI Traffic Simulation. Video. Retrieved from https://www.youtube.com/watch?v=d-xCRbhOlIQ.

The University of Reading Statistical Services Centre. (2000). Informative Presentation of Tables, Graphs, and Statistics. Retrieved from http://www.reading.ac.uk/ssc/resources/Docs/PresentationOfTablesGraphsAndStatistics.pdf.

Toledo, T., H. Koutsopolous, A. Davol, M. Ben-Akiva, W. Burghout, I. Andreasson, T. Johansson, and C. Lundin. (2003). Calibration and Validation of Microscopic Traffic Simulation Tools: Stockholm Case Study. Transportation Research Record 1831, pp. 65-75.

Transportation Research Board. (2000). Highway Capacity Manual 2000. Washington, DC: Transportation Research Board.

Transportation Research Board. (2010). Highway Capacity Manual 2010. Washington, DC: Transportation Research Board.

Transportation Research Board. (2016). Highway Capacity Manual, 6th ed. Washington, DC: Transportation Research Board.

Treiber, M., & Kesting, A. (2013). Traffic Flow Dynamics - Data, Models and Simulation. Heidelberg: Springer.

TRL Software - TRANSYT. (2018). TRANSYT. Retrieved from https://trlsoftware.co.uk/products/junction_signal_design/transyt.

Tu, H., H. Van Lint, and H. van Zuylen. (2008). Travel Time Reliability Model on Freeways. Proceedings of the 87th Annual Meeting of Transportation Research Board, Washington, DC.

Turner, S. (2002). Defining and Measuring Traffic Data Quality. Federal Highway Administration, Office of Policy.

Turner, S., W. Eisele, R. Benz, and D. Holdener. (1998). Travel Time Data Collection Handbook. TRID, FHWA-PL-98-035.

United Nations Economic Commission for Europe. (2009). Making Data Meaningful Part 2: A Guide to Presenting Statistics. Geneva. http://www.unece.org/fileadmin/DAM/stats/documents/writing/MDM_Part2_English.pdf.

United States Department of Transportation. About Ramp Metering. Accessed October 10, 2019. https://ops.fhwa.dot.gov/freewaymgmt/ramp_metering/about.htm.

United States Department of Transportation. Analysis, Modeling, and Simulation for Traffic Incident Management Applications. Accessed October 10, 2019. https://ops.fhwa.dot.gov/publications/fhwahop12045/2_synth.htm.

United States Department of Transportation. Recurring Traffic Bottlenecks: A Primer. Accessed October 10, 2019. https://ops.fhwa.dot.gov/publications/fhwahop18013/index.htm.

United States Department of Transportation. (2018). Connected Vehicle Pilot Deployment Program. Retrieved from https://www.its.dot.gov/pilots/.

University of Florida - McTrans. (2018). McTrans Transyt-7F. Retrieved from http://mctrans.ce.ufl.edu/mct/index.php/hcs/transyt-7f/.

University of Leicester. (n.d.). Presenting Numerical Data. Accessed January 2, 2019. https://www2.le.ac.uk/offices/ld/resources/numerical-data/numerical-data.

Van Aerde, M. (1998). QueensOD–Release 2.10–Users Guide: Estimating Origin Destination Traffic Demands from Link Flow Counts. Blacksburg, VA: Michel Van Aerde & Associates, Ltd.

van Wageningen-Kessels, F., H. Van Lint, K. Vuik, and S. Hoogendoorn. (2015). Genealogy of Traffic Flow Models. EURO Journal on Transportation and Logistics 4 (4), 445-473.

Virginia Department of Transportation. (2015). Traffic Operations and Safety Analysis Manual (TOSAM). Version 1.0. Accessed September 28, 2018. http://www.virginiadot.org/business/resources/TOSAM.pdf.

Vortisch, P., and P. Wagner. (2013). The MULTITUDE Traffic Simulation Primer. Joint Research Centre of the European Commission.

Washington State Department of Transportation. (n.d.). The PRICE Is Right!." Retrieved from http://www.wsdot.wa.gov/NR/rdonlyres/4D0BFFF3-E59B-45EE-ABB4-1F00F762C5DE/0/ MacDonaldRiceHandout.pdf.

Wisconsin Department of Transportation. Doug MacDonald - Rice and Traffic Congestion. April 20, 2007. YouTube video, 6:19. https://www.youtube.com/watch?v=8G7ViTTuwno.

Wunderlich, K. (2002). Seattle 2020 Case Study, PRUEVIN Methodology. Mitretek Systems.

Wunderlich, K. V., M. Vasudevan, and P. Wang. (2016). Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software 2019. Update to the 2004 Version. FHWA-HOP-18-036. Washington, DC: Federal Highway Administration.

Wunderlich, K., V. Alexiadis, and P. Wang. (2017). Scoping and Conducting Data-Driven 21st Century Transportation System Analyses. TRID, FHWA-HOP-16-072.

Wunderlich, K., M. Vasudevan, and P. Wang. (2018). Traffic Analysis Tools Volume III Update: Guidelines for Applying, PowerPoint slides used to describe the updates to Volume III. Washington, DC: Federal Highway Administration.

Xie, K., X. Wang, K. Ozbay, and H. Yang. (2013). Crash Frequency Modeling for Signalized Intersections in a High-Density Urban Road Network. Proceedings of the 4th

International Conference on Road Safety and Simulation, Rome, Italy, October 23–25, 2013.

Yelchuru, B., R. Kamalanathsharma, X. Zhou, P. Mirchandani, P. Li, B. Nevers, B. Abdelghany, P. Rinelli, H. Mahmassani, T. Bauer, and J. Ma. (2016). Analysis, Modeling, and Simulation (AMS) Testbed Development and Evaluation to Support Dynamic Mobility Applications (DMA) and Active Transportation and Demand Management (ATDM) Programs—Dallas Testbed Analysis Plan. Washington, DC: FHWA-JPO-16-373.

Zegeer, J., Bonneson, J., Dowling, R., Ryus, P., Vandehey, M., Kittelson, W., … Elefteriadou, L. (2014). Incorporation of Travel Time Reliability into the Highway Capacity Manual, SHRP2 Report S2-L08-RW-1. Washington, DC: Transportation Research Board

Zhang, X., Y. Wang, N. Nihan, and M. Hallenbeck. (2003). Development of a System to Collect Loop Detector Event (Individual Vehicle) Data. Transportation Research Record: Journal of the Transportation Research Board, No. 1855, pp. 168-175.

Zhou, X., Lu, J., Lei, H., Reynolds, W., Jia, A., & Li, M. (2014). DTALite.

BIBLIOGRAPHY

Barceló, J. (2010). Fundamentals of Traffic Simulation. Vol. 145, p. 439. New York: Springer.

Bazzan, A. L., and Klügl. F. (2013). Introduction to Intelligent Systems in Traffic and Transportation. Synthesis Lectures on Artificial Intelligence and Machine Learning. 7(3), 1-137.

Bazzan, A. L., and Klügl, F. (2014). A Review on Agent-Based Technology for Traffic and Transportation. The Knowledge Engineering Review, 29(3), 375-403.

Berto, F. and Tagliabue J. (Fall 2017). Cellular Automata. Stanford Encyclopedia of Philosophy. Retrieved from https://plato.stanford.edu/archives/fall2017/entries/cellular-automata.

Bing, A. J. (2010). Guidebook for Implementing Passenger Rail Service on Shared Passenger and Freight Corridors. Vol. 657. Transportation Research Board.

Carson, J. L. (2010). Best Practices in Traffic Incident Management. Washington, DC: United States Department of Transportation, Federal Highway Administration, Office of Operations.

Chaudhary, N. A., Tian, Z., Messer, C. J., and Chu, C. L. (2004). Ramp Metering Algorithms and Approaches for Texas. Texas Transportation Institute Technical Report.

Ciuffo, B., Punzo, V., and Montanino, M. (2014). Global Sensitivity Analysis Techniques to Simplify the Calibration of Traffic Simulation Models. Methodology and Application to the IDM Car-Following Model. IET Intelligent Transport Systems, 8(5), 479-489.

Cortés, C. E., Burgos, V., and Fernández, R. (2010). Modelling Passengers, Buses and Stops in Traffic Microsimulation: Review and Extensions. Journal of Advanced Transportation, 44(2), 72-88.

Cortés, C., Pagès, L., and Jayakrishnan, R. (2005). Microsimulation of Flexible Transit System Designs in Realistic Urban Networks. Transportation Research Record: Journal of the Transportation Research Board, (1923), 153-163.

Deniz, O., Celikoglu, H. B., and Gurcanli, G. E. (October 2012). Overview to Some Incident Detection Algorithms: A Comparative Evaluation with Istanbul Freeway Data. Proceedings of the 12th International Conference Reliability and Statistics in Transportation and Communication.

Environmental Protection Agency. (2002). Users Guide to MOBILE 6.0: Mobile Source Emission Factor Model. No. EPA420-R-02-001. Washington, DC.

Federal Highway Administration. (2006). Next Generation Simulation (NGSIM) Fact Sheet. Interstate 80 Freeway Dataset. No. FHWA-HRT-06-137. Washington, DC: Retrieved from https://www.fhwa.dot.gov/publications/research/operations/06137/.

Federal Highway Administration. (2018). Next Generation Simulation (NGSIM) Program. Retrieved from https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm.

Gettman, D., Pu, L., Sayed, T., Shelby, S., and Siemens, I. T. S. (2008). Surrogate Safety Assessment Model and Validation. No. FHWA-HRT-08-051. Washington, DC: Federal Highway Administration, Office of Safety Research and Development.

Ghariani, N., Elkosantini, S., Darmoul, S., and Said, L. B. (May 2014). A Survey of Simulation Platforms for the Assessment of Public Transport Control Systems. International Conference on Advanced Logistics and Transport (ICALT), p. 85-90. IEEE.

Hollander, Y., and Liu, R. (2008). The principles of Calibrating Traffic Microsimulation Models. Transportation, 35(3), 347-362.

Hou, Y., Zhao, Y., Hulme, K. F., Huang, S., Yang, Y., Sadek, A. W., and Qiao, C. (2014). An Integrated Traffic-Driving Simulation Framework: Design, Implementation, and Validation. Transportation Research Part C: Emerging Technologies, 45, 138-153.

Johansson, F. (2013). Microscopic Modeling and Simulation of Pedestrian Traffic. Doctoral dissertation. Linköping University Electronic Press.

Keller, M., and Kljun, N. (2007). ARTEMIS Road Emission Model-Model Description. EU project ARTEMIS deliverable13, INFRAS Report, 1-199.

Kelton, W.D., Sadowski, R.P. and Sadowski, D.A. (2010). Simulation with Arena. Chap. 2, Fundamental Simulation Concepts. McGraw Hill.

Kim, W., and Rilett, L. R. (2005). Improved Transit Signal Priority System for Networks with Nearside Bus Stops. Transportation Research Record, 1925(1), 205-214.

Kittelson & Assoc, Inc., Parsons Brinckerhoff, Inc., KFH Group, Inc., Texas A&M Transportation Institute, and Arup. (2013). Transit Capacity and Quality of Service Manual. 3rd ed. Transit Cooperative Highway Research Program (TCRP) Report 165. Washington DC: Transportation Research Board.

Lee, J., Shalaby, A., Greenough, J., Bowie, M., and Hung, S. (2005). Advanced Transit Signal Priority Control with Online Microsimulation-Based Transit Prediction Model. Transportation Research Record, 1925(1), 185-194.

Lieu, H. (2003). Revised Monograph on Traffic Flow Theory. Washington, DC: United States Department of Transportation, Federal Highway Administration. http://www.fhwa.dot.gov/publications/research/operations/tft/.

List, G., Mashayekhi, M., Bertini, R. L., and Nevers, B. (2016). Development of a Transportation System Simulation Manual Framework: From Theory to Practice. Transportation Research Record: Journal of the Transportation Research Board, (2561), 45-52.

Ma, J., Dong, H., and Zhang, H. (2007). Calibration of Microsimulation with Heuristic Optimization Methods. Transportation Research Record: Journal of the Transportation Research Board, (1999), 208-217.

Ma, X., Huang, Z., and Koutsopoulos, H. (2014). Integrated Traffic and Emission Simulation: A Model Calibration Approach Using Aggregate Information. Environmental Modeling & Assessment, 19(4), 271-282.

Martin, P. T., Perrin, J., Hansen, B., Kump, R., and Moore, D. (2001). Incident Detection Algorithm Evaluation. Utah Department of Transportation.

Möller, D. P. (2014). Introduction to Transportation Analysis, Modeling and Simulation: Computational Foundations and Multimodal Applications. Springer.

Nisan, N., Roughgarden, T., Tardos, E., and Vazirani, V. V. (2007). Algorithmic Game Theory. New York: Cambridge University Press.

Ntziachristos, L., Samaras, Z., Eggleston, S., Gorissen, N., Hassel, D., and Hickman, A. J. (2000). Copert III: Computer Programme to Calculate Emissions from Road Transport, Methodology and Emission Factors. Technical Report 49. Copenhagen: European Energy Agency (EEA).

Parkany, E., and Xie, C. (2005). A Complete Review of Incident Detection Algorithms & Their Deployment: What Works and What Doesn't. No. NETCR 37, NETC 00-7.

Russell, S. J., and Norvig, P. (2010). Artificial Intelligence (A Modern Approach).

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004). Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. John Wiley & Sons.

Stathopoulos, F., and Noland, R. (2003). Induced Travel and Emissions from Traffic Flow Improvement Projects. Transportation Research Record: Journal of the Transportation Research Board, (1842), 57-63.

Stevanovic, A., Stevanovic, J., Zhang, K., and Batterman, S. (2009). Optimizing Traffic Control to Reduce Fuel Consumption and Vehicular Emissions: Integrated Approach with VISSIM, CMEM, and VISGAOST. Transportation Research Record: Journal of the transportation research board, (2128), 105-113.

Transportation Research Board of the National Academies. (2014). SHRP 2 Project L02: Establishing Monitoring Programs for Travel Time Reliability. Washington, DC.

Transportation Research Board of the National Academies. (2014). SHRP 2 Project L08: Incorporating Travel Time Reliability into the Highway Capacity Manual. Washington, DC.

Transportation Research Board. (2000). Highway Capacity Manual. HCM 2000, chap. 31, p. 31-1. Washington, DC: Transportation Research Board.

Twaddle, H., Schendzielorz, T., and Fakler, O. (2014). Bicycles in Urban Areas: Review of Existing Methods for Modeling Behavior. Transportation Research Record, 2434(1), 140-146.

Wang, F. Y. (2010). Parallel Control and Management for Intelligent Transportation Systems: Concepts, Architectures, and Applications. IEEE Transactions on Intelligent Transportation Systems, 11(3), 630-638.

Whitman, Requardt & Associates LLP. (2015). I–270 Study Modeling Calibration Methodologies. Memorandum. Baltimore, Maryland.

Winsberg, E. (Summer 2018). Computer Simulations in Science. Stanford Encyclopedia of Philosophy. Retrieved from https://plato.stanford.edu/archives/sum2018/entries/simulations-science.

Young, W., Sobhani, A., Lenné, M. G., and Sarvi, M. (2014). Simulation of Safety: A Review of the State of the Art in Road Safety Simulation Modelling. Accident Analysis & Prevention, 66, 89-103.

U.S. Department of Transportation
Federal Highway Administration