JOINT TRANSPORTATION
RESEARCH PROGRAM

INDIANA DEPARTMENT OF TRANSPORTATION
AND PURDUE UNIVERSITY

Using Emerging and Extraordinary Data Sources to Improve Traffic Safety

Andrew P. Tarko, Qiming Guo, Raul Pineda-Mendez, Mario A. Romero

SPR-4302 • Report Number: FHWA/IN/JTRP-2021/04 • DOI: 10.5703/1288284317283

RECOMMENDED CITATION

Tarko, A. P., Guo, Q., & Pineda-Mendez, R. (2021). *Using emerging and extraordinary data sources to improve traffic safety* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2021/04). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284317283

AUTHORS

Andrew P. Tarko, PhD
Professor of Civil Engineering and Director of Center for Road Safety
Lyles School of Civil Engineering
(765) 494-5027
tarko@purdue.edu
Corresponding Author

Qiming Guo
Raul Pineda-Mendez
Graduate Research Assistants
Lyles School of Civil Engineering
Purdue University

ACKNOWLEDGMENTS

This study was funded by the Indiana Department of Transportation through the Joint Transportation Research Program at Purdue University under Grant No. SPR-4302. The constant feedback from members of the standing advisory committee contributed significantly to this study. Special thanks to Michael Holowaty, Edward Cox, Gregory Katter, Shuo Li, Mark McMahan, Taylor Ruble, and Jeremy VanVleet. High-quality data is a crucial asset in research. The authors would like to acknowledge Jonatan Weaver and Beth Hall from the Indiana State Climate Office, Justin Ferri from KMJ Consulting Inc., and Yaddel Ezeribe from the Center for Advanced Transportation Technology Laboratory at the University of Maryland.

JOINT TRANSPORTATION RESEARCH PROGRAM

The Joint Transportation Research Program serves as a vehicle for INDOT collaboration with higher education institutions and industry in Indiana to facilitate innovation that results in continuous improvement in the planning, design, construction, operation, management and economic efficiency of the Indiana transportation infrastructure. https://engineering.purdue.edu/JTRP/index_html

Published reports of the Joint Transportation Research Program are available at http://docs.lib.purdue.edu/jtrp/.

NOTICE

COVER IMAGE

Fischer, E. (2015, May 27). *Unsignalized leg of partially signalized intersection* [Photograph]. Flickr. https://www.flickr.com/photos/24431382@N03/18810884261

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/IN/JTRP-2021/04		

4. Title and Subtitle	5. Report Date
Using Emerging and Extraordinary Data Sources to Improve Traffic Safety	January 2021
	6. Performing Organization Code
7. Author(s)	**8. Performing Organization Report No.**
Andrew P. Tarko, Qiming Guo, and Raul Pineda-Mendez and	FHWA/IN/JTRP-2021/04
9. Performing Organization Name and Address	**10. Work Unit No.**
Joint Transportation Research Program Hall for Discovery and Learning Research (DLR), Suite 204 207 S. Martin Jischke Drive West Lafayette, IN 47907	**11. Contract or Grant No.** SPR-4302
12. Sponsoring Agency Name and Address	**13. Type of Report and Period Covered**
Indiana Department of Transportation (SPR) State Office Building 100 North Senate Avenue Indianapolis, IN 46204	Final Report **14. Sponsoring Agency Code**

15. Supplementary Notes

Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. Abstract

The current safety management program in Indiana uses a method based on aggregate crash data for conditions averaged over several-year periods with consideration of only major roadway features. This approach does not analyze the risk of crashes potentially affected by time-dependent conditions such as traffic control, operations, weather and their interaction with road geometry. With the rapid development of data collection techniques, time-dependent data have emerged, some of which have become available for safety management. This project investigated the feasibility of using emerging and existing data sources to supplement the current safety management practices in Indiana and performed a comprehensive evaluation of the quality of the new data sources and their relevance to traffic safety analysis. In two case studies, time-dependent data were acquired and integrated to estimate their effects on the hourly probability of crash and its severity on two selected types of roads: (1) rural freeways and (2) signalized intersections. The results indicate a considerable connection between hourly traffic volume, average speeds, and weather conditions on the hourly probability of crash and its severity. Although some roadway geometric features were found to affect safety, the lack of turning volume data at intersections led to some counterintuitive results. Improvements have been identified to be implemented in the next phase of the project to eliminate these undesirable results.

17. Key Words	18. Distribution Statement
safety management, crash risk, injury severity, rural freeways, signalized intersections, emerging data	No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	78 including appendices	

EXECUTIVE SUMMARY

Introduction

In the current practice of traffic safety analysis, counts of crashes aggregated over several years, typically 3 to 5 years, are analyzed with count data models to estimate the effects of traffic volume and major design elements on roadway safety performance. This method of analysis takes advantage of data aggregation to address the rarity of crash occurrence, but it fails to reflect the diversity of operational conditions in short periods, which may be critical in identifying crash causality and effective safety counter measures. In the time of big data, emerging data sources provide various opportunities to perform novel safety analysis and, as a result, they call for the evolution of traditional safety management procedures. Among the emerging data, the most promising is high-resolution, time-dependent data including hourly traffic volumes, speed, and weather conditions. Thanks to their low aggregation level, the factors of crash probability and severity in short intervals, such as one hour, can be analyzed and estimated.

This pilot study investigated the feasibility of using emerging data sources to improve traffic safety. The research consisted of three components: (1) evaluation of the usability and reliability of emerging data sources; (2) example analysis of hourly crash probability and severity for two cases—rural freeways and signalized intersections using static and temporal data (geometry, traffic, weather, other); and (3) identification of the potential limitations and challenges of future implementation to safety management practice.

Findings

The data sources maintained by INDOT and FHWA are reliable and easy to access, while data sources offered by commercial companies may be less reliable and may involve considerable costs. The preliminary analysis of the obtained sample data showed the possibility of linking various time-dependent data, volume, speed features, and weather conditions to traffic safety. Nevertheless, restructuring, standardizing, and linking data from various sources is complex and time-consuming. Several analysis examples of time-dependent data were conducted and the key findings are listed as follows:

1. Hourly traffic volumes were forecasted using probe density, time indicators, AADT, travel speed characteristics, and weather conditions. The hourly volume estimation/prediction model explained 90% of the sample variability. This indicates its good performance.
2. Time-dependent variables were found to be significant in estimating crash risk and severity. The most pertinent determinants were travel speed variation, decreasing operating speed, congestion level, scattered rain, and freezing temperatures.
3. Some of the detailed roadway features, such as barriers, street lighting, and road curves, were found to significantly affect the probability of a crash and the severity of its outcome. Some of the effects of intersection design elements were found to be counterintuitive due to the lack of turning volumes among the variables used in the models.
4. The limited size of the interaction sample precluded confirming the significance of a few roadway design elements and temporal traffic conditions that were expected to be safety factors.
5. The obtained models revealed safety-critical conditions such as potentially slippery pavement from near freezing temperatures combined with large speed variation on approaches to signalized intersections. Identification of critical combinations of adverse conditions helps understand crash causality.

Implementation

The revealed limitations of emerging data sources pointed out the following data needs: traffic volumes and speeds on local roads, turning volumes at road intersections, and high-resolution road geometric data. The two studied cases showed the potential for identifying safety-critical operational conditions. Time-dependent data may help traffic safety engineers identify these conditions, their frequency, and their overall contribution to crash occurrence and severity, while also devising effective safety countermeasures and estimate their safety benefit and cost effectiveness. Practical and efficient methods of scanning selected types of roads over time to identify the high-risk conditions are under development in the follow-up SPR-4540 project.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Well-known shortcomings in traffic safety analysis include underreporting of crashes, shortage of safety-related information, and data inconsistency. Crash occurrence and its various factors (e.g., collision location, traffic operations and control, weather, and road and roadside conditions) are not always correctly reflected.

The introduction of emerging data sources is fully expected to improve safety management by providing broader insights into the relationships between traffic and road conditions and the resulting safety outcomes. For example, the combined effect of travel speed variation and poor weather is generally understood, but a practical application of variable speed limits is challenging, and its safety benefits are unclear. Application such as this could be rectified by a better understanding of the relationship between weather conditions, speed, and safety.

Traditional safety management practices use limited amounts of data that are widely available at the system level and are usually restricted to the exposure and general geometric characteristics of road segments and intersections. Unfortunately, other contributing factors, such as traffic operations, traffic control, weather conditions, and high-density roadway features, are usually not addressed. The objective of this research project was to identify the currently available emerging data sources that offer more information about these other contributing factors that increase crash risk.

Customarily, crash data are aggregated and analyzed over long periods, usually 3 to 5 years. This project pursued a novel more disaggregated approach where the hourly collision risk and injury severity are considered. This methodological change can facilitate the inclusion of previously unobservable variables such as hourly volume, weather, and operating speed. It is worth mentioning that the disaggregated results can be aggregated back to the original annual trends used in current safety management programs.

Disaggregate analysis tools based on individual hours can reveal a broader group of factors behind high numbers of crashes. Among the advantages of short-term safety analysis, the formulation of operational countermeasures like signal timing or law enforcement intensity is visible. The usefulness of the basic knowledge of the relationships between safety and all the factors is vital for proactively warning drivers or connected and autonomous vehicles about changes in safety conditions. To the end-users, the results also need to be aggregated back for the regular periods and conditions. However, the current project focused on extracting safety effects using disaggregated data. An ongoing research project (SPR-4540) aims to incorporate time-dependent data in proactive safety management practices, which is a natural follow-up to the fundamental insight gained from this current project.

This project's primary objective was to identify new and emerging data sources that could be used in conjunction with conventional data to support safety analysis. Other objectives include the following:

1. Identify emerging data sources that could be used in conjunction with the traditional data to support safety analysis.
2. Evaluate the quality of the new data and analyze the relevance of the new data with crash frequency and severity.
3. Propose a road to detecting safety problems with the new data and conventional safety data combined.

The remainder of this report is organized as follows:

- *Section 2. Data Sources* provides a comprehensive list of new sources. Each data source is described, and its potential use in safety management is discussed. A narrower group of data sources ultimately is selected for disaggregate safety analysis.
- *Section 3. Methodology* introduces the research approach and describes the statistical background. The proposed method is tested in two case studies (see Sections 4 and 5).
- *Section 4. Rural Freeways* and *Section 5. Signalized Intersections* are the two case studies where the details on data preparation and safety analysis results are presented and their implications for safety management are discussed.
- *Section 6. Recommendations for Safety Management* offers specific recommendations for proper data acquisition, storage, and management. *Section 7. Closure* presents the results and conclusions and proposes future research directions.

2. DATA SOURCES

In the first phase of this project, potential data sources were identified that could be used to improve current safety management practices. This section introduces the investigated data sources and provides suggestions on their usability and reliability.

2.1 Data Sources Evaluation

This project involved a search of potential safety-related data sources. The following subsections found 16 potential data sources useful for safety management, along with descriptions of the data, their potential value (general relation to roadway safety management) and the decision of whether they were selected for analysis in this project. Links to the data sources are also provided when available.

1. Traffic Count Database System (TCDS).

   Link
   https://indot.ms2soft.com/tcds/tsearch.asp?loc=Indot&mod=

   Description
   This data source is maintained by INDOT and provides hourly traffic counts on both permanent detector stations and selected 48-hour traffic survey spots in the Indiana roadway network. The permanent stations are located on freeways/interstates and state roads and provide

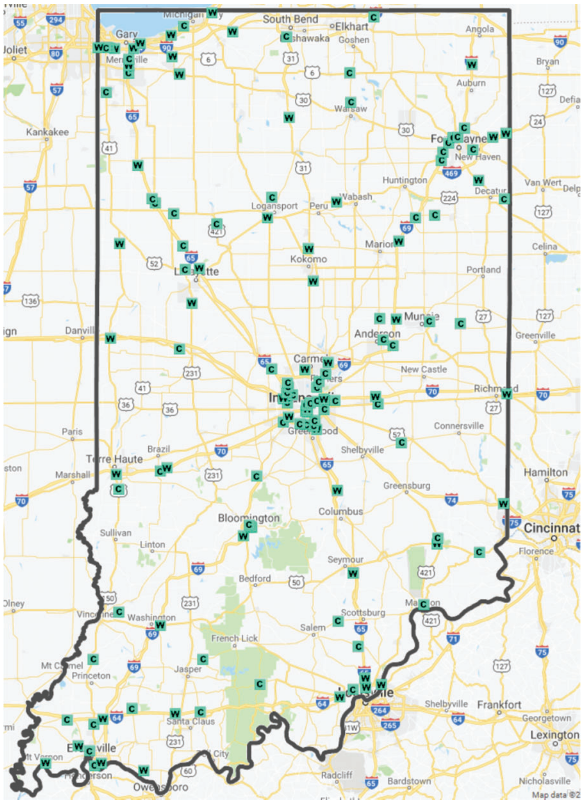


Figure 2.1 Permanent traffic count stations in Indiana.

consistent and usually reliable traffic counts (Figure 2.1) while the traffic surveys, which are performed irregularly, are used as complementary information to the permanent station data. This data source has high coverage on freeways and state roads, but there is no data on most local roads.

The traffic count data from permanent stations usually include 15-minute interval traffic counts specified by lane, traffic direction, and vehicle type (passenger car and truck). Some of the data also include the distribution of speeds, but these speeds are presented here in 5-mph bins.

Usage
Reliable and consistent hourly observations of traffic counts (volume) from permanent stations are the basis for estimating hourly volumes on roads, which is critical as the exposure factor in safety analysis. Although some of the detectors at these permanent stations provide speed data, the data is not consistent, and the precision (5 mph) does not meet the expected resolution.

Decision
SELECTED. The hourly traffic counts on permanent stations will be used to estimate the traffic volume prediction models, which will be applied to other segments without hourly volumes but with AADT values.

2. Automated Reporting Information Exchange System (ARIES)

Description
ARIES is the State of Indiana's crash repository. The crash data are generated through first responder crash reports and are collected within ARIES. Data are available from 2007 to present. These data include crash details such as vehicle information, road conditions, crash severity, weather conditions, location, date, and time.

The Indiana Traffic Records Coordinating Committee (TRCC) offers continuous input into the formulation and updating of ARIES. Its member agencies include the Indiana State Police (ISP), local and county police departments, INDOT, the National Highway Traffic Safety Administration (NHTSA), the Department of Natural Resources (DNR), the Bureau of Motor Vehicles (BMV), the Indiana Criminal Justice Institute (ICJI), the Fatal Accident Reporting System (FARS), and the ISP Commercial Vehicle Enforcement Division.

Usage
This data source serves as the basis of Indiana's roadway safety management practices.

Decision
SELECTED.

3. Highway Performance Monitoring System (HPMS)

Link
https://www.fhwa.dot.gov/policyinformation/hpms/shapefiles.cfm

Description
This data source is an inventory of U.S. highways, which is updated every year, is maintained by the Federal Highway Administration (FHWA). This inventory includes data on the functional classification, pavement condition, speed limit, cross-sectional, and longitudinal geometric characteristics of each road segment. The records are a combination of universal and sample data that depend on the hierarchy of the observed road. The higher the hierarchy is, the higher the resolution of the data.

Usage
The geometric characteristics of road segments can be used as explanatory variables in safety models. With high coverage and good maintenance from FHWA, this data source could support system-wide safety management analysis for interstates/freeways.

Decision
SELECTED.

4. National Performance Management Research Data Set (NPMRDS)

Link
https://ops.fhwa.dot.gov/perf_measurement/

Description
NPMRDS is a historical archive of average travel times in five-minute increments covering the National Highway System (NHS) and includes data from 2012 to present day. NPMRDS offers free access to state DOTs, and it is easy to maintain and has already been used in traffic efficiency/delay/performance analysis.

Before 2017, the passenger times were based on a collection of HERE probe data, sourced from a variety of consumer sources; and the freight times are based on probe data from ATRI. The variables included in the

NPMRDS are TMC code, state, county, date, epoch, travel time for all vehicles/passenger vehicles/freight vehicles, length of TMC, latitude and longitude, and corresponding shapefiles based on NHS. Some values were missing.

After 2017, NPMRDS turned to INRIX for travel time data. According to FHWA webinars information, the latest NPMRDS (2017 and 2018) data are also 5-minute increments and have separate passenger vehicle and freight vehicle travel times, but 87% of the new TMC segment length are within 0.1 mile compared to the 2-mile average length in the former NPMRDS data (2012–2016). The latest NPMRDS data (2019) have vehicle counts at 5-minute increments except for the travel time data.

Usage
The relationship between various definitions of speed and safety have been widely investigated and confirmed in previous research. The speeds provided by NPMRDS are high-resolution average mean speeds on segments; and the relation of these speeds to safety was promising in preliminary analysis. Moreover, along with the traffic volume, the "density" of TMC segments can be derived, which is a good estimate of the congestion level and could also have important effects on safety.

Decision
SELECTED. This data source covers mostly interstates and freeways so it could be used in the analysis of freeway segments.

5. INRIX

Link
http://inrix.com/wp-content/uploads/2018/10/Indiana-Case-Study_FINAL.pdf

Description
INRIX provides their speed data through JTRP with INDOT to evaluate the Indiana roadway system. This data source is currently maintained by INRIX but is available through JTRP.

According to the INRIX website, the GPS probe vehicle-based speed data are collected every 60 seconds and the average length of the data collection segments was 0.5 mile in 2014 through 2017. Because the INRIX speed data are collected from GPS probe vehicles or cell phones, there are certain time periods for which there are no observations, but the company claims to have interpolated these observations with prediction of historical data.

The INRIX speed data table includes location (INRIX segment ID), date, epoch, length, speed, and data quality index (a three-level categorical variable indicating the number of probe vehicles used to estimate the speed could be served as a proximity for volume under low volume circumstances), but the speed is not separated by vehicle type.

This data source could be viewed as an "advanced" version of NPMRDS in terms of resolution and coverage. It provides speed data with higher temporal resolution (consistent 1-minute data collection interval), spatial resolution (consistent average 0.5-mile segment length), and higher coverage (most of the state roads and some of the local roads). One challenge of using this data source would be the relatively large data size and computing demands.

Usage
Like NPMRDS, this data source is promising for analyzing the relationship of the speed to roadway safety. The data could be used to analyze more micro-level traffic state changes and their relationship to safety.

Decision
SELECTED. This data source provides the possibility of analyzing state roads and some local roads that are not covered in NPMRDS and therefore the safety of signalized intersections.

6. INDOT Traveler Information System (Traffic Wise–CARS 511)

Link
https://indot.carsprogram.org/#roadReports?timeFrame=TODAY&layers=roadReports%2CwinterDriving%2CweatherWarnings%2Cflooding%2CallReports

Description
The INDOT CARS program contains information about road conditions, closures, and width/weight restrictions. This information along with weather warnings, provides traffic speeds to motorists in Indiana through an INDOT interactive map (Figure 2.2). The original data are not available through their webpage, but they should be stored on INDOT servers.

Usage
In past safety management research projects, road closures and width restrictions were rarely considered when calibrating safety performance functions or building safety models due to the difficulty of assessing them for a long time (crash analysis time period). Consistent roadway conditions therefore were assumed when performing safety management. However, road closures and restrictions can influence exposure to crashes, and some problematic road segments may have been identified as safe if they were closed frequently for maintenance. This data source thus could be used to correct those misclassified risky segments or intersections and improve the quality of current safety management system.

Decision
NO. This data source is useful but departs from the major analysis framework of this report. Furthermore, the data were received too late to be included due to this project's time constraint so it will be analyzed in the follow-up project.

7. Waze (user reported events)

Link
https://www.waze.com/en/events

Description
Waze is a widely used navigation app that collects users' locations when they are using the app and allows them to report events such as a crash, police car enforcement, alert, and traffic jam. Table 2.1 shows the Waze coding for an alert event as an example. The Waze app records the location and time stamp of the event, the user's

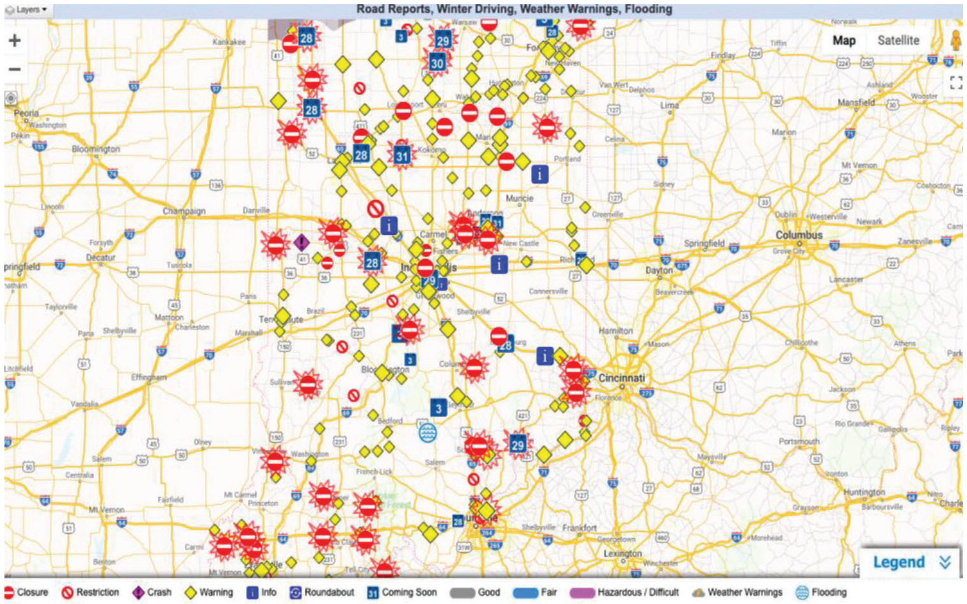


Figure 2.2 INDOT traveler information system.

TABLE 2.1
Waze alert event coding format

Element	Description
PubMillis	Publication date
Location	X Y–long-lat
Magvar	Event direction (driver heading at the report time)
Type	Event type
Report description	Report description by user
Street/city/county	Street/city/county names
Road type	Road type (coded)
Report rating	User rank between 1–6
Jam uuid	If the alert is connected to jam–jam ID
Reliability	Reliability scale

description, the users' ranking, whether the alert is related to a traffic jam, and the estimated reliability of the event. Because the users are reporting these events voluntarily, the quality of the "report description" may vary and sometimes unreliable. Waze established their Connected Citizens Program (CCP) and cooperates with world-wide users and city planners to provide data to analyze the traffic performance if the government agencies are cooperating with them.

Usage
Waze user-reported crashes provide a possible resource for unreported crashes since the crashes are usually reported by other road users instead of the drivers involved in crashes. However, the reported crashes by Waze users also can be unreliable because serious crashes that cause traffic jams likely will be investigated and reported by law enforcement.

Other user reported events, like jams and alert could be useful to cross validate the traffic operation data, while not confidential enough to conduct a network system level safety management analysis.

Decision
NOT AVAILABLE. INDOT currently does not cooperate with Waze; therefore, it would be difficult and cost-prohibitive to obtain this data.

8. High-Resolution Lidar Trajectory Data from TScan

Description
A TScan system is being developed by the Purdue University Center for Road Safety to collect high resolution Lidar/image information to identify traffic conflicts with trajectory recognition technology. The system is capable of recognizing traffic conflicts automatically at the intersection or short segment levels. By analyzing traffic conflicts or vehicle/pedestrian trajectories directly, TScan users can derive rich information about traffic safety problems in any investigated area.

Usage
The TScan trajectory data are promising in two aspects: safety audit and countermeasure effectiveness evaluation. On-site safety audits are usually carried out by experience

safety experts; and while historical crash data are useful in these audits, fast traffic conflict monitoring results could be a strong supplement to traditional practices. Countermeasure evaluation is usually time consuming since engineers must wait for at least 3 years to accumulate enough crash data to conduct the before and after comparison. The TScan high-resolution trajectory data collected in several days before and after the implementation of a countermeasure could provide reliable results and thus quick feedback on whether a countermeasure is working.

Decision
WILL BE CONSIDERED WHEN AVAILABLE. The TScan system will be applied to certain intersections as soon as it is ready to test the feasibility of the proposed methods.

9. Electronic Citation and Warning System (eCWS)

Link
https://www.in.gov/judiciary/admin/2655.htm

Description
The eCWS system give law enforcement officers statewide the ability to produce tickets electronically at the time of a traffic stop. The system is currently being used by ISP officers statewide and officers in over 490 other law enforcement agencies. The application is available in a laptop computer version for patrol cars. A hand-held portable version for motorcycle officers and water or foot patrol officers is currently being used by ISP motorcycle troopers, Indiana Excise Police, Indiana DNR, and over 20 local law enforcement agencies in the field. Data can be transmitted electronically to the appropriate law enforcement agencies and courts as well as state and federal agencies.

Considering the privacy issue, the data obtained from the BMV is not complete and may contain fuzzy location information (e.g., only county ID, no street/road names). Bloomington, Indiana provides 2 years of citation data (2016–2018) to the public (https://data.bloomington.in.gov/dataset/citations), which include the date, time, location, district, cited person's age/sex/race, and offense code. The provided location records are generally precise enough to identify certain street segments or intersections. These data were used in a preliminary case study to investigate the citation and crash relationship.

Usage
Because dynamic enforcement information is usually impossible to obtain, it has never been used in any safety performance analysis before now. This project set out to prove whether distribution of citations could be evidence of any underlying interaction effects between enforcement and safety performance.

Decision
DISCARDED. In the preliminary analysis using Bloomington citation data, the clustering of citations and crashes were confirmed; however, the marginal effect of citations was too small to make clear inferences of crash occurrences using citation data.

10. INDOT Roadside Weather Information System (INDOT-RWIS)

Links
https://rwis.indot.in.gov/

Description
INDOT maintains a roadside weather recording system (RWIS) across Indiana (44 stations in total). The system has atmospheric and surface sensors that update the conditions every 10 minutes for air temperature, dewpoint, humidity, wind direction, wind speed, and especially pavement temperature. The RWIS is helpful when decisions must be made to broadcast travel warnings and when to begin snow/ice removal.

Usage
The uniqueness of the RWIS roadside weather data lies in the pavement surface sensors, which are specially designed to monitor pavement conditions and can provide critical information about crash-prone icy pavement conditions in winter. This data source is limited in terms of stations locations as most of them are around bridges of interstates or freeways, making it difficult to interpolate and expand the conditions data to the entire road network.

Decision
DISCARDED IN THIS ANALYSIS. However, the RWIS should be considered for safety management practices that focus on bridges. It would be more appropriate for analyzing safety performance in a separate project specifically for bridges under adverse weather conditions.

11. INClimate

Link
https://iclimate.org/

Description
The Indiana State Climate Office (INClimate) is the state archive of official daily and hourly weather observations recorded throughout Indiana, which is administered by state climatologists at Purdue University. INClimate not only assists in providing climate observations and summaries on-line but also interprets and applies this data to solve climate related problems at hand for businesses and government agencies. The available data are air temperature, precipitation, humidity/dew point, wind speed and direction, soil temperature, degree days, air pressure, and solar radiation. The temporal resolution of these data are as relevant as 30-minute intervals at the county level depending on the various data sets.

Usage
Most of the weather data provided by INClimate are daily weather records or rough data by stations (neither state-wide nor systematic). The 30-minute county level weather data is the most relevant. This data source is promising for analyzing the relationship between crash risk and temporal climate changes.

Decision
SELECTED. The INClimate staff responded in a timely manner and provided us with high quality weather data (gridded hourly precipitation, gridded daily temperature, and station-based hourly temperatures) from 2014–2019.

12. Commercial Sources of Households and Businesses Data

Description
The commercial marketing data providers sell detailed location, households, and business information in certain areas, which is mainly aimed to help businesses make marketing decisions and is usually available for downtown areas. There are various data providers in this market and they usually charge based on the size of the area and the time periods requested.

Usage
The household and business information could be used to represent land-use intensity and to infer pedestrian intensity. Therefore, these data potentially could be useful in estimating pedestrian crash exposure.

Decision
NOT SELECTED FOR THIS ANALYSIS. However, these data could be useful in other pedestrian crash related projects.

13. StreetLight

Link
https://www.streetlightdata.com/

Description
StreetLight is a company that provides big data about transportation mobility by using smartphones as sensors to measure activities on all streets. According to their website, they have collected multi-mode data for five million miles of roadway and 10 million blocks covering 100% of U.S. traffic analysis zones. The price for the data varies by the scale of the area, precision required, and time periods.

Usage
According to a company consultant, they can obtain the movement patterns of people in selected areas, which would be helpful in determining traffic volume by directions and turning volumes at intersections (the absence of turning volume was identified as a primary limitation according to the intersection analysis of this report).

Decision
NOT SELECTED FOR THIS ANALYSIS. Due to time constraints, StreetLight could not be used in this project, but it will be considered the follow up project.

14. Waze (vehicle trajectory data)

Links
https://medium.com/louisville-metro-opi2/how-we-do-free-traffic-studies-with-waze-data-and-how-you-can-too-a550b0728f65
https://gduer.github.io/Collision-Prediction-in-Louisville-KY/

Description
This vehicle trajectory data is also provided by the Waze CCP, which is the same source as the 7th data type above. One good application example of their vehicle trajectory data was found in GitHub. Researchers from the University of Pennsylvania obtained Waze trajectory data with the help of the Louisville Metro and used it to predict the safety performance of Louisville. They shared their data processing procedure and report. According to their statements in the report, the Waze trajectory data was low-resolution and was sampled at an average rate of one point per 2 minutes. However, the safety performance measured by their study focused on large traffic zones rather than specific infrastructures and was not for safety management system purposes.

Usage
The low-resolution trajectory data from Waze may be more suitable to identify OD information or general safety performance but not for safety management improvement projects, which require much more precise location accuracy. Another limitation of the data is the selection bias that young people may be more likely to use such apps and only their driving trajectories would be tracked in terms of speed, acceleration, and deceleration.

Decision
NOT AVAILABLE. INDOT currently does not cooperate with Waze; therefore, it would be difficult and cost-prohibitive to obtain this data.

15. National Weather Service

Link
https://www.weather.gov/

Description
The National Weather Service (NWS) is an agency of the U.S. federal government that is tasked with providing weather forecasts, warnings of hazardous weather, and other weather-related information to organizations and the public for the purposes of protection, safety, and general information. It is a part of the National Oceanic and Atmospheric Administration (NOAA) branch of the Department of Commerce.

The weather data provided by NWS are usually raw without pre-processing. For example, there are four stations around airports in Indiana with 15-minute weather data and hundreds of county level weather data stations that measure daily temperature and precipitation only. To use these weather data, researchers must implement meteorology methods to estimate the hourly temperatures or precipitation rates around local road infrastructures.

Usage
Although there are methods to interpolate and predict more precise weather conditions from different levels of station data in meteorology, too much work is required and is not in the scope of this project. Cooperating with professional researchers in meteorology area, such as INClimate in the 11th data source would be a better choice.

Decision
NOT SELECTED. INClimate will be used instead.

16. Weather Underground

Link
https://www.wunderground.com/

Description
Weather Underground provides historical weather data freely through their website at zip code area precision. However, the provided data are only reliable for daily weather reports, but not for hourly weather analysis.

Usage
A great deal of effort is required to gather useful and consistent weather information from this data source, but it could be used as a reference in analysis.

Decision
NOT SELECTED. INClimate will be used instead.

2.2 Summary of Data Sources

Table 2.2 shows the data inventory discussed in this section, which includes traditional data such as crashes, geometric characteristics, and traffic counts, and emerging data such as time-dependent speed data, weather data, and user-reported data. Among the 16 listed data sources, 7 included useful data selected for analysis in this project.

- *Crash data:* ARIES provides rich information about crashes, including the involved vehicles and the occupants. The system is well maintained and used for the current Indiana roadway safety management program.
- *Traffic volume data:* The INDOT TCDS provides consistent and reliable observed traffic counts for the Indiana road network. The data are collected at 15-minute intervals and is specified by lanes and vehicle types. The major limitation to TCDS is its low coverage of local roads, which prevents analysis of local segments or intersections.
- *Speed data:* Two data sources were selected for speed in this project: NPMRDS and INRIX. NPMRDS provides good coverage on freeways/interstates and consistent segmentation across years. Although its temporal resolution is only collected at 5-minute intervals compared to the lower than 1-minute intervals of INRIX, it is sufficient for hourly speed analysis. INRIX provides much better coverage of the road network by providing speed features on state roads and some local roads. The major limitation of INRIX is its frequently changing segmentations. INRIX updates their segmentations every 6 months to increase the roads and to correct wrong segmentations. These updates are good in terms of improving the data quality year by year, but they introduce a great deal of work to link their segmentations every time to other data sources used in this analysis. Additional limitations of INRIX speed data include systematic bias compared to traditional loop detectors and the use of historical speed data when actual speed measurements are not available (Sharma et al., 2017).
- *Geometry data:* HPMS is as a reliable source of freeways/interstates geometry data. It is updated every year and maintained by Federal Highway Administration. However, since this data source only covers freeways/interstates, the geometry parameters of other road types must be collected manually through Google Maps. The research team received the INDOT road inventory geodatabase recently near the end of this project. The road inventory geodatabase provides good coverage of INDOT roads. Although the INDOT collected data is not as rich as HPMS data, it does include the basic information of speed limit, AADT, barrier settings, number of lanes, and street light locations. This data source also was partly used in this project.
- *Weather data:* Three data sources were identified at the beginning of this project, but INClimate was finally selected for the analysis. The weather data preprocessing involved a large volume of meteorology knowledge. Among the three possible data sources, INClimate is the only one that applies meteorology methods and provides gridded virtual station weather estimations. These gridded estimations further can be interpolated to the

TABLE 2.2
Data sources summary

ID	Data Sources	Collected Data	Available in this Project	Reliable, Proper Maintained	Selected for Analysis
1	Traffic Count Database System	Volume	×	×	×
2	Automated Reporting Information Exchange System	Crash	×	×	×
3	Highway Performance Monitoring System	Geometry	×	×	×
4	National Performance Management Research Data Set	Speed	×	×	×
5	INRIX	Speed	×	×	×
6	INDOT Traveler Information System	Road closure, warnings	×	×	
7	Waze (user reported events)	User-reported events		×	
8	TScan	Lidar trajectory	×	×	×
9	Electronic Citation and Warning System	Citation		×	
10	INDOT Roadside Weather Information System	Roadside weather	×	×	
11	INClimate	Weather	×		×
12	Commercial Sources of Households and Businesses Data	Marketing, land use		×	
13	StreetLight	Turning volume		×	
14	Waze (vehicle trajectory data)	Trajectory		×	
15	National Weather Service	Weather	×	×	
16	Weather Underground	Weather	×	×	

local weather conditions on segments or intersections (refer to Section 4.3). The structure of the selected data (variables and description) are provided in Appendix A.

3. METHODOLOGY

3.1 Analysis Framework

In traditional roadway safety management practices, the safety performance of roadways is evaluated require years of safety data accumulation, such as crash records. Post-crash safety evaluation is costly and inadequate for possible quick changes to roadway infrastructure considering the rapid development of connected and autonomous vehicles. Also, safety countermeasures are not intuitive because "hot spot analysis" is insufficient for detecting the causes of the aggregated crashes.

The approach proposed in this report departs from traditional safety analysis. The probabilities of crashes were estimated by considering temporary conditions in hourly intervals with and without crashes. The proposed approach begins with collecting the high-resolution time-dependent data discussed in Chapter 2. Using the traffic volume, speed, and weather conditions in hourly intervals, the probabilities of crashes were investigated at a microscopic level and the factors that lead to higher crash risk identified with the models, thus prompting possible countermeasures.

The proposed analysis framework consists of the following four steps, as shown in Figure 3.1.

1. Estimate the hourly operational data (volume, speed, temperature, precipitation) and collect the fixed geometric setting parameters.
2. Link the various sources of data by time stamps and spatial proximity.
3. Develop statistical models to predict the probability of crash at the hourly level.
4. Identify high risk conditions from the model outputs.

The proposed analysis framework is a general approach and the operational conditions will be different for different roadway components. To demonstrate the applicability of the proposed approach, two typical roadway components, rural freeway segments and signalized intersections were analyzed in this project. The traffic on rural freeway segments, which usually experiences unimpeded flow, are prone to run-off-road crashes. Therefore, barriers and horizontal curves were carefully recorded in the data collection process. For signalized intersections, the traffic on the approaching segments to the intersections (stop and go patterns) and the traffic within the intersections (yield to through traffic) are quite different, so the crashes linked to intersections were divided into two categories: (1) same approach crashes and (2) different approach crashes.

Although both roadway components were analyzed with the same type of statistical models, as explained in Section 3.2, given the differences in sampling, data collection, data linking, and model interpretation, they are discussed separately in Sections 4 and 5.

3.2 Statistical Modeling

The safety risk was analyzed at the disaggregated level (hourly). Considering the rarity of crashes, it was more reasonable to model the hourly probability of a crash on a segment or an intersection, rather than crash counts. A binary response was assumed in the model (i.e., the crash happened or not) in a 1-hour window. In one interval, there were more crashes than one on the same road element. In this case, the hourly record was duplicated with the same attributes so that the effects were counted twice in the analysis.

To model a binary response, logistic regression is the commonly used model, which is advanced insofar as explaining the odds ratios of different effects. In this project, a two-step modeling approach was proposed: (1) use a probability of crash model to estimate the response "crash or not" and (2) use a probability of severe crash model to estimate the response "severe or not given crash happened."

$$\log\left(\frac{P}{1-P}\right) = \eta = X\beta$$

$$P = \frac{e^{\eta}}{1+e^{\eta}}$$

The statistical settings of the proposed models are shown in the above equations, where P is the prob-

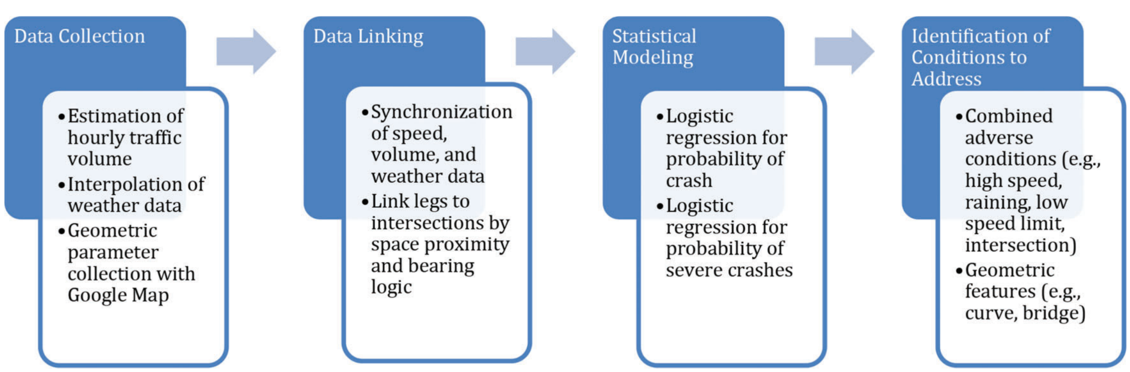


Figure 3.1 Analysis framework.

ability of crash/severe crash and X represents the safety-related factors (e.g., mean speed, volume, freezing temperature, etc.). SAS software was used to estimate the logistical models (Allison, 2012).

4. RURAL FREEWAYS

In this chapter, the safety analysis results for rural freeways are presented and their practical implications for safety management discussed. Data preparation tasks such as sampling, missing values imputation, and aggregation are discussed in detail. The specific sections include sampling description, estimation of hourly traffic volumes from vehicle probes and other factors, hourly temperatures interpolation, data collection of roadway characteristics, data linking process, and statistical analysis results.

4.1 Sampling

There are 2,546 one-way miles of rural freeways in Indiana. A sample of 133 miles of unidirectional rural freeways was selected based on data availability. Traffic volume data availability was the most limiting factor in selecting road segments. High-quality volume counts are essential for safety analysis. Only 9 out of the 70 statewide permanent loop detectors were found on rural freeways. The spatial distribution of the sample is presented in Figure 4.1. Out of the 133 unidirectional miles, there are 25 miles on I-64, 23 miles on I-65, 39 miles on I-69, 40 miles on I-70, and 6 miles on I-74.

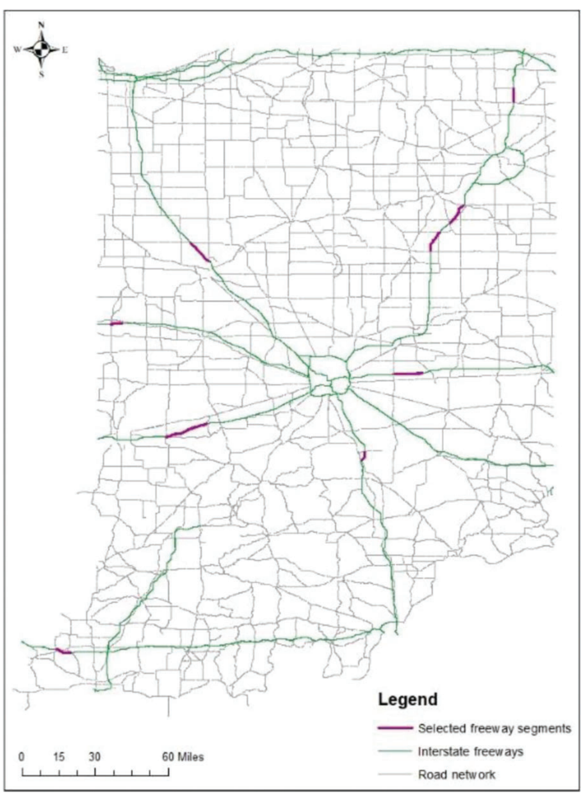


Figure 4.1 Selected rural freeway segments.

The selected road sections were divided into 532 segments with an average length of 0.25 miles (1,320 ft.), which is the shortest possible segment length that is suitable for safety analysis after considering various factors. First, individual vehicles, on average, spend 12–13 seconds on each analysis segment. Second, these road segments are homogeneous, and roadway characteristics tend to be constant. Significant road changes, such as bridges, tend to fall inside these sections. Last, on monotonous highways, it has been found that drivers can see as far as 1,500 ft.

There were 2,091 crashes assigned to the selected rural freeways from 2014 to 2018. Based on the crash severity, 85.8% of the crashes were classified as property damage only, 6.6% were potential or minor injury, and 7.6% were fatal or incapacitating. On average, each sample segment had 3.9 crashes over 5 years. The number and distribution of the crashes were suitable for statistical safety analysis.

4.2 Traffic Volume Estimation

Traffic volume plays a crucial role in safety analysis since it reflects the number of vehicles exposed to a potential crash. Yet, traffic volume data are expensive to collect and requires installing and maintaining counting stations. Traditionally, traffic volume data are collected using different devices, such as loop detectors, magnetic tubes, traffic cameras, and Bluetooth stations. Once the data are collected, it needs to be checked for quality control and stored in a data system accessed by the end-user. Recent developments in smartphones and other GPS instruments and increasing use of them have produced an opportunity for traffic engineers to forecast traffic conditions in real-time using vehicle probes.

Various datasets have been assembled to record the travel time along selected road sections. One of these datasets is FHWA's NPMRDS, which was established in 2014 and later updated in 2017. NPMRDS condenses travel times along all the road segments in the NHS. Road segments for which travel times are reported are usually defined between two consecutive ramps. Three travel times are reported every 5 minutes for all vehicles, passenger cars, and trucks. Unlike INRIX speed data, NPMRDS is purely observational; and no historical data is included.

Even though the concept of estimating traffic volume from vehicle probes is not new, this project provides one of only a few analyses with observational data. In 2017, Anuar and Cetin proposed a methodology to estimate traffic flow on freeways using probe vehicle trajectory data and shockwave theory (Anuar & Cetin, 2017). Their microsimulation showed that the estimated volumes had an average error of -4% when there was no congestion and 0% when analyzing congested traffic. Such efforts are becoming more common as probe vehicle data become more widely available.

In this project, two sets of models were developed to estimate the hourly traffic volume. The first set of models uses all 70 statewide permanent loop detectors to assess the prediction accuracy of multiple factors, including vehicle probes, roadway features, operating speed characteristics, weather conditions, and time-dependent indicators—the previous permits using alternative data sources to predict hourly volumes on roads without counting stations. The second set of models uses a similar set of predictors to impute missing volumes for the selected sample. Missing information is a common issue due to detector system maintenance, road construction, and other unexpected disruptions. For example, in the modeling dataset, 34% of the 788,400 observations had missing volumes.

4.2.1 Out-of-sample forecast

Hourly traffic volume was estimated based on vehicle probes density, roadway features, travel speed characteristics, weather variables, and seasonal indicators. Data from a sample of 138 road segments were selected. These segments have permanent traffic monitoring stations, with 70.3% of the stations located on urban roads and 29.7% on rural roads. The distribution of the selected road segments by functional classification is presented in Table 4.1.

Hourly volume data from 2017 was linked to the speed, weather, and roadway features. The working dataset was comprised of a total of 628,227 observations. The summary statistics of the available data are presented in Table C.1.

Multiple linear regression analysis was used to connect the predictor variables to the observed hourly volumes. The dependent variable was transformed by taking the logarithm of the volume. This transformation was necessary to agree with the model assumption of normally distributed residuals. The resulting predicted volumes were converted back to the original variable; and a stepwise selection procedure was used to select the most promising predictors to maximize the models' prediction power. Linear correlations of the significant predictors were checked to avoid multicollinearity issues. In each case, after an initial model was fit, the outliers were identified and removed to further improve the model's prediction power.

TABLE 4.1
Distribution of sample road segments by functional classification

Functional Classification	Number of Segments
Interstates	70
Other freeways and expressways	8
Principal arterials	54
Minor arterials	4
Major collectors	2

Four models were estimated: urban freeways, rural freeways, urban non-freeways, and rural non-freeways. Freeways include both interstates and other expressways; and non-freeways include arterials and major collectors. The least-squares estimates of the resulting models are presented in Table C.2 through Table C.5.

A 10-fold cross-validation approach was used to check the models' transferability to segments that did not have counting stations. Data from approximately 10% of road sections were removed for each fold, and the models were fitted with the remaining sites. The model prediction was later tested with the data from the removed locations. The proposed validation approach replicates the prediction accuracy's reality if the estimated models were applied to the entire system.

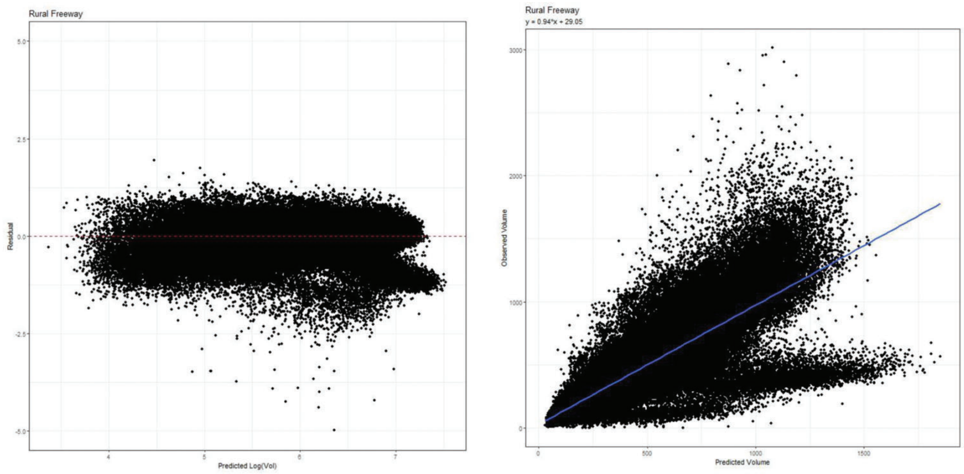
Four performance measures are proposed to evaluate the overall quality of the estimated models and are presented below:

- *Bias:* This is the estimated slope of the observed versus the predicted scatterplot. The closer this value is to 1, the lower the discrepancy. If this value is below 1, the model is underestimating the traffic volume; and bias values above 1 indicate an overestimation of the volume.
- *Residual standard error:* This is the standard deviation of the model residuals in vehicles per hour. It is the average number of vehicles that depart from the real traffic volume.
- *Percentage of volume:* This is the average value of the residual standard error divided by the observed traffic volume. In other words, the percent of the traffic volume that is taken by the residual term.
- *Adjusted R-squared:* This is the adjusted sum of squares of the model over the total sum of squares. This value is adjusted based on the number of variables included in the model.

The performance measures of the models are summarized in Table 4.2. Two methods are compared. On the one hand, in-sample prediction accuracy estimates the model performance if it were to be used for

TABLE 4.2
Summary of results volume forecast models

Road Type	Method	Bias	Residual Standard Error	Percent of Volume	Adjusted R-squared
Urban freeway	In	0.9617	502.4	22.00	0.8565
Rural freeway	In	0.9914	121.5	19.27	0.9004
Urban non-freeway	In	0.9399	169.7	28.76	0.8106
Rural non-freeway	In	0.9686	79.3	32.91	0.7411
Urban freeway	Out	0.5227	846.6	47.30	0.5913
Rural freeway	Out	0.7847	240.7	40.24	0.5667
Urban non-freeway	Out	0.3098	233.7	71.59	0.5229
Rural non-freeway	Out	0.7392	118.7	69.56	0.4197



(a) Residuals vs. Predicted Values (b) Observed vs. Predicted Volumes

Figure 4.2 Out-of-sample residuals of the rural freeway model.

the same population of roads. On the other hand, out-of-sample prediction accuracy uses the previously described validation method to estimate the model performance if it were to be used somewhere else.

The results indicate that the overall performance drastically decreased when using the estimated models to predict hourly volumes outside of the sample population; and this effect was significantly larger on urban roads compared to rural roads. A possible explanation of this phenomenon is that rural roads tend to be more consistent than urban roads. Another problematic element is the fact that vehicle probe counts are constrained for high volumes since the counts of 10 or more probes are grouped. It is recommended to work with INRIX and FHWA to obtain a better proxy for hourly volume, which may be the exact number of vehicle probes used to calculate speed.

The residuals diagnostic plots for rural freeways are presented in Figure 4.2. The right scatterplot helps check the assumptions of a homoscedastic normal distribution around zero for residuals. The left scatterplot displays the prediction potential; and the closer this plot is to 45 degrees (blue line), the higher the prediction power. As can be seen from both scatterplots, there were two groups of observations in our dataset. After clustering analysis, it was found that the lower group of marks were linked to road segments near Indianapolis. After removing this data, all the model assumptions were valid.

The results in this analysis were comparable to the results in the reported literature. In 2018, Hou et al. examined the feasibility of using sampled commercial probe data from INRIX combined with continuous counters to estimate hourly volume (Hou et al., 2018). They also included additional information, such as speed characteristics from GPS traces, road features, and weather conditions, as modeling inputs. A neural network model's resulting accuracy had an R2 varying from 0.61 to 0.94 with a median of 0.82, and a mean absolute percentage error (MAPE) ranging from 14% to 48% with a median of 27%. Additionally, the proposed analysis approach has an advantage over machine learning algorithms as it provides inferential conclusions by clearly stating the analytical form of the model.

4.2.2 Missing volume imputation

Missing traffic counts are common due to issues such as detector maintenance, road construction, and temporary system components failure. In this project, 18 freeways segments were selected for safety analysis. Volumes, speeds, weather variables, roadway features, and crash records were acquired and linked to the sample road segments. The working dataset was comprised of a total of 788,400 observations. There were 34% missing traffic volumes, 16% missing travel speed characteristics, and 2% missing hourly temperatures. Most notably, the missing traffic volume data was the main limiting factor.

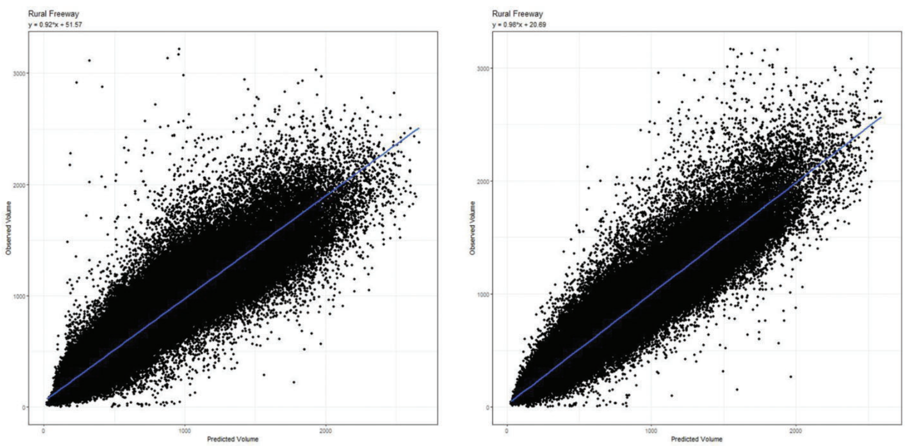
Following the statistical framework presented in Section 4.2.1, two regression models were used to impute the missing traffic volume records. The first model used data from 2014 to 2016, while the second model used data from 2017 and 2018. The reason for this separation the analysis into two periods was that the vehicle probe density from the NPMRDS speed dataset became available in 2017.

The summary characteristics of the two imputation models are shown in Table 4.3. From 2014 to 2016, 32% of the data were derived from the model. From 2017 to 2018, 36% of the observations were predicted using the model. The 2017–2018 model demonstrated a marginal improvement compared to the 2014–2016 model, which may have been due to the inclusion of vehicle probe densities as predictors.

The distribution of the models' type III sum of squares, which is the variability explained by each variable, is shown in Table C.6 and Table C.8. The distribution of the sum of squares ranks the relevance of each piece of information based on its contribution

TABLE 4.3
Summary statistics of volume imputation models

Characteristic	Model 1 (2014–2016)	Model 2 (2017–2018)
Total number of observations	473,472	315,360
Number of observations used in the model	303,277	214,209
Number of observations predicted	170,195	101,151
R-square	0.9144	0.9204
Average error	115.96 veh/h (19.33%)	119.96 veh/h (18.42%)
Bias	Slight underestimate	No significant bias



(a) (b)

Figure 4.3 Hourly volumes observed versus hourly volumes predicted with the imputation models: (a) period 2014–2016 and (b) period 2017–2018.

to the predicted hourly volumes. In descending order, the most pertinent variables were time indicators, AADT by type of vehicle, vehicle probes count, travel speed characteristics, and weather variables.

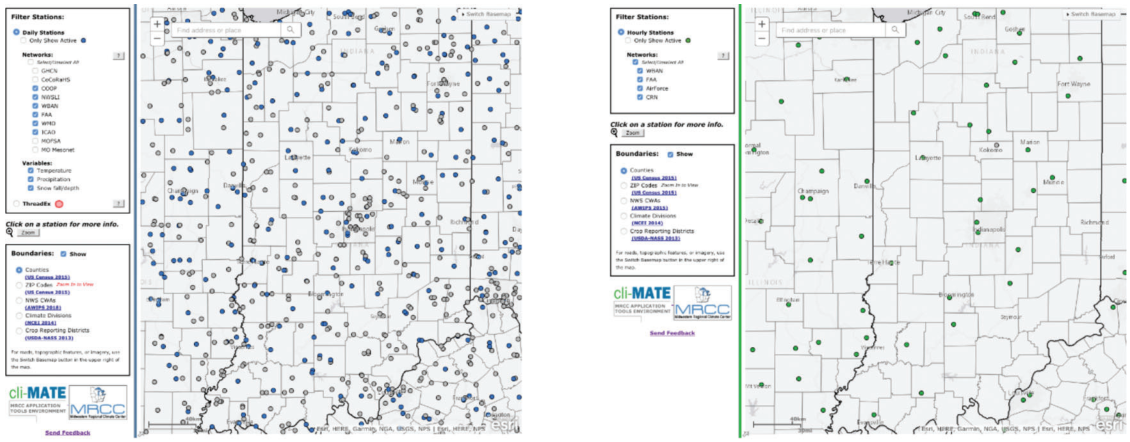
The parameter estimates of the resulting models are presented in Table C.7 and Table C.9. The factors that tend to increase the hourly volume include AADT for passenger cars, AADT for heavy trucks, vehicle probe counts, travel speed variability, intermediate traffic, and congested traffic indicator based on speed. On the other hand, the parameters that tend to decrease the hourly volume include precipitation, temperature, icy conditions, and downtrend speed indicator.

The models' scatterplots of observed versus predicted hourly volumes are presented in Figure 4.3. The statistical assumptions regarding homoscedastic normally distributed residuals, constant variance, and expected zero value were fulfilled. The 2014–2016 model presented a larger number of outlying observations above the blue line, which indicated that high volumes occasionally were underestimated. The results improved slightly in the 2017–2018 model with the inclusion of vehicle probe counts. However, to further improve the prediction power of the models, it was necessary to improve the quality of the raw data. In practical terms, the inclusion of unobserved factors, such as leading towards or against a major city or the improvement of existing variables, such as vehicle probe counts not as ranges but as the actual values, were among the enhancements. The models were found to be acceptable, and any error at that point was not expected to affect correct estimation of the overall safety effects.

4.3 Weather Interpolation

Temperature and precipitation are presumed to affect safety. Therefore, it is crucial to have a high-quality source of weather information. Traditionally, safety analysis uses data from weather stations. Weather stations collect and store climatological variables daily; and some selected urban weather stations near airports can produce hourly observations of weather conditions. Temporal aggregation is a fundamental limitation for disaggregate safety analysis; and a combination of data sources is used to overcome such restrictions.

NOAA offers publicly available weather data through its National Centers for Environmental Information (NCEI). Depending on the level of aggregation, different station densities can be found. Daily weather stations offer data on snowfall, snow depth, temperature (maximum, minimum, and average), and liquid precipitation. Figure 4.4a displays the available daily weather stations with the gray dots denoting inactive stations with some historical data and the blue dots



(a) Daily Weather Stations (b) Hourly Weather Stations

Figure 4.4 Map of weather stations across the region.

denoting active stations as of 2019. Hourly stations (Figure 4.4b) provide information about temperature and precipitation and are usually located at airports or near large cities.

INClimate is the state archive of official daily and hourly weather observations recorded throughout Indiana. INClimate maintains an online library of many recent daily and hourly observations from both manual and automated networks. Two of the datasets offered by INClimate are PRISM and multi-sensor precipitation estimates (MPE). PRISM is a gridded dataset with virtual stations spaced every 2.5 miles which contains information about hourly temperature. This data has interpolated observed temperatures spatially using statistical parameterization algorithms to account for terrain influences. The MPE dataset provides hourly liquid precipitation. Like PRISM, MPE's virtual stations are spaced every 2.5 miles. This dataset combines observational precipitation data with derived estimates from the Doppler radar network and satellites to offer a gridded characterization of precipitation. In total, there are 5,724 virtual gridded weather stations in Indiana.

The high density of the gridded weather stations network makes it possible to obtain weather condition variables reported near crash locations. On the other hand, traditional weather stations are more precise and reliable. Gridded daily data from PRISM and observed weather conditions from hourly weather stations were used in this analysis to interpolate hourly temperatures in a gridded fashion. The overall process is presented in Figure 4.5. This approach permitted exploiting the advantages of each data source to conduct the safety analysis. The resulting data were used in the traffic safety analysis of rural roads far from urban centers and weather stations.

4.3.1 Temperature interpolation process

The steps of the temperature interpolation process are as follows.

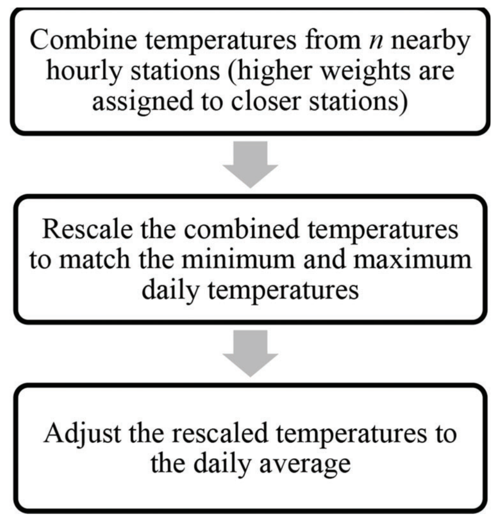


Figure 4.5 Temperature interpolation process.

Let t_h^s be the temperature at station s during hour h, d_s the distance from station s to the target road (or the nearest gridded station), t_{min} the minimum temperature during a day on the target road, t_{max} the maximum temperature during a day on the target road, and t_{avg} the average temperature during a day on the target road.

First, the distance weights are calculated for n close stations as

$$w_s = \left(1 - \frac{d_s}{\sum_i d_i}\right)/(n-1),\ i = 1, \ldots, n$$

where w_s is a weight for station s when combining temperatures from n stations, and d_s is the distance from station s to the target road.

Second, the combined temperature is estimated as

$$t_h^c = \sum\nolimits_s w_s t_h^s$$

where t_h^c is the temperature during hour h combined from n stations, w_s is the weight for station s, and t_h^s is temperature observed at station s during hour h.

Third, the combined temperatures are rescaled to match the minimum and maximum temperatures reported on the target road.

$$t_h^r = t_{min} + \frac{t_h^c - t_{min}^c}{t_{max}^c - t_{min}^c}(t_{max} - t_{min})$$

where t_h^r is the temperature in hour h rescaled and shifted for t_{min} and t_{max}, t_h^c is the temperature in hour h combined from n stations, t_{min} is the minimum temperature during a day on the road, t_{max} is the maximum temperature during a day on the road, t_{min}^c is the minimum combined temperature during a day, and t_{max}^c is the maximum combined temperature during a day.

The next step of adjusting the rescaled temperatures to the daily average temperature requires explanations. It is assumed that the total adjustment $24\left(t_{avg} - t_{avg}^r\right)$ needed for the rescaled temperatures is distributed among individual temperatures in proportion to the difference between the rescaled temperature and the closest measured extremum (minimum, maximum) temperature of the day. This way, the extremum temperatures are not adjusted since they are measured.

$$a_h = 24\left(t_{avg} - t_{avg}^r\right)\frac{\min\left\{\left|t_{min} - t_h^r\right|, \left|t_{max} - t_h^r\right|\right\}}{\sum_h \min\left\{\left|t_{min} - t_h^r\right|, \left|t_{max} - t_h^r\right|\right\}}$$

where a_h is the adjustment of temperature t_h^r to reconcile the daily average rescaled temperatures t_h^r, t_h^r is the temperature in hour h rescaled, t_{min} is the minimum temperature during a day on the road, t_{max} is the maximum temperature during a day on the road, t_{avg} is the average temperature during a day on the road, and t_{avg}^r is the average rescaled temperature.

The rescaled temperatures are adjusted to obtain the final values.

$$t_h^a = t_h^r + a_n$$

where t_h^a is the temperature in hour h adjusted for t_{avg}, a_h is the adjustment of temperature t_h^r to reconcile the daily average rescaled temperatures t_h^r, and t_h^r is the temperature in hour h rescaled.

4.3.2 Indiana case study results

Following the described interpolation procedure, gridded hourly temperatures were created from 2014 to 2018 in Indiana; and 98% of the hourly temperatures were successfully interpolated with a conservative rule of thumb on missing values. Each year, hourly weather stations were checked for missing information. Stations with more than 10% of the data missing were discarded from the analysis.

The total number of hourly temperature observations was 1,191,978. There were 210,390 missing values (15%); and while missing values behave differently depending on the time, some stations consistently present missing information and others do not provide any data at specific years. After discarding the weather stations with more than 10% missing data, the resulting dataset consists of 1,018,623 observations with 2% missing data. The number of weather stations and types of observations by year are summarized in Table 4.4. Each road was linked to three nearby weather stations. This process was repeated every year. The summary statistics of the distances are presented in Table 4.5.

TABLE 4.4
Missing hourly weather data by year

Item	2014	2015	2016	2017	2018
Number of valid stations	24	26	22	22	24
Available data	206,326	224,233	191,107	190,759	206,198
Total data	210,240	227,760	193,248	192,720	210,240
Missing data	2%	2%	1%	1%	2%

TABLE 4.5
Summary statistics of distances to nearby weather stations

Rank	Mean	Standard Deviation	Minimum	Maximum
1	17.13	6.45	6.78	37.79
2	33.11	8.54	15.32	81.16
3	40.56	12.01	26.80	90.47

Note: All distances are in miles.

Daily gridded temperature statistics (average, maximum, and minimum) were obtained from the PRISM dataset, which provides daily values for 5,724 virtual weather stations. These virtual weather stations have been consistent over the years. There are a total 10,417,330 observations in the dataset.

The method described in Section 4.3.1 was applied to interpolate the hourly temperatures using daily statistics and temperature measures at the three nearest weather stations. For each road, the nearest virtual weather station was assigned. A separate analysis found that five nearby stations' interpolated daily statistics changed 1 to 2 degrees °F, and the results were intuitive in most cases. However, in exceptional situations (<1%), the adjustment factor was susceptible and produced unrealistic temperature values. A revised version of the proposed methodology is presented below:

1. Consider n neighboring weather stations and form all possible combinations of the profiles from neighboring stations. For example, assuming three neighboring profiles, you will have seven varieties: three singles, three pairs, and one triplet.
2. Combine multiple temperature profiles using the method described in Section 4.3.1.
3. Rescale and shift each of the single or combined temperature profiles to match the minimum and maximum temperatures at the target location.
4. Adjust each of the rescaled profiles to the average value at the subject location using the method described in Section 4.3.1.

5. The final solution is the adjusted profile, which is (a) do not violate the min-max constraints, and (b) the variance of the adjustment is the lowest.
6. If no solution is found in step 5, select the profiles with the lowest variance of the adjustments and set the temperatures that violate the extremum at the extremum values. It might skew the average temperature at the subject location, but this error is negligible.

This approach weakens the arbitrary assumption about how distance affects other stations' adequacy for calculations and which stations should be considered. The smallest adjustment needed for the combined/rescaled profile shape indicates the best profile and station choice.

4.4 Roadway Features Data Collection

Information about safety-relevant road features was gathered from Google Earth's historical imagery. The collected Google data supplements the existing geometry data from road inventories. The extracted data was separated into two types: point features and longitudinal elements. This distinction is central for future data preprocessing and data analysis tasks.

Point features occur at a unique location. For a given road segment, they can be denoted by their frequency or density (frequency over distance). Some examples of such features include overpassing roads, crash attenuators, and signs.

Longitudinal elements arise along a portion of the road. Their beginning and end locations are straightforwardly marked. Depending on their extent, one longitudinal element could affect one or several road segments. Longitudinal elements can be represented by the target segment's proportion or the distance from the segment to the element ends. Some examples of such elements include road barriers, bridges, and curves.

Additionally, some roadway features were measured at specific points, but they were transformed into longitudinal elements. Some examples of such features include median width, shoulder width, and pavement type.

Finally, a small group of features were measured once between consecutive ramps and assigned to multiple road segments. These features include the number of lanes and posted speed limits by vehicle type.

A complete list of the extracted roadway features is presented below:

- Attenuators
  - Location
  - Type: fitch barrels or crash cushions
- Barriers
  - Beginning and end coordinates
  - Type: cable, concrete, or guardrail
  - Location: median or roadside
  - Distance offset from the road edge
- Bridges
  - Beginning and end coordinates
- Curves
  - Beginning and end coordinates
  - Radius measured at the curve's center without spirals
  - Deflection angle
  - Direction: left or right
- Number of lanes
- Lighting
  - Beginning and end coordinates
- Medians
  - Width measured from the road edges
- Overpassing roads
  - Location
- Pavement
  - Type: asphalt or concrete
- Prohibited U-turn signs
  - Location
- Ramp auxiliary lanes
  - Beginning and end coordinates
  - Type: tapered or parallel
  - Traffic direction: in or out
- Rumble strips
  - Beginning and end
  - Location: median or roadside
- Shoulder
  - Width
  - Location: median or roadside
- Speed Limits
  - Posted speed limit for passenger cars in mph
  - Poster speed limit for heavy trucks in mph

4.5 Data Linking

The information from the various data sources was linked to produce a dataset used for the statistical analysis. Figure 4.6 shows the data structure diagram, which presents the main characteristics of each data table and displays the linking process. A short list of variables is shown for each data table. Some variables were used as identifiers to connect data tables (in bold), while others were used as predictors in the statistical safety analysis.

In terms of crash data, the vehicle travel directions were linked to the crash records. The combined data table was later connected to the modeling dataset.

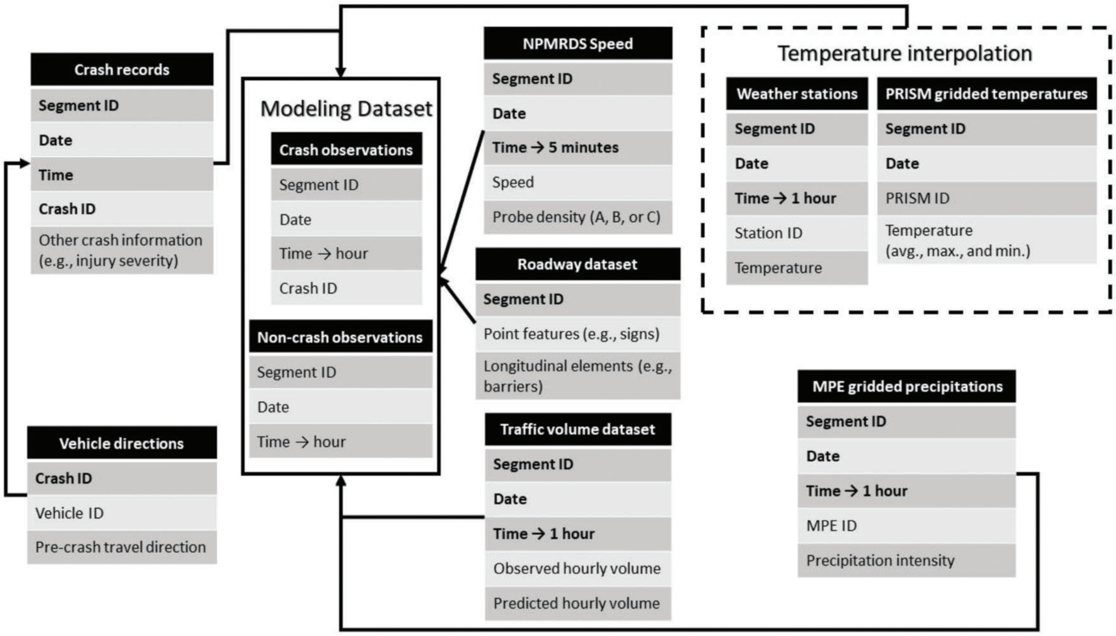


Figure 4.6 Building a sample for the rural freeway model.

Travel directions are a crucial element to evaluate the potential causes of the events that lead to the crash and associated injuries. For example, head-on collisions may significantly affect specific road elements, such as median, barrier, and ditch, compared to read-end collisions.

In terms of the crash contributing factors, information about the travel speed characteristics, roadway features, weather condition variables, and traffic volume was connected to the modeling dataset. A preliminary step was needed for temperature data where daily gridded temperatures and hourly temperatures measured at weather stations were combined to produce gridded hourly temperatures (see Section 4.3). Observations that occur 3 hours after a crash were removed. This step is necessary to avoid the use of post-crash observations in the statistical analysis. This type of condition does not represent the target phenomenon since it reflects the crash's impact on unaltered traffic conditions.

Once the data linking process was completed, the resulting modeling dataset was formed according to two types of observations: crashes and non-crashes. The summary statistics of the modeling dataset are presented in Table B.1. A 1:30 ratio between the two types of records was enforced via sampling non-crashes based on the number of crashes. Since crash observations are very uncommon compared to non-crashes, this sampling was needed to estimate the effects of the contributing factors on the hourly crash risk and injury severity. The constant term of the fitted model was later adjusted to represent the original conditions before the sampling.

4.6 Safety Analysis Results

The statistical analysis results are presented and discussed in this section. The estimated safety effects of time-dependent variables such as speed, traffic volume, weather conditions, and fixed roadway features such as geometry, pavement, and the speed limit are presented. The practical effects of the significant predictors on the hourly crash risk and the probability of severe outcomes also are summarized and discussed.

Disaggregate speed and volume data have been found to improve crash prediction on freeway segments. Dutta and Fontaine (2019) confirmed the benefits of using disaggregate data from multiple sources, specifically loop detectors and vehicle probes (INRIX), in a set of negative binomial regression models (Dutta & Fontaine, 2019). They found that including speed and volume increased the mean absolute prediction error by 11% for rural freeways and 20% for urban freeways. The current project elaborates further on this idea by including a broader group of speed, volume, weather, and roadway characteristics and testing their effects on the crash risk and injury severity. Two logistic regression models were used simultaneously to analyze the effects of multiple factors on crash occurrence and injury severity on a sample of rural freeway segments. Table D.1 and Table D.2 present the summary statistics of the maximum likelihood estimates for hourly crash probability and conditional injury probability, respectively. To further understand the practical impacts of the individual variables, the average marginal effects (AMEs) were calculated, which are summarized in Table D.3 and Table D.4.

A 95% confidence interval of each AME also is presented to check its statistical significance.

4.6.1 Crash risk

The hourly probability of crash was estimated as a function of traffic volume, roadway features, weather conditions, operating speed characteristics, and seasonal indicators. Table 4.6 summarizes the results, and a detailed description of the model estimates and marginal effects are provided in Table D.1 and Table D.3, respectively.

The following several factors were found to improve traffic safety. Overpassing roads reduced the hourly crash risk by 0.45% on average. A 1.78% decrease in collision probability was linked to a 1% increase in the segment proportion with a median guardrail. Segments with median barriers placed up to 30 ft. from the road edge experienced a 0.34% reduction in the crash risk. An additional median point hazard (barrier end) produced a 0.42% reduction in the collision probability. Increasing the average travel speed by 1 mph was linked to a 0.33% reduction in hourly crash risk. Lastly a 1% increase in speed trend was connected to a 0.72% decrease in the collision risk.

Other factors were also discovered to reduce traffic safety. For example, the exposure variables, such as hourly volume and AADT, were found, as expected, to increase the crash probability. The effect of hourly volume was found to be higher than the effect of AADT. A 1,000 veh/h increase in traffic volume was connected to a 0.82% increase in the hourly collision probability. A 1% increase in the length proportion of curve-to-segment for a moderate curve increased the hourly crash probability by 0.40%. A 1% increase in this proportion on a segment with a sharp curve was linked to a 0.62% increase in hourly crash probability. A 1% increase in the proportion of segment with a median cable barrier was connected to a 0.52% increase in crash probability. Likewise, a 1.87% increase in hourly collision probability was found when the segment proportion with a roadside guardrail increased by 1%.

Increasing the segment proportion with concrete pavement by 1% was linked to a 1.42% increase in crash risk. The presence of auxiliary lanes was found to be associated with the higher hourly collision probability.

TABLE 4.6
The crash frequency factors identified for Indiana rural freeways

Variable	Improves Safety	Reduces Safety	Counter Intuitive Result
Hourly traffic volume (1,000 veh/h)		×	
AADT cars (1,000 veh/day)		×	
AADT trucks (1,000 veh/day)		×	
Overpassing road	×		
Moderate curve (5.5 to 13.9 degrees)		×	
Sharp curve (14 degrees or more)		×	
Segment proportion with median cable barrier		×	
Segment proportion with median guardrail	×		×
Segment proportion with roadside guardrail		×	
Median barrier offset < 30 ft.	×		×
Segment proportion with concrete pavement		×	
Segment proportion with entering ramp		×	
Segment proportion with exiting ramp		×	
Posted speed limit reduced by 5 mph		×	×
Number of point hazards in the median	×		×
Number of point hazards in the roadside		×	
Average roadside shoulder width (ft.)		×	×
Light rain (precipitation < 0.098 in)		×	
Freezing temperature (temperature < 32°F)		×	
Average hourly travel speed (mph)	×		?
Hourly travel speed standard deviation (mph)		×	
Hourly speed trend–beta	×		
Downtrend speed indicator (beta < -5/60)		×	
Intermediate traffic		×	
Congested traffic		×	
Friday		×	
Sunday		×	
Year = 2014	×		
Year = 2015	×		
Year = 2017		×	
6:00am to 11:59am		×	
12:00pm to 5:59pm		×	
6:00pm to 11:59pm		×	

This effect was larger on segments with exit ramps. A 1% increase in the segment proportion with an exit ramp auxiliary lane produced a 4.11% increase in crash risk. The segments with a posted speed limit reduced by 5 mph experienced an increase in hourly collision probability by 0.81%. A single roadside point hazard was associated with a 0.34% increase in crash probability.

Rain and freezing temperatures were found to reduce the safety level. While light-intensity rain affected safety, no significant difference was observed between moderate to heavy rain and dry weather conditions. The latter finding may be due to the formation of a thin film of water on the surface, which, combined with dust, can cause safety issues related to low friction. As rainfall intensifies, however, the pavement surface becomes cleaner. Another possible explanation for these results is related to driver risk perception. Drivers trade off their perceived crash risk and aggressive driving behavior; and this trade-off may not start until a certain level of rain intensity is reached.

Speed characteristics also were associated with increases in hourly crash risk. A 1 mph increase in the standard deviation of hourly travel speed produced a 0.16% increase in the hourly collision probability. A speed downtrend that associates a growing congestion was connected to a 0.54% increase in crash probability. Intermediate and congested traffic based on the travel speeds were found to increase the crash risk. The effect was greater under moderate traffic conditions, where a 3.61% increase in the hourly collision probability was observed.

In 2015, Roshandel et al. (2015) conducted a systematic review of the impact of real-time traffic characteristics on freeway crash occurrence. They found 13 studies that were conducted from 2001 to 2012. The most typical data sources for these studies were loop detectors and vehicle trajectories. Their summarized results showed that crash probability increased with the increase in speed variations, speed difference between two consecutive segments, and average traffic volume. The most pronounced effect was produced by speed variance. A 1-mph increase was linked to a 22.6% increase in the crash occurrence odds ratio. This finding could be explained with dangerous transitions from low to high speed traffic. On the other hand, the increase in travel speed by 1 mph was found to decrease the odds ratio of a crash by 4.8%. The speed increase could be connected with the growing presence of non-congested high-speed conditions with fewer interactions among vehicles.

The intuitive results consistent with the current understanding of the crash mechanism may be used to support the selection of safety countermeasures. On the other hand, the counterintuitive results should also be retained in the models to improve the safety prediction of the model applied to Indiana roads and driver population. The counterintuitive results may be caused by variables missing in the model and correlated with the counterintuitive one.

4.6.2 Injury severity

The hourly conditional probability of severe injury given a crash was modeled in this project as a function of traffic volume, roadway features, weather conditions, speed characteristics, and other predictors. The maximum likelihood estimates and average marginal effects are summarized in Table D.2 and Table D.4, respectively. A summary of the results is presented in Table 4.7.

Roadway features such as barriers, curves, and lighting were found to affect injury severity. A 1% increase in the segment proportion with a median guardrail was connected to a 15.12% increase in the conditional probability of a severe crash' and increasing the segment proportion with a mild curve by 1% was linked to a 6.81% increase in hourly injury risk. Lighting was found to have a significant positive effect on safety; and increasing the segment proportion with lighting by 1% was associated with a 12.87% decrease in the probability of a severe crash. Lastly, reducing the posted speed limit by 5 mph produced a 7.02% reduction in the hourly injury risk.

Weather variables have been found to have intuitive effects on crash risk, but their effects on injury severity

TABLE 4.7
The crash severity factors identified for Indiana rural freeways

Variable	Improves Safety	Reduces Safety	Counter Intuitive Result
Segment proportion with median guardrail		×	?
Mild curve (3.5–5.4 degrees)		×	
Segment proportion with lighting	×		
Speed limit reduced by 5 mph	×		
Temperature (F)		×	
Light rain (precipitation < 0.098 in)	×		×
Freezing temperature (temperature < 32)	×		×
Average hourly travel speed (mph)		×	
Hourly travel speed standard deviation (mph)		×	
Downtrend speed indicator (beta $<$ -5/60)		×	
Intermediate traffic		×	
Friday	×		

are more difficult to interpret. The presence of light rain is related to a 2.48% reduction in hourly injury risk. Freezing temperatures produced a 1.19% decrease in the probability of a severe outcome crash. Lastly, a 10-degree increase in hourly temperature was linked to a 0.6% increase in the hourly injury risk.

Yu and Abdel-Aty (2014) analyzed crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. They found that the speed variation and steep grades increased the injury severity while snow and temperature reduced the probability of severe outcomes. The findings of the current project were similar to those of Yu and Abdel-Aty.

The negative effect of temperature on injury severity was interpreted as severe crashes that are less likely to happen during warm weather. During snow events, adverse weather conditions such as low visibility and high precipitation diminish the driving conditions; but at the same time, poor weather conditions may result in lower speeds and more cautious driving behavior on freeways.

Speed characteristics were successfully associated with crash injury severity. A 1-mph increase in the average hourly travel speed was linked to a 0.11% increase in the probability of a severe outcome crash. Increasing the standard deviation of hourly travel speed by 1-mph produced a 1.21% jump in the hourly injury risk. The highest marginal effect on safety was observed for the presence of downtrend speeds. A 4.34% increase in the conditional probability of a severe crash was associated with decreasing speed trends, usually resulting from growing congestion. Lastly, intermediate traffic was found to increase the hourly injury risk by 3.38%. No significant differences were observed between non-congested and fully congested conditions.

There were also counterintuitive results. Safety management insights are gathered from intuitive statistical findings with robust methodological support. Counterintuitive effects can only help improve the overall model predicting power. An additional limitation of crash injury severity models is related to the crash data. Crash probability models use crashes and non-crashes, while injury severity models are restricted only to crash observations. It is expected that injury severity models are more sensitive to changes in the sample, and a broader analysis period may be needed to minimize the variability of the estimates.

5. SIGNALIZED INTERSECTIONS

In typical crash analysis, crashes located within 250 ft. of intersection centroids are considered to be intersection-related crashes. Although many crashes may happen on segments linked to intersections, they are believed to be triggered or influenced by the presence of intersections. One obvious example is the rear-end crash, which is common around intersections due to vehicle queues that originate at intersections. Some crashes happen at intersections themselves and they involve vehicles from different approaches. For rear-end crashes that occur on one specific leg, it may be appropriate to use only the traffic volume on the intersection leg where the crash occurs. For crashes involving vehicles from two approaches, the traffic volumes and the design of the two involved approaches are important. For example, a reduced sight distance on one of the approaches may be a contributing factor.

In this project, vehicle crashes at intersections were divided into two categories: (1) approach crashes, which happen on the same approach segments to the intersection and (2) inside-intersection crashes, which happen inside the intersection and involve vehicles traveling from different approaches. The analysis framework is shown in Figure 5.1. Single vehicle crashes were not analyzed in this project considering their rarity around intersections (approximately 4% in the sample data).

For approach crashes, the roadway geometric features, the dynamic operational parameters (volume, speed, and weather conditions) of *the same approach*, and the intersection level features are used to estimate the safety models.

For inside-intersection crashes, the roadway geometric features, the dynamic operational parameters (volume, speed, and weather conditions) of *the two involved approaches*, and the intersection level features are used in model estimation.

5.1 Sampling

The number of state intersections in the current Indiana safety management system divided by rural and urban, signalized and unsignalized, four-leg and non-four-leg are summarized in Table 5.1. Ideally, a stratified sampling of the intersections by different categories is preferred so that their various safety performance and relationships with the investigated factors could be investigated separately. However, the coverage of the speed and volume data limited the analysis on all types of intersections.

The INRIX speed data and the AADT from the INDOT data inventory did not cover most of the local roads. For the intersection of two local roads and the intersection of local and state roads, the missing data issue on their legs was quite common. Thus, in this project, the four-leg signalized intersection of two state roads with eight directional segments connected at intersections and covered by the INRIX speed data and the INDOT AADT data inventory were analyzed.

Out of the 110 four-leg signalized intersections of two state roads, 69 of them were correctly assigned with their approaching and exiting legs. The ID and latitude and longitude of these intersections and the linked segment IDs are available in the appendices. The spatial distribution of the selected intersection is shown in Figure 5.2. There were 443 approach crashes and 251 inside-intersection crashes in the sample data.

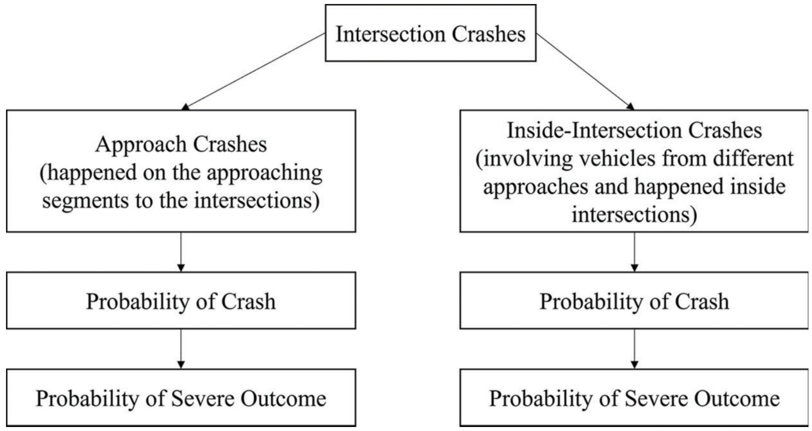


Figure 5.1 Intersection crashes analysis framework.

TABLE 5.1
Indiana state intersection numbers

	Four-Leg		Non-Four-Leg	
	Signalized	**Unsignalized**	**Signalized**	**Unsignalized**
Urban	1,036	3,556	382	5,646
Rural	154	4,669	65	8,661

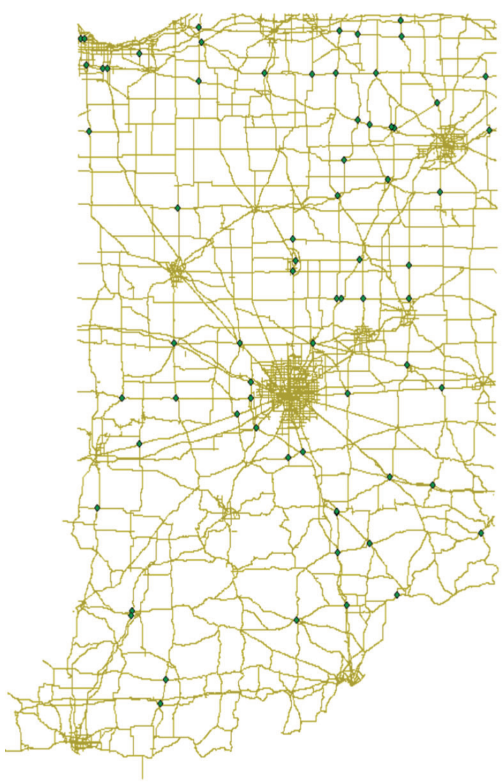


Figure 5.2 Selected sample intersections.

5.2 Traffic Volume Estimation

Unlike rural freeway segments, the traffic volumes for intersections refer to the volumes on all the linked approaching and exiting legs; and for each intersection, the volume prediction model was applied eight times to obtain the full view of the volume distribution on the different legs.

The traffic volume estimation for the signalized intersection legs follows the same method used for rural freeways. The same data source was used, which was the permanent volume collection stations in the INDOT traffic count system. The time of day, weather conditions, and AADT were applied as explanatory variables to estimate the linear regression models. The major difference in this estimation was the data sample. Namely, the interstate/freeway samples were removed according to the inherited road functional class in the volume table because all the traffic on the legs connected to intersections were interrupted traffic while the flow of traffic on the interstates/freeways was unimpeded and continuous.

5.3 Weather Interpolation

The weather interpolation for signalized intersections used the same data source and method as rural freeway segments as discussed in Section 4.3. The same weather conditions within the intersection areas (center of intersections and about one-quarter mile length legs) were assumed. The latitude and longitude of the intersection centroids were used as the coordinates to estimate hourly precipitation and temperature.

5.4 Roadway Features Data Collection

The roadway feature data collection of intersections involved two parts: (1) the intersection level geometric parameters and (2) the segment level parameters. The intersection level geometric parameters include the variables that describe the overall intersection characteristics, such as the size and the skewed angle of the legs. The segment level parameters describe the designs and settings of the approaching segments linked to intersections such as the median type, speed limit, exclusive left turn, etc.

Although the INDOT geometric data inventory tables provided some of the basic roadway features, such as the number of lanes and speed limits, more detailed geometric design parameters were collected. For example, for the intersection legs, traffic engineers usually increase the lanes for left-turn traffic so the number of lanes may change near intersections. These changes (e.g., increase one lane versus two lanes) can have different impacts on intersection safety and these features therefore were collected and coded in the proper ways.

All the roadway features that may affect safety and are available through Google Earth were manually collected for the sample intersections. The variable names and descriptions are shown below, and the coded features for the sample intersections are included in the appendices.

Intersection level parameters

- Residential area or not
- Degree of skewed angle (degree)
- Longer diagonal length of the intersection (ft.)

Segment level parameters

- General location
  - Bearing of the segment (8 directions)
  - Exist close intersection or not (500 ft.)
  - Exist close driveways (100 ft.)
- Cross-section setting
  - Number of lanes before the intersection stop line
  - Number of lanes on segment part
  - Center separation type (no, flushed, raised, depressed)
  - Center separation width (ft.)
  - Shoulder type (no, unpaved, paved)
  - Shoulder width (ft.)
  - Exist sidewalk on the approaching segment or not
- Curves
  - Exist curves
  - Radius of the curve (ft.)
- Turning settings
  - Left turn setting (exclusive left turn lane * left turn traffic signal)
  - Right turn setting (exclusive right turn lane)
  - Channelization for right turn or not
- Others
  - Exist roadside blocking objects that may affect sight distance (right turn)
  - Exist push button for pedestrians
  - Clear lane marking or not

5.5 Data Processing and Linking

The data preparation of intersections is much more complex than rural freeway segments, considering the four legs (four approaching segments and four exiting segments) linked to the intersections and the changes on the segmentations in the INRIX speed data. The data preparation consisted of two major parts: (1) generating non-crash samples and (2) linking data from the various sources to the crash/non-crash table.

5.5.1 Generating non-crash samples

According to the proposed analysis framework, the hours of the intersections when crashes happened are the "1" observations in the logistic regression model. The conditions that yield the occurrence of crashes are compared to the non-crash samples, which are the "0" observations. Due to the relatively small probability of crash in one intersection within a 1-hour period, there could be a very severe unbalance issue if all the non-crash hours are used as samples in the model estimation.

The typical way to handle this unbalanced response issue is to assume a moderate ratio between the crash and non-crash samples, which will allow the effects of the safety-related factors to remain the same signs and significance levels but with adjusted coefficients. According to the literature review, 1:30 is the commonly applied ratio, which was adopted in this analysis.

For crash records, the unit (vehicle) table with pre-crash driving directions of all involved vehicles was connected to the crash table by the crash ID. If the crash-involved vehicles were driving on the same approach (approach crashes), which are usually rear-end or sideswipe crashes, only the common approach segment was recorded. If the crash involved vehicles were driving on multiple approaches (inside-intersection crashes), only the approach segments of the two vehicles with major responsibility were recorded.

The non-crash samples were generated separately for approach crashes and inside-intersection crashes. Random intersection IDs and hours were generated and compared to all the crash hours on the corresponding intersections. If there were any crashes within 2 hours before or after a crash hour, the generated non-crash sample was replaced with the one out of the crash hour's range. This checking assures that all the generated non-crash samples were not affected by the potential influence of the existing crashes.

The discussed methods yielded two separate tables (approach crashes and inside-intersection crashes) both with crashes:non-crash = 1:30. These two tables include the Intersection ID and Segment ID indicating the location information and the time when the crash/non-crash happened along with the time information.

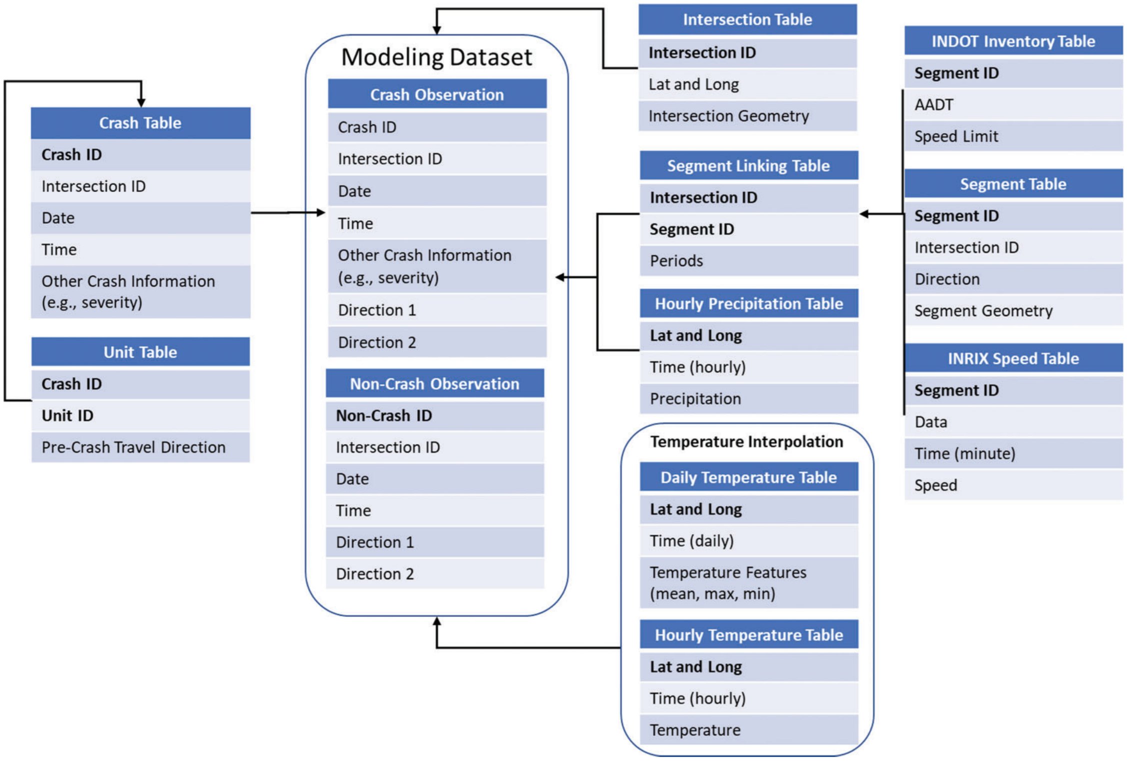


Figure 5.3 Signalized intersection data structure diagram.

The location and time information were used as the basis for various data sources linking.

5.5.2 Linking various data sources

The following are eight tables with various sources of data that need to be linked to the crash/non-crash table:

1. The intersection table with collected intersection level roadway features (unique ID: Intersection ID).
2. The segment table with collected segment level roadway features (unique ID: Segment ID).
3. The segment linking table with linking information between the segments and intersections (unique ID: Paired Intersection ID and Segment ID).
4. The INDOT inventory data table with each segment's AADT and speed limit data (unique ID: Segment ID).
5. The INRIX speed table with each segment's hourly speed (unique ID: Segment ID and Time).
6. The hourly precipitation table with hourly gridded precipitation data (unique ID: Latitude and Longitude and Time).
7. The daily temperature table with daily gridded temperature data (unique ID: Latitude and Longitude and Time).
8. The hourly temperature table with hourly temperature data on sparse stations across Indiana (unique ID: Latitude and Longitude and Time).

Figure 5.3 shows the data structure diagram. The diagram presents the main characteristics of each data table and displays the linking process.

The segment linking table includes the variable periods because the INRIX data changes their segmentations every 6 months and given the 3 years of data (2017–2019) used for this analysis, there were 6 distinct periods. A great deal of effort, with the help of ARCGIS spatial analyzing tools, was required to obtain the linking information between the segments and intersections. First, all the segments that were within 500 ft. of an intersection centroid were selected; second, segments that were too short (50 ft., mostly auxiliary lanes) were removed; third, the bearings of the left segments were checked and compared with each other, and only paired segments with matched bearings were kept; last, all the linked segments and their corresponding intersections were plotted in ARCGIS for final manual checking.

The INRIX speed table was sampled at the 1-minute level in the raw data set (i.e., 60 observations in the crash/non-crash hour). Various speed features were extracted from the speed profile of the hour, including the mean, maximum/minimum, standard deviance, estimated slope, uptrend, and downtrend indicators.

The hourly precipitation table and daily temperature table were gridded with hourly precipitation data and daily temperature data using the longitude and latitude of the virtual stations as the unique ID. The hourly temperature table contains the hourly temperature data from the sparse weather stations. The methods discussed in Section 4.3 were used to interpolate the weather data, after which the coordinates of the intersection and the time of the crash/non-crash were used to acquire the corresponding hourly temperature or precipitation.

TABLE 5.2
Summary of the effects of approach crashes risk on signalized state road intersections

Feature/Variable	Improves Safety	Reduces Safety	Counter Intuitive Result
Hourly leg traffic volume (1,000 veh/h)		×	
Standard deviation of speed (mph)		×	
Below-freezing temperatures (23°F–32°F)		×	
Near-freezing temperatures (32°F–41°F)		×	
Hourly average approach delay(s)		×	
Exclusive left-turn lane without left-turn arrow		×	×
Exclusive left-turn lane with left turn arrow		×	×
Urban PSL = 20, 25 mph (Ref = Urban 30, 35 mph)		×	×
Urban PSL = 40, 45 mph	×		×
Urban PSL ≥ 50 mph		×	
Rural PSL = 20, 25 mph	×		×
Rural PSL = 30, 35 mph		×	
Rural PSL = 40, 45 mph		×	
Rural PSL ≥ 50 mph	×		×
Workday 6:00am–9:00am (Ref = Workday 12:00am–6:00am)		×	
Workday 9:00am–12:00pm		×	
Workday 12:00pm–3:00pm		×	
Workday 3:00pm–6:00pm		×	
Workday 6:00pm–12:00am	×		
Weekend 6:00am–9:00am (Ref = Workday 12:00am–6:00am)		×	
Weekend 9:00am–12:00pm		×	
Weekend 12:00pm–3:00pm	×		
Weekend 3:00pm–6:00pm		×	
Weekend 6:00pm–12:00am		×	

5.6 Safety Analysis Results

The crashes on signalized state road to state road intersections were divided into approach crashes and Inside-intersection crashes. they were analyzed with two logistic regression models separately. The following interpretation of the models focuses on the effects of the factors and possible countermeasures. The statistical estimates can be found in Appendix C, Signalized Intersections (Table D.5, Table D.6, Table D.7, and Table D.8).

5.6.1 Crash risk

5.6.1.1 Approach crashes. Table 5.2 presents the effects of the factors on approach crashes on signalized state road-state road intersections.

The hourly leg traffic volume representing the exposure of crash risk increase the probability of crashes as expected.

The standard deviation of speed on the approaching legs to the intersection was obtained from the 60 1-minute speed observations in the crash/non-crash hour, which reflects the speed changes over time. When the vehicles in the queue experience large speed changes, the probability of crashes increased.

The temperature data were divided into categorical variables by 9°F intervals. The results show that both the below-freezing temperature (23°F–32°F) and the near-freezing temperature (32°F–41°F) had significant positive impacts on the occurrence of crashes. This result is reasonable because people tend to make more mistakes when the operating conditions change. When the temperature goes below the freezing point (32°F), the precipitation will change from rain to snow and lead to slippery pavement, and drivers who are not aware of this change may not maintain proper headways and could experience more rear-end crashes. However, when the temperature changing process is completed and drivers are aware of the risky weather conditions, they tend to drive more carefully to compensate for this change of risk. Therefore, no significant safety deterioration for the lower temperature groups was observed.

The hourly average approach delay was obtained from the difference between the observed and expected travel time ($\frac{Segment\ Length}{Observed\ Speed} - \frac{Segment\ Length}{Speed\ Limit}$), where the speed limit was used as the expected speed of passing the segment. This measure reflected the overall congestion level within the observing hour. When the delay is large, there are more vehicles and longer queues behind the stop line, and both of these factors can increase the probability of crashes.

The setting of the left-turn was categorized in three scenarios: (1) no exclusive left-turn lane, (2) exclusive left-turn lane without left-turn arrow, and (3) exclusive left-turn lane with left turn arrow. It was expected that the introduction of an exclusive left-turn lane and phasing would improve safety, but the results indicated that both of these settings reduced safety. This counter-intuitive result is probably caused by the omission of turning volumes in the model. Engineers design exclusive left-turn lanes or phasing for intersection approaches that have high left-turn volumes. Since the turning volume data were missing (only total volumes available), the effects of left-turn volume were included in the effects of other turn-related variables. The estimated effects of the left-turn variables might be intuitive if the left-turn volumes were available and included in the model.

The speed limit effects were checked separately for urban and rural intersections. For urban intersections, compared to the most common speed limit settings on signalized intersections (urban, posted speed limit = 30, 35), relatively high speed limit settings (≥ 50 mph) and low settings (20, 25 mph) both were associated with reduced safety performance, but the effect of setting (rural, posted speed limit > 50) was inverse, probably due to the small sample.

The effect of "hour of the day" and "weekend or not" were jointly examined, the estimates of different "weekday/weekend" + "day period" indicated that the "hour of the day" effects were consistent with the volume changes over the day. When there were more vehicles on the road during certain time periods, the probability of crashes for those time periods would increase. The probability of crash was the highest between 9:00am and 12:00pm on the weekend, while the probability was the highest between 6:00am and 9:00am on the weekdays.

Although the results for the approach crashes model are mostly self-explanatory, only a few geometric variables are included among the significant intersection variables and, consequently, only a few geometry improvements can be suggested. Fortunately, the hourly crash probability model has revealed conditions when the crash risk tends to be high (e.g., hours when temperatures are near the freezing point, hours when speed exhibits large variation). Indirect countermeasures that might eliminate or mitigate the risk include warning messages and better signal coordination with upstream signalized intersections.

5.6.1.2 Inside-intersection crashes. Table 5.3 presents the effects of the factors on inside-intersection crashes. As discussed at the beginning of this section, the volume estimates on the two involved intersection legs were both included in the inside-intersection crashes model. The two volume estimates were both positive and significant as expected. They reflect the effects of volume on the probability of crash. The coefficient for the crossing road volume was about two times the coefficient for the major road volume, indicating that the volumes on the crossing roads introduced a much higher crash risk than those on the major roads.

Separating traffic directions on the crossing road approach (divisional island, median) was found to

TABLE 5.3
Summary of the effects of inside-intersection crashes risk on signalized state road intersections

Feature/Variable	Improves Safety	Reduces Safety	Counter Intuitive Result
Hourly leg volume on major road (1,000 veh/h)		×	
Hourly leg volume on crossing road (1,000 veh/h)		×	
Flushed center separation on crossing road	×		
(Ref = No separation)			
Raised center separation on crossing road		×	
Depressed center separation on crossing road		×	
Urban PSL = 20, 25 mph on crossing road		×	×
(Ref = Urban PSL = 30, 35 mph)			×
Urban PSL = 40, 45 mph on crossing road		×	×
Urban PSL ≥ 50 mph on crossing road	×		×
Rural PSL = 30, 35 mph on crossing road	×		×
Rural PSL = 40, 45 mph on crossing road		×	×
Rural PSL ≥ 50 mph on crossing road		×	
Workday 6:00am–9:00am		×	
(Ref = Workday 12:00am–6:00am)			
Workday 9:00am–12:00pm		×	
Workday 12:00pm–3:00pm		×	
Workday 3:00pm–6:00pm		×	
Workday 6:00pm–12:00am	×		
Weekend 6:00am–9:00am		×	
(Ref = Workday 12:00am–6:00am)			
Weekend 9:00am–12:00pm		×	
Weekend 12:00pm–3:00pm		×	
Weekend 3:00pm–6:00pm	×		
Weekend 6:00pm–12:00am		×	

significantly affect safety performance. Compared to the undivided approach of the crossing road, the flushed separation improved the safety while the raised and depressed separations reduced the safety. These results seem to be counterintuitive. From the safety point of view, physical separations (raised and depressed) better defined the turning paths of vehicles at intersections and they should have a positive effect on safety unless curbed separating elements are incorrectly used or designed. The likely cause of the counterintuitive result, however, is similar to the cause of counterintuitive results for left-turn treatments with missing left-turn traffic volumes in the model. The physical separations on crossing road approaches may be used on roads with higher left-turn movements than where flushed separations are used.

The effects of speed limit settings and different time periods were mixed, which reflects the spatial and temporal distribution of the sample crashes. Their effects were difficult to interpret considering the relatively small samples of the inside-intersection crashes.

The inside-intersection crashes model confirmed the importance of exposure (traffic volume) on the probability of crash, but it lacked variables that could be used to infer useful safety countermeasures.

5.6.2 Injury severity

The hourly conditional probability of severe injury given crash was modeled as a function of traffic volume, roadway features, weather conditions, speed characteristics, and other predictors. While various variables were tested, only a few factors were found to be significantly related to severe injury crashes. The effects are summarized in Table 5.4 and Table 5.5 respectively for approach crashes and inside-intersection crashes.

5.6.2.1 Approach crashes. The three identified factors that significantly increased the probability of severe approach crashes were near-freezing temperature, hourly precipitation of more than 0.1 inch, and standard deviation of the speed.

The two weather-related variables are intuitive because near-freezing temperatures imply adverse pavement conditions, while precipitation higher than 0.1 inch implies moderate or heavy rain. Both of the conditions may increase the crash risk without a proper risk perception adjustment by drivers and thus cause more severe crashes.

The standard deviation of travel speed over time reflects the general variability of traffic conditions on the approach to the intersection. If there are large deviations of speed across time, the outcomes of the approach crashes (mostly rear-end crashes) could be more severe. This finding was consistent with the results for rural freeway segment in Section 4.6.2, Injury severity.

5.6.2.2 Inside-intersection crashes. The severity model for inside-intersection crashes included two variables that significantly increased the probability of a crash severe outcome: (1) the mean speed on the major road and (2) the number of lanes on the major road. The number of lanes on the crossing road was found to improve safety, which is counterintuitive.

The effect of mean speed is intuitive since higher speed implies more kinetic energy transferred between vehicles and to their occupants—the well-known factor of injury and damage severity. The number of lanes on the major road reflects the intersection size and the increased presence of heavy vehicles—another well-known factor of crash severity outcome.

The findings of the injury severity models imply that adverse weather conditions and improper speed controls will lead to more severe crashes on intersections. The same engineering countermeasures that could inform drivers of risky weather conditions could help improve safety as well. Also, prior intersection notice and speed control measures (e.g., gradually decreased lane width on approaching intersections segments) could even out drivers' speed profiles, avoid speeding, and large speed variation.

TABLE 5.4
Summary of the effects of severe approach crash on signalized state road intersections

Feature/Variable	Improves Safety	Reduces Safety	Counter Intuitive Result
Near-freezing temperatures (32°F–41°F)		×	
Precipitation > 0.1 in		×	
Standard deviation of speed (mph)		×	

TABLE 5.5
Summary of the effects of severe inside-intersection crashes on signalized state road intersections

Feature/Variable	Improves Safety	Reduces Safety	Counter Intuitive Result
Average hourly speed on major road (mph)		×	
Number of lanes on major road		×	
Number of lanes on crossing road	×		?

6. RECOMMENDATIONS FOR SAFETY MANAGEMENT

Emerging data were used to estimate the short-term effects of traffic volume, travel speed characteristics, weather conditions, and roadway features on the crash probability and injury severity. Two case studies at rural freeway segments and signalized intersections provided insights into the use of new and emerging data sources to supplement the current safety management practices in Indiana. This section discusses the main conclusions, data limitations, and suggestions for future work and provides a roadmap for potential application in Indiana's safety management system.

6.1 Main Findings

Short-term safety analysis shows promise. Hourly crash risk and injury severity can be estimated based on the existing and emerging data if supplemented with volume directional split on segments and turning movement volumes at intersections.

The traffic volumes in hourly intervals could be estimated using INRIX's probe density, time indicators, AADT by vehicle type, travel speed characteristics, and weather conditions. The obtained volume models were evaluated with out-of-sample validation methods and then used to impute the hourly volumes missing in permanent detectors count data. The imputation operation is critical for system-wide implementation of short-term safety analysis where hourly traffic volumes are limited and important. The estimated directional splits need to be improved with land use data pertaining to major cities reachable through rural arterial segments and to Traffic Analysis Zones reachable via rural and urban road segments.

Particularly at intersections, hourly turning volumes play a critical role in short-term safety analysis. The lack of turning traffic information causes other variables to compensate for the missing data, thereby producing counterintuitive results. For example, in this project, an exclusive left-turn lane was associated with an increase in rear-end crash probability on the approach. The lack of left-turning volume data caused this counterintuitive finding where exposure was missing. In addition to loop detectors, other data sources should be further evaluated, such as StreetLight, and high-resolution counts. A follow-up study is necessary to assess their usefulness as a proxy information for missing turning volumes. Additional data sources need to be identified and evaluated for signalized and unsignalized intersections.

Other time-dependent variables such as travel speed characteristics and weather conditions were found to be useful in estimating the hourly risk and severity of crashes. The most pertinent determinants were travel speed variation, decreasing operating speed, congestion level, scattered rain, and freezing temperatures. Furthermore, roadway features such as road barriers, speed limit, street lighting, presence and degree of curves, pavement type, and shoulder significantly affected the crash risk and injury severity.

Interactions among various time-dependent variables and roadway characteristics were investigated, but no considerable effect was found. It is expected that a larger sample may produce significant interaction effects. Also, selecting a particular functional form may reduce interaction terms' significance while prioritizing the main effects' importance. It is anticipated that the combined impact of trucks, curves, congestion, and poor weather conditions can be better addressed with alternative functional forms. Alternative safety models may provide a broader perspective of the effects of time-dependent and roadway characteristics on the short-term crash risk and injury severity. The present analysis was very useful because it describes valuable significant connections. However, the results from its multiple models can be combined to provide a more comprehensive view of the studied phenomenon.

6.2 Data Limitations and Recommendations

This project revealed several limitations of new and emerging data sources and their application in safety management. Addressing these limitations is essential for applying the proposed approach at the system level.

The detailed roadway data departs from the existing road inventories in the sense that minor changes in cross-sectional and longitudinal elements were detected with exact precision. Two comprehensive road inventories were obtained for a small sample of rural freeways and signalized intersections using Google Earth's imagery. The retrieval of detailed roadway features data is time-consuming, yet these features were found to influence safety. A system-wide implementation may require investing resources to obtain a similar dataset for all state-administered roads. Also, a regular update of the database may require spending additional resources. While this project provided a methodology to gather high-resolution roadway data, another good source of information is the Digital Terrain Model (DTM) data. Indiana recently completed a 4-year project to provide 1-foot resolution orthophotography and elevation data for Indiana. These data are available to the public via the Indiana Map.

Hourly traffic volumes are essential for short-term safety evaluation. The proposed estimation of hourly traffic volumes identified two central problems: directional split at road segments and turning flows at road crossings. While directional splits can be addressed using area-level land use data, turning volumes at intersections and freeway ramps are missing. Additional alternative data sources need to be considered, and their effectiveness assessed to resolve this issue. Data obtained from vehicle tracking (StreetLight) and counting stations (high-resolution count data) systematically captured at signalized intersections would help.

Vehicle probe densities were to be found useful in estimating hourly traffic volumes in places where counting stations are not available. This project's results constitute a first step towards their potential use in calculating hourly volumes at the system level. However, to improve the current models' prediction ability, additional raw data quality improvements are needed. INRIX's speed data offer a right-bounded representation of vehicle probe density as the actual number of probes is not provided. Our analysis confirmed the importance of tackling this issue since high hourly volumes were more challenging to predict. Better quality probe density information could help improve estimating traffic volumes. It is recommended to work with INRIX and the FHWA to study the possibility of obtaining the exact number of vehicle probes used to calculate operating travel speeds.

As new and emerging data sources are introduced into Indiana's safety management system, potential data management issues may arise. First, granular information increases data warehousing requirements. Most of the presented data sources are large in size, and sufficient disk space needs to be purchased. Second, standard reproducible data quality checks need to be installed and tested. This project provided the first insights towards standard quality controls. However, it is necessary to involve external stakeholders to learn the routine quality control procedures for various disciplines (e.g., climate analysis). Third, external commercial data sources need to be carefully considered and data ownership issues discussed. The data sources presented here are available at no cost to INDOT via partnerships between government agencies and private companies (e.g., INRIX, Purdue's INClimate, University of Maryland's CATT Lab). Lastly, preliminary and inferential data analysis models require technical personnel to be familiar with data processing and statistical analysis. Supplementary tools need to be provided to safety engineers to use the proposed methods at their full potential.

6.3 Application to Safety Management

The proposed short-term safety models presented here can be converted to traditional safety analysis units to facilitate their implementation. This conversion can be accomplished via statistical simulation. Assuming that the observed volumes, speed characteristics, and weather conditions are typical, one can obtain the expected annual number of crashes by severity level by adding all the hours for 1 year. In this project, a 5-year analysis was used to avoid the return to the mean effect. Additionally, if one is interested in the short-term safety consequences of sudden changes in weather or travel speed, the following rationale should be followed. For a specific location, hourly crash probabilities and injury severities are calculated assuming normal conditions (i.e., dry weather, non-congested traffic). Then, the marginal effects of time-dependent predictors can estimate the change in crash risk and injury severity.

The results from this project can supplement the current safety management system in Indiana in different forms. The most notable developments are related to data analysis. First, when identifying high-crash locations, the overall traffic and aggregate roadway features are typically used. While this approach is adequate to represent the overall crash rates, it fails to acknowledge the dynamic factors that affect traffic operations. Specifically, the combined effects of disaggregate roadway features, hourly volumes, travel speeds, and weather conditions on safety are missing. Including these factors during the early safety inspection balances the overall crash trends with particular site conditions. It is expected that a more open network screening may select more suitable candidates for safety improvements.

Second, the proposed consideration of time-dependent factors is critical to approach the actual causes of crashes. Notably, statistical path analysis (causal) will help understand the mechanism behind crash probability and injury severity. An intermediate study of harmful events is proposed. The harmful events component adds depth to the analysis by connecting the crash onset with its outcome. Zou et al. (2014) proposed and implemented this approach to study factors affecting barrier-relevant (BR) crashes. The type of harmful events and their occurrence during a BR crash, and the severity of the BR crash outcomes were considered. The proposed approach permits evaluating the significance of contributing factors to different crash characteristics.

Lastly, among the main advantages of short-term safety analysis, the formulation and evaluation of operational countermeasures are evident. Proactive countermeasures such as advanced signal warnings and variable speed limits are targeted to extreme weather events and the presence of growing queues. These effects were estimated within this project and can be used for the economic evaluation of countermeasures. Additional tools are needed to facilitate the implementation of statistical models by traffic engineers.

7. CLOSURE

Traditional safety management practices utilize a restricted set of data sources that are available at the system level. Crash rates are usually assessed based on aggregated exposure variables (e.g., AADT and length) and major roadway features (e.g., number of legs and functional classification). However, other contributing factors to crashes, such as traffic operations, traffic control, weather conditions, and disaggregate roadway features, are often discounted. This project served as a feasibility study of short-term safety analysis to supplement the current safety management practices in Indiana.

Multiple emerging data sources were identified, and historical data were acquired, which then were used in conjunction with conventional data to conduct a short-term safety analysis. A comprehensive evaluation of the quality of the new data and a relevance analysis to crash probability and severity was performed. This project is a fundamental first step to introducing a roadmap for detecting safety problems with the new and emerging data and conventional safety data combined.

A short-term analysis was employed in two case studies for rural freeways and signalized intersections. The results from a statistical analysis suggested a strong connection between the time-dependent factors related to traffic volume, travel speed, and weather and the hourly probability of crash and injury severity. Static roadway features also were shown to have a significant effect on hourly crash risk and outcome severity. These findings open the door for short-term proactive safety management that can supplement the current practices by calculating the safety effects of time-dependent variables. By controlling such variables, it is possible to calculate the safety benefits of implementing various operational countermeasures.

While the proposed new approach's benefits are clear, the process is data-intensive and requires further work before implementation at the system level. With this in mind, another ongoing research project (SPR-4540) aims to incorporate this time-dependent data into pro-active safety management practice. SPR-4540 is a natural follow-up from the fundamental knowledge gained from the current project and will conduct a pilot study for implementation on the studied road types (signalized intersections and rural freeways). This new project also is anticipated to have a strong emphasis on the implementation of specific tools for the end-user to assist with a future proactive safety management system in Indiana.

8. REFERENCES

Anuar, K., & Cetin, M. (2017). Estimating freeway traffic volume using shockwaves and probe vehicle trajectory data. In *Transportation Research Procedia*, *22*, 183–192. https://doi.org/10.1016/j.trpro.2017.03.025

Allison, P. D. (2012). *Logistic regression using SAS: Theory and application* (2nd ed.). SAS Institute.

Dutta, N., & Fontaine, M. D. (2019). Improving freeway segment crash prediction models by including disaggregate speed data from different sources. *Accident Analysis and Prevention*, *132*, 105253. https://doi.org/10.1016/j.aap.2019.07.029

Hou, Y., Young, S. E., Sadabadi, K., Sekuła, P., & Markow, D. (2018, March). *Estimating highway volumes using vehicle probe data–proof of concept* (Conference paper NREL/CP-5400-70938). National Renewable Energy Lab. https://www.nrel.gov/docs/fy18osti/70938.pdf

Roshandel, S., Zheng, Z., & Washington, S. (2015, June). Impact of real-time traffic characteristics on freeway crash occurrence: Systematic review and meta-analysis. *Accident Analysis & Prevention*, *79*, 198–211. https://doi.org/10.1016/j.aap.2015.03.013

Sharma, A., Ahsani, V., & Rawat, S. (2017). *Evaluation of opportunities and challenges of using INRIX data for real-time performance monitoring and historical trend assessment* (Report No. SPR-P1(14) M007). Iowa State University Center for Transportation Research and Education. https://lib.dr.iastate.edu/ccee_reports/24

Yu, R., & Abdel-Aty, M. (2014, March). Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. *Safety Science*, *63*, 50–56. https://doi.org/10.1016/j.ssci.2013.10.012

Zou, Y., Tarko, A. P., Chen, E., & Romero, M. A. (2014). Effectiveness of cable barriers, guardrails, and concrete barrier walls in reducing the risk of injury. *Accident Analysis & Prevention*, *72*, 55–65. https://doi.org/10.1016/j.aap.2014.06.013

APPENDICES

Appendix A. Data Items and Descriptions of Selected Data Sources

Appendix B. Roadway Data Collection

Appendix C. Traffic Volume Models

Appendix D. Safety Analysis Models

APPENDIX A. DATA ITEMS AND DESCRIPTIONS OF SELECTED DATA SOURCES

A.1 Traffic Count Database System

Table A.1 Description of data items available in INDOT's Traffic Counts Database System (TCDS)

Data Item Name	Data Item Type	Example	Description
Station ID	Number	950102_NEG	The unique volume station ID, separated by directions
Date	Date	10/1/2017	The date when the volume data was collected
Hour	Number	1	The hour of the day, 1, 2, …, 24
P1	Number	32	The observed volume in the first quarter of the hour
P2	Number	21	The observed volume in the second quarter of the hour
P3	Number	23	The observed volume in the third quarter of the hour
P4	Number	30	The observed volume in the last quarter of the hour
Vol	Number	106	The sum of the four quarter volumes, i.e., hourly volume

Note: There are several tables in the traffic count database in raw data and the volume values were organized as "one row for one day" format, the data preprocessing was applied to obtain the table presented here.

A.2 Automated Reporting Information Exchange System

Table A.2 Description of data items in Crash Table

Data Item Name	Data Item Type	Example	Description
MstrRecNbrTxt	Number	902849881	Indicates the State Repository Unique Identifier, called Master Record Number (MRN)
CollDayWeekCde	Text	2	1 - Sunday 2 - Monday 3 - Tuesday 4 - Wednesday 5 - Thursday 6 - Friday 7 - Saturday
CollTimeMilitaryTxt	Text	1711	0000 to 2359
MotorVehInvolvedNmb	Number	2	Number of involved motor vehicles
TrailersInvolvedNmb	Number	0	Number of involved trailers
InjuredNmb	Number	0	Number of injuries
DeadNmb	Number	0	Number of death
LatDecimalNmb	Number	38.959	Latitude of the crash location
LongDecimalNmb	Number	-85.835	Longitude of the crash location
RdwyClassCde	Text	02	01 – Interstate 02 – US Route 03 – State Road 04 – County Road 05 – Local/City Road 06 – Other
SchoolZoneInd	Text	N	Indicates if the crash occurs within a school zone
ConstructInd	Text	N	Indicates whether or not the crash occurred within a construction zone or in a traffic "back-up" outside of, but due to a construction zone
LightCondCde	Text	01	Describes the light conditions at the time and place of the crash: 01 - DAYLIGHT 02 - DAWN/DUSK 03 - DARK (LIGHTED) 04 - DARK (NOT LIGHTED) 05 - UNKNOWN

Table A.3 Description of data items in Crash Table (cont.)

Data Item Name	Data Item Type	Example	Description
WeatherCde	Text	02	Describes the primary atmospheric condition at the time and place of the crash: 01 - CLEAR 02 - CLOUDY 03 - RAIN 04 - SNOW 05 - SLEET/HAIL/FREEZING RAIN 06 - FOG/SMOKE/SMOG 07 - SEVERE CROSS WIND 08 - BLOWING SAND/SOIL/SNOW
SurfaceCondCde	Text	01	Describes the road surface conditions at the time and place of the crash: 01 - DRY 02 - WET 03 - MUDDY 04 - SNOW/SLUSH 05 - ICE 06 - LOOSE MATERIAL ON ROAD 07 - WATER (STANDING OR MOVING)
PrimaryFactorCde	Text	13	Indicates the primary cause of the crash: 01 - ALCOHOLIC BEVERAGES 02 - ILLEGAL DRUGS 03 - DRIVER ASLEEP OR FATIGUED 04 - PRESCRIPTION DRUGS 05 - DRIVER ILLNESS 06 - UNSAFE SPEED 07 - FAILURE TO YIELD RIGHT OF WAY 08 - DISREGARD SIGNAL/REG SIGN 09 - LEFT OF CENTER 10 - IMPROPER PASSING 11 - IMPROPER TURNING 12 - IMPROPER LANE USAGE 13 - FOLLOWING TOO CLOSELY 14 - UNSAFE BACKING 15 - OVERCORRECTING/OVERSTEERING 16 - RAN OFF ROAD 18 - WRONG WAY ON ONE WAY 19 - PEDESTRIAN ACTION 20 - PASSENGER DISTRACTION 21 - VIOLATION OF LICENSE RESTRICTION 22 - JACKKNIFING 23 - CELL PHONE USAGE 24 - OTHER TELEMATICS IN USE 25 - OTHER (DRIVER) - EXPLAIN IN NARRATIVE 26 - NONE (DRIVER) 27 - DRIVER DISTRACTED - EXPLAIN IN NARRATIVE 28 - SPEED TOO FAST FOR WEATHER CONDITIONS 35 - ENGINE FAILURE OR DEFECTIVE 36 - ACCELERATOR FAILURE OR DEFECTIVE 37 - BRAKE FAILURE OR DEFECTIVE 38 - TIRE FAILURE OR DEFECTIVE 39 - HEADLIGHT DEFECTIVE OR NOT ON 40 - OTHER LIGHTS DEFECTIVE 41 - STEERING FAILURE 42 - WINDOW/WINDSHIELD DEFECTIVE 43 - OVERSIZE/OVERWEIGHT LOAD 44 - INSECURE/LEAKY LOAD 45 - TOW HITCH FAILURE 46 - OTHER (VEHICLE) - EXPLAIN IN NARRATIVE

			47 - NONE (VEHICLE) 55 - GLARE 56 - ROADWAY SURFACE CONDITION 57 - HOLES/RUTS IN SURFACE 58 - SHOULDER DEFECTIVE 59 - ROAD UNDER CONSTRUCTION 60 - SEVERE CROSSWINDS 61 - OBSTRUCTION NOT MARKED 62 - LANE MARKING OBSCURED 63 - VIEW OBSTRUCTED 64 - ANIMAL/OBJECT IN ROADWAY 65 - TRAFFIC CONTROL INOPERATIVE/MISSING/OBSC 66 - OTHER (ENVIRONMENTAL) - EXPLAIN IN NARR 67 - NONE (ENVIRONMENTAL) 68 - UTILITY WORK
DamageEstimateCde	Text	03	Total dollar estimate of all damage (including vehicles and property) sustained in the crash
MannerCollCde	Text	01	Describes the type of crash: 01 - REAR END 02 - HEAD ON 03 - REAR TO REAR 04 - S a.m.E DIRECTION SIDESWIPE 05 - OPPOSITE DIRECTION SIDESWIPE 06 - RAN OFF ROAD 07 - RIGHT ANGLE 08 - LEFT TURN 09 - RIGHT TURN 10 - LEFT/RIGHT TURN 11 - BACKING CRASH 12 - OTHER - EXPLAIN IN NARRATIVE 13 - NON-COLLISION
TimeNotifiedMilitaryTxt	Text	1711	The time when the crash was notified
TimeArrivedMilitaryTxt	Text	1720	The time when the policy arrived the spot
colldte	Date/time	1/2/2017	The collision date

Note: Only variables that were used during analysis are included in the table here.

Table A.4 Description of data items in Unit Table

MstrRecNbrTxt	Text	902849883	Indicates the State Repository Unique Identifier, called Master Record Number (MRN)
UnitNmb	Number	1	Identifies the unit number
UnitTypeCde	Text	04	Indicates the type of unit involved
OccupsNmb	Number	1	The number of occupants of the vehicle
SpeedLimitTxt	Text	25	Recorded speed limit of the road
VehUseCde	Text	01	01 – Personal (Farm, Company) 02 – Commercial (Buses, Taxis, Common and Contract Carrier) 03 – Rental, not leased 04 – School 05 – Police 06 – Fire 07 – Ambulance 08 – Military 09 – Highway Department 10 – Other Government (Postal, etc.) 11 – Public Utilities (Gas, Electric, etc.) 12 – Other
RoadTypeCde	Text	05	01 – One Lane (One Way) 02 – Two Lanes (One Way) 03 – Multi-Lanes (One Way) 04 – Two Lanes (Two Way) 05 – Multi-Lane Undivided (Two Way) 06 – Multi-Lane Undivided 2-way left (Two Way) 07 – Multi-Lane Divided 3 or more (Two Way) 08 – Alley 09 – Private Drive

TravDirCde	Text	NW	N – North S – South E – East W – West NW – Northwest NE – Northeast SW – Southwest SE – Southeast
PreCollAction	Text	06	01 – Going Straight 02 – Backing 03 – Changing Lanes 04 – Overtaking/Passing 05 – Turning Right 06 – Turning Left 07 – Making U Turn 08 – Merging 09 – Entering Traffic Lane 10 – Leaving Traffic Lane 11 – Parked 12 – Slowing or Stopped in Traffic 13 – Unattended Moving Vehicle 14 – Avoiding Object in Roadway 15 – Starting in Traffic 16 – Driving Left of Center 17 – Crossing the Median

Note: Only variables that were used during analysis are included in the table here.

A.3 Highway Performance Monitoring System (H p.m.S)

Table A.3 Description of data items available in Highway Performance Monitoring System (H p.m.S)

Data Item Type	Data Item No.	Database-Specific Data Item Name	Data Item Name	Description
Inventory	1	F_System	Functional System	The FHWA approved Functional Classification System.
	2	Urban_Code	Urban Code	The U.S. Census Urban Area Code.
	3	Facility_Type	Facility Type	The operational characteristic of the roadway.
	4	Structure_Type	Structure Type	Roadway section that is a bridge, tunnel or causeway.
	5	Access_Control	Access Control	The degree of access control for a given section of road.
	6	Ownership	Ownership	The entity that has legal ownership of a roadway.
	7	Through_Lanes	Through Lanes	The number of lanes designated for through-traffic.
	8	HOV_Type	Managed Lane Operations Type	The type of managed lane operations (e.g., HOV, HOT, ETL, etc.).
	9	HOV_Lanes	Managed Lanes	Maximum number of lanes in both directions designated for managed lane operations.
	10	Peak_Lanes	Peak Lanes	The number of lanes in the peak direction of flow during the peak period.
	11	Counter_Peak_ Lanes	Counter Peak Lanes	The number of lanes in the counter-peak direction of flow during the peak period.
	12	Turn_Lanes_R	Right Turn Lanes	The presence of right turn lanes at a typical intersection.
	13	Turn_Lanes_L	Left Turn Lanes	The presence of left turn lanes at a typical intersection.
	14	Speed_Limit	Speed Limit	The posted speed limit.

Data Item Type	Data Item No.	Database-Specific Data Item Name	Data Item Name	Description
	15	Toll_Charged	Toll Charged	Identifies sections that are toll facilities regardless of whether or not a toll is charged.
	16	Toll_Type	Toll Type	Indicates the presence of special tolls (i.e., High Occupancy Toll (HOT) lane(s) or other managed lanes).
Route	17	Route_Number	Route Number	The signed route number.
	18	Route_Signing	Route Signing	The type of route signing.
	19	Route_Qualifier	Route Qualifier	The route signing descriptive qualifier.
	20	Alternative_Route_Name	Alternative Route Name	A familiar, non-numeric designation for a route.
Traffic	21	AADT	Annual Average Daily Traffic	Annual Average Daily Traffic.
	22	AADT_Single_Unit	Single Unit Truck and Bus AADT	Annual Average Daily Traffic for single-unit trucks and buses.
	23	Pct_Peak_Single	Percent Peak Single-Unit Trucks and Buses	Peak hour single-unit truck and bus volume as a percentage of total AADT.
	24	AADT_Combination	Combination Truck AADT	Annual Average Daily Traffic for Combination Trucks.
	25	Pct_Peak_Combination	Percent Peak Combination Trucks	Peak hour combination truck volume as a percentage of total AADT.
	26	K_Factor	K-factor	The design hour volume (30th largest hourly volume for a given calendar year) as a percentage of AADT.
	27	Dir_Factor	Directional Factor	The percent of design hour volume (30th largest hourly volume for a given calendar year) flowing in the higher volume direction.
	28	Future_AADT	Future AADT	Forecasted AADT.
	29	Signal_Type	Signal Type	The predominant type of signal system on a sample section.
	30	Pct_Green_Time	Percent Green Time	The percent of green time allocated for through-traffic at intersections.
	31	Number_Signals	Number of Signalized Intersections	A count of at-grade intersections where traffic signals are present.

Data Item Type	Data Item No.	Database-Specific Data Item Name	Data Item Name	Description
	32	Stop_Signs	Number of Stop Sign-Controlled Intersections	A count of at-grade intersections where stop signs are present.
	33	At_Grade_Other	Number of Intersections, Type - Other	A count of at-grade intersections, where full sequence traffic signal or stop sign traffic control devices are not present, in the inventory direction.
Geometric	34	Lane_Width	Lane Width	The measure of existing lane width.
	35	Median_Type	Median Type	The type of median.
	36	Median_Width	Median Width	The existing median width.
	37	Shoulder_Type	Shoulder Type	The type of shoulder.
	38	Shoulder_Width_R	Right Shoulder Width	The existing right shoulder width.
	39	Shoulder_Width_L	Left Shoulder Width	The existing left shoulder width.
	40	Peak_Parking	Peak Parking	Specific information about the presence of parking during the peak period.
	41	Widening_Obstacle	Widening Obstacle	Obstacles that prevent widening of the existing roadway for additional through lanes.
	42	Widening_Potential	Widening Potential	The number of through lanes that could be potentially added.
	43	Curves_A through Curves_F	Curve Classification	Curve classification data.
	44	Terrain_Type	Terrain Type	The type of terrain.
	45	Grades_A through Grades_F	Grade Classification	Grade classification data.
	46	Pct_Pass_Sight	Percent Passing Sight Distance	The percent of a Sample Panel section meeting the sight distance requirement for passing.
Pavement	47	IRI	International Roughness Index	IRI is the road roughness index most commonly used worldwide for evaluating and managing road systems. Road roughness is the primary indicator of the utility of a highway network to road users. IRI is defined as a statistic used to estimate the amount of roughness in a measured longitudinal profile.
	48	PSR	Present Serviceability Rating	Present Serviceability Rating (PSR) for pavement condition.
	49	Surface_Type	Surface Type	Surface type on a given section.

Data Item Type	Data Item No.	Database-Specific Data Item Name	Data Item Name	Description
	50	Rutting	Rutting	Average depth of rutting. A rut is defined as longitudinal surface depressions in the asphalt pavement derived from measurements of a profile transverse to the path of travel on a highway lane. It may have associated transverse displacement. Asphalt pavement is defined as pavements where the top-most surface is constructed with asphalt materials.
	51	Faulting	Faulting	Faulting is defined as a vertical misalignment of pavement joints in Portland Cement Concrete Pavements (Jointed Concrete Pavement). Jointed Concrete Pavements are defined as pavements where the top-most surface is constructed of Portland cement concrete with joints. It may be constructed of either reinforced or unreinforced (plain) concrete.
	52	Cracking_Percent	Cracking Percent	Cracking is defined as a fissure or discontinuity of the pavement surface not necessarily extending through the entire thickness of the pavement.
	54	Year_Last_Improv	Year of Last Improvement	The year in which the roadway surface was last improved.
	55	Year_Last_Construction	Year of Last Construction	The year in which the roadway was constructed or reconstructed.
	56	Last_Overlay_Thickness	Last Overlay Thickness	Thickness of the most recent pavement overlay.
	57	Thickness_Rigid	Thickness Rigid	Thickness of rigid pavement.
	58	Thickness_Flexible	Thickness Flexible	Thickness of the flexible pavement.
	59	Base_Type	Base Type	The base pavement type.
	60	Base_Thickness	Base Thickness	The thickness of the base pavement.
	61	Climate_Zone	Climate Zone	Climate zone as defined by the four LTPP climate zone descriptions.
	62	Soil_Type	Soil Type	Soil type as defined by AASHTO soil classes.
Inventory	63	County_Code	County Code	The County Federal Information Processing Standard (FIPS) code.

Data Item Type	Data Item No.	Database-Specific Data Item Name	Data Item Name	Description
Special Networks	64	NHS	National Highway System	A Roadway that is a component of the National Highway System (NHS).
	65	STRAHNET_Type	Strategic Highway Network	Roadway section that is a component of the Strategic Highway Network (STRAHNET).
	66	Truck	National Truck Network	Roadway section that is a component of the National Truck Network (NN) as defined by 23 CFR 658.
	67	Future_Facility	Future National Highway System	An unbuilt roadway (or section) of the National Highway System (NHS), including intermodal connectors.
Inventory	68	Maintenance_Operations	Maintenance & Operations	The legal entity that maintains and operates a roadway.
Traffic	69	Capacity	Capacity	The capacity of the roadway as estimated by the State DOT or local agency.
Inventory	70	Dir_Through_Lanes	Directional Through Lanes	The number of lanes designated for through-traffic, for a given direction of travel on a divided highway section.

A.4 National Performance Management Research Data Set (N p.m.RDS)

Table A.4 Description of data items available in National Performance Management Research Data Set (N p.m.RDS)

Data Item Name	Data Item Type	Example	Description
datasource	Text	N p.m.RDS (Passenger vehicles)	The data set this record comes from. This field is only included when choosing to merge the data sets into a single CVS file.
tmc_code	Text	107-12541	The unique 9-digit value identifying the TMC segment.
measurement_tstamp	Date	5/1/2017 12:00:00 a.m.	Date of data record, in “MM/DD/YY HH:NN:SS A” format. The date is in the local time of TMC segment to which the record pertains.
speed	Number	40	Speed is recorded in mph as an integer. The harmonic average speed for all reporting vehicles on the segment.
average_speed	Number	45	The historical average speed for the roadway segment for that hour of the day and day of the week in miles per hour.
reference_speed	Number	50	The calculated "free flow" mean speed for the roadway segment in miles per hour. This attribute is calculated based upon the 85th-percentile point of the observed speeds on that segment for all time periods, which establishes a reliable proxy for the speed of traffic at free-flow for that segment.
travel_time_minutes	Number	3	Travel time recorded in minutes as an integer. It is the ratio between the segment length and the harmonic average speed for all reporting vehicles on the segment.
data_density	Text	C	Data density indicator, where: A = 1 to 4 reporting vehicles B = 5 to 9 reporting vehicles C = 10 or more reporting vehicles

A.5 INRIX

Table A.6 Description of data items available in INRIX

Data Item Name	Data Item Type	Example	Description
Tstamp	Date/time	2019-09-09 00:00:40	The time variable in “YYYY-MM-DD HH:NN:SS” format, precise to second, recording the speed data collection time. The data recording resolution is by-minute.
Xdid	Number	393894578	The INRIX segment ID corresponding with those used in the INRIX segment shapefiles.
Speed	Number	21	Mean speed (mph) on the segment
Score	Categorical	30	Three categories, “10,” “20,” and “30,” recording the number of probe vehicles, 10 for lower than 10 vehicles, 20 for vehicle number between 10 and 30, 30 for probe vehicle number greater than 30.
Cvalue	Number	100	The ‘Confidential Value’ evaluating the confidence of the record, estimated by number of probe vehicles in the given time period, the maximum is 100.

A.6 Indiana State Climate Office (INClimate)

Table A.7 Description of data items available in Indiana State Climate Office (INClimate)

Data Item Name	**Data Item Type**	**Example**	**Description**
Lat	Number	37.74242	The latitude of the virtual station
Lon	Number	-88.19418	The longitude of the virtual station
Time	Date/Time	2018-01-01 01 :00:00	The time variable in "YYYY-MM-DD HH:NN:SS" format, precise to hour. The data recording resolution is by-hour.
Precipitation	Number	3.5	The precipitation in millimeters (volume/area), the unit is mm, representing the hourly precipitation intensity.
Temperature	Number	40	The temperature in Fahrenheit

Note: The raw weather data were stored in several large matrix for precipitation and temperature of different years; the processed weather data is presented here.

APPENDIX B. ROADWAY DATA COLLECTION

B.1 Rural Freeways

Table B.1 Descriptive statistics of modeling dataset for rural freeways.

Variable	N	Miss	Mean	Std. Dev.	Min	Max
Roadway Features						
Annual average daily traffic (veh/day)	65,305	0	31,949.83	8,742.82	13,152.00	47,253.00
Single-Unit Truck and Bus AADT (veh/day)	65,305	0	1,464.02	809.78	300.00	3,126.00
Combination truck AADT (veh/day)	65,305	0	9,806.15	3,243.56	4,913.00	15,500.00
Passenger car AADT (1,000 veh/day)	65,305	0	22.14	6.82	8.24	39.64
Heavy truck AADT (1,000 veh/day)	65,305	0	9.81	3.24	4.91	15.50
Segment length (ft)	65,305	0	1,321.01	10.98	1,300.56	1,343.30
Crash attenuator on the median	65,305	0	0.109	0.311	0.000	1.000
Crash attenuator on the roadside	65,305	0	0.005	0.073	0.000	1.000
Overpassing road indicator	65,305	0	0.109	0.312	0.000	1.000
Prohibited U-turn indicator	65,305	0	0.113	0.317	0.000	1.000
Curve deflection angle (deg)	65,305	0	5.470	11.784	0.000	46.255
Curve A indicator (Under 3.5 degrees)	65,305	0	0.017	0.129	0.000	1.000
Curve B indicator (3.5 – 5.4 degrees)	65,305	0	0.033	0.179	0.000	1.000
Curve C indicator (5.5 – 8.4 degrees)	65,305	0	0.040	0.195	0.000	1.000
Curve D indicator (8.5 – 13.9 degrees)	65,305	0	0.058	0.234	0.000	1.000
Curve E indicator (14.0 – 27.9 degrees)	65,305	0	0.034	0.182	0.000	1.000
Curve F indicator (28 degrees or more)	65,305	0	0.094	0.291	0.000	1.000
Curve length inside segment (ft)	65,305	0	207.464	406.514	0.000	1,342.980
Total curve length (ft)	65,305	0	566.849	1,032.530	0.000	4,572.730
Segment proportion with curve	65,305	0	0.157	0.308	0.000	1.000
Segment proportion with median cable barrier	65,305	0	0.693	0.429	0.000	1.000
Segment proportion with median concrete barrier	65,305	0	0.018	0.081	0.000	1.000
Segment proportion with roadside concrete barrier	65,305	0	0.019	0.084	0.000	1.000
Segment proportion with median guardrail	65,305	0	0.050	0.142	0.000	1.000
Segment proportion with roadside guardrail	65,305	0	0.147	0.263	0.000	1.000
Segment proportion with bridge	65,305	0	0.017	0.078	0.000	1.000
Segment proportion with lighting	65,305	0	0.018	0.095	0.000	0.901
Segment proportion with asphalt pavement	65,305	0	0.885	0.280	0.000	1.000
Segment proportion with concrete pavement	65,305	0	0.115	0.280	0.000	1.000
Segment proportion with parallel entering ramp	65,305	0	0.023	0.124	0.000	1.000
Segment proportion with parallel exiting ramp	65,305	0	0.014	0.077	0.000	0.758
Segment proportion with tapered entering ramp	65,305	0	0.004	0.050	0.000	0.818
Segment proportion with tapered exiting ramp	65,305	0	0.002	0.024	0.000	0.353
Segment proportion with entering ramp	65,305	0	0.028	0.133	0.000	1.000
Segment proportion with exiting ramp	65,305	0	0.016	0.081	0.000	0.758
Segment proportion with median rumble strip	65,305	0	0.972	0.112	0.000	1.000

Variable	N	Miss	Mean	Std. Dev.	Min	Max
Segment proportion with roadside rumble strip	65,305	0	0.965	0.131	0.000	1.000
Average median cable barrier offset (ft)	65,305	0	56.373	87.700	11.925	768.980
Minimum median cable barrier offset (ft)	48,538	16,767	25.971	13.075	10.760	44.920
Maximum median cable barrier offset (ft)	48,538	16,767	34.536	16.660	13.090	105.460
Average median concrete barrier offset (ft)	65,305	0	73.643	83.773	4.120	768.980
Minimum median concrete barrier offset (ft)	6,565	58,740	7.930	5.447	3.760	28.750
Maximum median concrete barrier offset (ft)	6,565	58,740	8.799	5.647	4.480	29.760
Average roadside concrete barrier offset (ft)	6,220	59,085	10.773	1.339	6.642	13.830
Minimum roadside concrete barrier offset (ft)	6,220	59,085	10.360	1.506	5.540	13.480
Maximum roadside concrete barrier offset (ft)	6,220	59,085	11.358	1.075	9.230	14.240
Average median guardrail offset (ft)	65,305	0	59.366	31.886	5.645	303.090
Minimum median guardrail offset (ft)	9,561	55,744	12.283	13.066	4.480	47.510
Maximum median guardrail offset (ft)	9,561	55,744	28.006	21.000	5.920	114.730
Average roadside guardrail offset (ft)	23,710	41,595	11.457	1.861	8.705	30.010
Minimum roadside guardrail offset (ft)	23,710	41,595	10.692	1.051	8.410	14.240
Maximum roadside guardrail offset (ft)	23,710	41,595	12.318	3.697	8.860	47.240
Segment proportion with median barrier	65,305	0	0.761	0.397	0.000	1.022
Segment proportion with roadside barrier	65,305	0	0.166	0.286	0.000	1.000
Average median barrier offset (ft)	52,723	12,582	28.452	14.338	5.645	73.665
Near median barrier indicator (offset < 30 ft)	65,305	0	0.449	0.497	0.000	1.000
Far median barrier indicator (offset ≥ 30 ft)	65,305	0	0.358	0.480	0.000	1.000
Number of median point hazards	65,305	0	0.199	0.407	0.000	2.000
Number of roadside point hazards	65,305	0	0.229	0.433	0.000	2.000
Average median width (ft)	65,305	0	79.272	81.227	50.470	768.980
Minimum median width (ft)	65,305	0	63.736	41.833	49.780	736.510
Maximum median width (ft)	65,305	0	95.569	131.270	50.710	802.350
Average median shoulder width (ft)	65,305	0	5.301	0.760	4.205	15.393
Minimum median shoulder width (ft)	65,305	0	4.849	0.549	3.690	6.150
Maximum median shoulder width (ft)	65,305	0	5.803	1.688	4.510	26.040
Average roadside shoulder width (ft)	65,305	0	11.265	0.762	9.125	13.980
Minimum roadside shoulder width (ft)	65,305	0	10.753	0.742	8.410	13.090
Maximum roadside shoulder width (ft)	65,305	0	11.790	0.955	9.840	15.070
Passenger car speed limit (mph)	65,305	0	69.681	1.223	65.000	70.000
Heavy truck speed limit (mph)	65,305	0	64.681	1.223	60.000	65.000
Operational Characteristics						
Average hourly travel speed (mph)	65,305	0	64.048	4.090	2.774	82.062
Standard deviation of hourly travel speed (mph)	65,285	20	2.561	1.939	0.000	27.681
Minimum hourly travel speed (mph)	65,305	0	60.034	6.256	0.622	82.062
Maximum hourly travel speed (mph)	65,305	0	68.373	4.670	5.227	98.000
Hourly travel speed range (mph)	65,305	0	8.339	6.129	0.000	77.953
Hourly speed trend	65,285	20	-0.008	0.455	-11.488	20.700
Downtrend speed indicator	65,285	20	0.355	0.478	0.000	1.000
Number of periods with probe density from 5 to 9	24,912	40,393	4.550	3.729	0.000	12.000
Number of periods with probe density < 5	24,912	40,393	2.704	3.784	0.000	12.000
Observed hourly traffic volume (veh/h)	43,413	21,892	650.990	423.100	13.000	3,123.000

Variable	N	Miss	Mean	Std. Dev.	Min	Max
Number of periods with probe density > 9	24,912	40,393	4.602	4.751	0.000	12.000
Forecasted hourly traffic volume (veh/h)	65,305	0	662.452	428.006	13.000	3,123.000
Instrumental traffic volume indicator	65,305	0	0.335	0.472	0.000	1.000
Free flow speed (mph)	65,305	0	66.092	1.749	34.459	73.994
Congestion index	65,305	0	0.032	0.056	0.000	0.957
Non-congested traffic state indicator	65,305	0	0.963	0.188	0.000	1.000
Congested traffic state indicator	65,305	0	0.003	0.058	0.000	1.000
Intermediate traffic state indicator	65,305	0	0.033	0.179	0.000	1.000
Crash indicator	65,305	0	0.032	0.176	0.000	1.000
Weather Conditions						
Precipitation intensity (in)	65,305	0	0.005	0.035	0.000	1.703
Light rain (precipitation < 0.098 in)	65,305	0	0.075	0.264	0.000	1.000
Moderate rain (precipitation 0.098–0.394 in)	65,305	0	0.013	0.111	0.000	1.000
Heavy rain (precipitation 0.394–1.969 in)	65,305	0	0.002	0.040	0.000	1.000
Violent rain (precipitation > 1.969 in)	65,305	0	0.000	0.000	0.000	0.000
Rain (precipitation > 0 in)	65,305	0	0.090	0.286	0.000	1.000
Temperature (F)	65,305	0	52.018	21.111	-18.261	98.293
Freezing temperature (temperature ≤ 32F)	65,305	0	0.192	0.394	0.000	1.000
Time Variables						
Year	65,305	0	2,015.890	1.500	2,014.000	2,018.000
Month	65,305	0	6.661	3.412	1.000	12.000
Weekend	65,305	0	0.286	0.452	0.000	1.000
Monday to Thursday	65,305	0	1.000	0.000	1.000	1.000
Friday	65,305	0	0.145	0.352	0.000	1.000
Saturday	65,305	0	0.141	0.348	0.000	1.000
Sunday	65,305	0	0.144	0.351	0.000	1.000

B.2 Signalized Intersections

Table B.2 Descriptive statistics of modeling dataset for signalized intersections

Variable	Description	N	Mean	Std Dev	Min	Max
Crash Severity						
InjuredNmb	Injury number of the crash	1,136	0.255	0.694	0	8
DeadNmb	Dead number of the crash	1,136	0.004	0.066	0	1
Intersection Related Variables						
StateIntID	Intersection ID	39,986	13,835.15	6488.34	967	23254
IntLong	Intersection Longitude	39,986	-86.158	0.732	-87.509	-84.871
IntLat	Intersection Latitude	39,986	40.341	1.031	38.169	41.720
IntSize	Size of intersection (longer diagonal distance in ft)	39,986	154.17	49.30	62	285.5
IntAng	Skewed angle of intersection (in degree)	39,986	84.76	8.30	60	90
Time Variables						
Datetime	Unix timestamp	39,986	1.5E+09	2.4E+07	1.5E+09	1.6E+09
Year	Year	39,986	2018	1	2017	2019
Month	Month	39,986	7.1	3.3	1.0	12.0

Variable	Description	N	Mean	Std Dev	Min	Max
Dayofweek	Day of week (1-Mon, 7-Sun)	39,986	4.0	2.0	1.0	7.0
Weather						
EstT	Estimated Temperature	39,986	53.54	20.85	-18.19	114.36
Pcp	Estimated precipitation	39,986	0.14	0.88	0.00	31.75
Major Road Traffic Operational Features						
Volume_1	Volume (veh/hr)	39,986	1,672	1631	20	15790
AADT_1	AADT (veh/hr)	39,986	15,443	9338.45	1346	55476
MeanSpd_1	Mean Speed (mph)	24,940	35.53	10.41	0.00	65.45
StdSpd_1	Standard deviation of speed (mph)	24,940	3.59	3.14	0.00	24.22
MaxSpd_1	Max Speed (mph)	24,940	40.78	11.46	0.00	83.00
MinSpd_1	Min Speed (mph)	24,940	28.44	11.78	0.00	62.00
SpdRange_1	Speed Range (mph)	24,940	12.34	10.74	0.00	64.00
Major Road Geometric Features						
CloseInt_1	Exist close intersection (500 ft)	39,986	0.42	0.49	0	1
IncreaseLane_1	Increased number of lanes at the intersection	39,986	1.21	0.71	0	3
IntLane_1	Number of lanes at the intersection	39,986	2.99	0.99	1	5
SegLane_1	Number of lanes on the segment	39,986	1.78	0.81	1	4
MedType_1	Median separation type at the intersection	39,986	1.54	0.80	1	3
MedWid_1	Median width	39,986	7.63	8.53	0	24
ShdType_1	Shoulder type	39,986	1.88	0.99	1	3
ShdWid_1	Shoulder width	39,986	2.92	3.73	0	12
SideWalk_1	Exist sidewalk	39,986	0.35	0.48	0	1
LeftLane_1	Exist left turn lane	39,986	0.89	0.32	0	1
LeftTurnPhase_1	Exist exclusive left phase	39,986	0.76	0.43	0	1
RightLane_1	Exist right turn lane	39,986	0.51	0.50	0	1
TriaIsland_1	Exist triangular island	39,986	0.09	0.29	0	1
RightTurnSightBlock_1	Exist sight distance blocking for right turn traffic	39,986	0.04	0.19	0	1
PedSignal_1	Exist pedestrian signals	39,986	0.33	0.47	0	1
CloseConnector_1	Exist close driveways/connectors near the intersection on the approach	39,986	0.66	0.48	0	1
Crossing Road Traffic Operational Features						
Volume_2	Volume (veh/hr)	39,986	1,064	1167	9	13,748
AADT_2	AADT (veh/hr)	39,986	9,974	7324	909	55,476
MeanSpd_2	Mean Speed (mph)	24,004	33.59	9.73	0.00	62.77

Variable	Description	N	Mean	Std Dev	Min	Max
StdSpd_2	Standard deviation of speed (mph)	24,004	2.86	2.81	0.00	21.64
MaxSpd_2	Max Speed (mph)	24,004	37.58	10.55	0.00	76.00
MinSpd_2	Min Speed (mph)	24,004	28.17	10.95	0.00	62.00
SpdRange_2	Speed Range (mph)	24,004	9.41	9.22	0.00	65.00
Crossing Road Geometric Features						
CloseInt_2	Exist close intersection (500 ft)	39,986	0.43	0.50	0	1
IncreaseLane_2	Increased number of lanes at the intersection	39,986	1.09	0.72	0	3
IntLane_2	Number of lanes at the intersection	39,986	2.63	0.97	1	5
SegLane_2	Number of lanes on the segment	39,986	1.55	0.71	1	4
MedType_2	Median separation type at the intersection	39,986	1.27	0.58	1	3
MedWid_2	Median width	39,986	4.41	6.59	0	24
ShdType_2	Shoulder type	39,986	1.77	0.97	1	3
ShdWid_2	Shoulder width	39,986	2.38	3.33	0	12
SideWalk_2	Exist sidewalk	39,986	0.35	0.48	0	1
LeftLane_2	Exist left turn lane	39,986	0.83	0.38	0	1
LeftTurnPhase_2	Exist exclusive left phase	39,986	0.67	0.47	0	1
RightLane_2	Exist right turn lane	39,986	0.41	0.49	0	1
TriaIsland_2	Exist triangular island	39,986	0.09	0.29	0	1
RightTurnSightBlock_2	Exist sight distance blocking for right turn traffic	39,986	0.04	0.19	0	1
PedSignal_2	Exist pedestrian signals	39,986	0.33	0.47	0	1
CloseConnector_2	Exist close driveways/connectors near the intersection on the approach	39,986	0.70	0.46	0	1
Other Variables						
Func_1	Major road functional class	39,986	3.23	0.66	2	7
SpdLimit_1	Major road speed limit	39,986	42.55	12.77	20	99
Func_2	Crossing road functional class	39,986	3.63	0.83	2	7
SpdLimit_2	Crossing road speed limit	39,986	40.80	10.88	20	99
Delay_1	Estimated Delay on Major Road	24,930	10.339	17.381	-33.263	247.740
Delay_2	Estimated Delay on Crossing Road	23,971	11.904	17.896	-32.294	310.649

APPENDIX C. TRAFFIC VOLUME MODELS

C.1 Out-of-Sample Prediction Models

Table C.1 Descriptive statistics of 2017 dataset

Variable	Mean	Std. Dev.	Minimum	Maximum
Average hourly travel speed (mph)	55.93	10.84	3.00	98.00
Hourly travel speed range (mph)	11.06	9.71	0.00	89.00
Periods with probe density < 5 (Na)	4.80	3.90	0.00	12.00
Periods with probe density from 5 to 9 (Nb)	2.49	3.33	0.00	12.00
Periods with probe density > 9 (Nc)	1.87	3.62	0.00	12.00
Traffic volume (veh/h)	854.74	1,042.00	0.00	9,280.00
Logarithm of volume	6.10	1.24	0.00	9.14
International Roughness Index - IRI (in/mi)	73.03	35.11	0.00	174.00
AADT cars (1,000 veh/day)	35.90	34.50	2.59	167.64
AADT trucks (1,000 veh/day)	5.94	6.73	0.01	37.71
Within city limits (1-yes, 0-no)	0.44	0.50	0.00	1.00
Nearest city area (1,000 ha)	1,161.43	1,836.87	2.87	6,325.86
Distance to city (miles)	2.45	3.81	0.00	20.63
Distance to downtown (miles)	12.77	12.06	1.55	66.98
Urban	0.44	0.50	0.00	1.00
Suburban	0.23	0.42	0.00	1.00
Rural	0.34	0.47	0.00	1.00
Interstates	0.54	0.50	0.00	1.00
Other freeways and expressways	0.05	0.22	0.00	1.00
Principal arterials	0.36	0.48	0.00	1.00
Minor arterials	0.03	0.18	0.00	1.00
Major collectors	0.01	0.09	0.00	1.00
Lanes = 2	0.13	0.33	0.00	1.00
Lanes = 3	0.01	0.10	0.00	1.00
Lanes = 4	0.62	0.49	0.00	1.00
Lanes = 5	0.02	0.14	0.00	1.00
Lanes = 6	0.16	0.37	0.00	1.00
Lanes = 8	0.07	0.25	0.00	1.00
Speed limit = 45 mph	0.05	0.22	0.00	1.00
Speed limit = 50 mph	0.07	0.26	0.00	1.00
Speed limit = 55 mph	0.34	0.47	0.00	1.00
Speed limit = 60 mph	0.18	0.38	0.00	1.00
Speed limit = 65 mph	0.04	0.19	0.00	1.00
Speed limit = 70 mph	0.31	0.46	0.00	1.00
Temperature (F)	53.78	17.95	-2.90	87.20

Variable	Mean	Std. Dev.	Minimum	Maximum
Precipitation (in)	0.09	0.35	0.00	7.00
January	0.08	0.27	0.00	1.00
February	0.07	0.26	0.00	1.00
March	0.06	0.24	0.00	1.00
April	0.07	0.25	0.00	1.00
May	0.08	0.26	0.00	1.00
June	0.07	0.26	0.00	1.00
July	0.09	0.28	0.00	1.00
August	0.09	0.29	0.00	1.00
September	0.09	0.28	0.00	1.00
October	0.10	0.30	0.00	1.00
November	0.09	0.29	0.00	1.00
December	0.11	0.31	0.00	1.00
Sunday	0.14	0.34	0.00	1.00
Monday	0.15	0.35	0.00	1.00
Tuesday	0.13	0.34	0.00	1.00
Wednesday	0.16	0.36	0.00	1.00
Thursday	0.15	0.36	0.00	1.00
Friday	0.14	0.34	0.00	1.00
Saturday	0.14	0.35	0.00	1.00
12:00 a.m. – 12:59 a.m.	0.04	0.19	0.00	1.00
1:00 a.m. – 1:59 a.m.	0.04	0.19	0.00	1.00
2:00 a.m. – 2:59 a.m.	0.04	0.19	0.00	1.00
3:00 a.m. – 3:59 a.m.	0.04	0.19	0.00	1.00
4:00 a.m. – 4:59 a.m.	0.04	0.19	0.00	1.00
5:00 a.m. – 5:59 a.m.	0.04	0.20	0.00	1.00
6:00 a.m. – 6:59 a.m.	0.04	0.20	0.00	1.00
7:00 a.m. – 7:59 a.m.	0.04	0.20	0.00	1.00
8:00 a.m. – 8:59 a.m.	0.04	0.21	0.00	1.00
9:00 a.m. – 9:59 a.m.	0.04	0.21	0.00	1.00
10:00 a.m. – 10:59 a.m.	0.04	0.21	0.00	1.00
11:00 a.m. – 11:59 a.m.	0.04	0.21	0.00	1.00
12:00 p.m. – 12:59 p.m.	0.04	0.21	0.00	1.00
1:00 p.m. – 1:59 p.m.	0.04	0.21	0.00	1.00
2:00 p.m. – 2:59 p.m.	0.04	0.21	0.00	1.00
3:00 p.m. – 3:59 p.m.	0.04	0.21	0.00	1.00
4:00 p.m. – 4:59 p.m.	0.04	0.21	0.00	1.00
5:00 p.m. – 5:59 p.m.	0.04	0.21	0.00	1.00
6:00 p.m. – 6:59 p.m.	0.04	0.20	0.00	1.00
7:00 p.m. – 7:59 p.m.	0.04	0.20	0.00	1.00

Variable	Mean	Std. Dev.	Minimum	Maximum
8:00 p.m. – 8:59 p.m.	0.04	0.20	0.00	1.00
9:00 p.m. – 9:59 p.m.	0.04	0.20	0.00	1.00
10:00 p.m. – 10:59 p.m.	0.04	0.19	0.00	1.00
11:00 p.m. – 11:59 p.m.	0.04	0.19	0.00	1.00

Table C.2 Least-square estimates of 2017 hourly volume model for rural freeways

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
Intercept	2.5500	0.0300	85.05	<.0001
AADT cars (1,000 veh/day)	0.0376	0.0003	113.36	<.0001
AADT trucks (1,000 veh/day)	0.0305	0.0007	42.5	<.0001
International Roughness Index - IRI (in/mi)	0.0002	0.00004	3.92	<.0001
Speed Limit = 50 mph	0.0817	0.0052	15.75	<.0001
Speed Limit = 60 mph	0.1220	0.0047	26.18	<.0001
Speed Limit = 70 mph	–	–	–	–
Nearest city area (1,000 ha)	-0.0001	0.000003	-21.46	<.0001
Temperature (F)	0.0011	0.0001	11.35	<.0001
Precipitation (in)	-0.0112	0.0028	-3.99	<.0001
Average hourly travel speed (mph)	0.0154	0.0004	35.51	<.0001
Hourly travel speed range (mph)	0.0021	0.0002	9.86	<.0001
Periods with probe density < 5 (Na)	0.0745	0.0007	101.54	<.0001
Periods with probe density from 5 to 9 (Nb)	0.1010	0.0009	116.28	<.0001
Na*Nb	0.0007	0.0001	7.1	<.0001
Periods with probe density > 9 (Nc)	0.1060	0.0010	107.58	<.0001
Nb*Nc	0.0006	0.0001	5.35	<.0001
January	-0.0784	0.0044	-17.99	<.0001
February	-0.0223	0.0045	-4.98	<.0001
March	0.0459	0.0047	9.71	<.0001
April	0.0638	0.0054	11.77	<.0001
May	0.1190	0.0054	21.99	<.0001
June	0.1530	0.0061	25.29	<.0001
July	0.1570	0.0060	26.18	<.0001
August	0.1150	0.0056	20.56	<.0001
September	0.0574	0.0055	10.46	<.0001
October	0.0769	0.0047	16.34	<.0001
November	0.0631	0.0041	15.29	<.0001
December	–	–	–	–
Sunday	-0.0928	0.0034	-27.3	<.0001
Monday	-0.1000	0.0034	-29.27	<.0001
Tuesday	-0.0766	0.0037	-20.89	<.0001

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
Wednesday	-0.0436	0.0036	-12.13	<.0001
Thursday	0.0057	0.0036	1.58	0.1150
Friday	0.1110	0.0036	30.81	<.0001
Saturday	–	–	–	–
12:00 a.m. – 12:59 a.m.	-0.2470	0.0063	-39.31	<.0001
1:00 a.m. – 1:59 a.m.	-0.4790	0.0063	-75.9	<.0001
2:00 a.m. – 2:59 a.m.	-0.6600	0.0063	-104.48	<.0001
3:00 a.m. – 3:59 a.m.	-0.7280	0.0063	-115.6	<.0001
4:00 a.m. – 4:59 a.m.	-0.6230	0.0063	-99.05	<.0001
5:00 a.m. – 5:59 a.m.	-0.3900	0.0063	-61.87	<.0001
6:00 a.m. – 6:59 a.m.	-0.0725	0.0064	-11.41	<.0001
7:00 a.m. – 7:59 a.m.	0.1960	0.0064	30.59	<.0001
8:00 a.m. – 8:59 a.m.	0.3850	0.0065	59.34	<.0001
9:00 a.m. – 9:59 a.m.	0.5420	0.0066	82.44	<.0001
10:00 a.m. – 10:59 a.m.	0.6630	0.0066	99.98	<.0001
11:00 a.m. – 11:59 a.m.	0.7460	0.0067	112.01	<.0001
12:00 p.m. – 12:59 p.m.	0.8030	0.0067	120.42	<.0001
1:00 p.m. – 1:59 p.m.	0.8430	0.0067	126.12	<.0001
2:00 p.m. – 2:59 p.m.	0.8850	0.0067	132.38	<.0001
3:00 p.m. – 3:59 p.m.	0.9350	0.0067	140.07	<.0001
4:00 p.m. – 4:59 p.m.	0.9670	0.0067	145.41	<.0001
5:00 p.m. – 5:59 p.m.	0.9430	0.0066	142.91	<.0001
6:00 p.m. – 6:59 p.m.	0.8560	0.0065	131.43	<.0001
7:00 p.m. – 7:59 p.m.	0.7230	0.0064	112.33	<.0001
8:00 p.m. – 8:59 p.m.	0.5740	0.0064	90.32	<.0001
9:00 p.m. – 9:59 p.m.	0.4120	0.0063	65.29	<.0001
10:00 p.m. – 10:59 p.m.	0.2130	0.0063	33.85	<.0001
11:00 p.m. – 11:59 p.m.	–	–	–	–

Table C.3 Least-square estimates of 2017 hourly volume model for urban freeways

Parameter	**Estimate**	**Std. Error**	**t Value**	**Pr. > |t|**
Intercept	3.9000	0.0181	215.14	<.0001
AADT cars (1,000 veh/day)	0.0107	0.0001	172.38	<.0001
AADT trucks (1,000 veh/day)	-0.0076	0.0002	-44.15	<.0001
International Roughness Index - IRI (in/mi)	0.0008	0.00003	28.53	<.0001
Speed Limit = 50 mph	0.4150	0.0069	59.96	<.0001
Speed Limit = 55 mph	0.1140	0.0033	34.48	<.0001
Speed Limit = 60 mph	-0.2750	0.0050	-55.52	<.0001
Speed Limit = 65 mph	-0.1270	0.0047	-26.89	<.0001
Speed Limit = 70 mph	–	–	–	–
Nearest city area (1,000 ha)	0.00004	0.000001	61.6	<.0001
Lanes = 4	0.0778	0.0053	14.76	<.0001
Lanes = 5	0.0320	0.0060	5.35	<.0001
Lanes = 6	0.1280	0.0032	39.99	<.0001
Lanes = 8	–	–	–	–
Average hourly travel speed (mph)	0.0090	0.0002	43.97	<.0001
Hourly travel speed range (mph)	0.0050	0.0001	37.06	<.0001
Periods with probe density < 5 (Na)	0.0543	0.0007	79.07	<.0001
Periods with probe density from 5 to 9 (Nb)	0.0842	0.0007	112.55	<.0001
Na*Nb	0.0020	0.0001	20.78	<.0001
Periods with probe density > 9 (Nc)	0.0997	0.0008	132.47	<.0001
Na*Nc	0.0046	0.0007	6.6	<.0001
Nb*Nc	-0.0003	0.0001	-2.76	0.0060
Na*Nb*Nc	-0.0005	0.0001	-3.77	0.0000
January	0.0419	0.0063	6.68	<.0001
February	0.0799	0.0068	11.82	<.0001
March	0.2160	0.0067	32.1	<.0001
April	0.1660	0.0066	25.01	<.0001
May	0.2580	0.0066	39.01	<.0001
June	0.2760	0.0070	39.34	<.0001
July	0.2870	0.0068	41.89	<.0001
August	0.2350	0.0069	34.22	<.0001
September	0.1820	0.0065	27.79	<.0001
October	0.0625	0.0059	10.62	<.0001
November	0.0313	0.0058	5.41	<.0001
December	–	–	–	–
Suburban | January	0.1180	0.0102	11.6	<.0001
Suburban | February	0.1340	0.0105	12.75	<.0001
Suburban | March	0.0565	0.0106	5.32	<.0001
Suburban | April	0.1240	0.0103	12.12	<.0001

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
Suburban | May	0.0526	0.0103	5.11	<.0001
Suburban | June	0.0335	0.0104	3.23	0.0010
Suburban | July	0.0134	0.0100	1.33	0.1820
Suburban | August	0.0410	0.0102	4.03	<.0001
Suburban | September	0.0456	0.0100	4.55	<.0001
Suburban | October	0.1740	0.0098	17.69	<.0001
Suburban | November	0.2200	0.0100	22.06	<.0001
Suburban | December	0.1960	0.0097	20.09	<.0001
Sunday	-0.1190	0.0032	-37.38	<.0001
Monday	-0.0379	0.0032	-11.77	<.0001
Tuesday	-0.0157	0.0034	-4.62	<.0001
Wednesday	0.0015	0.0033	0.45	0.6510
Thursday	0.0363	0.0033	10.92	<.0001
Friday	0.1270	0.0033	38.05	<.0001
Saturday	–	–	–	–
Suburban | Sunday	0.0307	0.0058	5.31	<.0001
Suburban | Monday	-0.0491	0.0058	-8.43	<.0001
Suburban | Tuesday	-0.0774	0.0060	-12.94	<.0001
Suburban | Wednesday	-0.0646	0.0058	-11.18	<.0001
Suburban | Thursday	-0.0520	0.0057	-9.06	<.0001
Suburban | Friday	-0.0318	0.0059	-5.42	<.0001
Suburban | Saturday	–	–	–	–
12:00 a.m. - 12:59 a.m.	-0.2510	0.0090	-27.78	<.0001
1:00 a.m. - 1:59 a.m.	-0.5220	0.0090	-57.81	<.0001
2:00 a.m. – 2:59 a.m.	-0.7110	0.0090	-78.72	<.0001
3:00 a.m. – 3:59 a.m.	-0.7470	0.0090	-82.78	<.0001
4:00 a.m. – 4:59 a.m.	-0.5860	0.0090	-64.86	<.0001
5:00 a.m. – 5:59 a.m.	-0.2250	0.0090	-24.84	<.0001
6:00 a.m. – 6:59 a.m.	0.1410	0.0091	15.58	<.0001
7:00 a.m. – 7:59 a.m.	0.4240	0.0091	46.63	<.0001
8:00 a.m. – 8:59 a.m.	0.5340	0.0091	58.59	<.0001
9:00 a.m. – 9:59 a.m.	0.6270	0.0092	68.43	<.0001
10:00 a.m. – 10:59 a.m.	0.7170	0.0092	78.03	<.0001
11:00 a.m. – 11:59 a.m.	0.7840	0.0092	85.29	<.0001
12:00 p.m. – 12:59 p.m.	0.8160	0.0092	88.77	<.0001
1:00 p.m. – 1:59 p.m.	0.8450	0.0092	91.97	<.0001
2:00 p.m. – 2:59 p.m.	0.8910	0.0092	96.91	<.0001
3:00 p.m. – 3:59 p.m.	0.9520	0.0092	103.62	<.0001
4:00 p.m. – 4:59 p.m.	1.0100	0.0092	110.2	<.0001
5:00 p.m. – 5:59 p.m.	1.0200	0.0091	111.82	<.0001

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
6:00 p.m. – 6:59 p.m.	0.9110	0.0091	99.92	<.0001
7:00 p.m. – 7:59 p.m.	0.7090	0.0091	77.99	<.0001
8:00 p.m. – 8:59 p.m.	0.5460	0.0091	60.29	<.0001
9:00 p.m. – 9:59 p.m.	0.3870	0.0090	42.81	<.0001
10:00 p.m. – 10:59 p.m.	0.2020	0.0090	22.39	<.0001
11:00 p.m. – 11:59 p.m.	–	–	–	–
Suburban | 12:00 a.m. – 12:59 a.m.	-0.0579	0.0108	-5.37	<.0001
Suburban | 1:00 a.m. – 1:59 a.m.	-0.1230	0.0108	-11.41	<.0001
Suburban | 2:00 a.m. – 2:59 a.m.	-0.1310	0.0108	-12.18	<.0001
Suburban | 3:00 a.m. – 3:59 a.m.	-0.0961	0.0108	-8.92	<.0001
Suburban | 4:00 a.m. – 4:59 a.m.	-0.0138	0.0108	-1.28	0.1990
Suburban | 5:00 a.m. – 5:59 a.m.	0.0596	0.0108	5.52	<.0001
Suburban | 6:00 a.m. – 6:59 a.m.	0.1670	0.0108	15.52	<.0001
Suburban | 7:00 a.m. – 7:59 a.m.	0.2080	0.0108	19.27	<.0001
Suburban | 8:00 a.m. – 8:59 a.m.	0.1600	0.0108	14.88	<.0001
Suburban | 9:00 a.m. – 9:59 a.m.	0.0472	0.0108	4.37	<.0001
Suburban | 10:00 a.m. – 10:59 a.m.	-0.0253	0.0108	-2.35	0.0190
Suburban | 11:00 a.m. – 11:59 a.m.	-0.0395	0.0108	-3.66	0.0000
Suburban | 12:00 p.m. – 12:59 p.m.	-0.0187	0.0108	-1.74	0.0820
Suburban | 1:00 p.m. – 1:59 p.m.	-0.0083	0.0108	-0.77	0.4430
Suburban | 2:00 p.m. – 2:59 p.m.	0.0075	0.0108	0.69	0.4870
Suburban | 3:00 p.m. – 3:59 p.m.	0.0476	0.0108	4.41	<.0001
Suburban | 4:00 p.m. – 4:59 p.m.	0.0787	0.0108	7.3	<.0001
Suburban | 5:00 p.m. – 5:59 p.m.	0.0808	0.0108	7.49	<.0001
Suburban | 6:00 p.m. – 6:59 p.m.	0.0485	0.0108	4.5	<.0001
Suburban | 7:00 p.m. – 7:59 p.m.	0.0322	0.0108	2.99	0.0030
Suburban | 8:00 p.m. – 8:59 p.m.	0.0204	0.0108	1.89	0.0590
Suburban | 9:00 p.m. – 9:59 p.m.	0.0397	0.0108	3.68	0.0000
Suburban | 10:00 p.m. – 10:59 p.m.	0.0487	0.0108	4.52	<.0001
Suburban | 11:00 p.m. – 11:59 p.m.	–	–	–	–

Table C.4 Least-square estimates of 2017 hourly volume model for rural non-freeways

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
Intercept	2.4100	0.0388	62.04	<.0001
AADT cars (1,000 veh/day)	0.1980	0.0033	59.2	<.0001
AADT trucks (1,000 veh/day)	-0.1110	0.0087	-12.7	<.0001
International Roughness Index - IRI (in/mi)	-0.0029	0.0003	-10.7	<.0001
Speed Limit = 50 mph	0.2750	0.0165	16.7	<.0001
Speed Limit = 55 mph	0.3140	0.0200	15.67	<.0001
Speed Limit = 60 mph	–	–	–	–
Nearest city area (1,000 ha)	-0.0005	0.0001	-5.09	<.0001
Temperature (F)	0.0012	0.0002	5.35	<.0001
Precipitation (in)	-0.0130	0.0056	-2.34	0.0190
Average hourly travel speed (mph)	-0.0021	0.0004	-4.8	<.0001
Hourly travel speed range (mph)	-0.0022	0.0002	-9.28	<.0001
Periods with probe density < 5 (Na)	0.0205	0.0009	21.89	<.0001
Periods with probe density from 5 to 9 (Nb)	0.0194	0.0072	2.68	0.0070
Na*Nb	-0.0002	0.0011	-0.16	0.8700
Na*Nb*Nc	0.0021	0.0017	1.22	0.2240
January	-0.0368	0.0094	-3.91	<.0001
February	0.0310	0.0099	3.15	0.0020
March	0.0139	0.0107	1.29	0.1960
April	0.0220	0.0112	1.97	0.0490
May	0.0945	0.0116	8.14	<.0001
June	0.1360	0.0130	10.48	<.0001
July	0.1130	0.0133	8.53	<.0001
August	0.1010	0.0124	8.2	<.0001
September	0.0755	0.0118	6.4	<.0001
October	0.0592	0.0104	5.68	<.0001
November	0.0325	0.0091	3.57	0.0000
December	–	–	–	–
Sunday	-0.1910	0.0077	-24.81	<.0001
Monday	0.0129	0.0076	1.7	0.0900
Tuesday	0.0527	0.0080	6.63	<.0001
Wednesday	0.0735	0.0077	9.58	<.0001
Thursday	0.0962	0.0078	12.41	<.0001
Friday	0.1830	0.0078	23.34	<.0001
Saturday	–	–	–	–
12:00 a.m. – 12:59 a.m.	-0.2820	0.0154	-18.33	<.0001
1:00 a.m. – 1:59 a.m.	-0.5930	0.0155	-38.16	<.0001
2:00 a.m. – 2:59 a.m.	-0.6520	0.0154	-42.31	<.0001
3:00 a.m. – 3:59 a.m.	-0.5540	0.0151	-36.68	<.0001

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
4:00 a.m. – 4:59 a.m.	-0.1220	0.0149	-8.2	<.0001
5:00 a.m. – 5:59 a.m.	0.4950	0.0145	34.28	<.0001
6:00 a.m. – 6:59 a.m.	0.9100	0.0145	63	<.0001
7:00 a.m. – 7:59 a.m.	1.1800	0.0145	81.12	<.0001
8:00 a.m. – 8:59 a.m.	1.2300	0.0146	84.35	<.0001
9:00 a.m. – 9:59 a.m.	1.2400	0.0146	84.47	<.0001
10:00 a.m. – 10:59 a.m.	1.2800	0.0146	87.44	<.0001
11:00 a.m. – 11:59 a.m.	1.3400	0.0146	91.92	<.0001
12:00 p.m. – 12:59 p.m.	1.3900	0.0146	95.23	<.0001
1:00 p.m. – 1:59 p.m.	1.4100	0.0146	96.53	<.0001
2:00 p.m. – 2:59 p.m.	1.4500	0.0146	99.31	<.0001
3:00 p.m. – 3:59 p.m.	1.5400	0.0146	105.72	<.0001
4:00 p.m. – 4:59 p.m.	1.6200	0.0144	111.98	<.0001
5:00 p.m. – 5:59 p.m.	1.5900	0.0144	110.27	<.0001
6:00 p.m. – 6:59 p.m.	1.3900	0.0144	96.85	<.0001
7:00 p.m. – 7:59 p.m.	1.1400	0.0143	79.48	<.0001
8:00 p.m. – 8:59 p.m.	0.9050	0.0144	62.7	<.0001
9:00 p.m. – 9:59 p.m.	0.6440	0.0146	44.09	<.0001
10:00 p.m. – 10:59 p.m.	0.3050	0.0150	20.39	<.0001
11:00 p.m. – 11:59 p.m.	–	–	–	–

Table C.5 Least-square estimates of 2017 hourly volume model for urban non-freeways

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
Intercept	3.9500	0.0226	174.71	<.0001
AADT cars (1,000 veh/day)	0.0518	0.0004	135.95	<.0001
AADT trucks (1,000 veh/day)	-0.1800	0.0049	-36.4	<.0001
International Roughness Index - IRI (in/mi)	-0.0020	0.0001	-20.23	<.0001
Speed Limit = 45 mph	0.0289	0.0061	4.74	<.0001
Speed Limit = 50 mph	-0.2640	0.0097	-27.2	<.0001
Speed Limit = 55 mph	0.0040	0.0061	0.65	0.5140
Speed Limit = 60 mph	–	–	–	–
Nearest city area (1,000 ha)	0.0000	0.0000	12.07	<.0001
Lanes = 2	0.4740	0.0182	26.1	<.0001
Lanes = 3	-0.0928	0.0127	-7.32	<.0001
Lanes = 4	0.3560	0.0132	27.03	<.0001
Lanes = 6	–	–	–	–
Temperature (F)	0.0008	0.0001	6.29	<.0001
Precipitation (in)	-0.0114	0.0035	-3.3	0.0010
Average hourly travel speed (mph)	-0.0056	0.0002	-28.21	<.0001
Hourly travel speed range (mph)	-0.0015	0.0001	-11.41	<.0001
Periods with probe density < 5 (Na)	0.0398	0.0006	65.93	<.0001
Na*Nb	-0.0010	0.0003	-3.77	0.0000
Periods with probe density > 9 (Nc)	0.1560	0.0157	9.94	<.0001
Na*Nc	-0.0290	0.0042	-6.97	<.0001
January	-0.1200	0.0085	-14.13	<.0001
February	-0.0393	0.0089	-4.44	<.0001
March	-0.0671	0.0089	-7.52	<.0001
April	-0.0343	0.0093	-3.7	0.0000
May	0.0232	0.0096	2.42	0.0150
June	0.0458	0.0104	4.39	<.0001
July	0.0074	0.0105	0.71	0.4770
August	0.0226	0.0096	2.36	0.0190
September	0.0069	0.0103	0.67	0.5040
October	0.0257	0.0101	2.55	0.0110
November	0.0808	0.0089	9.07	<.0001
December	–	–	–	–
Suburban | January	-0.1940	0.0180	-10.75	<.0001
Suburban | February	-0.2030	0.0182	-11.18	<.0001
Suburban | March	-0.1700	0.0185	-9.21	<.0001
Suburban | April	-0.1410	0.0181	-7.78	<.0001
Suburban | May	-0.1780	0.0182	-9.73	<.0001
Suburban | June	-0.1760	0.0183	-9.61	<.0001

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
Suburban | July	-0.1630	0.0179	-9.07	<.0001
Suburban | August	-0.1720	0.0176	-9.74	<.0001
Suburban | September	-0.1860	0.0180	-10.31	<.0001
Suburban | October	-0.1960	0.0182	-10.72	<.0001
Suburban | November	-0.2850	0.0180	-15.82	<.0001
Suburban | December	-0.2440	0.0173	-14.11	<.0001
Sunday	-0.2040	0.0079	-25.72	<.0001
Monday	-0.0553	0.0073	-7.62	<.0001
Tuesday	-0.0377	0.0074	-5.07	<.0001
Wednesday	-0.0075	0.0071	-1.05	0.2920
Thursday	0.0118	0.0072	1.64	0.1000
Friday	0.1120	0.0074	15.19	<.0001
Saturday	–	–	–	–
Suburban | Sunday	0.0069	0.0100	0.68	0.4940
Suburban | Monday	0.0674	0.0092	7.3	<.0001
Suburban | Tuesday	0.0916	0.0095	9.62	<.0001
Suburban | Wednesday	0.0740	0.0091	8.15	<.0001
Suburban | Thursday	0.0759	0.0091	8.31	<.0001
Suburban | Friday	0.0446	0.0094	4.74	<.0001
Suburban | Saturday	–	–	–	–
12:00 a.m. – 12:59 a.m.	-0.4560	0.0125	-36.54	<.0001
1:00 a.m. – 1:59 a.m.	-0.7890	0.0125	-63.06	<.0001
2:00 a.m. – 2:59 a.m.	-0.9660	0.0124	-77.61	<.0001
3:00 a.m. – 3:59 a.m.	-0.8200	0.0122	-67.2	<.0001
4:00 a.m. – 4:59 a.m.	-0.3160	0.0121	-26.1	<.0001
5:00 a.m. – 5:59 a.m.	0.2730	0.0118	23.15	<.0001
6:00 a.m. – 6:59 a.m.	0.7240	0.0116	62.63	<.0001
7:00 a.m. – 7:59 a.m.	0.9700	0.0115	84.47	<.0001
8:00 a.m. – 8:59 a.m.	1.0600	0.0114	92.73	<.0001
9:00 a.m. – 9:59 a.m.	1.0900	0.0114	95.73	<.0001
10:00 a.m. – 10:59 a.m.	1.1500	0.0114	101.06	<.0001
11:00 a.m. – 11:59 a.m.	1.2200	0.0114	106.97	<.0001
12:00 p.m. – 12:59 p.m.	1.2900	0.0114	112.99	<.0001
1:00 p.m. – 1:59 p.m.	1.3400	0.0114	117.69	<.0001
2:00 p.m. – 2:59 p.m.	1.4300	0.0114	125.19	<.0001
3:00 p.m. – 3:59 p.m.	1.5100	0.0114	131.85	<.0001
4:00 p.m. – 4:59 p.m.	1.5300	0.0114	133.72	<.0001
5:00 p.m. – 5:59 p.m.	1.4600	0.0114	128.03	<.0001
6:00 p.m. – 6:59 p.m.	1.2900	0.0114	113.69	<.0001
7:00 p.m. – 7:59 p.m.	1.0800	0.0114	94.24	<.0001

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
8:00 p.m. – 8:59 p.m.	0.8450	0.0115	73.3	<.0001
9:00 p.m. – 9:59 p.m.	0.6140	0.0117	52.52	<.0001
10:00 p.m. – 10:59 p.m.	0.3490	0.0119	29.35	<.0001
11:00 p.m. – 11:59 p.m.	–	–	–	–
Suburban | 12:00 a.m. – 12:59 a.m.	0.1250	0.0204	6.16	<.0001
Suburban | 1:00 a.m. – 1:59 a.m.	0.2040	0.0205	9.96	<.0001
Suburban | 2:00 a.m. – 2:59 a.m.	0.0977	0.0202	4.84	<.0001
Suburban | 3:00 a.m. – 3:59 a.m.	0.0375	0.0196	1.92	0.0550
Suburban | 4:00 a.m. – 4:59 a.m.	-0.1230	0.0192	-6.38	<.0001
Suburban | 5:00 a.m. – 5:59 a.m.	-0.1920	0.0188	-10.19	<.0001
Suburban | 6:00 a.m. – 6:59 a.m.	-0.2380	0.0183	-13.05	<.0001
Suburban | 7:00 a.m. – 7:59 a.m.	-0.1530	0.0179	-8.55	<.0001
Suburban | 8:00 a.m. – 8:59 a.m.	-0.0962	0.0178	-5.4	<.0001
Suburban | 9:00 a.m. – 9:59 a.m.	-0.0224	0.0178	-1.26	0.2070
Suburban | 10:00 a.m. – 10:59 a.m.	-0.0318	0.0177	-1.79	0.0730
Suburban | 11:00 a.m. – 11:59 a.m.	-0.0218	0.0177	-1.23	0.2190
Suburban | 12:00 p.m. – 12:59 p.m.	-0.0099	0.0177	-0.56	0.5770
Suburban | 1:00 p.m. – 1:59 p.m.	-0.0187	0.0177	-1.06	0.2910
Suburban | 2:00 p.m. – 2:59 p.m.	-0.0633	0.0177	-3.56	0.0000
Suburban | 3:00 p.m. – 3:59 p.m.	-0.0922	0.0177	-5.19	<.0001
Suburban | 4:00 p.m. – 4:59 p.m.	-0.0688	0.0178	-3.87	0.0000
Suburban | 5:00 p.m. – 5:59 p.m.	-0.0251	0.0178	-1.41	0.1600
Suburban | 6:00 p.m. – 6:59 p.m.	-0.0113	0.0180	-0.63	0.5320
Suburban | 7:00 p.m. – 7:59 p.m.	-0.0420	0.0183	-2.3	0.0220
Suburban | 8:00 p.m. – 8:59 p.m.	-0.0611	0.0186	-3.29	0.0010
Suburban | 9:00 p.m. – 9:59 p.m.	-0.0754	0.0189	-3.99	<.0001
Suburban | 10:00 p.m. – 10:59 p.m.	-0.0781	0.0193	-4.05	<.0001
Suburban | 11:00 p.m. – 11:59 p.m.	–	–	–	–

C.2 Missing Volume Imputation Models

Table C.6 Type III sum of squares summary of 2014–2016 hourly volume model for rural freeways

Source	DF	Type III SS	Mean Square	F Value	Pr. > F
AADT cars (1,000 veh/day)	1	21,151.88	21,151.88	348,270.00	<.0001
AADT trucks (1,000 veh/day)	1	4,478.54	4,478.54	73,740.10	<.0001
Light rain (precipitation < 0.098 in)	1	29.18	29.18	480.49	<.0001
Moderate rain (precipitation 0.098 – 0.394 in)	1	0.68	0.68	11.19	0.0008
Heavy rain (precipitation 0.394 – 1.969 in)	1	0.35	0.35	5.79	0.0161
Violent rain (precipitation > 1.969 in)	1	0.22	0.22	3.61	0.0575
Temperature (F)	1	16.05	16.05	264.26	<.0001

Freezing temperature (temperature ≤ 32°F)	1	1.73	1.73	28.41	<.0001
Average hourly travel speed (mph)	1	460.62	460.62	7,584.20	<.0001
Downtrend speed indicator	1	2.17	2.17	35.81	<.0001
Congested traffic state indicator	1	76.39	76.39	1,257.69	<.0001
Year	2	58.87	29.43	484.62	<.0001
Month	11	911.50	82.86	1,364.37	<.0001
Hour	23	61,544.70	2,675.86	44,058.50	<.0001
Friday*Hour	24	1,088.19	45.34	746.55	<.0001
Saturday*Hour	24	1,134.20	47.26	778.12	<.0001
Sunday*Hour	24	6,648.78	277.03	4,561.39	<.0001

Table C.7 Least-squares estimates of 2014–2016 hourly volume model for rural freeways

Parameter	**Estimate**	**Std. Error**	**t Value**	**Pr. > |t|**
Intercept	3.3037	0.0132	251.14	<.0001
AADT cars (1,000 veh/day)	0.0370	0.0001	590.14	<.0001
AADT trucks (1,000 veh/day)	0.0443	0.0002	271.55	<.0001
Light rain (precipitation < 0.098 in)	-0.0407	0.0019	-21.92	<.0001
Moderate rain (precipitation 0.098 – 0.394 in)	-0.0149	0.0045	-3.35	0.0008
Heavy rain (precipitation 0.394 – 1.969 in)	-0.0282	0.0117	-2.41	0.0161
Violent rain (precipitation > 1.969 in)	0.4235	0.2229	1.90	0.0575
Temperature (F)	-0.0009	0.0001	-16.26	<.0001
Freezing temperature (temperature ≤ 32F)	-0.0105	0.0020	-5.33	<.0001
Average hourly travel speed (mph)	0.0167	0.0002	87.09	<.0001
Downtrend speed indicator	-0.0055	0.0009	-5.98	<.0001
Congested traffic state indicator	0.5276	0.0149	35.46	<.0001
Year = 2014	-0.0369	0.0012	-31.03	<.0001
Year = 2015	-0.0158	0.0011	-14.09	<.0001
Year = 2016	–	–	–	–
January	-0.1317	0.0027	-48.61	<.0001
February	-0.0886	0.0025	-35.50	<.0001
March	0.0560	0.0022	25.12	<.0001
April	0.1141	0.0023	50.05	<.0001
May	0.1615	0.0024	66.01	<.0001
June	0.2227	0.0027	82.71	<.0001
July	0.2283	0.0026	86.45	<.0001
August	0.2047	0.0027	76.62	<.0001
September	0.1386	0.0025	54.40	<.0001
October	0.1351	0.0022	60.17	<.0001
November	0.0702	0.0021	32.76	<.0001
December	–	–	–	–

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
12:00 a.m. – 12:59 a.m.	-0.3264	0.0041	-80.08	<.0001
1:00 a.m. – 1:59 a.m.	-0.5215	0.0041	-128.01	<.0001
2:00 a.m. – 2:59 a.m.	-0.6232	0.0041	-152.88	<.0001
3:00 a.m. – 3:59 a.m.	-0.5670	0.0041	-139.14	<.0001
4:00 a.m. – 4:59 a.m.	-0.3105	0.0041	-76.20	<.0001
5:00 a.m. – 5:59 a.m.	0.1496	0.0041	36.70	<.0001
6:00 a.m. – 6:59 a.m.	0.5584	0.0041	137.05	<.0001
7:00 a.m. – 7:59 a.m.	0.8143	0.0041	199.81	<.0001
8:00 a.m. – 8:59 a.m.	0.9011	0.0041	220.75	<.0001
9:00 a.m. – 9:59 a.m.	0.9820	0.0041	240.11	<.0001
10:00 a.m. – 10:59 a.m.	1.0489	0.0041	255.99	<.0001
11:00 a.m. – 11:59 a.m.	1.0856	0.0041	264.47	<.0001
12:00 p.m. – 12:59 p.m.	1.1095	0.0041	269.63	<.0001
1:00 p.m. – 1:59 p.m.	1.1559	0.0041	280.78	<.0001
2:00 p.m. – 2:59 p.m.	1.2075	0.0041	293.13	<.0001
3:00 p.m. – 3:59 p.m.	1.2656	0.0041	307.29	<.0001
4:00 p.m. – 4:59 p.m.	1.2824	0.0041	312.01	<.0001
5:00 p.m. – 5:59 p.m.	1.2123	0.0041	295.62	<.0001
6:00 p.m. – 6:59 p.m.	1.0042	0.0041	245.52	<.0001
7:00 p.m. – 7:59 p.m.	0.8055	0.0041	197.63	<.0001
8:00 p.m. – 8:59 p.m.	0.6256	0.0041	153.67	<.0001
9:00 p.m. – 9:59 p.m.	0.4730	0.0041	116.30	<.0001
10:00 p.m. – 10:59 p.m.	0.2441	0.0041	59.99	<.0001
11:00 p.m. – 11:59 p.m.	–	–	–	–
Friday	12:00 a.m. – 12:59 a.m.	0.1348	0.0065	20.68	<.0001
Friday	1:00 a.m. – 1:59 a.m.	0.1403	0.0065	21.54	<.0001
Friday	2:00 a.m. – 2:59 a.m.	0.1429	0.0065	21.95	<.0001
Friday	3:00 a.m. – 3:59 a.m.	0.1251	0.0065	19.23	<.0001
Friday	4:00 a.m. – 4:59 a.m.	0.0890	0.0065	13.66	<.0001
Friday	5:00 a.m. – 5:59 a.m.	0.0368	0.0065	5.65	<.0001
Friday	6:00 a.m. – 6:59 a.m.	0.0108	0.0065	1.65	0.0979
Friday	7:00 a.m. – 7:59 a.m.	0.0025	0.0065	0.38	0.7047
Friday	8:00 a.m. – 8:59 a.m.	0.0347	0.0065	5.31	<.0001
Friday	9:00 a.m. – 9:59 a.m.	0.0784	0.0065	12.03	<.0001
Friday	10:00 a.m. – 10:59 a.m.	0.1185	0.0065	18.20	<.0001
Friday	11:00 a.m. – 11:59 a.m.	0.1510	0.0065	23.11	<.0001
Friday	12:00 p.m. – 12:59 p.m.	0.1807	0.0065	27.78	<.0001
Friday	1:00 p.m. – 1:59 p.m.	0.2010	0.0065	30.86	<.0001
Friday	2:00 p.m. – 2:59 p.m.	0.2115	0.0065	32.54	<.0001
Friday	3:00 p.m. – 3:59 p.m.	0.2083	0.0065	32.01	<.0001

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
Friday | 4:00 p.m. – 4:59 p.m.	0.2100	0.0065	32.26	<.0001
Friday | 5:00 p.m. – 5:59 p.m.	0.2439	0.0065	37.45	<.0001
Friday | 6:00 p.m. – 6:59 p.m.	0.3046	0.0065	46.86	<.0001
Friday | 7:00 p.m. – 7:59 p.m.	0.2998	0.0065	46.12	<.0001
Friday | 8:00 p.m. – 8:59 p.m.	0.2662	0.0065	40.94	<.0001
Friday | 9:00 p.m. – 9:59 p.m.	0.2178	0.0065	33.44	<.0001
Friday | 10:00 p.m. – 10:59 p.m.	0.2039	0.0065	31.19	<.0001
Friday | 11:00 p.m. – 11:59 p.m.	0.1769	0.0065	27.12	<.0001
Saturday | 12:00 a.m. – 12:59 a.m.	0.1845	0.0065	28.49	<.0001
Saturday | 1:00 a.m. – 1:59 a.m.	0.1066	0.0065	16.48	<.0001
Saturday | 2:00 a.m. – 2:59 a.m.	0.0049	0.0065	0.76	0.4476
Saturday | 3:00 a.m. – 3:59 a.m.	-0.1057	0.0065	-16.33	<.0001
Saturday | 4:00 a.m. – 4:59 a.m.	-0.2419	0.0065	-37.43	<.0001
Saturday | 5:00 a.m. – 5:59 a.m.	-0.4073	0.0065	-62.99	<.0001
Saturday | 6:00 a.m. – 6:59 a.m.	-0.4846	0.0065	-74.89	<.0001
Saturday | 7:00 a.m. – 7:59 a.m.	-0.3905	0.0065	-60.31	<.0001
Saturday | 8:00 a.m. – 8:59 a.m.	-0.1597	0.0065	-24.66	<.0001
Saturday | 9:00 a.m. – 9:59 a.m.	0.0058	0.0065	0.90	0.3664
Saturday | 10:00 a.m. – 10:59 a.m.	0.1005	0.0065	15.52	<.0001
Saturday | 11:00 a.m. – 11:59 a.m.	0.1200	0.0065	18.53	<.0001
Saturday | 12:00 p.m. – 12:59 p.m.	0.0858	0.0065	13.26	<.0001
Saturday | 1:00 p.m. – 1:59 p.m.	0.0233	0.0065	3.60	0.0003
Saturday | 2:00 p.m. – 2:59 p.m.	-0.0341	0.0065	-5.26	<.0001
Saturday | 3:00 p.m. – 3:59 p.m.	-0.1018	0.0065	-15.68	<.0001
Saturday | 4:00 p.m. – 4:59 p.m.	-0.1439	0.0065	-22.18	<.0001
Saturday | 5:00 p.m. – 5:59 p.m.	-0.1366	0.0065	-21.12	<.0001
Saturday | 6:00 p.m. – 6:59 p.m.	-0.0596	0.0065	-9.19	<.0001
Saturday | 7:00 p.m. – 7:59 p.m.	-0.0176	0.0065	-2.73	0.0064
Saturday | 8:00 p.m. – 8:59 p.m.	0.0120	0.0065	1.85	0.0641
Saturday | 9:00 p.m. – 9:59 p.m.	0.0068	0.0065	1.04	0.2964
Saturday | 10:00 p.m. – 10:59 p.m.	0.0019	0.0065	0.29	0.7723
Saturday | 11:00 p.m. – 11:59 p.m.	-0.0347	0.0065	-5.36	<.0001
Sunday | 12:00 a.m. – 12:59 a.m.	-0.0514	0.0066	-7.79	<.0001
Sunday | 1:00 a.m. – 1:59 a.m.	-0.1821	0.0066	-27.56	<.0001
Sunday | 2:00 a.m. – 2:59 a.m.	-0.3610	0.0067	-54.24	<.0001
Sunday | 3:00 a.m. – 3:59 a.m.	-0.5473	0.0066	-82.39	<.0001
Sunday | 4:00 a.m. – 4:59 a.m.	-0.7933	0.0066	-119.85	<.0001
Sunday | 5:00 a.m. – 5:59 a.m.	-1.0164	0.0066	-153.59	<.0001
Sunday | 6:00 a.m. – 6:59 a.m.	-1.0713	0.0066	-162.30	<.0001
Sunday | 7:00 a.m. – 7:59 a.m.	-0.9310	0.0066	-141.26	<.0001

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
Sunday | 8:00 a.m. – 8:59 a.m.	-0.6057	0.0066	-92.03	<.0001
Sunday | 9:00 a.m. – 9:59 a.m.	-0.3096	0.0066	-46.96	<.0001
Sunday | 10:00 a.m. – 10:59 a.m.	-0.0868	0.0066	-13.18	<.0001
Sunday | 11:00 a.m. – 11:59 a.m.	0.0629	0.0066	9.55	<.0001
Sunday | 12:00 p.m. – 12:59 p.m.	0.1229	0.0066	18.67	<.0001
Sunday | 1:00 p.m. – 1:59 p.m.	0.1160	0.0066	17.64	<.0001
Sunday | 2:00 p.m. – 2:59 p.m.	0.0928	0.0066	14.07	<.0001
Sunday | 3:00 p.m. – 3:59 p.m.	0.0654	0.0066	9.94	<.0001
Sunday | 4:00 p.m. – 4:59 p.m.	0.0543	0.0066	8.27	<.0001
Sunday | 5:00 p.m. – 5:59 p.m.	0.0864	0.0066	13.15	<.0001
Sunday | 6:00 p.m. – 6:59 p.m.	0.1811	0.0066	27.55	<.0001
Sunday | 7:00 p.m. – 7:59 p.m.	0.2153	0.0066	32.68	<.0001
Sunday | 8:00 p.m. – 8:59 p.m.	0.1798	0.0066	27.32	<.0001
Sunday | 9:00 p.m. – 9:59 p.m.	0.0731	0.0066	11.12	<.0001
Sunday | 10:00 p.m. – 10:59 p.m.	-0.0335	0.0066	-5.09	<.0001
Sunday | 11:00 p.m. – 11:59 p.m.	-0.1442	0.0066	-21.96	<.0001

Table C.8 Type III sum of squares summary of 2017–2018 hourly volume model for rural freeways

Source	DF	Type III SS	Mean Square	F Value	Pr. > F
AADT cars (1,000 veh/day)	1	12,095.70	12,095.70	216,434.00	<.0001
AADT trucks (1,000 veh/day)	1	655.62	655.62	11,731.30	<.0001
Light rain (precipitation < 0.098 in)	1	8.27	8.27	147.92	<.0001
Moderate rain (precipitation 0.098 – 0.394 in)	1	1.16	1.16	20.76	<.0001
Freezing temperature (temperature ≤ 32F)	1	44.75	44.75	800.69	<.0001
Average hourly travel speed (mph)	1	83.62	83.62	1,496.17	<.0001
Hourly travel speed range (mph)	1	1.71	1.71	30.65	<.0001
Intermediate traffic state indicator	1	6.08	6.08	108.74	<.0001
Congested traffic state indicator	1	13.08	13.08	233.99	<.0001
Periods with probe density < 5 (Na)	1	275.82	275.82	4,935.41	<.0001
Periods with probe density from 5 to 9 (Nb)	1	662.87	662.87	11,861.00	<.0001
Na*Nb	1	38.12	38.12	682.09	<.0001
Periods with probe density > 9 (Nc)	1	701.16	701.16	12,546.20	<.0001
Nb*Nc	1	10.81	10.81	193.49	<.0001
Year	1	2.97	2.97	53.09	<.0001
Month	11	1,035.81	94.16	1,684.93	<.0001
Hour	23	20,736.14	901.57	16,132.20	<.0001
Friday*Hour	24	922.30	38.43	687.63	<.0001
Saturday*Hour	24	711.55	29.65	530.50	<.0001
Sunday*Hour	24	2,668.86	111.20	1,989.80	<.0001

Table C.9 Least-squares estimates of 2017–2018 hourly volume model for rural freeways

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
Intercept	2.7979	0.0284	98.65	<.0001
AADT cars (1,000 veh/day)	0.0328	0.0001	465.22	<.0001
AADT trucks (1,000 veh/day)	0.0268	0.0002	108.31	<.0001
Light rain (precipitation < 0.098 in)	-0.0275	0.0023	-12.16	<.0001
Moderate rain (precipitation 0.098 – 0.394 in)	-0.0227	0.0050	-4.56	<.0001
Freezing temperature (temperature ≤ 32F)	-0.0482	0.0017	-28.30	<.0001
Average hourly travel speed (mph)	0.0154	0.0004	38.68	<.0001
Hourly travel speed range (mph)	0.0008	0.0001	5.54	<.0001
Intermediate traffic state indicator	0.0656	0.0063	10.43	<.0001
Congested traffic state indicator	0.3577	0.0234	15.30	<.0001
Periods with probe density < 5 (Na)	0.0573	0.0008	70.25	<.0001
Periods with probe density from 5 to 9 (Nb)	0.0869	0.0008	108.91	<.0001
Na*Nb	0.0016	0.0001	26.12	<.0001
Periods with probe density > 9 (Nc)	0.0924	0.0008	112.01	<.0001
Nb*Nc	0.0008	0.0001	13.91	<.0001
Year = 2017	0.0078	0.0011	7.29	<.0001
Year = 2018	–	–	–	–
January	-0.1414	0.0026	-54.15	<.0001
February	-0.0811	0.0027	-30.51	<.0001
March	0.0286	0.0026	11.12	<.0001
April	0.0317	0.0027	11.89	<.0001
May	0.0999	0.0026	37.86	<.0001
June	0.1416	0.0027	52.83	<.0001
July	0.1537	0.0026	59.82	<.0001
August	0.1139	0.0026	44.62	<.0001
September	0.0603	0.0026	23.24	<.0001
October	0.0695	0.0025	27.82	<.0001
November	0.0478	0.0025	19.14	<.0001
December	–	–	–	–
12:00 a.m. – 12:59 a.m.	-0.2635	0.0047	-55.91	<.0001
1:00 a.m. – 1:59 a.m.	-0.4724	0.0047	-100.17	<.0001
2:00 a.m. – 2:59 a.m.	-0.6076	0.0047	-128.64	<.0001
3:00 a.m. – 3:59 a.m.	-0.6251	0.0047	-132.79	<.0001
4:00 a.m. – 4:59 a.m.	-0.4701	0.0047	-100.03	<.0001
5:00 a.m. – 5:59 a.m.	-0.1224	0.0047	-26.03	<.0001
6:00 a.m. – 6:59 a.m.	0.2729	0.0047	57.78	<.0001
7:00 a.m. – 7:59 a.m.	0.5490	0.0048	115.54	<.0001
8:00 a.m. – 8:59 a.m.	0.6608	0.0048	138.13	<.0001
9:00 a.m. – 9:59 a.m.	0.7274	0.0048	150.71	<.0001

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
10:00 a.m. – 10:59 a.m.	0.7925	0.0049	163.13	<.0001
11:00 a.m. – 11:59 a.m.	0.8378	0.0049	171.89	<.0001
12:00 p.m. – 12:59 p.m.	0.8663	0.0049	177.40	<.0001
1:00 p.m. – 1:59 p.m.	0.8994	0.0049	183.85	<.0001
2:00 p.m. – 2:59 p.m.	0.9468	0.0049	193.46	<.0001
3:00 p.m. – 3:59 p.m.	1.0016	0.0049	204.57	<.0001
4:00 p.m. – 4:59 p.m.	1.0345	0.0049	211.53	<.0001
5:00 p.m. – 5:59 p.m.	1.0079	0.0049	207.56	<.0001
6:00 p.m. – 6:59 p.m.	0.8783	0.0048	182.65	<.0001
7:00 p.m. – 7:59 p.m.	0.7038	0.0048	147.71	<.0001
8:00 p.m. – 8:59 p.m.	0.5235	0.0047	110.77	<.0001
9:00 p.m. – 9:59 p.m.	0.3756	0.0047	79.85	<.0001
10:00 p.m. – 10:59 p.m.	0.1972	0.0047	41.99	<.0001
11:00 p.m. – 11:59 p.m.	–	–	–	–
Friday | 12:00 a.m. – 12:59 a.m.	0.1138	0.0075	15.23	<.0001
Friday | 1:00 a.m. – 1:59 a.m.	0.1137	0.0075	15.19	<.0001
Friday | 2:00 a.m. – 2:59 a.m.	0.1208	0.0075	16.15	<.0001
Friday | 3:00 a.m. – 3:59 a.m.	0.1120	0.0075	14.95	<.0001
Friday | 4:00 a.m. – 4:59 a.m.	0.0896	0.0075	12.01	<.0001
Friday | 5:00 a.m. – 5:59 a.m.	0.0496	0.0075	6.64	<.0001
Friday | 6:00 a.m. – 6:59 a.m.	0.0181	0.0075	2.43	0.0152
Friday | 7:00 a.m. – 7:59 a.m.	0.0043	0.0074	0.57	0.5677
Friday | 8:00 a.m. – 8:59 a.m.	0.0314	0.0075	4.21	<.0001
Friday | 9:00 a.m. – 9:59 a.m.	0.0741	0.0075	9.94	<.0001
Friday | 10:00 a.m. – 10:59 a.m.	0.1152	0.0075	15.44	<.0001
Friday | 11:00 a.m. – 11:59 a.m.	0.1513	0.0075	20.30	<.0001
Friday | 12:00 p.m. – 12:59 p.m.	0.1787	0.0075	23.91	<.0001
Friday | 1:00 p.m. – 1:59 p.m.	0.1955	0.0075	26.22	<.0001
Friday | 2:00 p.m. – 2:59 p.m.	0.2150	0.0075	28.86	<.0001
Friday | 3:00 p.m. – 3:59 p.m.	0.2215	0.0075	29.72	<.0001
Friday | 4:00 p.m. – 4:59 p.m.	0.2201	0.0075	29.44	<.0001
Friday | 5:00 p.m. – 5:59 p.m.	0.2504	0.0075	33.54	<.0001
Friday | 6:00 p.m. – 6:59 p.m.	0.3146	0.0075	42.16	<.0001
Friday | 7:00 p.m. – 7:59 p.m.	0.3323	0.0075	44.54	<.0001
Friday | 8:00 p.m. – 8:59 p.m.	0.3218	0.0075	43.07	<.0001
Friday | 9:00 p.m. – 9:59 p.m.	0.2950	0.0075	39.48	<.0001
Friday | 10:00 p.m. – 10:59 p.m.	0.2803	0.0075	37.41	<.0001
Friday | 11:00 p.m. – 11:59 p.m.	0.2553	0.0075	33.99	<.0001
Saturday | 12:00 a.m. – 12:59 a.m.	0.2357	0.0074	31.78	<.0001
Saturday | 1:00 a.m. – 1:59 a.m.	0.1886	0.0074	25.45	<.0001

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
Saturday | 2:00 a.m. – 2:59 a.m.	0.1038	0.0074	13.98	<.0001
Saturday | 3:00 a.m. – 3:59 a.m.	0.0290	0.0074	3.92	<.0001
Saturday | 4:00 a.m. – 4:59 a.m.	-0.0768	0.0074	-10.38	<.0001
Saturday | 5:00 a.m. – 5:59 a.m.	-0.2292	0.0074	-30.99	<.0001
Saturday | 6:00 a.m. – 6:59 a.m.	-0.3471	0.0074	-46.77	<.0001
Saturday | 7:00 a.m. – 7:59 a.m.	-0.3295	0.0074	-44.45	<.0001
Saturday | 8:00 a.m. – 8:59 a.m.	-0.1591	0.0074	-21.49	<.0001
Saturday | 9:00 a.m. – 9:59 a.m.	0.0251	0.0074	3.40	0.0007
Saturday | 10:00 a.m. – 10:59 a.m.	0.1483	0.0074	19.99	<.0001
Saturday | 11:00 a.m. – 11:59 a.m.	0.1932	0.0074	26.06	<.0001
Saturday | 12:00 p.m. – 12:59 p.m.	0.1858	0.0074	25.05	<.0001
Saturday | 1:00 p.m. – 1:59 p.m.	0.1463	0.0074	19.74	<.0001
Saturday | 2:00 p.m. – 2:59 p.m.	0.0966	0.0074	13.02	<.0001
Saturday | 3:00 p.m. – 3:59 p.m.	0.0477	0.0074	6.42	<.0001
Saturday | 4:00 p.m. – 4:59 p.m.	-0.0010	0.0074	-0.13	0.8952
Saturday | 5:00 p.m. – 5:59 p.m.	-0.0103	0.0074	-1.39	0.1651
Saturday | 6:00 p.m. – 6:59 p.m.	0.0477	0.0074	6.41	<.0001
Saturday | 7:00 p.m. – 7:59 p.m.	0.1111	0.0074	14.92	<.0001
Saturday | 8:00 p.m. – 8:59 p.m.	0.1685	0.0075	22.59	<.0001
Saturday | 9:00 p.m. – 9:59 p.m.	0.1980	0.0074	26.62	<.0001
Saturday | 10:00 p.m. – 10:59 p.m.	0.2146	0.0075	28.76	<.0001
Saturday | 11:00 p.m. – 11:59 p.m.	0.1898	0.0075	25.39	<.0001
Sunday | 12:00 a.m. – 12:59 a.m.	0.1633	0.0074	21.99	<.0001
Sunday | 1:00 a.m. – 1:59 a.m.	0.0826	0.0075	11.03	<.0001
Sunday | 2:00 a.m. – 2:59 a.m.	-0.0688	0.0075	-9.18	<.0001
Sunday | 3:00 a.m. – 3:59 a.m.	-0.2070	0.0075	-27.65	<.0001
Sunday | 4:00 a.m. – 4:59 a.m.	-0.4086	0.0075	-54.56	<.0001
Sunday | 5:00 a.m. – 5:59 a.m.	-0.6338	0.0075	-84.93	<.0001
Sunday | 6:00 a.m. – 6:59 a.m.	-0.7660	0.0075	-102.38	<.0001
Sunday | 7:00 a.m. – 7:59 a.m.	-0.7410	0.0075	-99.27	<.0001
Sunday | 8:00 a.m. – 8:59 a.m.	-0.5252	0.0075	-70.42	<.0001
Sunday | 9:00 a.m. – 9:59 a.m.	-0.2438	0.0074	-32.85	<.0001
Sunday | 10:00 a.m. – 10:59 a.m.	-0.0223	0.0074	-3.00	0.0027
Sunday | 11:00 a.m. – 11:59 a.m.	0.1431	0.0074	19.34	<.0001
Sunday | 12:00 p.m. – 12:59 p.m.	0.2229	0.0074	30.08	<.0001
Sunday | 1:00 p.m. – 1:59 p.m.	0.2350	0.0074	31.62	<.0001
Sunday | 2:00 p.m. – 2:59 p.m.	0.2194	0.0074	29.67	<.0001
Sunday | 3:00 p.m. – 3:59 p.m.	0.1889	0.0074	25.52	<.0001
Sunday | 4:00 p.m. – 4:59 p.m.	0.1733	0.0074	23.44	<.0001
Sunday | 5:00 p.m. – 5:59 p.m.	0.1838	0.0074	24.89	<.0001

Parameter	Estimate	Std. Error	t Value	Pr. > |t|
Sunday | 6:00 p.m. – 6:59 p.m.	0.2436	0.0074	32.98	<.0001
Sunday | 7:00 p.m. – 7:59 p.m.	0.2874	0.0074	38.95	<.0001
Sunday | 8:00 p.m. – 8:59 p.m.	0.2900	0.0074	39.41	<.0001
Sunday | 9:00 p.m. – 9:59 p.m.	0.2073	0.0074	28.14	<.0001
Sunday | 10:00 p.m. – 10:59 p.m.	0.1125	0.0074	15.15	<.0001
Sunday | 11:00 p.m. – 11:59 p.m.	-0.0005	0.0074	-0.07	0.9442

APPENDIX D. SAFETY ANALYSIS MODELS

D.1 Rural Freeways

Table D.1 Maximum likelihood estimates of hourly crash probability model

Parameter	Estimate	Std. Error	Chi-Square	Pr. > ChiSq
Intercept	2.0748	0.8416	6.08	0.0140
Hourly traffic volume (1,000 veh/h)	0.2995	0.0959	9.76	0.0020
AADT cars (1,000 veh/day)	0.0101	0.0050	4.02	0.0450
AADT trucks (1,000 veh/day)	0.0570	0.0095	36.05	<.0001
Overpassing road	-0.1622	0.0850	3.65	0.0560
Moderate curve (5.5 – 13.9 degrees)	0.1455	0.0784	3.44	0.0640
Sharp curve (14 degrees or more)	0.2258	0.0736	9.42	0.0020
Segment proportion with median cable barrier	0.1906	0.0890	4.58	0.0320
Segment proportion with median guardrail	-0.6460	0.2398	7.26	0.0070
Segment proportion with roadside guardrail	0.6817	0.0926	54.17	<.0001
Median barrier offset < 30 ft	-0.1228	0.0512	5.75	0.0170
Segment proportion with concrete pavement	0.5180	0.0855	36.69	<.0001
Segment proportion with entering ramp auxiliary lane	0.5501	0.1550	12.60	<.0001
Segment proportion with exiting ramp auxiliary lane	1.4950	0.2238	44.64	<.0001
Speed limit reduced by 5 mph	0.1480	0.0439	11.38	0.0010
Number of point hazards in the median	-0.1534	0.0664	5.33	0.0210
Number of point hazards in the roadside	0.0830	0.0618	1.80	0.1800
Average roadside shoulder width (ft)	0.1224	0.0346	12.48	<.0001
Light rain (precipitation < 0.098 in)	0.2553	0.0902	8.01	0.0050
Freezing temperature (temperature ≤ 32)	0.2460	0.0625	15.47	<.0001
Ice conditions	0.5450	0.1445	14.22	<.0001
Average hourly travel speed (mph)	-0.1402	0.0112	158.05	<.0001
Standard deviation of hourly travel speed (mph)	0.0586	0.0100	34.08	<.0001
Hourly speed trend	-0.2613	0.0366	51.00	<.0001
Downtrend speed indicator (beta < - 5/60)	0.1967	0.0569	11.96	0.0010
Average speed under intermediate traffic (mph)	0.1029	0.0129	63.12	<.0001
Average speed under congested traffic (mph)	0.0938	0.0239	15.41	<.0001
Intermediate traffic	-4.9279	0.7870	39.21	<.0001
Congested traffic	-5.2223	0.8494	37.80	<.0001
Friday	0.1332	0.0689	3.74	0.0530
Sunday	0.1408	0.0651	4.68	0.0310
Year = 2014	-0.5354	0.0638	70.33	<.0001
Year = 2015	-0.1770	0.0688	6.61	0.0100
Year = 2017	0.1159	0.0686	2.86	0.0910
6:00 a.m. – 11:59 a.m.	0.5447	0.0880	38.36	<.0001

Parameter	Estimate	Std. Error	Chi-Square	Pr. > ChiSq
12:00 p.m. – 17:59 p.m.	0.6037	0.1100	30.14	<.0001
18:00 p.m. – 23:59 p.m.	0.4639	0.0856	29.39	<.0001

Table D.2 Maximum likelihood estimates of hourly conditional probability of severe injury crash model

Parameter	Estimate	Std. Error	Chi-Square	Pr. > ChiSq
Intercept	-2.9990	0.2267	174.97	<.0001
Segment proportion with median guardrail	1.3215	0.4725	7.82	0.0050
Mild curve (3.5 – 5.4 degrees)	0.5946	0.3202	3.45	0.0630
Segment proportion with lighting	-1.1238	0.5530	4.13	0.0420
Speed limit reduced by 5 mph	-0.3066	0.1276	5.77	0.0160
Temperature (F)	0.0056	0.0030	3.53	0.0600
Ice conditions	-0.7935	0.3183	6.22	0.0130
Standard deviation of hourly travel speed (mph)	0.1057	0.0176	36.16	<.0001
Downtrend speed indicator (beta < - 5/60)	0.3793	0.1372	7.64	0.0060
Intermediate traffic	-1.4938	0.9153	2.66	0.1030
Average speed under intermediate traffic (mph)	0.0312	0.0172	3.28	0.0700
Friday	-0.2896	0.1900	2.32	0.1280

Table D.3 Average marginal effects of hourly crash probability model

Parameter	Estimate	Std. Error	95% Confidence Interval	
Hourly traffic volume (1,000 veh/h)	0.82	0.26	0.31	1.34
AADT cars (1,000 veh/day)	0.03	0.01	0.00	0.05
AADT trucks (1,000 veh/day)	0.16	0.03	0.11	0.21
Overpassing road	-0.45	0.23	-0.90	0.01
Moderate curve (5.5 – 13.9 degrees)	0.40	0.22	-0.02	0.82
Sharp curve (14 degrees or more)	0.62	0.20	0.22	1.02
Segment proportion with median cable barrier	0.52	0.24	0.04	1.00
Segment proportion with median guardrail	-1.78	0.66	-3.07	-0.48
Segment proportion with roadside guardrail	1.87	0.26	1.37	2.38
Median barrier offset < 30 ft	-0.34	0.14	-0.61	-0.06
Segment proportion with concrete pavement	1.42	0.24	0.96	1.89
Segment proportion with entering ramp auxiliary lane	1.51	0.43	0.68	2.35
Segment proportion with exiting ramp auxiliary lane	4.11	0.62	2.90	5.32
Speed limit reduced by 5 mph	0.81	0.24	0.34	1.29
Number of point hazards in the median	-0.42	0.18	-0.78	-0.06
Number of point hazards in the roadside	0.23	0.17	-0.11	0.56
Average roadside shoulder width (ft)	0.34	0.10	0.15	0.52
Light rain (precipitation < 0.098 in)	1.11	0.20	0.72	1.51

Freezing temperature (temperature ≤ 32°F)	0.88	0.16	0.56	1.19
Average hourly travel speed (mph)	-0.33	0.03	-0.38	-0.28
Standard deviation of hourly travel speed (mph)	0.16	0.03	0.11	0.22
Hourly speed trend	-0.72	0.10	-0.92	-0.52
Downtrend speed indicator (beta < - 5/60)	0.54	0.16	0.23	0.85
Intermediate traffic	3.61	0.30	3.04	4.19
Congested traffic	1.30	2.36	-3.32	5.92
Friday	0.37	0.19	-0.01	0.74
Sunday	0.39	0.18	0.04	0.74
Year = 2014	-1.47	0.18	-1.82	-1.13
Year = 2015	-0.49	0.19	-0.86	-0.12
Year = 2017	0.32	0.19	-0.05	0.69
6:00 a.m. – 11:59 a.m.	1.50	0.24	1.02	1.97
12:00 p.m. – 17:59 p.m.	1.66	0.30	1.07	2.26
18:00 p.m. – 23:59 p.m.	1.28	0.24	0.81	1.74
All values in %				

Table D.3 Average marginal effects of hourly conditional probability of severe injury crash model

Parameter	Estimate	Std. Error	95% Confidence Interval	
Segment proportion with median guardrail	15.12	5.40	4.55	25.70
Mild curve (3.5 – 5.4 degrees)	6.81	3.66	-0.37	13.98
Segment proportion with lighting	-12.87	6.33	-25.27	-0.46
Speed limit reduced by 5 mph	-7.02	2.92	-12.74	-1.30
Temperature (F)	0.06	0.03	0.00	0.13
Light rain (precipitation < 0.098 in)	-2.48	0.97	-4.37	-0.59
Freezing temperature (temperature ≤ 32°F)	-1.19	0.39	-1.96	-0.41
Average hourly travel speed (mph)	0.11	0.06	-0.01	0.22
Standard deviation of hourly travel speed (mph)	1.21	0.20	0.82	1.60
Downtrend speed indicator (beta < - 5/60)	4.34	1.57	1.26	7.42
Intermediate traffic	3.38	2.30	-1.14	7.89
Friday	-3.31	2.17	-7.58	0.95
All values in %				

D.2 Signalized Intersections

Table D.4 Maximum likelihood estimates of hourly crash probability model (intersection Approach Crashes)

Variables	Estimate	Std. Error	P-value
Intercept	-5.0557	0.1883	<.0001
Hourly Volume (1,000 veh)	0.3527	0.0238	<.0001
Standard Deviance of Speed	0.0516	0.0168	0.0022
Temperature between 23°F–32°F	0.322	0.1628	0.048
Temperature between 32°F–41°F	0.3137	0.1476	0.0335
Delay in Second	0.0126	0.00244	<.0001
Exclusive left turn lane without left turn arrow (ref = no exclusive left turn lane)	0.1575	0.1779	0.3759
Exclusive left turn lane with left turn arrow(ref = no exclusive left turn lane)	0.2817	0.1338	0.0353
Urban SpdLimit = 20, 25 (ref = Urban Spdlimit = 30, 35)	0.1685	0.2255	0.4549
Urban SpdLimit = 40, 45 (ref = Urban Spdlimit = 30, 35)	-0.2037	0.1957	0.2977
Urban SpdLimit ≥ 50 (ref = Urban Spdlimit = 30, 35)	0.4587	0.263	0.0812
Rural SpdLimit = 20, 25 (ref = Urban Spdlimit = 30, 35)	-1.2128	0.9012	0.1784
Rural SpdLimit = 30, 35 (ref = Urban Spdlimit = 30, 35)	0.5155	0.1912	0.007
Rural SpdLimit = 40, 45 (ref = Urban Spdlimit = 30, 35)	0.2154	0.161	0.1811
Rural SpdLimit ≥ 50 (ref = Urban Spdlimit = 30, 35)	-0.0181	0.1995	0.9276
Workday 06:00 a.m. – 09:00 a.m. (ref = Workday 00:00 a.m. – 06:00 a.m.)	0.6567	0.1237	<.0001
Workday 09:00 a.m. – 12:00 a.m. (ref = Workday 00:00 a.m. – 06:00 a.m.)	0.1256	0.1341	0.3491
Workday 12:00 a.m. – 15:00 p.m. (ref = Workday 00:00 a.m. – 06:00 a.m.)	0.1462	0.1219	0.2304
Workday 15:00 p.m. – 18:00 p.m. (ref = Workday 00:00 a.m. – 06:00 a.m.)	0.0384	0.1267	0.7618
Workday 18:00 p.m. – 00:00 a.m. (ref = Workday 00:00 a.m. – 06:00 a.m.)	-0.4842	0.1595	0.0024

Weekend 06:00 a.m. – 09:00 a.m. (ref = Weekend 00:00 a.m. – 06:00 a.m.)	0.1017	0.3148	0.7467
Weekend 09:00 a.m. – 12:00 a.m. (ref = Weekend 00:00 a.m. – 06:00 a.m.)	0.7097	0.2077	0.0006
Weekend 12:00 a.m. – 15:00 p.m. (ref = Weekend 00:00 a.m. – 06:00 a.m.)	-0.0874	0.2579	0.7347
Weekend 15:00 p.m. – 18:00 p.m. (ref = Weekend 00:00 a.m. – 06:00 a.m.)	0.3083	0.2215	0.1639
Weekend 18:00 p.m. – 00:00 a.m. (ref = Weekend 00:00 a.m. – 06:00 a.m.)	0.1118	0.2245	0.6184

Table D.5 Maximum likelihood estimates of hourly conditional probability of severe injury crash model (intersection Approach Crashes)

Variables	Estimate	Std. Error	P-value
Intercept	-2.134	0.2577	<.0001
Temperature between 32°F –41°F	0.8617	0.3187	0.0069
1,Precipitation higher than 1; 0, else	1.3179	0.5181	0.011
Standard Deviance of Speed	0.0559	0.0392	0.1537

Table D.6 Maximum likelihood estimates of hourly crash probability model (intersection Inside-intersection Crashes)

Variables	Estimate	Std. Error	P-value
Intercept	-3.8095	0.1512	<.0001
Major road: Hourly Volume (1,000 veh)	0.0804	0.0436	0.0649
Crossing road: Hourly Volume (1,000 veh)	0.1546	0.0426	0.0003
Flushed center separation (ref = no median)	-0.4515	0.157	0.004
Raised center separation (ref = no median)	0.4206	0.1371	0.0022
Depressed center separation (ref = no median)	0.2433	0.181	0.1789
Crossing road: Urban SpdLimit = 20, 25 (ref = Urban SpdLimit = 30, 35)	0.703	0.2036	0.0006
Crossing road: Urban SpdLimit = 40, 45 (ref = Urban SpdLimit = 30, 35)	0.3876	0.2108	0.0659
Crossing road: Urban SpdLimit ≥ 50 (ref = Urban SpdLimit = 30, 35)	-0.3724	0.4146	0.369
Crossing road: Rural SpdLimit = 20, 25 (ref = Urban SpdLimit = 30, 35)	-0.7907	0.2544	0.0019
Crossing road: Rural SpdLimit = 30, 35 (ref = Urban SpdLimit = 30, 35)	0.0669	0.1369	0.6249
Crossing road: Rural SpdLimit = 40, 45 (ref = Urban SpdLimit = 30, 35)	0.0717	0.1805	0.691
Crossing road: Rural SpdLimit ≥ 50 (ref = Urban SpdLimit = 30, 35)	0.2196	0.1744	0.2079
Workday 09:00 a.m. – 12:00 a.m. (ref = Workday 00:00 a.m. – 06:00 a.m.)	0.4677	0.1554	0.0026
Workday 12:00 a.m. – 15:00 p.m. (ref = Workday 00:00 a.m. – 06:00 a.m.)	0.3216	0.1638	0.0496
Workday 15:00 p.m. – 18:00 p.m. (ref = Workday 00:00 a.m. – 06:00 a.m.)	0.3722	0.1615	0.0212
Workday 18:00 p.m. – 00:00 a.m. (ref = Workday 00:00 a.m. – 06:00 a.m.)	-0.7293	0.2291	0.0015
Weekend 06:00 a.m. – 09:00 a.m. (ref = Weekend 00:00 a.m. – 06:00 a.m.)	0.0755	0.3465	0.8274
Weekend 09:00 a.m. – 12:00 a.m. (ref = Weekend 00:00 a.m. – 06:00 a.m.)	0.1781	0.2912	0.5408
Weekend 12:00 a.m. – 15:00 p.m. (ref = Weekend 00:00 a.m. – 06:00 a.m.)	0.2604	0.2926	0.3734
Weekend 15:00 p.m. – 18:00 p.m. (ref = Weekend 00:00 a.m. – 06:00 a.m.)	-0.0565	0.3307	0.8644
Weekend 18:00 p.m. – 00:00 a.m. (ref = Weekend 00:00 a.m. – 06:00 a.m.)	0.1468	0.262	0.5753

Table D.7 Maximum likelihood estimates of hourly conditional probability of severe injury crash model (intersection Inside-intersection Crashes)

Variables	**Estimate**	**Std. Error**	**P-value**
Intercept	-3.4292	0.9209	0.0002
Major Road: mean speed	0.0341	0.018	0.0588
Major Road: Number of Lane	0.6922	0.2292	0.0025
Crossing Road: Number of Lane	-0.5838	0.2149	0.0066

About the Joint Transportation Research Program (JTRP)

On March 11, 1937, the Indiana Legislature passed an act which authorized the Indiana State Highway Commission to cooperate with and assist Purdue University in developing the best methods of improving and maintaining the highways of the state and the respective counties thereof. That collaborative effort was called the Joint Highway Research Project (JHRP). In 1997 the collaborative venture was renamed as the Joint Transportation Research Program (JTRP) to reflect the state and national efforts to integrate the management and operation of various transportation modes.

The first studies of JHRP were concerned with Test Road No. 1—evaluation of the weathering characteristics of stabilized materials. After World War II, the JHRP program grew substantially and was regularly producing technical reports. Over 1,600 technical reports are now available, published as part of the JHRP and subsequently JTRP collaborative venture between Purdue University and what is now the Indiana Department of Transportation.

Free online access to all reports is provided through a unique collaboration between JTRP and Purdue Libraries. These are available at http://docs.lib.purdue.edu/jtrp.

Further information about JTRP and its current research program is available at http://www.purdue.edu/jtrp.

About This Report

An open access version of this publication is available online. See the URL in the citation below.

Tarko, A. P., Guo, Q., & Pineda-Mendez, R. (2021). *Using emerging and extraordinary data sources to improve traffic safety* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2021/04). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284317283