Verification and Calibration of Microscopic Traffic Simulation Using Driver Behavior and Car-Following Metrics for Freeway Segments

PUBLICATION NO. FHWA-HRT-24-103

APRIL 2024

U.S. Department of Transportation
Federal Highway Administration

Research, Development, and Technology
Turner-Fairbank Highway Research Center
6300 Georgetown Pike
McLean, VA 22101-2296

FOREWORD

This research was led by the Federal Highway Administration under Transportation Pooled Fund (TPF) study TPF-5(361), Second Strategic Highway Research Program (SHRP2) Naturalistic Driving Study Pooled Fund: Advancing Implementable Solutions (https://www.pooledfund.org/Details/Study/613), to develop novel, multidisciplinary solutions based on the recorded natural behavior of vehicle operators interacting with infrastructure and other vehicles (National Cooperative Highway Research Program, 2023). Data from the SHRP2 naturalistic driving study (https://insight.shrp2nds.us/) were used in this project to better understand car-following behavior from passenger vehicles on freeway segments (Virginia Tech Transportation Institute, 2020). A calibration process is proposed for microscopic traffic simulation to ensure that simulated vehicle-to-vehicle interactions reflect naturalistic behavior. The process complements calibration practices focused on macroscopic metrics.

The analysis produced detailed driving behavior distributions in terms of three main metrics: spacing between the instrumented vehicle and a leader in the same lane, acceleration of the instrumented vehicle, and acceleration change rate of the instrumented vehicle. The research team explored the potential benefits of incorporating measures derived from naturalistic data into traditional safety modeling. At a macroscopic scale, results from crash frequency prediction models showed consistent and significant effects associated with increases in multivehicle crash frequencies when the variance of the density increased, the variance of the speed increased, and the mean spacing decreased. The research team developed an open-source tool, the Naturalistic Assessments of Car-Following Trajectories tool, to implement the processes described in the report. This research will be of interest to traffic operations and highway safety professionals, and others with an interest in mitigating disruptions to vehicle flow and improving safety.

Brian P. Cronin, P.E.
Director, Office of Safety and Operations
Research and Development

Notice

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The U.S. Government assumes no liability for the use of the information contained in this document.

Non-Binding Contents

Except for the statutes and regulations cited, the contents of this document do not have the force and effect of law and are not meant to bind the States or the public in any way. This document is intended only to provide information regarding existing requirements under the law or agency policies.

Quality Assurance Statement

The Federal Highway Administration (FHWA) provides high-quality information to serve Government, industry, and the public in a manner that promotes public understanding. Standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information. FHWA periodically reviews quality issues and adjusts its programs and processes to ensure continuous quality improvement.

Disclaimer for Product Names and Manufacturers

The U.S. Government does not endorse products or manufacturers. Trademarks or manufacturers' names appear in this document only because they are considered essential to the objective of the document. They are included for informational purposes only and are not intended to reflect a preference, approval, or endorsement of any one product or entity.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA-HRT-24-103	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Verification and Calibration of Microscopic Traffic Simulation Using Driver Behavior and Car-Following Metrics for Freeway Segments		5. Report Date April 2024 6. Performing Organization Code:
7. Author(s) Juan C. Medina (ORCID: 0000-0001-5302-8814), Arman Malekloo (ORCID: 0000-0003-0188-3098), Kristin Kersavage (ORCID: 0000-0002-3601-7766), Richard, J. Porter (ORCID: 0000-0001-8535-3451), Xiaoyue (Cathy) Liu (ORCID: 0000-0002-5162-891X)		8. Performing Organization Report No.
9. Performing Organization Name and Address University of Utah 201 Presidents Circle Salt Lake City, UT 84112 VHB® 940 Main Campus Drive, Suite 500 Raleigh, NC 27606		10. Work Unit No. 11. Contract or Grant No. DTFH6116D00040L
12. Sponsoring Agency Name and Address Office of Safety and Operations Research and Development Federal Highway Administration 6300 Georgetown Pike McLean, VA 22101-2296		13. Type of Report and Period Covered Final Report; December 2019–February 2023 14. Sponsoring Agency Code HRSO-20
15. Supplementary Notes		

16. Abstract

This research leveraged the Second Strategic Highway Research Program naturalistic driving study (NDS) datasets (https://insight.shrp2nds.us/) to improve understanding of driving behavior, particularly car-following behavior from passenger vehicles on freeway segments under good weather and daytime conditions, and proposed a calibration process for microscopic traffic simulation to ensure that simulated vehicle-to-vehicle interactions reflect naturalistic behavior (Virginia Tech Transportation Institute, 2020). The proposed calibration complements typical calibration practices focused on macroscopic metrics such as travel time, delay, and queues. More than 1,600 h of car-following data from more than 1,700 unique drivers were used as a basis to create driving behavior distribution targets and represent naturalistic driving in terms of vehicle spacing, acceleration, and jerk at different speed levels. The research team developed the portable, open-source Naturalistic Assessments of Car-Following Trajectories tool to read generic trajectories from microscopic simulation, extract car-following behavior, and perform statistical comparisons against the NDS targets to guide the calibration process. In addition, the research team explored the potential benefits of incorporating macroscopic measures derived from NDS into traditional safety modeling, which indicated that an increase in the traffic density variance, increase in the speed variance, and decrease in the mean vehicle spacing had significant effects associated with increases in multivehicle crash frequencies. Potential improvements to vehicle conflict data analysis using the Surrogate Safety Assessment Model were also evaluated, pointing to conflict frequencies and locations that more closely resemble those from observed crash events when simulated sites were calibrated for vehicle-to-vehicle interactions (FHWA, 2022).

17. Key Words	18. Distribution Statement
Naturalistic driving study, SHRP2, car-following, microscopic simulation, calibration, safety modeling	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161. https://www.ntis.gov

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	68	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 1,000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2,000 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 (2,000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	2.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for 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 ACRONYMS

AADT	annual average daily traffic
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
CDF	cumulative distribution function
CLI	command line interface
CVM	Cramer–Von Mises
DKW	Dvoretzky–Kiefer–Wolfowitz
DOT	department of transportation
GPS	Global Positioning System
HSIS	Highway Safety Information System
KML	keyhole markup language
K-S	Kolmogorov–Smirnov
MV	multiple vehicle
NACT	Naturalistic Assessments of Car-Following Trajectories
NDS	naturalistic driving study
RID	Roadway Information Database
SAS	Statistical Analysis System
SHRP2	Second Strategic Highway Research Program
SSAM	Surrogate Safety Assessment Model
TPF	transportation pooled fund
USGS	United States Geological Survey
VTTI	Virginia Tech Transportation Institute
vpd	vehicles per day

EXECUTIVE SUMMARY

This research leveraged Second Strategic Highway Research Program naturalistic driving study (NDS) datasets to improve the understanding of driving behavior, particularly of car-following behavior from passenger vehicles on freeway segments under good weather and daytime conditions, and proposed a calibration process for microscopic traffic simulation to ensure that simulated vehicle-to-vehicle interactions reflect naturalistic behavior (Virginia Tech Transportation Institute, 2020). The proposed calibration is intended to complement typical calibration practices focused on macroscopic metrics such as travel time, delay, and queues. Final datasets for analysis included more than 1,600 h of car-following data from more than 1,700 unique drivers in three States. The analysis produced detailed driving behavior distributions in terms of three main metrics: spacing between the instrumented vehicle and a leader in the same lane, acceleration of the instrumented vehicle, and acceleration change rate (jerk) of the instrumented vehicle. In addition, the research team obtained different distributions for a wide range of traffic conditions represented by driving speeds and categorized using overlapping ranges between 5 mph and 85 mph. The distributions served as targets to verify if simulation outputs display a naturalistic car-following behavior.

The research team developed the portable, open-source Naturalistic Assessments of Car-Following Trajectories (NACT) tool to implement the necessary steps to transform generic trajectory data from simulation, produce leader-follower pairs, and perform statistical comparisons with the NDS targets.[1]

The research team also explored the potential benefits of incorporating measures derived from NDS into traditional safety modeling. At a macroscopic scale, results from crash frequency prediction models showed consistent and significant effects associated with increases in multivehicle crash frequencies when the variance of the density increased, the variance of the speed increased, and the mean spacing decreased. In addition, at the microscopic level, the research team analyzed the effects of calibrating simulation using the NDS targets in terms of potential improvements for safety analysis. Comparisons of simulated vehicle conflict data in low-traffic conditions using the Surrogate Safety Assessment Model were preliminary but pointed to encouraging results (Federal Highway Administration (FHWA), 2022). Overall, the research team observed that conflict frequencies and their locations resembled slightly more closely those from observed crash events when simulated sites were calibrated using the NDS targets. However, the research team recommended additional research to evaluate a comprehensive set of traffic conditions, including high-traffic conditions and high-demand fluctuations to better quantify to benefits of calibration.

[1]The NACT tool was developed as a part of this FHWA project.

CHAPTER 1. INTRODUCTION AND BACKGROUND

Microscopic traffic simulation is an invaluable tool for assessing the traffic performance of highway and street systems (Dowling, Skabardonis, and Alexiadis, 2004). Simulation enables not only the replication of existing traffic scenarios, but also the analysis of conditions when data are difficult to collect and the evaluation of "what if" scenarios with alternative geometries or traffic demands. However, building and setting up simulated scenarios is a complex task that requires detailed knowledge of the simulation tool at hand, as well as field data to calibrate simulation parameters to represent local traffic behavior. Thus, calibration involves adjusting microscopic model parameters to minimize differences between metrics produced by the simulation and those observed or representative from the field (Zhu, Wang, and Tarko, 2018).

Traditionally, calibration of microscopic traffic simulation is conducted to reproduce overall system performance measures, such as travel time, delay, and queues (Dowling, Skabardonis, and Alexiadis, 2004; Ge and Menendez, 2012; Gomes, May, and Horowitz, 2004; Park and Qi, 2006; Lownes and Machemehl, 2006). Such target measures represent macroscopic traffic characteristics that are the result of interactions between individual vehicles modeled in the simulation at a microscopic level.

Some State department of transportation (DOT) guidelines suggest calibration of microscopic simulation to match macroscopic outcomes and provide recommendations on the type of parameters to be adjusted and a range of typical values; examples include calibration documents from DOTs in Florida, Oregon, Virginia, and Wisconsin, where the guidance is intended to cover targets for macroscopic metrics without mention of microscopic-level calibration (Florida DOT, 2021; Oregon DOT, 2011; Virginia DOT, 2020; Wisconsin DOT, 2019). In addition, the 2019 updated version of *Traffic Analysis Toolbox Volume III—Guidelines for Applying Traffic Microsimulation Modeling Software* provides generalized steps for calibration at local and system-wide levels using macroscopic target variables and includes driver characteristics as a data source with potential for calibration, even though driver characteristics are difficult to collect (Wunderlich, Meenakshy, and Peiwei, 2019). Additionally, *Traffic Analysis Toolbox Volume III—Guidelines for Applying Traffic Microsimulation Modeling Software* describes a group of variables related to driver behavior to be calibrated in the simulation toward meeting macroscopic targets but does not offer further guidelines to ensure adequate microscopic behavior at the individual vehicle level.

Although macroscopic targets reflect certain field conditions, such calibration does not evaluate whether interactions between individual vehicles reflect naturalistic behavior. Lack of microscopic-level calibration is largely because collecting this type of data is difficult, impractical, and costly, but enhancing the simulation's ability to emulate a wider range of traffic scenarios and to support new applications such as surrogate safety analysis could be beneficial (Balakrishna et al., 2007; Kaths, Keler, and Bogenberger, 2021).

Microscopic target metrics include vehicle headways, spacing, speed, and acceleration distributions. Limited headway, spacing, and speed data can be collected for specific freeway sections at sensor stations, but tracking following behavior is only possible through probe or naturalistic driving studies (NDS). Naturalistic datasets provide real-world driving data and

present unique opportunities over several applications, including the calibration of car-following models (Sangster, Rakha, and Du, 2013).

Not surprisingly, given the potential benefits of NDS and the advance of onboard technology for vehicle data collection, recent years have seen an expansion in the use of NDS for analyzing microscopic behavior (Zhu, Wang, and Tarko, 2018; Abbas et al., 2011; Higgs and Abbas, 2014; Higgs, 2011; Sangster, Rakha, and Du, 2013). For example, Sangster, Rakha, and Du (2013) used processed data from 8 drivers, including more than 2,000 car-following events from the 100-car study by the Virginia Tech Transportation Institute (VTTI), to calibrate 4 car-following models (VTTI, 2020). These models include the Gipps, intelligent driver, Gaxis-Herman-Rothery, and Rakha-Pasumarthy-Adjerid models (Gipps, 1981; Treiber, Hennecke, and Helbing, 2000; Chandler, Herman, and Montroll, 1958; Rothery et al., 1964; Van Aerde and Rakha, 1995). Zhu, Wang, and Tarko also calibrated car-following models but used a genetic algorithm approach and driving data collected by the NDS in Shanghai, China (Zhu, Wang, and Tarko, 2018). This study used vehicle spacing as the measure of performance in the car-following models, which was extracted from 2,100 car-following periods by 42 drivers. Lastly, other researchers used NDS to explore the influence of environmental conditions on car-following behavior, focusing on how to incorporate these aspects to better calibrate specific simulation models (Geng et al., 2016; Hammit, James, and Ahmed, 2018).

The availability of the Second Strategic Highway Research Program (SHRP2) NDS has created new opportunities for calibrating microscopic traffic simulation (NCHRP, 2023; VTTI, 2020). Naturalistic data from the SHRP2 NDS allows the study of vehicle traversals on a large scale from a significant number of drivers traveling in real-world roadway conditions. A large body of literature relates to the SHRP2 NDS, with the project background, data description, dictionaries, and summary analyses available from the InSight data access website (VTTI, 2020).

Additional datasets collected as part of NDS and compiled in the Roadway Information Database (RID) enhance data usability, providing information related to roadway characteristics, geolocation of roadway segments, work zones, etc. (Iowa State University, 2023). In particular, this research made extensive use of roadway layers for spatial data processing, as well as crash datasets.

This research leverages SHRP2 NDS data to improve understanding of driving behavior, particularly car-following behavior from passenger vehicles on freeway segments under good weather and daytime conditions, which can be directly used to calibrate car-following models and traffic simulation (NCHRP, 2023). Longitudinal vehicle-to-vehicle interactions (i.e., car-following behaviors) are at the center of the study and provide the basis for calibrating simulation trajectories at a microscopic level.

The research team used a two-part approach to conduct the research. Part A focused on identifying and estimating suitable driver behavior metrics so that the research team could assess the potential of such metrics to characterize freeway traffic conditions at a microscopic level. Part A also used an NDS dataset that was readily available to the research team, which sped up the initial process to determine the feasibility of a larger, more complete analysis covering additional conditions in part B (NCHRP, 2023).

Part B expanded on the analysis of driving behavior metrics conducted in part A by covering freeway segments and traffic conditions that may not have been well represented in the initial dataset. Part B activities included identifying, requesting, and postprocessing additional NDS traversals. The research team used these supplemental data, combined with the initial dataset from part A, to generate characterizations of driver behavior for a wide range of traffic conditions on several freeway segments. The research team used these behaviors as targets to assess whether trajectories from microscopic simulation exhibit a naturalistic behavior, allowing for a definition of a proposed microscopic calibration procedure.

This report summarizes the research conducted through parts A and B and introduces a separate stand-alone document containing a guideline to use the Naturalistic Assessments of Car-Following Trajectories (NACT) tool.[2] Researchers can use the NACT tool to extract car-following behavior from user-specified simulation output files, process the input file, and compare the summary data to the NDS targets. The NACT tool serves as a key element in the proposed calibration process.

MAIN OBJECTIVE

This research's main objective is to characterize microscopic driving behavior and derive car-following metrics from NDS datasets collected on freeway segments in good weather conditions to develop model-independent guidelines that can enhance the calibration and verification of microscopic traffic simulation (VTTI, 2020).

This research is intended to provide practitioners and researchers with new criteria to evaluate simulation from a microscopic point of view, complementing typical calibration efforts using macroscopic system performance measures.

SECONDARY OBJECTIVE

A secondary, but complementary, objective is to capitalize on the availability of associated RID data to evaluate whether enhanced microscopic traffic simulation resulting from the primary objective would also enhance the ability of the simulation to produce metrics that can be used to better evaluate safety outcomes with safety surrogates (Iowa State University, 2023).

This report is organized into the following seven chapters:

- Chapter 1 provides introduction and background information and lists the overall objectives of this project as defined in the initial proposal.

- Chapter 2 describes data collection, including the data initially processed in part A, the data requested in part B, and the necessary transformations to begin the extraction of car-following metrics from NDS.

[2]The NACT tool was developed as a part of this FHWA project.

- Chapter 3 includes the data processing steps once the NDS datasets were preprocessed and contains a description of the extraction of leader-follower pairs and the treatment of traversal data to extract metrics that describe car-following behaviors (VTTI, 2020).

- Chapter 4 summarizes the extracted NDS targets in terms of vehicle spacing, acceleration, and jerk for all speed groups.

- Chapter 5 describes the proposed process to incorporate the tool into calibration, the required fields for extraction of behaviors from simulated traversals, and the comparisons proposed to use the NDS targets and determine adequacy of the simulation in terms of car-following behavior.

- Chapter 6 describes the exploration conducted in relation to the secondary outcomes related to safety analysis both at macroscopic and microscopic scales.

- Chapter 7 offers conclusions and recommendations.

CHAPTER 2. DATA COLLECTION AND PREPARATION

As described in chapter 1, Introduction and Background, the project included two main parts: part A and part B. Part A used datasets readily available to the research team, and part B complemented the initial datasets to strengthen the initial outcomes.

For both part A and part B datasets, the research team obtained site geometric characteristics and traffic demands, including the following information:

- Site and geometric characteristics: Site length, ramp spacing, number of lanes upstream and downstream of ramps, presence of auxiliary or high-occupancy vehicle lane, and speed limit.

- Traffic demands: Annual average daily traffic (AADT) on mainline and ramps for years when NDS data were collected.

The research team enhanced data processing by generating keyhole markup language (KML) files for each site, allowing for precise measurements and geographic referencing (Google® for Developers, n.d.). Figure 1 shows a sample image from one of the sites in North Carolina that includes the site boundaries. All KML files are available from the research team on request.

Original map: © 2018 Google® Earth™: Modified by FHWA (see Acknowledgments section) (Google®, 2023).

Figure 1. Map. Aerial view of a sample site in North Carolina with site boundaries denoted by a polygon.

PART A DATASETS

The initial datasets from part A included about 17,500 traversals from 56 sites—22 sites located in Washington and 34 sites located in North Carolina. The research team conducted site selection for part A specifically within those two States and focused on freeway segments with closely spaced ramps and similar ramp configurations, mostly diagonal ramps. The research team filtered part A data to only include traversals from instrumented vehicles traversing through the sites, removing instances when the instrumented vehicles entered or exited the site using the ramps. About one-third of the data were still available for the initial exploration phase.

Traversals allowed the study of vehicle interactions near interchanges, opening opportunities to capture car-following data under low-traffic conditions, as well as under higher traffic conditions and under the effect of disturbances due to entering and exiting traffic. Such a combination of scenarios would have been difficult to encounter at more isolated sites.

Analysis from part A confirmed the adequacy of the traversals and produced consistent results in terms of spacing, speed, and acceleration data. However, while more than 100 h of car-following data were available from part A, most conditions represented high-speed scenarios, including only about 5 h of combined car-following data with sustained speeds lower than 50 mph. This finding prompted the need to expand part A datasets, particularly to strengthen the data under lower-speed conditions.

PART B DATASETS

During part B, the research team identified and selected new sites to cover the gaps the team identified in part A. The RID was the primary source of data to identify and select candidate sites (Iowa State University, 2023). Given the need to expand traversal data with higher traffic conditions with lower speeds, the research team prescreened freeway segments with high AADT, considering sites in all six geographical areas and States where NDS data were collected (Tampa, FL; Bloomington, IN; Durham, NC; Buffalo, NY; State College, PA; and Seattle, WA). From this effort, the research team identified sections with high traffic demands and standard freeway sections with parallel ramps from sites in the Seattle-Tacoma, WA, area and the Tampa, FL, area. Segment selection favored locations in the upper range of AADTs and avoided innovative interchange configurations and ramps requiring significant speed reduction, such as loop ramps.

The research team carried out exploration for part B to determine a sample and select several traversals from the total available data at these locations. The research team provided VTTI a list of selected roadway segments by using their link identifications from the RID and received back a series of snapshots indicating that each of the 18 routes in both States (Washington and Florida) had more than 5,000 traversals and a wide distribution of driver age (Iowa State University, 2023). Each route from part B may include multiple interchanges, and thus the route can contain multiple individual sites, where one site is simply confined to a segment between two consecutive interchanges. Therefore, routes could be significantly longer than individual sites, and, in the research team's case, the routes ranged between 3 and 10 mi. Overall, all 18 routes represented more than 100 mi of roadways.

A review of the snapshots from the new routes from part B resulted in the preselection of about 15,000 traversals that VTTI processed for final extraction. Selected traversals covered good weather conditions and a random selection of times of day (all selected traversals occurred between 7 a.m. and 8 p.m.), avoiding typical nighttime conditions. Additionally, less than 5 percent of the traversals occurred before 8 a.m. or after 6 p.m. between October and March, reducing partial dark conditions in the dataset.

On receipt of the traversal data, the research team began a series of preprocessing steps to prepare the data for car-following behavior extraction. These steps were similar to those followed in part A and are described in the following sections.

NDS DATA PREPROCESSING

Datasets from parts A and B were initially subjected to quality checks for consistency and completeness. The research team digitized some geographic information (e.g., site pavement marking lines, gore locations) about each site so that the instrumented vehicle could be used for analysis in MathWorks® MATLAB® (MathWorks, 2022). In this case, the United States Geological Survey (USGS) high-resolution orthoimagery gathered from the national map interface extracted the approximate coordinates of the pavement markings and gore and taper locations in ArcGIS® (USGS, 2016; Esri, 2015).

The research team primarily used pavement marking lines, or edge lines, to identify the relative location of an instrumented vehicle along a site over the duration of a traversal. By matching a Global Positioning System (GPS) coordinate to the closest point on the pavement marking line, the position of the instrumented vehicle can be identified in reference to the points of interest and from the beginning of the site.

The data the research team received from VTTI were in a standard file format (comma-separated values) and included all fields and time series data for each traversal (VTTI, 2020). Once the data were read in MATLAB using internal functions (e.g., textscan, xlsread), the research team completed two basic functions: identified columns of time-series data associated with specific variables, and performed basic data cleaning based on identifiable problems (MathWorks, 2022).

After reading the data, the research team sorted the data in a MATLAB structure array (MathWorks, 2022). The structure array creates a single object that may contain multiple fields to store many different types of data associated with that object. In this case, each traversal acts as an object, and the different time-series variables were assigned to the object using unique field names. This data structure also made it possible to create a group of structure arrays, where a single variable object can be used to store and pass a large number of traversals and their associated data.

Some basic data-handling processes were also necessary in preparation for analyzing the time-series data from the instrumented vehicles. The most common circumstance involved handling missing data in time-series data, primarily due to data points collected at different frequencies for different variables (i.e., asynchronous series). This circumstance was the case when the research team analyzed lane position data (collected every 0.1 s) to detect a lane change and wanted to associate it with the GPS location of the vehicle (collected approximately

every 1 s). Alternatively, an observation might have just been dropped, leaving a small, empty space in a longer stream of data. The research team used simple linear interpolation to estimate values for variables at time stamps between observations as long as the data gap was at most 2 s, with larger gaps in data resulting in the time series not being further considered in the analysis. This procedure was necessary for analyzing most time-series variables, including lane position, GPS location, speed, and radar variables.

The research team also defined data transformations to make raw data simpler to work with. For example, the team developed a function to translate radar range and range rate of change into vehicle positions and speeds for nearby objects captured via radar. Since speed data could be available from multiple sources (GPS and instrumented vehicle), another function identifies the speed variable that contains the most information, performs some common-sense checks (e.g., check for unrealistic speeds), and sets the information aside for further analysis.

Some amount of processing was necessary for analyzing the location of time-series data based on GPS coordinates. This processing generally involved changing the projection of the GPS coordinate data for use with local projections, interpolating coordinate data, and calculating the relative position or distance for the available coordinate data for each traversal.

Lastly, the research team set vehicle location controls to identify anomalies in vehicle traversals and prevent unexpected metrics. For example, an instrumented vehicle may travel beyond the selected start and end points of a study site, or the trajectories may use a ramp to exit and then reenter the site. These situations were uncommon, but their removal was essential in maintaining the quality of the analysis.

After interpolating the vehicle position, the research team used an algorithm developed for NDS data to give each vehicle a relative location or position at each study site. To provide this common reference, the research team assigned each site a roadway edge line, which allowed each trajectory to be reprojected along the edge line so that distances could be calculated with a common starting point and the distances could all be calculated along the same path (removing variations in distance due to small lateral position changes).

CHAPTER 3. NDS CAR-FOLLOWING DATASETS

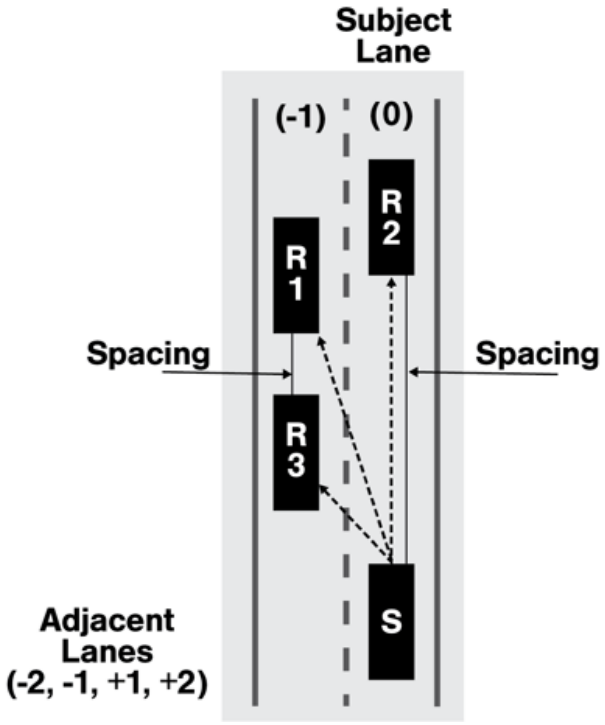
This analysis focused on car-following behavior on freeway segments, specifically when vehicles are traveling through mainline lanes. This focus excludes the behavior of drivers on entry or exit ramps, drivers using acceleration or deceleration segments as they leave or enter ramps, and drivers in the process of changing lanes.

Time series for each traversal contained essential data to extract the car-following behavior, including vehicle trajectory (i.e., vehicle coordinates), instrumented vehicle speed, and processed radar data identifying objects detected by the front-facing radar, their longitudinal and lateral range, and approximate lane width with respect to the instrumented vehicle.

To extract car-following behavior, the research team analyzed the radar data in combination with the instrumented vehicle trajectory to develop leader-follower time series. This new time series included vehicle spacing between the instrumented vehicle acting as a follower, and a downstream vehicle in the same lane acting as a leader. The conceptual process to extract leader-follower data is described in the Leader-Follower Time Series section.

LEADER-FOLLOWER TIME SERIES

By combining information from the radar and constructing traversals from the instrumented vehicles, the research team could analyze car-following behavior in terms of vehicle spacing between a vehicle acting as a follower and a vehicle ahead in the same lane acting as a leader. Along with spacing, changes in vehicle position every 0.1 s provide an indication of speed and a corresponding acceleration, completing a picture for the vehicle kinetics in the longitudinal direction and describing car-following behavior from a microscopic standpoint. The basic logic to estimate vehicle spacing for the instrumented vehicle and radar objects is illustrated in figure 2, where the dashed arrows represent the range measured from the instrumented vehicle (*S*) to nearby radar objects (*R1*, *R2*, and *R3*) and is the basis for the spacing estimation. Spacing can be calculated provided that vehicles are traveling in the same lane (e.g., *R1* and *R3*, *S* and *R2*), as identified by an indicator where 0 is the subject lane, and −2, −1, +1, or +2 are adjacent lanes to the left (indicated with a "−") and right (indicated with a "+") of the instrumented vehicle.



Source: Federal Highway Administration (FHWA).

Figure 2. Illustration. Conceptual representation of spacing calculation using radar data from subject instrumented vehicle (S).

The research team's initial calculations included metrics such as spacing for leader-follower pairs in adjacent lanes, as shown in figure 2; however, the final datasets developed NDS metrics only for pairs involving the instrumented vehicle (i.e., pairs when the leader was in the subject lane, or lane 0). Observations from adjacent lanes were beneficial to enhance the number of traversals in the analysis, but were inherently less reliable than those involving the instrumented vehicle because both the spacing between the two detected vehicles and their exact distance from the instrumented vehicle highly depended on what part of the vehicle was detected by the radar (e.g., the rear bumper, or the middle portion of the vehicle body, or even its front corner), increasing uncertainty. Additionally, driver information was not available, further limiting posterior analysis.

With the expansion of the datasets from part A to part B, a wider set of traffic conditions were expected to be covered by a larger number of traversals, also reducing the need to rely on additional data from adjacent lanes. The process to estimate traffic conditions from the trajectory data was essential in understanding how to analyze the car-following behavior and is described in the Traffic State Estimations section of this report.

TRAFFIC STATE ESTIMATIONS

Given the dependency of car-following behavior on existing traffic conditions, the extraction of data from NDS and the development of target metrics required contextualization and quantification of such conditions. However, since the NDS datasets do not provide such

estimations, the research team conducted exploration to identify potential metrics for this purpose (VTTI, 2020).

From the fundamental traffic diagram, speed, flow, and density were candidate metrics to describe the traffic conditions NDS drivers experienced along the traversals. From those three metrics, only speed was readily available, requiring additional analysis to develop estimates for flow and density.

The research team borrowed a method to obtain these estimates from ideas commonly applied to probe vehicle data, where, in this case, the instrumented vehicle and radar objects act as probe vehicles (Seo, Kusakabe, and Asakura, 2015). The methodology is based on definitions of speed and density using an area within the time-space diagram as shown in the equations in figure 3.

$$\widehat{q_A} = \frac{\sum_{n \in N(A)} d_n(A)}{\sum_{n \in N(A)} a_n(A)} \qquad \widehat{k_A} = \frac{\sum_{n \in N(A)} t_n(A)}{\sum_{n \in N(A)} a_n(A)} \qquad \widehat{v_A} = \frac{\sum_{n \in N(A)} d_n(A)}{\sum_{n \in N(A)} t_n(A)}$$

Figure 3. Equation. Flow, density, and speed for traffic state estimations.

Where:

$\widehat{q_A}$ = estimated flow.
$\widehat{k_A}$ = estimated density.
$\widehat{v_A}$ = estimated speed.
$d_n(A)$ = distance traveled.
$t_n(A)$ = time spent by a known vehicle n in segment A.
$a_n(A)$ = the time-space region in A observed by combinations of known vehicles.

The research team applied the summation terms in these equations over all known vehicles N observed in each segment A. Known vehicles include both instrumented vehicles and radar objects.

To explore the relationships between the three fundamental metrics, the research team divided traversals into speed groups using the speed of the instrumented vehicle to generate the following general classifications of traffic:

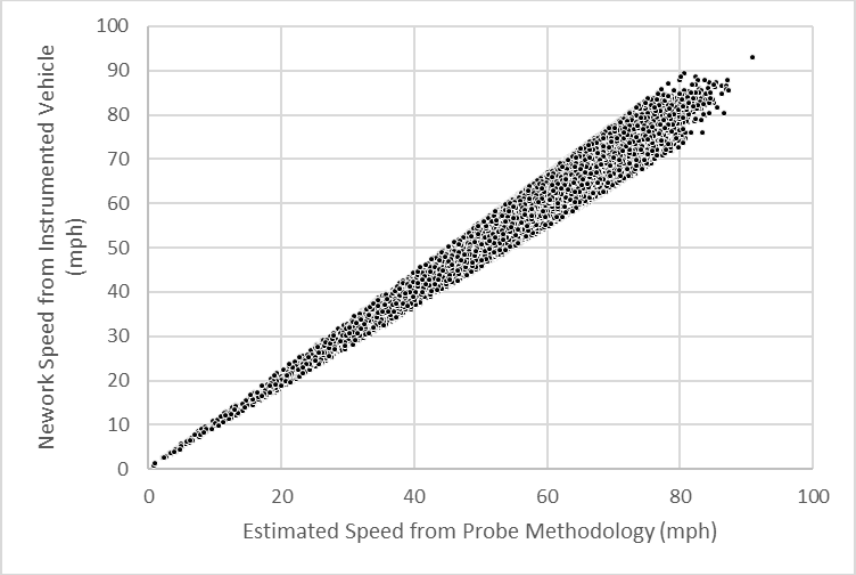
- Low traffic (speeds higher than 60 mph).
- Medium traffic (speeds between 40 and 60 mph).
- High traffic (speeds lower than 40 mph).

The research team partitioned traversals based on their sustained speed, similar to the process applied to identify the period of sustained following from the leader-follower pairs. Thus, each partition is considered to provide a representation of the traffic conditions at the time of the traversal.

The research team applied the methodology to estimate flow, density, and speed to each partition. The team obtained more than 50,000 partitions from this exercise, representing 780 h of driving including 637 h in low traffic, 115 h in medium traffic, and 28 h in high traffic.

Results provided consistent trends, with the speed estimations being the most reliable among the three, followed by density and estimates, which was expected, because the flow could amplify poor observations because flow is the one metric that uses two estimated values (i.e., distance traveled and area in the space-time diagram), whereas speed and density use direct measurements of time.

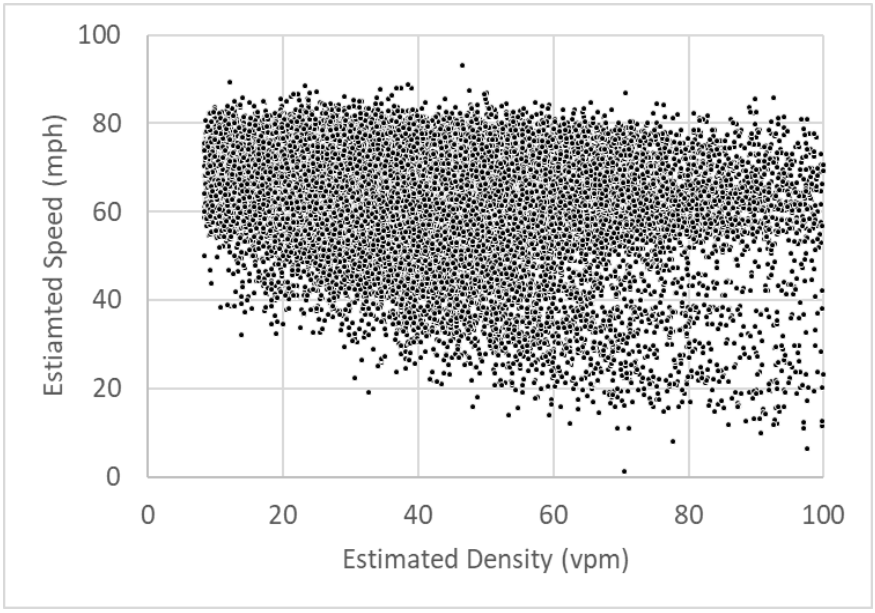
Verification of the speed estimates confirms the consistency of the calculations, as shown in figure 4, where the speeds from the methodology are compared to the network speeds from NDS.



Source: FHWA.

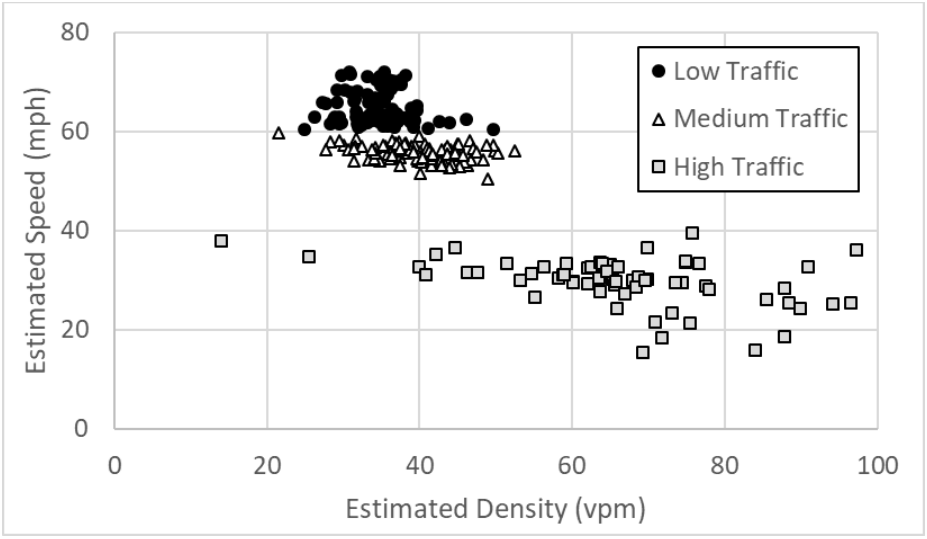
Figure 4. Graph. Comparison of calculated speeds from probe methodology and NDS speeds.

Similarly, the research team analyzed estimated densities to identify trends in relation to the speed data. The a priori expectation was to observe a decrease of speed with an increase of density, with the data confirming this trend, particularly when aggregated at the site level. Figure 5-A and figure 5-B show individual density-speed pairs and the density-speed aggregates by site, respectively.



Source: FHWA.
vpm = vehicles per mile per lane.

A. Example of observations from individual partitions or traversals.



Source: FHWA.

B. Example of observations at site level.

Figure 5. Graphs. Speed-density plots from NDS data.

Speeds lower than 8 mph are not observed, even though a large number of individual speed values at specific time stamps were this value and even zero, because each point represents the average sustained speed within the selected speed bins, potentially including averages along the complete traversal. In addition, aggregation of traffic condition categories based on speeds resulted in more stratified speed levels compared to the density estimates in figure 5-B. Additional subcategorization of conditions by speed will likely provide smoother transitions

from highest to lowest speeds, but, in turn, will affect the accuracy of the density estimations, as the area in the denominator of figure 3 will become increasingly smaller, becoming more susceptible to noise or lack of observations. This analysis helped support the idea of identifying general traffic density by using speed, as expected from the fundamental diagram of traffic flow, and supported the use of speed to represent traffic conditions when characterizing driving behavior.

In addition, through this analysis, the research team generated NDS targets using a number of speed bins instead of producing metrics for specific low-, medium-, or high-traffic conditions. Speed bins provide a more flexible set of conditions and an intuitive breakdown of traffic conditions, compared to traffic classifications such as low, medium, or high traffic, and even further breakdowns such as medium-high traffic. After further analysis of alternatives (including traffic condition levels), the research team defined the following 10 speed bins for NDS target extraction:

- 5–20 mph.
- 15–25 mph.
- 20–35 mph.
- 30–40 mph.
- 35–50 mph.
- 45–55 mph.
- 50–65 mph.
- 60–70 mph.
- 65–80 mph.
- 75–85 mph.

The research team analyzed traffic conditions and corresponding car-following behavior for observations when a leader-follower pair fit into one of the speed bins. Furthermore, the team only analyzed leader-follower pairs if their behavior was observed for at least 10 s consecutively. This condition was aimed at targeting only car-following instances where the follower had established a following behavior, removing instances of short leader-follower interactions. Setting a minimum time for leader-follower analyses is consistent with previous NDSs, where minimum thresholds have been used to characterize following behavior (Zhu, Wang, and Tarko, 2018; LeBlanc et al., 2013; Chong et al., 2013; Fernandez, 2011).

The speed bins overlap and the lowest speed bin starts from 5 mph. Overlapping bins increase flexibility to analyze driver behavior, as larger portions of the traversals could fit in one of the bins with sustained speeds within the bin boundaries. For example, without the overlapping bin 30–40 mph, a driver traveling between 32 mph and 38 mph would not fit a single bin, and the driver's behavior would have to be split, potentially leading to removing the data as a whole due to short following duration.

The research team set the minimum speed of 5 mph to prevent confounding effects closely related to standstill conditions, as opposed to active following behavior, which also allowed the development of preliminary distributions of standstill distance, even though development of preliminary distributions of standstill distance was not a primary outcome of the study.

The Final NDS Dataset section describes the final dataset that the research team used for analysis, before describing the NDS targets obtained for simulation (NCHRP, 2023).

FINAL NDS DATASET

The final NDS dataset combined part A and part B sites, for a total of 104 urban and suburban freeway sites located in 3 out of the 6 States with NDS data (North Carolina, Washington, and Florida), and included a wide variety of characteristics, including speed limits, number of lanes, and traffic conditions (NCHRP, 2023). Summary statistics for site characteristics are shown in table 1. Site selection in part B focused on higher AADT values and conventional freeway segments without loop ramps. Most of the explored sites in Bloomington, IN; Buffalo, NY; and State College, PA had much lower AADTs than in Seattle, WA, and Tampa, FL, resulting in the last two States being selected for part B datasets.

Table 1. Summary statistics selected sites for data analysis.

Characteristics	Number of Sites	Average	Maximum	Minimum
Directional AADT (vehicles per day)	104	56,440	101,333	15,833
Speed limit (mph)	104	62.1	70	55
Number of lanes	104	3.4	5	2
Presence auxiliary lane	104	0.34	1	0
Site length (miles)	104	1.62	4.21	0.35

Overall, the research team extracted leader-follower pairs from 15,963 traversals encompassing data from 1,738 unique drivers in 3 States. Driver demographics followed a similar distribution to those in the NDS complete datasets in terms of driver age, as shown in table 2 (NCHRP, 2023). This distribution does not follow the same age breakdown as the average U.S. driving population, so the research team also performed additional analysis to provide alternative characterizations of driver behavior that more closely represent national averages.

Table 2. Driver age distribution in final NDS dataset (*n* = 1,738).

Driver Age (yr)	Proportion (percent)
16–24	36
25–34	16
35–44	7
45–54	9
55–65	10
65–74	11
75+	11

Table 3 shows a detailed breakdown of the data in terms of total sustained car-following time used to extract the NDS targets. Total car-following times reflect final driving times already filtered by data quality checks and meeting the minimum duration of 10 s within 1 of the 10 predefined speed bins. Note the significant contribution of longer traversals from the part B datasets compared to those in part A.

Table 3. Summary final dataset and total car-following times by project phase and State.

Project Phase	State	Sites (no.)	Unique Drivers	Total Sustained Car-Following Time (hours)
Part A	NC	34	506	55.72
Part A	WA	22	575	46.32
Part B	FL	26	550	754.3
Part B	WA	22	582*	746.2

*A total of 475 drivers in Washington from part A datasets also appeared among the 582 drivers in part B. The total number of unique drivers in parts A and B combined when excluding this overlap is 1,738.
NC = North Carolina; WA = Washington; FL = Florida.

CHAPTER 4. EXTRACTION OF NDS TARGETS

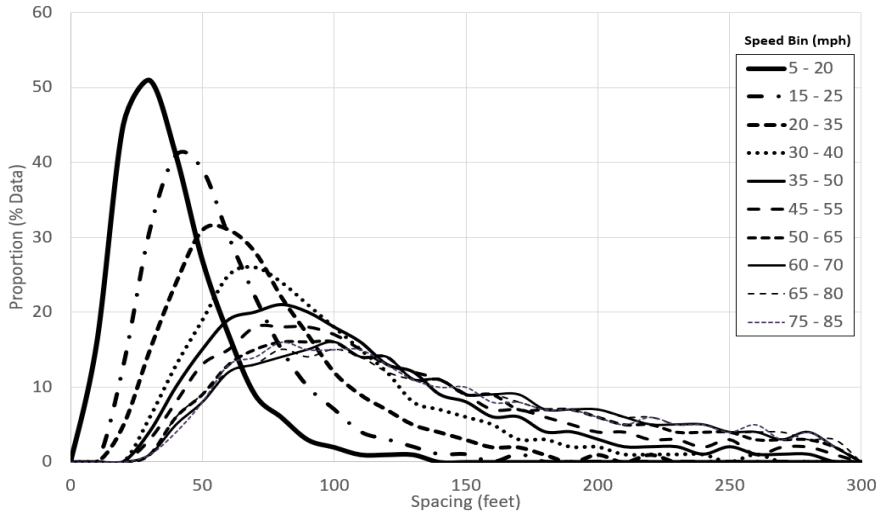
After defining the final dataset for analysis, the research team extracted car-following data from the leader-follower files for the periods of sustained behavior and within a given speed bin. Metrics of interest included spacing, acceleration, and jerk, with the spacing being the primary target to be developed and applied for simulation calibration. Previous naturalistic studies have used spacing as a preferred metric to analyze car-following behavior (Zhu, Wang, and Tarko, 2018; Sangster, Rakha, and Du, 2013; Punzo and Montanino, 2016).

VEHICLE SPACING

The final dataset provided sufficient data to extract behavior at individual speed-group levels while ensuring similar driver distributions to those shown in table 2, reducing possible bias due to dissimilar participant composition by speed groups. The number of spacing observations within each speed group ranged from more than 360,000 for the bin with fewest data points (i.e., 30–40 mph) to more than 6 million points for the bins with the largest number of points (i.e., 50–65 mph, and 60–70 mph). High-traffic conditions, reflected by speed bins with the lowest speeds had significant number of points, with more than 900,000 data points for the 5- to 20-mph bin.

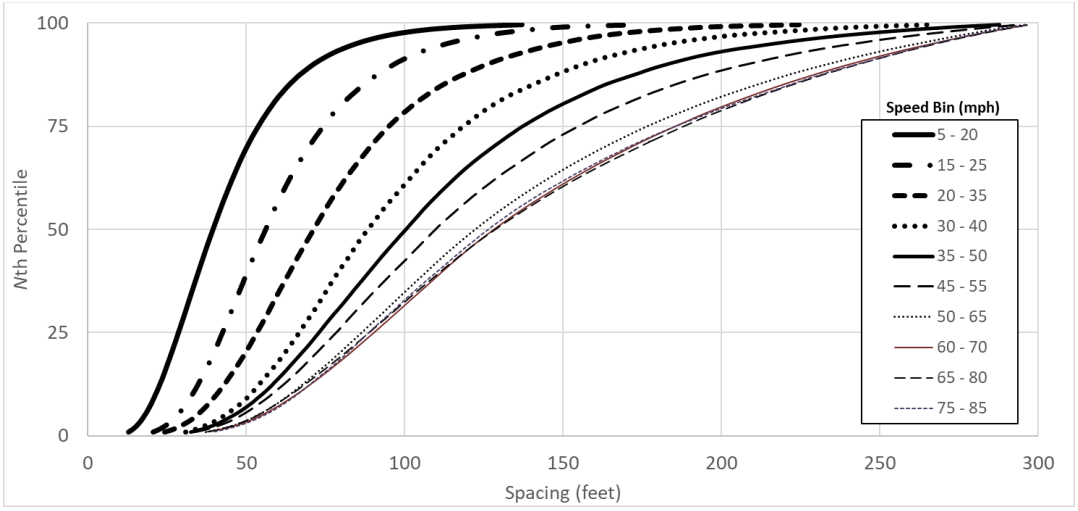
The research team developed detailed empirical distributions of vehicle spacing for each speed bin using the final NDS dataset (including parts A and B) (NCHRP, 2023). The distributions were then represented using 220 percentile points to serve as the main NDS targets. The distributions provide enough detail to differentiate percentiles with a precision greater than one quarter of a percent. Different from previous research, which often focused on expected or average spacing, the extracted data allow for specific comparisons of observed versus expected spacing at any percentile.

Figure 6 and figure 7 show illustrations of the NDS spacing targets in terms of the probability density function, as well as the cumulative distribution function (CDF) for each bin, respectively. Key percentiles of the spacing distributions from each speed group are shown in table 4.



Source: FHWA.

Figure 6. Graph. Empirical probability density plot of spacing for all speed groups.



Source: FHWA.

Figure 7. Graph. Empirical cumulative distribution plot of spacing for all speed groups.

Table 4. Key percentiles for vehicle spacing (foot) distributions from NDS by speed group.

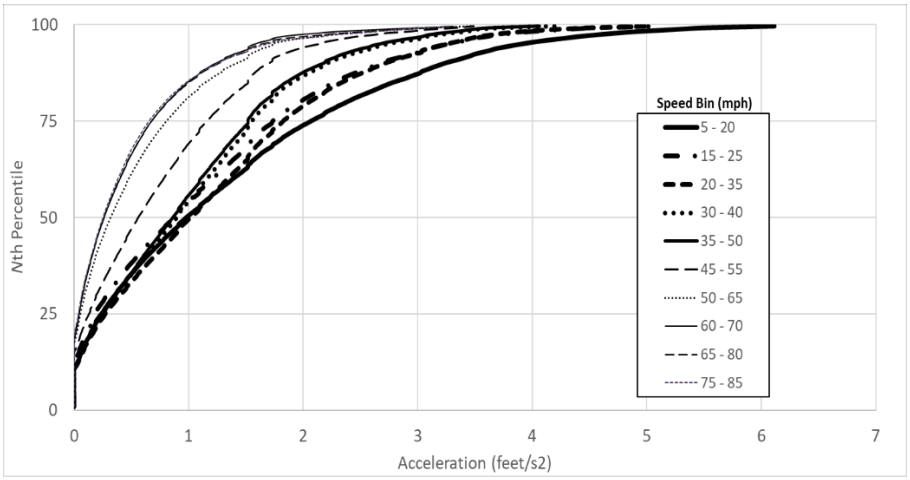
*N*th Percentile	Speed Bin (mph)									
	5–20	**15–25**	**20–35**	**30–40**	**35–50**	**45–55**	**50–65**	**60–70**	**65–80**	**75–85**
1	12.9	20.6	24.2	30.8	32.6	33.5	37.2	38.1	37.8	39.9
5	17.9	28.2	34.3	43.0	45.9	48.3	53.2	54.9	53.6	55.7
10	21.3	32.9	40.4	51.0	54.9	57.8	63.5	65.9	64.0	66.1
25	28.6	42.4	53.2	66.6	73.0	78.2	86.3	90.4	88.7	88.7
50	39.5	56.1	71.2	88.3	100.5	109.6	122.1	128.2	128.6	126.4
75	53.8	74.3	95.0	118.6	137.4	154.9	176.6	185.5	188.1	185.5
90	71.3	96.7	126.1	156.5	183.0	208.2	233.5	240.4	243.1	242.7
95	84.9	113.5	148.7	183.7	217.1	242.4	262.1	266.6	267.7	267.1
99	118.6	151.1	197.4	244.6	273.9	286.1	291.2	292.4	292.6	291.8

Note the consistency of the spacing values from figure 6 and figure 7, which gradually increase with the speed, as expected. These results could not have been obtained empirically with low sample sizes or without careful data processing to define car-following behavior and to filter out data inaccuracies.

ACCELERATION AND DECELERATION

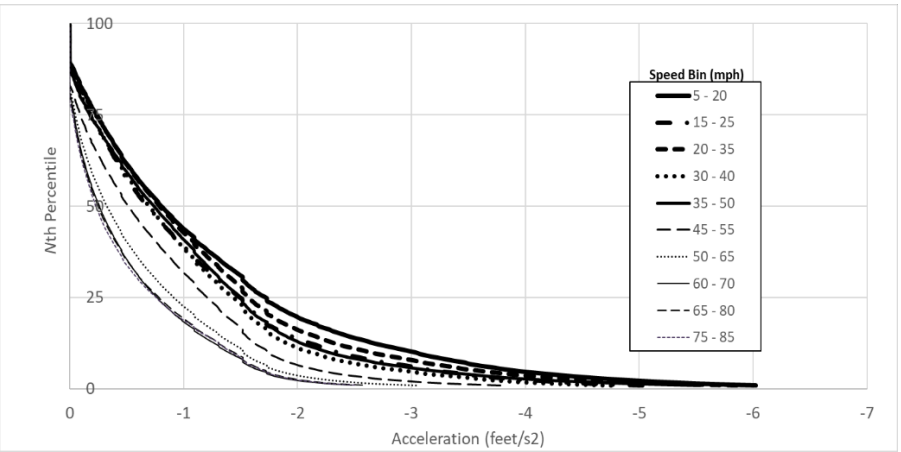
In terms of vehicle kinematics, acceleration is the next natural microscopic variable after analyzing vehicle location (i.e., spacing) by speed groups. However, differences in acceleration are expected to be subtler, particularly at higher speeds. The research team produced empirical cumulative distributions from NDS based on the speed rate change using a time window of 0.5 s. Therefore, the research team observed acceleration at a lower resolution than speed (e.g., the network speed was recorded at a rate of 10 Hz). This decision was partially a result of typical human driving, where conservative estimates would still exceed 0.5 s of perception-reaction times and therefore the ability to adjust acceleration more than once. Higher frequency of acceleration changes would also incorporate larger effects from speed measurement errors.

Figure 8 and figure 9 show the CDFs for positive acceleration and negative acceleration (deceleration), respectively. Similar to the spacing data, the acceleration distributions display organized and consistent patterns. Larger accelerations are more common at lower speeds (lower than 50 mph), with a 50th percentile of about 1 ft/s^2 compared to about 0.3 ft/s^2 for the 50- to 65-mph group. These acceleration values are relatively small compared to maximum accelerations a driver can typically apply, but they are well within typical ranges specified for oscillation acceleration parameters. Recognize that accelerations obtained from NDS correspond to car-following behavior observations, excluding overtaking or lane-changing maneuvers or free-flowing conditions.



Source: FHWA.

Figure 8. Graph. Empirical cumulative distribution plot of acceleration for all speed groups.



Source: FHWA.

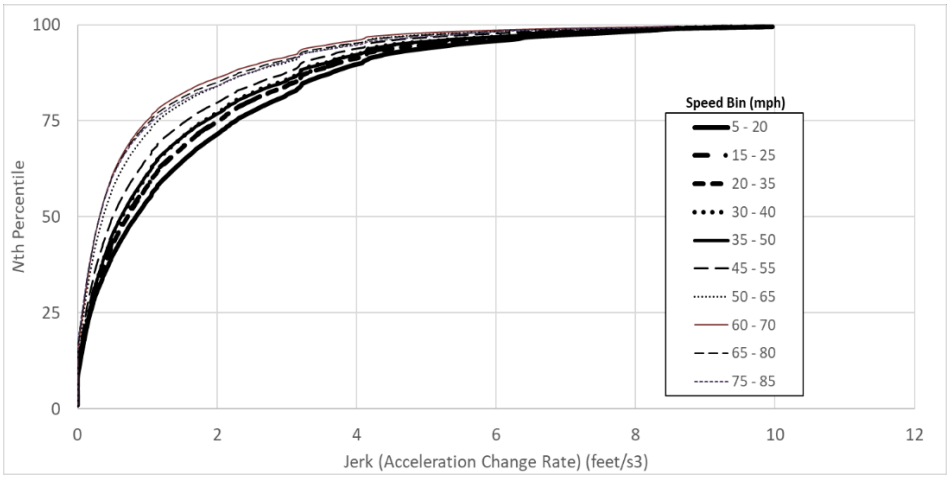
Figure 9. Graph. Empirical cumulative distribution plot of negative acceleration (deceleration) for all speed groups.

The research team also found that both positive and negative acceleration distributions from instrumented vehicles closely mirror each other, indicating an oscillatory behavior of a follower with respect to the leader, typically expected from car-following conditions.

Overall, the acceleration data also display consistency and follow expectations, reassuring confidence on the NDS data and the postprocessing methods. However, as a calibration metric it does not provide stronger separation markers between speed bins than spacing. Acceleration targets are a potential secondary metric if calibration efforts are conducted beyond spacing distributions. In such cases, comparisons between NDS targets and simulation trajectories would be conducted first, as described in the proposed process, and then a secondary microscopic calibration using NDS acceleration targets would take place. However, benefits in conducting both detailed calibrations (using spacing and acceleration) instead of only using spacing may not warrant the increase in the process complexity. If an analyst wants to incorporate acceleration, a recommended alternative is using central tendency measures for the evaluation of accelerations (in place of complete distributions), such as the median and the interquartile range.

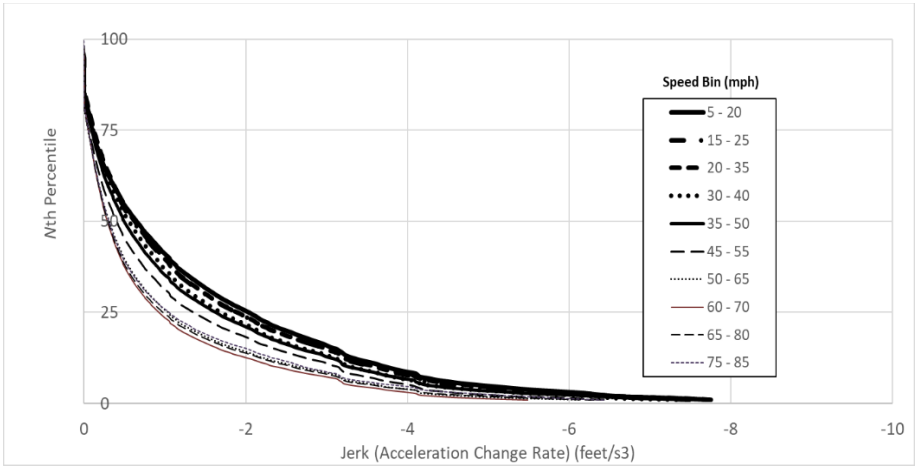
ACCELERATION RATES (JERK)

As an alternative to acceleration data, the research team also analyzed the rate of acceleration change (i.e., jerk) to establish potential NDS targets for simulation. The calculations for jerk values used two consecutive acceleration values. Since acceleration values were estimated every 0.5 s, jerk values reflect variations in the acceleration between such values. Figure 10 and figure 11 show the empirical distribution of the jerk values for both acceleration and deceleration, respectively. As expected from the acceleration distributions, larger jerk values are associated with lower speeds. Sudden maneuvers at lower speeds can produce larger longitudinal changes in acceleration compared to vehicles already traveling at freeway speeds. Also, data are consistent and indicate gradual decreases of jerk values with speed; however, as a differentiating metric between speed group, jerk values do not have as much power as acceleration, and even less than spacing. Distributions of jerk values are more similar to each other as the speeds reach 50 mph or higher, and small differences are observed between lower speed groups.



Source: FHWA.

Figure 10. Graph. Empirical cumulative distribution plot of jerk for all speed groups.

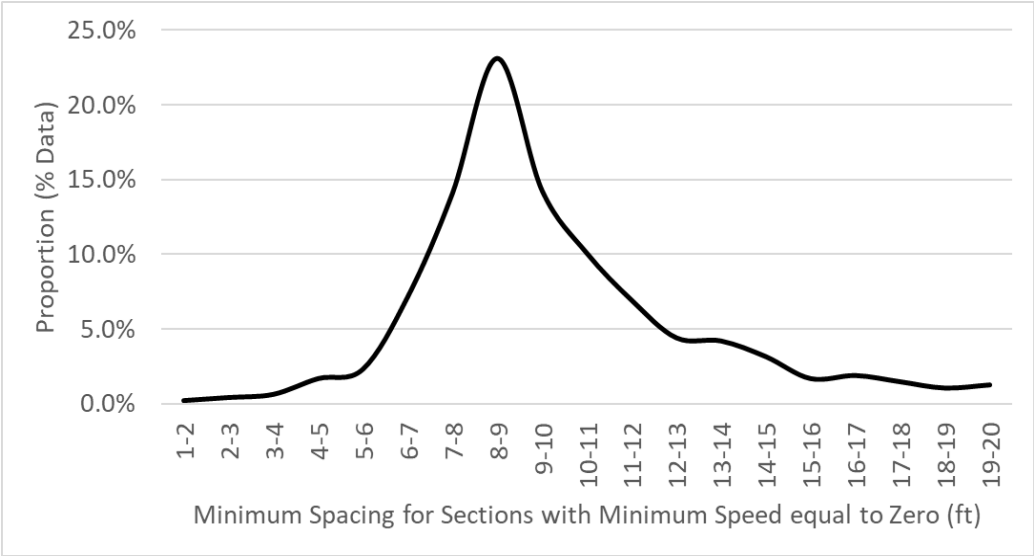


Source: FHWA.

Figure 11. Graph. Empirical cumulative distribution plot of jerk (for deceleration) for all speed groups.

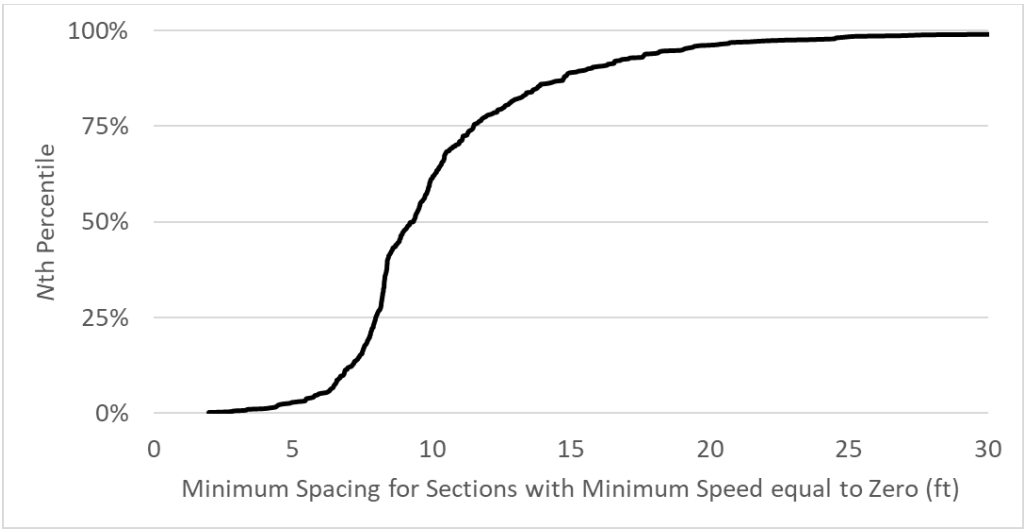
STANDSTILL DISTANCE

In addition to the car-following spacing, the research team also generated distributions to illustrate the standstill distance when speeds of the instrumented vehicle dropped to zero. Standstill distance does not reflect active following behavior, but may serve as a reference to calibrate such conditions in microscopic simulation. Figure 12 and figure 13 show the probability density function and the CDFs extracted from the NDS data. The 50th percentile of the distribution is equivalent to 9.3 ft, which is similar to what other studies have found, but may seem higher than some default simulation parameters (Houchin et al., 2015; Zhu, Wang, and Tarko, 2018).



Source: FHWA.

Figure 12. Graph. Probability distribution plot of standstill distance.



Source: FHWA.

Figure 13. Graph. Cumulative distribution plot of standstill distance.

CHARACTERIZATION AND CURVE FITTING FOR SPACING DISTRIBUTIONS

Given the results of the NDS target extraction, where the vehicle spacing was a preferred metric to differentiate behavior for the defined speed levels, the research team decided to further investigate the potential fit of the empirical spacing distributions to known distributions including normal, lognormal, Weibull, and gamma. Adequate fit to a known distribution would bring important advantages in data analysis, in particular, in terms of statistical comparisons to sample values drawn from simulation. Known distributions would also carry advantages to characterize behavior in car-following models, as well as changes in such behavior under different following speeds (i.e., under different traffic conditions).

For this purpose, the research team analyzed the complete NDS car-following datasets for each speed level using the SAS® software (NCHRP, 2023; SAS Institute, 2020). The data were segmented and loaded in SAS using two alternative options for exploration: a compact segmentation where a single observation was obtained from each speed group and each traversal, and a detailed segmentation where a single observation was obtained from each instance of sustained following within one of the speed groups. So, for a given speed level, option 1 represented each trip with a single-spacing value, whereas option 2 could include multiple values, each from a single sustained interaction with a leader vehicle. The research team also tested two different metrics summarizing the spacing behavior, so, for a given following event with sustained speeds, spacing could be represented either by the mean or the median of the observed behavior. In essence, all defined datasets represented different versions of central tendency measures related to vehicle spacing.

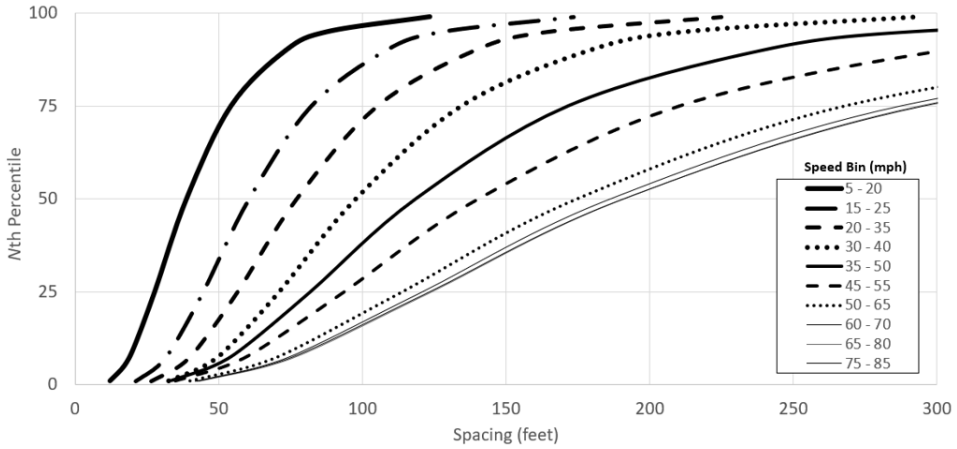
Results from the modeling indicated a closer relationship between the spacing distributions and families of lognormal distributions. Also, mean values of the observed behavior provided improved fit over sets where the median was tested, and similar outcomes were obtained with the compact and the detailed segmentations when the mean was used as a metric of choice.

Table 5 shows a summary of the fitted distributions for the compact and the detailed segmentation cases using the mean spacing as a metric of choice. Note the similarity in the distribution parameters for the two cases, even though some speed groups have a significantly larger number of observations, pointing to the possibility of using either segmentation for additional analysis. These results also indicated that intravariation of the metric for a given traversal had limited effects in the characterization of the driving behavior, hinting at consistent behavior within traversals (i.e., spacing from the same driver on the same site and under the same traffic conditions).

Table 5. Family of fitted lognormal distributions for all speed groups.

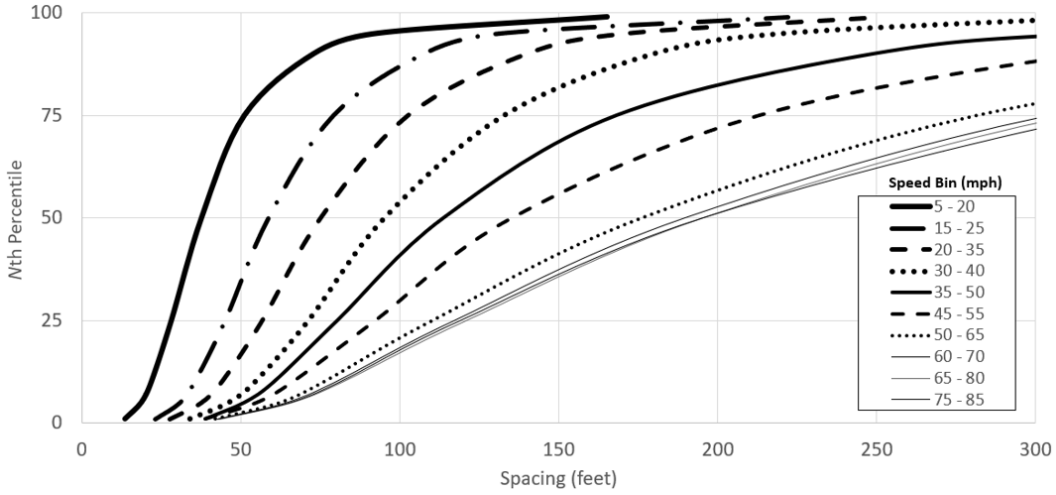
Speed Group	Compact Segmentation Parameters					Detailed Segmentation Parameters				
	Number of Observations	Scale	Shape	Mean	Standard Deviation	Number of Observations	Scale	Shape	Mean	Standard Deviation
0–20	1,328	3.68	0.472	44.13	22.04	3,516	3.66	0.497	44.00	23.29
15–25	1,228	4.11	0.430	66.57	30.04	3,184	4.10	0.453	67.09	31.98
20–35	1,703	4.37	0.432	87.14	39.48	4,318	4.34	0.461	85.64	41.68
30–40	1,498	4.58	0.457	108.33	52.25	2,515	4.58	0.473	109.29	54.77
35–50	2,741	4.78	0.538	137.61	79.65	4,177	4.78	0.548	137.83	81.59
45–55	4,305	4.94	0.580	165.86	104.82	6,021	4.94	0.600	168.11	110.60
50–65	19,981	5.18	0.605	214.40	142.67	28,163	5.16	0.633	213.73	150.13
60–70	22,649	5.26	0.611	232.22	156.29	33,005	5.23	0.641	228.64	163.00
65–80	19,121	5.29	0.605	238.34	158.56	26,565	5.25	0.639	233.91	166.11
70–85	4,772	5.30	0.602	239.52	158.41	7,059	5.25	0.647	234.93	169.47

Figure 14 shows a visualization of the family of lognormal curves for the detailed segmentation in comparison to those from the original datasets in figure 15. The research team constructed these curves using similar segments as the distributions from figure 6 before the distributions were trimmed to a maximum of 300 ft, but they also represent more condensed data because each sustained following event is summarized by a mean value, instead of including each observation found in the NDS datasets. Nonetheless, the figures illustrate the ability of lognormal distributions to reproduce similar behaviors to those observed in the field (VTTI, 2020).



Source: FHWA.

Figure 14. Graph. Lognormal spacing distributions for detailed segmentation for all speed groups.



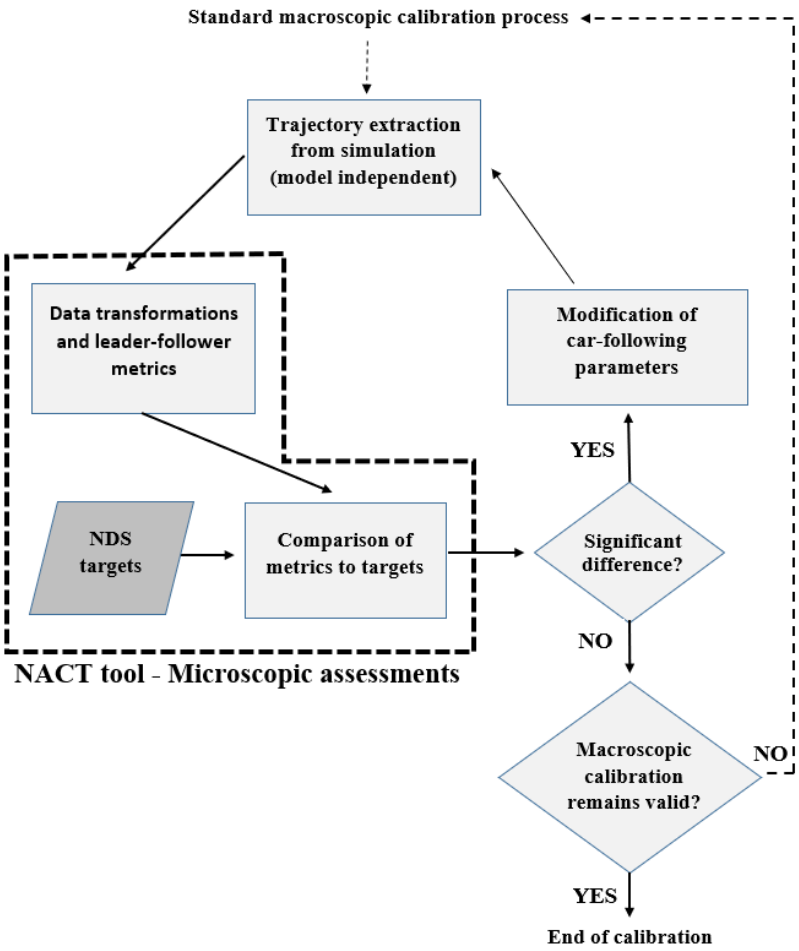
Source: FHWA.

Figure 15. Graph. Empirical spacing distributions for detailed segmentation for all speed groups.

CHAPTER 5. PROPOSED CALIBRATION PROCESS

Based on results described in chapter 4, the proposed calibration of simulation at the car-following data level is based on vehicle spacing distributions for specific speed groups. In this case, this research team defined 10 overlapping groups to provide flexibility and accommodate a large range of speeds from low- to high-traffic conditions.

The research team developed a custom stand-alone tool on Python® version 3.10.6 to process generic vehicle trajectory data from a simulation and analyze that data in relation to the naturalistic behavior targets extracted from the NDS datasets (Python Software Foundation, 2021; VTTI, 2020). This NACT tool is a compiled application in the form of an executable file that contains all the required dependencies to run on Microsoft® Windows®-based machines (Microsoft, 2023). A guideline for the NATC tool (*Guideline for the Naturalistic Assessments of Car-Following Trajectories Tool*) to document the NATC tool's usage is also part of the deliverables of this research. The tool is intended to be part of a proposed calibration process, as shown in the schematic representation in figure 16.



Source: FHWA.

Figure 16. Flowchart. Proposed use of microscopic calibration to complement standard calibration.

On selection of user-defined trajectory data from microscopic simulation, the NACT tool identifies and analyzes leader-follower pairs to build vehicle spacing distributions for corresponding traffic conditions (e.g., speed bins identified for the NDS targets). Then the tool compares such distributions to the target benchmarks.

Goodness-of-fit testing quantifies whether simulation trajectories produce similar spacing and accelerations as those in real-world conditions. If comparisons result in significant differences, the process loops back for further modification of the car-following parameters. Otherwise, the simulation can be considered calibrated from a microscopic standpoint.

The start of the microscopic calibration process (i.e., use of the NACT tool) is tied to standard calibration at the macroscopic level, specifically at the point where macroscopic calibration is complete so that a verification of the trajectories at the microscopic level can be conducted as a second calibration stage. The integration of calibration at the microscopic level can be viewed as a bi-level analysis scenario, where the first level deals with calibration of macroscopic metrics and the second level deals with microscopic metrics. Analogous to findings in previous studies

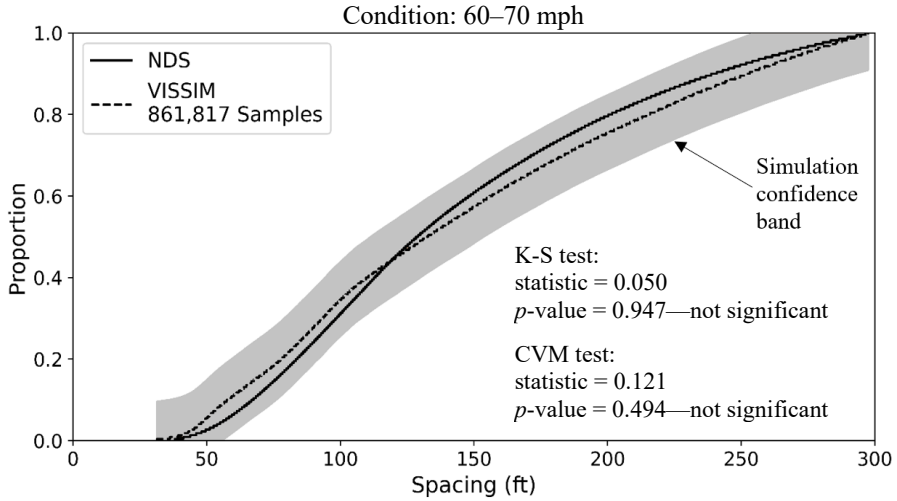
where calibration at different levels can affect each other, conducting a final verification of the validity of the initial macroscopic calibration after the microscopic stages are complete is important (Hale et al., 2021). The last decision point before the calibration is considered complete in the diagram (figure 16) illustrates the importance of verifying the validity of the initial macroscopic calibration.

Thus, the proposed microscopic comparisons of spacing are aimed not only at improving initial simulation parameter choice, but mostly as a verification of the driving behavior after typical macroscopic calibration steps.

The proposed process scales down the spacing distributions from simulation for each speed group, such that they can be compared to the NDS targets. Then, the NACT tool performs statistical comparisons and generates graphical representations of the NDS and the simulation distributions.

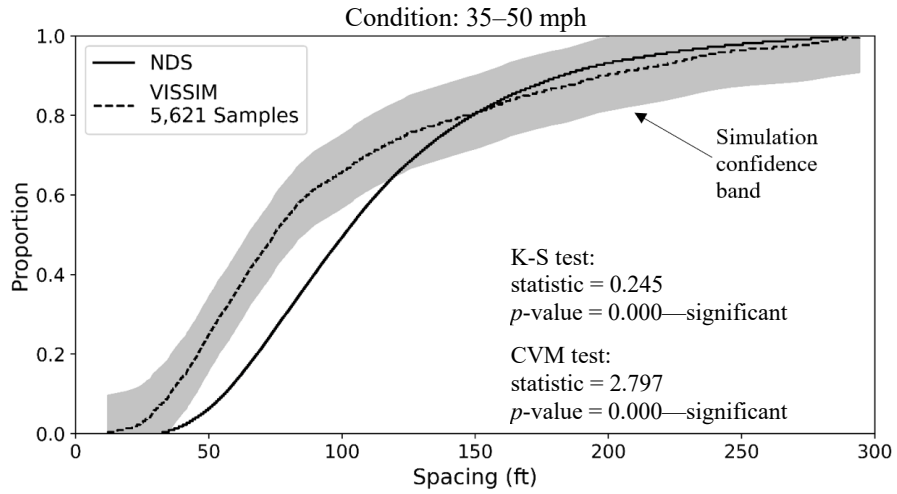
The Kolmogorov–Smirnov (K-S) test and the Cramer-Von Mises (CVM) test are applied to the data, where the two distributions being compared are those from the NDS target and the simulation data Comparisons are made for each of the speed groups and summarized in an output file produced by the NACT tool.

In addition, for each speed bin, K-S and CVM statistics and their p-value are reported on a figure generated by the application along with the CDF plots of the spacing distribution. For enhanced guidance, the tool also draws a confidence band using the Dvoretzky–Kiefer–Wolfowitz (DKW) Inequality with a 95-percent confidence level. The bands indicate sections of the distribution where the K-S value is more likely to differ from the NDS targets. Examples of the generated plots are shown in figure 17 and figure 18, where spacing distributions from simulation are identified as having similar distributions and being significantly different, respectively. Note the confidence band from DKW around the simulation series.



Source: FHWA.

Figure 17. Graph. Sample NACT tool output—similar simulation and NDS spacing distributions.



Source: FHWA.

Figure 18. Graph. Sample NACT tool output—significantly different simulation and NDS spacing distributions.

Outputs from the NACT tool provide guidance to identify speed groups and portions of the distributions where the largest differences exist between simulation trajectories and NDS targets. Simulation users are expected to take these outputs into consideration and iterate through the calibration process, as shown in figure 16.

While the NACT tool analyzes input data to generate comparisons between simulation and NDS targets, the tool does not provide further guidance on how to adjust specific car-following parameters to improve goodness of fit. The tool is model agnostic and compares generic vehicle trajectory data without knowledge of the underlying simulation package.

The guideline document, which was a deliverable of this project, provides details on the NACT tool and examples using a specific package to illustrate usage through different iterations of simulation runs and the effects of specific parameters on vehicle spacing.

CHAPTER 6. EXPLORING THE INCORPORATION OF NDS DATA TO SAFETY MODELING

This chapter presents additional exploration on the potential of incorporating NDS-derived variables into safety analysis at macroscopic and microscopic scales. First, the research team evaluated NDS macroscopic variables as part of crash prediction models in the form of density, speed, and spacing; and second, the research team compared safety-related outcomes from microscopic simulation before and after calibration of car-following models using the NDS targets. Although preliminary, these applications offer initial indications of some of the implications of using NDS targets in safety analysis.

NDS VARIABLES IN MACROSCOPIC CRASH PREDICTION MODELS

The analysis of the data included documenting the mean, median, variance, and 85th percentile of the NDS variables analyzed. These variables included:

- Density in vehicles per mile per lane, as calculated from traffic state estimates.
- Speed in miles per hour, as measured from speedometer or estimated network speeds.
- Spacing for leader-follower pairs, as calculated from radar datasets.

The research team obtained crash data for the prediction models from a combination of sources, including the RID and State files from the Highway Safety Information System (HSIS) (Iowa State University, 2023; FHWA, n.d.). In particular, the vehicle files from HSIS data were needed to identify the direction of travel for crash data in North Carolina.

The research team calibrated crash prediction models using conventional generalized linear modeling techniques with an assumed negative binomial error distribution. The team adopted the Hoerl function in addition to the conventional exponential model form for relating crash frequency to NDS variables (Hauer, 2015). The research team used the Hoerl function in addition to the conventional exponential model form because the Hoerl function allows the model to be increasing and then decreasing (or vice versa) rather than just increasing or decreasing only.

The general approach to developing the models was to first estimate a base model that did not include NDS variables. The base models' purpose was to have simple models based on site characteristics that affect crash frequency so that the addition of other variables can be assessed in that context. Then, each NDS variable (density, speed, and spacing) was added to the model using either the mean, median, variance, or 85th percentile value. The research team used the improvement in the individual variable significance and the Akaike Information Criterion (AIC) to assess the models' overall goodness of fit (Toga, 2015). The research team used AIC because AIC provides measures of model performance that account for model complexity. Smaller values of AIC generally represent a better overall goodness of fit. Model outcomes also included a secondary metric, the Bayesian Information Criterion (BIC) to inform the model fit. Similar to AIC, lower BIC indicates a better goodness of fit.

Table 6 presents summary statistics for sites in Florida, North Carolina, and Washington. Table 7 through table 9 provide the mean, median, variance, and 85th percentile for each site for density, speed, and leader-follower spacing for individual sites in Florida, North Carolina, and Washington. The tables also provide the total number of observed multiple vehicle (MV) crashes and the total number of observed MV daytime crashes for the 3-yr study period.

Table 6. Summary statistics for sites in Florida, North Carolina, and Washington.

State	Number of Sites	Metric	Variable					
			AADT (vpd)	On Ramp AADT (vpd)	Off Ramp AADT (vpd)	1/Ramp Spacing (mi)	Upstream Lanes	MV Daytime Crashes and MV Total Crashes
FL	26	Minimum	43,454	2,875	300	0.26	2.00	0.53
		Maximum	69,667	21,625	23,667	2.64	4.00	0.92
		Average	56,046	9,233	8,662	0.86	3.00	0.76
NC	34	Minimum	20,167	1,767	973	0.97	2.00	0.56
		Maximum	82,500	17,333	20,667	8.05	4.00	1.00
		Average	42,566	5,645	6,174	2.45	2.64	0.82
WA	44	Minimum	15,833	2,667	1,700	0.40	2.00	0.50
		Maximum	101,333	27,667	29,333	7.28	5.00	1.00
		Average	67,080	9,244	9,389	1.66	3.34	0.79

vpd = vehicles per day.

Table 7. Summary of Florida variables.

State	Site ID	MV Total Crashes (no.)	MV Daytime Crashes (no.)	Mean Density (veh/mile)	Median Density (veh/mile)	Variance Density (veh/mile)^2	85th-Per Density (veh/mile)	Average Speed (mph)	Median Speed (mph)	Variance Speed (mph)2	85th-Per Speed (mph)	Average Spacing (ft)	Median Spacing (ft)	Variance Spacing (ft)^2	85th-Per Spacing (ft)
FL	FL 1	81	64	37.83	33.94	528.58	51.23	66.88	68.57	62.14	72.50	201.86	181.59	9,931.62	306.79
FL	FL 2	79	63	37.65	34.25	405.49	53.11	68.13	69.65	58.98	74.34	201.94	174.39	11,631.50	311.43
FL	FL 3	28	18	36.08	33.36	258.41	51.33	70.27	70.34	20.98	74.59	229.23	199.90	16,790.09	370.63
FL	FL 4	34	24	34.50	31.64	184.65	48.05	70.15	70.66	29.62	75.06	212.33	192.05	10,776.47	322.24
FL	FL 5	13	8	35.40	31.13	562.45	51.69	69.19	69.16	32.41	75.09	251.28	213.19	22,617.57	419.88
FL	FL 6	36	24	35.10	30.74	336.13	52.59	71.64	72.06	32.23	76.36	228.10	197.61	15,939.53	352.66
FL	FL 7	82	71	37.15	33.81	363.16	54.38	71.09	71.61	28.93	76.14	206.66	181.37	12,761.15	320.37
FL	FL 8	73	53	36.98	33.68	384.46	50.82	69.28	69.92	34.98	74.18	202.43	175.31	11,303.30	311.61
FL	FL 9	7	4	35.58	30.11	444.78	55.50	69.93	70.18	28.95	75.18	245.33	220.62	20,010.84	410.31
FL	FL 10	34	18	34.66	31.32	255.92	48.06	70.28	70.39	22.96	75.00	218.85	198.57	11,588.91	336.86
FL	FL 11	51	32	30.87	27.46	331.00	42.90	71.89	72.11	19.65	76.09	248.93	237.70	12,756.48	350.51
FL	FL 12	63	43	29.56	26.27	243.18	40.97	71.31	71.41	21.63	75.68	250.16	243.10	10,157.70	351.21
FL	FL 13	71	59	37.95	33.81	460.83	55.32	68.76	70.54	72.58	75.45	201.01	177.19	11,861.98	305.68
FL	FL 14	105	89	35.04	29.93	366.10	53.68	70.64	71.05	35.96	76.40	231.08	203.34	17,096.59	375.15
FL	FL 15	141	125	37.86	34.48	317.39	54.00	67.52	69.12	64.11	74.00	188.88	165.03	10,665.97	290.24
FL	FL 16	69	45	30.37	26.41	253.52	41.62	71.33	71.50	22.02	75.77	261.36	244.95	14,548.94	390.21
FL	FL 17	58	45	33.82	28.95	507.63	48.65	70.47	71.15	29.78	75.24	231.07	215.26	12,591.41	338.85
FL	FL 18	121	99	35.84	31.45	393.70	53.06	67.53	69.34	66.15	73.73	221.80	199.29	15,267.72	355.94
FL	FL 19	54	44	38.22	30.33	647.27	60.73	68.04	70.92	148.19	76.02	225.78	192.76	20,318.61	382.48
FL	FL 20	84	64	36.06	30.28	421.11	52.06	67.62	69.95	77.95	74.90	214.84	193.18	13,477.15	349.74
FL	FL 21	73	61	40.03	35.88	473.83	60.91	62.78	62.66	38.50	68.34	209.02	173.25	16,931.43	345.07
FL	FL 22	23	19	38.39	33.22	466.01	57.95	63.66	63.24	30.41	69.08	223.35	185.58	19,681.64	379.78
FL	FL 23	32	29	37.17	33.75	367.90	56.24	61.94	61.78	31.14	67.46	211.81	173.53	16,629.11	358.87

State	Site ID	MV Total Crashes (no.)	MV Daytime Crashes (no.)	Mean Density (veh/mile)	Median Density (veh/mile)	Variance Density (veh/mile)^2	85th-Per Density (veh/mile)	Average Speed (mph)	Median Speed (mph)	Variance Speed (mph)2	85th-Per Speed (mph)	Average Spacing (ft)	Median Spacing (ft)	Variance Spacing (ft)^2	85th-Per Spacing (ft)
FL	FL 24	20	17	36.42	29.67	705.12	52.84	64.07	64.26	35.43	69.54	223.57	200.21	16,544.63	355.61
FL	FL 25	25	23	37.25	31.01	594.37	57.27	64.71	65.08	60.44	71.11	236.18	203.92	22,128.24	400.74
FL	FL 26	58	49	40.45	33.84	659.90	63.63	63.20	64.25	69.82	69.46	204.23	175.98	16,015.89	334.46

veh = vehicle.

Table 8. Summary of North Carolina variables.

State	Site ID	MV Total Crashes (no.)	MV Daytime Crashes (no.)	Mean Density (veh/mile)	Median Density (veh/mile)	Variance Density (veh/mile)^2	85th-Per Density (veh/mile)	Average Speed (mph)	Median Speed (mph)	Variance Speed (mph)^2	85th-Per Speed (mph)	Average Spacing (ft)	Median Spacing (ft)	Variance Spacing (ft)^2	85th-Per Spacing (ft)
NC	NC 1	20	20	42.49	37.29	631.45	64.47	60.69	60.34	19.12	64.86	209.10	171.83	20,336.21	373.56
NC	NC 2	28	19	32.60	28.03	256.88	51.17	66.85	67.38	17.33	70.48	259.41	240.07	17,536.80	416.46
NC	NC 3	45	37	35.11	29.69	341.58	51.32	64.97	66.01	42.68	70.08	226.72	227.95	16,979.62	325.50
NC	NC 4	44	35	36.62	35.48	296.71	52.02	61.63	62.17	25.41	66.51	208.45	171.68	16,219.85	367.79
NC	NC 5	125	97	29.59	28.03	206.93	41.38	62.09	62.30	13.16	65.45	270.38	259.96	16,539.88	398.31
NC	NC 6	56	41	36.23	30.18	625.39	54.80	66.79	67.27	18.13	70.46	233.26	213.86	18,963.63	392.10
NC	NC 7	26	21	34.57	33.20	270.40	44.48	65.47	66.36	33.36	69.95	244.07	224.86	18,567.34	361.46
NC	NC 8	80	73	32.81	29.30	255.43	46.89	67.50	68.24	18.70	71.37	242.88	195.94	20,199.21	430.09
NC	NC 9	54	44	32.90	28.67	363.26	44.25	66.93	68.53	50.52	71.66	255.11	246.67	16,865.42	368.60
NC	NC 10	26	26	31.53	25.34	239.46	51.86	65.14	65.81	32.74	69.99	226.35	218.08	18,856.23	353.26
NC	NC 11	43	25	32.29	27.83	305.33	44.62	65.22	65.40	17.49	69.77	242.13	233.28	17,316.33	392.46
NC	NC 12	29	27	34.12	30.41	329.95	54.92	63.79	66.52	78.89	69.51	246.39	207.40	22,112.64	431.94
NC	NC 13	10	10	33.35	27.66	514.12	49.16	64.63	65.16	29.08	69.57	240.35	210.11	20,936.18	407.65
NC	NC 14	16	15	30.24	27.55	253.46	44.15	64.59	65.08	26.11	68.52	258.89	221.97	19,140.80	402.87
NC	NC 15	4	3	42.43	29.10	3,664.44	57.22	57.73	58.28	14.58	60.74	220.21	198.18	14,754.12	346.81
NC	NC 16	N/A	N/A	32.12	30.01	246.93	43.78	59.17	59.04	16.09	63.32	246.03	209.73	18,457.44	380.70
NC	NC 17	25	21	33.50	32.16	235.08	45.61	62.61	63.28	39.52	67.13	253.06	223.96	15,253.64	387.97

State	Site ID	MV Total Crashes (no.)	MV Daytime Crashes (no.)	Mean Density (veh/mile)	Median Density (veh/mile)	Variance Density (veh/mile)^2	85th-Per Density (veh/mile)	Average Speed (mph)	Median Speed (mph)	Variance Speed (mph)^2	85th-Per Speed (mph)	Average Spacing (ft)	Median Spacing (ft)	Variance Spacing (ft)^2	85th-Per Spacing (ft)
NC	NC 18	28	24	37.00	32.63	578.86	50.17	57.02	59.23	83.42	63.42	238.04	214.65	19,107.67	366.98
NC	NC 19	35	34	39.80	36.69	504.06	54.15	56.14	57.18	43.53	61.85	200.96	183.47	15,016.09	324.64
NC	NC 20	15	15	38.79	33.09	546.88	60.78	56.28	56.31	51.39	62.57	228.75	181.41	21,406.74	396.41
NC	NC 21	32	28	34.81	28.07	358.38	52.43	58.28	58.70	24.20	63.55	224.40	207.25	14,850.14	360.88
NC	NC 22	9	8	34.76	28.02	521.06	52.14	61.63	61.12	14.83	64.62	228.00	206.54	14,349.55	361.04
NC	NC 23	9	5	36.86	32.04	384.99	55.98	59.55	59.74	25.28	63.56	224.26	196.04	17,604.68	357.41
NC	NC 24	21	19	33.53	31.34	329.74	46.91	56.79	56.82	25.25	61.63	228.01	208.05	12,653.56	365.84
NC	NC 25	15	11	29.33	25.82	185.54	44.13	60.08	60.31	26.61	64.17	245.57	213.52	17,435.93	391.29
NC	NC 26	17	10	27.46	24.72	211.03	39.06	66.11	66.67	13.40	69.60	293.39	271.99	21,898.81	485.98
NC	NC 27	68	57	35.32	30.84	258.39	52.05	66.47	67.13	29.20	71.69	231.51	186.10	22,263.39	403.16
NC	NC 28	8	6	29.90	27.93	159.55	38.22	68.06	68.19	19.03	71.44	256.09	225.96	16,321.99	417.54
NC	NC 29	50	42	35.05	32.89	276.30	47.72	64.53	65.08	34.92	69.28	206.24	175.38	13,927.48	315.78
NC	NC 30	113	97	30.48	28.24	180.96	41.34	61.16	61.91	22.55	65.12	252.96	217.31	13,643.29	376.47
NC	NC 31	97	88	34.90	31.83	236.58	48.30	61.09	61.95	26.58	65.05	232.25	202.62	15,244.21	350.18
NC	NC 32	18	16	26.79	22.30	184.63	38.25	66.11	65.32	16.65	70.87	304.85	282.78	20,665.44	470.47
NC	NC 33	5	3	42.55	30.61	3,047.64	61.94	60.56	60.96	19.40	64.67	243.84	212.80	16,667.43	391.89
NC	NC 34	92	72	30.40	27.84	206.56	43.53	68.18	68.23	21.81	72.13	250.99	214.66	18,374.50	421.81

N/A = not applicable.

Table 9. Summary of Washington variables.

State	Site ID	MV Total Crashes (no.)	MV Daytime Crashes (no.)	Mean Density (veh/mile)	Median Density (veh/mile)	Variance Density (veh/mile)^2	85th-Per Density (veh/mile)	Average Speed (mph)	Median Speed (mph)	Variance Speed (mph)^2	85th-Per Speed (mph)	Average Spacing (ft)	Median Spacing (ft)	Variance Spacing (ft)^2	85th-Per Spacing (ft)
WA	WA 1	26	20	36.96	32.67	404.40	55.14	61.87	61.56	11.16	65.37	214.34	189.15	16,628.30	366.67
WA	WA 2	15	13	34.28	29.99	252.83	48.65	60.89	60.75	18.01	65.14	226.09	186.17	17,940.27	376.84
WA	WA 3	19	13	36.01	33.96	433.03	49.17	59.40	59.98	48.70	63.99	228.86	178.26	20,074.69	384.54

State	Site ID	MV Total Crashes (no.)	MV Daytime Crashes (no.)	Mean Density (veh/mile)	Median Density (veh/mile)	Variance Density (veh/mile)^2	85th-Per Density (veh/mile)	Average Speed (mph)	Median Speed (mph)	Variance Speed (mph)^2	85th-Per Speed (mph)	Average Spacing (ft)	Median Spacing (ft)	Variance Spacing (ft)^2	85th-Per Spacing (ft)
WA	WA 4	88	78	46.23	41.01	825.56	61.32	48.35	51.62	123.60	58.40	133.37	122.79	5,927.99	210.99
WA	WA 5	114	84	36.58	34.75	183.86	51.90	58.70	58.87	15.59	63.00	195.05	170.56	10,875.48	300.63
WA	WA 6	45	36	32.28	28.43	208.20	48.44	62.45	62.15	13.84	65.85	237.45	229.89	15,616.58	386.07
WA	WA 7	88	71	38.25	35.54	351.56	52.72	55.88	58.25	64.69	61.50	198.87	145.85	16,906.74	341.92
WA	WA 8	8	7	29.75	25.66	220.85	44.15	56.74	56.81	16.79	60.99	250.16	235.27	16,983.96	365.63
WA	WA 9	27	20	35.60	32.16	359.24	53.94	60.08	60.88	35.39	64.09	221.58	179.22	18,779.73	364.27
WA	WA 10	23	16	35.18	34.37	222.08	51.35	60.50	60.24	25.71	65.12	217.98	168.11	19,442.84	345.67
WA	WA 11	11	8	35.11	31.65	388.67	52.55	60.37	61.04	15.68	63.94	238.12	183.03	22,167.93	431.10
WA	WA 12	63	54	36.75	34.04	387.60	56.01	58.99	59.33	34.71	63.11	226.10	200.63	17,275.88	361.47
WA	WA 13	2	2	25.79	22.21	175.03	41.07	62.25	62.60	9.91	64.98	300.90	284.63	21,952.99	479.15
WA	WA 14	71	36	34.32	29.42	276.37	48.71	61.91	62.33	41.93	66.59	215.76	198.82	10,538.85	318.28
WA	WA 15	13	11	32.00	29.47	224.32	43.21	58.88	58.98	13.28	62.07	226.60	201.19	13,629.36	339.02
WA	WA 16	6	5	30.15	26.95	134.73	40.78	60.57	60.60	10.84	63.23	258.06	210.45	18,078.48	411.52
WA	WA 17	80	74	40.90	37.91	223.45	56.24	55.33	56.80	52.04	60.74	179.06	145.13	13,270.17	259.04
WA	WA 18	5	5	31.26	28.25	282.27	48.72	58.12	57.93	10.88	61.50	266.71	231.54	24,007.10	472.34
WA	WA 19	62	52	36.33	34.12	398.84	47.21	56.98	58.42	42.63	62.08	200.82	170.32	14,958.46	309.99
WA	WA 20	3	2	27.96	21.67	335.20	48.02	58.85	58.96	10.65	62.66	292.89	276.75	22,470.42	477.17
WA	WA 21	8	4	34.15	31.53	259.45	50.33	58.93	58.65	16.03	63.43	217.39	173.51	18,635.07	336.06
WA	WA 22	3	2	27.70	25.43	211.73	37.12	58.94	59.34	12.11	62.06	271.99	232.61	19,178.15	439.21
WA	WA 23	84	61	31.18	27.07	242.18	43.20	63.65	63.48	10.38	66.89	263.46	241.76	17,827.16	422.73
WA	WA 24	140	86	34.16	29.01	872.08	45.11	64.05	64.32	16.87	67.20	245.53	229.21	15,121.58	369.59
WA	WA 25	166	88	34.38	29.92	537.90	49.45	62.86	63.42	25.03	66.50	238.62	223.48	14,836.17	365.22
WA	WA 26	88	62	33.62	29.58	311.42	48.86	61.22	61.89	23.85	65.09	250.49	215.54	20,738.36	416.05
WA	WA 27	84	78	37.01	34.33	240.78	52.58	58.81	59.90	35.62	63.53	201.54	174.86	13,053.32	324.05
WA	WA 28	40	35	39.06	35.75	337.68	58.18	59.72	60.53	32.13	64.33	204.47	170.07	15,713.17	338.99
WA	WA 29	54	46	36.66	34.28	254.43	52.52	59.34	60.46	42.96	64.16	210.72	175.33	16,800.60	352.10

State	Site ID	MV Total Crashes (no.)	MV Daytime Crashes (no.)	Mean Density (veh/mile)	Median Density (veh/mile)	Variance Density (veh/mile)^2	85th-Per Density (veh/mile)	Average Speed (mph)	Median Speed (mph)	Variance Speed (mph)^2	85th-Per Speed (mph)	Average Spacing (ft)	Median Spacing (ft)	Variance Spacing (ft)^2	85th-Per Spacing (ft)
WA	WA 30	67	50	37.94	34.45	405.23	52.56	59.42	61.41	56.77	65.62	201.94	168.56	13,029.27	319.52
WA	WA 31	74	59	34.86	32.24	171.73	47.47	60.31	61.70	37.28	65.42	210.05	184.87	11,996.15	323.73
WA	WA 32	83	59	33.84	31.24	160.03	45.03	61.77	62.43	22.55	65.72	216.84	194.56	10,716.79	326.40
WA	WA 33	137	122	33.67	30.97	229.85	47.65	62.00	62.56	25.86	66.50	223.91	201.99	15,442.70	352.86
WA	WA 34	301	249	38.02	35.31	264.77	51.98	58.36	59.53	41.87	63.08	181.66	155.43	10,730.29	281.16
WA	WA 35	157	142	36.71	33.41	267.20	51.85	59.37	61.94	78.25	65.59	199.90	169.79	14,241.03	317.73
WA	WA 36	54	47	36.45	33.05	361.19	53.32	59.51	61.74	76.99	65.43	219.04	178.73	20,603.30	390.48
WA	WA 37	177	169	37.44	33.75	326.83	53.26	58.72	61.62	97.27	65.59	203.92	170.36	17,224.14	336.05
WA	WA 38	121	107	34.29	30.80	255.78	48.59	59.54	60.43	44.49	64.28	225.17	203.58	16,253.09	367.16
WA	WA 39	29	26	37.43	33.48	339.76	54.34	57.35	59.19	58.18	62.51	221.27	182.16	19,648.45	400.20
WA	WA 40	124	105	38.64	35.29	378.14	55.15	58.84	60.00	41.39	63.97	196.54	162.74	14,072.04	319.55
WA	WA 41	82	49	36.18	31.23	430.53	53.22	61.69	62.18	32.82	66.28	225.38	193.67	18,810.20	382.66
WA	WA 42	58	40	38.70	34.79	448.04	57.69	60.74	61.36	24.67	64.71	213.25	172.90	17,711.05	356.91
WA	WA 43	98	86	37.98	33.30	551.13	54.22	59.75	61.18	54.19	64.97	202.50	166.62	17,285.14	337.55
WA	WA 44	85	75	41.26	33.87	850.12	61.71	56.82	59.95	118.52	64.30	188.07	146.24	19,789.22	330.94

The research team only applied the Hoerl form for the NDS variables (Hauer, 2015). The research team included the ratio of MV daytime to MV total crashes in the MV daytime crash model to account for the fact that while the model was predicting the daytime crashes, the AADT, on the on- and off-ramp AADTs included in the model were for the whole day. The base models have the model forms shown in figure 19 and figure 20.

$$MV\ Crashes = exp^{Intercept} \times AADT^{\beta_1} \times exp^{\beta_2 \times \left(\frac{1}{Ramp\ Spacing}\right)} \times exp^{\beta_3 \times UP_{Lanes}} \times On\ Ramp\ AADT^{\beta_4} \times Off\ Ramp\ AADT^{\beta_5} \times Length \times Years$$

Figure 19. Equation. Base model for all multivehicle crashes.

$$MV\ Daytime\ Crashes = exp^{Intercept} \times AADT^{\beta_1} \times exp^{\beta_2 \times \left(\frac{1}{Ramp\ Spacing}\right)} \times exp^{\beta_3 \times UP_{Lanes}} \times On\ Ramp\ AADT^{\beta_4} \times Off\ Ramp\ AADT^{\beta_5} \times exp^{\beta_6 \times \left(\frac{MV\ Daytime\ \ Crashes}{MV\ Total\ \ Crashes}\right)} \times Length \times Years$$

Figure 20. Equation. Base model for daytime multivehicle crashes.

Where:

Length (miles) = crossroad to crossroad length.
Ramp Spacing (miles) = spacing between on-ramps and off-ramps.
UP_{Lanes} = number of through lanes upstream of segment.

Table 10 presents the base model for MV total and MV daytime crashes.

Table 10. Base model.

Variable	MV Crashes		MV Daytime Crashes	
	Estimate	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**
Intercept	−20.7833	<0.0001	−22.2779	<0.0001
β_1	1.7974	<0.0001	1.7877	<0.0001
β_2	0.1164	0.0018	0.0999	0.0075
β_3	−0.2372	0.0105	−0.2365	0.0086
β_4	0.2499	0.0030	0.2483	0.0024
β_5	0.1953	0.0054	0.1810	0.0105
β_6	N/A	N/A	1.9315	<0.0001
β_8 (FL)	*−0.1233*	*0.3147*	*−0.1368*	*0.2538*
β_8 (NC)	*0.6952*	*<0.0001*	*0.6659*	*<0.0001*
β_8 (WA)	*0*	*N/A*	*0*	*N/A*
Dispersion	0.1699		0.1517	
AIC	847.6553		799.7545	
BIC	870.7345		825.3980	

Pr = probability; ChiSq = chi-squared.
Note: Italics identify the indicator variable for the state.

The research team added the NDS variables to the Base model using the Hoerl function or conventional exponential form, as shown in figure 21 and figure 22 (Hauer, 2015). One of the terms of the Hoerl function (i.e., β_6 or β_7 for the MV total crash models, and β_7 and β_8 for the MV daytime crash model) was removed from the model if the term was not significant.

$$MV\ Crashes = exp^{Intercept} \times AADT^{\beta_1} \times exp^{\beta_2 \times \left(\frac{1}{Ramp\ Spacing}\right)} \times exp^{\beta_3 \times UP_{Lanes}} \times On\ Ramp\ AADT^{\beta_4} \times Off\ Ramp\ AADT^{\beta_5} \times NDS\ Variable^{\beta_6} \times exp^{\beta_7 \times NDS\ Variable/1000} \times Length \times Years$$

Figure 21. Equation. Base model for all multivehicle crashes with NDS variable.

$$MV\ Daytime\ Crashes = exp^{Intercept} \times AADT^{\beta_1} \times exp^{\beta_2 \times \left(\frac{1}{Ramp\ Spacing}\right)} \times exp^{\beta_3 \times UP_{Lanes}} \times On\ Ramp\ AADT^{\beta_4} \times Off\ Ramp\ AADT^{\beta_5} \times exp^{\beta_6 \times \left(\frac{MV\ Daytime\ Crashes}{MV\ Total\ Crashes}\right)} \times NDS\ Variable^{\beta_7} \times exp^{\beta_8 \times NDS\ Variable/1000} \times Length \times Years$$

Figure 22. Equation. Base model for daytime multivehicle crashes with NDS variable.

Where *NDS Variable* is mean, median, variance, or 85th percentile of density (vehicle/mi), speed (mph), or spacing (ft).

Table 11 through table 16 present details of alternate models for MV total and MV daytime crashes including the NDS variables. For each NDS variable, the research team developed separate models for mean, median, variance, and 85th percentile. Based on the individual variable significance as well as the AIC and BIC values, the recommended model for each NDS variable is identified in bold text. Variables statistically insignificant at the 10-percent level have their p-values identified with an asterisk (*).

For the density models, the coefficients for all variables display the expected direction of effect (i.e., increase in volume associated with more crashes, longer ramp spacing associated with fewer crashes, more upstream lanes associated with fewer crashes) and are statistically significant at the 90-percent level or better (except for upstream lanes for the median of density model). However, looking at the individual models for mean, median, variance, and 85th percentile, the variance of density models presents comparatively higher individual variable significance and AIC values on the lower end of the four models. AIC values can only be compared within the same dependent variable (i.e., MV total crashes and MV daytime crashes).

Table 11. Density models (MV crashes).

Variable	Mean Density		Median Density		Variance Density		85th-per Density	
	Estimate	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**
Intercept	−22.7857	<0.0001	−22.6117	<0.0001	**−21.7712**	**<0.0001**	−23.7286	<0.0001
β_1	1.6247	<0.00091	1.4671	<0.0001	**1.6378**	**<0.0001**	1.6805	<0.0001
β_2	0.1026	0.0060	0.0990	0.0068	**0.1101**	**0.0025**	0.0953	0.0106
β_3	−0.1873	0.0503	−0.1488	0.1257*	**−0.1771**	**0.0562**	−0.1901	0.0403
β_4	0.2852	0.0008	0.3151	0.0002	**0.2759**	**0.0010**	0.2805	0.0007
β_5	0.1826	0.0086	0.1834	0.0074	**0.1984**	**0.0036**	0.1786	0.0095
β_6	0.9984	0.0861	1.3670	0.0119	**0.4417**	**0.0278**	1.0158	0.0279
β_7	n/a	n/a	n/a	n/a	**−0.7018**	**0.0069**	n/a	n/a
β_8 (FL)	*−0.1546*	*0.2054*	*−0.1392*	*0.2429*	***−0.1836***	***0.1380***	*−0.1705*	*0.1602*
β_8 (NC)	*0.7350*	*<0.0001*	*0.7568*	*<0.0001*	***0.7307***	***<0.0001***	*0.7497*	*<0.0001*
β_8 (WA)	*0*	*N/A*	*0*	*N/A*	***0***	***N/A***	*0*	*N/A*
Dispersion	0.1635		0.1585		**0.1583**		0.1596	
AIC	846.7485		843.4986		**844.3329**		844.9653	

Note: Italics identify the indicator variable for the state.

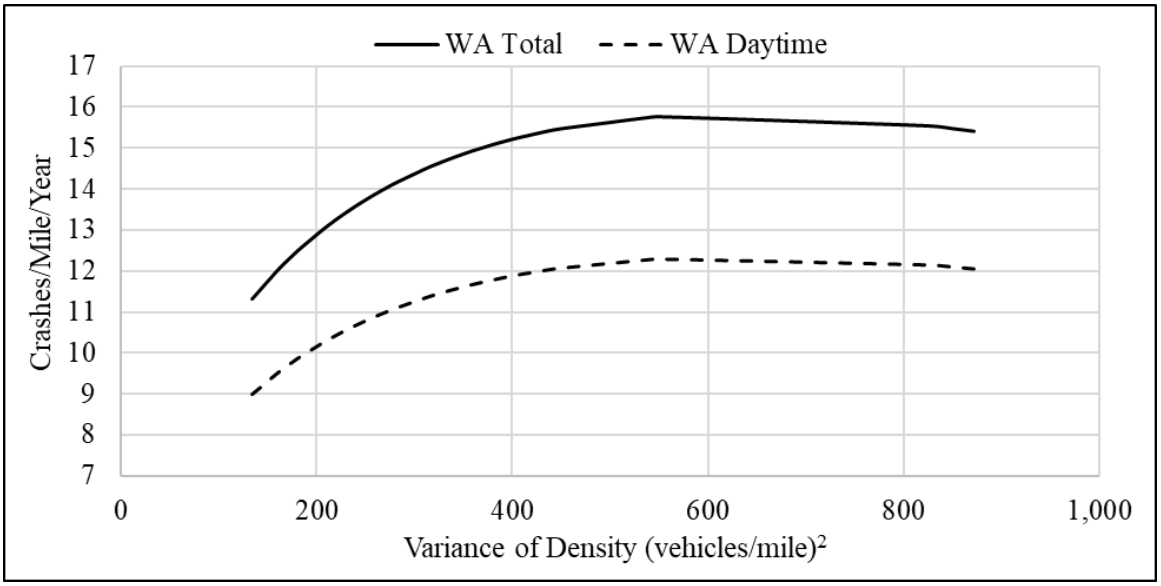
Table 12. Density models (MV daytime crashes).

Variable	Mean Density		Median Density		Variance Density		85th-per Density	
	Estimate	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**
Intercept	−24.0979	<0.0001	−23.9338	<0.0001	**−23.2810**	**<0.0001**	−24.9302	<0.0001
β_1	1.6361	<0.0001	1.4880	<0.0001	**1.6495**	**<0.0001**	1.6874	<0.0001
β_2	0.0905	0.0148	0.0895	0.0143	**0.0965**	**0.0086**	0.0860	0.0198
β_3	−0.1917	0.0394	−0.1550	0.1032	**−0.1838**	**0.0429**	−0.1934	0.0325
β_4	0.2793	0.0008	0.3064	0.0002	**0.2722**	**0.0009**	0.2736	0.0007
β_5	0.1730	0.0135	0.1774	0.0105	**0.1870**	**0.0069**	0.1723	0.0130
β_6	1.8141	<0.0001	1.7115	<0.0001	**1.8371**	**<0.0001**	1.7247	<0.0001
β_7	0.9091	0.1144	1.2693	0.0211	**0.4154**	**0.0370**	0.9298	0.0460
β_8	N/A	N/A	N/A	N/A	**−0.6549**	**0.0175**	N/A	N/A
β_9 (FL)	*−0.1604*	*0.1781*	*−0.1431*	*0.2214*	***−0.1902***	***0.1177***	*−0.1720*	*0.1467*
β_9 (NC)	*0.7083*	*<0.0001*	*0.7348*	*<0.0001*	***0.7017***	***<0.0001***	*0.7256*	*<0.0001*
β_9 (WA)	*0*	*N/A*	*0*	*N/A*	***0***	***N/A***	*0*	*N/A*
Dispersion	0.1465		0.1429		**0.1436**		0.1433	
AIC	799.2959		796.5372		**797.8315**		797.8720	

Note: Italics identify the indicator variable for the state.

Among all models developed for the NDS variables, only the variance of density model used the Hoerl function where the variance of density was used twice in the model (an ln variable and a scaled variable) (Hauer, 2015). For both the MV total and MV daytime crashes, the ln variable has a positive sign, and the scaled variable has a negative sign, which indicates that the crashes will typically increase with an increase in variance of density; however, the crashes will plateau at a certain variance of density and then start decreasing.

To further investigate this relationship, figure 23 presents a plot of predicted crashes versus the variance of density for both MV total and MV daytime crashes. These plots assumed average values for all other variables except for the variance of density. These plots show that for both MV total and MV daytime crash models in table 11 and table 12, the crashes plateau and then start to decrease slightly.



Source: FHWA.

Figure 23. Graph. Plots for protected crashes versus variance of density.

For the speed models, the coefficients for all variables except speed display the expected direction of effect (i.e., increase in volume associated with more crashes, longer ramp spacing associated with fewer crashes, more upstream lanes associated with fewer crashes) and are statistically significant at the 90-percent confidence level or better. When looking at speed, the direction of effect for the mean and median speed variables is counterintuitive but both are statistically insignificant. The 85th percentile of speed variable displays the expected direction of effect but is also statistically insignificant.

The variance of speed models in table 13 and table 14 present comparatively higher individual variable significance and AIC values on the lower end of the four models (alongside the variance of speed displaying the correct direction of effect while being statistically significant).

Table 13. Speed models (MV crashes).

Variable	Mean Speed		Median Speed		Variance Speed		85th-per Speed	
	Estimate	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**
Intercept	−14.1213	0.0154	−17.1823	0.0029	**−19.9201**	**<0.0001**	−21.2160	0.0010
β_1	1.7726	<0.0001	1.7944	<0.0001	**1.6540**	**<0.0001**	1.7977	<0.0001
β_2	0.0932	0.0238	0.1049	0.0101	**0.0823**	**0.0298**	0.1175	0.0034
β_3	−0.2042	0.0324	−0.2195	0.0221	**−0.2092**	**0.0192**	−0.2390	0.0130
β_4	0.2542	0.0023	0.2520	0.0027	**0.2445**	**0.0024**	0.2497	0.0031
β_5	0.1692	0.0202	0.1820	0.0124	**0.1773**	**0.0093**	0.1967	0.0070
β_6	−1.5320	0.2237*	−0.8545	0.5045*	**0.2595**	**0.0031**	0.1015	0.9437*
β_7	n/a	n/a	n/a	n/a	**n/a**	**n/a**	n/a	n/a
β_8 (FL)	*0.0596*	*0.7587*	*−0.0222*	*0.9095*	***−0.2298***	***0.0632***	*−0.1365*	*0.5398*
β_8 (NC)	*0.7990*	*<0.0001*	*0.7519*	*<0.0001*	***0.7077***	***<0.0001***	*0.6888*	*<0.0001*
β_8 (WA)	*0*	*N/A*	*0*	*N/A*	***0***	***N/A***	*0*	*N/A*
Dispersion	0.1660		0.1684		**0.1554**		0.1700	
AIC	848.1914		849.2115		**841.0452**		849.6504	

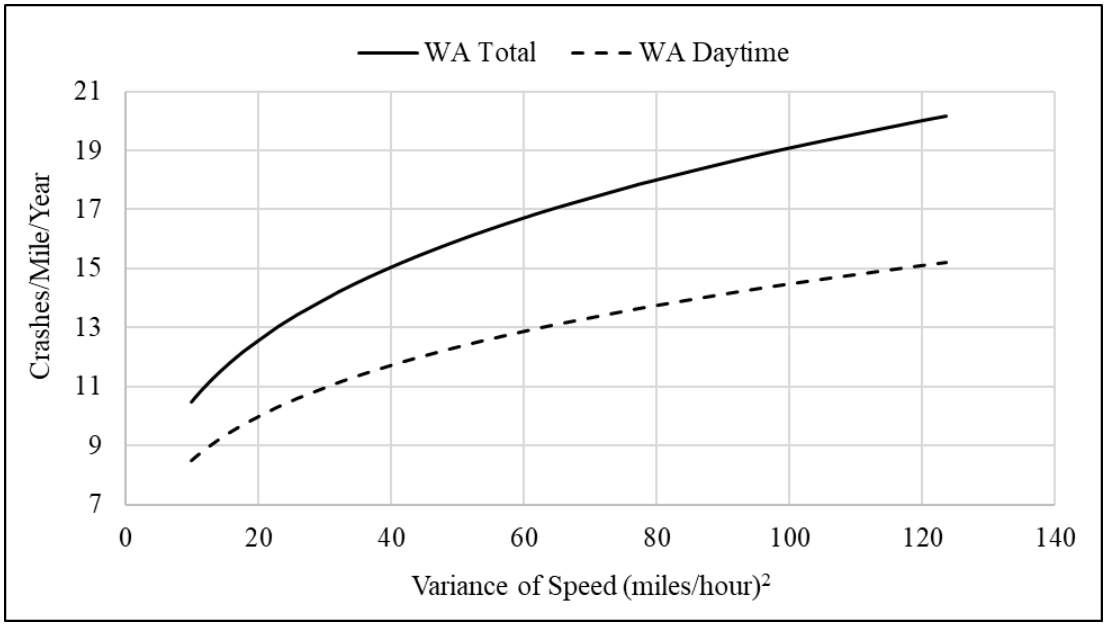
Note: Italics identify the indicator variable for the state.

Table 14. Speed models (MV daytime crashes).

Variable	Mean Speed		Median Speed		Variance Speed		85th-per Speed	
	Estimate	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**
Intercept	−16.5698	0.0054	−19.2213	0.0008	**−21.4063**	**<0.0001**	−23.0958	0.0003
β_1	1.7689	<0.0001	1.7862	<0.0001	**1.6666**	**<0.0001**	1.7881	<0.0001
β_2	0.0838	0.0365	0.0916	0.0213	**0.0776**	**0.0393**	0.1015	0.0099
β_3	−0.2079	0.0262	−0.2214	0.0180	**−0.2122**	**0.0159**	−0.2400	0.0105
β_4	0.2495	0.0021	0.2491	0.0023	**0.2405**	**0.0025**	0.2481	0.0024
β_5	0.1630	0.0245	0.1713	0.0183	**0.1761**	**0.0112**	0.1833	0.0118
β_6	1.8011	<0.0001	1.8829	<0.0001	**1.6056**	**0.0002**	1.9417	<0.0001
β_7	−1.3006	0.3100*	−0.7211	0.5707*	**0.2299**	**0.0106**	0.1910	0.8932*
β_8	n/a	n/a	n/a	n/a	**n/a**	**n/a**	n/a	n/a
β_9 (FL)	*0.0241*	*0.9034*	*−0.0488*	*0.8033*	***−0.2222***	***0.0685***	*−0.1620*	*0.4674*
β_9 (NC)	*0.7601*	*<0.0001*	*0.7159*	*<0.0001*	***0.6928***	***<0.0001***	*0.6534*	*<0.0001*
β_9 (WA)	*0*	*N/A*	*0*	*N/A*	***0***	***N/A***	*0*	*N/A*
Dispersion	0.1488		0.1504		**0.1427**		0.1518	
AIC	800.7348		801.4343		**795.2509**		801.7365	

Note: Italics identify the indicator variable for the state.

Figure 24 presents a plot of predicted crashes versus the variance of speed for both MV total and MV daytime crashes. This plot assumed average values for all other variables except for the variance of speed.



Source: FHWA.

Figure 24. Graph. Plots for protected crashes versus variance of speed.

For the spacing models, the coefficients for all variables display the expected direction of effect (i.e., increase in volume associated with more crashes, longer ramp spacing associated with fewer crashes, more upstream lanes associated with fewer crashes, increased spacing associated with fewer crashes) and are statistically significant at the 90-percent confidence level or better (except for the variance of spacing variable). However, looking at the individual models for mean, median, variance, and 85th percentile in table 15 and table 16, the mean of spacing model presents comparatively higher individual variable significance and AIC values on the lower end of the four models.

Table 15. Spacing models (MV crashes).

Variable	Mean Spacing		Median Spacing		Variance Spacing		85th-per Spacing	
	Estimate	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**
Intercept	**−9.9089**	**0.0178**	−14.4080	<0.0001	−17.3658	<0.0001	−12.2687	0.0035
β_1	**1.5088**	**<0.0001**	1.5992	<0.0001	1.7339	<0.0001	1.5615	<0.0001
β_2	**0.1011**	**0.0046**	0.1000	0.0065	0.1261	0.0009	0.1146	0.0014
β_3	**−0.1561**	**0.0937**	−0.1752	0.0624	−0.2240	0.0159	−0.1699	0.0735
β_4	**0.2789**	**0.0005**	0.2751	0.0008	0.2430	0.0038	0.2591	0.0016
β_5	**0.1868**	**0.0056**	0.1819	0.0081	0.2036	0.0036	0.1964	0.0040
β_6	**−1.5068**	**0.0035**	−0.8551	0.0217	−0.2880	0.2383*	−1.0616	0.0223
β_7	**N/A**	**N/A**	N/A	N/A	N/A	N/A	N/A	N/A
β_8 (FL)	***−0.1335***	***0.2559***	*−0.1112*	*0.3520*	*−0.1483*	*0.2296*	*−0.1453*	*0.2239*
β_8 (NC)	***0.7911***	***<0.0001***	*0.7811*	*<0.0001*	*0.6957*	*<0.0001*	*0.7455*	*<0.0001*
β_8 (WA)	***0***	***N/A***	*0*	*N/A*	*0*	*N/A*	*0*	*N/A*
Dispersion	**0.1540**		0.1596		0.1671		0.1592	
AIC	**841.4435**		844.5288		848.2704		844.5548	

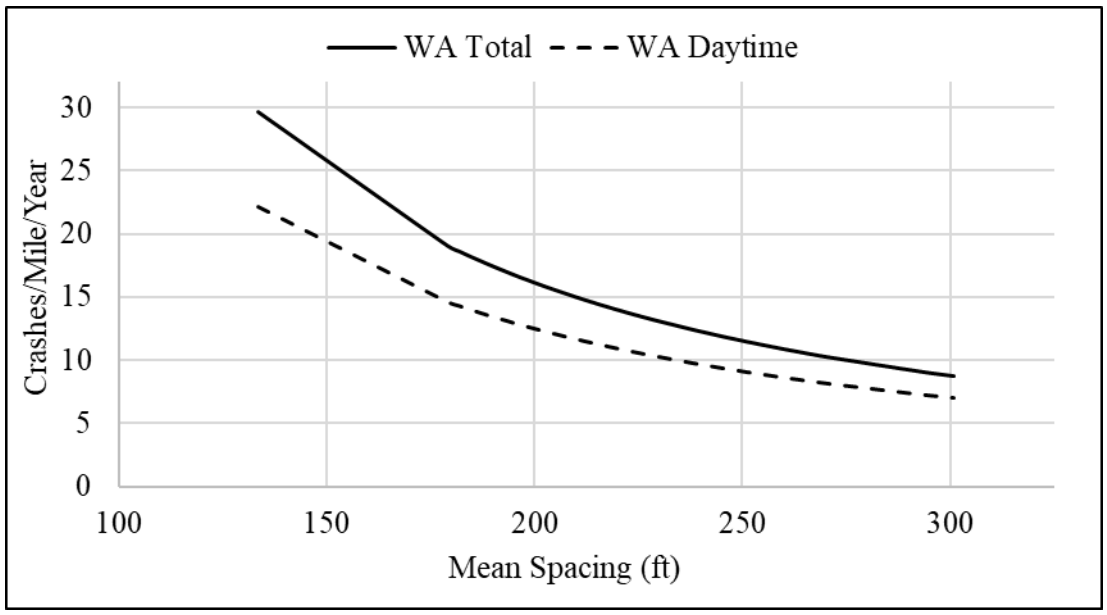
Note: Italics identify the indicator variable for the state.

Table 16. Spacing models (MV daytime crashes).

Variable	Mean Spacing		Median Spacing		Variance Spacing		85th-per Spacing	
	Estimate	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**	**Estimate**	**Pr > ChiSq**
Intercept	**−12.0171**	**0.0050**	−16.3754	<0.0001	−18.6410	<0.0001	−14.3372	0.0005
β_1	**1.5256**	**<0.0001**	1.6133	<0.0001	1.7204	<0.0001	1.5729	<0.0001
β_2	**0.0947**	**0.0080**	0.0924	0.0116	0.1096	0.0037	0.1024	0.0045
β_3	**−0.1626**	**0.0747**	−0.1805	0.0504	−0.2242	0.0128	−0.1762	0.0560
β_4	**0.2732**	**0.0006**	0.2685	0.0008	0.2427	0.0028	0.2574	0.0013
β_5	**0.1872**	**0.0066**	0.1791	0.0101	0.1902	0.0072	0.1890	0.0067
β_6	**1.5865**	**0.0002**	1.6442	0.0001	1.9696	<0.0001	1.7874	<0.0001
β_7	**−1.4134**	**0.0077**	−0.7820	0.0426	−0.3114	0.1940*	−0.9917	0.0308
β_8	**N/A**	**N/A**	N/A	N/A	N/A	N/A	N/A	N/A
β_9 (FL)	***−0.1364***	***0.2388***	*−0.1158*	*0.3257*	*−0.1670*	*0.1673*	*−0.1550*	*0.1856*
β_9 (NC)	***0.7732***	***<0.0001***	*0.7589*	*<0.0001*	*0.6655*	*<0.0001*	*0.7194*	*<0.0001*
β_9 (WA)	***0***	***N/A***	*0*	*N/A*	*0*	*N/A*	*0*	*N/A*
Dispersion	**0.1392**		0.1441		0.1486		0.1425	
AIC	**794.8728**		797.7253		800.0756		797.1918	

Note: Italics identify the indicator variable for the state.

Figure 25 presents a plot of predicted crashes versus the mean of spacing for both MV total and MV daytime crashes. This plot assumed average values for all other variables except for the mean of spacing.



Source: FHWA.

Figure 25. Graph. Plots for protected crashes versus mean of spacing.

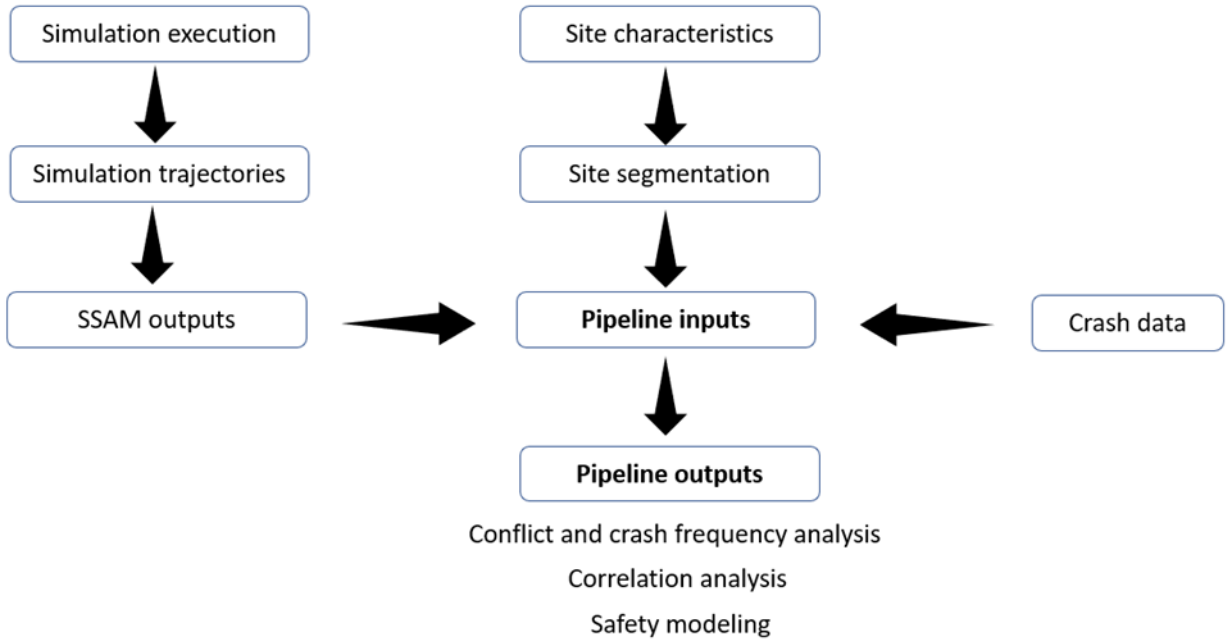
The analysis in this section provides promising results on the measurable effects of NDS-related variables for macroscopic safety analysis. Even though the most direct application of NDS measures is expected at a microscopic scale, the significance of density, speed, and spacing on the models opens opportunities to evaluate additional alternatives and to conduct research at different scales. For example, incorporating metrics on crash prediction at certain times of day (i.e., different temporal scales) or for specific sections that often exhibit large or sudden changes in traffic conditions (i.e., different spatial scales) could be reflected in variables similar to those from NDS evaluated in this report.

NDS-CALIBRATED CAR-FOLLOWING PARAMETERS AND VEHICLE CONFLICT ESTIMATIONS

This second exploration presents results from an initial evaluation of the effects of calibrating car-following parameters in the estimation of vehicle conflicts as identified by the Surrogate Safety Assessment Model (SSAM) (FHWA, 2022). This analysis determines if trends in conflict data produced by SSAM are likely to have a closer association to crash events than those obtained with default simulation parameters. Important caveats must be noted in this exercise, because the correlation of crash data and such surrogate measures is not well-established, and precise metrics or expected outcomes may not represent the full potential of improved calibration in microscopic safety modeling.

This exercise introduces a pipeline the research team built using open-source code in Python, making use of automated scripts to load a simulation file, extract trajectory data, run the

trajectory data through SSAM, and analyze the data for a given site or for sections within a site (FHWA, 2022; Python Software Foundation, 2021). Figure 26 shows a schematic representation of the pipeline.



Source: FHWA.

Figure 26. Illustration. Schematic representation of pipeline for processing trajectories and conflicts from SSAM.

In the simulation execution step, the research team instructs the automated script to access and set up desired parameters in VISSIM® using the COM interface (VISSIM, n.d.). The COM API allows setting parameter values in the simulation and running a scenario to output the desired vehicle trajectory data before data ingestion in SSAM (FHWA, 2022). Then, the command line interface (CLI) of SSAM is accessed to automatically input the trajectories and output the conflicts given user-defined time-to-collision and post-encroachment time.

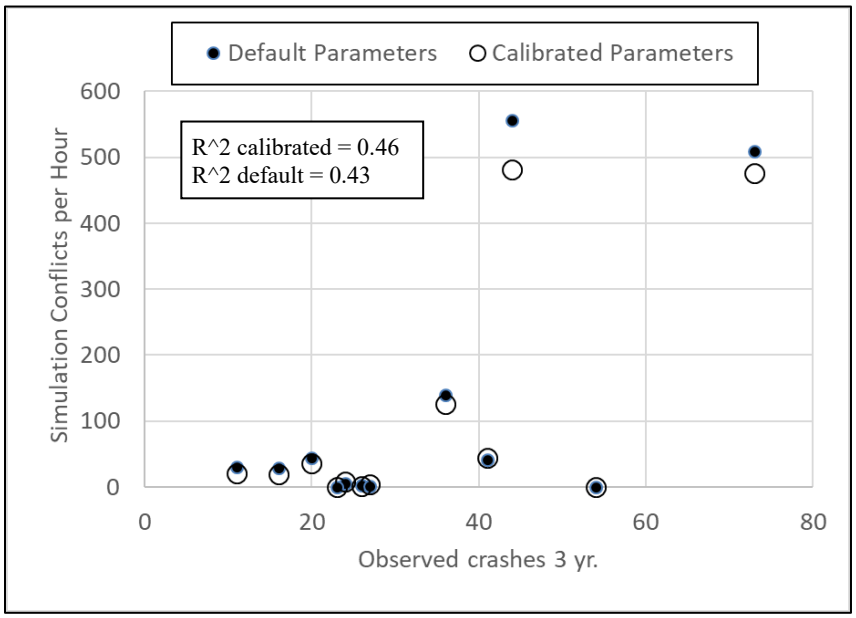
As part of the input data, the research team uses SUMO NetConvert CLI to extract VISSIM's links and connectors and construct a virtual road geometry for later analysis (SUMO, n.d.; VISSIM, n.d.). A series of site segmentation procedures is also carried out, using a separate input file containing the *X*-*Y* coordinates of key points within a site, including the beginning of the site, on-ramp location, off ramp location, and end of the site. The research team segmented sites in the following sections: upstream from entrance gore, weaving section, and downstream from exit gore. Additional or different sections can be defined in the X-Y coordinate input file to accommodate different site configurations and analysis areas.

Then, the research team applies a geographical matching process to assign conflicts from SSAM and crash events to site sections (FHWA, 2022). The user can use the pipeline outputs to estimate metrics such as conflicts per crashes, and correlation analysis for predefined sections.

Initial simulation runs used input values equivalent to 8 percent of the AADT, producing very low frequency of conflicts. While these outcomes were not beneficial for an evaluation of associations between conflicts and crashes, the outcomes confirmed that the simulation sites were built correctly, and vehicles traversed sites exhibiting reasonable behavior. Based on these initial outcomes, the research team decided to increase hourly volumes to 12 percent of the AADT, aiming at obtaining a larger sample of conflicts to analyze.

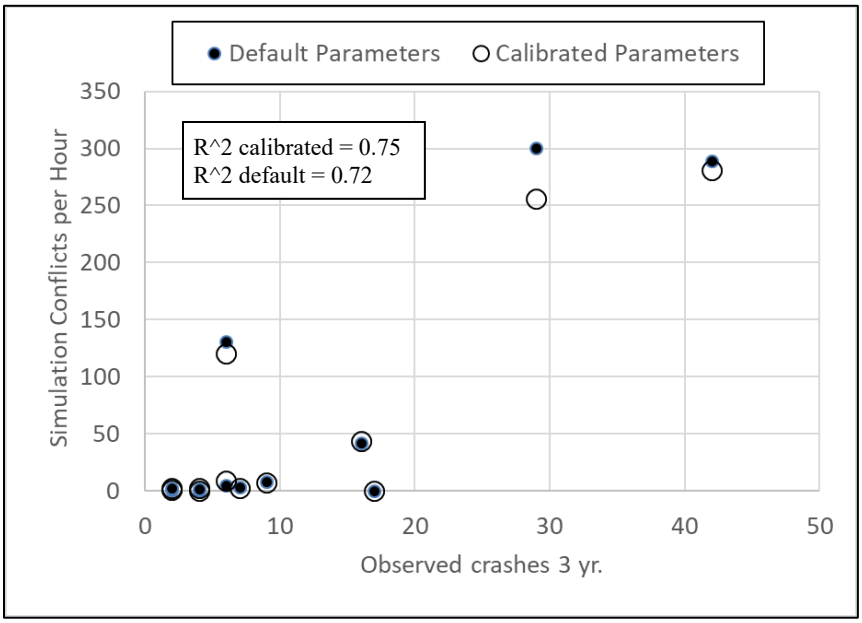
The research team first ran 12 simulation scenarios with default parameters, and then with parameters calibrated using NDS. The research team processed 10 runs for calibrated and noncalibrated parameters, extracting usable data for 2 h of simulation time from each set. For this exploration, all simulation scenarios had the same parameters, because they represented the same type of freeway segments and have similar configurations. However, indepth calibration of simulated scenarios should be conducted with as much field data as possible to represent observed conditions accurately.

Figure 27 shows the average hourly conflict frequencies and the cumulative 3-yr crash frequencies for the 12 sites investigated. Overall, results showed consistent positive trends between conflicts and crashes, with a slightly stronger association between assuming a linear relation, as indicated by the regression R^2 in the figures. In addition, the research team observed a stronger association for data extracted between the gores (figure 27-B) compared to those including sitewide data (figure 27-A). This information points to more consistent safety performance along this section within the sites, even though that section also has the highest crash frequencies.



Source: FHWA.

A. Example of sitewide conflicts and crashes.



Source: FHWA.

B. Example of conflicts and crashes between gores.

Figure 27. Graphs. Simulation conflicts versus crash frequencies for sample sites.

In terms of the effects of calibration, the research team observed a small effect in the tested sites when the car-following parameters were adjusted more closely to NDS data. The research team also observed only a slight improvement in the associations between crashes and conflicts. These results are not surprising, given that the simulations in this exercise represent basic freeway segments under expected normal operations, without special events that could generate disturbances in traffic.

The research team recommends more detailed evaluations to further assess conflict and crash associations before and after calibration. Such scenarios would require more specific volume data to represent typical traffic fluctuations and the corresponding safety performance observed within the modeled periods. Researchers can take advantage of recent freeway data collection systems to gather the necessary data for such evaluation.

CHAPTER 7. CONCLUSIONS AND RECOMMENDATIONS

This research leveraged SHRP2 NDS datasets to extract vehicle-level car-following measures from freeway segments using large samples, with more than 1,600 h of car-following data from more than 1,700 unique drivers and developed distributions to serve as targets to support the calibration of microscopic traffic simulation (VTTI, 2020).

The proposed calibration process complements current calibration practices, where the simulation car-following parameters are modified to produce similar values of macroscopic measures of performance such as travel time, delay, and queues compared to those observed in the field. The additional calibration steps are intended to produce vehicle-to-vehicle interactions from the simulation that reflect a naturalistic behavior. Verification of vehicle-to-vehicle interactions is a particularly important task, as simulation parameters may often provide reasonable default car-following settings, but such behavior could be modified during the macroscopic-based calibration, generating substantially different (and unintended) behavior.

NDS datasets were subject to custom post-processing analysis to identify leader-follower pairs and construct time series of the instrumented vehicle kinematics while in car-following conditions (VTTI, 2020). The research team characterized driving behavior in terms of three main metrics: spacing between the instrumented vehicle and a leader in the same lane, acceleration of the instrumented vehicle, and acceleration change rate (jerk) of the instrumented vehicle. The research team contextualized the driving behavior in terms of the instrumented vehicle speed, capturing changes in traffic conditions. The research team established the adequacy of using car-following speed as a proxy to determine traffic conditions by investigating and confirming that the field data produced a consistent relation between speed and density in scenarios with low-, medium-, and high-traffic demands.

The NDS targets developed in this research provide a new level of detail on naturalistic driving behavior for speeds ranging from 5 mph to 85 mph, and therefore cover a wide range of traffic conditions. The research team also conducted an analysis to identify a distribution of expected standstill distances, providing additional findings that could also be helpful for simulation calibration. The high degree of consistency of the empirical distributions is notable for all three metrics and speed levels, which supports the validity of the process to identify the necessary sites and freeway segments, specific traversals, and NDS metric extraction algorithms developed by the team to produce large and balanced samples to characterize driving behavior.

The research team identified vehicle spacing as the preferred metric for microscopic calibration because it provided the largest separation between speed groups, and therefore the most unambiguous targets for different traffic conditions. The NDS targets also expand the extent of potential comparisons with simulation outputs, because complete distributions for specific traffic conditions can be evaluated in addition to simpler central tendency metrics such as mean or median values.

Given the need to transform trajectory data to generate leader-follower pairs, to develop time series of vehicle spacing, and to perform comparisons with NDS targets, the research team developed a portable open-source tool to complete such tasks. The NACT tool reads generic

trajectories from microscopic simulation and produces outputs to direct a successful microscopic calibration process. The NACT tool exports files with figures contrasting the NDS and the simulation spacing percentiles along with a 95-percent confidence level band, as well as results of statistical tests comparing the distributions. This research also produced a guideline on the use of the NACT tool, which is a separate stand-alone document.

While the NACT tool analyzes input data to generate comparisons between simulation and NDS targets, the tool does not provide further guidance on how to adjust specific car-following parameters to improve goodness of fit. The tool is model agnostic and compares generic vehicle trajectory data without knowledge of the underlying simulation package.

As a secondary objective, the research team evaluated the potential benefits of incorporating macroscopic measures derived from NDS datasets, such as speed, spacing, and density in safety modeling (VTTI, 2020). For each of these metrics, their mean, median, variance, and 85th percentile were evaluated as contributing factors of well-constructed crash frequency models, indicating potential benefits worth further exploration. In particular, the research team identified increase in the density variance, increase in the speed variance, and decrease in the mean spacing as having consistent and significant effects associated to increases in multivehicle crash frequencies. Final recommended models quantify the effects of these variables in relation to other contributing factors. These findings suggest that the implications of NDS-derived metrics on safety analysis support further efforts to conduct more focused research on this topic. Questions related to the effects of vehicle spacing on safety-related events are of particular interest.

An additional exploration of the safety implications of NDS datasets on safety was also conducted at the microscopic level (VTTI, 2020). This work used a custom pipeline to automate the evaluation of potential improvements to generate more naturalistic vehicle conflict data when simulation scenarios are calibrated using NDS targets compared to default simulation parameters. The research team used the SSAM tool as a conflict identification tool part of the pipeline, where both frequency and location of conflicts are compared to those from crash history (FHWA, 2022).

An initial exploration of 12 sites revealed that under normal conditions without disturbances, the number of conflicts followed similar trends before and after calibration. Slightly lower conflict frequencies were observed for the calibrated scenarios, and in both cases, conflicts showed a positive relationship with crash frequencies, which indicates that calibration did not degrade the operation of simulation sites as vehicles moved under constant demand, and a small but positive effect of calibration reduced conflicts in such normal conditions.

Further indepth analysis to precisely determine the effects of conflicts, as well as their associations with crashes before and after calibration, requires detailed demand fluctuations and observed safety performance within the simulated periods. The research team recommends a careful experimental design to systematically evaluate the effects of calibration in the frequency of conflicts and their associations with crash data under different saturation levels and traffic varying conditions. Such evaluation is likely to require a large number of sites but could provide invaluable information on the full extent of the effects of calibration in safety assessments using surrogate measures from SSAM (FHWA, 2022).

The NDS targets developed in this research were extracted from the driving behavior of passenger cars on freeway segments under good weather and daytime conditions. The research team recommends additional research to adjust driving behavior (if needed) for additional vehicle types and environmental factors that may affect the spacing distributions at each speed level. Additionally, different segment types, such as on ramps or off ramps**,** are also recommended to be analyzed to verify potential adjustments to the observed car-following behavior in mainlines.

ACKNOWLEDGMENTS

The original map in figure 1 is the copyright property of Google® Earth™ and can be accessed from https://www.google.com/earth (Google, 2023). The authors inserted the blue polygon to illustrate the extent of one of the selected sites along I–440. The image includes location, scale, and elevation.

REFERENCES

Abbas, M., B. Higgs, Z. Adam, and A. M. Flintsch. 2011. "Comparison of Car-Following Models when Calibrated to Individual Drivers using Naturalistic Data." Report No. 11-3750. Presented at the *Transportation Research Board 90th Annual Meeting*. Washington, DC: Transportation Research Board (TRB).

Balakrishna, R., C. Antoniou, M. Ben-Akiva, H. N. Koutsopoulos, and Y. Wen. 2007. "Calibration of Microscopic Traffic Simulation Models: Methods and Application." *Transportation Research Record* 1999, no. 1: 198–207. https://journals.sagepub.com/doi/10.3141/1999-21, last accessed October 10, 2023.

Chandler, R.E., R. Herman, and E. W. Montroll. 1958. "Traffic Dynamics: Studies in Car Following." *Operations Research* 6 no. 2: 165–184.

Chong, L., M. Abbas, A. M. Flintsch, and B. Higgs. 2013. "A Rule-Based Neural Network Approach to Model Driver Naturalistic Behavior in Traffic." *Transportation Research Part C: Emerging Technologies* 32: 207–223.

Dowling, R., A. Skabardonis, and V. Alexiadis. 2004. *Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software*. Report No. FHWA-HRT-04-040. Washington, DC: FHWA. https://ops.fhwa.dot.gov/trafficanalysistools/tat_vol3/Vol3_Guidelines.pdf, last accessed October 10, 2023.

Esri. 2015. *ArcGIS Pro (ArcGIS for Desktop)* (software). Version 10.x. https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview?rsource=%2Fsoftware%2Farcgis%2Farcgis-for-desktop%2Findex.html, last accessed October 10, 2023.

FHWA. n.d. "Highway Safety Information System (HSIS)" (web page). https://highways.dot.gov/research/safety/hsis, last accessed October 10, 2023.

FHWA. 2022. "Surrogate Safety Assessment Model Overview" (web page). https://highways.dot.gov/research/safety/ssam/surrogate-safety-assessment-model-overview, last accessed October 10, 2023.

Fernandez, A. 2011. "Secondary Tasks in Steady State Car Following Situations." Master's thesis. Chalmers University of Technology. https://odr.chalmers.se/server/api/core/bitstreams/d14abe4d-57cd-4e9f-bb39-8f38fa3ca845/content, last accessed October 10, 2023.

FDOT. 2021. *Traffic Analysis Handbook*. Tallahassee, FL: Florida DOT. https://fdotwww.blob.core.windows.net/sitefinity/docs/default-source/planning/systems/systems-management/document-repository/traffic-analysis/traffic-analysis-handbook_05-2021.pdf, last accessed October 10, 2023.

Ge, Q. and M. Menendez. 2012. "Sensitivity Analysis for Calibrating VISSIM in Modeling the Zurich Network." Presented at the *12th Swiss Transport Research Conference* (STRC). Monte Verità, Ascona, Switzerland: STRC. https://www.strc.ch/2012/Ge_Menendez.pdf, last accessed October 10, 2023.

Geng, X., H. Liang, H. Xu, and B. Yu. 2016. "Influences of Leading-Vehicle Types and Environmental Conditions on Car-Following Behavior." *IFAC-PapersOnLine* 49, no. 15: 151–156.

Gipps, P. G. 1981. "A Behavioral Car-Following Model for Computer Simulation." *Transportation Research Part B: Methodological* 15, no. 2: 105–111.

Google® for Developers. n.d. *Keyhole Markup Language* (software). https://developers.google.com/kml, last accessed October 10, 2023.

Google. 2023. "Google® Earth™" (web page). https://www.google.com/earth, last accessed October 12, 2023.

Gomes, G., A. May, and R. Horowitz. 2004. *Calibration of VISSIM for a Congested Freeway*. California Partners for Advanced Transit and Highways Research Report No. UCB-ITS-PRR-2004-4. Berkeley, CA: University of California at Berkeley. https://horowitz.me.berkeley.edu/Publications_files/All_papers_numbered/Gomes_may_horowitz_PATH_report04.pdf, last accessed October 10, 2023.

Hale, D. K., X. Li, A. Ghiasi, D. Zhao, F. Khalighi, M. Aycin, and R. M. James. 2021. *Trajectory Investigation for Enhanced Calibration of Microsimulation Models*. Report No. FHWA-HRT-21-071. Washington, DC: FHWA.

Hammit, B., R. James, and M. Ahmed. 2018. "A Case for Online Traffic Simulation: Systematic Procedure to Calibrate Car-Following Models Using Vehicle Data," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC*). Piscataway, NJ, 3785–3790.

Hauer, E. 2015. *The Art of Regression Modeling in Road Safety*. New York, NY: Springer Publishing.

Higgs, B. 2011. "Application of Naturalistic Truck Driving Data to Analyze and Improve Car Following Models." Master's thesis. Virginia Polytechnic Institute and State University (Virginia Tech). https://vtechworks.lib.vt.edu/handle/10919/36089, last accessed October 10, 2023.

Higgs, B., and M. Abbas. 2014. "Multi-Resolution Comparison of Car-Following Models Using Naturalistic Data." Report No. 14-4528. Presented at the *Transportation Research Board 93rd Annual Meeting*. Washington, DC: TRB.

Houchin, A., J. Dong, N. Hawkins, and S. Knickerbocker. 2015. "Measurement and Analysis of Heterogenous Vehicle Following Behavior on Urban Freeways: Time Headways and

Standstill Distances." In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. Piscataway, NJ, 888–893.

Iowa State University. 2023. "Center for Transportation Research and Education SHRP 2—Roadway Information Database" (web page). https://ctre.iastate.edu/shrp2-rid/, last accessed October 10, 2023.

Kaths, H., A. Keler, and K. Bogenberger. 2021. "Calibrating the Wiedemann 99 Car-Following Model for Bicycle Traffic." *Sustainability* 13, no. 6: 3487.

LeBlanc, D. J., S. Bao, J. R. Sayer, and S. Bogard. 2013. "Longitudinal Driving Behavior with Integrated Crash-Warning System: Evaluation from Naturalistic Driving Data." *Transportation Research Record* 2365, no. 1: 17–21.

Lownes, N. E., and R. B. Machemehl. 2006. "VISSIM: A Multi-Parameter Sensitivity Analysis." In *IEEE Winter Simulation Conference*. Piscataway, NJ, 1406–1413. https://ieeexplore.ieee.org/document/4117765, last accessed October 10, 2023.

MathWorks, Inc. 2022. *MATLAB* (software). Version 2018a.

Microsoft®. 2023. *Microsoft Windows®* (software). https://www.microsoft.com/en-us/windows, last accessed October 10, 2023.

NCHRP. 2023. *SHRP 2 Naturalistic Driving Study Pooled Fund: Advancing Implementable Solutions*. TPF-5(361). Washington, DC: National Cooperative Highway Research Program (NCHRP). https://www.pooledfund.org/Details/Study/613, last accessed March 13, 2024.

Oregon DOT. 2011. *Protocol for Vissim Simulation*. Salem, OR: Oregon DOT. https://www.oregon.gov/odot/Planning/Documents/APMv2_Add15A.pdf, last accessed October 10, 2023.

Park, B., and H. Qi. 2006. "Microscopic Simulation Model Calibration and Validation for Freeway Work Zone Network - A Case Study of VISSIM." In *IEEE 2006 Intelligent Transportation Systems Conference*. Piscataway, NJ, 1471–1476. https://ieeexplore.ieee.org/document/1707431, last accessed October 10, 2023.

Punzo, V., and M. Montanino. 2016. "Speed or Spacing? Cumulative Variables, and Convolution of Model Errors and Time in Traffic Flow Models Validation and Calibration." *Transportation Research Part B: Methodological* 91: 21–33.

Python Software Foundation. 2021. Python (software). Version 3.10. http://www.python.org, last accessed October 10, 2023.

Rothery, R. W. R., R. Silver, R. Herman, and C. Torner. 1964. "Analysis of Experiments on Single-Lane Bus Flow." *Operations Research* 12, no. 6: 913–933.

Sangster, J., H. Rakha, and J. Du. 2013. “Application of Naturalistic Driving Data to Modeling of Driver Car-Following Behavior.” *Transportation Research Record* 2390: 20–33.

SAS Institute Inc. 2020. *SAS/ACCESS*® (software). Version 9.4.

Seo, T., T. Kusakabe, and Y. Asakura. 2015. “Estimation of Flow and Density Using Probe Vehicles with Spacing Measurement Equipment.” *Transportation Research Part C: Emerging Technologies* 53: 134–150.

SUMO. n.d. *NetConvert* (software). https://sumo.dlr.de/docs/netconvert.html, last accessed October 10, 2023.

Toga, A. W., ed. 2015. *Brain Mapping: An Encyclopedic Reference. Volume 1: Acquisition Methods, Methods and Modeling*. Amsterdam, Netherlands: Elsevier.

Treiber, M., A. Hennecke, and D. Helbing. 2000. “Congested Traffic States in Empirical Observations and Microscopic Simulations.” *Physical Review E* 62, no. 2: 1805–1824.

USGS. 2016. *National Map Downloader* (software). Version 2.0. https://apps.nationalmap.gov/downloader/#/, last accessed October 10, 2023.

Van Aerde, M., and H. A. Rakha. 1995. “Multivariate Calibration of Single Regime Speed–Flow–Density Relationships.” In Pacific Rim TransTech Conference. 1995 Vehicle Navigation and Information Systems Conference Proceedings. 6th International VNIS. A Ride into the Future. Piscataway, NJ, 334–341.

VDOT. 2020. *VDOT Vissim User Guide*. Richmond, VA: Virginia DOT. https://www.virginiadot.org/business/resources/VDOT_Vissim_UserGuide_Version2.0_Final_2020-01-10.pdf, last accessed October 10, 2023.

Virginia Tech Transportation Institute. 2020. “InSight Data Access Website - SHRP2 Naturalistic Driving Study” (web page). https://insight.shrp2nds.us/, last accessed October 10, 2023.

VISSIM. n.d. “VISSIM” (web page). https://www.vissim.no/, last accessed October 10, 2023.

WisDOT. 2019. “Traffic Analysis and Modeling: Microscopic Simulation Traffic Analysis.” In *Traffic Engineering, Operations & Safety Manual*, chapter 16, Section 20. Madison, WI: Wisconsin DOT. https://wisconsindot.gov/dtsdManuals/traffic-ops/manuals-and-standards/teops/16-20.pdf, last accessed October 10, 2023.

Wunderlich, K. E., V. Meenakshy, and W. Peiwei. 2019. *Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software 2019 Update to the 2004 Version*. Report No. FHWA-HOP-18-036. Washington, DC: FHWA. https://ops.fhwa.dot.gov/publications/fhwahop18036/index.htm, last accessed October 10, 2023.

Zhu, M., X. Wang, and A. Tarko. 2018. "Modeling Car-Following Behavior on Urban Expressways in Shanghai: A Naturalistic Driving Study." *Transportation Research Part C: Emerging Technologies* 93: 425–445. https://www.sciencedirect.com/science/article/abs/pii/S0968090X18308635, last accessed October 10, 2023.

Recycled
Recyclable

Recommended citation: Federal Highway Administration,
Verification and Calibration of Microscopic Traffic Simulation Using Driver Behavior and Car-Following Metrics for Freeway Segments
(Washington, DC: 2024) https://doi.org/10.21949/1521775

HRSO-20/04-24(WEB)E