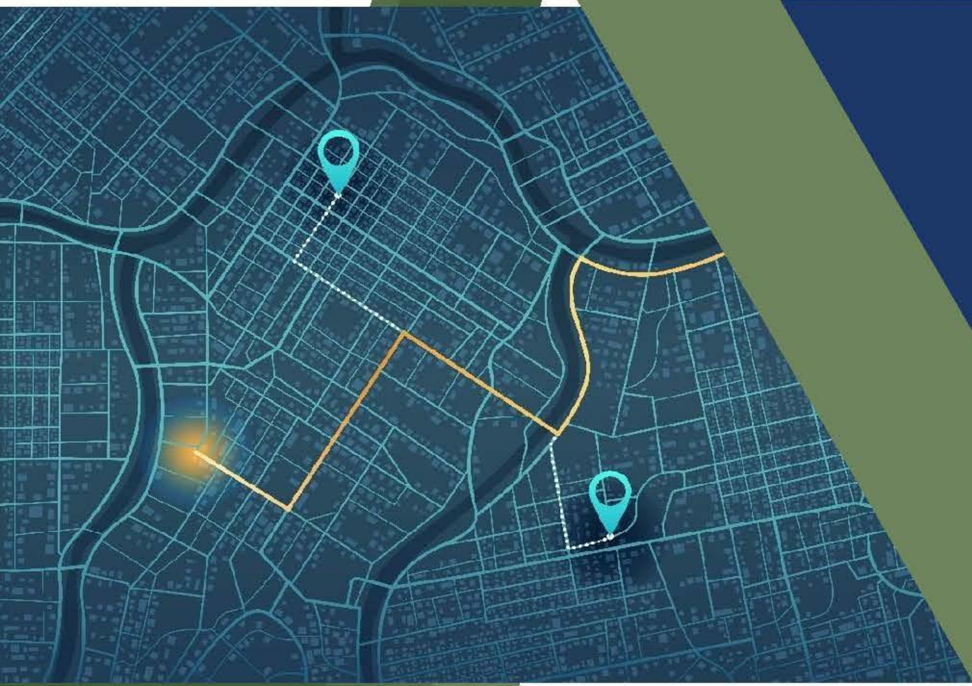

FINAL
REPORT

PROJECT J2

APRIL 2023

Real-Time Data-Based Decision Support System for Arterial Traffic Management

Mohammed Hadi, Ph.D., Florida International University
Lily Elefteriadou, Ph.D., University of Florida
Hector Mata, Ph.D., Florida International University
Ehsan Amini, Ph.D., University of Florida

STRIDE | Southeastern Transportation Research, Innovation, Development and Education Center

UF | Transportation Institute UNIVERSITY of FLORIDA

TECHNICAL REPORT DOCUMENTATION PAGE

<table>
<tr><td>1. Report No.
Project J2</td><td>2. Government Accession No.</td><td>3. Recipient's Catalog No.</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
Real-Time Data-Based Decision Support System for Arterial Traffic Management</td><td>5. Report Date
4/20/2023</td></tr>
<tr><td>6. Performing Organization Code</td></tr>
<tr><td colspan="2">7. Author(s)
Mohammed Hadi, Ph.D., Florida International University
Lily Elefteriadou, Ph.D., University of Florida
Hector Mata, Ph.D., Florida International University
Ehsan Amini, Ph.D., University of Florida</td><td>8. Performing Organization Report No.
STRIDE Project J2</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address

Florida International University, Department of Civil and Environmental Engineering, 10555 W Flagler Street, EC 3680 Miami, FL 33174

University of Florida, Department of Civil and Coastal Engineering, PO Box 116580, Gainesville, FL 32611</td><td>10. Work Unit No.</td></tr>
<tr><td>11. Contract or Grant No.
Funding Agreement Number 69A3551747104</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address

University of Florida Transportation Institute/ Southeastern Transportation Research, Innovation, Development and Education Center (STRIDE) 365 Weil Hall, P.O. Box 116580 Gainesville, FL 32611

U.S Department of Transportation/Office of Research, Development & Tech
1200 New Jersey Avenue, SE, Washington, DC 20590</td><td>13. Type of Report and Period Covered
8/15/2018 to 4/20/2023</td></tr>
<tr><td>14. Sponsoring Agency Code</td></tr>
<tr><td colspan="3">15. Supplementary Notes N/A</td></tr>
<tr><td colspan="3">16. Abstract - This study developed and evaluated a TRPS strategy based on supervised and unsupervised machine learning combined with signal timing optimization to address the issues with traditional TRPS. The strategy fills an important gap in providing a proactive traffic control that makes use of Automated Traffic Signal Performance Measures (ATSPM) measures-based data that are becoming available sources, including high-resolution controller data. This study also explored a methodology and evaluated multiple algorithms for the short-term prediction of the traffic state for the next half an hour. The results revealed that the ANN algorithm produced the best results in terms of accuracy and areas under the curve. Thus, this study used the ANN prediction model in the remaining task to implement and evaluate the prediction to support the activation of the signal timing plans in real-time operations. The study then assessed the performance of the predictive TRPS based on clustering and prediction by evaluating five different scenarios of signal timing plan selection. The results of the project case study showed that the predictive TRPS method can decrease the travel time by 7 percent compared to existing traffic signals, 4% compared to optimizing for a fixed signal timing plan based on a signature day for the whole database, and 17% compared to optimizing signal timing for a random day in the data. This shows that the TRPS based on traffic pattern identification and prediction has the potential to improve traffic performance compared to other assessed optimization scenarios.</td></tr>
<tr><td colspan="2">17. Key Words
arterials, traffic management, signal timing, optimization</td><td>18. Distribution Statement
No restrictions</td></tr>
</table>

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
N/A	N/A	56 pages	N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

DISCLAIMER

ACKNOWLEDGEMENT OF SPONSORSHIP AND STAKEHOLDERS

This work was sponsored by a grant from the Southeastern Transportation Research, Innovation, Development, and Education Center (STRIDE) at the University of Florida. The STRIDE center is funded through the U.S. Department of Transportation's University Transportation Centers Program. Additional funds were provided by the Alabama Department of Transportation (ALDOT).

LIST OF AUTHORS

Lead PI:

1. **Mohammed Hadi, Ph.D., PE**
   Professor, Department of Civil & Environmental Engineering
   Florida International University
   (305) 348-0092 | hadim@fiu.edu
   ORCID # 0000-0003-2233-8283

Co-PI:

1. **Lily Elefteriadou, Ph.D.**
   Professor, Department of Civil, Construction and Environmental Engineering
   University of Florida
   (321) 298-4360 | elefter@cd.ufl.edu
   ORCID # 0000-0003-4045-3365

Additional Researchers:

1. ***Hector D. Mata***
   Graduate Student
   Department of Civil and Environment Engineering
   Florida International University
   (701) 400-9637 | hmata010@fiu.edu
2. ***Ehsan Amini***
   Graduate Student
   Department of Civil and Environment Engineering
   University of Florida
   786-479-5724 | amini.e@ufl.edu

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ACS-Lite	Adaptive Control Software Lite
AFD	Arterial Fundamental Diagram
ALDOT	Alabama Department of Transportation
ANN	Artificial Neural Network
ATSPM	Automated Traffic Signal Performance Measures
AUC	Area Under the Curve
BIC	Bayesian Information Criteria
CLARA	Clustering Large Applications
CNN	Convolutional Neural Network
DBI	Davies-Bouldin Index
DI	Dunn Index
DOT	Department of Transportation
DT	Decision Tree
EB	Eastbound
EMD	Empirical Mode Decomposition
FDOT	Florida Department of Transportation
FHWA	Federal Highway Administration
FPR	False Positive Rate
GMM	General Mixture Model
GNB	Gaussian Naïve Bayes
GOR	Green Occupancy Ratio
HCM	Highway Capacity Manual
HCS	Highway Capacity Software
HRC	High Resolution Controller Data
ITS	Intelligent Transportation Systems
KNN	K-Nearest Neighbors
LA-ATCS	Los Angeles Adaptive Traffic Control System
LOS	Level of Service
LSTM	Long-Short Term Memory Algorithm
MLR	Multinomial Logistic Regression
MPE	Mean Percentage Error
OD	Origin-Destination Matrix
OPAC	Optimized Policies for Adaptive Control
PCR	Precision-Recall Curve
QOD	Queue Over Detector
RF	Random Forest
RHODES	Real-Time Hierarchical Optimized Distributed Effective System
RITIS	Regional Integrated Transportation Information System
RMSE	Root Mean Square Error
ROC	Receiving Operating Characteristic

RPART	Recursive Partitioning and Regression Tree
SCATS	Sydney Coordination Adaptive Traffic System
SCOOT	Split Cycle Offset Optimization Technique
SDD	Speed Standard Deviation
STAMP	Statewide Arterial Management Program
SUR	Split Utilization Ratio
SVC	Support Vector Classification
TAT	Traffic Analysis Toolbox
TDIDT	Top-Down Induction Decision Tree Algorithm
TOD	Time-of-Day
TPR	True Positive Rate
TRPS	Traffic Responsive Plan Selection
TSC	Traffic Signal Control
TSM&O	Transportation System Management and Operations
TT	Travel Time
TTC	Time to Collision
VI	Volatility Index
WB	Westbound
WCSS	Within Clusters Sum of Squares

ABSTRACT

Although adaptive signal control is a powerful strategy to address the day-to-day variation in traffic demands, most intersections in the United States are still operating under time-of-day (TOD) strategies due to the high cost and the additional requirements associated with the systems. In addition, adaptive signal control may not be beneficial to address all operation performance issues. Traffic responsive plan selection (TRPS) strategies have been proposed since the 1970s as an alternative to TOD that can address some of the issues associated with day-to-day variations in traffic patterns. The requirements of these strategies are much lower than those of adaptive signal control strategies. However, there are several limitations and issues associated with TRPS that have limited the adoption of these strategies in This study developed and evaluated a TRPS strategy based on supervised and unsupervised machine learning combined with signal timing optimization to addresses the issues with traditional TRPS. The strategy fills an important gap in providing a proactive traffic control that makes use of Automated Traffic Signal Performance Measures (ATSPM) measures-based data that are becoming available sources including high resolution controller data.

This study also explored a methodology and evaluated multiple algorithms for the short-term prediction of the traffic state for the next half an hour. The traffic states are predicted as belonging to one of the three clusters identified based on the results of the cluster analysis. This prediction in real-time operations can be used to activate the signal timing plan developed for the signature day for the cluster that represent the predicted state. The results revealed that the ANN algorithm, produced the best results in terms of accuracy and areas under the curve. Thus, this study used the ANN prediction model in the remaining task for the implementation and evaluation of the prediction to support the activation of the signal timing plans in real-time operations.

The study then assessed the performance of the predictive TRPS based on clustering and prediction by evaluating five different scenarios of signal timing plan selection. *The results for the project case study showed that the predictive TRPS method can decrease the travel time by 7 percent compared to existing traffic signals, 4% compared to optimizing for a fixed signal timing plan based on a signature day for the whole database, and 17% compared to optimizing signal timing for a random day in the data. This shows that the TRPS based on traffic pattern identification and prediction has the potential of improving traffic performance compared to other assessed optimization scenarios.*

EXECUTIVE SUMMARY

BACKGROUND

In most cases, traffic signal management agencies have used predetermined plans that are changed based on time-of-day (TOD) and day-of-week (mostly weekday vs. weekend) schedules. The TOD plans, sometime referred to as pre-timed plans, are selected for each period (e.g., a.m. peak, p.m. peak, or off-peak) plan using signal timing optimization tools combined with fine-tuning of signal timing based on field observations. In most cases, one plan is developed for each peak period and used throughout the year, based on very limited amount of data, although it is possible to implement different signal timing plans for different seasons. The assumption is that similar traffic patterns generally occur during the same times each day. TOD plans do not work well when there are large day-to-day variations in traffic conditions through the year due to demand variations and events like incidents and adverse weather.

Traffic responsive plan selection (TRPS) and traffic adaptive systems are two types of signal control strategies that have been used to address the day-to-day variation in traffic patterns. TRPS involves the real-time selection of timing plans from a library of pre-stored plans that are developed off-line based on traffic measurements, rather than based on TOD.

In an early work (Hadi 1990), identified several issues with traditional TRPS including:

- The need for the near-term prediction of traffic flow parameters for use in the plan selection rather than using traffic flow parameters that may change in the next period,
- The difficulty in designing the plans to be stored in the TRPS plan library,
- The need for a method to weight the data collected from different detectors,
- The need for a method to set the plan activation thresholds,
- The need for installing and maintaining additional detectors, and
- The need to limit the number of plans switching in a peak period to reduce the delays due to the transition interval between the plans.

In recent years, data have started to be available from multiple new sources including high resolution controller data, advanced detection technologies such as microwave detectors and video image detection, automatic-vehicle based identification technologies such as those based on Bluetooth readers, third party crowdsourcing data, and connected automated vehicles data. Many agencies have started to use the data from these systems to estimate what is referred to as the Automated Traffic Signal Performance Measures (ATSPMs). ATSPMs is defined by the FHWA as “a suite of performance measures, data collection and data analysis tools to support objectives and performance-based approaches to traffic signal operations, maintenance, management and design to improve the safety, mobility and efficiency of signalized intersections for all users” (FHWA 2022b). The ATSPMs include several performance measures, some of which can be used as inputs to a new generation of TRPS to allow the implementation of better and more cost-effective strategies and signal control plans including better setting of the plans and better activation of plans in real-time operations based on traffic responsive and adaptive strategies. Supervised and non-supervised data mining/machine learning algorithms combined with the ATSPMs will allow the application of pro-active traffic control strategies that are based on predicted traffic conditions in real-time operations. Such strategies will take advantage of ATSPM measures that were not available for traditional TRPS.

GOAL AND OBJECTIVES

The goal of the study is to develop and evaluate a proactive TRPS strategy that use ATSPM measures as input and select the signal timing plans for implementation based on traffic flow parameters predicted for the near-term future. The specific objectives are:

- Develop and assess methods to categorize the traffic conditions in a peak period in traffic patterns that best represent the day-to-day variations in traffic flow parameters.
- Identify a method to select signature days that best represent the identified patterns for use when optimizing the signal control plans that will be stored in the signal timing plan library as part of the TRPS implementation
- Develop and assess a model for near-term prediction of traffic patterns in real-time operations for use in TRPS
- Determine the benefits of the proactive strategies developed in this study.

The proposed approach will address the issues with traditional TRPS as follows:

- The need for the near-term prediction: This study will investigate various machine learning techniques for short term prediction of traffic conditions for use to activate the plans in TRPS.
- The difficulty in designing the plans to be stored in the TRPS plan library: The developed method will identify signature days that represent the traffic patterns in the network considering the day-to-day variation in traffic flow parameters throughout the year.
- The need for a method to set the weights on the detectors: The developed method will implicitly consider the importance of the data measurements from each detector when categorizing and predicting the traffic states based on traffic measurements using machine learning.
- The need for a method to set the activation thresholds: The developed method will associate a signal timing plan with each identified traffic pattern eliminating the need for setting the thresholds.
- The need for installing and maintaining additional detectors: The developed method will use performance measures estimated as part of the ASTPMs
- The need to limit the number of transitions between the plans: The developed system will limit the number of transitions between plans to one or two.

METHODOLOGY OVERVIEW

To accomplish the goal and objectives, this study uses advanced machine learning approach combined with signal timing optimization models based on ATSPM data to select the best timing plan to develop and activate in a proactive TRPS framework.

The proposed methodology involves first the categorization of peak period traffic conditions throughout the year using cluster analysis. The next step is to identify a signature (representative) day for each cluster to use in developing the signal timing plan. Then, models are developed using various data mining/machine learning techniques for short-term prediction of traffic flow for use to support signal timing plan activation in real-time operations. Finally, traffic analysis models are used to assess the performance of the proposed TRPS compared to other strategies. Below is an overview of each of these steps.

Cluster Analysis: Cluster analysis was implemented to identify traffic patterns that are representative of the traffic conditions, considering the variations in the day-to-day variations in traffic conditions throughout the years. The goal of the clustering algorithm was to categorize the days with similar traffic patterns within the analysis period (the AM peak). A proven and widely used clustering algorithm referred to as the k-means algorithm was used in this study to produce the clusters. The inputs to the clustering algorithm included vehicle counts, travel times, Green Occupancy Ratio (GOR), and Signal Utilization Ratio (SUR); aggregated at 15-minute intervals.

Identification of the Signature Days: The next step is to identify the signature day for each cluster. The signature day is identified as the best day that represents the traffic conditions. The methodology to accomplish this is the based on that proposed in the Traffic Analysis Toolbox Volume III (Wunderlich 2019). In order to identify the signature day for each cluster, a 15-minute profile analysis was performed across all days considering all travel conditions (clusters), at multiple locations (intersections) for the key measures. The algorithm implemented to find the signature day can be find in Chapter 4.

Development of the Signature Days: As stated earlier, the next step in the methodology is to use the results of the cluster analysis as inputs to a data analytic-based prediction model that uses data mining/machine learning to predict the traffic state in the short-term future in real-time operations. The investigated prediction models are classification algorithms that predict the traffic state in the next 30 minutes as belonging to one of the pre-identified clusters. For that effect, a 15-minute data across all days at multiple locations (intersections) is used as input for the prediction algorithms. The study implemented and evaluated seven different algorithms data mining/machine learning algorithms. The evaluated algorithms are the Decision Tree (DT), Random Forest (RF), Gaussian Naïve Bayes (GNB), Multinomial Logistic Regression (MLR), Support Vector Classification (SVC), K-nearest neighbors (KN), and Artificial Neural Network (ANN).

Implementation and Assessment of the Developed TRPS: This study used five scenarios to investigate the improvement in system performance due to the TRPS strategy based on the traffic patterns identification. In each scenario, the utilized traffic signal timing is optimized for different traffic patterns. The performance of the timing plans are assessed based on their performance for ten days randomly selected for use in the evaluation

RESULTS

The methodology of this study was applied to an arterial located in South Florida. This location had the advantage of having data available from different sources. This helped with the implementation of the clustering procedure and facilitated the identification of three signature days that clearly represent the traffic states at that location during the AM peak. The three states represent relatively low, medium, and heavy volumes that can be used as inputs to signal optimization models to identify signal timing plans that can be used as the plans to select from in systems that use the TRPS control. Further examination of the resulting volumes, plans, and the resulting performance can be done as described in Chapter 5 to determine if the resulting plans are significantly different to justify utilizing all of them in TRPS control. In some cases, for example, it may be determined that only two of the three plans can be justified for this purpose.

This study also explored a methodology and evaluated multiple algorithms for the short-term prediction of the traffic state for the next half an hour. The traffic states are predicted as belonging to one of the three clusters identified based on the results of the cluster analysis. This prediction in real-time operations can be used to activate the signal timing plan developed for the signature day for the cluster that represent the predicted state. The results revealed that the ANN algorithm, produced the best results in terms of accuracy and areas under the curve. Thus, this study used the ANN prediction model in the remaining task for the implementation and evaluation of the prediction to support the activation of the signal timing plans in real-time operations.

The study then assessed the performance of the predictive TRPS based on clustering and prediction by evaluating five different scenarios of signal timing plan selection. The results for the project case study showed that the predictive TRPS method can decrease the travel time by 7 percent compared to existing traffic signals, 4% compared to optimizing for a fixed signal timing plan based on a signature day for the whole database, and 17% compared to optimizing signal timing for a random day in the data. This shows that the TRPS based on traffic pattern identification and prediction has the potential of improving traffic performance compared to other assessed optimization scenarios.

REFERENCES

Hadi, M. "Improved Strategies for Traffic Responsive Control in Arterial Signal Systems," Ph.D. Dissertation,. Gainesville, FL, 1990.

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND

Traffic congestion along arterial streets is increasingly becoming a critical issue that needs to be addressed by transportation agencies. Compared to the relatively mature management of freeways, arterial traffic operations and management are lagging behind. To address such a gap, a strong emphasis has been placed on arterial traffic management at the national and state levels. For example, the Federal Highway Administration (FHWA) has developed an Arterial Management Program that is dedicated to the reduction of recurring congestion along arterial streets (Federal Highway Administration 2022). The Transportation Systems Management and Operations (TSM&O) program of the Florida Department of Transportation (DOT) initiated the Statewide Arterial Management Program (STAMP) to improve the mobility of arterial transportation system (Florida Department of Transportation 2021).

The selection of signal timing control strategies and the setting of signal control plan parameters play a crucial role in determining the network performance The signal management agencies select these strategies and plans to improve performance measures such as delays, queues, throughputs. In most cases, traffic signal management agencies have used predetermined plans that are changed based on time-of-day (TOD) and day-of-week (mostly weekday vs. weekend) schedules. The TOD plans, sometime referred to as pre-timed plans, are selected for each period (e.g., a.m. peak, a p.m. peak, or off-peak) plan using signal timing optimization tools combined with fine-tuning of signal timing based on field observations. In most cases, one plan is developed for each peak period and used throughout the year, based on very limited amount of data, although it is possible to implement different signal timing plans for different seasons. The assumption is that similar traffic patterns generally occur during the same times each day. TOD plans do not work well when there are large day-to-day variations in traffic conditions through the year due to demand variations and events like incidents and adverse weather.

Traffic responsive plan selection (TRPS) and traffic adaptive systems are two types of signal control strategies that have been used to address the issue mentioned above with TOD plan activation. Traffic responsive signal control involves the selection of timing plans from a library of pre-stored plans that are developed off-line based on data collected in real-time, rather than based on TOD. The interval for the reselection of the plans can be set by the user (e.g., every 15 minute, 30 minutes, one hour, etc.) Traditionally, the utilized data in traffic responsive control includes volume and/or occupancy measurements from a limited number of system detectors (advance detectors or departure-side detectors) that are assumed to provide data that reflect the traffic conditions in the network. The data from the detectors can be weighted, although there is no good guidance of how to set these weights. Different signal system vendors use different algorithms for TRPS. Still, there are no good methodologies implemented in practice for setting the thresholds to switch between the traffic plans based on detector data. It has been reported that when setting up the TRPS, considerable effort is needed to identify the best locations for the vehicle detectors, set the parameter values and weights associated with those detectors, establish appropriate thresholds and associated plans, and fine tune the system parameters (Federal Highway Administration 2015).

In an early work, (Hadi 1990) identified several issues with traditional TRPS including:

- The need for the near-term prediction of traffic flow parameters for use in the plan Selection rather than using traffic flow parameters that may change in the next period,
- The difficulty in designing the plans to be stored in the TRPS plan library,
- The need for a method to weight the data collected from different detectors,
- The need for a method to set the plan activation thresholds,
- The need for installing and maintaining additional detectors, and
- The need to limit the number of plans switching in a peak period to reduce the delays due to the transition interval between the plans.

Adaptive traffic signal control systems have been developed as a more powerful signal control strategy compared to TOD and TRPS strategies. These systems involve the use of data from much larger numbers of detectors compared to TRPS that provide traffic information for most segments and/or turn movements in a network. The adaptive system adjusts the signal control parameters in real-time at short time intervals (e.g., every phase and/or every cycle). This allows the signal timing to adapt to short term as well as long term fluctuation in traffic flow parameters, resulting in improvement in system performance. The systems generally implement frequent but small changes in signal timing parameters, thus avoiding the delays due to transitions between plans that are significantly different. However, adaptive systems require high level of reliability and accuracy of the detection system. Most adaptive systems are not proactive in that they do not predict near-time changes in traffic conditions. Some adaptive systems may be also slow in recognizing and reacting to high surges in demands and drops in capacity such as during special events and incidents. Many of these systems are optimized and fine-tuned to adapt to typical changes in traffic patterns and cannot adequately react to other conditions.

Although adaptive signal control is a powerful strategy, most intersections in the United States are still operating under TOD strategies. Adaptive signal control is much more expensive to implement than TOD and TRPS. Thus, if a TRPS strategy can be developed and implemented to addresses the issues with traditional TRPS, such strategy can fill an important gap in providing a proactive traffic control strategies, particularly for those locations that do not have adaptive signal control.

In recent years, data have started to be available from multiple new sources collection sources including high resolution controller data, advanced detection technologies such as microwave detectors and video image detection, automatic vehicle based identification technologies such as those based on Bluetooth readers, third party crowdsourcing data, and connected automated vehicles data. Many agencies have started to use the data from these systems to estimate what is referred to as the Automated Traffic Signal Performance Measures (ATSPMs). ATSPMs is defined by the FHWA as "a suite of performance measures, data collection and data analysis tools to support objectives and performance based approaches to traffic signal operations, maintenance, management and design to improve the safety, mobility and efficiency of signalized intersections for all users" (US Department of Transportation 2022). The ATSPMs include several performance measures, some of which can be used as inputs to a new generation of can be used as inputs to TRPS to allow the implementation of better and more cost-effective strategies and signal control plans including better setting of the plans and better activation of plans in real-time operations based on traffic responsive and adaptive strategies. Supervised and non-supervised data mining/machine learning algorithms combined with the ATSPMs will allow the application of pro-active traffic control strategies that are based on predicted traffic conditions in real-time operations. Such strategies will take advantage of ATSPM measures that were not available for traditional TRPS.

1.2 GOAL AND OBJECTIVES

The goal of the study is to develop and evaluate a proactive TRPS strategy that use ATSPM measures as input and select the signal timing plans for implementation based on traffic flow parameters predicted for the near-term future. To accomplish this goal, the researchers will use an advanced machine learning approach combined with signal timing optimization models to capture the dynamic changes in arterial network and select the best timing plan to develop and activate. The specific objectives are to:

- Develop and assess methods to categorize the traffic conditions in a peak period in traffic patterns that best represent the day-to-day variations in traffic flow parameters.
- Identify a method to select signature days that best represent the identified patterns for use when optimizing the signal control plans that will be stored in the signal timing plan library as part of the TRPS implementation
- Develop and assess a model for near-term prediction of traffic patterns in real-time operations for use in TRPS
- Determine the benefits of the proactive strategies developed in this study.

The proposed approach will address the issues with traditional TRPS identified by (Hadi 1990) (See Section 1.1 above), as listed below.

- The need for the near-term prediction: This study will investigate various machine learning techniques for short term prediction of traffic conditions for use to activate the plans in TRPS.
- The difficulty in designing the plans to be stored in the TRPS plan library: The developed method will identify signature days that represent the traffic patterns in the network considering the day-to-day variation in traffic flow parameters throughout the year.
- The need for a method to set the weights on the detectors: The developed method will implicitly consider the importance of the data measurements from each detector when categorizing and predicting the traffic states based on traffic measurements using machine learning.
- The need for a method to set the activation thresholds: The developed method will associate a signal timing plan with each identified traffic pattern eliminating the need for setting the thresholds.
- The need for installing and maintaining additional detectors: The developed method will use performance measures estimated as part of the ASTPMs
- The need to limit the number of transitions between the plans: The developed system will limit the number of transitions between plans to one or two.

1.4 METHODOLOGY OVERVIEW

As stated in Section 1.3, to accomplish the goal and objectives, this study uses advanced machine learning approach combined with signal timing optimization models based on ATSPM data to select the best timing plan to develop and activate in a proactive TRPS framework. This section provides an overview of the study methodology. More details about the methodology can be found in Chapters 4 and 5.

The proposed methodology involves first the categorization of peak period traffic conditions throughout the year using cluster analysis. The next step is to identify a signature (representative) day for each cluster to use in developing the signal timing plan. Then, models are developed using various data mining/machine learning techniques for short-term prediction of traffic flow for use to support signal timing plan activation in real-time operations. Finally, traffic analysis models are used to assess the performance of the proposed TRPS compared to other strategies. Below is an overview of each of these steps.

Cluster Analysis: Cluster analysis was implemented to identify traffic patterns that are representative of the traffic conditions, considering the variations in the day-to-day variations in traffic conditions throughout the years. The goal of the clustering algorithm was to categorize the days with similar traffic patterns within the analysis period (the AM peak). A proven and widely used clustering algorithm referred to as the k-means algorithm was used in this study to produce the clusters. The inputs to the clustering algorithm included vehicle counts, travel times, Green Occupancy Ratio (GOR), and Signal Utilization Ratio (SUR); aggregated at 15-minute intervals.

Identification of the Signature Days: The next step is to identify the signature day for each cluster. The signature day is identified as the best day that represents the traffic conditions. The methodology to accomplish this is the based on that proposed in the Traffic Analysis Toolbox Volume III (Wunderlich 2019). In order to identify the signature day for each cluster, a 15-minute profile analysis was performed across all days considering all travel conditions (clusters), at multiple locations (intersections) for the key measures. The algorithm implemented to find the signature day can be find in Chapter 4.

Development of the Signature Days: As stated earlier, the next step in the methodology is to use the results of the cluster analysis as inputs to a data analytic-based prediction model that uses data mining/machine learning to predict the traffic state in the short-term future in real-time operations. The investigated prediction models are classification algorithms that predict the traffic state in the next 30 minutes as belonging to one of the pre-identified clusters. For that effect, a 15-minute data across all days at multiple locations (intersections) is used as input for the prediction algorithms. The study implemented and evaluated seven different algorithms data mining/machine learning algorithms. The evaluated algorithms are the Decision Tree (DT), Random Forest (RF), Gaussian Naïve Bayes (GNB), Multinomial Logistic Regression (MLR), Support Vector Classification (SVC), K-nearest neighbors (KN), and Artificial Neural Network (ANN).

Implementation and Assessment of the Developed TRPS: To evaluate traffic performance with the developed TRPS method, we first optimized the traffic signals (using HCS7) for the signature day of each of the two clusters, the signature day of the entire database, and a random day in the database. The objective of the optimization was to minimize the delay in the network. Then, 10 sample days were randomly selected from the database and the average travel time in the network was calculated and compared under five different traffic signal timing. These five scenarios are: 1- the existed signal timing in the field, 2- Optimized signal timing for the signature day of the detected cluster for that day, 3- Optimized signal timing for the signature day of the predicted cluster for that day, 4- Optimized signal timing for the signature day of the entire database, 5- Optimized signal timing for a randomly selected day in the database.

1.5 OVERVIEW AND REPORT ORGANIZATION

In the remainder of this report, Chapter 2 presents the literature review. Chapter 3 provides the descriptions of the selected study locations for use in the case studies. Chapter 4 presents the methodology and results for the identification of the traffic patterns and signature days using an unsupervised clustering technique and the development of a model to predict the traffic state I real-time operations. Chapter 5 describes the methodology and results of the assessment of the pro-active TRPS developed in this study.

CHAPTER 2: LITERATURE REVIEW

As stated in Chapter 1, this study developed and assessed methods to categorize the traffic conditions in a peak period in traffic patterns that best represent the day-to-day variations in traffic flow parameters using cluster analysis. The first section in this presents a review of the use of cluster analysis for traffic pattern identification. This study also developed and assessed model for near-term prediction of traffic patterns in real-time operations for use in TRPS. Section 2.2 presents a review of past studies that used data mining and machine learning to predict the traffic state in the near future in real-time operations. Finally, this study used traffic analysis tools to estimate the benefits of the produced methodology. Section 2.3 presents a review of signal control strategies, while the last section summarizes the conclusions from the literature review.

2.1 REVIEW OF CLUSTER ANALYSIS FOR TRAFFIC PATTERN IDENTIFICATION

As discussed in Chapter 1, this study investigates the use of clustering as an unsupervised machine learning technique to categorize the traffic into traffic patterns. Several studies explored the use of clustering for the identification and classification of traffic states based on traffic measurements (Xia 2012), introduced an agglomerative clustering method that can identify congestion levels based on traffic characteristics such as flow, speed, and occupancy measures. The study used data collected from loop detectors located on Interstate 80 in the Bay Area, CA. The obtained traffic measurements were aggregated at 5-minute intervals for the analysis. The objective was to classify the traffic data into clusters of homogeneous characteristics by minimizing the inter-cluster data point distances while maximizing the intra-cluster data point distances. Their study concluded that the algorithm performed well in the identification of the traffic states on freeways by being able to provide an optimal fit based on an assessment using the Bayesian Information Criteria (BIC) along with the ratio of change as a dispersion measurement. The study reported that the test results were satisfactory for both real-time traffic monitoring and off-line traffic system performance evaluation.

Azimi and Zhang investigated pattern recognition methods using three clustering approaches to categorize freeway traffic conditions: the k-means, fuzzy C-means, and Clustering Large Applications (CLARA) algorithms (Azimi 2010). To perform the study, the researchers utilized field data collected from loop detectors located along US-290 in Austin, TX. The data consisted of timestamp, volume, occupancy, speed, and number of trucks and were aggregated at 15-minute intervals. The density was computed using the fundamental relationship from the measured speed and flow values. To facilitate the comparison of the produced traffic states with the level of service (LOS) according to the Highway Capacity Manual (HCM) classification, a total number of six clusters were introduced. The data were normalized and used as an input for the clustering algorithms. Then the data were clustered based on their density values. The comparison of the produced clusters with the level of service criteria specified in the HCM revealed that k-means performed better in terms of being consistent with the HCM LOS classification. Once the best clustering method was identified, the study performed a further categorization of the cluster corresponding to the LOS F into three separate subgroups. This time the input features considered for the categorization were speed and flow, and only the k-means and CLARA algorithms were used for this secondary analysis given the inconsistent results of Fuzzy C-means in the initial phase of the project. The possibility given by the clustering algorithms for classifying the oversaturated flow conditions provided an alternative for the analysis of the congested regime. The study demonstrated that for the subcategorization of LOS F, density is not a valid criterion, but attributes related to the theoretical shock wave speed are better for this type of classification.

Wu and Liu evaluated the impacts of signal operations on the Arterial Fundamental Diagram (AFD) by analyzing signal-based occupancy data from point detectors located on a major arterial in the Twin Cities, MN area (Wu 2011). The study utilized high resolution data and analyzed individual vehicle trajectories to assess the effect of the green-to-cycle (g/C) ratio, signal coordination, turning movements, and queue over detector (QOD) on the AFD. The analysis used a cycle-based approach to define the AFD based on flow and occupancy data from detectors located at a signalized intersection. The use of the cycle-based approach showed how signal operations interrupt the traffic operations for each cycle. Therefore, the cycle-based AFD was able to depict transitions between traffic states from cycle-to-cycle including under saturation, saturation, and oversaturation. The AFD was based on the flow-occupancy relationship producing cycle-based flow-occupancy diagrams utilizing data from both stop line and advance detectors at the selected intersections for the AM peak, PM peak, and off-peak time periods. The AFDs for different times of the day over two weeks were consistent in showing that different capacity values appear in the AFD for the morning and afternoon peaks. The results also revealed that the queue over detector (QOD) has an important effect on the AFD. When a queue spills back it can produce much higher occupancy readings while leaving the flow values unchanged and producing saturated and over-saturated areas of the AFD for low flow rate areas, with high occupancy values may result from the queue that builds up during the red signal period. By removing the QOD effect, the researchers produced a more stable form of the AFD that was used to quantify the effects of signal operations. The study concluded that not only does the g/C ratio constrain the capacity of a signalized approach, but poor signal coordination and turning movements also have a significant effect on the capacity.

Kianfar and Edara studied the application of diverse clustering techniques for the categorization of traffic flow data into free-flow and congested regimes (Kianfar 2013). The application is based on a framework that consists of clustering the fundamental traffic flow parameters (speed, flow, and occupancy) obtained from traffic sensor data from two major US metropolitan areas (San Louis, MI, and Twin Cities, MN). Three types of clustering algorithms were implemented for this study including the connectivity-based clustering, the k-means algorithm as a centroid-based clustering, and a general mixture model (GMM) algorithm as a distribution-based clustering algorithm. The objective of the study was to test different combinations of traffic variables to be used as input to the models and compare the results to evaluate the accuracy of the different algorithms with the aim to use the output (partitioned data) to generate the corresponding fundamental diagrams. The study identified the best clustering approach as the one with the minimum Davies-Bouldin Index (DBI), maximum Dunn Index (DI), and with a Silhouette Coefficient very close to one. The results of the experiment revealed that either the combination of speed and occupancy or the use of such parameters separately produced the best results for the algorithms in terms of accuracy. It was also found that the k-means and the hierarchical clustering algorithms performed better than the GMM algorithms by achieving higher accuracy for the same input features. The k-means algorithm was the one with the highest accuracy overall. Finally, the study used the output of the partitioned data form k-means to plot a flow-occupancy diagram and fit a linear regression model to define the flow-occupancy relationship. The experiment demonstrated that clustering is an effective way to categorize the traffic data into free flow and congested regimes.

Sun and Zhou derived a multi-regime fundamental traffic relationship based on cluster analysis for the segmentation of speed-density data. Three different datasets were utilized to develop and test the model (Sun 2005). The datasets included occupancy, flow, and speed for each lane collected at every 20 seconds from loop detectors and video image detectors installed at multiple locations in San Antonio, TX. Since the raw data as it was collected from loop detectors did not include densities, the traffic densities were derived from the occupancies based on flow rate and space mean speed. Once the density was computed the data were standardized to make the relative weight of each one of the traffic variables equal for the computation of the relative distances when implementing the clustering algorithm. For clustering the data, a k-means algorithm was implemented and the results were assessed with several alternatives regarding the numbers of clusters as well as standardized and non-standardized features. The study concluded that the k-means algorithm is an effective way for partitioning the traffic data for the development of speed-density models. The application of the clustering method produced clusters that can be visualized based on the speed-density relationship. After the segmentation, the subsets were individually fit to a regression model to produce an accurate representation of the speed-density relationship for each cluster. The experiment also revealed that the original (non-standardized data) also works well by producing clusters that give an accurate representation of different traffic states.

Previous research also explored the implementation of clustering algorithms to data collected from arterial streets. For example, Yang et al., utilized a spectral clustering algorithm to analyze the traffic state variations at the network level based on speed data (Yang 2017). With the implemented clustering approach, the authors identified five different traffic states that were later related to different locations and type of road section. The study highlights the importance of the knowledge on the spatiotemporal diversity of the network in combination with the clustering algorithm for further discovery and classification of traffic patterns. The authors also recommend the utilization of the clustering output for network level traffic predictions. Another study by Theofilatos, implemented an expectation maximization clustering algorithm to classify the traffic into multiple regimes in urban arterials for safety analysis (Theofilatos 2017). The study used the average occupancies, standard deviation of occupancies, and incident data as input for the implemented clustering algorithm. The implementation resulted in the identification of nine significant traffic regimes that were further analyzed using Bayesian logistic regression to model the likelihood of an incident and its severity.

Gu et al. employed a k-means clustering algorithm to analyze arterial traffic flow (Gu 2016). The study utilized high-resolution controller data and video images from multiple locations (intersections) along the network to categorize the existing traffic patterns before and after the closure and reopening of an arterial corridor to show how the traffic patterns evolve after the road closes or reopens. The study concluded that the employed clustering approach utilizing traffic counts and occupancies from high resolution data is effective in classifying different traffic patterns and has the potential to be used on large networks. A study by Mosammat in 2021 combined the use of high-resolution controller data and travel time measurements in a two-level clustering technique using the k-means algorithm (Tariq 2021). First, travel time data were classified into four separate clusters that represented different levels of congestion, then in a second clustering level the cluster that represented the highest congestion (peak period) was further partitioned using the green occupancy ratios derived from event-based controller data.

2.2 REVIEW OF DATA MINING AND MACHINE LEARNING TO PREDICT TRAFFIC STATE

As stated in the overview of the methodology proposed in this paper presented in Chapter 1, the next step is to use the results from clustering as input to train data mining/machine learning models for short term prediction of the traffic state in real-time operations. This section presents a review of the use of data mining and machine learning to predict the traffic state in traffic engineering literature.

Some studies have utilized a combination of unsupervised (clustering) and supervised learning for the prediction of the traffic states. In a study by Azizi and Hadi (2021) , a freeway segment was utilized as a use case to propose a methodology that includes the utilization of disturbance metrics including the number of oscillations, and Time to Collision (TTC) as input to a clustering algorithm for the off-line categorization of the traffic states. Once the traffic states were identified the study investigated the implementation of machine learning based classifiers for the recognition and ultimately the prediction of the traffic state based on available historical data generated from simulation. The study concluded that the proposed disturbance metrics were significant variables for the prediction model. Other studies, implemented neural network architecture models to predict the traffic state. For example, Hosseini et al., (2019) utilized time-space diagrams constructed from connected vehicles data in combination with a convolutional neural network (CNN) for the prediction of the traffic state which was defined based on the density flow relationship. The traffic states ranged from free flow to fully congested. For evaluation purposes, the study compared the predictions made by the CNN-based model with the prediction made using other algorithms including a multilayer perceptron, vector regression, and autoregressive moving average models. The study concluded that one of the benefits of the proposed architecture resides in its capability to capture the interaction between individual vehicles and their impact on the traffic stream that are otherwise not easily perceived by other types of architectures. According to the study, the CNN demonstrated a better performance in its prediction capabilities when compared with the other models.

To solve the complications inherent to the traffic fluctuations and signal control on arterials, Li and Ban developed a deep learning-based method for short-term traffic volume prediction of all movements at signalized intersections (Li 2019). The authors proposed a model that integrates a Convolutional Neural Network (CNN) with a Long-Sort Term Memory (LSTM) algorithm. In their model, a CNN is first implemented to account for the spatial dependencies of traffic flow by transferring the lane-based volume data from each intersection to a 2D image that is equivalent to an origin-destination (OD) matrix. The OD matrices form multiple adjacent intersections are stacked together and then used as an input to the CNN. The LSTM takes the CNN output as an input for each time step and is trained to learn the temporal dependencies of the data. The output of the model was validated using simulation. The results of the evaluation show that the combined CNN-LSTM model outperforms several other models in terms of prediction accuracy.

Iqbal and Hadi (2017),**Error! Reference source not found.** developed a model to predict the breakdown probability on urban arterial streets utilizing ITS data collected from detectors located along Glades Road in Boca Raton, FL. The study defined the breakdown occurrence in arterials based on the HCM 2010 threshold for LOS F on urban segments, which can be otherwise explained as the point where the speed decrease to a value that is less than 30% of the base free flow speed. The developed model utilized a 10-minute time horizon for the breakdown prediction to provide the facility operator with enough time to implement countermeasures that lower the probability of breakdown occurrence once a high probability of breakdown is predicted by the model. To address the complexities of the many parameters associated with signal control and traffic movements in arterials, the model utilizes not only data from point detectors along the road but also data form automatic vehicle identification technologies that was used as input for the model. The model utilizes a combination of decision tree and binary logistic regression algorithms to predict the breakdown probability. First, the top levels of the decision tree are built using a top-down induction decision tree algorithm (TDIDT), then a Recursive Partitioning and Regression Tree (RPART) algorithm is implemented to construct the lower levels of the tree. It was found that using multiple algorithms to construct the tree rather than a single algorithm helps to increase the proportion of breakdown data points at the higher levels of the tree, allowing a more effective performance of the RPART at the lower levels of the tree. Once the tree is fully developed, the final step consists in the implementation of a logistic regression model by fitting the data at the end nodes of the tree to improve the classification of the breakdown depending on the node attribute values. Additionally, a random forest analysis was used to identify the features that had a higher contribution to the prediction. It was found that the downstream occupancy, downstream speed, upstream occupancy, and upstream speed are the features that have a higher predictive power in the model. Root Mean Square Error (RMSE), and Mean Percentage Error (MPE) were the two metrics employed to validate the model. It was concluded that the model was able to classify conditions with high probability of breakdown occurrence for a 10-minute time horizon. Also, in the validation phase, the model revealed a satisfactory performance by achieving an RMSE of 13.6% and MPE of 11%.

Elfar et al. (2018), performed a study where three factors were identified as the main causes of traffic breakdown. Those factors were described by the authors as high traffic loads, bottlenecks, and disturbances caused by individual drivers such as abrupt braking or lane-change maneuvers. The authors noted that the first two factors are easily identifiable and measurable using traditional detectors data. However, the third factor is more difficult to observe because they occur at an individual vehicle level and thus, they are not easily identified by traditional sensors. Therefore, the study focused on the capability of machine learning based models to predict traffic congestion based on individual vehicles trajectory data available form connected vehicles technology. The study employed logistic regression, random forest, and neural network algorithms to develop prediction models for both offline and online operations. The vehicle trajectory data set used for the model was collected form available data form US 101 in Los Angeles, CA. The dataset includes information such as speed, acceleration, location, and headways at 0.1 seconds resolution. The data were first preprocessed to estimate traffic flow, density and mean speed aggregated at 10 second-time steps. Also, the segment was divided into sections whereby the speed standard deviation (SDD) was computed for the average speeds of every individual vehicle for all sections. The authors recognized that the SDD could be a good measure for the level of traffic disturbance provoked by individual vehicles given that in microscopic models the increase in SDD among individual vehicles is a good indicator for anticipating traffic breakdown according to (Treiber 2006). The first step of the analysis consisted of the identification of traffic states using the k-means clustering algorithm. For simplicity, during the study only two traffic states were identified (uncongested and congested). With the traffic states being identified, the next step consisted of the implementation of the prediction algorithms for 10 sec and 20 sec time horizons. Also, different levels of market penetration of connected vehicles were considered for the model implementation. The accuracy for the prediction of the congested traffic state was measured separately from the accuracy for prediction of the uncongested traffic state. The results showed that high accuracy (between 89% and 93%) was achieved when predicting for shorter time horizons. Regarding the prediction of the congested state, the three algorithms showed satisfactory results achieving accuracy scores from 94% to 97%. The accuracy for the prediction of the uncongested state, however, was much lower, achieving accuracy scores from 68% to 85%. The results also showed that the logistic regression and random forest algorithms were more accurate in prediction compared to the investigated neural network. In reference to the market penetration levels the models performed well under partially connected traffic streams achieving accuracy scores from 88% to

92% for low and medium connectivity percentages. For full connectivity, the achieved accuracy ranged from 92% to 96% for the congested state and achieved a significantly lower accuracy for the uncongested state with scores ranging from 68% to 83%.

Adu-Gyamfi and Zhao (2018), introduced a methodology for traffic speed prediction in urban arterials using a combination of LSTM Neural Network with an Empirical Mode Decomposition (EMD) algorithm. The model is intended to adaptively perform pattern recognition on historical traffic flow data using the EMD algorithm for multiscale pattern recognition. The EMD pattern recognition algorithm obtains information from the historical traffic speed fluctuation patterns to later serve as a guide to train the LSTM model to help achieve higher accuracy in speed prediction. Data obtained from US 50 in northern Virginia was used to feed the model. The dataset includes 15-minute traffic counts and average speeds collected from over 200 detectors located at 38 different intersections. A time-varying volatility index (VI) was added to the dataset to represent the variability of the traffic speeds across different scales. To compute the VI, the EMD algorithm is used to extract the underlying traffic speed patterns at each detector over the full period of analysis. Once the VIs are produced, the patterns detected at all detectors are re-grouped into high-frequency patterns that represent short-term events and low-frequency patterns that represent the general (common) traffic speed patterns. Once the high and low frequency VIs were aggregated, the LSTM model was implemented by testing multiples architectures by varying the number of historical input features as well as the output vector shape. The experiment also includes the testing of the model with and without the use of the high and low frequency VI to assess the benefits of the EMD algorithm. The study found that the speed prediction error ranged from 2 to 6 mph with an average of 3 mph. The study concluded that the use of the EMD algorithm to add volatility information to the model could significantly improve the model capability to learn and predict traffic speed patterns in about 35% on average.

2.3 REVIEW OF TRAFFIC SIGNAL CONTROL STRATEGIES

One of the most effective and critical methods to control traffic and create safe and fast travel is traffic signal control (TSC). Since the introduction of TSC in 1913 in Cleveland, Ohio, U.S.A. (Mueller 1970), research has been conducted to improve their safety and efficiency. TSC regulates vehicle movements based on a signal phase sequence that periodically repeats.
Three types of intersection TSC problems are defined in the literature review: isolated intersections, arterial networks, and general networks (Eom 2020). A single intersection that works separately is an isolated intersection, while a sequence of consecutive intersections forms an arterial. A general network includes several intersections that are not necessarily all along the same axis.

Four general TSC strategies have been developed: fixed time, actuated, responsive, and adaptive. In the fixed-time (or pre-timed) strategy, the cycle length, phase plan, and duration are predetermined based on the historical traffic information. This strategy assumes the traffic demand remains nearly constant, and the optimal signal timing plan can be calculated based on this demand. Several fixed-time plans may be deployed for different hours of the day; however, this strategy cannot handle traffic fluctuations very well, especially when demand variability is large.

The actuated control strategy uses sensor data (typically from loop detectors) and applies simple logic and rules such as maximum green time, green time extension, and gap out to change the traffic signal timings based on real-time traffic conditions. While this strategy is flexible based on the traffic demand at the intersections, it cannot react to significant changes in traffic patterns. Any traffic signal control system that collects the traffic data analyzes and optimizes the traffic signal performance and then modifies signal timing in response to that is described as adaptive signal control technology (Mueller 1970), (Eom 2020), (Lo 2002). The main objective of these systems is to minimize travel time and reduce the number of stops through the corridors. While these systems are expensive, and their installation is time-consuming, previous studies have shown that they can improve traffic performance. Several types of adaptive strategies have been developed, such as Insync, SynchroGreen, Split Cycle Offset Optimization Technique (SCOOT), Sydney Coordinated Adaptive Traffic System (SCATS), Los Angeles Adaptive Traffic Control System (LA-ATCS), Real-Time Hierarchical Optimized Distributed Effective System (RHODES), Optimized Policies for Adaptive Control (OPAC), and Adaptive Control Software Lite (ACS-Lite). Each of these strategies requires a different type of detection and equipment, and uses a different algorithm to calculate split, cycle, offsets, and phase sequences. The following paragraphs briefly describe InSync and SynchroGreen, as examples of adaptive signal control (Dell 1995).

Insync Adaptive Traffic Signal Control System: The InSync adaptive traffic control system was released in 2008 by Rhythm Engineering (Lee 2017a). According to a Federal Highway Administration report, adaptive signal control systems, continuously adjust daily signal schedules to account for traffic demand, respond quickly to any change in traffic patterns, and gradually make the travel time reliability better in the corridor (Lee 2017b). InSync uses the idea of states, where each state is a phase or pair of phases that occur concurrently but without conflicting with each other (Rhythm Engineering 2014). The InSync adaptive signal control system adjusts dynamically the signal states, sequences, and/or the green time to take into account the traffic demand at any time.

After collecting all available detection data, a "greedy" optimization algorithm's logic is applied at the "local" intersection level to reduce the overall delay. In this algorithm, tokens are given out to vehicles approaching the intersection on red, and the vehicles receive a token every five seconds they are stopped at the intersection. The algorithm seeks to reduce the number of tokens distributed in order to reduce the delay at the intersection. The InSync system's goal is to ensure the delay along the corridor is minimized by establishing speed lines through the corridor and optimizing the traffic signal times locally (Chandra 2010). In order for the vehicles traveling at the desired speed to pass through the corridor without stopping, the global optimizer ensures the platoons of vehicles through the corridor move with minimum delay (Whitelock 2014). Then at each intersection, by calculating the optimal phase combination, the optimizer tries to minimize delay to serve all demands at the intersection. The timing plan or cycle length requirements for this process don't need to be fixed (Siromaskul 2010). Since 2008 the InSync adaptive signal control has been installed at various corridors. These installations report 4-42% travel time reduction (TJKM Transportation Consultants 2011), (E. Hathaway 2012), (Inc. 2012), (Sprague 2012), (J. N. Hutton 2010).

Synchrogreen Adaptive Traffic Signal Control System: SynchroGreen takes into the account side street, pedestrian, and mainline traffic to provide an adaptive response to the dynamic traffic demands. SynchroGreen has the following characteristics: 1) adjusts traffic signal timing in real-time based on current traffic demand 2) uses three optimization engines to achieve better traffic flow and green time allocation 3) is compatible with the current infrastructure for traffic control, including numerous common traffic controllers and many types of detection 4) provides users with the option to choose from a variety of ways to support balanced traffic flow, progression bandwidth, and critical movements 5) interacts with SimTraffic and Synchro to model and test various system parameters before deployment (Trafficware 2012).

The main objective of the SynchroGreen algorithm is to maximize mainline progression bandwidth while minimizing network delay. SynchroGreen additionally offers three alternative adaptive control modes: a) The Balanced Mode offers acceptable mainline bandwidth while distributing green time fairly; b) Mainline progression is prioritized in the Progression mode; and c) The detected critical movements are given more weight in the Critical Movement Mode. In summary, SynchroGreen uses a more traditional approach to signal control optimization and adapts the phase allocation (splits), period (cycle length), and start time (offsets) in real-time based on the traffic conditions.

TRPS strategies have been proposed since the 1970s as an alternative to TOD that can address some of the issues associated with day-to-day variations in traffic patterns. The requirements of these strategies are much lower than those of adaptive signal control strategies Hadi (1990) presented a detailed review of the different variations of TRPS and classified the TRPS algorithms into two types. In the first, the timing plans are selected based on comparison function of volume and occupancy measured based on few detectors located at strategic locations in the network. The value of the function is calculated in real-time operations and compared to preset transfer thresholds to switch between the plans. In the second type, preset transfer thresholds are used to select the cycle length, offsets, and splits separately based on volume measurements from few specific detectors in the network. There are variations of how this second method is applied. For example, one system selects the cycle length based on main street traffic volume, the offsets based on the difference between the inbound and outbound volume levels, and the split based on the difference or ratio between the side street and main street volumes. There are several limitations and issues associated with TRPS that have limited the adoption of these strategies as listed below (Hadi 1990).

- The need for the near-term prediction of traffic flow parameters for use in the plan selection rather than using traffic flow parameters that may change in the next period,
- The difficulty in designing the plans to be stored in the TRPS plan library,
- The need for a method to weight the data collected from different detectors,
- The need for a method to set the plan activation thresholds,
- The need for installing and maintaining additional detectors, and
- The need to limit the number of plans switching in a peak period to reduce the delays due to the transition interval between the plans.

2.4 REVIEW OF TRAFFIC SIGNAL CONTROL STRATEGIES

The following conclusions can be drawn from the review of literature on cluster analysis for traffic state identification.

- Clustering methods have proven to be a useful tool for pattern recognition in traffic analysis. Clustering works by grouping similar patterns into clusters whose members are more similar to each other than to members of other clusters. Such grouping is directly applicable to signal control as timing plans can be designed off-line for each traffic pattern identified using cluster analysis. The identified clusters can also be used in combination with other data mining/machine learning techniques to predict the traffic patterns in real-time operations, as is done in this study.
- Generally, traffic pattern clustering has been based on fundamental macroscopic traffic variables including flow, density/occupancy, and speed/travel time. More detailed measures based on high resolution data and vehicle trajectories have also proven to be useful to categorize the traffic states using clustering.
- Overall, the k-means cluster algorithm, which is the most widely used algorithm, has been demonstrated to produce better categorization of the traffic states than other types of investigated algorithms such as the hierarchical, centroid based, and density-based algorithms. Hence, k-means is selected in this study to produce the categorization of the traffic states for the use case.
- Valid ways to evaluate the produced clusters include several statistic measures and traffic measures including the visualization of the clusters using the fundamental diagram.

The following conclusions can be drawn from the review of literature on data mining and machine learning techniques to predict traffic state in real-time operations.

- Traffic state prediction plays a significant role in Intelligent Transportation Systems. However, it's used in signal control has been limited. There is a need for additional research to develop and assess methods for using data-based prediction models as part of proactive signal control in real-time operations.

- A popular approach for predicting the traffic states is a combination of unsupervised and supervised learning methods where, first the traffic data is categorized in different traffic states using clustering, and then a prediction algorithm is implemented to determine the probability of occurrence of a given traffic state (cluster) in the short term, given the existing conditions expressed as a set of traffic variables.
- Among the most used traffic variables as input for the prediction models are speed, travel time, volume, density, and occupancy. Additional metrics have been derived and proven useful for prediction purposes. Such metrics include different variants of the standard deviation of speed, time to collision (TTC), and number of oscillations.

The following conclusions can be drawn from the review of literature on signal control.

- Four general TSC strategies have been implemented including fixed time, actuated, responsive, and adaptive.
- When there is a high variation in traffic demand, agencies have implemented traffic responsive and adaptive strategies.
- The adaptive strategies have the potential of improving system performance. However, these strategies require additional resources and costs to design, install, and maintain. In addition, adaptive traffic systems may not be appropriate for all locations. Thus, there is a need to explore the use of next generation traffic responsive strategies that utilize ATSPMs already estimated by an increasing number of TSC around the nation.

CHAPTER 3: DESCRIPTION OF THE STUDY LOCATIONS AND THEIR DATA

In this study, two locations were initially selected as case studies for the application and the evaluation of the methodology developed in this study. The first case study is Newberry Road in Gainesville, FL, and the second is NW 119th Street in Miami, FL. This chapter describes each location and presents the issues encountered in the collected data sets.

3.1 NEWBERRY ROAD CASE STUDY

Along most of its length, Newberry Road has three lanes in each direction. As shown in Figure 1, the facility contains six intersections from Newberry Rd @ I-75 W to Newberry Rd @ 66th St. The speed limit of Newberry Road varies between 35 and 45 mph. The City of Gainesville provided the traffic data, including average travel time for each direction from BlueToad, as well as traffic volume and turning movements for every 15 minutes obtained from Iteris VantageLive video image detection system, in addition to signal timing information. The data obtained for this study is from January 1st, 2019 to August 28th, 2019.

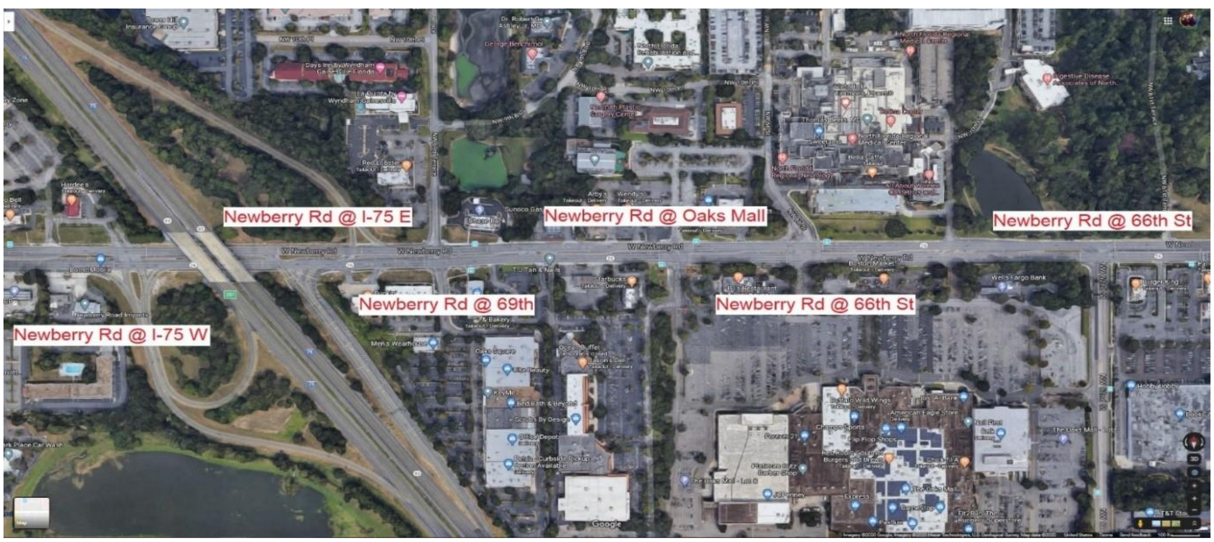

FIGURE 1. NEWBERRY ROAD STUDY CORRIDOR, IN GAINESVILLE, FL

One of the main issues with the data obtained from this site was the high number of records with missing values. Also, blank records in consecutive time steps were consistently present during long periods. For example, there were missing records for the whole month of January at the first intersection (NewberryRd@1-75 W), around 40 consecutive days with missing records at the second intersection (NewberryRd@1-75 E) during July and August, and so on. Significant periods with consecutive blank records result in having to eliminate them from the analysis. Generally, there were many data points with missing records in between detector readings during the peak hour. This means that even though the data resolution was 5 minutes, there were many instances where the records were available only every hour, or there were other instances where the data were available only every other record (i.e., every 10 minutes instead of every 5 minutes). Missing values could be estimated by interpolation but when the failure lasted for more than two consecutive records , estimating the missing values accurately is not possible. Therefore, taking such days out of the dataset was preferred to avoid biased results. Dismissing records comes with a cost, namely reducing the sensitivity of the model by either having a much lower number of instances to train the model or by losing examples that represented a particular traffic pattern in the real world but will be not present in the dataset. In total, 70 days' worth of data had to be dismissed. Also, the reported travel times in the database were found not to logically match the traffic volume. For example, Figure 2 shows the relationship between traffic volume and travel time on May 9, 2019. from 3 pm to 6 pm. Based on this figure, while the traffic volume slightly decreases in the eastbound direction and does not have many variations for the westbound direction, the travel times in both directions increase dramatically. However, the measured traffic volume is much lower than the capacity of these intersections. Since there is no logical correlation between travel time and traffic volume at this site, we decided not to use this site in our analysis.

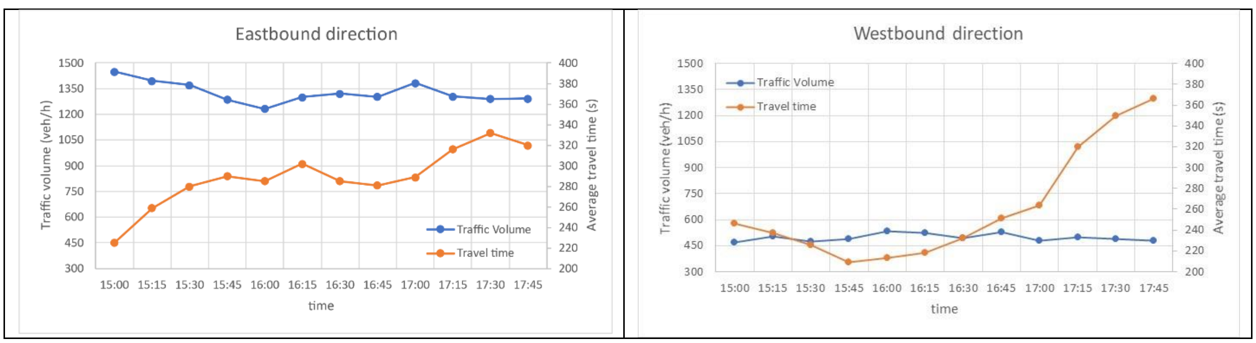

FIGURE 2. RELATIONSHIP BETWEEN TRAFFIC VOLUME AND TRAVEL FOR THE NEWBERRY ROAD SITE

3.2 NW 119TH ST CASE STUDY

Given the challenges regarding data collection faced in the first location, as described in Section 3.1, the team opted to analyze a second location to test the proposed methodology. The second location is NW 119 Street in Miami Dade County, which is an important east-west arterial in Miami, FL. The analyzed segment is a two-mile segment between NW 32nd Avenue and NW 7th Avenue, containing nine signalized intersections, as shown in Figure 3. The speed limit for the entire segment is 40 mph. Available data for this location includes high-resolution controller data (HRC) from all the signalized intersections within the site, travel time data for both directions from the Regional Integrated Transportation Information System (RITIS) database (obtained from HERE a third-party vendor data), as well as data from Bluetooth detectors. The utilization of the HRC from the signalized intersections allowed the derivation of metrics such as the green occupancy ratios (GOR), split utilization ratio (SUR), and vehicle counts from each approach.

FIGURE 3. NW 119TH STREET IN MIAMI DADE, FL

The obtained data for this site were found to be of good quality with almost no missing records. Also, the data from different sources or databases for this location are consistent in terms of showing the same patterns during the analyzed period.

The main challenge that was identified with the data is related to the configuration of the sensors on the through movements of the main approaches, since they were not located at ideal locations to estimate the performance metrics for all links. It was determined that it is impossible to get accurate vehicle counts and other metrics for the through movements from the stop bar detectors at most intersections in the EB direction. This is due to the configuration and dimension of the loop detectors installed at the stop bar locations, which may result in several vehicles being counted as a single one if they are over the detector at the same time. Exit detectors at the main approaches were used as alternative detector locations for vehicle count purposes to overcome this problem. Exit detectors are installed around 200 ft to 300 ft downstream of several intersections. Cross street detectors, where available, were also utilized to get approximate vehicle counts of the through movements from the northbound and southbound cross street approaches. By using a combination of exit detectors on the main approaches and cross street detectors, it was determined that it was possible to get an accurate representation of the traffic patterns for the purpose of this study considering that the intersections with available exit detectors happen to be the major intersections in this segment. These intersections are NW 32nd Ave, NW 27th Ave, NW 22nd Ave, and NW 17th Ave. Due to the difficulty in estimating the turn movement volumes because of the detector configurations, the turn movement volumes at the analyzed locations were determined based on the total link volumes using turn movement proportions movements as reported in the final report of the 2018 Operational Analysis study by District Six. (Florida Department of Transportation 2018).

The data were first filtered to include only the data points belonging to the peak period. Through a preliminary analysis, it was observed that the study location presented a pattern of recurring congestion during both the AM peak (weekdays from 7:00 am to 10:00 pm) and PM peak (weekdays from 4:00 pm to 8:00 pm) with the AM peak being the most critical one of the two referenced periods by having higher volumes and travel times than the PM peak. Figure 4 compares the average travel times in the eastbound (EB) and westbound (WB) directions. As the figure depicts, the travel time in the EB direction is significantly higher and has more variability than the travel time in the WB direction, in the AM peak. This is due to the fact that many vehicles use NW 119th St to access the I-95 south during the morning peak, as many of the drivers are heading to downtown Miami. Considering the described factors, it was decided to utilize the GOR, SUR, travel times, and volumes from the EB direction as features in the clustering algorithm as they are more likely to explain the traffic states or levels of congestion that may occur during the morning peak period.

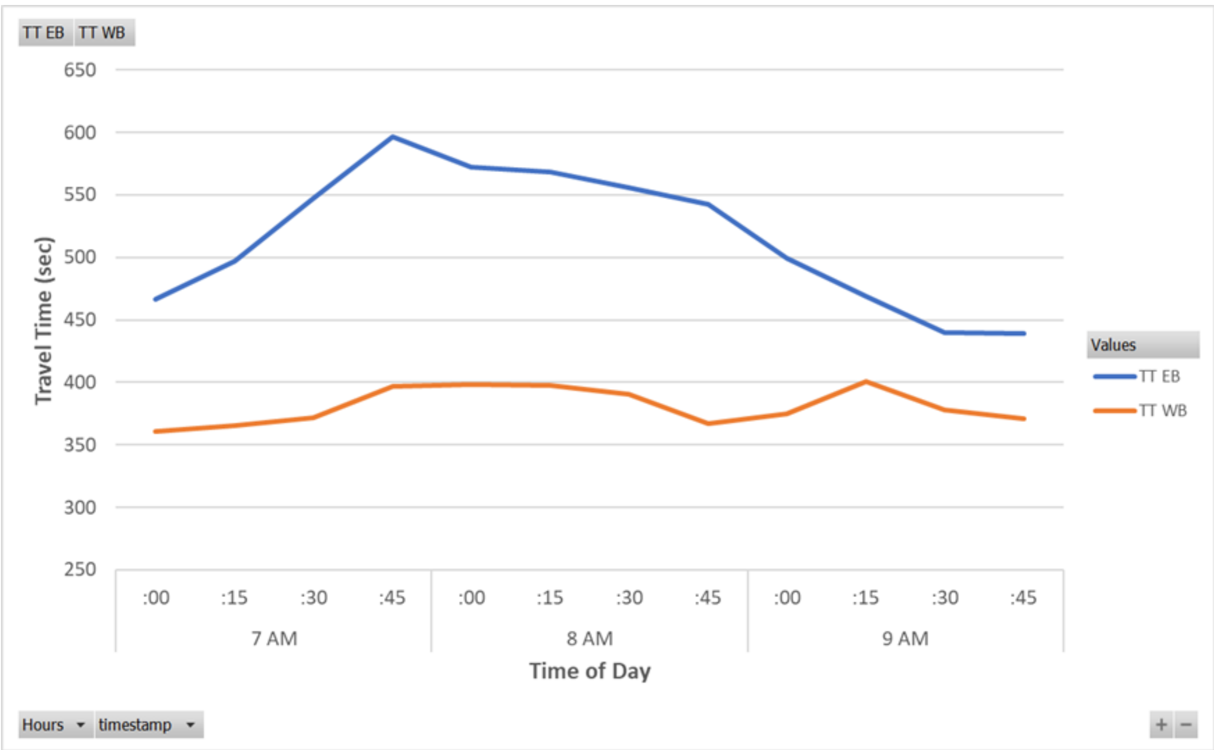

FIGURE 4. AVERAGE TRAVEL TIMES DURING THE AM PEAK ON THE 119TH STREET

CHAPTER 4: IDENTIFICATION OF TRAFFIC STATES FOR PLAN DEVELOPMENT AND ACTIVATION

This chapter presents more details about the proposed methodology to identify the traffic patterns for which signal plans can be designed and activated and the results from applying the methodology. The chapter first describes the method and results of categorizing the traffic data based on clustering analysis and presents the process of identifying a signature (representative) day for each cluster to use in developing the signal timing plan. Then, it describes the development of and assessment of predictive models using various data mining/machine learning techniques based on the results of cluster analysis. Due to the issues observed in the database of the Newberry Rd. site, it was not possible to obtain clusters that represented distinctive traffic states for that location. Thus, the rest of this document presents results only for NW 119th St case study.

4.1 MEASURES OBTAINED BASED ON HIGH- RESOLUTION CONTROLLER

In recent years there has been an increasing interest in utilizing high-resolution controller data that includes signal timing and detection at the highest resolution of the controller (0.1 seconds), in combination with data from other sources to support signal optimization systems. The high-resolution controller data provides significant support of the operation and maintenance of traffic signals by allowing the identification of capacity utilization level, determining progression quality, estimating performance measures, and assessing detection and communication malfunctions. The measures estimated based on high resolution controller data can be used for daily operations including setting basic parameters, identification of detection problems, and estimating impacts under non-recurrent conditions. The measures can also be used for off-line modeling and optimization of the signals and for prioritizing signal improvement needs and to communicate the system status to the decision makers.

Examples of derived measurements based on high-resolution controller data include approach delay, Purdue phase termination, volume/capacity (v/c) ratio, green occupancy ration (GOR), and split utilization ratio (SUR). The last two, (the GOR and SUR) have been utilized in this study for the identification of the traffic states. These two measures can be categorized as capacity utilization performance measures. Below is a description of two measures calculated based on high-resolution controller data and utilized in this study.

.

Green Occupancy Ratio (GOR)

GOR is a measure that is intended to reflect the phase utilization. It is defined as the stop bar detector occupancy during the green interval. Higher values of GOR reflect higher utilization of the green time. This measure can be used under different detector configurations. However, it requires stop bar detection for the movements. Higher values of GOR reflect higher usage of the green time. This value increases to values above 0.5 in the peak periods.

Split Utilization Ratio (SUR)

The split utilization ratio (SUR) measure is derived for each intersection movement to allow the assessment of congestion levels in all directions. The SUR is defined as the ratio of the number of vehicles passing the detector to the maximum number of vehicles that can pass during the effective green time.

4.2 CLUSTER ANALYSIS

Cluster analysis was implemented in this study to identify traffic patterns that are representative of the traffic conditions present at the site, considering the variations in the day-to-day variations. The categorization of traffic patterns using clustering was implemented to support data modeling for short-term prediction as described in Section 4.4 and signal timing plan development and implementation, as described in Chapter 5. The goal of the clustering algorithm was to categorize the days with similar traffic patterns within the analysis period (the AM peak).

As described in the previous studies reviewed in this report, the k-means algorithm has been proven to be effective in categorizing the traffic states using real-world and or simulation data. Therefore, the k-means algorithm was adopted in this study to produce the corresponding clusters. The K-means algorithm starts by generating “k” centroids randomly from the data points and assigning each data point to the nearest centroid. Once all data points have been assigned, the centroids are recomputed and relocated such that the intra-cluster distance is minimized, and the inter-cluster distance is maximized (Tan 2016). This process repeats until either the maximum number of iterations are reached, or when the data points stop changing cluster (centroid assignment does not change). If the k-means algorithm works properly, the obtained clusters are expected to represent accurately the traffic states at the study location. Vehicle counts, travel times, GOR, and SUR from the selected intersections were utilized as the input for the clustering algorithm. The data from all intersections and segments for all days were compiled together and aggregated in intervals of 15 minutes for the analysis period which is the AM peak period in the EB and NB directions.

The K-means algorithm requires the specification of the number of clusters as an input to the algorithm. This study obtained the number of cluster (k) based on the elbow technique and also based on examining the results from the cluster analysis. The elbow technique is a plot that depict the total within clusters sum of squares (WCSS) for each value of k. The k value is selected at the point in the graph where the decrease in WCSS stop being significant as the value of k. The output of the clustering procedure allowed the categorization of the data points into three different clusters or traffic states representing uncongested, intermediate, and congested conditions. Clustering has been used to identify operational patterns for use in combination with traffic management and traffic simulation modeling. For example, Xia and Chen (2007) used K-means clustering to identify the traffic flow phases based on traffic density and speed data aggregated in 15 minutes. The best and most extensive example of the utilization of cluster analysis in transportation engineering is its use in efforts funded by the Federal Highway Administration (FHWA) to assess AMS Testbeds. Although the K-means clustering method has been widely used, there are several other clustering methods, each with its advantages and disadvantages. Some of these methods are K-prototypes, K-medoids, Hierarchical Clustering, clustering with dimension reduction using principal component analysis (PCA), fuzzy clustering, Gaussian mixture models (GMM) clustering, and clustering using Wavelet transformation, among others.

4.3 IDENTIFICATION OF THE SIGNATURE DAYS

The next step sought to identify the signature day for each cluster. The signature day is identified as the best day that represents the traffic conditions. The methodology to accomplish this is the based on that proposed in the Traffic Analysis Toolbox Volume III (Wunderlich 2019). In order to identify the signature day for each cluster, a 15-minute profile analysis was performed across all days considering all travel conditions (clusters), at multiple locations (intersections) for the key measures. The algorithm implemented to find the signature day is and is as follows:

1. For a particular key measure, list the analyzed locations (intersections)
   - Let *M* be the set of measures, considered over *J* (set of locations).
   - $N_{cluster}$ is going to represent the number of days in each cluster.
   - $M_{i,j}(t)$ is the value of measure on day *i*, in time interval *t,* at location *j*.
2. The average value for each 15-minute time interval across all days in the travel condition for each location is computed for each measure as:

$$\bar{m}_{t,j} = \frac{\sum_i m_{i,j}(t)}{N_{cluster}} \ \forall m, t, j \tag{2}$$

3. The percentage difference between the average value and the value observed on a particular day is computed by means of:

$$\dot{m}_{i,j}(t) = \frac{\sqrt{(\bar{m}_{t,j} - \bar{m}_{i,j(t)})^2}}{\bar{m}_{t,j}} \ \forall m, t, j \tag{3}$$

4. Finally, the signature day is identified as the individual day that minimizes the difference between the individual day and the average values, considering all the selected locations and measures. The signature day is then identified as:

$$i^* = min_i \left[\sum_m \sum_i \sum_t \dot{m}_{i,j}(t) \right] \qquad (4)$$

For example, suppose we have a cluster of 30 data points representing the performance of different locations or intersections over a period of time. We want to identify a signature day that can be used as a day with performance that best represents the performances of all days in the cluster. To do this, we compute the average value of each measure for each time interval across all intersections. Then, for each day, we calculate the distance between the daily values and the centroid of each cluster using a distance metric such as the Euclidean distance. We then identify the day, which we refer to as the signature day, where the distance to the centroid is the smallest relative to other days within the cluster.

4.4 CLUSTERING RESULTS

The above-described methodology allowed the identification of three signature days (one for each cluster at the study location). November 27, 2019, was identified as the signature day for Cluster 1, which is the cluster with the lowest volumes (uncongested conditions). November 4th, 2019 was identified as the signature day for Cluster 2, which represents the medium volume cluster. November 20th, 2019 was identified as the signature day for Cluster 3, which represents the clusters with the highest volumes (congested conditions).

Figure 4 shows the volumes of the signature days for all approaches on three of the intersections within the study segment. Note that due to space limitations, only three intersections are shown in the figure. As Figure 4 shows that the produced clusters from the clustering procedure implemented for NW 119th St confirms clearly represent different traffic states in the facility. As indicated by the through volumes in the EB direction (the peak direction), Cluster 1 exhibits the lowest volumes of all intersections relative to the other clusters, whereas Cluster 3 exhibits the highest volumes overall for the EB through movements at all intersections. The difference in volume is significant. For example, for the intersection of NW 19th Street with NW 32nd Avenue, Cluster 1 shows an average hourly volume of 1,740 vph during the AM peak for the EB through movement. Cluster 2, for the same intersection and same movement, shows a significantly higher average hourly volume (2,068 vph) which is 328 vph (or 19%) higher than Cluster 1. Finally, Cluster 3 shows an hourly volume of 2,292 vph, which is 11% higher than the volume of Cluster 2 and a 31% higher than the volume of Cluster 1.

The signature days provided the average hourly volumes for each traffic level (cluster) at each approach in the analyzed intersections, that is, the intersections where vehicle count was possible due to the existence of exit detectors. For the rest of the intersections in the system, the vehicle counts were estimated using an intersection volume balancing tool (Wisconsin DOT 2018) which is based on the Furness method where the average volumes for each cluster are used as target values to balance the volumes in the rest of the intersections in the system. In order to do that, the Furness method employs a gravity model where the volumes are arranged into an OD matrix form. Then the rows and columns are factored by multiplying the values in the matrix cells by the ratio of the desired to actual values (Wiscosin DOT 2019), (Ren 2009). The software kept iteratively multiplying the rows and columns of the OD matrix until the row and column sums meet the targets, or the error (deviation between desired to actual values) is small enough to be tolerated.

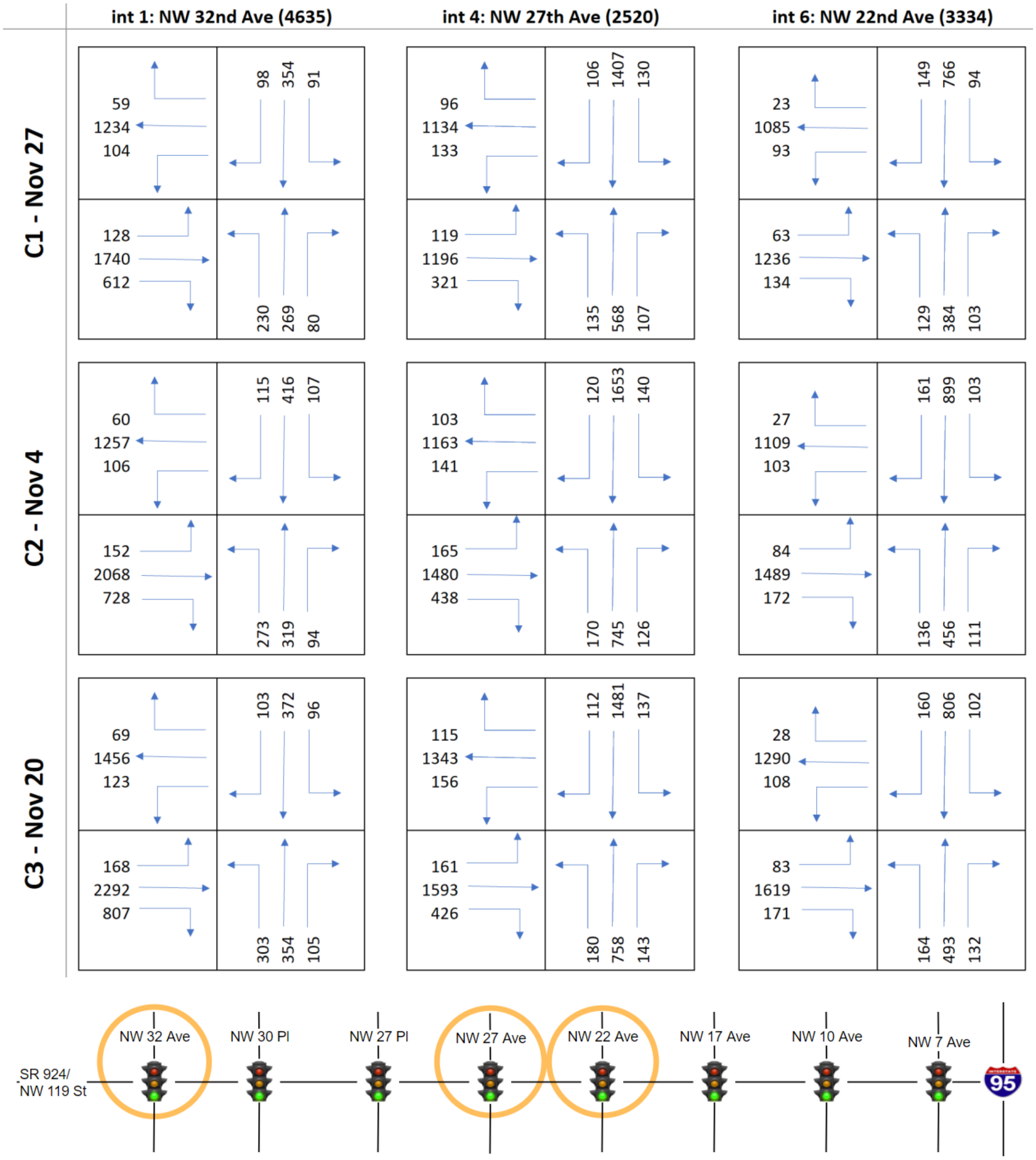

FIGURE 5. OUTPUT OF THE CLUSTERING PROCEDURE: AVERAGE VOLUMES OF THE SIGNATURE DAYS AT SELECTED INTERSECTIONS

4.5 IMPLEMENTATION AND EVALUATION OF THE PREDICTION ALGORITHMS

As stated earlier, the next step in the methodology is to use the results of the cluster analysis described in the previous sections as inputs to a data analytic-based prediction model that uses data mining/machine learning to predict the traffic state in the short-term future in real-time operations. The investigated prediction models are classification algorithms that predict the traffic state in the next 30 minutes as belonging to one of the pre-identified clusters, identified in the previous section. For that effect, a 15-minute data across all days at multiple locations (intersections) is used as input for the prediction algorithms. The study implemented and evaluated seven different data mining/machine learning approaches. The evaluated approaches to predict the traffic states categorized based on cluster analysis are the Decision Tree (DT), Random Forest (RF), Gaussian Naïve Bayes (GNB), Multinomial Logistic Regression (MLR), Support Vector Classification (SVC), K-nearest neighbors (KN), and Artificial Neural Network (ANN). Below are descriptions of these approaches.

Decision Trees and Tree Ensembles

The decision tree (DT) is a very popular supervised machine learning tool that can be used for both classification and regression. DT has been widely used in transportation engineering research literature. A DT can classify measurements and can also estimate the probability of an instant belonging to a particular class (for example, the probability of traffic breakdown occurrence). The DT algorithms search for the dominant attribute from all attributes. Then, this most dominant attribute is put on the top of the tree as decision node. A similar process is repeated for the other attributes at the next level of the DT. DT has been one of the most popular data mining techniques. It can work with high dimensional data, be developed in an efficient manner, and does not require any domain knowledge or parameter setting (Han and Kamber, 2006) (Tan et al., 2016). The results are also easy to present, and are well understood by humans. DT generally has good accuracy, but the accuracy may vary depending on the data. Scalability issues have been identified with the popular DT algorithms for very large data sets, and algorithms have been proposed for use on very large data sets. However, this should not be an issue for most transportation system and management and operations (TSMO) applications with currently available data.

Bayesian Classification

Bayesian classification uses the Bayes' probability theorem to predict the class membership probabilities. Naïve Bayesian classifiers assume the effect of each attribute on the classification is independent of each other to simplify the required computation. As an extension of this approach, the Bayesian Belief Networks allow the consideration of the dependencies between the attributes. For example, incident duration and severity can both affect the probability of secondary incidents. However, incident duration and severity are related to each other and thus

Bayesian Belief Networks rather than Naïve Bayesian classifiers could be used. It was reported that the performance of Naïve Bayesian classifiers can be comparable to classification DTs and some neural networks.

Artificial Neural Network

ANN is a very power machine learning technique and is considered to be the core of deep learning. It can deal with very the complex and large classification, recognition, prediction, and recommendation of action tasks. In an analogy to human brain structure, the ANN consists of nodes (emulating neurons) that are assembled in layers, and links that connect these nodes. The most basic form is a single node referred to as a perceptron, which has inputs and outputs that can be trained (e.g., to solve classification problems). However, most utilized ANN are multi-layer. The most common ANN is a supervised learning method referred to as the multilayer perception (MLP). An MLP consists of one input layer, one or more intermediate layers referred to as hidden layers, and an output layer. When the ANN has two or more hidden layers, it is called a deep neural network. MLP is trained using an optimization process referred to as the backpropagation training algorithm. In many applications, even in cases of complex functions, it has been shown that a single hidden layer is sufficient, as long as it has enough neurons. However, deep networks have proven to be much more efficient in modeling complex functions with fewer numbers of neurons (Geron, 2017).

Support Vector Machine

SVM is a powerful supervised machine learning tool allowing classification, regression, and outlier detection. The SVM algorithm classifies the data instances in a manner that minimizes the possibility of the misclassification when used to classify new instances not used in the training. Thus, it is less susceptible to overfitting compared to DTs. Model parameters can be selected to reduce the impacts of outliers on the training. linear SVM classifiers separate the instances into different classes by straight lines. In some cases, linear SVM is not sufficient, and the nonlinear SVM classification has been used. Thus, the option of nonlinear classifiers has to be assessed to determine if it produces better results. Classification and regression using SVM are highly accurate, since SVM algorithms can deal with nonlinear decision boundaries. They are much less likely to over-fit the model to the training data, and can provide a compact description of the learned model (Han and Kamber, 2006). However, the computation associated with SVM is slow and not efficient for large data.

K-Nearest Neighbor

KNN works by using the values or classes of the “nearest neighbors” to the data point to find its value or for classification. For classification, the classes of all neighbors of the data point are identified and counted. The class with the highest count is assigned to the data point. For regression, the value of the data point is determined by averaging the values of all neighbors. The user must select the specific distance metric to use in determining the nearest neighbor.

The classification algorithms were evaluated in terms of the accuracy as well as the areas under the curve of the receiving operating characteristic (AUC ROC), and precision recall curve (AUC

PRC). In order to assess the accuracy of the predictions in a classification model, one of the first metrics to look at is the accuracy score, which is computed as the ratio of the correct predictions made by the model and the total number of predictions, that is, the fraction of the correct predictions over $n_{samples}$ as follows:

$$accuracy\ (y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(\hat{y}_i = y_i) \tag{1}$$

The AUC ROC is determined based the plot of the True Positive Rate (TPR) versus the False Positive Rate (FPR) and reflects how well the model can distinguish between two different and is a measure of how well the model is predicting when there are imbalanced classes. Imbalanced classes are a common problem in machine learning and occurs when the data points in one class is significantly higher than the observation in other classes and can affect the performance of the machine learning algorithms.

In order to prepare the input data for the prediction algorithms, the records were filtered and aggregated for every 30 minutes. This allows the use of the models to predict the traffic state (identified by the predetermined cluster) for the next 30 minutes based on the available historical data during the previous “n” 30-minute periods since the start of the AM peak. For example, considering that the AM peak started at 7:00 am, data for the 7:00 am to 7:30 am interval across multiple locations are used as input to predict the of the traffic state that will likely occur from 7:30 am to 8:00 am. Then, in the second prediction interval during the AM peak, volume data from 7:00 am to 7:30 am and 7:30 am to 8:00 am are used to predict the traffic state that will likely occurs from 8:00 am to 8:30 am, and so on. This will allow the implementation of traffic responsive control plans that are activated based on the prediction of traffic conditions in the next 30 minutes.

The results of the evaluation of the implemented prediction models are summarized in Table 1. Higher values of the Accuracy, ROC AUC, and PR AUC indicate better performance of the prediction model. By observing the summary of the results depicted in Table 1, it can be concluded that in general, the predictions become more accurate and the areas under the curves become larger as more data are incorporated into the input, as data from more 30-minute intervals are incorporated in the training data used as input to the model in later 30-minute intervals of the AM peak period. In some cases, Table 1 shows that algorithms such as the MLR, KNN, and ANN can produce predictions with a high degree of accuracy even with just half an hour input data (for the first prediction interval). It is clear from the table as well that, the ANN model performed better than the other algorithms by producing the best results overall for each stage of the prediction.

TABLE 1. SUMMARY OF THE EVALUATION OF THE IMPLEMENTED PREDICTION MODELS

Data Input Period	***7:00 am - 7:30 am***			***7:00 am - 8:00 am***			***7:00 am - 8:30 am***		
Predicting Period	***7:30 am - 8:00 am***			***8:00 am - 8:30 am***			***8:30 am - 9:00 am***		
Algorithm	**Accuracy**	**ROC AUC**	**PR AUC**	**Accuracy**	**ROC AUC**	**PR AUC**	**Accuracy**	**ROC AUC**	**PR AUC**
Decision Tree (DT)	0.375	0.530	0.349	0.750	0.810	0.646	0.750	0.810	0.646
Random Forest (RF)	0.500	0.710	0.580	0.875	0.900	0.870	0.875	0.910	0.831
Gaussian Naïve Bayes (GNB)	0.500	0.570	0.453	0.750	0.880	0.778	0.975	0.930	0.873
Multinomial Logistic Regression (MLR)	0.750	0.870	0.847	0.875	0.950	0.900	0.875	0.900	0.789
Support Vector Classification (SVC)	0.500	0.830	0.384	0.750	0.980	0.899	0.675	0.790	0.747
K-Nearest Neighbors (KN)	0.750	0.770	0.620	0.875	0.820	0.577	0.875	0.860	0.868
Artificial Neural Network (ANN)	0.750	0.880	0.820	0.875	0.900	0.789	0.875	0.930	0.883

Note: The algorithm in bold font indicates a better performance model.

4.6 CONCLUSIONS

As mentioned earlier, the clustering task proved to be challenging due to the lack of variability and the inconsistencies detected in the data from the Gainesville location. Several blank records in the data made it necessary to eliminate a large number of records. This affected the model sensitivity as data from a large proportion of the periods were lost. Another aspect that contributed to the complexity of the analysis in the first location was the fact there was only one data source available with nothing to validate against and the quality of the data was determined to be suspicious based on careful examinations of the clustering results.

The South Florida location had the advantage of having data available from different sources including high-resolution controller data. This helped with the implementation of the clustering procedure and facilitated the identification of three signature days that clearly represent the traffic states at that location during the AM peak. The three states represent relatively low, medium, and heavy volumes that can be used as inputs to signal optimization models to identify signal timing plans that can be used as the plans to select from in systems that use the TRPS control. Further examination of the resulting volumes, plans, and the resulting performance can be done as described in Chapter 5 to determine if the resulting plans are significantly different to justify utilizing all of them in TRPS control. In some cases, for example, it may be determined that only two of the three plans can be justified for this purpose.

This study also explored a methodology and evaluated multiple algorithms for the short-term prediction of the traffic state for the next half an hour. The traffic states are predicted as belonging to one of the three clusters identified based on the results of the cluster analysis. This prediction in real-time operations can be used to activate the signal timing plan developed for the signature day for the cluster that represent the predicted state. The results revealed that the ANN algorithm, produced the best results in terms of accuracy and areas under the curve. Thus, the ANN prediction model will be used in the implementation and evaluation of the prediction to activate the signal timing plans, described in Chapter 5.

CHAPTER 5: EVALUATION OF THE PREDICTIVE TRAFFIC RESPONSIVE SIGNAL CONTROL

5.1 INTRODUCTION

This chapter presents the evaluation of a predictive TPRS developed based on traffic clustering and prediction, as explained in the previous chapter. First, the research team randomly selected ten days and obtained the respective real-world traffic data. These data were used as an evaluation dataset and are discussed in the first subsection of this chapter. Next, we defined scenarios to evaluate the traffic control plan development and selection. For these scenarios, the study optimized the timing plans for the traffic signals based on the traffic volumes of the signature day of each cluster as well a random day among the evaluation dataset. The process of signal optimization is explained next. Finally, the results of the comparison between different scenarios are presented.

5.2 EVALUATION DATASET

Table 2 shows the data obtained for ten days randomly selected to evaluate the efficiency of the traffic patterns identification and prediction methodology which explained in Chapter 4. This table presents the cluster that each day belong to in addition to the cluster that is predicted from applying the prediction model in an emulated real-time environment. Given that the centroids of cluster 2 and cluster 3 were relatively close to each other (i.e., similar clusters), it was decided to merge these clusters for the purposes of the evaluation, thus Error! Reference source not found. refers to cluster 2 or cluster 3 indistinctively as cluster 2. Under the said consideration, Table 1 also shows that the prediction algorithm could accurately predict the traffic cluster for eight out of ten days. Please, note that the prediction of different clusters than the true clusters is expected to be for those days that are on the boundary between the two clusters and thus the impact of the misclassification can be less significant as confirmed by the results presented later in this chapter.

TABLE 2. TRUE AND PREDICTED CLUSTERS FOR TEN RANDOM DAYS

Date	True Cluster	Predicted Cluster	Accuracy of the Prediction Algorithm
11/1/2019	2	2	T
11/5/2019	1	2	F
11/7/2019	2	2	T
11/8/2019	2	2	T
11/13/2019	2	2	T
11/15/2019	1	1	T
11/19/2019	1	2	F
11/21/2019	1	1	T
11/25/2019	1	1	T
11/26/2019	1	1	T

Accuracy of prediction algorithm	80%

5.3 EVALUATED SCENARIOS

This study used five scenarios to investigate the improvement in system performance due to the TRPS strategy based on the traffic patterns identification and as developed in this scenario. In each scenario, the utilized traffic signal timing is optimized for different traffic patterns. The performance of the timing plans are assessed **Error! Reference source not found.** based on their performance for the ten days randomly selected for use in the evaluation. The timing plans in the five scenarios are calculated based on the followings.

- Scenario 1 - Base condition using the existed signal timing in the field: In this scenario, the existing signal timing is modeled in the network. Scenario 1 is used as a benchmark to determine whether the recommended process provides improved performance compared to the existing signal timing.

- Scenario 2 - Optimized signal timing based on the signature days of the true clusters: This scenario is meant to show the performance of the network when using plans optimized for the signature days of the true clusters. This means that this scenario assumes that the prediction algorithm is able to accurately predict the traffic patterns in all days. Thus, this scenario test the performance of the clustering in identifying the traffic patterns but not the performance of the predictive model.

- Scenario 3- Optimized signal timing based on the signature days of the predicted clusters: This scenario is meant to show the performance of the network when using plans optimized for the signature days of the predicted clusters. This scenario evaluates the impacts of using signal timing plans that may not be optimal for those days, for which the predicted clusters are different from the true clusters (in our case study two of the ten days). Thus, this scenario test the combined performance of the clustering and the predictive model.

- Scenario 4 - Optimized signal timing for the signature day of the entire database: In this scenario, a signature day was selected for the entire database, representing the traffic demand for the entire database. In this scenario, the utilized traffic signal timing is that optimized for this database-wide signature day. This scenario is designed to evaluate the benefits of activating signal timing plans based on the predicted clusters versus having a fixed timing plan optimized for the whole year. This scenario test the performance of using the signature day identification methodology based on the whole year data rather than using a random day in the year or a day selected based on limited amount of data in the optimization of signal timing. However, this scenario does not test the performance of the clustering or the predictive model.

- Scenario 5 - Optimized signal timing for a randomly selected day in the database: In this scenario, the signal timing plan is optimized based on the data for a day selected randomly among the ten sample days (November 13 is randomly selected in the case study). This scenario is meant to assess the traffic performance when the traffic signals are designed and optimized based on collected traffic data from a random day or a day selected based on limited amount of data, as may be used in common practice.

5.4 SIGNAL TIMING OPTIMIZATION

Software programs such as TRANSYT-7F, PASSER II, SYNCHRO, Highway Capacity Software (HCS), and PTV Vistro have been used to optimize the timing plans for signalized intersections and evaluate the overall performance of the network. This study used the seventh version of the HCS (HCS7) to develop the optimal signal timing plans for the evaluated scenarios and assess the performance. The HCS7 follows the procedures of the sixth edition of the highway capacity manual to evaluate the traffic performance.

To calculate the optimal signal plan, the network and intersections were coded in the HCS-Streets module (McTrans Center, 2021). As recommended by the HCS7, the GA optimization algorithm was used to optimize the cycle length, followed by splits and offsets. The existing signal plan was input as the initial signal timing plan in the optimization. The optimization uses this information as the initial timing in the optimization process. The overall delay was selected as the objective function for the optimization algorithm. The GA optimization parameters were set as recommended by the HCS7.

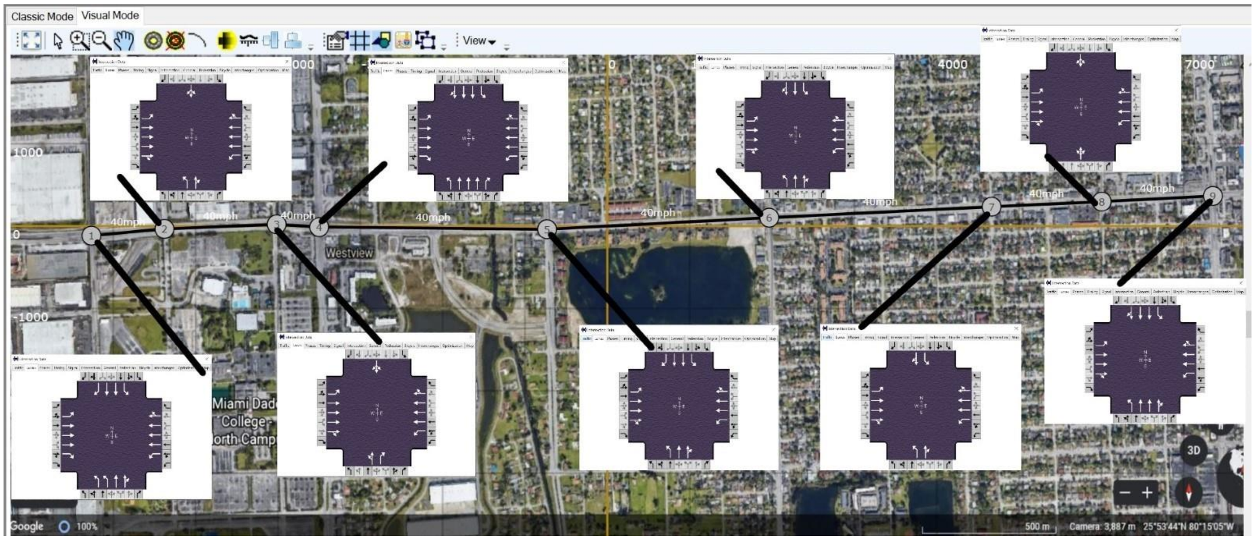

FIGURE 6. EXISTING INTERSECTIONS AND GEOMETRY OF119TH STREET AS INPUT TO HCS7

5.5 SCENARIO EVALUATION

The average travel time in the system was used as the performance measure to evaluate and estimate the benefits of the traffic clustering and prediction method to support signal control, **Error! Reference source not found.** provides the average travel time for the ten sample days and for each of the five evaluated scenarios described above. As shown, Scenarios 2 and 3 have the same average performance across all scenarios and similar performance for the days with wrong prediction, confirming that the patterns for these two days (November 5th and November 19th) fall on the boundary between the two clusters. Scenario 2 and 3 produced the lowest travel times among the evaluated scenarios with 7% improvement compared to the existing plan (Scenario 1), 4% compared to optimizing for a fixed signal timing plan based on a signature day for the whole data (Scenario 2), and 17% improvement compared to optimizing signal timing for a random day in the data.   This shows that the TRPS based on traffic pattern identification and prediction, as assessed in the evaluation of Scenario 4, has the potential of improving traffic performance compared to other assessed optimization scenarios.

For all days except two days (November 13 and November 19) the optimized signal timing for the signature day of the entire database provides a lower travel time than the existing traffic signal timing. This signifies the importance of selecting a representative day and optimize signal timings based on date collected from multiple days and processing the data to identify the best day to use in the optimization.

TABLE 3. AVERAGE TRAVEL TIME FOR EACH SCENARIO FROM HCS7.

Date	Existing situation (Sc.1)	Optimized for true cluster (Sc.2)	Optimized for predicted cluster (Sc.3)	Optimized for the signature day (Sc.4)	Optimized for the random day (Sc.5)

	Average travel time (s)				
11/1/2019	386	351	351	375	483
11/5/2019	350	319	327	331	336
11/7/2019	382	348	348	355	394
11/8/2019	389	362	362	367	415
11/13/2019	388	367	367	420	576
11/15/2019	392	368	368	368	399
11/19/2019	398	388	377	420	461
11/21/2019	400	369	369	369	507
11/25/2019	390	369	369	371	377
11/26/2019	378	354	354	358	358
Weighted average TT over ten days	385	360	359	375	433
Ratio of Scenario TT / existing TT	1	0.93	0.93	0.97	1.12

CONCLUSIONS

Although adaptive signal control is a powerful strategy to address the day-to-day variation in traffic demands, most intersections in the United States are still operating under TOD strategies sue to the high cost and the additional requirements associated with the systems. In addition, adaptive signal control may not be beneficial to address all operation performance issues. TRPS strategies have been proposed since the 1970s as an alternative to TOD that can address some of the issues associated with day-to-day variations in traffic patterns. The requirements of these strategies are much lower than those of adaptive signal control strategies. However, there are several limitations and issues associated with TRPS that have limited the adoption of these strategies in This study developed and evaluated a TRPS strategy based on supervised and unsupervised machine learning combined with signal timing optimization to addresses the issues with traditional TRPS. The strategy fills an important gap in providing a proactive traffic control that makes use of ATSPM measures-based data that are becoming available sources including high resolution controller data.

This study used k-means clustering, a widely used clustering algorithm to identify the traffic patterns to use in signal timing plan development. The cluster analysis identified three traffic patterns or states that have different demands in the AM peak of the case study used in the project. The three states represent relatively low, medium, and heavy volumes that can be used as inputs to signal optimization models to identify signal timing plans that can be used as the plans to select from in systems that use the TRPS control. Further examination of the resulting volumes, plans, and the resulting performance indicates that only two patterns should be used. When the demands from the three patterns were used in signal timing optimization, it was found that signal timing optimization software produced similar signal timings for two of the three identified patterns. Thus, these two patterns were combined in one pattern and only two patterns were used in the optimization and real-time activation of the signal timing plans.

This study also explored a methodology and evaluated multiple algorithms for the short-term prediction of the traffic state for the next half an hour. The traffic states are predicted as belonging to one of the patterns identified based on the results of the cluster analysis. This prediction can be used in real-time operations to activate the signal timing plan developed for the signature day of the cluster that represent the predicted state. The evaluated algorithms to predict the traffic states categorized based on cluster analysis are the Decision Tree (DT), Random Forest (RF), Gaussian Naïve Bayes (GNB), Multinomial Logistic Regression (MLR), Support Vector Classification (SVC), K-nearest neighbors (KN), and Artificial Neural Network (ANN). The results revealed that the ANN algorithm, produced the best results in terms of various prediction performance measures.

The performance of the predictive TRPS based on clustering and prediction was assessed by evaluating five different scenarios of signal timing plan selection. The results showed that the predictive TRPS method can decrease the travel time by 7 percent compared to existing traffic signals, 4% compared to optimizing for a fixed signal timing plan based on a signature day for the whole database, and 17% compared to optimizing signal timing for a random day in the data. This shows that the TRPS based on traffic pattern identification and prediction has the potential of improving traffic performance compared to other assessed optimization scenarios. For eight of the ten days used in the evaluation, the optimized signal timing for the signature day of the entire database provides a lower travel time than the existing traffic signal timing. This signifies the importance of selecting a representative day and optimize signal timings based on date collected from multiple days and processing the data to identify the best day to use in the optimization.

REFERENCES

Adu-Gyamfi, Y., and Zhao, M. 2018. "Traffic Speed Prediction for Urban Arterial Roads Using Deep Neural Networks." *International Conference on Transportation and Development.*

Azimi, M., and Zhang, Y. 2010. "Categorizing Freeway Flow Conditions by Using Clustering Methods." *Transportation Research Record* Vol. 2173, No. 1. pp. 105-114.

Azizi, L., and Hadi, M. 2021. "Using Traffic Disturbance Metrics to Estimate and Predict Freeway Traffic Breakdown and Safety Events." Edited by SAGE Journals. *Transportation Research Record. Journal of the Transportation Research Board* 2675 (10). https://journals.sagepub.com/doi/10.1177/03611981211012422.

Chandra, R., and Gregory, C. 2010. "Insync adaptive traffic signal technology: Real-time artificial intelligence delivering real-world results." *InSync White Pap* p. 26.

Dell, P., and Mirchandani, B. 1995. "REALBAND: An approach for real-time coordination of traffic flows on networks." *Transp. Res. Rec.* Vol. 1494, pp. 106–116.

E. Hathaway, T. Urbanik, and S. Tsoi. 2012. "Memorandum: Cornell Road InSync System Evaluation. Kittelson Assoc. Memo. Transp. Oper. Innov. Demonstr. Eval. Cornell Road InSync Syst. Eval."

Elfar, A., Talebpour, A., and Mahmassani, H. 2018. *Machine Learning Approach to Short-Term Traffic Congestion Prediction in a Connected Environment.* Evanston, IL: Civil and Environmental Engineering. Northwestern University.

Eom, M., and Kim, B.-I. 2020. "The traffic signal control problem for intersections: a review." *Eur. Transp. Res. Rev.* Vol. 12, no. 1, pp. 1–20.

Federal Highway Administration. 2022. "Arterial Management Program." *Office of Operations. 21st Century Operations Using 21st Century Technologies.* https://ops.fhwa.dot.gov/arterial_mgmt/.

Federal Highway Administration. 2015. *Signal timing Manual, Second Edition.* Washington, DC: FHWA.

Florida Department of Transportation. 2018. *Final Operational Analysis for SR 924/NW 119th Street from NW 32 Avenue to NW5 Avenue.* Miami, FL: FDOT, District Six.

Florida Department of Transportation. 2021. *Statewide Arterial Management Program (STAMP).* Tallahassee, FL: FDOT.

Geron, A. (2017). Hands-on Machine Learning with Scikit-learn and tensor-flow (1st ed.). Sebastopol, CA., O'Reilly Media, Inc.

Gu, Y., Qian, Z., Xie, F., and Smith, S. 2016. "An Unsupervised Learning Approach for Analyzing Traffic Impacts Under Arterial Road Closures: Case Study of East Liberty in Pittsburgh." *Journal of Transportation Engineering* Vol. 142, No. 9, p. 401.

Hadi, M. 1990. *Improved Stategies for Traffic Responsive Control in Arterial Signal Systems.* Gainesville, FL: Ph.D. Dissertation.

Han, J. and Kamber, M. (2006). Data Mining: Concepts and Techniques. 2nd edition. San Francisco: Morgan Kaufmann Publishers.

Hosseini, M., and Talebpour, A. 2019. "Traffic Prediction Using Time-Space Diagram: A Convolutional Neural Network Approach." Edited by Transportation Research Record. *Journal of the Transportation Research Board* (SAGE Journals). https://journals.sagepub.com/doi/pdf/10.1177/0361198119841291.

Inc., Kimley Horn and Associates. 2012. "InSync Evaluation. Before and After Study." *Prepared for Pinellas County, FL.*

Iqbal, S., Hadi, M., and Xiao, Y. 2017. *Predicting Arterial Breakdown Probability: A Data Mining Approach.* Miami, FL: Department of Civil and Environmental Engineering. Florida International University.

J. N. Hutton, C. D. Bokenkroger, and M. M. Meyer. 2010. "Evaluation of an adaptive traffic signal system: route 291 in Lee's Summit, Missouri."

Kianfar, J., and Edara, P. 2013. "A Data Mining Approach to Creating Fundamental Traffic Flow Diagram." *Procedia-Social Behavior Science* Vol. 1934, No. 1, pp. 430-439.

Lee, S., Wong, S. C., and Varaiya, P. 2017a. "Group-based hierarchical adaptive traffic-signal control part I: Formulation." *Transp. Res. part B Methodol.* Vol. 105, pp. 1–18.

Lee, S., Wong, S. C., and Varaiya, P. 2017b. "Group-based hierarchical adaptive traffic-signal control Part II: Implementation." *Transp. Res. part B Methodol.* Vol. 104, pp. 376–397.

Li, W., and Ban, X. 2019. *A Deep Learning Approach for Lane-Based Short-Term Traffic Volume Prediction at Signalized Intersections.* Seattle, USA: University of Washington.

Lo, H. K., and Chow, H. F. 2002. "Adaptive traffic control system: Control strategy, prediction, resolution, and accuracy." *J. Adv. Transp.* Vol. 36, no. 3, pp. 323–347.

McTrans Center, University of Florida Transportation Institute. (2021). Highway Capacity Software. User Guide. Streets Module. https://mctranswordpresssa.blob.core.windows.net/uploads/2022/01/Streets.pdf

Mueller, E. A. 1970. "Aspects of the history of traffic signals." *IEEE Tranportation and Veicles Technologies* Vol. 19, no. 1, pp. 6–17.

Ren, J., and Rahman, A. 2009. "Automatically Balancing Intersection Volumes in a Highway Network." *Transportation Research Board.* Washington, DC: 12th TRB Transportation Planning Applications Conference.

Rhythm Engineering. 2014. *InSync's Performance.* http://rhythmtraffic.com/insyncs-performance.

Siromaskul, S., and Selinger, M. 2010. "Insync: The next generation of adaptive signal systems."

Sprague, D. 2012. "Adaptive signal timing comparison between the InSync and QuicTrac adaptive signal systems installed in Colorado." *Colorado. DTD Applied Research and Innovation Branch.*

Sun, L., and Zhou, J. 2005. "Development of Multiregime Speed-Density Relationships by Cluster Analysis." *Transportation Research Record* Vol. 1934, No. 1, pp. 64-71.

Tan, P.-N., Steinbach, M. and Kumar, V. 2016. *Introduction to Data Mining.* London: Pearson Education, Inc.

Tariq, M. T., Hadi, M., and Saha, R. C. 2021. "Using High-Resolution Controller Data in the Calibration of Traffic Simulation Models." *Transportation Research Record. Journal of the Transportation Reseach Board.*

Theofilatos, G. Yannis, E. I. Vlahogianni, and J. C. Golias. 2017. "Modeling the Effect of Traffic Regimes on Safety of Urban Arterials: The Case Study of Athens." *Journal of Transportation Engineering (English Edition)* Vol. 4, No. 3, pp. 240-251.

TJKM Transportation Consultants. 2011. "“valuation of Main Street Adaptive Traffic Signal System." *Report Prepared for the City of Salinas.*

Trafficware. 2012. "SynchroGreen Technical White Paper." Accessed May 23, 2014. http://www.trafficware.com/wp-content/uploads/2014/04/SynchroGreen-Technical-White-Paper_10-03-2013.pdf.

Treiber, M., Kesting, A., and Helbing, D. 2006. "Undrestanding Widely Scattered Traffic Flows, the Capacity Drop, and Platoons as Effects of Variance-Driven Time Gaps." *Physical Review* 74 (1): 016123.

US Department of Transportation. 2022. *Federal Highway Administration. Office of Operations.* July 29. https://ops.fhwa.dot.gov/arterial_mgmt/performance_measures.htm.

Whitelock, R. 2014. "Adaptive Traffic Control Systems (ATCS)-Part 5: SynchoGreen." *IMSA J.*

Wisconsin DOT. 2018. *Traffic Analysis, Modeling and Data Management. Intersection Volume Balancing Tool (version 5.8.1).* Madison, WI: State of Wisconsin Department of Transportation.

Wiscosin DOT. 2019. *Traffic Engineering, Operations & Safety Manual. Traffic Analysis and Modeling. Section 5, Data Assembly and Preparation.* Madison, WI: WDOT.

Wu, X., Liu, X., and Geroliminis, N. 2011. "An Empirical Analysis on the Arterial Fundamental Diagram." *Transportation Research Record Part B* Vol 45, No.1, pp. 255-266.

Wunderlich, K. E., Vasudevan, M., and Wang, P. 2019. "Guidelines for Applying Traffic Microsimulation Modeling Software. Traffic Analysis Toolbox (TAT) Volume III: 2019 Update to the 2004 Version. No. FHWA-HOP-18-036." Federal Highway Administration, Washington, DC.

Xia, J., and M. Chen. (2007). A Nested Clustering Technique for Freeway Operating Condition Classification. Department of Civil Engineering, University of Kentucky, Lexington, KY.

Xia, J., Huang, W., and Gou, J. 2012. "A Clustering Approach to Online Traffic State Identification Using ITS Data." *KSCE J. Civil Engineering.* Vol. 16 No. 3, pp. 426-432.

Yang, S., Wu, J., Qi, G., and Tian, K. 2017. "Analysis of Traffic State Variation Patterns for Urban Road Network Based on Spectral Clustering." *Advanced Mechanics and Engineering* Vol. 9, No. 9, p. 168.