
Exploring Large-Scale Crowdsourced Connected Vehicle Data (CVD) to Support Alabama Transportation Decision-Making

A Final Report to

The Alabama Department of Transportation
1409 Coliseum Boulevard
Montgomery, AL 36110

by

Praveena Penmetsa, Ph.D.
Zihe Zhang, Ph.D.
Vamshi Chaitanaya Ananimalla, Ph.D.
Alex Hainen, Ph.D.
Jun Liu, Ph.D.

The University of Alabama
2013 SCIB Building, Tuscaloosa, AL 35487

August 3, 2026

DISCLAIMER STATEMENT

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of Alabama, Alabama Department of Transportation or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation. Comments contained in this paper related to specific testing equipment and materials are not considered an endorsement of any commercial products and/or services; no such endorsement is intended or implied.

ACKNOWLEDGEMENTS

The research team gratefully acknowledges ALDOT's support, and the guidance provided throughout the project. The contents of this report reflect the views of the authors and do not necessarily represent the official views or policies of ALDOT.

This project was initiated under the leadership of Dr. Jun Liu, Associate Professor of Civil, Construction and Environmental Engineering at The University of Alabama, who served as Principal Investigator. Dr. Liu passed away during the course of this project. His foundational contributions to the research design, analytical framework, and case study development are reflected throughout this report. The research team dedicates this work to his memory and is grateful for his scholarly leadership and commitment to advancing transportation research in Alabama. The project was completed under the direction of Dr. Praveena Penmetsa as Principal Investigator, with co-investigator support from Dr. Alex Hainen and Dr. Jeremy Pate of the Alabama Transportation Institute (ATI) at The University of Alabama. The team also thanks Dr. Steven Jones, Director of the Alabama Transportation Institute, for his institutional support and guidance. Dr. Zihe Zhang, Dr. Emmanuel Kofi Adanu, Dr. Abhay Lidbe, and Dr. Vamshi Chaitanaya Annimalla contributed as collaborators on the project.

The team also recognizes the important contributions of the students: Xiao Zou, Yang Yu, Sunday Okafor, Jiayi Kong, Yanfang Su and Ningzhe Xu.

ABSTRACT

Connected vehicle data (CVD) provides high-resolution, continuous observations of vehicle trajectories, speeds, and driving events, offering significant advantages over traditional data sources in terms of spatial accuracy, temporal granularity, and behavioral insights. This project explores the potential of large-scale CVD to support data-driven transportation decision-making in Alabama in transportation operation, transportation safety, and transportation planning. The primary objective of this study is to develop a comprehensive analytical framework and practical applications that demonstrate how CVD can be integrated with existing transportation datasets, such as CARE crash data, ALGO incident data, HPMS roadway inventory data, and points of interest (POI) data, to enhance decision-making in transportation planning, operations, and safety. The project establishes scalable methods for processing, aggregating, analyzing, and visualizing large-scale trajectory data, while also addressing key challenges related to data integration and usability for agency applications.

This study presents seven major applications of CVD to support transportation decision-making; each illustrated with corresponding case studies and associated methodologies. First, in traffic signal evaluation and optimization, a case study on intersection delay analysis compares performance across signalized and unsignalized intersections, as well as peak and non-peak periods, providing insights for improving intersection efficiency. Second, in corridor and network-level mobility analysis, multiple case studies examine corridor speed and delay patterns, traffic disruptions under extreme weather, event-related traffic dynamics, and freeway congestion characteristics using trajectory data, illustrating the capability of CVD to capture spatiotemporal mobility patterns. Third, in traffic safety analysis, this study presents applications including driving volatility-based safety assessment, correlation analysis between hard braking events and crashes, and proactive safety screening using behavioral indicators, demonstrating the value of CVD in identifying high-risk locations and supporting preventive safety strategies. Fourth, in travel demand and origin-destination analysis, a case study illustrates how CVD can be used to explore trip patterns and visualize demand distributions to inform planning decisions. Fifth, in transportation accessibility analysis, case studies evaluate access to essential services and healthcare facilities, revealing spatial disparities and delays, and supporting equity-oriented planning. Finally, in the context of connected and automated vehicles, the report outlines a conceptual framework and discusses behavior-informed decision support and vehicle-to-infrastructure integration, emphasizing the role of CVD in enabling future intelligent transportation systems.

This project offers both a methodological framework and practical guidance for integrating CVD into transportation decision-making processes. The findings underscore the value of leveraging CVD to support data-driven decisions in transportation operations, safety, and planning. The proposed framework and accompanying case studies provide actionable insights to facilitate the adoption of CVD by transportation agencies and practitioners in Alabama and beyond.

Key Words:

Connected Vehicle Data, Traffic Operations, Transportation Safety, Travel Demand Analysis, Transportation Accessibility, Data Fuse, Transportation Decision-Making

Table of Contents

DISCLAIMER STATEMENT	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
List of Figures	ix
List of Tables	xii
1. Introduction.....	1
2. Review of Related Work.....	2
2.1 Transportation Planning or Travel Demand Analysis	2
2.2 Traffic Operations	5
2.3 Connected/Automated Vehicles.....	6
2.4 Transportation Safety	6
2.5 Transportation Accessibility and Environment	7
2.6 Summary	9
3. Project Framework and Objectives	10
3.1 Project Scope and Overall Framework	10
3.2 Project Main Objectives.....	11
4. Data	12
4.1 Data Sources	12
4.1.1 Wejo Data	12
4.1.2 Other Major Data Sources	13
4.2 Data Processing.....	14
5. Applications in Traffic Signal Evaluation and Optimization	16
5.1 Case Study 1 – Intersections Delaying Analysis	16
5.1.1 Study area.....	16
5.1.2 Methodology	16
5.1.3 Results and Analysis	16
6. Applications in Corridor or Networkwide Mobility Analysis	19
6.1 Case Study 1 – Corridor Speed and Delay Analysis in Tuscaloosa	19
6.1.1 Study Corridors and Data Overview	19
6.1.2 Data Preparation and Corridor Trip Extraction	20
6.1.3 Performance Measures and Baseline Construction.....	22

6.1.4 Baseline Corridor Traffic Patterns	25
6.1.5 Key Findings	30
6.2 Case Study 2 – Traffic Disruption under Extreme Weather	30
6.2.1 Event Selection and Comparison Strategy	30
6.2.2 Weather-Related Traffic Disruptions	30
6.2.3 Appendix	31
6.2.4 Key Findings	36
6.3 Case Study 3 – Corridors Traffic Pattern	37
6.3.1 Event Selection and Comparison Strategy	37
6.3.2 Corridor-Level Traffic Responses during Selected Game-Day Events	37
6.3.3 Spatial Evidence from Point-Speed and Segment-Level Mapping	40
6.3.4 Key Findings	41
6.4 Case Study 4 – Freeway Congestion Analysis Using DCVP Data	42
6.4.1 Empirical Setting	42
6.4.2 Freeway DCVP Data	60
6.4.3 Trajectory Alignment Using K-D Tree Algorithm	60
6.4.4 Traffic Characteristics Distribution	62
6.4.5 Clustering Mechanism for Congestion Identification	62
6.4.6 Shockwave Speed and Queue Length Analysis	64
6.4.7 Driver Behavior Analysis During Congestion	64
6.4.8 Route Alternatives and Behavioral Patterns	65
6.4.9 Comparative Analysis of Primary and Alternate Routes	66
6.4.10 Congestion Scoring Framework	67
6.4.11 Congestion Score Patterns and Unreported Disruptions	68
6.4.12 Discussion	70
7. Applications in Traffic Safety Analysis	71
7.1 Case Study 1 – Driving Volatility-based Proactive Safety Management	71
7.1.1 Study Scope	71
7.1.2 Methodology	71
7.1.3 Results	76
7.1.4 Summary and Conclusions	86
7.2 Case Study 2 – Statewide Correlation Analysis of Hard Braking Events and Road Traffic Crashes	86

7.2.1 Study Scope	86
7.2.2 Methodology	86
7.2.3 Results.....	87
7.2.4 Summary and Conclusions	88
7.3 Case Study 3 – Integrating Hard Braking Event Data for Proactive Road Network Safety Screening.....	88
7.3.1 Study Scope	88
7.3.2 Methodology	88
7.3.3 Results.....	88
7.3.4 Summary and Conclusions	90
8. Applications in Travel Demand or Origin-Destination Analysis	91
8.1 Case Study 1 – Trip Origin and Destination Analysis.....	91
8.1.1 Study Scope	91
8.1.2 Methodology	91
8.1.3 Travel Demand Visualization and Summary	92
9. Applications in Transportation Accessibility Analysis.....	94
9.1 Case Study 1 – Accessibility to Essential Services	94
9.1.1 Study Scope	94
9.1.2 Methodology	94
9.1.3 Results.....	95
9.2 Case Study 2 – Hospital Accessibility Analysis Using DCVP Data	96
9.2.1 Empirical Setting	96
9.2.2 Data and Method.....	98
9.2.3 Identifying and Measuring Travel Delays	98
9.2.4 Route Disparities Between ORS and DCVP Data.....	100
9.2.5 Observations from Delay Classified Trips.....	103
9.2.6 Spatial Analysis of Delay Categories	104
9.2.7 Identified Key Regions for Accessibility Improvements.....	109
10. Applications in Connected/Automated Vehicles	100
10.1 Conceptual Framework.....	100
10.2 Behavior-Informed CAV Decision Support	101
10.3 V2I Integration at Intersections	101
11. Summary, Implication, and Recommendations	102

11.1 Project Summary.....	102
11.2 Implications.....	103
11.3 Recommendations and Future Work	104
References.....	107

List of Figures

Figure 3.1 Overall Methodology Framework	10
Figure 5.1 The distribution of vehicles delaying on signalized intersections versus delaying on unsignalized intersections	17
Figure 5.2 Delaying distribution during peak hours vs non-peak hours.....	17
Figure 5.3 Spatial distribution of signalized and unsignalized intersections and their associated delays	18
Figure 6.1 Study Corridors in Tuscaloosa	20
Figure 6.2 Daily Trip Counts in March 2021.	20
Figure 6.3 Corridor Trip Extraction and Filtering Workflow.....	23
Figure 6.4 Hourly Traffic Performance on McFarland Blvd E	26
Figure 6.5 Hourly Traffic Performance on 15th Street.....	27
Figure 6.6 Daily Traffic Variation in March 2021	29
Figure 6.7 Hourly Traffic Performance on McFarland Blvd E on March 17, 2021	34
Figure 6.8 Hourly Traffic Performance on McFarland Blvd E on March 25, 2021	35
Figure 6.9 Corridor-Level Traffic Performance on 15th Street on March 23 relative to the March 16 baseline	39
Figure 6.10 Point-Speed Maps on 15th Street during March 19 and March 26 Game-Day Comparisons	41
Figure 6.11 Segment-Level Average Speed on 15th Street from 14:00 to 17:00 on March 23 ...	41
Figure 6.12 I-20/59 Corridor Study Area, spanning from Exit 68 to Exit 104.....	60
Figure 6.13 (a). Alignment of DCVP data journeys using the K-D Tree algorithm: Journey which traveled the full length of the study corridor (left) and journey that traveled only part of the corridor (right). (b). Assigning the initial position to a jour	61
Figure 6.14 Observed distribution of vehicle travel distances along the corridor.	62
Figure 6.15 Vehicle speed distribution in the corridor.	62
Figure 6.16 Congestion Clustering and Dynamics: (a) DBSCAN cluster identification, (b) LOESS regression through inflection points, (c) Accumulating and dissipating shockwave, and (d) Queue length over time.	63
Figure 6.17 (a). Hard braking events during a congestion event. (b). Hard acceleration events during a congestion event. (c). Unlatched seatbelt events during a congestion event. (d). Latched seatbelt events during a congestion event.	64
Figure 6.18 Route alternatives between Exit 79 and Exit 97 on the I-20/59 study corridor.	65
Figure 6.19 Vehicles taking the Exit 79 alternate route.....	65

Figure 6.20 Comparison of traffic metrics on the I-20 route (left column) and alternate route (right column) during 3 different incidents.....	66
Figure 6.21 (a) 3D Time-space diagram. (b) Volume between Threshold 1 to 2. (c) Volume ...	67
Figure 6.22 Volumes and weighted score.....	68
Figure 6.23 Time-series of congestion scores over the study period.....	69
Figure 6.24 Frequency distribution of congestion scores.	69
Figure 6.25 Congestion events visible in DCVP data but unreported in ATMS incident records.	70
Figure 7.1 Data Integration Framework.....	72
Figure 7.2 Procedure for calculating segment-level and intersection-level volatility	73
Figure 7.3 Extends of accelerations and decelerations across speeds	76
Figure 7.4 Location-based driving volatility at intersection level	77
Figure 7.5 Location-based driving volatility at segment level	77
Figure 8.1 Data Processing Framework.....	92
Figure 8.2 Within-TAZ Trip Counts in Tuscaloosa County.....	92
Figure 8.3 Inter-TAZ Trip Counts in Tuscaloosa County	93
Figure 8.4 Distribution of Average Travel Time and Distance Between TAZ Pairs	93
Figure 9.1 Methodology Framework	94
Figure 9.2 Trip Counts to Individual POIs by Category.....	95
Figure 9.3 Average Travel Time and Distance by Destination Category.....	96
Figure 9.4 Study hospitals in Tuscaloosa and Northport, Alabama.	97
Figure 9.5 Isochrone map highlighting areas reachable under 10 minutes, 30 minutes, and 60 minutes to study hospitals.	97
Figure 9.6 Scatter plot of trips showing variations in actual travel time and distance from ORS estimates.....	98
Figure 9.7 (a). Percentage of trips within different journey time intervals. (b). Frequency of trips by distance. (c). Number of trips over a 24-hour period.....	99
Figure 9.8 (a). Actual trip trajectory obtained from DCVP data (DB4IoT). (b). Estimated travel time and distance from ORS.	101
Figure 9.9 Region-wise distribution of actual vs. estimated travel times for trips to and from the hospital.....	102
Figure 9.10 (a). Trips after classification. (b). Comparison of travel time gaps by type of delay.	103
Figure 9.11 (a). Spatial distribution of stopped delays from hospitals. (b). Spatial distribution of stopped delays to hospitals.....	104

Figure 9.12 (a). Spatial distribution of excess travel time caused by intermediate stops from hospitals. (b). Spatial distribution of excess travel time caused by intermediate stops to hospitals.	1
Figure 9.13 (a). Spatial distribution of excess travel time due to route deviations from hospitals. (b). Spatial distribution of excess travel time due to route deviations to hospitals.	1
Figure 9.14 (a). Places with highest travel time deviation due to route deviation (to hospital). (b). Places with highest stopped delay (to hospital). (c). Places with most stops (to hospital).	109
Figure 9.15 (a). Route trajectories from Knoxville, Alabama, showing multiple alternate paths taken by drivers to access the hospital. (b). Route trajectories from West Blocton, Alabama, illustrating alternative hospital access routes.	98
Figure 10.1 Conceptual Framework	100

List of Tables

Table 2.1 Selected studies related to application of trajectory data on transportation planning. ...	3
Table 2.2 Selected studies related to application of trajectory data on traffic operation.....	4
Table 2.3 Selected studies related to application of trajectory data on connected/automated vehicles.	6
Table 2.4 Selected studies related to application of trajectory data on Traffic Safety	7
Table 2.5 Selected studies related to application of trajectory data on transportation accessibility and environment.....	8
Table 4.1 Sample of processed DCVP data (captured at 1--3 second intervals) illustrating the structure and temporal resolution of trajectory points used for congestion analysis.	12
Table 5.1 Peak hour and non-peak hour distribution.....	18
Table 6.1 Study Corridors and Buffer Settings.....	21
Table 6.2 Minimum and Maximum Allowable Travel Times.....	22
Table 6.3 Definitions of Performance Measures and Aggregation Levels.....	24
Table 6.4 Weather Conditions on March 17, 2021	31
Table 6.5 Weather Conditions on March 25, 2021	32
Table 6.6 Summary of Corridor-Level Traffic Responses during Selected Game-Day Events...	38
Table 6.7 Sample of processed DCVP data (captured at 1 to 3 second intervals) illustrating the structure and temporal resolution of trajectory points used for congestion analysis.	60
Table 7.1 Descriptive statistics of variables at the intersection level	78
Table 7.2 Descriptive statistics of variables at segment level	79
Table 7.3 Model Results at Intersection Level	81
Table 7.4 Model Results at Segment Level	83
Table 7.5 Summary Statistics of HBE and Crashes by Functional Class.....	87
Table 7.6 Spearman and Pearson Correlation Coefficients by Functional Class	87
Table 7.7 Spearman Correlation — HBE vs. Crash Types at Segment Level (Tuscaloosa County Arterials)	89
Table 7.8 Number of Hotspot Segments by Confidence Level	89
Table 7.9 Proportion of Crash-Based Hotspots Also Identified by HBE	89
Table 7.10 Total Rank Difference (TRD) of Common Hotspots — Lower = Better Agreement	89
Table 9.1 Number of trips to and from hospitals across the study period.	98
Table 9.2 Sample DCVP-ORS records used for analysis, illustrating differences between estimated (ORS) and actual (DCVP) travel metrics for selected trips.	98

1. Introduction

Advancements in information technologies have led to the rapid proliferation of data generated by mobile devices and electronic sensors, creating unprecedented opportunities for transportation researchers and practitioners to examine transportation systems at a much finer spatial and temporal resolution. Among these emerging data sources, connected vehicles equipped with satellite navigation systems provide crowdsourced, high-resolution vehicle trajectory and driving event data across real-world roadway networks. This data offers significant potential for transportation agencies to better understand system performance and support data-driven decision-making.

Large-scale crowdsourced connected vehicle data (CVD) enables a wide range of applications, including analyzing regional travel demand patterns, identifying congestion hotspots, evaluating location-specific safety performance, and assessing accessibility and environmental impacts. Compared to traditional data sources, CVD provides continuous, high-frequency observations of vehicle movements and driving behaviors, allowing for more comprehensive and dynamic insights into transportation systems. Despite their potential, large-scale CVD have not yet been widely adopted by transportation agencies. One of the key barriers is the lack of practical guidance and documented applications demonstrating how such data can be effectively utilized for planning, operations, and policymaking. As transportation agencies anticipate broader access to these data in the near future, there is a critical need to better understand their characteristics, limitations, and integration with existing data systems, as well as to identify agency-specific needs and priorities.

To address this gap, this project focuses on exploring CVD covering roadways across Alabama and developing practical applications to support transportation decision-making. The CVD leveraged in this study comprised of vehicle movement and driving event data, covering more than 95% of roadways in Alabama. Compared to traditional mobile-based sensing data, connected vehicle data offer several key advantages. First, they provide higher spatial accuracy due to the integration of onboard GPS and cellular antennas. Second, they include richer data attributes beyond location, such as vehicle speed and vehicle events (hard-braking events, etc.). Third, they are reported at a high and consistent frequency, typically every few seconds, enabling detailed temporal analysis of traffic conditions and driving behavior.

The primary objective of this project is to evaluate the characteristics and potential applications of large-scale CVD in Alabama, identify data needs from local transportation agencies, and develop methodologies and use cases that demonstrate how these data can support decision-making. Specifically, this project aims to: (1) conduct a preliminary analysis of connected vehicle data across Alabama roadways; (2) develop methods for processing, aggregating, analyzing, and visualizing CVD, as well as integrating these data with existing transportation datasets such as CARE crash data, ALGO incident data, HPMS roadway inventory data, and points of interest (POI) data; (3) develop applications that leverage CVD to inform transportation planning, traffic operations, connected and automated vehicle strategies, highway safety, accessibility, and environmental analysis; and (4) present a series of case studies demonstrating the use of CVD in major traffic event management and local transportation planning and engineering practice.

Through these efforts, this project seeks to provide actionable guidance and practical examples to facilitate the adoption of CVD by transportation agencies, ultimately supporting more efficient, resilient, and data-driven transportation systems in Alabama and beyond.

2. Review of Related Work

In recent times, thanks to the development of various data collection and storage technology, numerous detailed large-scale trajectory datasets have become available. These datasets include Global Positioning System (GPS) traces sourced from cellular phones and automotive vehicles, Call Detail Records (CDR) in an anonymized format from cellular service providers, vehicle trajectory data taken from drone in video data formation. These trajectory data contain huge spatial-temporal information about the travelers and vehicles in the transportation system and provide a great opportunity for transportation researchers and practitioners to harness these resources for enhanced analysis methodologies and to address crucial queries to assist transportation management and decision-making. Since different datasets contain different information and are associated with specific data format, transportation researchers and practitioners may think of using different datasets to achieve their different management goals. The scope of this project is to explore large-scale Crowdsourced Connected Vehicle Data (CVD) to Support Alabama Transportation Decision-Making. Specifically, CVD from Wejo is used to explore the potential applications. Wejo's CVD provides both vehicle movement data (i.e., vehicle trajectory data) and vehicle movement data. Before conducting Wejo's CVD data application, it is essential to conduct a synthetic literature review to review current research and applications using different types of trajectory data in transportation management to gain a comprehensive understanding of the function of such type of data.

The utilization of trajectory data within road transportation is amalgamated into five distinct domains, each elaborated upon in individual subsections in the subsequent segments of this report. It is worth noting that the following review does not seek to provide an exhaustive overview of the literature, but highlight some of the relevant work and their limitation to illustrate applications of trajectory data in different areas of transportation engineering which is expecting to shed lights to the application of Wejo CVD data to conduct case studies and assist ALDOT transportation management.

2.1 Transportation Planning or Travel Demand Analysis

Unraveling human travel behaviors by large-scale trajectory data provide valuable information for transportation planning or travel demand analysis. A profound comprehension of historical and real-time travel patterns facilitates the formulation of pertinent transportation policies and the delineation of efficacious traffic management strategies, such as contingency evacuation plans during severe weather conditions. Conventionally, the primary approach to amass data on human travel behavior has been through travel surveys. This methodology, although exhaustive in terms of extracting trip-related details like purpose and mode of transport, often incurs significant costs and typically results in a constrained sample size. In contrast, large-scale trajectory data, enriched with spatial-temporal nuances of travel movements, offers a promising repository to decode the intricacies of both vehicular and pedestrian movement. Table 2.1 encapsulates selected studies emphasizing the application of trajectory data in transportation planning. These studies harness diverse sources of large-scale trajectory data including GPS loggers and GPS-enabled mobile phones (Dabiri et al., 2020), Call Detail Records (CDRs) from cellular phones (Zhao et al., 2022), GPS data from taxis (Gong et al., 2019) and private vehicles (Huang et al., 2019; Fan et al., 2019). Such data underpins varied applications, ranging from the derivation of O-D matrices, gauging human behavior during evacuations, estimating vehicle miles travel, to identifying both travel patterns and modes of transport.

Table 2.1 Selected studies related to application of trajectory data on transportation planning.

Authors, year	Application	Products	Data type	Case study area	Data amount	Method
Toole et al.,2015	O-D Matrix Derivation	An interactive visualization platform	Call detail records (CDRs) from mobile phones	Multiple cities (Boston and San Francisco, USA, Lisbon and Porto, Portugal, and Rio de Janeiro, Brazil)	1 week - 70 gigabytes (for Boston)	
Zhao et al.,2022	Human behavior during evacuation		Cell Phone GPS data	Sonoma County, CA	44,211,050 records, or a total of 5338 residents	
Fan et al.,2019	Vehicle mile traveled estimation	A method for estimating vehicle mile traveled	GPS vehicle trajectory data	Maryland	19.8 million GPS trajectories	Apache Spark-based geo-computing framework
Dabiri et al.,2020	Identification of travelers' transportation modes	A semi-supervised and supervised methods for identifying modes	GPS data from loggers and GPS-phones	GeoLife project	182 users with 17,621 trajectories	
Huang et al.,2019	Identification of travelers' behavior pattern	Algorithms to identify individual travel patterns of private car users A two-layer framework to identify activities of taxi trip and estimate the return trips and successive activities	Private car trajectory dataset	NA	25,564 private cars	DBSACN
Gong et al.,2019	Daily activity analysis		Taxi GPS dataset	New York, US; Shenzhen, China		Bayesian inference and Monte Carlo simulation

Table 2.2 Selected studies related to application of trajectory data on traffic operation.

Authors, year	Application	Products	Data type	case study area/Program name	Data amount	Method
Dabiri et al.,2020	Identify the vehicle's class	Algorithm to identify the vehicle's class	Private car GPS dataset	Maryland	21 million GPS trajectories	A deep Convolutional Neural Network for Vehicle Classification (CNN-VC) Four deep learning models including convolutional neural network, recurrent neural network, long short-term memory, and gated recurrent unit
Sun et al.,2019	Traffic congestion prediction	Algorithm to predict segment level speed	Taxi GPS trajectory data	Chengdu, China	2000 taxis	
Kan et al.,2019	Traffic congestion detection	Algorithm to detect turn-level congestions in multiple dimensions	Taxi GPS trajectory data	Wuhan, China	6658 taxis	Clustering approach
Kong et al.,2016	Traffic Flow Prediction	Algorithm to estimate and predict traffic congestion	Taxi GPS trajectory data	Beijing, China	12,000 taxis	A Fuzzy Comprehensive Evaluation method with dynamic adaptive weight

Qian et al.,2020	Citywide speed reduction/emission analysis	NA	Major transportation network companies' (TNCs) Uber GPS trajectory data	New York City (NYC)		NA
Kan et al.,2022	Traffic congestion estimation	NA	Taxi GPS trajectory	Wuhan, China	730, 000 trips	NA
Abdelraouf et al.,2023	Traffic estimation	NA	Wejo & Roadside sensor-based data	4 expressways in Orlando, Florida for 14 days	61.3 million GPS points generated from 580,864 unique trips	A deep learning-based sequence-to-sequence architecture called Seq2seq GCN-LSTM

While cell phone data has wide coverage and includes many users, it often lacks consistent details, making it hard to track travel routes. On the other hand, GPS data from vehicles, especially commercial ones like taxis, is easier to analyze for travel patterns. However, these commercial vehicles only cover a small part of all road users. Private car GPS data would be ideal for such travel pattern studies, however, the limited accessibility to such datasets curtails its widespread academic adoption. Huang et al. (2019) show the first effort to explore individual travel patterns of private car users via studying private car trajectory big data, which provides new insight into the research of using private vehicle GPS data in human travel activities, traffic management and urban planning.

2.2 Traffic Operations

Regarding traffic operations, vehicle trajectory data is extensively utilized. Such data mainly serves three purposes in traffic operations: 1) the historical trajectory data aids in assessing the efficiency of a traffic system in selected time range; 2) the real-time trajectory data provide opportunity to predict the future traffic conditions during real-time traffic operations; 3) Identify the traffic patterns in both historical and real-time cases. Table 2.2 presents the selected studies related to application of vehicle trajectory data in traffic operation. As mentioned above, due to the data availability, most studies use commercial vehicle trajectory, such as Taxi GPS data, TNC vehicle GPS data (Sun et al.,2019; Kan et al.,2019; Kong et al.,2016; Qian et al.,2020; Kan et al.,2022). Few studies use private vehicle GPS data (Dabiri et al.,2020; Abdelraouf et al.,2023). The applications cover vehicle class identification (Dabiri et al.,2020), traffic congestion detection (Kan et al.,2019) and prediction (Sun et al.,2019), traffic flow prediction (Kong et al.,2016), traffic speed estimation (Qian et al.,2020).

2.3 Connected/Automated Vehicles

Table 2.3 synthesis some selected studies related to application of trajectory data on connected/automated vehicles. In these studies, the trajectory data provided by Next-generation-simulation (NGSIM) datasets are commonly used to develop, calibrate, and/or validate countless algorithms and models to assist applications in connected and autonomous vehicles. Different from GPS trajectory collected from probe vehicles with a penetration rate, NGSIM contains all trajectory data extracted from video files from all vehicles passing the sites. It serves as great research for traffic microsimulation studies.

2.4 Transportation Safety

Traditional safety analysis is based on the observed crash frequency in road segments or intersections, which can be regarded as post-safety event analysis. Given the random features related to the occurrence of traffic crashes, pre-crash conditions and high-risk events (or near-miss events) should be given more attention in road safety management, which could proactively reduce

Table 2.3 Selected studies related to application of trajectory data on connected/automated vehicles.

Author, year	Application	Data type	Case study area/Program name	Data amount
Qu et al., 2022	CAV - trajectory prediction	Next-generation-simulation (NGSIM) datasets	California I-80 highway	836 vehicles, with 1,549,918 trajectory records
Lin et al., 2021	CAV - HDV longitudinal trajectory prediction in a mixed flow environment	Next Generation Simulation US 101 dataset	US-101	45 min of vehicle trajectories over a 0.3-mi freeway segment with five lanes.
Li et al., 2020	CAV - calibrate microscopic traffic flow (MTF) models	Next Generation Simulation (NGSIM) trajectory datasets	I-80, Lanker shim Boulevard, and US-101	every one-tenth of a second

the likelihood of crashes. Vehicle trajectory data can provide spatial and temporal traffic dynamic information related to crashes and high-risk events. Therefore, many researchers use trajectory data to assist in proactive traffic safety management. The topics covered include real-time crash prediction (Zhang et al., 2022), real-time conflict prediction (Islam & Aty, 2023; Xie et al., 2019), Time-to-Collision (TTC) estimation (Hu et al., 2022), high-risk event detection (Yu et al., 2021), and crash frequency prediction (Zhou et al., 2022). The trajectory data used in these studies includes trajectories from connected vehicles (i.e., Wejo data), trajectories from probe vehicles (i.e., taxi GPS data), and trajectories extracted from video. Table 2.4 presents the selected studies related to the application of trajectory data in Traffic Safety.

2.5 Transportation Accessibility and Environment

In addition to their direct function in aiding transportation planning and travel demand analysis, large-scale trajectory data, capturing both human and goods movements, offer a valuable resource for quantifying facility accessibility and conducting transportation-related environmental analyses. Table 2.5 presents some selected studies related to application of trajectory data on transportation accessibility and environment. Regarding emission analysis, trajectory data offers a means to more accurately measure emissions stemming from transportation activities. For example, Li et al. (2019) studied spatial and temporal dynamic emission patterns based on taxi GPS data in Beijing. The relationships between emissions and road densities are studied. And policy implications of results include the region function planning, traffic planning, the public transport improvement, optimal design of charging stations and new energy vehicles promotion were given based on the results. In another study, Yao et al., (2023) decomposes the relationship between trip and emissions through the trip chain perspective by analyzing massive truck trajectory data in China. Results indicate that the real-world emissions.

Table 2.4 Selected studies related to application of trajectory data on Traffic Safety

Author, year	Application	Data type	Case study area/ Program name	Data amount	Method
Zhang et al., 2022	Real-time crash potential prediction	Wejo	Orlando, Florida area, including State Roads 417, 429, 528, and Florida Turnpike,	200 freeway basic segments, The average length of road segments is around 0.54 miles.	A bidirectional long short- term memory (LSTM) model with two convolutiona l layers
Islam & Aty, 2023	Real-time conflict prediction	Wejo	Orange County of Orlando, Florida	2,568,421 GPS points from 14,753 unique journeys 110,000 vehicles from 60 sub- datasets (4 K resolution, 25 HZ) at six different locations on freeways near Cologne, German, during 2017 and 2018	A long-short term memory (LSTM) model
Hu et al., 2022	Time-to- collision (TTC) estimation	Drone - based HighD dataset	German		NA

Yu et al., 2021	High-risk events detection	Drone - based HighD dataset	German	44,500 driven kilometers and 147 driven hours	NA
Xie et al., 2019	Estimate rear-end crash conflict risk	Traffic video data	Brooklyn, New York City	70-hour	NA
Xu et al., 2019	Crash frequency prediction on the expressways	Taxi GPS data	Shanghai, China	NA	NA
Zhou et al., 2022	Modeling speed mean and variance along the structured roadways	Taxi GPS data	Chengdu, China	5,755 taxis	NA

Table 2.5 Selected studies related to application of trajectory data on transportation accessibility and environment.

Author, year	Application	Data type	case study area/ Program name	Data amount
Li et al., 2018	Emission Estimation	taxi GPS trajectory	Beijing, China	12409 taxis
Yao et al., 2023	Emission Estimation	Truck GPS trajectory	China	2297.8556 million with 2,000 trucks
Sun et al., 2021	Emission Estimation	Truck GPS trajectory	China	4300 taxis
Huang et al., 2021	Accessibility Measure	Taxi GPS data	Shenzhen, China	NA
Zhou et al., 2023	Accessibility	Mobile phone- based Baidu trajectory data	Zhuhai, China	201,124 trajectory data

from heavy-duty diesel trucks vary significantly by trip purpose, such as load status, empty-load rate, driving behavior, etc. Taxis, integral to urban transportation, facilitate door-to-door conveyance, underpinning various urban activities. Consequently, taxi trajectory data frequently serves as a proxy for estimating urban accessibility. Huang et al. (2020), for instance, utilized such data to gauge accessibility, employing refined impedance and attraction functions. Their findings highlighted diminished urban accessibility, albeit with heightened accessibility in the southern central regions during evening peak times. This was attributed to minimal passenger wait times and heightened destination allure. Distinct spatiotemporal factors influenced these accessibility variations. While taxi GPS data is widely adopted in accessibility studies, a limitation is its representation: it chiefly mirrors the accessibility experienced by taxi patrons. For a more holistic,

less biased perspective, data encompassing a broader traveler demographic is preferable. Employing the Baidu trajectory dataset from Zhuhai and considering dynamic consumption travel flows, Zhou et al. (2023) crafted a research framework to dissect the spatial interplay between commercial and residential zones, further assessing the commercial accessibility across diverse social demographics. Their results illuminated that the emergence of spatial association communities is occasionally constrained by natural impediments and administrative delineations.

2.6 Summary

In conclusion, this chapter synthesizes the selected studies from the recent 5 years that use large-scale trajectory data to assist in transportation management and decision-making. The topics covered include applications and research in transportation planning and management, traffic operations, connected and autonomous vehicles, traffic safety, and transportation accessibility and environment. The main findings of literature are shown below:

- The large-scale trajectory datasets used in these studies include Global Positioning System (GPS) from cellular phones, trajectory data from commercial vehicles, trajectory data from connected vehicles, Call Detail Records (CDR) from cellular service providers and drone-based vehicle trajectory video data.
- The automotive vehicles used to provide GPS data are mostly commercial vehicles including taxis and vehicles providing mobility service associated with transportation Network Company (TNC).
- The GPS vehicle trajectory data can be used to reveal & predict traffic conditions. When additional datasets (e.g., crash data, detector data) are available, more proactive traffic management can be implemented, such as, real-time crash risk prediction, real-time volume estimation.
- Trajectory data from cellular service providers are usually used to analyze travel demand and human traveler behavior.
- Drone-based video data which can provide trajectories for all the vehicles in road segments are usually used in surrogate safety analysis and trajectory prediction.
- The research using Crowdsourced Connected Vehicle Data (CVD), such as Wejo, is limited to safety analysis, but the scope is limited to selected freeways with limited length.
- Regional-level and state-level analysis on traffic operation/safety is often conducted based on crowd-sourced traffic data (such as segment-based speed) provided by companies, such as HERE, INRIX, but not on trajectory data directly.

3. Project Framework and Objectives

3.1 Project Scope and Overall Framework

The scope of this project is limited to connected vehicle (CV) data covering specific time periods, geographic areas, vehicle classes, and roadway types within Alabama. The project develops a set of applications leveraging Wejo data to inform key transportation decision-making domains, including travel demand analysis, traffic operations, highway safety, connected and automated vehicles (CAVs), and transportation accessibility. For each application, a case study is conducted to demonstrate how Wejo data can support data-driven decision-making. Figure 3.1 illustrates the overall methodological framework adopted in this project.

The workflow begins with data collection, primarily leveraging connected vehicle (CV) data (e.g., Wejo), supplemented by external datasets such as traffic, safety, and roadway inventory data. The CV data then undergoes processing and pre-analysis, followed by in-depth analysis to extract relevant mobility and behavioral insights.

A key step involves data integration, where CV data are combined with traditional transportation datasets to enhance contextual understanding. Inputs from transportation agencies are incorporated to ensure that the analysis aligns with practical needs and policy priorities. Based on these integrated data, the project advances to application development, targeting multiple domains such as travel demand, operations, safety, and accessibility. For each application, case studies are conducted to demonstrate real-world use and validate the analytical approaches. Finally, the framework produces recommendations and guidance to support data-driven transportation decision-making and implementation.

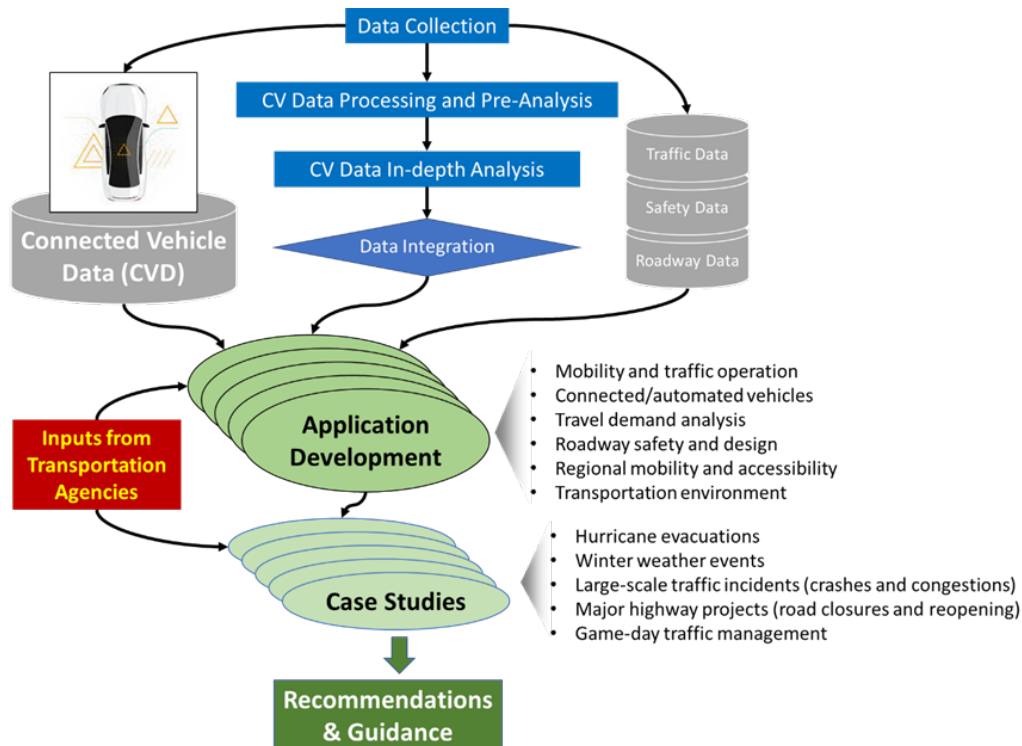


Figure 3.1 Overall Methodology Framework

3.2 Project Main Objectives

The primary objective of this project is to develop a comprehensive framework for leveraging CV data to support data-driven transportation decision-making in Alabama. Specifically, the project aims to:

- Conduct a preliminary analysis of CV data collected from connected vehicles operating on Alabama roadways to understand its characteristics, coverage, and potential applications.
- Develop scalable methods for processing, aggregating, analyzing, and visualizing CV data, as well as integrating CV data with existing Alabama transportation datasets (e.g., CARE crash data, ALGO incident data, traffic count data, and roadway inventory data).
- Design and implement a set of applications that utilize CV data to inform key transportation domains, including transportation planning, traffic operations, connected and automated vehicles (CAVs), highway safety, transportation accessibility, and environmental analysis.
- Conduct case studies to demonstrate the practical use of CV data in real-world scenarios, such as major traffic event management and local transportation planning and engineering practices.

4. Data

4.1 Data Sources

4.1.1 Wejo Data

This project utilizes disaggregated connected vehicle probe (DCVP) data to support a broader range of transportation analyses. The data are provided by Wejo, a commercial vendor of connected vehicle data (CVD), which offers both vehicle movement (trajectory) data and driving event data. These data enable detailed analyses of individual vehicle behavior, corridor-level mobility, and network-wide traffic patterns. The connected vehicle dataset used in this project was sourced via the DB4IoT platform and covers multiple time periods across Alabama’s roadway network.

Table 4.1 Sample of processed DCVP data (captured at 1--3 second intervals) illustrating the structure and temporal resolution of trajectory points used for congestion analysis.

Journey ID	Event Timestamp	Speed (MPH)	Lat	Long	Trip Odometer (m)	Distance on Corridor (Miles)
1	9/27 6:47:46	62.2	33.175	-87.401	8309.7	13.07
1	9/27 6:47:49	66.5	33.175	-87.4	8309.76	13.13
1	9/27 6:47:52	73	33.176	-87.399	8309.82	13.19
1	9/27 6:47:55	77.2	33.176	-87.398	8309.88	13.25
2	9/27 15:50:27	63.6	33.17	-87.538	8729.8	5.04
2	9/27 15:50:30	62.2	33.17	-87.537	8729.85	5.09
2	9/27 15:50:33	62.2	33.17	-87.536	8729.9	5.15
2	9/27 15:50:36	61.5	33.17	-87.536	8729.96	5.2
3	9/27 17:41:55	79.4	33.173	-87.517	3656.2	6.28
3	9/27 17:41:58	80.1	33.173	-87.516	3656.27	6.35
3	9/27 17:42:01	80.1	33.174	-87.515	3656.33	6.42
3	9/27 17:42:04	80.9	33.174	-87.514	3656.4	6.49

Through the partnership between The University of Alabama (UA) and ALDOT, and as part of the Eastern Transportation Coalition pilot that uses Wejo data to monitor and manage emergency evacuations during the hurricane season, the Alabama Transportation Institute (ATI) at UA has invested in acquiring more than 34 TB of Wejo vehicle movement and driving event data covering over 95% of Alabama streets and 99.99% of highways and primary roads. Compared with traditional mobile-based sensor data, the Wejo CVD exhibits several distinguishing characteristics. First, it offers greater spatial accuracy: because OEM telematics use vehicle-integrated GPS and cellular antennas, location data are typically accurate to within a 3-meter radius. Second, it carries more data attributes beyond position, including vehicle start/stop events, hard stops, windshield wiper status, and seat belt usage, which support not only mobility and safety analyses but also weather-responsive and compliance-related studies. Third, it provides a consistent high reporting frequency, with trajectory points recorded every 1–3 seconds, enabling the reconstruction of detailed vehicle maneuvers. Through the MoonShadow DB4IoT visualization platform, users can directly observe real-time locations of individual probe vehicles and generate aggregated and

disaggregated views such as vehicle heading, segment-level speed, point-level speed, and trip origins and destinations.

The DCVP dataset comprises high-resolution GPS trajectory records collected at 1–3 second intervals. Key data attributes include timestamps, vehicle speed, distance traveled, heading, and geographic coordinates (latitude and longitude). Each record is associated with a unique journey identifier, enabling the reconstruction of complete trip trajectories from individual GPS points. Table 4.1 presents a representative sample of the processed DCVP dataset used for congestion analysis. The data have been spatially aligned to the study corridor using a predefined reference path, with each row corresponding to an individual trajectory point. The Initial Offset denotes the starting position of each trip along the reference path, while the Distance on Corridor represents the aligned longitudinal position used for constructing time–space diagrams and performing clustering-based analyses. This structured dataset forms the foundation for downstream methodologies, including congestion detection and traffic shockwave characterization.

The dataset used for freeway congestion analysis (Chapter 6, Case Study 4) includes over 895,000 journeys and approximately 220 million records collected over a one-year period (August 2020 to July 2021) on the I-20/59 corridor. The connected vehicle penetration rate during the study period was estimated to be approximately 9–10 percent. For the corridor-level mobility analyses (Chapter 6, Case Studies 1–3), connected vehicle trajectory data collected during March 2021 were used. The raw dataset was organized hierarchically by date and hour, with each daily folder containing 24 hourly subfolders of anonymized GPS trajectory records. Across the study period, the data exhibited a stable and interpretable weekly rhythm consistent with expected urban mobility patterns.

4.1.2 Other Major Data Sources

In addition to the Wejo connected vehicle data, this project draws on several Alabama transportation datasets that are routinely used by ALDOT and partner agencies. These datasets are integrated with the Wejo data in the case studies that follow (Chapters 5–9) to enable location-based analyses of safety, mobility, travel demand, and accessibility.

CARE Crash Data

The Critical Analysis Reporting Environment (CARE) is a crash data analysis platform originally developed for problem identification and countermeasure development in traffic safety. CARE is maintained by ALDOT with support from the Center for Advanced Public Safety (CAPS) at The University of Alabama, and it stores statewide traffic crash records including crash location and time, crash type, injury severity, and other standard crash features. In this project, CARE crash data are used to provide ground-truth safety outcomes against which connected-vehicle-based surrogates (e.g., hard braking events from Wejo) are benchmarked. Access to CARE data is governed by a Non-Disclosure Agreement, and only non-identifiable crash attributes are used in the analyses reported here.

ALGO Incident Data

ALGO is a traffic incident management and reporting platform used in Alabama to log roadway incidents such as crashes, stalled vehicles, debris, and construction-related events, along with their reported locations, start and clearance times, and incident classifications. In this project, ALGO incident records are used as an independent reference for validating congestion events detected

from Wejo trajectory data, most directly in the freeway congestion case study (Chapter 6, Case Study 4), where the congestion scoring framework is compared against reported incidents to distinguish documented disruptions from unreported congestion episodes observed in the Wejo data. Beyond validation, ALGO records provide temporal context for interpreting anomalies in corridor-level mobility metrics.

Points of Interest (POI) Data

Points of Interest data describe the location and category of destinations that attract travel, such as schools, hospitals and other health and medical facilities, retail stores, dining and drinking establishments, workplaces, recreational sites, and similar activity locations. In this project, Foursquare POI data are used as the primary source because they provide detailed, hierarchical category labels and broad coverage across Alabama. POI records are geocoded with latitude/longitude and are used in the accessibility and origin–destination case studies to classify trip destinations, to enable category-level accessibility measures (e.g., travel time and distance to essential services), and to quantify the relative attractiveness of individual destinations based on observed Wejo trips.

HPMS Roadway Inventory Data

The Highway Performance Monitoring System (HPMS) is a national roadway information system administered by the Federal Highway Administration, with state-level data maintained by ALDOT. HPMS provides a shapefile-based roadway network along with roadway inventory attributes such as functional classification (e.g., Interstate, Other Principal Arterial, Minor Arterial, Major Collector, Local Road), annual average daily traffic (AADT), number of lanes, area type (urban/rural), and segment geometry. In this project, the 2017 HPMS shapefile is used as the canonical road network to which Wejo trajectory points, hard braking events, and CARE crashes are spatially joined. HPMS attributes also serve as explanatory variables for safety screening and for stratifying mobility and safety findings by functional class and area type.

Other Supporting Datasets

Several additional datasets are used in specific case studies and are introduced here for completeness. OpenStreetMap (OSM) is used to extract signalized and unsignalized intersection locations and local road geometry for the volatility and intersection delay analyses. Traffic Analysis Zone (TAZ) boundaries from the Tuscaloosa-area MPO are used to aggregate Wejo-derived origins and destinations for travel-demand analysis. American Community Survey (ACS) population estimates are used to normalize crash and hard braking counts at the census-tract level for statewide screening. Residential-location data (from parcel and address layers) and Open Route Service (ORS) travel-time estimates are used in the accessibility case studies as reference benchmarks against observed Wejo travel patterns.

4.2 Data Processing

The high-resolution Wejo trajectory and driving event data are spatially mapped to the underlying road network and then linked to other data sources based on their geographic location. Because the raw trajectory records contain only latitude and longitude coordinates and a traditional spatial join over hundreds of millions of points is computationally prohibitive, a K-D Tree algorithm (a k-dimensional binary tree for nearest-neighbor search) is used as the primary spatial mapping algorithm, reducing the nearest-neighbor search complexity to approximately $O(\log n)$. The K-D

Tree algorithm is applied, for example, to align individual Wejo trajectory points to a corridor reference path, to match large-scale driving records and hard braking events to their nearest intersections and road segments, and to connect trip origins and destinations to residential locations and points of interest. Once the Wejo data are anchored to the road network through the K-D Tree spatial mapping, they are joined with the other major data sources introduced in Section 4.1.2 based on their spatial location: CARE crash records are merged with the same road segments or intersections through shared spatial references (route, milepost, or coordinates); HPMS attributes (AADT, functional class, number of lanes, area type, etc.) are attached to the HPMS segments to provide roadway context; and POI records are linked to trip destinations through spatial proximity. Temporal alignment is also performed as needed, for example by extracting pre-event traffic conditions around a crash or by aggregating trajectory records to hourly or daily intervals. Detailed, application-specific data processing procedures (including the exact spatial buffers, time windows, aggregation rules, and filtering criteria) are documented in the corresponding case study sections of each application chapter.

5. Applications in Traffic Signal Evaluation and Optimization

Contributors: Jun Liu, Jiayi Kong, Yanfang Su

The CV data can show detailed vehicle trajectories (speeds and locations) approaching intersections. Through analyzing the vehicle trajectories, hard brakes in dilemma zone can be identified, delay experienced by each probe vehicle can be measured, and energy consumption and emissions can also be computed based on the vehicle trajectories. With all this information, strategies to improve traffic signals can be developed and evaluated with objectives to reduce hard brakes, delay, and/or energy consumption/emissions. This section provides a case study that leverages CV data to quantify the intersection delay.

5.1 Case Study 1 – Intersections Delaying Analysis

5.1.1 Study area

Intersections play a crucial role in traffic systems by managing the flow of vehicles, pedestrians, and cyclists where roads meet. There are mainly two types of intersections: signalized intersections and unsignalized intersections.

5.1.2 Methodology

Vehicles are affected by intersections' control signs and traffic flows when driving through, leading to delays that are distributed across a range of values. To identify the intersection-delaying from Wejo trajectory data, this study applies the following methodology.

Each intersection is set to form a circle-shaped polygon with a 500ft radius as a buffer zone. This study matches each vehicle's trajectory records to the road and determines if the vehicle is currently passing through an intersection (and which intersection it is passing through). Then each trajectory's intersection activity is extracted for further analysis. After computing the time vehicles spend on passing through intersections as well as the travel time without intersection-delay, the delaying time for each intersection on each vehicle passing activity is calculated.

The delaying time calculation process can be written as the formula below:

$$\begin{aligned} \text{Delaying} &= \text{Actual Passing Through Time} - \text{Ideal Passing Through Time} \\ &= (t_{\text{leave}} - t_{\text{enter}}) - \frac{\text{Distance}(\text{leave}, \text{enter})}{\text{speed limit}} \end{aligned}$$

5.1.3 Results and Analysis

Unsignalized Intersection vs signalized intersection

Figure 5.1 shows the distribution of vehicles delaying on signalized intersections versus delaying on unsignalized intersections. Analysis shows that vehicles are spending much more time passing through signalized intersections compared to unsignalized intersections.

Peak Hours versus non-peak hours

To explore the traffic congestion status on peak hours, this study divides typical daytime into two categories as shown in Table 5.1.

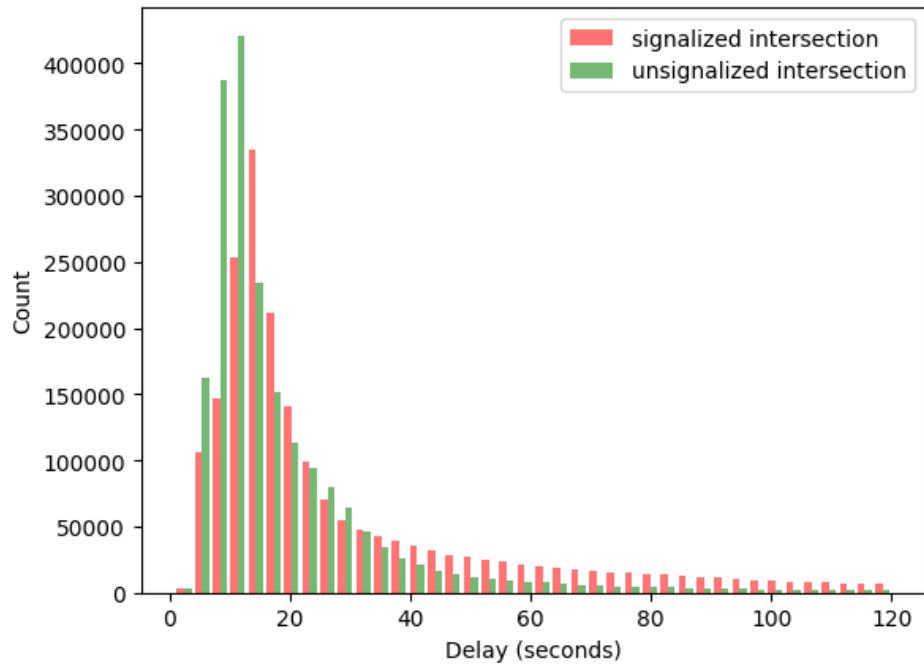


Figure 5.1 The distribution of vehicles delaying on signalized intersections versus delaying on unsignalized intersections

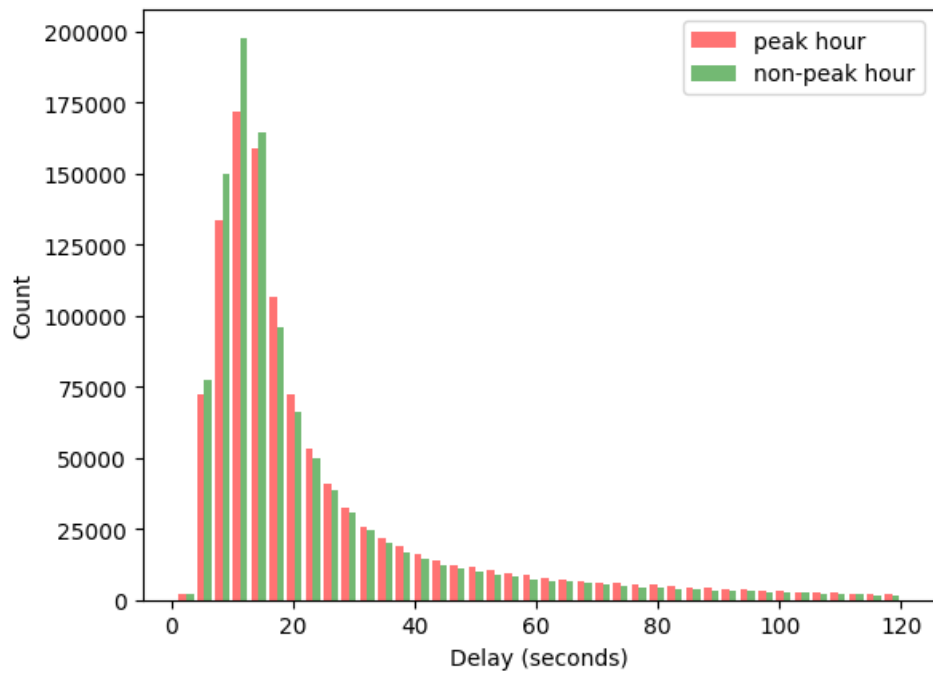


Figure 5.2 Delaying distribution during peak hours vs non-peak hours

Table 5.1 Peak hour and non-peak hour distribution

Time Range	Definition
7:00 AM ~ 9:59 AM, 4:00 PM to 6:59 PM	Peak Hour
6:00AM ~ 7:00 AM, 10:00 AM ~ 3:59 PM, 7:00 PM ~ 8:00 PM	Non-peak hour (daytime)

Figure 5.2 shows the delaying difference between peak hours and non-peak hours. It is apparent that vehicles are facing more jammed traffic in peak hours, which leads to longer passing time through intersections.

Analysis on intersection delays comparison

After calculating the intersection-passing-delays, the overall intersection delaying map is represented in Figure 5.3, where larger circles represent higher delaying intersections:

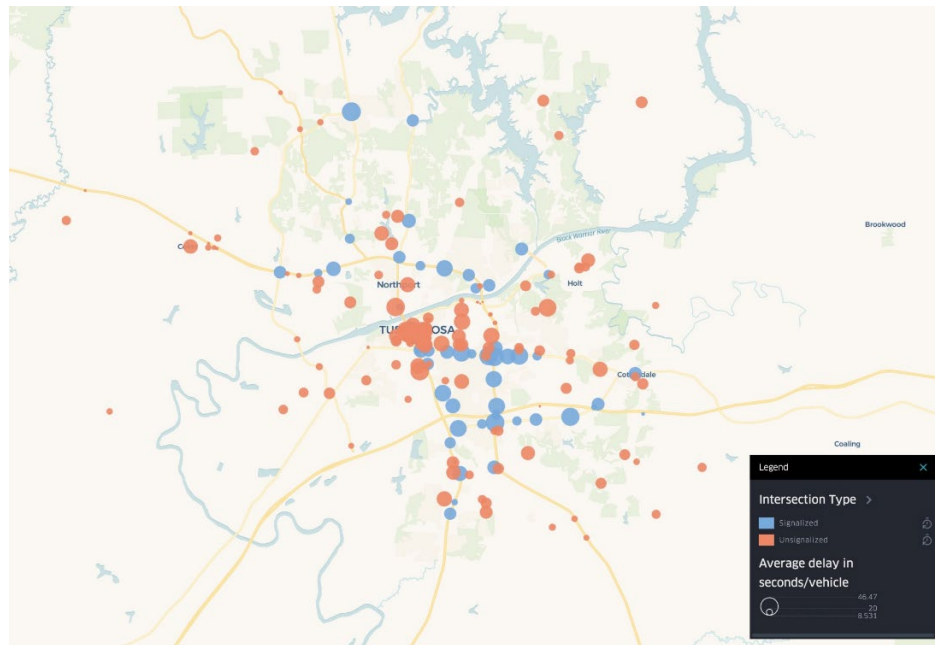


Figure 5.3 Spatial distribution of signalized and unsignalized intersections and their associated delays

6. Applications in Corridor or Networkwide Mobility Analysis

Contributors: Jun Liu, Zihe Zhang, Yang Yu

The CV data contains high-resolution probe vehicle speed information in road networks. The data makes it easy to evaluate the traffic flow characteristics throughout a corridor or network, and also to identify the locations and time periods where and when speeds are significantly affected (i.e., areas or times of congestion). Both the spatial and temporal (seasonal) traffic speed patterns are important information for agencies to develop traffic operation management and improve policies.

6.1 Case Study 1 – Corridor Speed and Delay Analysis in Tuscaloosa

6.1.1 Study Corridors and Data Overview

This case study examines corridor-level traffic performance in Tuscaloosa using Wejo connected vehicle trajectory data. The purpose of this section is to establish baseline traffic conditions for three representative roadway corridors before examining disruption patterns under extreme weather and game-day conditions in the following sections. The analysis focuses on five corridor-level performance measures: travel time, average speed, delay, delay rate, and trip count. The three study corridors were selected to represent distinct functional roles within the Tuscaloosa roadway network and to capture different forms of traffic activity associated with both routine travel and event-related circulation.

- McFarland Blvd E is the primary arterial corridor in the study area and functions as a high-capacity east–west through route. The selected segment extends from the intersection with 15th Street to the final intersection before the roadway splits into ramps toward I-20/I-59. Because of its multilane, high-speed configuration, this corridor captures both regional through traffic and traffic surges associated with major events.
- 15th Street serves as an important connector between the University of Alabama campus area and McFarland Blvd. The selected segment begins at Hackberry Lane, a key access point for campus circulation and baseball event parking, and extends eastward to McFarland Blvd. This corridor reflects a mixture of commuter-oriented traffic and event-distributed movements.
- University Blvd E represents a more localized access corridor. The selected segment extends from 2nd Avenue to the last intersection before McFarland Blvd and is closely associated with the baseball stadium and nearby parking areas. Its narrower cross section and more local access function make it useful for examining lower-volume but activity-sensitive traffic conditions.

Figure 6.1 shows the locations of the three corridors and their surrounding activity context.

The analysis uses Wejo CV trajectory data collected during March 2021. The raw dataset was organized hierarchically by date and hour, with each daily folder containing 24 hourly subfolders of anonymized GPS trajectory records. For consistency in day-level comparison, the first and last dates in the available archive were excluded because they contained incomplete data coverage. Across the remaining study period, the data exhibited a stable and interpretable weekly rhythm. Weekday traffic volumes remained relatively consistent, Fridays showed the highest trip activity, and weekends, especially Sundays, displayed markedly lower volumes. This pattern is consistent with expected urban mobility behavior in a university-oriented city, where weekday travel is shaped by commuting and campus activity while weekend demand is substantially reduced. As shown in Figure 6.2, two dates, March 17 and March 25, exhibited noticeably lower trip volumes than adjacent weekdays; these anomalies are examined later in the weather-related case analysis.

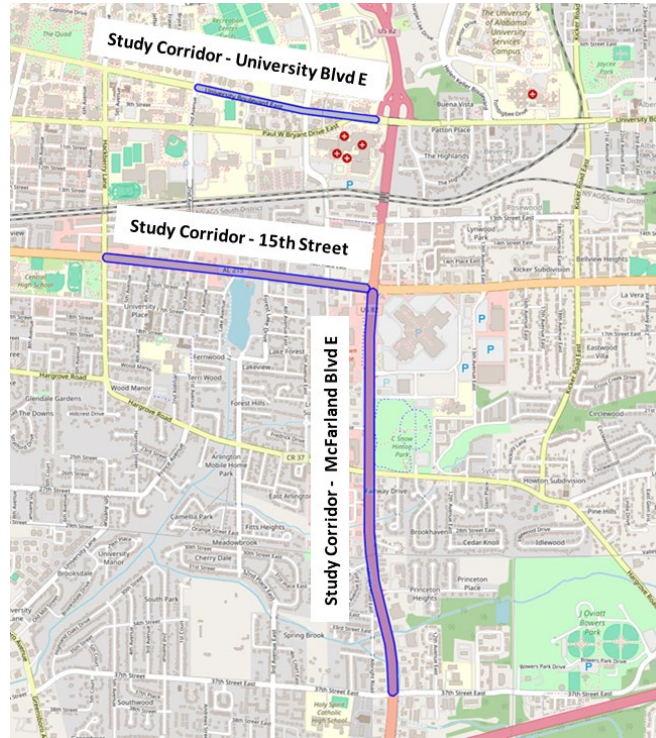


Figure 6.1 Study Corridors in Tuscaloosa

March						
03-01 31,561	03-02 32,166	03-03 35,396	03-04 35,978	03-05 37,796	03-06 31,887	03-07 26,793
03-08 34,951	03-09 35,706	03-10 35,967	03-11 36,626	03-12 38,584	03-13 29,754	03-14 24,329
03-15 29,507	03-16 30,274	03-17 25,369	03-18 32,594	03-19 34,286	03-20 29,629	03-21 25,445
03-22 35,891	03-23 32,428	03-24 36,740	03-25 23,431	03-26 39,455	03-27 31,543	03-28 24,665
03-29 36,773	03-30 35,284	03-31 26,771				
Mon	Tue	Wed	Thu	Fri	Sat	Sun

Figure 6.2 Daily Trip Counts in March 2021.

Taken together, the selected corridors and the March 2021 trajectory dataset provide an appropriate basis for developing corridor-level baseline traffic profiles in Tuscaloosa. The next subsection describes the preprocessing and trip extraction procedures used to convert the raw Wejo trajectory records into cleaned corridor-level trip data for performance measurement and comparison.

6.1.2 Data Preparation and Corridor Trip Extraction

The raw Wejo trajectory data were processed through a structured pipeline to generate cleaned corridor-level trip records for subsequent performance analysis. Because the original data were stored in separate hourly folders, all hourly files for each date were first merged using the unique journey identifier so that each vehicle trip could be reconstructed as a continuous trajectory. This step was necessary because a non-negligible portion of trips extended across hour boundaries. For example, on March 1, 2021, 13.78% of journeys spanned two consecutive hours and 0.84% extended across three or more hours. Without this integration step, such trips would have been

fragmented, leading to biased estimates of travel time and speed. After trip reconstruction, all timestamps were converted from UTC to local time to ensure consistency with local traffic activity and later event-based comparisons.

To isolate traffic movements associated with the selected corridors, polygon buffers were constructed around each corridor centerline and used as spatial filters for the point-level GPS data. The buffer sizes were differentiated according to corridor class and roadway geometry, with a wider buffer assigned to McFarland Blvd E and narrower buffers used for 15th Street and University Blvd E. This design balanced two competing objectives: retaining corridor-related vehicle movements while limiting contamination from nearby but unrelated roadway activity. After applying the corridor buffers, approximately 2% of all GPS points and 15–16% of unique journey IDs were matched to the monitored corridors during the 30-day study period, indicating that the selected segments captured a meaningful subset of vehicle trips within the broader Tuscaloosa network. Table 6.1 summarizes the buffer settings used for the three study corridors, along with the corresponding roadway context and justification for spatial matching.

Table 6.1 Study Corridors and Buffer Settings

Corridor	Buffer Radius (meters)	Justification
McFarland Blvd E	35 m	Wide arterial with 6+ lanes and high-speed traffic
15th Street	25 m	Collector road with moderate width and event overflow
University Blvd E	20 m	Narrower urban streets, closer to pedestrian areas

A second-stage screening procedure was then applied to identify valid traversing trips for each corridor. For all selected corridors, start and end zones were defined near the upstream and downstream endpoints, and a trip was retained only if its earliest matched point fell within the start zone and its latest matched point fell within the end zone. This procedure ensured that the retained observations represented complete corridor traversals rather than partial entries or short local movements within the buffer area. To further reduce noise, trips with calculated travel times shorter than 10 seconds were removed. This threshold excluded implausibly short traversals caused by GPS jitter, segmentation artifacts, or incomplete point sequences.

Following trip extraction, additional filtering was performed to remove implausible travel times while preserving realistic congestion conditions. On the lower end, a minimum allowable travel time was defined based on corridor length and a threshold speed equal to 1.5 times the posted speed limit. The lower-bound travel time was calculated as

$$T_{\min} = \frac{D}{1.5 v_{\text{PSL}}} \quad (1)$$

where D is the corridor distance and v_{PSL} is the posted speed limit expressed in consistent units. This conservative threshold was used to eliminate unrealistically fast trips while retaining valid high-speed movements under uncongested conditions.

For excessively slow trips, a hybrid IQR–FHWA thresholding framework was applied. When the hourly sample size was sufficient, an interquartile range criterion was used to identify upper-end outliers. Under lower-sample conditions, the threshold defaulted to an FHWA-informed delay definition so that meaningful congestion would not be removed simply because of limited observations. In all cases, the final upper bound was selected conservatively to preserve slow but

plausible trips. As an additional safeguard, trips with travel times greater than 1,800 seconds were removed as likely artifacts of trajectory discontinuity or segmentation failure. This combined strategy reduced extreme noise while maintaining the sensitivity of the dataset to genuine corridor congestion. Figure 6.3 illustrates the overall workflow used to reconstruct, extract, and filter valid corridor trips from the raw Wejo trajectory records. The resulting minimum and maximum allowable travel times for the three corridors are listed in Table 6.2.

Table 6.2 Minimum and Maximum Allowable Travel Times

Corridor	Distance (m)	Threshold Speed (mph)	Max Time (s)	Min Time (s)
University Blvd E	1094	35	291.6	46.7
15th Street	1536	45	318	51.0
McFarland Blvd E	2374	50	442.8	70.8

The cleaned trip dataset produced through these steps provided the foundation for corridor-level calculation of travel time, speed, delay, and volume measures. The next subsection defines these performance metrics and describes how they were aggregated to construct the March 2021 baseline traffic profiles for the three study corridors.

6.1.3 Performance Measures and Baseline Construction

Based on the cleaned corridor trip dataset, five performance measures were computed to characterize traffic conditions along the three study corridors: travel time, average speed, delay, delay rate, and trip count. Travel time was defined as the elapsed time between the first and last matched GPS point of each valid corridor trip. Average speed was derived from the corridor length and observed travel time. Trip count was defined as the number of valid trips retained within each temporal aggregation unit. Together, these measures capture both traffic demand and operational performance, allowing the analysis to distinguish between changes in traffic volume and changes in travel efficiency.

Among these measures, delay and delay rate were used as the primary indicators of corridor inefficiency. Delay was defined as the nonnegative difference between the observed travel time and the speed-limit-based travel time, expressed as

$$\text{Delay} = \max(0, T_{\text{obs}} - T_{\text{SL}}) \quad (2)$$

where T_{obs} is the observed corridor travel time and T_{SL} is the reference travel time under the posted speed limit. The speed-limit-based travel time was calculated as

$$T_{\text{SL}} = \frac{D}{v_{\text{PSL}}} \quad (3)$$

where D is corridor length and v_{PSL} is the posted speed limit. This formulation ensures that only excess travel time beyond the nominal free-flow expectation is counted as delay.

To support comparison across corridors with different lengths and speed limits, a normalized delay rate was further calculated as

$$\text{Delay Rate} = \frac{\text{Delay}}{T_{\text{SL}}} \quad (4)$$

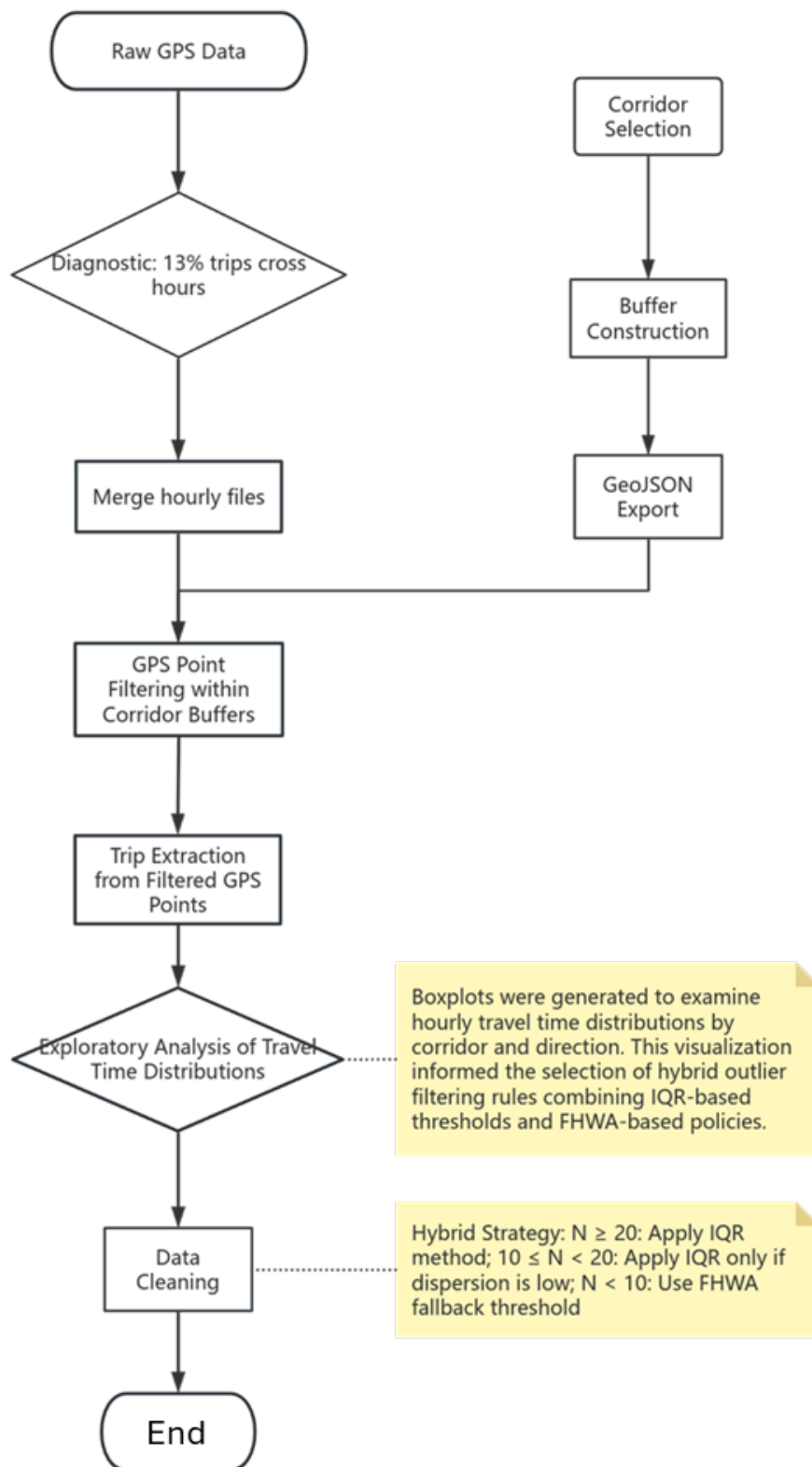


Figure 6.3 Corridor Trip Extraction and Filtering Workflow

Unlike absolute delay, delay rate provides a dimensionless measure of relative inefficiency and therefore allows more direct comparison between corridors with different geometric and operational characteristics. This normalization is important in the present study because McFarland Blvd E, 15th Street, and University Blvd E differ substantially in roadway function, corridor length, and posted speed environment. As a result, delay rates were particularly useful in the cross-corridor comparison and in the later identification of disruption intensity under weather and event conditions.

To describe traffic conditions at multiple temporal scales, the trip-level performance measures were aggregated into four complementary levels: hourly summaries, hourly-by-weekday summaries, weekday-level summaries, and daily summaries. The hourly aggregation captured short-term intra-day dynamics and revealed the timing of directional peaks or abrupt congestion changes. The hourly-by-weekday aggregation was used to identify recurring temporal structures, such as commuter peaks, midday congestion, and weekday versus weekend differences. Weekday-level summaries provided a broader view of directional variation across the week, while daily

Table 6.3 Definitions of Performance Measures and Aggregation Levels

Measure / Level	Definition	Unit	Purpose in Analysis
Travel Time	Elapsed time between the first and last matched GPS point of a valid corridor trip	Seconds	Captures observed traversal time along the corridor
Average Speed	Corridor length divided by observed travel time for each valid trip	MPH	Represents the overall operating speed of corridor traffic
Delay	Nonnegative difference between observed travel time and speed-limit-based travel time	Seconds	Measures excess travel time beyond nominal free-flow conditions
Delay Rate	Delay divided by the speed-limit-based travel time	Dimensionless	Normalizes corridor inefficiency to support comparison across corridors with different lengths and speed limits
Trip Count	Number of valid corridor trips retained within each aggregation unit	Counts	Reflects traffic demand and supports interpretation of changes in operational conditions
Hourly Aggregation	Aggregation of trip-level measures by corridor, direction, and hour	—	Used to capture short-term intra-day dynamics and identify peak-period variation
Hourly-by-Weekday Aggregation	Aggregation of hourly measures by weekday and hour	—	Used to identify recurring temporal structures such as directional commuter peaks and midday congestion
Weekday-Level Aggregation	Aggregation of measures across each weekday category	—	Used to compare broader weekday patterns and directional variation
Daily Aggregation	Aggregation of measures by calendar day	—	Used to examine day-to-day stability and identify anomalous dates within the study month

summaries supported month-level screening for stable baseline patterns and anomalous dates. This multi-level aggregation framework enabled the analysis to move from trip-level observations to interpretable corridor traffic profiles. Table 6.3 summarizes the definitions of the performance measures and the aggregation levels used to construct the baseline traffic profiles.

These measures and aggregation levels established the March 2021 baseline traffic profiles for the three study corridors. This baseline framework serves two purposes. First, it characterizes the normal traffic signatures of corridors with different functional roles in the Tuscaloosa network. Second, it provides the reference structure required for the subsequent analyses of weather-related and game-day disruptions, where observed corridor conditions are compared against expected baseline patterns rather than interpreted in isolation.

6.1.4 Baseline Corridor Traffic Patterns

The baseline traffic patterns derived from the March 2021 Wejo data reveal clear differences across the three study corridors in both temporal structure and operational characteristics. These differences reflect the distinct functional roles of McFarland Blvd E, 15th Street, and University Blvd E within the Tuscaloosa roadway network. At the same time, the baseline results provide a necessary reference for distinguishing routine directional and time-of-day variation from the weather-related and event-related disruptions examined in the following sections.

Representative Hourly Patterns by Corridor

Among the three corridors, McFarland Blvd E shows the clearest arterial traffic signature. The hourly results indicate a pronounced directional pattern, with average speed declining and delay-related measures increasing during both the morning and afternoon periods. As shown in Figure 6.4, these effects are especially visible between 7:00 and 9:00 AM and again between 3:00 and 6:00 PM. The southbound direction experiences slightly higher delay rates during the afternoon peak, indicating stronger congestion accumulation in that direction. This pattern is consistent with the corridor's role as a major arterial carrying both commuter-oriented and broader through traffic.

A different baseline structure is observed on 15th Street. Rather than being dominated by a sharp directional commuter peak, this corridor shows a broader period of reduced speed and elevated delay centered around the middle of the day. As shown in Figure 6.5, the most notable inefficiency occurs between approximately 11:00 AM and 3:00 PM, particularly in the westbound direction. This pattern suggests that 15th Street is influenced by a mixture of campus access, local circulation, and connector traffic between the university area and McFarland Blvd, rather than by a single concentrated commuter surge. Relative to McFarland Blvd E, the congestion profile on 15th Street is therefore more distributed across the daytime period and more indicative of localized access-related pressure.

University Blvd E shows lower overall trip volumes than the other two corridors, but its hourly traffic measures still reveal interpretable operational variation. Although absolute delay levels are generally smaller, localized periods of speed reduction and delay increase remain visible, indicating that even this lower-volume corridor experiences meaningful fluctuations tied to surrounding activity. Its narrower roadway form and access-oriented role make it less suitable for absorbing surges in demand and more sensitive to localized interference.

Taken together, the hourly results indicate that corridor-level traffic performance in Tuscaloosa cannot be interpreted uniformly across facilities. McFarland Blvd E reflects a high-capacity arterial

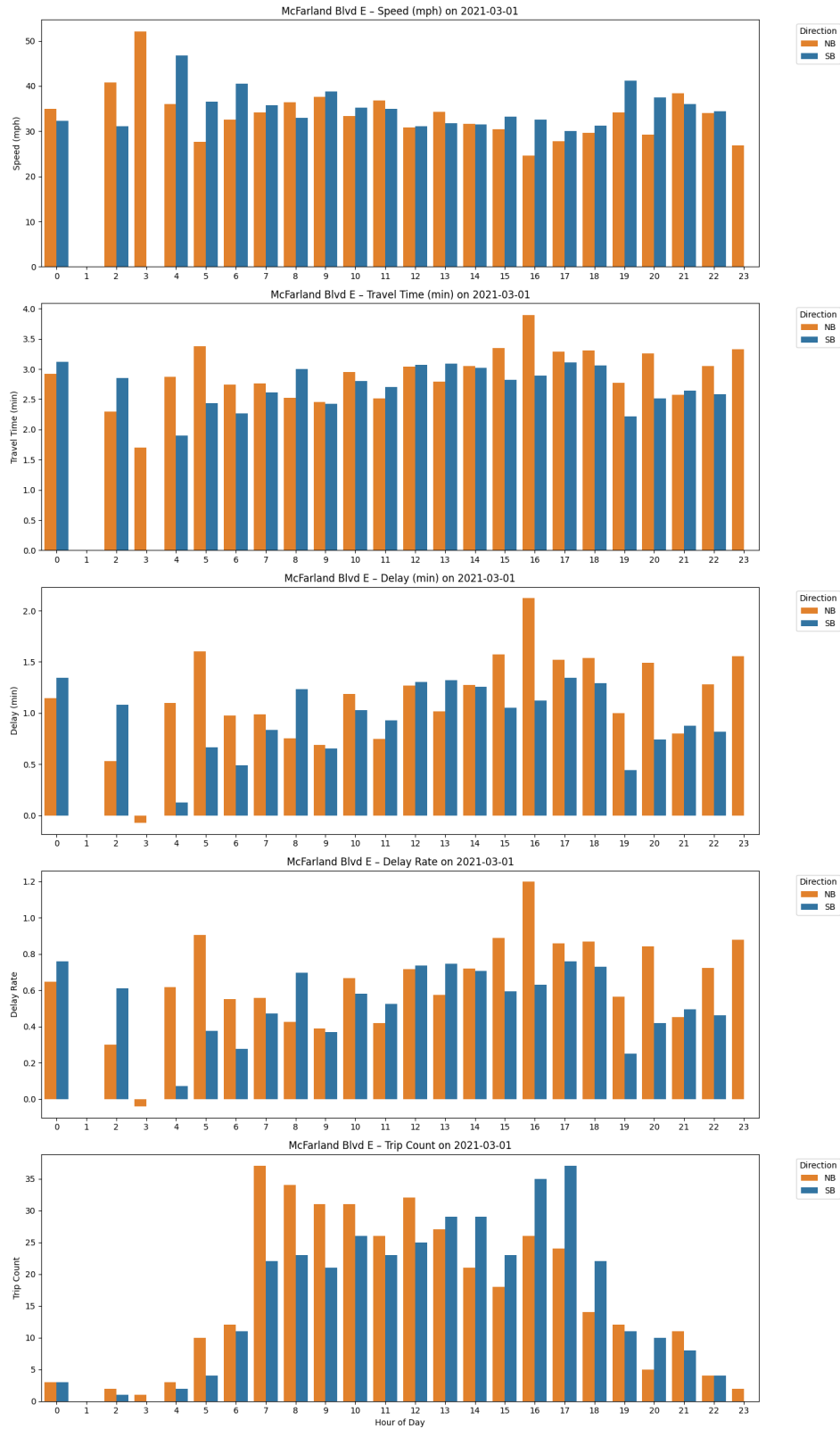


Figure 6.4 Hourly Traffic Performance on McFarland Blvd E

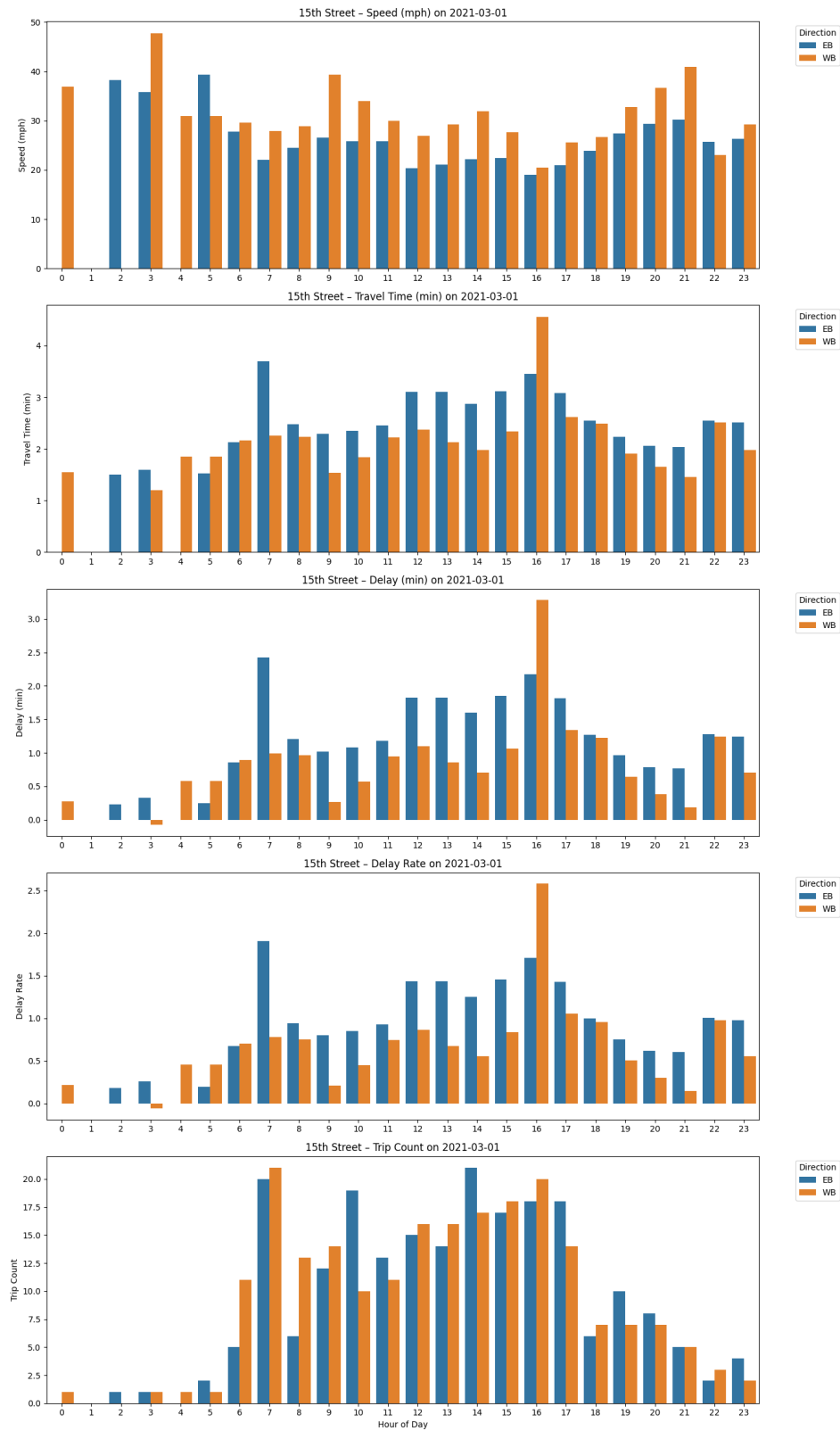


Figure 6.5 Hourly Traffic Performance on 15th Street

profile, 15th Street reflects a mixed commuter and connector profile with stronger midday inefficiency, and University Blvd E reflects a lower-volume but activity-sensitive local access profile.

Recurring Weekday and Directional Patterns

The baseline structure becomes clearer when the hourly results are interpreted in relation to recurring weekday and directional variation. On McFarland Blvd E, the directional asymmetry observed in the hourly results is consistent with a corridor shaped by concentrated peak-period demand. The northbound direction is more active during the morning period, whereas the southbound direction experiences stronger delay accumulation later in the day, especially during higher-demand weekdays. This recurring contrast indicates that the observed baseline is not a one-day artifact, but a stable directional pattern associated with the corridor's arterial function.

On 15th Street, the recurring pattern is less about sharp directional commuting peaks and more about sustained inefficiency across the middle part of the day. This broader temporal footprint suggests that the corridor is influenced by a wider mix of traffic purposes, including campus access, local circulation, and movements linking the university area with adjacent arterials. The baseline weekday structure on 15th Street is therefore more consistent with a connector corridor exposed to localized activity pressure than with a purely commute-driven arterial facility.

Although University Blvd E has lower sample volumes, its weekday structure still supports the same general interpretation. The corridor does not display the same level of directional peak concentration as McFarland Blvd E, nor the same sustained midday congestion as 15th Street, but it does show repeatable local fluctuations that are consistent with a lower-volume access corridor. For this reason, the weekday interpretation in this section is presented through text synthesis rather than through an additional figure.

Overall, the weekday-based interpretation reinforces the corridor hierarchy observed in the hourly plots. McFarland Blvd E is characterized most strongly by directional peak-period demand, 15th Street by more distributed midday inefficiency, and University Blvd E by localized, lower-volume variation.

Daily Variation and Anomalous Days

Daily aggregation provides a month-level view of baseline stability and helps distinguish ordinary temporal variation from externally influenced conditions. As shown in Figure 6.6, traffic activity during March 2021 follows a clear and interpretable rhythm, with weekday conditions remaining relatively stable and weekend volumes consistently lower. This pattern indicates that the observed corridor traffic behavior is not random from day to day but instead reflects a regular temporal structure suitable for baseline comparison.

In addition to confirming overall month-level stability, the daily results also identify a small number of dates that depart noticeably from the normal weekday pattern. In particular, March 17 and March 25 show visibly lower trip activity than adjacent weekdays. These deviations are important because they demonstrate that the baseline framework is sensitive enough to detect externally affected conditions rather than simply smoothing them into the monthly average. As a result, these two dates are treated in the next section as weather-related disruption cases rather than as ordinary weekday variation. The daily analysis serves two purposes. First, it confirms that March 2021 provides a sufficiently stable empirical baseline for corridor-level traffic assessment.



Figure 6.6 Daily Traffic Variation in March 2021

Second, it identifies the specific anomalous dates that warrant separate disruption-focused analysis in the following section.

6.1.5 Key Findings

The baseline analysis shows that the three study corridors exhibit clearly different traffic signatures in both operational performance and temporal structure. McFarland Blvd E functions most strongly as a high-capacity arterial with pronounced directional peak-period effects, while 15th Street shows a broader midday congestion profile consistent with its connector role between campus activity and the arterial network. University Blvd E carries lower overall traffic volumes but still displays stable local activity-related variation. At the same time, the March 2021 data provide a sufficiently consistent baseline for corridor-level comparison, while also highlighting a small number of anomalous dates that justify separate disruption-focused analysis in the following sections.

6.2 Case Study 2 – Traffic Disruption under Extreme Weather

Contributors: Jun Liu and Yang Yu

6.2.1 Event Selection and Comparison Strategy

March 17 and March 25, 2021, were selected as weather-related disruption cases for detailed analysis on McFarland Blvd E. In the March 2021 baseline presented in the previous section, both dates appeared as anomalous weekdays with visibly lower traffic activity than adjacent normal weekdays. Weather records further indicate that both dates were affected by adverse conditions, including severe thunderstorms, rainfall, fog, and reduced visibility. Detailed hourly weather information for these two dates is provided in Appendix Table 6.4 and Table 6.5. Taken together, the anomaly identified in the baseline data and the documented weather conditions make March 17 and March 25 appropriate case days for examining weather-related traffic disruption on the study corridor.

To evaluate the traffic effects of these weather events, each case day was compared with a matched weekday baseline constructed from three normal dates with the same day-of-week pattern. Baseline dates were selected to exclude other known weather disturbances and special-event conditions so that the comparison would reflect ordinary corridor operation under typical conditions. The analysis focused on McFarland Blvd E and compared hourly changes in average speed, travel time, delay, delay rate, and trip count between each weather day and its matched baseline. This comparison design makes it possible to interpret the observed traffic response as a deviation from expected weekday conditions rather than as isolated day-level variation.

6.2.2 Weather-Related Traffic Disruptions

The two selected weather days did not produce the same level of traffic disruption on McFarland Blvd E. March 17 showed a clearer and more sustained deterioration in corridor performance, while the impact on March 25 was weaker and more localized. As shown in Figure 6.7 and Figure 6.8, the difference between the two dates is visible across speed, delay, delay rate, and trip count, indicating that adverse weather did not translate into a uniform traffic response even on the same corridor.

On March 17, the most evident disruption occurred between approximately 10:00 AM and 1:00 PM. During this period, average speed in both directions fell below the matched baseline, while

delay and delay rate increased substantially. The northbound direction showed the strongest deterioration, with delays exceeding 1.5 minutes and delay rate rising above 1.1 relative to the

6.2.3 Appendix

Table 6.4 Weather Conditions on March 17, 2021

Time	Weather	Visibility
12:53 am	Mostly cloudy.	9 mi
1:09 am	Passing clouds.	9 mi
1:31 am	Partly cloudy.	9 mi
2:27 am	Partly cloudy.	10 mi
2:53 am	Passing clouds.	9 mi
3:53 am	Mostly cloudy.	10 mi
4:53 am	Passing clouds.	10 mi
5:25 am	Passing clouds.	10 mi
5:53 am	Partly cloudy.	10 mi
6:40 am	Light rain. Mostly cloudy.	10 mi
7:16 am	Partly sunny.	10 mi
7:32 am	Low clouds.	10 mi
7:53 am	Low clouds.	9 mi
8:53 am	Low clouds.	7 mi
9:41 am	Fog.	5 mi
10:04 am	Fog.	2 mi
10:28 am	Fog.	2 mi
10:53 am	Fog.	3 mi
11:25 am	Heavy rain. Fog.	1 mi
11:42 am	Light rain. Fog.	3 mi
11:53 am	Light rain. Low clouds.	7 mi
12:33 pm	Overcast.	10 mi
12:53 pm	Overcast.	10 mi
1:28 pm	Light rain. Fog.	5 mi
1:53 pm	Thunder showers. Broken clouds.	7 mi
2:18 pm	Thunder showers. Fog.	6 mi
2:38 pm	Heavy rain. Fog.	2 mi
3:04 pm	Strong thunderstorms. Fog.	1 mi
3:10 pm	Strong thunderstorms. Fog.	0 mi
3:31 pm	Thunderstorms. Fog.	1 mi
3:53 pm	Thunder showers. Cloudy.	10 mi
4:18 pm	Thunder showers. Mostly cloudy.	9 mi
4:36 pm	Broken clouds.	9 mi
4:53 pm	Broken clouds.	8 mi
5:53 pm	Thunderstorms. Fog.	5 mi
6:09 pm	Light rain. Mostly cloudy.	8 mi
6:53 pm	Thunder showers. Mostly cloudy.	10 mi
7:53 pm	Thunderstorms. Mostly cloudy.	9 mi
8:18 pm	Thunderstorms. Mostly cloudy.	7 mi
8:31 pm	Strong thunderstorms. Fog.	2 mi
8:55 pm	Thunderstorms. Fog.	2 mi
9:08 pm	Heavy rain. Fog.	3 mi
9:20 pm	Light rain. Fog.	6 mi
9:53 pm	Light rain. Overcast.	10 mi
10:53 pm	Passing clouds.	10 mi

11:53 pm	Overcast.	10 mi
----------	-----------	-------

Table 6.5 Weather Conditions on March 25, 2021

Time	Weather	Visibility
12:53 am	Rain. Fog.	6 mi
1:53 am	Light rain. Overcast.	9 mi
2:27 am	Thunder showers. Fog.	5 mi
2:42 am	Rain. Fog.	4 mi
2:53 am	Thunder showers. Fog.	3 mi
3:21 am	Heavy rain. Fog.	2 mi
3:31 am	Light rain. Fog.	5 mi
3:53 am	Rain. Fog.	5 mi
4:36 am	Light rain. Fog.	4 mi
4:53 am	Light rain. Fog.	4 mi
5:17 am	Mostly cloudy.	10 mi
5:53 am	Light rain. Mostly cloudy.	8 mi
6:53 am	Light rain. Overcast.	8 mi
7:03 am	Low clouds.	8 mi
8:01 am	Thunderstorms. Fog.	6 mi
8:16 am	Thunder showers. Fog.	6 mi
8:23 am	Fog.	6 mi
8:53 am	Heavy rain. Fog.	1 mi
9:11 am	Light rain. Fog.	3 mi
9:53 am	Thunder showers. Overcast.	7 mi
10:04 am	Overcast.	7 mi
10:18 am	Thunder showers. Overcast.	8 mi
10:53 am	Thunder showers. Fog.	6 mi
11:21 am	Light rain. Low clouds.	7 mi
12:00 pm	Fog.	6 mi
12:08 pm	Thunder showers. Broken clouds.	7 mi
12:17 pm	Thunder showers. Scattered clouds.	7 mi
12:37 pm	Thunderstorms. Partly sunny.	9 mi
12:57 pm	Thunder showers. Cloudy.	7 mi
1:31 pm	Thunderstorms. Mostly cloudy.	8 mi
1:53 pm	Thundershowers. Fog.	6 mi
2:08 pm	Thunder showers. Fog.	2 mi
2:17 pm	Thunderstorms. Fog.	2 mi
2:33 pm	Light rain. Fog.	4 mi
2:58 pm	Thunderstorms. Fog.	1 mi
3:20 pm	Thunderstorms. Fog.	4 mi
3:38 pm	Strong thunderstorms. Fog.	1 mi
4:19 pm	Thunder showers. Fog.	2 mi
4:35 pm	Thunder showers. Fog.	1 mi
4:53 pm	Thunder showers. Fog.	6 mi
5:05 pm	Thunder showers. Fog.	4 mi
5:09 pm	Rain. Fog.	2 mi
5:20 pm	Light rain. More clouds than sun.	7 mi
5:53 pm	Light rain. Partly sunny.	8 mi
6:53 pm	Mostly cloudy.	9 mi
7:53 pm	Thunder showers. Overcast.	8 mi
8:10 pm	Partly cloudy.	9 mi
8:53 pm	Partly cloudy.	10 mi

9:53 pm	Passing clouds.	10 mi
10:53 pm	Overcast.	9 mi
11:53 pm	Mostly cloudy.	8 mi



Figure 6.7 Hourly Traffic Performance on McFarland Blvd E on March 17, 2021



Figure 6.8 Hourly Traffic Performance on McFarland Blvd E on March 25, 2021

normal weekday profile. At the same time, trip count remained broadly stable, which suggests that the observed performance degradation was not primarily driven by a sharp reduction in traffic demand. Instead, the corridor experienced a clear decline in operating efficiency under adverse weather while continuing to carry a comparable level of activity. This temporal pattern is consistent with the documented period of heavy rain, fog, and reduced visibility on that date.

The March 25 case showed a different pattern. Although adverse weather was also documented on that date, the traffic response on McFarland Blvd E was more limited in both duration and intensity. As shown in Figure 6.8, the most visible changes occurred during the early-to-mid afternoon period, when trip count declined between about 1:00 PM and 4:00 PM. However, the associated increases in delay and delay rate were modest, and in the northbound direction these measures remained close to or below the matched baseline. Compared with March 17, this pattern indicates a weaker disruption effect on corridor performance, with lower demand likely offsetting part of the operational pressure created by adverse weather. The shorter and less concentrated period of deterioration is also consistent with the weather record for March 25, which shows adverse conditions but less pronounced sustained traffic degradation than on March 17.

The two case days show that weather-related traffic disruption on McFarland Blvd E was event-specific rather than uniform. March 17 produced a clearer efficiency loss under continuing traffic activity, whereas March 25 showed a milder response accompanied by some reduction in volume. This contrast indicates that the traffic effect of adverse weather depends not only on the presence of storms or low visibility, but also on when those conditions occur and how they overlap with ongoing corridor demand.

6.2.4 Key Findings

The contrast between March 17 and March 25 shows that adverse weather does not produce a fixed or uniform traffic effect, even in the same corridor. In this case, the difference in traffic response was closely related to the timing, duration, and concentration of the weather conditions relative to ongoing corridor activity. On March 17, the most severe deterioration occurred during a period when trip volume remained broadly stable, indicating that the corridor continued to carry active demand while operating conditions worsened. On March 25, by contrast, the observed weather effects were more limited and coincided with some reduction in trip count, which likely reduced the pressure on corridor performance. These two cases therefore suggest that the operational effect of adverse weather depends not only on the presence of rain, fog, or reduced visibility, but also on how those conditions overlap with active travel demand on the corridor.

The comparison also shows that delay rate was more informative than absolute delay in identifying short-lived weather-related inefficiency. Because delay rate expresses excess travel time relative to the speed-limit-based travel time, it provides a normalized measure of deterioration that is easier to interpret across changing traffic conditions. In the March 17 case, elevated delay rate accompanied relatively stable trip activity, making the loss of corridor efficiency more visible than raw delay alone. In the March 25 case, the smaller change in delay rate was consistent with the more moderate overall disruption. For this reason, delay rate provided the clearest single indicator of weather-related operational degradation in the present case study.

6.3 Case Study 3 – Corridors Traffic Pattern

Contributors: Jun Liu and Yang Yu

6.3.1 Event Selection and Comparison Strategy

This section focuses on selected game-day cases within the March 2021 study period to examine how event-related activity affected corridor traffic conditions near the University of Alabama area. The purpose is not to cover all possible game-day conditions, but to identify the most representative and interpretable cases from the available March 2021 observations. Within these cases, the observed impacts were associated primarily with medium-scale event activity near the stadium area and nearby access corridors. Accordingly, the findings in this section are intended to describe the selected March 2021 cases examined here rather than to generalize to larger event conditions such as football-scale game days.

Two comparison settings were selected for detailed analysis. The primary comparison is March 26, 2021, a baseball event day, versus March 19, 2021, a softball event day. This pair was chosen because both dates fall on Fridays and therefore provide a useful basis for examining how different event types and event intensities were associated with different corridor responses under a similar weekday context. A secondary comparison uses March 23, 2021, a baseball event day, and March 16, 2021, a non-event Tuesday baseline. This pair serves as a supporting reference for distinguishing event-related slowdown from ordinary weekday variation. Across both settings, corridor-level performance was evaluated using average speed, travel time, delay, delay rate, and trip count. Segment-level speed mapping was then applied only to the strongest disturbance case in order to identify where the most pronounced localized slowdown occurred within the corridor network.

6.3.2 Corridor-Level Traffic Responses during Selected Game-Day Events

Corridor-level comparisons show that, within the selected March 2021 game-day cases examined in this report, event-related traffic effects were concentrated on specific access-oriented corridors rather than distributed uniformly across all three study facilities. Among the three corridors, 15th Street exhibited the clearest and most consistent event-related response, while McFarland Blvd E generally remained close to baseline conditions. University Blvd E showed more limited or mixed deviations depending on the comparison setting. Table 6.6 summarizes the main corridor-level responses observed in the two selected comparison settings.

One corridor-level disturbance was observed in Friday in comparison between March 26, 2021, a baseball day, and March 19, 2021, a softball event day. Under this comparison, 15th Street showed substantially lower speed, higher delay, higher delay rate, and greater trip activity on March 26 than on March 19, especially during the period surrounding game start. In contrast, the March 19 softball case remained much closer to typical Friday conditions, with only modest deviations in the same metrics. This comparison indicates that the two Friday game-day cases did not generate the same corridor response and that the baseball event produced a distinctly stronger operational effect on 15th Street. At the same time, McFarland Blvd E remained comparatively stable, while University Blvd E showed weaker and less consistent changes, reinforcing the view that the strongest event-related pressure was concentrated on the corridor most directly associated with access to the stadium area.

Another comparison between March 23, 2021, a baseball event day, and March 16, 2021, a non-event Tuesday baseline, provides supporting evidence for the same interpretation. Relative to the ordinary weekday reference, 15th Street again showed the clearest event-related slowdown, with

Table 6.6 Summary of Corridor-Level Traffic Responses during Selected Game-Day Events

Comparison Setting	Corridor	Relative Response	Interpretation
March 26 baseball vs March 19 softball	15th Street	Stronger slowdown, higher delay and delay rate, and higher trip count on March 26	Clearest event-related disturbance among the selected March 2021 cases
March 26 baseball vs March 19 softball	McFarland Blvd E	Minor deviation from reference conditions	Event-related impact remained limited on the arterial corridor
March 26 baseball vs March 19 softball	University Blvd E	Mixed to modest change	Weaker event-related signal than on 15th Street
March 23 baseball vs March 16 non-event Tuesday	15th Street	Visible slowdown relative to the weekday baseline	Event-related effect remained detectable against ordinary weekday conditions
March 23 baseball vs March 16 non-event Tuesday	McFarland Blvd E	Little change relative to baseline	Corridor remained close to ordinary weekday conditions
March 23 baseball vs March 16 non-event Tuesday	University Blvd E	Limited and inconsistent deviation	No clear corridor-level disturbance pattern

reduced speed, increased delay and delay rate, and a noticeable increase in trip activity, particularly in the eastbound direction toward the stadium-facing side of the corridor. By contrast, McFarland Blvd E remained close to baseline, and University Blvd E showed little evidence of a strong corridor-level disturbance. This Tuesday-to-Tuesday comparison is useful because it confirms that the event-related signal on 15th Street was not limited to gameday versus gameday contrasts but could also be detected against a normal weekday reference without weather-related confounding. Figure 6.9 shows the corridor-level performance of 15th Street on March 23 relative to the matched March 16 non-event Tuesday baseline.

The corridor-level evidence identifies 15th Street as the most event-sensitive corridor in the selected March 2021 cases and establishes the March 26 baseball event as the strongest and most interpretable disturbance case. At the same time, the March 23 baseball comparison confirms that event-related slowdown remains detectable even under a controlled weekday baseline. These findings provide the operational context for the spatial analysis in the next subsection, which examines where the strongest slowdown was concentrated within the local corridor network.

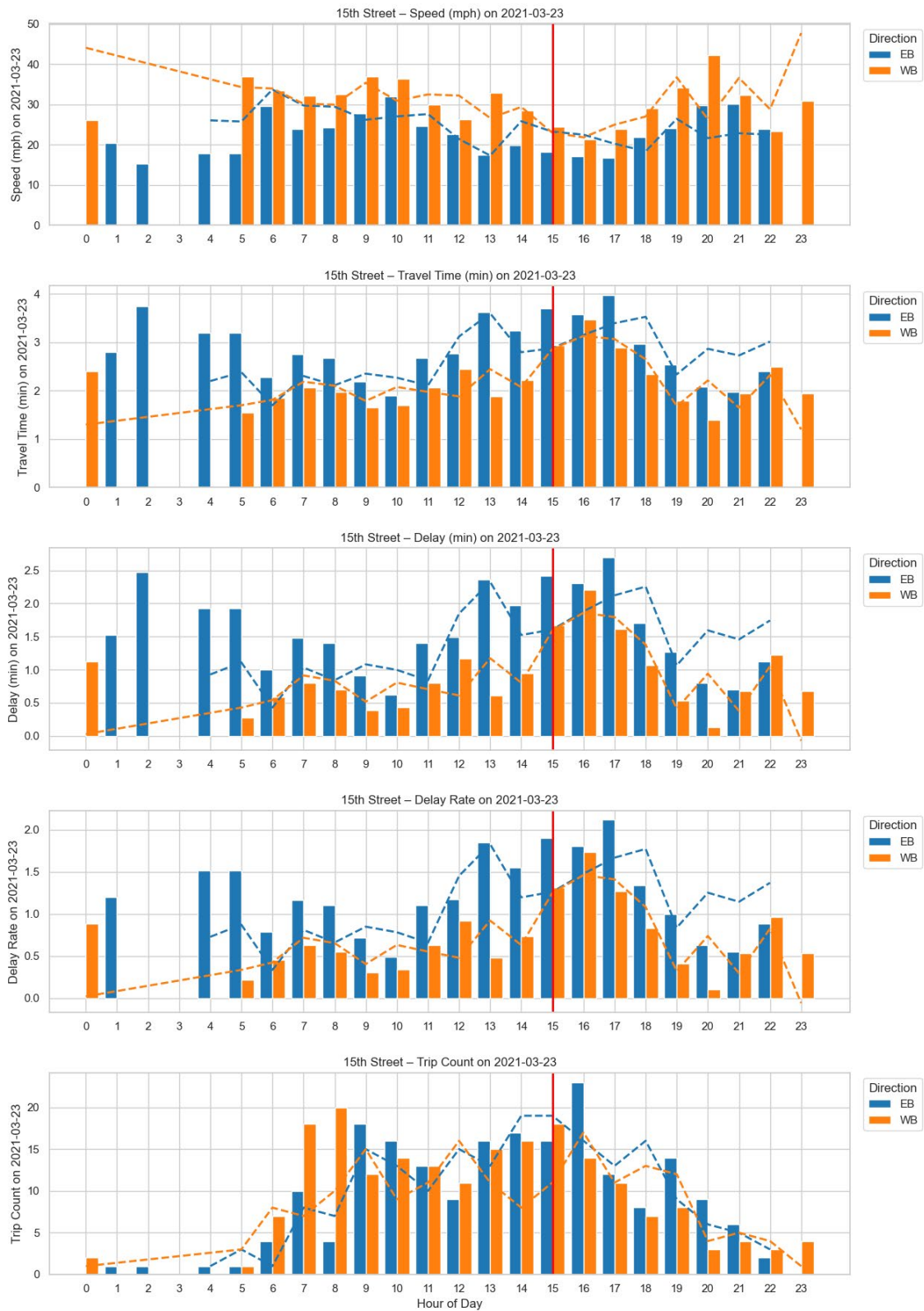


Figure 6.9 Corridor-Level Traffic Performance on 15th Street on March 23 relative to the March 16 baseline

6.3.3 Spatial Evidence from Point-Speed and Segment-Level Mapping

The corridor-level comparisons in the previous subsection establish that 15th Street was an event-sensitive corridor in the selected March 2021 cases. To further examine how this event-related effect appeared in space, two complementary map-based views were used. The first compares point-speed conditions between the March 19 softball case and the March 26 baseball case under a similar Friday setting. The second examines how segment-level average speed evolved on March 23 from 14:00 to 17:00 during the event period itself. Together, these two spatial views help distinguish between cross-case spatial contrast and within-event-day spatial evolution on 15th Street.

As shown in Figure 6.10, the Friday comparison between March 19 and March 26 indicates that the March 26 baseball event produced a denser and more spatially concentrated cluster of low-speed points on 15th Street than the March 19 softball event. The contrast is visible at both 16:00 and 18:00, but it becomes clearer at 18:00, when March 26 shows a larger share of stopped, near stopped, and severely congested observations near the eastern portion of the corridor and around the access area connected to McFarland Blvd. By comparison, the March 19 maps display a weaker and less concentrated low-speed pattern during the same late-afternoon period. This spatial contrast supports the corridor-level result that the March 26 baseball case generated a stronger operational effect than the March 19 softball case under a comparable Friday context.

Figure 6.11 provides a different but complementary view by showing segment-level average speed conditions on March 23 from 14:00 to 17:00. Unlike the Friday comparison in Figure 6.10, this figure is not a baseline-relative change map; instead, it shows how speed conditions evolved across corridor segments within the event day itself. The maps indicate that lower-speed conditions became more evident as the afternoon progressed, with the eastern section of 15th Street showing the clearest concentration of slower segments. The pattern is especially visible around 15:00 and 16:00, when multiple adjacent segments near the eastern access-oriented portion of the corridor display more pronounced slowdown than the rest of the facility. By 17:00, some easing is visible, but the eastern side still remains slower than the upstream portion of the corridor. This within-day evolution suggests that the March 23 event-related effect was spatially concentrated rather than uniformly distributed along 15th Street.

Overall, the two spatial figures reinforce the same interpretation. Figure 6.10 shows that the March 26 baseball case produced a stronger late-afternoon concentration of low-speed conditions than the March 19 softball case, while Figure 6.11 shows that, within the March 23 event day, slower conditions developed most clearly on the eastern, access-oriented portion of 15th Street during the mid-to-late afternoon period. These map-based results explain why 15th Street emerged as the most event-sensitive corridor in the preceding corridor-level analysis and show that, in the selected March 2021 cases, the strongest game-day effects were concentrated near venue-access segments rather than distributed evenly across the wider corridor network.

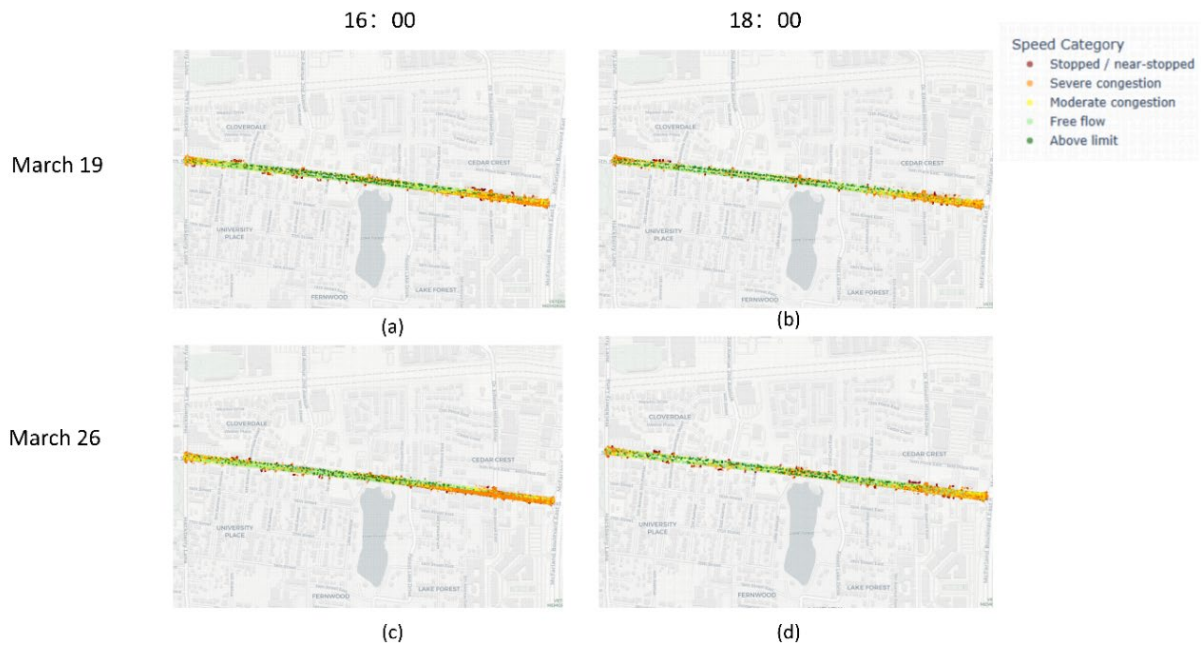


Figure 6.10 Point-Speed Maps on 15th Street during March 19 and March 26 Game-Day Comparisons

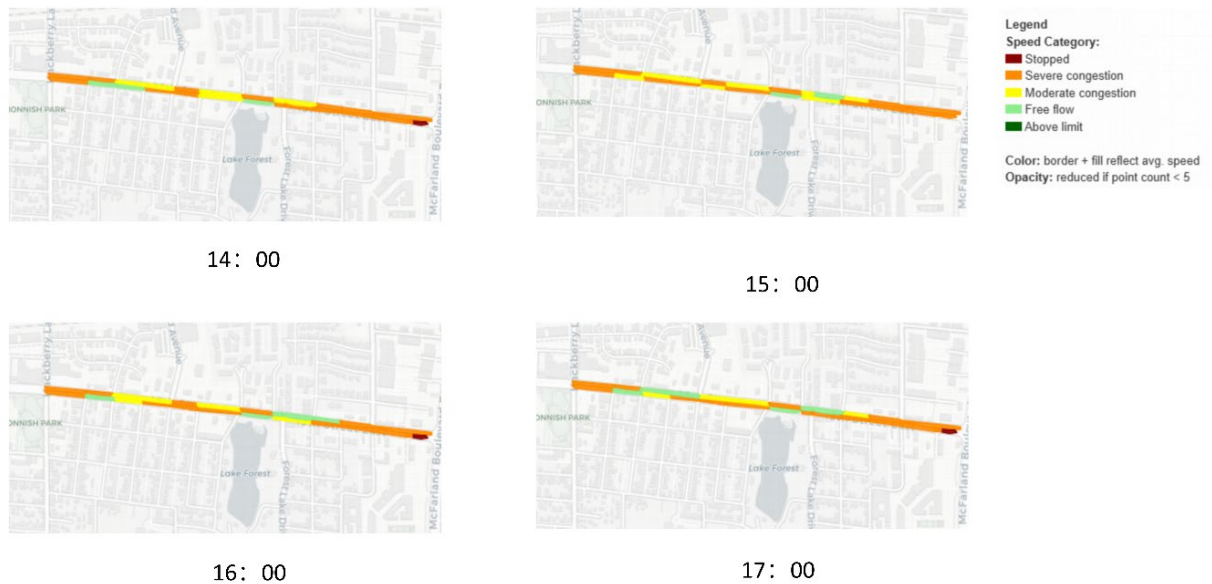


Figure 6.11 Segment-Level Average Speed on 15th Street from 14:00 to 17:00 on March 23

6.3.4 Key Findings

The selected March 2021 game-day cases indicate that event-related traffic effects were primarily localized on corridors and segments directly connected to venue access rather than appearing as uniform corridor-wide or systemwide degradation. This interpretation should be limited to the cases examined in this section, which mainly represent medium-scale stadium activity during the March 2021 study period, rather than larger event conditions such as football-scale game days. Within these selected cases, McFarland Blvd E generally remained close to baseline conditions, while 15th Street consistently showed the clearest event-related response. This pattern suggests that corridor function and proximity to the stadium access area were more important than simple network position in determining the magnitude of observed disruption.

Across the two comparison settings, 15th Street emerged as the most event-sensitive corridor, and the March 26 baseball event produced the strongest observable disturbance among the selected March 2021 cases. The corridor-level comparisons showed stronger slowdown, higher delay and delay rate, and greater trip activity on 15th Street during the baseball case than during the softball or non-event reference cases. The map-based evidence further clarified that these effects were spatially concentrated, with the most visible slowdown occurring on the eastern, access-oriented portion of 15th Street and near key approach intersections connected to the stadium area. Together, these findings show that the selected game-day impacts in this study were not uniformly distributed across the local corridor network, but instead reflected localized, event-driven pressure on the facilities most directly involved in venue access and arrival circulation.

6.4 Case Study 4 – Freeway Congestion Analysis Using DCVP Data

Contributors: Vamshi Chaitanaya Anнимalla, Alex Hainen, Praveena Penmetsa

In addition to the corridor-level analysis presented above, this section presents a comprehensive freeway congestion analysis using disaggregated connected vehicle probe (DCVP) data on the I-20/59 corridor in Alabama. Traditional traffic data sources, such as sensors (including radar, Bluetooth, and inductive loops) and aggregated probe vehicle data, provide insights but remain constrained by limited coverage, fragmentation, and low granularity. This study uses DCVP data to provide high-resolution insights into congestion dynamics. Advanced trajectory alignment and clustering techniques identify congestion clusters, analyze queue lengths, and examine shockwave propagation. Driver behavior analysis, such as sudden deceleration and illegal turns, reveal safety risks and secondary crash potential. A comparative analysis of freeway and alternate routes evaluates the impact of congestion on travel efficiency. Additionally, a congestion scoring framework quantifies severity, validated against regional incident data to detect disruptions not tied to reported incidents.

6.4.1 Empirical Setting

The study corridor is a segment of the I-20/59 freeway in Alabama, stretching from Exit 68 to Exit 104. This corridor spans approximately 36 miles and includes several critical exits and interchanges between Tuscaloosa and Birmingham, making it prone to congestion. Figure 6.12 illustrates the study corridor and highlights the extents considered for analysis. This section was selected as a representative case to study congestion dynamics and evaluate management strategies in high-traffic freeway conditions.

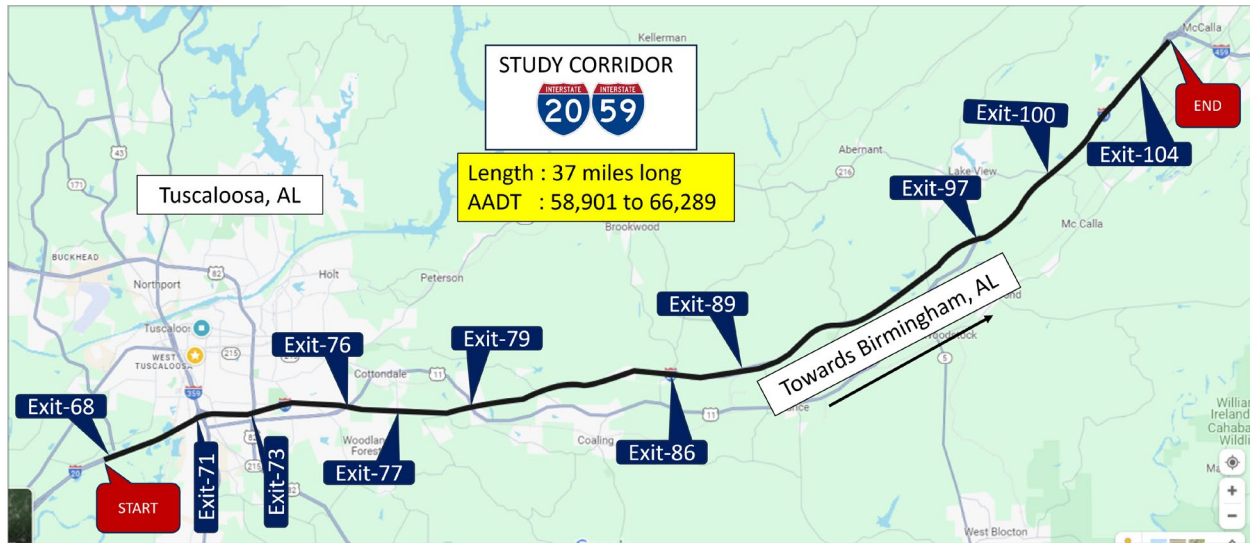


Figure 6.12 I-20/59 Corridor Study Area, spanning from Exit 68 to Exit 104

6.4.2 Freeway DCVP Data

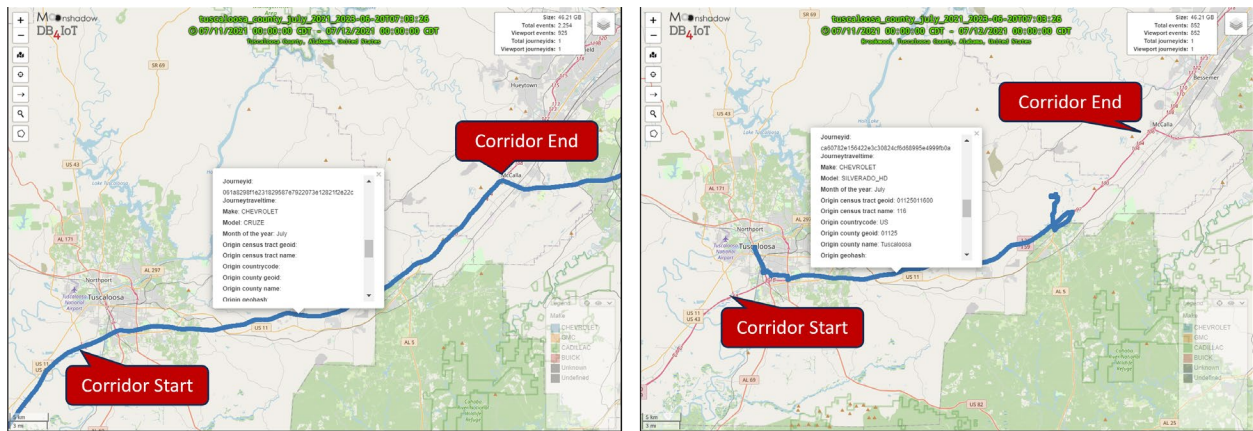
Disaggregated connected vehicle probe (DCVP) data were sourced from a commercial vendor via the DB4IoT platform, collected over one year (August 2020 to July 2021). The dataset comprises over 895,000 journeys and 220 million records with a temporal resolution of 1 to 3 second intervals. Key attributes include timestamps, vehicle speed, distance, heading, and geographical coordinates. The connected vehicle penetration rate during the study period was approximately 9 to 10 percent. A reference journey spanning the entire corridor length (Figure 6.13(a)) was used as a baseline for spatially aligning other journeys that covered only partial segments. Table 6.7 displays a sample of the processed DCVP dataset used for congestion analysis.

Table 6.7 Sample of processed DCVP data (captured at 1 to 3 second intervals) illustrating the structure and temporal resolution of trajectory points used for congestion analysis.

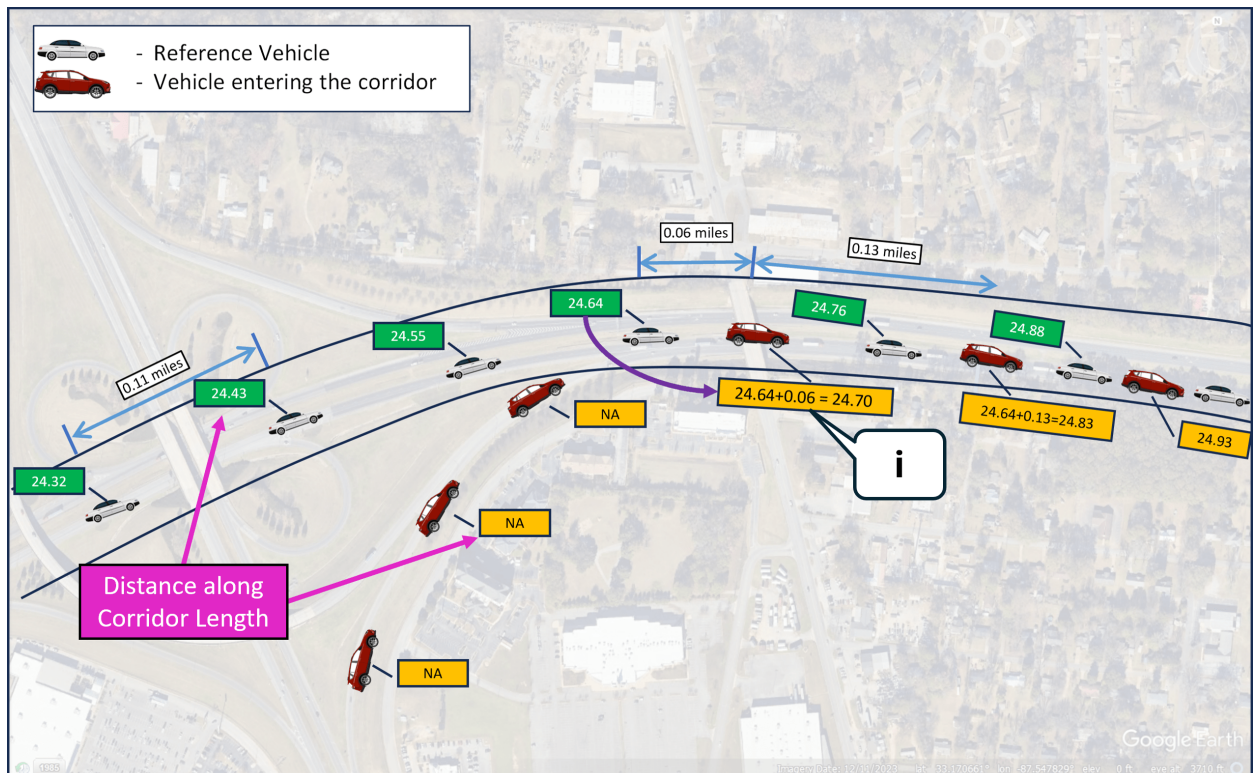
ID	Event Timestamp	Speed (MPH)	Lat	Long	Trip Odometer (m)	Initial Offset (Miles)	Distance on Corridor (Miles)
1	9/27 6:47:46	62.2	33.175	-87.401	8309.70	6.36	13.07
1	9/27 6:47:49	66.5	33.175	-87.400	8309.76	6.36	13.13
1	9/27 6:47:52	73.0	33.176	-87.399	8309.82	6.36	13.19
2	9/27 15:50:27	63.6	33.170	-87.538	8729.80	0.38	5.04
2	9/27 15:50:30	62.2	33.170	-87.537	8729.85	0.38	5.09
3	9/27 17:41:55	79.4	33.173	-87.517	3656.20	1.24	6.28
3	9/27 17:41:58	80.1	33.173	-87.516	3656.27	1.24	6.35

6.4.3 Trajectory Alignment Using K-D Tree Algorithm

A K-D Tree algorithm was used to align all journeys to a fixed corridor reference path without relying on external datasets like road network shapefiles or mile-marker data. A reference journey spanning the entire corridor length was identified, and a K-D tree was constructed from its



(a)



(b)

Figure 6.13 (a). Alignment of DCVP data journeys using the K-D Tree algorithm: Journey which traveled the full length of the study corridor (left) and journey that traveled only part of the corridor (right). (b). Assigning the initial position to a jour

coordinates to enable efficient nearest-neighbor searches. For each vehicle journey, the first observed point along the corridor was identified, and the K-D tree located the closest reference point. The geodesic distance was then calculated, and the cumulative distance associated with the nearest reference point was assigned as an offset. This approach reduced nearest-neighbor search complexity to $O(\log n)$, providing both scalability and precision compared to ArcGIS Pro, which

proved computationally inefficient for large-scale DCVP data. Callout (i) in Figure 6.13(b) illustrates the snapping of the entry point to its corresponding reference coordinate. GPS points deviating significantly from the reference path were excluded, and journeys with large temporal gaps were segmented.

6.4.4 Traffic Characteristics Distribution

The speed distribution across the corridor shows a mean speed of 72.2 MPH with a standard deviation of 10.9 MPH. The vehicle speed distribution (Figure 6.15) is right-skewed, indicating generally steady freeway conditions with occasional slowdowns. Congestion severity thresholds were defined based on standard deviations below the mean: one standard deviation below the mean (61.3 MPH) for mild congestion, two standard deviations (50.3 MPH) for moderate to severe congestion, and three standard deviations (39.2 MPH) for extreme congestion. The travel distance distribution (Figure 6.14, callout (i)) shows a significant peak near 37 miles, indicating the predominance of inter-city travel between Tuscaloosa and Birmingham.

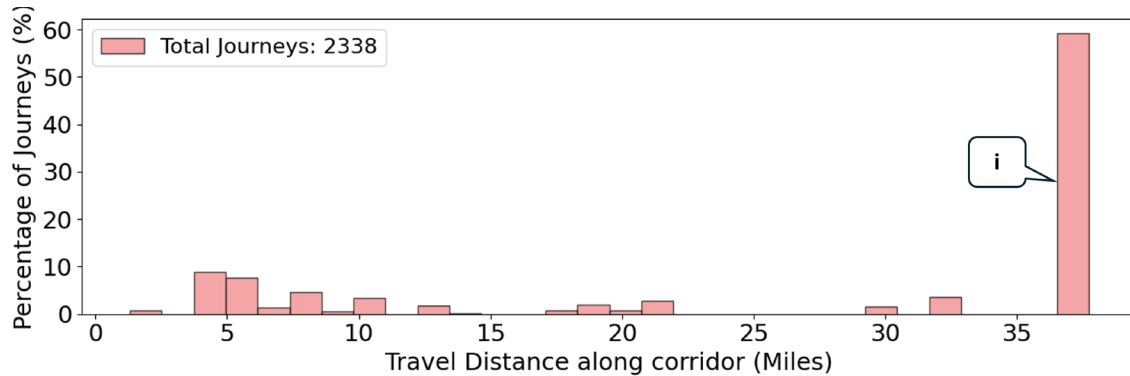


Figure 6.14 Observed distribution of vehicle travel distances along the corridor.

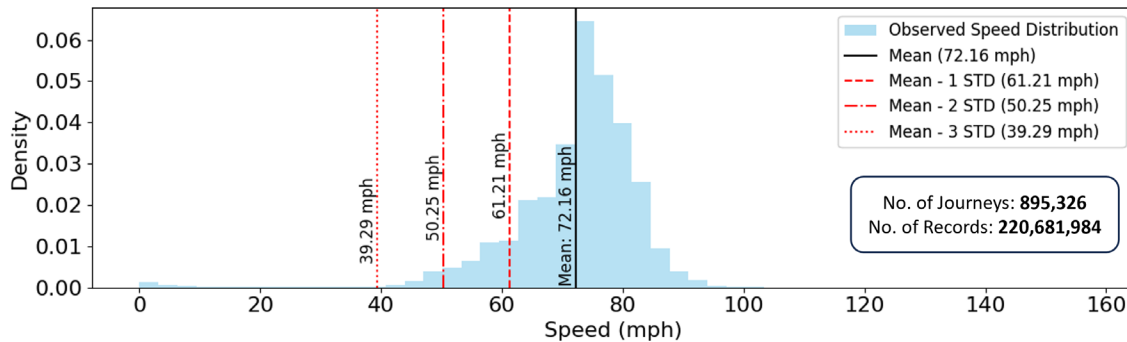


Figure 6.15 Vehicle speed distribution in the corridor.

6.4.5 Clustering Mechanism for Congestion Identification

A Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was applied to identify congestion clusters from the two-dimensional coordinates of timestamp (scaled to hours from midnight) and cumulative distance along the corridor, as visualized in Figure 6.16. The epsilon parameter was set between 0.3 and 1 (adjusted based on data availability) and MinPts between 5 and 10. The algorithm classifies core points based on neighborhood density and assigns

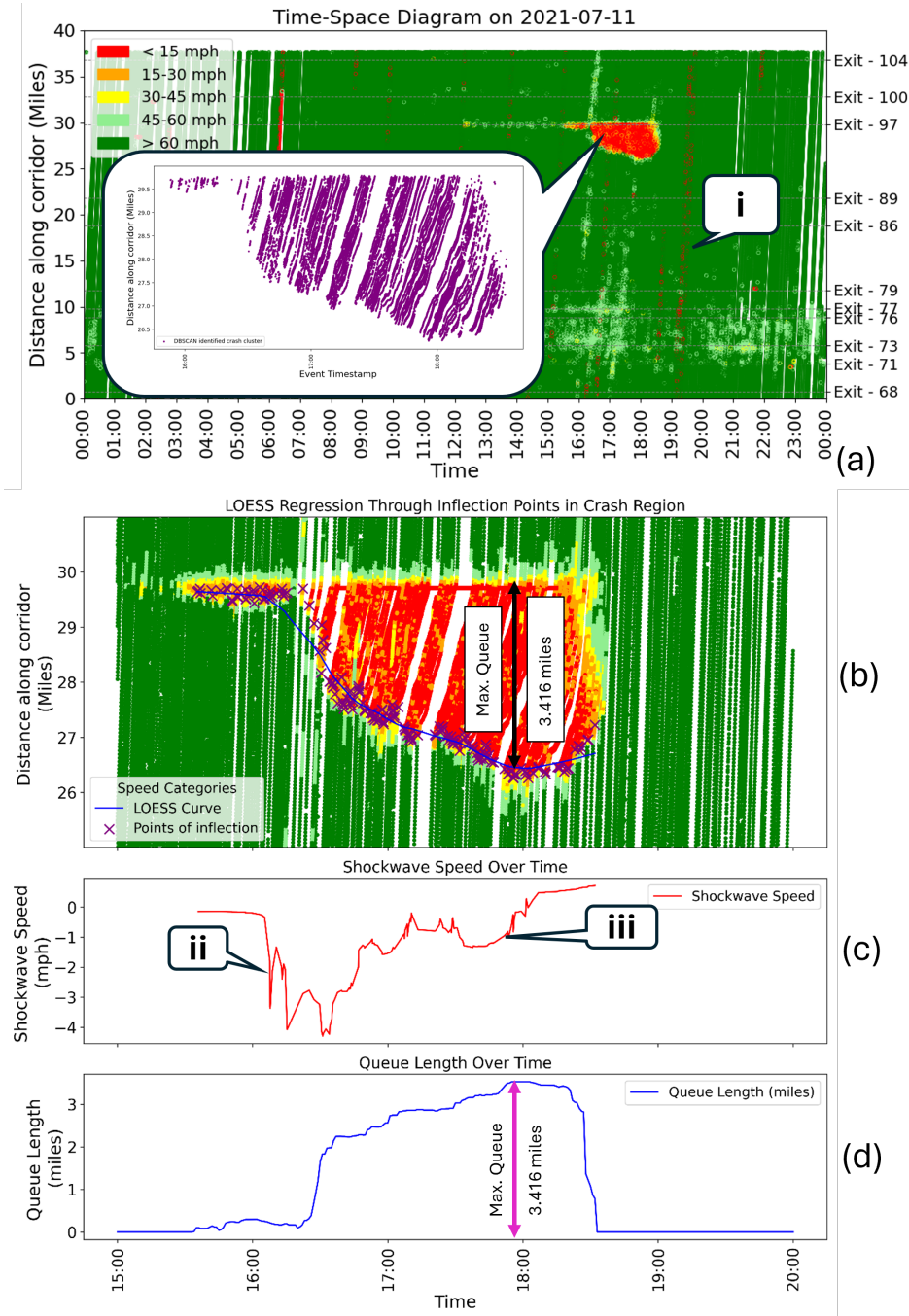


Figure 6.16 Congestion Clustering and Dynamics: (a) DBSCAN cluster identification, (b) LOESS regression through inflection points, (c) Accumulating and dissipating shockwave, and (d) Queue length over time.

all reachable points within distance epsilon to the same cluster. Vehicles that maintained consistently low speeds across the full corridor length, illustrated in callout (i) of Figure 6.16 (a), were excluded from the analysis. An IQR-based refinement step filters the largest identified cluster using the Interquartile Range method, retaining points within bounds of Q1 minus 1.5 times the

IQR and Q3 plus 1.5 times the IQR. This eliminates outliers and excludes slow-moving long-distance trips from being misclassified as congestion.

6.4.6 Shockwave Speed and Queue Length Analysis

Inflection points were identified where vehicle speeds dropped below a 15 MPH threshold, indicating congested conditions. A LOESS (Locally Estimated Scatterplot Smoothing) regression was applied to these inflection points, and the gradients of the resulting curve were used to estimate shockwave speeds (Figure 6.16 (c)). Negative gradients, as shown in callout (ii), indicate the onset of traffic shockwaves (backward propagation of congestion due to crashes or slowdowns), while positive gradients, highlighted in callout (iii), represent shockwave dissipation as speeds transition toward free-flow conditions. A LinearND interpolator was used to fill gaps in the DCVP dataset around DBSCAN-identified clusters, ensuring spatial-temporal continuity for precise measurement of queue lengths. As shown in Figure 6.16 (d), during major crash events, queue lengths reached up to 7 miles and recovery times extended to 6 hours.

6.4.7 Driver Behavior Analysis During Congestion

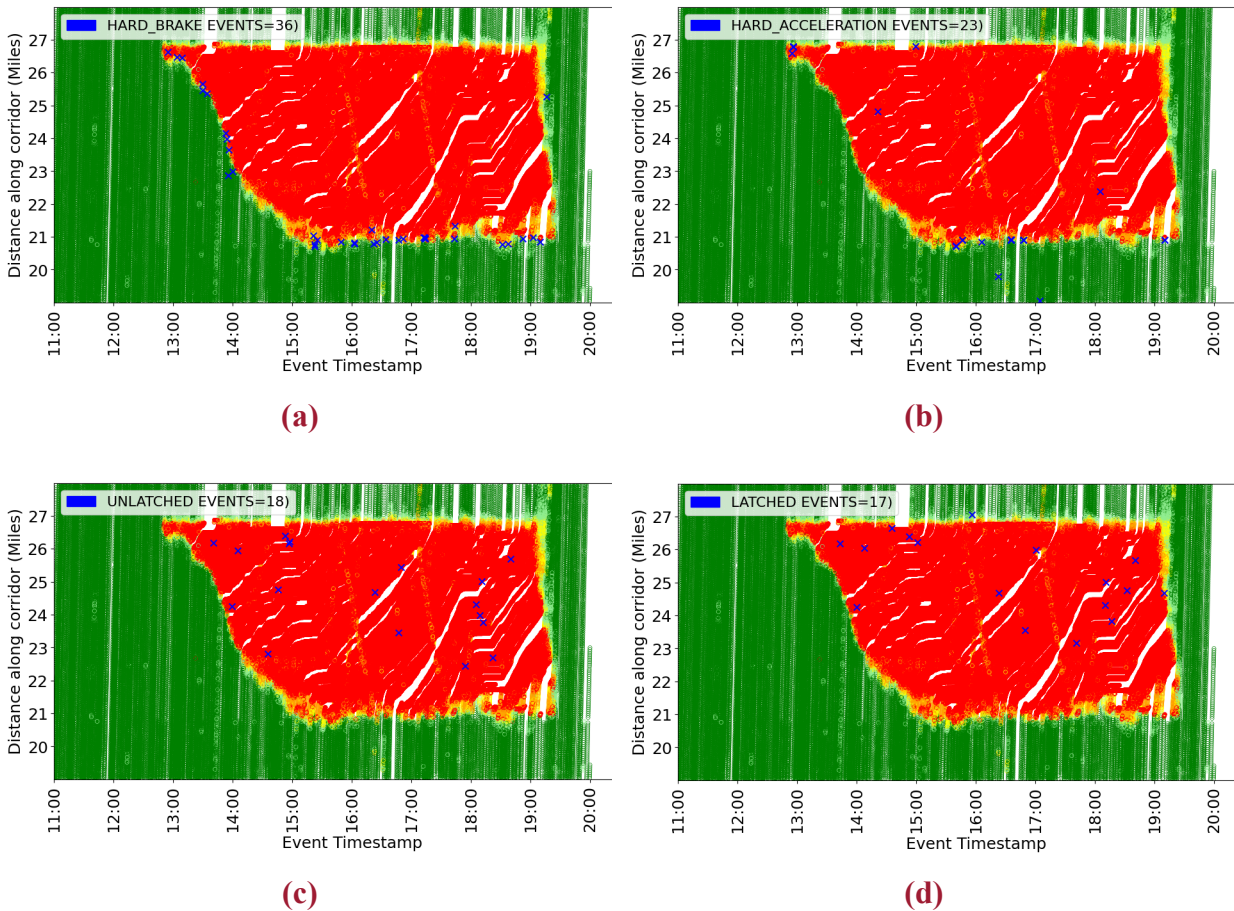


Figure 6.17 (a). Hard braking events during a congestion event. (b). Hard acceleration events during a congestion event. (c). Unlatched seatbelt events during a congestion event. (d). Latched seatbelt events during a congestion event.

acceleration, jerk) to analyze driver behavior during congestion events, as shown in Figure 6.17. Figure 6.17(a) depicts hard braking events concentrated at the leading edge of the congestion zone, where drivers encountered the back of the queue with abrupt slowdowns, indicating a high-risk zone for secondary crashes. Figure 6.17(b) shows hard acceleration events concentrated at the rear of congestion and during dissipation, reflecting driver impatience attempting to recover lost time. Figure 6.17(c) highlights seatbelt unlatching events observed during prolonged congestion, likely caused by driver discomfort, posing significant safety risks during potential secondary collisions. Figure 6.17(d) shows latched seatbelt events, reflecting that most drivers who unlatched their seatbelt re-latched once traffic resumed.

6.4.8 Route Alternatives and Behavioral Patterns

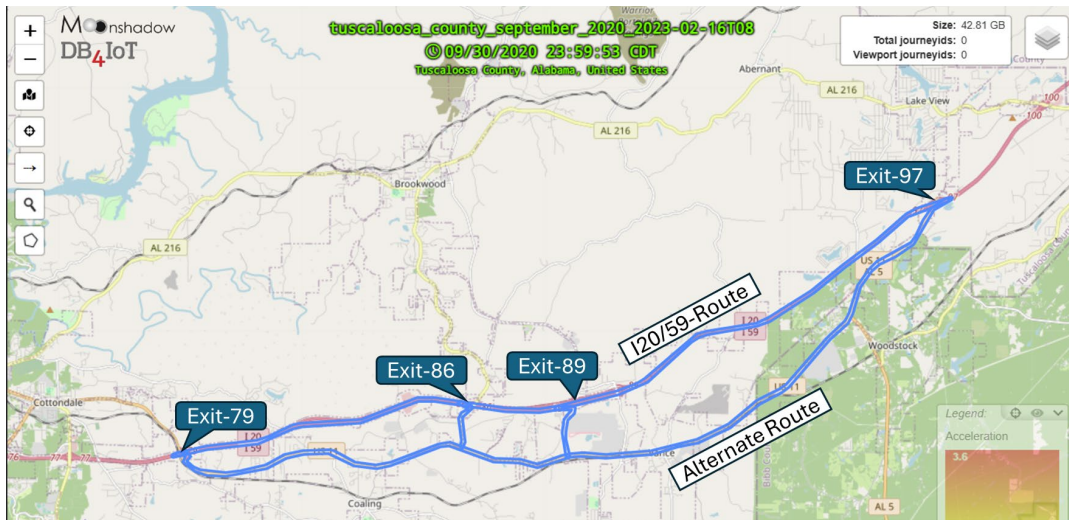


Figure 6.18 Route alternatives between Exit 79 and Exit 97 on the I-20/59 study corridor.



Figure 6.19 Vehicles taking the Exit 79 alternate route.

Analysis of route alternatives during congestion events revealed that drivers tend to select upstream exits for alternative routes when congestion queues extend beyond a certain point, as illustrated in Figures 6.18 and 6.19. Figure 6.18 illustrates the I-20/59 corridor, highlighting potential exit alternatives drivers use to bypass congestion. Figure 6.19 reflects how vehicles are distributed across various available routes in response to a crash downstream on the freeway. The decision is influenced by the extent of congestion ahead, the availability of feasible alternate paths, and traffic diversions.

6.4.9 Comparative Analysis of Primary and Alternate Routes

Figure 6.20 presents the traffic performance comparison between the primary I-20 route and an alternate route during three different incidents, revealing stark differences in traffic performance metrics. On the I-20 route, Day 2 exhibits a notable increase in average travel time, as highlighted

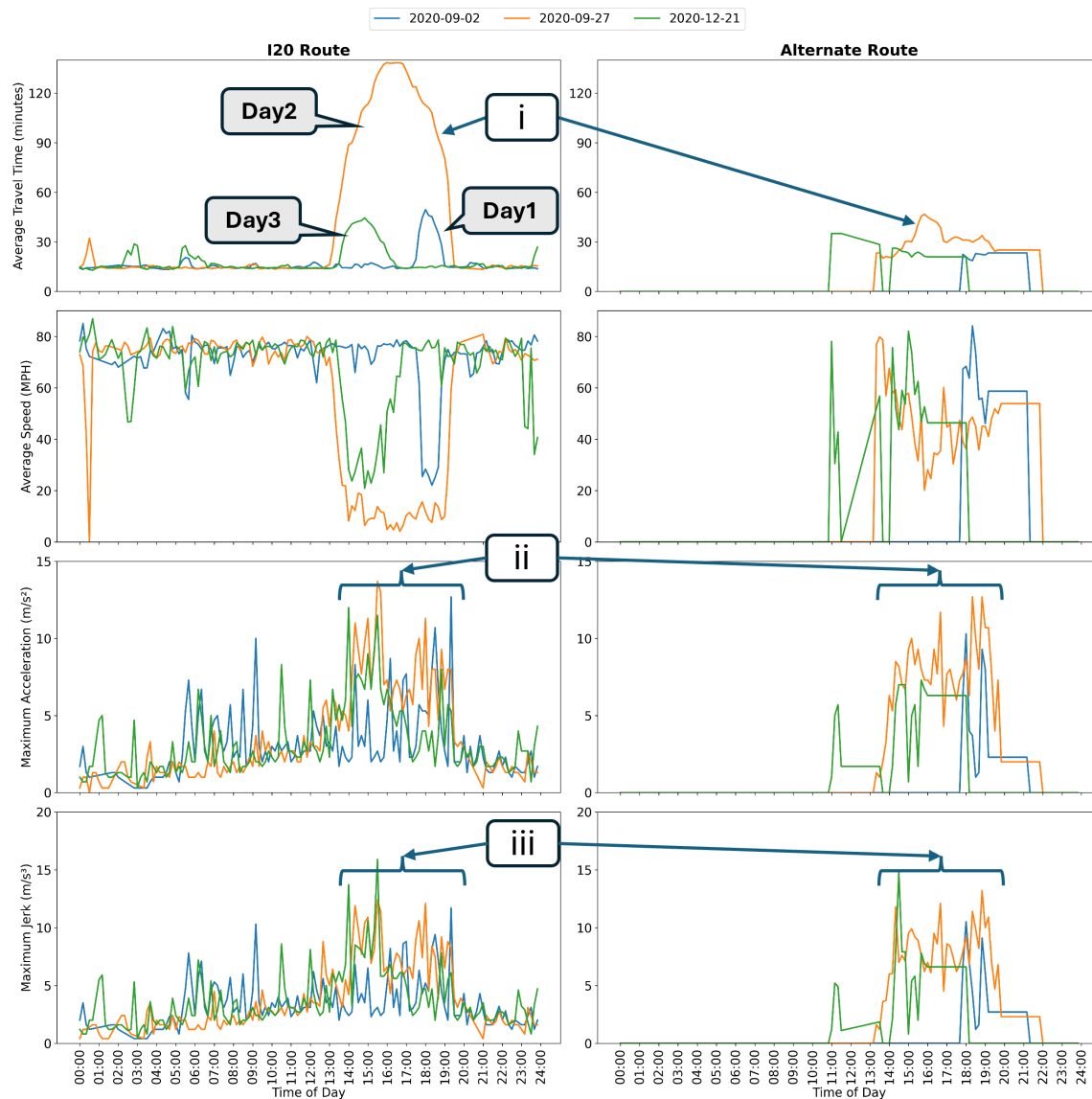


Figure 6.20 Comparison of traffic metrics on the I-20 route (left column) and alternate route (right column) during 3 different incidents.

by callout (i), indicating severe post-noon congestion. The maximum acceleration and maximum jerk values also spike on Day 2, as seen in callouts (ii) and (iii), indicating stop-and-go traffic due to congestion. On the alternate route, average travel time increased as more vehicles diverted from the I-20 route to avoid congestion, as seen in callout (i). Average speed on the alternate route decreased as higher traffic volumes led to slower conditions. Days 1 and 3 exhibited less severe patterns with lower travel times and higher speeds. Rerouting provided temporary relief on the freeway but caused congestion spillover on alternate routes.

6.4.10 Congestion Scoring Framework

A congestion scoring framework was developed to quantify congestion severity by integrating speed deficits over time and space using Simpson's rule, as shown in Figure 6.21(a). The time-space dataset was segmented into 10-minute intervals. Speed deficits were calculated as the difference between observed speeds and three congestion thresholds: Threshold 1 ($T1 = \text{Mean} \text{ minus } 1 \text{ STD} = 61.3 \text{ MPH}$) for moderate congestion, Threshold 2 ($T2 = \text{Mean} \text{ minus } 2 \text{ STD} = 50.3$

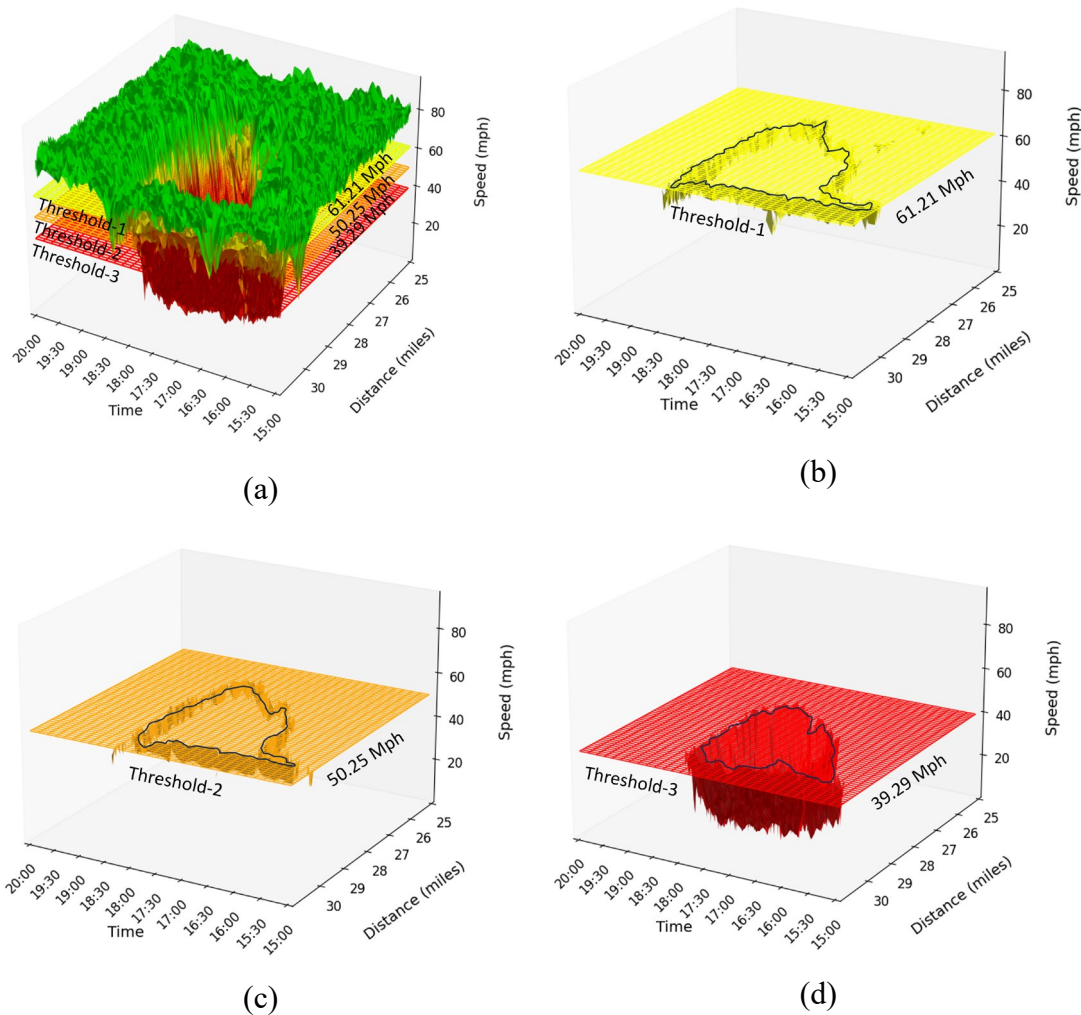


Figure 6.21 (a) 3D Time-space diagram. (b) Volume between Threshold 1 to 2. (c) Volume between Threshold 2 to 3 (d) Volume below Threshold 3.

MPH) for severe congestion, and Threshold 3 ($T3 = \text{Mean} - 3 \text{ STD} = 39.2 \text{ MPH}$) for extreme congestion. Figures 6.21(b) through 6.21(d) visualize congestion volumes across thresholds. Volumes in units of mph times miles times minutes were calculated for each severity level and then weighted: moderate congestion received 50 percent weight, severe congestion 30 percent, and extreme congestion 20 percent, reflecting the higher frequency and impact of moderate congestion while accounting for the critical but less frequent effects of severe and extreme congestion. Figure 6.22 illustrates this scoring methodology; as highlighted by callout (i), extreme congestion (below Threshold 3) dominates, reaching a maximum volume exceeding 2,000 mph times miles times minutes, indicating critical bottleneck conditions.

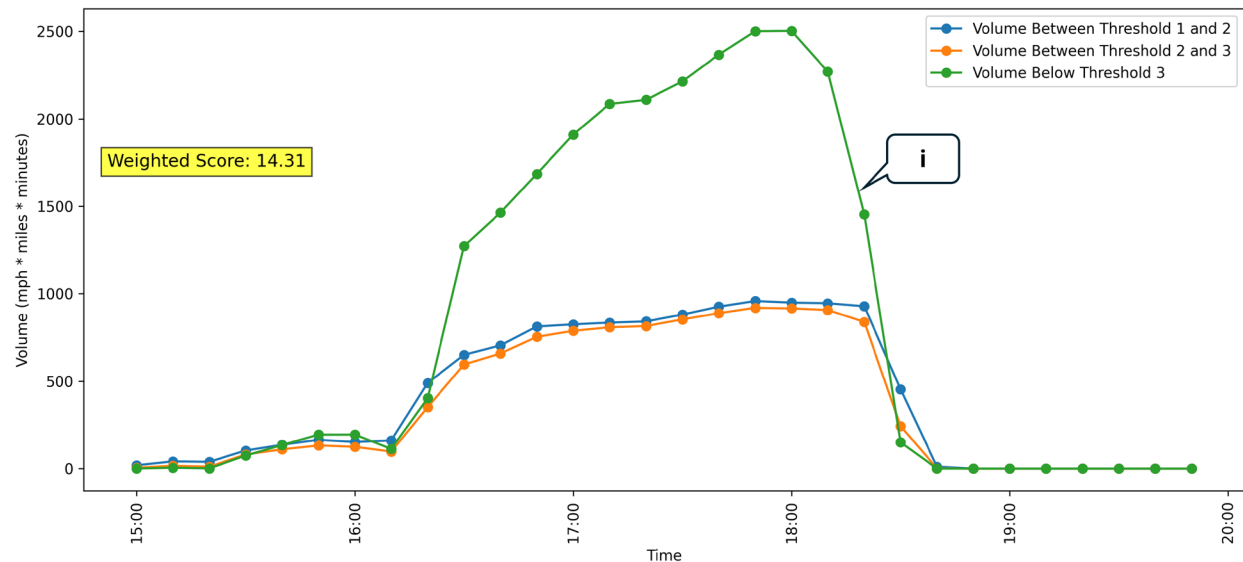


Figure 6.22 Volumes and weighted score.

6.4.11 Congestion Score Patterns and Unreported Disruptions

The congestion scoring framework was applied across the full study period from August 2020 to July 2021. As shown in Figure 6.23, daily congestion scores revealed variations in congestion severity across incidents. A sharp spike in November 2020 reached a peak score of 32.41, linked to a truck crash requiring more than 6 hours to clear per ATMS incident logs. December 2020 showed a peak congestion score of 23.49, and February 2021 reached 25.41. The year 2020 experienced higher and more frequent peaks in congestion scores compared to 2021, which saw relatively smoother traffic flow in early months. Figure 6.24 illustrates the frequency distribution of congestion scores for the study period. Most congestion scores were concentrated below 10, indicating that most congestion events dissipated quickly and suggesting effective traffic management strategies.

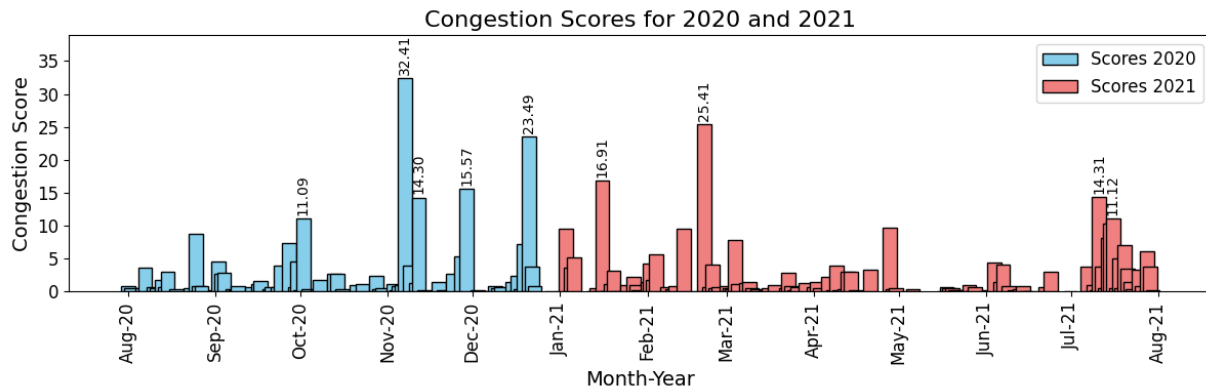


Figure 6.23 Time-series of congestion scores over the study period.

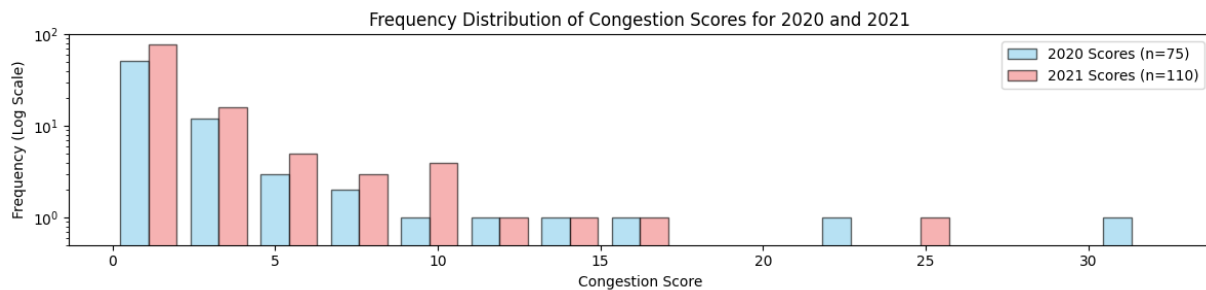


Figure 6.24 Frequency distribution of congestion scores.

Cross-validation of DCVP-derived congestion scores with ATMS (Advanced Traffic Management System) incident data revealed 74 significant congestion events detected exclusively through DCVP data that were not reported in ATMS incident records, as shown in Figure 6.25. In contrast, 111 events were validated against RTMC incident records. Several high-scoring DCVP-only events were clustered near Exit 73, a critical entry point where Tuscaloosa drivers merge onto the freeway toward Birmingham. The repeated occurrence of high-scoring events in this area suggests persistent bottlenecks arising from capacity constraints, merging inefficiencies, or localized demand surges. These findings indicate systemic traffic flow challenges not captured by traditional incident-based monitoring systems.

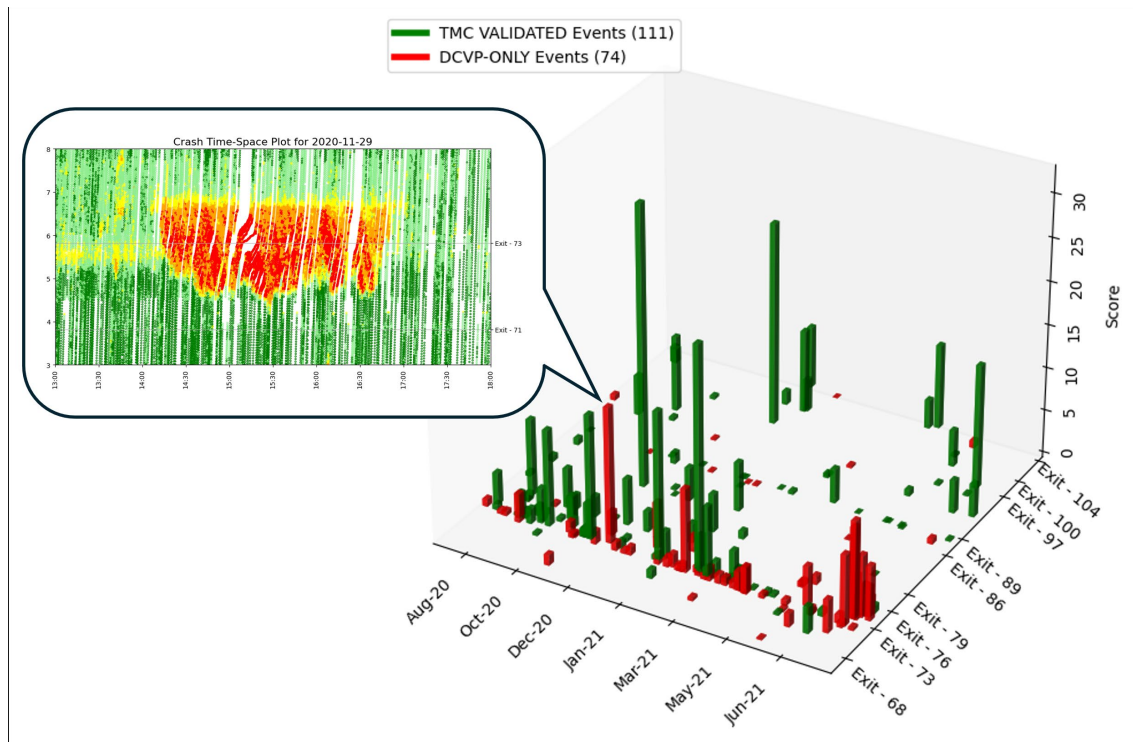


Figure 6.25 Congestion events visible in DCVP data but unreported in ATMS incident records.

6.4.12 Discussion

The freeway congestion analysis demonstrates that DCVP data offers near-continuous, high-resolution insights into congestion dynamics that overcome the limitations of traditional traffic datasets. The K-D Tree algorithm efficiently aligned DCVP data without relying on external datasets, while DBSCAN with IQR filtering effectively identified congestion clusters while handling noisy datasets. The LOESS regression on inflection points enabled precise shockwave speed estimation, and the LinearND interpolator ensured spatial-temporal continuity for queue length measurement. During major crash events, maximum queue lengths reached 7 miles with recovery times extending to 6 hours, demonstrating severe disruption potential.

Driver behavior insights from DCVP event data revealed hard braking concentrated at congestion leading edges (indicating secondary crash risk), hard acceleration at the rear and during dissipation (reflecting driver impatience), seatbelt unlatching during prolonged congestion (safety concerns), and illegal U-turns and median crossings (infrastructure gaps). Route diversion patterns showed that drivers selected upstream exits during severe congestion, and while the primary I-20 route experienced significant delays, alternate routes experienced spillover congestion from diversions. The congestion scoring framework with three threshold levels and Simpson's rule integration, combined with cross-validation against ATMS incident data, identified 74 significant unreported congestion events, highlighting the unique capability of DCVP data to uncover latent traffic challenges and inform data-driven traffic management.

7. Applications in Traffic Safety Analysis

Traditionally, researchers and practitioners use traffic crash data to evaluate and report on the safety performance at a site. A large number of unsafe instances (e.g., near-miss) that do not lead to a reportable event (i.e., crash) are not included in traffic safety analysis. With high-resolution CV data, unsafe non-crash events can be identified. For example, hard breaks may be treated as near-miss events. There are various reasons for hard brakes which may be related to the front traffic, dilemma zone at a signalized intersection, pedestrians/bicyclists, animals, and hidden or unexpected road conditions (sharp curve, lane end and traffic incident) without proper advanced warning. The CV data can be used to identify the traffic safety issues in the network and inform traffic safety improvement.

7.1 Case Study 1 – Driving Volatility-based Proactive Safety Management

Contributors: Jun Liu, Xiao Zou, Ziheng Zhang

7.1.1 Study Scope

Traditional traffic safety analysis relies heavily on reported crash data to evaluate the safety performance of roadways. While crash records are important, they capture only a small fraction of unsafe conditions, leaving out numerous near-miss events and risky driving behaviors that never escalate into reportable incidents. With the growing availability of high-resolution CV data, it is now possible to detect and analyze unsafe, non-crash events such as hard braking, which may indicate driver responses to hidden hazards, poor roadway design, or unpredictable traffic behavior.

The team leverages crowdsourced vehicle trajectory data to conduct location-based driving volatility analysis, a behavior-based safety assessment method that reflects real-time fluctuations in driving patterns. Volatility serves as a surrogate indicator of crash risk and provides transportation agencies with the ability to identify high-risk locations proactively, even before crashes occur. The team aims to demonstrate how these insights can inform Intelligent Transportation System (ITS) based highway safety management practices.

7.1.2 Methodology

Figure 7.1 presents the data integration framework used for location-based traffic safety analysis. Roadway geometry, including intersections and segments, was extracted from OpenStreetMap to define the analysis units. Crash records from the CARE crash database were geocoded to identify crash locations and linked with road networks to count the crash frequency at intersection-level and segment-level. Road environmental attributes were extracted from the Highway Performance Monitoring System (HPMS) database. CV data, including trajectory and speed information, were processed to calculate location-level driving volatility based on extreme maneuver behavior. These components were integrated to produce a comprehensive dataset that includes crash frequency and driving volatility at each location. The resulting data were used as inputs for statistical modeling to assess the relationship between behavioral driving patterns and crash risk across the roadway network.

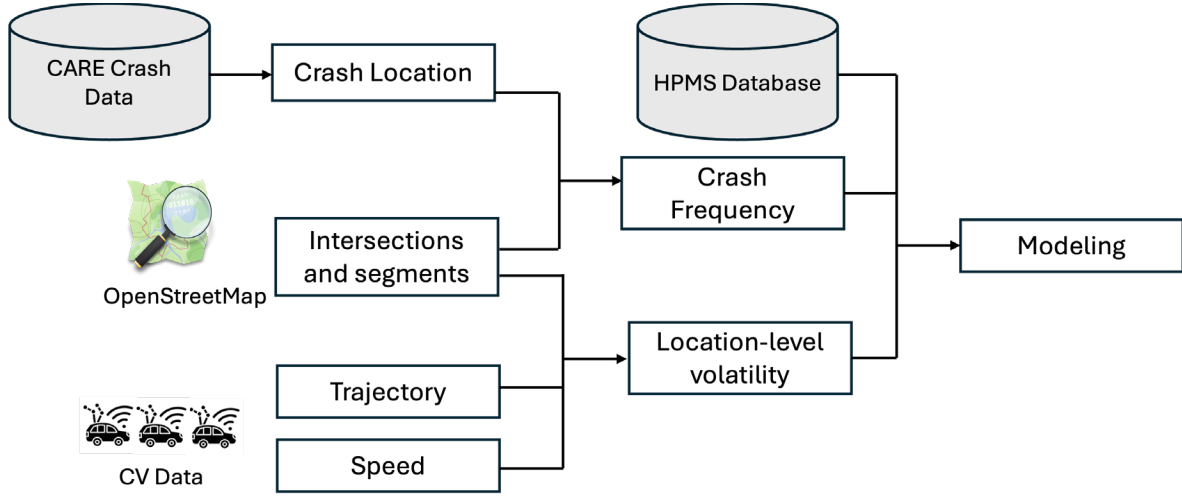


Figure 7.1 Data Integration Framework

Location-based driving volatility

Driving volatility (Fu et al., 2022; Gu et al., 2023; Liu et al., 2015; Liu & Khattak, 2020; Y. Zhang et al., 2023) is a concept of measuring the fluctuations and variations in a driver's behavior while operating a vehicle. It includes the degree of inconsistency or instability in driving patterns, which can be influenced by a variety of factors including environmental conditions, road characteristics, traffic density, and the driver's physiological and psychological state. High driving volatility is often characterized by abrupt changes in speed, sudden braking, rapid lane changes, and inconsistent steering. The team conducts driving volatility analysis at both segment-level and intersection-level. Location-based driving volatility can be determined by identifying the number of extreme maneuvers performed and then computing the percentage of time devoted to these maneuvers relative to the total maneuver time within a specific road segment or intersection.

Location – based Volatility Score %

$$= \frac{\text{Time of (Extreme Maneuvers} > \text{Threshold}_k)}{\text{Total Driving Time at a Location}} \times 100 \quad (1)$$

where, a location is a specific road segment or intersection. Specifically, at the intersection level, whether an intersection is signalized can significantly affect driving behaviors and crash frequencies. For example, signalized intersections typically include different patterns of vehicular stops and starts, which can influence the likelihood and severity of accidents compared to unsignalized ones. Therefore, it is important to differentiate between signalized and unsignalized intersections. The team quantified the driving volatility at signalized and unsignalized intersections, respectively. Threshold_k is the threshold used to identify extreme maneuvers at speed range k . It can be calculated by Equation 2:

$$\text{Threshold}_k = \mu_k + n \quad (2)$$

where, μ_k and σ_k represent the mean and standard deviation of the acceleration or deceleration at speed range k , respectively. n indicates the multiplier of the standard deviation.

Calculation of segment-level and intersection-level volatility

One of the most challenging tasks of this task is calculating location-based volatility at both the segment and intersection levels. The main difficulty lies in matching large-scale vehicle movement data—over 90 million driving records—with the corresponding road segments or intersections. Due to the substantial data volume, traditional spatial joins require considerable computational resources. Moreover, there is a discrepancy in the geographic referencing systems used: vehicle movement data utilizes coordinates, whereas road segments are identified using line-string references. Consequently, there is a need to develop an efficient method to manage these disparities and compute location-based volatility accurately. The framework for calculating location-based volatility is depicted in Figure 7.2.

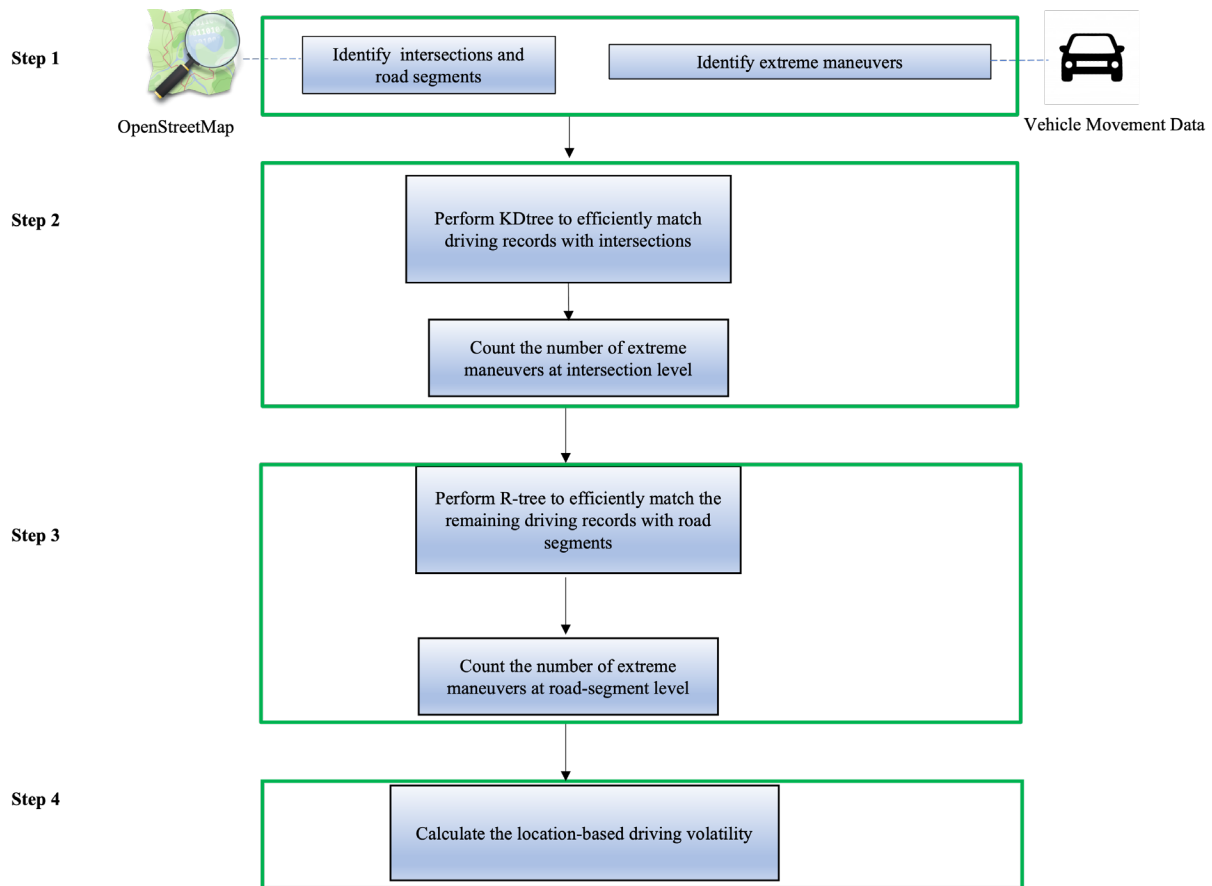


Figure 7.2 Procedure for calculating segment-level and intersection-level volatility

Step 1: Using OpenStreetMap (Haklay & Weber, 2008), 278 signalized and 2,255 unsignalized intersections, along with 11,292 road segments, were initially identified in Tuscaloosa. However, error checking revealed several duplicate intersection nodes, particularly among the signalized ones. To address this, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was employed. This clustering algorithm helps to group closely located points and retain only one representative per cluster. After applying DBSCAN, 167 signalized and 2,025 unsignalized intersections were identified. Additionally, using Equation (2), whether a driving record indicates an extreme maneuver can also be determined.

Step 2: Based on the identified intersections, a K-D Tree algorithm was employed to efficiently match large-scale driving records to their nearest intersections. The K-D Tree is a highly effective spatial search method, particularly suited for large datasets. The detail of this algorithm is described in (Maneewongvatana & Mount, 1999). The general idea is that the kd-tree is a binary tree, each of whose nodes represents an axis-aligned hyperrectangle. Each node specifies an axis and splits the set of points based on whether their coordinate along that axis is greater than or less than a particular value. Consistent with previous studies (Gu et al., 2023; Kamrani et al., 2017), a threshold of 150 feet was used. This approach allowed for accurate counting of extreme maneuvers at each specific intersection after matching driving records to their nearest points of interest.

Step 3: R-tree algorithm was used to efficiently match large-scale vehicle movement data with corresponding road segments, which are geometrically represented as line strings. The R-tree, a spatial data structure, optimizes the querying process by using Minimum Bounding Rectangles (MBRs). An MBR is the smallest rectangle that fully contains a spatial object or group of objects, aligning with the coordinate axes. By grouping nearby objects within MBRs, the R-tree narrows down the search area, significantly reducing the computational load. This hierarchical organization allows the algorithm to rapidly identify relevant road segments by testing for overlaps with MBRs, rather than checking every individual segment, making it particularly effective in handling large datasets. Following this, the number of extreme maneuvers at the segment level was quantified.

Step 4: Location-based driving volatility at both segment and intersection levels can be calculated by Equation (1). After processing, 2192 intersection data and 8422 road segment data were used.

Modeling location-based volatility and crashes

Zero-Inflated Negative Binomial (ZINB) regression is often applied to model count data with excessive zeros. Preliminary descriptive statistics show that 71 percent of intersections and 83 percent of road segments had zero crashes. A ZINB model is a modified Negative Binominal regression model. It captures both the excess zero observations and the nonzero observations, by estimating two separate models simultaneously. The first-step model is a binary logit model for predicting the probability of zero crashes given a set of independent variables. The second-step model is a negative binomial regression model that estimates the crash frequencies. Formally, the probability function of ZINB model has two components:

$$\Pr(y_i = k) = \begin{cases} \pi_i + (1 - \pi_i)(1 + \alpha\mu_i)^{1/\alpha}, & k = 0 \\ (1 - \pi_i) \frac{\Gamma[(1/\alpha) + k]}{\Gamma(1/\alpha)\Gamma(k + 1)} \left[\frac{1/\alpha}{(1/\alpha) + \mu_i} \right]^{1/\alpha} \left[\frac{\mu_i}{(1/\alpha) + \mu_i} \right]^k, & k > 0 \end{cases} \quad (3)$$

where,

π_i = predicted probability of zero crashes at location i

μ_i = predicted crash frequency at location i

y_i = observed crash frequency at location i

k = possible crash frequency at a location i

α = over-dispersion parameter

The coefficients of explanatory variables can be estimated through predicting the π_i and μ_i by

$$\pi_i = \Pr(\mu_i = 0) = \frac{\exp(X_i\beta_1)}{1 + \exp(X_i\beta_1)} \quad (4)$$

$$\mu_i = \exp(X_i\beta_2 + \varepsilon_i) \quad (5)$$

where,

X_i = intersection or road segment i's explanatory variables (e.g., location-based volatility and speed limit)

β_1 = a vector of estimated coefficients from binary logit regression model

β_2 = a vector of estimated coefficients from the negative binomial regression model

$\exp(\varepsilon_i)$ = a gamma-distributed error term.

The Vuong test will be performed to examine whether ZINB regression model is better than NB regression model (Greene, 1994; Vuong, 1989).

$$Vuong = \frac{\sqrt{N}\bar{m}}{S_m} \quad (6)$$

where, N is the number of observations, $\bar{m}$ is the mean and S_m is the standard deviation of

$$m_i = \ln \frac{\widehat{Pr}_1(y_i = k)}{\widehat{Pr}_2(y_i = k)} \quad (7)$$

where, $\widehat{Pr}_1(y_i = k)$ and $\widehat{Pr}_2(y_i = k)$ are predicted probabilities of $(y_i = k)$ using the negative binomial (NB) and ZINB models. If *Vuong* is larger than 1.96, ZINB model is better than NB model. If *Vuong* is smaller than -1.96, NB model fits better. If $|Vuong|$ is smaller than 1.96, neither is preferred.

To further evaluate the role of driving volatility, the team also developed models without the volatility variable to compare their performance. These comparisons allowed us to determine whether including volatility significantly improves the model's ability to explain safety outcomes.

In addition to the ZINB model, the team also developed an Ordinary Least Squares (OLS) model using driving volatility as the dependent variable. The OLS equation can be expressed as follows:

$$Volatility = \alpha + \beta X \quad (8)$$

where, α is the model's intercept; β is the vector of estimated coefficients for explanatory variables. This model explores whether driving volatility could be a useful predictor in cases where crash data is not available. If driving volatility can indicate risk levels, it may serve as a practical alternative or complement to traditional crash data for assessing road safety. This would be particularly valuable in areas with limited crash records since it can help identify and mitigate potential safety risks based on observed driving behaviors.

7.1.3 Results

Measured location-based driving volatility

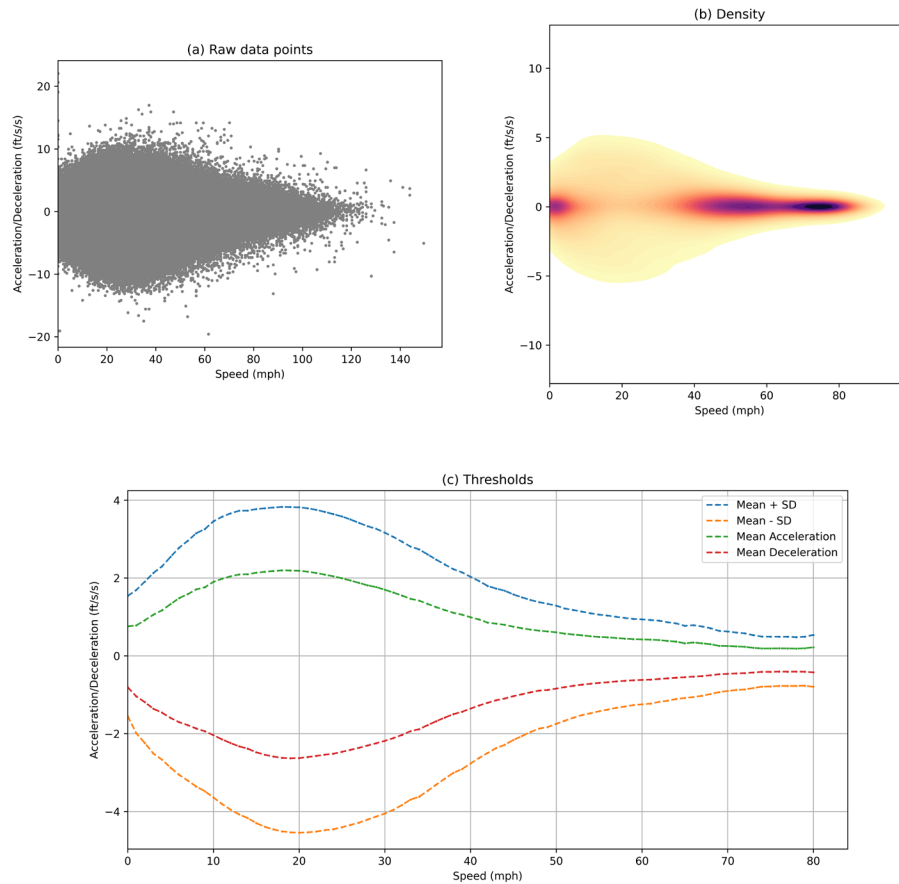


Figure 7.3 Extends of accelerations and decelerations across speeds

Figure 7.3 presents three plots derived from vehicle movement data. Figure 7.3-a displays the density patterns of driving records across all road types—including intersections, city roads, and highways. It highlights a notable trend where speeds below 30 mph often show greater speed fluctuations compared to those above 60 mph. Figure 7.3-b categorizes the density of driving records by acceleration levels across different speeds, identifying three predominant driving conditions: near-zero speeds typical at the start, end, and at intersections of trips; speeds ranging from 45 to 55 mph on city roads; and speeds around 70 mph on interstate highways. Figure 7.3-c presents the mean acceleration/deceleration and thresholds (mean + standard deviation for acceleration and mean - standard deviation for deceleration) for each speed range (every 1 mph). Peaks in mean accelerations and decelerations are observed at approximately 18 mph and 20 mph, respectively. Beyond 70 mph, both acceleration and deceleration magnitudes diminish and stabilize, with minimal speed changes, likely influenced by the 70-mph speed limit on most Tuscaloosa highways and the widespread use of cruise control for maintaining consistent speeds.

The location-based driving volatility at both intersection and road segment level was calculated. Some locations have limited number of driving records. The more data records a location has, the more accurately the driving performance in a location can be reflected. The team only reported location-based volatility for intersections or road segments that have more than 500 valid records. More driving records are needed for a complete picture of location-based driving volatility. Figure 7.4 shows the location-based volatility at the intersection level, where a larger red circle indicates higher volatility. According to Google Street View, the intersection exhibiting the most significant driving volatility features a notably complex configuration, characterized by multiple merging lanes where a smaller road converges with a main thoroughfare. This complexity is likely a contributing factor to the observed volatility. Figure 7.5 shows location-based volatility at the segment level, with varying colors representing different levels of volatility. The segments exhibiting the highest volatility are located on a city highway, characterized by a left turn lane that lacks traffic signals. This layout may prompt drivers to engage in "extreme" maneuvers, where speed and acceleration significantly change.

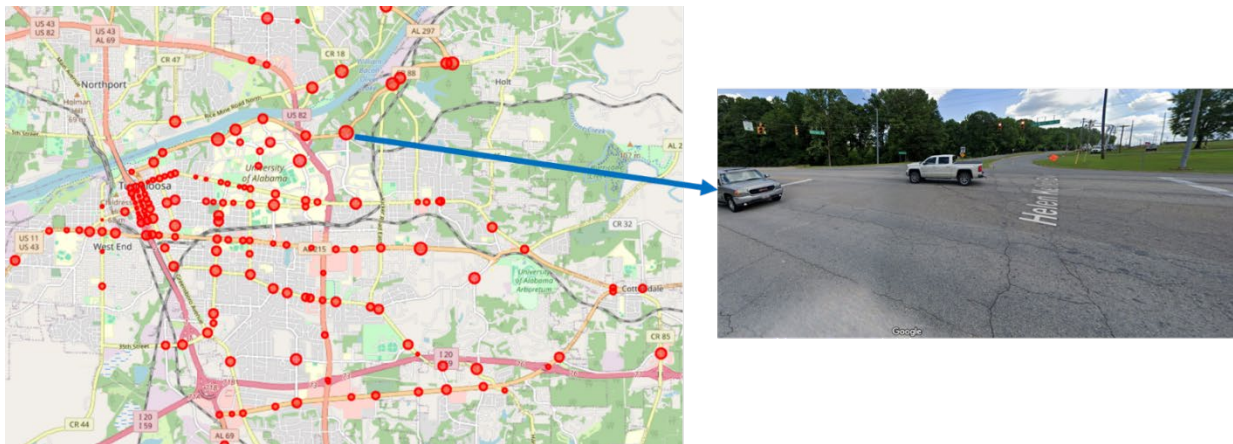


Figure 7.4 Location-based driving volatility at intersection level

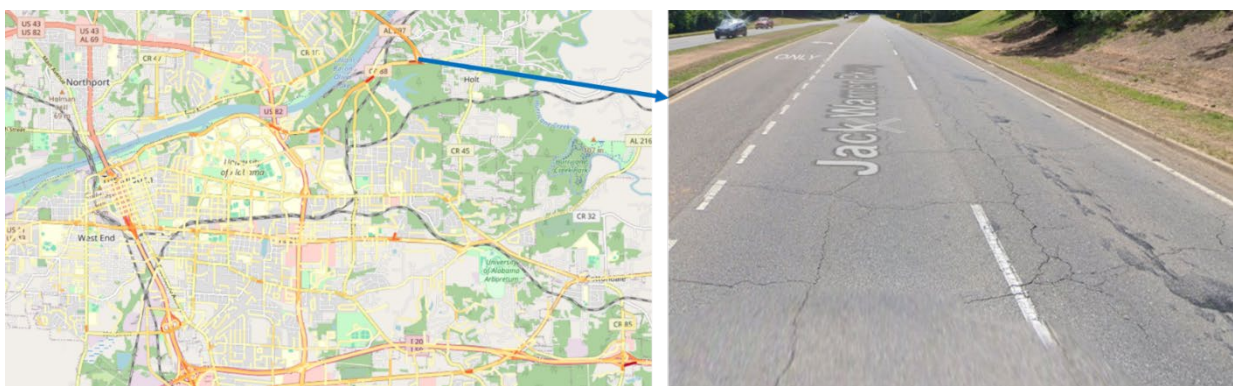


Figure 7.5 Location-based driving volatility at segment level

Modeling Results

Table 7.1 and Table 7.2 show the descriptive statistics of variables including in the model at intersection level and segment level, respectively. The mean crash frequency of all signalized and unsignalized intersection in Tuscaloosa in 2021 is 8.78 and 0.52, respectively. As expected, the crash frequency at signalized intersections is significantly higher than at unsignalized intersections. This finding is consistent with previous studies (Abdel-Aty & Haleem, 2011; Kamrani et al., 2017). The reason for this difference may be due to the increased traffic volume and complexity of maneuvers at signalized intersections, which can lead to more opportunities for collisions. Additionally, the mean crash frequency at road segments is 0.43, which is lower than the crash frequencies at both signalized and unsignalized intersections. This lower mean crash frequency at road segments may be due to the generally simpler traffic conditions and fewer conflict points compared to intersections. Regarding location-based driving volatility, the average volatility score

Table 7.1 Descriptive statistics of variables at the intersection level

		Signalized (N=167)				Unsignalized (N= 2025)			
Variable		Mean	S.D.	Percentage	Freq.	Mean	S.D.	Percentage	Freq.
Crash Frequency (dependent variable)		8.78	10.22			0.52	1.69		
Location-based driving volatility score (%)		23.04	8.48			8.93	8.17		
Annual average daily traffic (AADT, per 1000 vehicles) on major road		17.31	11.24			10.93	9.86		
Annual average daily traffic (AADT, per 1000 vehicles) on minor road		11.69	8.42			8.27	7.39		
Number of lanes on major road	<=2			39.52%	66			67.01%	1357
	3~4			54.49%	91			27.21%	551
	>=5			5.99%	10			5.78%	117
Number of lanes on minor road	<=2			60.48%	101			80.40%	1628
	3~4			37.13%	62			16.74%	339
	>=5			2.40%	4			2.86%	58
Speed limit on major road (mph)	<=35			13.17%	22			21.73%	440
	40~50			19.16%	32			13.48%	273
	>=55			5.39%	9			3.41%	69
	Unknown			62.28%	104			61.38%	1243
Speed limit on minor road (mph)	<=35			11.98%	20			16.89%	342
	40~50			5.39%	9			7.41%	150
	>=55			1.20%	2			1.28%	26
	Unknown			81.44%	136			74.42%	1507

Table 7.2 Descriptive statistics of variables at segment level

Variable		Mean	S.D.	Percentage	Frequency
Crash Frequency (dependent variable)		0.43	3.48		
Road segment length (m)		171.32	199.61		
Location-based driving volatility score (%)		6.71	9.18		
Annual average daily traffic (AADT, per 1000 vehicles)		10.55	9.86		
Lane width (ft)	<=11			0.52%	44
	12~13			1.70%	143
	>=14			0.20%	17
	Unknown			95.58%	8218
Median type	No median			1.89%	159
	Unprotected			0.21%	18
	Curbed			0.20%	17
	Positive Barrier – rigid			0.07%	6
	Positive Barrier – semi-rigid			0.05%	4
	Unknown			97.58%	8218
Number of lanes	<=2			71.05%	5984
	3~4			24.31%	2047
	>=5			4.64%	391
	<=35			16.67%	1404
Speed limit (mph)	40~50			10.73%	904
	>=55			2.60%	219
	Unknown			70.00%	5895
Signalized intersections on two sides	Neither			89.90%	7571
	One			8.89%	749
	Both			1.21%	102

at signalized intersections, unsignalized intersections, and road segments is 23.04%, 8.93%, and 6.71%, respectively. This indicates that signalized intersections have the highest level of driving volatility, likely due to the frequent stops and starts required by traffic signals, as well as the complex interactions between vehicles, pedestrians, and cyclists. Unsignalized intersections, while still exhibiting notable volatility, have lower scores due to simpler traffic flow and fewer forced stops, which reduce extreme maneuvers. Road segments show the lowest driving volatility, reflecting the generally smoother and more predictable traffic patterns on these stretches, where vehicles maintain a more consistent speed and direction. Other variables, such as Average Annual Daily Traffic (AADT), the number of lanes, and speed limits, are also considered in the model. Descriptive analysis highlights the importance of considering driving volatility in traffic safety analyses, as areas with higher volatility scores tend to correspond with higher crash frequencies.

Understanding the relationship between driving behavior and crash risk at different types of locations can inform targeted interventions and improvements in traffic management and infrastructure design to enhance road safety. [Table 7.3](#) and [Table 7.4](#) present the model results at the intersection level and segment level, respectively. Overall, both models demonstrate a good fit according to the Vuong test results, with values larger than 1.96, indicating that the ZINB model fits the data better than traditional negative binomial models. We also calculated the Variance Inflation Factor (VIF) to assess the multicollinearity among the independent variables. For each independent variable, VIF can be calculated by:

$$VIF(X_j) = \frac{1}{1-\hat{R}_j^2} \quad (8)$$

Where $\hat{R}_j^2$ is the square of multiple R^2 for the regression of X_j on the other covariates. [Table 7.3](#) and [Table 7.4](#) present the VIF values for independent variables. After removing variables with VIF larger than 10, all independent variables in both the segment-level and intersection-level model, have a VIF of less than 5, indicating insignificant collinearity between variables. Since the road feature data miss a lot of important information for some variables, such as speed limits and lane width, the p-values are not so good. Despite these challenges, by developing two sets of models, the incorporation of driving volatility into analysis proves beneficial. The ZINB models without driving volatility produce reasonable results, demonstrating expected relationships such as the positive correlation between AADT and crash frequencies. However, as shown in [Table 7.3\(i\)](#) and [Table 7.4\(i\)](#), models including driving volatility consistently yield higher R-squared values and lower Akaike Information Criterion (AIC) scores across both segment and intersection levels.

These results indicate that driving volatility adds significant information to the models, enhancing our understanding of roadway safety performance. Specifically, at the intersection level, the R-squared value improves from 0.121 to 0.147, and the AIC decreases from 4296 to 4175 with the inclusion of driving volatility. At the segment level, the R-squared value increases from 0.024 to 0.076, and the AIC drops from 4985 to 4730. These improvements indicate that driving volatility adds significant explanatory power to the models, enhancing our understanding of roadway safety performance.

To complement the ZINB model results, OLS model was also fitted, with driving volatility as the dependent variable. The purpose of this model was to explore factors contributing to driving volatility and assess its predictive utility, especially for locations where crash data may be unavailable. As shown in [Table 7.3\(ii\)](#) and [Table 7.4\(ii\)](#), the OLS model results were generally consistent with the findings from the ZINB model, reinforcing the relevance of driving volatility as an indicator of traffic safety. This alignment across models further supports the robustness of driving volatility as a useful measure for understanding crash risks.

While the p-value for driving volatility is statistically significant at the segment level, it is not consistently significant at the intersection level. This difference suggests that volatility is more critical for segment-level safety analysis. At intersections, where drivers frequently adjust their behavior in response to signals or signage, driving behavior changes more dynamically due to expected factors such as hard braking, rapid accelerations, and frequent stops. These instantaneous changes can obscure the relationship between volatility and crash risk, making volatility a less reliable predictor at intersections. In contrast, on road segments, which lack these frequent controls,

Table 7.3 Model Results at Intersection Level**(i) ZINB model results (y = crash frequency)**

Variable		VIF	Model with volatility				Model without volatility			
			Negative Binomial (count model)		Logit (zero inflated model)		Negative Binomial (count model)		Logit (zero inflated model)	
			β	P-value	β	P-value	β	P-value	β	P-value
Intercept			-0.320	0.039	-1.430	0.008	-0.440	0.020	-0.863	0.164
Location-based driving volatility score (%)		1.220	0.016	0.779	-2.480	< 0.001	-	-	-	-
Annual average daily traffic (AADT, per 1000 vehicles) on major road		4.110	0.020	0.084	0.004	0.913	0.020	0.050	0.010	0.770
Annual average daily traffic (AADT, per 1000 vehicles) on minor road		3.550	-0.013	0.250	-0.074	0.038	-0.010	0.320	-0.050	0.080
Traffic control	Non - signalized	Base	Base		Base		Base		Base	
	Signalized	1.250	2.048	< 0.001	-4.490	0.746	2.050	< 0.001	-10.510	0.028
Number of lanes on the major road	<=2	Base	Base		Base		Base		Base	
	3~4	2.990	0.124	0.445	-0.040	0.931	-0.010	0.950	-0.840	0.053
	>=5	3.760	0.370	0.441	1.060	0.477	0.623	0.148	0.710	0.390
Number of lanes on the minor road	<=2	Base	Base		Base		Base		Base	
	3~4	2.880	0.140	0.422	-0.469	0.393	0.240	0.210	0.030	0.050
	>=5	3.350	0.900	0.111	0.462	0.770	0.710	0.200	0.190	0.180
Speed limit on major road (mph)	<=35	Base	Base		Base		Base		Base	
	40~50	2.970	-0.410	0.095	0.620	0.432	-0.480	0.060	0.430	0.564
	>=55	1.870	0.480	0.200	1.330	0.252	0.430	0.210	0.630	0.380
	Unknown	3.450	0.250	0.189	1.300	0.056	0.110	0.600	0.490	0.306
Speed limit on minor road (mph)	<=35	Base	Base		Base		Base		Base	
	40~50	2.870	0.080	0.807	0.900	0.334	0.230	0.510	1.120	0.257
	>=55	1.620	-1.700	0.037	-0.280	0.893	-1.610	0.050	0.670	0.751
	Unknown	3.330	0.040	0.862	-0.340	0.668	0.330	0.170	0.673	0.320
Summary Statistics	Vuong test = 5.62 Number of Observations: 1734									
Model Comparison	Pseudo R-square	0.147					0.121			
	Log Likelihood	-2056.700					-2119.500			
	LL at null	-2412.300					-2412.300			
	AIC	4175.440					4296.960			

(ii) OLS model results (y = location-based driving volatility)

Variable		VIF	β	p-value
Intercept			9.842	< 0.001
Annual average daily traffic (AADT, per 1000 vehicles) on major road		4.110	-0.634	0.175
Annual average daily traffic (AADT, per 1000 vehicles) on minor road		3.550	0.024	0.952
Traffic control	Non - signalized	Base	Base	
	Signalized	1.250	12.860	< 0.001
Number of lanes on the major road	<=2	Base	Base	
	3~4	2.990	1.826	0.021
	>=5	3.760	-1.010	0.609
Number of lanes on the minor road	<=2	Base	Base	
	3~4	2.880	-0.108	0.900
	>=5	3.350	-1.005	0.609
Speed limit on major road (mph)	<=35	Base	Base	
	40~50	2.970	-1.476	0.153
	>=55	1.870	1.450	0.360
	Unknown	3.450	-0.161	0.837
Speed limit on minor road (mph)	<=35	Base	Base	
	40~50	2.870	1.337	0.320
	>=55	1.620	-4.198	0.109
	Unknown	3.330	1.330	0.122
Summary Statistics	Number of Observations: 1734			
	R-squared: 0.178			
	Log Likelihood: -6186.2			
	AIC: 12400			

Table 7.4 Model Results at Segment Level

(i) ZINB model results (y = crash frequency)

Variable		VIF	Model with volatility				Model without volatility			
			Negative Binomial (count model)		Logit (zero inflated model)		Negative Binomial (count model)		Logit (zero inflated model)	
			β	p-value	β	p-value	β	p-value	β	p-value
Intercept			-1.120	0.304	-7.200	0.002	-0.370	0.160	1.820	<0.001
Location-based driving volatility score (%)		1.105	0.220	0.009	-4.080	<0.001	-	-	-	-
Road segment length(m)		1.015	0.720	<0.001	0.200	0.024	0.750	<0.001	0.270	0.040
Annual average daily traffic (AADT, per 1000 vehicles)		2.242	0.020	0.006	-0.010	0.439	0.020	0.017	-0.190	0.002
Lane width (ft)	<=11	1.030	1.040	0.384	8.980	0.578	0.920	0.490	-0.470	0.750
	12~13	1.820	-0.800	0.404	6.670	0.678	-1.410	0.160	-1.420	0.380
	>=14	Base	Base		Base		Base		Base	
	Unknown	>10								
Median type	No median	Base	Base		Base		Base		Base	
	Unprotected	1.450	0.350	0.782	4.150	0.234	2.070	0.120	1.000	0.830
	Curbed	1.220	1.950	0.142	-3.020	0.310	2.900	0.040	1.400	0.740
	Positive Barrier – rigid	1.230	1.320	0.376	-9.600	0.934	2.500	0.140	0.650	Nan
	Positive Barrier – semi-rigid	1.170	0.300	0.864	-8.800	0.934	1.040	0.600	0.980	0.900
	Unknown	>10								
Number of lanes	<=2	Base	Base		Base		Base		Base	
	3~4	1.570	-0.560	0.003	-0.610	0.118	-0.480	0.005	0.640	0.400
	>=5	1.874	-0.040	0.899	0.290	0.701	0.170	0.640	1.020	0.390
Speed limit (mph)	<=35	Base	Base		Base		Base		Base	
	40~50	1.750	-0.650	0.070	0.810	0.201	-0.710	0.070	0.670	0.360
	>=55	1.370	-0.720	0.141	0.670	0.513	-0.700	0.180	-2.160	0.090

	Unknown	1.840	- 0.450	0.063	- 0.600	0.144	- 0.730	0.007	- 5.660	0.002
Signalized intersections on two sides	Neither	Base	Base		Base		Base		Base	
	One	1.080	- 0.300	0.066	- 1.540	0.069	0.280	0.100	0.080	0.920
	Both	1.060	0.380	0.346	0.120	0.887	0.440	0.200	- 9.200	0.950
Summary Statistics	Vuong test = 4.490 Number of Observations: 3967									
Model Comparison	Pseudo R-square	0.076					0.024			
	Log Likelihood	-2326.500					-2456.500			
	LL at null	-2517.600					-2517.600			
	AIC	4730.960					4985.030			

(ii) OLS model results (y = location-based driving volatility)

Variable		VIF	β	p-value
Intercept			1.368	< 0.001
Road segment length(m)		1.015	0.001	< 0.001
Annual average daily traffic (AADT, per 1000 vehicles)		2.242	0.002	0.157
Lane width (ft)	<=11	1.030	0.134	0.426
	12~13	1.820	-0.034	0.764
	>=14	Base	Base	
	Unknown	>10		
Median type	No median	Base	Base	
	Unprotected	1.450	1.039	< 0.001
	Curbed	1.220	0.214	0.409
	Positive Barrier – rigid	1.230	0.926	0.006
	Positive Barrier – semi-rigid	1.170	-0.239	0.550
	Unknown	>10		
Number of lanes	<=2	Base	Base	
	3~4	1.570	-0.105	0.001
	>=5	1.874	0.168	0.022
Speed limit (mph)	<=35	Base	Base	
	40~50	1.750	0.261	< 0.001

	≥ 55	1.370	0.432	< 0.001
	Unknown	1.840	0.370	< 0.001
Signalized intersections on two sides	Neither	Base	Base	
	One	1.080	0.579	< 0.001
	Both	1.060	0.298	< 0.001
Summary Statistics	Number of Observations: 3967			
	R-squared: 0.116			
	Log Likelihood: -4433.1			
	AIC: 8898			

driving behavior tends to be more stable and reflective of overall volatility, making it a more meaningful predictor of crash frequency.

According to the results from the count model (negative binomial), which indicate how predictors affect the expected crash frequency, the coefficients of driving volatility are positive at both intersection level and segment level, suggesting that crashes are more likely to occur at locations with higher volatility. This finding supports the hypothesis that higher driving volatility, characterized by more frequent and severe maneuvers, increases the likelihood of crashes. In addition to driving volatility, AADT also shows a positive relationship with crash frequency in both models. Higher AADT indicates increased traffic volume, which elevates the likelihood of vehicle interactions and potential conflicts. This increased interaction can lead to more frequent collisions, as higher traffic volumes exacerbate the complexity of traffic flow and the chances of driver error. At the intersection level, signalized control is positively associated with crash frequency. This positive relationship may be due to the complex traffic patterns and higher vehicle interactions inherent at signalized intersections. Signalized intersections often manage higher traffic volumes and a mix of turning movements, pedestrian crossings, and other interactions, which can increase the potential for collisions despite the presence of traffic control devices. At the segment level, the influence of signalized intersections on crash frequency is also significant. If both ends of a road segment are signalized intersections, the coefficient for crash frequency is even more positive. This implies that road segments bookended by signalized intersections experience higher crash frequencies, likely due to the increased stop-and-go traffic and potential for abrupt maneuvers. The frequent stops at signalized intersections can lead to sudden acceleration and deceleration on the intervening road segment, creating a volatile driving environment prone to crashes.

In summary, the findings from both the ZINB and OLS models emphasize the value of incorporating location-based driving volatility into safety analyses, especially for segment-level analysis. Driving volatility provides additional explanatory power in understanding crash risks on road segments, where it is less influenced by frequent, momentary adjustments due to traffic controls, as seen at intersections. These results indicate that while volatility is a useful predictor of crash frequency overall, it is particularly effective in segment-level safety analysis where driving patterns are more stable and less disrupted by immediate external factors.

7.1.4 Summary and Conclusions

The team developed a behavior-based traffic safety analysis framework using large-scale connected vehicle (CV) data. Traditional crash-based methods often overlook near-miss events and non-crash driving behaviors that may indicate elevated safety risk. To address this limitation, the analysis quantified location-based driving volatility, defined by the frequency of extreme maneuvers such as hard braking and rapid acceleration, as a surrogate safety indicator. Volatility was calculated at both the segment and intersection levels. Statistical modeling integrated volatility metrics with traditional traffic and roadway characteristics such as AADT, lane count, and signal presence. Results showed that locations with higher driving volatility tended to experience more crashes, with segment-level volatility demonstrating particularly strong predictive power.

By identifying patterns of risky driving behavior in near real-time, the framework enables transportation agencies to shift from reactive crash-based analysis to proactive safety management. High-volatility locations can be prioritized for safety interventions—such as speed control measures, signal timing adjustments, or improved signage—before crashes occur. In summary, the use of connected vehicle data to assess driving volatility offers a scalable, forward-looking tool for improving roadway safety performance and advancing ITS-based traffic safety management strategies.

7.2 Case Study 2 – Statewide Correlation Analysis of Hard Braking Events and Road Traffic Crashes

Contributors: Praveena Penmetsa, Sunday Okafor, Steven Jones, Emmanuel Kofi Adanu

7.2.1 Study Scope

This case study presents a statewide correlation analysis of Wejo hard braking event (HBE) data and road traffic crashes across different roadway functional classes in Alabama. Traditional road safety analysis relies on historical crash records, which are inherently reactive and typically require three to five years of data to support statistically robust analysis. This study examines whether HBE from connected vehicles can serve as a proactive surrogate for crash data on a statewide basis, enabling more timely safety monitoring (Okafor et al., 2024).

The study used approximately six months of HBE data collected between August 2020 and June 2021 alongside crash data from the same periods. The analysis covered the entire state of Alabama using the 2017 HPMS roadway network shapefile, yielding 172,674 unique road segments. A total of 92,269 crashes and 5,669,194 HBE events were included in the correlation analysis. The study distinctly examined HBE-crash correlations separately across roadway functional classes, providing ALDOT with targeted guidance on where CVD-based screening is most applicable.

7.2.2 Methodology

Three datasets were integrated: crash data from the CARE database, HBE data from Wejo, and roadway inventory data from the 2017 HPMS shapefile. HBE is defined as deceleration below -2.67 m/s^2 (-8.75 ft/s^2). Both the HBE and crash data were geocoded in ArcGIS using latitude and longitude coordinates and spatially joined to roadway segments using a 100-foot search radius. After spatial joining, each segment carried counts of HBE and crashes that occurred during the study period.

Spearman's rank-order correlation was used as the primary statistical method, as it is appropriate for non-normally distributed data — a characteristic observed for both HBE and crash counts. For

comparison, Pearson correlation was also estimated. Correlation strength thresholds followed Evans (1996): very strong (0.80–1.00), strong (0.60–0.79), moderate (0.40–0.59), weak (0.20–0.39), and very weak (0.00–0.19). Analyses were conducted separately by functional class and by area type using the HPMS classification.

7.2.3 Results

Table 7.5 presents summary statistics of HBE and crashes by functional class. Other Principal Arterials exhibited the highest combined counts of HBE (2,643,447) and crashes (35,576), associated with frequent intersection conflicts and lack of access control.

Table 7.5 Summary Statistics of HBE and Crashes by Functional Class

Functional Class	Total Segment Length (ft)	HBE Count	Crash Count
Interstate	2,376.09	191,807	11,127
Other Freeways & Expressways	76.37	23,852	1,065
Other Principal Arterials	5,694.74	2,643,447	35,576
Minor Arterials	6,377.35	1,506,245	25,657
Major Collectors	15,867.34	1,303,060	18,826
Local Roads	7.60	783	18
Total	30,399.49	5,669,194	92,269

Table 7.6 presents the Spearman and Pearson correlation results by functional class. Based on the Spearman rank-order correlation results, a moderate correlation between HBE and crashes was observed on Other Principal Arterials ($\rho = 0.49$), Minor Arterials ($\rho = 0.42$), and Major Collectors ($\rho = 0.47$). Interstates and local roads exhibited weak correlations ($\rho = 0.24$ and 0.30 , respectively). All coefficients were statistically significant at the 0.05 level.

Table 7.6 Spearman and Pearson Correlation Coefficients by Functional Class

Functional Class	Spearman ρ	Spearman Strength	Pearson ρ	Pearson Strength
Interstate	0.24	Weak	0.19	Negligible
Other Freeways & Expressways	0.40	Moderate	0.30	Low
Other Principal Arterials	0.49	Moderate	0.79	High
Minor Arterials	0.42	Moderate	0.61	Moderate
Major Collectors	0.47	Moderate	0.50	Moderate
Local Roads	0.30	Weak	0.31	Low

The moderate correlation on arterials and collectors is consistent with the presence of intersection conflicts, stop-and-go traffic patterns, and frequent access points — conditions that generate both HBE and crash risk. On interstates, traffic is access-controlled with fewer conflict points, leading to a weaker relationship between HBE and crashes. Urbanized area results showed that Mobile, Birmingham, and Huntsville exhibited the strongest correlations among the urbanized areas on arterials and collectors. Rural areas consistently showed weaker correlations ($\rho = 0.26$ across all functional classes).

7.2.4 Summary and Conclusions

These findings suggest that HBE data can serve as a practical surrogate for crash-based safety screening on Other Principal Arterials, Minor Arterials, and Major Collectors across Alabama. HBE is not recommended as a crash surrogate for interstates due to the weak correlation observed. Transportation agencies can apply HBE-based screening on arterials and collectors to conduct more frequent safety reviews without waiting for multi-year crash data accumulation.

7.3 Case Study 3 – Integrating Hard Braking Event Data for Proactive Road Network Safety Screening

Contributors: Praveena Penmetsa, Sunday Okafor, Steven Jones, Emmanuel Kofi Adanu

7.3.1 Study Scope

This case study examines the potential of integrating Wejo HBE data as crash surrogates for proactive road network safety screening on arterial roadway segments in Tuscaloosa County, Alabama. Road network screening is traditionally the first step in road safety management programs, relying on historical crash data to identify sites that benefit from safety improvements. This approach is inherently reactive and requires approximately three to five years of crash records (FHWA, 2022; HSM, 2010). This study applied established network screening techniques from the 2010 Highway Safety Manual using both historical crash data and HBE from CVD to identify high-risk arterial segments (Okafor et al., 2025). The study used three months of HBE data (October–December 2020), yielding approximately 4.4 million unique events, and four years of crash records (2017–2020) comprising 278,518 spatially joined arterial crashes. The study area focused on Tuscaloosa County arterial segments, selected based on the highest spatial overlap between crash and HBE clusters identified through statewide spatial autocorrelation analysis.

7.3.2 Methodology

The methodology proceeded in three stages. First, global and local spatial autocorrelation (Moran's I) was estimated at the census tract level using K-nearest neighbor spatial weights ($k = 8$), with crash and HBE data normalized per 1,000 people using American Community Survey population estimates. Second, Spearman's rank-order correlation was conducted at the segment level to quantify the monotonic relationship between HBE and crashes by type and severity. Third, Getis-Ord G^*i spatial statistics were applied at the segment level to identify hotspots at 90%, 95%, and 99% confidence levels, separately for HBE and each crash type. The identified hotspots were then compared to determine the proportion of crash-based hotspots also captured by the HBE-based screening. Both crashes and HBE were geocoded and spatially joined to the 2017 HPMS arterial segment network using a 100-foot search radius. Segments shorter than 80.47 meters were excluded to avoid data coding errors.

7.3.3 Results

The global Moran's I statistics were 0.20 for crashes and 0.21 for HBE (both $p < 0.01$, based on 999 permutations), confirming statistically significant positive spatial autocorrelation — census tracts with high crash or HBE rates cluster together. Local spatial autocorrelation analysis identified Tuscaloosa County as a high-high cluster for both crashes and HBE.

The Spearman correlation results at the segment level are presented in Table 7.7 All crash types were significantly correlated with HBE at the 99% confidence level. The three strongest correlations were for all crashes ($\rho = 0.69$, strong), no-injury crashes ($\rho = 0.66$, strong), and rear-end crashes ($\rho = 0.62$, strong).

Table 7.7 Spearman Correlation — HBE vs. Crash Types at Segment Level (Tuscaloosa County Arterials)

Crash Type / Severity	Spearman ρ	Correlation Strength
All crashes	0.69	Strong
No-injury crashes	0.66	Strong
Rear-end crashes	0.62	Strong
Side impact 90°	0.52	Moderate
Side impact angled	0.51	Moderate
Severe (fatal and injury)	0.50	Moderate
Possible injury	0.49	Moderate
Sideswipe same direction	0.47	Moderate
Suspected minor injury	0.41	Moderate
Angle opposite direction	0.37	Weak
Head-on crashes	0.27	Weak
Single vehicle crashes	0.17	Very weak
Fatal crashes	0.11	Very weak

Table 7.8 presents the hotspot count by confidence level. HBE identified 216 total hotspots, substantially more than any crash type, reflecting its greater sensitivity. Table 7.9 shows that significant proportions of crash-based hotspots were concurrently identified as HBE hotspots. The highest overlap was for rear-end crashes (96.83%), followed by all crashes (86.25%), no-injury crashes (85.37%), and severe crashes (78.22%).

Table 7.8 Number of Hotspot Segments by Confidence Level

Category	90%	95%	99%	Total
HBE	40	46	130	216
All crashes	13	21	46	80
Rear-end crashes	14	12	37	63
No-injury crashes	21	16	45	82
Severe crashes	13	33	55	101

Table 7.9 Proportion of Crash-Based Hotspots Also Identified by HBE

Crash Type	Crash-Based Hotspots	Common Hotspots with HBE	Proportion (%)
All crashes	80	69	86.25
Rear-end crashes	63	61	96.83
No-injury crashes	82	70	85.37
Severe crashes	101	79	78.22

Table 7.10 Total Rank Difference (TRD) of Common Hotspots — Lower = Better Agreement

Crash Type	Top 10 TRD	Top 10 SD	Top 20 TRD	Top 20 SD	Top 30 TRD	Top 30 SD
All crashes	165	17.8	293	13.7	409	11.6
Rear-end crashes	152	17.7	250	13.0	376	13.3
No-injury crashes	166	18.1	304	14.0	425	13.2
Severe crashes	115	13.9	316	14.3	480	12.1

A ranking consistency analysis using Total Rank Difference (TRD) assessed how similar HBE and crash data ranked the common hotspot locations (Table 7.10). For the top 20 and top 30 hotspots, rear-end crashes showed the lowest TRD (250 and 376, respectively), indicating the strongest agreement in ranking between HBE and crash-based screening. For the top 10 hotspots, severe crashes had the lowest TRD (115). Average TRD decreased as the number of evaluated top sites increased for all crash types except severe crashes, indicating that ranking agreement improves as the hotspot list expands.

7.3.4 Summary and Conclusions

These findings demonstrate that HBE data can effectively replicate crash-based network screening using only three months of CVD compared to the three to five years of crash data typically required. Rear-end crashes show the strongest alignment with HBE-based screening, which is physically intuitive since hard braking is the evasive action in a potential rear-end collision. The results support the use of HBE for frequent, proactive safety screening of arterial road segments, enabling ALDOT to identify high-risk locations and prioritize safety interventions before crashes accumulate.

8. Applications in Travel Demand or Origin-Destination Analysis

Transportation planning projects or travel demand studies often rely on travel surveys to derive the trip-making patterns or use traffic count data (from detectors) to estimate trips origins and destinations. Transportation planners also use regional travel demand models to perform travel analysis. However, the travel surveys and regional travel models which are developed based on travel surveys are often outdated. Further, regional travel models only cover the metropolitan areas with a population above 50,000, and stateside travel models only cover major routes in a state. The CV data contains complete spatiotemporal trajectories of vehicle trips from their origins and destinations. The trip origins and destinations of probe vehicles can be directly observed at individual locations, and/or aggregated to represent the zonal level trip-making dynamics which are critical information for traffic impact and economic analysis of projects.

8.1 Case Study 1 – Trip Origin and Destination Analysis

Contributors: Jun Liu and Xiao Zou

8.1.1 Study Scope

This case study examines the origin and destinations distribution in Tuscaloosa using Wejo connected vehicle trajectory data. The purpose of this section is to establish the data processing framework to conduct origin–destination (O–D) based on CV data and explore the trip distance and travel time distribution.

8.1.2 Methodology

To support origin–destination (O–D) analysis, this component of the project utilizes large-scale connected vehicle (CV) trajectory data that capture detailed spatiotemporal movement patterns of individual trips across the road network. The goal is to extract reliable trip-level information—including origin, destination, travel time, and travel distance—and aggregate these insights to the zone-level for further analysis.

Figure 8.1 illustrates the data processing workflow. The connected vehicle data are first loaded using a parallel processing structure (e.g., Dask DataFrame) to efficiently handle the high data volume. Each record in the dataset includes a JourneyID, which uniquely identifies a single vehicle trip. The processing steps are as follows:

- *Trip Identification:* Trajectory records are grouped by JourneyID. For each group, the earliest location is identified as the origin, and the latest location as the destination.
- *Trip Metrics Calculation:* Based on the timestamp and coordinates, travel time and travel distance are calculated for each trip.
- *TAZ-Level aggregation:* A Traffic Analysis Zone (TAZ) shapefile for Tuscaloosa County is used to spatially link each origin and destination point to its corresponding zone. Trips are then aggregated at the TAZ level to capture zone-based travel patterns.
- *O–D Matrix Construction:* The number of trips between each TAZ pair is counted to form an O–D matrix. In addition to trip counts, the average travel time and average travel distance between each TAZ pair are also computed.

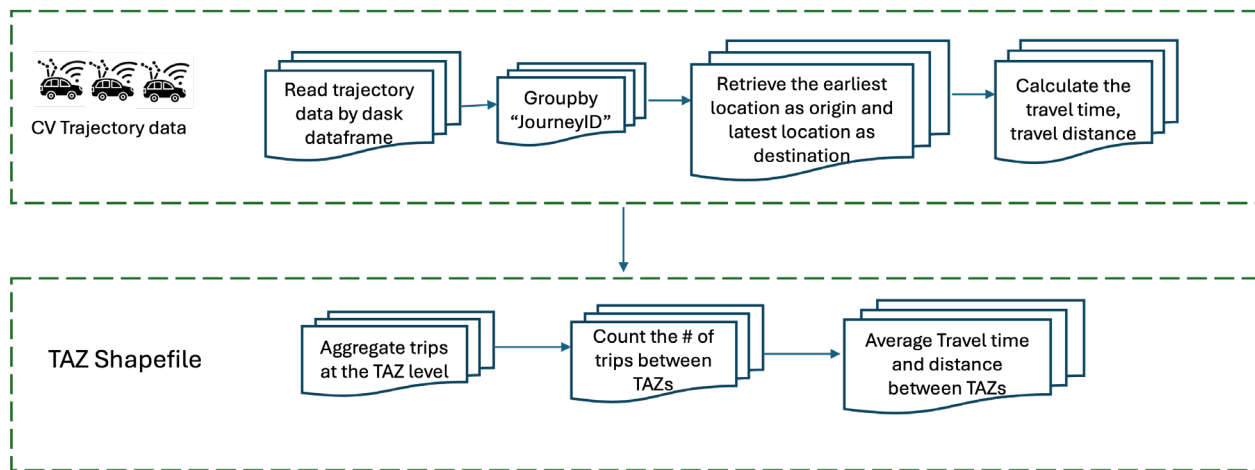


Figure 8.1 Data Processing Framework

8.1.3 Travel Demand Visualization and Summary

To summarize travel patterns in Tuscaloosa County, a set of visualizations was developed. The processed dataset includes 435 TAZs and 71,628 TAZ-to-TAZ O-D pairs. Figure 8.2 shows the total number of within-TAZ trips, where both the origin and destination fall within the same zone. Figure 8.3 displays inter-TAZ trip flows as directional lines between zones. Line color indicates trip volume, with blue representing high trip counts, red for medium, and green for low. Only the top 1,000 TAZ-to-TAZ OD pairs (out of 71,628 total) are visualized to improve clarity. Figure 8.4 presents the distribution of average travel time and average trip distance between all TAZ pairs. Most trips fall within a moderate travel range, with a positively skewed distribution indicating a larger number of short-distance trips and fewer long-distance ones.

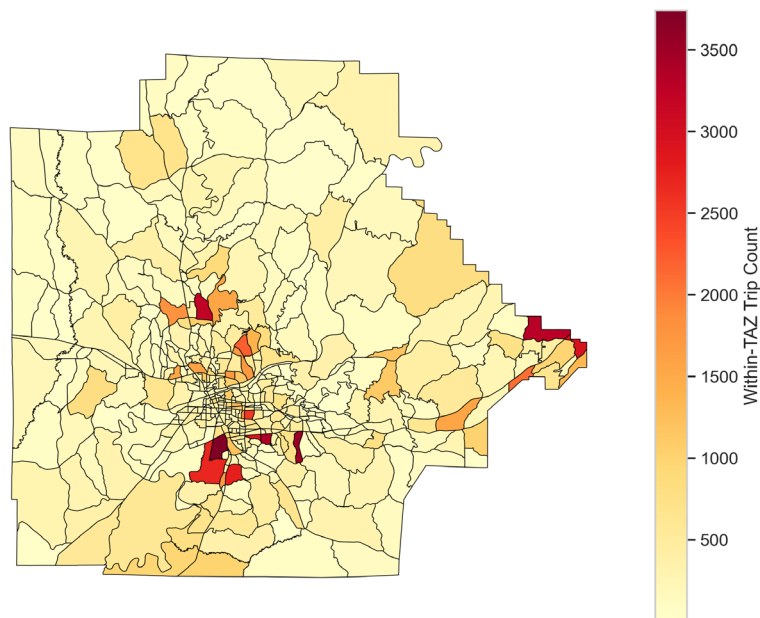


Figure 8.2 Within-TAZ Trip Counts in Tuscaloosa County

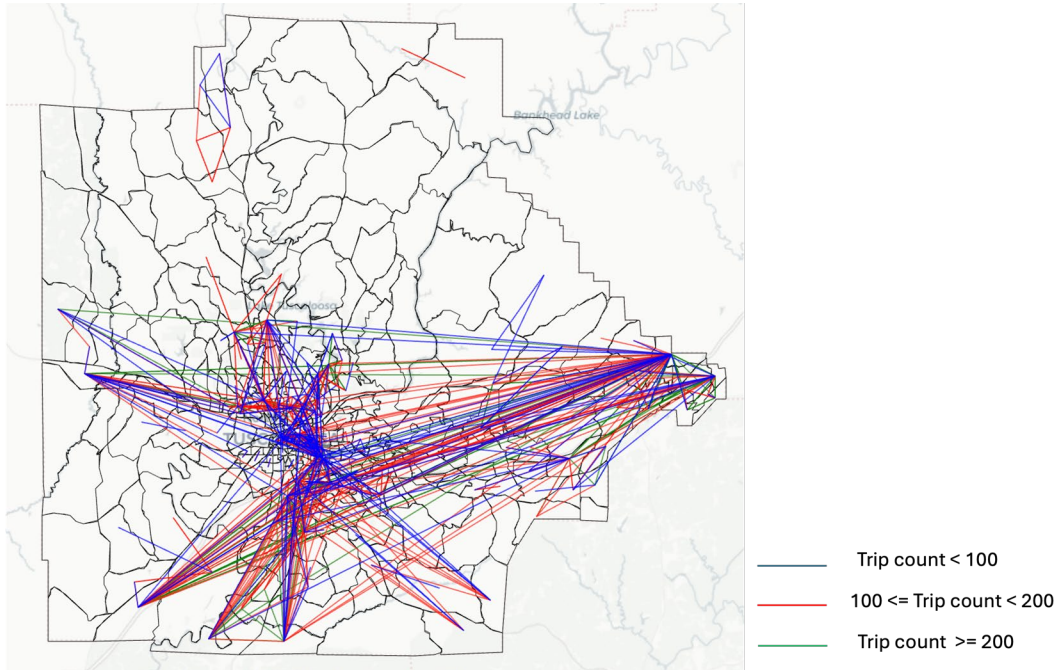


Figure 8.3 Inter-TAZ Trip Counts in Tuscaloosa County

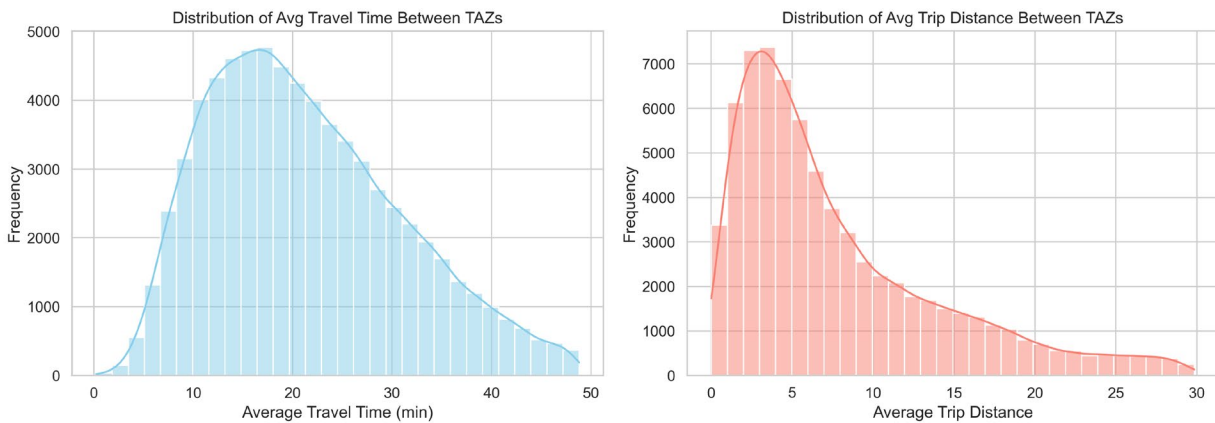


Figure 8.4 Distribution of Average Travel Time and Distance Between TAZ Pairs

9. Applications in Transportation Accessibility Analysis

As described above, the CV data provides insights into the origins and destinations of probe vehicle trips. Through analyzing the land use of the origins and destinations, trip-making time and speed, trip duration (and associated travel cost), transportation agencies are able to evaluate the transportation accessibility (by car) of a region to various socioeconomic opportunities such as employment, healthcare, education, and recreation.

9.1 Case Study 1 – Accessibility to Essential Services

Contributors: Jun Liu and Xiao Zou

9.1.1 Study Scope

The case study area is limited to Tuscaloosa County, focusing on residential populations and their access to essential services such as healthcare, grocery stores, education facilities, and emergency services.

9.1.2 Methodology

This section of the project focuses on evaluating transportation accessibility from residential areas to selected destinations using CV data in Tuscaloosa County. Accessibility is assessed through actual travel time and travel distance observed from vehicle trips. The trip-level origin–destination data used in this analysis were generated from the OD processing described in the previous section. Figure 9.1 presents the processing workflow used in the analysis.

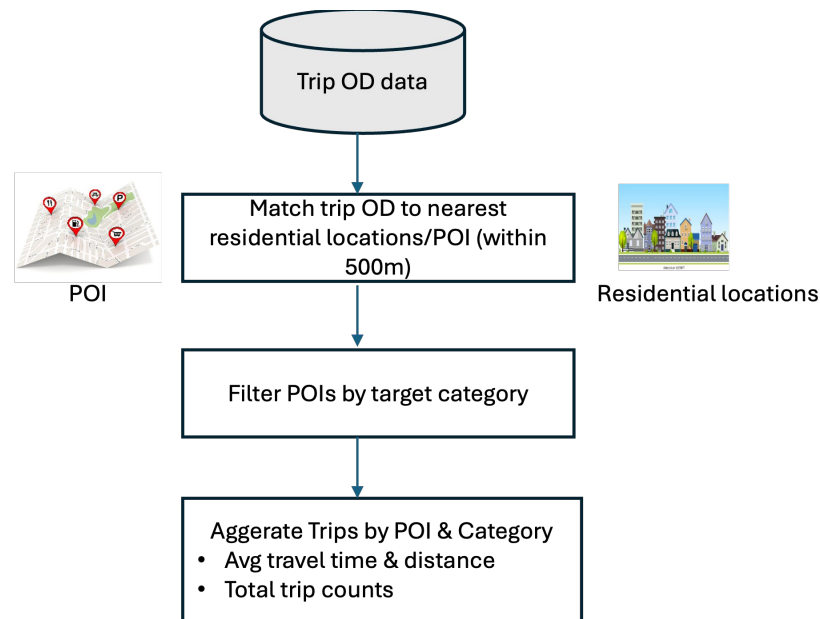


Figure 9.1 Methodology Framework

In the first step, the team use spatial nearest-neighbor search to identify the closest residential location to each trip origin and the closest Points of Interest (POI) to each trip destination. To perform this efficiently at scale, we use a KD-tree to build spatial indices for the residential and

POI datasets. This allows rapid nearest-neighbor querying with time complexity of approximately $O(\log n)$, significantly improving performance over brute-force distance checks. Only matches within a 500-meter threshold are retained. Out of the original 863,637 trips, 150,873 trips were successfully matched to both a residential origin and a POI destination, indicating that these trips are likely to represent travel from home to a specific place. Foursquare POI data includes detailed and hierarchical category labels. For this phase of the project, four key destination types were selected to represent access to socioeconomic opportunities: School; Health and Medicine; Dining and Drinking; Food and Beverage Retail. These categories were chosen for their relevance to daily life and mobility planning. However, the framework allows for flexibility, and additional destination types can be explored in future analyses based on specific policy or research needs. Finally, Trips are aggregated by POI and destination category. For each type of destination, the following accessibility indicators are computed: average travel time from residential origins; average travel distance; total number of trips.

9.1.3 Results

To illustrate accessibility from residential areas to selected destinations, two sets of visualizations were created based on the matched trips between residential locations and selected POIs. Figure 9.2 maps the trip count to individual POIs, with each circle representing a POI location. The size of each circle reflects the total number of matched trips, and different colors indicate the POI categories. A few destinations stand out with much higher trip volumes. For example, two Walmart stores attracted significantly more trips than other POIs. Several large health and medicine facilities also showed high trip counts, highlighting their importance in daily travel patterns across the region. Figure 9.3 presents the average travel time and average trip distance for each destination category.

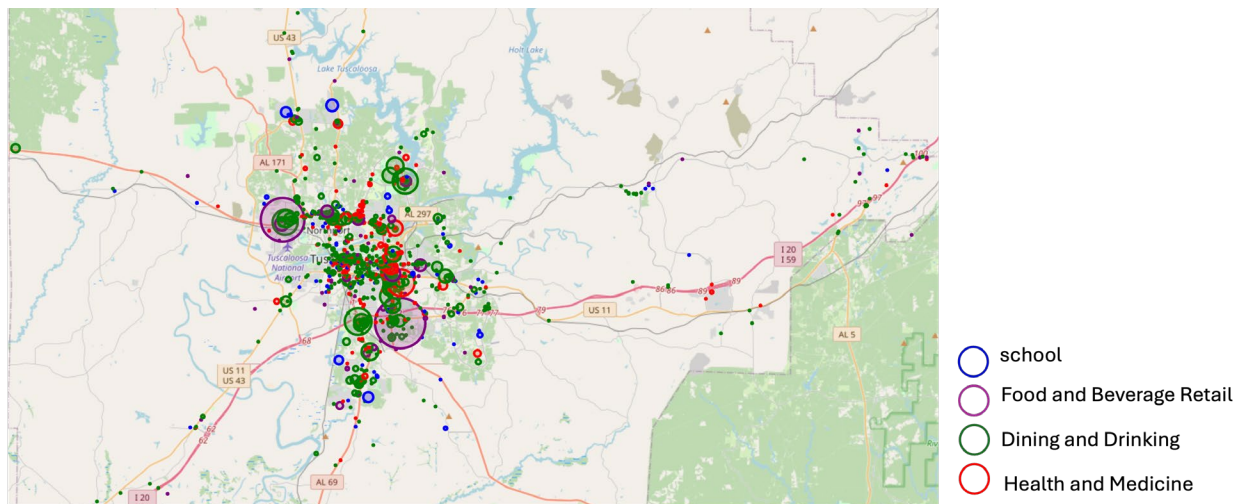


Figure 9.2 Trip Counts to Individual POIs by Category

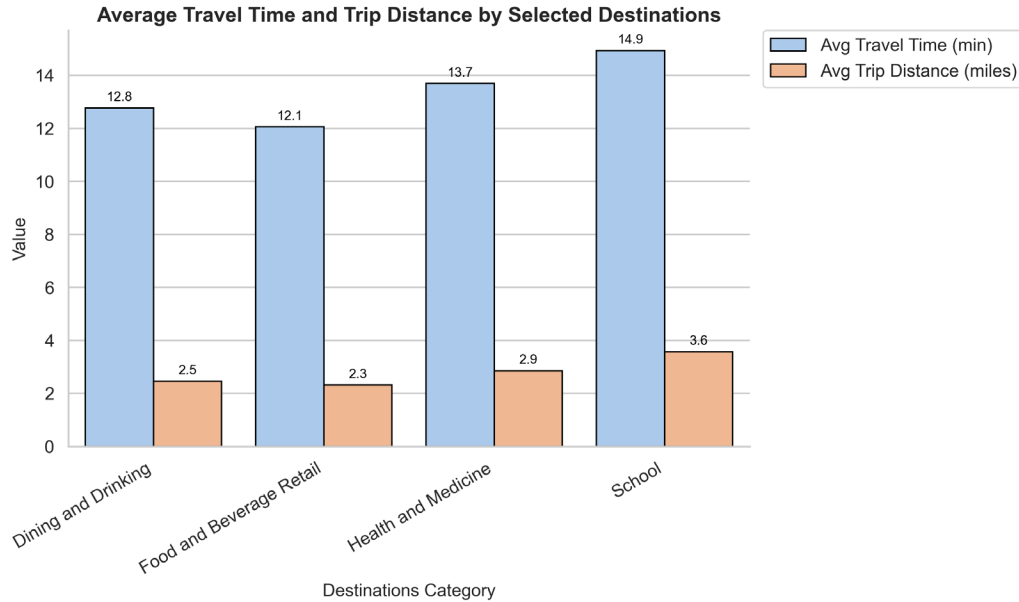


Figure 9.3 Average Travel Time and Distance by Destination Category

9.2 Case Study 2 – Hospital Accessibility Analysis Using DCVP Data

Contributors: Vamshi Annimalla, Praveena Penmetsa, Abhay Lidbe, Alex Hainen

This section presents a detailed analysis of observed hospital travel access conditions using disaggregated connected vehicle probe (DCVP) data. Spatial accessibility to healthcare facilities ensures timely medical care. Traditional models, including distance-based measures and travel time estimates, provide broad insights into potential accessibility but often overlook real-world travel dynamics such as congestion and route deviations. These approaches work at city or regional scales but under-resolve turn-by-turn variation along specific intersections and routes. To fill this gap, this study leverages DCVP data to analyze observed hospital travel access conditions by integrating empirical travel patterns against Open Route Service (ORS) estimates to attribute excess travel time to specific mechanisms.

9.2.1 Empirical Setting

The study focuses on two regional hospitals in Tuscaloosa, Alabama: Tuscaloosa Druid City Hospital Regional Medical Center (TDCH) and Northport Medical Center (NDCH). These hospitals are the primary healthcare providers serving the city and surrounding rural areas. Figure 9.4 identifies the locations of the two study hospitals. Figure 9.5 shows an isochrone map depicting areas reachable within 10, 30, and 60 minutes of travel to the study hospitals. Areas reachable within 10 minutes remain distinct for each hospital. The 30-minute and 60-minute zones overlap, suggesting patients from farther distances may have similar access to either hospital. However, travel time variations are influenced by road infrastructure, congestion, and signal delays, not solely distance. Urban areas generally experience shorter durations, while rural communities face extended travel times due to lower road density, fewer direct routes, and traffic impediments.

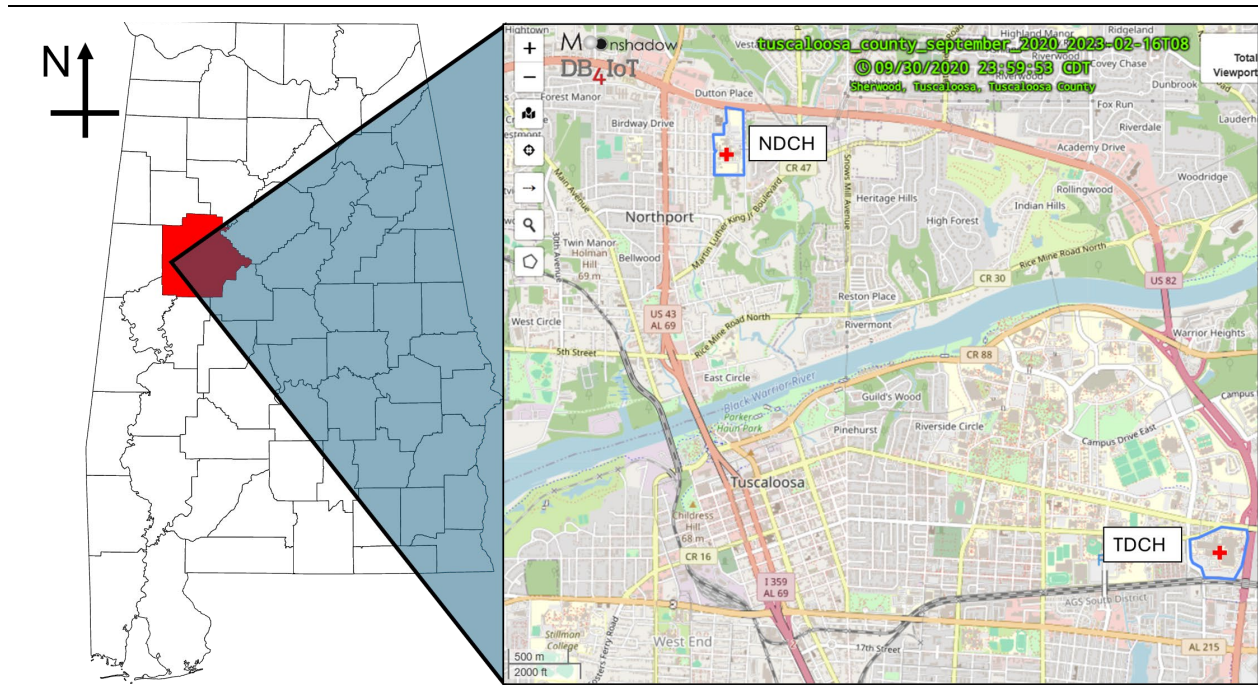


Figure 9.4 Study hospitals in Tuscaloosa and Northport, Alabama.

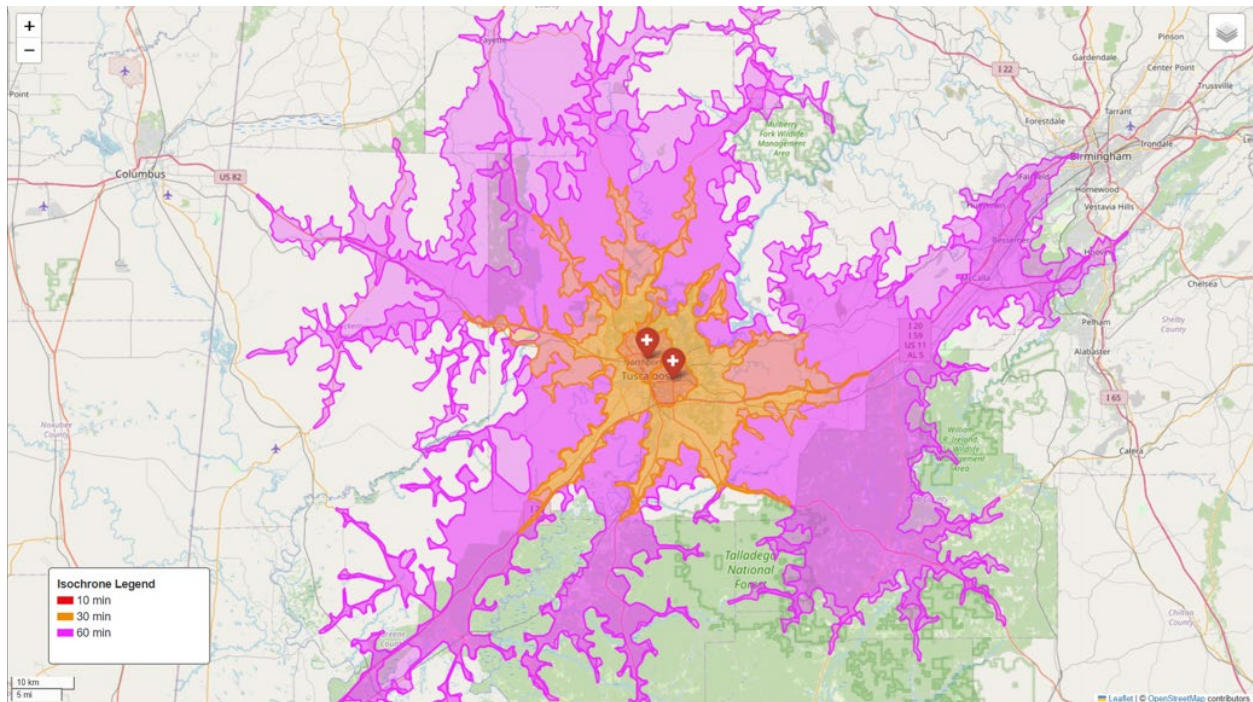


Figure 9.5 Isochrone map highlighting areas reachable under 10 minutes, 30 minutes, and 60 minutes to study hospitals.

9.2.2 Data and Method

DCVP data were acquired from a commercial vendor and processed through the DB4IoT platform. The study period spans September 2020 to January 2021, encompassing 54,743 trips (14,573 to and from NDCH and 40,170 to and from TDCH), as summarized in Table 9.1. Vehicle speed and location were captured at 3-second intervals. Destination polygons were drawn around the hospitals including parking lots, and trips were filtered to include only those with hospitals as origin or destination. The other end of each trip was aggregated at the ZIP-code level for anonymization. Table 9.2 summarizes the dataset, including recorded coordinates, actual and estimated travel distances, and journey times.

Table 9.1 Number of trips to and from hospitals across the study period.

Hospital	Sep 2020	Oct 2020	Nov 2020	Dec 2020	Jan 2021	Total
NDCH	2,860	3,087	2,566	3,086	2,974	14,573
TDCH	8,039	8,165	7,161	8,421	8,384	40,170
Total	10,899	11,252	9,727	11,507	11,358	54,743

Table 9.2 Sample DCVP-ORS records used for analysis, illustrating differences between estimated (ORS) and actual (DCVP) travel metrics for selected trips.

Journey ID	Direction	Lat	Long	Est. Time (min)	Est. Dist (mi)	Actual Time (min)	Actual Dist (mi)
1	Origin	33.233	-87.570	15.83	4.63	59.00	18.34
2	Origin	34.074	-86.864	108.80	100.79	100.63	96.09
5	Dest	32.905	-86.635	95.75	88.69	92.75	97.59
8	Dest	33.603	-86.283	99.65	91.49	102.38	92.15
11	Origin	34.495	-87.728	134.68	101.85	129.92	96.74

9.2.3 Identifying and Measuring Travel Delays

Trips were classified into four delay categories based on deviations from estimated travel conditions and speed progression. Stopped delay was calculated as the sum of time intervals where

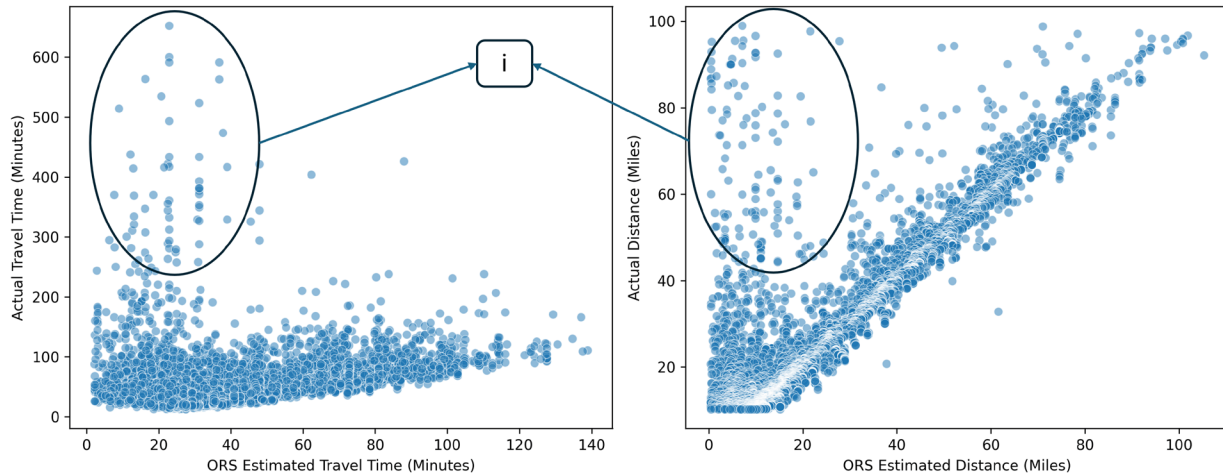
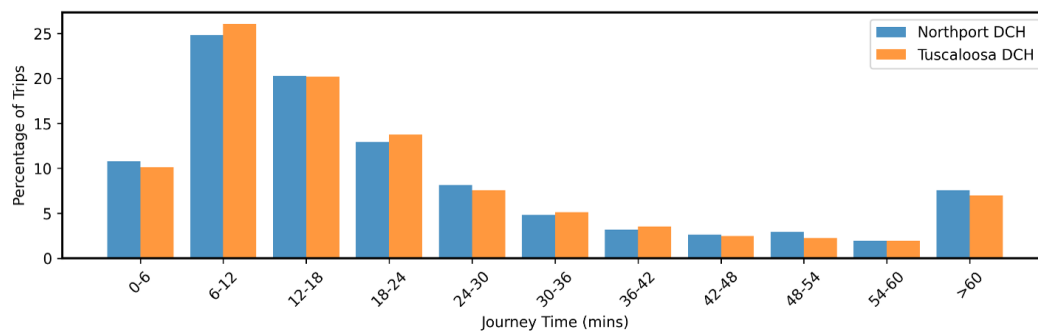


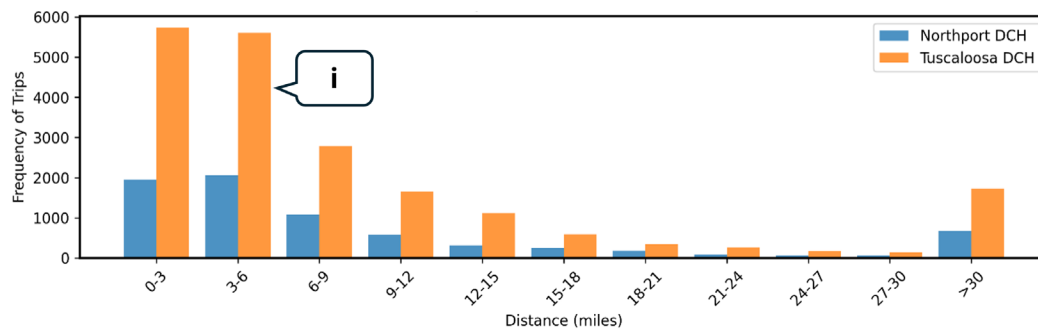
Figure 9.6 Scatter plot of trips showing variations in actual travel time and distance from ORS estimates.

the vehicle speed was at or below 5 MPH, with a minimum duration of 10 seconds required to classify a stop event. Total travel time was measured only during active movement, excluding idle periods at journey start and end. Travel time and distance differences between actual DCVP data and ORS estimates were computed to quantify deviations.

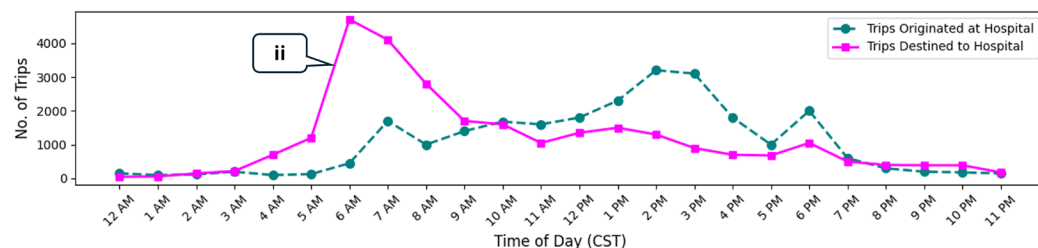
Route choice delays were classified when the actual distance significantly exceeded the estimated distance by more than 25 percent. Stopped delay (congestion) was identified when the distance was similar (within 10 percent of estimated) but stopped time was significant. Intermediate stops were identified when the distance exceeded 15 percent of estimated, the actual speed exceeded the median speed, and the travel time exceeded 25 percent of the estimated time. Normal trips were those with a travel time index (TTI) less than 1.2, indicating the actual time was within 20 percent of the estimated time.



(a)



(b)



(c)

Figure 9.7 (a). Percentage of trips within different journey time intervals. (b). Frequency of trips by distance. (c). Number of trips over a 24-hour period.

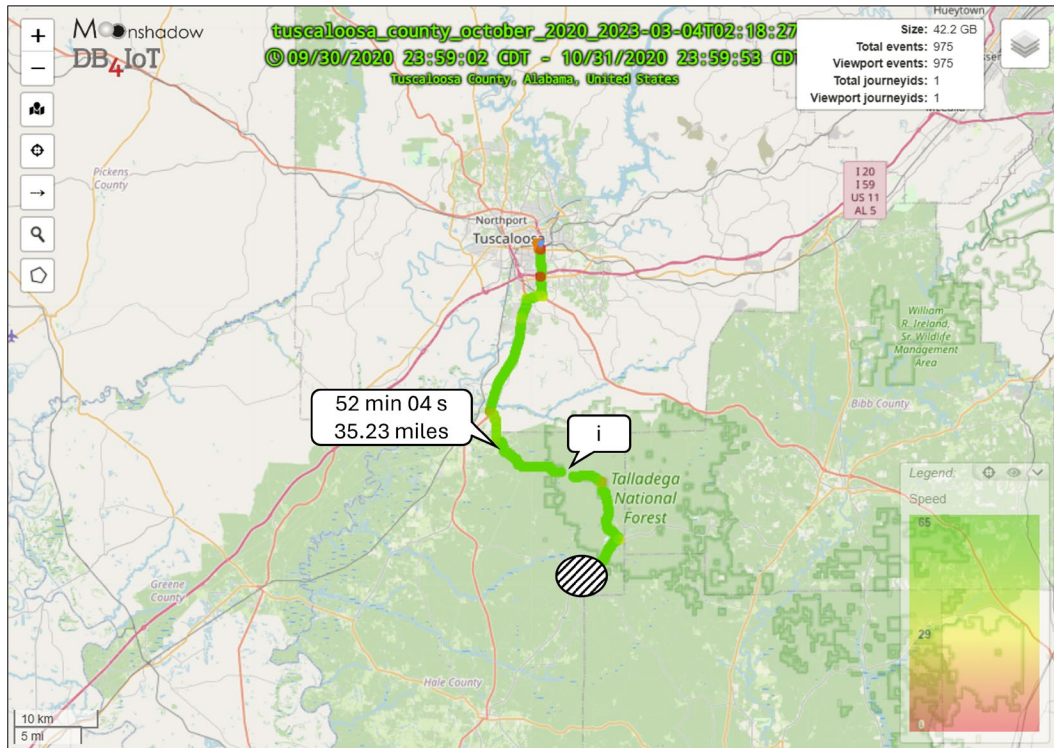
As shown in Figure 9.6, comparison of actual travel times and distances from DCVP data against ORS estimates reveals several deviations, with many trips experiencing significantly longer actual travel times (callout (i)). The distance comparison reveals a stronger correlation, with most points aligning along the diagonal, though some deviations suggest route choice variations. These unusual delays are attributed to voluntary route choices, congestion effects, or unobserved incidents.

The journey time distribution (Figure 9.7(a)) shows a significant concentration of trips within the 6 to 18 minute range, with longer travel times also present, particularly for NDCH. TDCH attracts a significantly higher number of trips across all distance ranges (Figure 9.7(b)); this is particularly evident in short-distance trips as pointed by callout (i), reinforcing its role as the primary choice for nearby patients. NDCH remains crucial for communities beyond Tuscaloosa, indicated by a considerable number of longer-distance trips. The time-of-day distribution (Figure 9.7(c)) reveals a strong morning peak (callout (ii)) for hospital-bound trips, suggesting higher travel demand during morning hours. Outbound trips exhibit a more dispersed pattern throughout the day.

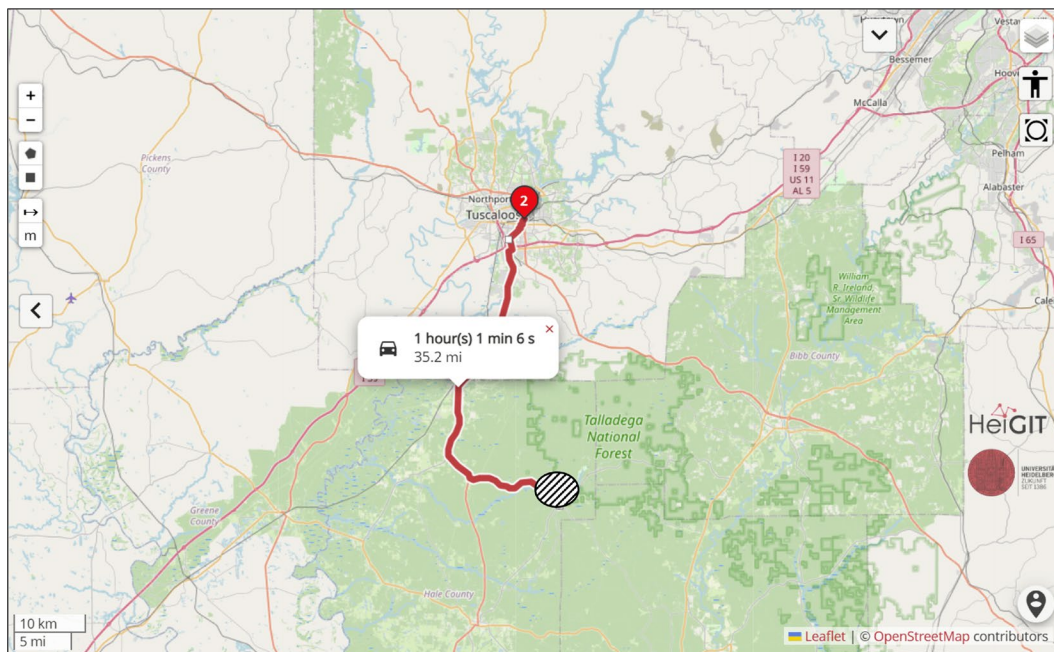
9.2.4 Route Disparities Between ORS and DCVP Data

A side-by-side comparison of estimated and actual travel conditions in Figures 9.8(a) and 9.8(b) for a trip from a rural location south of Tuscaloosa to the study hospital shows that ORS estimates represent an idealized scenario based on static routing and road-class speed profiles, while actual travel is influenced by congestion, traffic signals, road conditions, and driver choices. As highlighted by callout (i) in Figure 9.8(a), gaps in the GPS trajectory data appear along certain sections of the trip, resulting from limited cellular coverage, satellite signal obstructions, or telematics power-saving features.

As shown in Figure 9.9, the distribution of actual and estimated travel times varies across different locations. Actual travel times from DCVP exhibit greater variability, while ORS estimates remain more uniform and narrow, representing best-recommended times under ideal conditions. Particularly high variation is evident for trips from areas close to hospitals as pointed by callout (i), possibly due to drivers making additional stops or adjusting routes. Longer-distance trips show lower variation, likely because they are planned in advance. Knoxville and West Blocton were identified as critical locations with significant median travel time gaps compared to ORS estimates (callout (ii)).



(a)



(b)

Figure 9.8 (a). Actual trip trajectory obtained from DCVP data (DB4IoT). (b). Estimated travel time and distance from ORS.

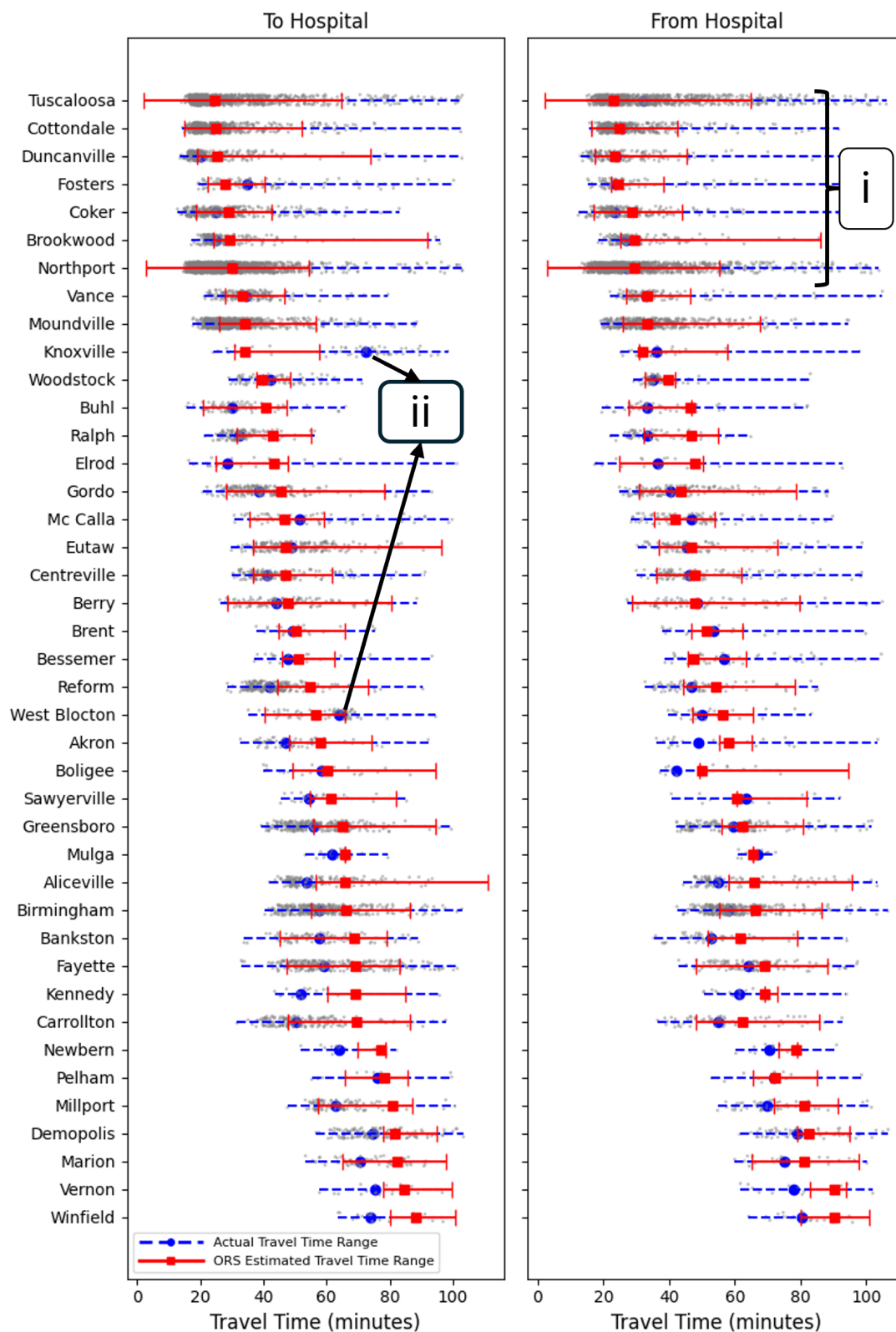


Figure 9.9 Region-wise distribution of actual vs. estimated travel times for trips to and from the hospital.

9.2.5 Observations from Delay Classified Trips

As shown in Figure 9.10(a), of all hospital-bound trips, 55.9 percent were classified as normal trips adhering to expected travel conditions. Congestion-related stopped delays affected 33.9 percent of trips, reflecting peak-hour demand and frequent signalized intersections along key corridors. Route deviations accounted for 9.1 percent of hospital-bound trips, while intermediate stops accounted for the remainder. Return trips showed lower congestion (29.5 percent) but higher route deviations (14.1 percent), suggesting drivers are less constrained by urgency when departing the hospital and more willing to take alternative routes or make discretionary stops. Figure 9.10(b) compares travel time gaps by delay type. Route delays exhibit the highest median and variability, suggesting route choices significantly impact travel times.

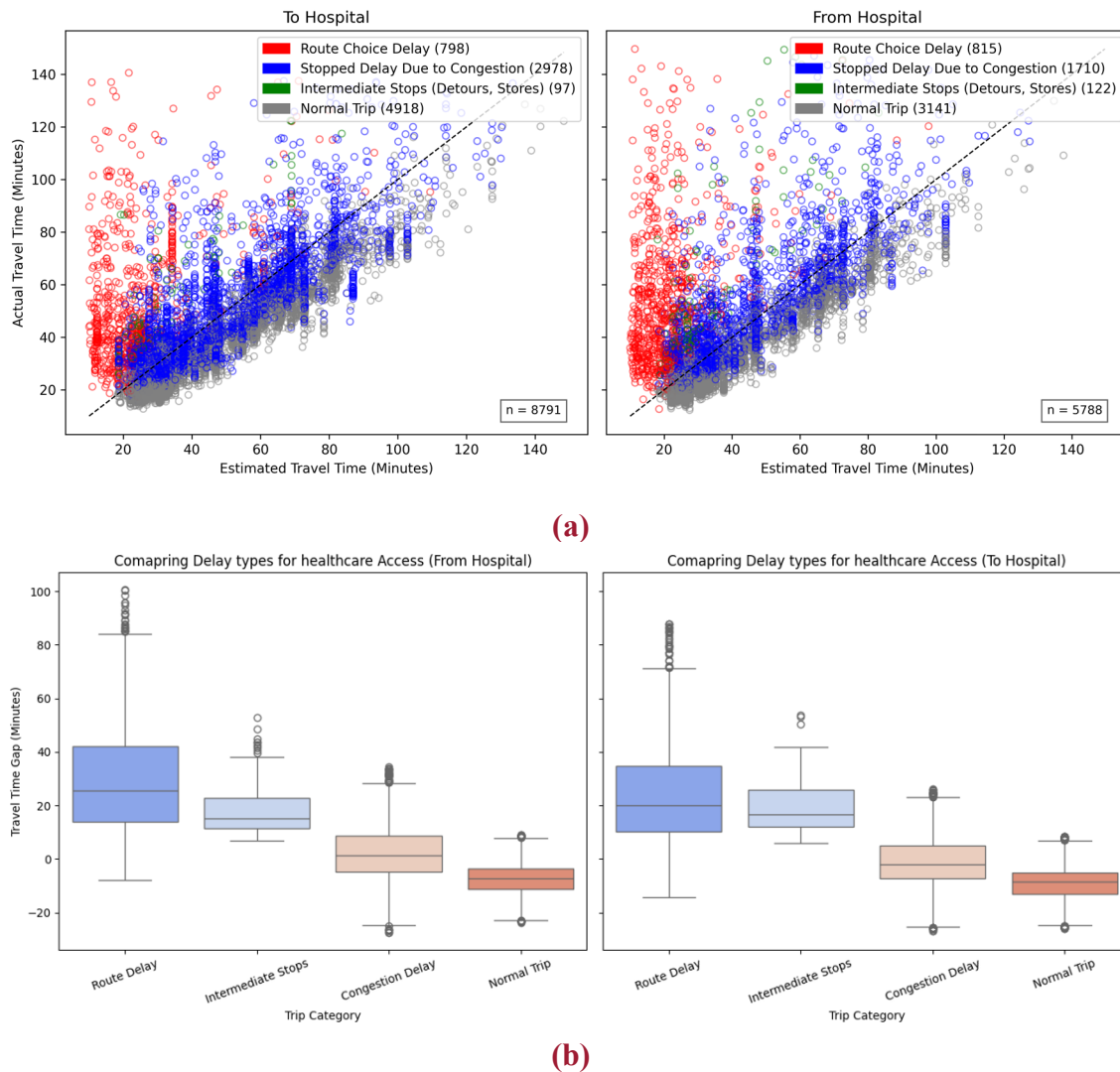


Figure 9.10 (a). Trips after classification. (b). Comparison of travel time gaps by type of delay.

9.2.6 Spatial Analysis of Delay Categories

Locations Experiencing Stopped Delay: The spatial distribution of stopped delays due to congestion and traffic signals (Figures 9.11(a) and 9.11(b)) shows that hospital-bound trips exhibit notably higher stopped delays than outbound trips, as shown by callout (i) in Figure 9.11(b). This

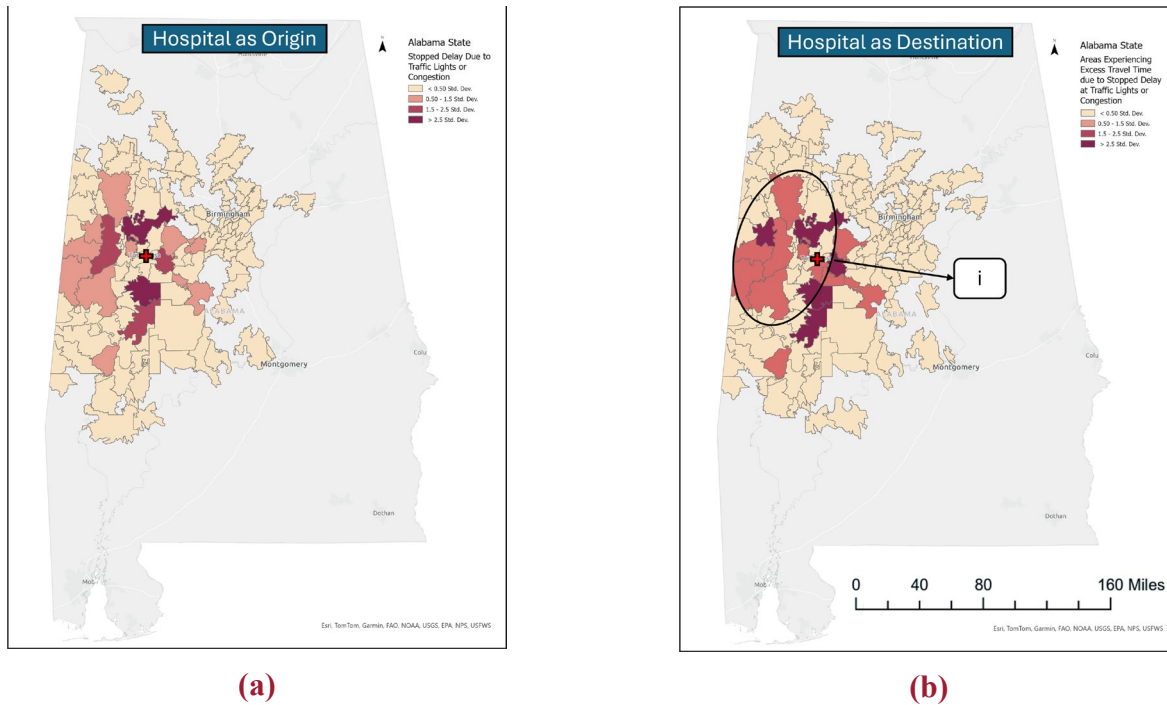
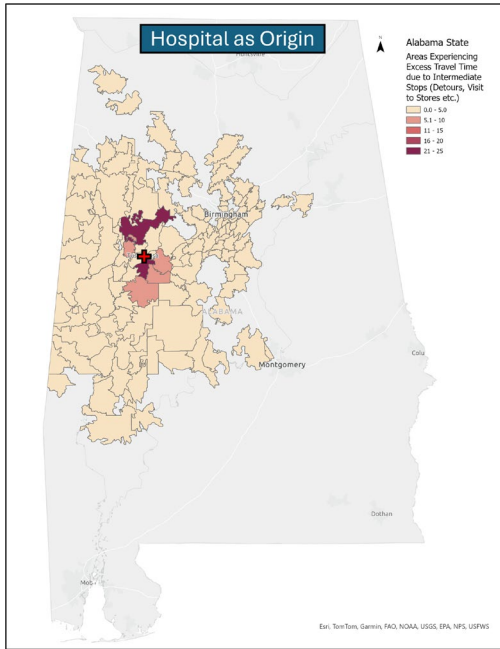


Figure 9.11 (a). Spatial distribution of stopped delays from hospitals. (b). Spatial distribution of stopped delays to hospitals.

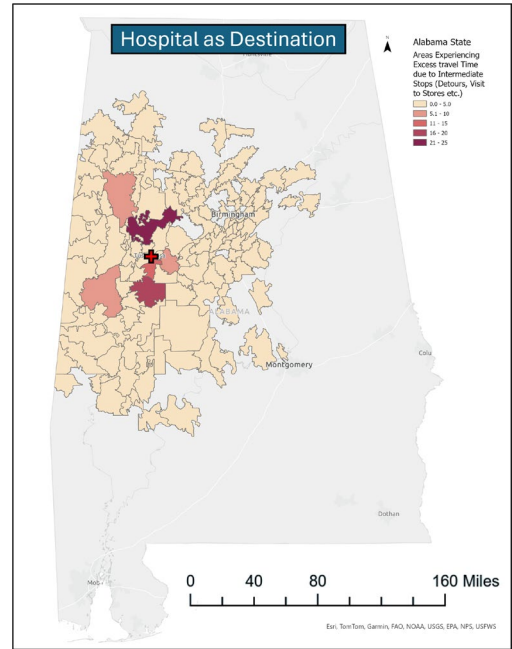
pattern is primarily influenced by peak-hour congestion, signalized intersections, and access routes to hospitals. Key locations with prolonged delays include Moundville, Cottondale, Knoxville, and West Blocton. The morning peak sees a surge in hospital-bound trips, aligning with commuting patterns and scheduled appointments, increasing road network demand and wait times at intersections. In contrast, outbound trips experience lower stopped delays due to less urgency and more dispersed distribution throughout the day.

Spatial Distribution of Intermediate Stops: For trips originating from hospitals (Figure 9.12(a)), Tuscaloosa and Northport exhibit the highest concentration of intermediate stops, suggesting post-hospital activities such as pharmacy or grocery visits. Moderate levels are observed in Cottondale and Moundville. For trips destined to hospitals (Figure 9.12(b)), intermediate stops are less frequent, though Tuscaloosa and Northport still show notable delays. Higher intermediate stops in Tuscaloosa and Northport are attributed to a familiarity effect, where local residents incorporate additional stops due to area knowledge, whereas travelers from farther locations prioritize direct hospital access.

Locations Experiencing Delay Due to Route Deviation: As shown in Figures 9.13(a) and 9.13(b), the darkest shaded regions, concentrated around Tuscaloosa, Northport, and Knoxville, indicate areas where trips exhibited significant deviations from estimated travel time. These deviations

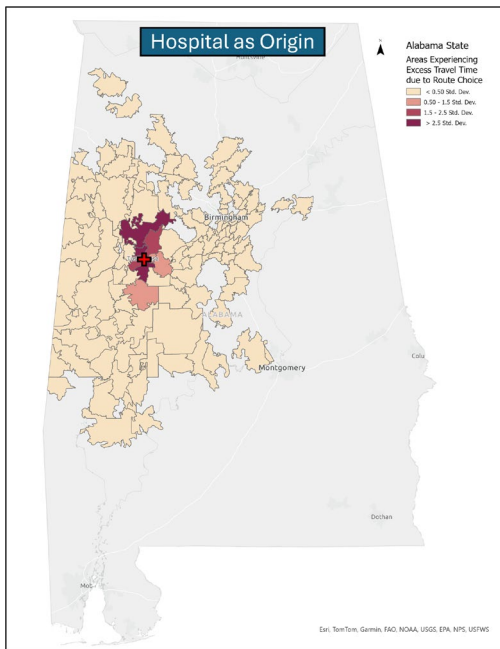


(a)

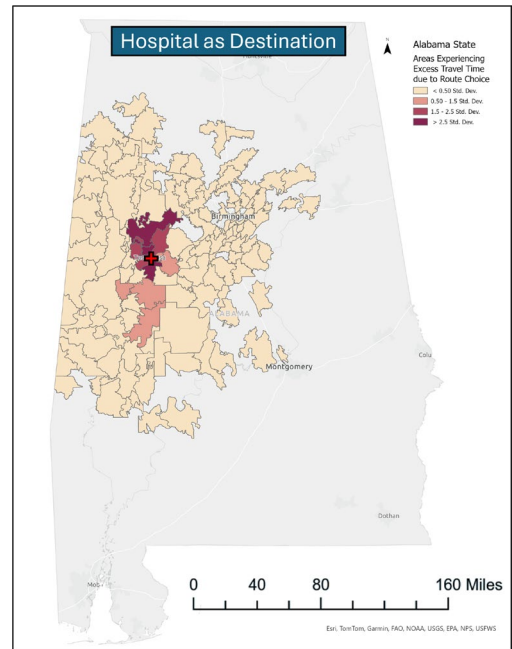


(b)

Figure 9.12 (a). Spatial distribution of excess travel time caused by intermediate stops from hospitals. (b). Spatial distribution of excess travel time caused by intermediate stops to hospitals.



(a)



(b)

Figure 9.13 (a). Spatial distribution of excess travel time due to route deviations from hospitals. (b). Spatial distribution of excess travel time due to route deviations to hospitals.

suggest drivers frequently opted for alternative routes influenced by observed congestion, familiarity with specific roads, or avoidance of high-traffic intersections. The pattern is more pronounced for trips originating from the hospital (Figure 9.13(a)), aligning with the expectation that individuals leaving a medical facility may be less constrained by urgency and more likely to use routes incorporating personal preferences or intermediate stops.

9.2.7 Identified Key Regions for Accessibility Improvements

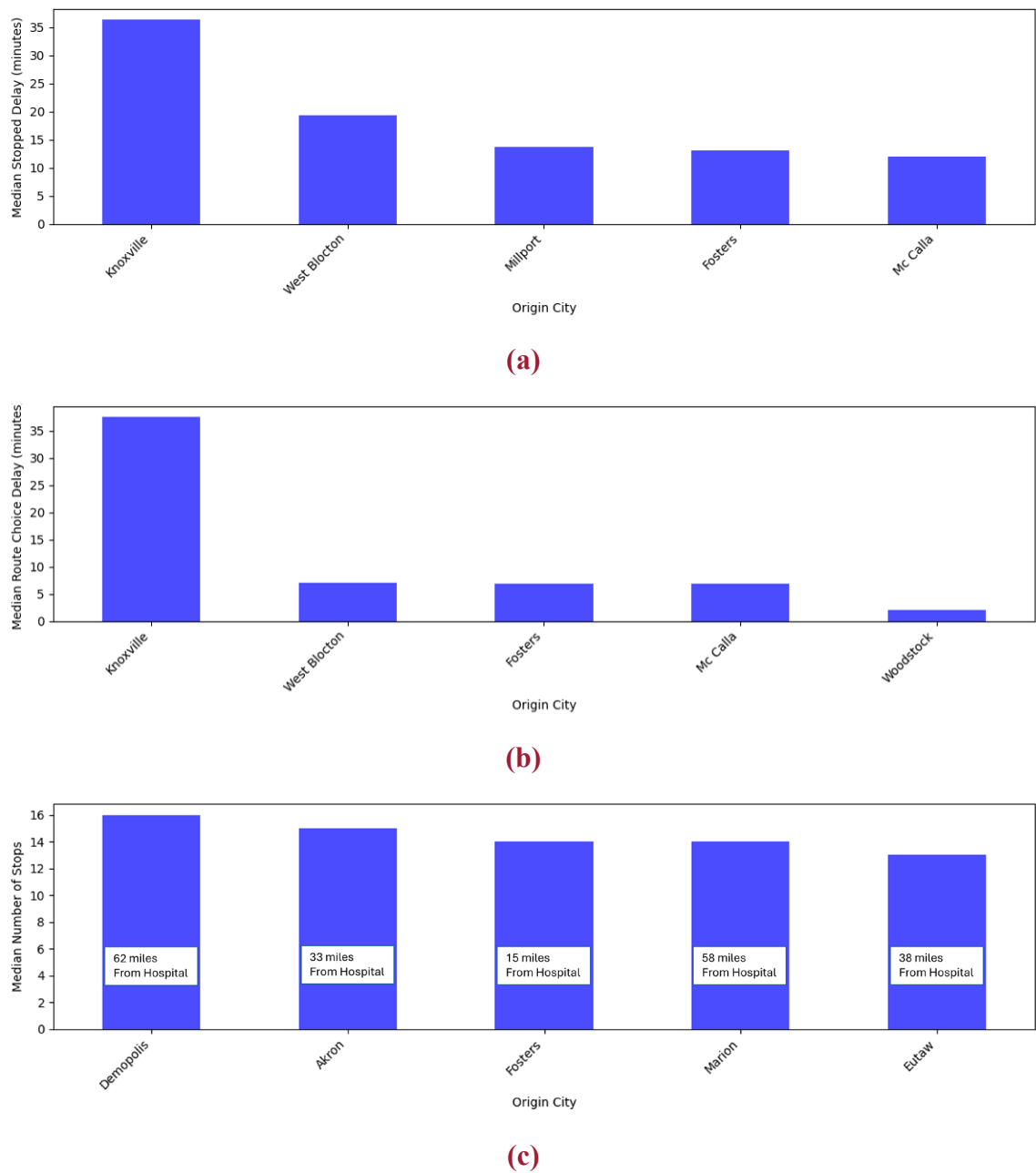
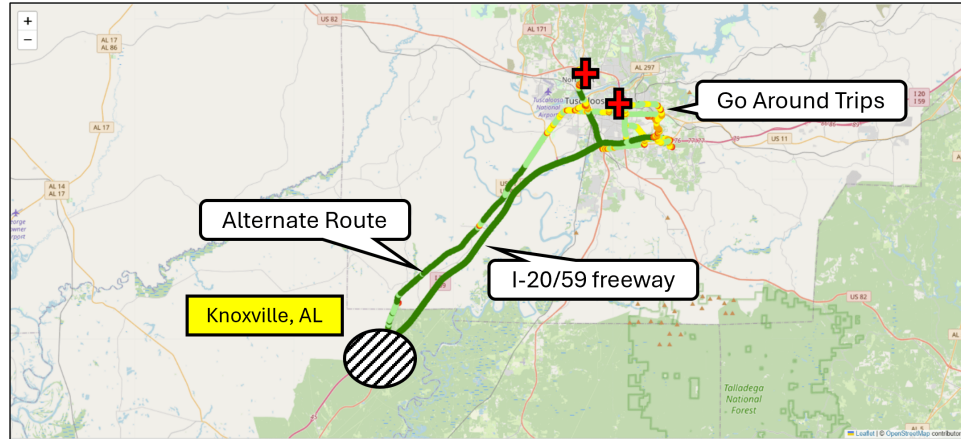
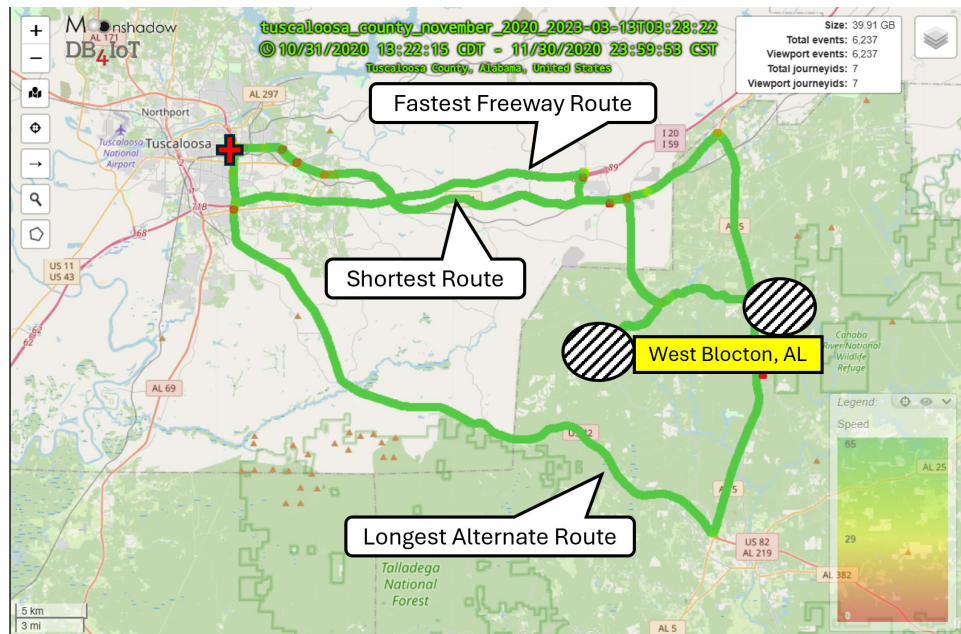


Figure 9.14 (a). Places with highest travel time deviation due to route deviation (to hospital). (b). Places with highest stopped delay (to hospital). (c). Places with most stops (to hospital).

The analysis focused on regions farther from hospitals (excluding Tuscaloosa and Northport due to the familiarity effect and extreme values), where travelers generally take direct routes but face significant delays, indicating broader accessibility challenges. Route deviations were particularly evident in Knoxville, West Blocton, and Fosters, with significant travel time discrepancies (Figure 9.14(a)). Stopped delays were most prominent in Knoxville and West Blocton due to recurring traffic bottlenecks (Figure 9.14(b)). Frequent stops increased for locations farther from the hospital, with Demopolis (62 miles) and Marion (58 miles) exhibiting the highest median stops (Figure 9.14(c)).



(a)



(b)

Figure 9.15 (a). Route trajectories from Knoxville, Alabama, showing multiple alternate paths taken by drivers to access the hospital. (b). Route trajectories from West Blocton, Alabama, illustrating alternative hospital access routes.

More stops for shorter distances (Fosters at 15 miles, Akron at 33 miles, Eutaw at 38 miles) raise accessibility concerns, suggesting detours or frequent congestion significantly impact hospital access beyond distance alone. A deeper examination of route deviations in Knoxville and West Blocton reveals a significantly higher gap in median actual versus estimated travel times, as pointed in Figure 9.15, indicating trips between these locations and hospitals are subject to considerable delays. Figure 9.15(a) shows multiple route selections by drivers from Knoxville, while Figure 9.15(b) illustrates the dispersed pattern of route trajectories from West Blocton, with numerous branches reflecting different route choices.

9.2.8 Discussion

This analysis demonstrates that high-resolution DCVP data provides empirical travel times and route-level diagnostics that complement traditional accessibility tools, revealing accessibility discrepancies overlooked by shortest-path assumptions and static models. Congestion-related stopped delays affect 33.9 percent of hospital-bound trips, representing the single largest impediment to hospital accessibility in the study region. More than half of hospital-bound trips (55.9 percent) align with estimated travel patterns, demonstrating that static models capture typical conditions adequately, but the remaining 44.1 percent experiencing delays highlights the need for dynamic analysis.

Significant accessibility challenges were identified in Knoxville and West Blocton, where actual travel times substantially exceed ORS estimates. The analysis demonstrates the need for dynamic traffic management, predictive routing, and improved medical transport strategies. Adaptive signal timing and congestion management along hospital corridors could address the 33.9 percent of delayed trips, while targeted roadway improvements in identified rural bottlenecks and real-time traveler information systems could minimize the 9.1 percent of route deviation delays. The methodology is applicable beyond healthcare to education, employment centers, and other essential services facing similar travel inefficiencies.

10. Applications in Connected/Automated Vehicles

This category of applications is related to the traffic operations that aim to improve mobility, safety and environment. With the vehicle trajectories approaching an intersection, vehicle-to-infrastructure (signals) applications can be developed to manage and assist vehicle maneuvers to avoid hard brakes and reduce delays. Besides, automated vehicle applications can be developed using CV data. With a large number of vehicle trajectories, historical vehicle driving decisions may be stored in automated vehicles as a reference to making new driving decisions at the same location and time. When a new driving decision is drastically different from historical driving decisions, driving decision errors need to be corrected and extra steps of driving decision validation will be needed. For example, when driving a curve or corner, typically, a vehicle needs to slow down first and then steer to inside of the curve or corner. If a new driving decision made or suggested by the automated vehicle is constant speed and steering straight or other side of the road, a driving decision error is identified and needs to be corrected.

10.1 Conceptual Framework

Connected and Automated Vehicles (CAVs) present a transformative opportunity to enhance mobility, safety, and environmental performance through real-time data sharing and intelligent decision-making. By utilizing connected vehicle (CV) data, CAV systems can be optimized to respond more effectively to dynamic roadway conditions. This section of the project focuses on how location-specific driving behavior patterns, extracted from large-scale CV data, can inform CAV operations, improve decision-making accuracy, and support vehicle-to-infrastructure (V2I) communication strategies.

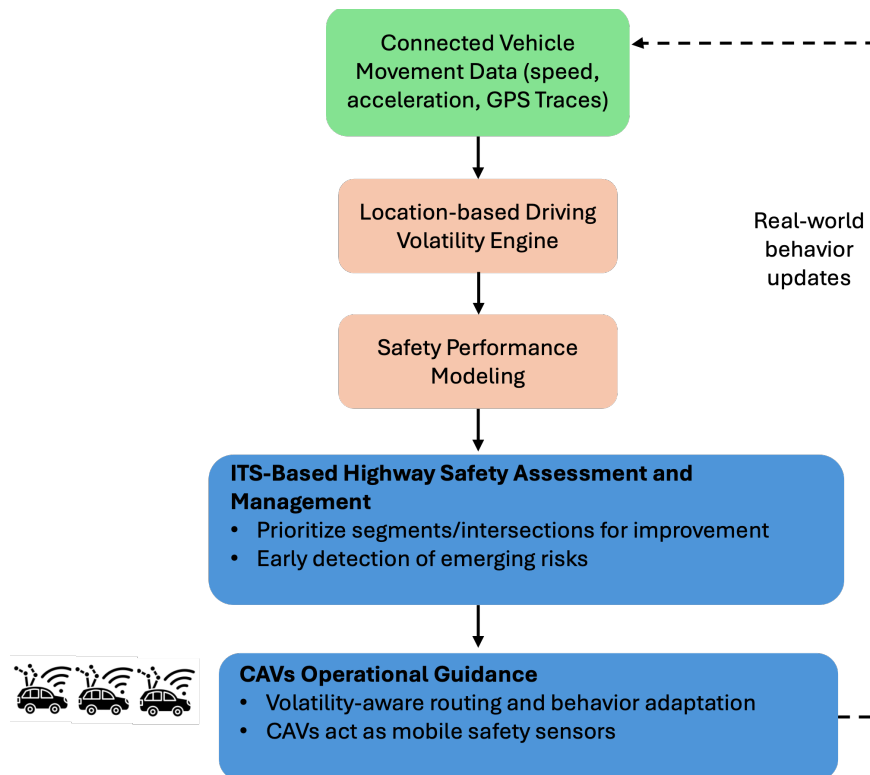


Figure 10.1 Conceptual Framework

Figure 10.1 The framework illustrates how connected vehicle (CV) movement data can be leveraged to support operational guidance for CAVs. High-resolution data—including speed, acceleration, and GPS traces—are first processed through a location-based driving volatility engine, which identifies segments and intersections with frequent extreme driving behaviors. These volatility metrics are then integrated into safety performance models to estimate crash risk and highlight potential safety concerns. The insight can support behavior-aware routing and maneuvering strategies. CAVs can adapt their driving behavior in real time and avoid high-risk locations. CAVs also act as mobile safety sensors, collecting and transmitting behavioral data back into the system. This real-time feedback loop allows continuous updates to the volatility assessment, enabling adaptive, behavior-aware safety monitoring and operational decisions.

10.2 Behavior-Informed CAV Decision Support

The proposed framework also provides meaningful support for the safe and adaptive operation of CAVs. As CAVs increasingly share the roadway with human-driven vehicles, their ability to interpret and respond to localized traffic conditions becomes essential for system-wide safety. Location-based driving volatility offers CAVs a valuable data stream for understanding behavioral risk in their surrounding environment. By integrating real-time volatility metrics into their decision-making processes, CAVs can adjust their behavior in response to areas where frequent abrupt maneuvers—such as hard braking or erratic acceleration—are detected. For instance, a CAV approaching a segment known for high driving volatility may choose to reduce speed, increase following distance, or shift to an alternate route. This behavior-aware routing enhances both vehicle-level safety and the overall stability of traffic flow. Beyond passive response, CAVs can actively contribute to the system. As mobile sensing platforms, they are well-equipped to collect and transmit high-resolution movement data that reflect current roadway conditions. This creates a continuous feedback loop, where real-world driving behavior captured by CAVs helps to update volatility assessments in near real time. The updated insights can be used not only for improving vehicle decisions but also for informing transportation agencies and ITS systems about evolving roadway risks. In this way, CAVs are not just data users, but also active participants in a connected safety ecosystem. Their role as both consumers and producers of behavior-based safety information enables more responsive, adaptive, and cooperative traffic environments which are key goals of intelligent and connected transportation systems.

10.3 V2I Integration at Intersections

The framework also supports vehicle-to-infrastructure (V2I) coordination, particularly at signalized intersections. By processing real-time vehicle approach data (speed, trajectory, distance), infrastructure systems can:

- Send signal phase and timing (SPaT) information to CAVs to reduce last-minute braking
- Predict dilemma zone entries and recommend appropriate deceleration
- Adjust signal timing dynamically to reduce vehicle delay and improve safety

These applications enhance both traffic flow and safety by enabling intelligent interaction between vehicles and infrastructure.

11. Summary, Implications, and Recommendations

11.1 Project Summary

This report documents the work carried out by the research team to explore the potential of large-scale crowdsourced connected vehicle data (CVD) to support data-driven transportation decision-making in Alabama. The project leverages multiple data sources, including Wejo CVD (disaggregated connected vehicle probe data), CARE crash data, ALGO incident data, HPMS roadway inventory data, Points of Interest (POI) data, and other supporting datasets (e.g., OpenStreetMap, TAZ boundaries, and ACS population estimates). The major tasks completed in this project include: (1) a comprehensive review of related work on state practice and scholarly research, (2) data collection and integration from multiple sources, (3) data processing and variable creation, and (4) the development of a series of analytical applications and case studies spanning traffic signal evaluation, corridor and network-wide mobility analysis, traffic safety, travel demand analysis, transportation accessibility, and connected/automated vehicle applications. The key findings and recommendations from each application area are summarized below.

Using Wejo trajectory data, the team quantified intersection delays by comparing actual passing times with ideal (free flow) travel times across signalized and unsignalized intersections. The analysis revealed that vehicles experience substantially longer delays at signalized intersections compared to unsignalized ones. Peak-hour delays were found to be notably higher than non-peak-hour delays, confirming the sensitivity of CVD-based metrics to temporal traffic variation. These findings demonstrate the value of CVD for identifying underperforming intersections and supporting signal timing optimization strategies.

Four case studies were conducted to demonstrate the capability of CVD for corridor and network-level mobility assessment. The corridor speed and delay analysis in Tuscaloosa established baseline traffic patterns using March 2021 Wejo data, revealing distinct directional and temporal patterns across three representative corridors. The extreme weather case study showed that weather events caused measurable disruptions to corridor performance, with reduced speeds and increased travel time variability. The game-day traffic analysis demonstrated how major events at the University of Alabama produced significant deviations from baseline conditions, with elevated congestion on key corridors before and after events. The freeway congestion analysis using DCVP data on the I-20/59 corridor employed advanced trajectory alignment and clustering techniques to identify congestion clusters, analyze shockwave propagation and queue lengths, and evaluate driver behavior during congestion. A congestion scoring framework was developed and validated against regional incident data, revealing unreported disruptions not tied to documented incidents.

Three case studies were conducted in the traffic safety domain. The driving volatility-based proactive safety management study developed a framework to compute location-based driving volatility at both segment and intersection levels using large-scale CVD. The results demonstrate that driving volatility metrics derived from CVD can effectively characterize the safety performance of roadway locations. The statewide correlation analysis of hard braking events and road traffic crashes established a significant positive relationship between hard braking event frequency and crash occurrence, supporting the use of hard braking data as a proactive safety indicator. The third case study on proactive road network safety screening integrated hard braking event data with crash records to identify high-risk locations across the road network, providing a practical screening tool for transportation agencies.

The travel demand case study demonstrated how CVD can be used to extract trip origin-destination patterns and visualize travel demand distributions. By reconstructing complete trip trajectories from Wejo data, the team was able to identify spatial patterns in travel demand and provide insights into regional trip-making behavior. These findings support transportation planning efforts by offering an alternative to traditional survey-based origin-destination data collection methods.

Two case studies addressed transportation accessibility. The first evaluated accessibility to essential services using CVD-derived travel patterns, revealing spatial disparities in access across different communities. The second case study focused on hospital accessibility using DCVP data, measuring travel delays to healthcare facilities and identifying route disparities between observed vehicle paths and standard routing service estimates. The analysis classified trips by delay severity and identified key regions where accessibility improvements are most needed. These findings provide evidence-based support for equity-oriented transportation planning and investment prioritization.

The report presents a conceptual framework for integrating CVD with connected and automated vehicle (CAV) systems. The framework describes how location-based driving volatility metrics can serve as real-time inputs for CAV decision-making, enabling adaptive behavior in response to localized traffic risk conditions. The vehicle-to-infrastructure (V2I) integration component demonstrates how volatility data at intersections can be communicated to approaching CAVs to support safer and more efficient navigation. This framework positions CVD as a foundational element in the development of intelligent and connected transportation systems.

11.2 Implications

The findings from this project have significant implications for transportation decision-making in Alabama, particularly in the areas of traffic operations, safety management, transportation planning, accessibility, and future connected and automated vehicle deployment. The applications developed in this study demonstrate that large-scale crowdsourced connected vehicle data (CVD), including raw trajectory data and driving event data, can provide ALDOT with a more continuous, spatially detailed, and behavior-based view of roadway system performance than many traditional data sources alone. By integrating CVD with other datasets including crash records, incident data, roadway inventory, and points of interest, transportation agencies can move toward a more proactive and evidence-driven approach to identifying operational problems, safety risks, and transportation service gaps across the network.

For traffic operations, this project shows that CVD can support performance monitoring at multiple spatial scales, from individual intersections to corridors and freeway networks. The intersection delay, corridor speed, weather disruption, game-day traffic, and freeway congestion case studies illustrate how trajectory data can reveal recurring congestion patterns, event-related deviations, unreported disruptions, travel time variability, and driver route choices. These findings imply that ALDOT can use CVD-derived measures to supplement existing traffic monitoring systems, prioritize locations for signal timing review, evaluate corridor reliability, and improve incident and special-event response strategies. Because CVD captures actual vehicle movements over time, it can also help agencies establish performance baselines and evaluate how roadway conditions change after operational improvements or external disruptions.

The safety applications in this project indicate that CVD can strengthen proactive safety management by identifying locations where risky driving behavior occurs before crashes are observed in large numbers. The demonstrated relationships among hard braking events, driving volatility, and crash occurrence suggest that behavioral indicators from CVD can serve as early warning measures for network screening. This has important implications for ALDOT's safety programs because it provides evidence for prioritizing locations for field review, countermeasure selection, and Highway Safety Improvement Program planning. Rather than relying only on historical crash frequency, agencies can incorporate near-crash and volatility-based indicators to better detect emerging safety concerns, especially at locations where crash data may be sparse or delayed.

The travel demand and accessibility analyses also show that CVD can expand the information available for long-range planning and equity-oriented decision-making. Origin-destination patterns derived from vehicle trajectories can help agencies understand actual travel behavior without depending entirely on household surveys or model estimates. Similarly, the accessibility case studies demonstrate how observed travel paths and delays can be used to evaluate access to essential services, including healthcare facilities. These findings have practical implications for identifying communities that experience longer travel times, route inefficiencies, or limited access to critical destinations. For Alabama, where rural and underserved areas may face distinct mobility challenges, CVD-based accessibility analysis can support more targeted investment decisions and improve the alignment between transportation planning and community needs.

Finally, this project highlights the strategic value of CVD as Alabama prepares for a more connected transportation future. The conceptual framework for connected and automated vehicle applications demonstrates how driving volatility, congestion, and roadway condition indicators could eventually support vehicle-to-infrastructure communication and real-time decision support. This implies that the methods developed in this report are not limited to retrospective analysis; they can serve as building blocks for future intelligent transportation systems. Overall, the project demonstrates that CVD can become a practical decision-support resource for ALDOT, helping the agency improve operational efficiency, advance proactive safety management, support equitable planning, and prepare for emerging transportation technologies.

11.3 Recommendations and Future Work

Based on the findings of this project, the following recommendations are offered for ALDOT and other transportation agencies seeking to leverage CVD for decision-making:

1. Continue and expand the acquisition of connected vehicle data to increase temporal and spatial coverage across Alabama's roadway network, as higher penetration rates will strengthen the reliability of CVD-based performance metrics.
2. Integrate the analytical tools and methods developed in this project into ALDOT's operational workflows, particularly the corridor-level performance monitoring, congestion scoring framework, and proactive safety screening approaches.
3. Explore the fusion of CVD with additional emerging data sources, such as drone-based traffic monitoring and LiDAR, to further enhance the spatial resolution and behavioral detail available for transportation analysis.

-
4. Leverage the hard braking and driving volatility indicators developed in this project as part of ALDOT's network safety screening and Highway Safety Improvement Program (HSIP) processes to proactively identify high-risk locations before crashes occur.
 5. Apply the accessibility analysis framework to evaluate equity in transportation service delivery, particularly for healthcare access in rural and underserved communities across Alabama.
 6. Prepare for the growing availability of connected vehicle data by investing in data infrastructure, storage capacity, and analytical capabilities. As vehicle penetration rates increase and new vehicle communication technologies emerge, the potential applications of CVD will continue to expand.
 7. Continue to explore the integration of CVD with connected and automated vehicle platforms, leveraging real-time driving volatility and traffic condition data to support safer and more efficient vehicle operations within Alabama's transportation network.

The analyses presented in this report illustrate the value of large-scale crowdsourced connected vehicle data for Alabama, but also point to a number of directions in which the work can be extended. Several promising avenues for continued research and implementation are outlined below.

First, the analytical framework can be broadened to additional application areas and case studies identified in the original project scope. In particular, a dedicated Work Zone application leveraging Wejo trajectory and driving event data could quantify the mobility and safety impacts of construction activity and support variable work zone management strategies. Environmental and emissions analyses using speed and driving-behavior signals could produce corridor- and network-level estimates of vehicle energy use and tailpipe emissions and inform sustainability planning. The case-study framework demonstrated through the game-day and freeway-incident analyses in this report can also be extended to additional special traffic events, including hurricane evacuations along coastal Alabama corridors, winter weather events, and planned or unplanned road closures and reopenings, building toward a more comprehensive CVD-based playbook for emergency and special-event management.

Second, the temporal, spatial, and demographic scope of the data used in future studies should be expanded. Most analyses in this report used one-month to one-year data windows at penetration rates of roughly nine to ten percent. Extending the analytical pipeline to multi-year windows, to higher-penetration datasets as the connected vehicle fleet grows, and to additional corridors and regions across Alabama would strengthen trend detection, support robust before-and-after evaluations of infrastructure investments, and improve the statistical reliability of safety and congestion metrics at finer spatial resolutions.

Third, the methodological pipeline developed in this project, spatial mapping, trajectory alignment, clustering-based congestion detection, driving-volatility indicators, and accessibility modeling provides a natural foundation for further research into machine learning and deep learning methods for anomaly detection, short-term congestion and incident prediction, and multi-source data fusion. Systematic validation of these methods against independent data sources (for example, loop detectors, Bluetooth probes, and future V2X feeds) and testing of their transferability to other corridors, regions, and state DOTs are also important next steps.

Fourth, the outputs of this project should be progressively integrated into ALDOT's operational systems. Future work can develop dashboards, standardized reporting templates, and application

programming interfaces that deliver CVD-derived performance measures (corridor mobility, congestion scores, hard-braking hotspots, accessibility indicators) directly to signal retiming, safety screening, incident response, traffic management center, and long-range planning workflows. Pilot deployments with selected ALDOT regions or districts would help validate the practical utility of these measures and refine them based on operator feedback.

Finally, the accessibility work presented in this report focused on healthcare access as an illustrative use case. Future research can broaden this framework to additional opportunity categories such as employment, education, and essential services, incorporate demographic and socioeconomic overlays, and quantify equity outcomes of specific transportation investments across urban, suburban, and rural Alabama. Together, these directions will continue to expand the operational and policy value of connected vehicle data for the state and position ALDOT to take full advantage of the growing availability of high-resolution mobility data.

References

- Abdel-Aty, M., & Haleem, K. (2011). Analyzing angle crashes at unsignalized intersections using machine learning techniques. *Accident Analysis & Prevention*, 43(1), 461–470.
- Abdelraouf, A., Abdel-Aty, M., & Mahmoud, N. (2022). Sequence-to-sequence recurrent graph convolutional networks for traffic estimation and prediction using connected probe vehicle data. *IEEE Transactions on Intelligent Transportation Systems*, 24(1), 1395-1405.
- Dabiri, S., Lu, C. T., Heaslip, K., & Reddy, C. K. (2019). Semi-supervised deep learning approach for transportation mode identification using GPS trajectory data. *IEEE Transactions on Knowledge and Data Engineering*, 32(5), 1010-1023.
- Dabiri, S., Marković, N., Heaslip, K., & Reddy, C. K. (2020). A deep convolutional neural network based approach for vehicle classification using large-scale GPS trajectory data. *Transportation Research Part C: Emerging Technologies*, 116, 102644.
- Dimitrijevic, B., Zhong, Z., Zhao, L., Besenski, D., & Lee, J. (2022, November). Assessing connected vehicle data coverage on new jersey roadways. In *2022 IEEE 7th International Conference on Intelligent Transportation Engineering (ICITE)* (pp. 388–393). IEEE.
- Fan, J., Fu, C., Stewart, K., & Zhang, L. (2019). Using big GPS trajectory data analytics for vehicle miles traveled estimation. *Transportation Research Part C: Emerging Technologies*, 103, 298-307.
- Feng, H., Bai, F., & Xu, Y. (2019). Identification of critical roads in urban transportation network based on GPS trajectory data. *Physica A: Statistical Mechanics and its Applications*, 535, 122337.
- Fu, X., Nie, Q., Liu, J., Khattak, A., Hainen, A., & Nambisan, S. (2022). Constructing spatiotemporal driving volatility profiles for connected and automated vehicles in existing highway networks. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 26(5), 572–585.
- Gong, S., Cartlidge, J., Bai, R., Yue, Y., Li, Q., & Qiu, G. (2020). Extracting activity patterns from taxi trajectory data: A two-layer framework using spatio-temporal clustering, Bayesian probability and Monte Carlo simulation. *International Journal of Geographical Information Science*, 34(6), 1210-1234.
- Greene, W. H. (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. NYU Working Paper No. EC-94-10.
- Gu, Y., Liu, D., Arvin, R., Khattak, A. J., & Han, L. D. (2023). Predicting intersection crash frequency using connected vehicle data: A framework for geographical random forest. *Accident Analysis & Prevention*, 179, 106880.
- Haklay, M., & Weber, P. (2008). Openstreetmap: User-generated street maps. *IEEE Pervasive Computing*, 7(4), 12–18.
- Hoseinzadeh, N., Gu, Y., Han, L. D., Brakewood, C., & Freeze, P. B. (2021, March). Estimating freeway level-of-service using crowdsourced data. In *Informatics* (Vol. 8, No. 1, p. 17). MDPI.
- Hu, Y., Li, Y., Huang, H., Lee, J., Yuan, C., & Zou, G. (2022). A high-resolution trajectory data driven method for real-time evaluation of traffic safety. *Accident Analysis & Prevention*, 165, 106503.
- Huang, H., Wu, J., Liu, F., & Wang, Y. (2020). Measuring accessibility based on improved impedance and attractive functions using taxi trajectory data. *Sustainability*, 13(1), 112.

-
- Huang, Y., Xiao, Z., Wang, D., Jiang, H., & Wu, D. (2019). Exploring individual travel patterns across private car trajectory data. *IEEE Transactions on Intelligent Transportation Systems*, 21(12), 5036-5050.
- Hunter, M., Saldivar-Carranza, E., Desai, J., Mathew, J. K., Li, H., & Bullock, D. M. (2021). A proactive approach to evaluating intersection safety using hard-braking data. *Journal of Big Data Analytics in Transportation*, 3(2), 81-94.
- Islam, Z., & Abdel-Aty, M. (2023). Traffic conflict prediction using connected vehicle data. *Analytic Methods in Accident Research*, 39, 100275.
- Kamrani, M., Wali, B., & Khattak, A. J. (2017). Can data generated by connected vehicles enhance safety?: Proactive approach to intersection safety management. *Transportation Research Record*, 2659(1), 80–90.
- Kan, Z., Kwan, M. P., Liu, D., Tang, L., Chen, Y., & Fang, M. (2022). Assessing individual activity-related exposures to traffic congestion using GPS trajectory data. *Journal of Transport Geography*, 98, 103240.
- Kan, Z., Tang, L., Kwan, M. P., Ren, C., Liu, D., & Li, Q. (2019). Traffic congestion analysis at the turn level using Taxis' GPS trajectory data. *Computers, Environment and Urban Systems*, 74, 229-243.
- Kong, X., Xu, Z., Shen, G., Wang, J., Yang, Q., & Zhang, B. (2016). Urban traffic congestion estimation and prediction based on floating car trajectory data. *Future Generation Computer Systems*, 61, 97-107.
- Li, T., Wu, J., Dang, A., Liao, L., & Xu, M. (2019). Emission pattern mining based on taxi trajectory data in Beijing. *Journal of Cleaner Production*, 206, 688-700.
- Li, Y., Chen, S., Du, R., Ha, P. Y. J., Dong, J., & Labi, S. (2020). Using empirical trajectory data to design connected autonomous vehicle controllers for traffic stabilization. *arXiv preprint arXiv:2010.05440*.
- Lin, L., Gong, S., Peeta, S., & Wu, X. (2021). Long short-term memory-based human-driven vehicle longitudinal trajectory prediction in a connected and autonomous vehicle environment. *Transportation Research Record*, 2675(6), 380-390.
- Liu, J., & Khattak, A. (2020). Informed decision-making by integrating historical on-road driving performance data in high-resolution maps for connected and automated vehicles. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 24(1), 11–23.
- Liu, J., Khattak, A., & Wang, X. (2015). The role of alternative fuel vehicles: Using behavioral and sensor data to model hierarchies in travel. *Transportation Research Part C: Emerging Technologies*, 55, 379–392.
- Liu, J., Wang, X., & Khattak, A. (2014). Generating real-time driving volatility information. 2014 World Congress on Intelligent Transport Systems.
- Maneewongvatana, S., & Mount, D. M. (1999). Analysis of approximate nearest neighbor searching with clustered point sets. *ArXiv Preprint Cs/9901013*.
- Peng, Z., Gao, S., Li, Z., Xiao, B., & Qian, Y. (2018). Vehicle safety improvement through deep learning and mobile sensing. *IEEE Network*, 32(4), 28-33.
- Qian, X., Lei, T., Xue, J., Lei, Z., & Ukkusuri, S. V. (2020). Impact of transportation network companies on urban congestion: Evidence from large-scale trajectory data. *Sustainable Cities and Society*, 55, 102053.

-
- Qu, D., Wang, S., Liu, H., & Meng, Y. (2022). A car-following model based on trajectory data for connected and automated vehicles to predict trajectory of human-driven vehicles. *Sustainability*, 14(12), 7045.
- Sun, M., Xue, C., Cheng, Y., Zhao, L., & Long, Z. (2021). Analyzing spatiotemporal daily travel source carbon emissions based on taxi trajectory data. *IEEE Access*, 9, 107012-107023.
- Sun, S., Chen, J., & Sun, J. (2019). Traffic congestion prediction based on GPS trajectory data. *International Journal of Distributed Sensor Networks*, 15(5), 1550147719847440.
- Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A., & González, M. C. (2015). The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*, 58, 162-177.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307-333.
- Xie, K., Ozbay, K., Yang, H., & Li, C. (2019). Mining automatically extracted vehicle trajectory data for proactive safety analytics. *Transportation Research Part C: Emerging Technologies*, 106, 61-72.
- Xu, C., Wang, X., Yang, H., Xie, K., & Chen, X. (2019). Exploring the impacts of speed variances on safety performance of urban elevated expressways using GPS data. *Accident Analysis & Prevention*, 123, 29-38.
- Yao, Z., Gan, M., Qian, Q., Qiao, Y., & Wei, L. (2023). Variation of truck emission by trip purposes: Cases by real-world trajectory data. *Transportation Research Part D: Transport and Environment*, 122, 103887.
- Yu, R., Han, L., & Zhang, H. (2021). Trajectory data-based freeway high-risk events prediction and its influencing factors analyses. *Accident Analysis & Prevention*, 154, 106085.
- Zhou, L., Wang, C., & Zhen, F. (2023). Exploring and evaluating the spatial association between commercial and residential spaces using Baidu trajectory data. *Cities*, 141, 104514.