1-2023

Network Effects of Disruptive Traffic Events

Juan Medina
University of Utah

Xiaoyue Cathy Liu
University of Utah

Recommended Citation
Medina, J. C, Liu, X. L Network Effects of Disruptive Traffic Events. NITC-RR-1082. Portland, OR: Transportation Research and Education Center (TREC), 2022. https://doi.org/10.15760/trec.285

NITC
NATIONAL INSTITUTE for
TRANSPORTATION and COMMUNITIES

Final Report 1082
January 2023

Photo by Denys Nevozhai on Unsplash

Network Effects of Disruptive Traffic Events

Juan Medina, Ph.D.

Xiaoyue Cathy Liu, Ph.D., P.E.

THE UNIVERSITY OF UTAH

Network Effects of Disruptive Traffic Events

Final Report

NITC-RR-1082

By

Principal Investigator:
Juan Medina, Ph.D.
University of Utah

Co-Principal Investigator:
Xiaoyue Cathy Liu, Ph.D., P.E.
University of Utah

for

National Institute for Transportation and Communities (NITC)
P.O. Box 751
Portland, OR 97207

NITC
NATIONAL INSTITUTE for
TRANSPORTATION and COMMUNITIES

January 2023

Technical Report Documentation Page

1. Report No. NITC-RR-1082	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Network Effects of Disruptive Traffic Events		5. Report Date January 2023 6. Performing Organization Code
7. Author(s) Juan Medina, University of Utah, 0000-0001-5302-8814 Xiaoyue Cathy Liu, University of Utah, 0000-0002-5162-891X		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Civil & Environmental Engineering University of Utah 110 Central Campus Drive, Suite 2000 Salt Lake City, UT		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 69A3551747112
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE, Washington, DC 20590		13. Type of Report and Period Covered 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Current traffic management strategies are based on expected conditions caused by recurring congestion (e.g., by time of day, day of week), and can be very effective when provisions are also given for reasonable variations from such expectations. However, traffic variations due to non-recurrent events (e.g., crashes) can be much larger and difficult to predict, making also challenging efforts to identify, measure, and forecast their disruptive effects. This project explores a proactive approach to deploy a tool for managing non-recurrent congestion by identifying and quantifying the effects of disruptive traffic events at a microscopic level using a comprehensive set of data sources. A combination of resources including detailed near-time crash records, high-resolution vehicle detection activations and deactivations, as well as traffic signal phasing and timing, are combined together to build an understanding of standard traffic patterns, store this knowledge, and compare it with new incoming data for event identification. The team explored the use of high-resolution data for this purpose at surface street and arterial levels, and the outcomes from model fittings in such scenarios. Upon deployment using virtual servers and interfaces developed by the University of Utah team, ingestion of daily data and event detection will build up of a library of events and their effects, and this process will continue over time to strengthen the knowledge base on the corridors analyzed. Further outcomes from this research could lead to detailed event-based spatio-temporal congestion and safety models, ultimately enabling informed proactive traffic management and safety countermeasures. This project uses the Salt Lake Valley as a testbed and could open new opportunities for research that relies on the integration of large and disaggregated datasets.
.

17. Key Words Traffic Control, Traffic Events, Traffic Management		18. Distribution Statement No restrictions. Copies available from NITC: www.nitc-utc.net	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 37	22. Price

ACKNOWLEDGEMENTS

This project was funded by the National Institute for Transportation and Communities (NITC; grant number 1082) a U.S. DOT University Transportation Center. The project also benefitted from matches from the University of Utah and the Utah Department of Transportation.

RECOMMENDED CITATION

Medina, J. C, Liu, X. L Network Effects of *Disruptive Traffic Events*. NITC-RR-1082. Portland, OR: Transportation Research and Education Center (TREC), 2022.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Current traffic management strategies are based on expected conditions caused by recurring congestion (e.g., by time of day, day of week), and could become very effective when provisions are also given for reasonable variations from such expectations. However, traffic variations due to non-recurrent events (e.g.. crashes) can be much larger and difficult to predict, rendering efforts to measure and forecast their disruptive effects a challenging task.

This project explores a proactive approach through a simple framework to deploy a tool for managing non-recurrent congestion by identifying and quantifying the effects of disruptive traffic events at a microscopic level using a comprehensive set of data sources. A combination of resources including high-resolution vehicle detector data, and traffic signal phasing and timing data, together with highly detailed crash records hold promise not only to identify events in the network, but also to monitor for anomalies. A number of additional elements can complement the proposed approach by including external factors such as weather or planned short-term construction.

Data preparation and management played a significant role in the project development, as the team had initially proposed the use of video images for event detection and quantification. However, exploration of this alternative did not yield consistency in the data acquisition due to limited and uncontrolled field of view, as well as image processing limitations using images from cameras with changing orientation and angles (i.e., PTZ cameras).

Data collection methods were later expanded to access the newly available high-resolution datasets from the Automated Traffic Signal Performance Measures (ATSPM) managed by the Utah DOT (UDOT). These datasets provide similar capabilities as those traditionally available for freeway systems, where detector and counting stations are regularly spaced along the facilities. Instead, ATSPM collects all vehicle detection activation and deactivation calls and signal controller phasing and signal indication, using a standardized set of enumerations for decoding and encoding.

Custom scripts were required to read and process the high-resolution data and extract metrics of interest that may characterize typical approach-level performance. These included time-series of vehicle volumes, measures of signal coordination such as vehicles arriving in green, and measures of speed and occupancy.

Data analysis incorporated proof of concepts of simple but practical algorithms that could play a significant role in regular monitoring, namely time-series thresholds, as well as dynamic predictions from autoregressive ARIMA models and machine-learning methods such as Long Short-Term Memory. Ultimately, Long Short-Term Memory neural networks

were preferred given their flexibility to replicate both well-defined trends as well as additional random variations.

Results from a selection of crash events along different corridors in the Salt Lake Valley showed the applicability of the approach in a monitoring environment, but additional validation is needed for this ongoing effort when bringing the proposed framework into a deployment mode. The monitoring tool will continue evolving to build a library of events and their impact on the traffic network, but the exploration conducted as part of this project set the stage for this long-term application that could be shareable to others as soon as the sample size reaches a significant level to make transferability more likely. For example, intersections with certain traffic patterns, number of lanes, and in the presence of an event type have experienced a specific range of effects in time and space.

Outcomes from this research are expected to lead to shareable event-based spatiotemporal congestion and safety models, ultimately enabling informed and proactive traffic management and safety countermeasures. This project uses the Salt Lake Valley as a testbed and could open new opportunities for research that relies on the integration of large and disaggregated datasets.

1.0 BACKGROUND AND OBJECTIVES

This project aims to take advantage of a combination of datasets generally not available for analysis of traffic conditions, particularly in the presence of disruptive events. Integrated datasets encompassing high-resolution vehicle detection data and detailed crash records constitute the core data source for this analysis. Different from traditional analysis of events or incidents where the occurrence of events is to be predicted or modeled, the primary efforts in this project are directed at measuring their impacts upon occurrence.

There is much to learn from the effects of incidents or other non-recurrent events. Most research in this area has been directed to freeways, where fixed-location and mobile sensor coverage has traditionally been more extensive than other surface streets. As sensor presence has become more ubiquitous and data collection capabilities continue advancing, new opportunities have opened for this type of research on surface streets and, particular for this research, along arterials.

Initial conceptualization of this project was aimed at the integration of video data from live feeds available to the research team, as well as detailed crash data reports. However, while the research progressed, the team redirected its main interests towards newly available datasets with high-resolution vehicle detections from the Automated Traffic Signal Performance Measures (ATSPM). ATSPM provides analytical tools and approaches derived from high-resolution traffic controller data (i.e., traffic signal states and vehicle detections) collected and transformed automatically to produce actionable performance measures (FHWA, 2020).

With access to these rich datasets, the team evaluated the pros and cons of the initial approach using video images, making the decision to invest the subsequent efforts into ATSPM data in place of video images. While video certainly had promising utilization, camera angles, camera height, and zoom levels were often not favorable to observe crash events and their effects as they occurred. In addition, given their changing zoom level and field of view, it was challenging to automate extraction of related data such as volume and queues as a result of a given incident. ATSPM offered a systematic data collection process at the individual vehicle level that was difficult to maintain outside of the team's main focus.

As the team hosts the State of Utah's crash database using a custom content management system, there is accessibility to crash events in near time and details that are often not available for research purposes. These include a complete set of coded values, crash diagrams, and narratives. Coupling detailed crash data with high-resolution vehicle detection and traffic signal states, the team can tackle questions surrounding not only the incident itself, but the magnitude and duration of its effects in space and at specific locations.

As the team continues collecting data and enhancing the ability to develop robust models for individual locations, research from this project has the potential to provide transferable insights to answer different types of questions.

The main objectives leading to the research project were the following:

1) To create a testbed for advanced real-time traffic mobility using unique capabilities that already exist at the University of Utah’s Traffic Lab, but are not yet integrated. It is anticipated for this testbed to support continued research on a wide range of topics inherent to proactive traffic management and mobility strategies;

2) To develop methods to develop congestion models that describe the temporal and spatial effects as a function of network characteristics and for a range of event types, including crashes (e.g., by crash severity, time of occurrence, and corresponding traffic demands) and special events (e.g., major sport and cultural events); and

3) To quantify the safety effects of those events by integrating crash data from a large geographical area, and to describe the temporal and spatial effects of crashes in traffic networks.

This project was designed with the intent of providing the groundwork for future research seeking to identify incident response alternatives based on individualized traffic and environmental context, anticipate their intended outcomes on network performance, and use such forecasts to select optimal response strategies.

2.0 METHODOLOGY

The data integration makes use of existing infrastructure at the University of Utah's Traffic Lab to create a testbed for network-wide traffic analysis. A unique source of data allows the team to access crash reports in near time (less than 24 hours after the events are reported). The UofU has developed a content management system for crash records through a related project titled "Crash Data Initiative" (UTAPS-CDI), including databases and web interfaces through which users can interact with the data.

In addition, the team has direct access to API services that query the ATSPM database in Utah. ATSPM covers hundreds of signalized intersections throughout the state, with the highest concentration of locations in the Salt Lake Valley. The general ATSPM interface allows the user to produce metrics given an intersection ID, a time frame, and specific parameters (if any) for selected metrics. While this interface is highly valuable for visualization purposes, it does not allow for further analysis of the raw data. More in-depth access can be gained through an authenticated account to download raw datasets and open opportunities for new analysis and data integration. The team uses a set account type to access raw datasets with individual vehicle detections and traffic signal states at a resolution of 0.1 seconds.

A number of considerations need to be addressed before the data can be properly integrated. First, not all intersections are defined identically, so additional work is required to map traffic signal and sensor settings to outputs. In general, ATSPM follows standard event codes to identify specific updates in sensor or traffic signal states so a sensor activation will always have the same coded number, but the output channel for such a sensor may be associated with a different lane depending on the approach and intersection.

Evaluation of crash data also requires a number of steps before integration, including quality control on the specific location of the crash. Crash data in its original form does not guarantee that the assigned coordinates of a crash (if any) are correct, and the precise location of the crash needs to be verified for general accuracy purposes. In addition, crashes are typically located along the roadway centerline, so a route number and a milepost can be associated with linear referencing systems. So, the use of narratives and crash diagrams is essential not only to confirm the crash location along the road, but also to identify the lane where the crash occurred, the circumstances of the crash, the location of impact, and the resting place of the vehicles after the crash occurred, when available.

Crash location and timing information is then integrated into the ATSPM raw files to identify the likely moment the crash occurred. It should be noted that crash times from law enforcement reports provide only an approximation of the actual timing of the crash, so “anomalies” in the ATSPM data will be used to identify the precise moment that traffic was disrupted as a consequence of the crash event. The need to detect an “anomaly” in traffic conditions implicitly requires the definition of a baseline or a characterization of “normal” traffic conditions not only to identify the moment when the crash likely occurred, but also the moment that traffic returned to “normality” so the duration of the disruption can be also quantified. Lastly, the magnitude of the changes in traffic conditions can also be characterized and modeled over time.

A number of methods to identify traffic disruptions have been proposed in previous research, as described in the literature review. They range from fixed thresholds to probabilistic analysis, time-series analysis, and machine-learning approaches. This research explores the adequacy of some of these methods under a variety of locations and traffic conditions, and focuses on a machine-learning application for online model development to maintain knowledge in terms of current expected traffic patterns and to identify and quantify effects of events at a given location.

It is important to highlight that unlike most previous research, this work is based on empirical data without the need for simplifications in simulated environments. Also, incident detection in terms of detection and false-alarm rates is not intended to be the main goal of the study, but the quantification of the effects given that starting point of the event is identified.

Moreover, associations between crash characteristics and resulting magnitude of traffic disruptions are also investigated. Both standard and non-standard elements from crash data are incorporated in the analysis. Examples of standard elements include coded values for well-defined crash characteristics, including crash severity, manner of collision, vehicle types, and vehicle maneuvers. On the other hand, non-standard elements include information on vehicles’ resting place post-collision and additional data from narratives and diagrams to help inform the post-crash scene.

Ultimately, the exploration of events and the magnitude of their disruptions is intended to be characterized. With enough observations of specific “types” of events, models therefore become more representative of the actual responses of the network.

The team has proposed the general framework in Figure 2.1 to accomplish the main objectives, and has explored the applicability of integrating these elements as part of this project. Full-scale deployment will continue after this exploratory phase and will require technical expertise, virtual and physical resources that are accounted for as part of the UTAPS-CDI long-term initiative at the University of Utah.

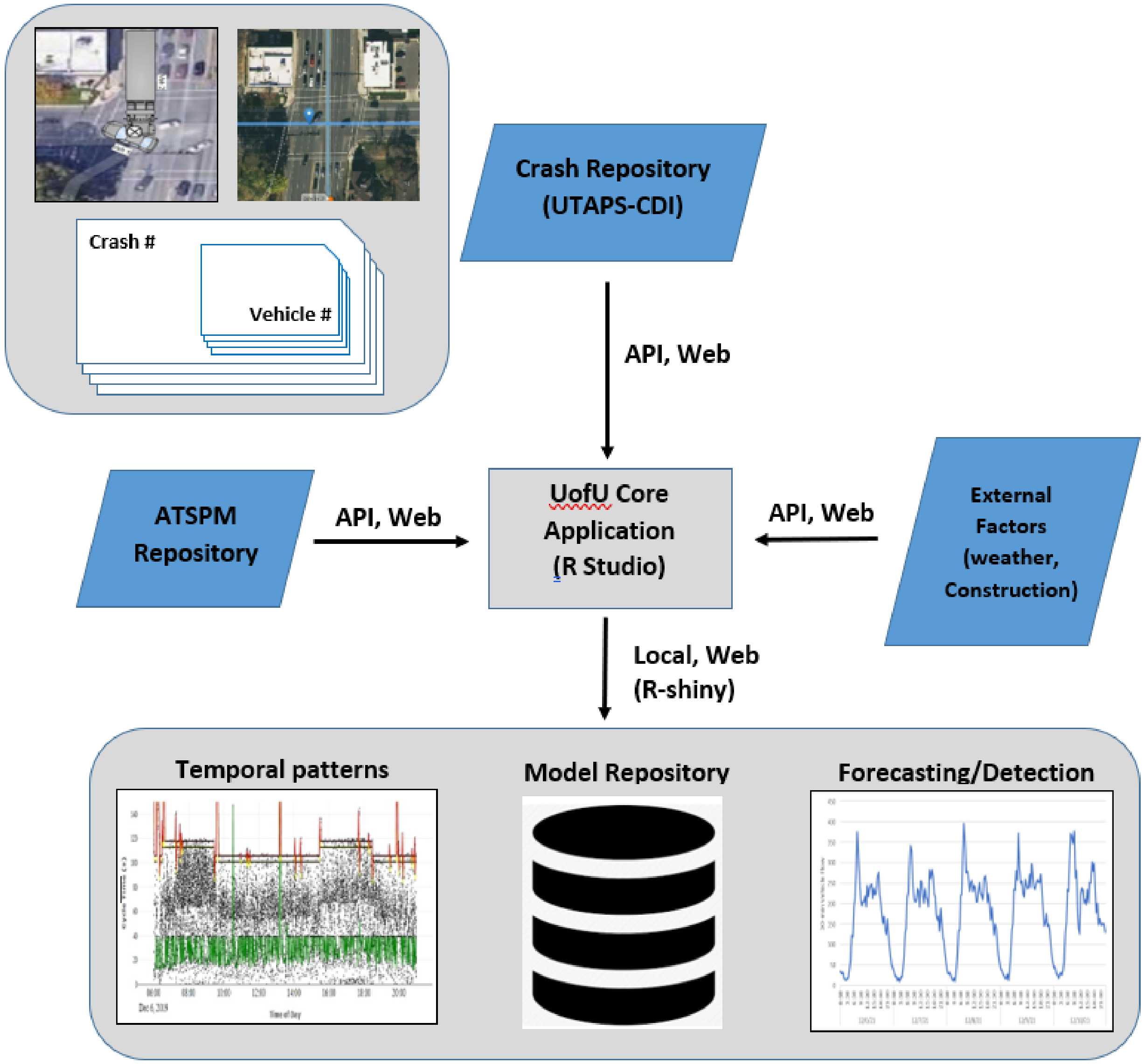


Figure 2-1: Schematic Representation of the Proposed Data Integration System

3.0 LITERATURE REVIEW

This chapter reviews previous studies on incident analysis on arterials and, more generally, on interrupted traffic conditions with the objective of identifying gaps in past data analysis and data integration initiatives. To our knowledge, efforts to conduct incident analysis from a combination of crash data and high-resolution signal and vehicle detection data have been rather limited, particularly using extensive field data. Moreover, studies aim to generally estimate the probability of an event (with little research dealing with the effects of such events, particularly on arterials), but not to build knowledge from a particular network to improve future long-term operation.

3.1 AUTOMATED INCIDENT DETECTION

An area more closely associated with the objectives of this work deals with incident detection. This is a well-developed area, mostly covering uninterrupted traffic (e.g., freeways), but it has also been addressed in the context of arterial roadways by a relatively small number of researchers (Chen et al., 2016). Evans et al. (2020) estimated that only about 10% of traffic incident algorithms have been developed for urban networks, with the remaining 90% targeting freeways. Moreover, only a few of them have been tested on real data or implemented in the field, and those implemented have difficulties with numerous challenges in real-world applications, including sport events or holidays.

Incident detection is mostly centered around identifying anomalies in traffic significant enough to trigger an event alert, so inherently it is a decision-making process. In order to detect such anomalies, an understanding of baseline conditions needs to be established over time and/or space. Such a baseline could be built upon preset thresholds, or dynamic values, or from more advanced algorithms using alternative data sources different from standard vehicle detection, including video processing.

Similar to our objectives, the effects of incidents in “normal” traffic are quantified to estimate the occurrence or the probability of occurrence of such events. However, incident detection research does not necessarily account for the duration and severity of the traffic disruption, so from this perspective the work proposed in this study could provide additional inputs to enhance incident analysis.

Comparative analysis is perhaps one of the main building blocks of automated incident detection, and includes well-known and now traditional algorithms such as those proposed by Payne and Tignor (1978), also known as the California algorithm, and

Chassiakos and Stephanedes (1993) by smoothing of occupancy measures to distinguish short-duration traffic homogeneities from incidents (i.e., a low-pass filter). A somewhat related approach was also proposed by Persaud (1990) with the McMaster algorithm, which separates a flow-occupancy diagram into areas corresponding to three different traffic conditions, and identifies incidents if specific changes in traffic are observed in a short time period or if sudden speed drops are detected. This is essentially realized in a two-step process: first identifying congestion, and then determining if the congestion is attributable to an incident.

However, these popular solutions were designed for freeway environments without a clear path for application along arterials. Nonetheless, some examples can be found in the literature related to urban networks with interrupted flows. Examples of different methods are provided below and include research using threshold-based approaches and discriminant analyses, Bayesian or belief networks, and machine-learning.

Threshold-based real-time incident detection was proposed by Ahmed and Hawas (2012), using standard vehicle detectors and measures of volume and speed over each traffic signal phase split, for a total of eight variables for four splits (later reduced to six variables given no volume during red phases). Standard regression analysis was conducted with limited success, where variables were significant in detecting an incident but had false-alarm rates over 10% using simulated data. The study indicated significant refinements needed beyond a standard regression method.

Sethi et al. (1995) presented an analysis of simulated data including fixed sensor locations along a nine-intersection arterial. Discriminant analysis was used to classify incidents, and basic concepts related to expected changes in traffic conditions due to an incident served as the core ideas for the proposed algorithms. These expectations included higher occupancy and slower flows upstream of an incident, and lower occupancy and flow speed downstream. Deviations and ratios of metrics derived from volume and occupancy were used as inputs. Results showed superior performance when combining both volume and occupancy and when the detection was mostly performed upstream of the incident, with only marginal improvements by adding downstream detectors. Probe vehicle data was also considered, and the deviation and ratios of travel times and speeds were used in the algorithm. Adequate prior probabilities were deemed important and to influence the detection rates and false-alarm rates, but they could only be estimated through sensitivity analysis given the lack of comprehensive field data.

Sermons and Koppelman (1996) also used discriminant analysis for incident detection, but derived data from field-collected vehicle GPS traces. Traffic conditions observed on the field indicated if there was an incident or not, serving as ground truth. Due to difficulty of collecting trajectories during actual crash incidents, short-term lane closures were assumed to be incidents, resulting in about one-third of the traversals having an incident (out of 154 traversals). Improvements were obtained when total travel time was decomposed into stopped and running times, and adding variables such as coefficient of variation of speed, speed noise, etc. may be marginally beneficial. Overall, large and accurate data collection for model development and validation was difficult to achieve in the context of actual non-recurrent incidents.

Zhang and Taylor (2006) presented an automated incident detection algorithm based on Bayesian (or belief) networks, where expert knowledge was combined with traffic data to generate a decision on the presence of incidents. More specifically, volume and occupancy were extracted from a subject site and also at a location upstream using both stop bar detectors for volume and advance detectors for occupancy. Simplified traffic states (e.g., low, medium, high) are used for evidential reasoning, and multiple scenario-specific Bayesian network approaches were used to deal with complicated arterial road incident detection problems. Simulation data was used to test the algorithm, resulting in low false-alarm rates (0.62%) with detection rates of up to 88%. Detection rates and alarm rates were also shown to be stable when the incident detection threshold was set between 65% and 80%.

In addition to traditional methods such as those mentioned above, more complex methods for incident detection seem to be needed. Machine-learning and less traditional approaches have evolved within the last two decades, including support vector machines (SVM) and more notoriously neural networks.

Classification using SVM has been proposed for incident detection. Wang et al. (2018) proposed a classification process using an adaptive booster classifier to identify data outliers, indicating the potential presence of an incident, and a SVM method to further classify the outlier into a set of categories. Simulated data was used to complement field data collected at a signalized intersection, and principal component analysis was used to reduce the dimensionality of the feature vector that contained volume averages over four consecutive cycles, with 80 cycles (i.e., dimensions) per day and movement. Outliers were classified into those due to recurrent congestion or due to non-recurrent congestion. The overall accuracy of the hybrid Adaptive Boosting SVM method produced an average of 92% prediction accuracy.

On a different approach, but featuring an additional module to adjust signal timing settings after an incident, Hawas et al. (2020) proposed the incident detection algorithm to be based on deviations with respect to historical records and their standard deviation. Sets of fuzzy variables indicated the occurrence of a potential incident via flags if the metrics fell in the “high” variation category. Parameters to define such categories were based on the central limit theorem, so values over 175% standard deviations of the normal distribution of detector readings were assigned a “high” label. Then, an incident index transformed the fuzzy variables to a numerical scale and the indices were added over the section being analyzed for a final score. The methods can be calibrated for improved detection and false-alarm rates, but it always depends on such calibration over the assumption of a given distribution of traffic volume variations and impacts caused by incidents. It is also noted that this study used simulation data for model calibration.

Ghosh and Smith (2014) implemented an arterial version of four machine-learning algorithms (three neural network-based and one SVM-based) typically used for freeway incident detection. The neural networks included a multi-layer, feed-forward neural network, a probabilistic neural network, and a fuzzy-wavelet radial basis neural network. The main objective of the research was to investigate the transfer of well-performing algorithms for extensive urban network usage under the general premise of pre-processing data to reduce the effect of signal presence by scaling traffic volume and occupancy data. Results were based on simulated data and provided an initial exploration with positive results, but significant work ahead prior to implementation.

Research focused on automated incident detection is extensive and mostly directed to freeway applications. However, examples can be drawn for arterial roadways using a range of methods including threshold-based or discrimination-based classifications, belief networks, SVM and neural networks, among others. Even though the examples illustrated above do not represent a comprehensive list of methods, a common denominator of such research emerges. That is, the complexity of arterial scenarios and the difficulty to test algorithms using field data, lead to most evaluations using simulation data only or limited field data supplemented by simulation.

Overall, machine-learning applications have gained momentum in the incident detection arena and, compared to more traditional parametric tools, they seem to provide additional flexibility not only in terms of time-series trend modeling, but also in terms of computational efficiency. Online applicability and modeling updating is an essential objective for the large-scale deployment of the proposed framework, and thus significant consideration will be given to this group of methods in testing and exploration phases.

3.2 ANALYSIS ON INCIDENT DURATION

Complementary to incident detection, analysis and prediction of incident duration and clearance times is also an important area of research when the objective is to quantify the impacts of traffic incidents. Similar to incident detection, most research in this area has been directed at freeway systems. Data granularity and availability seem to be an issue for analysis on arterials, with only a few studies dealing with such applications. Likewise, limitations on crash data not only stem from the usage of high-level crash characteristics but also on the accuracy of geographical and/or environmental attributes. In addition, traffic data may be limited to conditions during the incident occurrence, without equivalent data to analyze “baseline” conditions.

Raub and Schofer (1997) compiled data from traffic events in the Chicago area and reported incident duration times by incident type based on the moment from which a dispatcher was notified until the officer reported clear of the scene. An average crash duration of 57 minutes and deviation of 35 minutes was estimated from reports, with injury crashes having a longer average clearance time of 71 minutes. Crashes were collected on arterials at times when they were at or near capacity, but traffic data was not available for further assessments even though some preliminary data from a parallel study indicated capacity reductions by more than 60% for crashes and more than 50% for disabled vehicles when one lane was blocked out of four lanes on an arterial. Crashes resulting in injuries generated the most severe scenario observed in the study, with only

32% of capacity available for traffic. Similarly, minor crashes with a disabled vehicle resulted in a remaining capacity as low as 43% of the arterial directional capacity. (Raub and Pfefer, 1998).

Challenges of prediction of duration of incidents are related to relationships between vehicle detectors and the event, and difficulties obtaining consistent traffic information associated with said event. Using machine-learning and a bi-level prediction framework combining classification and regression, Mihaita et al. (2019) analyzed crash reports and traffic data (15-minute flows when the crash occurred, one-hour flow before the crash, and the ratio between the two metrics) to obtain predictions of incident durations on arterials. Also, among other related features, the affected lanes, hour of the day, and speed limit are significant factors in the incident duration. Crash duration was obtained directly from crash reports (as stated by the officer), showing a skewed distribution with an average of 30 minutes and a longer tail to the right including about 10% of records with over 100 minutes of duration. Gradient-boosted decision trees and extreme-boosted decision trees provided the best performance to classify crash durations as either lower or higher than 45 minutes until clearance time.

Several studies have found associations between incident duration and crash characteristics. As pointed out by Nam and Mannering (2000), longer crash durations were linked to higher severities and greater number of involved vehicles. Additional factors increasing clearance time were identified by Chung (2010) for the time of crash; by Junhua et al. (2013) when crashes blocked traffic lanes; in terms of season and weather conditions by Dimitriou and Vlahogianni (2015) and Vlahogianni and Karlaftis (2013); and roadway type by Gu et al. (2021), with crashes taking longer to clear in minor arterials compared to urban arterials and collectors. Gu et al. (2021) pointed out that previous research has also identified errors in crash duration data given the difficulty to observe the crash site from the moment the crash occurred until traffic recovered to pre-crash conditions (Garib et al., 1997; Khattak, Schofer & Wang, 1995). Use of crash duration can be adjusted to reflect the time frame until the response vehicle leaves the scene, as various researches did in the past. Their own study reported average durations in urban arterials and collectors of less than 80 minutes, with most crashes cleared within 50 and 100 minutes.

This research will address a number of limitations in the literature, mainly the expansion of analysis to a corridor or a network-wide scale, where specific locations will be monitored beyond current capabilities of installed systems that use ATSPM metrics. In addition, the idea of model building and storage is new and recognizes the variability of patterns at different locations. These efforts focus on existing techniques, and the development of an application that integrates data sources for long-term use by agencies and researchers alike.

4.0 DATA COLLECTION & MANAGEMENT TOOLS

Throughout the development of this project, the team considered a number of data sources to identify traffic states along urban arterials and to quantify the effects of individual, non-recurring events. Initially, and as a result of capabilities of data processing at the Utah Traffic Lab, the team proposed use of video images from the state's traffic monitoring CCTV camera network. The cameras cover a significant portion of the network managed by UDOT, and at the surface street level they are located at significant intersections within the Salt Lake Valley.

The Utah Traffic Lab was equipped to collect video images, so the team proceeded to identify key high crash frequency locations, where crashes would be more likely to be captured. A total of eight camera feeds were selected and set to record for multiple weeks at a time, so the data collection opened opportunities to capture crash events as they occurred. As the team received a crash report at one of these locations, the video was reviewed to determine if the crash was captured on video. Capturing events in this fashion was expected to provide data to quantify the extent of the effects of such events, so the team could analyze and characterize them.

As this process continued, it was noticed that the field of view of the cameras was not effective at capturing these events, and most of the events provided effective metrics in terms of changes in traffic conditions. For example, even in the case that an event was captured, the extent of queue or the effects on the approaches upstream of the crash were not visible.

4.1 UTAH'S AUTOMATED TRAFFIC SIGNAL PERFORMANCE MEASURES (ATSPM)

After a long period of data collection, the team explored different and more systematic metrics. At the arterial level, UDOT had rolled out ATSPM, as a result of a joint research program that included Indiana DOT, Purdue University, Econolite, PEEK, and Siemens. The ATSPM produces metrics to monitor traffic operations, such as traffic volumes, signal progression or traffic speed, and presents data analysis outputs using a web interface. An example of these outputs is the Purdue Coordination Diagram (Day et al., 2010), to visualize the temporal relationship between the coordinated phase indications and vehicle arrivals on a cycle-by-cycle basis.

To produce these metrics, the system uses detailed vehicle sensor and signal phasing and timing data. Given the large scale of datasets, when such high-resolution data is collected, it is essential for the signal control cabinets to be equipped with advanced communications to transfer the data to a storage location. In Utah, this is performed via fiber-optic communication, so the system is capable of transmitting raw data for ATSPM's calculation to UDOT's Traffic Operations Center (TOC), where it is stored and managed. A series of standardized codes are defined for ATSPM for the system to identify specific sensor calls or signal changes. This protocol is defined by the Traffic Signal Hi Resolution Data Logger Enumerations (Sturdevant et al., 2012).

Interfaces from ATSPM allow a user to query the large database stored at the TOC for specific metrics at a given signal, date, and time of day. UDOT uses the ATSPM system in many different ways, including monitoring arterial networks for signal coordination improvements, special events, and even for maintenance requirements as abnormal detector data in identified.

A sample image adapted from the ATSPM interface is shown in Figure 4.1, where an intersection was queried for traffic volumes over a span of four hours and the figure shows plots with these outcomes for the different directions of traffic of interest. In this particular case, the image shows volumes for northbound and southbound and with a maximum resolution of five minutes.

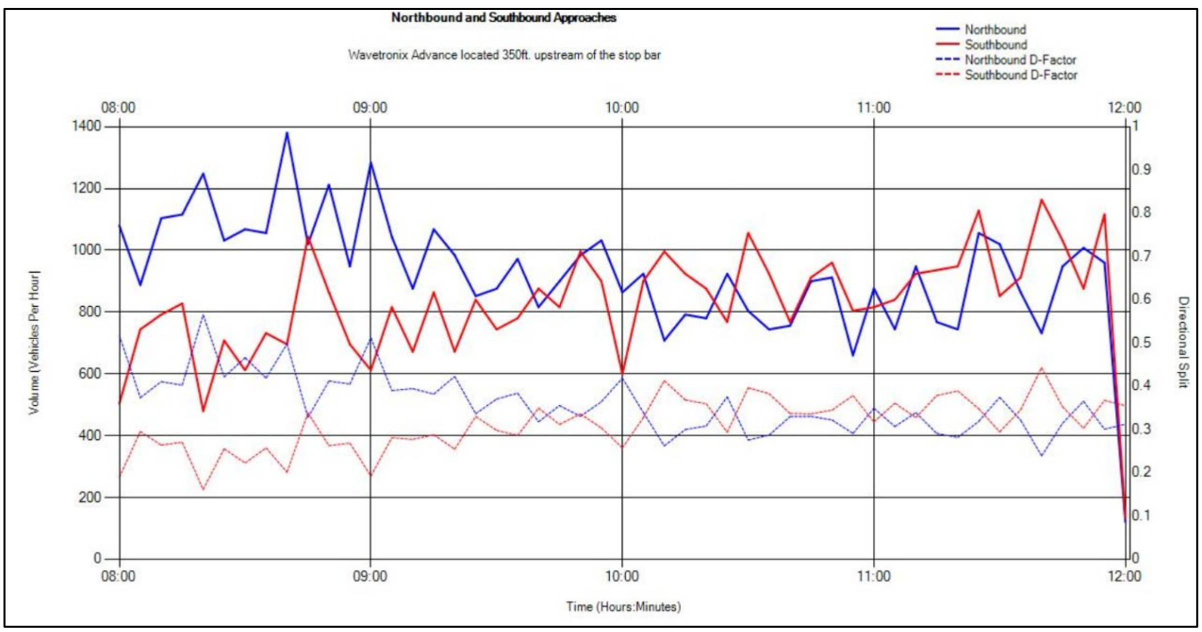


Figure 4-1: Sample Image from ATSPM Interface Showing Volumes for Northbound and Southbound

Similar outcomes are provided for a number of different metrics, including vehicle arrivals in red, delay, turning movement counts, etc. An interesting feature is the visualization of vehicle arrivals with respect to the signal cycle, coined as the Purdue Coordination Diagram, where an engineer can observe areas of interest for improving coordination and overall signal efficiency. Figure 4.2 illustrates an example of the diagram for the same intersection and time period shown in Figure 4.1, in this case displaying the northbound direction of travel. The diagram shows the moment within the cycle that each vehicle activation is received, where the continuous lines show the beginning of each signal indication (color coded for green, yellow, or red). Also, the signal plan is being identified, as well as the arrivals on green (AoG), the percentage of green time (GT), and the platoon ratio (PR).

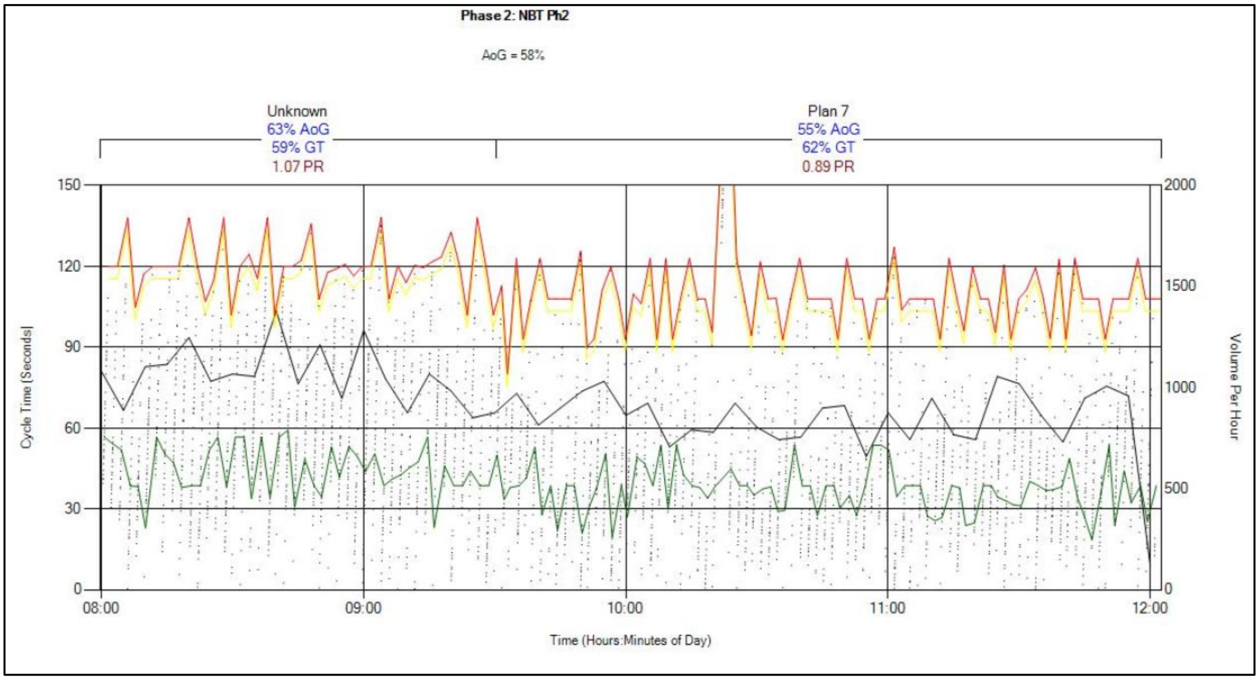


Figure 4-2: Sample Image of the Purdue Coordination Diagram

However, a significant drawback of these interfaces is evident when additional details or data post-processing is needed for custom analysis. For example, the data to produce the Purdue Coordination Diagram or the traffic volumes is not available as a summary in an output that can be consumed. Extracting details from the default plots is not practical and, thus, an analyst is limited to visualizations and manual processing of each individual approach. Furthermore, a combination of adjacent locations for comparison purposes is also difficult at levels finer than standard ATSPM metrics, such as those listed for the Purdue Coordination Diagram.

Fortunately, sometime after UDOT launched the main set of interfaces, it made available a password-protected extension for downloading raw controller data using the Logger enumerations mentioned above. In combination with the documentation to decode the enumerations, as well as the mapping of detector and signal phasing from each intersection, raw datasets could allow an analyst to expand data exploration in any research direction.

It is noted that the main interfaces also provide the mapping of sensors for each of the approaches and the signal phases. For example, for a specific approach, say the northbound approach, the mapping indicates the protected and permitted phases of the approach, as well as the lane-by-lane detector phase and channel allocation as they are produced in the raw data files.

Given these opportunities, the research team developed custom code to make use of the raw ATSPM datasets so analysis of volumes and occupancies could be conducted to study the effects of events in the network. Together with detailed information about events, and in particular crash events, now an analysis of the time-series data is possible. A caveat in this process is related to limitations to extract data, as individual files are limited to 2^{20} rows (a total of 1,048,577 rows), which are filled by outputs from a standard arterial intersection in two or three days' worth of data.

A sample image of one of the custom interfaces is shown in Figure 4.3, where a diagram similar to the Purdue Coordination Diagram is generated along with a representation of approach volumes and vehicle arrivals in green. While the outputs are seemingly similar to those from ATSPM, now the team has access to all details to the data and complete flexibility for analysis purposes. The custom code and relatively simple ad-hoc variations have a wide range of potential applications in traffic operations and safety.

Figure 4-3. Sample Image from Custom Code Showing Signal Timing and Vehicle Activations (top) and Volume and Arrival in Green Trends (bottom)

4.2 CRASH DATASETS

A second piece essential to the analysis in this project is related to crash data and, in particular, to detailed reports beyond typical crash-level information. The team has direct access to complete crash records in the state and, in fact, is responsible for the content management system that produces local and state-wide data and metrics to local and regional agencies. These functions are part of the Utah Transportation and Public Safety – Crash Data Initiative (UTAPS-CDI), and constitute a significant capability in terms of crash and event analysis.

As part of UTAPS-CDI, the team can produce internal queries to verify and cross-validate crash event details not always available to third parties. First, the team verifies the precise location of all crash records in the state, ensuring that coordinates, route numbers, and mileposts correctly represent those of the actual crash event. Location information in the original crash reports is the most commonly reviewed set of items, as a significant percentage of records would need coordinate updates and, most importantly, because the records are not by default associated to a centerline within the state's GIS roadway layers.

Second, the team has access to all coded values related to the vehicle files, including their maneuver, manner of collision, and direction of travel. While these fields are commonly available from the state's databases, the UTAPS-CDI team can verify this information with the help of detailed narratives of events as written by the officers in the complete reports. It is noted that while the team can internally verify such details, narratives are restricted and not available for consumption outside of the UTAPS-CDI repositories.

Lastly, crash diagrams are also available to the team and play an important role when confirming the point of crash in relation to roadway features, help verify the vehicle's travel directions, and point out the resting places of vehicles within the intersection area.

Different from standard crash-related analyses conducted at an intersection level, elements such as the travel direction are key to correctly assign or review potential crashes for analysis. For example, if there is interest in events along a specific corridor because of potential effects on the main road, the travel direction of a left-turning vehicle may indicate the direction before or after the maneuver is completed. The two possible situations result in a very different crash event of the vehicle. Consider the case of a northbound vehicle turning left against opposing through traffic traveling southbound. If the direction of the turning vehicle is coded as westbound (i.e., after the turn is completed), then the crash could be interpreted as involving a vehicle traveling west and turning left towards the south direction.

Thus, overall, the team had a very favorable level of access to detailed data that could be essential to characterize the effects of events in the network.

4.3 DATA INTEGRATION AND PROPOSED FRAMEWORK

Access to both raw datasets from ATSPM and crash data allowed the team to analyze the two data sources together. First, identification of a crash of interest can be done from a custom query builder developed by the team as part of UTAP-CDI. Then, exports will feed scripts coded in R to read the crash information, and process raw datasets from ATSPM to investigate changes in traffic flow in terms of volume rates, arrivals with respect to the signal timing, and occupancy.

Moreover, the team envisions the integration of these tools in a process that allows for a framework to monitor for traffic flow anomalies in the network. Such tools would help operators to understand not only instances of non-recurring events, but also previously unidentified disturbances outside of expected fluctuations.

A general overview of the framework is shown in Figure 2.1, where the data elements described above are integrated to produce a monitoring system along a specific corridor. ATSPM outputs could enable capabilities to monitor traffic through alerts using performance thresholds, but no explicit public-facing mechanism is currently in operation to maintain models or to provide dynamic thresholds for this purpose.

5.0 DATA ANALYSIS

After the identification of datasets necessary to build a framework to analyze non-recurrent events on arterials, the team began an exploration of potential sites for analysis, and experimenting with the steps to post-process the raw ATSPM datasets.

As described above, the team used custom processes to manage the enumerations in the raw datasets. A number of metrics were identified as having potential to flag variations in traffic patterns as a result of a crash event, including traffic volume over a given time window (ranging from five- to 30-minute windows), variations in a measure of speed, and variations in arrivals or in occupancy at the stop bar.

The process starts with the identification of sites with minimum detection requirements for the analysis. Individual movement volumes can be captured at stop bar and advance counting zones, so either one of these channels could be used in the analysis. In our case, preference was given to stop bar counting zones, particularly for through movements along the subject arterial, but both locations are recommended to be used for a more complete picture in cases where queues can reach the detector zones.

In addition, presence zones can be used to estimate proxy metrics for speed and occupancy. Speed can be approximated by the duration of the detection calls during the green indication, and after discarding a few seconds from the start of the phase where vehicles could be experiencing start-up loss time (e.g., queues are being processed at rates lower than a typical saturation headway). In addition, vehicle arrivals could also be investigated by analyzing the distribution of vehicle spacing at the presence zones.

A significant portion of intersections investigated in the ATSPM system in Utah had lane-by-lane stop bar zones (both presence and count zones) as well as count capabilities for the complete approach at the advance locations. It is noted that most vehicle detection along arterials is collected for ATSPM using microwave detection, as opposed to inductive loops or video detection systems.

5.1 SAMPLE LOCATIONS FOR ANALYSIS

The team identified available corridors to explore the data integration. Given that ATSPM operates on a large proportion of the intersections managed by the state, the number of intersections suitable for testing was considerably large. This is also further indication of the significant potential of the proposed monitoring framework.

The team settled on the 700 East corridor along an important arterial in the north-south direction of the central-city area within Salt Lake City. Figure 5.1 shows the selected corridor inside the dashed box of the zoomed-in area, together with other ATSPM locations in the Salt Lake Valley pointed by the blue markers.

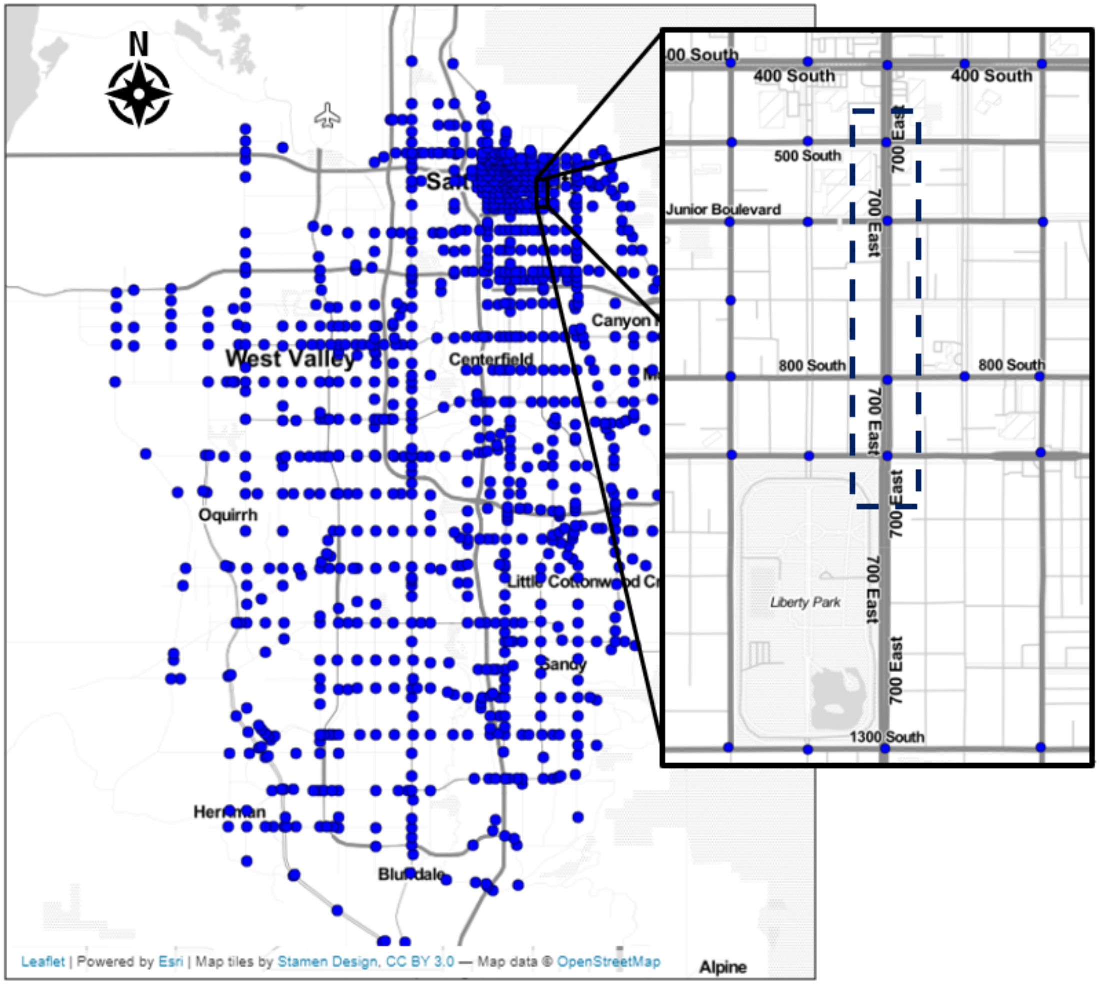


Figure 4-1. ATSPM Locations in the Salt Lake Valley and Focus Corridor (700 E)

Data selection was limited to events between December, 2021 and May, 2022, so the team had multiple crashes to evaluate. Within the selected period, there were a total of 25 crashes within the dashed box in Figure 5.1, where their severity ranged from no injury to suspected minor injury crashes. Among those, one pedestrian crash and one motorcycle-involved crash were observed. Table 5-1 shows a summary of the crash data in the selected sample.

Table 1 Selected Crashes for Analysis on 700 E Corridor

Crash Sequence	Date	Crash Severity	Manner of Collision	Notes
1	12/10/21	Suspected Minor Injury	Angle	Red light running - At least one vehicle disabled
2	12/15/21	No Injury	Rear End	Not at intersection - no vehicles disabled
3	12/25/21	Possible Injury	Angle	Opposing left-turn - both vehicles disabled
4	1/7/22	No Injury	Angle	No narrative
5	1/19/22	Suspected Minor Injury	Angle	Red light running - Both vehicles disabled
6	1/22/22	No Injury	Angle	Opposing left-turn - No indication of disabled vehicles
7	2/11/22	No Injury	Single Vehicle	
8	2/27/22	Possible Injury	Angle	Red light running - Both vehicles disabled
9	3/2/22	No Injury	Angle	Opposing left-turn - at least one vehicle disabled
10	3/2/22	Possible Injury	Angle	Red light running - At least one vehicle disabled
11	3/25/22	No Injury	Angle	Opposing left-turn - both vehicles disabled
12	3/30/22	Suspected Minor Injury	Angle	Opposing left-turn - motorcycle
13	4/7/22	No Injury	Rear End	No narrative
14	4/8/22	No Injury	Single Vehicle	Pedestrian crash - no injury
15	4/12/22	Possible Injury	Sideswipe	No narrative
16	4/18/22	Possible Injury	Rear End	No vehicles disabled
17	4/18/22	Possible Injury	Angle	Opposing left-turn - both vehicles disabled
18	4/22/22	Suspected Minor Injury	Single Vehicle	
19	4/24/22	Suspected Minor Injury	Angle	Red light running - Both vehicles disabled
20	4/28/22	Suspected Minor Injury	Angle	Red light running - Both vehicles disabled
21	5/3/22	Possible Injury	Rear End	One vehicle disabled
22	5/6/22	No Injury	Rear End	No vehicles disabled
23	5/19/22	Possible Injury	Rear End	No narrative
24	5/20/22	Possible Injury	Single Vehicle	No narrative
25	5/23/22	No Injury	Sideswipe	Minor crash with emergency vehicle

5.2 DATA EXTRACTION AND POST-PROCESSING

Direct extraction of datasets from the ATSPM interface is straight forward and only requires identifying the signal controller ID from a master list embedded in the HTML of the platform, or by manually selecting a single location from the provided map.

Extraction of metrics requires prior mapping of the detector channels and decoding of the logged enumerations. For example, event code 82 indicates a “Detector On” status and event code 81 provides the corresponding “Detector Off,” so for the same detector channel a sequence of timestamped activations and deactivations can be processed.

So, by accounting for all activations along through movements within a time frame, the framework can piece together traffic flow patterns at a given location. Moreover, data from multiple days can be combined to generate time-series from such volumes.

Let’s take an intersection part of the selected sample corridor as an example. In particular we discuss traffic flows for the intersection of 700 E and 600 S, which has included in the selected area from Figure 5.1, and is shown in Figure 5.2. At this intersection, 700 E runs in the north-south direction and has a symmetrical cross-section with three through lanes and exclusive left turn and right-turn lanes in both directions. Similarly, the minor street (600 S) has two through lanes in the eastbound, one through lane in the westbound, and both directions have exclusive left and right through lanes.

Figure 5-2. Aerial View of Sample Intersection at 700 E and 600 S

To illustrate the proposed framework, Figure 5.3 shows the northbound 30-minute traffic flow at 700 E and 600 S, where multiple weekdays can be seen together for continuous 24-hour periods. It is noted that standard ATSPM interfaces do not provide capabilities to develop this analysis, so custom post-processing is needed.

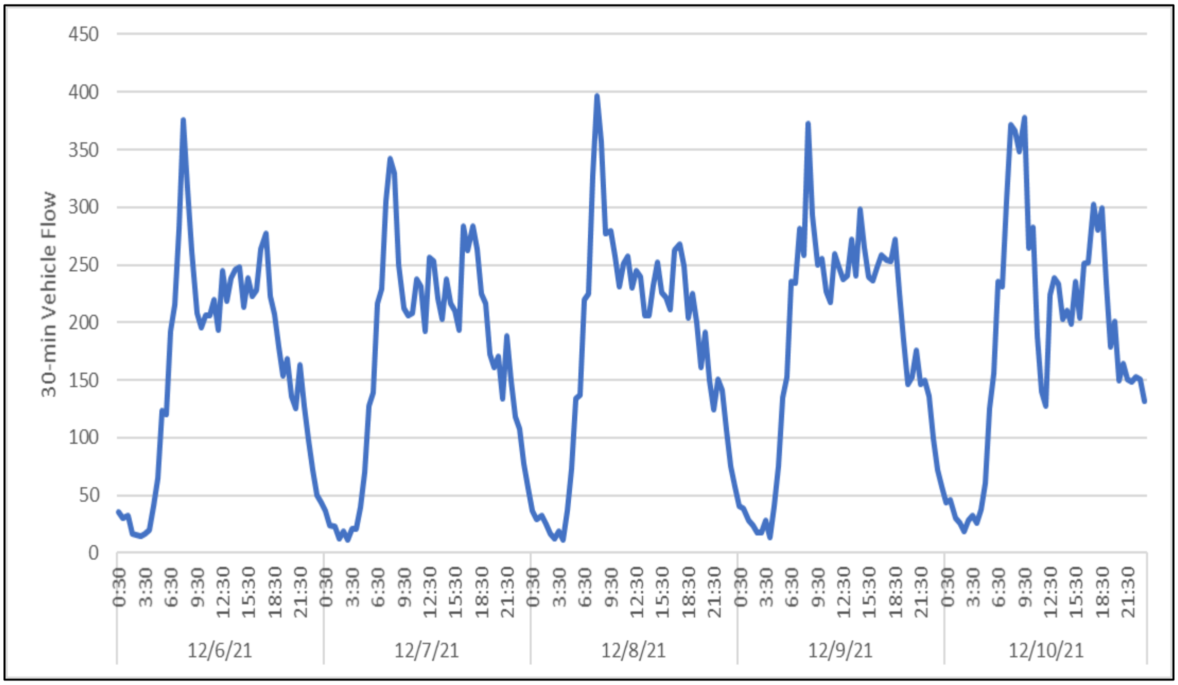


Figure 6-3. 24-Hour Weekday Vehicle Flow Pattern – Northbound of 700e & 600s

Also, note from Figure 5.3 that a potential new pattern is observed the last day of the series at some point towards the end of the morning. This was actually the result of a crash event that was further investigated. Based on details from the crash report, the manner of collision of this crash was classified as an “angle”, and resulted in a disabled vehicle resting in the middle of the intersection with one suspected minor injury.

A different perspective of the same intersection approach is seen in Figure 5.4, where data from the same days are displayed for a time window between 9 a.m. and 1 p.m., covering the period where the event is observed. It is now more apparent that an event occurred on 12/10/21. Two main observations are drawn from the figure. First, traffic flow patterns are highly consistent between weekdays from the same time period, seasonality, and isolated from weather events, as expected. Second, the process is suitable for event identification given an algorithm that can identify a baseline trend and a confidence interval.

In addition, as the team gathered crash events and reviewed complete crash records, including diagrams and narratives from this and other locations among the 700 E corridor, it was apparent that incident signatures (or the effects of cashes) could vary from seemingly large effects along the corridor to no measurable effect in terms of changes in traffic flow.

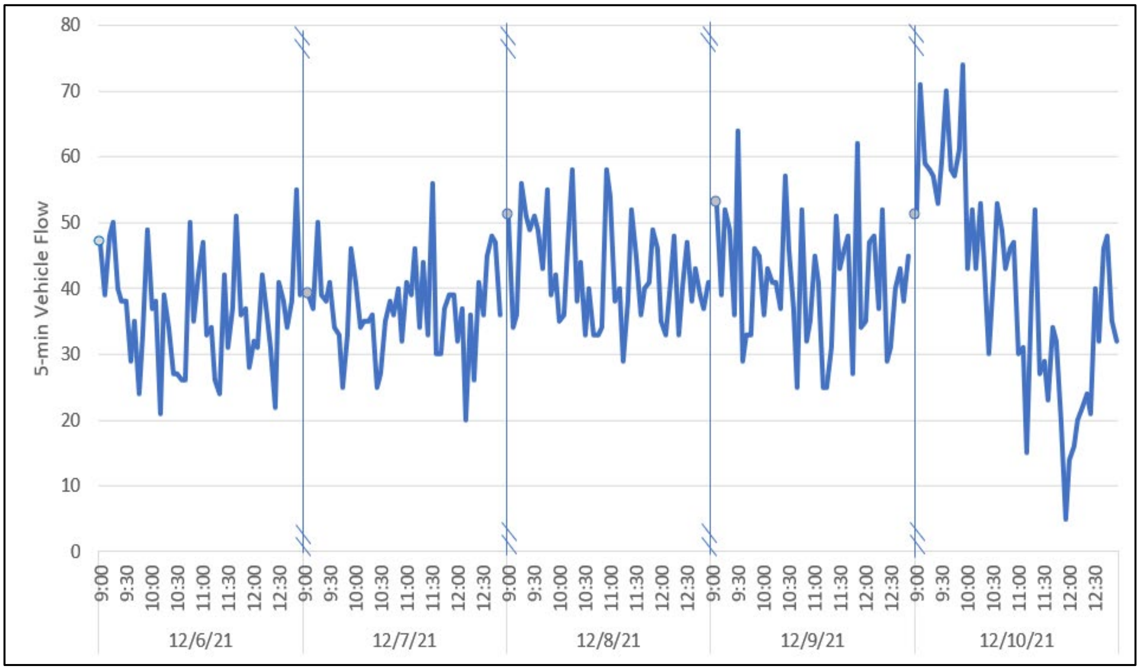


Figure 7-4. 4-hr Time Window Vehicle Flow Pattern – Northbound of 700e & 600s

It is noted that valuable information was extracted in part from the narratives and diagrams contained in the crash reports. For example, Figure 5-5 shows the crash diagram for the crash event on the last day of Figure 5-4, where the narrative also indicated that one of the involved vehicles was disabled and resting in the middle of an intersection.

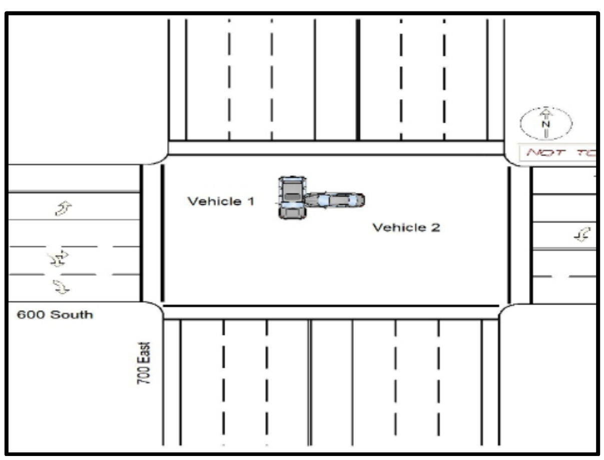


Figure 8-5. Crash Diagram from Sample Crash

Additional metrics were also explored to identify complementary flags that could help identify changes in traffic patterns. Even if traffic flow is being processed at a given rate, perhaps disturbances could be measured in terms of speed or density variations. Preliminary exploration showed that meaningful outcomes could be produced from such metrics and are also subject to the same metrics and time-series analysis described below for vehicle flows.

5.3 TIME-SERIES FORECASTING IN THE PROPOSED FRAMEWORK

A number of approaches can be implemented to forecast or "backcast" time-series data, ranging from traditional parametric approaches including ARIMA methods to simulation, or non-parametric methods using machine-learning. Automation of such processes was considered in the analysis, as well as the flexibility to capture a wide range of daily traffic patterns.

Taking advantage of advances in machine-learning, an application based on Long Short-Term Memory (LSTM) was used to demonstrate a time-series forecast process that can be implemented in the framework and can be operated while online. LSTMs are part of the larger recurrent neural networks field and can be characterized by their ability to learn patterns and sequences of unknown length (Malhotra et al, 2015). In particular for anomaly detection in time-series data, LSTM networks prevent issues with varying or decaying gradients over time by using multiplicative gates that maintain error flows through the states of "memory cells." Significant documentation on LSTM has been produced in the literature, including detailed descriptions of the network formulations (Graves et al., 2013).

LSTMs have recently remerged in the transportation field as efficient methods to forecast or impute traffic demands and calculate crash risk (Zhuo et al., 2017; Abbas et al., 2018; Mackenzie et al., 2018; Yuan et al., 2019; Saroj et al., 2021), particularly for applications with a focus on online monitoring, resource intensive, and big data processing. Therefore, LSTMs hold promise for the development of traffic pattern models that can be updated online as the system receives newly captured data from traffic controllers in the field.

It is noted that the main objective of using LSTMs is to characterize and maintain up-to-date patterns under common conditions using recent historical data, so that a baseline is available for detection and quantification of potential events in the network.

In our discussion, the use of LSTMs can be mainly thought as a tool for traffic demands, but their application extends to other measures of performance discussed above, including speed and occupancy estimates from detector data, as well as expected signal progression at different times of day.

The implementation of LSTM within the proposed framework followed an open source code implementation using the deep learning KERAS library in combination with

TENSORFLOW. Although these libraries were originally python-based, the particular implementation used in this project was based on the R versions of the libraries (Keras and Tensorflow v. 2.9.0).

To explore the implementation of LSTMs in our context, inputs for network training and forecasting were based on historical vehicle flows arriving at a subject intersection. Both five-minute and 30-minute aggregation intervals were explored. Sample sizes from multiple days were used for training, with a minimum of eight days from the nearest neighbors to be selected for model building. For example, if the system is about to scan traffic patterns for anomalies using data for a Thursday, the eight most recent weekdays will be selected or, in the case of delayed estimates, the four weekdays before and four weekdays after said Thursday are used. It is noted that traffic patterns with high variability may require additional training data, but typical recurrent patterns along arterials showed consistent results following this data selection.

Among different parameters in the LSTM formulation, the layer size provides an indication for the model complexity. Larger layer sizes are capable of producing larger weight combinations and more precise patterns, but with a risk of overfitting the training data if the network is much larger than needed. Standard network sizes were deemed appropriate for the application at hand, where a layer size of 50 was enough to replicate observed patterns when traffic volumes were aggregated at five-minute and 30-minute intervals. Figure 5-6 illustrates the progression of the model and its ability to reproduce greater details with an increase in the layer size. However, almost negligible benefits are observed above a layer size of 50.

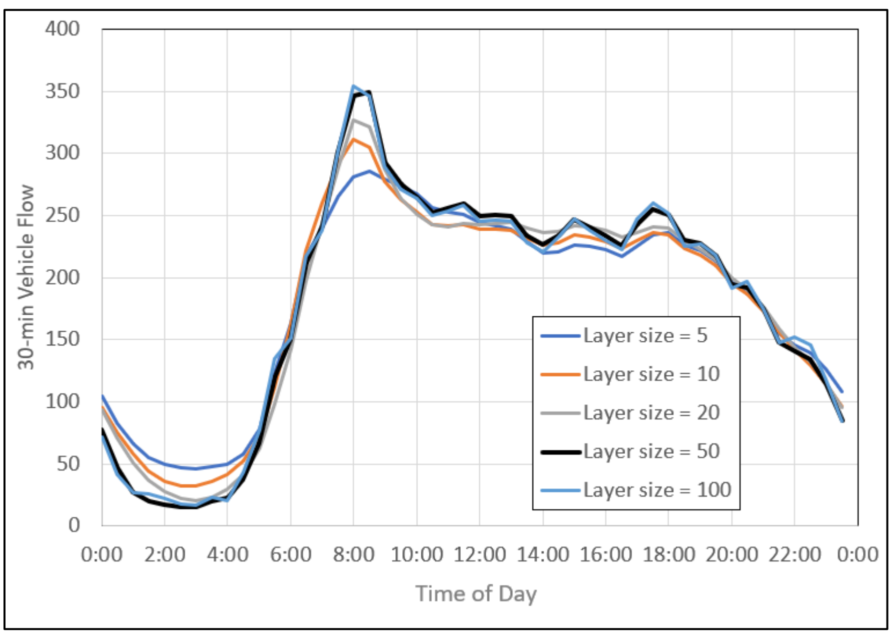


Figure 9-6. Effect of LSTM Layer Size on Reproduction of Detailed Vehicle Flow Patterns

To evaluate the implementation of LSTM for event detection, a k-fold cross validation approach was followed and produced performance measures when comparing model predictions with observed data points. This allowed for a quantification of the expected model accuracy based on historical input data and, therefore, to evaluate newly collected data in light of such expectations.

Figure 5.7 shows three standard metrics of model performance and their relative spread, and a second series indicating the actual evaluation of a data point from a day with a confirmed crash event. The measures of performance are standard practice to model evaluation and included the mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean squared error (RMSE), defined as follows:

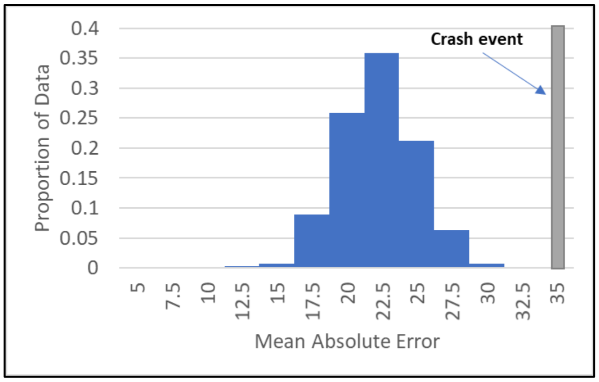
$$MAE = \frac{1}{n}\sum_{1}^{n}|y_i - \hat{y}_i|$$

$$MAPE = \frac{100}{n}\sum_{1}^{n}\frac{y_i - \hat{y}_i}{y_i}$$

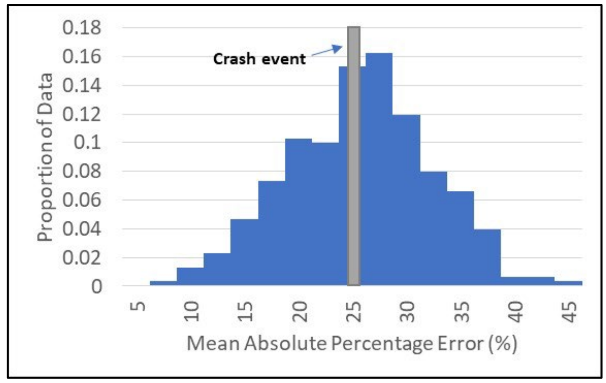
$$RMSE = \sqrt{\frac{\sum_{1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

Where y_i is a value observed in the time-series and $\hat{y}_i$ is an estimated time-series value from the LSTM model.

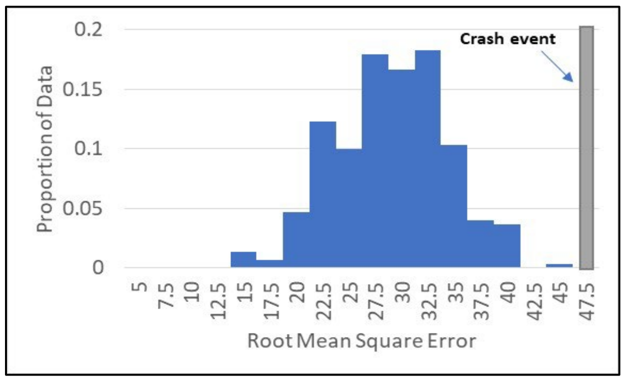
It is noted that not all three measures of performance are needed to flag an anomaly for the detection to be successful. In this particular case, two of our three measures indicate an observation away from the expected performance distribution, pointing to potential rules based in a number of minimum flags needed for a positive event to be communicated by the system.



a) MAE Distribution



b) MAPE Distribution



c) RMSE Distribution

Figure 10-7. Expected Performance of LSTM Modeled Vs Observed Data in Relation to Data Point with a Confirmed Crash Event

A similar approach was also followed to evaluate the distribution of individual points throughout the time-series. This is different from the central tendency metrics illustrated above, and refers to the point locations along the time-series with respect to the expected spread based on the LSTM models and their variance. Figure 5.8 shows an example of the confidence bands from LSTM and a time-series with a confirmed crash event. It is important to highlight that portions of the time-series outside of the confidence bands do not necessarily indicate the occurrence of an event, but the cumulative proportion of such points may be used as an indication instead. The proposed idea is analogous to the analysis performed by the Cramér–von Mises test, but applied in the context of a time-series instead of a cumulative distribution function.

Finally, it is noted that measures of performance are not only expected to serve as incident detection flags, but also as indicators of incident severity. Likewise, deviations found from expected performance can be measures in terms of time, providing a measure of event duration. Both severity and duration of the effects of an event could also be cross-referenced with confirmed data from detailed crash reports, specially during validation of system performance after the physical deployment of the framework.

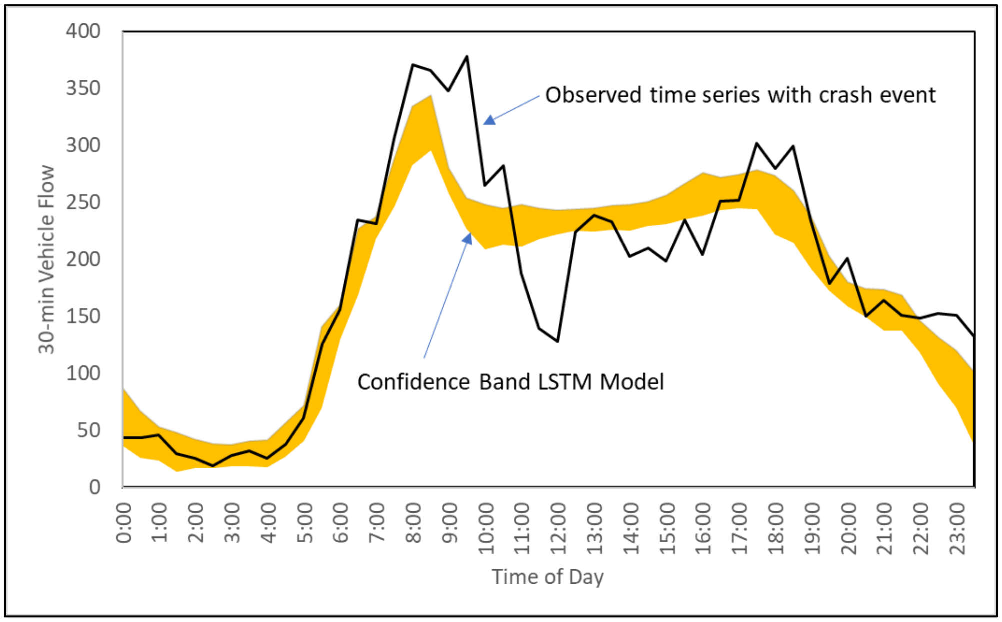


Figure 11-8. LSTM Model Confidence Bands and Time-Series with Crash Event

5.4 PLANNED LONG-TERM DEPLOYMENT AND EVENT CHARACTERIZATION

Using the framework described above for event detection, the long-term deployment of the virtual infrastructure for event monitoring is underway and will be supported through continuous commitments as part of the UTAPS-CDI initiative. It is expected for the system to use internal virtual servers within the University of Utah network but with web-based interfaces for information dissemination through dashboards. The online server application from R-studio, called Shiny, is the preferred method of choice for the interface, and will directly interact with the elements developed as part of this project.

After deployment and a validation phase, expected outcomes include a library of models and recurrent traffic patterns that systematically scan for new datasets in a search for potential events. Near-time crash data from UTAPS-CDI will be used as ground truth for confirmed events, but there is potential to uncover additional event data from underreported events related to short-term lane closures, special cultural and sporting events, as well as unexpected uncategorized events.

Long-term plans for the infrastructure include potential to support a number of additional performance measures, including travel time reliability and crash risk prediction, among others.

6.0 CONCLUSIONS AND FUTURE WORK

Efforts from this project contributed to the development and testing of a scalable process to verify, identify, and quantify effects of event data on traffic flow by integrating high-definition vehicle detector activations/deactivations and traffic signal data, in combination with detailed crash data. The framework is centered on a flexible application for surface streets, using open source code that can incorporate database connections, access to API services, and powerful libraries to model time-series from traffic data using machine-learning.

The team used newly available high-definition datasets from the Automated Traffic Signal Performance Measures (ATSPM) as one of the main assets for the application. ATSPM provides unique opportunities for research and its use is currently limited to unscheduled monitoring of traffic operations, with limited applications in the safety domain. The infrastructure, communications system, and data collection and management behind ATSPM required significant asset acquisitions and needs large operation expenses. Thus, leveraging such resources for extended applications provides significant added value at a very low cost.

The framework deployment within the University of Utah network will also leverage existing resources associated with the Utah Transportation and Public Safety – Crash Data Initiative (UTAPS-CDI), which is a continuing effort between the Utah DOT and Department of Public Safety and, thus, is an ideal set-up in terms of personnel expertise and to ensure project continuation.

Today's automated applications and system response programs based on incident detection are exclusive for freeway systems, and to the authors' knowledge, there are no similar initiatives underway to integrate detailed crash records, high-resolution ATSPM datasets, and efficient algorithms to produce an arterial monitoring system with a strong safety component. Therefore, the framework proposed through this project paves the way for new avenues for research with a strong technological component that can adapt to new data sources and advances in connectivity.

7.0 REFERENCES

Abbas, Zainab, Ahmad Al-Shishtawy, Sarunas Girdzijauskas, and Vladimir Vlassov. "Short-term traffic prediction using long short-term memory neural networks." In 2018 IEEE International Congress on Big Data (BigData Congress), pp. 57-65. IEEE, 2018.

Ahmed, Faisal, and Yaser E. Hawas. "A threshold-based real-time incident detection system for urban traffic networks." *Procedia-Social and Behavioral Sciences* 48 (2012): 1713-1722.

Chassiakos, Athanasios P., and Yorgos J. Stephanedes. "Smoothing algorithms for incident detection." *Transportation Research Record* 1394 (1993): 8-16.

Chen, Z., Liu, X. C., & Zhang, G. (2016). Non-recurrent congestion analysis using data-driven spatiotemporal approach for information construction. *Transportation Research Part C: Emerging Technologies*, *71*, 19-31.

Chung, Younshik. "Development of an accident duration prediction model on the Korean Freeway Systems." *Accident Analysis & Prevention* 42, no. 1 (2010): 282-289.

Dimitriou, Loukas, and Eleni I. Vlahogianni. "Fuzzy modeling of freeway accident duration with rainfall and traffic flow interactions." *Analytic Methods in Accident Research* 5 (2015): 59-71.

Evans, Jonny, Ben Waterson, and Andrew Hamilton. (2020) "Evolution and future of urban road incident detection algorithms." *Journal of transportation engineering, Part A: Systems* 146, no. 6.

Federal Highway Administration. 2020. Website, Automated Traffic Signal Performance Measures. URL: https://ops.fhwa.dot.gov/publications/fhwahop20002/ch2.htm. Accessed May, 2022.

Garib, A., A. E. Radwan, and HJJoTE Al-Deek. "Estimating magnitude and duration of incident delays." *Journal of Transportation Engineering* 123, no. 6 (1997): 459-466.

Ghosh, Bidisha, and Damien P. Smith. "Customization of automatic incident detection algorithms for signalized urban arterials." *Journal of Intelligent Transportation* Systems 18, no. 4 (2014): 426-441.

Graves, Alex, Abdel-rahman Mohamed, and Geoffrey Hinton. "Speech recognition with deep recurrent neural networks." In 2013 IEEE international conference on acoustics, speech and signal processing, pp. 6645-6649. Ieee, 2013.

Hawas, Yaser E., Mohammad Sherif, and Md Didarul Alam. "Optimized multistage fuzzy-based model for incident detection and management on urban streets." *Fuzzy Sets and Systems* 381 (2020): 78-104.

Junhua, Wang, Cong Haozhe, and Qiao Shi. "Estimating freeway incident duration using accelerated failure time modeling." *Safety science* 54 (2013): 43-50.

Khattak, Asad Jan, Joseph L. Schofer, and Mu-han Wang. *A simple procedure for predicting freeway incident duration*. Transportation Research Board, 1994.

Li, Ruimin, Francisco C. Pereira, and Moshe E. Ben-Akiva. "Overview of traffic incident duration analysis and prediction." *European transport research review* 10, no. 2 (2018): 1-13.

Mackenzie, Jonathan, John F. Roddick, and Rocco Zito. "An evaluation of HTM and LSTM for short-term arterial traffic flow prediction." IEEE Transactions on Intelligent Transportation Systems 20, no. 5 (2018): 1847-1857.

Malhotra, Pankaj, Lovekesh Vig, Gautam Shroff, and Puneet Agarwal. "Long short term memory networks for anomaly detection in time series." In Proceedings, vol. 89, pp. 89-94. 2015.

Mihaita, Adriana-Simona, Zheyuan Liu, Chen Cai, and Marian-Andrei Rizoiu. "Arterial incident duration prediction using a bi-level framework of extreme gradient-tree boosting." *arXiv preprint arXiv:1905.12254* (2019).

Nam, Doohee, and Fred Mannering. "An exploratory hazard-based analysis of highway incident duration." *Transportation Research Part A: Policy and Practice* 34, no. 2 (2000): 85-102.

Payne, Howard J., and Samuel C. Tignor. "Freeway incident-detection algorithms based on decision trees with states." *Transportation Research Record* 682 (1978).

Persaud, Bhagwant N., Fred L. Hall, and Lisa M. Hall. "Congestion identification aspects of the McMaster incident detection algorithm." *Transportation Research Record* 1287 (1990).

Raub RA. Occurrence of Secondary Crashes on Urban Arterial Roadways (1997). Transportation Research Record. 1581(1):53-58. doi:10.3141/1581-07

Raub, Richard A., and Joseph L. Schofer. "Managing incidents on urban arterial roadways." *Transportation research record* 1603, no. 1 (1997): 12-19.

Raub, Richard A., and Ronald C. Pfefer. "Vehicular flow past incidents involving lane blockage on urban roads: A preliminary exploration." *Transportation research record* 1634, no. 1 (1998): 86-92.

Sarker, Afrid A., Alireza Naimi, Sabyasachee Mishra, Mihalis M. Golias, and Philip B. Freeze. "Development of a secondary crash identification algorithm and occurrence pattern determination in large scale multi-facility transportation network." *Transportation Research Part C: Emerging Technologies* 60 (2015): 142-160.

Saroj, Abhilasha J., Angshuman Guin, and Michael Hunter. "Deep LSTM recurrent neural networks for arterial traffic volume data imputation." Journal of Big Data Analytics in Transportation 3, no. 2 (2021): 95-108.

Sermons, M. William, and Frank S. Koppelman. "Use of vehicle positioning data for arterial incident detection." *Transportation Research Part C: Emerging Technologies* 4, no. 2 (1996): 87-96.

Sethi, Vaneet, Nikhil Bhandari, Frank S. Koppelman, and Joseph L. Schofer. "Arterial incident detection using fixed detector and probe vehicle data." *Transportation Research Part C: Emerging Technologies* 3, no. 2 (1995): 99-112.

Vlahogianni, Eleni I., and Matthew G. Karlaftis. "Fuzzy-entropy neural network freeway incident duration modeling with single and competing uncertainties." *Computer-Aided Civil and Infrastructure Engineering* 28, no. 6 (2013): 420-433.

Wang, Li-Li, Henry YT Ngan, and Nelson HC Yung. "Automatic incident classification for large-scale traffic data by adaptive boosting SVM." *Information Sciences* 467 (2018): 59-73

Yuan, Jinghui, Mohamed Abdel-Aty, Yaobang Gong, and Qing Cai. "Real-time crash risk prediction using long short-term memory recurrent neural network." Transportation research record 2673, no. 4 (2019): 314-326.

Zhang, Kun, and Michael AP Taylor. "Effective arterial road incident detection: a Bayesian network based algorithm." *Transportation Research Part C: Emerging Technologies* 14, no. 6 (2006): 403-417

Zhuo, Qinzheng, Qianmu Li, Han Yan, and Yong Qi. "Long short-term memory neural network for network traffic prediction." In 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), pp. 1-6. IEEE, 2017.