NDOT Research Report

Report No. 512-13-803

Traffic Prediction and Responses through Data Mining and Data Stream Processing

March 2015

Nevada Department of Transportation

1263 South Stewart Street

Carson City, NV 89712

Disclaimer

This work was sponsored by the Nevada Department of Transportation. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of Nevada at the time of publication. This report does not constitute a standard, specification, or regulation.

Traffic Prediction and Responses through Data Mining and Data Stream Processing Project 14-29

Final Report

Prepared for: Nevada Department of Transportation

Prepared by: Applied Engineering Management Corporation

March 26, 2015

TECHNICAL REPORT DOCUMENTATION PAGE

ACKNOWLEDGEMENTS

Applied Engineering Management Corporation would like to acknowledge the support of the project panel for this study including:

- Ish Garza, Nevada Department of Transportation (NDOT)
- Janelle Thomas, NDOT
- Jae Pullen, NDOT
- Seth Daniels, NDOT
- Ken Chambers, NDOT
- Reed Gibby, University of Nevada, Las Vegas Transit Center University of Nevada, Las Vegas/University of Nevada, Reno
- Brian Hoeft, Regional Transportation Commission of Southern Nevada (RTCSNV)/Freeway & Arterial System of Transportation (FAST)
- Gang Xie, RTCSNV/FAST
- Jim Poston, Regional Transportation Commission of Washoe County
- Chung Tran, Federal Highway Administration Resource Office

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

It is estimated that congestion costs \$121 billion in wasted fuel and lost productivity in urban America. In Las Vegas, the Urban Mobility Report estimated congestion to cost drivers \$931 million in 2011.¹ It is also estimated that the average driver in the Reno-Carson City area spends \$1,698 annually driving on deficient and congested roadways.² The traditional solution to remedy congestion historically has been to add roadway capacity, however this method in most cases is infeasible and cost-prohibitive due to reduced gas tax revenues and constrained budgets. Intelligent Transportation Systems (ITS), a broad term that encompasses the use of advanced communications, data collection, and computational technologies to improve transportation efficiency, are often viewed as cost-effective methods for increasing system efficiency and decreasing congestion.

This research is intended to assess the current ITS infrastructure and capabilities within NDOT and explore the possible products that could be generated and the anticipated benefits that would support a proactive traffic management approach. This research will address several critical NDOT research areas including:

- ITS infrastructure assessment,
- Potential improvements in traffic management and ITS infrastructure,
- Data-driven decision making, and
- Risk-based analysis to address congestion and safety issues.

Currently, the most advanced ITS technology deployment in Nevada lies with the Regional Transportation Commission of Southern Nevada's (RTCSNV) Freeway and Arterial System of Transportation (FAST) as such this location has served as the primary location for analysis for this study. The project focused on four specific tasks to assess the existing ITS deployments to support proactive traffic management:

1. Reviewed and described the NDOT ITS architecture and traffic data

In this step, an analysis of the NDOT ITS infrastructure such as sensor location, sensor technologies, sensor network architecture, and data collected including format and attributes was completed.

2. Reviewed and assessed NDOT ITS architecture and the traffic data generated In this step, an analysis of the NDOT traffic data in correlation with the ITS infrastructure information gathered was performed. This analysis followed a data curation approach in

¹ Schrank, D., Eisele, B., & Lomax, T. (2012). *TTI's 2012 Urban Mobility Report: Powered by INRIX Traffic Data.* Texas A&M Transportation Institute. Retrieved from http://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/mobility-report-2012.pdf

² TRIP. (2013). *Nevada Transportation By The Numbers: Meeting the State's Need for Safe and Efficient Mobility.* TRIP. Retrieved from http://www.tripnet.org/docs/NV_TRIP_by_The_Numbers_Report_Apr_2013.pdf

combination with advanced data visualization to explore the NDOT infrastructure traffic data; its completeness, quality and usability.

3. Demonstrated prediction capability using NDOT data

In this step, it was shown that the NDOT ITS traffic data collected can be augmented by ancillary data, such as weather and city events, can be used to train prediction models, can generate congestion predictions that could potentially be used within an operational context for proactive traffic management.

4. Performed gap analysis and Strength/Weakness/Opportunities/Threats (SWOT) analysis In this step, a comparison between the assessment of the current NDOT ITS infrastructure done in Task 2 and the prediction development and implementation needs uncovered in Task 3 was performed and results showing NDOT ITS'S advantages, its weaknesses, what it could exploit or benefit from implementing traffic prediction methods, and what could cause future challenges in the context of this research was performed.

A review of the FAST Architecture revealed a range of sensors used to collect information on the system performance and status that are connected to the Traffic Management Center (TMC) through a variety of methods including fiber and wireless connections. Task 2 revealed that 24,704,277 records from 242 FAST sensors passed data appraisal and selection process for completeness, quality and usability of data. This represents about 70% of all data produced by FAST traffic sensors in 2013. Of these 24,704,277 valid traffic data records approximately 57% were produced by Wavetronix radar sensors connected to the TMC using serial and fiber connections; 27% of the valid traffic data records were produced by ISS radar sensors connected to the TMC using wireless and fiber connections with the remaining valid traffic data records generated by a range of known and unknown sensors.

Next, a traffic prediction demonstration was conducted for one of the sensors in the FAST system utilizing a year's worth of traffic condition data as well as ancillary data including weather data and event data gathered from the Las Vegas Convention Center. A traffic prediction model was developed that predicted the probability of traffic being in one of the three traffic phases: Free Flow, Synchronized Flow, and Wide Moving Jam, for each 15-minute interval over the course of the year. The demonstration supported the use of traffic data prediction to enable proactive traffic management by allowing TMC operators to understand the expected conditions for each 15-minute interval throughout any particular day of interest. An example application of prediction for proactive traffic management within a TMC was provided including potential mockups of TMC operator displays to allow operators to move prediction windows forward and back depending on their desired window of management. The demonstration also revealed that additional data that could be used to further enhance traffic prediction included: additional years of traffic data (for example 2014 which should now be available for analysis), construction data, incident data, and improvements in sensor reliability to close gaps in traffic data.

AEM CORPORATION 3

INTRODUCTION

Background

It is estimated that congestion costs \$121 billion in wasted fuel and lost productivity in urban America. In Las Vegas, the Urban Mobility Report estimated congestion to cost drivers \$931 million in 2011.³ It is also estimated that the average driver in the Reno-Carson City area spends \$1,698 annually driving on deficient and congested roadways.⁴ The traditional solution to remedy congestion historically has been to add roadway capacity, however this method in most cases is infeasible and cost-prohibitive due to reduced gas tax revenues and constrained budgets. Intelligent Transportation Systems (ITS), a broad term that encompasses the use of advanced communications, data collection, and computational technologies to improve transportation efficiency, are often viewed as cost-effective methods for increasing system efficiency and decreasing congestion.

While ITS technologies have been successful in reducing congestion in specific areas (e.g., traffic incident management and electronic toll collection), the value of traffic management systems has not been fully capitalized to date. For example, traffic management systems generate volumes of data daily that is most often not utilized by highway agencies for applications other than short term, reactive traffic management. There is, however, a potential to utilize this data to better understand system performance and proactively anticipate and reduce system breakdowns and more efficiently manage those breakdowns when they occur. Still, many agencies are overwhelmed by the data that is generated from their systems. In order to fully consume the data and information generated from ITS technologies, advanced computing strategies should be utilized. In addition, historically traffic has been difficult and expensive to monitor, data are usually sparse due to poor equipment performance, and collection is sometimes manual, which increases the time and cost required to process traffic data. A monitoring plan that includes high-frequency data processed quickly can be used to develop traffic prediction tools and responses to unpredictable incidents.

To collect and use high-frequency data, the Nevada Department of Transportation (NDOT) may require an update to its existing monitoring plans, near real-time transmittal of data, and the computing capacity to ingest and analyze disparate sources of data. The benefits of ITS include insight into:

- traffic patterns,
- congestion prediction,
- anomaly detection,

³ Schrank, D., Eisele, B., & Lomax, T. (2012). *TTI's 2012 Urban Mobility Report: Powered by INRIX Traffic Data.* Texas A&M Transportation Institute. Retrieved from http://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/mobility-report-2012.pdf

⁴ TRIP. (2013). *Nevada Transportation By The Numbers: Meeting the State's Need for Safe and Efficient Mobility.* TRIP. Retrieved from http://www.tripnet.org/docs/NV_TRIP_by_The_Numbers_Report_Apr_2013.pdf

- travel time reduction, and
- rapid response to unpredictable events such as accidents.

This research is intended to assess and explore the current ITS infrastructure and capabilities within NDOT and explore the possible products that could be generated and the anticipated benefits. This research will address several critical NDOT research areas including:

- ITS infrastructure assessment,
- potential improvements in traffic management and ITS infrastructure,
- data-driven decision making, and
- risk-based analysis to address congestion and safety issues.

Currently, the most advanced ITS technology deployment in Nevada lies with the Regional Transportation Commission of Southern Nevada's (RTCSNV) Freeway and Arterial System of Transportation (FAST) as such this location has served as the primary location for analysis for this study.

Problem Statement

To address the requirements of real-time transmittal of data, NDOT's current monitoring and data needs were assessed, recommendations regarding the incorporation of data mining and data stream processing into the decision-making process developed, and a computational system to ingest and process high-frequency data for real-time traffic monitoring, prediction, and response needs demonstrated. Below is a list of the steps taken during this project, along with a brief description, to address each identified project objective:

1. Describe the NDOT ITS architecture and traffic data

In this step, an analysis of the NDOT ITS infrastructure such as sensor location, sensor technologies, sensor network architecture, and data collected including format and attributes was completed.

2. Assess NDOT ITS architecture and the traffic data generated

In this step, an analysis of the NDOT traffic data in correlation with the ITS infrastructure information gathered was performed. This analysis followed a data curation approach in combination with advanced data visualization to explore the NDOT infrastructure traffic data; its completeness, quality and usability.

3. Demonstrate prediction capability using NDOT data

In this step, it was shown that the NDOT ITS traffic data collected can be augmented by ancillary data, such as weather and city events, can be used to train prediction models, can generate congestion predictions that could potentially be used within an operational context for proactive traffic management.

4. Perform gap analysis and Strength/Weakness/Opportunities/Threats (SWOT) analysis

In this step, a comparison between the assessment of the current NDOT ITS infrastructure done in step 2 and the prediction development and implementation needs uncovered in step 3 was performed and results showing NDOT ITS'S advantages, its weaknesses, what it could exploit or benefit from implementing traffic prediction methods, and what could cause future challenges in the context of this research was performed.

Organization of Report

In the main body of this report, reviews the findings of the analysis of the NDOT ITS architecture and traffic data collected and generated, the tools and approaches used to assess the NDOT ITS architecture and data generated, the development of a preliminary prediction model, and the gap and SWOT analysis approaches, findings and recommendations. Following the main body of the report, detailed visualizations of the NDOT ITS architecture and traffic data generated during its assessment are presented in a set of appendices including information for every sensor in the FAST network in which data was provided. The report is organized as follows:

- Section 2 NDOT ITS Infrastructure and Traffic Data
- Section 3 Assessment of FAST ITS Architecture
- Section 4 Prediction Demonstration
- Section 5 Gap and SWOT Analysis
- Section 6 Conclusions and Recommendations
- Appendix A Completeness Assessment of 2013 FAST Dataset
- Appendix B Calendar Heatmaps of 2013 FAST Dataset
- Appendix C Hexagonal Binning Speed -Flow Plots of 2013 FAST Dataset

NDOT ITS INFRASTRUCTURE AND TRAFFIC DATA

The study of the NDOT ITS infrastructure originally focused on two particular areas: District 1 and District 2. The RTCSNV, headquartered in Las Vegas, is located within District 1. The RTCSNV is both the transit authority and the transportation-planning agency for Southern Nevada. NDOT District 2 covers northwest Nevada, with headquarters in Sparks. Both districts were of particular interest to NDOT as they handle the visit of many out of state tourists throughout the year travelling on I-15 and I-80. [Figure 1](#page-16-2) shows a map of NDOT's districts.

Figure 1. Map of the Nevada DOT districts

NDOT District 1: RTCSNV-FAST Las Vegas Area

RTCSNV-FAST is an organization that oversees arterial and freeway monitoring and management in the Las Vegas metropolitan area and is designed to monitor and control traffic through the use of traffic sensors, closed-circuit television cameras, traffic signals, ramp meters, dynamic message signs, and land use control signals.⁵ Of particular importance is the interactive dashboard that provides travelers and decision makers with information regarding traffic incidents, traffic speeds, archived data, detector data, wind conditions, and additional information on the system performance.⁶

⁵ *Freeway & Arterial System of Transportation (FAST*). (2015). Retrieved from Regional Transportation Commission of

Southern Nevada: http://www.rtcsnv.com/planning-engineering/freeway-arterial-system-of-transportation-fast/

⁶ *Performance Monitoring & Measurement System*. (n.d.). Retrieved from FAST Dashboard: http://bugatti.nvfast.org/

NDOT District 2: NNROC Reno Area

NDOT's District 2 Road Operations Center manages the freeway infrastructure in the Reno metropolitan area. The Northern Nevada Regional Road Operations Center (NNROC), is the planned, collocated regional traffic management center (TMC). The NNROC will serve as the northern Nevada counterpart to FAST in the Las Vegas area as reflected in the Northern Nevada Regional ITS Architecture. Once in place, the NNROC will duplicate all connection between the FAST TMC and NDOT's District 2 Road Operations Center. District 2 possess a similar ITS data system as RTCSNV- FAST and has implemented a similar dashboard capability as FAST. This original project scope included analysis of District 2's traffic data, however, only one month of historical traffic data for I-80 was available for analysis. Unfortunately, one month of historical data is not sufficient to perform a relevant and comprehensive analysis of ITS infrastructure nor to support traffic prediction which requires a minimum of one year's worth of traffic data to provide a reasonable prediction of traffic conditions. As a result, the study focused its efforts on the data available through Las Vegas FAST for the analysis.

ASSESSMENT OF FAST ITS ARCHITECTURE

FAST is one of the first truly integrated ITS organizations in the country. Since 2004, FAST has been administered by the RTCSNV and NDOT. Both the RTCSNV and NDOT fund and contribute to FAST's operations and management.

The RTCSNV staff is responsible for two major areas that make up the system: the Arterial Management Section and the Freeway Management Section. FAST is designed to both monitor and manage traffic to optimize performance in the Las Vegas metropolitan area. Traffic control requires detection of traffic conditions through the use of various sensors including video image detection and inductive loop detection. Visual verification of traffic conditions is also possible through closed-circuit television cameras that are fed to the TMC and posted to the public website.

Below is a list of the ITS equipment currently being used by FAST:

- Radar-based traffic flow sensors
- Inductive loop traffic flow sensors
- Closed-circuit television cameras (CCTV)
- Portable detectors
- CMS boards
- Road weather information system (RWIS) stations
- Variable Speed Limit (VSL)
- Ramp meters

FAST Data Collection Architecture

Traffic data is collected through several different types of sensors using different types of communication networks. Two types of sensors are currently in use: Image Sensing System (ISS) radar sensors and Wavetronix radar sensors in combination with wired and wireless network connections. [Figure 2](#page-19-0) shows a diagram representing the manner in which each sensor type is connected to the TMC. Each sensor connects to the main controller using either a serial or a wireless Ethernet connection. Ultimately all sensors end up merging into a main controller that is connected to the TMC using a fiber connection. Prior to this, each sensor connects to the main controller using either a serial or a wireless Ethernet connection. Three sensor/connection combinations are in place, which are represented by the color of each of the sensors in [Figure 2.](#page-19-0) These combinations are:

Figure 2. ITS Sensor Connection Diagram

In addition to these sensors, several locations in Las Vegas are also utilizing loop detectors to gather traffic data. This type of detector is not considered by FAST as a future viable option for traffic monitoring due to its short lifetime and its costly repair needs, however many of these sensors still exist in the system. Connection setup information for these sensors are not clearly defined, consequently both these sensors and their connection to the TMC were classified as "unknown" within the analysis.

A map showing the distribution of each type of sensor across the FAST roadway network is shown i[n Figure 3.](#page-20-1) The green lines represent road segments equipped with Wavetronix radar sensors using serial and fiber connections back to the TMC. The red lines represent the road segments equipped with Wavetronix radar sensors using wireless and fiber connections back to the TMC. The blue lines represent the road segments with Image Sensing System (ISS) radar sensors using wireless and fiber connections back to the TMC. The unknown sensors with unknown connections are not geocoded and could not be mapped.

Figure 3. Distribution of FAST Sensor Type

FAST Current Use of ITS Data

FAST makes use of its traffic data primarily through its advanced dashboard.⁷ The FAST dashboard is a web application that monitors and presents the metropolitan Las Vegas freeway's traffic condition both in real time and historically, through maps, daily peak speeds, time of day speeds, average speed, and congestion plots from different perspectives. The FAST dashboard is deployed on a traditional, on premise web server and database server hosted at the RTCSNV. Below are screenshots of the FAST

⁷ *Performance Monitoring & Measurement System*. (n.d.). Retrieved from FAST Dashboard: http://bugatti.nvfast.org/

dashboard. This project leveraged, as much as possible, the accomplishments of the RTCSNV-FAST dashboard and relied on its organization and classification of the FAST data to perform the analyses. [Figure 4](#page-21-0) shows an example of the FAST mapping and performance visualization capability.

Figure 4. Screenshot of the FAST ITS Dashboard⁸

⁸ *Performance Monitoring & Measurement System*. (n.d.). Retrieved from FAST Dashboard: http://bugatti.nvfast.org/

Figure 5. FAST Dashboard Use at TMC⁹

[Figure 5](#page-22-1) shows how the FAST dashboard is currently implemented as part of the RTCSNV-FAST TMC to monitor and manage traffic in real-time. The FAST dashboard software has been migrated and deployed at NDOT's District 2 TMC in Reno to provide similar data and video analysis capability.

FAST Data Description

In order to perform a complete analysis of the FAST ITS infrastructure, in addition to its physical architecture details, the AEM team needed to acquire speed, volume, flow, incident, and construction data generated by FAST to characterize the system behaviors and performance. The FAST dashboard team provided data, extracted from the FAST dashboard database archive, for the entirety of 2013.

Records within the archive contained the data of 499 sensors aggregated at a 15-minute interval. [Table 1](#page-23-3) provides a breakdown of the number of devices in operation and records by sensor type.

⁹ *Performance Monitoring & Measurement System*. (n.d.). Retrieved from FAST Dashboard: http://bugatti.nvfast.org/

	Number of	Number of
Sensor Type	Devices on Road	Records
Wavetronix Radar Sensor	208	19,401,481
ISS Radar Sensor	121	9,787,761
Unknown	170	5,975,263
Total	499	35,164,505

Table 1. Number of Devices and Records by Sensor

Traffic Data

Traffic data was obtained from the FAST database archives and contained speed, volume, flow and additional sensor statistics for all highway segments currently monitored by FAST for the entire year of 2013. [Table 2](#page-24-1) describes the traffic data fields and data types obtained from FAST. This project focused on traffic information representative of the entire roadway and did not drill down to the lane level. The date and time information from the Date Time Stamp, DayOfWeek, Date Value, HourIdx, Holiday fields; traffic conditions information from the Volume, Occupancy and Speed fields; infrastructure information from the Detector ID, Roadway ID, Segment ID, Device ID; and quality information from the Poll Count, Failure, Polling Period, Invalid fields were all extracted.

Construction/Work Zone Data

The 2013 construction and lane closure data for each of the road segments that FAST oversees were requested. The construction information is managed by NDOT, independent of RTCSNV-FAST, and is scattered and organized across multiple databases. Construction data was therefore not available in an exploitable format at the time of this project. If construction data had been available, it would have been used in the context of prediction and could have been a valuable input as it is a reliable predictor of slow traffic in the vicinity of the construction zone.

Incident Data

FAST is equipped with an advanced traffic incident management (TIM) system and records and analyzes traffic data and video. Traffic incident data are most often considered to be the results of a prediction (outcome), such as the risk of incident at a specific location and time rather than an input (feature) to a prediction model, as they are occurring somewhat randomly. Incidents can also be used as a predictor in the context of a short time prediction. As a predictor, incidents become irrelevant when dealing with long term prediction as it is impossible to pinpoint where they will occur. The focus of this project was on traffic conditions and it was decided to omit incident data as part of this work. In parallel with this project, AEM is currently working on another project with FAST, focused on TIM, and is leveraging the rich incident data features offered by FAST for this current study.

Table 2. FAST Traffic Data Representation

Integer – Numeric Value

Datetime – Calendar date and time of day

Character – Character string with a maximum length of 255 characters

Assessment Approach

In order to assess the FAST ITS infrastructure, a process similar to that used commonly when curating data was followed. Data curation is described as the activities required to maintain data long-term such that it is available for reuse and preservation. The process follows the steps described below, the approach within this project mainly focused on validating speed and volume data. An assessment of occupancy was not conducted within this process.

Data Acquisition Planning

In this step, the data processing and storage capabilities needed to store, review and analyze the data, were determined. Amazon Web Services and Google Cloud Services were used, greatly reducing the cost of processing and analyzing the FAST dataset.

Data Reception and Metadata Acquisition

In this step, the raw data provided by FAST was collected and stored. The raw data was also assigned additional metadata required to complete the dataset for analysis. In addition, the geolocation of FAST sensors as well as a hierarchical representation of each sensor network connectivity was added to the raw dataset.

Data Appraisal and Selection

During this step, the raw FAST data was appraised and assessed for its ability to support traffic prediction based on a selection policy composed of the following three steps:

1. Completeness

The completeness step consisted of assessing the quantity and distribution of data produced for every month of the year by all 499 sensors. The sensors that did not produce a regular amount of data across the year were identified. Both valid and invalid data points were included in this step.

2. Quality

The quality step consisted of assessing the quality of the remaining data from step 1 following the recommendation of the FAST dashboard team. Every record for every sensor was checked to verify that out of the 15 possible reads in each of the 15-minute interval, at least 10 were successful. That is to say that each of the records satisfying the quality requirement is the result of a 66% successful sensor pull rate.

3. Usability

It was observed during step 1 and step 2 that some records in the raw data set were displaying unrealistic volumes or speeds. For example, some records showed an average speed above 100 mile per hour (mph) over a 15-minute interval or an average number of vehicles per hour (vph) exceeding 40,000 over a 15-minute interval. It was also noticed that some records listed an average volume of 0 vph and an average speed greater than 0 mph or an average volume greater than 0 vph and an average speed of 0 mph.

In the usability step, a maximum limit was defined for the average speed and volume. A realistic speed limit of 90 mph for average speed and 13,200 vph for average volume was set. Records exceeding these limits were omitted along with records from sensors where more than 20% of the records showed a similar inconsistency.

At the end of each of these three steps, a summary table, custom map, and treemap¹⁰ visualization were generated to display the impact of each assessment step on the 499 sensors of the FAST network. The remaining FAST sensors data was then imported into a NoSQL database in preparation for the prediction demonstration.

Tools

The assessment process described above was completed using only Open Source software and cloud services from Amazon Web Services and Google Cloud Services without relying on any in-house hardware infrastructure. Below is a list of the cloud services and Open Source software utilized in the analysis.

Amazon Web Service (AWS)

Amazon Simple Storage Service (S3)

Amazon S3 is an online file storage cloud service offered by Amazon Web Services. Amazon S3 provides storage through web services interfaces and was the first publicly available web service in the United States. It was used to store the FAST raw sensor data.

Amazon Relational Database Service (RDS) for MySQL

Amazon RDS is a distributed relational database cloud service. It provides a relational database for use in applications. It is aimed at simplifying the setup, operation, and scaling of a relational database. It was used to explore and integrate metadata to the un-curated FAST sensor dataset.

Google Cloud Services

Google Cloud storage

Google Cloud Storage is a cloud service for storing and accessing data on the Google Cloud infrastructure. The service combines the performance and scalability of Google's Cloud with advanced security and sharing capabilities. It is comparable to the Amazon S3 online storage service and was used to store the curated FAST sensor dataset.

¹⁰ *Treemapping*. (2015, February 26). Retrieved from Wikipedia: The Free Encyclopedia: http://en.wikipedia.org/wiki/Treemapping

Google BigQuery

BigQuery is a cloud service that enables interactive analysis of massively large datasets working in conjunction with Google Storage. It supports the SQL language and was used to perform large scale data queries necessary to prepare and train the FAST prediction model demonstration.

Open Source Software

R Programing Language

R is a programming language and software environment for statistical computing and graphics. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. R was used principally as a client to a relational database to perform statistical operations, generate visualizations as well as develop and test prediction models.

Visualization

In order to represent the results of each step of the analysis and see the effects of each step on the entire FAST infrastructure several custom visualizations, capable of displaying millions of records in a format manageable by the human eye, were designed.

Maps

Maps were used not only to overlap the quality and performance of FAST sensors over the road network and visualize where traffic sensor failures and successes occurred in 2013, but also to try to uncover potential patterns or any additional information that could help explain why some sensors performed poorly. Maps were generated using R and the Google map web service. [Figure 6](#page-28-1) shows the initial map of the FAST sensor network; each blue dot on the map represents a geocoded FAST sensor. Only 331 FAST sensors were geocoded out of the 499 FAST sensors given for analysis, therefore only 331 sensors are displayed on the map.

Figure 6. FAST Sensor Initial Map

Treemaps

Within the field of information visualization and computing, tree mapping is a method for displaying hierarchical (tree-structured) data as a set of nested rectangles. Each branch of the tree is given a rectangle, which is then tiled with smaller rectangles representing sub-branches of the tree structure. A leaf node's rectangle has an area proportional to a specified dimension on the data. Often the leaf nodes are colored to show a separate dimension of the data. When the color and size dimensions are correlated in some way with the tree structure, one can often easily see patterns that would be difficult to spot in other ways.

Treemaps are also advantageous in the fact that, by construction, they make efficient use of space. As a result, they can legibly display thousands of items on the screen simultaneously. [Figure 7](#page-29-0) shows an example of a treemap displaying the world population and gross national income broken down by a hierarchy of continent and nations. The size of the square is proportional to the size of the nation's population and the color of each square is representative of the nation's gross national income.

Figure 7. Treemap Example of the World's Population and Gross National Income

A similar visualization paradigm was applied to the FAST dataset and its data collection infrastructure hierarchy (shown in [Figure 8\)](#page-30-0). The main rectangle was decomposed into smaller rectangles, each representing the primary network connection (fiber in blue, unknown in black), then subdivided into even smaller rectangles, each representing the secondary network connection (Ethernet wireless and serial in grey, unknown in black), and finally filled with even smaller rectangles, each representing a single sensor (green rectangles). In this case the size of the green rectangle represents the number of records collected during 2013.

A color was then applied to each of the sensor rectangles based on the performance for each of the three data appraisals and selection steps described previously.

Figure 8. FAST Treemap Description

[Figure 9](#page-31-0) is the initial representation of the FAST treemap, in which each of the green rectangle tile color will change after each step of the data appraisal and selection process, which displays the performance and quality of each specific FAST sensor and its data.

0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 Online

Figure 9. Initial FAST Treemap

Assessment Results

In this section, using tables, maps, and treemap visualizations, the results of each step of the data appraisal and selection process are summarized.

Step 1 Results: FAST Traffic Data Completeness Evaluation

In this step, all 499 FAST sensors were assessed for completeness using the criteria expressed below:

- Generating data for all 12 consecutive months of the year 2013
- Producing good or bad data
	- o Good data is defined as data for every month with less than 20% of that month being flagged by FAST as invalid
	- o Bad data is defined as everything else

It should be noted that construction data could have been leveraged here to identify if the lack of data in a few consecutive months was due to sensor defects or the results of construction. Figures 10-13 contain four examples of FAST sensor behaviors observed during the completeness assessment step.

Figure 10. FAST Sensor 388.1.245 for 2013

[Figure 10](#page-32-2) shows that sensor 388.1.245 had a consistent data count level between 10,000 and 12,000 data points across 2013 with less than 20% of bad data each month (in red). This is the desired behavior for each FAST sensor in order to obtain data complete enough to be used in prediction.

Figure 11. FAST Sensor 352.1.318 for 2013

[Figure 11](#page-33-0) shows that sensor 352.1.318 had a consistent data count level between 7,000 and 8,000 data points for the first five months of the year with very little error, less than 20% of the data each month is bad (in red), however the sensor stops producing data abruptly in June until the end 2013. This anomaly can be the result of a failure of the sensor, network, or even the beginning of a construction project. This sensor behavior unfortunately does not meet the requirements for prediction despite its relatively high amount of good data as during the year the sensor did not report data for a significant portion of the year.

Figure 12. FAST Sensor 179.2.275 for 2013

[Figure 12](#page-34-0) shows that sensor 179.2.275 had an inconsistent data count level ranging from 0 to 1,500 data points across 2013 with no apparent pattern. The bad data in red is sometimes found to be close to 20%. While this pattern across 2013 may somehow be explained, it is doubtful that a FAST sensor would see no traffic for two months unless construction is involved. Data from these sensors were flagged as incomplete and therefore unusable for prediction.

Figure 13. FAST Sensor 126.2.80 for 2013

[Figure 13](#page-35-0) shows that sensor 126.2.80 had small and inconsistent data count levels for a few months across 2013, as well as 100% bad data in other months. This sensor behavior is clearly showing that the data cannot be considered for any analysis or prediction of the behavior of the roadway at this location.

A report containing a completeness histogram such as the ones discussed previously for all 499 FAST sensors across the year 2013 has been generated for FAST to better understand the behavior of the sensor that were rejected during this step (See Appendix A). In addition, [Figure 14](#page-36-0) shows the completeness scoring for each of the 499 FAST sensors on the initial treemap visualization. The performance of each sensor on the treemap is expressed using a gradient of color ranging from dark green to yellow. The sensors that provided more than 80% of good data consistently across 2013 are shown in dark green whereas the sensors that provided more than 20% of bad data and were inconsistent across 2013 are shown in colors ranging from pale green to yellow, depending on their severity of their lack of completeness. It is interesting to note that the majority of the completeness failures are distributed between the ISS radar sensors/wireless connection zone and the unknown sensors/unknown network connection zone of the treemap.

[Figure 15](#page-37-0) shows a map of the 331 geocoded FAST sensors, each dot represents a sensor. The color of each dot indicates its completeness score. FAST sensors that provided more than 80% of good
data consistently across 2013 are shown as blue dots whereas FAST sensor that failed the completeness step are shown in colors ranging from purple to red. It is also interesting to examine the location of the sensors failing completeness. They are mainly distributed along I-15 outside of Las Vegas and at the main intersection in Las Vegas at the crossings between I-15, I-215, and I-95.

NumberOfMonth

Figure 14. FAST Sensor Treemap after Completeness Assessment

Figure 15. FAST Sensor Map after Completeness Assessment

Completeness Assessment Outcome

Of the 35,164,505 records (499 sensors) provided by FAST, 7,280,003 records (or 211 sensors) failed to pass the completeness assessment and were removed from further analysis. 27,884,502 records remained to be assessed by the subsequent steps. [Table 3](#page-37-0) shows a breakdown of the completeness failure per sensor by network connection type. It appears that sensors using wireless network connections are more likely to fail completeness than sensors using serial network connections. Also notable is the fact that close to 50% of the unknown sensors with unknown network connections failed the completeness assessment.

Table 3. FAST Sensor Completeness Assessment Results

Step 2 Results: FAST Traffic Data Quality Evaluation

In this step, the remaining 288 FAST sensors were assessed for quality using the criteria expressed below:

 FAST sensors data is pulled every minute for 15-minute intervals then averaged and recorded. For the average to be reliable, there needs to be a limited amount of error during the 15 data pulls made. Based on FAST guidance, records can be trusted if they have 10 or more successful pulls out of 15 attempts, or 66% successful. Records with less than 66% successful pulls are not deemed viable.

[Figure 16](#page-39-0) and [Figure 17,](#page-40-0) the map and treemap of the remaining FAST sensor after the completeness assessment, are used to display the quality of the remaining FAST sensors. A gradient of color, similar to the ones used in step 1, are used for both the treemap and map in this step. Note that the sensors that failed the completeness step have been removed from the map and set to pale yellow on the treemap.

Figure 16. Map of FAST Sensors that Passed Completeness Assessment

Figure 17. Treemap of FAST Sensors that Passed Completeness Assessment

[Figure 18](#page-41-0) shows the quality scores of each of the 288 remaining FAST sensors on the initial treemap visualization. The quality of each sensor on the treemap is expressed using a gradient of color ranging from dark green to yellow. A dark green shows a 100% successful sensor pull for each 15 minute interval and a pale yellow shows a 0% successful sensor pull for each 15-minute interval. The cutoff used for quality is 66%, which corresponds to a light medium green. Only 24 sensors did not pass the 66% quality threshold. These 24 sensors are mainly located in the unknown sensor/network connection and ISS radar sensor/wireless network connection zones of the treemap. It should also be noted that some of the failing sensors do not appear on the treemap. This is due to the amount of data produced by theses sensors is too small to be perceived at the bottom right corner of each rectangle zone on the treemap.

[Figure 19](#page-42-0) shows a map of the 264 remaining geocoded FAST sensors, each dot represents a sensor. The color of each dot indicates its quality score. FAST sensors that provided more than a 66% successful sensor pull, consistently across 2013, are shown as blue dots whereas FAST sensors that showed less than a 66% successful sensor pull are shown in colors ranging from purple to red. It was found that no apparent pattern for quality failure is appearing on the map. Sensors seem to fail quality assessment across the entire network with various rates of failure.

Figure 18. FAST Sensor Treemap after Quality Assessment

Figure 19. FAST Sensor Map after Quality Assessment

Quality Assessment Outcome

From the 27,884,502 remaining records that passed the completeness assessment, 1,341,366 records (or 24 sensors) failed to pass the quality assessment and were removed from further analysis. 26,543,136 records remained to be assessed by the subsequent step. [Table 4](#page-43-0) shows a breakdown of the quality failure per sensor/network connection type. [Table 4](#page-43-0) clearly shows that quality failure is occurring mainly with ISS radar sensors with wireless connections and unknown sensors with unknown network connection. Except for a 20% drop when using wireless, Wavetronix radar sensors in combination with both serial and wireless network connection are not showing many quality issues.

Table 4. FAST Sensor Quality Assessment Results

Step 3 Results: FAST Traffic Data Usability Evaluation

In this step, the 264 FAST sensors remaining from the quality step were assessed for usability. Usability in this project was defined as the ability for the FAST traffic data to be used as an input to a traffic prediction model. More precisely, can the traffic data from "good" FAST sensors be used in calculations to generate reliable future predictions? After further review of the FAST traffic sensor data for completeness and quality, it was observed, while sampling and exploring the FAST dataset, that some records were showing abnormal data. Some records were showing a speed greater than 0 mph and a volume of 0 vph, or a speed of 0 mph and a volume greater than 0 vph. While the two values passed the previous two assessments, they cannot happen in reality, as there cannot be a nonzero average speed with no volume or no speed reported with a non-zero volume. Training traffic prediction models using such data would result in the prediction model learning and leveraging unrealistic traffic behaviors to make future predictions, therefore diminishing its accuracy. The presence of such unrealistic records needs to be limited to guarantee a maximum accuracy when developing traffic prediction models. Based on the review of prediction algorithms tolerance for noisy data, the team decided that a maximum of 20% of the records generated by each sensors could be tolerated without affecting accuracy.

Following the previous findings, the following conditions were used to filter the remaining sensors:

- Remove records showing Speed > 0 and Volume = 0 or Speed = 0 and Volume > 0
- Select FAST sensors where only 20% or less of the records have speed/flow discrepancies

[Figure 20](#page-44-0) and [Figure 21](#page-45-0) contain the map and treemap of the remaining FAST sensors after quality assessment and are used to display the usability of the remaining FAST sensors. A similar gradient of color for both the treemap and map was used for this step. Note that the sensors that failed the quality step were removed from the map and set to a pale yellow on the treemap.

Figure 20. Map of FAST Sensors that Passed Quality Assessment

Figure 21. Treemap of FAST Sensors that Passed Quality Assessment

[Figure 22](#page-46-0) shows the usability scores of each of the 264 remaining FAST sensors on the initial treemap visualization. The usability of each sensor on the treemap is expressed using a gradient of color ranging from dark green to yellow. A dark green shows that 100% of the records had no speed/flow discrepancies and a pale yellow shows that all records had speed/flow discrepancies. The cutoff used for usability is a maximum of 20% speed/flow discrepancies for each sensor, which corresponds to 80% of sensor data without speed/volume discrepancies. This threshold is represented on the treemap as a medium green. Unfortunately 22 sensors did not pass the 80% usability threshold and are not easily visible on the treemap. While these sensors data are consistent and of good quality, they have very few records in comparison with the rest of the sensors, which positions them to the bottom right of the medium rectangles on the treemap where the smallest tiles are positioned (see red circles) making them impossible to see.

[Figure 23](#page-47-0) shows a map of the 264 remaining geocoded FAST sensors with each dot representing a sensor. The color of each dot indicates its usability score. FAST sensors that provided more than 80%

of usable speed and volume data across 2013 are shown as blue dots whereas the FAST sensors that were found to have speed/volume discrepancies are shown in colors ranging from purple to red. It can be seen that sensors with speed/volume discrepancies are occurring almost everywhere on the FAST road network, but few sensors exceed the 20% discrepancy limit.

Figure 22. FAST Sensor Treemap after Usability Assessment

Figure 23. FAST Sensor Map after Usability Assessment

Usability Assessment Outcome

Of the 26,543,136 remaining records that passed the quality assessment, 1,838,859 records (or 22 sensors) failed to pass the usability assessment and were removed from analysis leaving 24,704,277 records (or 242 sensors) that passed the usability assessment step. [Table 5](#page-48-0) shows a breakdown of the usability failure per sensor/network connection type. [Table 5](#page-48-0) clearly shows that usability failures are occurring across all sensor/network connection combinations with the failure rate being about 30% except for the Wavetronix radar sensors with wired connections which are showing a 15% drop in record count and the unknown/unknown sensor/network combinations which are showing a 70% drop.

Primary	Secondary	Device	Total	Records		Drop Percent Remaining
Network	Network		Records	Passed	Passed	Sensors
Fiber		Serial/Ethernet Wavetronix radar sensor 16,407,332 14,062,011 14.29% 85.71%				127
Fiber	Wireless	ISS radar sensor	9,787,761	6,679,735 31.75% 68.25%		66
Fiber	Wireless	Wavetronix radar sensor	2,994,149	2,126,525 28.98% 71.02%		16
Unknown	Unknown	Unknown	5,975,263	1,836,006 69.27% 30.73%		33
		Total	35,164,505 24,704,277			242

Table 5. FAST Sensor Usability Assessment Results

FAST Traffic Data Assessment Summary

[Table 5](#page-48-0) showed that 24,704,277 records from 242 FAST sensors passed data appraisal and selection process for completeness, quality and usability of data. This represents about 70% of all data produced by FAST traffic sensors in 2013. These 24,704,277 traffic data records are coming from the sensors displayed i[n Figure 24](#page-49-0) with more than half, about 57%, of these records come from the Wavetronix radar sensors connected to the TMC using serial and fiber connections. The ISS radar sensors connected to the TMC using wireless and fiber connections represent only 27% of the final dataset. The Wavetronix radar sensors using wireless and fiber connections and unknown sensors using unknown connections represent 8.6% and 7.4% of the final dataset, respectively. It appears that wireless connections are significantly less reliable than serial ones when using the Wavetronix radar sensors. While this difference may be the result of less reliable network hardware, it may also be the result of either magnetic or landscape disturbances. [Figure 24](#page-49-0) shows the location of the geocoded FAST sensors that passed all three assessments. It can be observed, by referring to [Figure 24,](#page-49-0) that many of wireless sensors and unknown sensors have been removed from the analysis. Overall about 15% of the Wavetronix radar sensors using serial/fiber connections, 29% of the Wavetronix radar sensors using wireless/fiber connections, about 32% of the ISS radar sensors using wireless/fiber connection, and close to 70% of the unknown sensors using unknown network connections have been removed from analysis after performing these three assessment steps.

Figure 24. FAST Sensors Remaining afrer all Three Assessments

FAST Traffic Data Failure Analysis

In this section an attempt to discover, using the same treemap and map visualizations as before, the possible failure patterns that lead to the results observed in [Figure 24.](#page-49-0)

[Figure 25](#page-50-0) shows a treemap organized in the same structure as the previous treemaps, except that it only shows the failing sensors and that the colors represent the assessment step at which the sensor failed, instead of an assessment score. Orange is used for sensors that failed the completeness assessment, purple is used for sensors that failed the quality assessment, and green is used for the sensors that failed the usability assessment. It can be seen that among the Wavetronix radar sensors using serial/fiber connections, the majority of the failures first occur in the completeness assessment and then in the usability assessment, few of these sensors failed the quality assessment. For ISS radar sensors using wireless/fiber connections, completeness assessment failure is the major cause of failure with a few sensors failing in the quality and usability assessments. Wavetronix radar sensors using wireless/fiber connections had no failures during the quality assessment, but an almost two-thirds, one-third split in failure between the completeness and usability assessments, respectively. Unknown sensors using unknown connections mostly failed the completeness assessment.

An interesting observation is that while the area of the unknown sensors, in the treemap, is roughly the same as the wireless sensors, there is a large number of smaller "tiles" representing sensors that collected very little data in 2013 in comparison to the other sensor/network combination.

Figure 25. Failure Treemap

Figures 26, 27 and 28 show the location of the geocoded FAST sensors that failed one of the three steps, for each of the sensors and connection combinations. For each of these maps, three different colors were used to indicate the failure type of the sensors. Sensors that failed the completeness assessment are represented as red dots, sensors that failed the quality assessment are represented as green dots, and sensors that failed the usability assessment are represented as blue dots.

[Figure 26](#page-51-0) shows the Wavetronix/serial/fiber that failed the assessments. It is observed that the failing sensors are all located on the upper part of I-15 and that the completeness failures are often located in between highway intersections. [Figure 27](#page-52-0) shows the Wavetronix/wireless/fiber that failed the assessments. It is observed that many of these sensors were not geocoded and do not appear on the map. The sensors visible on the map are showing a concentration of completeness failure in a string of sensors on I-215, east of the I-15 and I-215 interchange. [Figure 28](#page-53-0) shows the ISS/wireless/fiber that failed the assessments. It is observed that these sensors are located in the southern portion of I-15 and that the majority failed the completeness assessment. While the failure appears to be randomly distributed across I-15, it appears that between the towns of Sloan and Jean a consecutive group of sensors all failed completeness. Because of the proximity of the failing sensors,

this failure pattern may not be related to the sensors themselves but may be to a poor wireless connection in that specific area which could be the result of a malfunctioning or improperly located antenna or even environmental wireless noise or disturbance. The unknown sensors using unknown connections were not geocoded, therefore did not have location information available so it was not possible to map these devices.

Figure 26. Map of Failing Wavetronix/Serial/Fiber

Figure 27. Map of Failing Wavetronix/Wireless/Fiber

Figure 28. Map of Failing ISS/Wireless/Fiber

Top 10 Performing FAST Sensors

[Table 6](#page-54-0) contains the top 10 FAST sensors in terms of data completeness, quality, and usability and their respective performances. It is noted that ISS radar sensors using wireless/fiber connections represent the majority of the top performing devices, followed by Wavetronix radar sensors using serial/fiber connections. No Wavetronix radar sensors using wireless connections are listed and despite the large amount of failure observed in the unknown sensors using unknown connections, one of them is listed among the top ten performers. [Figure 29](#page-54-1) shows the location of the top ten sensors that have a geocoded location. It can be seen that all of these sensors are located in Las Vegas along I-15.

					Secondary	Primary	
RoadwayID	DetectorID	Completeness	Quality	Usability	Network	Network	Device
							Wavetronix
528	528.2.87	100.00%	99.19%	96.76%	Fiber	Serial/Ethernet	radar sensor
349	349.3.152	100.00%	99.17%	96.73%	Fiber	Wireless	ISS radar sensor
							Wavetronix
78	78.2.20	100.00%	99.12%	96.73%	Fiber	Serial/Ethernet	radar sensor
526	526.1.144	100.00%	99.16%	96.72%	Fiber	Wireless	ISS radar sensor
354	354.2.144	100.00%	97.72%	96.72%	Fiber	Wireless	ISS radar sensor
							Wavetronix
48	48.2.83	100.00%	99.18%	96.71%	Fiber	Serial/Ethernet	radar sensor
350	350.1.158	100.00%	95.18%	96.71%	Fiber	Wireless	ISS radar sensor
350	350.2.159	100.00%	98.17%	96.70%	Fiber	Wireless	ISS radar sensor
524	524.1.85	100.00%	99.18%	96.70%	Unknown	Unknown	Unknown
355	355.3.155	100.00%	98.92%	96.70%	Fiber	Wireless	ISS radar sensor

Table 6. Top 10 FAST Sensors

Figure 29. Location of Top 10 Performing FAST Sensors

Additional Visualizations

In order to gain more insight into the behavior of each of the sensors, two more visualizations were generated to allow NDOT to further examine the previous results and isolate the potential origin of these performances. An example of each visualization is shown below, however a visualization for each sensor was also generated for all 499 sensors (See Appendix B and C).

Hexagonal Binning Plot

Hexagon binning is an alternative to a scatterplot used when dealing with very large datasets where points have a tendency to stack up on top of each over on the plot and therefore not allowing the reviewer to visualize the true number of points at that location. The underlying concept is to split the plot into multiple hexagons, count how many points are present within each hexagon and color code each hexagon based on the point density. [Figure 30](#page-55-0) shows an example of hexagonal binning plot where a set of dark hexagons with a large concentration of points are able to be singled out. A regular scatterplot would not have been able to reveal them as they are embedded within a cloud of low to medium density.

Figure 30. Example of a Hexagonal Binning Plot

[Figure 31](#page-56-0) contains the same kind of plot applied to the speed-flow diagram of sensor 428.3.344, located on I-15 SB near Hotel Rio Drive and Martin Luther King Boulevard for the entire year 2013. Dark blue hexagons show a low density (500 records or less) whereas light blue are showing high density (between 1,500 and 2,000 records). This visualization is interpreted to show that the majority

of the traffic at that location occurs at a speed of 70 mph and a volume ranging from 1,000 to 2,000 vph and that breakdown and congestion, although they have occurred, are not common.

Figure 31. Example of Hexagonal Binning Speed-Flow Plot

Calendar Heatmap

A heatmap is a graphical representation of data where the individual values contained in a matrix are represented as colors. A calendar heatmap is a heatmap that uses a matrix representing calendar values such as month, year or days. A similar visualization of traffic behavior was encountered while visiting RTCSNV-FAST during the interview phase of this project. These were being built by hand using Microsoft Excel. Using the 2013 FAST dataset and the visualization tools developed in this project allowed the calendar maps to be generated automatically. In this case the matrix used represents the month, week, and day of the week for the year 2013 and the color of each tile of the matrix is color coded to show the average speed that day for the entirety of 2013 in one plot. [Figure 32](#page-57-0) contains an

example of this calendar heatmap for Roadway ID 437: I-15 NB near Decatur Road and Washington Avenue. It is observed that the middle of the week during the last three months of the year experienced slower traffic at that location than other times of the week. These kinds of patterns can be of great value when trying to build predictions. While these patterns may be known to TMC operators due to their seasonality or correlation with events happening in town at the time, they can be leveraged by prediction algorithms, in combination with datasets such as weather conditions or event calendars, to learn the behavior of this part of the road network and predict eventual recurrence. Grey represents the lack of usable data during these days of 2013.

Figure 32. Example Calendar Heatmap

In the next section, traffic prediction is explored utilizing the data deemed acceptable for further analysis.

PREDICTION DEMONSTRATION

This section demonstrates how the 2013 FAST data that passed all three of the previous assessments can be used to make traffic condition predictions on the Las Vegas roadway network. To do so, a FAST sensor among the top 10 performing sensors was selected and an experimental predictive model using its data was built then tested using historical data, the results and performance of the predictive model were then assessed and analyzed.

Having only one year of data is somewhat restrictive to build and test a prediction model capable of capturing the behavior of traffic at a specific location of the road network. Some traffic patterns that occur only once a year are difficult to identify using only one year of data and traffic anomalies that only occurred that year may be taken as baseline traffic behavior by the prediction model. This means that the performance of this experimental prediction model should not be judged too severely and that the goal of this demonstration is to help understand what is needed to develop more reliable and robust traffic prediction models in the future.

NDOT Vision and Need for Traffic Prediction

During meetings with NDOT and RTCSNV-FAST stakeholders, discussions were held regarding the various traffic management challenges, both recurring and nonrecurring, and how forecasting, or prediction, could be used to alleviate these challenges. The stakeholders also discussed peak period strategies, such as ramp-meters and signal timing plans that could be enabled in advance as a response to a traffic congestion prediction. These peak period response strategies are not considered real time responses and require 30 minutes or more to be implemented, therefore the prediction models that a TMC could use would need to forecast traffic conditions enough time ahead of the time it would take to implement the peak period strategies.

Stakeholders identified both recurring and nonrecurring events and noted the, often, random nature of traffic issues. While it is impossible to predict randomness, it is possible to investigate and learn if some traffic issues are correlated with specific events, such as weather conditions, day of year, or large conferences. NDOT and RTCSNV-FAST stakeholders noted that the following recurring events, taking place in NDOT's District 2, have a significant impact on traffic:

- Reno Rodeo 3rd week of June
- Hot August Nights 2nd week of August
- Lake Tahoe Shakespeare Festival August
- Burning Man End of August
- Sparks Rib Cook-off 1st week of September
- Reno Balloon Races 2nd week of September
- Street Vibrations 3rd week of September

In Southern Nevada (NDOT District 1), many events in Las Vegas also impact traffic. Events such as National Association of Stock Car Auto Racing (NASCAR) events, which brings 140,000 people to the Las Vegas area, the Electric Daisy Convention, which brings about 300,000 people to Las Vegas, and other special events including peak periods around holiday weekends.

NDOT staff and RTCSNV-FAST stakeholders mentioned the following locations as possible candidates for a prediction demonstration:

- NDOT's District 1:
	- o I-15 / US 95 Interchange (Spaghetti Bowl), rush hour
	- o McCarran Boulevard near the airport
	- \circ I-15 near the California border (especially weekend traffic between Las Vegas and California)
- NDOT's District 2:
	- \circ State Route 447, a 2-lane highway that services the Burning Man event
	- \circ I-80 near California border, truck parking issues related to weather-related closures heading into California
	- o I-80 in Sparks (Sparks Boulevard and Vista Boulevard, near the Keystone Shopping Center)

Prediction Approaches

Prediction models can typically be separated into two groups: regression models and classification models. Based on the 15-minute time resolution of the 2013 FAST dataset, the experimental prediction model developed in this study was done using a classification approach rather than a regression one. Regression models are designed to predict a single continuous variable of the environment it emulates. In the case of the FAST dataset, the continuous variable could be the speed or volume of traffic at a specific location. Classification models on the other hand are attempting to categorize the future conditions based on an already defined set of categories in the training dataset.¹¹

In the case of the FAST dataset, the categories could be one of the phases in the three-phase traffic theory; free flow (F), synchronized flow (S) or wide moving jam (J). Both of these models require some inputs such as date, time, weather, and any additional data that appear to be influential to the outcome of the prediction model. Traditionally forecasted weather conditions are always included as an input in prediction models.

The FAST dataset provides historical traffic conditions that have a 15-minute resolution. Speed and volume in the dataset are averages of the observed speed and volume over a 15-minute

 \overline{a}

¹¹ Hashemi, S., Almasi, M., Ebrazi, R., & Jahanshahi, M. (2012). Predicting the Next State of Traffic by Data Mining Classification Techniques. *International Journal of Smart Electrical Engineering, 1*(3), 180-192. Retrieved from http://www.sid.ir/en/VEWSSID/J_pdf/5068520120304.pdf

interval. This is a rather high aggregation level when it comes to traffic variable variations. Changes that could be picked up by the model could potentially be blended in and smoothed within a 15 minute interval resulting in much more difficult traffic pattern to identify and leverage. This means that a regression model designed to predict speed or volume reliably would be difficult to achieve.

Also the usefulness of predicting the speed or the volume may be limited as traffic behaviors cannot be expressed using only one variable, and the implementation of specific peak period strategies may be more difficult. For these reasons, it was decided to focus on a classification approach and develop a prediction model using the three-phase traffic theory as a method of classification for the FAST data. The 2013 FAST dataset was classified and a traffic phase was assigned to each sensor for each 15-minute interval, the prediction model was then trained to predict the most likely traffic phase the sensor will experience in the future.

While travel time appears to be an ideal measure, it is the product of a calculation using speed at various locations and can be just as difficult to predict as speed can be using the FAST data.

Predictive Algorithms

Traffic researchers have investigated over the years a wide variety of pattern recognition algorithms ranging from k-Nearest Neighbor¹² and Hidden Markov Model¹³ to complex Neural Network¹⁴ in an attempt to predict future traffic conditions. Researchers have improved and fine-tuned these algorithms to try to handle the stochastic nature of traffic, but, to this day, reliable real-time prediction models are not possible without the addition of a parallel traffic simulation component. The spatiotemporal and stochastic nature of traffic makes it very difficult to develop good prediction models using traditional algorithms only, especially when attempting to predict traffic conditions at a minute or second level. Various pattern recognition algorithms that have been or could be used to predict traffic conditions are discussed in the following paragraphs.

 \overline{a}

¹² Zheng, Z., & Su, D. (2014, June). Short-term traffic volume forecasting: A k-nearest neighbor approach enhanced by constrained linearly sewing principle component algorith. *Transportation Research Part C: Emerging Technologies, 43*(Part 1), 143-157. Retrieved from https://www.researchgate.net/publication/260804708_Shortterm_traffic_volume_forecasting_A_k-

nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm

¹³ Necula, E. (2014). Dynamic Traffic Flow Prediction Based on GPS Data. *2014 IEEE 26th International Conference on Tools with Artifical Intelligence (ICTAI)* (pp. 922-929). Limassol: IEEE. Retrieved from

http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6984576&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs _all.jsp%3Farnumber%3D6984576

¹⁴ Kumar, K., Parida, M., & Katiyar, V. K. (2013). Short term traffic flow prediction for a non urban highway using Artificial Neural Network. *Procedia-Social and Behavioral Sciences, 104*, 755-764. Retrieved from http://www.sciencedirect.com/science/article/pii/S1877042813045618

k-Nearest Neighbor Algorithms

In pattern recognition, the k-Nearest Neighbors algorithm (or k-NN) is a non-parametric method used for classification and regression. It makes its prediction by combining the values of the closest historical data points that have happened under similar conditions. The k-NN algorithm is among the simplest of all machine-learning algorithms. It is referred to as lazy learning, where the outcome (or prediction) is only approximated locally in the dataset. This simplicity unfortunately makes it very sensitive to noisy and unbalanced datasets.

Artificial Neural Networks

Artificial Neural Networks, also known as "Artificial neural nets", "neural nets", or ANN for short, are computational tools modeled on the interconnection of the neuron in the nervous systems of the human brain. ANNs mimic the biological neural networks found in the human brain. ANNs are network systems constructed from atomic components known as "neurons". Artificial neural nets are a type of non-linear processing system that is ideally suited for a wide range of tasks, especially tasks where there is no existing algorithm for task completion, such as the prediction of nonlinear phenomena. ANN can be trained to solve certain problems using a teaching method and sample data. With proper training, ANNs are capable of generalizations and are able to identify patterns in noisy and unbalanced datasets. ANNs are notoriously difficult to optimize as the weights and thresholds used in each of the "neurons" do not have a direct meaning in the real world.

Random Forest

With the advent of commodity cluster computing (cloud), a new kind of prediction algorithm, based on traditional prediction methods, but leveraging large parallel computation capabilities, has been developed. The Random Forest algorithm is one of these algorithms. Random Forest is an ensemble learning method for classification or regression that operate by constructing a multitude of decision trees at training time on subsets of the training sets and then outputting a mean of individual trees' predictions. Random Forest leverages the scale of cloud computing to correct the decision trees' habit of over fitting to their training set. Random Forest is simple yet capable of dealing with noisy and unbalanced datasets. Random Forests' algorithms and its variations are the main algorithms used to perform prediction online such as advertising selection, book recommendations, and other marketing predictions. They are also used in law enforcement and emergency response to predict crime and incidents. [Figure 33](#page-62-0) shows a graphical representation of the Random Forest algorithm where several decision trees are combined into an ensemble to make a prediction.

Figure 33. Random Forest Graphical Representation

Based on the limited amount funding and time available in this study to develop and tune an experimental traffic prediction algorithm, the team decided to choose the Random Forest algorithm to build the experimental prediction demonstration as it is simple, capable of dealing with noisy data and unbalanced dataset and as already been implemented as part of many open source and commercial prediction software solutions.

Detector Selection

An attempt was made to identify a sensor that provided reliable data and that was close to the area of interests listed by NDOT staff and RTCSNV-FAST stakeholders; near either the I-15 / US 95 Interchange (Spaghetti Bowl), the McCarran Boulevard near the airport, or on I-15 near the California border. The closest and most usable sensor found among the top 10 performing sensors is detector 350.1.158 located on I-15 southbound near Spring Mountain Road and Decatur Road. It is an ISS radar sensor that uses a wireless/fiber connection to relay its data to the FAST TMC. Its scores based on the three data assessment steps are shown in [Table 7.](#page-63-0)

[Figure 34](#page-63-1) shows the location of the sensor 350.1.158 on a map of Las Vegas the lone blue dot on the map. [Figure 35](#page-64-0) and [Figure 36](#page-65-0) show sensor 350.1.158's hexagonal binning speed-flow diagram and its calendar heatmap. [Figure 35](#page-64-0) shows a clear concentration of traffic around 65 mph and between 20 and 2,000 vph, a small concentration of data points around 25 mph and between 0 and

1,000 vph, and a small number of data points scattered between the two main clusters of traffic data points. According to the three-phase traffic theory, the first and largest cluster of data points is considered to be the free flow (F) phase at this location, the second and bottom cluster is considered the wide moving jam (J) phase, and the scattered data points between the F and J phases and the transition data points between the F and J phase is considered to be the synchronized flow (S) phase. This classification of the FAST dataset was used to build a traffic prediction model.

Figure 34. Location of FAST Sensor 350.1.158

[Figure 36](#page-65-0) shows a calendar heatmap of the speed recorded by sensor 350.1.158 across the year 2013. The speed data in this heatmap is aggregated by day and the color gradient used for each day tile on the heatmap is representative of the average speed at this location across an entire day. The gradient ranges from green, which represents an average speed of 70 mph across the entire day, to

red, which represents speeds 50 mph and below. The color gradient range on this heatmap was kept between 50 and 70 mph so as to highlight days with speed averages below 50 mph. These low speed averages appear in bright red on the calendar heatmap.

Figure 35. Hexagonal Binning Speed-Flow Plot of Sensor 350.1.158

It is observed on the calendar heatmap that for the first three weeks of January 2013 the tiles are red. This pattern is rather atypical of traffic on interstates unless construction is occurring. By comparing the hexagonal binning speed flow plot [\(Figure 35\)](#page-64-0) and the calendar heatmap [\(Figure 36\)](#page-65-0), it can be seen that the bottom cluster of data points, the J phase based on the three-phase traffic theory, has an average speed of about 25 mph and that these data points are most likely the slow speed observed on the calendar heatmap during the first three weeks of January.

Since construction information was not available, beginning and end dates of construction could not be correlated to what was observed. Without the ability to tell the prediction model that data points during these three weeks are representative of the traffic behavior at this location when construction is occurring, the prediction is skewed and produces a biased mode of traffic, but this behavior can be solved by adding an additional year of data and a construction input variable when developing the prediction model. Despite this possible bias, development of the prediction model was pursued in order to demonstrate the process and the outcomes that such a prediction model can provide.

Figure 36. Calendar Heatmap of Sensor 350.1.158

Clustering and Classification

Before being able to develop, that is to say, train, a prediction algorithm such as a Random Forest classification to learn the behavior of traffic at a specific location, it was needed to classify traffic at that location into significant states or phases for each of the data points collected by the sensor.¹⁵

To do so, a process called clustering, which attempts to discover closely correlated data points, was used. Data points that share similar values and tend to be located next to each other on a graph are grouped together to form one or more clusters. To cluster sensor 350.1.158 data points, a three dimensional clustering method based on speed, volume, and occupancy was used. [Figure 37](#page-67-0) shows the clusters of data points discovered by the clustering algorithm. A total of seven distinct clusters were discovered by the algorithm when attempting to group all 2013 data points by their speed, volume, and occupancy. Each of the ovals in [Figure 37](#page-67-0) represents the average radius of a cluster. While five of the clusters are all contiguous and representative of the F traffic phase, it can be seen that the algorithm was able to distinguish the J traffic phase in green, and the S traffic phase in purple. To simplify the prediction model and adjust the clusters to the three-phase traffic theory, the five F traffic phase clusters, represented in purple, red, light blue, orange and blue on [Figure 37,](#page-67-0) were merged into a single cluster this allowed for the assignment of a traffic phase to each of the 2013 data points recorded by sensor 350.1.158.

The clustering step transformed the raw FAST sensors data into a dataset correlating the three possible traffic phases at sensor 350.1.158 to each of the 15-minute interval, hour, day, month, day of the week, and holiday values found within the FAST data. This new data set was used to train a prediction model capable of guessing the traffic phases based on the minute, hour, day, month, day of week, and holiday.

 \overline{a}

¹⁵ Xia, J., Huang, W., & Guo, J. (2012). A Clustering Approach to Online Freeway Traffic State Identification Using ITS Data. *KSCE Journal of Civil Engineering, 16*(3), 426-432. doi:10.1007/s12205-012-1233-1

Figure 37. FAST Sensor 350.1.158 Clustering Algorithm Results

Ancillary Data

Despite the recurring nature of traffic patterns over time, time only is not a sufficient enough input to accurately predict the traffic phase at a specific location and time. The prediction model could benefit from discovering and learning correlations of traffic phases with data other than time and date to help design a more precise prediction model. This type of data is often referred to as ancillary data. Weather data is such a dataset, it varies across time and often influences human behaviors based on its values (temperature, humidity, etc). This is true for traffic in Las Vegas, for example people in Las Vegas may or may not decide to perform certain activities depending on the outside temperature. Additional ancillary data can also be used to build more precision into the prediction model, in the case of Las Vegas, its touristic activities and its numerous recurring events also influence traffic. The events locations, and durations may be of value to build a more precise traffic prediction model.

In this section, a description of ancillary data researchers were able to gather and use to provide more precision to the traffic prediction model is provided. As mentioned previously, weather data and Las Vegas events data was used to add additional information to the FAST dataset that could improve the prediction success of the developed model.

Weather Data

The possibilities of using the Nevada Road Weather Information System (RWIS) stations data was explored, but issues with missing data on some stations and challenges arose when trying to obtain one year of low resolution (hourly) weather data for the stations. As an alternative weather dataset, data from the National Weather Service archives for the McCarran Airport in Las Vegas (KLAS) was used. ¹⁶ [Figure 38](#page-70-0) shows a map of the Las Vegas area displaying the area the weather station monitors next to the McCarran Airport. The data available is from the National Oceanic and Atmospheric Administration (NOAA) web page for KLAS weather station for each day of 2013.¹⁷ The weather data provided several measures including: temperature, dew point, humidity, sea level pressure, visibility, wind and gust speed, precipitation, and wind direction. [Table 8](#page-69-0) contains the various weather measures available from the NOAA dataset and their data type. Daily observations from this weather station dataset were correlated using the date and added to the transformed FAST dataset thereby adding additional input values from which the prediction algorithm could uncover and learn patterns.

 \overline{a}

¹⁶ *METAR/Synop Information for KLAS (LAS) in McCarran International Airport, NV, US*. (2014). Retrieved from Weather Quality Reporter: http://weather.gladstonefamily.net/site/KLAS

¹⁷ *Las Vegas, McCarran International Airport*. (2012). Retrieved from National Weather Service: http://w1.weather.gov/data/obhistory/KLAS.html

Table 8. Sample Data from KLAS

Date – Calendar date

Integer – Numeric value

Character – Character string of variable length with a maximum length of 255 characters Float – Numeric, decimal value

Figure 38. Location of KLAS Weather Station¹⁸

Las Vegas Events Data

In order to obtain data relevant to events in Las Vegas during the year 2013, the team contacted the Las Vegas Convention Center. The Convention Center provided researchers with a calendar of its 2013 activities including venues, beginning date, end date, and expected attendance. The calendar was provided as an edited Adobe document that was extracted and cleaned as a table containing all 2013 events, their beginning date, their end date, their actual location, and their venue, as well as, the attendance. [Table 9](#page-71-0) shows a description of the columns of this table.

 \overline{a}

¹⁸ *McCarran International Airport, Las Vegas, NV*. (2015). Retrieved from Google Maps: https://www.google.com/maps/@36.0872362,-115.1752943,11z

[Table 9](#page-71-0) was then transposed, reorganized, and correlated with the transformed FAST dataset using the date. This step added a column for each venue location and their attendance on that day. [Table 10](#page-71-1) shows a two row sample representation of the final table following the transposition, reorganization, and correlation.

Table 10. Simplified Example of Final FAST Prediction Dataset for Sensor 350.1.158

Date	Time	Temperature	Bellagio	Mirage	Traffic Phase
01/02/2013	18:15	54	12,500	27,000	Free Flow
01/02/2013	19:00	53	12,500	27,000	Free Flow

With the addition of ancillary data, the dataset for sensor 350.1.158 contained a total of 94 values (columns) including the date, time, weather data, events data columns, and 31,563 observations for each and every 15-minute intervals of 2013 (rows). It was noted that the dataset has less columns than there are 15-minute intervals in a year (35,040), due to some of the invalid or erroneous sensor data was removed during this assessment.

Feature Analysis

While weather and event data were gathered and added to the FAST dataset to augment the prediction capability of the developed prediction model, some of the ancillary data columns may not have any correlation with the traffic phases at the location of sensor 350.1.158 and resulted in complicating the development of a prediction model without providing any additional value to the modeling process. To allow the prediction algorithms to focus on relevant patterns and correlations,
some data needed to be removed from the dataset. To identify which data columns do not appear to vary in correlation with the variation of sensor 350.1.158's traffic phases, the team used a process called feature selection analysis. The features or predictors are the columns of the sensor 350.1.158 dataset other than the traffic phase column, which is referred to as the outcome.

To identify the features relevant to the changes of traffic phases in sensor 350.1.158's dataset, the team performed a variance analysis between each of the features (columns other than traffic phases) and the outcome (traffic phase). Features that showed near zero variance with the traffic phase were removed from the dataset.

As a result of this analysis, out of the original 94 columns of the FAST sensor dataset, only 20 were selected to have an impact on the traffic phase. Below is a list of these columns (or features):

- Date and time
	- o Month
	- o Day
	- o Hour
	- o Minute (15-minute interval)
	- o DayOfWeek
- Weather data
	- o Mean Temperature in Fahrenheit
	- o Mean Wind Speed in MPH
- Events data
	- o Attendance at the Bellagio
	- o Attendance at the Caesars Palace
	- o Attendance at the Tropicana Las Vegas
	- o Attendance at the Mandalay Bay Resort And Casino
	- o Attendance at the Bally's Las Vegas
	- o Attendance at the Circus Circus Hotel Casino and Theme Park
	- o Attendance at the Monte Carlo Resort and Casino
	- o Attendance at the Venetian Resort Hotel Casino
	- o Attendance at the ARIA Resort And Casino
	- o Attendance at the Cosmopolitan of Las Vegas
	- o Attendance at the South Point Hotel Casino And Spa
	- o Attendance at the Westin Las Vegas Hotel Casino And Spa
	- o Attendance at the MEET Las Vegas

This final 20 feature dataset was then used to train the experimental prediction model.

Training and Optimization of Prediction Model

After the removal of the 74 features of the dataset showing no influence on sensor 350.1.158 traffic phases from the final dataset, the development of the prediction model began. The development of a prediction model is often referred to as training because the model is being adjusted by the prediction algorithm as it is exposed to the historically observed predictors and outcomes.

The developed prediction model was trained using the Random Forest algorithm and a randomly chosen set of two-thirds of sensor 350.1.158's final dataset. The remaining one-third was used to test the validity of the generated prediction algorithm. During this process, the prediction model was optimized using two accuracy prediction processes called cross validation and bootstrapping.

Cross Validation

Cross validation is a process used when developing prediction models to avoid over fitting the model to the data. As the prediction model is optimized, using a subset of the original dataset, the model was tested on another subset of the original dataset to estimate its validity. In the case of the final FAST dataset, the testing set representing one-third of the final dataset was used for cross validation.

Bootstrapping

Bootstrapping, or bootstrap aggregating, is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. It also reduces variance and helps to avoid over fitting. It is usually applied to decision tree methods.

It was noted by the research team that it is not ideal to split a dataset containing only one year of data as some traffic patterns of interest that only happen once a year may be compromised by the split even if done randomly across the dataset. It would have been preferable to have two or more years' worth of data to train and test the prediction model.

Testing of Prediction Model

[Table 11](#page-74-0) and [Figure 39](#page-75-0) contain the comparison between the observed traffic phase of the testing dataset and the results of the prediction algorithms generated when the predictors of the testing dataset were used as input. Both [Table 11](#page-74-0) and [Figure 39](#page-75-0) contain a matrix called a confusion matrix which compares the actuals outcome of the test dataset versus predicted outcome. The matrix is designed to quickly show how the prediction model is performing. [Table 11](#page-74-0) the confusion matrix of sensor 350.1.158's prediction model. It is observed that a total of 9,308 traffic phases were predicted by the model. The predicted traffic phases that matched the actual traffic phases are shown in green and are on the diagonal. It can be seen that out of 8,703 records with an actual traffic phase of Free Flow (F), 8,672 were predicted as Free Flow (F) and 31 were predicted as Synchronized Flow (S) by the

model. This represents a Free Flow classification error of 0.36%. It was observed that all 605 records with an actual traffic phase of Wide Moving Jam (J) were classified as Wide Moving Jam (J) by the model with no error. The model was not as successful dealing with Synchronized Flow (S). Out of the 163 records with an actual traffic phase of Synchronized Flow, it was predicted that 104 records were in Free Flow (F) and only 59 were in Synchronized Flow phase (S). This represents an error of 63.8% which is too high to allow the model to be trusted when predicting Synchronized Flow.

Table 11. Confusion Matrix for Sensor 350.1.158 Prediction Model

The poor performance of the Synchronized Flow detection can be explained by several factors. First there are only a limited number of records in this phase as it is a transition phase between Free Flow and Wide Moving Jam. Data points categorized to be in the Synchronized Flow cluster are rather dispersed and border the Free Flow cluster (see [Figure 37\)](#page-67-0). This makes it difficult for the algorithm to find patterns capable of differentiating Synchronized Flow and Free Flow. Training the prediction model with a dataset at a lower resolution than 15-minute interval aggregation or adding a trend component as an input to the model may help dissociate the two clusters.

The lack of error shown by the prediction model when detecting Wide Moving Jam is rather abnormal. It can be explained by the fact that the only Wide Moving Jams occurred during a period of potential construction during the first three weeks of January 2013, this pattern was easily isolated and learned by the Random Forest Algorithm when building the prediction model associating only one possible traffic phase, Wide Moving Jam, with any days during the first three week of January as a result. This shows a strong bias in the model that could be eliminated by the addition of January data from years when construction was not ongoing if this information could be verified.

[Figure 39](#page-75-0) is similar to [Table 11](#page-74-0) but instead of showing actual and predicted counts, it normalized the outcomes of the prediction between 0 and 1. A gradient between dark blue (0) and light blue (1) is used instead of an actual value, this allows the performance of the model to be easily seen, when data point counts vary greatly between each category (phases in this case). I[n Figure 39,](#page-75-0) it is clearly seen that most of the actual Synchronized Flow (S) is predicted as Free Flow (F), light blue tile on the bottom right, and only some predicted records are correctly predicted as Synchronized Flow, medium blue tile on top right.

Figure 39. Heatmap of Sensor Prediction Model Confusion Matrix

Use Case of Traffic Phase Prediction Model

Despite the lack of precision in detecting the Synchronized Flow traffic phase, revealed in [Table 11](#page-74-0) and [Figure 39,](#page-75-0) it was decided to rely on the detection performance of the two other traffic phases, Free Flow and Wide Moving Jam, to test the model on a set of three full days to show how it could be used to plan ahead for a particular weekend, or, perhaps, week. Since data for 2014 was not available, a dataset was created using data from 2013. The models inputs were collected, including weather and events data, for a three-day period from January 22nd to January 24th, 2013. Noise, or variation, was added to the original dataset to avoid testing on the exact same data that was used to train the model. The level of noise was kept to a minimum of plus or minus 20% of the original data. The outcome of the prediction model using this noisy data can be observed in [Figure 40.](#page-77-0)

The model predicted the probability that traffic will be in one of the three traffic phases, Free Flow, Synchronized Flow, and Wide Moving Jam, for each 15-minute interval between midnight on

January 22nd and midnight on January 24th, 2014. [Table 12](#page-76-0) shows an example of the probability output of the prediction model on at 1:00 am on January 22nd, 2014.

[Figure 40](#page-77-0) shows a matrix of each 15-minute interval over the three day test period. The color inside each of the matrix tile represents the observed traffic state during the 15-minute interval. Orange represents Wide Moving Jam, green represents Free Flow. It is noted that some tiles are missing from the matrix, this is because these time intervals were removed during the FAST data assessment process. Figures 41, 42 and 43 are showing the probability that the model calculated for each of the 15-minute intervals over the three day test period. It is observed that [Figure 41](#page-78-0) shows a high probability, between 0.9 and 0.6, of being in the Wide Moving Jam traffic phase between 0:00am and 11:45am on January 22nd, but starting at 12:00pm the probability decreases to about 0.2 for the rest of the day. The following days are showing a near 0 probability of traffic being in the Wide Moving Jam traffic phase. [Figure 42](#page-79-0) shows the probability of being in the Synchronized Flow phase. It is observed that only a few 15-minute intervals, between 0:00am and 4:00am on January 23rd and January 24th, show a probability of being in the Synchronized Flow phase and the probability for these intervals is rather low with a maximum of 0.004. [Figure 43](#page-80-0) shows the probability of being in the Free Flow phase. It is observed that the prediction model shows a low probability, less than 0.3, on January 22nd from 0:00am to 11:30am. From 11:30am to 12:00pm the probability increases to 0.4 and 0.5, respectively. Past 12:00pm on January 22nd, the probability of being in Free Flow rises to 0.8 until the end of the day and is predicted to be between 0.9 and 1 for the last two days.

Figure 40. Traffic State Between 01/22/2014 and 01/24/2014 to be Predicted

Figure 41. Probability of Being in the Wide Moving Jam (J) Traffic Phase

Figure 42. Probability of Being in the Synchronized Flow (S) Traffic Phase

Figure 43. Probability of Being in the Free Flow (F) Traffic Phase

Figure 44. Side by Side Comparison of Three Day Prediction

[Figure 44](#page-81-0) shows a side by side comparison of each of the predicted traffic phases and the original traffic phase prediction. These prediction values could easily be used to generate visualizations, such as calendar heatmaps or roadmaps, showing colored road segments used in combination with a time slider, along with a color gradient, representing the most probable traffic phase that road segment will be in at any given time of day. [Figure 45](#page-82-0) and [Figure 46](#page-83-0) contains examples of such prediction visualizations. [Figure 45](#page-82-0) contains a prediction visualization using a map where the various road segments on the map are color coded based on their predicted traffic phase. The user can then use the horizontal slider to advance the map in time and observe the phase changes as time changes.

Figure 45. Prediction Map Mockup

[Figure 46](#page-83-0) shows a second example of a prediction visualization. This visualization uses a traditional time-space plot approach often found in traffic engineering. The plot is organized as a matrix where each row represents a specific section of a highway, or a sensor on a corridor, and each column represents a 15-minute interval. Each tile of the matrix is color coded based on the predicted traffic phase. The user can use both the horizontal and vertical slider to explore the predicted evolution of traffic along the corridor over the next few hours or days.

Figure 46. Time-Space Plot Mockup

In the next section, a gap and SWOT analysis is conducted to further explore the possibility of utilizing proactive traffic prediction for traffic management.

GAP AND SWOT ANALYSIS

The findings from the ITS architecture assessment and the experimental prediction demonstration fed into two analyses: a Gap analysis and a SWOT analysis. This section of the report describes the approaches used for conducting these analyses. The focus of both the gap and SWOT analysis will revolve around the FAST capabilities as most of the infrastructure and data used in this study was provided by RTCSNV.

Gap Analysis Approach

The research team followed the following four-step process to complete the gap analysis:

- 1. Establish the fundamental questions for the gap analysis.
- 2. Create a gap analysis template.
- 3. Populate the template with findings from NDOT interviews and in-person meetings, TMC site visits, and a review and assessment of NDOT/RTCSNV FAST and the data it collected in 2013, as well as an attempt to build prediction model using this data.
- 4. Make comparisons and assess where gaps exist.

First, the project team established the following fundamental questions with which to assess the existing the infrastructure used by FAST, the data it generated, and how it is being used. These questions included:

- Where is NDOT/RTCSNV currently in terms of traffic data collection and analysis (e.g., what are the current traffic data sources, what are the current networking technologies being used, how is the traffic data stored and organized, how is the traffic data being used with regard to traffic management activities)? Answering these questions will establish NDOT's current performance in the area of data sources and transmission techniques.
- Where does NDOT/RTCSNV want to be with respect to traffic data collection and analysis (e.g., what improvements to processes and/or tools does NDOT/RTCSNV wish to make, what data do NDOT/RTCSNV need/want, how and for what purposes does NDOT/RTCSNV want to use these data)? Answering these questions will help to establish NDOT/RTCSNV goals for future performance, or potential performance, in the area of traffic data and analysis.

A gap analysis template was created to answer many of these fundamental questions based on the work completed in the information gathering, data collection, and analysis tasks. [Table 13](#page-85-0) shows the gap analysis template. The template contains the two fundamental questions, as well as four elements on which to assess each question:

- Traffic data sources and networking capabilities
- Traffic data collection and storage
- Traffic data analysis and prediction

Traffic prediction integration with current traffic management practice

Each cell indicates the source from where the information/findings was drawn to answer the questions for each of the assessment elements. Information sources included: conversations with the NDOT project management team, findings from the project kick-off meeting and TMC site visits, and the review and assessment of the FAST infrastructure and data.

Once information was drawn from the various sources and the template was populated with this information, the following comparison was made to determine if gaps exist and if so in what areas:

 Where NDOT is currently (current performance) *versus* Where NDOT wants to be in the future in terms of traffic data collection, analysis, and prediction (potential performance).

The findings of the gap analysis are found following the SWOT Analysis Approach section.

SWOT Analysis Approach

A SWOT analysis is a strategic planning method used to evaluate the Strengths,

Weaknesses/limitations, Opportunities, and Threats (SWOT) involved in a project. It involves stipulating the objective of the project and then identifying the internal and external factors that are advantageous and disadvantageous to achieving that objective. A multi-step methodology was used for conducting the SWOT analysis:

- 1. State the objective of the project.
- 2. Create a SWOT analysis template.
- 3. Populate template with findings from discussions with NDOT staff, from the project kick-off meeting, TMC site visits, and review/assessment of the 2013 FAST dataset and the prediction demonstration.
- 4. Conduct analysis.

[Table 14](#page-87-0) shows the SWOT analysis template, which illustrates the four quadrants of the SWOT analysis for this project. As with the gap analysis, the team used findings from the information gathering, data collection, and analysis tasks to populate the quadrants of the template.

Finally, the strengths, weaknesses, opportunities, and threats were analyzed with respect to the same four assessment elements that were used in the gap analysis:

- Traffic data sources and networking capabilities
- Traffic data collection and storage
- Traffic data analysis and prediction
- Traffic prediction integration with current traffic management practice

The findings from the SWOT analysis are discussed in detail following the Results of Gap Analysis section.

Results of Gap Analysis

This section presents the findings from the gap analysis and provides insight as to investments needed to achieve the goal of proactive traffic management.

Current ITS Performance Versus Potential ITS Performance

A comparison of NDOT/RTCSNV's current ITS performance to its potential ITS performance was made for each of the four assessment elements defined in the gap analysis template [\(Table 13\)](#page-85-0) and they are described here.

Traffic Data Sources and Networking Capabilities Where does NDOT/RTCSNV want to be?

NDOT/RTCSNV would like to have an ITS capable of producing data that can be used to predict traffic conditions that could lead to corrective actions. This translates into at least 80% of all collected data being complete, valid, and usable.

Where is NDOT/RTCSNV now?

By reviewing the results of the 2013 FAST dataset assessment in [Table 5,](#page-48-0) it can be concluded that about 70% of the dataset passed the three assessments: Completeness, Quality, and Usability. This is lower than the 80% target but not greatly different. Further investigation of [Table 5](#page-48-0) reveals that depending on the type of sensors, network connections being used to collect the data, and location the percentage of usable data varies greatly. For example, Wavetronix radar sensors connected through a serial, then fiber connection, which are mostly located along I-15 inside the Las Vegas metropolitan area, shows a percentage of 85.71% of usable traffic data on average, which is above the 80% threshold. This usable data represents almost 40% of the total amount of data collected by FAST in 2013. The loss of usable data is mainly concentrated in the unknown sensors or legacy sensors whose locations are unknown and could not be mapped in [Figure 3.](#page-20-0) The unknown sensors and legacy sensors with unknown network combinations represent only 30% of the usable data collected in 2003, as shown i[n Table 5.](#page-48-0) While this is a significant loss, the unknown sensors, legacy sensors, and their connections are considered to be sensors that will eventually be replaced. They account for about 17% of the total amount of data collected by FAST in 2013, which is not sufficient to generate a loss of 20% even if all sensors failed. The main reason why the 2013 FAST dataset failed to pass the 80% threshold is the lack of reliability related to road sensors using wireless connection. In [Table 5](#page-48-0) it is observed that both Wavetronix radar sensors and ISS radar sensors using a wireless connection are showing a percentage of usable records of 68.25% and 71.02%, respectively. If the performance of the Wavetronix radar sensors using both serial connections and wireless connections are compared, it is seen that the use of wireless connections generate a drop of almost 18% in usable records. Also, review of [Figure 27](#page-52-0) and [Figure 28](#page-53-0) it is observed that wireless connected road sensors are failing in small consecutive groups that are collocated. A small group of Wavetronix radar sensors with a wireless connection that failed the completeness assessment can be observed on I-215 east of the I-15/I-215 intersection and a group of ISS radar sensors that failed the completeness assessment can be observed on I-15 toward California between the towns of Sloan and Jean. These "pockets" of wireless

sensor failure had a significant role in the 2013 wireless connected sensors performance, however these "pockets" may not be related to the performance of the system itself, but to environmental wireless disturbance or the presence of construction for several months at that location. Following these observations, the NDOT/RTCSNV ITS performance was assessed and it was found that, while close, it is only partially meeting the future requirements for prediction. It is recommended that an investigation into the wireless sensors performance be conducted, along with the modernization of the unknown and legacy sensors with unknown network connections, to close this gap. [Table 15](#page-89-0) includes a summary of this analysis.

Table 15. Traffic Data Sources and Networking Capabilities

Traffic Data Collection and Storage

Where does NDOT/RTCSNV want to be?

NDOT/RTCSNV would like to be able to collect, organize and store current traffic data at a low resolution (1-minute or less), ancillary data such as weather, construction, and city events attendance data for several years while retaining the ability to query, visualize and analyze the data in its entirety.

Where is NDOT/RTCSNV now?

NDOT/RTCSNV is currently collecting data using a traditional transactional (OLTP) / analytical (OLAP) relational database architecture. Traffic is captured by sensors at a 1-minute interval, then sent to a controller that aggregates it at a 15-minute interval. The aggregated data is then recorded into a relational database table, where it is assessed for defect and quality, transformed (date time format, unit conversion, etc.) and loaded into an analytical database. This process is known as an extract, transform, and load process (ETL). The analytical database contains multiple lookup tables populated with metadata, such a geo-location of road sensors, allowing the collected data to be merged with metadata on demand and visualized through the many charts on the FAST web dashboard. The NDOT/RTCSNV FAST database is limited in space, it cannot retain more than a few years of data within the database, and regularly removes and archives aging data to flat files. The NDOT/RTCSNV data collection system also allows a direct connection to live sensor data using a XML web service-like connection. The data published thought this XML web service is refreshed on a minute basis.

While the ETL process currently used by the NDOT/RTCSNV FAST database to upload data performs quality checks on its data feed, these checks are not sufficient to detect and isolate certain records, such as the ones failing completeness, which could compromise the ability to build reliable prediction models. The NDOT/RTCSNV FAST database may also be challenged by the complexity of database queries needed to be run to detect and isolate such records. These queries typically require heavy calculations, grouping and comparison on millions of records (an entire year worth of data), and may be detrimental to the sustainment of a responsive FAST web dashboard when ran.

The NDOT/RTCSNV FAST database system collects and manages, in correlation with road sensor data, traffic incidents data captured by TMC operators observing incident scenes using one or more of the video feeds available on the FAST dashboard. The FAST dashboard also appears to collect some weather data. Some features on the FAST dashboard, such as the wind over Hoover Dam Bridge, show that some weather data is collected and recorded, but the weather data does not appear to have been integrated with the traffic data at the data model level. Construction data is not present on the NDOT/RTCSNV FAST database system, it is managed by the maintenance office of NDOT Region 1. Las Vegas events data including attendance data or hotel occupancy data is also not present in the NDOT/RTCSNV FAST database system.

As was discussed in the prediction demonstration section, prediction cannot be done solely using historical traffic data, date, and time. Ancillary data, such as weather and Las Vegas event attendance data, are needed to supplement historical traffic data and to obtain a reliable prediction model. Based on these observations, while the NDOT/RTCSNV FAST system is efficiently collecting, storing and using traffic data and video, it was found that the NDOT/RTCSNV ITS data collection and storage is not meeting the future requirements for prediction. It lacks the presence of construction data integrated with the traffic data and other necessary data feeds, such as weather and city events data, to develop reliable prediction models. It also lacks the ability to store and process very large amounts of traffic data and ancillary data in order to isolate prediction worthy records. This kind of large database storage and analytical capability often requires investment that may exceed the NDOT/RTCSNV funding ability, but cloud services could offer a less expensive alternative. [Table 16](#page-91-0) contains a summary of this analysis.

Table 16. Traffic Data Collection and Storage Capabilities

Traffic Data Analysis and Prediction

Where does NDOT/RTCSNV want to be?

NDOT/RTCSNV would like to improve its current live traffic data analysis (FAST dashboard) with a reliable traffic prediction feature to improve its ability to anticipate upcoming traffic conditions and react as early as possible.

Where is NDOT/RTCSNV now?

NDOT/RTCSNV is currently making use of its collected ITS data through its advanced FAST dashboard. The FAST dashboard is a web application that monitors and presents the Las Vegas Metropolitan Freeway's traffic condition, both in real time and historically, through maps, daily peak speeds, time of day speeds, average speeds, and congestion plots from different perspectives. The FAST dashboard is deployed as a traditional, on premise hardware and uses a web server and database server as its main software components. The FAST dashboard also offers other features such as live camera feeds, video walls, dynamic message sign monitoring, variable message signs, historical incident analysis, ramp meter monitoring, historical traffic animations, and corridor plotting. Experimental features such as 3D plotting and trending travel time along corridors have also been deployed in an attempt to try to further explore and uncover the Las Vegas traffic patterns. These features are clearly very useful in understanding the traffic patterns around Las Vegas and can be used to further refine the TMC operators understanding, but they remain reactive rather than proactive. The travel time trend calculation, using a simple moving average calculation, is a proactive measure and is a simple form of prediction. A moving average is sometimes valuable to use in anomaly detection, but is often difficult to use to forecast a stochastic phenomenon such as traffic as they smooth and eliminate irregularity by design.

Based on these observations, while the NDOT/RTCSNV FAST dashboard provides very impressive visualizations and monitoring capabilities, it was found that the NDOT/RTCSNV traffic data analysis and prediction is only partially meeting the future requirements for prediction. It lacks the testing and implementation of more robust prediction algorithms such as the Random Forest algorithm used in the experimental prediction demonstration section that, while requiring more data than FAST currently collects, was able to predict traffic changes more precisely in both time and space. [Table 17](#page-92-0) includes a summary of this analysis.

Table 17. Traffic Data Analysis and Prediction Capabilities

Traffic Prediction Integration with Current Traffic Management Practice

Where does NDOT/RTCSNV want to be?

NDOT/RTCSNV would like to leverage its traffic prediction capability to implement mitigation strategies such as signal timing or ramp metering ahead of detected traffic conditions in an attempt to reduce the effect of the upcoming traffic conditions.

Where is NDOT/RTCSNV now?

NDOT/RTCSNV is currently experimenting mitigation strategies by communicating to drivers, using Variable Message Signs, the expected travel times calculated using a moving average. While this is experimental, implementation does not include mitigation strategies such as ramp metering or signal timing changes. This approach is an early attempt to mitigate traffic conditions by influencing driver behavior ahead of a forecasted change in travel time. This method relies heavily on the reaction of drivers to be effective and may not be ideal. It is not clear if ramp metering or signal timing are being initiated based on travel time prediction yet as its implementation is still experimental and its lack of precision if implemented may lead to unexpected traffic disruption. No tools on the FAST dashboard

seem to be helping the TMC operators correlate the current expected travel time with the specific location where mitigation strategies could be implemented.

Based on these observations, it was found that the NDOT/RTCSNV traffic prediction integration with TMC practices is only partially implemented. This conclusion derives directly from the fact that non-reliable, test prediction methods are currently in place and no experimentation involving the use of advanced prediction models and the correlated application of mitigation strategies to measure their effectiveness has been conducted. [Table 18](#page-93-0) includes a summary of this analysis.

Table 18. Traffic Prediction Integration with Current Traffic Management Practice

The information and data gathered from work conducted in the assessment of the 2013 FAST dataset and prediction demonstration were compared to where NDOT/RTCSNV currently is in terms of traffic data sources and networking capabilities; traffic data collection and storage capabilities; traffic data analysis and prediction capabilities; traffic prediction integration with current traffic management practice capabilities; and where NDOT/RTCSNV wants to be (potential performance). The findings were used to identify the gaps in the data collection, networking, data storage and analysis and to suggest other areas for improvement.

The gap analysis provided a foundation for measuring the investment of time, money, and human resources required to achieve the outcome of incorporating traffic prediction into the NDOT/RTCSNV TMC's decision-making process and demonstrating an ITS system that includes the ingestion and processing of ancillary data to augment its capabilities.

Results of SWOT Analysis

This section presents the findings of a SWOT analysis performed for each of the four assessment elements used in the gap analysis:

- Traffic data sources and networking capabilities
- Traffic data collection and storage
- Traffic data analysis and prediction
- Traffic prediction integration with current traffic management practice

What was learned about the NDOT/RTCSNV's current data sources and networking analysis capabilities evaluation, along with what was learned from the 2013 FAST data assessment and prediction demonstration, was used to populate each of the four quadrants from the SWOT analysis template [\(Table 14\)](#page-87-0). These four quadrants are:

- Strengths characteristics of NDOT/RTCSNV's traffic data collection, storage and analysis that provide an advantage over others.
- Weaknesses characteristics of NDOT/RTCSNV's traffic data collection, storage and analysis that place it at a disadvantage relative to others.
- Opportunities attributes of the environment (e.g., best practices, new/innovative methodologies/ technologies for collecting, transmitting, and processing data) that could help to incorporate prediction into NDOT/RTCSNV's decision-making process.
- Threats attributes of the environment (e.g., lessons learned, resource limitations, costs) that could impede the incorporation of prediction into NDOT's decision-making process.

[Table 19](#page-99-0) shows a summary of the Strength, Weakness, Opportunity, and Threat analysis.

Traffic Data Sources and Networking Capabilities

Strengths

NDOT/RTCSNV is currently operating and maintaining a network of 499 traffic sensors. This network of sensors generates about 35 million data points per year of which about 70%, or 24 million data points, is already usable for prediction analysis.

Weaknesses

Out of the 499 NDOT/RTCSNV traffic sensors, 257 are providing unusable data. There are no reliable and easily identifiable construction data sources. Most of the 257 failing sensors are legacy sensors, such as loop detectors, which have an unknown connection type and location. These sensors perform poorly and generate on average only 30% of usable traffic data. Also among the 257 sensors are sensors of various vendors using a wireless connection to the NDOT/RTCSNV'S TMC. It was observed that in comparison to traffic sensors using wired connections, wireless traffic sensors show on average a 17% loss of usable data.

Opportunities

It was found that 242 of 499 traffic sensors are providing usable data for traffic prediction models. Opportunities exist to identify and develop reliable sources of construction data to be implemented within the developed traffic prediction model. The opportunity to investigate the cause of failure for sensors using a wireless connection is also present. When investigating the distribution of wireless

connected sensor that failed the assessment, it appeared that the failing sensors were in small groups of consecutive sensors along a particular roadway.

Threats

The only solution to remedy the poor performance of the 170 unknown and legacy traffic sensors of the NDOT/RTCSNV network is to replace them with new and more reliable traffic sensors. This gain in data usability will only be possible after a significant infrastructure investment. Also, despite the currently observed performance of best performing traffic sensors, their performance will decrease over time as traffic sensors' measurements often drift over time. This drift is a possible threat to the high usability of the current traffic sensor data and each traffic sensors data needs to be monitored continuously and checked for consistency to detect early sensor drift and remedy them as quickly as possible.

Traffic Data Collection and Storage *Strengths*

The NDOT/RTCSNV ITS system is already collecting, aggregating, organizing and archiving multiple years of traffic data at a 15-minute resolution. This, in combination with the high usability of its data, allows for direct exploitation of the data without modifying or augmenting the current collection and storage system. The NDOT/RTCSNV ITS system also provides access to 1-minute resolution, live traffic data which could be leveraged quickly should the 15-minute resolution fail to provide detailed enough traffic behaviors. The NDOT/RTCSNV ITS system also collected very detailed incident data and maintains a historical incident database that could be leveraged during analysis and prediction development.

Weaknesses

While more than one year's worth of traffic data is stored on the NDOT/RTCSNV ITS system, the data is still fairly recent (2012 and up) and the amount of traffic data archived varies between sensors as they were installed over several years. The analysis and prediction success could be limited if sensors are too new to have more than one year's worth of data. There is no construction data past, present or future stored and integrated with the traffic data on the NDOT/RTCSNV ITS system. This lack of construction data is a problem when trying to develop prediction as construction is very influential to traffic behavior and needs to be distinguished from regular congestion. The NDOT/RTCSNV ITS system also stores weather data, but its resolution is rather limited.

Opportunities

The NDOT/RTCSNV ITS system could leverage existing RWIS data feeds to develop a more detailed weather dataset, which could translate to better prediction when developing forecasting models.

Construction data for Region 1 exists, but is not centralized to the knowledge of the researchers. Centralizing and integrating past, present and future construction could greatly benefit the precision of traffic prediction by allowing the prediction model to avoid identifying construction traffic as normally occurring congestion patterns and to detect nearby congestion triggered at the location of the construction zone.

The NDOT/RTCSNV ITS system is based on a traditional relational database and flat file storage architecture and could easily be augmented by additional resources to handle these additional datasets. Should the data collected exceed the capacity of the current on premise hardware, NDOT/RTCSNV could leverage the current offering of cloud storage services to reduce the cost of system augmentation.

Threats

While augmentation of storage can easily be accommodated, database scalability may be more costly and challenging. The addition of new data may require the need for better performing database and lead to more expensive hardware and licenses. Also the integration of new sets of data such as construction and weather could require a data model redesign and a significant impact on the current operation of the FAST database.

Traffic Data Analysis and Prediction

Strengths

The NDOT/RTCSNV is currently leveraging the prediction usability of its collected traffic data through the use of advanced visualizations and analysis within the FAST dashboard. The FAST dashboard offers complete data analysis capabilities of the NDOT/RTCSNV ITS system, including traffic data, incident data, real-time and historical plots, and video as well as experimental features such as, 3D plots and expected travel time calculation. This accomplishment demonstrates a strong intent on behalf of the NDOT/RTCSNV to explore prediction capabilities.

Weaknesses

Despite the strong NDOT/RTCSNV intent, it's ITS system still lacks data to effectively develop prediction models. As shown in this study, traffic data alone is not sufficient to develop effective traffic prediction models and additional data such as construction data, weather data and event attendance data, or any other correlated ancillary data, could be tracked, collected, and leveraged to develop reliable traffic prediction models. The NDOT/RTCSNV ITS system does not integrate such data and does not implement them as part of its travel time calculation. Also the NDOT/RTCSNV is using simple calculation such as moving average to establish expected travel time, while this method is common, it often generates misleading results as it is smoothing the stochastic nature of traffic.

Opportunities

The NDOT/RTCSNV could try to leverage newly created web prediction services to help with the adoption of more complex prediction algorithms as part of its ITS management and decision making process. Cloud services such as the Google Prediction API¹⁹ and BigML²⁰ machine learning services could be explored and evaluated at limited cost before investing in larger more costly prediction solutions such as SAS²¹. The existing relationship between the NDOT/RTCSNV and the Las Vegas Casino Regional Commission should be explored to obtain and gather historical and upcoming event data and their expected and previous attendance. NDOT's RWIS stations could also be leveraged to collect and integrate more detailed historical weather data.

Threats

Prediction algorithms, or cloud prediction services, sometimes over fit their models to the data they are trained on and subsequently generate unreliable predictions. The expertise of professional prediction model developers may be needed to achieve reasonable prediction models. This will of course increase the cost of the prediction model development, but does not necessarily guarantee success. Often to find a suitable prediction model, multiple models are developed in parallel using various approaches and ranked based on a set of test datasets to determine which is best. This method while effective often necessitates more time and additional hardware to run the multiple model trainings in parallel. Another possible outcome is that the prediction model, while successful, may not bring any additional value to the NDOT/RTCSNV institutional knowledge by matching what is already known from TMC operators of the road network behavior.

Traffic Prediction Integration with Current Traffic Management Practice *Strengths*

The NDOT/RTCSNV has shown a desire to incorporate traffic prediction as part of its traffic management operations and explore proactive traffic management techniques by publishing its experimental travel time calculation to drivers through variable message sign displays.

Weaknesses

 \overline{a}

Variable message sign displays may be efficient at relaying information to drivers, but they can't control a driver's behavior as well as ramp meters or signal timings would. Drivers may ignore message signs and reduce the efficiency of the mitigation strategy.

²¹ SAS Institute Inc. (n.d.). *Predictive Analytics*. Retrieved from SAS:

¹⁹ Google . (2014). *Google Prediction API*. Retrieved from Google Cloud Platform: https://cloud.google.com/prediction/docs

²⁰ BigML. (2015). *About Us*. Retrieved from BigML: https://bigml.com/

http://www.sas.com/hu_hu/insights/analytics/predictive-analytics.html

Opportunities

In addition to variable message sign announcements, mitigation strategies such as early ramp metering or signal timing changes at the proximity of the highway entrances upstream from the traffic conditions and triggered as a response to a less than desirable traffic prediction should be explored and their efficiency assessed. Innovative approaches that do not directly affect the roadway system could also be explored; mitigation strategies such as collaborating with hotels, casinos and conference centers to try to delay or advance drivers departures may also be viable and effective options. Funding for such exploratory work could be obtained from a Federal Highway Administration's (FHWA) Technology and Innovative Deployment Program, which funds efforts to accelerate the implementation and delivery of new innovations and technologies that result from highway research and development to benefit all aspects of highway transportation.²²

Threats

Should the implementation of mitigations strategies, in response to traffic predictions, lead to poor results or unexpected complications; there will be a loss of credibility in the eyes of the public and public officials.

[Table 19](#page-99-0) summarizes the findings of the SWOT analysis.

 \overline{a}

²² *Technology and Innovation Deployment Program*. (2013). Retrieved from FHWA: http://www.fhwa.dot.gov/map21/factsheets/tid.cfm

Table 19. Summary of SWOT Analysis

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

This research was intended to assess and explore the potential of currently deployed ITS infrastructure within the FAST program in the Las Vegas metropolitan area to support proactive traffic management.

Four primary tasks were completed to complete this study including:

- Described the NDOT ITS architecture and traffic data.
- Assess NDOT ITS architecture and the traffic data generated.
- Demonstrate prediction capability using NDOT data.
- Perform gap analysis and Strength/Weakness/Opportunities/Threats (SWOT) analysis.

The study revealed that there is potential for proactive traffic management in the Las Vegas Resort Corridor given the rich number of sensors deployed, the quality of the data generated from these sensors, and the type of information that can be generated from the sensors. Detailed information was provided regarding areas of the system that could benefit from improvements in communications to help further improve the usefulness of the available sensors to support proactive traffic management. For example, it was noted that in some cases, wireless communication technology does not appear to be consistently able to support sensor networks and should be further investigated to improve overall sensor system performance. In addition, it was noted that additional ancillary information such as construction activity tied to sensor locations would be helpful to help isolate periods of reporting that are currently void of any information. For example, sensor 350.1.158 located on I-15 southbound near Spring Mountain Road and Decatur Road appears to have been offline for the first three weeks of January 2013, however, without specific information related to construction zone activity, it makes it difficult to isolate these dark periods, which can skew results of the traffic prediction model. Additional ancillary information that can further enhance modeling efforts to improve traffic prediction include:

- Data for less aggregated time periods for example, if data for 1-minute or 5-minute periods could be obtained, predictions can be improved to similar time interval windows.
- Event information given the nature of the Las Vegas Resort Corridor, any information related to regularly scheduled events could help to improve traffic prediction.
- Detailed weather information from RWIS detailed RWIS data from multiple stations across the Las Vegas area to the California border could help to improve traffic prediction.

In total, the project findings support the use of traffic prediction to provide proactive traffic capabilities to the FAST program. Next, specific recommendations are made to assist NDOT in its decision-making process regarding proactive traffic management.

Recommendations

The following are recommendation of additional work that NDOT could perform to further move it's ITS infrastructure toward the integration and proactive use of prediction technologies.

Traffic Data Sources and Networking Capabilities

- NDOT could perform a similar assessment using a year or more of its I-80 traffic data and extend its assessment of its traffic data usability outside of the Las Vegas area.
- NDOT could create a consolidated database containing the historical and planned construction information necessary for prediction model development.
- NDOT could develop a detailed historical weather dataset intended for prediction development based on the content of its RWIS database.
- NDOT could modify the assessment traffic sensors performed in this study to rapidly and more effectively pre-qualify new traffic detection hardware.
- NDOT/RTCSNV could leverage the speed/volume diagram, completeness diagram and calendar heatmap to assess the need for complex traffic sensor or predict model development at each of the 499 sensor location. This data-driven acquisition or investment approach could allow NDOT/RTCSNV to focus funding sooner on the most vulnerable part of its network.
- NDOT could modify the assessment approach adopted in this study to analyze real-time (1 minute) traffic sensors and detect anomalies and early drifting of the road sensors from historically observed behaviors. This data-driven approach to traffic sensor maintenance could allow resources and funding to be targeted quickly to the misbehaving sensors. This would help NDOT maintain a more constant traffic data quality and usability across its entire network.

Traffic Data Collection and Storage

 NDOT/RTCSNV could explore the use of cloud storage and analytical capabilities as an alternative to storing compressed files on premise. This would increase storage capacity at equivalent cost while allowing for immediate accessibility and analysis of the data.

Traffic Data Analysis and Prediction

 NDOT/RTCSNV could develop a traffic prediction project focusing on a small set of traffic sensors of interest. This prediction project could leverage existing RWIS, construction and events attendance data. If possible, several years' worth of data should be used to develop the prediction to help prediction algorithms differentiate recurring traffic patterns from anomalies, such as construction and incidents. Prediction models generated could be designed to produce either regression (predicted speed, occupancy or volume) or classification (predicted traffic phase) outcomes. The usability and performance of each of these models could then be assessed. Additionally, historical incident data could be added to prediction models to further

refine its ability to forecast traffic conditions when an incident is already occurring at a specific location on the road network.

Traffic Prediction Integration with Current Traffic Management Practice

- Following the success of a reliable traffic prediction model, NDOT could investigate the usability of prediction outcomes to support TMC activities planning using visualizations or tools such as the ones suggested in this study.
- NDOT could then investigate the effectiveness of various TMC congestion mitigation strategies, such as ramp-meters and signal timing plans, in response to prediction outcomes and fine tune the implementation and timing of such countermeasures.

APPENDIX A

Completeness Assessment of 2013 FAST Dataset

APPENDIX B

Calendar Heatmaps of 2013 FAST Dataset

Speed Time−Series Calendar Heatmap for Roadway ID: 7

Speed Time−Series Calendar Heatmap for Roadway ID: 43

Speed Time−Series Calendar Heatmap for Roadway ID: 77

Speed Time−Series Calendar Heatmap for Roadway ID: 132

Speed Time−Series Calendar Heatmap for Roadway ID: 197

Speed Time−Series Calendar Heatmap for Roadway ID: 347

Speed Time−Series Calendar Heatmap for Roadway ID: 391

Speed Time−Series Calendar Heatmap for Roadway ID: 392

Speed Time−Series Calendar Heatmap for Roadway ID: 394

Speed Time−Series Calendar Heatmap for Roadway ID: 395

Speed Time−Series Calendar Heatmap for Roadway ID: 399

Speed Time−Series Calendar Heatmap for Roadway ID: 400

Speed Time−Series Calendar Heatmap for Roadway ID: 402

Speed Time−Series Calendar Heatmap for Roadway ID: 403

Speed Time−Series Calendar Heatmap for Roadway ID: 418

Speed Time−Series Calendar Heatmap for Roadway ID: 436

Speed Time−Series Calendar Heatmap for Roadway ID: 531

Speed Time−Series Calendar Heatmap for Roadway ID: 541

Speed Time−Series Calendar Heatmap for Roadway ID: 542

Speed Time−Series Calendar Heatmap for Roadway ID: 543

Speed Time−Series Calendar Heatmap for Roadway ID: 562

Speed Time−Series Calendar Heatmap for Roadway ID: 583

Speed Time−Series Calendar Heatmap for Roadway ID: 584

APPENDIX C Hexagonal Binning Speed-Flow Plots of 2013 FAST Dataset

