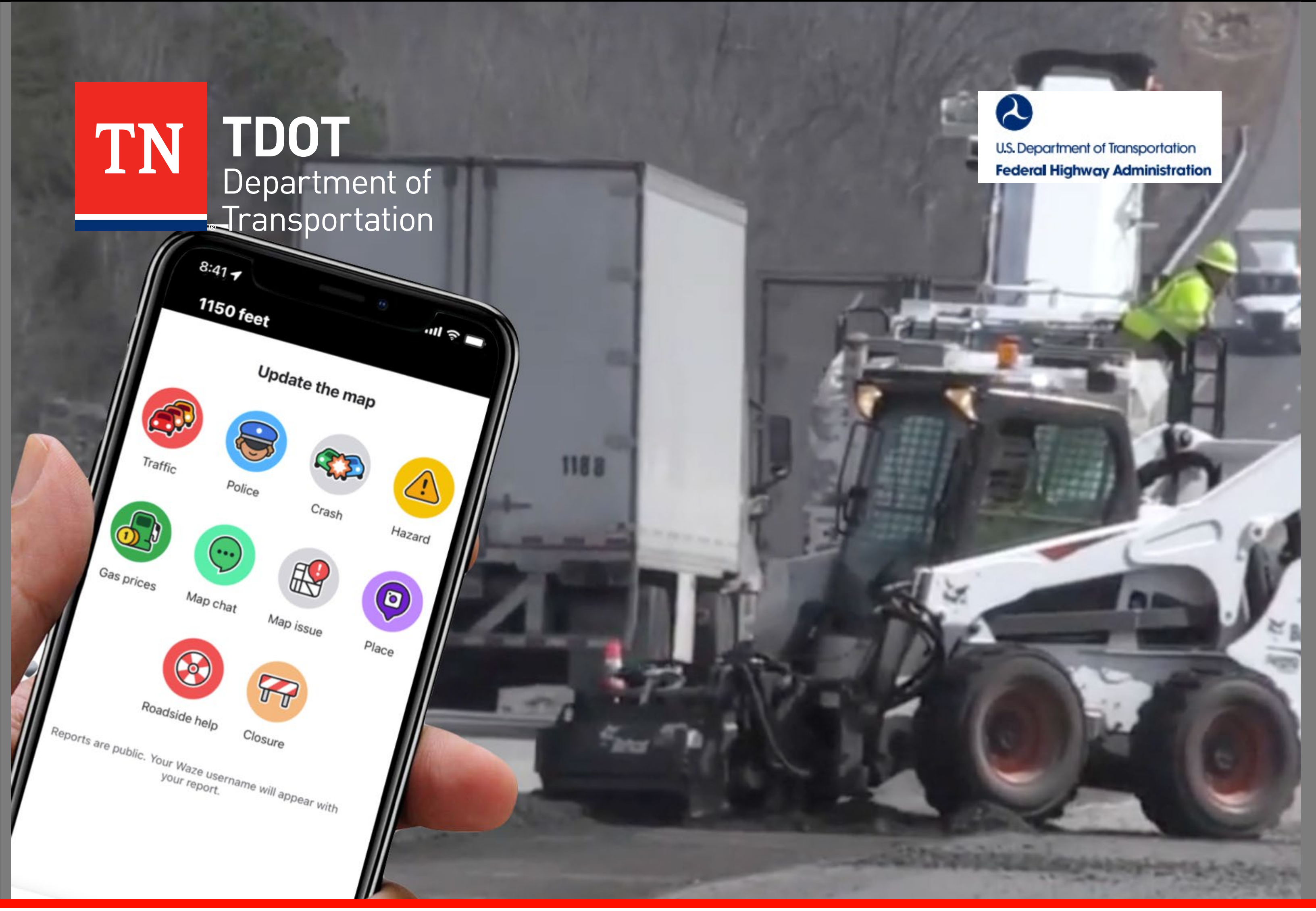


Waze Data Reporting

Research Final Report from The University of Tennessee | Yangsong Gu, Hairuilong Zhang, Candace Brakewood, and Lee D. Han | May 30, 2022

Sponsored by Tennessee Department of Transportation Long Range Planning Research Office & Federal Highway Administration

DISCLAIMER

This research was funded through the State Planning and Research (SPR) Program by the Tennessee Department of Transportation and the Federal Highway Administration under ***RES2019-11 Waze Data Reporting***.

This document is disseminated under the sponsorship of the Tennessee Department of Transportation and the United States Department of Transportation in the interest of information exchange. The State of Tennessee and the United States Government assume no liability of its contents or use thereof.

The contents of this report reflect the views of the author(s) who are solely responsible for the facts and accuracy of the material presented. The contents do not necessarily reflect the official views of the Tennessee Department of Transportation or the United States Department of Transportation.

Technical Report Documentation Page

1. Report No. RES2019-11	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle *WAZE Data Reporting*		5. Report Date May 30, 2022 6. Performing Organization Code
7. Author(s) Yangsong Gu, Hairuilong Zhang, Candace Brakewood, and Lee D. Han		8. Performing Organization Report No.
9. Performing Organization Name and Address Center for Transportation Research The University of Tennessee Knoxville, TN, 37996-4133		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Tennessee Department of Transportation 505 Deaderick Street, Suite 900 Nashville, TN 37243		13. Type of Report and Period Covered 14. Sponsoring Agency Code

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

16. Abstract
This study evaluated the quality of crowdsourced Waze data (including reports and speed) and explored promising use scenarios of Waze data to facilitate the development of intelligent transportation in Tennessee. To this end, the thoroughly assessed Waze reports quality in terms of spatiotemporal accuracy and coverage. The study found Waze users reported crash events about 2.2 minutes sooner, on average, than reports of the same events recorded in the state's Locate/IM incident log. The reported crash locations per Waze are on average 6 feet from the Locate/IM log reported by the officials. It is found that 26% of crashes reported in Waze was matched with 67% Locate/IM crash reports, with the rest 74% reports pointing to unreported incidents. Waze speed is affected by the Wazers behaviors and tends to be slightly higher than detector speed in free-flow status. This study evaluated several novel use scenarios such as secondary crash detection, end of queue detection and tracking, level of service evaluation, work zone monitoring, wildlife hazards and crashes, and pothole detection and maintenance. Results show that Waze is a suitable data source for incident management, level of service evaluation, work zone management, roadway maintenance management, etc. when properly used and in cooperation with the agency's other information sources.

17. Key Words **WAZE DATA, QUALITY ASSESSMENT, USE SCENAIROS**	18. Distribution Statement		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages xxx	22. Price

Acknowledgement

The research team is grateful for the sponsorship and guidance from Tennessee Department of Transportation (TDOT) and Federal Highway Administration (FHWA). Traffic data provided by TDOT, and Google/WAZE were essential to the efforts and findings presented herein. A special thank you goes to TDOT Research Office for their support in completing this report.

We should also mention that Dr. Yuandong Liu, Dr. Zhihua Zhang, and Dr. Nima Hoseinzadeh contributed significantly to this research endeavor during their doctoral studies at the University of Tennessee.

Executive Summary

This study sought to explore and assess Waze data as a viable data/information source for TDOT traffic operation and management purposes. Waze crowdsourced data are obtained whenever and wherever Wazers travel. This has the potential of providing an additional information source independent to the TDOT's existing infrastructure, such as TMCs, RDS detectors, and HELP trucks, to monitor traffic condition in real time. The study examined the data quality through comparing Waze data with ground truth, studied use cases reported by other agencies, and explored the feasibility and proposed deployment scenarios for TDOT, including queueing management, performance measurement, incident management, roadway maintenance, and work zone management.

Key Findings

- In general, Waze incident reports are timelier, by 2.2 minutes on average for crashes and 7.8 minutes for disabled vehicles, than reports from other means. Waze reported incident locations are typically very accurate, within 6 feet on average, to the verified crash sites.
- Crashes and disabled vehicles logged in the Locate/IM system only cover a small portion of all Waze reports of the same type. Waze reports are particularly useful, in terms of spatial coverage, in areas outside of the major urban areas where FSSP programs do not cover.
- The availability and accuracy of Waze speed data on a specific roadway link depends on having sufficient Waze users traversing that section of roadway. Roadways with higher traffic volume tend to have better data Waze data quality.
- Use scenarios successfully demonstrated in this study include using Waze for EOQ detection, secondary crashes identification, roadway performance LOS estimation, incident data supplement, incident management assistance, pavement (pothole) management, and work zone management.

Key Recommendations

- TDOT could employ real-time WAZE report and data for disseminating incident, stop-and-go traffic, and work zone information.
- TDOT could expand its FSSP services beyond the existing major urban areas into rural area-based incident statistics from WAZE reports and utilizing real-time WAZE reports for dispatching decisions.
- TDOT could supplement its Locate/IM incident records with WAZE reported incidents. Spatial-temporal clustering assessment could be performed to identify hazardous scenarios and problematic spots.
- TDOT should consider systematic data imputation procedures for all its multitudes of data including RDS, WAZE, NPMRDS, WIM, etc. to improve data quality.
- TDOT should consider use scenarios evaluated and proposed in this study for implementation to save lives and save money.

Table of Contents

List of Tables

List of Figures

Glossary of Key Terms and Acronyms

AADT	Average Annual Daily Traffic
A&M	Agricultural and Mechanical
AM	Ante meridiem, which means "before noon."
API	Application Programming Interface
APP	Application software
ATMS	Advanced Traffic Management System
Blvd.	Boulevard
BTI	Buffer Time Index
CCP	Connected Citizens Program of Waze
COV	Covariance
DBSCAN	Density-based spatial clustering of applications with noise
DC	Districted of Columbia
DDOT	District Department of Transportation of Washington DC
DelDOT	Delaware Department of Transportation
DMS	Dynamic Message Sign
DOT	Department of Transportation
EOQ	End of Queue
ETRIMS	Enhanced Tennessee Roadway Information Management System
FHWA	Federal Highway Administration
FSSP	Freeway Service Safety Patrol
GPS	Global Positioning System
HELP	TDOT's Highway Incident Response Unit
i.e.	Latin id ist or "that is"
INRIX	A private company that provides location-based traffic data and software-as-a-service analytics
IQR	Interquartile Range
JSON	A data file format system
KYTC	Kentucky Transportation Cabinet

Locate/IM	Locate Incident Management
LOS	Level of Service
MAE	Mean Absolute Error
MOE	Measure of Effectiveness
MPH	Miles per hour
NB	Northbound
NPMRDS	National Performance Management Research Data Set
PCMS	Portable Changeable Message Signs
PDI	Pavement Distress Index
PM	Post Meridiem, which means “after noon.”
PTI	Planning Time Index
PTQ	Protect the Queue
RDS	Radar Detection System
RMSE	Root Mean Square Error
RTMS	Remote Traffic Microwave Sensors
SB	Southbound
SD	Standard Deviation
SE	Standard Error
SPR	State Planning and Research Program
St.	Street
ST-DBSCAN	Spatiotemporal Density-based spatial clustering of applications with noise
TDOT	Tennessee Department of Transportation
THP	Tennessee Highway Patrol
TMC	Traffic Management Centers
TN	Tennessee
TRB	Transportation Research Board
TTI	Travel Time Index
USA	United States of America

UTC	Coordinated Universal Time
UUID	Universally Unique Identifier
W4C	Waze for Cities
MAPE	Mean Absolute Percent Error
WAZE	A subsidiary of Google company that collects real-time traffic information like travel times, and traffic incidents from users.
Wazer	Waze User
WARP	Waze Analytics Relational-database Platform
WIM	Weigh in Motion
XML	Extensible Markup Language

Chapter 1 Introduction

Over the past two decades, the Tennessee Department of Transportation (TDOT) [1] has invested significantly towards real-time traffic management and improved substantially motorist safety and operational efficiency. To augment TDOT's existing HELP Freeway Safety Service Patrols (FSSP) service, the Traffic Management Centers (TMCs) in Tennessee's major metropolises, the hundreds of freeway traffic cameras, Dynamic Message Signs (DMS) and the Protect the Queue program for major events, TDOT came to agreement with Waze for Cities (W4C)[1] to share data. The agency could potentially use the crowd-sourced travel data and traffic report to enhance real-time traffic incident management and advance the understanding of spatiotemporal incident trends on the State's most travelled roadways for infrastructural investment decisions and traffic management strategies in general.

As a subsidiary of Google company, Waze collects real-time traffic information like travel times, and traffic incidents from Waze users ("Wazers"). Distinct from the Google Map, Waze app allows Wazers to report different kinds of roadway incidents, e.g., accidents, traffic jams, speed, police traps, and so son, whereas Wazers' location and identify are anonymous In turn, via collecting real-time traffic information, Waze is not only able to share the traffic status and incidents with other Wazers, but also to dynamically detour and reroute for Wazers. Waze has its own built-in credit mechanism to evaluate the reliability of reports. Wazers who often report true incidents tend to have high reliability [3].

1.1 Scope and Objectives

The goal of this project is to thoroughly assess the Waze data for insights and strategies that could enhance TDOT's existing traffic operation, management, and planning programs. To that end, three objectives were identified:

Objective 1. Waze data assessment and analysis– An extensive study of various Waze data, primarily the incident reports and link-based travel time data, will be performed to assess their accuracy, coverage, and reliability in comparison with TDOT's existing Locate/IM, RTMS and NPMRDS data.

Objective 2. State of the practice Waze use cases and studies – The research team will sustain a continuous effort reviewing and keeping track of successful with Waze for cities program (WCP) at other state DOT or transportation agencies and evaluate how such might be transplanted to Tennessee. Various use cases and deployment experience will be studied and be presented to TDOT.

Objective 3. Waze scenarios for TDOT development – Based on current needs of TDOT and successful Waze implementation identified in objective 2, a list of use scenarios will be identified and compared for potential development and deployment in Tennessee. In consultation with TDOT, the research team will further identify promising scenarios for refinement and

[1]The Waze for Cities (W4C) program, formerly known as Connected Citizens Program (CCP) [2] Wazeopedia. https://wazeopedia.waze.com/wiki/USA/Waze_for_Cities (accessed Feburary 2, 2022)..

development. A report on these use scenarios, how they could be implemented in Tennessee, and their pros and cons will be presented to TDOT.

1.2 Methodology and Outcomes

Waze data have many promising applications, yet the data quality is a concern. Data redundancy and unreliability are two major shortcomings of Waze data. To overcome this, the researchers employed spatial-temporal thresholds to filter the Waze reports based on ground truth. The thresholds are carefully selected based on the match rates pattern. After matching Waze reports with ground-truth data, the first report is elected to assess the spatial and temporal accuracy. Besides, the match rate is further used to assess the event completeness. To assess the speed, the researchers investigated the arithmetic difference between Waze speed and Bluetooth and RTMS detector speed for different speed levels and time of the day. The second predominant approach for exploring use scenarios is spatiotemporal DBSCAN (Density-based spatial clustering of applications with noise) clustering. The basic spatiotemporal clustering is implemented to discover similar Waze reports and filter out noise reports (so-called false reports). Upon the basic model, the researchers extended the model to dynamic ST-DBSCAN, and network-based ST-DBSCAN for powering real-world applications.

1.3 Report Organization

This remainder of this report is organized as the following sequence. Chapter 2 presents a synoptic review of State-of-the-practice use scenarios and studies. Chapter 3 describes the methodologies for Waze data quality assessment and use scenarios. Chapter 4 elaborates the findings and proposed the promising use scenarios for TDOT development. Chapter 5 closes the report with conclusions and recommendations. The core section of Waze data processing is followed by the below framework (see

Figure 1-1).

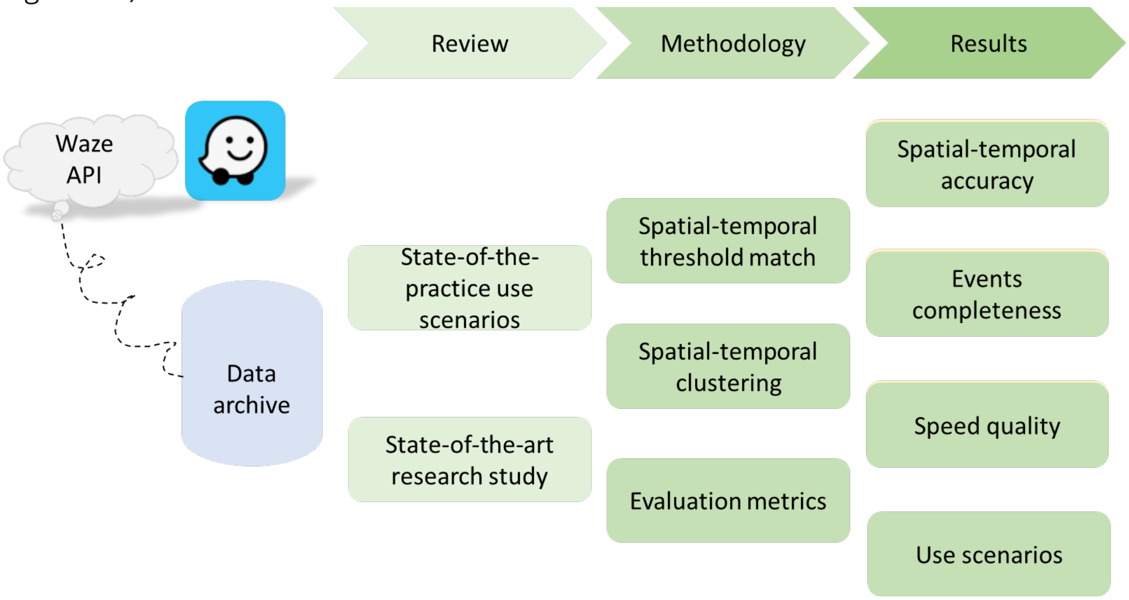


Figure 1-1 Framework of processing and analyzing Waze data

Chapter 2 Literature Review

This chapter introduced Waze data and its associated challenges. Then a series of state-of-the-practice use scenarios are collected which provides the insights on developing use scenarios for TDOT. Besides, the state-of-the-art research studies are also summarized for proposed studies.

2.1 Waze Data

Nowadays, mobile crowdsourcing becomes increasingly popular due to the proliferation of smart devices. A promising example in transportation is the navigation app Waze, which has over 100 million monthly active users worldwide [4]. Wazers are encouraged to share the real-time traffic status, accidents, and other roadway incident information through the app report portal. TDOT has established a partnership with Waze through the Waze for Cities program to share the data. Under the agreement, TDOT can access real-time incident reports and speed data. The major incident reports mainly include accident, jam, weather hazard, and road closed reports. Each major incident report is further divided into different subtypes. For instance, the accident alert has “ACCIDENT_MINOR” and “ACCIDENT_MAJOR” subtypes. This finer description of the incident reflects the perception of Wazers who are in the incident context, which can help drivers understand the severity of incidents. Besides, a typical incident report usually contains user identification, timestamp, street name, latitude and longitude, reliability and confidence of Wazers’ report, and road types. In addition to miscellaneous traffic reports, Waze traffic view enables transportation agencies to add own-defined routes to a watchlist. Once the route appears on the watchlist, Waze will update the speed, real-time travel time, historical travel time, and jam level of designated route every 2 minutes.

2.2 Challenges of Crowdsourced Waze Data

Although crowdsourced data has inviting characteristics, such as cost-effective, wide coverage and so on, the unreliability and the redundancy have been two major concerns when they are employed. Specifically, the unreliability refers to false alarms. This is likely to happen when Wazers intentionally report nonexistent incidents, or they misclassify the incident types. Therefore, some studies have been conducted to evaluate the truthiness of Waze reports. For instance, Goodall and Lee (2019) evaluated the accuracy of Waze crash and disabled vehicle reports along a 2.7-miles section of urban freeway by comparing Waze reports with ground truth from video footage. Among 40 crashes associated with the Waze reports, 2 (5%) were confirmed false alarms. By contrast, disabled vehicle reports have more false alarms than crash reports. Of 560 disabled vehicle reports, 131 (23%) were identified as false alarms. Apart from the primary report, the false alarms should get attention as they will interfere the decision of emergency response. The redundancy is due to multiple Wazers reporting the same incident.

To mitigate the data redundancy, some existing studies proposed different approaches. The majority of them applied a threshold-based approach as these redundant reports tend to locate at the vicinity area of incident locale and be close to time of incident occurrence. Therefore, when the distance difference and time difference between Waze reports and ground truth are within the threshold, they are assumed to report on the same event. In other words, they are redundant reports. For instance, Liu et al. [5] enumerate all possible combinations of spatiotemporal thresholds to match Waze crash and disabled vehicle reports with official incident reports (i.e., Locate IM). They found that matching rates would stabilize when time threshold exceeds 30 minutes and space threshold exceeds 1.5 miles for crash cases. The time and space threshold for disabled vehicles are 50 minutes and 1 mile, respectively. It is worth noting that the time and space thresholds vary for different events as their impact on traffic status are totally different.

2.3 State-of-the-practice use scenarios

Many transportation agencies have successfully partnered with Waze through the W4C program. Waze data plays a valuable role in different use scenarios. Herein, several successful cases are reviewed and summarized.

Iowa Department of Transportation

Iowa DOT has used Waze data as a source of incident detection in the Advanced Traffic Management System (ATMS) since September 2015 [1]. Waze reports detection ranked fourth in 23 detection sources, accounting for 13.4% of all ATMS records. Additionally, it was reported that approximately 12% of incidents are solely identified by Waze reports [6]. In addition, Iowa DOT improved 5-1-1 services by broadcasting traffic jam and crash detected by Waze reports (see

Figure 2-1) [7, 8].

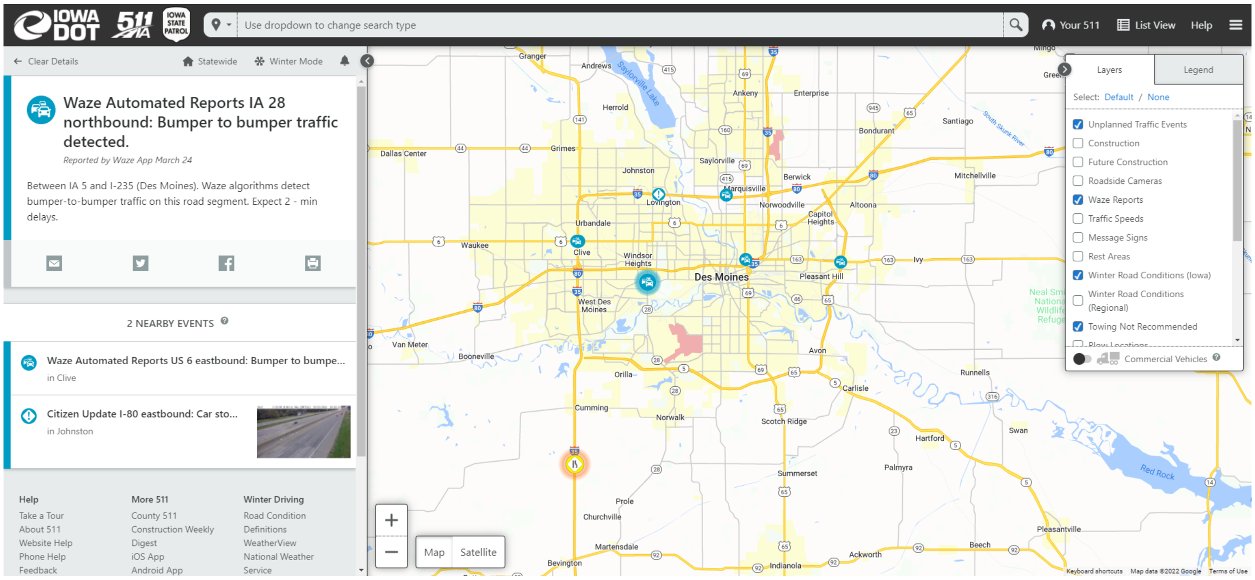


Figure 2-1 Iowa DOT 5-1-1 service map

Lake County DOT

Lake County DOT extensively employed Waze travel times for routing and signal management system. Lake County incorporate its regions' road network in the Waze Traffic View tool using approximately 600 directional road segments [9]. Meanwhile, they routinely post Waze travel times on portable changeable message signs in the areas with the greatest travel time variability. Additionally, Lake County used Waze travel time data for automated traffic signal performance measures such as delay, travel time index, and so on, thereby performing signal coordination and timing. The free and frequent Waze travel time data not only offset the signal system's cost by eliminating field data collection, but also yield benefits for frequently conducting signal coordination. A real-world case in December 2019 demonstrated that crowdsourced data informed signal timing resulted in far less travel times than law enforcement directed detours [9].

Louisville, Kentucky

In 2015, Louisville started to share road closure information with Waze through W4C partner portal, enabling Wazers to bypass active road closures. Besides, the City of Louisville established a Waze Analytics Relational-database Platform (WARP) to help region prompt the data-centered management of traffic and signal operations at a cost-effective way. This WARP platform can perform before and after signal retiming evaluation by Waze jam reports, which significantly saves labor and cost for field data collection [10].

Kentucky Transportation Cabinet (KYTC)

Many workers and motorists lost their lives in work zone crashes every year. The Work Zone safety has been one of the top priorities in transportation agencies. Kentucky Transportation Cabinet blends Waze traffic speeds, incident reports with traditional data source to monitor the crash and traffic jam near work zones. The developed work zone monitoring dashboard (see **Figure 2-2**) is widely used by operators, construction safety officers, and planners [11].

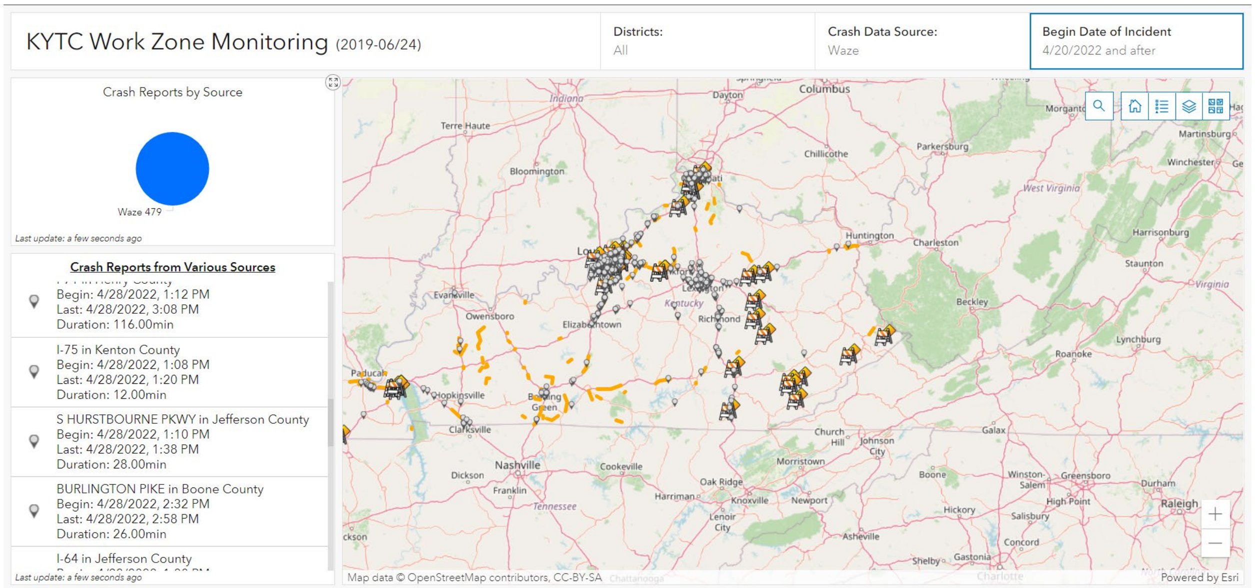


Figure 2-2 Kentucky Transportation Cabinet Work Zone Monitoring Website

Washington DC DOT

In *Washington DC*, which was one of the earliest cities to work with WAZE, WAZE users can use a unique audio feature in the app that enables them to report potholes to the District Department of Transportation (DDOT), enabling DDOT to quickly find and fix potholes [12].

Delaware DOT (DelDOT)

DelDOT connects Waze crowdsourced data with 511 services via a DelDOT travel advisory app. Users can see information on Waze incidents and traffic slowdowns. In turn, 511 services disseminate real-time traffic information from crowdsourced Waze data to users [13].

2.4 State-of-art research study

In addition to these examples of successful implementation by transportation agencies, there are also numerous academic research studies that have recently used Waze data. This brief review highlights some of the most recent and relevant prior literature.

Waze data has been tested the accuracy, coverage, and timeliness by many partners. One TRB paper evaluated the reliability and coverage of Waze data for the state of Iowa [6]. After analyzing one year of Waze data and comparing it to the Iowa's advanced traffic management system (ATMS), the researchers concluded that WAZE data has broad coverage (including approximately 43% of ATMS crash and congestion reports) and is received in a timely manner (on average, 9.8 minutes earlier than probe data). Another research team from Virginia Transportation Research Council [14] compared Waze crash and disabled vehicle records with video ground truth. They found 33% (13 out of 40) Waze accident reports were primary reports, and 5 (2 out of 40) were labeled as false alarms. For disabled vehicles, 22% (125 out of 560) of Waze reports were primary reports and 23% (131 out of 560) were false alarms. Among 13 matched crashes, Wazers reported crash prior to Transportation operations center operators' awareness. These evaluation practices suggest that Waze has merits of sufficient geographic accuracy, timely reporting, and broad coverage.

Waze data recently have been employed in flood risk management. A research paper assessed Waze flood Incident Reports for Norfolk, Virginia. The study found out that 71.7% (502 out of 697) of Waze flood reports can be considered trustworthy based on their topographic, environmental, temporal and peer reporting characteristics [15]. A paper from nature scientific report evaluated the application of Waze flood reports in detecting flash floods. The research team compared the Waze flood reports with local storm reports and found that Waze reports can provide real-time situational alerts timely and accurately [16]. Those compelling evidence demonstrate that Waze flood reports can potentially complement existing flood observation in flood management.

Waze data contribute road safety analysis. The researchers from Texas A&M University explored the potential of using Waze incident reports to identify high-risk road segments. The study found that adding Waze incident reports can help identify more high-risk segments than solely relying on police reports. In addition, Waze incident reports indicated high correlation with police reports, which can be used a s strong predictor in crash prediction [17]. Besides, Waze reports are also useful for incident early detection. Senarath et. al. [18] employed Waze accident and jam reports to detect the crash. Their proposed model can detect the incident ahead of 5.92 minutes on average than police report. These application of Waze data can bring huge benefits to emergency response operation.

Waze data facilitate traffic systems management and operations. One paper presented at the 2017 annual meeting of the Transportation Research Board (TRB) analyzed Waze data for Washington DC to evaluate *traffic speeds on urban arterial*s and compared it to data from the local Vehicle Detection System and INRIX [19]. The results suggest that Waze data provides an accurate, low-cost alternative to measure traffic speeds after correcting for some biases in the data. Another TRB paper [20] interviewed staff from state DOTs to assess *how different agencies are utilizing crowdsourced data* at their traffic management centers. A key finding from this study was that Waze W4C program has been “extremely popular and beneficial for many state DOTs/TMCs, particularly in mining information from areas with poor coverage,” such as rural areas with limited cameras or other means of identifying incidents.

Chapter 3 Methodology

This chapter presents the Waze data collection effort, the methodology for assessing the quality of the collected Waze data, and the approaches for considering various use scenarios. The quality of the Waze data was assessed in several categories including event location accuracy, event time accuracy, event report completeness, and speed accuracy. The methodologies of these assessments are described in the following sections.

3.1 Data Collection

Before all the assessment and analysis efforts could commence, the collection of data, specifically Waze data, was essential to this study. To that end, the research team began to acquire and archive real-time Waze data at essentially minute-by-minute frequency on a dedicated server. There are two main types of information records from Waze that were archived in the respective log files.

Waze Event Log **–** Waze disseminates event information based on crowdsourced reports from Wazers with subtypes, including construction, stopped vehicle, heavy traffic, accident (also known as crash to some), etc., in correspondence with Waze app report buttons, see Figure 3-1. A sample event log in its native XML format is shown in **Figure 3-2**.

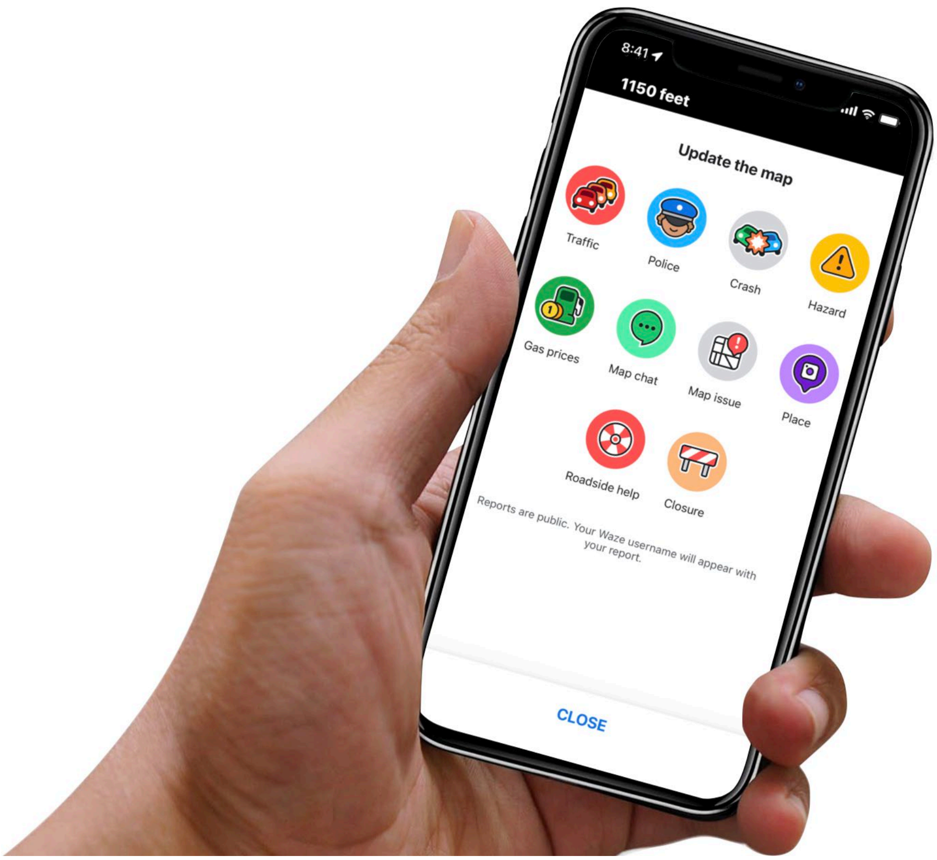
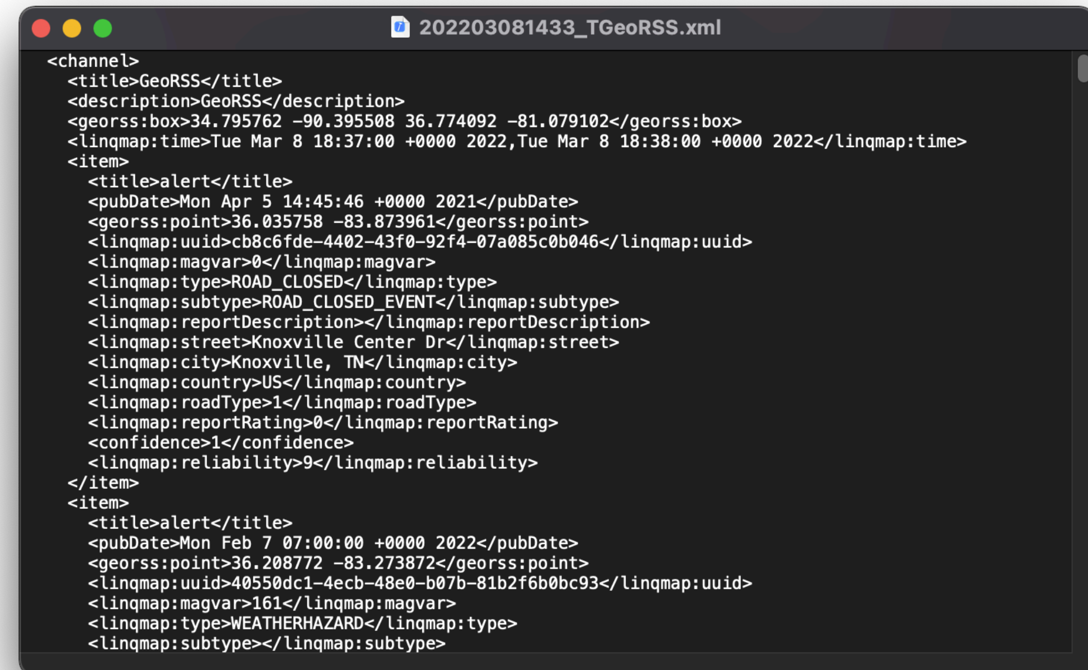


Figure 3-1 Waze app screen of reportable events

Since Waze app refreshes traffic information (i.e., speed and reports) every two minutes, the researchers download and store Waze data every minute to prevent any missing reports. Each file contains information on all reports within the *Waze for Cities* partner's geographical area of interest (i.e., the entire state of Tennessee) that were active during the reporting period (i.e., one minute).As is shown in **Figure 3-2**, the contents of the file are organized in individual report blocks. Each includes a report title, report publication date, geolocation, report type, street, city, state, country, etc. After parsing the XML files and aggregating the records over time, one could compile a more comprehensive log file like the one shown in **Figure 3-3** with the ability to sort, query, and examine a number of events reported by Wazers along different roadways in different cities at different time and day.



```
202203081433_TGeoRSS.xml
<channel>
  <title>GeoRSS</title>
  <description>GeoRSS</description>
  <georss:box>34.795762 -90.395508 36.774092 -81.079102</georss:box>
  <linqmap:time>Tue Mar 8 18:37:00 +0000 2022,Tue Mar 8 18:38:00 +0000 2022</linqmap:time>
  <item>
    <title>alert</title>
    <pubDate>Mon Apr 5 14:45:46 +0000 2021</pubDate>
    <georss:point>36.035758 -83.873961</georss:point>
    <linqmap:uuid>cb8c6fde-4402-43f0-92f4-07a085c0b046</linqmap:uuid>
    <linqmap:magvar>0</linqmap:magvar>
    <linqmap:type>ROAD_CLOSED</linqmap:type>
    <linqmap:subtype>ROAD_CLOSED_EVENT</linqmap:subtype>
    <linqmap:reportDescription></linqmap:reportDescription>
    <linqmap:street>Knoxville Center Dr</linqmap:street>
    <linqmap:city>Knoxville, TN</linqmap:city>
    <linqmap:country>US</linqmap:country>
    <linqmap:roadType>1</linqmap:roadType>
    <linqmap:reportRating>0</linqmap:reportRating>
    <confidence>1</confidence>
    <linqmap:reliability>9</linqmap:reliability>
  </item>
  <item>
    <title>alert</title>
    <pubDate>Mon Feb 7 07:00:00 +0000 2022</pubDate>
    <georss:point>36.208772 -83.273872</georss:point>
    <linqmap:uuid>40550dc1-4ecb-48e0-b07b-81b2f6b0bc93</linqmap:uuid>
    <linqmap:magvar>161</linqmap:magvar>
    <linqmap:type>WEATHERHAZARD</linqmap:type>
    <linqmap:subtype></linqmap:subtype>
```

Figure 3-2 Sample event log file in XML format

Waze Speed Report – The second piece of Waze information is related to segment speed and travel time. Depending on the highway segments selected by the specific *Waze for Cities* partner of interest (i.e., TDOT in this study), travel time data for these segments based on aggregated Wazers' Global Positioning System (GPS) samples within each segment are disseminated in JSON files, see **Figure 3-4**. Details in terms of the physical descriptions of the segment, segment length, and so on are provided in the JSON files, along with the travel time of the most recent period and the historical travel time for this time of day and day of the week.

TDOT Radar Detector System (RDS) Data – In addition to Waze data, additional data were acquired for the purpose of this study. These include TDOT's own RDS traffic data from hundreds of stations in the state's major urban areas, see **Figure 3-5**. The contents consist of lane-by-lane traffic volume, average speed, and occupancy data for the past 30 seconds.

	D	E	F	G	H	I	J	K	M	N
1	type	confidence	reliability	subtype	street	city	Lat	Long	localtime	localdate
135115	WEATHERHAZARD	0	5	HAZARD_ON_ROAD_CONSTRUCTION	I-24 W	Nashville, TN	36.105397	-86.717538	2/22/20 13:51	2/22/20
135116	WEATHERHAZARD	0	6	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.035664	-86.644121	2/22/20 13:57	2/22/20
135117	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.069803	-86.689657	2/22/20 14:16	2/22/20
135118	WEATHERHAZARD	0	6	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.311442	-86.827687	2/22/20 14:31	2/22/20
135119	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.163455	-86.762002	2/22/20 14:47	2/22/20
135120	ACCIDENT	0	5	ACCIDENT_MINOR	I-24 W	Nashville, TN	36.16441	-86.763039	2/22/20 14:47	2/22/20
135121	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.295965	-86.810938	2/22/20 14:54	2/22/20
135122	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.036677	-86.645257	2/22/20 15:00	2/22/20
135123	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.097964	-86.709459	2/22/20 15:07	2/22/20
135124	JAM	0	5	JAM_STAND_STILL_TRAFFIC	I-24 W	Nashville, TN	36.157019	-86.759722	2/22/20 15:15	2/22/20
135125	JAM	0	5	JAM_HEAVY_TRAFFIC	I-24 W	Nashville, TN	36.160908	-86.760225	2/22/20 15:17	2/22/20
135126	ACCIDENT	0	5		I-24 W	Nashville, TN	36.164861	-86.763503	2/22/20 15:25	2/22/20
135127	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.108546	-86.720725	2/22/20 15:29	2/22/20
135128	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.250936	-86.784088	2/22/20 15:54	2/22/20
135129	JAM	0	5	JAM_HEAVY_TRAFFIC	I-24 W	Nashville, TN	36.140769	-86.729793	2/22/20 16:02	2/22/20
135130	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.074282	-86.692682	2/22/20 16:14	2/22/20
135131	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.103544	-86.71482	2/22/20 16:16	2/22/20
135132	JAM	0	5	JAM_HEAVY_TRAFFIC	I-24 W	Nashville, TN	36.133527	-86.726733	2/22/20 16:18	2/22/20
135133	JAM	0	5	JAM_HEAVY_TRAFFIC	I-24 W	Nashville, TN	36.183312	-86.775028	2/22/20 16:25	2/22/20
135134	JAM	0	5	JAM_STAND_STILL_TRAFFIC	I-24 W	Nashville, TN	36.129002	-86.727288	2/22/20 16:29	2/22/20
135135	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.057076	-86.676545	2/22/20 16:37	2/22/20
135136	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.277445	-86.800029	2/22/20 16:41	2/22/20
135137	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.309835	-86.824917	2/22/20 16:47	2/22/20
135138	WEATHERHAZARD	0	6	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.163722	-86.762292	2/22/20 16:47	2/22/20
135139	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.177491	-86.772758	2/22/20 16:48	2/22/20
135140	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.258566	-86.787523	2/22/20 16:55	2/22/20
135141	JAM	0	5	JAM_STAND_STILL_TRAFFIC	I-24 W	Nashville, TN	36.136825	-86.726759	2/22/20 17:09	2/22/20
135142	JAM	0	5	JAM_HEAVY_TRAFFIC	I-24 W	Nashville, TN	36.188731	-86.775387	2/22/20 17:19	2/22/20
135143	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.162746	-86.761282	2/22/20 17:27	2/22/20
135144	JAM	0	5	JAM_STAND_STILL_TRAFFIC	I-24 W	Nashville, TN	36.136229	-86.726755	2/22/20 17:39	2/22/20
135145	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.248103	-86.783636	2/22/20 17:44	2/22/20
135146	JAM	0	5	JAM_MODERATE_TRAFFIC	I-24 W	Nashville, TN	36.132799	-86.726736	2/22/20 17:45	2/22/20
135147	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.140889	-86.730269	2/22/20 17:46	2/22/20
135148	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.290012	-86.8097	2/22/20 17:54	2/22/20
135149	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.296045	-86.810955	2/22/20 17:55	2/22/20
135150	JAM	0	5		I-24 W	Nashville, TN	36.140632	-86.729253	2/22/20 18:13	2/22/20
135151	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.309243	-86.823703	2/22/20 18:14	2/22/20
135152	WEATHERHAZARD	0	5	HAZARD_ON_SHOULDER_CAR_STOPPED	I-24 W	Nashville, TN	36.163881	-86.762464	2/22/20 18:26	2/22/20

Figure 3-3 Sample Waze event log of hazards, accidents, and jams

202203081433_JSON

{"usersOnJams":[{"wazersCount":2677,"jamLevel":0},{"wazersCount":0,"jamLevel":1},
{"wazersCount":10,"jamLevel":2},{"wazersCount":75,"jamLevel":3},
{"wazersCount":24.5,"jamLevel":4}],"routes":[{"toName":"2036134","historicTime":55,"line":
[{"x":-83.54968104148107,"y":35.86529268274319},{"x":-83.54774331779099,"y":35.864845607829885},
{"x":-83.54774331779099,"y":35.864845607829885},{"x":-83.5473939038962,"y":35.86475954158312},
{"x":-83.5473939038962,"y":35.86475954158312},{"x":-83.54706809971978,"y":35.86468426370773},
{"x":-83.54706809971978,"y":35.86468426370773},{"x":-83.54625999999999,"y":35.864501000000004},
{"x":-83.54625999999999,"y":35.864501000000004},{"x":-83.545663595199,"y":35.864377203287},
{"x":-83.545663595199,"y":35.864377203287},{"x":-83.54525847817325,"y":35.86429781210299},
{"x":-83.54525847817325,"y":35.86429781210299},{"x":-83.545081001312,"y":35.86426303176201},
{"x":-83.5449026939672,"y":35.8642338769677},{"x":-83.5449026939672,"y":35.8642338769677},
{"x":-83.54478832697865,"y":35.86421517696948},{"x":-83.54478832697865,"y":35.86421517696948},
{"x":-83.544616,"y":35.864187},{"x":-83.54459989180287,"y":35.86418473742612},
{"x":-83.54459989180287,"y":35.86418473742612},{"x":-83.544203142958,"y":35.864129009656},
{"x":-83.544203142958,"y":35.864129009656},{"x":-83.543962552621,"y":35.864095216058},
{"x":-83.5439625526210 1,"y":35.864095216058},{"x":-83.5435328896114,"y":35.864053147897806}],"bbox":
{"minY":35.864053147897806,"minX":-83.54968104148107,"maxY":35.86529268274319,"maxX":-83.543532889611
4},"name":"","fromName":"2038171","length":573,"jamLevel":3,"id":9351,"time":122,"type":"STATIC","jam
s":[]},{"toName":"2099949","historicTime":87,"line":[{"x":-83.570163,"y":35.849422},
{"x":-83.57019300329256,"y":35.849091429891814},{"x":-83.57019300329256,"y":35.849091429891814},
{"x":-83.57020012079752,"y":35.84901301048334},{"x":-83.570275,"y":35.848188},
{"x":-83.570275,"y":35.848188},{"x":-83.57038299999999,"y":35.846984},
{"x":-83.57038299999999,"y":35.846984},{"x":-83.570483,"y":35.84587},{"x":-83.570483,"y":35.84587},
{"x":-83.57061,"y":35.844561},{"x":-83.57061,"y":35.844561},
{"x":-83.57071284426117,"y":35.843758166117325},{"x":-83.57071284426117,"y":35.843758166117325},
{"x":-83.570804,"y":35.843061},{"x":-83.570804,"y":35.843061},
{"x":-83.57085071225181,"y":35.84236387907421},{"x":-83.57085071225181,"y":35.84236387907421},
{"x":-83.570909,"y":35.841494},{"x":-83.570909,"y":35.84149399999999},
{"x":-83.5709279421221,"y":35.84105297356577},{"x":-83.5709279421221,"y":35.84105297356577},
{"x":-83.571022421122,"y":35.83985871559598},{"x":-83.571022421122,"y":35.83985871559598},

Figure 3-4 Sample Waze travel time JSON file

Other Data – Other data acquired for the purpose of this study include traffic incidents recorded in the state's Locate/IM files, crash records in the state's Enhanced Tennessee Roadway Information Management System, National Performance Management Research Data Set, and TDOT's pothole records.

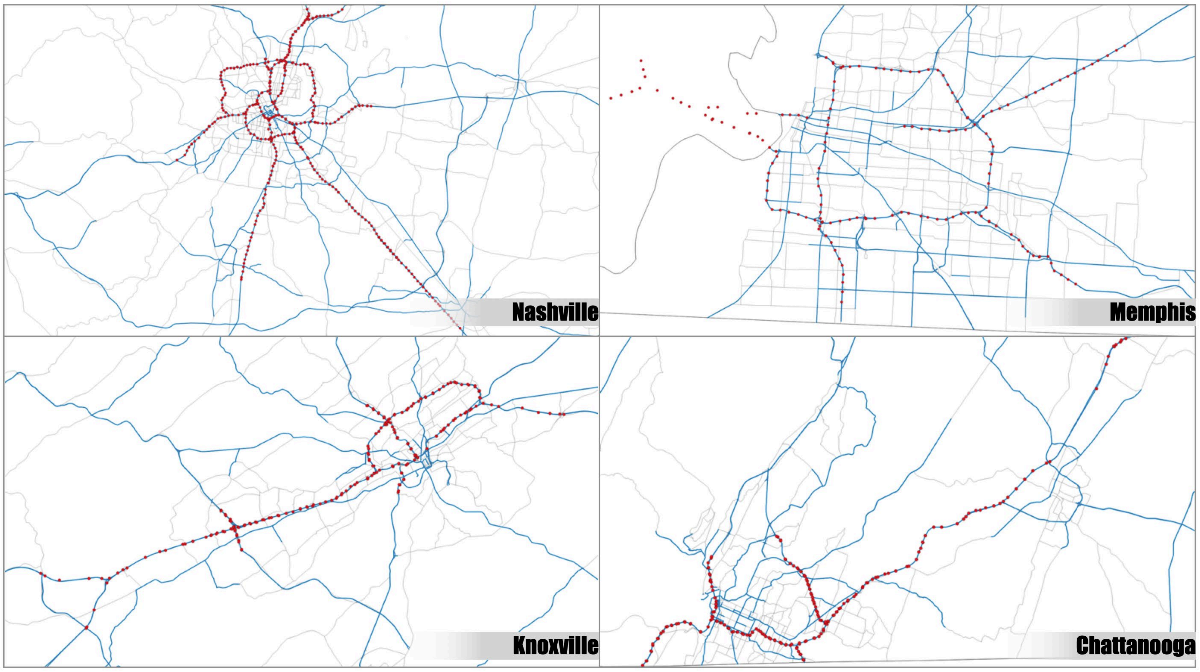


Figure 3-5 TDOT RDS station locations in major urban areas

3.2 WAZE Data Quality Assessment

As mentioned in literature section, data redundancy and unreliability are two big concerns of Waze data. For one thing, different Wazers might report the same incident prior or post the incident. For another, Wazers might misreport the incident. Therefore, to filter these potential redundant reports and noise reports, the researchers test different combinations of spatial and temporal thresholds for matching Waze reports with ground truth (e.g., Locate IM). This matching process can be divided into three steps, which are:

Step 0. Prepare data.

The Step 0 mainly focuses on processing the Waze location and timestamp information. Specifically, we projected the Waze reports to road networks, and compute the mile markers of Waze reports. Similarly, we also labeled the ground truth data using mile markers. This makes easy for computing spatial difference, namely, the distance between two reports, for the reports on the same route. Besides, the original timestamp of Waze reports (e.g., "*pubDate*") was recorded by UTC (Coordinated Universal Time) time zone, so they were converted to local time to align with ground-truth data.

Step 1. Filter Waze reports by an array of possible spatial and temporal thresholds.

For each combination of spatial and temporal thresholds, we computed the difference of time and mile markers for the reports on the same route. Then, we applied both spatial and temporal thresholds to filter Waze reports. Reports are seen to be trustworthy if and only if both differences are smaller than their thresholds. For each report in ground-truth dataset, the filtering result could be: 1) no Waze reports 2) one or more Waze reports. Especially, we further distinguished the reports in second case by primary report which was made at the earliest and secondary reports which came after a primary report.

Step 2. Determine an optimal combination of spatial and temporal thresholds.

This step is to elaborately select a combination of spatial and temporal thresholds for Waze data quality assessment. This step is essential as it would influence the match rate (i.e., Waze contribution) and the tag of Waze data (e.g., false alarms and positive alarms). In experiment, the researchers hypothesized that the Waze match rates should exhibit a critical point from which the match rates would not increase significantly. The corresponding threshold values at critical point leads to the optimal combination.

3.3 Spatiotemporal clustering

Spatiotemporal clustering has been a major research field of spatiotemporal data mining and knowledge discovery. Spatiotemporal clustering generally deals with five different types of data: spatial-temporal event, geo-referenced variables, geo-reference time series, moving objects and trajectories. In this study, Waze reports belong to the spatial-temporal event. Thus, the clustering purpose is to find out clusters in which the Waze reports are in close to others both in time and in space.

Event-based spatiotemporal clustering approaches can be divided into three types: spatial scan methods, distance-based methods, and density-based methods. Spatial scan methods filter data by moving spatiotemporal cylinders of which radius is determined by spatial distance and height is determined by time interval. The cluster is made by the cylinder where inside density if higher than the outside density. Distance-based methods cluster data by spatial and temporal proximity by assuming that the times of events occurrence is random across the location and event distribution follows a certain distribution. This is not applicable for Waze case since Waze reports are not randomly distributed. The last approach, density-based methods intend to find densely clustered objects. In this project, the spatiotemporal DBSCAN (Density-based spatial clustering of applications with noise), hereinafter referred to as ST-DBSCAN, is developed and improved for exploring Waze reports [21]. There are several merits of applying this approach: 1). It can help identify false alarms which do not have minimum number of neighbor reports, 2). It does not require one to specify the number of clusters a priori which is usually hard to estimate, 3). It can find arbitrarily shaped clusters. This merit helps identify incidents located on complex network.

ST-DBSCAN was extended from the classical DBSCAN by adding a temporal dimension. **Figure 3-6** depicts the traditional DBSCAN (a) and ST-DBSCAN (b) where A, B, C, D, E and N are data points. In traditional DBSCAN clustering, threshold epsilon (ϵ) and minimum number of points (hereinafter, refer to as $minPts$) are two critical parameters. Threshold epsilon determines the neighbors of a data point and $minPts$ ensures a certain of density of a cluster. For traditional DBSCAN, the search area of each data point is a circle, with radius being threshold. However, in ST-DBSCAN, the threshold is split into spatial and temporal thresholds, which forms the square search area. That is, a point (e.g., A) is the neighbor of another point (e.g., C) only if their spatial and temporal distance meet the threshold criteria. The algorithm starts with a core point which has at least $minPts$ neighbors. For instance, if $minPts$ takes 2 and spatial and temporal thresholds are illustrated in Figure 3-6(b), then points A, B, C are labeled as core points. Then the algorithm proceeds to look for adjacent core points to form a cluster. Since A, B, C are all core points, and they are close in both spatial and temporal thresholds, they comprise a cluster. Last, the algorithm assigns the rest points by proximity rule. If the point spatially and temporally distant from all core points, then it is viewed as a noise point, as point N in Figure 3-6 (b) illustrates. For more information about the ST-DBSCAN, readers can refer to our research team's paper [21].

Like common thresholds matching approach, ST-DBSCAN needs to elaborately select parameters for thresholds. Generally, there are two ways. One is based on the domain knowledge and the other is based on clustering pattern. The former is easy to implement while it might not produce the optimal clusters. The latter considers the pattern of clustering results such as number of clusters, variation of intra-cluster, distance between inter-clusters and so on. In addition to choosing the thresholds from domain knowledge in some case studies, the researchers also proposed a novel threshold selection approach in dynamic ST-DBSCAN.

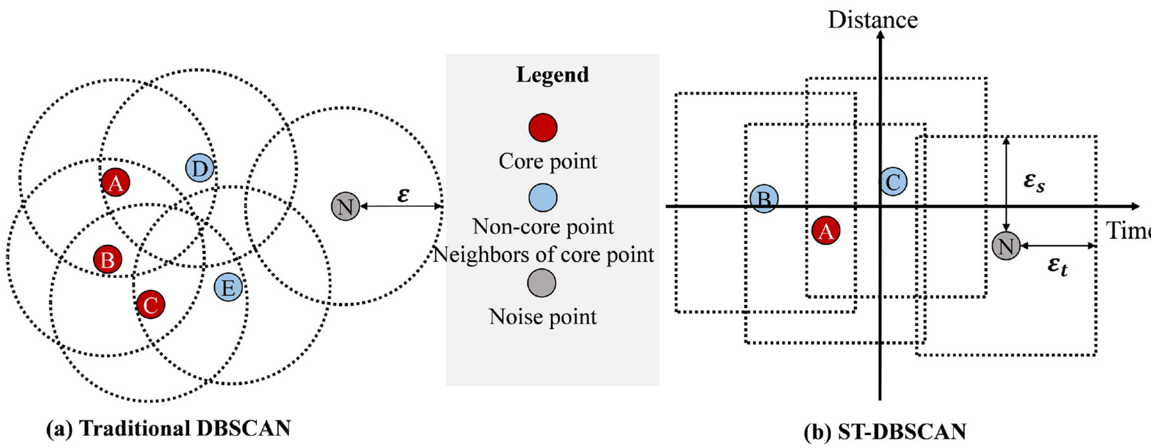


Figure 3-6 Traditional density-based spatial clustering of applications with noise (DBSCAN) versus spatiotemporal (ST)-DBSCAN

The novel approach obtains the thresholds from many samples. We introduced two concepts as illustrated in **Figure 3-7.**

Definition 1. Close Reports: for a report A, and two reports B and C, if report C has a shorter spatial and temporal distance to report A than report B, then report C is closer to report A than report B (**Figure 3-7 (a)**).

Definition 2. For report A and C, if there exists no other report that is closer to C than A, then report C is the nearest report of A.

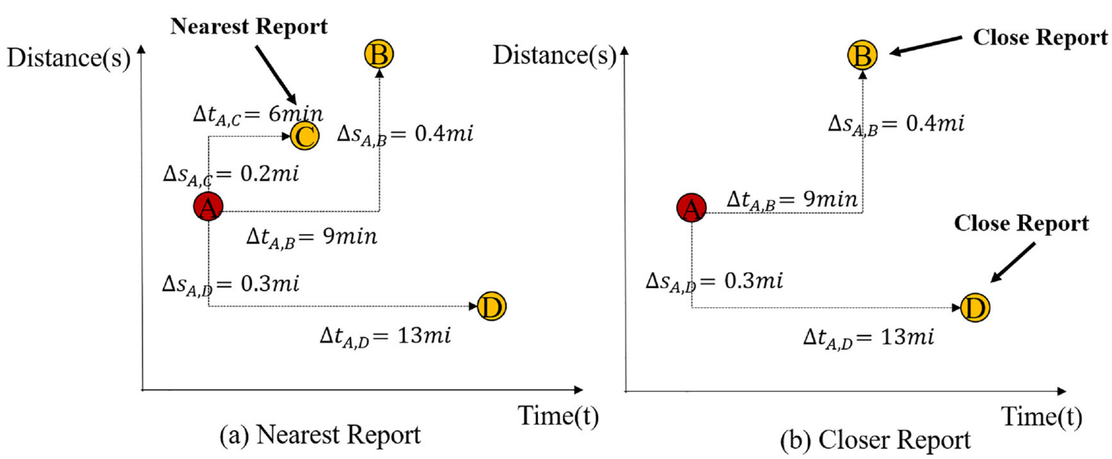


Figure 3-7 Nearest and closer report

Based on above two definitions, we enumerated all reports and identified the nearest report for each report in each cluster. Then we computed all the space and time difference between each report and their nearest report, which forms a big threshold pool. From the threshold pool, we selected the thresholds by n^{th} percentile value.

3.4 Evaluation metrics

There are multiple criteria for comparing two different pairwise points. We take different strategies for evaluating Waze data. When dealing with Waze reports data, we compare reports with ground truth discretely. Thus, the simplest method is based on their subtraction. For instance, we have a Waze report on crash at 10:00 AM and a police report on the same crash at 10:05 AM. Then, we can claim that Waze report the crash is sooner than police by 5 minutes. Such substruction can be easily applied to similar attributes such as the spatial difference between Waze reports and police reports. However, Waze speed is time series data which is a continuous variable. We take Mean Absolute Error (MAE), Mean Absolute Percent of Error (MAPE) and Root Mean Square Error (RMSE), which are the most common accuracy evaluation measure for time series speed data [22]. From mathematic perspective, MAE is an arithmetic average of the absolute difference between Waze data and ground truth. MAPE expresses the accuracy as a ratio of difference. It is an arithmetic average of the ratio of absolute difference to the ground truth. RMSE refers to square root of the mean of the squares of the difference between Waze and ground truth. The higher values of these metrics, the higher deviation from the ground truth. The minimum value of those metrics is 0, which means the perfect match of two datasets.

Chapter 4 Results and Discussion

4.1 Waze Data Quality Assessment Results

As mentioned in section **3.2** WAZE Data Quality Assessment, Waze data quality was assessed based on the spatiotemporal accuracy of the events and the completeness of these events in comparison with TDOT's Locate/IM of verified incident records.

4.1.1 Spatiotemporal Accuracy Assessment of Waze Reports

Waze log includes a collection of crowdsourced events, which are defined with event type, event location, event time, and so on, see **Figure 3-3**. Each verified event or incident is documented in TDOT's Locate/IM incident log file, with location and time information. The same event is also typically reported by multiple eyewitnesses via Waze application. These Waze reports typically have slightly different time stamp with different location coordinates. Spatiotemporal clustering techniques were employed to identify crowdsourced reports on the same event to compare against TDOT's Locate/IM records. Two primary types of events were used for comparisons, i.e., crashes and stopped vehicles, to have coverage over severe non-recurring incidents as well as non-critical disabled/abandoned vehicle situations. In terms of temporal accuracy, typically the first Waze report marks the earliest and typically the closest time to the occurrence of the event. On the other hand, the spatial accuracy of Waze reports depends on where the Wazer initiates the reporting process, which could take place upstream, when they saw the event, or downstream, after they have passed the event location.

This study specifically compared two types of incidents for spatiotemporal accuracy purposes. The first type are accidents, which include crashes, overturned vehicles, and vehicles on fire. The second type are stopped vehicles, which include disabled or abandoned vehicles on the shoulder or the median. Five months-worth of Waze data from 2017 were used for comparison against TDOT Locate/IM data from the same period for the entire I-40 in Tennessee. A total of over 8,000 Waze accidents were reported compared against about 2,000 Locate/IM events. Clustering analyses were performed to identify the unique events reported by both Waze and Locate/IM for comparison purposes.

The difference between the time and location of the reports are plotted in **Figure 4-1**, where the horizontal axis represents the difference in time (in minutes) of the Locate/IM report and the first Waze report of the same accident. A negative value means Waze report came in before Locate/IM report while a positive value means Locate/IM came in before Waze report. The vertical axis represents the difference in location (in miles) of the Locate/IM report and the Waze reports of the same accident. A negative value means Waze report is upstream of the Locate/IM reported location and vice versa. Both the time and distance differences exhibit bell-shaped distribution. On average, the first Waze report comes in about 2.2 minutes sooner than the Locate/IM event time. On average, Waze reported accident locations are only 6 feet upstream of the Locate/IM reported location. This is within the margin of error of GPS systems.

Similarly, reports of stopped vehicles for the same duration on I-40 were analyzed and plotted in **Figure 4-2** in identical fashion. On average, the first Waze report comes in about 7.8 minutes sooner than the Locate/IM event time. Also, Waze reported accident locations are on average 132 feet upstream of the Locate/IM reported location. These are greater differences in both event time and location compared to those for accidents. A likely explanation is that Wazers may feel more compelled to report more serious incidents, such as accidents, than less serious events, such as vehicles on the shoulder.

In a nutshell, this study finds Wazers appear to report incident more promptly than those documented in Locate/IM and with acceptable accuracy in terms of incident location.

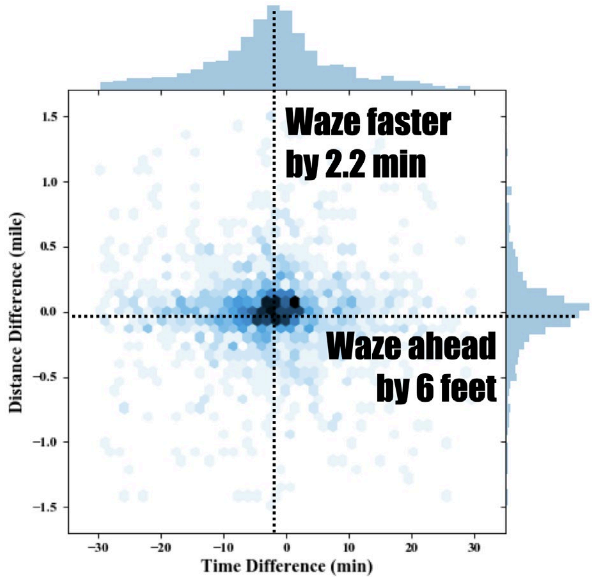


Figure 4-1 Differences in accident report time and distance between Waze and Locate/IM

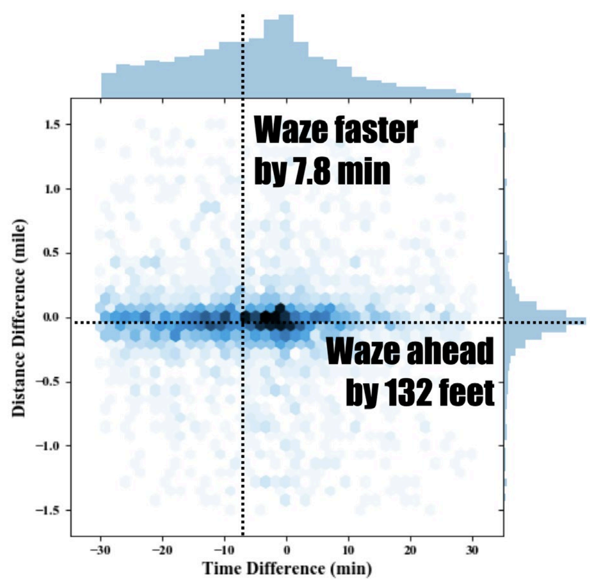


Figure 4-2 Differences in stopped vehicle report time and distance between Waze and Locate/IM

4.1.2 Event Completeness Assessment

The Waze-type of crowdsourced data are at the mercy of the crowd. A severe event in heavily trafficked area is likely to be reported multiple times by different individuals while a less serious event on less travelled roads may receive few or even no report. In general, Waze reports can cover a large geographical area wherever Wazers roam, see **Figure 4-3**. On the other hand, TDOT Locate/IM incident log files tend to contain verified incidents on roadways under TDOT's jurisdiction. Again, matching analyses were performed to identify Waze events, which are compared against Locate/IM incidents for coverage completeness assessment.

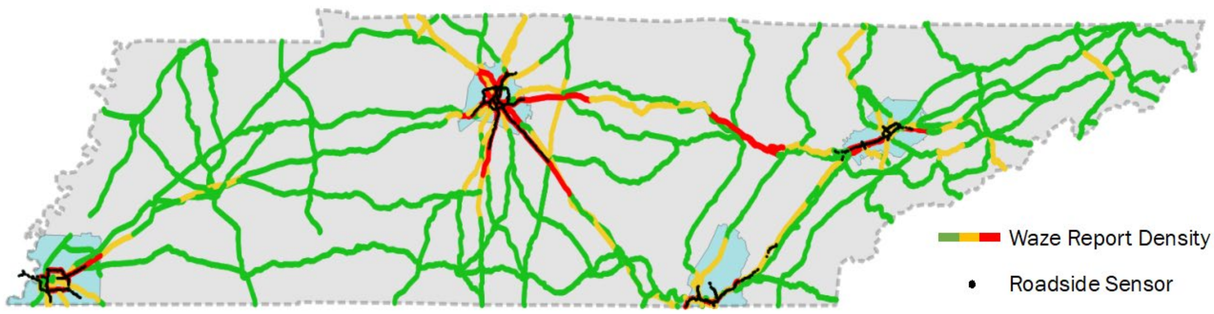


Figure 4-3 Sample Waze event report density vs TDOT RDS station locations

The events represented in **Figure 4-1** and **Figure 4-2** are those matched in both Waze and Locate/IM reports. Not all Locate/IM events are reported by Wazers and vice versa, see **TABLE 4-1**. For the five months' worth of data used in this part of the study, Wazers reported 8,068 crashes and 93,707 stopped vehicles while the Locate/IM recorded only 2,052 crashes and 5,459 stopped vehicles. Employing spatiotemporal clustering techniques to match these events, this study found that about two-third of the crashes reported by Locate/IM are reported in Waze crash reports. The study also found that about 85% of the stopped vehicles reported by Locate/IM are also reported in Waze reports, see **TABLE 4-1**.

TABLE 4-1 TOTAL INCIDENT RECORDS AND MATCHED RECORDS BETWEEN WAZE AND LOCATE/IM

	Waze Reports		Locate/IM Reports	
	Vehicular Crash	Stopped Vehicle	Vehicular Crash	Stopped Vehicle
Total number of Reports	8,068	93,707	2,052	5,459
Number of Matched Reports	2,066	13,203	1,374	4,674
Percent Matched	26%	14%	67%	85%

4.1.3 Crowdsourced Speed on Surface Streets Assessment

When Wazers keep the app open for navigation, they can implicitly tell Waze app the current speed on a specific stretch of the road. The road speed is measured by such real-time updates from Wazers along with historical information [23]. Hence, the speed estimation is partly influenced by number of Wazers on the road. The term 'surface street' refers to an interrupted signalized roadway with intersections where usually less probe vehicles are observed than highways. Hence, examining the quality of Waze speed on surface streets is necessary yet challenging. To this end, the study conducted a case study in Sevierville, Tennessee where there are many surface and signalized streets, and many Bluetooth stations are installed. Bluetooth speed data is employed as ground truth, which is independent of Waze data. Under the Waze for Cities Program, the researchers added 36 exact segments paired with Bluetooth Stations through TDOT user portal as **Figure 4-4** shows. One-month data was collected for both Waze and Bluetooth speed data in January 2019, and they were aggregated at 5-minutes level for comparison purposes. Since the number of active Wazers varies at different time of the day and traffic status, each day is divided into four categories: AM Peak (7 AM to 10 AM), Mid-day (10 AM to 4 PM), PM Peak (4 PM to 7 PM), and Rest of Day (7 PM to 7 AM). Furthermore, the speed is divided into five groups: (0-<20mph), (20-<30mph), (30-<40mph), (40-<50mph) and (≥ 50mph) to capture the impact of different traffic status. Further, MAE, MAPE, and RMSE, as mentioned in section 3.4, are calculated for Waze speed validation.

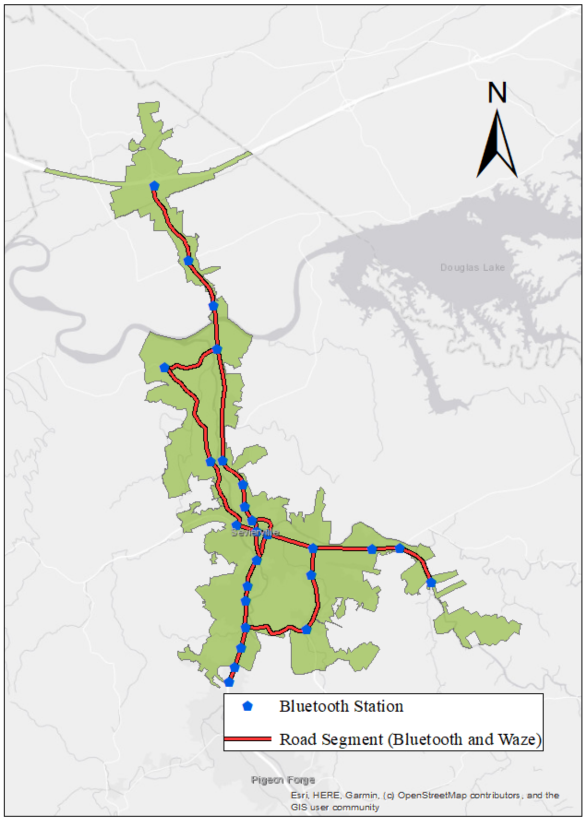


Figure 4-4 Case study location in Sevierville, TN

Figure 4-5 compared Waze and Bluetooth speed data over a month for four randomly selected segments using 5-min aggregated speed data. The probability scatter plots of data from two data source are plotted for different time of day, including AM-Peak, Mid-day, PM peak and Rest of day. The darker hexagons represent the higher density of data points and the lighter color hexagons denote the lower density of data points. The diagonal dashed line (y=x) represents the perfect match between two datasets. Thus, the more probability on (or close) to this line, the higher the accuracy of Waze speed data. As shown in the figures, for all segments in AM Peak, Mid-day, and PM Peak when more probe vehicles are present, the dark color surrounds the perfect match line. In contrast, during rest of day (e.g., night hours and early morning), the higher density hexagons deviate from the perfect match line evidently. Furthermore, the Waze speed of Parkway, Veterans Blvd, and Main St. exhibit a linear behavior. This issue is caused by the small number of probe vehicles observed so it uses historical data that also have a small range of values (or sometimes constant value). However, the Waze speed data from Route 66 is less linear during night hours. Route 66 ends at Interstate 40, which has more probe vehicles during night hours than other segments.

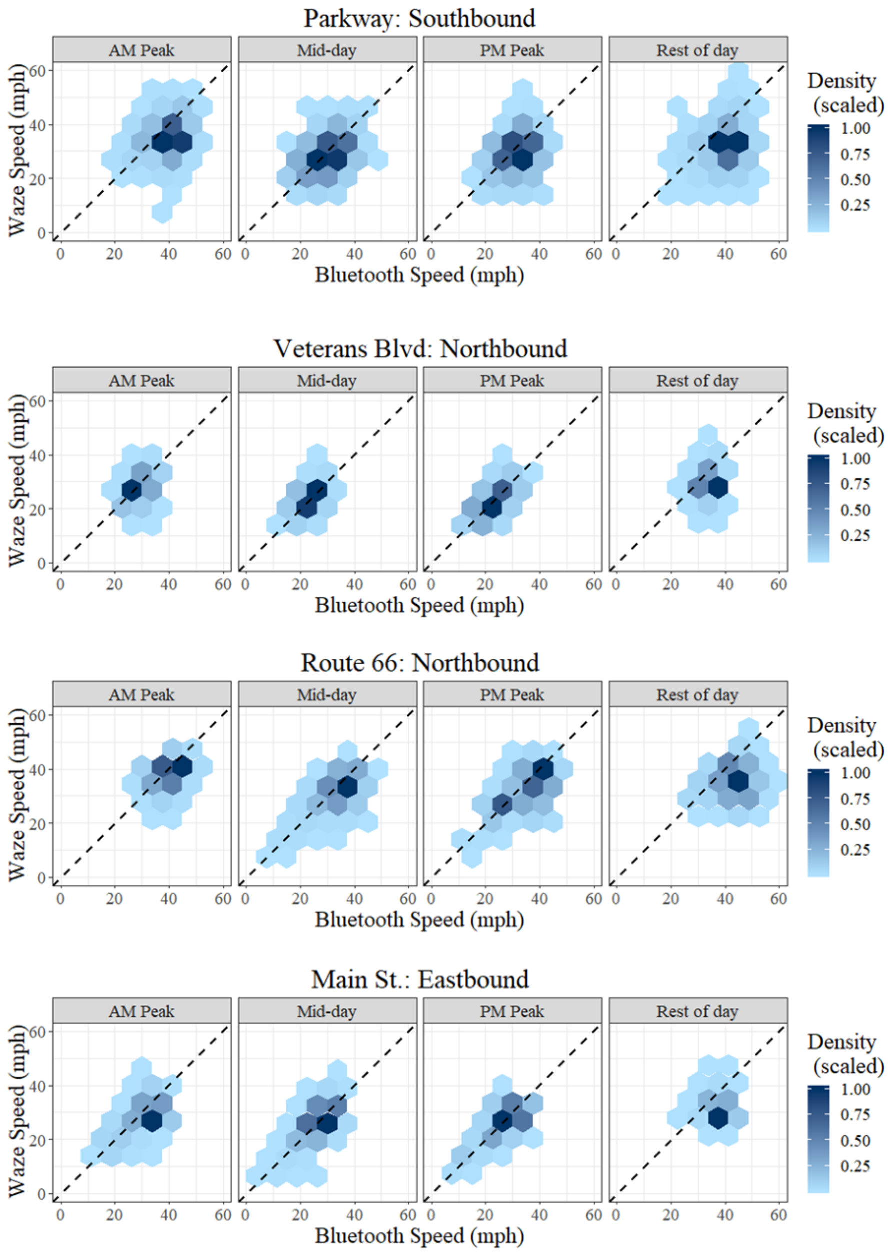


Figure 4-5 Probability Density plot: Waze speed vs. Bluetooth speed

In addition, **TABLE 4-2TABLE 4-2** shows the results for different speed levels. In the daytime (AM Peak, Mid-day and PM Peak), the difference between Waze and ground-truth data sightly decreases as speed increases. In terms of percentage of difference, MAPE shows the same result. However, during nighttime (Rest of Day), the value of error is higher than daytime. Classifying speed shows that in higher speed categories Waze has a slightly higher rate of accuracy. Also, in daytime and peak hours, Waze speed data seem to be more accurate than nighttime and nonpeak hours.

This report only recaps the most important findings from research of Waze speed accuracy on surface streets. For more research findings, please refer to the research team's paper entitled

"Quality of location-based crowdsourced speed data on surface streets: A case study of Waze and Bluetooth speed data in Sevierville, TN" [24].

TABLE 4-2 ACCURACY ANALYSIS OF WAZE SPEED DATA FOR DIFFERENT SPEED LEVEL

Speed category (mph)	Daytime (AM peak, Mid-day and PM Peak)			Nighttime (Rest of Day)		
	RMES (mph)	MAE (mph)	MAPE (%)	RMES (mph)	MAE (mph)	MAPE (%)
0-<20	3.3	3	13.2	4.1	3.4	21.8
20-<30	4.5	3.6	14.1	4.3	3.7	14.4
30-<40	3.8	3.1	8.8	5.7	5.3	14.7
40-<50	4.7	4.2	8.6	8.0	7.5	17.1
≥ 50	3.9	3.7	7.1	11.5	11.1	21.2

4.1.4 Crowdsourced Speed on freeways assessment

Vehicles traveling on interstate highways are participating in uninterrupted flow. Travel speed pattern are significantly different from surface streets. The research team is among the first to examine the Waze speed accuracy on freeways. To thoroughly evaluate and explore Waze speed, the study conducted by research team used a 10.8 mile stretch of interstate 40, from mile marker 374.3 to 385.1, in Knoxville, Tennessee (see **Figure 4-6**) to compare the speed measurements from Waze and Remote Traffic Microwave Sensors (RTMS) over a two-month period. The RTMS stations collect point-speed every 30 seconds while Waze updates speed every 1 minutes. To keep the consistency of two speed source and preserve the speed resolution, the RTMS data was aggregated to one minute by taking the arithmetic mean of vehicles' speed passing a sensor station in one minute. The validation approach is mainly focused on quantifying the speed difference between RTMS speed and Waze speed at different periods and speed levels.

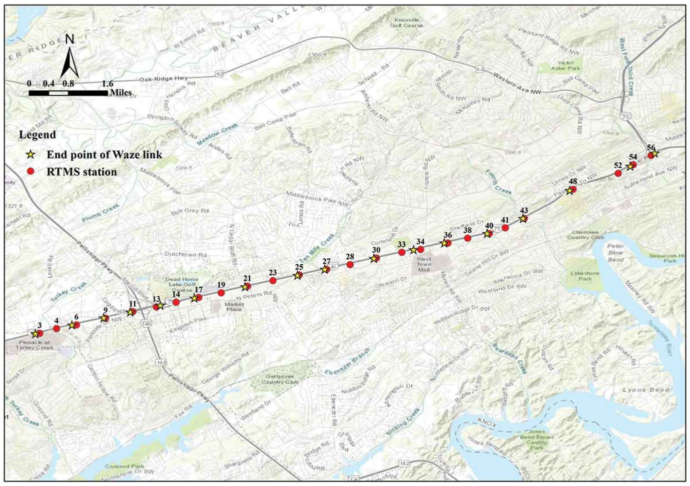


Figure 4-6 Map of the selected RTMS stations and corresponding Waze links along I-40 Eastbound in Knoxville, Tennessee

The research team extracted all the speed data for sixteen designated Waze links and corresponding RTMS stations in July and August 2019. **Figure 4-7** shows the scatter plot for total 1,506,414 observations. The diagonal dash line indicates the perfect match of two speed source. The Pearson correlation test shows a relatively high correlation (r=0.65). The histogram of horizontal axis reveals that RTMS speed has a bell shape distribution and most of the speed data are concentrated at 60 mph. In contrast, the distribution of Waze speed is more zigzag and most of speed data focus near 70 mph, which is significantly higher than RTMS speed. It is worth noting that a considerable number of repeated values for Waze were found, suggesting that Waze may not be able to report real-time speed every-minute. Waze speeds tend to be higher than RTMS speeds for high speeds, while RTMS speeds are similar to or even higher than Waze speeds for low speeds. That may be caused by sample bias in calculating the speed for Waze.

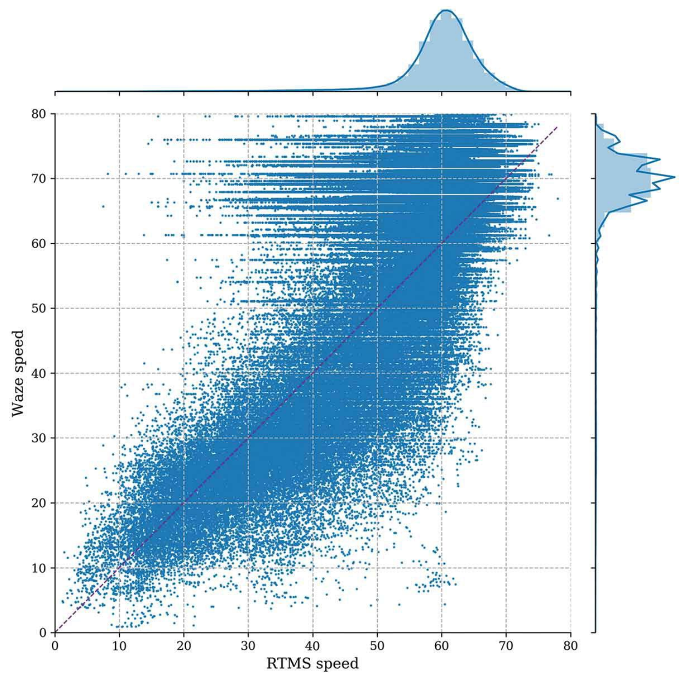


Figure 4-7 Scatter plot for all speed observations from both Waze and RTMS

Figure 4-8 presents the absolute speed difference at different speed levels: (0-45 mph, 45-55 mph, and greater than 55 mph). Particularly, 55 mph is the posted speed limit for road segments in study area. From the figure, we can observe that as the Waze speed increases, the interquartile ranges become smaller, indicating less variation in speed difference for high Waze speeds. The median speed differences for three groups, 0-45 mph, 45-55 mph, and greater than 55 mph, are 6.5 (17.0%), 3.4 mph (6.3%), and -8.9 (-14.6%), respectively. Also, this figure reaffirms that Waze speed is greater than RTMS speed for high speeds while it is likely to be less than RTMS speed for low speeds.

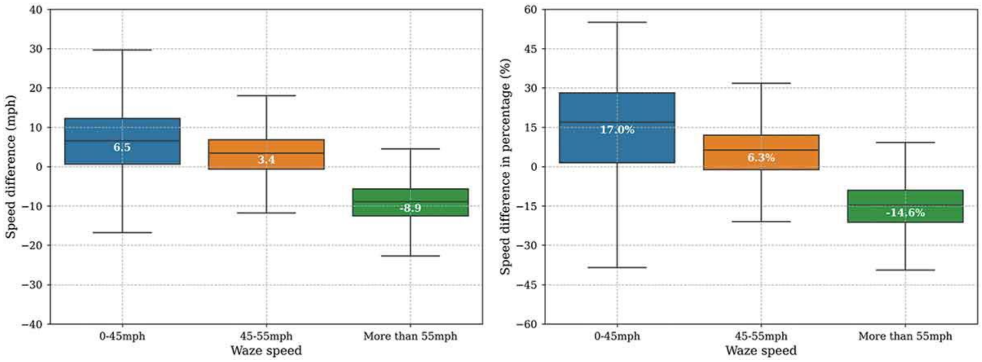


Figure 4-8 Boxplots of the absolute speed difference (left) and absolute relative speed difference (right)

4.2 Use Scenarios

Because of the crowdsourcing nature of Waze data and its uniqueness in complementing existing infrastructure-based traffic data, such as TDOT's RTMS, there are scenarios where transportation agencies could benefit from incorporating Waze data into their resources. To that end, this study investigated an array of use cases regarding Waze reports and speed and explored a selected number of them to demonstrate the potential benefits and challenges.

4.2.1 Queue protection

Traffic queues, especially queues caused by non-recurrent events such as incidents, are unexpected to high-speed driver's approaching the end of queue (EOQ). Detecting EOQ is of importance to prevent rear-end or secondary crashes. Most existing EOQ detection methods highly rely on the congestion pattern (e.g., speed) extracted from fixed traffic sensor data. Thus, these detection methods are restricted by the spatial and temporal resolutions of traffic sensors. However, Crowdsourced Waze can contribute to detecting the EOQ dynamically and widely. To test the feasibility of this promising use scenario, we took Waze jam and accident reports collected between August 2017 and December 2017 on Knoxville highways. ST-DBSCAN was employed to detect EOQ dynamically.

Figure 4-9 is a time-space diagram showing a collection of actual Waze reports in the aftermath of a crash on I-40 in Knoxville, TN area. All circle points belong to the same cluster and each circle represents either a jam or an accident report at a specific location and a specific time. The circles of various shades of red are events of congestion/queue from stand-still traffic to heavy traffic, to moderate traffic. The white half-circles are reports of accident, or crash. The first accident report (half-circle) came in just after 15:10 somewhere downstream of mile marker 377.5 while there had been multiple congestion reports 10 minutes before that in that area. The queue propagated upstream over time as the arrow shows. The green points represent the EOQ at different time. Furthermore, we compared the Waze-based EOQ detection with fixed sensor-based EOQ detection (see blue dash line in **Figure 4-9**) for 385 detection cases. The results found that sensor based EOQ reports the EOQ is only 1.1 minutes earlier than the proposed method based on Waze data. In addition, Waze generates 1.9 EOQ detection points every mile which is slightly higher than traffic sensor-based detection. This small detection time difference suggests that the proposed method based on crowdsourced Waze data can be a perfect surrogate approach to detect EOQ in real time.

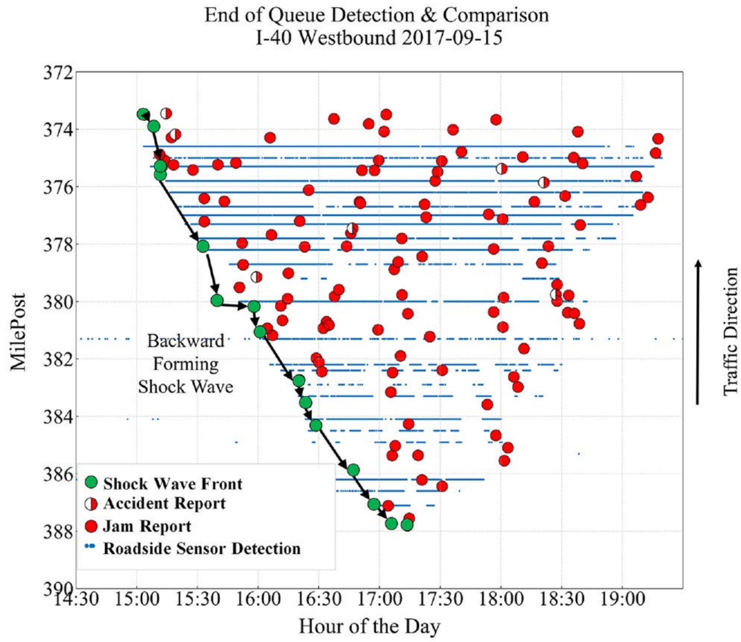


Figure 4-9 Example of backward-forming shock wave detection

4.2.2 Secondary Crash identification

Secondary crash refers to the crash resultant from a primary crash. FHWA estimated that approximately 20 percent of all incidents were secondary crashes in 2013 (FHWA, 2004 #17). A better understanding of secondary crashes would benefit traffic incident management, and this requires accurate detection of secondary crashes. Most conventional approaches identify the secondary crash through speed pattern at incident locale, which relies on the speed data from different detectors. Waze is rarely applied for such scenario previously. The researchers took a stretch of Interstate 40 in Knoxville as an example and examined the feasibility of this application in real word. The method was mainly based on ST-DBSCAN as mentioned in section 3.3 Spatiotemporal clustering. Notably, the spatial and temporal thresholds were selected according to the domain knowledge which are 1 mile and 30 minutes.

Figure 4-10 shows an example of identification result. All the points are from the same cluster in the aftermath of the primary crash. Five Wazers reported accident in a short spell of time, and many Wazers reported the jam upstream of the crash reports. According to the TDOT crash report, the primary crash occurred at 14:15 at milepost 391 of Interstate 40 and the secondary crash took place at 14:32 at milepost 389 of Interstate 40 eastbound. However, the Wazers reported the primary crash and secondary crash at 14:14 and 14:29 which were 1 minute and 3 minutes sooner than office report, respectively. In addition, among 708 crashes, the proposed approach correctly identified 17 out of 39 secondary crashes which were observed from speed contour plot. This finding verified the potential of Waze in detecting secondary crash. For more information about this study, readers can refer to the research team's paper [25].

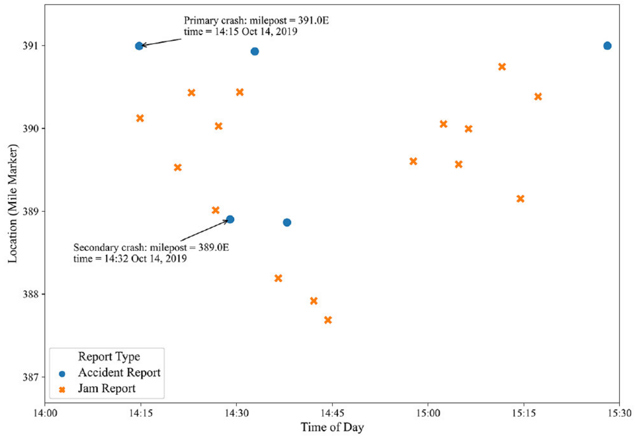


Figure 4-10 An example of a primary–secondary crash relationship captured by Waze Reports

4.2.3 Performance measurements

Level of Service (LOS) is an important indicator relating the quality of roadway traffic service. For interstate highways, the Highway Capacity Manual (HCM) regulates analytical methods to assess LOS based on traffic density, speed, and highway characteristics (Manual, 2010 #21). In real word, it is costly and unrealistic to obtain reliable density data on every road and in large networks using fixed location sensors and cameras. However, Waze, as a type of Crowdsourced data, could provide a low-cost and alternative solution to assess LOS for both freeways and intersections. To examine the feasibility of this use scenario, the researchers took a stretch of Interstate 40 (i.e., 1.5 miles around mile marker 385 westbound) in Knoxville as an example. One month (i.e., October 1, 2019, to October 31, 2019) of Waze reports and speed data, along with fixed sensor speed data were prepared. The Waze reports were aggregated by hours and speed data were employed to calculate travel time reliabilities such as travel time index, average speed and so on. These performance measurements were further mapped to the LOS ground truth which was primarily estimated from densities.

Figure 4-11 presents the boxplot of the average speed and number of alerts with respect to six LOS services. The average speed decreases as LOS service level increases. The number of Waze alerts distinctly increases with respect to low LOS service. **Figure 4-12** ranks the importance of each measurement in estimating LOS services. Notably, Waze speed is the leading factor of LOS, followed by number of Waze alerts. Both Waze indicators have more effect on LOS than other travel time reliability indicators. This evidence makes this application applicable that Waze speed and alerts could potentially become surrogate measurements for estimating freeway LOS.

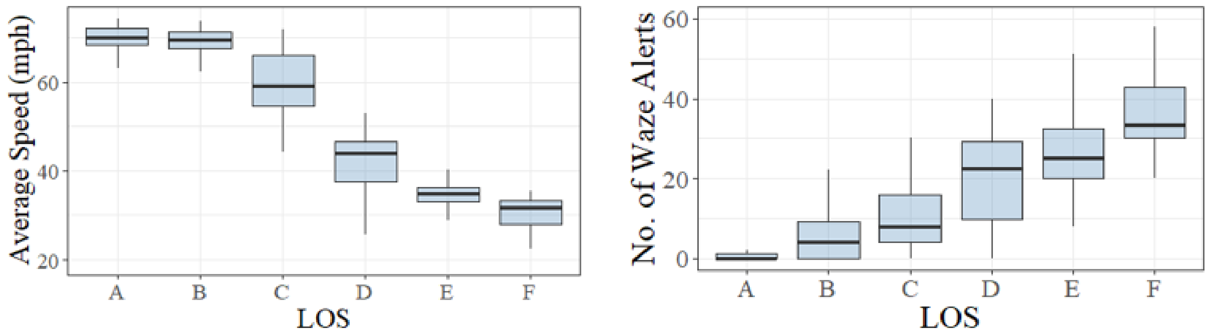


Figure 4-11 Boxplot of different measures in each LOS category

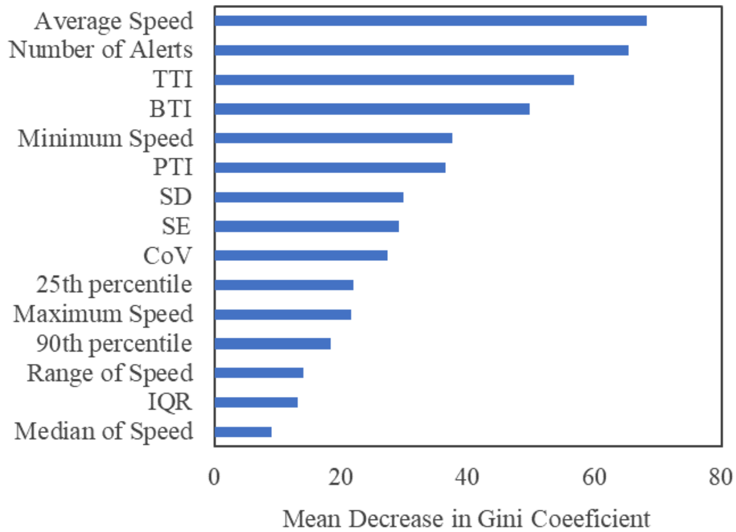


Figure 4-12 Variable importance in estimating LOS

LOS can also be applied to surface streets, to describe major signalized intersections. Different from the freeway, LOS on surface streets is usually determined by average delay during a short period (Manual, 2010 #21). The speed is the primary source of estimating delay whereas there are no speed detectors in many intersections or arterials. Hence, estimating delay is difficult in such conditions. However, Waze data could shine lights on this surface street LOS evaluation. There are several reasons:

- ***Wide coverage and huge flexibility.*** With Waze for cities program, transportation agencies can designate routes for any intersections within specified region. This makes Waze data superior to fixed sensors as transportation agencies can monitor LOS at intersections where no traffic detectors are installed.
- ***Frequent updates.*** Waze speed and report data can be collected every 2 minutes across the day. Thus, the delay and LOS level can be updated every 2 minutes. Many Wazers are active across the day (see **Figure 4-13**), which can provide sustainable data source for LOS estimation. What's more, the operators can frequently evaluate the signal-related project by high resolution LOS at special occasions such as signal retiming, signal coordination, and so on.
- ***Save labor save cost.*** With Waze automatic feed, we would have hundreds or even more witnesses at intersections. Thus, our operators can get the traffic status in house. This advantage saves a lot of labor and cost for collecting data at field.

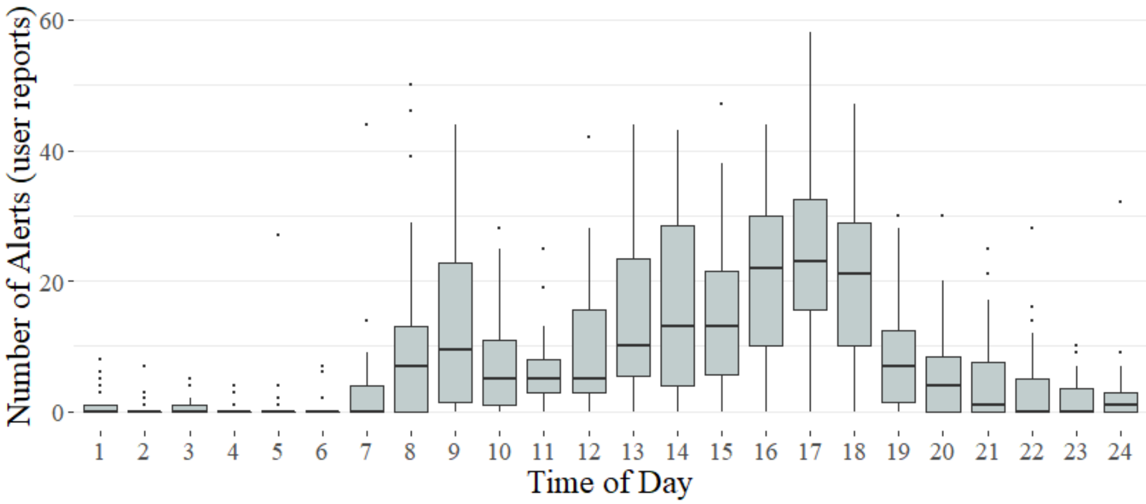


Figure 4-13 Boxplot of number of alerts at different time of the day

4.2.4 Incident management

As mentioned in section 2.3, many states have started to incorporate Waze reports with traditional data sources for incident management. In practice, Waze data can help TDOT in the following three fronts.

Supplement event database.

In addition to crash and disabled vehicle, the Locate/IM backs up many incidents such debris, congestion and so on. Those incidents are mainly reported by highway helpers and police. Meanwhile, Wazers reports also cover hazard on road, weather, road closure and so on. **Error! Reference source not found.** lists the five major event types and their corresponding terminology in two datasets. **Figure 4-14 Venn diagram of the incident database.** depicts the relationship between all incidents, Waze and Locate/IM database. The red shaded area (B) denotes both reports from Waze and Locate/IM. The area A refers to the incidents only reported on Locate/IM. The area C refers to the incidents only reported by Waze. Outside of two data sources, there are some incidents neither detected by Waze or TDOT operators. Therefore, the Waze could complement the event database as listed in **TABLE 4-3**. The enhanced event database would help TDOT make do further operation and planning projects.

TABLE 4-3 EVENTS MAP BETWEEN WAZE AND LOCATE/IM

Event type	Waze	Locate/IM
Crash	Major, minor accident	Single/multi vehicle crash
Congestion	Moderate, heavy, standstill, light traffic	Congestion

Road Hazard	Hazard on road, shoulder, roadkill	Debris on the Roadway
Road Closed	Road closed hazard/construction/event	Road work scheduled
Weather Hazard	Weather hazard heat	Weather
	Hazard on shoulder car stopped,	Disabled vehicle, abandoned

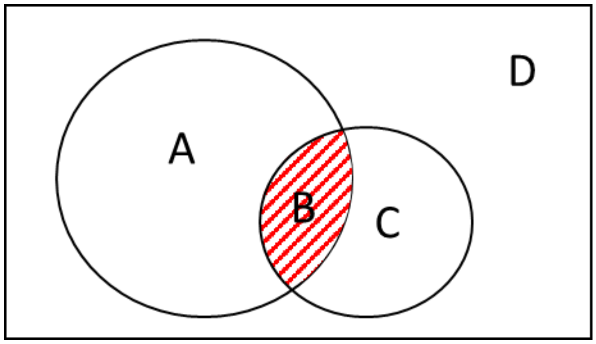


Figure 4-14 Venn diagram of the incident database

Connect with HELP truck

TDOT HELP trucks (see **Figure 4-15**) play an important role in giving fast emergency help to motorists, clearing incidents quickly, removing road hazards, easing traffic congestions, and so on. HELP trucks operate on assigned city core highways every day. In addition to finding incident by regular patrol, they also respond to motorists from phone call *(*THP (*847)*) [26]. Waze can facilitate this request and respond process. As demonstrated in section 4.1.1, the primary Waze crash and stopped vehicle reports have relatively good spatial and temporal accuracy on average. Once Waze is linked with HELP truck, this will build a two-way informative channel. For one thing, Waze can timely inform HELP operators the location and time of incidents, which help them make decisions such as rerouting and optimize help resources in peak hours. For another thing, HELP can inform Wazers when HELP trucks are responding to emergency incidents or stopping at the scene. This would alert the Wazers to slow down to create a safe and fast channel for HELP operators.

Figure 4-15 TDOT help truck and the operator, from TDOT webpage[2]

Freeway Service Safety Patrol Expansion

In the long term, Waze data can also serve for Freeway Service Safety Patrol Expansion. Note that Waze reports reflect the driver's complaints about the different roadway incidents. Hence, the higher density reports area, the more attention that HELP truck should pay attention to when they consider expanding road service. **Figure 4-16** illustrates the expansion scenarios of FSSP in Nashville city area. Using one month data (October 2021) as an example, the researchers calculated the number of major incidents including accidents, jam, weather, and road closed reports for each isochrone levels. **Figure 4-17** shows the cumulative number of Waze reports wrapped by each isochrone map. The cumulative plot suggests different critical points after which the number of Waze reports do not increase significantly. For instance, the number of accident reports tend to stabilize after expanding 5.5 miles from original HELP patrol area, meaning that most crashes concentrated on this buffer area. Therefore, decision makers can take this point into consideration when considering the expansion of patrol area.

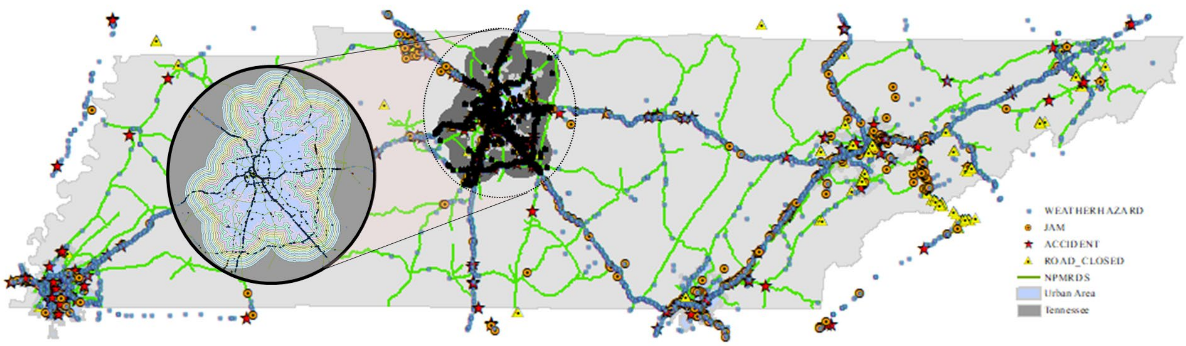


Figure 4-16 An isochrone map showing the coverage of Waze reports for Nashville, TN

[2]TDOT HELP program, https://www.tn.gov/tdot/traffic-operations-division/how-does-the-help-program-work-.html

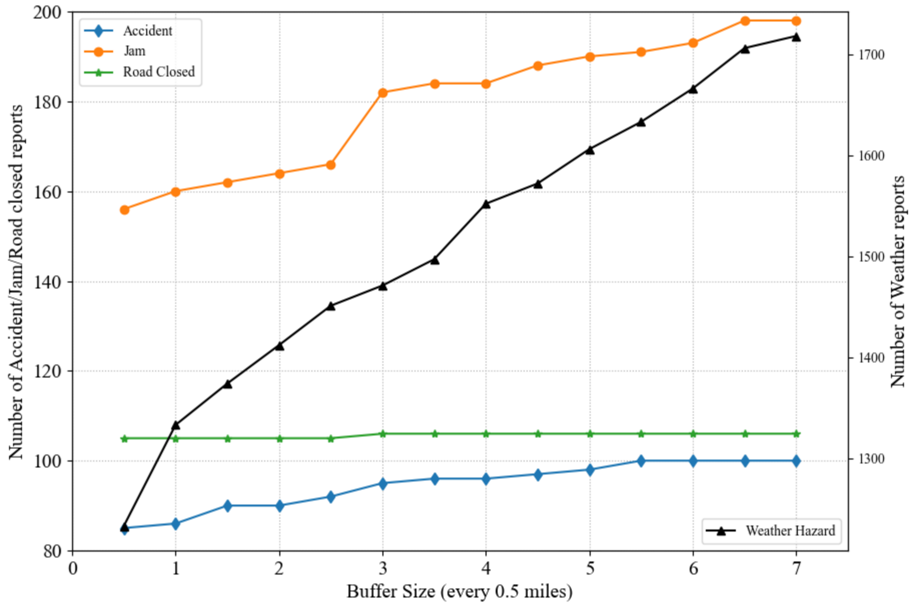


Figure 4-17 Number of different Waze reports covered by different isochrone service areas

4.2.5 Roadway maintenance

Roadway maintenance would benefit from the Waze special reports including pothole and different weather event reports. Specifically, Waze could be applied for:

Pothole detection

At the beginning of 2022, TDOT launched a plan for pothole repair after several winter storms hit Tennessee and an abundance of potholes emerge on highways and state routes. The statewide expenditures for pothole patching in the first 19 days of 2022 is $3.32M which accounts for over one-third of the total budget (i.e., $9.16M) of 2022. Besides, TDOT had to spend around $8M every year repairing potholes in the past three years [1]. However, in current practice, the majority of potholes are inspected by annual laser scanning [27] while the minority of them are reported by motorists through the TDOT maintenance request[3]. The former way can generate a reliable diagnosis of pavement condition while it involves massive labor and cost. The latter way is at the mercy of the user's perception of potholes. Compared to laser scanning, motorists cannot have a good sense of the location of potholes especially when they pass the potholes rapidly. However, the crowdsourced reports from motorists might have a wider range and more frequent scanning of the road than TDOT annual detection.

The research team took Interstate 40 in Tennessee as an example and investigated the feasibility of applying Waze pothole reports to fix potholes. The entire year of 2021 reports were analyzed, and the detection results were compared the ground truth from the work request in TDOT pavement maintenance system. **Figure 4-18** shows the results of pothole detection by applying ST-DBSCAN on Waze pothole reports. The upper left figure demonstrates three types of points: clustered reports which are close to each other in terms of space and time, detected potholes

[3] TDOT maintenance request form: https://www.tn.gov/tdot/maintenance/maintenance-request.html

that were aggregated from reports in the same cluster, and noise reports which are false alarms. The right bar chart shows the number of potholes detected for each month. The number of potholes surged from January to March and mitigate after March. This is because some areas were patched by TDOT crews timely during the outbreak of potholes. Further, the lower left scatter plot shows the comparison between detected potholes (marked as 'x') and TDOT pothole patch request (red dashed line, work request labeled the start and end log mile of a work activity.). From the figure, it is worth noting that TDOT pothole requests mostly concentrated the first four months. For rest of the year, there were many potholes reported by Wazers while no work request was found at vicinity area. Besides, the potholes presented on and off (sort of periodically) at some locations such as log mile 1 (i.e., mile marker 193) and 23 (i.e., mile marker 215), indicating potholes tended to occur frequently in these areas even the repair work had been done. Pavement resurfacing work was suggested instead of spot patching.

Compared with TDOT annual scanning and citizen's form request, the Waze reports directly pinpoint the time and location of potholes, which could help TDOT crews more quickly and efficiently identify the broken roads.

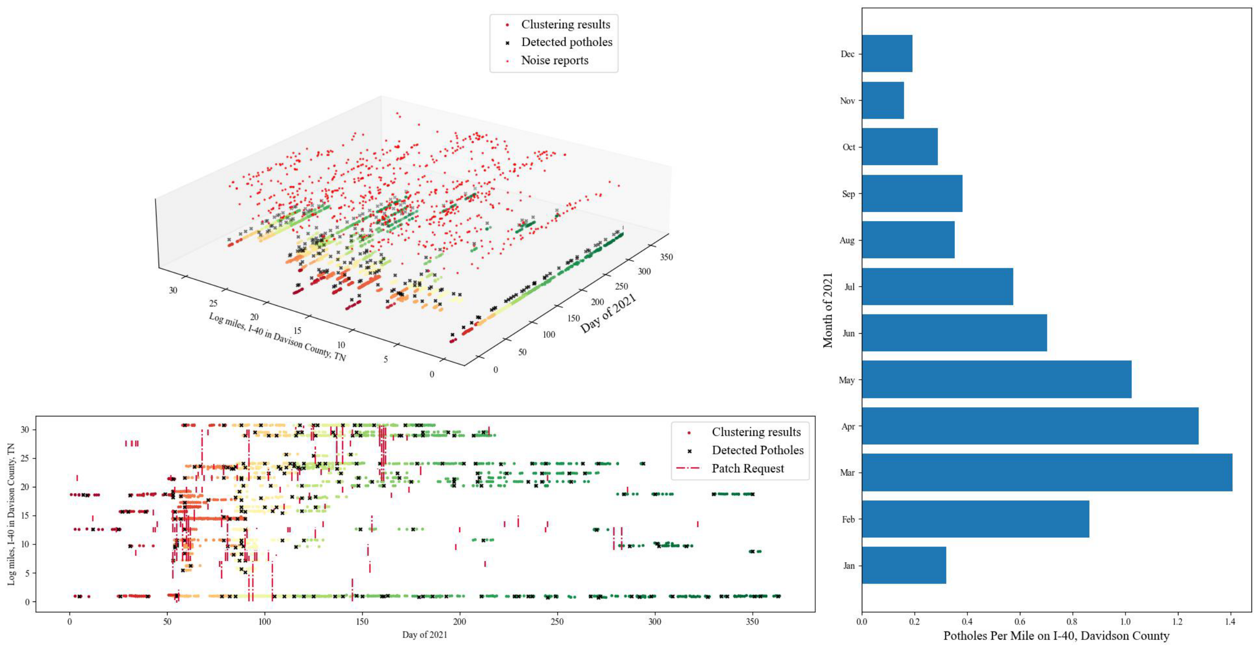


Figure 4-18 Pothole detection results for I-40, Davidson County, TN

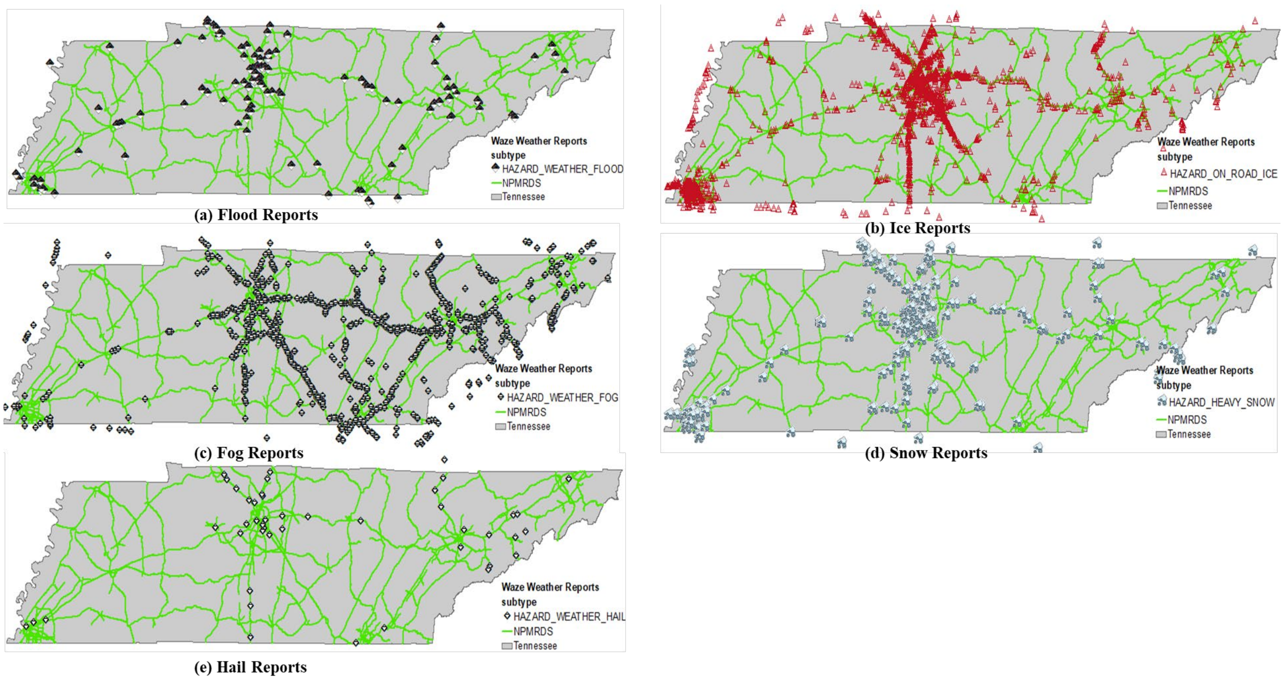


Figure 4-19 Extreme Weather Reports Collected in February 2021

Snow/Ice removal

State of Tennessee suffers from different natural disasters and severe weather every year. These weather-related incidents also pose great threats to motorists lives and properties. **Figure 4-19** shows the five types of weather reports: flood, ice on the road, fog, snow on the road, and hail collected in February 2021. Ice, fog, and snow have the relatively higher density than flood and hail reports. City of Nashville is likely to have all those weather events during the winter. Depending on the impact of weather events on road traffic, TDOT can set up different countermeasures. For instance, TDOT can effectively deploy crews to plow snow and ice based on the information from Wazers' snow and ice reports. In addition, fog reports can be injected into low visibility warning systems and dynamic message signs. For flood reports, they can be used for evaluating flood risk and making evacuation scenarios.

Pavement Condition Evaluation

TDOT evaluates the pavement condition every year. The data used for evaluation is collected by laser scanning once a year. The researchers aggregated the Waze pothole reports and compared them with pavement condition index from TDOT report. **Figure 4-20** is a scatter plot with x-axis being number of pothole reports normalized by AADT and y-axis being pavement distress index (PDI) estimated from field data. The higher PDI, the better the pavement condition, vice versa. A simple linear regression line was added to the plot and the correlation suggested that the number of potholes is negatively associated with the PDI. P-value is extremely small, which means this correlation is statistically significant. Therefore, the number of potholes could potentially be an important indicator for pavement condition assessment. Compared with yearly pavement data for estimating PDI, the Waze data could provide more frequent reporting of pavement conditions.

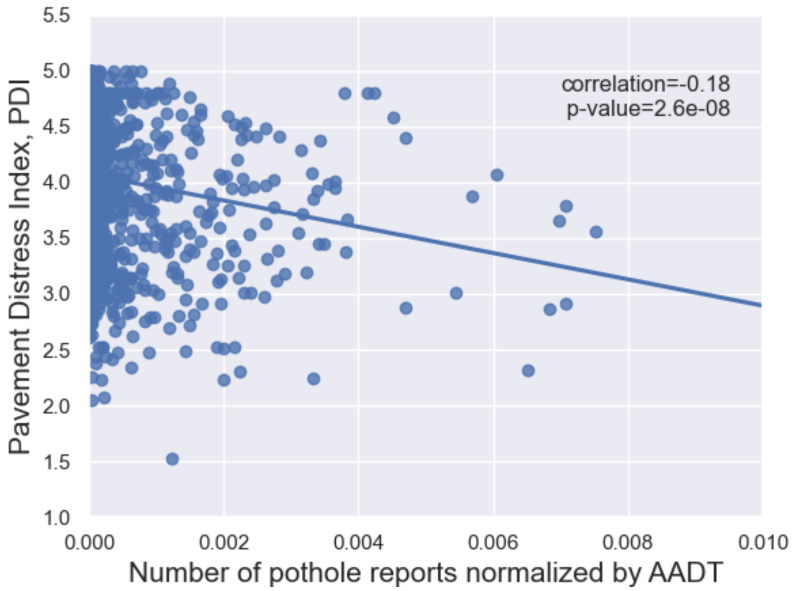


Figure 4-20 Scatter plot of pothole reports per AADT with respect to PDI.

4.2.6 Work zone management

Highway Work zones are hazardous both for motorists who drive through the array of signs, barrels, and lane changes, and for workers who build, repair and maintain the bridges and highways. Hence, improving zone traffic management is essential to motorists and worker safety. TDOT can incorporate with Waze to enhance the work zone management by the following measures. For one thing, Waze partner hub allows partner to share private data with Waze[4]. Thus, TDOT can push work zone construction information (e.g., location and time of activities, work type, duration and so on) to Waze app. This work log information would complement the Waze roadside work reports so that Wazers could be alerted to slow down when they approach the Work zone areas. For another thing, Waze traffic view enables operators to add any routes within specified region to monitor the route speed. Speed monitoring based on Waze has many prominent merits compared with other speed source. For instance, Waze provides segment speed instead of point speed obtained from fixed sensor. To the researcher's knowledge, there are two important speed data sources in addition to Waze. NPMRDS releases speed data aggregated at 5 minutes once a month, and RDS sensor data releases speed data at the end of day. However, Waze updates speed every 2 minutes in real time, which enables operators to deploy variable speed limits and strategically schedule the work.

As a showcase, the researchers utilized Waze to monitor speed change of Interstate 55 during the closure of Memphis Hernando DeSoto bridge. Two routes were designated and added to Waze traffic watchlist, as shown in **Figure 4-21**. Using the historical and real-time travel time, the researchers visualized the travel time change of interstate 55 from the bridge closure to reopening (see **Figure 4-22**). The travel time updates not only provided the important information for detour and rescheduling the travels back then, but also help operators deploy operational strategies. It should be noted that Southbound travel time (see I-55S in **Figure 4-22**) backed to

[4] https://support.google.com/waze/partners/answer/10618039?hl=en&ref_topic=10616686

normal after Jun 14, 2021, after TDOT paved a second lane onto the next ramp near Crump Boulevard[5]. The speed monitoring explicitly showed the benefits of adding a new lane for easing traffic.

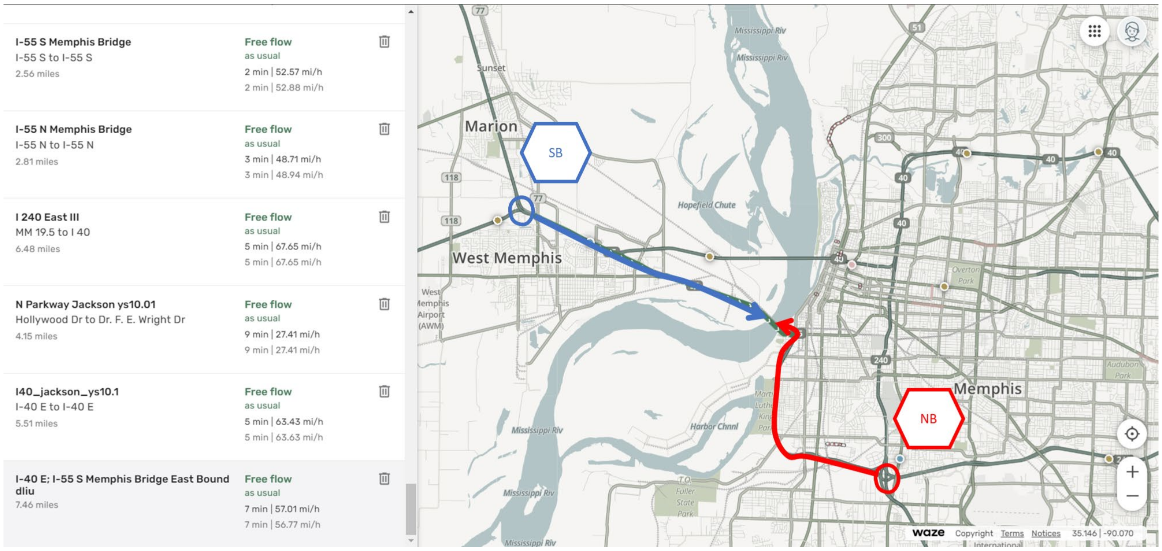


Figure 4-21 Add I-55 to Waze traffic watchlist

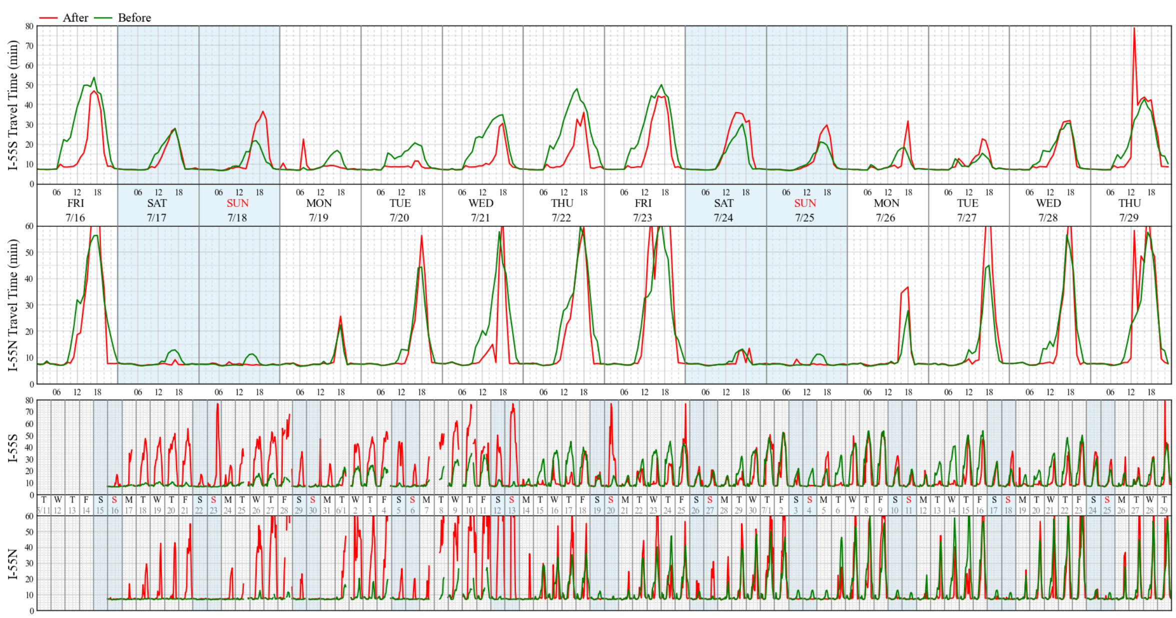


Figure 4-22 Monitor travel time change before and after the closure of Hernando DeSoto Bridge

4.3 Discussions

[5] https://wreg.com/news/added-second-lane-on-i-55-south-at-crump-relieves-traffic-congestion-on-both-sides-of-the-river/

4.3.1 Findings and implications

The prevalence of crowdsourced Waze has become available to transportation agencies under the Waze for cities program, which breaks through the ways of collecting traffic conditions and events. Existing pilot projects and research have shown many merits of Waze data over conventional traffic sensors, yet the data quality issues including redundancy and unreliability should be examined before further implication. This project helped TDOT archive Waze report and speed data every 1 minute, 24 hours a day and seven days a week. Based on the archived data and real time Waze feed, the researchers assessed the quality of Waze data, and explored different use scenarios. The following sections will provide the new findings and recommendations for real-world applications of Waze data.

Waze spatiotemporal accuracy

Through the spatiotemporal comparison between Waze reports and locate/IM records, the researchers found that:

- Primary Waze crash report comes in about 2.2 minutes sooner than locate/IM event time on average. The reported accident locations are only 6 feet upstream of the locate/IM geotag.
- Primary Waze stop vehicles report comes in about 7.8 Minutes sooner than locate/IM event time on average. The reported stopped vehicle locations are 132 feet upstream of the locate/IM geotag.

Event completeness

- 26% of Waze accident reports can be successfully matched with 67% Locate/IM crash reports.
- 14% of Waze stopped vehicle reports can be successfully matched with 85% Locate IM disabled vehicle reports.

It is worth noting that 74% of Waze accident reports and 86% of Waze stopped vehicle reports do not find proximate Locate/IM records. Those unmatched Waze reports could include false reports or truth reports that were not recorded by Locate/IM. To utilize these unmatched reports for complementing Locate/IM data, a spatiotemporal clustering could be conducted, and the results could be verified via other data sources.

Speed quality

The researcher assessed Waze speed by comparing it with Bluetooth speed on surface streets and RDS speed on freeways. We found that:

Speed on surface streets:

- Overall, Waze speed is well aligned with Bluetooth speed for four randomly selected segments on Sevierville surface streets.
- The Waze speed is perfectly matched with the Bluetooth speed during the daytime, while for the rest of the day tends to be lower than Bluetooth speed. This is because the fewer Wazers are active on the road at night than at daytime.

In addition, the daytime speeds also have higher accuracy than nighttime speed for different speed levels.

Speed on freeways:

- In general, Waze speed demonstrates the high correlation (r=0.65) with RDS detected speed.
- Waze speed is higher than RDS speed when Waze speed is smaller than 55 mph which is posted speed limit in study area. On average, for Waze speed 0-45 mph, it is 6.5 mph higher than RDS speed; for Waze s peed group 45-55 mph, it is 3.4 mph higher than RDS speed. On the contrary, Waze speed is lower than RDS speed when Waze speed is higher than 55 mph. On average, it is reported that Waze speed is -8.9 mph lower than RDS speed.

Use scenarios

The researchers explored different use scenarios of Waze data for queue warning, incident management, performance measurement, roadway maintenance and work zone management. The researchers found some promising results.

End of Queue detection

- The EOQ identification time from Waze reports is close to the EOQ detection time from traffic sensor data, with only 1.1 minutes difference on average.
- By using Waze reports, we can dynamic update EOQ. On average, Waze generate 1.9 EOQ detection points every mile which is slightly higher than traffic sensor based EOQ detection frequency (i.e., 1.8). This finding shows the good capability of Waze on detecting EOQ.

TDOT has been dedicated to launching **Protect the Queue** campaign since June 2013. Alerting drivers of the slow-moving traffic ahead is one of the essential tasks at the incident locale. In practice, the proposed method can be widely implemented and embedded into the EOQ warning system, especially the place where traffic sensors are not deployed. Accurately and timely detecting the EOQ will help TDOT improve the effectiveness of alert system.

Secondary crash identification

- For a specific secondary crash, the Waze reported the primary crash and secondary crash 1 minute and 3 minutes earlier than police report, respectively.
- The capability of identifying secondary crash was also investigated. Among 708 crashes, using Waze reports, the researchers identified 17 secondary crashes which includes several secondary crashes that was not identified by speed contour plot method.
- Despite the identification rate is affected by the number of Wazers, the Waze data have potential to serve as an alternative data source for detecting secondary crash.

TDOT has been promoting “quick clearance” of highway incidents. HELP trucks patrol the most heavily traveled freeways in Knoxville, Chattanooga, Nashville, and Memphis seven days a week. They can respond to incident quickly and help clear the incident to avoid secondary crashes. Identifying potential secondary crashes could help truck operators prevent and handle the potential secondary crashes timely and accurately.

Performance measurements

The researcher also explored the Waze speed and alerts for performance measurements. The researchers are among the first to apply Waze speed and alerts for LOS assessment. Taking freeway LOS as an example, the major findings are:

- Waze average speed and number of alerts have distinct average value at each LOS level.
- Waze average speed and number of alerts are two leading factors of LOS, which is more important than travel time reliability metrics.
- This finding suggests that Waze data could mitigate the difficulty of LOS evaluation by eliminating density and flow data which are needed in HCM approach.

The traffic performance measurements are of importance as they help transportation agencies to monitor traffic status. In real-world application, Waze real-time data is very promising to be used for evaluating any traffic improvement projects. For instance, TDOT has been building a smarty corridor on Interstate 24[6]. The operators can use the Waze speed and alerts data to assess the traffic change before and after the project like ramp metering and variable message sign. In addition, the Waze data have great potential on the surface streets. The key criterion of LOS at intersections is delay which is obtained from speed and traffic volume. The Waze data can alter the conventional sensor data that is hard to get for all intersections. What's more, signal operators can deploy signal retiming efficiently and frequently in house via monitoring the performance change from Waze data.

Incident management

- Waze can complement TDOT Locate/IM database not only for crashes and disabled vehicles, but also for congestion, road closure, and road debris and weather events.
- A two-way information delivery channel can be created between Waze and HELP patrol.
- The Waze alerts provides evidence and insights with freeway service patrol expansion.

Roadway maintenance

The researchers explored the Waze reports for pothole detection, pavement condition evaluation and Snow/Ice removal. Some meaningful results were found which proved the Waze's capabilities in roadway management.

- Waze detected many more potholes than TDOT maintenance requests.
- Most of TDOT maintenance requests concentrated on winter and spring. While Waze reports suggest that potholes occur across the year including summer and fall.
- More interestingly, for a short stretch of the road, the Waze reports reveal the seasonal failure, which can potentially reflect the vulnerability of the road section.
- Pavement performance based on the number of potholes is correlated to pavement condition index estimated from TDOT report, which could be used for monitoring pavement conditions frequently.
- Snow/ICE and other weather events are not investigated for a specific case, but the use suggestions were provided in this report.

Waze Pothole reports enrich the way of collecting road pothole information. Compared to TDOT road scanning and citizens' reporting, Waze pothole can timely discover the potholes and share

---

[6] https://www.tn.gov/tdot/projects/region-3/i-24-smart-corridor.html

the accurate the location on all roadways. This will help TDOT repair the potholes efficiently and quickly and prevent motorists from hitting the potholes. Furthermore, Waze pothole reports can be used to speed up the pavement condition evaluation which was done once a year in the past. Lastly, the snow and ice reports help maintenance crews to find and clear the road snow and ice more quickly. More importantly, they can be fed to 5-1-1 service to warn motorists.

Work zone management

The goal of applying Waze in work zone management is to help crews to set up warning signs to have traffic slow down. With the help of Waze partner hub, the crews can add segments around the work zone area and continuously monitor speed. The speed could be used for setting up variable speed limit sign, dynamic message sign, speeding warnings and so forth. In a real word case, the researchers monitored travel time change before and after the Memphis bridge closure, which helped motorists in detours. The travel time also indicated the remarkable benefit after TDOT added a new lane to the exit ramp.

4.3.2 Limitations and improvement

Although the researchers demonstrated many inviting advantages and promising applications of Waze data for transportation management, there are several limitations in these studies. To mitigate this shortage, the corresponding improvements are also provided for future implementation.

Quality for other events remains unknown

In this project, the researchers showcased the quality of Waze crash and stopped vehicles reports. While for other Waze reports mentioned in section 4.2.4, their quality remains to be verified before further application. The spatial and temporal accuracy of different report types should be distinct as the different events have different impact on traffic status. Therefore, the threshold for matching Waze reports and ground-truth should be investigated.

Unmatched reports are not examined

As shown in section 4.1.2, 74% of Waze accident reports and 86% of Waze stopped vehicle reports do not find proximate Locate/IM records. The unmatched reports should be explored and verified by other data source such as social media. The spatiotemporal clustering algorithm could be implemented for mitigating the duplicated and false reports.

Wazers population affects the speed and detection capabilities

Like many other probe vehicle-based speed data, Waze speed collocation is also relied on Wazers. The crowdsourced data are collected by multiple individuals that are a sample of population. The greater number of active Wazers on the road, the more accurate of speed. If there is no Wazers observed during the update time interval, the speed will be missed, and historical data will be replaced. Therefore, the future study could focus on Waze speed imputation and calibration.

For EOQ detection and secondary crash identification, sufficient Waze accident or Jam reports are needed to guarantee the frequent report. In future research, their performance will be evaluated for different traffic jam conditions. If necessary, Waze data should integrate the conventional data source to make the proposed method adaptive to more scenarios.

4.3.2 Lessons for making full use of Waze data

In transportation field, the terminate goal of collecting Waze data is to make full use of Waze data for transportation operation and management so that multiple users including operators in management center and motorists on the road can benefit from the traffic project. Therefore, the researchers summarized and provided the following suggestions learned from existing successful applications and proposed applications in this project.

Data archive

Archiving the data is imperative to support continuous study and agency's planning. Without the historical data, it's hard to know what traffic looked like in the past, let along how to improve the traffic management. Besides, data archive makes data sharing possible in the future.

Data integration

As shown in the cases studies, Waze data have many advantages such as coverage and real-time feed over established data source, yet it also has some shortages like dependency on Wazers. Thus, it is better to integrate Waze with existing data source to enhance the overall capabilities. Taking EOQ as an example, the RTMS detectors are separated by a certain interval where the speed between detectors is unknown. Using detectors speed information to detect the EOQ is accurate, yet it is highly limited by the density of detectors. However, with the Waze reports, the location resolution could be improved. Thus, combing them for EOQ detection would augment the overall detection benefit.

Data conflation

Data conflation is a challenge work but would benefit for data analysis and long-range planning. For Waze reports, it can be directly projected to existing geometry files. The speed data are segmented based which might be different from the geometry of TDOT network files. Thus, conflation makes speed available at TDOT network segmentation level. With the conflation, many road attributes like AADT, number of lanes are available, which will facilitate traffic analysis.

Real-time application

The crowdsourced data can be access through partner hub with no or few time lags. Due to the potential redundancy and unreliability of the Waze data, real-time preprocessing is required for applying Waze data for real-time scenarios. The time of preprocessing influences the value of Waze reports. For emergent events like the accident, the earlier detection from Waze reports save time for response. Thus, a series of proactive measures should be initiated when the crash report is monitored.

Chapter 5 Conclusion

Under crowdsourcing context, ubiquitous road users have become real-time moving sensors which can provide low-cost, wide-coverage, valuable data on traffic operation, roadway conditions, travel patterns and so on. Leverage the opportunity from Waze for cities program, this project aims to help TDOT archive Waze feed, assess Waze data quality, review the existing use scenarios, and provide the promising use scenarios for TDOT management and operations programs.

Firstly, an extensive quality assessment was conducted for both Waze reports and link-based speed. A spatial-temporal thresholds-based approach was proposed to match the Waze reports and ground-truth from Locate/IM. The results show that the first Wazer reports the event earlier than official record on average, and their reported location is also close to the locale. Speed was evaluated for freeways and surface street respectively. The difference between Waze speed and Bluetooth speed, RTMS detector speed were quantified at each speed level and different time of the day. The results indicate that Waze speed was well aligned with Bluetooth speed on surface street at daytime. Waze speed is likely to be distinct from the Bluetooth speed at night. Besides, the Waze data have a smaller percentage of error in higher speed ranges. The Waze speed collected on freeways suggest that the Waze speed tend to be higher than the RTMS detector speed when Waze speed is lower than the post speed limit, while the Waze speed is faster than RTMS detector speed when traffic is moving beyond the posted speed limit. The speed quality is influenced by the active Wazers on the road.

Secondly, several use scenarios were examined for TDOT development. In severe traffic conditions, the queue-end detection algorithm shows that Waze data is a reliable alternative for roadside sensor data in the application of queue-end detection. Besides, the Waze data contributes to identifying secondary crashes. The proposed network ST-DBSCAN clustering method can be integrated with commonly used method for more accurate identification of secondary crashes. In addition, the Waze data can be used to construct performance indicators. The researchers used the number of reports and Waze speed to evaluate LOS on freeways and found that they are two leading indicators of LOS levels. This finding could potentially eliminate the collection of density and flow data in HCM method. This work can also be extended to surface street LOS evaluation in a similar fashion. For the incident management, we proposed that other types of events could be utilized to complement the Locate/IM database. The real-time Waze reports can be fed to HELP patrol, leading to faster and more accurate responses to traffic incidents and other emergent events. In terms of roadway maintenance, Waze reports are a very timely and low-cost source of collecting geolocated potholes. What's more, a correlation test indicates that the pothole reports can be used for pavement condition evaluation. Finally, the Waze data can be used to monitor the speed around the work zone areas as demonstrate by the case of Memphis bridge closure a w TDOT can push the work zone schedule to Waze to alert travelers to detour or slow down when approaching the work zone areas.

However, there are still many other applications not investigated in this project, yet they could facilitate the transportation management and operations. For instance, the future work could investigate the signal retiming based on Waze data. Meanwhile, utilizing Waze data should not stop by treating them as a competitive role. Instead, such crowdsourced data are compensating conventional data in many ways. The integration process should focus on both methods and data source aspects. In general, by examining the feasibility of real-world applications using Waze data, TDOT is positioning itself to better leverage the rapidly putting new technologies to positively improve the traffic management and operations.

References

[1] TDOT. "TDOT Launches Plan for Pothole Repair." https://www.tn.gov/tdot/news/2022/1/21/tdot-launches-plan-for-pothole-repair.html (accessed Feb, 12, 2022).
[2] Wazeopedia. https://wazeopedia.waze.com/wiki/USA/Waze_for_Cities (accessed Feburary 2, 2022).
[3] Wikipedia. "Waze." https://en.wikipedia.org/wiki/Waze (accessed October 6, 2022).
[4] Waze. "Waze company factsheet." https://assets.brandfolder.com/p31v19-dpmnts-ci9it3/original/Waze%20Company%20Factsheet.pdf (accessed Feburary 2, 2022).
[5] Y. Liu, N. Hoseinzadeh, L. D. Han, and C. Brakewood, "Evaluation of Crowdsourced Event Reports for Real Time IMplementationâ€"Spatial and Temporal Accuracy Analyses," 2019.
[6] M. Amin-Naseri, P. Chakraborty, A. Sharma, S. B. Gilbert, and M. Hong, "Evaluating the reliability, coverage, and added value of crowdsourced traffic incident reports from Waze," *Transportation research record,* vol. 2672, no. 43, pp. 34-43 %@ 0361-1981, 2018.
[7] (2019). *DES MOINES METROPOLITAN AREA INTEGRATED CORRIDOR MANAGEMENT (ICM)*. [Online] Available: https://iowadot.gov/desmoinesicm/pdf/DesMoines-ICM-ProgramConOps.PDF
[8] "Crowdsourcing for Operations Case Study Lake County, Illinois Division of Transportation," ed: U.S. Department of Transportation Federal Highway Adminstration, 2020.
[9] J. R. Effinger. "Integration of Crowdsourced Data into Automated Traffic Signal Performance Measures (ATSPMs)." https://www.fhwa.dot.gov/innovation/everydaycounts/edc_6/docs/crowdsourcing_case_study_lake_county.pdf (accessed October 6, 2022).
[10] FHWA. (2019) Crowdsourcing for operations case study City of Louisville, Kentucky.
[11] K. T. Cabinet. "Work Zone Monitoring." https://transops.s3.amazonaws.com/uploaded_files/Kentucky%20Slides%20-FHWA-NOCoE%20Webinar-May%202021%20EDC6.pdf (accessed Feburary 2, 2022).
[12] F. Kunkle, "Can Waze speed up emergency response times? Perhaps, as data-sharing becomes two-way street," ed. The Washington Post, 2016.
[13] "EDC-5 Innovations." https://www.fhwa.dot.gov/innovation/everydaycounts/edc_5/docs/crowdsourcing_applications.pdf (accessed Feburary 2, 2022).
[14] N. Goodall and E. Lee, "Comparison of Waze crash and disabled vehicle records with video ground truth," *Transportation research interdisciplinary perspectives,* vol. 1, pp. 100019 %@ 2590-1982, 2019.
[15] S. Praharaj, F. T. Zahura, T. D. Chen, Y. Shen, L. Zeng, and J. L. Goodall, "Assessing trustworthiness of crowdsourced flood incident reports using Waze data: a Norfolk, Virginia case study," *Transportation research record,* vol. 2675, no. 12, pp. 650-662 %@ 0361-1981, 2021.
[16] C. Lowrie, A. Kruczkiewicz, S. N. McClain, M. Nielsen, and S. J. Mason, "Evaluating the usefulness of VGI from Waze for the reporting of flash floods," *Scientific reports,* vol. 12, no. 1, pp. 1-13 %@ 2045-2322, 2022.
[17] X. Li, B. Dadashova, S. Yu, and Z. Zhang, "Rethinking highway safety analysis by leveraging crowdsourced waze data," *Sustainability,* vol. 12, no. 23, p. 10127, 2020.
[18] Y. Senarath, S. Nannapaneni, H. Purohit, and A. Dubey, "Emergency incident detection from crowdsourced waze data using bayesian information fusion," 2020: IEEE, pp. 187-194 %@ 1665419245.

[19] R. Nune, W. Zhong, K. Fu, and J. X. Tao, "Exploring social traffic data for evaluating urban arterial congestion," ed, 2017.

[20] A. Kumar, C. L. Ross, A. Karner, and R. Katyal, "Traffic and Incident Management Using Crowdsourced Data: Challenges and Lessons Learned from Practice," ed, 2018.

[21] Y. Liu, Z. Zhang, L. D. Han, and C. Brakewood, "Automatic traffic queue-end identification using location-based Waze user reports," *Transportation research record,* vol. 2675, no. 10, pp. 895-906 %@ 0361-1981, 2021.

[22] S. R. Hargrove, H. Lim, L. D. Han, and P. B. Freeze, "Empirical evaluation of the accuracy of technologies for measuring average speed in real time," *Transportation Research Record,* vol. 2594, no. 1, pp. 73-82 %@ 0361-1981, 2016.

[23] Waze, "Under the Hood: Real-time ETA and How Waze Knows You're on the Fastest Route," vol. 2020, ed, 2020.

[24] N. Hoseinzadeh, Y. Liu, L. D. Han, C. Brakewood, and A. Mohammadnazar, "Quality of location-based crowdsourced speed data on surface streets: A case study of Waze and Bluetooth speed data in Sevierville, TN," *Computers, Environment and Urban Systems,* vol. 83, pp. 101518 %@ 0198-9715, 2020.

[25] Z. Zhang, Y. Liu, L. D. Han, and P. B. Freeze, "Secondary crash identification using crowdsourced Waze user reports," *Transportation research record,* vol. 2675, no. 10, pp. 853-862 %@ 0361-1981, 2021.

[26] Lynnette, "Drivers In Tennessee: Dial *THP For Roadside Assistance & To Report Road Rage Or A Driver Under The Influence," ed: FRANKLIN / NASHVILLE TN GUIDE.

[27] B. Huang and X. Jia, "Quality Improvement and Application of TDOT Pavement Management Systems (PMS) Data," 2015.