VTRC
Virginia Transportation
Research Council

We Bring Innovation to Transportation

Methods to Analyze and Predict Interstate Travel Time Reliability

http://www.virginiadot.org/vtrc/main/online_reports/pdf/22-r2.pdf

XIAOXIAO ZHANG, Ph.D.
Graduate Research Assistant
Department of Engineering Systems and Environment
University of Virginia

MO ZHAO, Ph.D.
Research Scientist
Virginia Transportation Research Council

JUSTICE APPIAH, Ph.D., P.E.
Senior Research Scientist
Virginia Transportation Research Council

MICHAEL D. FONTAINE, Ph.D., P.E.
Associate Director
Virginia Transportation Research Council

Final Report VTRC 22-R2

VIRGINIA TRANSPORTATION RESEARCH COUNCIL
530 Edgemont Road, Charlottesville, VA 22903-2454
vtrc.virginiadot.org

Standard Title Page - Report on Federally Funded Project

1. Report No.: FHWA/VTRC 22-R2	2. Government Accession No.:	3. Recipient's Catalog No.:

4. Title and Subtitle:
Methods to Analyze and Predict Interstate Travel Time Reliability

5. Report Date:
July 2021

6. Performing Organization Code:

7. Author(s):
Xiaoxiao Zhang, Ph.D., Mo Zhao, Ph.D., Justice Appiah, Ph.D., P.E., Michael D. Fontaine, Ph.D., P.E.

8. Performing Organization Report No.:
VTRC 22-R2

9. Performing Organization and Address:
Virginia Transportation Research Council
530 Edgemont Road
Charlottesville, VA 22903

10. Work Unit No. (TRAIS):

11. Contract or Grant No.:
114313

12. Sponsoring Agencies' Name and Address:

Virginia Department of Transportation	Federal Highway Administration
1401 E. Broad Street	400 North 8th Street, Room 750
Richmond, VA 23219	Richmond, VA 23219-4825

13. Type of Report and Period Covered:
Final

14. Sponsoring Agency Code:

15. Supplementary Notes:
This is an SPR-B report

16. Abstract:

The Moving Ahead for Progress in the 21st Century Act (MAP-21) defined requirements for system reliability performance measures. Under MAP-21, state departments of transportation are responsible for reporting travel time reliability and for setting targets and showing progress toward those targets. In order to know how to improve travel time reliability and what to expect from investments in transportation infrastructure, these agencies need a better understanding of the factors that affect travel time reliability and methods to predict future travel time reliability. The purpose of this study was to quantify the factors influencing travel time reliability and investigate how to account for these factors in setting reliability targets and communicating progress.

To achieve these objectives, this study developed models to estimate quantiles (the 50th, 80th, and 90th) of travel time distributions to quantify the effects of travel time reliability impact factors and predict select reliability measures. First, linear quantile mixed models (LQMMs) were built using both data maintained by the Virginia Department of Transportation (VDOT) and crowdsourced event data. Model results using the crowdsourced data were unstable and difficult to interpret because of data quality issues such as unbalanced spatial density, duplicate reporting, and inconsistent event classification because of individual observer bias. The results using VDOT-maintained data were more reliable and interpretable. Those models showed that frequencies of non-recurrent events, such as incidents and weather, were correlated with higher travel time percentiles. The LQMM was compared with the trend line approach, a common prediction method used in practice, and the results showed that LQMMs significantly improved the accuracy of predictions over the trend line approach based on mean absolute percent error. Generalized random forest (GRF) models were also tested as an alternative prediction method. GRF models improved the prediction accuracy over LQMMs for the 50th and 80th percentiles, but the accuracy was slightly worse for the 90th percentile. In addition, the GRF models could also reflect the impact of variables that were removed from LQMMs because of insignificance, such as the presence of safety service patrols.

Before-after studies were conducted to illustrate the application of LQMMs and GRF models. LQMMs captured the changes in the 90th percentile travel times better, and GRF models captured the changes of level of travel time reliability better in most cases. GRF models were more sensitive to the reliability changes caused by non-recurrent events, such as incidents or work zones, and could reflect the impact of variables that were removed from LQMMs because of insignificance.

The study recommends that VDOT use the GRF model for predicting travel time reliability on interstate highways. In addition, further research is recommended to extend the GRF models to meet the requirements of MAP-21 federal target setting.

17 Key Words:
Travel Time Reliability, Target Setting, LOTTR, Quantile Regression, Generalized Random Forests

18. Distribution Statement:
No restrictions. This document is available to the public through NTIS, Springfield, VA 22161.

19. Security Classif. (of this report):	20. Security Classif. (of this page):	21. No. of Pages:	22. Price:
Unclassified	Unclassified	71	

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

FINAL REPORT

METHODS TO ANALYZE AND PREDICT INTERSTATE TRAVEL TIME RELIABILITY

Xiaoxiao Zhang, Ph.D.
Graduate Research Assistant
Department of Engineering Systems and Environment
University of Virginia

Mo Zhao, Ph.D.
Research Scientist
Virginia Transportation Research Council

Justice Appiah, Ph.D., P.E.
Senior Research Scientist
Virginia Transportation Research Council

Michael D. Fontaine, Ph.D., P.E.
Associate Director
Virginia Transportation Research Council

In Cooperation with the U.S. Department of Transportation
Federal Highway Administration

Virginia Transportation Research Council
(A partnership of the Virginia Department of Transportation
And the University of Virginia since 1948)

Charlottesville, Virginia

July 2021
VTRC 22-R2

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Virginia Department of Transportation, the Commonwealth Transportation Board, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation. Any inclusion of manufacturer names, trade names, or trademarks is for identification purposes only and is not to be considered an endorsement.

Copyright 2021 by the Commonwealth of Virginia.
All rights reserved.

ABSTRACT

The Moving Ahead for Progress in the 21st Century Act (MAP-21) defined requirements for system reliability performance measures. Under MAP-21, state departments of transportation are responsible for reporting travel time reliability and for setting targets and showing progress toward those targets. In order to know how to improve travel time reliability and what to expect from investments in transportation infrastructure, these agencies need a better understanding of the factors that affect travel time reliability and methods to predict future travel time reliability. The purpose of this study was to quantify the factors influencing travel time reliability and investigate how to account for these factors in setting reliability targets and communicating progress.

To achieve these objectives, this study developed models to estimate quantiles (the 50th, 80th, and 90th) of travel time distributions to quantify the effects of travel time reliability impact factors and predict select reliability measures. First, linear quantile mixed models **(**LQMMs) were built using both data maintained by the Virginia Department of Transportation (VDOT) and crowdsourced event data. Model results using the crowdsourced data were unstable and difficult to interpret because of data quality issues such as unbalanced spatial density, duplicate reporting, and inconsistent event classification because of individual observer bias. The results using VDOT-maintained data were more reliable and interpretable. Those models showed that frequencies of non-recurrent events, such as incidents and weather, were correlated with higher travel time percentiles. The LQMM was compared with the trend line approach, a common prediction method used in practice, and the results showed that LQMMs significantly improved the accuracy of predictions over the trend line approach based on mean absolute percent error. Generalized random forest (GRF) models were also tested as an alternative prediction method. GRF models improved the prediction accuracy over LQMMs for the 50th and 80th percentiles, but the accuracy was slightly worse for the 90th percentile. In addition, the GRF models could also reflect the impact of variables that were removed from LQMMs because of insignificance, such as the presence of safety service patrols.

Before-after studies were conducted to illustrate the application of LQMMs and GRF models. LQMMs captured the changes in the 90th percentile travel times better, and GRF models captured the changes of level of travel time reliability better in most cases. GRF models were more sensitive to the reliability changes caused by non-recurrent events, such as incidents or work zones, and could reflect the impact of variables that were removed from LQMMs because of insignificance.

The study recommends that VDOT use the GRF model for predicting travel time reliability on interstate highways. In addition, further research is recommended to extend the GRF models to meet the requirements of MAP-21 federal target setting.

FINAL REPORT

METHODS TO ANALYZE AND PREDICT INTERSTATE TRAVEL TIME RELIABILITY

Xiaoxiao Zhang, Ph.D.
Graduate Research Assistant
Department of Engineering Systems and Environment
University of Virginia

Mo Zhao, Ph.D.
Research Scientist
Virginia Transportation Research Council

Justice Appiah, Ph.D., P.E.
Senior Research Scientist
Virginia Transportation Research Council

Michael D. Fontaine, Ph.D., P.E.
Associate Director
Virginia Transportation Research Council

INTRODUCTION

Increasing congestion levels have become a major concern for roadway users. In the past, many measures, such as delay, volume-to-capacity ratio (v/c ratio), and level of service, were developed to reflect the magnitude of congestion (Aftabuzzaman, 2007). The Federal Highway Administration (FHWA) has established that meaningful congestion performance measures must be based on travel times because they are easily understood by practitioners and the public and are applicable from both the user and facility perspectives of performance (Cambridge Systematics, 2005). Although average travel times are widely used in traffic operations and planning, they represent only typical situations and do not show the entire breadth of user experiences. Unexpected interruptions, including those caused by incidents, inclement weather, special events, and work zones, cause deviations from average travel times. This variability is also extremely important to roadway users and transportation agencies. The concept of travel time reliability was developed to quantify the variability in travel times and has become a critical factor in the evaluation of transportation networks. Travel time reliability has been defined by FHWA (2006) as "the consistency or dependability in travel times, as measured from day-to-day and/or across different times of the day." This reliability definition was used for this study.

Transportation professionals widely acknowledge that although capacity expansion is an obvious solution for congestion, it is extremely difficult and expensive to implement. As a result, improvement strategies incorporating transportation systems management and operations

(TSM&O) are often preferred alternatives since they may optimize traffic flow through better management of existing capacity. These TSM&O approaches often impact extreme values of the travel time distribution differently than average travel time. Travel time reliability measures are capable of capturing how much unexpected events influence traffic. These measures will help transportation professionals (1) identify the needs and related strategies that would address those needs (e.g., incident management); (2) quantify the benefits of TSM&O strategies that may be obscured by using average congestion measures; and (3) promote better project prioritization and decision making.

Travel time reliability is also critical to various roadway users, including drivers, transit riders, and freight shippers, influencing decisions about where, when, and how travel is made (Abdel-Aty et al., 1995; Cambridge Systematics, 2013; Elefteriadou et al., 2008; Van Lint and Van Zuylen, 2005). Since unexpected events can create unreliable travel times, travelers and shipping agencies need to plan extra time to arrive on time. However, early arrivals are also a waste of time that could have been spent on other activities. From an economic perspective, this extra time has values beyond the average travel time used in traditional economic analyses (Cambridge Systematics, 2013).

The Moving Ahead for Progress in the 21st Century Act (MAP-21) features the establishment of a performance- and outcome-based program that requires states to invest resources to achieve targets that facilitate the progress toward national goals. One of the six national goals is "System reliability—To improve the efficiency of the surface transportation systems." To assess progress toward national goals, FHWA issued a set of rulemakings to establish performance measures, specify related data requirements and calculation methodologies, and set target requirements. The performance measures adopted to address the system reliability goals for the National Highway System (NHS) were as follows:

- *Interstate travel time reliability measure*: percent of person-miles traveled on the interstate that is reliable.

- *Non-interstate travel time reliability measure*: percent of person-miles traveled on the non-interstate NHS that is reliable.

- *Freight reliability measure*: truck travel time reliability index.

To comply with the rulemaking, state departments of transportation (DOTs) and metropolitan planning organizations need to establish 2- or 4-year targets for the travel time reliability performance measures and report progress toward those targets. It is essential to have an effective analytical framework for travel time reliability as a prerequisite to set proper targets. Currently, targets are set based on either historic trend lines or change rates of other congestion measures. Prediction models that explicitly account for changes in underlying impact factors are not available (The National Capital Region Transportation Planning Board [TPB], 2018). The trend line method cannot predict reliability changes in the opposite direction of the fitted lines, and it also is not sensitive to changes in reliability impact factors. Adopting the change rates of other congestion measures could be biased since there is no guarantee that reliability would change in the same way as other measures. As a result, there is a need to develop improved target setting methods that can account for the variety of factors that could influence future travel time reliability.

PURPOSE AND SCOPE

The purpose of this study was to develop methods to predict travel time reliability that account for important factors influencing recurring and nonrecurring congestion for interstate highways. Such methods could be used to improve the target setting process required by MAP-21 and to assess potential impacts of future projects.

The objectives were as follows:

- Develop practical models that could be used (1) to predict travel time reliability at an acceptable accuracy level and (2) to analyze the importance and quantify the influence of impact factors.
- Compare and contrast the value of VDOT-maintained and private crowdsourced event data sources and evaluate how various factors and data sources affect travel time reliability modeling and prediction.
- Demonstrate model application through before-after case studies.

Travel time reliability and its influencing factors were studied for all interstate highways in Virginia using data from 2017-2019. The reliability metrics used were level of travel time reliability (LOTTR) and the 90th percentiles of travel time distributions. LOTTR is required in 23 CFR Part 490 (*Federal Register*, 2017), and the 90th percentile is frequently used by practitioners and researchers to reflect extreme situations.

METHODS

The following tasks were conducted to meet the study objectives:

1. Review the literature.
2. Collect and prepare data.
3. Develop models for travel time reliability analysis and prediction.
4. Perform before-after studies to demonstrate model application.

Task 1. Review the Literature

A review of the literature was conducted to summarize the latest developments on travel time reliability measures, metrics, and influencing factors. Relevant databases such as the Transportation Research International Documentation (TRID) and Transport databases were used to identify sources for the literature review. This task produced a summary of travel time reliability metrics and measures and documented the findings on reliability modeling approaches, influencing factors, and operational countermeasures.

Task 2. Collect and Prepare Data

This study considered various factors that could impact reliability, including geometric features, weather, incidents, work zones, traffic demand, parallel managed lanes, and safety service patrols (SSPs). Factors were selected for consideration based on the results of the literature review and discussions with the study's technical review panel (TRP).

Data Types and Sources

The following sections introduce the data sources, the data formats of each factor examined, and the preparation process.

Travel Time

Probe travel time data from INRIX were used in this study. VDOT uses probe vehicle data frequently to measure highway system performance and to provide traveler information. INRIX uses the traffic message channel (TMC) standard to define roadway segments spatially. The TMC network used for this analysis is shown in Figure 1. Two segment types are defined: internal and external. Internal segments represent a stretch of road within an interchange (e.g., between an exit ramp and an entrance ramp) and are defined as interchange segments in this study. External segments represent a stretch of road between interchanges, referred to as freeway segments in this study.

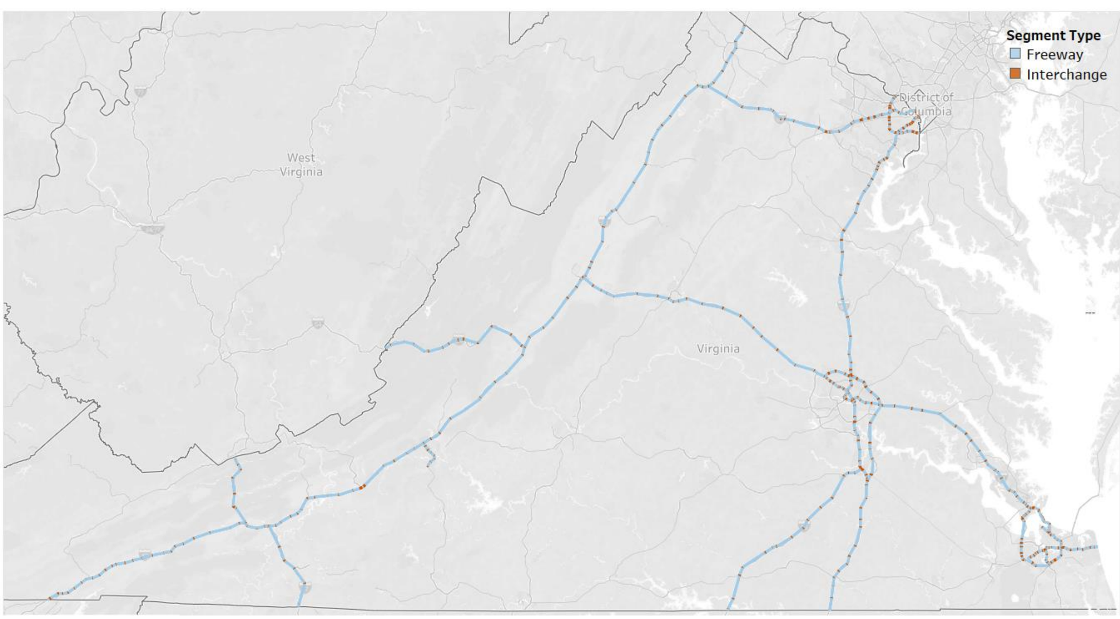


Figure 1. Traffic Message Channel Probe Vehicle Data Coverage in the Study Area

Figure 2 shows an example of the difference between internal and external TMCs. In the figure, internal segments are prefixed with the letters "P" or "N" for the positive and negative directions, respectively, and external segments are prefixed with a "+" or a "-" sign.

Travel time data were collected from January 2017 to December 2019 for all interstate highways. For this study, only travel times during the AM peak (6 to 10) and PM peak (4 to 8) periods on weekdays were considered since those are the time periods with the greatest travel time variability. There were 40,512 hours (about 0.34%) with missing data points across all TMCs. Since these hours represented a small portion of overall data, they were not used in the analysis. TMC segments less than 0.1 miles long (about 8.5% of total TMCs) were also removed since those segments had noisier data. This screening produced a total of 1,853 TMC segments across both directions that were used in the analysis.

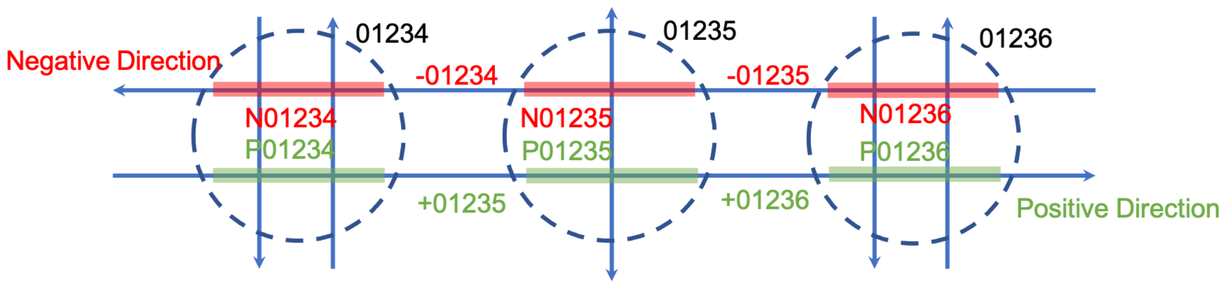


Figure 2. Traffic Message Channel Segments With Internal and External Segments. Adapted from RITIS (2021).

Geometric Features

The geometric factors considered in this study were as follows:

- *TMC segment length.* The metadata associated with each TMC include TMC length information.
- *Number of through lanes.*
- *Urban/rural designation.* The number of through lanes and area types were obtained from the VDOT roadway inventory.
- *Presence of parallel high occupancy vehicle (HOV), high occupancy toll (HOT), or express lanes.* VDOT's Roadway Network System was used for collecting this information.

Weather

The Local Climatological Data (LCD) from the National Centers for Environmental Information were used for hourly weather conditions. The LCD identify more than 20 types of weather conditions, including rain, snow, drizzle, hail, fog, storm, and so on. Although they all could have some impact on travel time reliability, this study selected and categorized weather factors using two variables:

1. *Accumulated hourly rain (liquid) precipitation volume:* the amount of drizzle, rain, and thunderstorm precipitation in inches to hundredths.

2. *Accumulated hourly snow (frozen) precipitation volume:* the amount of snow, snow grains, snow pellets, ice pellets, and hail precipitation in inches to hundredths.

Incidents and Crashes

The impacts of incidents and crashes were counted using the event frequencies. These data were derived from two sources: VDOT-maintained data, and crowdsourced data. VDOT-maintained data are verified data generated through official mechanisms. Two VDOT-maintained incident and crash data sources were used:

1. *Crash data.* VDOT has a crash data query tool, named the Crash Analysis Tool (VDOT, 2021a), which provides crash information such as crash severity, location, and time. Crashes with different severity types (e.g., severe injury, visible injury, non-visible injury, and property damage only [PDO]) are counted separately to create four independent variables. These data are derived from police crash reports and are verified.

2. *Non-crash incidents.* Non-crash incident events were collected using VaTraffic. VaTraffic is a VDOT operations and incident management database that provides information on various activities such as traffic incidents and roadway maintenance. These data include information on vehicle breakdowns and hazards. For VaTraffic, breakdown events typically involve disabled vehicles on the shoulder, and hazards represent non-crash disruptive events such as vehicle fires.

Crowdsourced crash and incident data generated from the traveler information application Waze were also evaluated. Waze provides information on travel times and unexpected events to users and allows users to report crashes, incidents, work zones, and other unusual events. These data may include minor incidents not recorded in VDOT databases, and they are not quality controlled or verified by VDOT. Further, it is possible that they could contain erroneous reports or duplicate reports of a single crash with different locations and timestamps. Waze provides crash information similar to that provided by VaTraffic except that there is no severity information. As a result, all types of crashes are counted in one independent variable. Unlike VaTraffic, which includes only police-reported crashes, the Waze dataset may include minor PDO crashes not reported to the police. Non-crash incident event definitions in Waze may also differ from those used in VaTraffic.

Figure 3 shows the event report interface from Waze. It can be seen that there are three categories under hazard events: On road, Shoulder, and Weather. Under each category, more specific event types could be chosen. The descriptions of hazard events in VaTraffic have mainly involved objects (usually vehicles) that are on fire. The hazard events in Waze data include more situations than in the VaTraffic data. Both data sources provide event attributes such as event type, location, and timestamps.

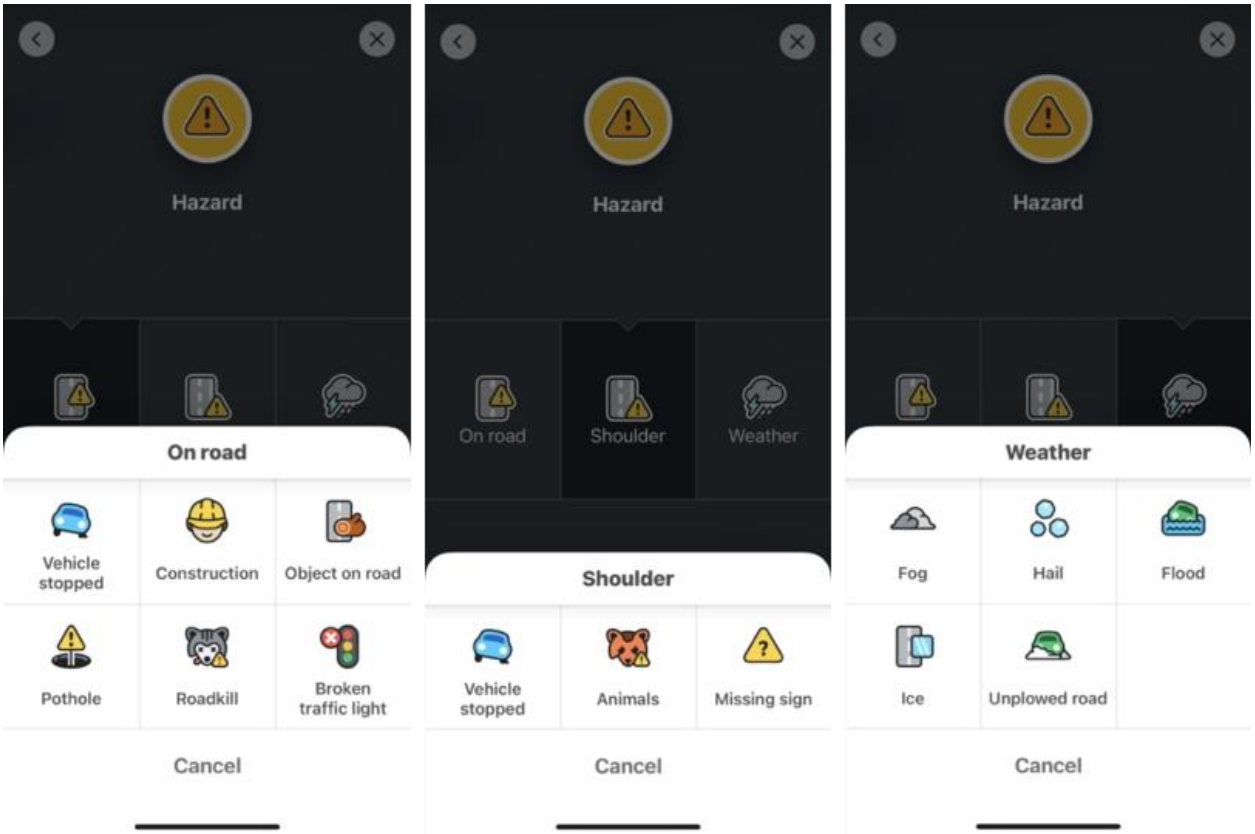


Figure 3. Waze Report Interface

Work Zones

Work zone event data were also collected from VaTraffic. Event attributes available are work zone type, location, start time, and duration. In addition, each event has a text description of work details such as what will be done, route information, scheduled times, and possible impacts to traffic (e.g., lane/shoulder closures or potential delays). Since lane and shoulder closures are expected to have more detectable effects on reliability, the frequencies of such events are used to reflect work zone impacts. Although Waze also provides work zone event data, these data are presented as point events. As a result, they could not be mapped spatially to the network, so Waze information was not used.

Traffic Demand

Previous research (Cambridge Systematics, 2013) has shown that the v/c ratio plays a vital role in determining travel time reliability, and thus it was considered in this study. Capacity was calculated in accordance with the *Highway Capacity Manual, Sixth Edition* (HCM-6) (Transportation Research Board, 2016) using Equation 1:

$$\begin{gathered} Base\ capacity = 2{,}200 + 10 \times (FFS - 50) \\ Capacity = Base\ capacity \times Lanes \times f_{hv} \times PHF \times f_p \times f_g \end{gathered} \qquad \text{[Eq. 1]}$$

where

FFS = free-flow speed

f_{hv} = the heavy vehicle adjustment factor depending on facility type, vehicle mix, and grade

PHF = the peak hour factor, calculated by the ratio of the peak 15-minute flow rate to the average hourly flow rate

f_p and f_g = adjustment factors for driver population and grades provided by VDOT's Traffic Engineering Division.

The capacity obtained is in vehicles per hour. Accordingly, the traffic volume was also aggregated at an hourly interval. VDOT's Traffic Engineering Division provided annual average daily traffic (AADT) estimated from continuous and short-term count stations. These values needed to be broken down to an hourly basis by different seasons, which was achieved by applying an hourly volume profile to the AADT. This volume profile was built using volume data collected at continuous count stations at 5- or 15-minute intervals and then assigned to nearby segments. The locations of the stations are shown in Figure 4. Missing data were filled in with the value from the previous time interval when they were less than 1 hour. When more than 1 hour of missing data needed to be filled in, the average value for the same hour for the same day of the week was used. To reflect traffic volume better, the percent of heavy vehicles (buses, single unit trucks, and tractor-trailer combinations) was also considered an impact factor. These data could be found from the AADT files.

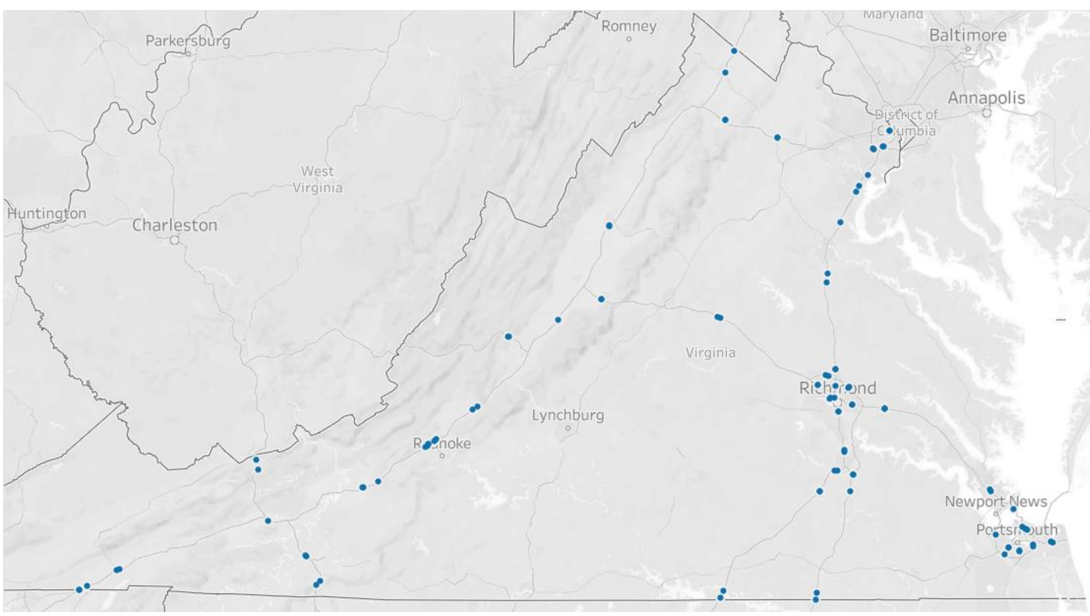


Figure 4. Continuous Count Station Location Map

Safety Service Patrol (SSP) Presence

SSPs are incident management trucks with personnel who can assist disabled vehicles, provide traffic control for crashes and incidents, and otherwise seek to reduce the impact of nonrecurring events on safety and mobility. This program provides benefits such as faster detection of incidents, scene management through temporary traffic control, and clearance of obstructions and debris on roadways. SSPs help minimize incident durations and reduce secondary crashes. The SSP schedules were obtained through VDOT. An SSP indicator variable was created to denote the presence of SSP service during the AM or PM peak period. As long as SSP service was available during any hour within the AM or PM peak period, the SSP variable was assigned a value of 1.

Data Combination and Conflation

After the data were assembled, they were combined following the data fusion process shown in Figure 5 and described here.

1. *Travel times were provided using the TMC network, but other VDOT data were provided using different spatial reference systems.* For example, traffic volume was recorded using VDOT's Traffic Monitoring System (TMS) segments. All data collected from or calculated based on VDOT datasets, including geometric features, v/c ratio, percentage of heavy vehicles, and SSP presence, were mapped to TMS segments. These two versions of network segmentation (TMC and TMS) were matched such that all data attributes could be linked to the TMCs. The matching process was adapted from the conflation approach developed by Schrank et al. (2018) and conducted through the spatial join function in ArcGIS.

2. *The LCD weather data were collected from weather stations.* Each station was matched with nearby TMC segments using the method proposed by Lan et al. (2019). The use of these weather station data to represent roadway weather conditions was based on the assumption that the weather conditions were the same on the roadways near the weather stations. This assumption could be inaccurate and potentially bias modeling results since there are only approximately 20 weather stations to cover more than 1,800 segments statewide. The farthest distance between TMC segments and their weather stations could be more than 40 miles. However, these were the best available data.

3. *After the TMS and TMC segments were matched in Step 1, each TMC was provided with start and end mileposts.* The collected incident and work zone data have the event locations presented as mileposts, which could be used to match the impacted TMC segments. Then, the timestamps of the events were used to count the frequency of the events during the AM and PM peak periods.

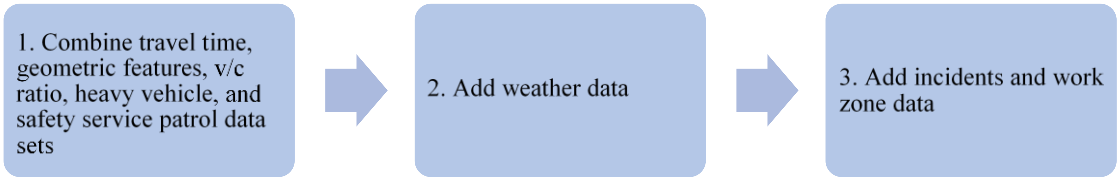


Figure 5. Data Conflation Overview

Task 3. Develop Models for Travel Time Reliability Analysis and Prediction

This study used two advanced statistical modeling techniques for travel time reliability analysis and prediction: linear quantile mixed models, and random forest models.

Linear Quantile Mixed Models (LQMMs)

After frequently used travel time reliability measures were reviewed, it was found that most of them rely on travel time distributions, especially upper tails such as the 80th, 90th, or 95th percentile. In studies focused on estimating travel time distributions, the authors often had difficulty choosing among different distributions or deriving a closed-form expression for the probability density functions. Quantile regression provides a convenient way to characterize any type of distribution with feasible estimation techniques. Since this method estimates quantiles rather than specific travel time reliability measures, it is flexible enough to obtain almost any measure from the estimated quantiles.

Target setting requires that travel time reliability is analyzed at the TMC segment level. As a result, the input data are prepared as a repeated measures dataset from different times of the day and days of the week on each TMC segment. Two correlations exist in this data structure: one from the observations of travel time within each TMC, and the other from the fact that the data represent all interstate highways. LQMMs are suitable for estimating quantile regression on such a data structure. They include TMC-level (subject-specific) characteristics through both random effects and fixed effects to reflect the relationship between the travel time reliability and its impact factors for interstate highways. Independent variables representing various factors that affect reliability are considered fixed effects, and their coefficients are estimated as part of the modeling.

To facilitate practical use, LQMMs with only location-shift random effects were considered in this study. To account for location differences and shape variations in the travel time distributions of different TMC segments, they were first clustered into homogenous groups, as discussed in the next section. Then, LQMM was applied to each cluster, where all independent variables were included as fixed effects and intercepts were included as random effects to distinguish individual TMCs. The clustering process is described here.

Travel Time Distribution Clustering

Clustering travel time distributions was done to identify similarity among travel time distributions and group homogeneous TMCs into clusters, allowing TMCs within the same cluster to be represented by one distribution shape. Weijermars and Van Berkum (2005) recommended some pre-classification to improve clustering results. For this study, the facility type (freeway segments or interchange) was used as a pre-classification factor, with the assumption that frequent traffic interactions because of merging and diverging at interchanges play an important role in distinguishing distributions. Thus, data were divided into freeway and interchange groups before clustering.

The main question to answer before applying any clustering algorithm is how to measure the similarity between cumulative distribution functions (CDFs) without knowing the exact functional expressions, which leads to the consideration of non-parametric methods of comparing distributions. Two non-parametric approaches were adopted in this study: the Kolmogorov-Smirnov (KS) test, and the Anderson-Darling (AD) test. They both quantify the differences between CDFs using defined statistical values. For any two TMCs with CDFs $F_n(x)$ and $F_m(x)$, the two-sample KS statistic is defined as $D_{n,m}$, which is the maximum distance between two distributions (Eq. 2):

$$D_{n,m} = \sqrt{\frac{nm}{n+m}} sup_{t\in\mathbb{R}} |\widehat{F_n(x)} - \widehat{F_m(x)}| \quad \text{[Eq. 2]}$$

where

m and *n* = the size of the two samples, respectively.

As mentioned, as an alternative option, the AD test was also adopted for this study because it is particularly sensitive to the differences at the tails of distributions by assigning larger weights (Thas, 2010). The two-sample AD statistic is defined in Equation 3:

$$A_{nm}^2 = \frac{nm}{n+m} \int_{-\infty}^{+\infty} [F_n(x) - F_m(x)]^2 \psi(H_{n+m}(x)) dH_{n+m}(x) \quad \text{[Eq. 3]}$$

where

$\psi(u) = \frac{1}{u(u-1)}, 0 \le u \le 1$ = the weight function
$H_{n+m}(x) = \frac{n}{n+m} F_n(x) + \frac{m}{n+m} F_m(x)$ = the empirical function of joint samples.

The KS and AD statistics were calculated between each pair of TMC segments to form a dissimilarity matrix. Each row of this matrix presented the dissimilarity between one TMC and the rest. An agglomerative hierarchical clustering method was then applied to rearrange rows in the KS and AD matrices so that similar rows could be paired together into clusters. The process starts with singleton clusters at the bottom level and continues to merge two clusters at a time to build a bottom-up hierarchy. There are several widely used methods to measure cluster distances: single linkage (McQuitty, 1964), complete linkage (King, 1967), group average (Bailey, 1994), and Ward's criterion (Sokal, 1966). After a thorough consideration of the advantages and disadvantages of those methods, the complete linkage and Ward's criterion methods were selected to use for this study.

Another technique for clustering travel time distributions is model-based functional data clustering. With this technique, the CDFs of travel times can be treated as functional data so that they can be approximated using dimension reduction techniques and represented by a set of finite parameters. Model-based clustering algorithms are then applied based on these parameters (Bouveyron et al., 2015).

Both the non-parametric and model-based clustering methods are considered data-driven. There are also non–data-driven ways to cluster TMC segments, such as by geometric features. In order to evaluate the benefit and necessity of data-driven approaches, clustering by the geometric features of TMC segments was also conducted for comparison. Figure 6 summarizes the methods used.

The goal of clustering travel time distributions was to find homogeneous clusters that could be represented by unique distributions to improve models built in the next task for reliability analysis. For this study, the performance of optimal clusters resulting from the different methods was evaluated by how close the 50th, 80th, and 90th percentile travel times within each cluster were using the Silhouette width, a weighted average of each observation's Silhouette value (Rousseeuw, 1987). The Silhouette value measures the degree of confidence in a particular clustering assignment on a scale of -1 to +1, with well-clustered observations having values near +1 and poorly clustered observations having values near -1. The three percentiles were selected as target variables of the reliability models because (1) the 50th and 80th percentiles are the two components of the federally mandated LOTTR measure (FHWA, 2017) and (2) the 90th percentile is preferred over the 95th percentile to reflect the benefit of operational improvements (Cambridge Systematics, 2013).

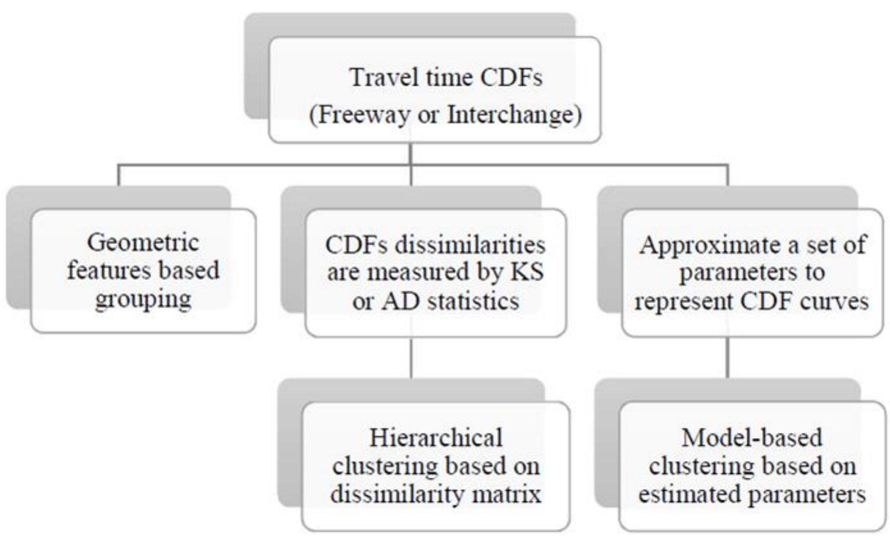


Figure 6. Clustering Method Summary. CDF = cumulative distribution function; KS = Kolmogorov-Smirnov; AD = Anderson-Darling.

Development of LQMMs

Classic linear regression estimates conditional means by minimizing the residual sum of squares. Similarly, quantile regression (Koenker and Bassett, 1978) estimates conditional quantiles by minimizing a sum of asymmetrically weighted absolute residuals.

Suppose a dataset (x_i^T, y_i), where i = 1, …; *N*, x_i^T are row vectors of a known design matrix *X* with *p* elements; and y_i are observations of the dependent variable. The linear conditional quantile functions (τth quantile) are defined as shown in Equation 4:

$$G_{y_i}(x_i) = x_i^T\beta, \qquad i = 1, \dots, N, \qquad \text{[Eq. 4]}$$

where

G_{y_i} = the inverse of a continuous distribution function F_{y_i}

β, specified to τ, = a column vector of length *p* estimated through solving the optimization problem shown in Equation 5 (Koenker and Bassett, 1978):

$$\hat{\beta}_\tau = \underset{\beta \in \Re^p}{\operatorname{argmin}} \left\{ \sum \rho_\tau (y_i - \mathbf{x}_i^T \beta) \right\} \qquad \text{[Eq. 5]}$$

where

$\rho_\tau(r) = \{(1-\tau)r,\ r < 0; \tau r,\ r \geq 0\}$ = the asymmetrically weighted L_1 loss function.

The linear mixed quantile functions with random intercepts are defined as shown in Equation 6:

$$G_{y_{ij|u_i}}(x_{ij}, u_i) = x_{ij}^T\beta + u_i, \qquad i = 1, \dots, N, \qquad j = 1, \dots, n_i \qquad \text{[Eq. 6]}$$

where

i = the i-th subject
j = the size of observations of a given subject.

Comparing Equation 6 to Equation 4, Equation 6 includes the location-shift random effect u_i. The objective function to estimate β is then defined as indicated in Equation 7:

$$\hat{\beta}_\tau = \underset{\beta \in \Re^p}{\operatorname{argmin}} \left\{ \sum \rho_\tau (y_{ij} - \mathbf{x}_{ij}^T \beta - u_i) \right\} \qquad \text{[Eq. 7]}$$

Geraci and Bottai (2007) developed a likelihood-based approach using an asymmetric Laplace distribution to estimate $\hat{\beta}_\tau$ by minimizing the objective function. This study used the R package "lqmm" (Geraci, 2014) to estimate the parameters of interest (η) and the random effects (u) using a method based on Gaussian quadrature developed by Geraci and Bottai (2014).

Modeling Strategy and Evaluation

The incident data were collected from two data sources: VDOT-maintained (VaTraffic) and crowdsourced (Waze) datasets. Although they convey similar information, VDOT-maintained data are expected to be more accurate but may be more limited in their coverage area since they depend on police crash reports and incident detection by official sources. On the other hand, crowdsourcing data do not go through any cleaning and validation. These data may have better coverage, but their accuracy is unknown. It was of great interest to compare the reliability models built with the traditional data source that has been used by most previous studies and those built with the emerging crowdsourced data source. As illustrated in Figure 7, depending on the incident data source, the input dataset was termed the "VaTraffic" or the "Waze" dataset.

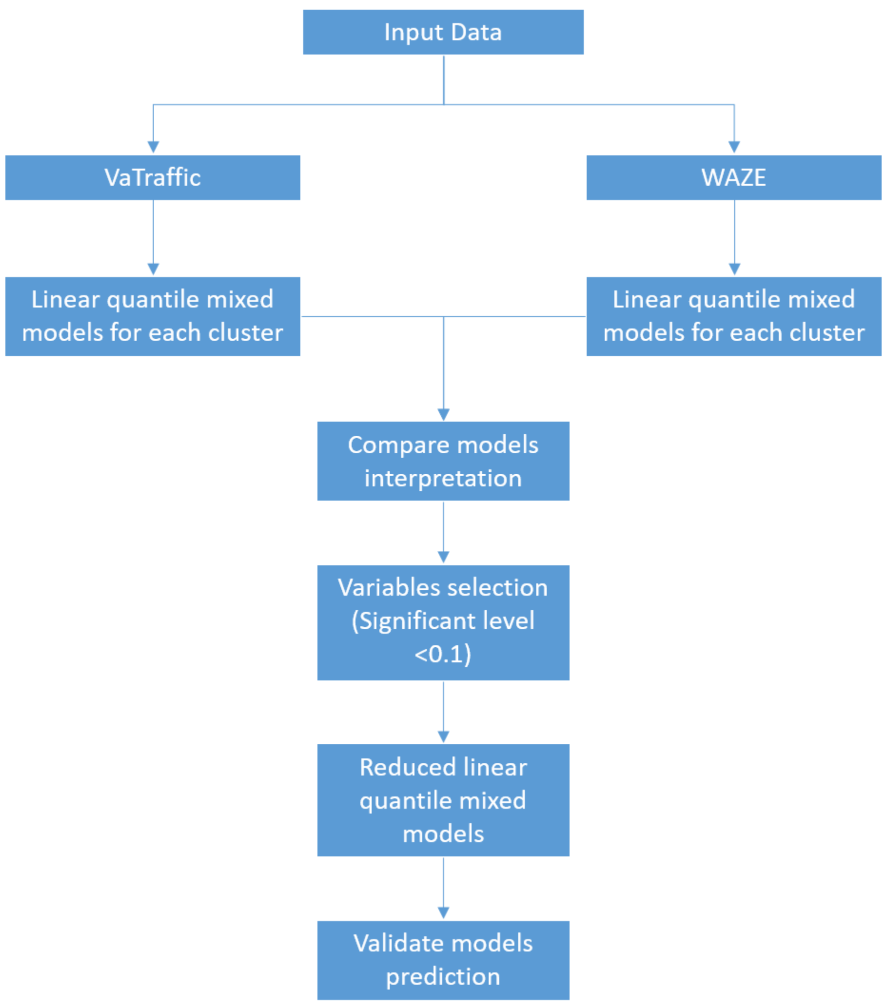


Figure 7. Linear Quantile Mixed Model Development

It is worth mentioning that the spatial definition of TMC segments changes periodically. Such changes could result in the same TMC code referring to different segment endpoints or even completely different locations at different times. The data used for this study were from 2017 to 2019, during which the TMC update in December 2018 caused about 30% of TMC segments to have more than a 0.5-mile length or location difference. To simplify data conflation, the training data and testing data were divided based on these two versions of the TMC map. Data from January 1, 2017, to December 3, 2018, were used as training data, and data from December 4, 2018, to December 31, 2019, were used as testing data.

Freeway and interchange segments were clustered into homogeneous groups according to the shapes of their travel time distributions in the previous step. LQMMs were then built for each cluster using VaTraffic and Waze training data, referred to as the VaTraffic and Waze models. The estimated coefficients from these models were used to investigate and quantify the contribution of impact factors on travel time reliability using different percentiles of travel times. Comparing the results from the VaTraffic and Waze models provided insights into the impacts of different input data sources.

To validate the prediction performance, variables with a significance level of more than 0.1 were eliminated from the input dataset to avoid potential overfitting issues. LQMMs were estimated again using the subset of significant variables ($p < 0.1$). The training error and testing error were both quantified using the following four measures, where training error was estimated using 5-fold cross-validation:

1. mean absolute error (MAE)
2. mean absolute percentage error (MAPE)
3. mean square error (MSE)
4. bias.

Random Forest Models

When the number of independent variables is relatively large, application of linear regression risks overfitting the data. With the popularity of machine learning methods, previous studies (Breiman, 2001; Meinshausen and Ridgeway, 2006) have demonstrated their capability to outperform linear regression models. Random forests, first introduced by Breiman (2001), is one of the most commonly used machine learning techniques with a reputation for good prediction accuracy and stability. It is nonparametric, so the linear assumption is relaxed. The interactive impacts among variables are considered during the model construction process without increasing the risk of overfitting. Meinshausen and Ridgeway (2006) compared the performance of linear quantile regression models with and without interaction terms and random forest models using various popular datasets. They found that random forest models showed better prediction accuracy than traditional quantile regression, especially for higher quantiles. These advantages fit the needs of this study since the travel time reliability measures often involved quantiles toward the tail portion of the travel time distribution. As a result, random forest models were built using the same set of input data as LQMMs. The prediction accuracy of LQMMs and random forests were then compared to assess their suitability for travel time reliability prediction.

Algorithm

The essential idea of random forests is to grow an ensemble of trees that consists of B single trees, $\{T_b\}_1^B$. For each tree, the following steps are taken:

1. Draw a bootstrap sample from the training data.

2. Using these bootstrapped data, conduct the following steps at each node until a leaf is created:

   a. Select a random subset of independent variables, denoted $mtry$, as the split variable candidates at each node. Often, $mtry$ is set as one-third of the total number of independent variables for regression models, and it will stay constant during the tree growing process. Also, independent variables can be selected multiple times at different nodes. It is worth mentioning that the choice of $mtry$ is the main parameter that needs tuning for random forests, even though results are typically nearly optimal over a wide range of this parameter (Breiman, 2001). The value of $mtry$ can be optimized using out-of-bag (OOB) errors.

   b. Determine the splitting variable and the threshold. The sum squared residuals (SSRs) using different values of each variable are calculated, and the value with the smallest SSR becomes the threshold. Then, the variable with the smallest SSR at its threshold becomes the split variable at that node. This partition is known as the Classification and Regression Tree (CART) rules.

   c. Repeat Steps 2a and 2b for continuing to grow the tree until some terminal criteria are met, which can be a minimum number of observations in a leaf or a predefined number of nodes.

3. Repeat Steps 1 and 2 for B times. In practice, at least 200 trees are recommended. More details of random forests are provided in Breiman (2001) and Meinshausen and Ridgeway (2006).

Athey et al. (2019) extended the algorithm for quantile estimation based on the work of Breiman (2001) and Meinshausen and Ridgeway (2006). Instead of using SSRs to measure the quality of a split, the authors used moment conditions in the form of Equation 8 to identify the best split that maximizes the heterogeneity of quantiles of interest.

$$\psi_\theta(Y_i) = q\mathbf{1}(\{Y_i > \theta\}) - (1-q)\mathbf{1}(\{Y_i \leq \theta\}) \qquad \text{[Eq. 8]}$$

where

q = the estimated quantile
θ = the estimation at X_i
Y_i = the observation at X_i.

Tuning Parameters

The parameter $mtry$, which refers to the number of randomly selected variable candidates used for node splitting, has a significant impact on prediction performance and variable importance ranking. Often, forests built with a lower $mtry$ value provide a better opportunity for exploiting variables with moderate effects on the targeted quantiles. Because the chance of such variables being selected simultaneously with variables having strong effects is also lower, they have a higher possibility to be used as the variable for node splitting. On the contrary, if the value of $mtry$ is set high, the variables with moderate effects will most likely be masked by the variables with strong effects. The disadvantage of selecting a lower $mtry$ value is that some trees are constructed by variables that actually do not have a significant impact on the targeted quantiles, which reduces the prediction accuracy after the results of all trees are averaged. In order to balance these two aspects, models using different values of $mtry$ were compared to obtain the optimal results. Most previous studies recommended using one-third of the total number of independent variables as the optimal $mtry$ for prediction accuracy. Bernard et al. (2009) also pointed out that the optimal $mtry$ value is highly related to the number of variables that significantly impact the response variable. In order to capture the performance trend along with the increased value of $mtry$, it was decided to use 5, 8, 11, 14, and 17 (the total number of variables) as the $mtry$ options in this study.

Although there are other tuning parameters, such as the minimum number of observations in a leaf node or number of total trees, they contribute a minimal amount to prediction changes (Probst et al., 2019). These parameters were set to commonly used values. The minimum node size was set to 10, and the number of trees was set to 2,000 for all models.

Model Evaluation and Comparison

Because the purpose of the splitting rule of random forests is to maximize the heterogeneity of the quantiles of interest, it should automatically separate samples with different travel time distributions. As a result, it is expected that clustering TMC segments into homogeneous groups with similar shapes of distributions might not be necessary when random forests are used. To test this hypothesis, random forest models were built for both the freeway and interchange TMC datasets and their clustered groups. The performance of these models was compared using the method described here to determine if the preferred models needed clustering as a prerequisite. Then, the performance of the preferred models using random forests was compared to the performance of LQMMs using the same method.

The method to compare performance was as follows. For LQMMs, cross-validation was used to evaluate and compare model performance. For random forests, no matter the bagging version or fraction of the full input data sampled, there are always data that are not selected to construct trees, called OOB observations. These data could be used the same way as testing data in the cross-validation method and establish the OOB predictions during tree growing. This is considered a preferred validation method to compare modeling results and goodness-of-fit over cross-validation since the process takes less time. The measures adopted to quantify model performance were the same as the ones used by LQMMs, i.e., MAE, MSE, MAPE, and bias.

Task 4. Perform Before-After Studies to Demonstrate Model Application

Overview of Selected Projects

In this task, before-after studies were conducted to evaluate the impacts of improvement projects on travel time reliability. The impacts were evaluated using the models developed in Task 3 to illustrate their application and effectiveness. The scope of this task was limited to several project types selected in conjunction with the TRP. Specifically, the three projects selected were as follows:

1. *I-81 Safety Service Patrols (before-after operational improvement).* As part of the safety improvement project along I-81, SSP coverage was extended temporally and spatially to cover more segments and for longer periods. There are 23 TMC segments (27 miles), as shown in Figure 8, with extended service areas or service periods. This project was selected by VDOT since the benefits of SSPs are of high interest. The changes started on July 1, 2019.

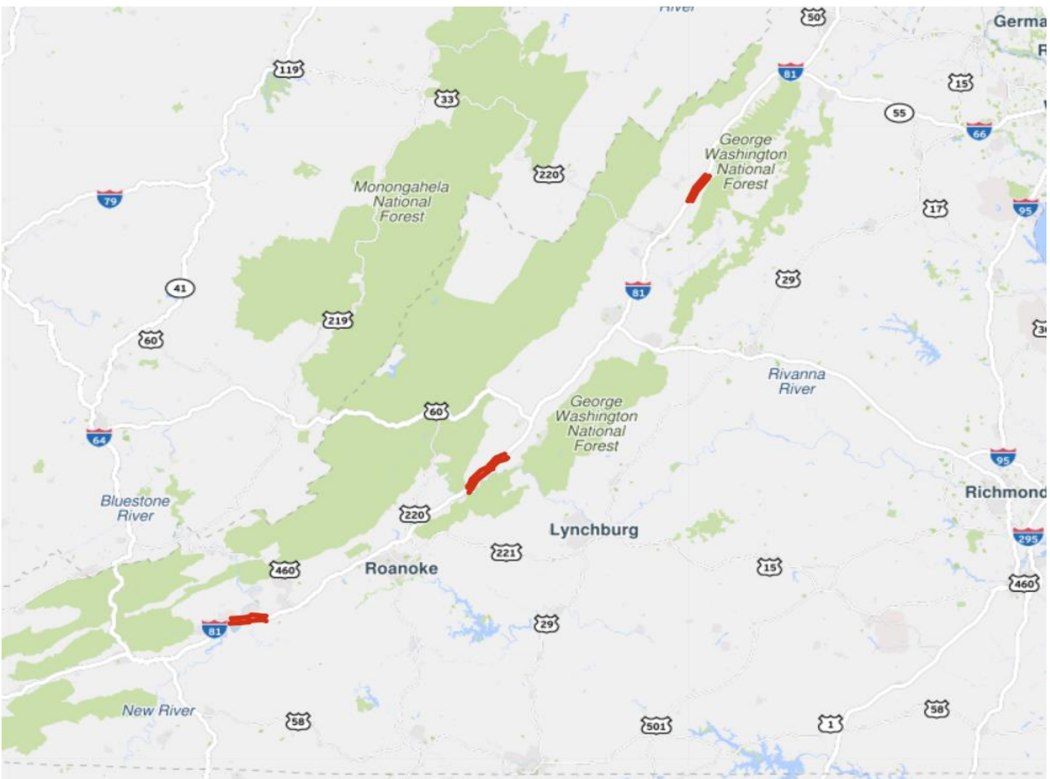


Figure 8. Traffic Message Channel Locations With New Safety Service Patrol Coverage or Extended Schedule

2. *I-64 Widening in Richmond (before-after capacity expansion).* To reduce congestion along I-64 in the Richmond area, the section of I-64 between Exit 200 (I-295 interchange) and Exit 205 (Bottoms Bridge) was widened from four to six lanes. The location of this section is shown in Figure 9. The construction of this project began in August 2017 and was completed in August 2019. Because of the impact of COVID-19 on traffic, data from 2020 were not considered. As a result, the "after" period was

set as August 2019 to December 2019. Accordingly, the "before" period was set from March 2017 to July 2017, resulting in 5 months for both periods. Since the data during the construction period were available, travel time reliability of the "during" period, from January 2018 to December 2018, was also included in the analysis.

Figure 9. I-64 Widening Location (VDOT, 2021b)

3. *I-64/I-264 Interchange Improvements—Phase II (before-during work zone impacts).* To optimize the traffic operations of the I-64/I-264 interchange, this project will reconfigure the roadway connections and add a new flyover. The project location is shown in Figure 10. The construction began in February 2018. As a result, the before period was set from February 2017 to February 2018, and the during period was set from March 2018 to March 2019.

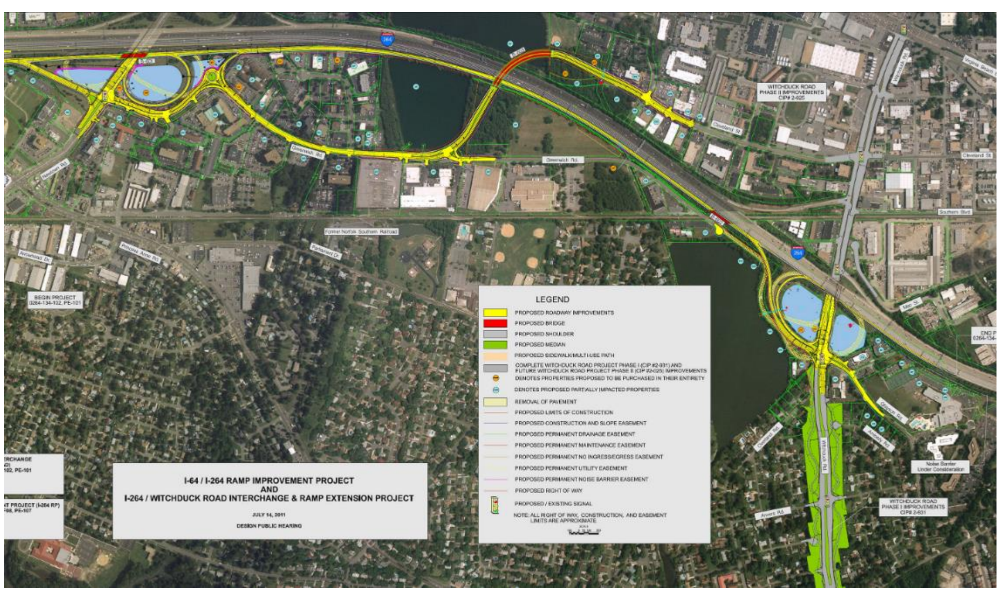


Figure 10. I-64/I-264 Interchange Improvement Project Location (VDOT, 2021c)

Evaluation Methodology

The optimal models using LQMMs and random forests were applied to the selected projects to demonstrate model application. Each model used two sets of input data for the after/during periods to predict travel time reliability. One set was the actual data, and the other was projected data based on the expected changes mapped to certain impact factors resulting from the implementation of the three projects. Using the actual data provides a way to show how well the models predicted reliability changes vs. actual observed data. Use of the projected data provides an example of how the models could be used in practice to assess proposed future projects that do not yet have any available after data.

The methods used for calculating the projected values of impact factors were as follows:

- *Geometric features.* These include TMC length, number of through lanes, presence of HOV/HOT/express lanes, and area type. Unless the studied projects create geometric changes, such as the addition of a lane to expand capacity or the conversion of general purpose lanes to HOV/HOT/express lanes, all factors remain the same as in the before period.

- *Weather factors.* The rain and frozen precipitation variables are assumed to have the same values as those on the same day of the month from the previous year.

- *Incidents and crashes.* The factors included in this category include the crash frequency with four severity levels, the frequency of breakdowns, and the frequency of hazards. A common method to project crash changes is to use safety performance functions (SPFs). According to SPFs, crash frequency is proportional to AADT and segment length, as shown in Equation 9:

  $$\begin{aligned} &Predicted\ crash\ frequency\ per\ year\ for\ freeway\ segment \\ &\quad = e^{\alpha} \times AADT_{one\ direction}^{\beta_1} \times Segment\ length_{one\ direction} \end{aligned} \quad \text{[Eq. 9]}$$

  Using the AADT and Virginia-specific parameters shown in Table 1 (Kweon and Lim, 2014), the ratio of SPF-predicted crash frequency between the before and after periods could be obtained. This ratio could then be multiplied by the observed crash frequency to get the projected values of all crashes. SPFs are applicable only to crashes. For projecting the frequency of breakdown and hazard events, it is assumed that they are directly proportional to AADT. The growth rate of AADT was applied to calculate the projected frequency of breakdowns and hazards.

- *Work zones.* Because of the different traffic environments for work zones, crash modification factors (CMFs) are applied to account for work zone effects. The two projects considered for the work zone study involved large-scale construction. The CMF was used under the scenario of daytime work zones with one or more lanes closed for freeways and expressways. According to Ullman et al. (2018), the CMFs under this condition are 1.66 for the total crash and 1.46 for the injury-related crash.

Table 1. SPF Virginia-Specific Parameters

Site Type	Total Crash		Fatal and Injury	
	α	β_1	α	β_1
Urban freeway, 4 lanes	-18.05	1.98	-18.27	1.88
Urban freeway, 6 lanes	-12.85	1.45	-15.64	1.6
Urban freeway segments within an interchange area, 4 lanes	-12.05	1.43	-12.53	1.35

SPF = safety performance function.

- It is assumed that the work zone lane closure information is uncertain at the time of conducting the before-after study when projected data are used. As a result, the two factors directly related to work zones, the frequency of lane closures and the frequency of shoulder closures, in the projected after-period data are set to zero, and the impact of work zones is reflected only through crash variables.

- *Traffic demand.* This includes the v/c ratio and heavy vehicle percentage. The after-period volumes are extrapolated by the ratio of AADT changes between the previous 2 years. Induced demands by capacity expansion are not considered. Capacity remains the same unless there are changes in the number of through lanes. It should be noted that the capacity reduction because of lane closures cannot be accounted for in the projected after-period data since, as mentioned previously, the work zone impact was considered only through crash changes.

- *SSP*. SSP schedules are planned in advance. So, the projected values of the SSP variable are assumed to be the same as the actual after-period data.

The actual and projected data for the after/during periods are then input into the proposed LQMMs and random forest models to predict the 50th, 80th, and 90th percentiles of travel time distributions. The corresponding values of LOTTR are also calculated. The before-after study is conducted by comparing the actual and predicted changes of travel time reliability (e.g., the LOTTR and the 90th percentile).

RESULTS AND DISCUSSION

Literature Review

Travel Time Reliability Measures

As there are multiple definitions of travel time reliability, various metrics and measures have been created to represent different aspects of reliability. The following section provides a summary of these metrics and measures and their applications and suitability under certain circumstances.

Travel Time Percentiles

Percentiles are well-known statistical measures and are easy to calculate, but different percentiles might be suitable for different reliability applications. The SHRP 2 Project L03 study (Cambridge Systematics, 2013) pointed out that the 95th percentile travel time may be too extreme to reflect improvements created by certain operational strategies. Instead, the 80th percentile may be a better performance measure for understanding the effects of operational improvements. A similar study by Zhang and Chen (2019) to quantify the impact of inclement weather and incidents on reliability suggested that the 95th percentile is more sensitive to weather and the 90th percentile is more sensitive to incidents. Lam and Small (2001) found that the difference between the 90th percentile and the median travel time is a good measure to assess the value of travel time reliability in a value pricing case study.

LOTTR

LOTTR is the reliability metric required by FHWA rulemaking (FHWA, 2017) to be reported periodically. It is calculated as the ratio of the 80th percentile to 50th percentile travel time of each reporting segment during four time periods (6 to 10 AM weekdays; 10 AM to 4 PM weekdays; 4 to 8 PM weekdays; 6 AM to 8 PM weekends) of the entire year. Segments with LOTTR values below 1.5 for all four time periods are classified as reliable. Otherwise, they are classified as unreliable if the threshold is exceeded for one or more periods. The travel time reliability measures are derived using Equation 10 for the interstate and non-interstate NHS:

$$TTRM = 100 \times \frac{\sum_{i=1}^{R} SL_i \times AV_i \times OF_j}{\sum_{i=1}^{T} SL_i \times AV_i \times OF_j} \qquad \text{[Eq. 10]}$$

where

$TTRM$ = travel time reliability measure

$SL_i =$ the segment length of interstate (or non-interstate) NHS reporting segment i

AV_i = annual traffic volume of reporting segment i, calculated as AADT × Directional factor × 365 (366 for leap year), where the directional factor is the factor for splitting AADT by direction with the default value of 0.5

OF_j = occupancy factor for vehicles on the NHS within a specified geographic area j within the state/metropolitan planning area

R= total number of interstate (or non-interstate) reporting segments with an LOTTR value below 1.50 for all four time periods

T = total number of interstate (or non-interstate) NHS reporting segments.

Variance and Standard Deviation

Both variance and standard deviation measures represent how spread out travel time distributions are. These are widely used in the transportation field because they are convenient to calculate (Dong and Mahmassani, 2009). Dowling et al. (2009) explored the correlation between standard deviation and other reliability measures, such as the buffer time index (BTI), misery index, planning index, and failure/on-time. The authors used 1 year's worth of 5-minute speed and volume data collected from four detector locations on two freeways in the San Francisco Bay Area. They concluded that standard deviation could be used as a proxy for other measures. The California DOT (1998) also adopted standard deviation to evaluate the travel time reliability of roadway links.

Buffer Time Index

The BTI is the ratio of the difference between the 95th percentile travel time and the average travel time over the mean travel time. It measures the extra time above the average travel time required to arrive on time 95% of the time, as shown in Equation 11. In the SHRP 2 Project L03 study (Cambridge Systematics, 2013), a thorough empirical analysis was conducted to find suitable reliability measures for highway improvement evaluations. The authors found that when travel time distributions were highly skewed, the median was a more robust estimation of central tendency than the mean. Pu (2011) confirmed this conclusion that median-based BTI is preferred over average-based BTI (see Eq. 11).

$$\mathrm{BTI} = \frac{\text{95th percentile travel time} - \text{Average or median travel time}}{\text{Average or median travel time}} \qquad \text{[Eq. 11]}$$

Planning Time Index

The planning time index (PTI) is calculated using the 95th percentile travel time divided by the free-flow travel time. It represents the extra time travelers should budget in addition to free-flow travel time to have a 95% probability of arriving on time. For example, a PTI value of 1.2 means that a 10-minute free-flow trip requires 12 minutes during the heaviest traffic. Although the BTI shows the additional travel time that needs to be budgeted, the PTI shows the total travel time required for on-time arrivals. List et al. (2014 a) concluded that PTI is recommended to measure reliability for daily, constrained trips. Such trips refer to those "for which the user experiences day-to-day variability in travel time (because of recurring congestion and incident or nonrecurring congestion) and desires to arrive at the destination at a fixed time (or within a small time window)" (see Eq. 12).

$$\mathrm{PTI} = \frac{\text{95th percentile travel time}}{\text{Free} - \text{flow travel time}} \qquad \text{[Eq. 12]}$$

Travel Time Index

Travel time index (TTI) is the ratio of peak-period travel time to the time required to make the same trip at free-flow speed. A TTI of 1.2 means that the travel time during the peak period is 20% longer than the free-flow travel time. The SHRP 2 Project L02 study (List et al., 2014b) recommended using the TTI to measure travel time reliability for occasional, constrained trips for passenger travelers. "Occasional, constrained trips" are defined as trips "for which the user does not experience day-to-day variability but does have temporal constraints on arrival time." SHRP 2 Project L08 (Zegeer et al., 2014) used a TTI of 1.33 for freeways and 2.50 for urban streets as the thresholds of congestion. When used for measuring multiple periods, the TTI can be used to compare measured travel time to free-flow travel time. In this case, TTI is treated as a congestion measure rather than a reliability performance metric (see Eq. 13).

$$\mathrm{TTI} = \frac{\text{Mean peak period travel time}}{\text{Free} - \text{flow travel time}} \qquad \text{[Eq. 13]}$$

Percent On-Time Travel and Misery Index

Percent on-time travel is the percentage of trips with travel times less than a particular threshold. Past thresholds have been set at 110% to 130% of average travel time (Cambridge Systematics, 2013). Depending on various contexts and use case scenarios, the Florida DOT (2000) adopted four levels of thresholds (greater than 5%, 10%, 15%, 20% over the average travel times) to define unreliable trips. Slightly different from on-time arrival and the Florida DOT's method, the misery index emphasizes how bad the worst days are by comparing the longest 20% trips to the average.

Factors Affecting Reliability

A National Cooperative Highway Research Program study (2003) identified seven factors that contribute to unreliable travel times: traffic incidents, inclement weather, work zones, special events, traffic control devices, fluctuations in demand, and inadequate base capacity. This study considered only freeways, so the impact of traffic control devices (e.g., traffic signals) was not included. Table 2 summarizes selected studies that attempted to estimate the impact of the other six factors. In most cases, previous studies examined the effect of only one or two of these factors and did not consider the impact of multiple factors present simultaneously.

Reliability Modeling Approaches

Although many studies have investigated the impacts of various factors on reliability, no models suitable for reliability prediction using all these factors simultaneously have been developed. The current state of the practice relies on historical trend lines or assumptions of relationships between reliability and traditional congestion measures. The TPB, which is the metropolitan planning organization for metropolitan Washington, summarized the current forecast methods used by VDOT, the Maryland DOT, and the District of Columbia DOT (TPB, 2018). Three methods were mentioned and used to analyze travel time data in the Washington, D.C., area and are illustrated in Figure 11:

1. *Extrapolation of measured performance.* DOTs set targets for future years through fitting travel time reliability of past years into trend lines (linear or exponential), as shown in Figure 11(a).

Table 2. Summary of Factors Affecting Reliability

Factor	Data Source	Reliability Measures	Methodology	Conclusions	Authors
Traffic incidents/ Crashes	Loop detectors	10th, 50th, and 90th percentiles	Descriptive analysis	The 10th, 50th, and 90th percentiles of travel time with crashes were higher by an average of 3%, 38%, and 75%, respectively, compared to travel time without crashes.	Tu et al. (2008)
	Loop detectors	Buffer time index and percentiles (10th, 50th, and 90th)	Descriptive analysis	1. A higher number of lanes increases the resilience of roadways to incident-caused unreliability. 2. Although shoulder incidents did not block any travel lanes, their impact on reliability was similar to that of single-lane incidents.	Wright et al. (2015)
	Loop detectors	Extra buffer time index (EBTI)	Tobit regression	The independent variables of "incident duration" and "more than one lane blocked" were significant for all incident types. In contrast, "multiple vehicles involved," "major incident," and "assistant requested" were significant only for crashes, hazards, and stationary vehicles, respectively.	Hojati et al. (2016)
	Loop detectors	90th percentile	Linear regression	The authors recommended two models depending on the data availability: one with fewer variables, including mean travel times, number of lanes, time of day, and direction, and the other with more variables, including number of crashes, state of repair, and road works in addition to the previous variables.	Charlotte and Sandra (2017)
Inclement weather	Taxi GPS data	Mode travel time, coefficient of variation (CoV)	Decision tree–based method	The study found that inclement weather improved travel time reliability, which was counterintuitive. This could partially be because of the reduction of taxi and limousine commission services because of adverse weather.	Yazici et al. (2013)
	EZPass roadside sensors	95th percentile, standard deviation, percent variation, buffer time, and skewness	Monte Carlo simulation (estimation of the travel time distribution)	As weather conditions went from clear to rain and snow, almost all reliability measures increased, meaning inclement weather exacerbates travel time reliability.	Caceres et al. (2012)
	Vehicle probe data	80th and 90th percentiles	Decision tree–based method	The magnitude of the negative impact on the average travel time was highly related to the level of recurrent congestion. It was stated that the 80th percentile travel time could be used to assist travelers with route planning and the 90th percentile travel time can be used by freight carriers to ensure just-in-time deliveries.	Zhang and Chen (2019)
Work zones	Vehicle probe data	95th percentile travel time, buffer time index, and planning time index	Paired sites comparison	Measures showed that work zones degraded reliability at a significant level. Specifically, the buffer time index increased on average by 48%, the planning time index increased on average by 18%, and the 95th percentile travel rate increased by an average of 16%.	Edwards and Fontaine (2012)
	NPMRDS	Work zone impact ratio (the ratio of travel time reliability measures of work zone and baseline)	Linear regression	Independent variables included the traffic volume per open lane, the total number of lanes, and the open ratio. The model predictions were poor for all reliability measures. The authors attributed this to a lack of information on various work zone details such as barrier types, posted speed limits, lateral distance to the obstacle, etc.	Srivastava et al. (2018)
Demand/ Capacity	Loop detectors	10th, 50th, and 90th percentiles	Descriptive analysis	The authors concluded that below a certain inflow (in this case, 1,400 veh/hr/ln), traffic crashes slightly increase median travel time by an average of 6%; above the inflow of 1,400 veh/hr/ln, crashes increase median travel time by average of 38%.	Tu et al. (2008)
	Loop detectors	Travel time variance	Linear regression	The authors concluded that the breakpoint in reliability generally occurs at a volume significantly lower than the theoretical capacity of the facility computed in accordance with the *Highway Capacity Manual*. The breakpoint volume appeared to be a dynamic value influenced by various factors that were not included in the study.	Dowling et al. (2009)

NPMRDS = National Performance Management Research Data Set.

2. *The use of travel demand model data*. Travel demand models project future traffic conditions and produce changes in congestion-related measures. Although reliability is not included in these measures, forecasting could be achieved through using the change rate of percentage of congested AM peak hour vehicle miles traveled, as shown in Figure 11(b). The impact of nonrecurring congestion is not explicitly modeled using this method.

3. *The average of the two*. The results of the previous methods are averaged, as illustrated in Figure 11(c).

These methods are relatively simple and provide estimates at a highly aggregated level. The benefits of operational and capacity improvement projects are not included in the target setting process in the aforementioned methods, and negative impacts from planned disruptive events such as work zones are also not included. Effects of factors impacting reliability are considered only through the use of observed historical trends and do not explicitly model the impact of items such as work zones, incidents, and weather.

It is noticeable that most reliability measures introduced earlier rely on travel time distributions. The starting point of estimating travel time distributions could be fitting travel time data through single-mode probability distributions.

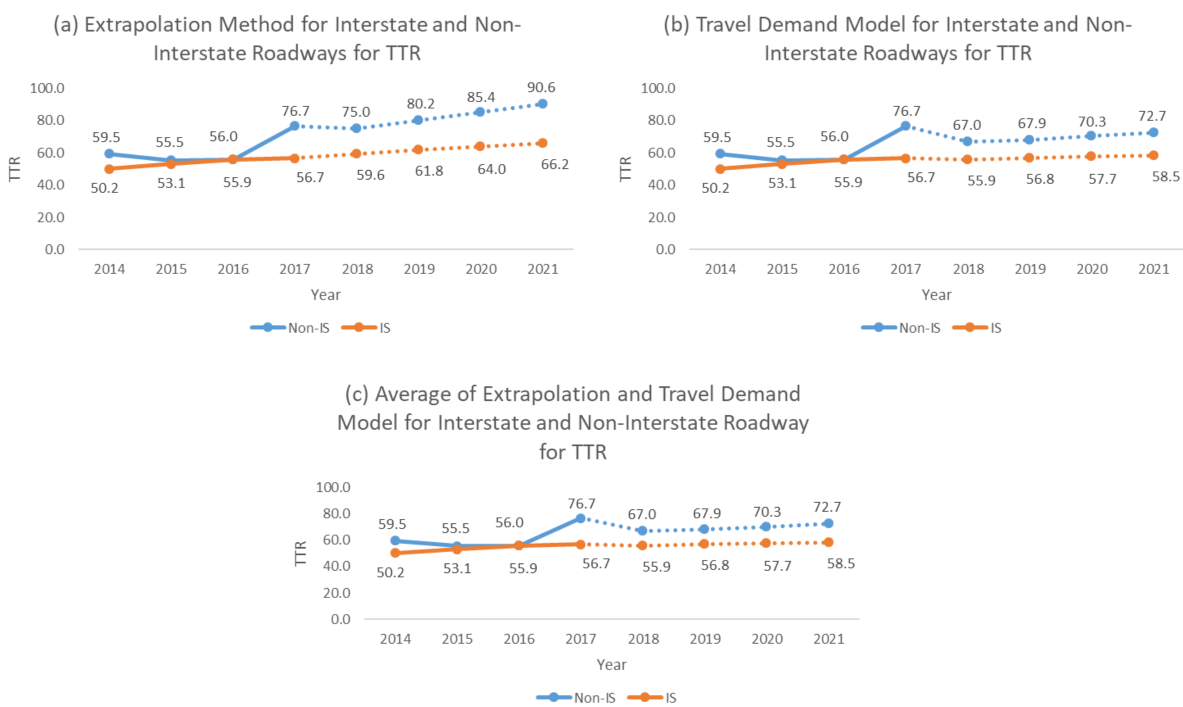


Figure 11. Performance Target Setting Methods. TTR = travel time reliability. Adapted From The National Capital Region Transportation Planning Board (2018).

A variety of distributions have been tested in previous studies, such as normal and lognormal (Pu, 2011); gamma and compound gamma (Kim and Mahmassani, 2015); Weibull (Arroyo and Kornhauser, 2005); generalized beta (Castillo et al., 2012); Halphen distribution (Delhome et al., 2017); and Burr distributions (Taylor, 2017). Emam and Al-Deek (2006) employed the AD goodness-of-fit statistics and 90th percentile of absolute error to evaluate the performance of four distribution types: Weibull, exponential, lognormal, and normal. The modeling results indicated that the lognormal distribution provided the best model fit. In addition, data from the same day of the week fit better than data collected across multiple weekdays because of the significant differences between traffic patterns across days. Li et al. (2006) suggested that a lognormal distribution best characterized the distribution of travel time when a large time window (e.g., more than 1 hour) was under consideration with the presence of congestion.

Little agreement could be found in previous studies on the type of probability distribution most appropriate for modeling travel times. One main reason behind this is the heterogeneity of the traffic environment. Authors developed more complex modeling methodologies to overcome the limitation of single-mode distributions, such as mixture/multi-state and non-parametric models. These models can better capture the heterogeneity by associating travel time distribution with different traffic conditions both temporally and spatially. Specifically, two levels of uncertainty can be considered here: (1) the probability of a given traffic condition (such as free flow, congestion onset, congestion, and congestion dissipation), and (2) travel time variation within each traffic condition (Park et al., 2011; Van Lint and Van Zuylen, 2005). However, unpredictable factors (e.g., weather, work zones, incidents) further complicate the classification of traffic states. In addition, adopting different goodness-of-fit measures to determine the best fit distribution types also contributes to the lack of consistent results. Plötz et al. (2017) demonstrated that different goodness-of-fit measures might lead to different conclusions with regard to the best fit distribution types even when the same travel time dataset is used. Several studies have used different multi-state models, such as the normal mixture model (Guo et al., 2015); gamma mixture model (Yang and Wu, 2016); and kernel density estimation using the Hasofer-Lind-Rackwitz-Fiessler algorithm (Yang et al., 2014). Yang and Wu (2016) believed that the selection of single-mode distribution types has little impact on measuring travel time reliability if a mixture model is used.

Guo et al. (2015) introduced two approaches to incorporate the influence of traffic volume on travel time reliability: (1) Bayesian mixed-effect travel time regression models, and (2) hidden Markov models. The Bayesian mixed-effect travel time regression model advanced the multi-state travel time reliability model of Park et al. (2011) by developing regressions on the proportions and distribution parameters for underlying traffic states using field data collected along a section on I-35 in San Antonio, Texas. The modeling results indicated a negative relationship between the proportion of free-flow state and traffic volume. There might exist only one travel time state for low traffic volume conditions, in which case single-state models would be sufficient. The estimation for the congested state indicated that the travel time under such conditions had substantial variability and is positively related to traffic volume. The hidden Markov model considered the interactions of consecutive segments compared to the former model that assumes all observations are independent. The modeling results showed that the traffic volume has a positive effect on the proportion of congested state and the mean parameters of such state. In terms of the model fitting, it was concluded that the hidden Markov model is superior with regard to interpreting the data without sacrificing model simplicity. Although the non-parametric and mixture models can provide a better fitting, it is not always practical to derive travel time reliability measures from these complex models, limiting their applications.

The HCM-6 provides a probability-based method to incorporate the impact factors in reliability analysis at a freeway facility or corridor level. The core element of this method is the scenario generator, which provides a large set of different combinations of impact factors (e.g., demand, weather, incidents) with their corresponding probabilities. Then, travel times are inferred for each scenario through the HCM FREEVAL tool to construct the travel time distribution. According to Tufuor et al. (2020), the HCM-6 method was validated only through simulations without calibration with empirical travel time data. Tufuor and Rilett (2020) compared the travel time distribution built by the HCM-6 method and the empirical distribution using data on a 1.16-mile testbed in Lincoln, Nebraska. The results indicated that these two distributions were statistically different, with the standard deviation of travel time distribution by the HCM-6 method being 67% less than that of the empirical distribution.

In Virginia, the *SMART SCALE Technical Guide* (Office of Intermodal Planning and Investment, 2021) provides guidance on how to evaluate travel time reliability at a corridor level five using factors: BTI, incident impact, incident frequency, weather impact, and weather frequency. The incident frequency and weather frequency are assigned scores from 0 to 5 and 0 to 2, respectively, depending on the actual times the events happened. The incident impact and weather impact are also defined using scores ranging from 0 to 2, which are determined from Table 10.9 in the *SMART SCALE Technical Guide.* To compute travel time reliability, the product of the incident impact and the incident frequency is added to the product of the weather impact and the weather frequency. This result is then multiplied by the BTI.

Data Collection and Preparation

The data used in this study were assembled from VDOT-maintained and crowdsourced sources, as discussed previously. Tables 3 and 4 list relevant variables and their factor categories for the VDOT-maintained (VaTraffic) and crowdsourced (Waze) datasets, respectively. The incident variables distinguish these two sets of input variables and their resulting models. The VDOT-maintained crash data provide the severity level of crashes so that each type of crash is counted separately; crowdsourced data count only the frequency of crashes with no indication of severity.

Cluster Analysis

Three quantiles, the 50th, 80th, and 90th, were set as the target variables. Table 5 shows a comparison of the efficiency of different clustering methods for the 80th percentile. The average Silhouette value by geometric feature was -0.47 and -0.75 for freeway and interchange segments, respectively, which indicated that this was not a preferable clustering method. For freeway segments, hierarchical clustering using the KS dissimilarity matrix and complete linkage had the best performance, and the optimal cluster size was 2.

Table 3. Variable Summary for VaTraffic Models

Factor Category	Factor	Variable Name
Geometric features	Traffic Message Channel length (miles)	miles
	Number of through lanes (count)	throu_lane
	Presence of parallel managed lanes (presence=1, otherwise=0)	Par_lane
	Area type (urban=0, rural=1)	rural
Weather	Frozen precipitation (inches to hundredths)	frozen_precip
	Rain precipitation (inches to hundredths)	rain_precip
Incident	Crash frequency by severity level (count)	Severe_Injury
		Visible_Injury
		Nonvisible_Injury
		PDO
	Frequency of breakdown (disabled vehicles) (count)	breakdown
	Frequency of hazard (fire related) (count)	hazard
Work zone	Shoulder closure (count)	shoulder_closure
	Lane closure (count)	lane_closure
Traffic demand	Volume-to-capacity ratio	vc_ratio
	Percentage of heavy vehicles	heavy_percent
Safety Service Patrol (SSP)	Presence of safety service patrol (SSP present=1, otherwise=0)	ssp

Table 4. Variable Summary for Waze Models

Factor Category	Factor	Variable Name
Geometric features	Traffic message channel length (miles)	miles
	Number of through lanes (count)	throu_lane
	Presence of parallel managed lanes (presence=1, otherwise=0)	Par_lane
	Area type (rural=0, urban=1)	rural
Weather	Frozen precipitation (inches to hundredths)	frozen_precip
	Rain precipitation (inches to hundredths)	rain_precip
Incident	Frequency of crashes (count)	WAZE_crashes
	Frequency of breakdown (count)	WAZE_breakdown
	Frequency of hazard (count)	WAZE_hazard
Work zone	Shoulder closure (count)	shoulder_closure
	Lane closure (count)	lane_closure
Traffic demand	Volume-to-capacity ratio	vc_ratio
	Percentage of heavy vehicles	heavy_percent
Safety Service Patrol (SSP)	Presence of safety service patrol (SSP present=1, otherwise=0)	ssp

For interchange segments, hierarchical clustering using the AD dissimilarity matrix and Ward's linkage had the best performance, and the optimal cluster size was also 2. The average Silhouette values of the 50th and 90th percentiles were also computed since the optimal cluster results might change. For this study, the highest average Silhouette value was from hierarchical clustering using a KS matrix with a cluster size of 2 for all three percentiles of freeway segments. For interchange segments, hierarchical clustering using an AD matrix with a cluster size of 6 had the highest value for the 50th percentile, and a KS matrix with a cluster size of 5 had the highest value for the 90th percentile. Since the Silhouette values did not differ too much from each other, it was decided to adopt the smallest cluster size to reduce the complexity of the next step.

Table 5. Average Silhouette Width of the 80th Percentiles of Travel Time

Clustering Approach	Linkage Method	Average Silhouette	
		Freeway Segment	Interchange Segment
Hierarchical: Kolmogorov-Smirnov	Ward's	0.27	0.13
	Complete	**0.49**	0.20
Hierarchical: Anderson-Darling	Ward's	0.36	**0.23**
	Complete	-0.04	0.21
Model-Based	N/A	0.34	-0.21
Geometric Feature	N/A	-0.47	-0.75

Bold font indicates the best performing method. N/A = not applicable.

The travel time distribution of freeway segments overall had a less curved shape than distributions of interchange segments. The model-based functional data clustering did not offer better results since typical applications of this type of clustering are often curves with peaks or time-series data. Theoretically, clustering using the AD matrix is expected to have better results than with the KS matrix. However, it was observed from the TMC distribution plots that the maximum distances were most likely to occur toward the tail of the distribution, which could be captured by the KS statistics. As a result, the theoretical advantage of AD statistics did not materialize.

Figures 12 and 13 show the distribution of each cluster for the final clustering results of freeway and interchange segments, respectively. Below the distribution plots, clustered segments are shown geographically in a map, which demonstrates reliability patterns by location.

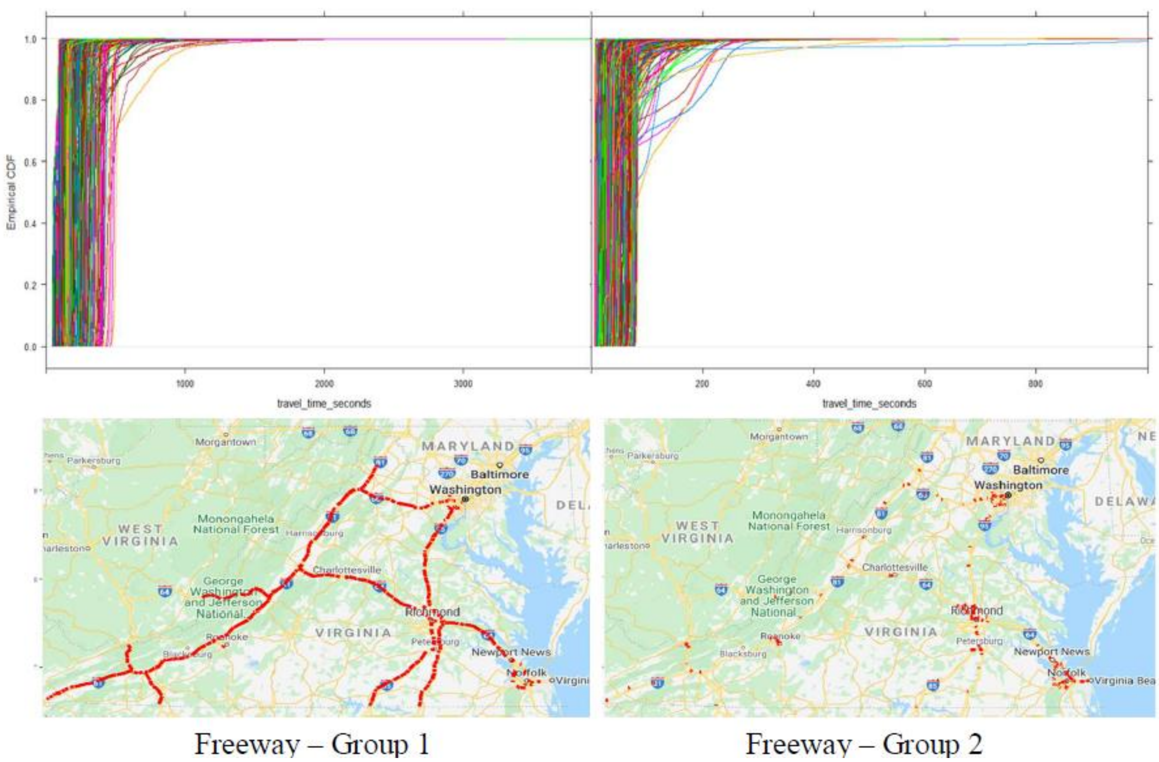
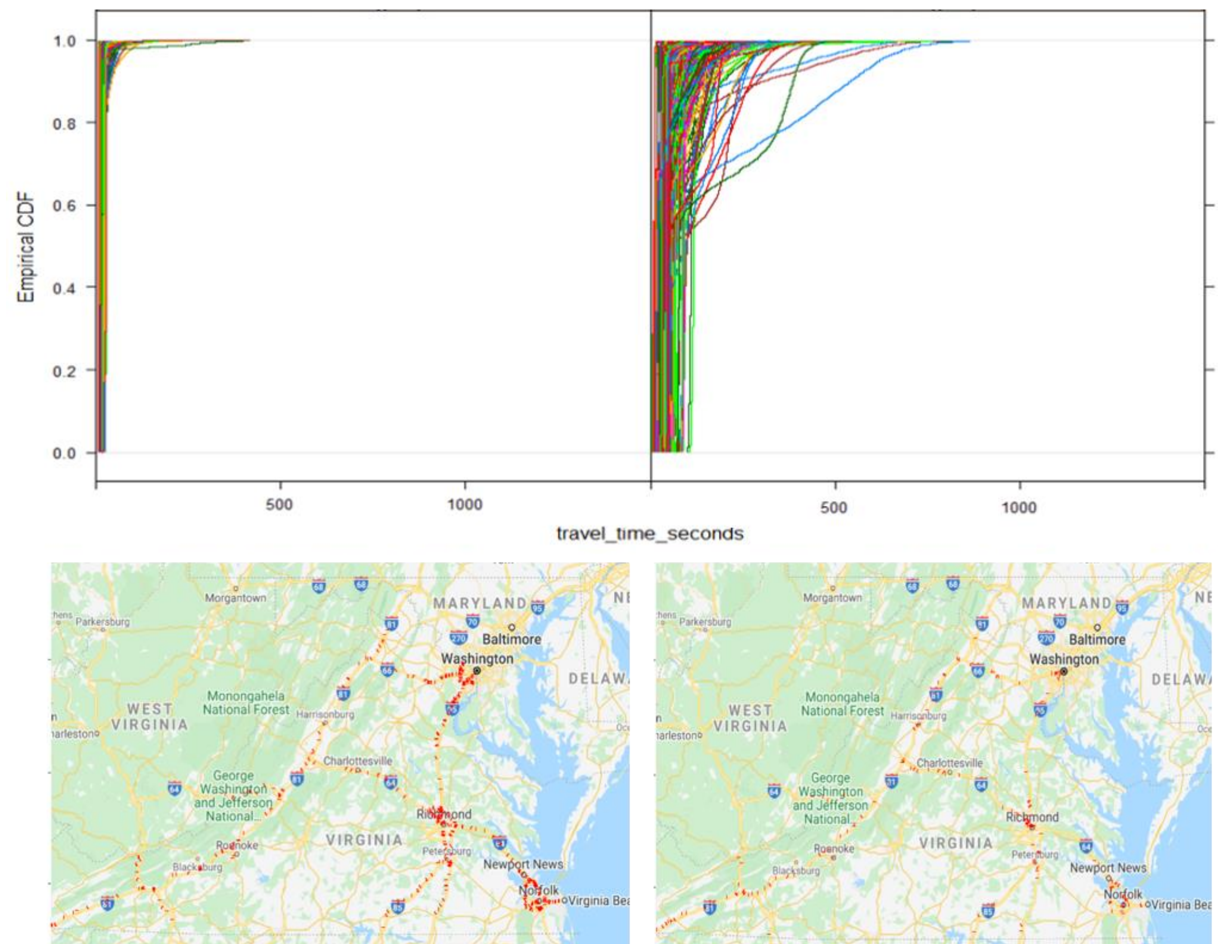


Figure 12. Final Clustering Results for Freeway Segments

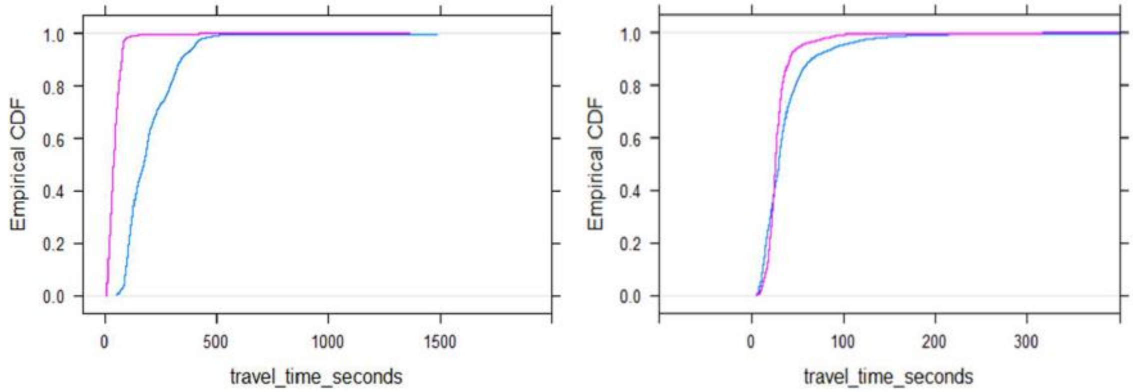


Interchange – Group 1

Interchange – Group 2

Figure 13. Final Clustering Results for Interchange Segments

For freeway segments, most segments in rural areas fell into clusters with relatively reliable travel times (Group 1), and segments along urban areas were likely to be in clusters with lower reliability levels (Group 2). For interchange segments, the urban/rural division was not as apparent as with freeway segments. Figure 14 summarizes the overall distribution of each cluster. It was observed that each cluster had a unique overall distribution shape, especially in the tails. Distributions with shorter tails indicated that travel times from those clusters were more reliable than distributions with longer tails. To quantify the reliability trends better, tables in Figure 14 also summarize the average and standard deviation of the 80th percentile travel times and the LOTTR of each cluster.



Cluster Group	Mean LOTTR	SD LOTTR
1	1.12	0.34
2	1.22	0.51

Cluster Group	Mean LOTTR	SD LOTTR
1	1.07	0.16
2	1.28	0.65

Figure 14. Overall Distributions and Reliability Comparison: left, freeway; right, interchange. LOTTR = level of travel time reliability; SD = standard deviation.

Reliability Analysis Using LQMMs

The assembled data were used to develop separate LQMMs for freeway segments and interchange segments based on the clusters from the previous step. Separate models were also developed using VaTraffic data and Waze data. The variables listed in Tables 3 and 4 were considered inputs. Also, interaction terms were considered between v/c ratio and non-recurrent events, such as weather, incidents, and work zones. The impact of influencing factors was interpreted through the estimated coefficients resulting from the LQMMs.

Freeway Segment Models

For freeway segments, two groups were defined by the cluster analysis. As discussed previously, Group 1 is more reliable with longer TMC segments in rural areas and Group 2 is less reliable with shorter segments in urban or suburban areas. Before quantile models were developed, a descriptive analysis was conducted to provide an overview of each input dataset, as shown in Tables 6 and 7.

To make the coefficients comparable and more easily interpretable, standardized values of coefficients were also calculated. These were calculated by multiplying the coefficient (b_i) by the standard deviation of its corresponding independent variable and dividing it by the standard deviation of the dependent variable. They represent the expected change in Y (in standardized units of Y where each "unit" is a statistical unit equal to 1 standard deviation) because of an increase in X_i of one of its standardized units with all other X variables unchanged. The model results are summarized in Table A1 in the Appendix.

Table 6. Input Data Summary for VaTraffic Freeway Models

Group	TMC No. (Miles)	Variable Name (Units)	Mean	Std. Dev.	Minimum	Maximum
Freeway-Group 1	475 (1548.77)	Miles (mile)	3.4571	1.8527	0.60	8.97
		throu_lane (count)	2.3884	0.6689	2.00	5.00
		frozen_precip (inches to hundredths)	0.0008	0.0180	0.00	1.44
		rain_precip (inches to hundredths)	0.0382	0.3003	0.00	20.13
		Severe_Injury (count)	0.0007	0.0259	0.00	2.00
		Visible_Injury (count)	0.0043	0.0675	0.00	3.00
		Nonvisible_Injury (count)	0.0006	0.0253	0.00	1.00
		PDO (count)	0.0159	0.1352	0.00	6.00
		shoulder_closure (count)	0.0075	0.0883	0.00	5.00
		lane_closure (count)	0.0052	0.0739	0.00	4.00
		Breakdown (count)	0.0419	0.2209	0.00	5.00
		Hazard (count)	0.0006	0.0246	0.00	2.00
		Rural (0/1)	0.5223	0.4995	0.00	1.00
		vc_ratio (number)	0.3167	0.1871	0.03	1.10
		Par_lane (0/1)	0.0413	0.1990	0.00	1.00
		heavy_percent (percent)	16.2440	8.3276	0.00	33.15
		SSP (0/1)	0.4845	0.4998	0.00	1.00
Freeway-Group 2	467 (231.25)	Miles (mile)	0.6331	0.3814	0.10	1.56
		throu_lane (count)	2.8894	0.8314	2.00	6.00
		frozen_precip (inches to hundredths)	0.0009	0.0226	0.00	1.44
		rain_precip (inches to hundredths)	0.0500	0.3970	0.00	20.13
		Severe_Injury (count)	0.0002	0.0144	0.00	1.00
		Visible_Injury (count)	0.0018	0.0424	0.00	2.00
		Nonvisible_Injury (count)	0.0003	0.0160	0.00	1.00
		PDO (count)	0.0059	0.0803	0.00	5.00
		shoulder_closure (count)	0.0022	0.0475	0.00	3.00
		lane_closure (count)	0.0017	0.0424	0.00	3.00
		Breakdown (count)	0.0125	0.1162	0.00	4.00
		Hazard (count)	0.0003	0.0167	0.00	2.00
		Rural (0/1)	0.1155	0.3196	0.00	1.00
		vc_ratio (number)	0.4271	0.1887	0.04	1.04
		Par_lane (0/1)	0.0902	0.2865	0.00	1.00
		heavy_percent (percent)	8.6772	7.3104	0.30	33.15
		SSP (0/1)	0.5316	0.4990	0.00	1.00

TMC = Traffic Message Channel; PDO = property damage only crash; SSP = safety service patrol.

To compare the magnitude of impacts, variables with significance levels higher than 90% for each model were plotted in Figures 15 and 16, with the standardized coefficients ordered from highest to lowest. The overall trends for both the VaTraffic and Waze models were that (1) the 50th, 80th, and 90th percentile travel times of Group 1 (more reliable) were impacted by more factors than Group 2 (less reliable), but each factor contributed almost equally in the two groups; and (2) when the same variable appeared in all three percentile models, the values of standardized coefficients increased toward the upper tail percentages. The first trend could be explained by the difference in the segment length of the two clusters. Group 1 had an average TMC length of 3.5 miles, providing a spatial opportunity for different events to occur at the same time. On the contrary, Group 2 had an average TMC length of only 0.63 miles, and impact factors were more likely to contribute separately on different TMC segments.

Table 7. Input Data Summary of Waze Freeway Model

Group	TMC No. (Miles)	Variable Name (Units)	Mean	Std. Dev.	Minimum	Maximum
Freeway-Group 1	475 (1548.77)	Miles (mile)	3.4571	1.8527	0.60	8.97
		throu_lane (count)	2.3884	0.6689	2.00	5.00
		frozen_precip (inches to hundredths)	0.0008	0.0180	0.00	1.44
		rain_precip (inches to hundredths)	0.0382	0.3003	0.00	20.13
		WAZE_crashes (count)	0.0720	0.4041	0.00	15.00
		WAZE_breakdown (count)	1.2197	1.8501	0.00	48.00
		WAZE_hazard (count)	0.2092	1.0275	0.00	322.00
		shoulder_closure (count)	0.0075	0.0883	0.00	5.00
		lane_closure (count)	0.0052	0.0739	0.00	4.00
		Rural (0/1)	0.5223	0.4995	0.00	1.00
		vc_ratio (number)	0.3167	0.1871	0.03	1.10
		Par_lane (0/1)	0.0413	0.1990	0.00	1.00
		heavy_percent (percent)	16.2440	8.3276	0.00	33.15
		SSP (0/1)	0.4845	0.4998	0.00	1.00
Freeway-Group 2	467 (231.25)	Miles (mile)	0.6331	0.3814	0.10	1.56
		throu_lane (count)	2.8894	0.8314	2.00	6.00
		frozen_precip (inches to hundredths)	0.0009	0.0226	0.00	1.44
		rain_precip (inches to hundredths)	0.0500	0.3970	0.00	20.13
		WAZE_crashes (count)	0.0348	0.2516	0.00	12.00
		WAZE_breakdown (count)	0.3405	0.7738	0.00	12.00
		WAZE_hazard (count)	0.0583	0.3424	0.00	39.00
		shoulder_closure (0/1)	0.0022	0.0475	0.00	3.00
		lane_closure (0/1)	0.0017	0.0424	0.00	3.00
		Rural (0/1)	0.1155	0.3196	0.00	1.00
		vc_ratio (number)	0.4271	0.1887	0.04	1.04
		Par_lane (0/1)	0.0902	0.2865	0.00	1.00
		heavy_percent (percent)	8.6772	7.3104	0.30	33.15
		SSP (0/1)	0.5316	0.4990	0.00	1.00

TMC = Traffic Message Channel; SSP = safety service patrol.

In terms of interpreting the modeling results for impact factors, the key findings were as follows:

- For the VaTraffic models, for both Group 1 and Group 2, the impacts of non-recurrent events such as incidents and weather changed from being significant only as interaction terms with v/c ratio to the impact factors alone being significant as the percentile being modeled increased.

- Non-recurrent events have a higher impact on Group 2 than Group 1 because the freeways near urban/suburban areas are already more congested than in rural areas and are more vulnerable to these random interruptions. This finding could also be supported by the fact that traffic demand variables, including v/c ratio and heavy vehicle percentage, are all significant and have a higher impact ranking in Group 1 but not in Group 2.

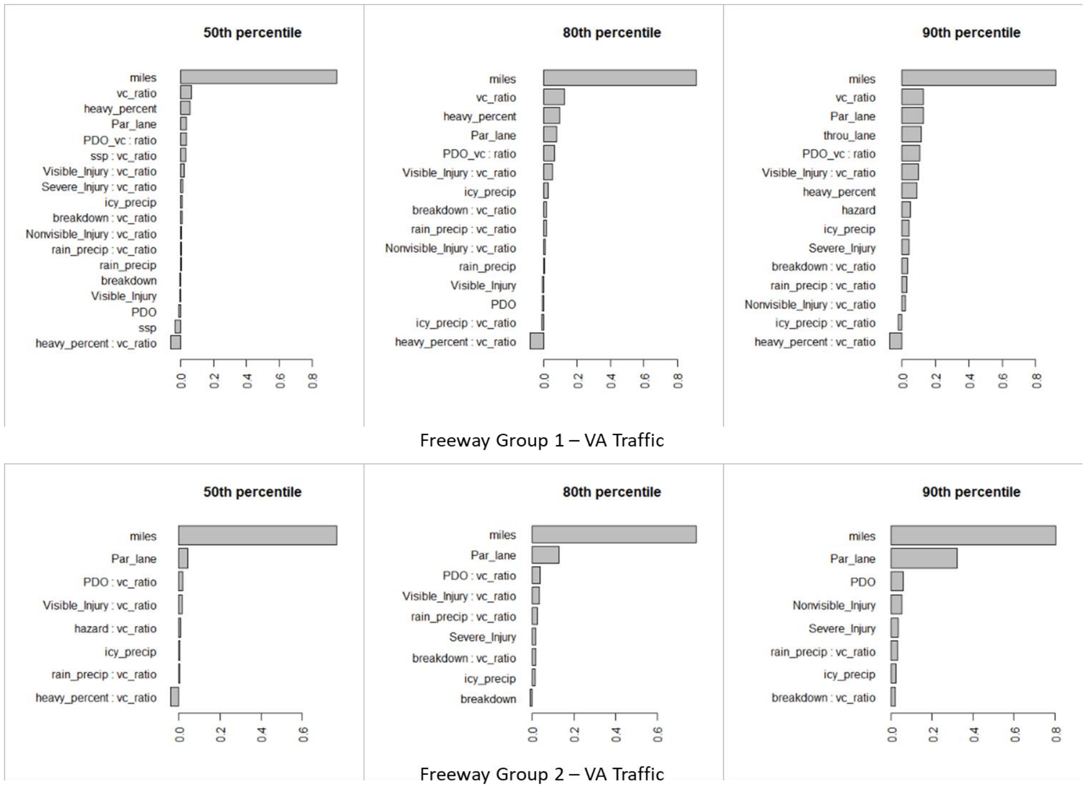


Figure 15. VaTraffic Freeway Segment Variable Coefficients Rankings. X-axis = Standardized Coefficients.

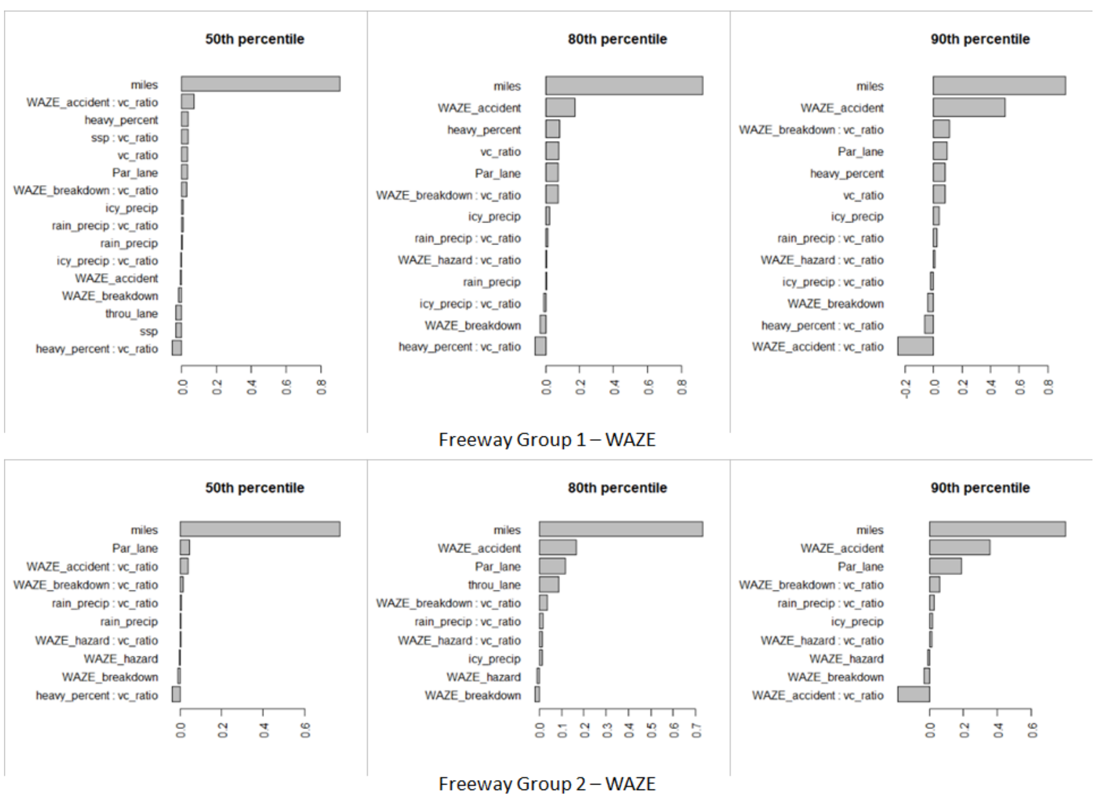


Figure 16. Waze Freeway Segment Variable Coefficients Rankings. X-axis = Standardized Coefficients.

Interchange Segment Models

The interchange segments were also clustered into two groups. There was no clear urban/rural division as seen in freeway segments. Similar to freeway segments, descriptive statistics are provided for interchange segments, as shown in Tables 8 and 9. Modeling results are summarized in Table A2 in the Appendix.

Table 8. Input Data Summary of VaTraffic Interchange Models

Group	TMC No. (Miles)	Variable	Mean	Std. Dev.	Minimum	Maximum
Interchange-Group 1	345 (163.23)	Miles (mile)	0.4784	0.2742	0.11	2.02
		throu_lane (count)	2.9360	0.8362	2.00	6.00
		frozen_precip (inches to hundredths)	0.0011	0.0260	0.00	1.44
		rain_precip (inches to hundredths)	0.0506	0.3827	0.00	20.13
		Severe_Injury (count)	0.0002	0.0142	0.00	1.00
		Visible_Injury (count)	0.0022	0.0475	0.00	3.00
		Nonvisible_Injury (count)	0.0003	0.0185	0.00	2.00
		PDO (count)	0.0072	0.0867	0.00	2.00
		shoulder_closure (count)	0.0024	0.0510	0.00	3.00
		lane_closure (count)	0.0018	0.0444	0.00	3.00
		Breakdown (count)	0.0166	0.1359	0.00	4.00
		Hazard (count)	0.0002	0.0133	0.00	1.00
		Rural (0/1)	0.0400	0.1960	0.00	1.00
		vc_ratio (number)	0.4855	0.1997	0.03	1.62
		Par_lane (0/1)	0.1304	0.3368	0.00	1.00
		heavy_percent (percent)	6.5769	6.0541	0.30	33.15
		SSP (0/1)	0.6162	0.4863	0.00	1.00
Interchange-Group 2	566 (244.93)	Miles (mile)	0.5184	0.2221	0.11	1.73
		throu_lane (count)	2.3731	0.6351	2.00	5.00
		frozen_precip (inches to hundredths)	0.0007	0.0150	0.00	1.36
		rain_precip(inches to hundredths)	0.0380	0.3109	0.00	20.13
		Severe_Injury (count)	0.0001	0.0112	0.00	1.00
		Visible_Injury (count)	0.0006	0.0250	0.00	2.00
		Nonvisible_Injury (count)	0.0001	0.0090	0.00	2.00
		PDO (count)	0.0026	0.0527	0.00	3.00
		shoulder_closure (count)	0.0016	0.0408	0.00	2.00
		lane_closure (count)	0.0009	0.0306	0.00	2.00
		Breakdown (count)	0.0064	0.0826	0.00	4.00
		Hazard (count)	0.0001	0.0106	0.00	1.00
		Rural (0/1)	0.5083	0.4999	0.00	1.00
		vc_ratio (number)	0.2815	0.1460	0.03	0.99
		Par_lane (0/1)	0.0169	0.1291	0.00	1.00
		heavy_percent (percent)	17.2953	7.7183	2.20	33.15
		SSP (0/1)	0.4507	0.4976	0.00	1.00

TMC = Traffic Message Channel; PDO = property damage only crash; SSP = safety service patrol.

Table 9. Input Data Summary for Waze Interchange Model

Group	TMC No. (Miles)	Variable	Mean	Std. Dev.	Minimum	Maximum
Interchange-Group 1	345 (163.23)	Miles (mile)	0.4784	0.2742	0.11	2.02
		throu_lane (count)	2.9360	0.8362	2.00	6.00
		frozen_precip (inches to hundredths)	0.0011	0.0260	0.00	1.44
		rain_precip (inches to hundredths)	0.0506	0.3827	0.00	20.13
		WAZE_crashes (count)	0.0508	0.3107	0.00	15.00
		WAZE_breakdown (count)	0.3867	0.9190	0.00	19.00
		WAZE_hazard (count)	0.0677	0.3909	0.00	64.00
		shoulder_closure (count)	0.0024	0.0510	0.00	3.00
		lane_closure (count)	0.0018	0.0444	0.00	3.00
		Rural (0/1)	0.0400	0.1960	0.00	1.00
		vc_ratio (number)	0.4855	0.1997	0.03	1.62
		Par_lane (0/1)	0.1304	0.3368	0.00	1.00
		heavy_percent (percent)	6.5769	6.0541	0.30	33.15
		SSP (0/1)	0.6162	0.4863	0.00	1.00
Interchange-Group 2	566 (244.93)	Miles (mile)	0.5184	0.2221	0.11	1.73
		throu_lane (count)	2.3731	0.6351	2.00	5.00
		frozen_precip (inches to hundredths)	0.0007	0.0150	0.00	1.36
		rain_precip (inches to hundredths)	0.0380	0.3109	0.00	20.13
		WAZE_crashes (count)	0.0112	0.1448	0.00	7.00
		WAZE_breakdown (count)	0.2226	0.6549	0.00	15.00
		WAZE_hazard (count)	0.0380	0.3655	0.00	64.00
		shoulder_closure (count)	0.0016	0.0408	0.00	2.00
		lane_closure (count)	0.0009	0.0306	0.00	2.00
		Rural (0/1)	0.5083	0.4999	0.00	1.00
		vc_ratio (number)	0.2815	0.1460	0.03	0.99
		Par_lane (0/1)	0.0169	0.1291	0.00	1.00
		heavy_percent (percent)	17.2953	7.7183	2.20	33.15
		SSP (0/1)	0.4507	0.4976	0.00	1.00

TMC = Traffic Message Channel; SSP = safety service patrol.

Variables with p-value < 0.1 for each model were ranked from highest to lowest according to the values of standardized coefficients, as shown in Figures 17 and 18. In contrast to the trend for freeway segments, the three percentiles of travel time distributions of Group 1 interchange segments (more reliable) were impacted by fewer factors than Group 2 (less reliable). Because there was not much length difference between the two groups and interchange segments are much shorter than freeway segments, there was less of a smoothing effect. More significant impact factors contribute to less reliable segments. This also further explains why segment length is included as one of the independent variables and travel times instead of travel rates are used as the target values. For example, if two crashes happened on each of a one-half-mile interchange and a 5-mile freeway segment, the maximum distance between the respective crashes could be 10 times longer for the freeway than the interchange segment. Hence, the impacts on travel times are most likely to be more significant for the latter.

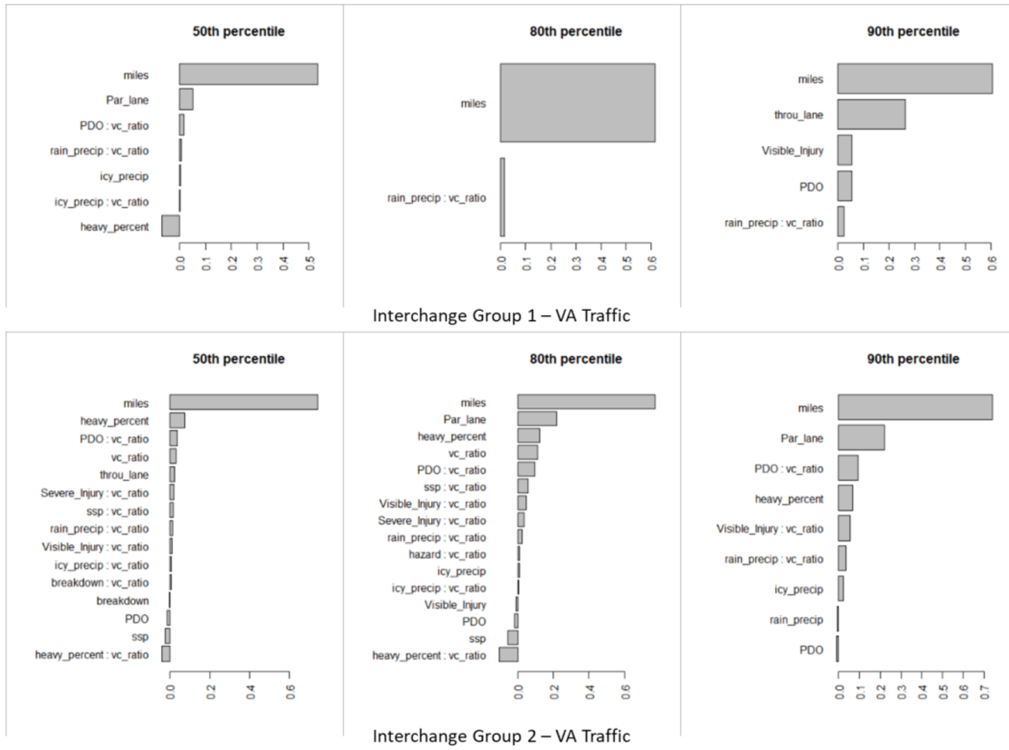


Figure 17. VaTraffic Interchange Variable Coefficient Ranking. X-axis = Standardized Coefficients.

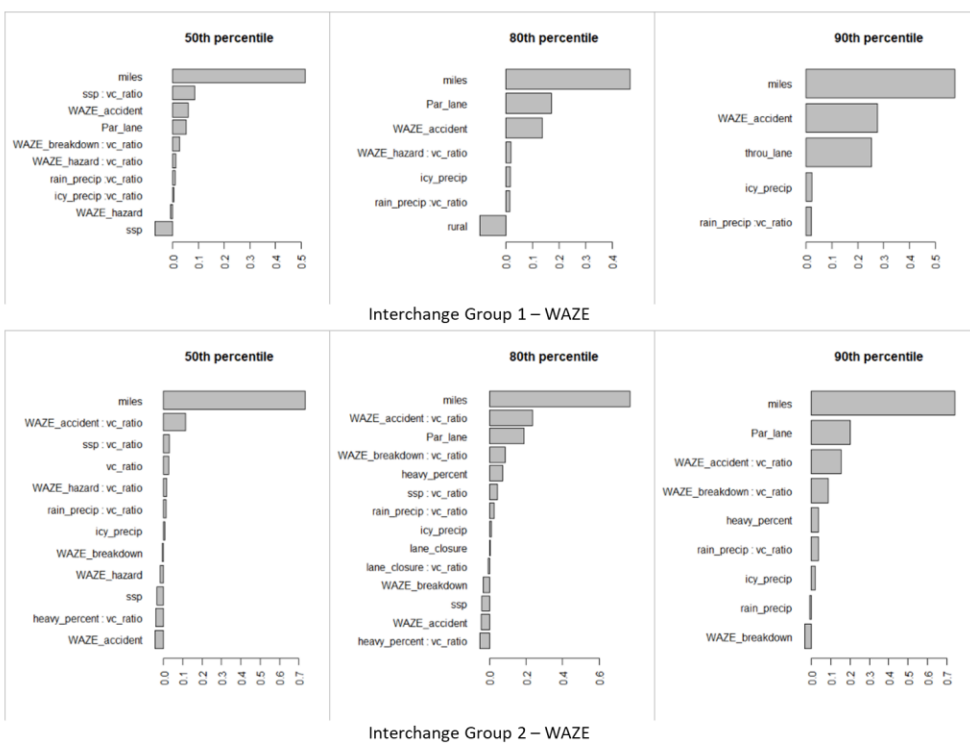


Figure 18. Waze Interchange Variable Coefficient Ranking. X-axis = Standardized Coefficients.

The longer the segments, the better ability they have to absorb some of the impacts. A similar trend previously mentioned about how the magnitude of the same variables increased with percentiles was also observed for both the VaTraffic and Waze models.

Comparing the VaTraffic models of different clusters, it was found that Group 1 was impacted mostly by more common, lower severity events, such as PDO crashes or rain. Traffic demand variables were also significant only as interaction terms. However, for Group 2, in addition to the influencing factors mentioned for Group 1, more extreme events (e.g., crashes with severe injury and snow) also showed significant impacts. Traffic demand variables were not only included as significant variables, but they were also among the highest coefficients as main factors. It is worth mentioning that the presence of SSP reduced 50th and 80th percentile travel times, indicating the reliability benefits of such service on unreliable interchange segments. Similarly, with freeway VaTraffic models, incident variables still had more impact than weather, and work zone variables were not included in any models. This observation could also be due to the relatively low frequency of these variables. The results from the Waze models were mostly intuitive but less interpretable than those from the VaTraffic models, especially for Group 2.

A comparison of the VaTraffic and Waze models showed that VaTraffic models provided more consistent interpretations of factors than Waze models. The main reasons that could explain such an observation are as follows:

- The Waze data used for modeling here were raw data that had not been validated or cleaned. They were collected directly from drivers' reports. Depending on different surrounding situations (e.g., duration of events, traffic volume, speeds, number of Waze users), the number of duplicate reports could vary significantly. Also, it is difficult to know if two events that happened on adjacent but separate TMC segments were duplicated. Although such duplication could be a potential way to reflect effects beyond segments where events actually occurred, without further studies on data validation, they were treated as individual events that might lead to an exaggeration of the impacts.

- The definitions of the non-recurrent incident/hazard events are different for VaTraffic and Waze data as discussed in the "Methods" section. Specifically, breakdown events in VaTraffic data refer to disabled vehicles, and hazard events usually involve objects (vehicle / roadside vegetation) on fire. These data are collected by people with proper training, whereas crowdsourcing data come from drivers' or passengers' subjective judgment.

- The number of Waze reports and users is not necessarily constant spatially. It could be proportional to traffic volume, number of Waze users, level of congestion, or population, which could impact the model results since the variable coefficients increase when the corresponding events occur more frequently.

Although the prediction accuracy of the VaTraffic and Waze models was similar, the VaTraffic models were selected as the preferred models because VaTraffic data are more stable over time. The number of Waze reports could increase as more users adopt the application, which may create additional requirements for re-estimation of the models as the data quality changes over time. The following sections provide the analysis of only VaTraffic models. Because the clustering results of freeway and interchange segments are both separated into the more/less reliable categories, models combining reliable and less reliable segments (regardless of freeway segment / interchange classification) were also analyzed. However, the prediction accuracy decreased in these unified models, and it was also difficult to identify patterns to explain impact factors. As a result, the models were kept separate as freeway segments and interchange segments in subsequent analyses.

Model Performance and Validation

The VaTraffic models were estimated again using only significant variables for prediction performance validation. Four measures were used to quantify the prediction accuracy: MAE, MSE, MAPE, and bias. Table 10 summarizes these measures using both testing data (2019) and training data (2017-2018) by 5-fold cross-validation. The resulting models often overestimated at a constant value, so bias was subtracted from predictions to correct the overestimation. Values in Table 10 with an asterisk indicates that the predicted values were adjusted by bias. In general, predictions using cross-validation training data performed better than those using testing data. This was a common observation since the training data used for prediction were drawn from the same dataset, thus drawing from the same distribution, to build the models.

Table 10. Linear Quantile Mixed Model Performance Validation

2019						2018 Cross-Validation				
	50th Percentile						50th Percentile			
	Freeway		Interchange				Freeway		Interchange	
	Group 1	Group 2	Group 1	Group 2			Group 1	Group 2	Group 1	Group 2
MAE	9.25	6.99*	10.16	3.7		MAE	6.79	2.82	5.46	1
MSE	318.41	197.82*	316.94	48.48		MSE	237.85	38.53	144.62	9.57
MAPE	6.63	26.88*	33.67	13.43		MAPE	5.04	9.09	15.1	3.4
Bias	5.37	3.21	3.76	-0.97		Bias	6.76	2.68	5.32	0.99
	80th Percentile						80th Percentile			
	Freeway		Interchange				Freeway		Interchange	
	Group 1	Group 2	Group 1	Group 2			Group 1	Group 2	Group 1	Group 2
MAE	13.94*	9.3*	12.96*	4.43		MAE	10.35*	4.95*	10.12*	1.30*
MSE	631.51*	268.33*	448.67*	62.39		MSE	428.49*	84.53*	291.37*	16.11*
MAPE	7.95*	28.47*	37.68*	15.28		MAPE	5.58*	12.70*	28.03*	3.50*
Bias	15.55	6.93	6.66	0.94		Bias	16.57	6.23	8.47	2.98
	90th Percentile						90th Percentile			
	Freeway		Interchange				Freeway		Interchange	
	Group 1	Group 2	Group 1	Group 2			Group 1	Group 2	Group 1	Group 2
MAE	22.50*	12.07*	19.32*	5.52*		MAE	19.81*	8.18*	16.65*	6.43*
MSE	1610.6*	402.53*	992.73*	155.68*		MSE	1370.36*	193.04*	770.67*	152.79*
MAPE	11.47*	33.60*	52.58*	16.28*		MAPE	9.78*	20.40*	43.94*	21.03*
Bias	28.81	11.23	12.14	-13.9		Bias	29.89	10.41	13.07	10.27

Asterisks indicate results where models were adjusted to correct for bias. MAE = mean absolute error; MSE = mean square error; MAPE = mean absolute percent error.

Also, predictions of lower percentiles had better accuracy than those of higher percentiles since the data at higher percentiles became less stable. Models of freeway segments provided better predictions than those of interchange segments. This was also as expected since interchange segments have shorter lengths than freeway segments. They are more vulnerable to input data inaccuracy such as mismatched non-recurrent events. Alternatively, the traffic conditions are often more complicated at interchange segments than freeway segments. It is possible that other factors or events happening on adjacent ramps could also impact interchange segments, which were not included in the models. Especially, factors representing the traffic conditions of the adjacent ramps are expected to have impacts on interchange reliability. However, ramp volumes and other ramp event data are not available. The worse performance for interchange segments may be primarily due to the inability to account fully for ramp flows.

Trend line prediction is a common method used by agencies for performance measure prediction. Trend line predictions were created using the 2-year data from 2017 to 2018 to build a fitted line through linear regression for the 50th, 80th, and 90th percentiles for each TMC. Predictions of these three percentiles in 2019 were made using the fitted lines. To illustrate the improvements the estimated models provide, the prediction performance was also compared with trend line predictions. Two measures were used to quantify the prediction errors. One was the actual time differences between actual values and predictions, and the other was the time differences divided by the actual percentile to obtain the prediction errors as a percentage. Figures 19 and 20 show the histogram comparison of these two measures from LQMMs and the trend line method. The red bars are the distribution of prediction errors using the developed models, and the green bars show the trend line method. The histograms of the prediction errors from the developed models have a narrower deviation and higher peaked values around 0, meaning a better prediction performance than the trend line method. Figures 21 and 22 compare the predicted values with the actual values. The y-axis of figures on the upper row is Predicted 50th, 80th, and 90th Percentiles Using LQMM, and the y-axis of figures on the lower row is Predicted 50th, 80th, and 90th Percentiles Using the Trend Line Approach. The closer these points are to the 45-degree line, the closer the prediction values to the actual ones. Overall, the points on the upper row figures are obviously closer to that line, indicating a better match between actual values and predicted values with the developed models than the trend lines.

Random Forest Models for Travel Time Reliability Predictions

Clustering Necessity Analysis

It was speculated earlier that clustering TMC segments into homogeneous groups with similar shapes of distributions might not be necessary when using random forests. To test this speculation, an initial step in this study was to develop, and compare the results for, random forest models based on both the entire freeway and interchange TMC segment groups and their respective clustered groups. The input data and division of the training and testing datasets were the same as for the LQMMs using VaTraffic data introduced previously. Generalized random forest (GRF) models were built using the R package “GRF” (Tibshirani et al., 2021). The trained models consisting of 2,000 trees were saved as quantile forest objects in R and were ready to make predictions using new input data. It was decided to use 5, 8, 11, 14, and 17 (the total number of variables) as the $mtry$ options discussed previously.

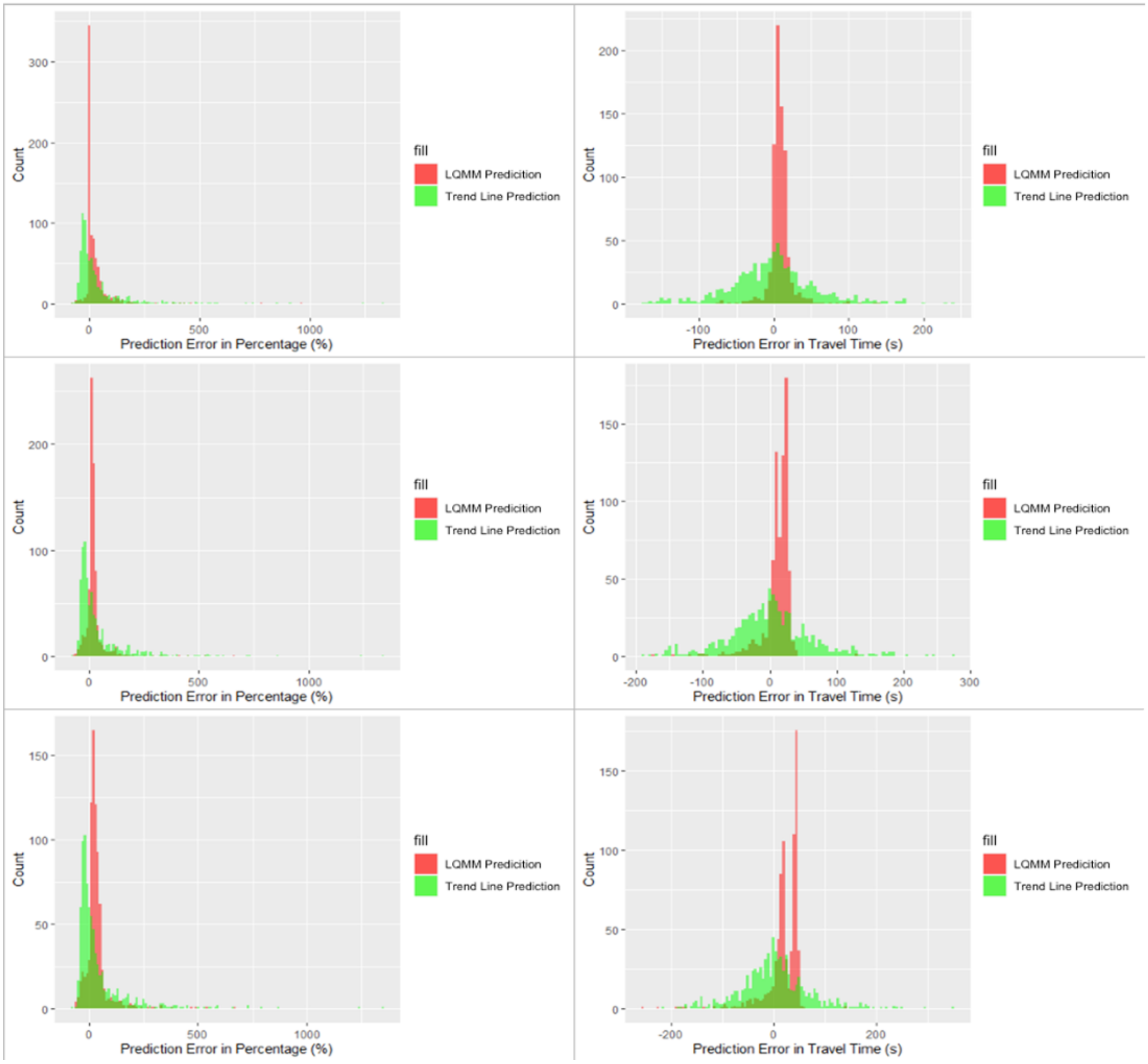


Figure 19. Freeway Prediction Error Comparison Between Linear Quantile Mixed Models and Trend Line. From top to bottom: 50th, 80th, and 90th percentile of travel times. The red bars are the distribution of prediction errors using the developed models, and the green bars show the trend line method.

The results of $mtry$ values of 5 and 8 have minimal differences, as do values of 11 and 14. Therefore, only the results of *mtry* values 5, 11, and 17 are included in Tables 11 and 12 for the sake of brevity. These two tables summarize the prediction performance for the 50th, 80th, and 90th percentiles of travel times for freeway and interchange segments, respectively. The highlighted cells indicate the best performance of models with different $mtry$ values, which was used for comparison with other groups.

For freeway segments, the model of the entire group had a performance similar to that of clustering Group 2 and better than that of clustering Group 1. For interchange segments, the trend was the opposite. This phenomenon indicated that clustering TMCs into homogeneous distribution shapes did not significantly impact the prediction performance of random forest models.

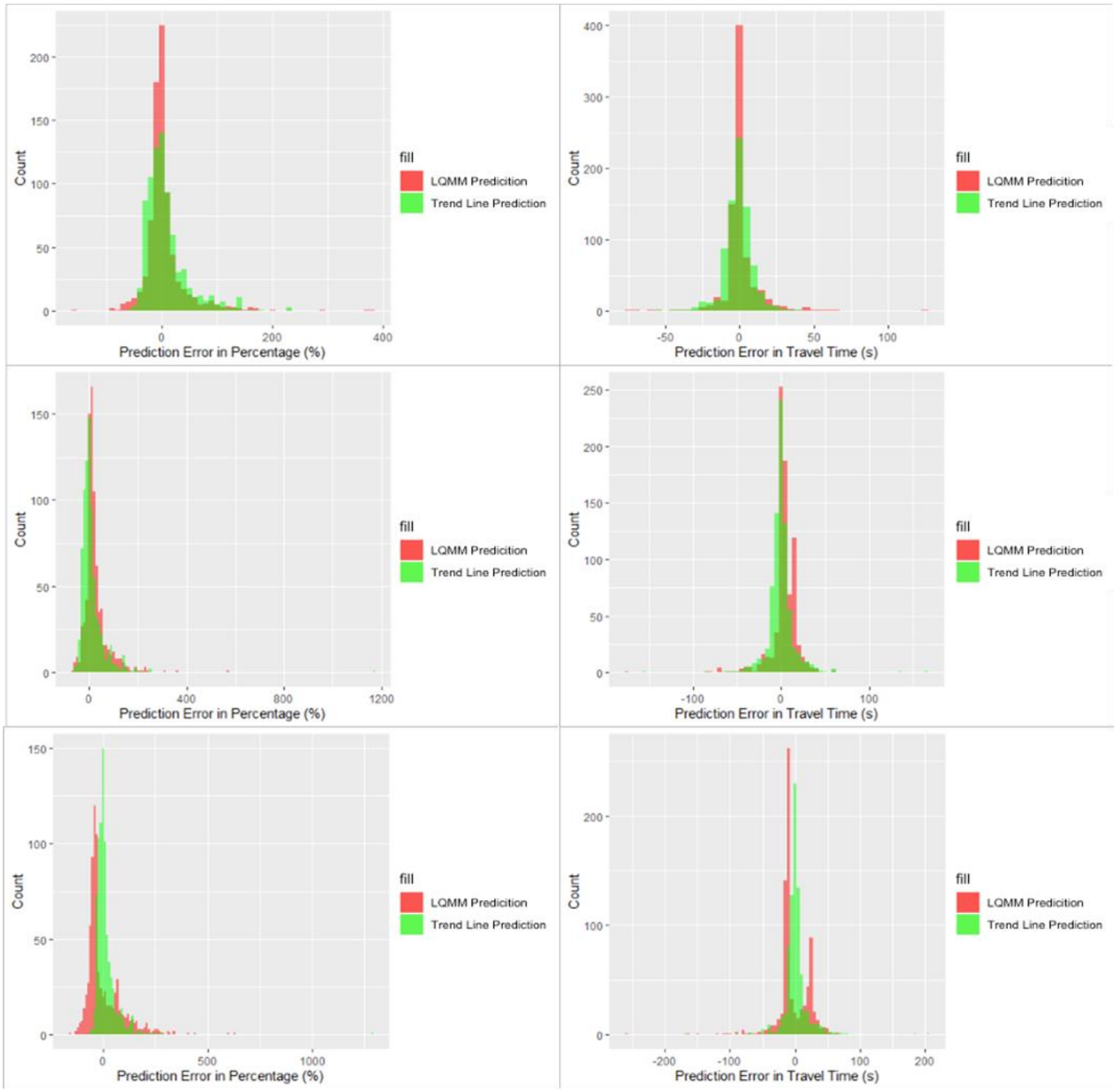


Figure 20. Interchange Prediction Error Comparison Between Linear Quantile Mixed Models and Trend Line. From top to bottom: 50th, 80th, and 90th percentile of travel times. The red bars are the distribution of prediction errors using the developed models, and the green bars show the trend line method.

It could either improve prediction when the models of the entire group and clustering Group 2 of both freeway and interchange segments were compared or decrease accuracy significantly when the models of the entire group and clustering Group 1 were compared. As a result, the clustering process is not necessary when the random forest models.

GRF Analysis Results

GRF models were built using *mtry* values of 5, 11, and 17 to predict the 50th, 80th, and 90th percentiles of travel times for all freeway and interchange segments. The prediction performance is shown in Table 13.

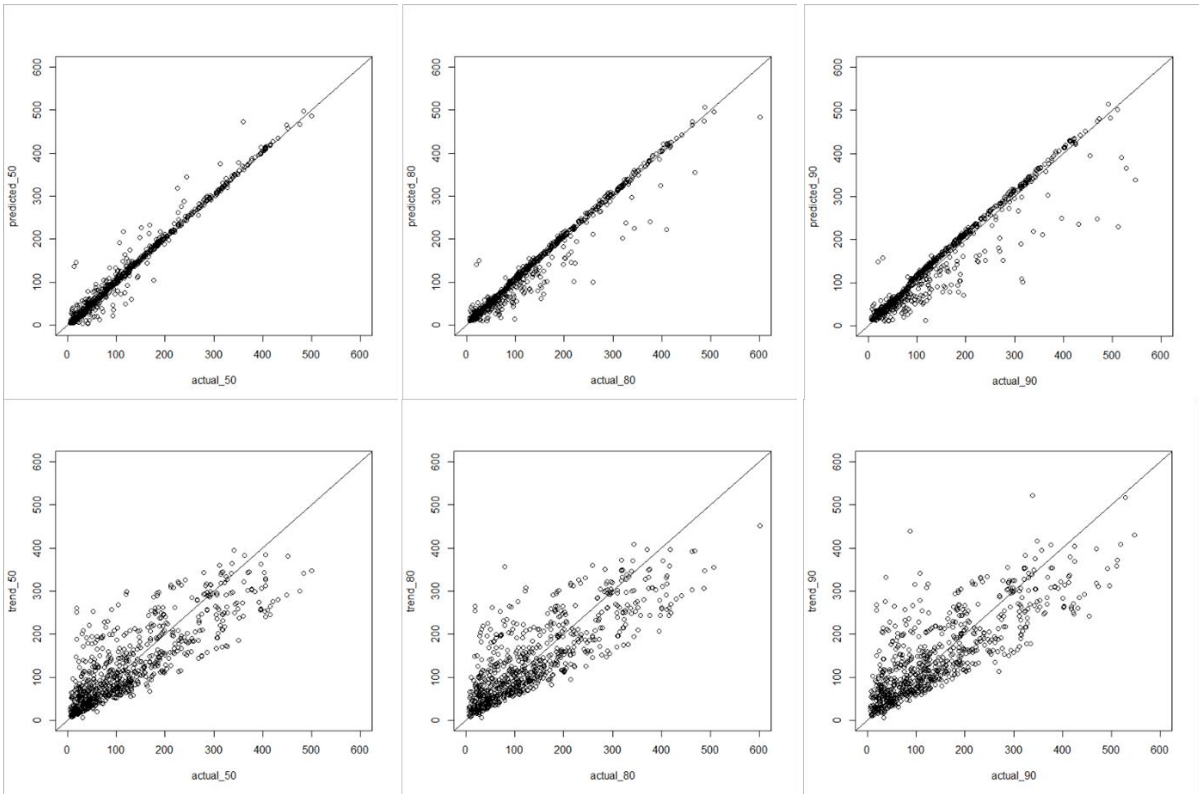


Figure 21. Predicted vs. Observed Travel Time Percentiles for Freeways: top, linear quantile mixed models; bottom, trend line.

Almost all four performance measures indicated that with the increase of $mtry$ values, the accuracy also increases. The same trend could be observed for both OOB and testing data. It was noticed that the OOB predictions sometimes had lower accuracy than the testing data. This is not uncommon for random forests using the "honest" method, which uses a portion of the sample obtained from the previous step for node splits and populates the remaining portion as observations in the leaf nodes. The bias of OOB predictions is reduced through this process. It also should be mentioned that since the value of $mtry$ used by GRF is drawn from a Poisson distribution with a mean equal to $mtry$, it is likely to change for each node splitting. As a result, even when set to the maximum number 17, the models should still maintain a certain level of stability.

After the optimal GRF models with $mtry$ equal to 17 were selected, the prediction results were compared to LQMMs. A similar comparison used previously was also used here, which is the histogram of the difference between predicted and actual values and the percentage of such difference. Figures 23 and 24 summarize the prediction errors using these two measures for freeway and interchange segments. It is clearly shown that GRF models reduced the prediction errors as compared to LQMMs. The histograms in green (GRF) are narrower, and their peaks are closer to zero than the red ones (LQMM), indicating a smaller deviation of errors and a lower value of mean errors.

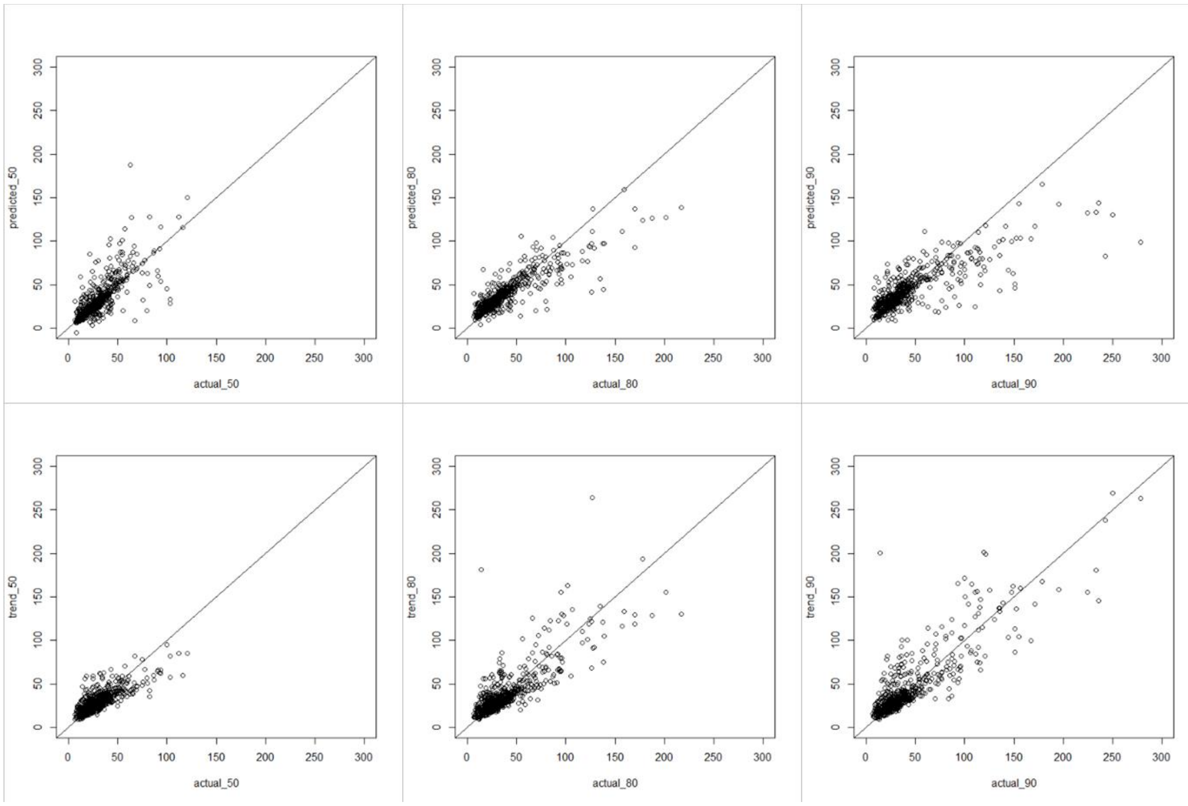


Figure 22. Predicted vs. Observed Travel Time Percentiles for Interchanges: top, linear quantile mixed models; bottom: trend line.

Table 11. Model Performance Comparison of Entire Dataset and Clustered Groups of Freeway Segments

	Freeway			**Freeway-Group 1**			**Freeway-Group 2**		
	50th								
	mtry=5	*mtry*=11	*mtry*=17	*mtry*=5	*mtry*=11	*mtry*=17	*mtry*=5	*mtry*=11	*mtry*=17
MAE	14.53	6.18	6.15	17.77	9.51	9.61	5.66	3.77	3.69
MSE	922.63	143.23	140.83	936.20	226.35	262.01	168.56	113.76	110.88
MAPE	21.39	7.28	6.73	15.45	9.64	10.49	20.10	9.91	9.39
Bias	-6.43	-0.35	-0.48	-2.09	5.39	5.07	-1.53	-0.82	-1.08
	80th								
	mtry=5	*mtry*=11	*mtry*=17	*mtry*=5	*mtry*=11	*mtry*=17	*mtry*=5	*mtry*=11	*mtry*=17
MAE	45.34	16.35	14.22	52.48	17.81	16.40	15.43	9.15	9.11
MSE	2848.02	885.64	886.50	3685.52	1001.58	982.33	495.33	357.00	379.04
MAPE	85.19	18.83	14.68	40.03	13.25	12.68	56.15	22.36	21.56
Bias	34.70	3.62	-0.11	40.94	7.02	4.18	7.44	-0.64	-1.16
	90th								
	mtry=5	*mtry*=11	*mtry*=17	*mtry*=5	*mtry*=11	*mtry*=17	*mtry*=5	*mtry*=11	*mtry*=17
MAE	92.76	28.34	22.14	102.42	27.74	23.48	24.95	14.19	13.80
MSE	10496.43	1779.16	1555.32	12605.24	1899.52	1833.35	968.73	573.06	606.38
MAPE	164.43	33.68	25.21	74.20	18.62	16.32	89.58	37.21	34.91
Bias	84.46	16.02	7.17	91.69	16.02	9.43	16.34	3.85	2.89

Orange highlighted cells represent the best performing value of *mtry*. MAE = mean absolute error; MSE = mean square error; MAPE = mean absolute percent error.

Table 12. Model Performance Comparison of Entire Dataset and Clustered Groups of Interchange Segments

	Interchange			Interchange-Group 1			Interchange-Group 2		
	50th								
	$mtry$=5	$mtry$=11	$mtry$=17	$mtry$=5	$mtry$=11	$mtry$=17	$mtry$=5	$mtry$=11	$mtry$=17
MAE	3.40	2.32	2.31	6.18	4.02	4.18	3.04	1.15	1.02
MSE	54.17	29.29	33.61	122.70	62.63	69.88	43.96	3.75	3.85
MAPE	13.86	8.28	7.65	25.27	13.18	12.81	14.89	5.02	4.12
Bias	-0.72	-0.20	-0.78	-1.61	-0.68	-0.63	0.24	0.74	0.55
	80th								
	$mtry$=5	$mtry$=11	$mtry$=17	$mtry$=5	$mtry$=11	$mtry$=17	$mtry$=5	$mtry$=11	$mtry$=17
MAE	10.35	7.23	7.48	20.36	14.00	15.18	6.49	2.35	2.06
MSE	369.43	275.40	345.89	896.79	555.90	694.36	79.66	31.45	34.22
MAPE	38.28	19.55	17.34	74.99	37.20	37.16	32.34	8.30	6.34
Bias	1.86	-0.69	-1.95	2.52	-1.01	-0.98	4.55	0.98	0.32
	90th								
	$mtry$=5	$mtry$=11	$mtry$=17	$mtry$=5	$mtry$=11	$mtry$=17	$mtry$=5	$mtry$=11	$mtry$=17
MAE	16.87	11.19	11.73	30.75	22.24	23.73	11.66	3.77	3.27
MSE	586.75	455.37	496.10	1416.72	1023.58	1271.65	195.80	63.18	69.08
MAPE	64.37	34.56	33.83	119.70	69.34	68.61	53.39	13.39	9.49
Bias	8.97	3.29	2.59	14.92	8.15	7.94	9.79	1.90	0.90

Orange highlighted cells represent the best performing value of *mtry*. MAE = mean absolute error; MSE = mean square error; MAPE = mean absolute percent error.

Before-After Studies to Demonstrate Model Application

The main purpose of the developed models is to assist in target setting, which requires that the models have the ability to predict reliability considering any changes in the impact factors of interest. Three projects encompassing different types of transportation improvement projects suggested and selected by the TRP were used for before-after case studies of model application.

Project 1: I-81 Corridor Safety Service Patrol (SSP)

Starting July 1, 2019, there was increased SSP coverage on I-81 both temporally and spatially. Ideally, at least 1 year of data from the before and the after periods are needed for evaluation since LQMMs were built based on clusters of homogenous shapes of distributions constructed using a full year's data. However, the traffic demand reduced significantly because of the COVID stay home order after March 2020. Assuming travel times on interstate highways would most likely be reliable following COVID-induced traffic reductions, including data from 2020 would not provide many valuable insights. The after period was then selected as July 2019 to December 2019, for a total of 6 months. To keep the duration similar, the before period was selected as January 2019 to June 2019.

Table 13. Prediction Performance of GRF Models

OOB						
50th Percentile						
	Freeway			**Interchange**		
	$mtry$=5	$mtry$=11	$mtry$=17	$mtry$=5	$mtry$=11	$mtry$=17
MAE	15.53	5.90	5.78	3.77	2.82	2.69
MSE	1038.68	168.95	157.18	77.24	52.28	52.75
MAPE	22.60	7.33	6.81	12.21	8.15	7.50
Bias	-8.23	-2.46	-2.56	-1.76	-0.79	-1.45
80th Percentile						
	$mtry$=5	$mtry$=11	$mtry$=17	$mtry$=5	$mtry$=11	$mtry$=17
MAE	46.51	17.91	15.63	10.67	8.30	8.55
MSE	2916.21	1004.39	987.24	440.60	370.49	460.19
MAPE	90.83	21.09	16.86	32.88	18.70	17.38
Bias	34.60	3.01	-1.00	1.07	-1.21	-2.52
90th Percentile						
	$mtry$=5	$mtry$=11	$mtry$=17	$mtry$=5	$mtry$=11	$mtry$=17
MAE	93.89	31.15	24.58	17.62	13.09	13.53
MSE	10637.82	2089.17	1829.77	788.35	754.10	782.34
MAPE	173.72	37.82	29.35	56.02	31.54	31.42
Bias	83.94	15.85	6.98	8.26	2.70	2.17
Test Data						
50th Percentile						
	$mtry$=5	$mtry$=11	$mtry$=17	$mtry$=5	$mtry$=11	$mtry$=17
MAE	14.53	6.18	6.15	3.40	2.32	2.31
MSE	922.63	143.23	140.83	54.17	29.29	33.61
MAPE	21.39	7.28	6.73	13.86	8.28	7.65
Bias	-6.43	-0.35	-0.48	-0.72	-0.20	-0.78
80th Percentile						
	$mtry$=5	$mtry$=11	$mtry$=17	$mtry$=5	$mtry$=11	$mtry$=17
MAE	45.34	16.35	14.22	10.35	7.23	7.48
MSE	2848.02	885.64	886.50	369.43	275.40	345.89
MAPE	85.19	18.83	14.68	38.28	19.55	17.34
Bias	34.70	3.62	-0.11	1.86	-0.69	-1.95
90th Percentile						
	$mtry$=5	$mtry$=11	$mtry$=17	$mtry$=5	$mtry$=11	$mtry$=17
MAE	92.76	28.34	22.14	16.87	11.19	11.73
MSE	10496.43	1779.16	1555.32	586.75	455.37	496.10
MAPE	164.43	33.68	25.21	64.37	34.56	33.83
Bias	84.46	16.02	7.17	8.97	3.29	2.59

Red font indicates the value of *mtry* with the best results. OOB = out-of-bag sample; MAE = mean absolute error; MSE = mean square error; MAPE = mean absolute percent error.

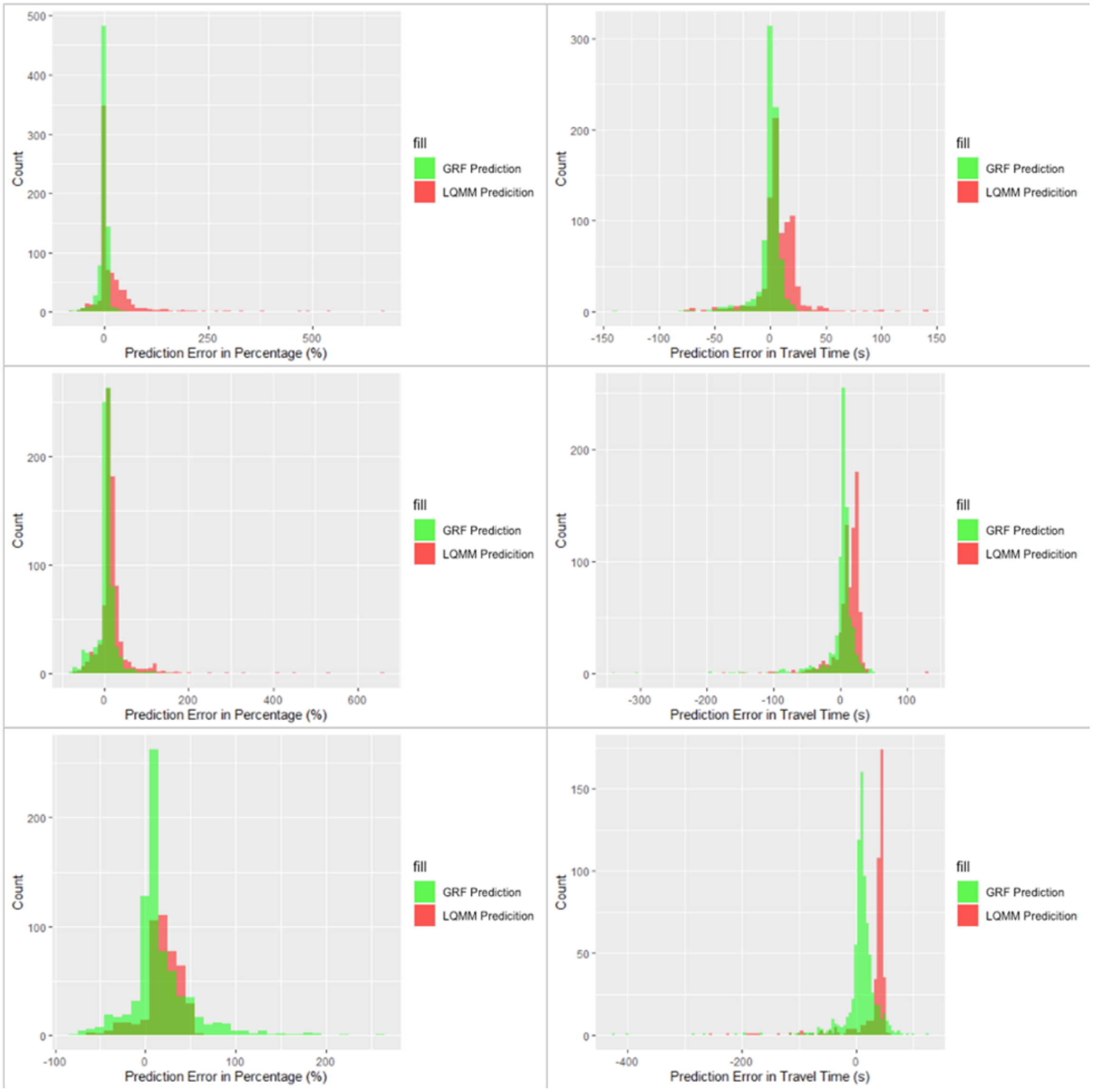


Figure 23. Comparison of Generalized Random Forests and Linear Quantile Mixed Models for Freeways. From top to bottom: 50th, 80th, and 90th percentiles.

Since SSP presence was not a significant variable in LQMMs, only GRF models were applicable. Also, since the changes in SSP coverage and schedule are planned in advance and the before and after periods are within the same year, there is no significant difference between actual and projected after-period data. So, the analysis of the after period used only the actual data as input for predictions and was then compared with that of the before period.

It was found by comparing the observed reliability measures of the before and after periods that the differences in reliability created by this project were very small in terms of LOTTR and the 90th percentile travel times. Specifically, for freeway segments, the average "after" LOTTR increased 0.001 (about 0.1%) and the average 90th percentile decreased 13.4 seconds (about 5.5%). For interchange segments, the average "after" LOTTR increased 0.002 (about 0.2%) and the average 90th percentile decreased 1.0 second (about 4.9%).

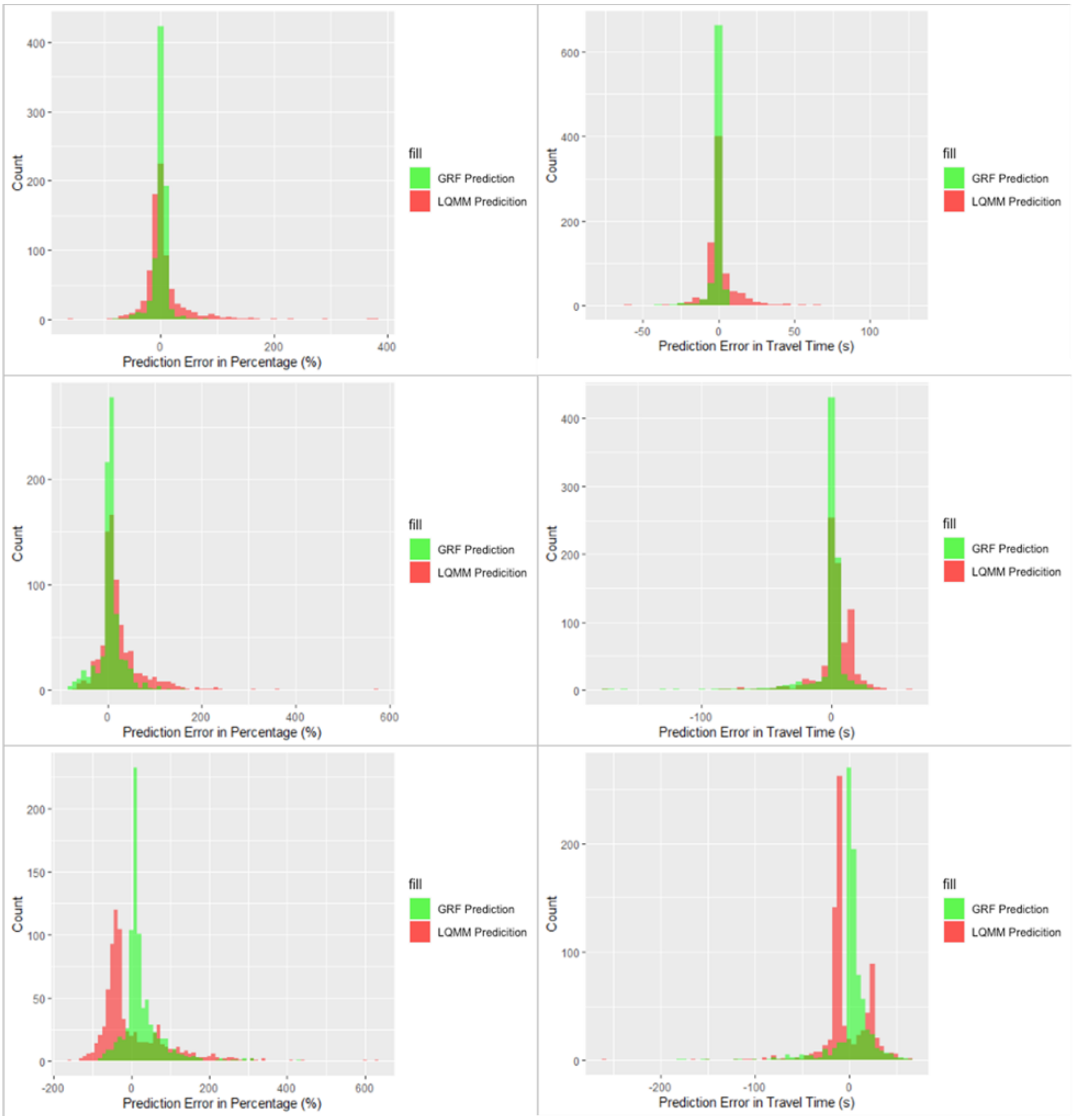


Figure 24. Comparison of Generalized Random Forests and Linear Quantile Mixed Models for Interchanges. From top to bottom: 50th, 80th, and 90th percentiles.

After the GRF models were applied, the changes between predicted reliability measures for the after period and the actual values of the before period were as follows:

1. *Freeway segments.* The predicted LOTTR of the after period was 0.06 higher than the actual value of the before period; the predicted 90th percentile was 22.80 seconds higher than the actual value of the before period.

2. *Interchange segments.* The predicted LOTTR of the after period was 0.04 higher than the actual value of the before period; the predicted 90th percentile was 1.36 seconds higher than the actual value of the before period.

The prediction accuracy is summarized in Table 14. The predicted difference in LOTTR for freeway and interchange segments was also minimal, aligning with the actual observation. There was not much change in the 90th percentile on interchange segments, and the model predictions indicated the same trend. Although the predicted difference in the 90th percentile for freeway segments was the opposite of the observations, if the bias of an average overestimation of 36.174 seconds listed in Table 14 is considered, the model could still capture a slight decrease in the after period.

Overall, the reliability did not improve much between the before and after periods, and the proposed models drew a similar conclusion. It should be noted that although this project did not create major changes in reliability, the benefits of SSP might lay in some other aspects, such as reduced secondary crashes or clearance durations that are outside the impact factors of this study.

Table 14. Model Prediction Accuracy of I-81 Safety Service Patrol Project

	Freeway		Interchange	
	LOTTR	**90th Percentile**	**LOTTR**	**90th Percentile**
MAE	0.058	36.174	0.038	2.397
MSE	0.005	1769.575	0.002	6.050
MAPE	5.770	17.483	3.747	12.967
Bias	0.058	36.174	0.038	2.397
Actual Difference	0.001	-13.400	0.002	-1.000
Predicted Difference	0.060	22.800	0.040	1.360

LOTTR = level of travel time reliability; MAE = mean absolute error; MSE = mean square error; MAPE = mean absolute percent error.

Project 2: Capacity Expansion on I-64

The widened section on I-64 was between Exit 200 (I-295 interchange) and Exit 205 (Bottoms Bridge). For the before-after analysis, interchanges at both ends were also considered in the analysis, as shown in Figure 25. Table 15 lists the six segments and the cluster group to which they belonged for the LQMM analysis. Freeway segments in Group 1 were more reliable, and interchange segments in Group 2 were less reliable.

The projected input data for the after and during construction periods were prepared in accordance with to the method discussed previously. Both the LQMMs and GRF models were applied for freeway and interchange segments. The prediction accuracy of the LQMMs and GRF models is summarized in Table 16. Predictions using the actual data of the after and during (work zone) periods were evaluated separately along with predictions using the projected data. The accuracy measures are the average of the after and during periods in this table. It could be concluded from the results that predictions of LOTTRs have better accuracy than the 90th percentile for both the LQMMs and GRF models.

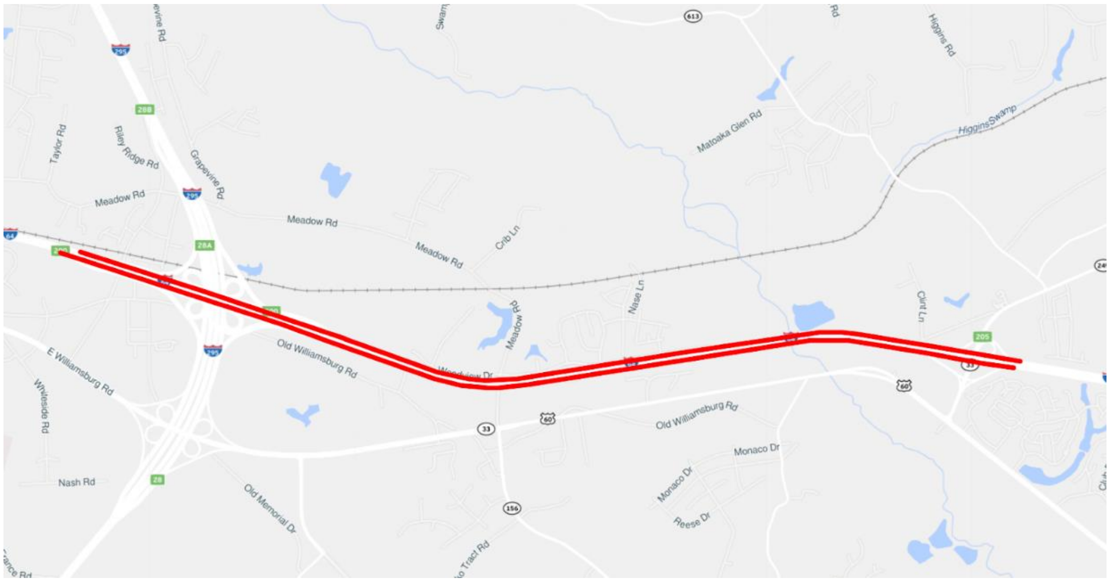


Figure 25. Traffic Message Channel Segments Used for I-64 Capacity Expansion Project

Table 15. Traffic Message Channel (TMC) Segment Summary for I-64 Capacity Expansion Project

TMC	Direction	Intersection	Cluster Group
110N04903	EASTBOUND	US-60/EXIT 200	interchange-2
110-04902	EASTBOUND	VA-33/VA-249/EXIT 205	freeway-1
110N04902	EASTBOUND	VA-33/VA-249/EXIT 205	interchange-2
110P04903	WESTBOUND	US-60/EXIT 200	interchange-2
110+04903	WESTBOUND	US-60/EXIT 200	freeway-1
110P04902	WESTBOUND	VA-33/VA-249/EXIT 205	interchange-2

Table 16. Prediction Accuracy Summary of Linear Quantile Mixed Models and Generalized Random Forest Models (I-64)

	I-64 Capacity Expansion LQMM							
	Freeway				Interchange			
	LOTTR	90th Percentile	Projected LOTTR	Projected 90th Percentile	LOTTR	90th Percentile	Projected LOTTR	Projected 90th Percentile
MAE	0.016	14.860	0.014	15.836	0.064	6.245	0.072	5.587
MSE	0.000	289.672	0.000	337.708	0.005	46.212	0.007	42.936
MAPE	1.530	6.985	1.404	7.536	6.256	13.654	7.028	12.190
Bias	0.016	1.489	0.014	4.193	0.064	6.245	0.072	2.787
	I-64 Capacity Expansion GRF							
MAE	0.033	15.778	0.044	17.022	0.077	14.548	0.105	17.161
MSE	0.001	355.006	0.002	395.511	0.009	258.439	0.015	333.097
MAPE	3.259	7.566	4.253	8.286	7.572	30.045	10.297	37.404
Bias	0.033	5.195	0.044	9.896	0.077	14.548	0.105	17.161

LQMM = linear quantile mixed model; LOTTR = level of travel time reliability; GRF = generalized random forest; MAE = mean absolute error; MSE = mean square error; MAPE = mean absolute percent error.

Interchange segments were shorter than freeway segments, so the non-scaled error measures (e.g., MAE, MSE, and bias) had higher values for freeway segments. The prediction accuracy was similar when the actual and the projected after/during input data were used. Although the GRF models showed better prediction accuracy across all interstate highways (both freeway and interchange segments), the LQMMs provided better predictions for this project. In addition to the accuracy comparison, it is necessary to examine the reliability changes predicted by the models to determine a model preference.

Tables 17 and 18 summarize the changes of actual (observed), predicted (model predicted using actual after data), and projected predicted (model predicted using projected input data based on methods discussed earlier) reliability metrics, including LOTTR and 90th percentile, for the before, during, and after periods. Table 17 shows the comparisons between the before and after periods, and Table 18 shows the comparisons between the before and during periods. Each section provides the actual (Rows 1 and 6), predicted (Rows 2, 3, 7, and 8), and projected predicted (Rows 4, 5, 9, and 10) reliability changes. The predicted (and projected predicted) differences between the before and after and the before and during periods were evaluated in two ways: (1) the changes between predicted (projected predicted) values of the after/during period and the actual observations of the before period; (2) the changes between predicted (projected predicted) values of the after/during period and the predicted values of the before period. The purpose of using two evaluations was to accommodate different application scenarios. For example, for evaluation of operational projects, the comparison between the actual values and predictions is commonly used, whereas for planning projects, the comparison between both predicted values may be more appropriate.

The main findings were as follows:

- When the results of prediction and projected prediction were compared, there was not much difference in the overall reliability trends. This observation means even when using extrapolation of input variables using the existing data, both the LQMMs and GRF models could still provide results similar to knowing the exact future values of these variables.

- As mentioned previously, the proposed models for interchange segments were less accurate than for freeway segments, directly impacting the before and after study results. Similarly, the prediction accuracy of the 90th percentile was lower than the 50th and 80th percentiles, so the LOTTR evaluation performed better.

- Both models overestimated the after/during period values, with GRF models having larger errors. It was noticed that most cases had minimal reliability change (less than 5.5%). In these cases, LQMMs reflected the trend that traffic was less reliable for the after/during periods closer than GRF models, especially for the 90th percentile. However, if the predicted before and after/during difference was used, the GRF models provided closer results to the actual trend.

Table 17. Before-After Changes in Reliability Metrics (I-64)

	I-64 Freeway				I-64 Interchange			
	LOTTR		90th Percentile		LOTTR		90th Percentile	
	LQMM	GRF	LQMM	GRF	LQMM	GRF	LQMM	GRF
(1) Actual Difference % (before - after)	0.006 (0.609%)		3.541 (1.808%)		0.007 (0.782%)		-3.261 (-5.453%)	
(2) Predicted Difference % (actual before-predict after)	0.026 (2.57%)	0.031 (3.10%)	19.890 (10.16%)	22.939 (11.72%)	0.072 (7.11%)	0.075 (7.45%)	5.022 (8.34%)	13.459 (22.51%)
(3) Predicted Difference % (predict before-predict after)	0.004 (0.40%)	-0.025 (-2.51%)	0.307 (0.16%)	-3.693 (-1.89%)	-0.003 (-0.32%)	-0.028 (-2.76%)	-2.214 (-3.70%)	-3.063 (-5.12%)
(4) Projected Difference % (actual before-projected predicted after)	0.024 (2.40%)	0.043 (4.28%)	23.570 (12.04%)	27.424 (14.01%)	0.083 (8.22%)	0.115 (11.38%)	5.113 (8.55%)	17.061 (28.54%)
(5) Projected Difference % (predict before-projected predicted after)	0.002 (0.23%)	-0.014 (-1.34%)	3.987 (2.04%)	0.791 (0.40%)	0.008 (0.79%)	0.012 (1.17%)	-2.124 (-3.55%)	0.539 (0.90%)

LQMM = linear quantile mixed model; LOTTR = level of travel time reliability; GRF = generalized random forest.

Table 18. Before-During Changes in Reliability Metric (I-64)

	I-64 Freeway				I-64 Interchange			
	LOTTR		90th Percentile		LOTTR		90th Percentile	
	LQMM	GRF	LQMM	GRF	LQMM	GRF	LQMM	GRF
(6) Actual Difference % (before - during)	0.013 (1.327%)		33.059 (16.883%)		0.008 (0.776%)		3.097 (5.180%)	
(7) Predict Difference % (actual before-predicted during)	0.025 (2.45%)	0.055 (5.42%)	19.688 (10.06%)	24.049 (12.28%)	0.072 (7.09%)	0.094 (9.36%)	7.304 (12.22%)	15.474 (25.88%)
(8) Predict Difference % (predicted before-predicted during)	0.003 (0.28%)	-0.002 (-1.90%)	0.105 (0.05%)	-2.583 (-1.32%)	-0.003 (-0.34%)	-0.009 (-0.85%)	0.067 (0.11%)	-1.048 (-1.75%)
(9) Project Difference % (actual before-projected predicted during)	0.024 (2.37%)	0.063 (6.25%)	21.416 (10.94%)	28.968 (14.79%)	0.076 (7.55%)	0.111 (10.94%)	7.287 (12.19%)	17.098 (28.60%)
(10) Project Difference % (predicted before-projected predicted during)	0.002 (1.93%)	0.006 (0.64%)	1.832 (0.94%)	2.335 (1.19%)	0.001 (0.12%)	0.007 (0.74%)	0.049 (0.83%)	0.576 (0.96%)

LQMM = linear quantile mixed model; LOTTR = level of travel time reliability; GRF = generalized random forest.

Project 3: Work Zone Impact: I-64/I-264 Interchange Improvements—Phase II

The project location is shown in Figure 10, and the TMC segments used for comparison are shown in Figure 26. Table 19 summarizes the location information of the relevant TMC segments. All freeway segments belong to cluster Group 2, and all interchange segments belong to cluster Group 1.

The analysis method was the same as for the previous case study. LQMMs of appropriate groups and GRF models were applied respectively using actual and projected data from the during period. The prediction accuracy of these models is summarized in Table 20. For both freeway and interchange segments, GRF models performed better in predicting LOTTR, and LQMMs performed better in predicting the 90th percentile. Again, the before-during reliability changes are shown in Table 21. The interchange segments had a much higher degradation of reliability than freeway segments in the during period. GRF models correctly predicted this trend for LOTTR, whereas LQMMs successfully captured the changes of the 90th percentile.

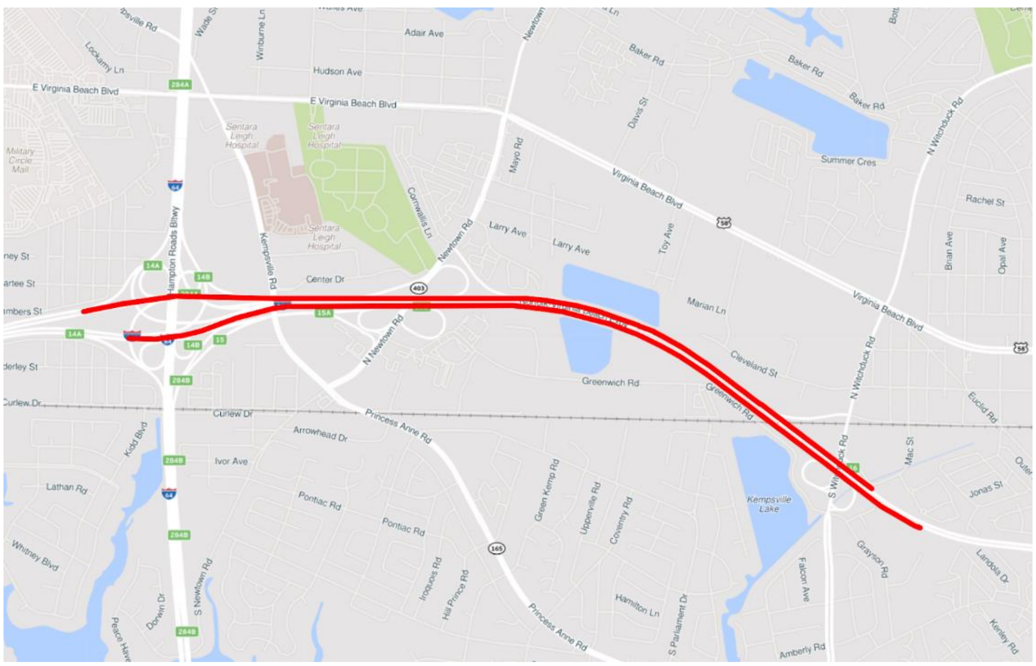


Figure 26. Traffic Message Channel Segments Considered for I-64/I-264 Interchange Improvements

Table 19. Summary of Traffic Message Channel (TMC) Segments for I-64/I-264 Interchange Improvement

TMC	Direction	Intersection	Cluster Group
110P04702	EASTBOUND	I-64/EXIT 14	interchange-1
110+04703	EASTBOUND	GREENWICH RD/NEWTOWN RD	freeway-2
110P04703	EASTBOUND	GREENWICH RD/NEWTOWN RD	interchange-1
110+04704	EASTBOUND	VA-190/WITCHDUCK RD/EXIT 16	freeway-2
110N04702	WESTBOUND	I-64/EXIT 14	interchange-1
110-04702	WESTBOUND	I-64/EXIT 14	freeway-2
110N04703	WESTBOUND	GREENWICH RD/NEWTOWN RD	interchange-1
110-04703	WESTBOUND	GREENWICH RD/NEWTOWN RD	freeway-2
110N04704	WESTBOUND	VA-190/WITCHDUCK RD/EXIT 16	interchange-1

Table 20. Summary of Prediction Accuracy of LQMMs and GRF Models (I-264)

	I-264 Capacity Expansion LQMM							
	Freeway				Interchange			
	LOTTR	90th Percentile	Projected LOTTR	Projected 90th Percentile	LOTTR	90th Percentile	Projected LOTTR	Projected 90th Percentile
MAE	0.115	3.482	0.121	3.167	0.278	5.268	0.334	6.569
MSE	0.014	18.285	0.016	15.603	0.126	55.901	0.151	70.993
MAPE	9.967	7.740	10.547	7.094	23.641	13.323	26.633	15.429
Bias	0.056	-1.645	0.064	-1.294	0.232	5.175	0.314	3.348
	I-264 Capacity Expansion GRF							
MAE	0.044	9.752	0.046	9.363	0.222	21.366	0.210	16.103
MSE	0.002	135.811	0.002	125.508	0.064	492.691	0.070	381.527
MAPE	3.774	24.144	3.930	23.210	16.879	44.304	15.410	32.087
Bias	0.044	9.752	0.046	9.363	0.009	21.366	-0.002	16.103

LQMM = linear quantile mixed model; LOTTR = level of travel time reliability; GRF = generalized random forest; MAE = mean absolute error; MSE = mean square error; MAPE = mean absolute percent error.

Table 21. Before-During Changes in Reliability Metric (I-264)

	I-264 Freeway				I-264 Interchange			
	LOTTR		90th Percentile		LOTTR		90th Percentile	
	LQMM	GRF	LQMM	GRF	LQMM	GRF	LQMM	GRF
Actual Difference % (before - during)	0.044 (3.939%)		0.615 (1.429%)		0.280 (24.868%)		7.338 (17.583%)	
Predicted Difference % (actual before-predicted during)	0.100 (8.93%)	0.088 (7.84%)	-1.030 (-2.39%)	10.367 (24.10%)	0.513 (45.49%)	0.289 (25.66%)	12.514 (29.98%)	28.704 (68.78%)
Predicted Difference % (predicted before-predicted during)	-0.003 (-0.30%)	-0.002 (-0.17%)	-0.081 (-0.19%)	0.387 (0.89%)	-0.095 (-8.42%)	0.011 (1.01%)	2.111 (5.06%)	5.263 (12.61%)
Projected Difference % (actual before-projected predicted during)	0.108 (9.66%)	0.090 (8.00%)	-0.679 (-1.58%)	9.978 (23.19%)	0.595 (52.77%)	0.278 (24.65%)	10.687 (25.61%)	23.442 (56.17%)
Projected Difference % (predicted before-projected predicted during)	0.004 (0.42%)	0 (0%)	0.270 (0.63%)	-0.001 (0%)	-0.013 (-1.14%)	0 (0%)	0.284 (0.68%)	0 (0%)

LQMM = linear quantile mixed model; LOTTR = level of travel time reliability; GRF = generalized random forest.

Summary

The results of the case studies suggest that the effects of operational projects, capacity expansion, and work zones are reasonably modeled using the approaches developed in this study. The LQMMs captured changes in the 90th percentile better than the GRF models, whereas the GRF models captured changes in LOTTR better in most cases. GRF models were also better at capturing the reliability changes caused by non-recurrent events, such as incidents or work zones. They also could assess impact factors that were not significant in LQMMs, such as SSP presence.

The reliability changes predicted using projected input data were similar (less than 5%) to those using actual after data, demonstrating how the models could be used in practice to assess proposed future projects that do not yet have any available after data.

CONCLUSIONS

- *Clustering segments into homogenous groups based on travel time distribution can improve the efficiency of further travel time reliability modeling*. Hierarchical clustering techniques were successfully used to consolidate the information contained in a multitude of statewide TMCs into four homogeneous clusters. This greatly enhances efficiency and facilitates efforts aimed at (1) understanding travel time reliability and its influencing factors; (2) predicting reliability; and (3) setting appropriate targets for the future.

- *Overall, VDOT-maintained data sources such as VaTraffic are currently better suited for reliability modeling than crowdsourced event data on interstates*. The prediction accuracies of VaTraffic and Waze models were similar. However, the incident data in Waze models suffer from issues such as unbalanced spatial density, duplicated reporting, and inconsistent event classification because of individual observer bias, which significantly impacts the modeling results, making them unstable and difficult to interpret. As a consequence, VaTraffic models are the preferred models because VaTraffic data are more stable over time.

- *LQMMs are significantly more effective for predicting the 50th, 80th, and 90th percentile travel times than the commonly used trend line method*. For freeway segments, the improvements in MAPE were 82%, 74%, and 68% for the 50th, 80th, and 90th percentiles, respectively. The prediction improvements for interchange segments were relatively modest at 15% (50th), 12% (80th), and 5% (90th).

- *The LQMMs and GRF models were able to reflect accurately actual changes in reliability created by improvement projects*. Effects of operations projects, capacity expansion, and work zones were successfully modeled using the approaches developed in this study.

- *The GRF approach is preferred over LQMM for reliability prediction.* GRF models performed better than LQMMs at predicting the federally mandated LOTTR measure (also 80th and 50th percentiles of travel times) and performed only slightly worse at predicting the 90th percentile. The GRF approach could also reflect the impact of variables that were removed from LQMMs because of insignificance, such as the presence of SSPs. Further, it does not require clustering of travel time distributions into homogeneous groups prior to model building, helping to reduce the work required to format data prior to modeling.

RECOMMENDATIONS

1. *The Virginia Transportation Research Council (VTRC) should develop detailed step-by-step guidance for VDOT's Traffic Engineering Division (TED), Operations Division (OD), and Transportation Mobility and Planning Division (TMPD) and the Office of Intermodal Planning and Investment (OIPI) so that they can use the data-driven models developed in this study for travel time reliability prediction on interstates.* This guidance will include detailed, specific information on data sources used, data formatting, conflation methods, technical specifications of computational resources, and programming code used to develop the models. This document will also include guidance on how practitioners can apply these models.

2. *VTRC should conduct additional research to adapt directly the preferred reliability prediction methods identified in this study to meet the requirements of MAP-21 federal target setting.* This will include expansion of the existing models to use the National Performance Management Research Data Set (NPMRDS) instead of INRIX commercial data; extension to weekday midday and weekend periods; and expansion of this analysis to cover non-interstate NHS routes. Any developed models should include the reliability impact factors that were identified as significant in this study.

3. *VDOT's TED and OD should explore new data sources* that could augment or improve *existing data sources that were identified as having limitations, such as weather data and work zone information.* Improved reliability predictions could be generated if the underlying data were further improved.

IMPLEMENTATION AND BENEFITS

Implementation

With regard to Recommendation 1, VTRC will work with VDOT's TED, OD, and TMPD, as well as the OIPI to develop the guidance document that will allow VDOT to incorporate the models into travel time reliability target setting procedures for interstates. VTRC will consult with these groups to determine the timeline for federal target setting and initiate a project to develop the "how to" document. This will occur no later than 6 months after the publication of this report but may occur sooner depending on the timeline for target setting.

With regard to Recommendation 2, VTRC will initiate two research projects. The first will expand the approaches in this study to other time periods on the interstates using the NPMRDS and will occur within 6 months of the publication of this report. The second will produce target setting approaches for non-interstate NHS routes and is already in the planned FY22 VTRC Work Program.

With regard to Recommendation 3, VTRC and VDOT's TED and OD will engage with emerging connected vehicle data providers to assess the feasibility of using data from on-road probes to determine weather conditions. Several providers are now selling connected car data that include factors such as windshield wiper usage and antilock brake activation that could provide microscopic level weather condition data, potentially eliminating the need for spatial extrapolation. The feasibility of using these sources will be determined within 1 year of the publication of this report. VDOT is already involved with the Work Zone Data Exchange (WZDx) initiative, which should improve the quality of available work zone data.

Benefits

Implementing the recommendations will ensure that the impacts of transportation improvement projects could be reflected by projecting related impact variables as illustrated in the before-after case studies. Predicting travel time reliability and capturing the expected changes in reliability more precisely will facilitate (1) setting more reasonable targets and tracking the progress toward their achievement; (2) identifying and prioritizing sites with high potential for reliability improvements; and (3) selecting cost-effective projects on interstates for funding through the SMART SCALE program.

The three study recommendations all directly support these benefits. Specifically, the benefits of implementing Recommendation 1 would be clear, direct information that VDOT divisions and OIPI could immediately use to implement the models developed in this study. The benefits of implementing Recommendation 2 would be methods that could be applied statewide for federal target setting. The benefits of implementing Recommendation 3 would be improved overall data quality for these data elements, supporting higher accuracy of the reliability models and a number of other applications.

ACKNOWLEDGMENTS

The authors express their gratitude to VDOT and FHWA for their support of this research. The authors are thankful to the following personnel who served on the TRP for this study: Mena Lockwood (VDOT's TED), Paul Szatkowski (VDOT's OD), Margit Ray (OIPI), Jungwook Jun (VDOT's TMPD), and Ramkumar Venkatanarayana (VTRC). The authors also thank Simona Babiceanu and Sanhita Lahiri (VDOT's TED) and Katie McCann (VDOT's OD) for sharing data and providing constructive feedback throughout the study.

REFERENCES

Abdel-Aty, M.A., Kitamura, R., and Jovanis, P.P. (1995). Investigating Effect of Travel Time Variability on Route Choice Using Repeated-Measurement Stated Preference Data.

Transportation Research Record: Journal of the Transportation Research Board, 1493, 39-45.

Aftabuzzaman, M. (2007). Measuring Traffic Congestion: A Critical Review. *The 30th Australasian Transport Research Forum*. https://www.australasiantransportresearchforum.org.au/papers/2007. Accessed December 1, 2019.

Arroyo, S., and Kornhauser, A.L. (2005). Modeling Travel Time Distributions on a Road Network. In *Proceedings of the Eighty-Fourth Annual Meeting of the Transportation Research Board*, Washington, DC.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized Random Forests. *Annals of Statistics*, 47(2), 1148-1178.

Bailey, K. (1994). Numerical Taxonomy and Cluster Analysis. *Typologies and Taxonomies*, 34, 24.

Bernard, S., Heutte, L., and Adam, S. (2009). Influence of Hyperparameters on Random Forest Accuracy. In *International Workshop on Multiple Classifier Systems*, 171-180. Springer, Berlin, Heidelberg.

Bouveyron, C., Côme, E., and Jacques, J. (2015). The Discriminative Functional Mixture Model for a Comparative Analysis of Bike Sharing Systems. *Annals of Applied Statistics*, 9(4), 1726-1760.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.

Caceres, N., Romero, L.M., and Benitez, F.G. (2012). Estimating Traffic Flow Profiles According to a Relative Attractiveness Factor. *Procedia-Social and Behavioral Sciences*, 54, 1115-1124.

California Department of Transportation, Transportation System Information Program. (1998). *California Transportation Plan: Transportation System Performance Measures: Final Report*. Sacramento.

Cambridge Systematics. (2005). *Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation*. FHWA-HOP-05-064. Federal Highway Administration, Washington, DC.

Cambridge Systematics. (2013). *Analytical Procedures for Determining the Impacts of Reliability Mitigation Strategies*. Transportation Research Board, Washington, DC.

Castillo, E., Nogal, M., Menendez, J., Sanchez-Cambronero, S., and Jimenez, P. (2012). Stochastic Demand Dynamic Traffic Models Using Generalized Beta-Gaussian Bayesian Networks. *IEEE Transactions on Intelligent Transportation Systems*, 13(2), 565-581.

Charlotte, C., and Sandra, B. (2017). Empirical Estimation of the Variability of Travel Time. *Transportation Research Procedia,* 25, 2769-2783.

Delhome, R., Billot, R., and El Faouzi, N.E. (2017). Travel Time Statistical Modeling With the Halphen Distribution Family. *Journal of Intelligent Transportation Systems,* 21(6), 452-464.

Dong, J., and Mahmassani, H.S. (2009). Flow Breakdown and Travel Time Reliability. *Transportation Research Record: Journal of the Transportation Research Board,* 2124(1), 203-212.

Dowling, R.G., Skabardonis, A., Margiotta, R.A., and Hallenbeck, M.E. (2009). Reliability Breakpoints on Freeways. Presented at the 88th Annual Meeting of the Transportation Research Board, Washington, DC.

Edwards, M.B., and Fontaine, M.D. (2012). Investigation of Travel Time Reliability in Work Zones with Private-Sector Data. *Transportation Research Record*: *Journal of the Transportation Research Board,* 2272, 9-18.

Elefteriadou, L., Xu, H., and Xie, L. (2008). *Travel Time Reliability Models.* University of Florida, Gainesville.

Emam, E.B., and Al-Deek, H. (2006). Using Real-Life Dual-Loop Detector Data to Develop New Methodology for Estimating Freeway Travel Time Reliability. *Transportation Research Record: Journal of the Transportation Research Board,* 1959(1), 140-150.

Federal Highway Administration. (2006). *Travel Time Reliability: Making It There on Time, All the Time.* Washington, DC.

Federal Highway Administration. (2017). System Performance/Freight/CMAQ Performance Measures Final Rule. Washington, DC.

Federal Register. (2017). National performance management measures; Assessing performance of the national highway system, freight movement on the Interstate System, and Congestion Mitigation and Air Quality Improvement Program [Internet], 82, 5970-6050.

Florida Department of Transportation. (2000). *The Florida Reliability Method in Florida's Mobility Performance Measures Program.* Tallahassee.

Geraci M. (2014). Linear Quantile Mixed Models: The Lqmm Package for Laplace Quantile Regression. *Journal of Statistical Software*, 57(13), 1-29.

Geraci, M., and Bottai, M. (2007). Quantile Regression for Longitudinal Data Using the Asymmetric Laplace Distribution. *Biostatistics,* 8(1), 140-154.

Geraci, M., and Bottai, M. (2014). Linear Quantile Mixed Models. *Statistics and Computing*, 24(3), 461-479.

Guo, F., Zhang, D., and Rakha, H. (2015). *Bayesian Travel Time Reliability Models*. Virginia Tech Transportation Institute, Blacksburg.

Hojati, A.T., Ferreira, L., Washington, S., Charles, P., and Shobeirinejad, A. (2016). Modelling the Impact of Traffic Incidents on Travel Time Reliability. *Transportation Research Part C: Emerging Technologies*, 65, 49-60.

Kim, J., and Mahmassani, H.S. (2015). Spatial and Temporal Characterization of Travel Patterns in a Traffic Network Using Vehicle Trajectories. *Transportation Research Procedia,* 9, 164-184.

King, B. (1967). Step-Wise Clustering Procedures. *Journal of the American Statistical Association*, 62(317), 86-101.

Koenker, R., and Bassett, G., Jr. (1978). Regression Quantiles. *Econometrica: Journal of the Econometric* Society, 33-50.

Kweon, Y.J., and Lim, I.K. (2014). *Development of Safety Performance Functions for Multilane Highway and Freeway Segments Maintained by the Virginia Department of Transportation.* VCTIR 14-R14. Virginia Transportation Research Council, Charlottesville.

Kwon, J., Barkley, T., Hranac, R., Petty, K., and Compin, N. (2011). Decomposition of Travel Time Reliability Into Various Sources: Incidents, Weather, Work Zones, Special Events, and Base Capacity. *Transportation Research Record: Journal of the Transportation Research Board*, 2229, 28-33.

Lam, T.C., and Small, K.A. (2001). The Value of Time and Reliability: Measurement From a Value Pricing Experiment. *Transportation Research Part E: Logistics and Transportation Review*, 37(2-3), 231-251.

Lan, C.L., Venkatanarayana, R., and Fontaine, M.D. (2019). Development of a Methodology for Determining Statewide Recurring and Nonrecurring Freeway Congestion: Virginia Case Study. *Transportation Research Record: Journal of the Transportation Research Board*, 2673(6), 566-578.

Li, R., Rose, G., and Sarvi, M. (2006). Using Automatic Vehicle Identification Data to Gain Insight Into Travel Time Variability and Its Causes. *Transportation Research Record: Journal of the Transportation Research Board*, 1945(1), 24-32.

List, G.F., Williams, B.M., Rouphail, N.M., Hranac, R., Barkley, T., Mai, E., and Karr, A.F. (2014a). *Establishing Monitoring Programs for Travel Time Reliability*. SHRP 2 Report S2-L02-RR-1. Transportation Research Board, Washington, DC.

List, G.F., Williams, B., Rouphail, N., Hranac, R., Barkley, T., Mai, E., Ciccarelli, A., Rodegerdts, L., Pincus, K., Nevers, B., Karr, A., Zhou, X., Wojowicz, J., Schofer, J., and Khattak, A. (2014b). *Guide to Establishing Monitoring Programs for Travel Time Reliability*. SHRP 2 Report S2-L02-RR-2. Transportation Research Board, Washington, DC.

McQuitty, L.L. (1964). Capabilities and Improvements of Linkage Analysis as a Clustering Method. *Educational and Psychological Measurement,* 24(3), 441-456.

Meinshausen, N., and Ridgeway, G. (2006). Quantile Regression Forests. *Journal of Machine Learning Research*, 7(6).

National Cooperative Highway Research Program. (2013). *NCHRP Report 510: Interim Planning for a Future Strategic Highway Research Program*. Transportation Research Board, Washington, DC.

Office of Intermodal Planning and Investment. (2021). *SMART SCALE Technical Guide*. Commonwealth of Virginia Transportation Board, Richmond.

Park, S., Rakha, H., and Guo, F. (2011). Multi-State Travel Time Reliability Model: Impact of Incidents on Travel Time Reliability. In *14th International IEEE Conference on Intelligent Transportation Systems*, 2106-2111.

Plötz, P., Jakobsson, N., and Sprei, F. (2017). On the Distribution of Individual Daily Driving Distances. *Transportation Research Part B,* 150, 213-227.

Probst, P., Wright, M.N., and Boulesteix, A.L. (2019). Hyperparameters and Tuning Strategies for Random Forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3), e1301.

Pu, W. (2011). Analytic Relationships Between Travel Time Reliability Measures. *Transportation Research Record: Journal of the Transportation Research Board,* 2254(1), 122-130.

RITIS. (2021). The Probe Data Analytics Suite. https://pda.ritis.org/suite/. Accessed March 16, 2021.

Rousseeuw, P.J. (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.

Schrank, D., Geng, L., Kang, D.H., and Kong, X. (2018). Technical Memorandum on Conflation and Geospatial Data Integration, Task 4—Syntheses Development (Synthesis 1 of 3). Texas A & M Transportation Institute, College Station.

Sokal, R.R. Numerical Taxonomy. (1966). *Scientific American*, 215(6), 106-117.

Srinivasan, K.K., Prakash, A.A., and Seshadri, R. (2014). Finding Most Reliable Paths on Networks With Correlated and Shifted Log-Normal Travel Times. *Transportation Research Part B: Methodological*, 66, 110-128.

Taylor, M.A. (2017). Fosgerau's Travel Time Reliability Ratio and the Burr Distribution. *Transportation Research Part B: Methodological,* 97, 50-63.

Thas, O. (2010). *Comparing Distributions.* Springer, New York.

The National Capital Region Transportation Planning Board. (2018). *System Performance Targets: Travel Time Reliability and Truck Travel Time Reliability*. The Metropolitan Washington Council of Governments, Washington, DC.

Tibshirani, J., Athey, S., Friedberg, R., Hadad, V., Hirshberg D., Miner, L., Sverdrup, E., Wager, S., and Wright, M. (2021). Package 'grf'.

Transportation Research Board. (2016). *Highway Capacity Manual, Sixth Edition.* Washington, DC.

Tu, H., Van Lint, H., and Van Zuylen, H. (2008). The Effects of Traffic Accidents on Travel Time Reliability. In *2008 11th International IEEE Conference on Intelligent Transportation Systems,* 79-84.

Tufuor, E., Rilett, L.R., and Zhao, L. (2020). Calibrating the Highway Capacity Manual Arterial Travel Time Reliability Model. *Journal of Transportation Engineering, Part A: Systems,* 146(12), 04020131.

Tufuor, E.O., and Rilett, L.R. (2020). Analysis of Component Errors in the Highway Capacity Manual Travel Time Reliability Estimations for Urban Streets. *Transportation Research Record: Journal of the Transportation Research Board,* 2674(6), 85-97.

Ullman, G.L., Pratt, M., Fontaine, M.D., Porter, R.J., and Medina, J. (2018). *NCHRP Research Report 869: Estimating the Safety Effects of Work Zone Characteristics and Countermeasures: A Guidebook.* Transportation Research Board, Washington, DC.

Van Lint, J.W.C., and Van Zuylen, H.J. (2005). Monitoring and Predicting Freeway Travel Time Reliability: Using Width and Skew of Day-to-Day Travel Time Distribution. *Transportation Research Record: Journal of the Transportation Research Board,* 1917(1), 54-62.

Virginia Department of Transportation. (2021a). VDOT Crash Analysis Tool. https://bit.ly/VDOTCrashTool_Public. Accessed March 16, 2021.

Virginia Department of Transportation. (2021b). I-64 Widening. https://i64widening.org/learn_more/segment_2.asp. Accessed March 16, 2021.

Virginia Department of Transportation. (2021c). I-64/I-264 Improvements. https://bit.ly/VDOTCrashTool_Public. Accessed March 16, 2021.

Weijermars, W., and Van Berkum, E. (2005). Analyzing Highway Flow Patterns Using Cluster Analysis. In *Proceedings of the 2005 IEEE Intelligent Transportation Systems*, 308-313.

Wright, B., Zou, Y., and Wang, Y. (2015). Impact of Traffic Incidents on Reliability of Freeway Travel Times. *Transportation Research Record: Journal of the Transportation Research Board*, 2484, 90-98.

Yang, S., Malik, A., and Wu, Y.J. (2014). Travel Time Reliability Using the Hasofer–Lind–Rackwitz–Fiessler Algorithm and Kernel Density Estimation. *Transportation Research Record: Journal of the Transportation Research Board*, 2442, 85-95.

Yang, S., and Wu, Y.J. (2016). Mixture Models for Fitting Freeway Travel Time Distributions and Measuring Travel Time Reliability. *Transportation Research Record: Journal of the Transportation Research Board*, 2594(1), 95-106.

Yazici, M.A., Kamga, C., and Mouskos, K.C. (2013). Analysis of Travel Time Reliability in New York City Based on Day-of-Week and Time-of-Day Periods. *Transportation Research Record*: *Journal of the Transportation Research Board*, 2308(1), 83-95.

Zegger, J., Bonneson, J., Dowling, R., Ryus, P., Vandehey, M., Kittelson, W., Rouphail, N., Schroeder, B., Hajbabaie, A., Aghdashi, B., Chase, T., Sajjadi, S., Margiotta, R., and Elefteriadou, L. (2014). *Incorporating Travel Time Reliability Into the Highway Capacity Manual*. SHRP 2 Report S2-L08-RW-1. Transportation Research Board, Washington, DC.

Zhang, X., and Chen, M. (2019). Quantifying the Impact of Weather Events on Travel Time and Reliability. *Journal of Advanced Transportation*, 2019.

APPENDIX

Table A1. Freeway Model Coefficients

VaTraffic												
	Group 1						Group 2					
	50th	Beta _50	80th	Beta _80	90th	Beta _90	50th	Beta _50	80th	Beta _80	90th	Beta _90
miles	54.919	0.950*	52.805	0.914*	53.099	0.919*	54.049	0.768*	55.361	0.786*	56.522	0.803*
throu_lane	0.210	0.001	10.413	0.065	17.962	0.112*	0.079	0.003	-0.097	-0.003	0.650	0.020
frozen_precip	50.495	0.009*	148.525	0.025*	248.805	0.042*	8.431	0.007*	17.923	0.015*	29.278	0.025*
rain_precip	1.561	0.004*	1.314	0.004*	1.359	0.004	0.352	0.005	0.111	0.002	0.533	0.008
Severe_Injury	-7.940	-0.002	38.876	0.009	169.264	0.041*	10.703	0.006	37.651	0.020*	64.437	0.035*
Visible_Injury	-9.424	-0.006*	-10.151	-0.006*	-13.840	-0.009	-2.501	-0.004	-4.718	-0.007	15.673	0.025
Nonvisible_Injury	-3.714	-0.001	-4.186	-0.001	-7.684	-0.002	-1.843	-0.001	-1.613	-0.001	84.671	0.050*
PDO	-8.519	-0.011*	-7.634	-0.010*	-6.879	-0.009	-1.617	-0.005	0.022	0.000	19.800	0.059*
shoulder_closure	-0.342	0.000	-0.056	0.000	0.239	0.000	0.052	0.000	-0.610	-0.001	-0.399	-0.001
lane_closure	-0.137	0.000	1.900	0.001	2.747	0.002	-0.393	-0.001	-0.647	-0.001	0.214	0.000
breakdown	-2.282	-0.005*	-2.325	-0.005	-3.012	-0.006	-0.872	-0.004	-2.166	-0.009	-2.331	-0.010
hazard	-1.153	0.000	38.311	0.009	217.313	0.050*	-8.377	-0.005	-1.819	-0.001	10.826	0.007
rural	-5.783	-0.027	-4.036	-0.019	-4.430	-0.021	2.089	0.025	-1.423	-0.017	-1.842	-0.022
vc_ratio	37.971	0.066*	71.861	0.126*	71.718	0.125*	6.140	0.043	2.297	0.016	4.785	0.034
Par_lane	19.564	0.036*	42.264	0.079*	67.267	0.125*	4.197	0.045*	12.050	0.129*	30.323	0.324*
heavy_percent	0.763	0.059*	1.247	0.097*	1.116	0.087*	-0.107	-0.029	-0.030	-0.008	0.125	0.034
ssp	-7.813	-0.037*	-5.438	-0.025	-6.763	-0.032	-1.696	-0.032	-1.282	-0.024	-1.635	-0.030
frozen_precip : vc_ratio	-22.254	-0.002	-143.995	-0.011*	-278.759	-0.022*	3.527	0.002	1.690	0.001	19.999	0.008
rain_precip : vc_ratio	4.643	0.006*	11.831	0.016*	19.215	0.026*	0.856	0.007*	3.183	0.025*	3.963	0.031*
Severe_Injury : vc_ratio	144.520	0.014*	256.476	0.025	308.795	0.030	-6.982	-0.002	-24.048	-0.007	-46.394	-0.013
Visible_Injury : vc_ratio	79.015	0.023*	184.318	0.053*	331.630	0.095*	19.516	0.017*	40.911	0.035*	45.371	0.039
Nonvisible_Injury : vc_ratio	60.149	0.007*	96.277	0.011*	168.545	0.019*	19.106	0.006	51.293	0.017	-44.945	-0.015
PDO_vc : ratio	61.230	0.036*	112.572	0.065*	179.580	0.104*	11.248	0.019*	23.165	0.038*	0.742	0.001
shoulder_closure : vc_ratio	0.838	0.000	1.545	0.001	1.342	0.000	-0.291	0.000	1.023	0.001	0.516	0.000
lane_closure : vc_ratio	4.144	0.001	-0.089	0.000	-2.571	-0.001	1.496	0.001	2.281	0.002	0.300	0.000
breakdown : vc_ratio	7.690	0.008*	16.574	0.017*	32.302	0.034*	2.375	0.006	7.082	0.017*	8.238	0.020*
hazard : vc_ratio	18.623	0.002	60.119	0.006	-150.314	-0.015	27.914	0.009*	22.913	0.007	16.674	0.005
heavy_percent : vc_ratio	-2.876	-0.061*	-3.850	-0.082*	-3.557	-0.076*	-0.518	-0.040*	-0.188	-0.014	-0.578	-0.044
ssp : vc_ratio	14.610	0.031*	14.769	0.032	19.961	0.043	1.217	0.012	1.972	0.020	2.101	0.021

ssp : breakdown	0.131	0.000	-2.118	-0.004	-3.906	-0.007	0.105	0.000	0.068	0.000	0.206	0.001
Waze												
	Group 1						**Group 2**					
miles	52.605	0.910*	53.432	0.925*	53.429	0.925*	54.172	0.769*	51.537	0.732*	56.211	0.798*
throu_lane	-5.262	-0.033*	5.140	0.032	13.841	0.087	-0.209	-0.007	2.831	0.088*	0.360	0.011
frozen_precip	54.541	0.009*	138.200	0.023*	224.258	0.038*	4.413	0.004	16.562	0.014*	21.210	0.018*
rain_precip	1.700	0.005*	1.516	0.004*	1.150	0.003	0.329	0.005*	0.430	0.006	0.580	0.009
WAZE_crashes	-2.666	-0.010*	45.341	0.171*	132.911	0.502*	-0.235	-0.002	17.649	0.165*	38.135	0.357*
WAZE_breakdown	-1.022	-0.018*	-1.882	-0.033*	-2.480	-0.043*	-0.277	-0.008*	-0.672	-0.019*	-1.057	-0.031*
WAZE_hazard	-0.134	-0.001	-0.222	-0.002	-0.274	-0.003	-0.238	-0.003*	-0.741	-0.010*	-0.672	-0.009*
shoulder_closure	0.266	0.000	0.591	0.001	0.719	0.001	-0.051	0.000	0.141	0.000	-0.520	-0.001
lane_closure	0.569	0.000	0.700	0.001	3.768	0.003	0.181	0.000	0.170	0.000	2.057	0.003
rural	2.922	0.014	-8.802	-0.041	-7.428	-0.035	0.567	0.007	1.393	0.017	-1.936	-0.023
vc_ratio	19.728	0.035*	45.238	0.079**	45.222	0.079*	5.296	0.037	6.604	0.046	5.372	0.038
Par_lane	17.684	0.033*	40.928	0.076*	51.690	0.096*	4.250	0.045*	10.871	0.116*	17.724	0.189*
heavy_percent	0.488	0.038*	1.075	0.084*	1.034	0.080*	-0.055	-0.015	0.062	0.017	0.135	0.037
ssp	-7.166	-0.033*	-4.183	-0.020	-4.846	-0.023	-1.485	-0.028	-0.091	-0.002	-1.390	-0.026
frozen_precip : vc_ratio	-39.298	-0.003*	-127.832	-0.010*	-239.506	-0.019*	9.009	0.004	6.576	0.003	20.042	0.008
rain_precip : vc_ratio	4.350	0.006*	9.490	0.013*	17.349	0.024*	0.867	0.007*	2.133	0.017*	3.510	0.028*
WAZE_crashes : vc_ratio	31.400	0.069*	-11.004	-0.024	-114.070	-0.250*	7.157	0.038*	-9.549	-0.051	-34.741	-0.186*
WAZE_breakdown : vc_ratio	3.324	0.029*	8.415	0.074*	12.583	0.111*	0.903	0.014*	2.354	0.036*	3.893	0.060*
WAZE_hazard : vc_ratio	0.659	0.002	1.707	0.006*	2.219	0.008*	0.693	0.005*	2.228	0.015*	2.246	0.015*
shoulder_closure : vc_ratio	-0.412	0.000	-1.662	-0.001	-2.663	-0.001	0.037	0.000	-0.415	0.000	0.356	0.000
lane_closure : vc_ratio	2.067	0.001	6.211	0.002	-2.735	-0.001	-0.012	0.000	-0.120	0.000	-2.690	-0.002
heavy_percent : vc_ratio	-2.589	-0.055*	-2.960	-0.063*	-2.884	-0.061*	-0.494	-0.038*	-0.302	-0.023	-0.662	-0.051
ssp : vc_ratio	17.232	0.037*	8.499	0.018	9.795	0.021	1.240	0.013	-1.276	-0.013	2.131	0.022
ssp : WAZE_breakdown	-0.139	-0.002	-0.256	-0.004	-0.343	-0.005	-0.079	-0.002	-0.039	-0.001	-0.080	-0.002

Coefficients with asterisks are significant at p = 0.05. Columns with names starting with "beta_" are standardized values.

Table A2. Interchange Model Coefficients

VaTraffic												
	Group 1						Group 2					
	50th	**Beta_50**	**80th**	**Beta_80**	**90th**	**Beta_90**	**50th**	**Beta_50**	**80th**	**Beta_80**	**90th**	**Beta_90**
miles	66.650	0.533*	76.866	0.615*	75.645	0.605*	54.943	1.274*	57.092	1.323*	54.886	1.272*
throu_lane	-0.163	-0.004	2.186	0.053	10.769	0.263*	0.656	0.033*	0.677	0.034	0.129	0.007
frozen_precip	5.420	0.004*	11.128	0.008	13.593	0.010	1.686	0.002	11.638	0.016*	23.604	0.033*
rain_precip	0.265	0.003	0.178	0.002	0.066	0.001	-0.049	-0.001	-0.138	-0.003	-0.386	-0.009*
Severe_Injury	36.253	0.015	25.884	0.011	61.535	0.026	-10.592	-0.009	-4.609	-0.004	38.462	0.034
Visible_Injury	-0.850	-0.001	19.175	0.027	40.445	0.056*	-1.328	-0.003	-5.296	-0.014*	-4.249	-0.011
Nonvisible_Injury	6.467	0.004	10.572	0.006	23.762	0.013	-2.243	-0.002	-1.137	-0.001	17.415	0.017
PDO	-0.297	-0.001	8.247	0.021	21.658	0.055*	-4.661	-0.023*	-5.668	-0.028*	-3.406	-0.017*
shoulder_closure	0.004	0.000	0.981	0.002	0.945	0.001	0.057	0.000	0.082	0.000	1.024	0.003
lane_closure	0.267	0.000	0.568	0.001	0.053	0.000	0.086	0.000	0.460	0.001	-0.081	0.000
breakdown	1.863	0.007	4.848	0.019	13.182	0.052	-0.704	-0.005*	-1.285	-0.009	0.337	0.002
hazard	4.426	0.002	-1.729	-0.001	7.268	0.003	-2.678	-0.003	-5.916	-0.006	-0.932	-0.001
rural	4.303	0.025	-4.866	-0.028	-3.199	-0.018	-0.020	0.000	-0.764	-0.015	-0.355	-0.007
vc_ratio	5.326	0.031	6.868	0.040	1.173	0.007	3.615	0.042*	12.754	0.146*	9.556	0.110
Par_lane	5.294	0.052*	10.787	0.106	14.356	0.141	3.535	0.062	27.912	0.486*	28.097	0.489*
heavy_percent	-0.387	-0.068*	-0.063	-0.011	0.575	0.102	0.160	0.071*	0.261	0.116*	0.145	0.064*
ssp	-3.406	-0.048	-2.768	-0.039	-5.801	-0.082	-0.727	-0.022*	-1.869	-0.057*	-0.427	-0.013
frozen_precip : vc_ratio	6.999	0.003*	3.937	0.002	-4.138	-0.002	22.289	0.015*	22.766	0.015*	18.882	0.013
rain_precip : vc_ratio	0.960	0.006*	2.549	0.017*	3.559	0.024*	1.574	0.020*	3.190	0.041*	4.668	0.060*
Severe_Injury : vc_ratio	-43.709	-0.011	1.743	0.001	-44.160	-0.012	68.725	0.031*	119.848	0.053*	66.847	0.030
Visible_Injury : vc_ratio	17.070	0.015	-0.225	0.000	-23.988	-0.021	18.466	0.026*	73.576	0.104*	83.918	0.119*
Nonvisible_Injury : vc_ratio	6.010	0.002	0.417	0.000	-11.216	-0.004	49.907	0.027	61.977	0.034	90.269	0.049
PDO : vc_ratio	11.187	0.018*	5.134	0.008	-12.803	-0.020	24.975	0.067*	65.461	0.176*	63.525	0.171*
shoulder_closure : vc_ratio	-0.411	0.000	-2.762	-0.002	-3.934	-0.003	0.030	0.000	-0.428	-0.001	-2.778	-0.004
lane_closure : vc_ratio	0.654	0.001	2.522	0.002	2.401	0.002	0.063	0.000	-0.560	-0.001	0.897	0.001
breakdown : vc_ratio	1.999	0.005	2.465	0.006	-7.257	-0.018	2.732	0.011*	5.309	0.021	-0.354	-0.001
hazard : vc_ratio	-5.476	-0.001	16.175	0.004	14.972	0.003	10.760	0.006	43.636	0.023*	34.172	0.018
heavy_percent : vc_ratio	-0.427	-0.024	-0.499	-0.028	0.470	0.026	-0.292	-0.037*	-0.749	-0.094*	-0.477	-0.060
ssp : vc_ratio	6.575	0.056	5.599	0.047	7.811	0.066	1.569	0.026*	4.907	0.081*	1.177	0.019
ssp : breakdown	-2.289	-0.007	-4.340	-0.013	-6.105	-0.019	-0.174	-0.001	-0.111	-0.001	-0.026	0.000

Waze												
	Group 1						Group 2					
miles	64.486	0.516*	58.326	0.467*	72.004	0.576*	54.330	0.733*	56.981	0.769*	54.916	0.741*
throu_lane	-0.630	-0.015	0.305	0.007	10.378	0.253*	-0.307	-0.012	0.602	0.023	0.128	0.005
frozen_precip	2.884	0.002	24.465	0.019*	29.529	0.022*	5.894	0.005*	12.521	0.011*	22.455	0.021*
rain_precip	0.133	0.002	0.442	0.005	0.261	0.003	-0.075	-0.001	-0.129	-0.002	-0.324	-0.006*
WAZE_crashes	6.508	0.059*	15.235	0.138*	30.453	0.276*	-5.067	-0.045*	-5.047	-0.044*	8.706	0.077
WAZE_breakdown	-0.019	-0.001	1.000	0.027	2.655	0.071	-0.186	-0.007*	-0.908	-0.036*	-0.870	-0.035*
WAZE_hazard	-0.668	-0.008*	-0.552	-0.006	-0.591	-0.007	-0.735	-0.016*	-0.309	-0.007	-0.101	-0.002
shoulder_closure	0.101	0.000	0.624	0.001	-0.927	-0.001	0.480	0.001	0.137	0.000	0.910	0.002
lane_closure	-0.440	-0.001	-2.052	-0.003	3.666	0.005	0.714	0.001	2.769	0.005*	1.762	0.003
rural	1.835	0.011	-17.071	-0.098*	-2.779	-0.016	-0.331	-0.010	-0.647	-0.020	-0.411	-0.013
vc_ratio	2.581	0.015	5.386	0.031	1.306	0.008	3.096	0.028*	4.436	0.039	4.055	0.036
Par_lane	5.201	0.051*	17.508	0.172*	11.569	0.114	3.708	0.029	24.201	0.190*	25.590	0.201*
heavy_percent	-0.034	-0.006	0.145	0.026	0.546	0.096	0.082	0.039	0.155	0.073*	0.079	0.037*
ssp	-4.915	-0.070*	-1.294	-0.018	-6.163	-0.088	-1.084	-0.033*	-1.336	-0.040*	-0.519	-0.016
frozen_precip : vc_ratio	9.435	0.004*	-16.975	-0.007	-26.224	-0.010	11.104	0.004	16.221	0.005	14.531	0.005
rain_precip : vc_ratio	1.194	0.008*	2.158	0.014*	3.110	0.021*	1.766	0.014*	3.213	0.025*	4.797	0.037*
WAZE_crashes : vc_ratio	0.120	0.001	-5.468	-0.031	-21.294	-0.122	22.532	0.112*	47.203	0.235*	30.650	0.153*
WAZE_breakdown : vc_ratio	1.572	0.027*	2.791	0.048	1.399	0.024	0.589	0.012	4.293	0.086*	4.281	0.086*
WAZE_hazard : vc_ratio	1.715	0.012*	2.669	0.019*	1.907	0.014	2.254	0.018*	1.478	0.012	0.986	0.008
shoulder_closure : vc_ratio	-0.253	0.000	-1.917	-0.002	1.970	0.002	-0.738	-0.001	-0.195	0.000	-2.734	-0.003
lane_closure : vc_ratio	1.978	0.001	7.478	0.006	-5.443	-0.004	-1.601	-0.001	-7.136	-0.005*	-3.650	-0.002
heavy_percent : vc_ratio	-0.640	-0.035	-0.529	-0.029	0.646	0.036	-0.296	-0.041*	-0.379	-0.053*	-0.238	-0.033
ssp : vc_ratio	10.042	0.085*	1.188	0.010	6.152	0.052	2.450	0.029*	3.817	0.045*	1.527	0.018
ssp : WAZE_breakdown	-0.489	-0.009	-1.730	-0.032	-2.004	-0.037	0.002	0.000	-0.266	-0.009	-0.212	-0.007

Coefficients with asterisks are significant at p = 0.05. Columns with names starting with “beta_” are standardized values.