Fall 2016

An artificial neural network model for predicting freeway work zone delays with big data

Bo Du
New Jersey Institute of Technology

ABSTRACT

AN ARTIFICIAL NEURAL NETWORK MODEL FOR PREDICTING FREEWAY WORK ZONE DELAYS WITH BIG DATA

by
Bo Du

Lane closures due to road reconstruction and maintenance have resulted in a major source of non-recurring congestion on freeways. It is extremely important to accurately quantify the associated mobility impact so that a cost-effective work zone schedule and an efficient traffic management plan can be developed. Therefore, the development of a sound model for predicting delays or road users is desirable.

A comprehensive literature review on existing work zone delay prediction models (i.e., deterministic queuing model and shock wave model) is conducted in this study, which explores the advantages, disadvantages, and limitations of different modeling approaches. The performance of those models seems restricted to predict congestion impact under space-varying (i.e., road geometry, number of lanes, lane width, etc.) and time-varying (i.e., traffic volume) conditions. To advance the delay prediction accuracy, a multivariate non-linear regression (MNR) model is developed first by incorporating big data to capture the relationship of speed versus the ratio of approaching traffic volume to work zone capacity for work zone delay prediction. The MNR model demonstrates itself able to predict spatio-temporal delays with reasonable accuracy.

A more advanced model called ANN-SVM is developed later to further improve the prediction accuracy, which integrates a support vector machine (SVM) model and an artificial neural network (ANN) model. The SVM model is responsible to predict work zone capacity, and the ANN model is responsible to predict delays. The ultimate goal of

ANN-SVM aims to predict spatio-temporal delays caused by a work zone on freeways in the statewide of New Jersey subject to road geometry, number of lane closure, and work zone duration in different times of a day and days of a week. There are 274 work zones with complete information for the proposed model development, which are identified by mapping data from different sources, including OpenReach, Plan4Safety, New Jersey Straight Line Diagram (NJSLD), New Jersey Congestion Management System (NJCMS), and INRIX. Big data analytics is used to examining this massive data for developing the proposed model in a reliable and efficient way.

A comparative analysis is conducted by comparing the ANN-SVM results with those produced by MNR, RUCM (NJDOT Road User Cost Manual approach), and ANN-HCM (the ANN model with work zone capacity suggested by Highway Capacity Manual). It is found that ANN-SVM in general outperforms other models in terms of prediction accuracy and reliability. To demonstrate the applicability of the proposed model, an analysis tool, which adapts to ANN-SVM, is developed to produce graphical information. It is worth noting that the analysis tool is very user friendly and can be easily applied to assess the impact of any work zones on New Jersey freeways. This tool can assist transportation agencies visualize bottlenecks and congestion hot spots caused by a work zone, effectively quantify and assess the associated impact, and make suitable decisions (i.e., determining the best starting time of a work zone to minimize delays to the road users). Furthermore, ANN-SVM can be applied to develop, evaluate, and improve traffic management and congestion mitigation plans and to calculate contractor penalty based on cost overruns as well as incentive reward schedule in case of early work competition.

AN ARTIFICIAL NEURAL NETWORK MODEL FOR PREDICTING FREEWAY WORK ZONE DELAYS WITH BIG DATA

by
Bo Du

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Transportation

John A. Reif, JR. Department of Civil and Environmental Engineering

January 2017

.

APPROVAL PAGE

AN ARTIFICIAL NEURAL NETWORK MODEL FOR PREDICTING FREEWAY WORK ZONE DELAYS WITH BIG DATA

Bo Du

Dr. Steven I-Jy Chien, Dissertation Advisor Date
Professor of John A. Reif, Jr. Department of Civil & Environmental Engineering, NJIT

Dr. Janice Daniel, Committee Member Date
Associate Professor of John A. Reif, Jr. Department of Civil & Environmental Engineering, NJIT

Dr. Joyoung Lee, Committee Member Date
Assistant Professor of John A. Reif, Jr. Department of Civil & Environmental Engineering, NJIT

Dr. Lazar Spasovic, Committee Member Date
Professor of John A. Reif, Jr. Department of Civil & Environmental Engineering, NJIT

Dr. Kyriacos Mouskos, Committee Member Date
Research Professor, Transportation Institute, City College of New York

BIOGRAPHICAL SKETCH

Author: Bo Du

Degree: Doctor of Philosophy

Date: January 2017

Undergraduate and Graduate Education:

- Doctor of Philosophy in Transportation,
  New Jersey Institute of Technology, Newark, NJ, 2017

- Master of Science in Transportation,
  New Jersey Institute of Technology, Newark, NJ, 2012

- Bachelor of Science in Transportation Engineering,
  Beijing University of Technology, Beijing, People's Republic of China, 2010

Major: Transportation

Publications:

Bo Du, Steven Chien, Joyoung Lee, Lazar Spasovic, and Kyriacos Mouskos. Artificial Neural Network Model for Estimating Temporal and Spatial Freeway Work Zone Delay Using Probe-Vehicle Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2573*, 2016, pp. 164-171.

Joyoung Lee, Zijia Zhong, Bo Du, Slobodan Gutesa, and Kitae Kim. Low-Cost and Energy-Saving Wireless Sensor Network for Real-Time Urban Mobility Monitoring System. *Journal of Sensors*, Vol. 2015, Article ID 685786, 8 pages, 2015. Doi: 10.1155/2015/684786

Bo Du and Steven Chien. Feasibility of Shoulder Use for Highway Work Zone Optimization. *Journal of Traffic and Transportation Engineering*, Vol. 1, No. 4, 2014, pp. 235-246.

Jiancheng Weng, Lili Liu, and Bo Du. ETC Data Based Traffic Information Mining Techniques. *Journal of Transportation Systems Engineering and Information Technology*, Vol. 10, No. 2, 2010, pp. 57-63.

Presentations:

Bo Du, Joyoung Lee, Kitae Kim, Steven Chien, Branislav Dimitrijevic. Freeway Traffic Flow Estimation Using Bluetooth-Based Probe Data. Presented at the 95th Transportation Research Board Annual Meeting, Washington, D.C., 2016.

Liuhui Zhao, Bo Du, and Steven Chien. Optimizing Work Zone Schedule with Floating Car Data Considering Traffic Diversion and Managed Lanes. Presented at the 95th Transportation Research Board Annual Meeting, Washington, D.C., 2016.

Bo Du, Liuhui Zhao, Amir Ibrahim, Juan Restrepo, and Joyoung Lee. Evaluating the Impact of Dynamic Shoulder Lane for Freeway Work Zone Congestion Mitigation. Presented at ITS America Annual Meeting & Expo, Pittsburgh, PA, 2015.

Bo Du, Joyoung Lee, Steven Chien, Branislav Dimitrijevic, and Kitae Kim. Short-term Freeway Work Zone Capacity Estimation Using Support Vector Machine Incorporated with Probe-vehicle Data. Presented at the 94th Transportation Research Board Annual Meeting, Washington, D.C., 2015.

Joyoung Lee, Zijia Zhong, Kitae Kim, Branislav Dimitrijevic, Bo Du, and Slobodan Gutesa. Examining Applicability of Small Quadcopter Drone for Traffic Surveillance and Roadway Incident Monitoring. Presented at the 94th Transportation Research Board Annual Meeting, Washington, D.C., 2015.

Joyoung Lee, Zijia Zhong, Brijesh Singh, Jeevanjot Singh, Branislav Dimitrijevic, Kitae Kim, Bo Du, Steven Chien, and Lazar Spasovic. WIMAP: Work Zone Interactive Monitoring Application. Presented at the 94th Transportation Research Board Annual Meeting, Washington, D.C., 2015.

Bo Du, Joyoung Lee, Branislav Dimitrijevic, Kitae Kim, and Steven Chien. Approach for Freeway Work Zone Capacity Estimation Incorporating Probe Vehicle Data. Presented at the 21st World Congress on Intelligent Transport Systems, the Cobo Convention Center, Detroit, Michigan, 2014.

Bo Du and Steven Chien. Impact of Shoulder Use and Capacity Reduction Factors on Highway Work Zone Optimization. Presented at the 93rd Transportation Research Board Annual Meeting, Washington, D.C., 2014.

This dissertation is dedicated to my beloved wife and family.

ACKNOWLEDGMENT

This dissertation would not have been possible without the help of many people in many ways. It is to them that I owe my sincere gratitude.

First of all, I would like to express my deepest appreciation to my advisor, Dr. Steven I-Jy Chien, for his unwavering support and encouragement. He provides me valuable and countless resources, insight, and intuition. His enthusiasm for research, kindness to people, and commitment of the highest standards will inspire me for the rest of my life.

I am very grateful to Dr. Joyoung Lee for his scientific advice and many insightful discussions on work zone capacity/delay predictions and simulation. I would like to thank the rest of my dissertation committee, Dr. Janice Daniel, Dr. Lazar Spasovic, and Dr. Kyriacos Mouskos for their interests in my research progress. Their valuable comments make the dissertation more solid and sound.

I am lucky to have worked in a friendly environment where I received friendship and help from a lot of former and current members in our program. I extend my thanks to Mr. Brijesh Singh and Ms. Liuhui Zhao for their help on coding and data analysis. My appreciation also goes to Mr. Branislav Dimitrijevic, Mr. Zijia Zhong, Mr. Mohammed Hasan, Dr. Kitae Kim, and Dr. Dejan Besenski for their informative discussion on various topics that helped me improve my knowledge in different areas and spark new ideas.

I also appreciate support from the New Jersey Department of Transportation, Intelligent Transportation System Resource Center, and New Jersey Institute of

Technology associated with the research project “Feasibility of Lane Closures Using Probe Data”.

Most of all, I would like to thank my family: my parents and parents-in-law, for their unconditional love and support; and my wife, Xinyue, for having faith on me and always encouraging me to challenge myself to move forward.

TABLE OF CONTENTS

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

Table **Page**

LIST OF TABLES
(Continued)

Table **Page**

LIST OF FIGURES

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF ACRONYMS

AADT	Average Annual Daily Traffic
ANN	Artificial Neural Network
BP	Back Propagation
BPR	Bureau of Public Roads
CORSIM	Corridor Simulation
DB	Database
DOT	Department of Transportation
ESA	Exhaustive Search Algorithm
FHWA	Federal Highway Administration
GPS	Geographic Positioning System
HCM	Highway Capacity Manual
iPeMs	Iteris Performance Management System
ITS	Intelligent Transportation System
LM	Levenberg-Marquardt
MAC	Medium Access Control
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MNR	Multivariate Non-linear Regression
MSD	Mean Squared Deviation
NJ	New Jersey
NJDOT	New Jersey Department of Transportation
NJCMS	New Jersey Congestion Management System
NJSLD	New Jersey Straight Line Diagram

QUEWZ	Queue and User Cost Evaluation of Work Zone
RBFNN	Radial Basis Function Neural Network
RFID	Radio-frequency Identification
RILCA	Rutgers Interactive Lane Closure Application
RMSE	Root Mean Square Error
RTMS	Remote Traffic Microwave Sensor
RUCM	Road User Cost Manual
SVM	Support Vector Machine
TMC	Traffic Message Channel
WIMAP-P	Work Zone Interactive Management Application - Planning
WZCAT	Work Zone Capacity Analysis Tool

LIST OF SYMBOLS

A, B	-	Coefficients of the ANN model
a_{ij}, b_{ij}, c, d	-	Coefficients of the MNR model
c_h	-	The hth actual observation value
$\hat{c}_h$	-	The hth predicted value
$\bar{c}$	-	The mean of observations
C	vphpl	Normal capacity
C_d	$/zone	Delay cost to road users caused by the work zone
C_w	vph	Work zone capacity
d	mi	Work zone length
d_i	mi	Distance from segment *i* to the work zone
D	veh-hr	Total queue delay caused by the work zone
e_r	-	The rth input variable of the ANN model
f	-	Function of the SVM-based work zone capacity prediction model
f_{HV}	-	The heavy-vehicle adjustment factor
g	-	Function of ANN-based work zone delay prediction model
h	-	Index of observation
H	-	Total number of observations
i	-	Index of segments in upstream of work zone
I	vphpl	The adjustment factor for type and intensity of work activity
j	-	Index of time interval
k	-	Kernel function
l	-	Total number of training data sets in the SVM model
l_i	mi	Length of freeway segment *i*

L_j	mi	Queue length at time *j*
m	-	Number of segments in upstream of work zone
M	-	Total number of input variables in the ANN model
n	-	Number of time intervals since the beginning of a zone till 2 hours after the work zone has been removed
N	-	Number of training vectors in the SVM model
N_o	-	Number of open lanes
N_T	-	Total number of lanes
p, q	-	Indices of training data set in SVM model
P_c, P_t	-	The percent of passenger cars and heavy vehicles
Q_j	vph	Traffic volume approaching the work zone at time *j*
R	vph	Manual adjustment for on-ramps
R^2	-	Coefficient of determination
R_o	-	The open-lane ratio (i.e., N_o/N_T)
s_{ij}	mph	Average speed of segment *i* at time *j* under normal condition
t	hr/veh	Added travel time
T	-	Transposition of the matrix
v_{ij}	-	Weighted speed of segment *i* at time *j*
v_w	mph	Work zone speed
v_u	mph	Unrestricted speed
V_{ij}	vph	Traffic volume of segment *i* at time *j*
x	-	Training vector in the SVM model
y_{ij}	mph	Average speed of segment *i* at time *j* under work zone condition
$\hat{y}_{ij}$	mph	Predicted average speed of segment *i* at time *j* under work zone condition

α_p, α_q	-	Vectors of Lagrange multipliers
β	-	Constant in the SVM model
γ	-	Parameter in kernel function
δ	-	Work Zone capacity reduction factor
λ	-	Regularization parameter in the SVM model
μ_c, μ_t	$/veh-hr	Values of travel time delay for passenger cars and heavy vehicles
ξ_p	-	Slack variable in the SVM model
τ_{ij}	-	Binary variable of segment *i* at time *j*
φ	-	Non-linear transformation from
ω	-	Vector of coefficients in the SVM model

CHAPTER 1

INTRODUCTION

1.1 Background and Problem Statement

Transportation systems, especially roadway networks, form an integral part for the movement of passengers and goods aiding in progressive economic development. Severe weather conditions, heavy usage, and growing demand deteriorate the condition and functioning of these interconnected road networks over time. This makes it necessary to conduct regular road rehabilitation and reconstruction projects, which require different configurations of lane closures depending on when and where these activities occur.

Closing a lane or even a shoulder of a road segment will cause disruptions in traffic flow, especially during peak hours. In the United States, 67% of federal funds were spent for roadway projects towards system preservation during 2011 and 2013 (Highway Statistics, 2013). These activities result in reduced travel time reliability and increased delays, crashes, wasted fuel, and frustration, which leads to increased road user costs; excess delay caused by lane closures in work zones is typically unavoidable. The U.S. road users lost approximately 552 million gallons of fuel and 482 million hours every year sitting in traffic jams caused by work zones (Facts and Statistics - Work Zone Delay, 2016). Furthermore, traveler delay is considered critical in making key decisions about staging and scheduling for roadway reconstruction projects.

The 1998 Federal Highway Administration (FHWA) report identifies this issue and recommends the development of a sound tool to predict and quantify work zone delays. Developing a method to predict the road user cost, delay and related traffic measures (i.e., speed, queue length, emissions, etc.) can aid in implementing appropriate counter

measures to mitigate the impacts, which is important for successful work zone management. Hence, developing a sound tool that provides reliable predictions for speed, delay and queue development due to work zone activities will help move traffic more efficiently and reduce motorist inconvenience by effectively planning work phasing and arranging detour routes.

The deterministic queuing concept has been widely used to predict work zone delay because of its simplicity, which uses approaching traffic volume and work zone capacity as inputs (Abraham and Wang, 1981; Dudek and Richards, 1982; Chien and Schonfeld, 2001; Weng and Meng, 2013; Du and Chien, 2014). One drawback of the deterministic queuing model is that it usually underestimated the delay if the stochastic nature of traffic flow and the heterogeneous geometric conditions either were oversimplified or neglected (Chien et al., 2002; Tang and Chien, 2008). In addition, the approaching traffic speed is assumed to be constant in many previous studies (Chien et al. 2002; Karim and Adeli 2003; Habtemichael et al. 2015). Therefore, the deterministic queuing model shall be advanced in order to improve its prediction accuracy.

Over the past two decades, the focus of modeling approaches in predicting work zone delay has transitioned from multivariate temporal correlation to multivariate spatio-temporal correlation and from parametric to non-parametric forms. As machine learning techniques can recognize patterns and adjust itself dynamically, the artificial neural network (ANN) models (e.g., radial basis function and multi-layer feed-forward neural networks, etc.) have been widely applied (Karim and Adeli, 2003; Jiang and Adeli, 2003; Ghosh-Dastidar and Adeli, 2006). However, previous ANN models have been limited by using spot speeds and traffic counts collected by loop detectors.

The transportation industry has been experiencing a wide variety of unprecedented massive traffic data obtained from different sources, such as infrastructure sensors, mobile devices, and floating cars. This new and rich data (big data) needs to be managed, communicated, interpreted, aggregated, and analyzed in a reliable and efficient way. The use of conventional data management tools is not able to uncover hidden patterns, correlations, and other insights, which would leave the huge amount of traffic data underutilized. Therefore, big data analytics, which creates richer and more complete picture of what's happening on the road, becomes a viable alternative for transportation engineers to analyze information efficiently and make decisions based on what they have learned.

For the freeway work zone impact analysis, leveraging big data analytics and advanced delay prediction methods (e.g., ANN models), the accuracy of predicted work zone speed and delay can then be elevated. The ability of big data analytics to work faster and stay agile gives transportation agencies a competitive edge they did not have before. In addition, it would help transportation agencies improve work zone scheduling, reduce delays and better serve motorists.

It is desirable to develop a sound model with big data for precisely predicting spatio-temporal work zone delay. Such a model helps transportation agencies have a good understanding of actual impacts of various highway reconstruction activities on a given network, and to be able to identify effective work zone mitigation measures. In addition, the proposed model can aid decision making by assessing the impacts of work zone activities in order to minimize disruptions to the traveling public. Furthermore, it can be utilized to accurately predict work zone road user costs more than the currently used

deterministic queuing models. More accurate road user cost prediction will allow for a more effective work zone congestion mitigation plan. This in turn will result in reduced travel delay, consumed fuel, and vehicle emissions.

1.2 Objective and Work Scope

The objective of this study is to develop a sound spatio-temporal freeway work zone delay prediction model with big data under various road geometric and work zone conditions.

To achieve the above objective, the limitations of previous and existing delay prediction models are thoroughly reviewed. Then, two prediction models are developed and evaluated under various road geometric and work zone conditions. The first model is a multivariate non-linear regression (MNR) model, which utilizes the big data collected from various data sources including work zone information, road geometry, directional traffic volumes, and floating-car speed data. To enhance the prediction accuracy and reliability, the second model, an Artificial Neural Network (ANN) model, is developed utilizing the same big data mentioned above. To further improve the prediction accuracy, the work zone capacity used in the ANN model is predicted using the Support Vector Machine (SVM) model.

The proposed freeway work zone delay prediction models are calibrated using historical traffic data on New Jersey freeways in years 2013 and 2014. To fill the gaps when no historical traffic data under work zone conditions are available during peak hours (i.e., 6-9 AM and 3-6 PM), a microscopic traffic simulation model (VISSIM) is used to simulate traffic data under various traffic conditions. To make the simulation closer to real-world condition, the traffic volume obtained from 2012 New Jersey Congestion Management System (NJCMS) database and speed data from INRIX are used to calibrate

the VISSIM network. Then, the performance of each model is evaluated based on the root mean square error (RMSE).

1.3 Organization

This dissertation is organized into six chapters. The flowchart of this study is shown in Figure 1.1. The focus of each chapter is briefly discussed below.

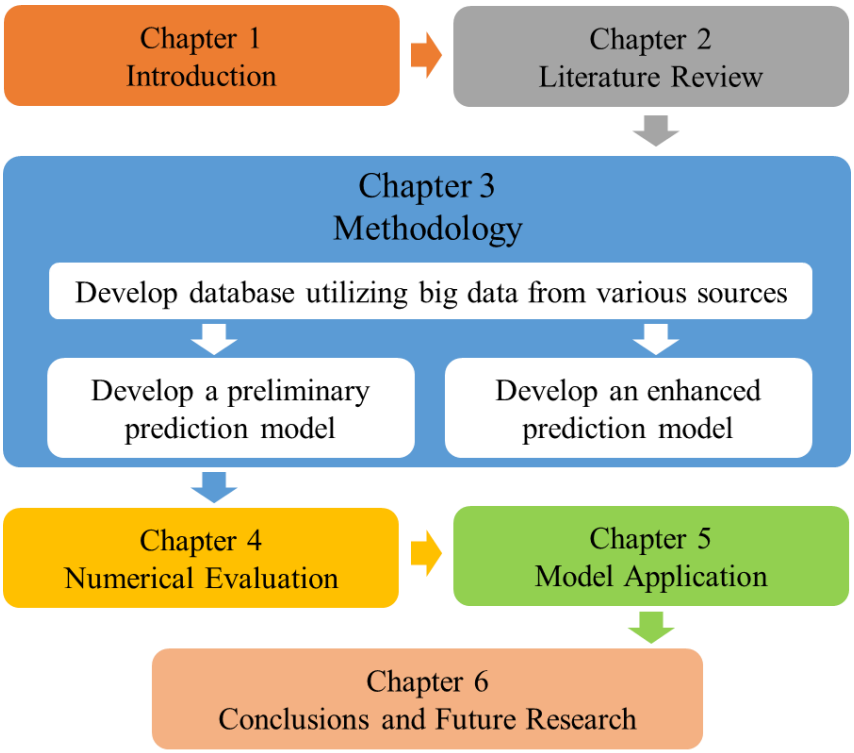


Figure 1.1 Organization of the Dissertation.

Chapter 1 introduces the background and the needs of developing a sound spatio-temporal freeway work zone delay prediction model, and discusses the objective and work scope of this research. Previous studies are reviewed and summarized in Chapter 2, which include work zone capacity and delay prediction methods, model evaluation

performances, and big data technologies which could be utilized in the practice of work zone impact analysis and management. Chapter 3 describes the data needed for the model development, followed by the formulations of the work zone delay prediction models integrated with big data. Chapter 4 presents the development and evaluation of the proposed models based on qualified freeway work zone data collected in years 2013 and 2014 on New Jersey freeways. The potential applications of the proposed models are presented in Chapter 5. Finally, research findings are concluded and suggestions to future studies are summarized in Chapter 6.

CHAPTER 2

LITERATURE REVIEW

The literature review in this chapter mainly focuses on the work zone delay prediction methods, model evaluation measures, and big data technologies. Since work zone capacity plays a key role affecting delays, this chapter first focuses on the existing studies on predicting restricted capacity caused by a work zone. Then, a detailed discussion of the models and tools for spatio-temporal work zone delay prediction is presented, followed by the reviews of big data technologies which could be utilized in the practice of work zone impact analysis and management.

2.1 Work Zone Capacity Prediction

Work zone capacity could be affected by many factors (e.g., number of opened lanes, work zone speed, approaching speed, work zone length, heavy vehicle percentage, etc.) which led to work zone capacity prediction more complex. Numerous relevant studies have focused on predicting work zone capacities based on field data, which can be generally categorized into three groups: parametric, simulation, and non-parametric approaches.

2.1.1 Parametric Approaches

Krammes and Lopez (1994) developed a regression model to predict the capacity for short-term freeway work zones using the data collected in 33 short-term work zones in Texas between 1987 and 1991. While the adjustment values of work intensity, presence of ramps, and heavy vehicles were considered in the prediction model, only a few factors affecting work zone capacity (i.e., work intensity, presence of ramps, and heavy vehicles)

were included. Kim et al. (2001) developed a new work zone capacity prediction methodology for freeways in Maryland based on more capacity reduction factors. These factors included the number of closed lanes, location of the closed lanes, heavy vehicle percentage, lateral distance to the open lanes, intensity of work activity, length and grade of the work zone. The authors reported that the developed regression model produced better capacity predicts as compared to the Krammes and Lopez model (1994).

Elefteriadou et al. (2007) developed regression models for predicting capacity of a highway work zone considering the effects of heavy vehicles, lighting and weather conditions. A model was developed for analyzing three types of work zone configurations in Florida (i.e., 2-to-1, 3-to-2 and 3-to-1 lane closures). It was found that the model predicted work zone capacity more accurately based on simulation data. In addition, the presence of heavy vehicles had a significant impact on the capacity of a work zone. The work zone capacity dropped about 8% when the percentage of heavy vehicles increased from 0% to 20%.

A more general model (Highway Capacity Manual, 2010) is applicable for predicting capacity of both short- and long-term work zones. The Highway Capacity Manual (HCM) recommended base capacity values for short-term work zones on freeways, which can be adjusted by using multiple reduction factors including percentage of heavy vehicles, intensity of work activity, lane width, and presence of ramps. While the HCM model is straightforward to apply, the work zones with shoulder closure were not considered. In addition, it is challenging for users to properly determine the reduction factors and examine delay impacts of shoulder closure work zones, thereby potentially result in in significant prediction errors caused by users' own judgments.

2.1.2 Simulation Approaches

Although the concepts of parametric approaches are widely accepted for predicting work zone capacity, the prediction results might not be accurate due to lack of proper data set. In addition, it is also challenging to examine the impact of different factors on work zone capacity by collecting field data under different work zone configurations. For example, it is unrealistic to set different work zone lengths, different speeds in the upstream of work zones, and different work zone durations in the field to cover all traffic and road geometric conditions. To evaluate the work zone impacts under various conditions, simulation models could be applied.

In the past, simulation models have been applied in various studies focusing on work zone capacity predictions (Chien et al., 2002; Heaslip et al., 2009; Chatterjee et al., 2009). Knowing that simulation models, once they are well calibrated, are capable of generating high fidelity traffic data given various work zone configurations, numerous research efforts exploiting simulation models have been conducted to predict work zone capacity. Heaslip et al. (2009) used CORSIM to develop a comprehensive database for various work zone scenarios, taking geometric, traffic, and work zone related factors into consideration. The simulated capacity was found to range between 1,288 vehicles per hour per lane (vphpl) and 1,982 vphpl, depending on the level of each parameter. However, the simulation models need to be well calibrated and require high levels of computational resources and time (Edara and Cottrell, 2007).

2.1.3 Non-Parametric Approaches

As discussed earlier, the parametric approaches usually provide low prediction accuracy because they cannot fully describe the complicated effects of influencing factors due to the

interaction effects and nonlinearities (Weng and Meng, 2013). In addition, the simulation approaches face a great challenge that it is time consuming to simulate traffic data for work zone capacity prediction. The computation time consumed by simulation approaches may increase rapidly as the road network expands and the work zone duration and number of vehicles increase.

To compensate for the deficiencies of the parametric and simulation approaches, numerous non-parametric approaches (e.g., artificial neural network model and support vector machine model) have been introduced to predict work zone capacity more accurately. Adeli and Jiang (2003) developed a radial-basis function neural network model to predict work zone capacity. This model took account for eleven different variables affecting work zone capacity. It was found that this neural network approach could provide higher prediction accuracy than parametric approaches.

Support vector machine (SVM) is a new pattern recognition technique developed by Vapnik (1995 and 1998). It has been recently applied to many traffic volume and work zone capacity prediction analyses (Zhang and Xie, 2008; Xie et al., 2010; Lord and Mannering, 2010; Boto-Giralda et al., 2010; Yu and Abdel-Aty, 2013; Du et al., 2015). SVM has two unique features enabling to produce outstanding performance. On one hand, SVM is based on structural risk minimization principle and has better generalization ability than traditional work zone capacity prediction approaches (Suykens et al., 2002; Du et al., 2015). It can reduce the chance of over-fitting and produce accurate predictions. On the other hand, a globally optimal solution is guaranteed regardless of the initial weights because the training of SVM is to solve a convex optimization problem (Scholkopf et al., 2000; Zhang and Xie, 2008).

2.2 Work Zone Delay Prediction

In this section, a summary of existing publications and reports related to the freeway work zone delay prediction is provided. Travel delay is defined as extra time motorists experience while traveling on a roadway segment due to the reduced capacity, such as work zone lane closures (Ullman et al., 2011; Weng and Meng, 2013; Habtemichael et al., 2015). Predicting the work zone travel delay plays a critical role in developing traffic management plan and calculating road user cost. This section first describes the factors affecting work zone delay prediction. Then, a detailed discussion of the commonly used models and tools available for work zone delay prediction is presented. Similar to work zone capacity prediction, numerous methods have been developed for predicting work zone delays, which can be generally categorized into three groups: parametric, simulation, and non-parametric approaches.

2.2.1 Factors Affecting Travel Speed

Many factors influencing travel speed caused by a work zone lane closure were identified in the previous studies, which can be classified as follows.

(1) Work zone related factors:

- *Number of closed lanes, total number of lanes, and lane closure location*. Several studies (Dudek and Richards, 1981; Krammes and Lopez, 1994, Dixon et al., 1997; Kim et al., 2001; Chien et al., 2002; Chung et al., 2012) pointed out that the travel speed approaching and through a work zone vary significantly with the number and location of lane closures and total number of lanes due to restricted capacity.

- *Intensity of work activity*. The intensity of work activity refers to the number of workers on the site, the number and size of work vehicles in use, and the proximity of the work activity to the travel lanes (HCM, 2010). The travel speed approaching and through a work zone may decreases as the intensity of work activity increases.

- *Starting/ending time and duration of the work zone*. The travel speed approaching and through a work zone varies significantly with starting/ending time and duration of the work zone (Chien and Schonfeld, 2001; Chien et al., 2002; Tang and Chien, 2002; Meng and Weng, 2012; Du and Chien, 2014).

(2) Traffic related factors:

- *Traffic volumes approaching a work zone*. The travel speeds approaching and through a work zone may decrease as the approaching traffic volumes increase (Dudek and Richards, 1981; Krammes and Lopez, 1994; Chien and Schonfeld, 2001; Chien et al., 2002; Tang and Chien, 2002; Du and Chien, 2014).

- *Work zone capacity*. The work zone capacity will affect the travel speeds approaching and through a work zone. The work zone capacity prediction methods were discussed in Section 2.1.

- *Heavy vehicle percentage in traffic stream*. Since heavy vehicles occupy more space and move more slowly than passenger cars on the roadway, a high heavy vehicle percentage may result in the decrease of the travel speeds approaching and through a work zone.

(3) Geometric related factors:

- *Road type (Rural/Urban)*. The travel speeds approaching and though a work zone will be affected by road types (i.e., rural and urban roadways).

- *Grade*. Kim et al. (2001) found that the presence of grades may exacerbate the flow constriction in work zones particularly in the presence of heavy vehicles which may result in travel speed reduction.

- *Effective lane width and lateral clearance of the work zone*. Both the restricted lane width and lateral clearance of the work zone will negatively affect the travel speeds approaching and through a work zone.

- *Presence of ramps*. The presence of ramps, especially the entrance ramp within the area approaching the work zone lane closure, can have a noticeable effect on work zone capacity for handling mainline traffic (HCM, 2010), which results in the reductions of the travel speeds approaching and through a work zone.

(4) Others:

- *Weather and light conditions*. Adverse weather (e.g., fog, snow, and rain) and bad light conditions have a negatively impact on travel speeds approaching and through

a work zone (Chien and Schonfeld, 2001; Chien et al., 2002; HCM, 2010; Weng and Meng, 2013).

- *Driver population*. Non-commuter driver populations do not display the same characteristics as do regular commuters (HCM, 2010), which may have an impact on the travel speeds approaching and through a work zone.

2.2.2 Parametric Approaches

The deterministic queuing theory has been a commonly used parametric approach for predicting work zone delay (Abraham and Wang, 1981; Dudek and Richards, 1982; Chien and Schonfeld, 2001; Weng and Meng, 2013; Du and Chien, 2014). It has been in practice for decades, and was implemented by both the federal and state transportation agencies (e.g., FHWA and various state DOTs in Alabama, Florida, Illinois, New Jersey, Ohio, Oklahoma, Washington, etc.). It is often depicted using the diagram shown in Figure 2.1, in which the shaded area is the queuing delay (veh-hr) caused by work zone lane closures. The critical inputs are the approaching volume, roadway capacity under normal and work zone conditions, and duration of the work zone (McCoy et al., 1980; Jiang, 2001; Chien and Schonfeld, 2001; Tang and Chien, 2002; Du and Chien, 2014). The pros and cons of the deterministic model are examined as follows.

In a study conducted by McCoy et al. (1980), the user delay was considered as the time lost while one is traveling through a construction and maintenance zone. The time lost is taken to be a function of the difference between the average overall speed of the two-lane two-way no-passing operation and that of the normal four-lane divided highway based on 1979 data. Since they do not consider the situation in which the approaching traffic volume exceeds the work zone capacity, queuing delay is not taken into account in their study.

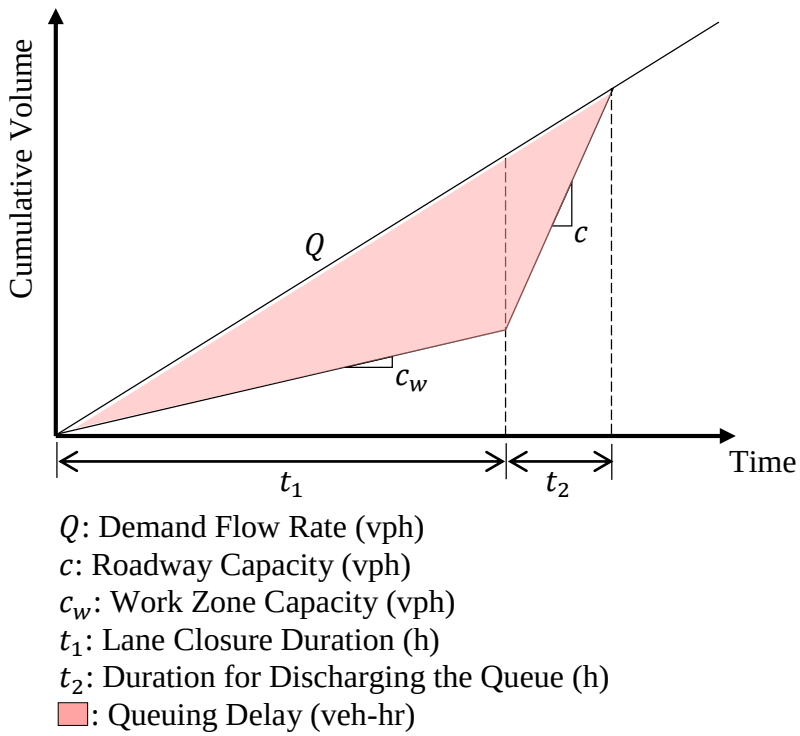


Figure 2.1 Queuing delay predicted by the deterministic model.

Chien and Schonfeld (2001) used deterministic queuing theory to predict user queuing delays caused by a work zone with single lane closure on a four-lane highway (two-lane per direction). In addition to the queuing delay, the moving delay incurred by vehicles traveling through the work zone was also included in the user delay function. However, the time-varying traffic volume and factors affecting work zone capacity were not considered in this paper.

Since work zone delay is significantly affected by volume-capacity ratio, light condition, heavy vehicle percentage, and lane width, Du and Chien (2014) formulated delay considering time-varying traffic pattern, work zone capacity adjustment factors and shoulder usage. A sensitivity analysis was conducted, and results suggested that shoulder use is needed, which increases work zone capacity and reduces user delay, especially

during peak hours. It is found that the traffic speeds under work zone and normal conditions with and without using the road shoulder were assumed (i.e., 45 mph and 55 mph, respectively), which may affect the accuracy of predicted delays. The deterministic queuing model is suitable for predicting a delay for planning purposes but fails to provide accurate prediction under traffic operations wherein there are time-varying and congested traffic conditions (Chung, 2011). These models have a limited ability to analyze the spatio-temporal congestion impacts caused by work zones.

Another well-known parametric approach for predicting work zone delay is based on the shockwave theory originally developed by Lighthill and Whitham (1955) and Richards (1956). The shockwave theory assumes that traffic flow is analogous to fluid flow and employs a flow-speed-density relationship to analyze the transition of traffic flow over space and time. The length of a physical queue can be determined based on a specified demand and capacity.

A shockwave-based model developed by Wirasinghe (1978) was applied to determine total delay upstream of an incident. The model was indicated in a time-space diagram by considering two traffic-flow states (i.e., free-flow and jam). Al-Deek et al., (1995) predicted delays caused by single and multiple incidents on Route 1-880 in California with the shockwave theory. This method seemed effective in determining temporal and spatial incident delays, but overestimated the maximum incident queue length.

Benekohal et al. (2013) used the shockwave theory to estimate queue and delay caused by a work zone. As concluded in that report the queue length and delay could be overestimated, especially under congested condition, because the shockwave speed was

approximated by interpolation from speed-flow curves rather than field data. Therefore, the shockwave theory seems is not a very reliable approach under congested traffic conditions (Habtemichael et al., 2015).

2.2.3 Simulation Approaches

Simulation approaches generally can be classified into macroscopic, mesoscopic, and microscopic approaches. Macroscopic simulation approaches are based on deterministic relationships of flow, speed, and density of the traffic stream (FHWA, 2006). Examples of macroscopic simulation approaches include Bottleneck Traffic Simulator (BTS) (Lin and Hall, 1991), Freeway Corridor Simulation Model (FREQ) (Smith et al., 1992), and TRANSYT-7F (Joseph et al., 1988; Schroeder et al., 2015).

Mesoscopic approaches combine properties of both macroscopic and microscopic simulation approaches, which assign vehicle types and driver behaviors as well as relationships with roadway characteristics (FHWA, 2006). Examples of mesoscopic approaches include Continuous Traffic Assignment Model (CONTRAM) (Taylor, 2003), and Dynamic Network Assignment Simulation Model for Advanced Road Telematics for Planning (DYNASMART-P) (Sbayti et al., 2002).

Microscopic simulation approaches simulate the movement of individual vehicles, based on theories of car-following and lane-changing (FHWA, 2006). Previous studies have been applying microscopic simulation approaches to quantify work zone delay (Chien et al., 2002; Meng and Weng, 2010; Chung et al., 2012). Knowing that well calibrated simulation models are capable of generating high fidelity traffic measures under various work zone configurations. CORSIM (Chien et al., 2002), PARAMICS (Wang et

al., 2002), and VISSIM (Edara et al., 2013) are among the most widely used microscopic simulation models.

Chien et al. (2002) proposed a method for approximating work zone delay by integrating simulation data obtained from Corridor Simulator (CORSIM) and the concept of the deterministic queuing model. The simulation model considered various geometric conditions and time-varying traffic distributions was applied to predict queuing delays on interstate I-80 in New Jersey. However, to make the model more applicable to other work zone configurations, it requires extensive calibration and validation.

Yang et al. (2008) used CORSIM to predict work zone delays due to reduced capacity. They found that predicting work zone delay with CORSIM was better than using deterministic queuing model because the simulation model can record the acceleration delay, deceleration delay, shockwave delay, and other factors that are ignored in deterministic queuing model. Since CORSIM can not simulate delays if a queue spillbacks beyond the entry nodes, traffic delay at work zones could be underestimated in congested conditions. In addition, tedious work is required to input data and long computation times may be needed for the simulation model.

Wang et al. (2002) used the microscopic traffic simulation software, PARAMICS, to predict the traffic delay to road users under different maintenance schedule. This model is again only applicable when there is no queue, that is, the approaching traffic volume is less than the work zone capacity.

Edara et al. (2013) developed a simulation model using VISSIM for predicting the traffic impacts (i.e., delay and queue length) of work zones under congested condition, which was calibrated using field data from two work zones in Missouri. Both work zones

involved a single lane closure on a three-lane section of freeway. The study found that VISSIM was appropriate for work zones in urban areas where lane closures may affect the traffic on neighboring roadways. Due to data limitations, the study recommended that the use of private sector data (e.g., INRIX) for predicting delay and queue length could generate a sufficiently large sample of work zones that could be used for calibration.

To develop a simulation model, a comprehensive historical traffic volume origin-destination trip tables and speed data, high computational resources, time-consuming parameter calibration and long running time are required (Edara and Cottrell, 2007).

2.2.4 Non-parametric Approaches

To overcome the limitations of parametric and simulation approaches, non-parametric approaches were introduced. Since the concept of McCulloch–Pitt neuron introduced in the early 1940s (Adeli and Hung, 1995), the artificial neural network (ANN) has been evolving towards more precise and powerful model for pattern recognition and prediction. Neural networks were inspired by the mechanisms by which real biological neurons work in the human brain. The decision making process of the brain is simulated by an artificial network of neurons manipulating data among the many nonlinear nodes operating in parallel.

In the transportation industry, artificial neural networks have been used to various traffic measures, such as traffic flow (Jiang and Adeli 2005; Kumar et al. 2015), freeway work zone capacity (Neubert et al. 2000; Karim and Adeli 2003), and work zone delay (Ghosh-Dastidar and Adeli 2006; Du et al. 2016). Zhang et al., (1997) use the simple back propagation (BP) neural network to simulate a macroscopic freeway traffic flow model. Park et al. (1998) use a radial-basis function neural network to forecast freeway traffic

flow. Suzuki et al. (2000) use a combination of the BP neural network and the Kalman filter and a macroscopic model to predict origin–destination (O–D) travel times and traffic flows.

For the work zone related analysis, Karim and Adeli (2003) developed a radial basis function neural network (RBFNN) model to predict the work zone capacity, delay and queue length, considering number of lanes, number of open lanes, work zone layout, length, lane width, heavy vehicle percentage, grade, speed, work intensity, darkness factor, and proximity of ramps. Based on the prediction results of three examples, the authors concluded that RBFNN model was acceptable for most practical purposes, but the sample size used to train the RBFNN model is marginal. A neural network needs to be trained by using large number of samples so that the prediction results could be closer to actual observations.

As discussed above, many studies have successfully applied ANN models (i.e., radial basis function and multi-layer feed-forward neural networks) to predict the freeway work zone capacity. However, these studies have not been directly applied to the work zone delay prediction problem.

Vemuri et al. (1998) presented a sigmoidal neural network model for short-term forecasting of traffic delays in highway construction zones using data from presence detectors. The method was based on a modular approach wherein data from adjacent detectors was processed for predicting the travel time between the two detectors. Simulation examples were used to illustrate the traffic delay prediction algorithm. The simulation results indicated that the proposed approach performs reasonably well on

simulated data, while the performance of the proposed approach on field data needs to be investigated.

Ghosh-Dastidar and Adeli (2006) presented a multi-layer feed-forward neural network model (i.e., Levenberg-Marquardt neural network model) for delay and queue length prediction at freeway work zones. The neural network model was trained using simulated data and tested using both simulated and real-world data. The computational model presented was applied to five examples of freeways with two and three lanes and one lane closure with varying entry flow or demand patterns. It is found that the actual traffic speed and volume patterns in freeway segments are not considered in this paper.

Du et al. (2016) developed a multi-layer feed-forward ANN model to predict work zone delay using the probe-vehicle data (i.e., speeds under normal and work zone conditions) subject to the condition when traffic volume and capacity information are missing. Based on the prediction results of three examples, it was found that the ANN model outperformed the deterministic queuing model in terms of the accuracy in predicting travel delays caused by reconstruction projects. If the approaching traffic volumes are at or near the work zone capacity, this accuracy of prediction of this ANN model is not promising because the relationship of approaching traffic volume and work zone capacity is not considered.

2.2.5 Tools for Work Zone Delay Prediction

Memmott and Dudek (1985) developed a model called Queue and User Cost Evaluation of Work Zone (QUEWZ), which has been commonly used to predict user costs resulting from work zone lane closures. The model was designed to evaluate work zones on freeways or multilane divided highways with up to six lanes in each direction, considering percentage

of heavy vehicles, lane closure configuration (e.g., number of open lanes, length of the lane closure, and capacity of the work zone, etc.), hourly traffic volumes, and queuing length. An enhanced QUEWZ, called QUEWZ-98 (Copeland, 1998), approximated the work zone capacity based on the HCM 2000 procedures and the excess emissions to determine the road user cost. QUEWZ-98 can identify lane closure schedules that minimize work zone related delay. It was reported that QUEWZ-98 is applicable to work zones on freeways or multilane divided highways (Benekohal et al., 2003).

QuickZone (Mitretek System, 2000) is a work zone delay impact analysis tool developed by the FHWA. It is a Microsoft Excel-based application that facilitates software customization through an open source code. This tool is capable of calculating the average traffic delay and maximum queue length that could result from lane closure or restriction in both urban and suburban work zones. It was found that QuickZone could deliver of highly comprehensive and detailed output, and adopt the approach of modeling traveler response to prevailing traffic conditions, such as route changes, peak spreading and mode shifts. The main limitation of QuickZone is its detailed data requirements for both the main line where the work zone is installed and alternative route diversion roadways upstream of the work zone. Thus, users may not be able to gather all the data inputs that are necessary to implement QuickZone (Batson et al., 2009).

Chitturi and Benekohal (2004) compared the performance of QUEWZ and QuickZone in predicting traffic delay at work zones using field data collected from 14 freeway work zones in Illinois. Field data were compared to the results from QUEWZ and QuickZone software. QUEWZ overestimated the capacity and average speed, but underestimated the average queue length. The queue lengths from QuickZone did not

match the field data, which in general underestimated the queue length as well as the total delay observed in the field. Especially as demand is less than capacity, QuickZone does not return any user delay because it does not consider the delay due to reduced speed within the study work zone.

The Work Zone Capacity Analysis Tool (WZCAT) analytical software program was developed by the Wisconsin Department of Transportation (2007). The main objective of this tool is to predict delays and queues for short-term work zone closure. WZCAT is developed based on the concept of deterministic queuing analysis through basic input/ output analysis. This tool was developed to function as an add-on program that operates within Microsoft Excel. Although WZCAT has a simple structure, it is not able to produce identical queuing patterns to the observed field data and significantly over predicts the queue length. Furthermore, the queuing pattern predicted was not similar to what was observed.

The Iteris performance management system (iPeMS) is a commercial traffic data collecting, processing, and analyzing tool to assist traffic engineers in assessing the performance of the freeway system. It is an enhanced model from PeMS, which was originally developed by the University of California, Berkeley, in cooperation with Caltrans. This tool collects real-time traffic data from deployed intelligent transportation system (ITS) sensors, saves them in a data storage, and presents this information in various forms to traffic operators and planners. It also allows users to query freeway traffic data and to compute various performance measures. iPeMS can assist with conducting simple to advanced traffic analyses, including Highway Capacity Manual analyses, Synchro analyses, and computer simulations. In addition to determine the spatio-temporal impacts

of the existing work zone lane closures on the freeway, iPeMS also provides travel time predictions, where the algorithms combine historical and real time data. The longest prediction period is 30 minutes from the starting time (Choe et al., 2002).

Rutgers Interactive Lane Closure Application (RILCA) is an interactive computer tool for planning lane closures for work zones developed for the New Jersey Turnpike Authority-Garden State Parkway division. Bartin et al. (2012) found that RILCA could provide various analyses and visualization options to plan lane closures interactively, obtain traffic volume information, determine the maximum queue length, and predict the time of clearance. However, the disadvantages of RILCA include the following:

- Oversimplified formulae to predict queue length and delay,
- No real-time traffic data, and
- Lane closure analyses on only the NJ Turnpike and GSP.

Chien et al. (2016) developed an on-line system analysis tool called the Work Zone Interactive Management Application - Planning (WIMAP-P), an easy-to- use and easy-to-learn tool for predicting the traffic impact caused by work zones on freeways and arterials. WIMAP-P is supported by a working database that was developed based on the data feeds from various sources, including OpenReach, Plan4Safety, New Jersey Straight Line Diagram (NJSLD), New Jersey Congestion Management System (NJCMS), and INRIX. The WIMAP-P system architecture comprises of three specific modules (i.e., a working database, a work zone speed prediction model, and an on-line software application) interacting together to generate the required results as shown below.

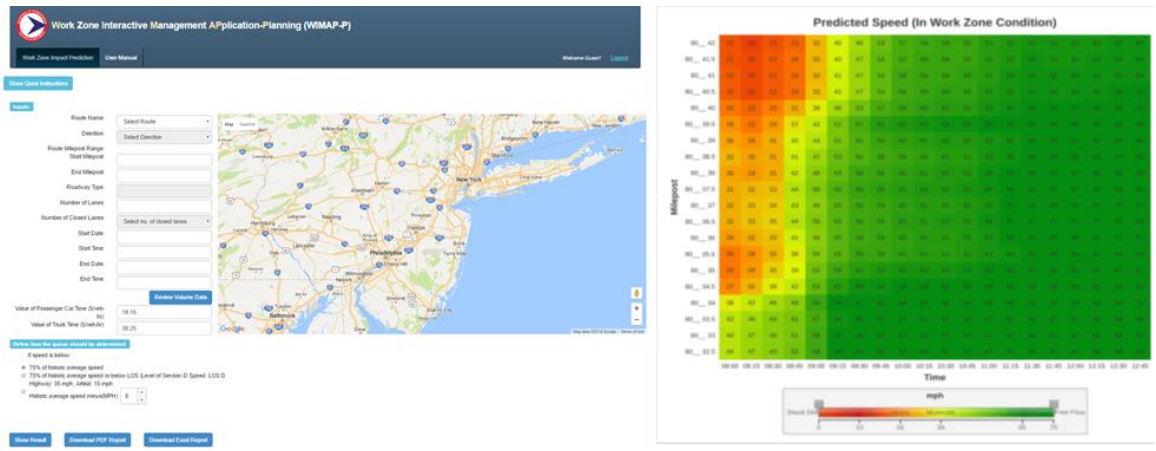


Figure 2.2 Work zone impact prediction module and result.

Source: Chien, S., L. Spasovic, J. Lee, K. Mouskos, and B. Du. Feasibility of Lane Closures Using Probe Data. Draft Report. New Jersey Department of Transportation, 2016.

An artificial neural network (ANN) and multivariate non-linear regression (MNR) models were developed based on 466 work zones, which are employed by WIMAP-P to predict speed caused by work zones on NJ freeways and arterials. The study found that the ANN model is slightly more accurate for predicting delays of historic work zones, but the MNR model demonstrates better reliability and consistency in predicting delays of work zones in places where there are no historic data. The graphical user interface of WIMAP-P can effectively facilitate data input and analysis in an efficient and reasonably intuitive manner while producing graphical results and customized reports. In addition to predict the spatio-temporal speed impact caused by work zones, WIMAP-P also computes the associated road user cost.

In addition to the tools discussed above, other work zone delay prediction tools were reviewed and summarized in Table 2.1.

Table 2.1 Work Zone Impact Prediction Tools

State	Tool	Finding
Tennessee (Tennessee DOT, 2006)	Lane Closure Decision Support System (LCDSS)	Developed a web-based tool to predict queue length and delay based on HCM methodology for lane closures on Tennessee roads.
Florida (Washburn et al., 2008)	N/A	Developed a non-linear regression model to predict work zone travel speed, saturation flow rate, queue delay, and queue length for two-lane roadway work zones (with a lane closure).
Alabama (Turner et al., 2009)	N/A	Presented the results of research done to determine the need for an update of the queue prediction portion of ALDOT's lane closure analysis tool, a Microsoft Excel-based "Lane Rental Model" whose work zone capacity values are based on the 1994 Highway Capacity Manual.
Oregon (Oregon DOT, 2010)	Web-based Work Zone Traffic Analysis (WZTA)	Predicted project and corridor work zone delays by using the deterministic methodology. The thresholds used by WZTA are based on decades of on-the-job experience, technical observation and engineering evaluation.

2.3 Performance Index for Model Evaluation

This section focuses on the model evaluation measures that are widely used in evaluation of transportation models, such as coefficient of determination (R^2), root mean squared error

(RMSE), and mean absolute percentage error (MAPE). These measures are discussed below.

The R^2 is a statistical measure of how close the data are to the fitted regression line (Forst, 2013). It is also known as the coefficient of determination. An R^2 of 1.0 indicates that the regression line fits the data perfectly (Meng and Weng, 2010). The R^2 can be defined as

$$R^2 = 1 - \frac{\sum_{h=1}^{H}(\hat{c}_h - c_h)^2}{\sum_{h=1}^{H}(c_h - \bar{c})^2} \tag{2.1}$$

where:

c_h = The hth actual observation value;

$\widehat{c_h}$ = The hth predicted value;

$\bar{c}$ = The mean of observations; and

H = The total number of observations.

The RMSE is a measure of the predictive success of the model and is a commonly referenced as providing an indication of the error of a model. It is usually calculated across all key observed and predicted count data points and may be calculated on an hourly basis, or across the full evaluation period depending on the focus of the project (New Zealand Transportation Agency, 2014). A smaller RMSE indicates greater accuracy of the model.

Many studies (Wild, 1997; Zhang and Xie, 2008; Hou et al., 2015; Du et al. 2016) have used RMSE for evaluating the accuracy of work zone traffic flow forecasting models. For the work zone delay prediction model evaluation, the RMSE can be used for denoting the variability between the predicted and observed speeds upstream of a work zone as

shown in Eq. 2.2. Note that the observed speeds may be obtained from the floating-car data as discussed in Section 2.4.

$$RMSE = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(\hat{c}_h - c_h)^2} \tag{2.2}$$

where:

c_h = The hth actual observation value;

$\hat{c}_h$ = The hth predicted value; and

H = The total number of observations.

The MAPE is another relative measure of error, which expresses accuracy as a percentage of the error as shown in Eq. 2.3. It has been used by several researchers (Wild, 1997; Park, 2002; Chu et al., 2005; Hou et al., 2015; Du et al., 2016).

$$MAPE = \frac{1}{h}\sum_{h=1}^{H}\left|\frac{\hat{c}_h - c_h}{c_h}\right| \times 100\%, \qquad c_h \neq 0 \tag{2.3}$$

where:

c_h = The ith actual observation value;

$\hat{c}_h$ = The ith predicted value; and

H = The total number of observations.

There are other measures similar to the measures covered in this section, such as root mean square percentage error (RMSPE), mean absolute error (MAE), mean absolute deviation (MAD), and mean squared deviation (MSD), which were used by other

researchers (Fudala and Fontaine, 2010; Meng and Weng, 2010; Hou et al., 2015). In this study, the RMSE and MAPE are selected for model evaluation purpose.

2.4 Big Data Technologies

The traffic volume and speed data used in the existing work zone delay prediction studies were usually collected by loop detectors or road tube counters. In case the loop detectors are not installed near the work zone location, it is very challenging to obtain volume and speed data for work zone impact analysis. A wide variety of unprecedented massive traffic data obtained from different sources (e.g., infrastructure sensors, mobile devices, floating cars, and toll tags) has become increasingly available. As the sources of big data provide a lower-cost approaching for collecting traffic volume and speed data (Burt et al., 2014), the loop detectors would wane as these big data technologies become more common and mature.

Previous studies (Haghani et al., 2009; Chen and Rakha, 2014; Elhenawy et al., 2014) indicated that the floating-car speed data (i.e., INRIX) are reliable for travel time prediction. One of the most notable activities of using floating-car technology would be the I-95 Corridor Coalition project (2010), which demonstrated that floating-car data were accurate under a variety of traffic conditions, including congestion caused by incidents. However, it is very challenging to use floating-car data for developing work zone delay models as actual volume data under normal and work zone conditions are missing. Despite an increasing attention in modeling work zone delay prediction, only few studies (Chung, 2011; Chung et al., 2012; Habtemichael et al., 2015) examined the spatio-temporal impacts of incidents with traffic volume and speed data collected by loop detectors. Therefore, it became desirable to interface floating-car data with an ANN framework that can precisely

predict spatio-temporal delays caused by work zone lane closures. Each of these floating-car technologies will be discussed below.

Bluetooth is an open, wireless communication platform used to connect myriad electronic devices. Many computers, car radios and dashboard systems, PDAs, cellular phones, headsets, or other personal equipment are, or can be, Bluetooth-enabled to streamline the flow of information between devices (KMJ Consulting, Inc., 2010).

Manufacturers typically assign unique Median Access Control (MAC) addresses to Bluetooth equipped devices. Bluetooth-based travel time measurement involves identifying and matching the MAC addresses of Bluetooth-enabled devices carried by motorists, passing a detector. The matchings of Bluetooth device can be used to measure arterial travel time, average running speed, and origin-destination patterns of travelers. Since MAC addresses are not tracked when the device is sold within the marketplace, these unique addresses can be detected and matched without establishing a relationship to personal or, otherwise, sensitive information, and thus, keeping the traveling public and their personal information anonymous (KMJ Consulting, Inc., 2010; Cambridge Systematics, Inc., 2012).

The sample size of data is also critical in providing accurate and up-to-date travel times. A research conducted by University of Maryland (Puckett and Vickich, 2010) suggests that a four percent detection rate is required for roadways of 36,000 AADT or greater. Roads with lower volumes would require a larger match percentage to attain an adequate sample. A study by Tarnoff et al. (2009) has discussed that 5 – 7 % of vehicles in a traffic stream have Bluetooth enabled devices, which would be considered an adequate sample size.

Bluetooth technology is new but rapidly maturing as the percentage of vehicles with Bluetooth devices (smartphones, in-vehicle connections, tablets, etc.) increases rapidly; it is also easy to install and maintain. With the cost per unit being relatively low, the predictions of travel times performed by Bluetooth technology have been compared to floating car methods and radio-frequency identification (RFID) as an accurate and cost-effective alternative (KMJ Consulting, Inc., 2010; Mendez, 2011).

Another commonly used floating-car technologies is toll tags, which used for electronic toll collection and deployed at various points on a roadway network to obtain average travel time and speed information. With technological advancements, the traffic data collection technologies utilizing floating-car concepts have improved rapidly in the past few years, in terms of geographic coverage, sample size, accuracy in detecting vehicle location, and data processing algorithms. These improvements engendered greater accuracy and reliability of predicted information, such as speed and travel time, based on the floating-car traffic data. There are four components in a toll tag travel time system: electronic tags, antennas, readers, and a central computing and communication facility (Cambridge Systematics, Inc., 2012). As a vehicle with an electronic tag passes underneath a toll tag reader, the time and toll tag identification number are recorded. If the same vehicle passes the next reader location, the travel time and average speed between the two locations can be determined. The toll tag identification number can be coded to protect privacy.

Sample size requirements for a toll tag travel time system depend on the market penetration of the toll tags. Ferman et al. (2005) suggests that a three percent penetration rate on freeways and 5 % on arterials is adequate. According to the New Jersey Turnpike

Authority (2008), more than 70 % of the vehicles registered in New Jersey have E-Z Pass toll tags. Similar to Bluetooth technology, toll tag readers are also mature due to their capability of providing a huge number of data points. With its simplicity in installation and maintenance, the percentage of toll transactions in New Jersey was predicted to be more than 70 % in 2010 (INRIX, 2008). The cost per unit is also relatively low.

Radar detection system is a non-intrusive radar-based system operating in the microwave band and needs to be mounted on a roadside pole above a certain height. The radar sensor provides per-lane presence, volume, occupancy, speed, as well as classification information in up to 12 user-defined detection zones. Output information is provided to existing controllers via contact closure and to other computing systems by serial port, IP communication port or by an optional radio modem. A single radar unit can replace multiple inductive loop detectors and the attendant controller. RTMS (Remote Traffic Microwave Sensor), one of the advanced radar detectors, is all-weather accurate and virtually maintenance-free. The detection range of one RTMS is up to 250 feet, which provides coverage for up to eight lanes of traffic (Image Sensing Systems, Inc., 2012). Microwave radar detector technology is mature, due to its ability to provide accurate spot speed data despite its inaccuracy for volumes. Radar units are easy to maintain, which can be conducted on radar units without closing traffic lanes (Cambridge Systematics, Inc., 2012).

Besides the floating-car technologies mentioned earlier, some commercial vendors also provide floating-car data, such as INRIX, TomTom, and HERE. They are based on GPS tracking systems, which capture vehicle movements nearly continuously within a very small time interval (e.g., 1 second) (Mudge et al., 2013). It is assumed that the travel

time of tagged vehicle in a vehicle stream represents well the true travel time of all (tagged and non-tagged) vehicles in the stream. These methods capture the actual volume of all tagged vehicles while the portion of non-tagged vehicles remains unknown.

The INRIX reported speeds are generated based on the data from a variety of sources including GPS-enabled vehicle fleets, smart phones, and connected cars equipped with GPS locator devices (Schrank et al., 2015). The primary INRIX data is from GPS-enabled vehicle fleets (e.g., delivery vans, taxi cabs, and long-haul trucks, etc.), which are supplemented by sensor-based data (Seymour et al., 2011). The collected data are compiled into an average speed profile for most freeways and arterials, which covers nearly 5 million miles of road, ramp and interchange in over 40 countries (INRIX, 2016). INRIX data attribute consists of three levels, which are real-time data for the specific segment, historical data (e.g., road reference speeds), and combination of real-time and historical data (Middleton et al., 2011).

When sample sizes are large, it is likely that the average speed reflects the percentages of different vehicle types in the traffic stream, resulting in less bias (Turner et al., 2011). Haghani et al. (2009) found that the INRIX data have a satisfying accuracy on freeways. White et al. (2010) suggests that because INRIX data is based largely on fleet-based GPS probe vehicles, its use may be an issue for arterials, due to reduced sample size and the fact that commercial vehicles operate differently than other vehicles in terms of their acceleration and deceleration characteristics. Data quality specifications are in effect when flow exceeds 500 vehicles per hour and apply to both freeways and arterials (Brewer, 2007). Unlike toll tag readers and Bluetooth-installed readers, INRIX requires no

installation or maintenance cost for transportation agencies. The data providers have great incentive to provide accurate data at a low cost.

2.5 Summary

This chapter presented a literature review focusing on work zone capacity and delay prediction approaches, model evaluation measures, and big data utilized in work zone impact analysis. The findings and conclusions on the comprehensive literature review have been identified.

Many research efforts predicted work zone capacity using parametric, non-parametric, and simulation approaches. However, the non-linear and interrelating factors affecting work zone capacity could not be fully described by parametric approaches and it is time consuming to calibrate and apply simulation models for work zone capacity prediction. Therefore, in order to improve the prediction accuracy and computational efficiency, there is a need to develop a creditable non-parametric model (i.e., SVM model) to uncover the non-linear relationships between influencing factors.

Similar to the approaches for the work zone capacity prediction, the approaches developed for predicting work zone delay have been classified into parametric, non-parametric, and simulation approaches. Each approach has demonstrated good performance for work zone delay prediction under situation suitable for it. To fulfill the objective of this research, a sound and reliable model should be developed to predict spatio-temporal freeway work zone delay leveraging big data.

Two well-known parametric methods developed for work zone delay prediction include the deterministic queuing models and the shockwave models. These two models do not take into account the dynamic changes of traffic volumes and speeds over space and

time and are therefore not very accurate and reliable for work zone delay prediction. Although the simulation approaches are able to provide the most detailed traffic data for work zone delay prediction, they suffer from several shortcomings as they usually require high computational resources, time-consuming parameter calibration, and long running time.

To compensate for the weak points of the parametric and simulation approaches, non-parametric approaches have been employed for the prediction of work zone delay in many previous studies. One of the most widely used non-parametric approaches is the ANN, which is data-driven, flexible, and able to model non-linear dependencies of the influencing factors.

With technological advancements, a wide variety of massive traffic data from infrastructure sensors and floating cars has become increasingly available. This new and rich data has made a way for big data analytics as an emerging method for predicting freeway spatio-temporal work zone delay. Although several studies (Vemuri et al. 1998; Ghosh-Dastidar and Adeli, 2006) used ANN to predict work zone delay but with limited traffic data from loop detectors or simulation, few studies were found assessing work zone spatio-temporal delay with big data. Hence, this dissertation tends to enhance the prediction accuracy, which focuses on capturing the relationship between approaching traffic volumes and work zone capacity predicted by a SVM model. With that, the spatio-temporal speed under a work zone condition can be predicted using an ANN model.

CHAPTER 3

METHODOLOGY

This chapter describes the data needed for the model development, followed by the equations of two spatio-temporal freeway work zone delay prediction models considering approaching traffic volumes, work zone capacity, and normal speeds prior to the work zone, including multivariate non-linear regression (MNR) model and the artificial neural network (ANN) model, are discussed in this chapter. The work zone capacity in the MNR model is approximated using reduction factors. While in the ANN model, the work zone capacity is predicted from SVM model. It is noted that the input variables of MNR and ANN models are determined based on Pearson and Spearman correlation tests and data availability. These models are explored, tested, and evaluated through historical freeway work zones for years 2013 and 2014 in New Jersey, which will be discussed in Chapter 4.

3.1 Database Development

This section presents the main characteristics of the database (DB) for the proposed model development. To develop a sound model for predicting speed/delay caused by an expected work zone with lane closures on freeways, it requires significant amount of data under different categories (Karim and Adeli, 2003; Edara and Cottrell, 2007; Taylor et al. 2007; Habtemichael et al. 2015):

- Historical work zone data: work zone related information, such as work zone type, date and locations of the work zone, lanes closed, length, duration, lane width, shoulder width, shoulder usage.
- Road geometry data: road type, number of lanes, distance, speed limit, grade, median/shoulder width, lateral clearance, and interchange locations.

- Traffic volume data: weekday and weekend passenger-car and heavy vehicle volumes collected in the field or from the sources of big data presented in Section 2.4.

- Floating-car data: traffic speeds for freeway segments under normal and work zone conditions, which are collected by floating-car technologies listed in Section 2.4.

- Crash records: crash location, severity, and start/end time.

The overview of these databases is shown in Figure 3.1.

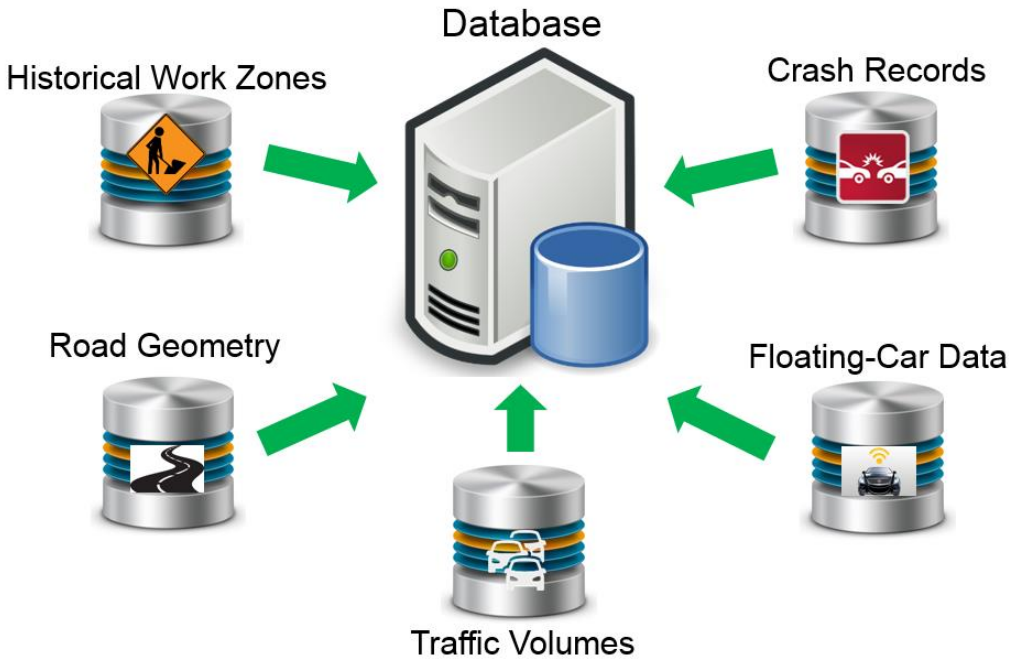


Figure 3.1 General database overview.

It is noted that the process to define the final DB is based on the availability and applicability to predict the freeway work zone speed and delay as required by the proposed prediction model. The DB can be developed using the advanced computing resources, which provided adequate of data storage and computing processing to handle the large data resources that are necessary to process and execute the proposed prediction model.

3.2 Model Formulation

Two models, including a multivariate non-linear regression (MNR) model and a multi-layer feed forward ANN model, are developed to predict the spatio-temporal delay caused by a planned work zone on freeways. This section describes the basic configurations of MNR and ANN models followed by identification of inputs for their implementation on freeways.

3.2.1 The Multivariate Non-Linear Regression Model

When a work zone is present especially during the peak hours, the traffic flow will be significantly interrupted due to an insufficient capacity. This will reduce the speed upstream from the work zone. The congestion will continue to propagate, the speed of this disturbance will dependent on the upstream traffic volume and capacity of the work zone segment. In order to predict the spatio-temporal speed under work zone condition, the work zone characteristics (e.g., work zone length and duration), road geometry (e.g., number of lanes and grade), and traffic condition (e.g., volume and speed) shall be taken into account (Karim and Adeli, 2003; Edara and Cottrell, 2007; Taylor et al., 2007; Habtemichael et al., 2015).

As mentioned in the literature review, the potential input variables of work zone delay prediction suggested by previous studies (Kim et al., 2001; Adeli and Jiang, 2003; Edara and Cottrell, 2007; Du et al., 2016) include approaching traffic volume, work zone capacity, work zone length, work zone duration, work zone lane width, to name a few. The actual choice of inputs is based on the Pearson and Spearman correlation tests and data availability. Briefly, the Pearson correlation evaluates the linear relationship between two variables while the Spearman correlation evaluates the non-parametric relationship

between the variables. The closer the value is to 1 or -1, the stronger the correlation between the variables. It will be discussed later in the next chapter that how the factors affecting speed of upstream work zone in this study are determined.

The general freeway MNR model is a non-linear function considering approaching traffic volumes, work zone capacity, and normal speeds prior to the work zone. For each freeway section *i* ($1 \leq i \leq m$) and for each specific time interval *j* ($1 \leq j \leq n$), a total of *m* x *n* speed observations are available during the observation period. As discussed in previous chapters, the big data analytics allows researchers to examine large amounts of traffic data to efficiently uncover hidden patterns, correlations and other insights. The speed for any particular time interval/section combination can be obtained either from loop detector stations or floating-car data sources (e.g., Bluetooth, INRIX, etc.) using big data analytics. From these speed observations, a spatio-temporal distribution of speeds under normal condition (i.e., condition in which there is no incident) prior to the work zone (denoted as s_{ij}) can be constructed as in Table 3.1.

Then associated with a work zone occurred on freeway section *i* at time *j*, a spatio-temporal speed matrix under work zone condition prior to the work zone (denoted as y_{ij}) can be constructed based on observed freeway traffic data (e.g., work zone speed, traffic flow approaching work zone, etc.). Table 3.2 shows the speed matrix under work zone condition.

Table 3.1 Speed Distribution under Normal Condition

Traffic Flow Direction →	Freeway Section	Time Interval					
		1	**2**	**...**	***j***	**...**	***n***
	1	s_{11}	s_{12}	...	s_{1j}	...	s_{1n}
	2	s_{21}	s_{22}	...	s_{2j}	...	s_{2n}
	⋮	⋮	⋮	...	⋮	...	⋮
	i	s_{i1}	s_{i2}	...	s_{ij}	...	s_{in}
	⋮	⋮	⋮	...	⋮	...	⋮
	m	s_{m1}	s_{m2}	...	s_{mj}	...	s_{mn}

Table 3.2 Speed Distribution under Work Zone Condition

Traffic Flow Direction →	Freeway Section	Time Interval					
		1	**2**	**...**	***j***	**...**	***n***
	1	y_{11}	y_{12}	...	y_{1j}	...	y_{1n}
	2	y_{21}	y_{22}	...	y_{2j}	...	y_{2n}
	⋮	⋮	⋮	...	⋮	...	⋮
	i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{in}
	⋮	⋮	⋮	...	⋮	...	⋮
	m	y_{m1}	y_{m2}	...	y_{mj}	...	y_{mn}

As traffic increases on the network, the resulting travel time and delay increase, especially when the approaching traffic volume is close to the work zone capacity. In an

effort to better represent speed reduction and delay due to work zone activities, the concept of the Bureau of Public Roads (BPR) function (Bureau of Public Roads, 1964) is adapted in this study to construct the corresponding speed-flow relationships based on historical work zone data to achieve reasonable congested weighted speeds. Therefore, the MNR model is formulated as follows.

$$y_{ij} = a_{ij} + b_{ij} \frac{s_{ij}}{1 + c\left(\frac{Q_j}{C_w}\right)^d} \tag{3.1}$$

where:

y_{ij} = The average speed of segment *i* at time *j* under work zone condition (mph);

s_{ij} = The average speed of segment *i* at time *j* under normal condition (mph);

i = The ith freeway segment in upstream of work zone ($1 \leq i \leq m$);

j = The jth time interval after work zone started ($1 \leq j \leq n$);

m = The number of freeway segments (e.g., Traffic Message Channels) upstream of work zone;

n = The number of time intervals (e.g., 15 minutes) since the beginning of a freeway work zone till 2 hours after the work zone has been removed;

Q_j = Traffic volume approaching the work zone at time *j* (vph);

C_w = Work zone capacity (vph);

a_{ij}, b_{ij} = Freeway model coefficients of segment *i* at time *j*; and

c, d = Arrays of freeway model coefficients.

In Eq. 3.1, arrays of coefficients *c* and *d* determine how fast the speed under work zone condition (y_{ij}) decreases from normal to congested conditions. Based on the information gathered in the database developed previously, the optimal values of a_{ij}, b_{ij}, *c*, and *d* can be determined by minimizing RMSE (as defined by Eq. 3.2) with an exhaustive search algorithm (ESA) (Hajdin and Lindenmann, 2007; Weng and Meng, 2012). The detailed step procedure of ESA will be discussed in Chapter 4. As discussed in Section 2.3, the lower the RMSE, the better is the model performance.

$$RMSE = \sqrt{\frac{1}{mn}\sum_{\forall i,j}(\hat{y}_{ij} - y_{ij})^2} \tag{3.2}$$

where:

$\hat{y}_{ij}$ = Predicted speed of segment *i* at time *j* (mph); and

y_{ij} = INRIX reported speed of segment *i* at time *j* under work zone condition (mph).

The work zone capacity (C_w) in Eq. 3.1 is approximated as a product of normal capacity (C), work zone capacity reduction factor (δ), total number of lanes (N_T), and open lane ratio (R_o). Thus,

$$C_w = C * \delta * N_T * R_o \tag{3.3}$$

where:

C = The normal capacity (vphpl);

δ = The work zone capacity reduction factor;

N_T = The total number of lanes; and

R_o = The ratio of the number of open lanes to the total number of lanes.

The optimal values of capacity reduction factors in Eq. 3.3 under different lane configurations (i.e., 2-lane, 3-lane, and 4-lane) can be determined by minimizing the corresponding RMSE with ESA.

It is worth noting that it is important to identify the spatio-temporal boundaries (i.e., *m* and *n* in Eq. 3.1) by comparing the predicted speed ($\hat{y}_{ij}$) to the normal speed (s_{ij}) for each segment *i* at time *j*, in a similar manner to previous studies (Chung, 2011; Chung et al., 2012; Du et al., 2016). There are two possible conditions that can be observed:

- If the predicted speed of segment *i* at time *j* is lower than or equal to 75% of its normal speed value (i.e., $\hat{y}_{ij} \leq 0.75s_{ij}$), that segment is considered as negatively affected by the work zone at time *j*. Adjacent upstream segments meeting this condition are joined together to form the queue.
- When predicted speeds on every upstream segment associated with a queue have returned to values greater than 75% of their normal speed values (i.e., $\hat{y}_{ij} > 0.75s_{ij}$) and prevailed for one time interval (e.g., 15 minutes), the work zone impact is considered cleared.

Therefore, the congestion status of upstream segment *i* at time *j* is associated with a binary variable denoted as τ_{ij}. If the congestion status is positively affected by the work zone, τ_{ij} = 1; otherwise, τ_{ij}= 0. Thus,

$$\tau_{ij} = \begin{cases} 1 & if\ \hat{y}_{ij} \leq 0.75s_{ij} \\ 0 & otherwise \end{cases} \tag{3.4}$$

Relative to the display of information in Table 3.2, an example of the negative effects (i.e., speed reduction) of the work zone can be identified diagrammatically as

shown in Table 3.3. The negative effects of the work zone will be propagated from the work zone location to upstream sections. Such a discontinuity between non-congested and congested traffic flows and speeds is the reason for instabilities, spreading of shock waves, and formation of congestion with stop-and-go waves. With the criteria defined above, the segments affected by the reconstruction project (light red cells in Table 3.3) can be identified. Then the spatio-temporal boundaries of the work zone can be determined accordingly. It should be pointed out that the criteria of determining spatio-temporal boundary can be adjusted based on user preference.

Table 3.3 Example of Freeway Sections Impacted by the Work Zone

Traffic Flow Direction →	**Freeway Section**	**Time Interval**									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**...**	***n***
	1	$\hat{y}_{11}$	$\hat{y}_{12}$	$\hat{y}_{13}$	$\hat{y}_{14}$	$\hat{y}_{15}$	$\hat{y}_{16}$	$\hat{y}_{17}$	$\hat{y}_{18}$	...	$\hat{y}_{1n}$
	2	$\hat{y}_{21}$	$\hat{y}_{22}$	$\hat{y}_{23}$	$\hat{y}_{24}$	$\hat{y}_{25}$	$\hat{y}_{26}$	$\hat{y}_{27}$	$\hat{y}_{28}$	...	$\hat{y}_{2n}$
	3	$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$	$\hat{y}_{35}$	$\hat{y}_{36}$	$\hat{y}_{37}$	$\hat{y}_{38}$	...	$\hat{y}_{3n}$
	4	$\hat{y}_{41}$	$\hat{y}_{42}$	$\hat{y}_{43}$	$\hat{y}_{44}$	$\hat{y}_{45}$	$\hat{y}_{46}$	$\hat{y}_{47}$	$\hat{y}_{48}$	...	$\hat{y}_{4n}$
	5	$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$	$\hat{y}_{55}$	$\hat{y}_{56}$	$\hat{y}_{57}$	$\hat{y}_{58}$	...	$\hat{y}_{5n}$
	6	$\hat{y}_{61}$	$\hat{y}_{62}$	$\hat{y}_{63}$	$\hat{y}_{64}$	$\hat{y}_{65}$	$\hat{y}_{66}$	$\hat{y}_{67}$	$\hat{y}_{68}$	...	$\hat{y}_{6n}$
	7	$\hat{y}_{71}$	$\hat{y}_{72}$	$\hat{y}_{73}$	$\hat{y}_{74}$	$\hat{y}_{75}$	$\hat{y}_{76}$	$\hat{y}_{77}$	$\hat{y}_{78}$	...	$\hat{y}_{7n}$
	8	$\hat{y}_{81}$	$\hat{y}_{82}$	$\hat{y}_{83}$	$\hat{y}_{84}$	$\hat{y}_{85}$	$\hat{y}_{86}$	$\widehat{y_{87}}$	$\hat{y}_{88}$	...	$\hat{y}_{8n}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...	⋮
	m	$\hat{y}_{m1}$	$\hat{y}_{m2}$	$\hat{y}_{m3}$	$\hat{y}_{m4}$	$\hat{y}_{m5}$	$\hat{y}_{m6}$	$\hat{y}_{m7}$	$\hat{y}_{m8}$	...	$\hat{y}_{mn}$

Set of freeway sections impacted by the work zone

As mentioned earlier, the accurate prediction of traffic delay is of utmost importance in supporting the efficient planning of work zones for transportation agencies (e.g., traffic management centers, metropolitan planning organizations and state DOTs). The predicted spatio-temporal speeds under work zone condition with the developed MNR model can be used for assessing work zone impacts (e.g., delay, delay cost, and queue length). The work zone delay (*D*) can be defined as the additional delay produced by the reduced speed caused by the work zone ($\hat{y}_{ij}$) over the normal speed (s_{ij}), which can be calculated by Eq. 3.4.

$$D = \sum_{\forall i,j} max\left\{ l_i \left[\frac{1}{\hat{y}_{ij}} - \frac{1}{s_{ij}} \right] V_{ij}, 0 \right\}, \qquad \forall \tau_{ij} = 1 \tag{3.5}$$

where:

D = The total queue delay caused by the work zone (veh-hr);

l_i = The length of freeway segment *i* (mi); and

V_{ij} = The traffic volume of segment *i* at time *j* (veh).

Consequently, the congested impact length (or called queue length) at time *j* can be measured as a summation of congested segments positively affected by the work zone at time *j*.

$$L_j = \sum_{\forall i,j} (\tau_{ij} l_i), \qquad \forall \tau_{ij} = 1 \tag{3.6}$$

where:

L_j = The queue length at time *j* (mi).

Note that the maximum queue length can thus be determined as the greatest L_j within the work zone period. In addition to delay and queue length, the MNR model determines the delay cost to road users caused by work zones. Considering the values of travel time delay for passenger cars and heavy vehicles, the delay cost is equal to the sum of delays consumed by passenger cars and heavy vehicles multiplied by the corresponding values of time. Thus,

$$C_d = D(P_c\mu_c + P_t\mu_t) \tag{3.7}$$

where:

C_d = The delay cost to road users caused by the work zone ($/zone);

P_c = The percent of passenger cars;

P_t = The percent of heavy vehicles;

μ_c = The value of travel time delay for passenger cars ($/veh-hr); and

μ_t = The value of travel time delay for heavy vehicles ($/veh-hr).

The percent of passenger cars and heavy vehicles can be obtained from available traffic counts database. The monetary values of travel time delay for passenger cars and heavy vehicles can be determined based on user preference.

3.2.2 The Artificial Neural Network Model

To enhance the MNR model for predicting freeway work zone delay, a multi-layer feed-forward ANN is proposed for predicting the spatio-temporal delays caused by a pre-scheduled freeway work zone. In the proposed ANN model, the SVM model is in charge of predicting the restricted capacity caused by a work zone.

As discussed in the literature review, SVM is very friendly to use and has the particular strength of overcoming the over-fitting problem and local minima. To develop the SVM model, the whole dataset processed in Section 3.1 was randomly split into three subsets for training (70%), validation (20%), and testing (10%). Based on previous studies (Kim et al., 2001; Adeli and Jiang, 2003; Edara and Cottrell, 2007; Du et al., 2015), the principal training vectors in the SVM model may include but not limited to: number of lanes, number of open lanes, work zone length, upstream traffic volume, heavy vehicle percentage, and average upstream speed. The actual choice of training vectors can be determined based on data availability and Pearson and Spearman test results.

The basic idea of SVM, as shown in Figure 3.2, is to map the training vectors mentioned previously into a higher dimensional space via a kernel function and then construct a separating hyper-plane with maximum margin (dash lines in Figure 3.2). Finally, a SVM model is created to predict the work zone capacity (C_w).

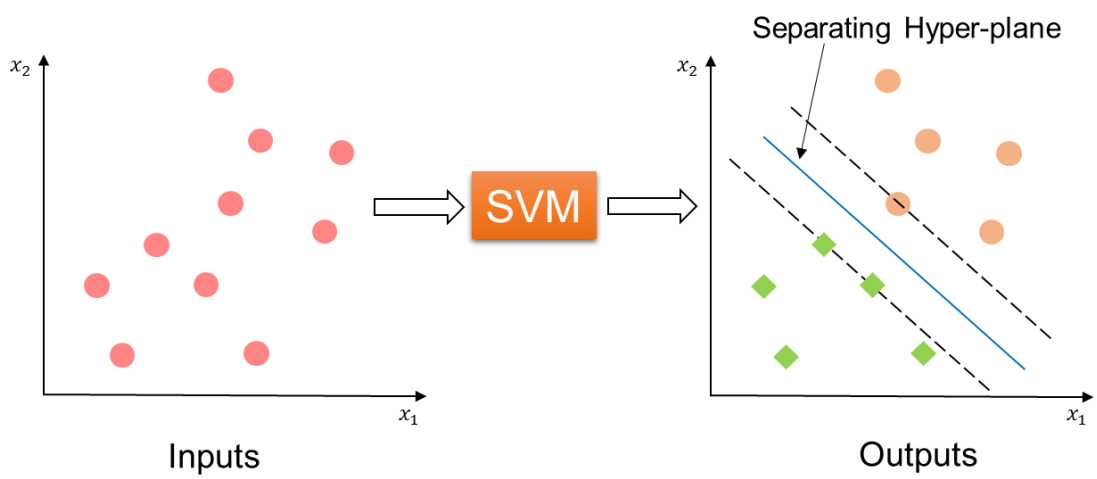


Figure 3.2 Basic concept of the SVM model.

To be more specific, given a non-linear training data set of l instance-label pairs (x_p, C_{wp}), $p = 1, \dots, l$ where l is the total number of training samples, $x_p \in R^N$ consists of N training vectors, and $C_{wp} \in R$ is the work zone capacity of the corresponding sample p. The non-linear relationship between x_p and C_{wp} can be linearized:

$$C_w = f(x) = \omega^T \phi(x) + \beta \tag{3.8}$$

where ω is the vector of coefficients, T is the transposition of the matrix, β is a constant, and ϕ is a non-linear transformation from R^N to a higher dimensional space. To find the value of ω and β, SVM requires the solution of the following optimization problem:

$$\min_{w,b,\xi_p} \frac{1}{2}\omega^T\omega + \lambda \sum_{p=1}^{l} \xi_p$$
$$subject\ to \begin{cases} C_{wp}\left(\omega^T x_p + \beta\right) \geq 1 - \xi_p \\ \xi_p \geq 0 \end{cases} \tag{3.9}$$

where ξ_p is a slack variable and λ is a regularization parameter. It is noted that Eq. 3.9 is known as an error function and the subscript p indicates the iteration number of training. That is, the training process of SVM tends to minimize the error function by updating the training vectors iteratively. It is also worth noting that an increasing λ places more weight on the slack variable ξ_p, meaning that the optimization attempts to make a stricter separation between classes. By solving for the Lagrangian dual of the above problem, a dual problem is introduced:

$$min \frac{1}{2}\sum_{p=1}^{l}\sum_{q=1}^{l} C_{wp}C_{wq}\alpha_p\alpha_q\, k(x_p, x_q) - \sum_{p=1}^{l}\alpha_p$$

$$subject\ to \sum_{p=1}^{l} C_{wp}\alpha_p = 0\,, \quad 0 \le \alpha_p \le \lambda, \quad p, q = 1,2, \dots, l \tag{3.10}$$

where α_p is the Lagrange multiplier vector and $k(x_p, x_q)$ is the kernel function. There are several types of kernel functions, including linear, polynomial, radial basis, and sigmoid kernel functions (Wu et al., 2004; Zhang and Xie, 2007; Xiao and Liu, 2012). One of the most widely used kernel functions is the radial basis function, which is used in this study and defined in Eq. 3.11.

$$k(x_p, x_q) = exp\left\{-\gamma\left|x_p - x_q\right|^2\right\} \tag{3.11}$$

where γ is a parameter. The sequential minimal optimization algorithm (Xiao and Liu, 2012) can be used to solve the constrained quadratic problem of Eq. 3.10 and get the final decision function as follows:

$$C_w = f(x) = \sum_{p=1}^{l}\left(\alpha_p{}^* - \alpha_q\right)\cdot k(x_p, x_q) + \beta \tag{3.12}$$

where $\alpha_p{}^*$ is the optimal value of Lagrange multiplier vector α_p.

Both optimization problems in Eqs. 3.9 and 3.10 are convex optimization problems, which means once the kernel function and the parameters λ and γ are determined, there will be a global optimal solution for w and b. It is noted that these

parameters can be determined by using ESA. The configuration of tuning parameters of SVM model is shown in Figure 3.3. More details will be discussed in Chapter 4.

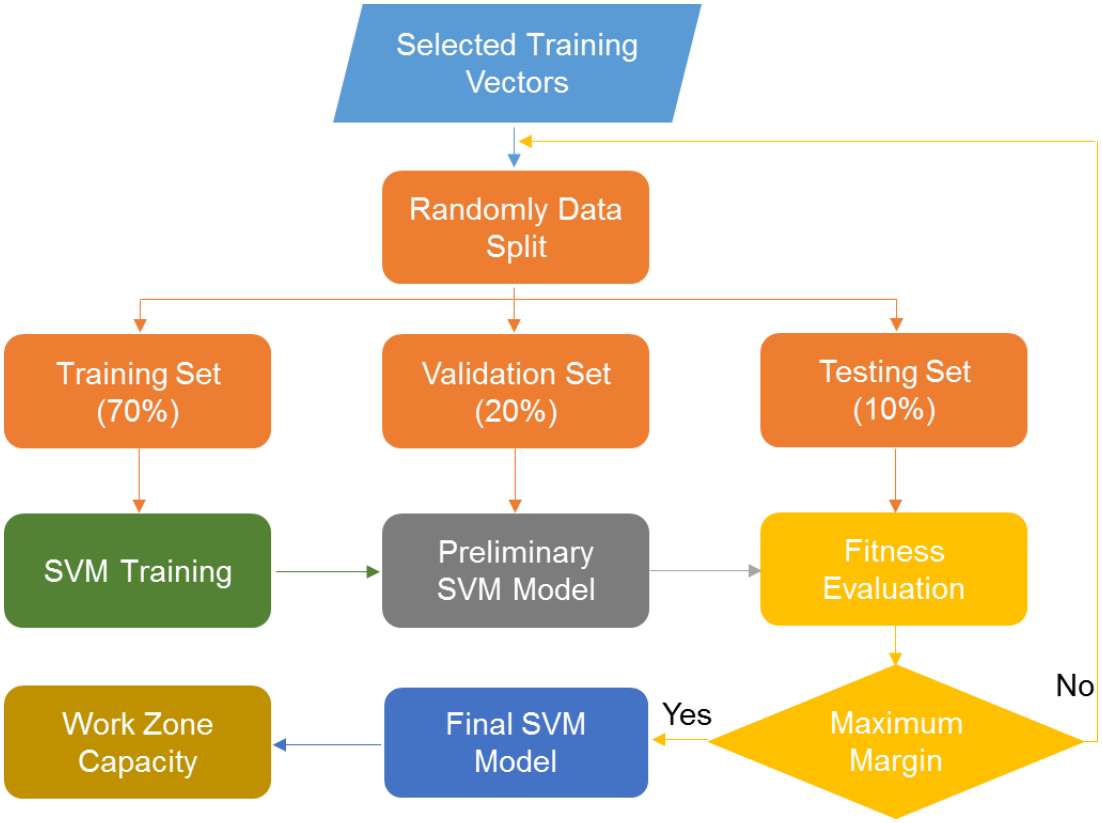


Figure 3.3 Parameter tuning of the SVM model.

After developing the SVM model, the next step is to develop the ANN model. As suggested by many researchers (Adeli and Jiang, 2003; Weng and Meng, 2013; Pan et al., 2015; Hajbabaie et al., 2015; Du et al. 2016), factors affecting the speeds of upstream a work zone include, but are not limited to, total number of lanes (N_T), number of open lanes (N_o), approaching traffic volume (Q_j), work zone capacity (C_w), heavy vehicle percentage (P_t), etc. Symbolically, the spatio-temporal work zone speed (y_{ij}) can be expressed as a

function of the selected input variables based on the Pearson and Spearman correlation tests:

$$y_{ij} = g\left(N_T, Q_j, C_w, P_t, \dots, e_r, \dots, e_M\right) \tag{3.13}$$

where:

y_{ij} = The average speed of segment *i* at time *j* under work zone condition (mph);

N_T = The total number of lanes;

Q_j = Traffic volume approaching the work zone at time *j* (vph);

C_w = Work zone capacity (vph);

P_t = The percent of heavy vehicles;

e_r = The *r*th input variable;

r = The element at input layer ($1 \leq r \leq M$); and

M = The total number of input variables.

The general configuration of the ANN model for predicting the work zone speed is shown schematically in Figure 3.4. It has an input layer with *M* nodes representing the *M* input variables included in the work zone delay prediction function defined by Eq. 3.13, multiple hidden layers with numerous neurons, and an output layer with one node representing predicted work zone speed of segment *i* at time *j* ($\hat{y}_{ij}$). The weights of input variables can be tuned based on training algorithms, such as Levenberg-Marquardt, Bayesian regularization, and scaled conjugate gradient algorithms (Chan, 2002; Karim and Adeli, 2003; Ghosh-Dastidar and Adeli, 2006). To improve the prediction accuracy, the training algorithm, and number of hidden layers and neurons of the ANN model must be carefully determined, which will be discussed in details in the next chapter. Similar to the

MNR model, the RMSE is selected as the primary criterion for determining the best ANN model based on the available work zone data.

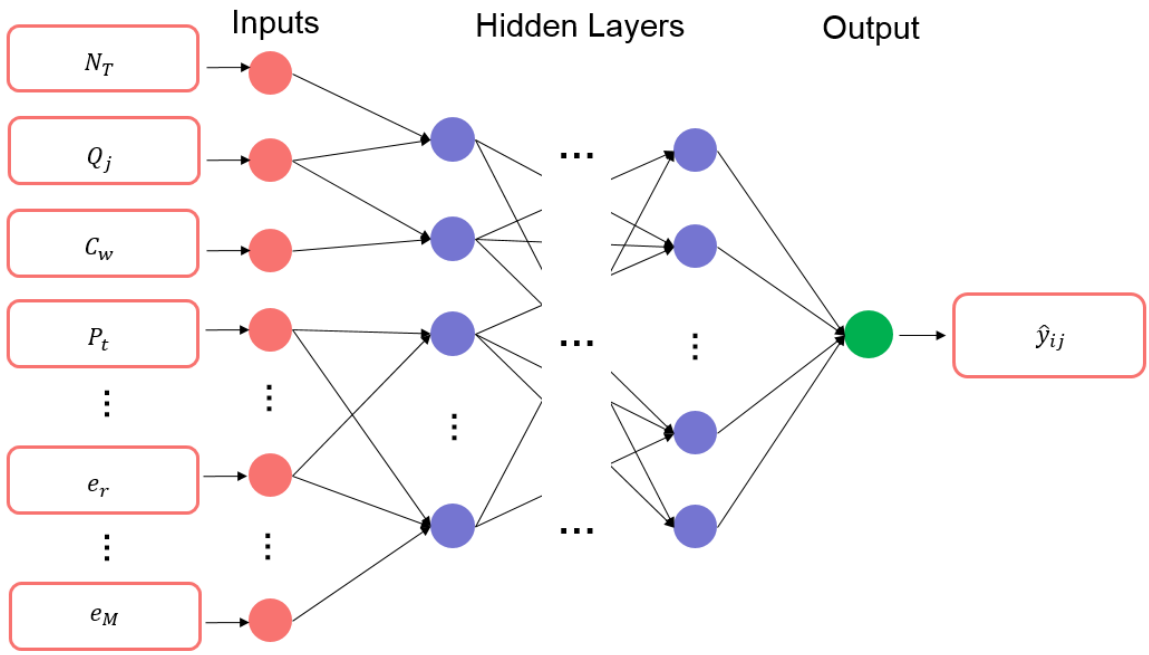


Figure 3.4 Configuration of a general ANN model.

With the predicted speed from the ANN model, the work zone delay (D), queue length (L_j), and delay cost (C_d) can be calculated using Eqs. 3.5 through 3.7.

3.3 Summary

In this chapter, two models (i.e., MNR and ANN models) are developed for freeway spatio-temporal work zone delay prediction. The MNR model is a non-linear parametric model considering approaching traffic volumes, work zone capacity, and normal speeds prior to the work zone. The input variables used for developing the MNR model can be determined based on the results of Pearson and Spearman correlation tests.

In addition to the MNR model, an ANN model is adapted to further enhance the freeway work zone delay prediction accuracy, which focuses on capturing the relationship between approaching traffic volumes and work zone capacity predicted by a SVM model. With the historical freeway work zone information and associated traffic data collected from various data sources, SVM is expected to predict the freeway work zone capacity reasonably well when it is similar to the historical profile. Similar to the MNR model, the actual choices of input variables considered in the proposed ANN model can be determined based on the Pearson and Spearman correlation tests and data availability. RMSE is selected as the primary criterion for determining the optimal model parameters of the MNR model, and suitable training algorithm, optimal number of hidden layers and neurons of the ANN model.

In Chapter 4, the parameters of these models will be determined and calibrated based on the available freeway work zone data for the years 2013 and 2014 in New Jersey. Apart from historical work zone data, the evaluation for peak-hour work zone cases will be conducted in microscopic simulation due to the absence of historical work zone data during peak hours (e.g., 6-9 AM and 3-6 PM).

CHAPTER 4

NUMERICAL EVALUATION

The freeway work zone data for years 2013 and 2014 in New Jersey are utilized here to develop the proposed models discussed in Chapter 3. The data collection and processing procedures are presented in Sections 4.1 and 4.2. Since all work zone data were incurred during off-peak periods, a microscopic traffic simulation model (VISSIM) discussed in Section 4.3 is developed to generate simulated traffic speeds under various work zone configuration and traffic conditions. The developments of the MNR and ANN models for spatio-temporal freeway work zone delay prediction are discussed in Section 4.4. Then the evaluation analysis is conducted to evaluate the model performance as shown in Section 4.5. Finally, three short-term freeway work zones on New Jersey freeways are used to further test model performance for assessing work zone impacts (e.g., delay, delay cost, and queue length) in Section 4.6.

4.1 Data Collection

To develop a sound model for predicting speed/delay caused by work zones with lane closures on freeways, both the quantity and quality of data from multiple sources are needed. Based on the available work zone data (years 2013 and 2014) in New Jersey, five data sources are identified and applied to develop a working database (see Figure 4.1).

- OpenReach DB: work zone type, location, starting/ending time, number of closed lanes, duration, and length of the work zone.
- NJSLD (New Jersey Straight Line Diagram) DB: road type, number of lanes, distance, speed limit, and interchange location.

- Plan4Safety DB: severity, location, and starting time of the accidents on New Jersey freeways.
- NJCMS (New Jersey Congestion Management System) DB: traffic volumes and heavy vehicle percentage.
- INRIX Speed DB: traffic speeds for freeway Traffic Message Channel (TMC) segments.

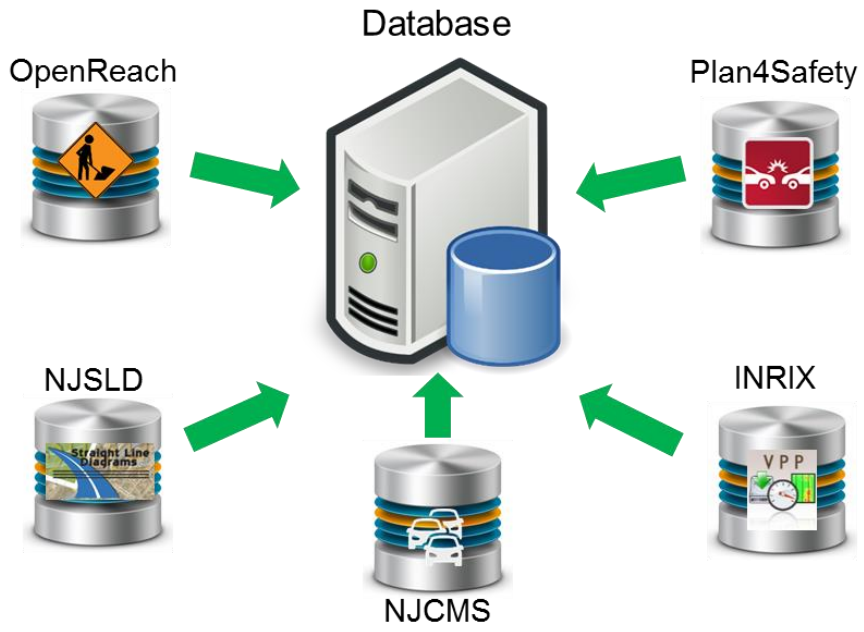


Figure 4.1 Data sources and working database.

4.1.1 OpenReach DB

The work zone data was extracted from TRANSCOM's incident reporting system called OpenReach (CoVal Systems, 2016). It receives work zone and other incident data from various sources including New Jersey Department of Transportation (NJDOT), which are then uploaded into the OpenReach DB for storage and dissemination to other TRANSCOM member agencies, traveler information providers, and the general public via the 511 traveler information system. It contains a list of work zones with location, starting and ending mileposts, description, duration and length as shown in Table 4.1. The

definition of each field can be found in Appendix A. There were more than 15,000 work zone events on New Jersey freeways in 2013 and 2014 recorded in the OpenReach DB. A data cleaning process was conducted to identify suitable freeway work zones for the model development, which will be discussed in the next section.

4.1.2 NJSLD DB

The roadway inventory and geometry data of each work zone event (e.g., standard route identifier (SRI), functional classification, total number of lanes, and presence of signalized intersections) was based on the most recent NJSLD DB (NJDOT, 2015). As shown in Figure 4.2, the NJSLD, initially designed as a planning tool is a one-dimensional graphical depiction of a section of roadway and its related data which includes the Interstate freeways, the US highways, and the State routes. The NJSLD information management system, including the data repository and software, is maintained by NJDOT's Bureau of Transportation Data Development. By using SRI and mileposts obtained from NJSLD, the travel speed within work zones and upstream of work zones can be identified. Further the main geometric characteristics of the work zone such as direction, speed limit, and number of lanes were used to develop the proposed models in this research.

Table 4.1 Sample Data Extracted from 2014 OpenReach

Event ID	Facility Name	Created Time	Closed Time	Event Type	Event Description	From Mile Marker	To Mile Marker
72747501	I-295	5/1/14 09:00	5/1/14 14:00	Construction	NJ DOT - TOC South: Construction, construction on I-295 southbound North of Exit 60 - I-195/NJ 129 (Hamilton Twp) to Exit 61 - Arena Dr (Hamilton Twp) right lane closed until 2:00 P.M.	60.5	61.4
72747901	I-80	5/1/14 09:00	5/1/14 15:00	Construction	NJ DOT - STMC: Construction, guard rail repairs on I-80 both directions between East of Exit 12 - CR 521/Hope-Blairstown Rd (Frelinghuysen Twp) and West of Exit 26 - US 46 (Mount Olive Twp) left lane closed for repairs until 3:00 P.M.	14	26
72748401	I-78	5/1/14 09:00	5/1/14 15:00	Construction	NJ DOT - STMC: Construction, pothole repair on I-78 both directions West of Exit 26 - CR 665/Rattlesnake Bridge Rd (Readington Twp) to East of Exit 41 - Dale Rd to Plainfield Ave (Watchung) right lane closed until 3:00 P.M.	26.7	42.7
72764701	I-80	5/1/14 20:00	5/2/14 06:00	Construction	NJ DOT - STMC: Construction, milling on I-80 eastbound between Exit 53 - NJ 23/US 46 (Wayne Twp) and Exit 57 - NJ 19 (Paterson) 3 left lanes closed for repairs until 6:00 A.M. 10-15 minute delay.	53.6	58.2

Source: CoVal Systems. Introduction to OpenReach: http://www.covalsystems.com/latest/openreach/openreach.html, accessed on July 10, 2016.

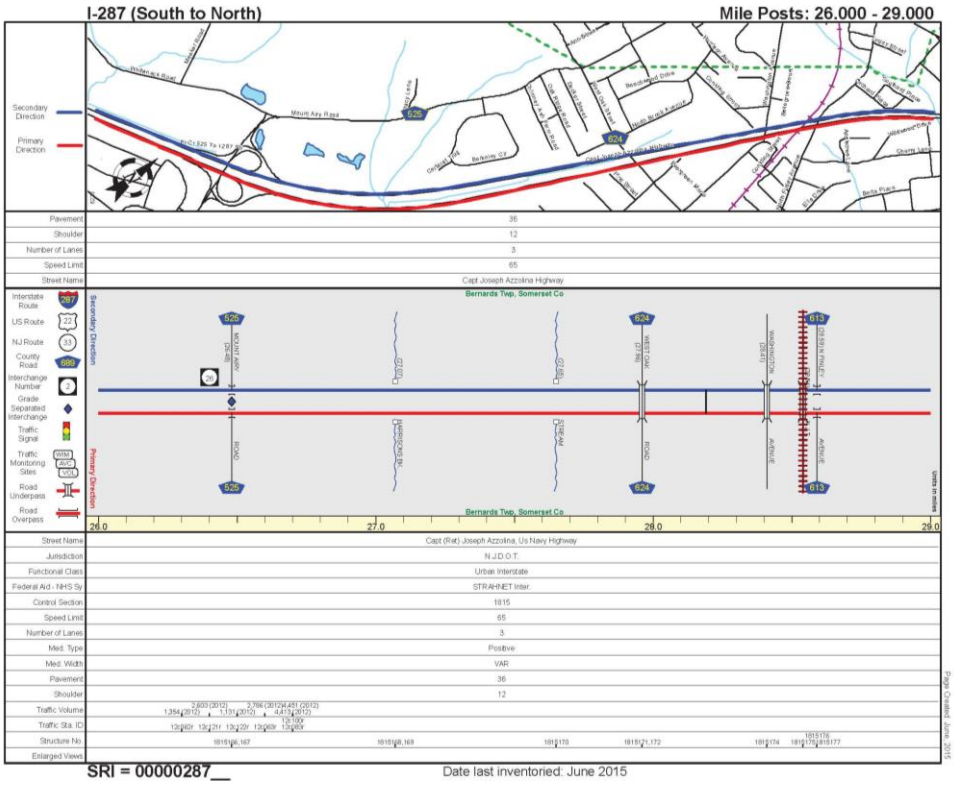


Figure 4.2 Sample data extracted from 2015 NJSLD DB.

Source: New Jersey Department of Transportation. 2015 Straight Line Diagrams Website: http://www.state.nj.us/transportation/refdata/sldiag/, accessed on Jul. 10, 2016.

4.1.3 Plan4Safety DB

Plan4Safety (Maher et al. 2016) is a multi-layered decision support program decision support tool created for the NJDOT to aid the studies conducted by transportation engineers, planners, enforcement, and decision makers in New Jersey's transportation and safety agencies. It helps to analyze crash data in geospatial and tabular forms. Similar to the NJDOT Crash Record, Plan4Safety provides crash location, date and time of the crash as shown in Table 4.2, which is used for screening out crash related work zone events in order to analyze mobility impacts purely caused by work zone activities. Nearly 11,000 freeway crashes out of more than 200,000 crashes in 2013 and 2014 on New Jersey highways were used for this screening assessment.

Table 4.2 Sample Data Extracted from 2014 Plan4Safety

DOT Web ID	Case Number	County	Crash Date	Crash Time	Severity	Crash Location	Location Direction	SRI	Milepost
2014090414-003375	14-003375	HUDSON	4/4/2014	17:08	Injury	I-280	South	00000280	14.92
20142019B130-2014-03979A	B130-2014-03979A	UNION	12/9/2014	12:56	Injury	I-78	West	00000078	53
20140414A310-2014-01381A	A310-2014-01381A	CAMDEN	7/12/2014	22:20	Property Damage	I-76	North	00000076	0.5
2014023314-03506	14-03506	BERGEN	3/7/2014	18:59	Property Damage	I-287	South	00000287	66

Source: Maher, A., M. Jafari, E. Bossett, M. O'Connell, and J. Buison. Plan4Safety Website: http://cait.rutgers.edu/tsrc/plan4safety, accessed on July 10, 2016.

4.1.4 NJCMS DB

The traffic flow data, necessary for the analysis of work zone impacts, were obtained from the most recent 2012 New Jersey Congestion Management System (NJCMS) (Chien and Ozbay, 2012). The NJCMS is data management and data analysis system used primarily by the Bureau of Systems Planning to predict congestion measures for New Jersey highways. The highway links in the NJCMS tables are identified by SRI or Route Name (e.g., I-80, or I-195), and by begin and end mileposts. The link information stored in NJCMS was tied to work zones identified in OpenReach DB using these unique link identifiers. Traffic flow data was then used to calculate link volumes in conjunction with work zone information for the model development.

4.1.5 Floating-Car Traffic Speed DB

The main traffic speed data that are used for model development in this research are historical speed data from INRIX (2016). The historical INRIX speed data is anonymously collected from GPS-enabled vehicles and mobile devices through Traffic Message Channel (TMC) and compiled into 1-minute-average speed measurements. This historic 1-minute speed data were aggregated into 15-minutes of speed data for each TMC located upstream of each work zone that was used to develop the proposed models. As shown in Figure 4.3, there are more than 1,200 TMCs in New Jersey covering interstate and express freeways. The INRIX raw data, which included more than 2 billion records, was collected for 24 hours a day over a 2-year period, from January 2013 to December 2014. This time period, including weekdays, weekends, peak, and non-peak hours, adequately reflected real traffic conditions before, during, and after work zone activities.

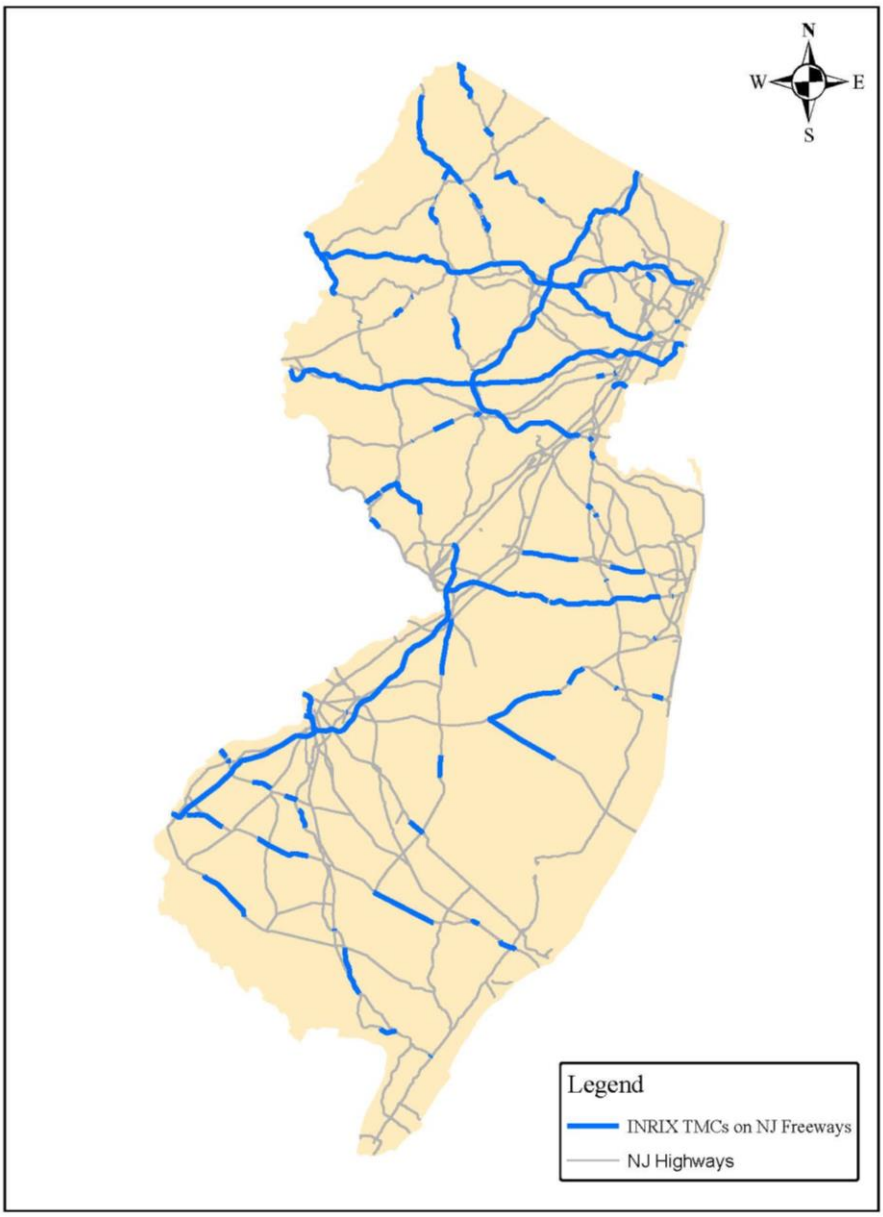


Figure 4.3 INRIX TMC locations for NJ freeways.

4.2 Data Processing

In order to identify work zones with full and accurate information needed for model development, Figure 4.4 illustrates a data cleaning procedure applied to identify work zone data suitable for developing the proposed model. This study leverages big data analytics to

process the massive amount of traffic data in an efficient and reliable way. A sample of SQL queries can be found in Appendix B.

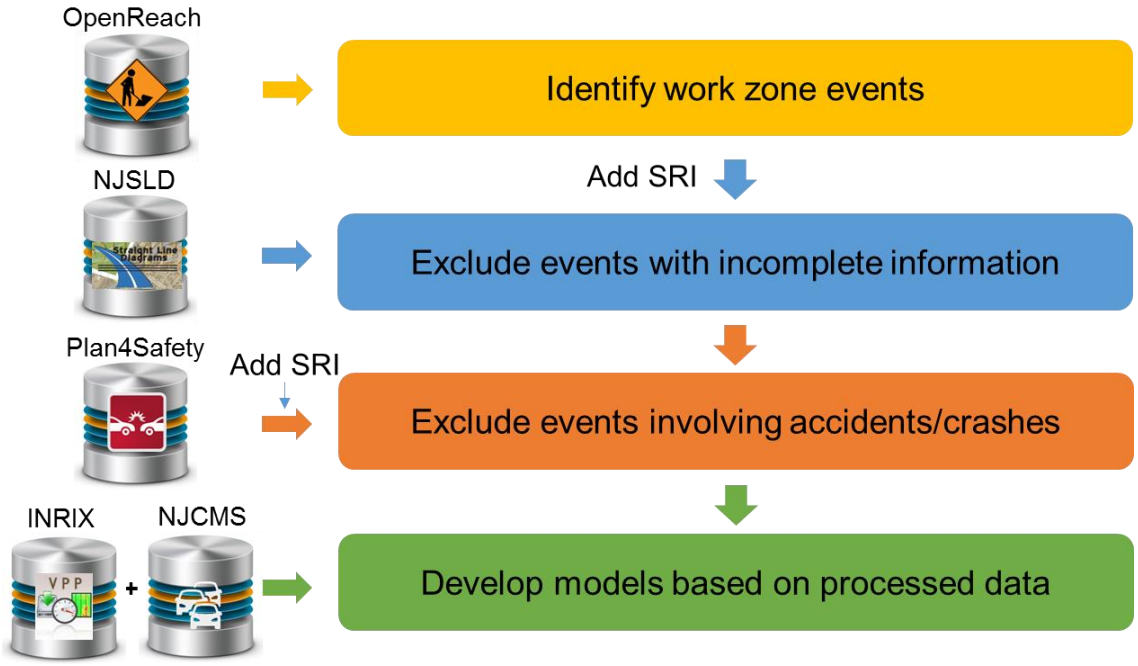


Figure 4.4 Data processing procedure.

- Step 1: Identify historical work zone events from the OpenReach incident database. Remove work zones with uncompleted information (e.g., missing work zone milepost, starting/ending date, and duration).
- Step 2: Add the standard route identifier (SRI), work zone direction, and number of lanes-closed information to each work zone based on the NJSLD database.
- Step 3: Neglect accident-related historical work zones by crosschecking accidents recorded in the Plan4Safety database.
- Step 4: Map the aggregated 15-minute speed data from INRIX for each TMC located in the upstream of each work zone identified in Step 3.

The major issues encountered during data processing are described below.

OpenReach: The DB was populated with additional information related to each work zone by adding the corresponding SRI, work zone direction and number of lanes-closed information (Table 4.3). Discrepancies in time and even the location of observed work zone and planned work zone reported in the OpenReach were observed (e.g., work zone created date was later than closed date, work zone facility name was not reported, number of closed lanes were not mentioned in "Event Description", etc.). Such inconsistent records were neglected in the model development process.

SLD: The attribute "number of lanes" in SLD often includes both directions and center turn lane in SLD. Since this attribute plays a key role in both model development and work zone impact analysis, the SLD records had to be manually updated. It is noted that only a few numbers of cases were observed to have this issue.

Plan4Safety: The SRI in Plan4Safety is different from the SLD as shown in Table 4.3. The DB developed contains manually fused data from the Plan4Safety and the SLD such that all the crash data had the same SRI consistent to the SLD SRI as illustrated in Table 4.4.

INRIX: INRIX reported speed was on a TMC basis, which has only starting and ending coordinates in the original data, while the corresponding work zones in OpenReach DB are based on the SLD and mileposts. These two data sources could not be cross-referenced with each other. Therefore, a conversion methodology to associate INRIX TMC information and SLD information needed to be developed. Hence, the DB in this research fused the INRIX TMC data to the SRI-based OpenReach data.

Table 4.3 Processed 2014 OpenReach Sample Data

Event ID	Facility Name	SRI*	Direction*	Created Time	Closed Time	Event Type	Event Description	From Mile Marker	To Mile Marker	Closure Lane*
72747501	I-295	00000295_S	Southbound	5/1/14 09:00	5/1/14 14:00	Construction	NJ DOT - TOC South: construction on I-295 southbound North of Exit 60 to Exit 61, right lane closed until 2:00 P.M.	60.5	61.4	Right lane
72747901	I-80	00000080__	Eastbound	5/1/14 09:00	5/1/14 15:00	Construction	NJ DOT - STMC: guard rail repairs on I-80 eastbound between East of Exit 12 and West of Exit 26, left lane closed for repairs until 3:00 P.M.	14	26	Left lane
72748401	I-78	00000078__	Eastbound	5/1/14 09:00	5/1/14 15:00	Construction	NJ DOT - STMC: pothole repair on I-78 eastbound between West of Exit 26 to East of Exit 41, right lane closed until 3:00 P.M.	26.7	42.7	Right lane
72764701	I-80	00000080__	Eastbound	5/1/14 20:00	5/2/14 06:00	Construction	NJ DOT - STMC: milling on I-80 eastbound between Exit 53 and Exit 57, left lane closed for repairs until 6:00 A.M. 10-15 minute delay.	53.6	58.2	Left lane

*: Manually added columns.

NJCMS: The SRI coded in NJCMS is non-directional, which has been manually corrected to match the directional SRI coded in SLD DB. With this correction, the link traffic volume by direction can be accurately stored in the working database.

After conducting the data cleaning procedure, there are 274 work zones qualified for developing the proposed model. Figure 4.5 illustrates the variation of work zone duration by number of lanes using a box plot. The median of work zone durations on 4-lane freeways is slightly greater comparing to those on 2-lane and 3-lane freeways. In addition, the duration distribution of selected historical work zones on 3-lane freeways is more dispersed than the other two categories.

It was also found that all 274 historical short-term work zones occurred during either night time or middle of day. In other words, no short-term work zones were found during peak hours (i.e., 6-9 AM and 3-6 PM). Therefore, the evaluation for the cases of work zone duration crossing peak-hour was conducted by comparing the predicted speeds to the speeds generated by microscopic simulation.

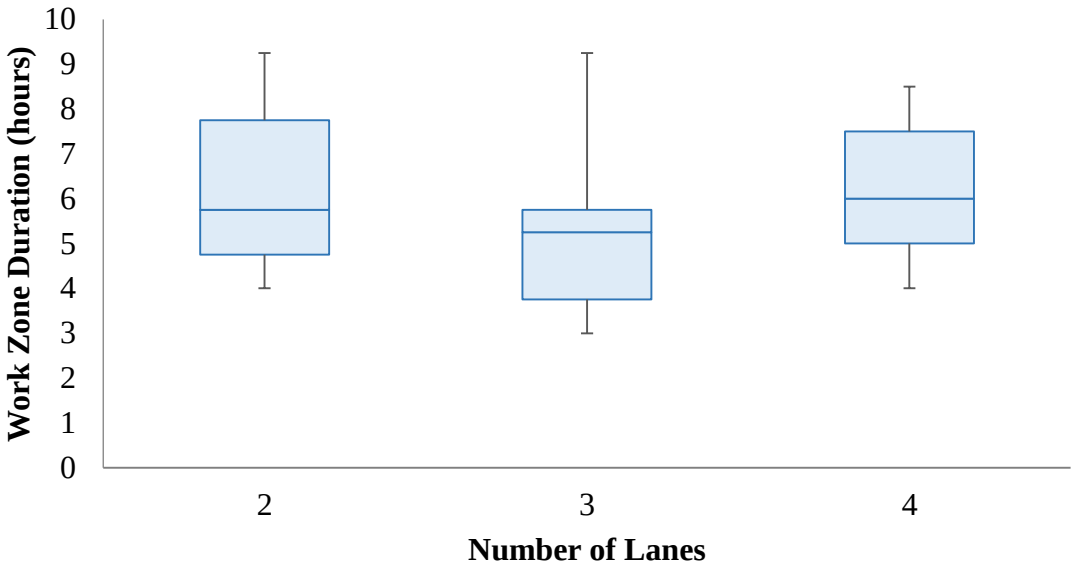


Figure 4.5 The box plot of work zone duration by number of lanes.

Table 4.4 Processed 2014 Plan4Safety Sample Data

DOT Web ID	Case Number	County	Crash Date	Crash Time	Severity	Crash Location	Location Direction	SRI	Modified SRI*	Milepost
2014090414-003375	14-003375	HUDSON	4/4/2014	17:08	Injury	I-280	South	00000280	00000280__	14.92
20142019B130-2014-03979A	B130-2014-03979A	UNION	12/9/2014	12:56	Injury	I-78	West	00000078	00000078_W	53
20140414A310-2014-01381A	A310-2014-01381A	CAMDEN	7/12/2014	22:20	Property Damage	I-76	North	00000076	00000076__	0.5
2014023314-03506	14-03506	BERGEN	3/7/2014	18:59	Property Damage	I-287	South	00000287	00000287_S	66

*: Manually added column.

4.3 Simulated Work Zone Data

Given that conducting reconstruction projects during peak hours in the real world would be impractical, VISSIM was selected to generate traffic data under normal and work zone conditions during peak hours based on its capabilities. VISSIM can simulate traffic operations in a complex and large-scale roadway network, in which every vehicle is modeled as a distinct object and follows a stochastic lane-change, car-following, and gap acceptance logic. The vehicle movement is updated in every 0.1 second (or a shorter interval depending on the computer capability) to regenerate the status of vehicles. The stochastic factors such as driver behavior characteristics, vehicle characteristics, and traffic characteristics are considered to simulate the relationships among vehicles on the links. Therefore, VISSIM is able to generate various traffic data (e.g., traffic speed and volumes on user-specified links aggregated in different time intervals under normal and work zone conditions) for evaluating the developed prediction model.

Three hypothetical work zones located on 2-lane, 3-lane, and 4-lane freeways were identified for generating simulated traffic data. The characteristics of them and the total number of simulation runs are shown in Table 4.5. The road geometry and lane configuration data were obtained from NJSLD DB. Traffic volumes were obtained from NJCMS DB. In the VISSIM network, a total of 20 sensors were placed at 0.5-mile spacing throughout the 10 miles upstream of each hypothetical work zone location. The simulated traffic volumes and speeds were measured in every 15 minutes at each data collection point. In order to consider traffic flow variations, five random seeded VISSIM simulation runs were made for each work zone combination. Therefore, a total of 1,980 simulation runs (440 + 660 + 880 = 1,980) were made.

Table 4.5 Characteristics of the Hypothetical Work Zones

Work Zone Location	No. of Lanes	Work Zone Duration (hr)	Work Zone Starting Time	No. of Closed Lanes	AADT	No. of Simulation Runs
I-295 NB, MP 27-27.5	2	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22	7 AM, 10 AM, 4 PM, 8 PM	0*, 1	45,269	440
I-80 WB, MP 42-43	3	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22	7 AM, 10 AM, 4 PM, 8 PM	0, 1, 2	80,476	660
NJ-42 NB, MP 12-12.8	4	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22	7 AM, 10 AM, 4 PM, 8 PM	0, 1, 2, 3	91,304	880

*: 0 represents shoulder closure.

The COM interface from VISSIM 7 using C# program was used to program the work zone lane closure combinations listed in Table 4.5. The VISSIM network, prior to conducting experimental analyses, had also been calibrated with field-collected traffic data to match the real-world conditions by tweaking car following and lane changing models parameters. Figures 4.6 through 4.8 show that the simulated traffic volumes are very close to the traffic counts obtained from NJCMS DB for all three cases.

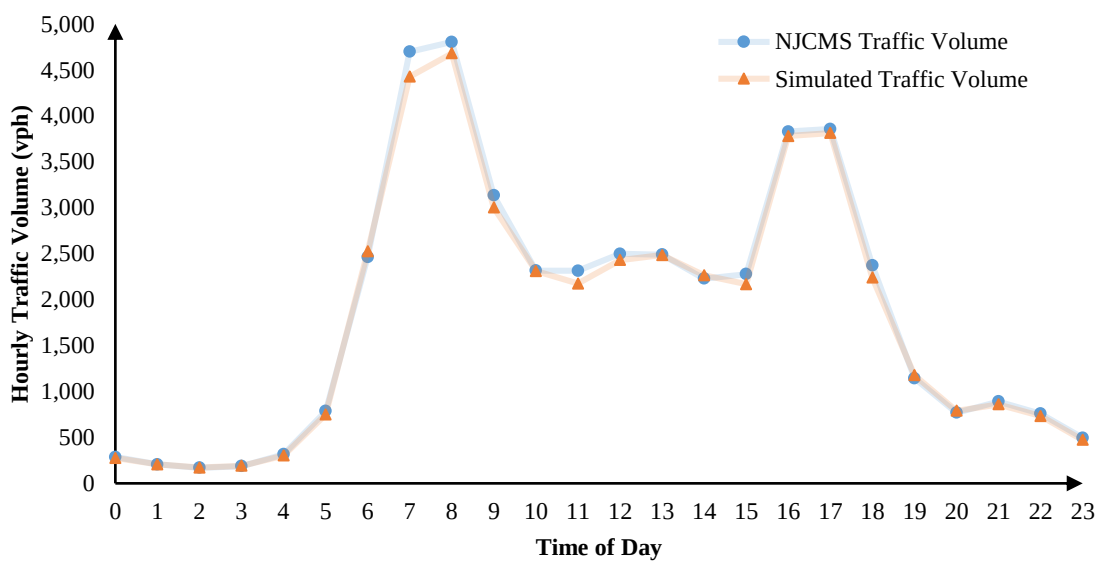


Figure 4.6 NJCMS and simulated traffic volumes at MP 27 on I-295 NB.

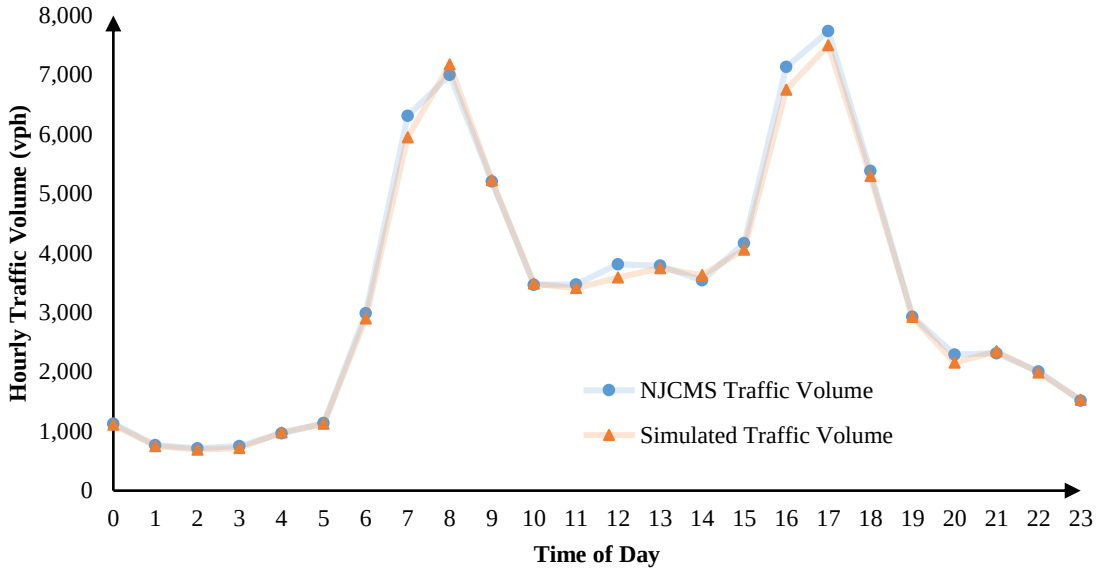


Figure 4.7 NJCMS and simulated traffic volumes on at MP 42 on I-80 WB.

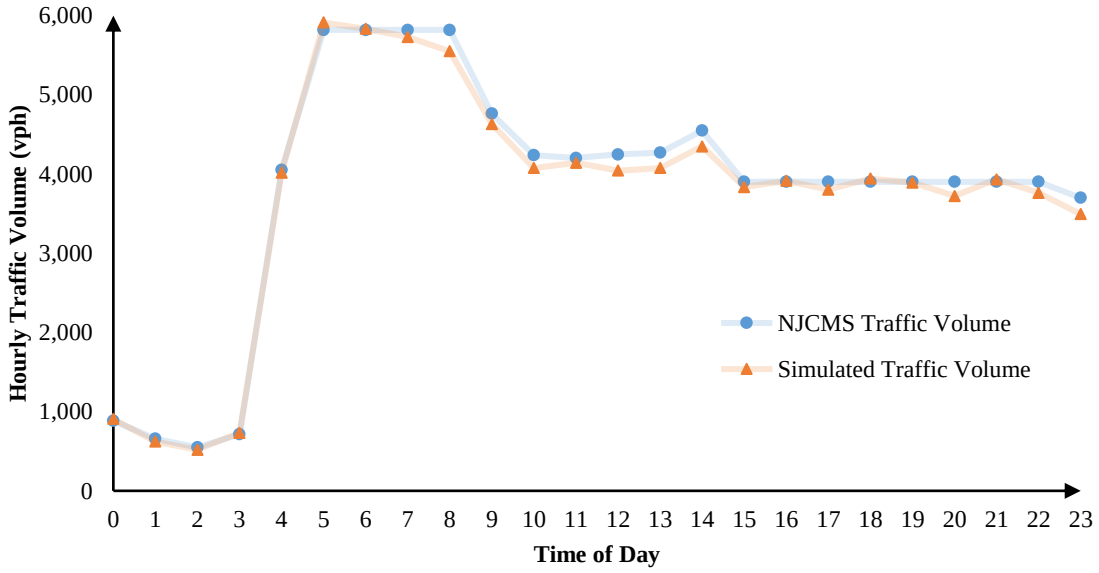


Figure 4.8 NJCMS and simulated traffic volumes at MP 12 on NJ-42 NB.

3.4 Model Development

In this section, the multivariate non-linear regression (MNR) model and the multi-layer feed forward artificial neural network (ANN) model are developed and discussed next.

4.4.1 The Multivariate Non-linear Regression Model

Based on the developed database in previous sections, a multivariate non-linear regression (MNR) model was developed to predict speeds caused by work zone lane closures. This section describes the configuration of the MNR model followed by identification of inputs for its implementation on freeways.

It is worth noting that the predicted speeds can be extended up to 10 miles upstream of the work zone and 2 hours after the work zone is removed, in the absence of any accidents during the analysis period. These limits were determined based on the 2013 and 2014 freeway work zone data collected in New Jersey as discussed in Sections 4.1 and 4.2. The formula for the MNR model whose development processes are discussed next.

As discussed in Chapter 3, by assessing the database developed in previous sections, the work zone capacity, approaching volume, and normal speed of upstream segment are selected as model inputs. Therefore, the freeway MNR model is formulated for different lane configurations namely, 2-lane, 3-lane and 4-lane, which is formulated as Eq. 4.1.

$$y_{ij} = a_{ij} + b_{ij} \frac{s_{ij}}{1 + c\left(\frac{Q_j}{C_w}\right)^d} \tag{4.1}$$

where:

y_{ij} = The average speed of segment *i* at time *j* under work zone condition (mph);

s_{ij} = The average speed of segment *i* at time *j* under normal condition (mph);

Q_j = Traffic volume approaching the work zone at time *j* (vph);

C_w = Work zone capacity (vph);

a_{ij}, b_{ij} = Freeway model coefficients of segment *i* at time *j*; and

c, d = Arrays of freeway model coefficients.

In Eq. 4.1, the work zone capacity (C_w) is approximated as a product of normal capacity (C), work zone capacity reduction factor (δ), total number of lanes (N_T), and open lane ratio (R_o). Thus,

$$C_w = C * \delta * N_T * R_o \tag{4.2}$$

where:

C = The normal capacity (vphpl);

δ = The work zone capacity reduction factor;

N_T = The total number of lanes; and

R_o = The ratio of the number of open lanes to the total number of lanes.

The freeway MNR model was based on a data set of 274 work zones, which were selected due to the completeness of their data that were deemed useful in developing the model. These 274 records covered a total of eight work zone types as shown in Table 4.6. For each lane closure type presented in Table 4.6 (i.e., shoulder, 1-lane, and 2-lane closures), the first randomly selected 70% of the qualified freeway work zones in 2013 and 2014 were used to develop the MNR model; the next 20% work zone data were used for validation; and the rest 10% work zone data were used for testing.

Table 4.6 Number of Work Zones for Various Types of Lane Closures

No. of Lanes	Shoulder Closure	1-lane Closure	2-lane Closure
2	10	62	N/A
3	30	104	12
4	7	36	13

The values of capacity reduction factors are shown in Table 4.7. Table 4.8 illustrates the optimal coefficients values of MNR with respect to different lane configurations, which were also determined based on 70% of the historical work zone data. It is noted that these parameters were determined by using ESA. The steps of the ESA processes are discussed below.

Step 1: Set c = 0.1.

Step 2: Set d = 1.

Step 3: Set δ = 1.

Step 4: Calculate $\hat{y}_{ij}$ using Eq. 4.4 based on 70% of the work zone data.

Step 5: Calculate RMSE using Eq. 3.2.

Step 6: Let $\delta = \delta - 0.05$. If $\delta > 0$, go to Step 4; otherwise, go to Step 7.

Step 7: Let $d = d + 0.1$. If $d < 4$, go to Step 3; otherwise, go to Step 8.

Step 8: Let $c = c + 0.05$. If $c < 1$, go to Step 2; otherwise, go to Step 9.

Step 9: Find the optimal combination of a_{ij}, b_{ij}, c, d, and δ with the least RMSE.

Table 4.7 Capacity Reduction Factors δ and R_o

Lane Closure Configuration	2-lane		3-lane		4-lane	
	δ	R_o	δ	R_o	δ	R_o
Shoulder Closure	0.9	1	0.95	1	0.95	1
1-lane Closure	0.5	0.5	0.6	0.66	0.7	0.75
2-lane Closure	-	-	0.5	0.33	0.6	0.5
3-lane Closure	-	-	-	-	0.5	0.25

Table 4.8 The MNR Model Coefficients

No. of Lanes	Coefficients	
	c	d
2	0.1	2.9
3	0.1	2.1
4	0.1	2.3

With the predicted work zone speed from MNR model, work zone delay, delay cost, and queue length can be calculated using Eqs. 3.5 - 3.7, accordingly.

4.4.2 The Artificial Neural Network Model

This section deals with a non-parametric approach for predicting short-term work zone delay on freeways by an artificial neural network (ANN) model, in which the work zone capacity is predicted by SVM. The SVM is developed using traffic data (i.e., traffic volume and speed) based on nine work zones in years 2014 and 2015. The speed data were gathered from INRIX, and traffic volumes were collected by Remote Traffic Microwave Sensors (RTMS). Since most work zones were conducted during off-peak periods only,

simulated data for peak period are applied using VISSIM. As discussed in Chapter 2, the RTMS provides per-lane presence, volume, occupancy, speed, as well as vehicle classification information. Currently there are nine RTMSs deployed around I-295/I-76/Route 42 Direct Connection project location, which collect traffic volume data (aggregated at 5 minutes) for 24 hours a day from November 2014 to July 2015.

Based on the combined dataset and the results of the Pearson and Spearman tests, six training vectors were selected for developing the SVM model in this research, which are number of lanes, number of open lanes, work zone length, upstream traffic volume, heavy vehicle percentage, and average upstream speed. These vectors were first randomly split into three subsets for the purposes of training, validation and testing. Then, the SVM model mapped these training vectors into a higher dimensional space using a radial basis kernel function. There are several types of kernel functions, including linear, polynomial, radial basis, sigmoid and automatic relevance determination kernel functions (Zhang and Xie, 2008; Yu and Abdel-Aty, 2013). Then the optimal hyper-planes were determined by maximizing the margins of the training vectors. Finally, a trained SVM model was created to predict the work zone capacity (C_w). More detailed information has been discussed in Chapter 3.

After the development of SVM, the next step is to develop an artificial neural network (ANN) model. Similar to the development of MNR, the factors affecting the speed upstream of the work zone were determined by the Pearson and Spearman correlation test, which include average speed of upstream work zone TMC segment *i* at time *j* under normal condition (s_{ij}); traffic volume approaching work zone at time j (Q_j); work zone capacity (C_w); and distance from segment i (d_i). Note that C_w is predicted by the SVM model.

Therefore, the average speed of upstream work zone TMC segment i at time j (y_{ij}) under work zone condition can be represented by Eq. 4.3:

$$y_{ij} = g\left(s_{ij}, Q_j, C_w, d_i\right) \tag{4.3}$$

As Q_j increases, the resulting travel time and delay increase, especially when it is close to the restricted capacity caused by a work zone lane closure. To represent the relationship among $s_{ij}, Q_j, and\ C_w$, the concept of the Bureau of Public Roads (BPR) function (Bureau of Public Roads, 1964) was adapted. It is assumed that the weighted speed of segment i at time j denoted as v_{ij} is the historic speed under normal condition multiplied by a reduction factor that is a function of approaching volume and work zone capacity ratio, which can be formulated as:

$$v_{ij} = \frac{s_{ij}}{1 + \mathrm{A}\left(\frac{Q_j}{C_w}\right)^B} \tag{4.4}$$

where:

s_{ij} = The speed of segment i at time j under normal condition (mph);

Q_j = The traffic volume approaching the work zone at time j (vph);

C_w = The work zone capacity (vph);

A, B = The arrays of model coefficients;

i = The ith freeway segment in upstream of work zone ($1 \leq i \leq m$);

j = The jth time interval after work zone started ($1 \leq j \leq n$);

m = The number of segments (e.g., TMCs) upstream of work zone; and

n = The number of time intervals (e.g., 15 minutes per interval) since the beginning of a freeway work zone till 2 hours after the work zone has been removed.

It will be discussed later how the optimal value of parameters *A* and *B* can be determined. With the weighted speed (v_{ij}) from Eq. 4.4 and the distance from segment *i* to the work zone (d_i), the work zone speed (y_{ij}) in Eq. 4.3 can be simplified as:

$$y_{ij} = g(v_{ij}, d_i) \tag{4.5}$$

The Neural Network Toolbox in MATLAB (2016) was used for developing the ANN model. As discussed earlier, there were 274 number of freeway work zones available, which were randomly divided into three groups (i.e., 70%, 20%, and 10% of total work zones, respectively) for training, validation, and testing purposes. It is worth noting that different divisions had been investigated and it was possible to get a minimum error using the above combination. The root mean square error (RMSE) formulated as Eq. 3.2 was used as an index to determine the optimal combination of *A* and *B* in Eq. 4.4, the suitable training algorithm, and optimal numbers of hidden layers and neurons by using ESA. The lower the RMSE value, the better is the model performance. The steps of the ESA processes are discussed below.

Step 1: Set $A = 0.1$.

Step 2: Set $B = 1$.

Step 3: Calculate v_{ij} using Eq. 4.4. Then predict work zone speed using single layer ANN model with 10 neurons.

Step 4: Calculate RMSE using Eq. 3.2.

Step 5: Let $B = B + 0.1$. If $B < 4$, go to Step 3; otherwise, go to Step 6.

Step 6: Let $A = A + 0.05$. If $A < 1$, go to Step 2; otherwise, go to Step 7.

Step 7: Find the optimal combination of *A* and *B* with the least RMSE.

By using ESA, the optimal values of *A* and *B* in Eq. 4.4 with respect to different lane configurations are illustrated in Table 4.9, which were determined based on single layer ANN models with 10 neurons with 70% of freeway work zone data.

Table 4.9 Calibrated Model Coefficients for Predicting v_{ij}

No. of Lanes	**Coefficients**	
	A	*B*
2	0.1	2.7
3	0.1	2.6
4	0.2	2.2

After determining the optimal values of *A* and *B*, the next step is to find the best training algorithm. Table 4.10 depicts the lowest RMSEs for the three training algorithms provided by MATLAB Neural Network Toolbox (2016) based on single layer ANN models with 10 neurons. By considering work zones on 3-lane freeways, it was found that the Levenberg-Marquardt (LM) algorithm (i.e., RMSE = 4.9 mph) was selected for its better efficiency and performance, compared to Bayesian Regularization (i.e., RMSE = 5.3 mph) and Scaled Conjugate Gradient (i.e., RMSE = 5.8 mph) algorithms.

Table 4.10 RMSEs of Various Training Algorithms in the ANN Model

No. of Lanes	RMSE (mph)		
	Levenberg-Marquardt (LM)	Bayesian Regularization (BR)	Scaled Conjugate Gradient (SCG)
2	5.9	6.3	6.5
3	4.9	5.3	5.8
4	6.3	6.6	6.9

Based on the selected LM algorithm, Table 4.11 shows the RMSEs of the 1-hidden-layer and 2-hidden-layer models for the work zones on 3-lane freeways. It was found that no substantial difference occurs by adjusting number of neurons or adding an extra layer in the ANN model. Hence a single layer ANN model with 10 neurons is sufficient to predict work zone speed with satisfactory accuracy along with the benefit of reduced computation time as compared to 2 or more layers ANN models. Similarly, one-layer LM ANN model with 10 neurons is satisfactory for work zones on both 2-lane (i.e., RMSE = 5.9 mph) and 4-lane (i.e., RMSE = 6.3 mph) freeways.

Table 4.11 RMSEs of Various ANN Models (3-lane Freeway)

ANN Models	No. of Neurons		RMSE (mph)
	Layer 1	Layer 2	
1-layer ANN	5	-	5.4
	10	-	**4.9**
	15	-	5.0
2-layer ANN	5	5	5.6
	10	10	5.3
	15	15	5.2

The finalized architecture of the proposed ANN model is shown in Figure 4.9. The ANN model consist of an input layer with two neurons representing the weighted speed (v_{ij}) and distance from upstream segment i (d_i), one optimized hidden layer with ten neurons and an output layer with one neuron representing predicted work zone speed ($\hat{y}_{ij}$). In the input layer, the predicted work zone capacity (C_w) from SVM model along with normal speed (s_{ij}) and approaching traffic volumes (Q_j) were used for calculating the weighted speed (v_{ij}). It is worth noting that the proposed ANN model can predict speeds up to 10 miles upstream of the work zone since the beginning of a freeway work zone till 2 hours after the work zone has been removed.

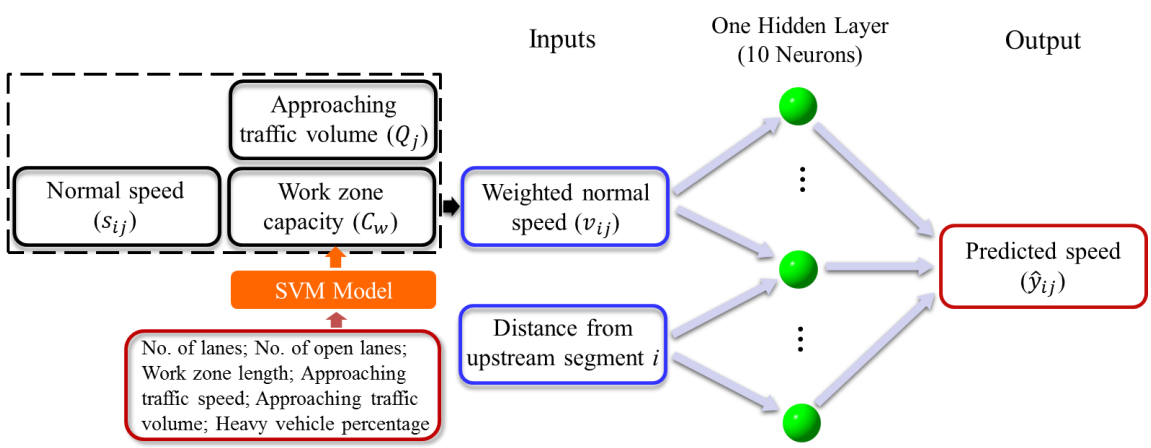


Figure 4.9 Configuration of the proposed ANN model.

Similar to MNR model, with the predicted work zone speed from ANN model, work zone delay, delay cost, and queue length can be calculated using Eq. 3.5 - 3.7 accordingly.

3.5 Model Evaluation

Based on historical work zone data for off-peak period, the performances of the two work zone delay prediction models (i.e., MNR and ANN) developed in previous section under various lane configurations (i.e., 2-lane, 3-lane, and 4-lane) and locations (i.e., North, Central, and South NJ) are assessed in this section.

First, a detailed analysis is conducted to assess the overall model performance of the MNR and ANN models for predicting delays caused by work zone activities on freeways. These two freeway work zone delay prediction models are evaluated using 10% (27) of 274 identified work zone records in 2013 and 2014. The steps taken to assess the model accuracy/reliability are listed below.

Step 1: Classify the randomly selected 27 freeway work zones by lane configuration (i.e., 2-lane, 3-lane, and 4-lane) and location (i.e., North, Central, and South NJ). The corresponding data distribution per lane and region of the selected work zones are illustrated in Table 4.12. Note that no qualified work zone was selected on 4-lane freeways in South NJ as the corresponding data for the years 2013 and 2014 were found to be insufficient to be included in the model development.

Table 4.12 Test Samples by Lane Configuration and Region

No. of Lanes	**Region**		
	North	Central	South
2	3	3	2
3	6	5	3
4	4	1	0

Step 2: Run each work zone with the freeway MNR and ANN models, respectively. Then compute the RMSE based on the predicted speeds versus the travel speeds reported from the corresponding INRIX speed database. The RMSEs under various lanes and types of lane closure by regions for two models are summarized in Table 4.13. It is found that the ANN model outperformed the MNR model for all lane configurations and regions. Table 4.13 also indicates that the ANN model yielded the lowest RMSE (RMSE = 4.9 mph) for testing historic work zones on 3-lane freeways against the 2-lane (RMSE = 5.9 mph) and 4-lane freeway (RMSE = 6.3 mph) because of more work zones available for model development.

Table 4.13 RMSE of the MNR and ANN Models (mph)

No. of Lanes	**North**		**Central**		**South**		**Overall**	
	MNR	ANN	MNR	ANN	MNR	ANN	MNR	ANN
2	8.8	5.8	9.3	6.2	5.5	5.4	8.2	5.9
3	6.3	4.6	5.6	4.9	5.5	5.3	5.9	4.9
4	6.7	6.4	6.2	5.8	N/A	N/A	6.6	6.3
Overall							**6.4**	**5.2**

Step 3: According to the results from Tables 4.13, the ANN model outperforms the MNR model in terms of smaller RMSE based on historical work zones during off-peak periods. From this step, the ANN model is further evaluated. Based on the RMSE associated with each test work zone, the average RMSEs were classified into 3 categories (i.e., < 5 mph, 5 - 10 mph, and 10 - 15 mph) by lane configuration and region as shown in

Table 4.14. Comparing the results by lane configuration, the ANN model on 3-lane freeways produced the most accurate and reliable (64% RMSE < 5 mph) predicts, followed by 2-lane (28% RMSE < 5 mph) and 4-lane (100% RMSE between 5 - 10 mph) freeways. Comparing the results by region, the predicted results of work zone delays in the Northern NJ is relatively stable and accurate (47% RMSE < 5 mph), followed by Southern NJ (40% RMSE < 5 mph) and Central NJ (32% RMSE < 5 mph). One possible reason for this is that there were more work zones on 3-lane freeways in Northern NJ available for model development.

Table 4.14 RMSE Distribution of the ANN Model

No. of Lanes	**RMSE Range**	**Region**		
		North	Central	South
2-lane	< 5 mph	33%	0%	50%
	5 - 10 mph	67%	100%	50%
	10 - 15 mph	0%	0%	0%
3-lane	< 5 mph	83%	60%	33%
	5 - 10 mph	17%	40%	67%
	10 - 15 mph	0%	0%	0%
4-lane	< 5 mph	0%	0%	0%
	5 - 10 mph	100%	100%	0%
	10 - 15 mph	0%	0%	0%
Overall	< 5 mph	47%	32%	40%
	5 - 10 mph	53%	68%	60%
	10 - 15 mph	0%	0%	0%

Step 4: To further demonstrate the model performance, the simulated data for work zones crossing peak hours were used for evaluating the performance of the ANN model. It is found in Table 4.15 that in general the ANN model could achieve satisfactory

performance for work zone speed prediction in terms of accuracy and stability during peak hours (i.e., 6.9 mph for 2-lane freeway, 5.3 mph for 3-lane freeway, and 6.7 for 4-lane freeway). This implies that the ANN model could generate prediction results with compatible accuracy when the trend of real-world traffic conditions during peak hours is similar with the simulated data. It is also found that as the number of closed lanes increases from shoulder closure to 2-lane closure, the RMSEs are slightly increased for all three lane configurations. This indicates that the traffic congestion during peak period could reduce the accuracy of the ANN model. Therefore, to improve the prediction accuracy, the actual traffic counts for peak period at the scenes of work zones should be collected from the field to replace the simulated data for further validation of the developed models.

Table 4.15 RMSE of the ANN Model based on Simulation Data

No. of Lanes	RMSE (mph)			
	Shoulder Closure	1-lane Closure	2-lane Closure	Overall
2	6.6	7.2	N/A	6.9
3	5.1	5.4	5.9	5.3
4	6.3	7.2	6.8	6.7

3.6 Case Studies

Overall, the evaluation results in Section 4.5 indicate that the ANN model is able to perform well in predicting freeway work zone delay under various lane configuration conditions and time of day. In this section, the ANN model is evaluated with new work zones in 2015, in which delay, delay cost, and maximum queue length were applied to

assess the model performance. Results from the proposed ANN model with the work zone capacity predicted by SVM (called ANN-SVM) are compared with the prediction results using other models:

- RUCM: The method suggested by the NJDOT Road User Cost Manual (NJDOT, 2015) (see Appendix C for more details);
- ANN-HCM: The proposed ANN model with work zone capacity suggested by HCM (2010) as formulated in Eq. 4.6; and
- ANN-SVM: The proposed ANN model with work zone capacity suggested by SVM.

$$C_w = (1600 + I) f_{HV} N_o - R \tag{4.6}$$

where:

C_w = The work zone capacity (vph);

I = The adjustment factor for type and intensity of work activity (vphpl);

f_{HV} = The heavy-vehicle adjustment factor indicated in the HCM;

N_o = The number of open lanes within the work zone; and

R = The manual adjustment for on-ramps (vph).

The characteristics of three short-term work zones performed in 2015 are shown in Table 4.16, which include time period, road geometry, and traffic pattern. Case 1 was a 2-mile long work zone with two-lane closure on a three-lane segment on I-78 westbound, which was performed between 11 PM to 6 AM next day in October 2015. Case 2 was a 0.3-mile long work zone with one-lane closure on a three-lane segment on NJ-21 southbound, which was performed between 10 AM and 3 PM in November 2015. Case 3 was a 0.2-mile long work zone with shoulder closure on a two-lane segment on I-280

eastbound, which was performed between 10 AN and 3 PM in December 2015. In addition, the work zone capacities suggested by the SVM model as well as the HCM method (2010) are summarized in Table 4.16. Due to the impacts of approaching traffic volume and speed are neglected in the HCM method, the predicted work zone capacity with HCM for Cases 1 and 2 are lower than those with SVM. While for Case 3, the predicted work zone capacity with HCM is greater than that with SVM. The hourly traffic distributions for all 3 cases are shown in Figure 4.10, which were used for calculating work zone delay and cost.

Table 4.16 Work Zone Characteristics

	Case 1	**Case 2**	**Case 3**
Location	I-78 WB	NJ-21 SB	I-280 EB
Milepost Range	47.3 - 49.3	4.2 - 4.5	14.1 - 14.3
Number of Lanes	3	3	2
Work Zone Length (mi)	2	0.3	0.2
Starting Time	11 PM, 10/2015	10 AM, 11/2015	10 AM, 12/2015
Ending Time	6 AM, 10/2015	3 PM, 11/2015	3 PM, 12/2015
Duration (hours)	7	5	5
Number of Closed Lanes	2	1	0*
$\boldsymbol{C_w}$ **with SVM (vph)**	1,524	3,222	3,798
$\boldsymbol{C_w}$ **with HCM (vph)**	1,395	2,976	3,910

*: Shoulder closure.

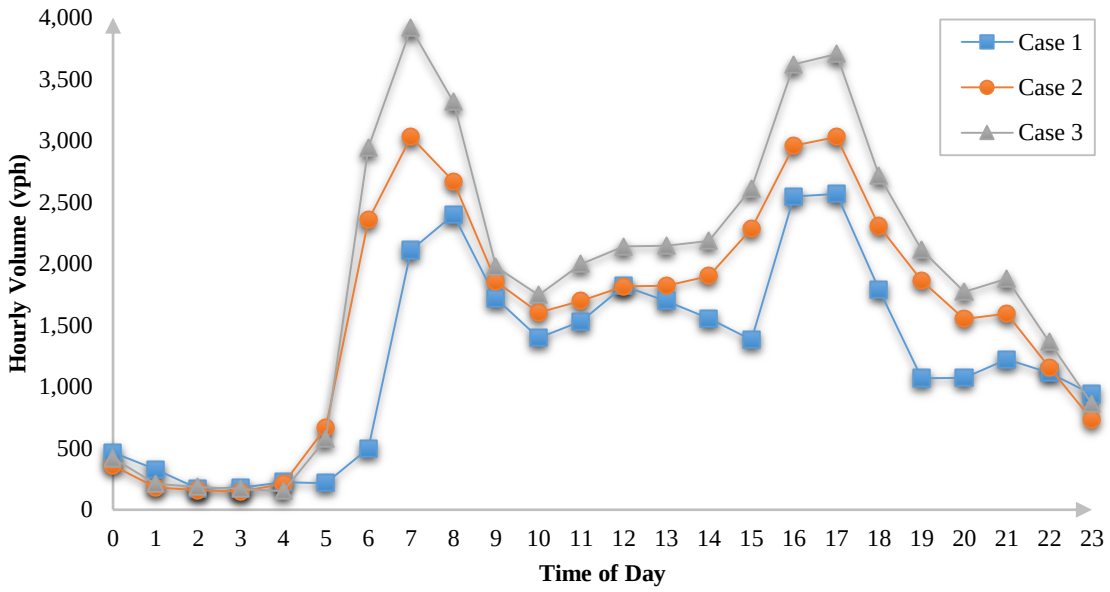


Figure 4.10 Hourly traffic distribution.

As summarized in Table 4.17, the delays with all the three models are compared to the "ground truth" information which is based on INRIX reported speeds. Note that the number in the parentheses represents the error percentage from predicted delay against ground truth delay, which indicates model performance in terms of prediction accuracy. As RUCM does not furnish the calculation details regarding work zones with shoulder closures on freeways, ANN-SVM is compared with ANN-HCM for Case 3. Apparently ANN-SVM outperforms both RUCM and ANN-HCM. Because ANN-SVM takes approaching traffic volume and speed variations into consideration, it is more applicable than other two models. The assumption of no queue under non-congested condition is a possible reason why the delays predicted by RUCM are underestimated for Cases 1 and 2.

In Table 4.17, delay cost is computed using Eq. 3.5. It is also worth noting that for Case 1, the error percentage differences of three models seem minor because of low traffic

volumes during nighttime. When work zones are placed in daytime with higher traffic volumes (i.e., Cases 2 and 3), ANN-SVM becomes very effective and outperforms other two models. In addition, the maximum queue lengths (approximated using Eq. 3.6) with the three models as well as the ground truth data are illustrated.

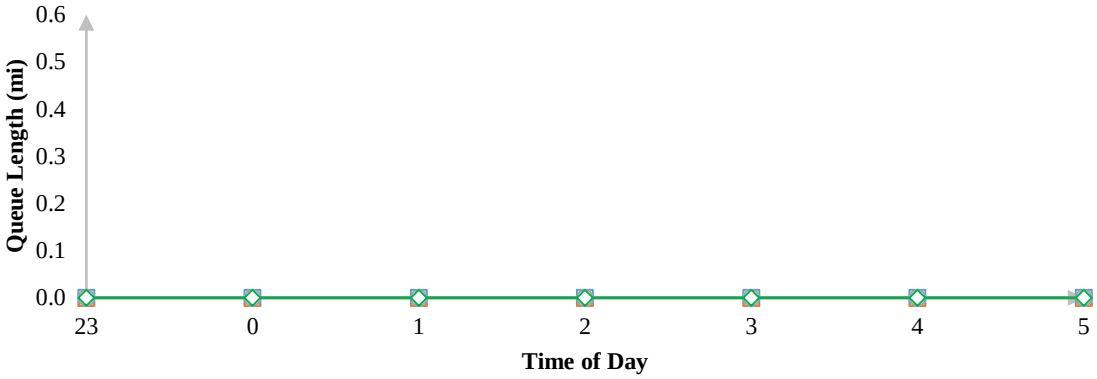
Table 4.17 Model Results Comparison

	Models	**Case 1**	**Case 2**	**Case 3**
Delay [a] (Error %)	RUCM	0 (100%)	0 (100%)	N/A
	ANN-HCM	62 (17%)	70 (6%)	72 (14%)
	ANN-SVM	59 (11%)	63 (5%)	81 (4%)
	Ground Truth [d]	53	66	84
Delay Cost [b]	RUCM	0	0	N/A
	ANN-HCM	1,350	1,524	1,562
	ANN-SVM	1,284	1,371	1,757
	Ground Truth [d]	1,153	1,437	1,822
Maximum Queue Length [c]	RUCM	0	0	N/A
	ANN-HCM	0	0.2	0.6
	ANN-SVM	0	0.2	0.6
	Ground Truth [d]	0	0.2	0.6

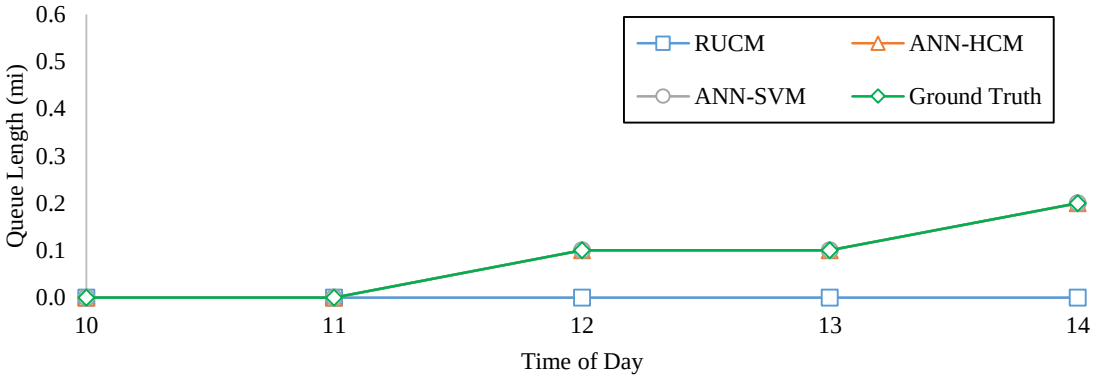
Note: [a] Delay: veh-hr; [b] Delay cost: $; [c] Queue length: miles; [d] INRIX speeds.

Figure 4.11 illustrates the variation of the queue lengths over time predicted by all the three models using Eq. 3.6, which are used to compare with the ground truth queue length. It is found that all these models performed well in Case 1 because of low traffic volumes during nighttime. However, for the daytime work zone with higher volumes (i.e.,

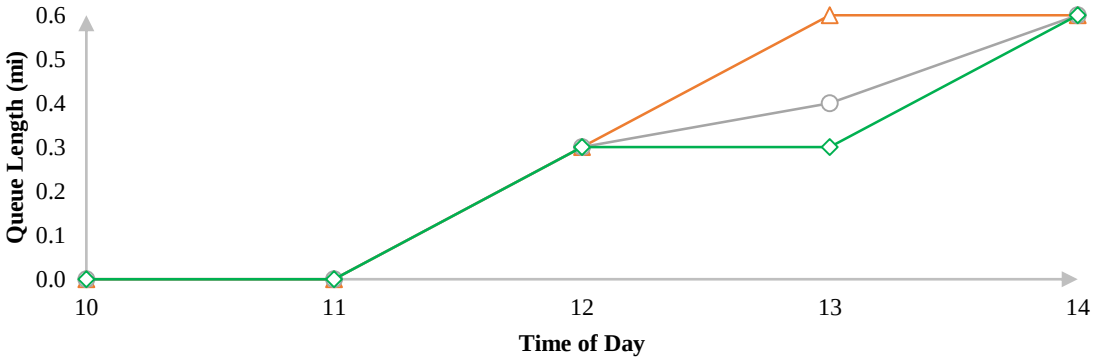
Cases 2 and 3), the queue length predicted by ANN-SVM is more accurate that other two models.



(a) Case 1



(b) Case 2



(c) Case 3

Figure 4.11 Temporal queue length distribution.

3.7 Summary

In this chapter, two models are developed for work zone delay prediction. The first model, the MNR model, is a non-linear model to capture spatio-temporal speed changes when non-recurrent congestion occurs caused by work zone activity. The prediction accuracy of the MNR model is acceptable as illustrated in Section 4.5. Regarding ANN-SVM, the evaluation results indicate that it is a better approach for work zone delay prediction because it can improve the accuracy of prediction results comparing to other models (i.e., MNR, RUCM, and ANN-HCM). The proposed ANN-SVM can predict the work zone impacts (i.e., delay, delay cost, and queue length) for the future work zone reasonably well when the traffic pattern is similar to the profile of the training data. The proposed ANN-SVM will be applied to various applications in the next chapter.

CHAPTER 5

MODEL APPLICATIONS

As discussed in Chapter 1, the objective of this study is to develop a sound spatio-temporal freeway work zone delay prediction model with big data under various road geometric and work zone conditions. Two freeway work zone delay prediction models (i.e., MNR and ANN models) have been developed in Chapter 3 and evaluated in Chapter 4. Comparing to RUCM, MNR, and ANN-HCM, ANN-SVM had demonstrated its performance in terms of prediction accuracy under various lane configuration and time of day.

In this chapter, the potential applications of ANN-SVM to support work zone planning and analysis on freeways are discussed. By employing ANN-SVM, a work zone delay prediction tool is developed in Section 5.1. Then, Case 2 presented in Section 4.6 is applied here for determining optimal the start time of a work zone that yields the least delay as well as cost in Section 5.2. Finally, ANN-SVM is applied to calculate the contractor penalty in terms of cost overruns as well as an incentive reward schedule in case of early work competition as shown in Section 5.3.

5.1 Work Zone Impact Analysis

By incorporating ANN-SVM, a work zone delay prediction tool (WZDPT) can be developed to post information graphically, which can aid transportation agencies to make proper decisions by assessing work zone activities in order to minimize disruptions to the traveling public. Depending upon the user inputs such as route, milepost range and direction, WZDPT can quickly locate the expected work zone on the map and apply ANN-SVM for work zone impact analysis. This further enhances the ease of use of

WZDPT, as users would not require any pre-requisite knowledge regarding road geometry condition for analysis.

A historical work zone on the Interstate Highway 80 (I-80) - one of the most congested and busiest highways in New Jersey - is selected for demonstrating the application of ANN-SVM. One out of three lanes was closed for repairs on I-80 eastbound between mileposts 34.0 and 34.5 from 9 AM to 3 PM on October 14, 2014 as shown in Figure 5.1. The traffic volumes are obtained from NJCMS (2012) as illustrated in Figure 5.2, which consists of an average 7% of heavy vehicles.

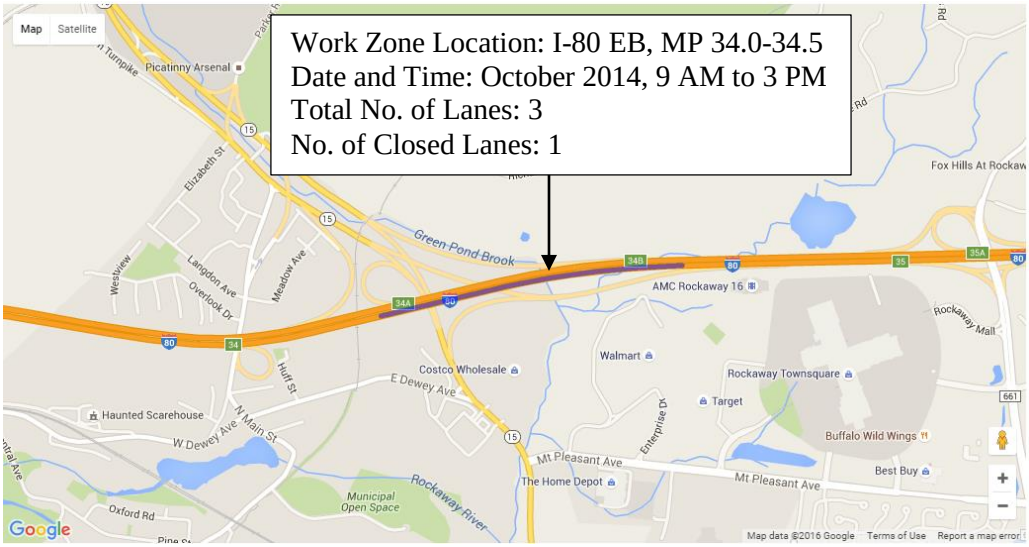


Figure 5.1 Work zone on I-80 in Wharton, NJ.

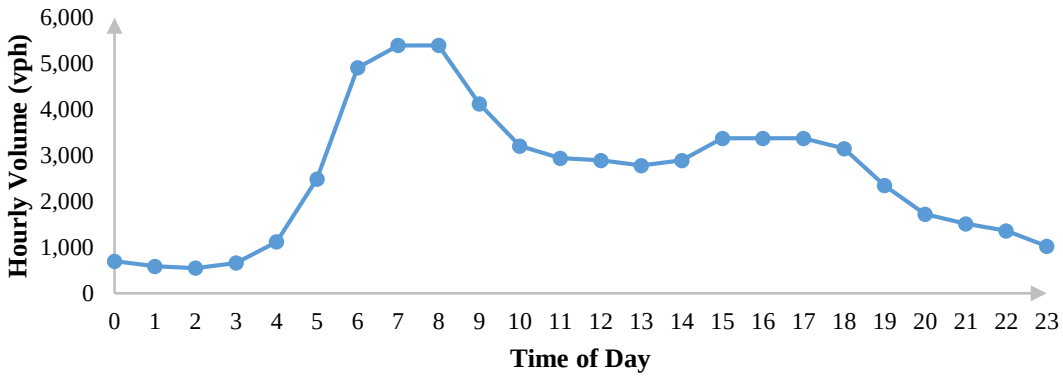


Figure 5.2 Hourly traffic distribution at MP 33.79 on I-80 EB.

Based on the work zone information discussed above, WZDPT will retrieve the roadway geometry information (e.g., number of lanes at work zone location) from the NJSLD DB and normal speeds in upstream of work zone from INRIX. Then WZDPT can quickly generate the normal speed (part a) and the predicted speed (part b) heat maps of I-80 eastbound work zone as shown in Figure 5.3. This enables user to compare spatio-temporal speed changes side-by-side and better assess the impact of the proposed reconstruction project.

By using Eq. 4.5, Figure 5.4 illustrates the predicted impacts of the 6-hour work zone on I-80 EB versus different lane closures (i.e., shoulder, 1-lane, and 2-lane) and work zone starting times (i.e., 3 AM, 9 AM, 3 PM, and 9 PM). The normal speeds and predicted work zone speeds are illustrated horizontally with respect to the number of lane closures and vertically with respect to the starting time of the work zone. The predicted work zone delays consistently increase as number of closed lanes increases, especially during peak periods. In addition, work zone delay impact is greater in the peak period than in the off-peak period (comparing heat maps in Rows 1 and 4. Moreover, the speed recovers slowly as the work zone end time approaches the peak period (compare heat maps in Row 1). The work zone delay prediction tool shows the capability of creating richer and more complete picture of what is happening on the road, which can be used as a viable alternative for transportation engineers to analyze information efficiently and make proper delay mitigation strategies.

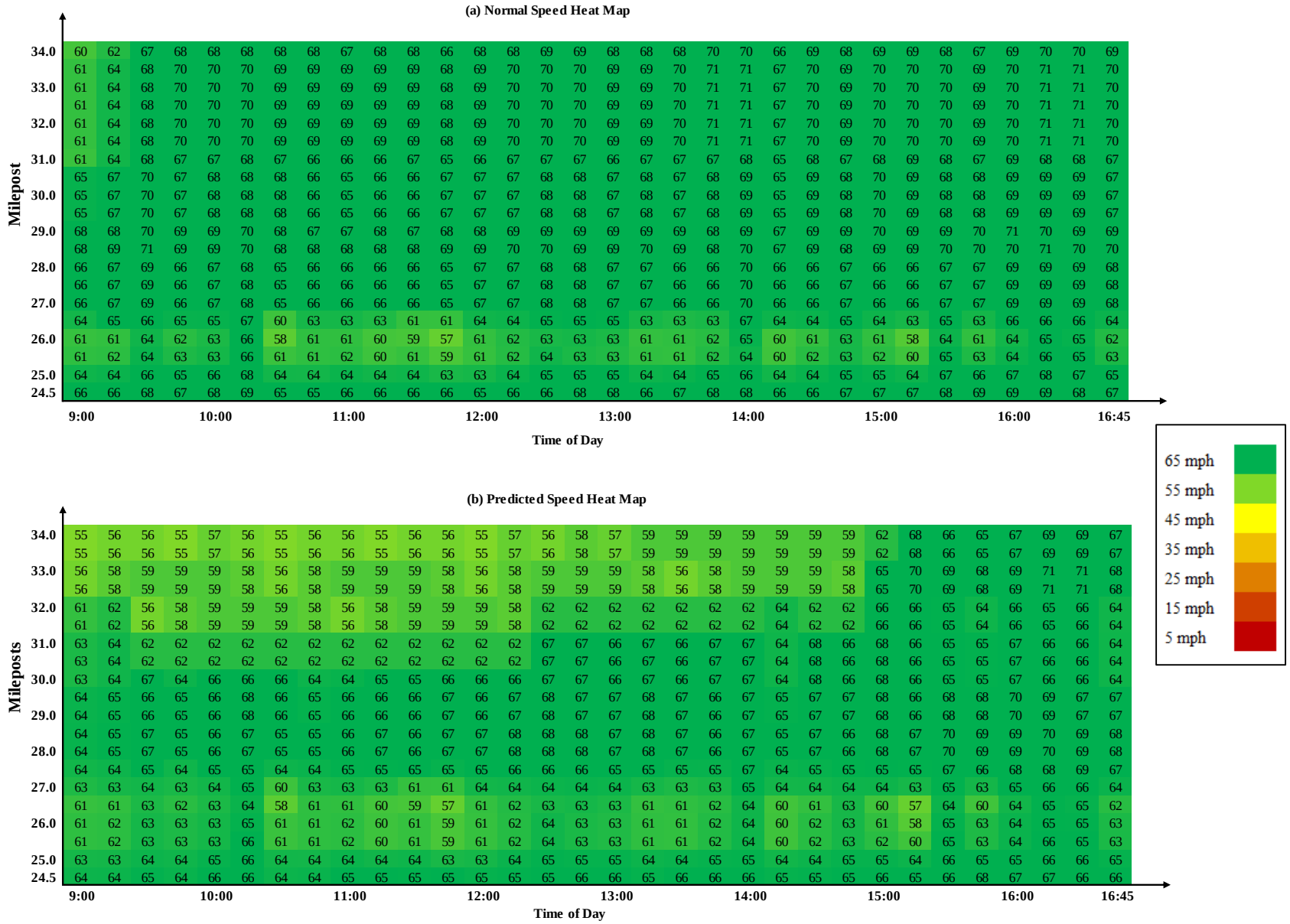


Figure 5.3 Comparison of predicted and actual speeds of the I-80 EB work zone site.

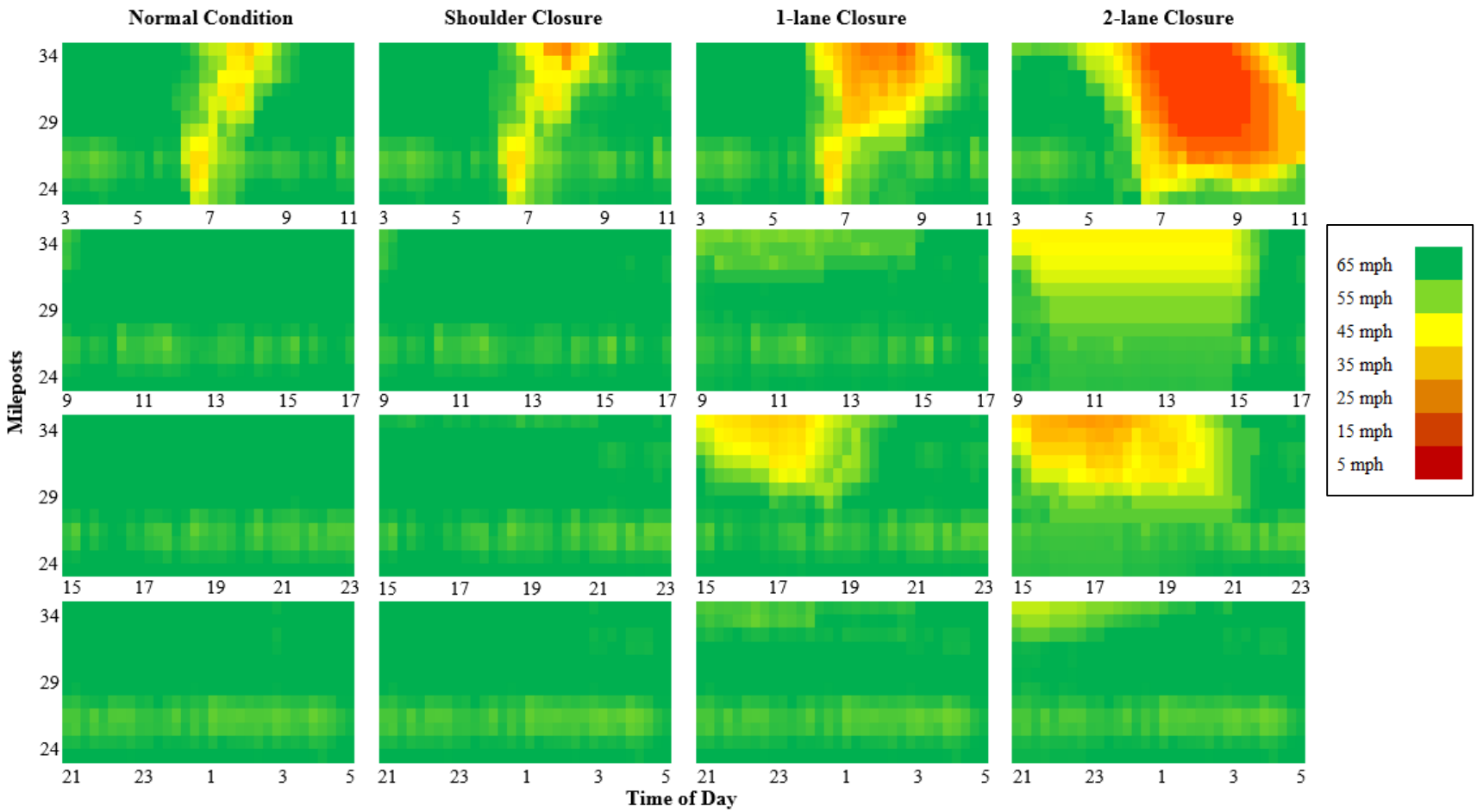


Figure 5.4 Comparison of predicted speeds with different work zone starting times and lane closure configurations.

In addition to examining the work zone impact prediction results, users can view the hourly volume distribution approaching the work zone obtained from NJCMS DB as shown in Figure 5.5, which allows users to examine the volume changes over space and time. If the traffic counts of a study work zone site are different from those that NJCMS summarized in the table, a user-specified parameter (in percentages) is offered to adjust the volumes.

Review Volume

Increase by Decrease by 0 Percent Apply Changes

ROUTE	SRI_CMS	BEGINMP	ENDMP	VOL 8:00 AM	VOL 9:00 AM	VOL 10:00 AM	VOL 11:00 AM	VOL 12:00 PM	VOL 1:00 PM	VOL 2:00 PM	VOL 3:00 PM	VOL 4:00 PM	VOL 5:00 PM
I-80	00000080__	33.58	34.02	5399	4128	3206	2944	2893	2784	2893	3375	3376	3375
I-80	00000080__	30.61	33.58	5176	4217	3275	3008	2956	2844	2957	3175	3176	3176
I-80	00000080__	28.91	30.61	4382	2888	2271	2239	2199	2118	2173	2352	2667	2856
I-80	00000080__	27.01	28.91	4701	2713	2163	2296	2256	2171	2177	2332	2869	3059
I-80	00000080__	26.25	27.01	4484	2815	2224	2252	2213	2130	2195	2378	2663	2852
I-80	00000080__	25.50	26.25	3550	1903	1529	1692	1665	1603	1665	1847	2003	2049
I-80	00000080__	25.04	25.50	3317	2212	1737	1702	1673	1610	1673	1773	1773	1773
I-80	00000080__	24.77	25.04	3317	2212	1737	1702	1673	1610	1673	1773	1773	1773
I-80	00000080__	24.36	24.77	3317	2212	1737	1702	1673	1610	1673	1773	1773	1773
I-80	00000080__	23.23	24.36	3317	2212	1737	1702	1673	1610	1673	1773	1773	1773

Figure 5.5 Hourly traffic volumes.

After reviewing traffic volume counts, users may select one of the three criteria below to determine the queue:

Criterion 1: 75% of historic average speed – The status of queue is positive at a segment whose speed falls below 75% of the historic average speed. The historic average speed is specific to the time of a day and the day of a week for each segment, and is calculated based on the speeds collected in 2014. More detailed information can be found in Chapter 3.

Criterion 2: 75% of historic average speed or LOS (Level of Service) D Speed – The status of queue is positive at a segment whose speed falls below 75% of the historic average speed or LOS D speed (i.e., 35 mph).

Criterion 3: Historic average speed – The status of queue is positive at a segment whose speed falls below the historic average speed. This measure will show predicted queue over space and time that is "worse than normal." Users are also able to enter an "offset" into this option.

For the 6-hour work zone conducted at 3AM with one lane closure on I-80 EB MP 34 - 34.5 (see speed heat map in Column 3 and Row 1 in Figure 5.4), Figure 5.6 shows the queue length distribution over time by using three criteria listed above. The work zone delay prediction tool provides user with flexibility in determining work zone impacts based on preferences and needs. Note that the queue by using Criterion 3 is determined for any time when speeds are 5 mph lower than normal speed. It is found that the queue length defined by Criterion 3 is longer than those defined by Criteria 1 and 2. The reason for this is that due to lane closure required by the planned work zone, the speed drops quickly as the traffic volume increases.

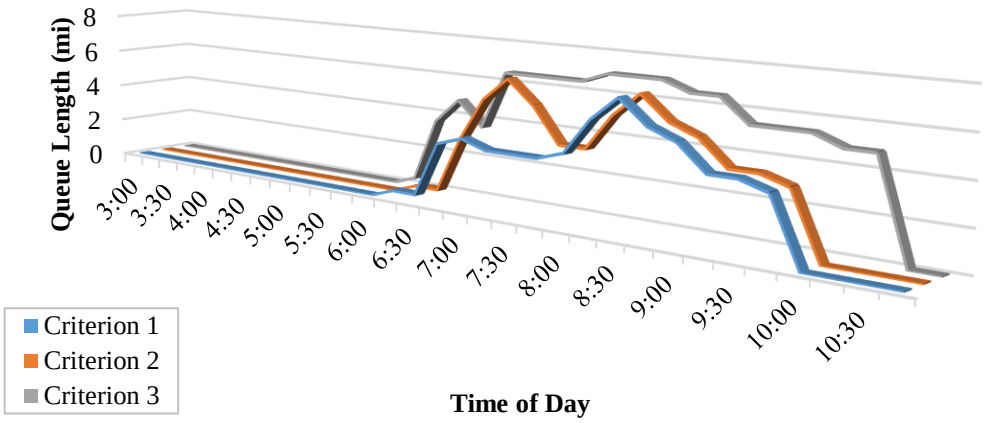


Figure 5.6 Temporal queue length distributions.

Furthermore, WZDPT allows the user to generate a report of lane closure impacts based on a default template. The report contains all the necessary information for the roadway segment of interest as well as the predicted delay, delay cost and queue length using Eqs. 3.5 through 3.7. For instance, a report generated for the lane closure of I-80 from milepost 34 to milepost 34.5 from 3:00 AM to 9:00 AM plus two hours after the work zone removed is illustrated in Figure 5.7. This report not only presents the impact of a proposed lane closure in a logical and concise manner, it also assists agencies and contractors in preparing project documentation. It is noted that the volume showed in the analysis report is the hourly volume approaching the work zone obtained from NJCMS DB.

Work Zone Impact Analysis Report

Work Zone Information

Route Name:	I-80	**Number of Closed Lanes:**	1
Start Milepost:	34.0	**Expected Start Date:**	10/14/2014
End Milepost:	34.5	**Expected Start Time:**	03:00
Direction:	Eastbound	**Expected End Date:**	10/14/2014
Number of Lanes:	3	**Expected End Time:**	09:00

Value of Time Parameters

Value of Passenger Car Time ($/veh-hr)*:	18.15	**Value of Truck Volume ($/veh-hr)*:**	30.25

Predicted Work Zone Impact

Total Queue Delay Cost ($):	12,761	**Total Queue Delay (veh-hr):**	659
Queue Determined by:	75% of historic average speed		

Predicted Hourly Impact

Date and Time	Approaching Car Volume (veh)*	Approaching Truck Volume (veh)*	Queue Delay (veh-hr)	Queue Delay Cost ($)	Maximum Queue Length (mi)
10/14/2014 03:00	601	59	0	0	0
10/14/2014 04:00	1,032	88	0	0	0
10/14/2014 05:00	2,280	209	0	0	0
10/14/2014 06:00	4,472	439	34	660	3.2
10/14/2014 07:00	4,882	517	192	3,724	3.6
10/14/2014 08:00	4,899	500	311	6,020	6.5
10/14/2014 09:00	3,757	371	122	2,357	4.8
10/14/2014 10:00	2,828	378	0	0	0
Grand Total			659	12,761	

Note: * Source: 2012 New Jersey Congestion Management System (NJCMS).

Figure 5.7 Work zone mobility impact report.

5.2 Work Zone Schedule Optimization

In this section, ANN-SVM is evaluated with the work zone in Case 2 (see Section 4.6) under various starting times and durations. Figure 5.8 shows the variation of delay cost versus start time for various work zone durations. Considering a 5-hour work zone, it is found that the most cost-effective starting time would be 12 AM. If this work zone must be performed during the daytime (i.e., between 6 AM and 6 PM), the suggested starting time would be 10 AM. It is also found that when the 5-hour work zone ends close to or at peak hours, the residual queue must wait for extra time to be cleared, which results in more delay and cost. As the duration is greater than 7 hours, the delay cost reaches the minimum at 10 PM because of light traffic volumes between 10 PM and 5 AM.

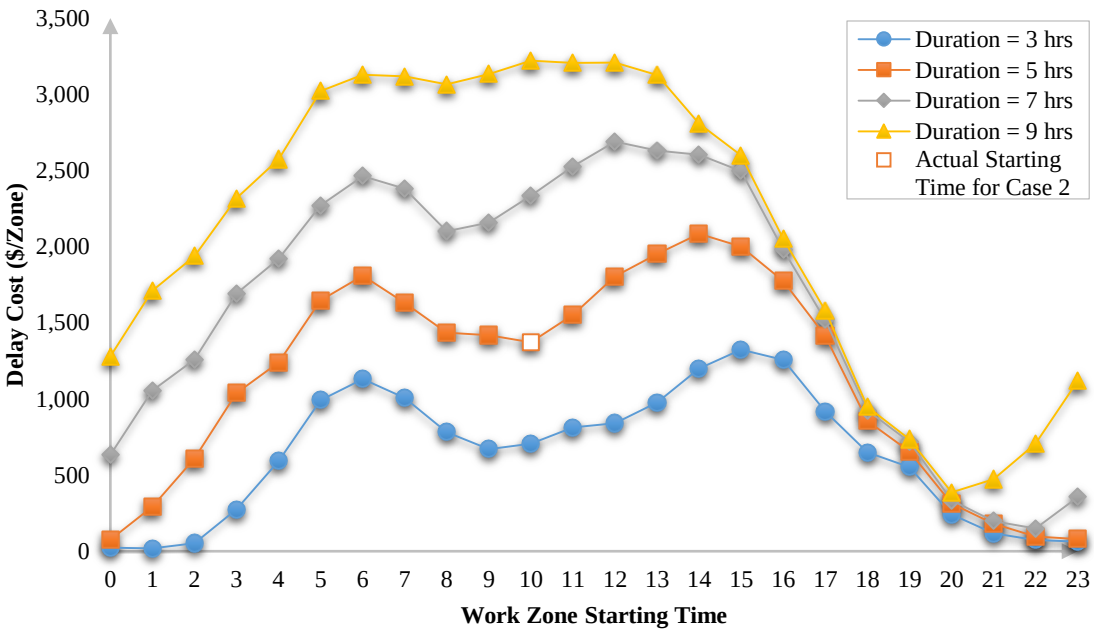


Figure 5.8 Delay cost vs. starting time for various work zone durations (Case 2).

Figure 5.8 also indicates that a work zone performed in the daytime with longer duration would raise the delay cost, especially if the work zone schedule crosses peak

hours. In general, low delay cost may be expected as the work zone is performed during the nighttime, albeit the labor cost is expected to be high. This also explains the work zone practices often seen in daily commutes.

Figure 5.9 illustrates and explores the relationship between delay cost and start time for various demand levels, varying from 80% to 150% of the original volume in Case 2. It is found that the delay costs are close and relatively low for the start time beginning with 11 PM or later until 3 AM (next day) because the traffic during the corresponding work zone time period is light. The delay cost significantly increases if the work zone duration crosses peak hours. The results would give transportation agencies a competitive edge by examining the delay costs versus work zone start and end times subject to different traffic distributions over space and time.

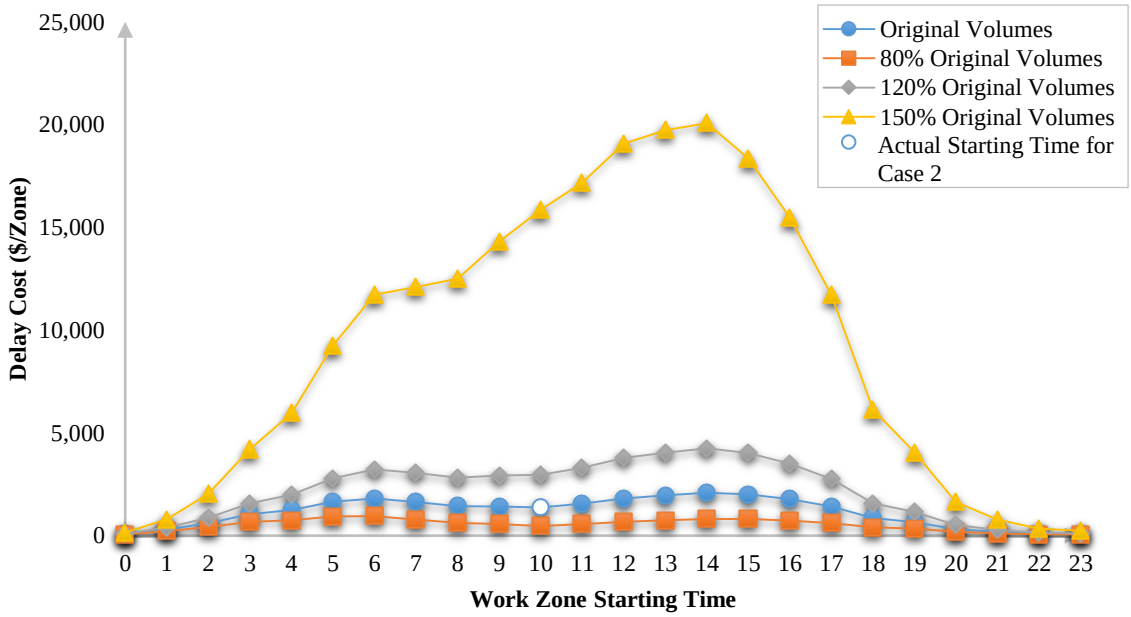


Figure 5.9 Delay cost vs. starting time for various traffic multipliers (Case 2).

5.3 Lane Rental Charge Determination

A FHWA report (FHWA, 2011) defines lane rental fee as a daily-base or hourly-base charge for the time period a lane is closed to through traffic for construction activities. This provision is intended to minimize the disruption of the work zone traffic and to encourage minimal use of lanes for construction activities. The delay costs for various starting times and durations, as shown in Figure 5.8, can be used as a guideline to form the basis for awarding or deducting payments to contractors for early and late project completions, respectively. For example, in Case 2, assuming that the contractor delays two hours to open the closed lane to traffic (i.e., takes seven hours instead of five hours to complete the work). If work zone started at 10 AM, the transportation agency could charge $964 in penalties to the contractor for late completion because of the cost incurred by the excess delay. Note that this charge may vary depending on the traffic volume distribution, work zone starting time, and duration of late work completion of the study site as shown in Figure 5.10.

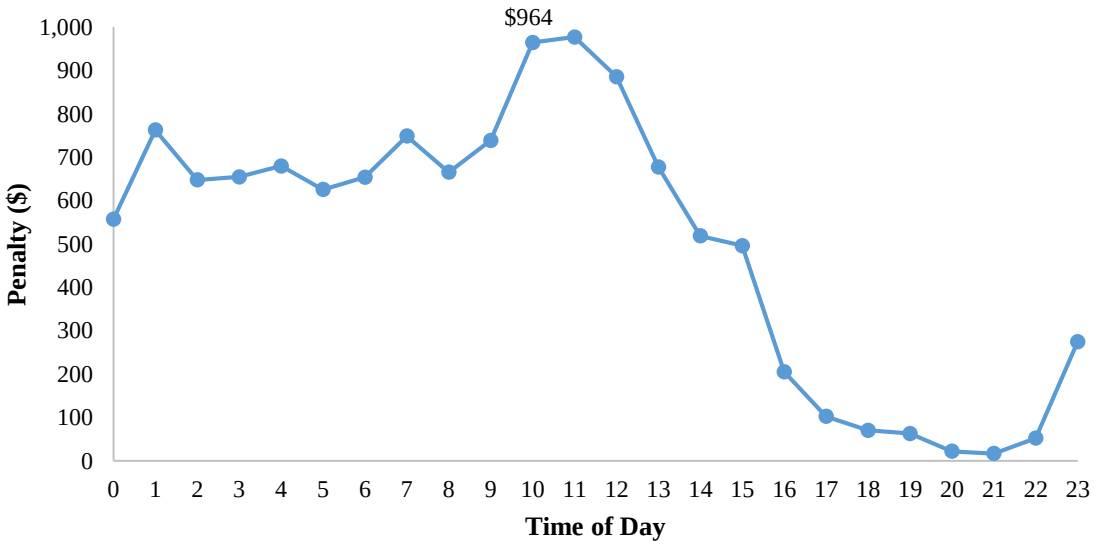


Figure 5.10 Penalty vs. starting time for 2-hour delayed completion (Case 2).

CHAPTER 6

CONCLUSIONS AND FUTURE RESEARCH

With increasing roadwork activities that are necessary to rehabilitate and revitalize the roadways in the United States, planning lane closures for roadwork has drastically demanded more accurate predictions on the impact of lane closures. It is crucial to be able to precisely predict the lane closure impacts to minimize both the cost and traffic congestion induced by roadwork. In response to this challenge, two models, the MNR and ANN models, for quantifying work zone delay were developed using big data in this research. In the MNR model, the work zone capacity was predicted using reduction factors based on historical work zones in years 2013 and 2014. While in the ANN model, the work zone capacity was approximated using the SVM model (called ANN-SVM). Subject to the limitation of work zone related traffic information collected from the field for peak period, a calibrated and validated simulation model was developed using VISSIM to generate traffic data for model development. The performance of each model was analyzed.

Then the proposed ANN-SVM model was embedded into a work zone delay prediction tool, which can be used to support state and local traffic construction, operations, planning staff, and construction contractors to:

- Quantify and display temporal-spatial corridor speed/delay predictions resulting from capacity decreases in work zones on New Jersey freeways and arterials.
- Identify delay impacts of alternative project phasing plans.
- Conduct tradeoff analyses between construction costs and delay costs.
- Examine the impacts of construction staging by location, time of day (peak versus off-peak), and season (summer versus winter).

- Assess travel demand measures and other delay mitigation strategies.
- Help establish work completion incentives.

For example, ANN-SVM could be used to calculate the costs of conducting work at night instead of during the day, to change the starting and ending times, and to compare the impact of several time schedules on traffic flow conditions, or to divert the traffic to one road versus another road during different phases of construction. The costs, traffic delays, and potential backups can be predicted for both an average day of work and for the whole life cycle of construction. This model can also analyze the advantages of various strategies for minimizing the projected traffic delays. These mitigation strategies might include the retiming of signals on detour routes to help traffic flow more smoothly, planning a media campaign to publicize the planned work zones, or using traveler information systems that allow drivers to plan ahead and choose other routes if possible.

6.1 Conclusions

While developing the work zone capacity and delay prediction models, a wealth of insights, challenges, areas of potential improvements, and opportunities available to agencies in the areas of work zone impact assessment, data collection, and performance measurement were identified, all of which are summarized below.

6.1.1 Spatio-temporal Work Zone Delay Prediction

In this study, an ANN-SVM was developed using big data to quantify delays incurred by work zones on New Jersey freeways, in which the restricted capacity (or called work zone capacity) was approximated using SVM. ANN-SVM was designed to adapt to the relationship of speed versus the ratio of approaching traffic volume to work zone capacity,

which has proven to be a robust work zone delay prediction model and achieves reasonable well prediction accuracy. The performance of ANN-SVM outperforms that with RUCM and ANN-HCM in predicting delay, delay cost, and queue length.

A work zone delay prediction tool integrated with ANN-SVM was developed to post information graphically, which can aid transportation agencies to make proper decisions by assessing work zone activities in order to minimize disruptions to the traveling public. It is worth noting that this easy-to-use and easy-to-learn tool does not require users to set various adjustment factors based on practical experience. It is very convenient for practitioners to assess the impact of work zones and determine the optimal work zone schedule which can yield the least delay and cost. Based on the predicted spatio-temporal speeds affected by an expected work zone, a proper traffic management plan (i.e., locations of changeable message signs, variable speed limits, and traffic detour management, etc.) may be prepared accordingly. ANN-SVM can assist work zone planners in designing optimal start and end time of work zone as function of time of day. In addition, it can be used to calculate contractor penalty in terms of cost overruns as well as incentive reward schedule in case of early work competition.

6.1.2 Big Data Analytics in Work Zone Impact Analysis

With technological advancement, the transportation industry has been experiencing a wide variety of unprecedented massive traffic data obtained from different sources, such as infrastructure sensors, mobile devices, and floating cars. This new and rich data (big data) needs to be managed, communicated, interpreted, aggregated, and analyzed in a reliable and efficient way. However, use of conventional data management tools is not able to uncover hidden patterns, correlations, and other insights, which would leave the huge

amount of traffic data underutilized. Therefore, big data analytics, which creates richer and more complete picture of what's happening on the road, becomes a viable alternative for transportation engineers to analyze information efficiently and make decisions based on what they've learned.

For the freeway work zone impact analysis, leveraging big data analytics and advanced freeway work zone delay prediction methods (e.g., ANN models) with big data, the accuracy of predicted work zone speed and delay can be then significantly improved, rather than predicting delay using traditional deterministic queuing method with the data captured by loop detectors. The ability of big data analytics to work faster and stay agile gives transportation agencies a competitive edge they did not have before. In addition, it would help transportation agencies improve work zone operations, reduce delay costs and better serve motorists.

6.1.3 Work Zone Data Deficiencies

The major issues founded during data processing procedures are as follows:

- Although the length of a work zone and the corresponding starting/ending times are initially set by NJDOT, this information is finalized by the contractor who demarcates the work zone. OpenReach DB needs to be updated based on the contractor's finalized work zone schedule.

- The traffic counts information at the scenes of work zones are important measures for predicting speed and delay, which is not available at most places. The hourly traffic volumes recorded in NJCMS DB are thus used for model development.

- The OpenReach and INRIX DBs do not include the SRI information. In addition, INRIX DB also lacks the mileposts of TMCs. This problem has been fixed manually in this study. This issue will occur as new TMCs on New Jersey freeways are defined.

6.2 Future Research

Future research to enhance ANN-SVM in prediction spatio-temporal work zone delay and its applications shall focus on the following aspects:

The actual traffic counts for peak period at the scenes of work zones should be collected from the field to replace the simulated data for further validation of the developed models. While using real world traffic counts, the sample size should be chosen in a way that assure that the collected data can reflect the actual work zone impacts on traffic flows under various lane configurations and work zone conditions. More accurate traffic counts information will substantially improve the reliability of the developed models and produce more accurate results regarding the upstream speed, queue delay and cost. Such extensions will allow the transportation engineers to identify the optimal start and end times of each work zone, which will further improve the traffic flow operation of each facility.

It is desirable to develop a self-updating database by gathering data from various sources in an automated manner wherever feasible. Modifying and standardizing the existing database with the inclusion of common fields of information, in order to facilitate effective communication between sources that would reduce the time required for manual processing and improve productivity.

Traffic Message Channels (TMCs) can play a key role in collecting mobility and safety data, identifying issues that arise, and providing information to the public regarding current work zones within its surveillance zone. INRIX has re-defined the length of the TMCs, which are now smaller. The performance of the proposed model in this study can be elevated if it utilizes these smaller TMCs, as it will more accurately predict the speed and queue length for each time interval.

The work zone capacity predicted from SVM can be applied in the MNR model to improve the prediction accuracy. In addition, the proposed model in this study can be further extended to include the network impact of a work zone. Such an expanded model may have functions including: (a) a network-wide work zone impacts prediction module; (b) an optimal work zone schedule module; and (c) a work zone optimal staging module.

APPENDIX A

OPENREACH DATA DEFINITION

In this appendix, the OpenReach data fields are identified in the list below.

Field	**Description**	**Data Stream Example**
EVENTID*	Event Identification	45675101
Facility Name*	Route Name	NJ 3
Created At Date Time*	Incident Start Date and Time	2/1/13 22:00
Closed At Date Time*	Incident End Date and Time	2/2/13 7:31
Event Type	Incident Type	Construction
Event Description	Description of the Incident	NJ DOT - STMC: Construction, construction on NJ 3 both directions between US 46 (Clifton) and West of CR 509/Broad St (Clifton) right lane closed until 7:00 A.M.
City From Name	The city at the start of the incident	Clifton
County From Name	The county at the start of the incident	Passaic
State From Name	The state at the start of the incident	NEW JERSEY
City To Name	The city at the end of the incident	Clifton
County To Name	The county at the end of the incident	Passaic
State To Name	The state at the end of the incident	NEW JERSEY
From Mile Marker*	Incident Starting Milepost	3.8
To Mile Marker*	Incident Ending Milepost	4.9
Final Duration	The Duration of the Incident	570
Latitude	The Latitude of the Incident	40.83257731
Longitude	The Longitude of the Incident	-74.14454447

*: Fields selected for database development.

APPENDIX B

A SAMPLE QUERY FOR DATA PROCESSING

Presented below is a sample SQL query used in the database development of this study:

```
CREATE NONCLUSTERED INDEX [day_week] ON
[dbo].[Interstate_Highway_Feb_2014] ([day_week] ASC) WITH (PAD_INDEX  = OFF,
STATISTICS_NORECOMPUTE  = OFF, SORT_IN_TEMPDB = OFF,
IGNORE_DUP_KEY = OFF, DROP_EXISTING = OFF, ONLINE = OFF,
ALLOW_ROW_LOCKS  = ON, ALLOW_PAGE_LOCKS  = ON) ON [PRIMARY]
go
================================================================
update [Interstate_Highway_Feb_2014] set time_range_fk_id= time_range.time_range_id
from time_range
where CONVERT(time, [measurement_tstamp], 102)  between min_interval and
max_interval
go
================================================================
update [Interstate_Highway_Feb_2014] set  dw=0 where (day_week=1 or day_week=7)
go
================================================================
update [Interstate_Highway_Feb_2014] set  dw=1 where dw is null
go
================================================================
CREATE CLUSTERED INDEX [ix_cluster3] ON [dbo].[Interstate_Highway_Feb_2014]
([tmc_code] ASC, [time_range_fk_id] ASC, [dw] ASC) WITH (PAD_INDEX  = OFF,
STATISTICS_NORECOMPUTE  = OFF, SORT_IN_TEMPDB = OFF,
IGNORE_DUP_KEY = OFF, DROP_EXISTING = OFF, ONLINE = OFF,
ALLOW_ROW_LOCKS  = ON, ALLOW_PAGE_LOCKS  = ON) ON [PRIMARY]
GO
================================================================
SELECT [tmc_code], [time_range_fk_id], COUNT(dw)as max_len
into   Interstate_Highway_Feb_2014_Maxrecords
FROM [Interstate_Highway_Feb_2014]
where [dw]=0
group by [tmc_code], [time_range_fk_id]
go
================================================================
SELECT [tmc_code], [time_range_fk_id], COUNT(dw) as max_len
into   US_Highway_feb_2014_wd_maxrecords
FROM [Interstate_Highway_Feb_2014]
where [dw]=1
group by [tmc_code], [time_range_fk_id]
```

```
go
==================================================================
SELECT *, ROW_NUMBER() OVER(PARTITION BY tmc_code, time_range_fk_id
ORDER BY speed ) AS "Row Number"
into Processed_Interstate_Highway_Feb_2014
FROM [Interstate_Highway_Feb_2014]
where dw=0
go
==================================================================
SELECT *, ROW_NUMBER() OVER(PARTITION BY tmc_code, time_range_fk_id
ORDER BY speed ) AS "Row Number"
into Processed_Interstate_Highway_Feb_2014
FROM [Interstate_Highway_Feb_2014]
where dw=1
go
==================================================================
update [Processed_Interstate_Highway_Feb_2014] set
[Processed_Interstate_Highway_Feb_2014].max_len=agg.max_len
from Interstate_Highway_Feb_2014_Maxrecords agg WITH (NOLOCK)
where [Processed_Interstate_Highway_Feb_2014].tmc_code = agg.tmc_code and
        [Processed_Interstate_Highway_Feb_2014].time_range_fk_id =
        agg.time_range_fk_id and
        [Processed_Interstate_Highway_Feb_2014].[dw]=0
go
==================================================================
update Processed_Interstate_Highway_Feb_2014  set percentile=round(CAST([Row
Number] AS float)/ CAST([max_len] AS float),6)
go
update Processed_Interstate_Highway_Feb_2014  set percentile=round(CAST([Row
Number] AS float)/ CAST([max_len] AS float),6)
go
==================================================================
select tmc_code, time_range_fk_id, avg(speed) as avg_speed, stdev(speed) as
stdev_speed, max(speed) as max_speed, min(speed) as min_speed, count(speed) as
count_speed
into Interstate_Highway_Feb_2014_Output
from   Processed_Interstate_Highway_Feb_2014 y
where  ([percentile]>=0.05 and [percentile]<=0.95)
group by [tmc_code], time_range_fk_id
go
==================================================================
SELECT [tmc_code], [time_range_fk_id], [min_interval], [max_interval], [avg_speed],
[stdev_speed], [max_speed], [min_speed], [count_speed]
FROM Interstate_Highway_Feb_2014_Output INNER JOIN [time_range] ON
[time_range_id]=[time_range_fk_id]
go
```

APPENDIX C

NJDOT RUCM APPROACH

The detailed procedure of NJDOT RUCM approach predicting work zone delay and cost is presented in this Appendix. Before conducting the computation, certain important criteria and assumptions must be identified:

- Average user cost per car hour is $18.15/veh-hr.
- Average user cost per truck hour is $30.25/veh-hr.
- The work zone speed is generally 10mph -15mph less than the unrestricted speed. The unrestricted speed is generally assumed the posted speed limit of the section operating in an unrestricted flow condition. Following this, the unrestricted speed of the studied segment as 65 mph; hence, the work zone speed is assumed as 50 mph.

Take Case 1 of Section 4.6 as an example, the selected section of the I-78 WB mainline is comprised of three lanes. The closure of two lanes was required for carrying out work zone operations, and all traffic operations were supported by the remaining one open lane. Capacity of the roadway in both normal and work zone scenarios are given in the NJDOT RUCM (NJDOT, 2015) as illustrated in Table C.1.

Table C.1 Traffic Capacities

Facility Type	Ideal Capacity
Freeway - 4 lanes	2,200 passenger cars per hour per lane
Freeway - 6 or more lanes	2,300 passenger cars per hour per lane
Multilane highway	2,200 passenger cars per hour per lane
Two-lane highway	1,400 passenger cars per hour per lane*
Signalized Intersection	1,900 passenger cars per hour of green per lane

*: For 50/50 volume, split by direction.

Work zone road capacity counted in vehicle/lane/hour is taken as the number of lanes multiplied by the capacity provided in Table C.2. With one lane closure on a 3-lane freeway, the work zone capacity is 1,200 vph. Table C.3 depicts the calculation procedure suggested by the NJDOT RUCM.

Table C.2 Measured Work Zone Capacity - Freeway Section

Number of Direction Lanes		Number of Studies	Average Capacity		Recommended Value (*) veh/lane/hour
Normal	Open		Vehicle per hour	Vehicle per lane per hour	
3	1	7	1,170	1,170	1,200
2	1	8	1,340	1,340	1,300
5	2	8	2,740	1,370	1,400
4	2	4	2,960	1,480	1,500
3	2	9	2,980	1,490	1,500
4	3	4	4,560	1,520	1,500

*: Values may be increased 100 veh/lane/hour when work zone is protected with Jersey barrier.

The queue rate is calculated as the difference between the hourly capacity of the facility and the unrestricted hourly demand during each hour of the day. The queuing rate is the hourly rate at which vehicles accumulate, or, if negative, dissipate from any queue that may exist. A physical queue develops when the queue rate is greater than zero. In this scenario, the approaching volume is too small compared to the capacity provided. Hence, either negative queue rates are obtained or no queue is formed.

Under unrestricted flow conditions, the number of vehicles that travel through the work zone is generally seen as the traffic demand on the facility during the hours when the work zone is in place. The total number of vehicles travelling through the work zone was 5,054 vph as shown in Table C.3. As shown in Table C.4, the added travel time caused by the work zone based on the NJDOT RUCM can be computed using the following formula:

$$t = \frac{d}{v_w} - \frac{d}{v_u} \tag{C.1}$$

where:

t = Added travel time (hr/veh);

d = Work zone length (mi);

v_w = Work zone speed (mph); and

v_u = Unrestricted speed (mph).

Table C.3 Analysis of the Work Zone

Work Zone:	I-78 WB MP 47.3 - 49.3			Normal Capacity:	6,900
Normal Speed (mph):	65	Percent Cars:	90	Work Zone Capacity:	1,200
Directional ADT:		Percent Truck:	10	Lanes Under Normal Operation:	3

3.1(A)	3.1(B)	3.1(C)	3.1(D)	3.1(E)	3.1(F)	3.1(G)	3.1(H)	3.1(I)	3.1(J)
Time Period (hour)	Hourly Traffic (%)	Vehicle Demand (vph)	Lanes Open (#)	Roadway Capacity (vph)	Queue Rate (vph)	Queued Vehicles (vph)	Work Zone Present? (Y or N)	Vehicles that Travel Work Zone (vph)	Vehicles that Travel Queue (vph)
12-1 AM	0.7	466	1	1,200	-734	0	Y	466	0
1-2	0.5	329	1	1,200	-871	0	Y	329	0
2-3	0.4	173	1	1,200	-1,027	0	Y	173	0
3-4	0.6	180	1	1,200	-1,020	0	Y	180	0
4-5	1.8	223	1	1,200	-977	0	Y	223	0
5-6	4.4	216	1	1,200	-984	0	Y	216	0
6-7	6.2	499	1	6,900	-6,401	0	N	0	0
7-8	7.2	2,108	1	6,900	-4,792	0	N	0	0
8-9	5.6	2,398	1	6,900	-4,502	0	N	0	0
9-10	5.0	1,717	1	6,900	-5,183	0	N	0	0
10-11	4.8	1,396	1	6,900	-5,504	0	N	0	0
11-12 PM	5.1	1,533	1	6,900	-5,367	0	N	0	0
12-1	5.3	1,817	1	6,900	-5,083	0	N	0	0
1-2	5.5	1,695	1	6,900	-5,205	0	N	0	0
2-3	5.6	1,555	1	6,900	-5,345	0	N	0	0
3-4	6.5	1,380	1	6,900	-5,520	0	N	0	0
4-5	6.9	2,547	1	6,900	-4,353	0	N	0	0
5-6	6.4	2,566	1	6,900	-4,334	0	N	0	0
6-7	5.9	1,789	1	6,900	-5,111	0	N	0	0
7-8	4.9	1,070	1	6,900	-5,830	0	N	0	0
8-9	4.0	1,073	1	6,900	-5,827	0	N	0	0
9-10	3.0	1,223	1	6,900	-5,677	0	N	0	0
10-11	2.1	1,113	1	6,900	-5,787	0	N	0	0
11-12	1.6	939	1	1,200	-261	0	Y	939	0
TOTALS	100.0	30,005						2,526	0

Table C.4 Work Zone Delay Calculation

Work Zone Length (mile)	Work Zone Speed (mph)	Unrestricted Speed (mph)	Work Zone Travel Time at Unrestricted Speed (hr/veh)	Work Zone Travel Time at Work Zone Speed (hr/veh)	Added Time to Travel Work Zone (hr/veh)
2	50	65	0.031	0.040	0.009

The delay cost is calculated for specific vehicle classes, and is the product of the percentage of class and the volume, additional travel time delay, and the average user cost per vehicle. Table C.5 shows the calculation based on the NJDOT RUCM. The reduction factor is used to accommodate for variations in traffic data, roadway capacities, and cost rates.

Table C.5 Work Zone Delay Cost Computation (NJDOT RUCM)

3.5(A)	3.5(B)	3.5(C)	3.5(D)	3.5(E)	3.5(F)	3.5(G)	3.5(H)
Road User Cost Component	Vehicle Class	Percent Class (%)	Total Vehicles (#)	Added Travel Length (mile/veh)	Added Time (hr/veh)	Cost Rate ($/veh-hr, $/mile)	Road User Cost ($)
Queue/Flagging Delay (Added Time)	CAR	90	2,526		0.000	18.15	0
	TRUCK	10	2,526		0.000	30.25	0
Queue/Flagging Idling VOC (Added Cost)	CAR	90	2,526		0.000	0.9695	0
	TRUCK	10	2,526		0.000	1.1150	0
Work/Flagging Zone Delay (Added Time)	CAR	90	2,526		0.009	18.15	371
	TRUCK	10	2,526		0.009	30.25	69
Circuity Delay (Added Time)	CAR	90	0		0.000	18.15	0
	TRUCK	10	0		0.000	30.25	0
Circuity VOC (Added Cost)	CAR	90	0	0.0		0.320	0
	TRUCK	10	0	0.0		0.640	0
Total Vehicles that Travel Queue:			0		Daily / Hourly Road User Cost		440
Total Vehicles that Travel Work Zone:			2,526		Calculated Road User Cost (CRUC)		330
Total Vehicles that Travel Detour:			0		Daily RUC (1) or Hourly RUC (0)		1
Percent Passenger Cars:			90%		Total Road User Cost (per Day)		330
Percent Trucks:			10%		Total Road User Cost (per minute)		

REFERENCES

Abraham, C. M. and J. J. Wang. Planning and Scheduling Work Zones Traffic Control. US Department of Transportation, Federal Highway Administration, 1981.

Adeli, H. Neural Networks in Civil Engineering: 1989-2000. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 16, No. 2, 2001, pp. 126-142.

Adeli, H., and S. L. Hung. *Machine Learning - Neural Networks, Genetic Algorithms, and Fuzzy Sets*. New York, NY: John Wiley and Sons, 1995.

Adeli, H., and X. Jiang. Neuro-fuzzy Logic Model for Freeway Work Zone Capacity Estimation. *Journal of Transportation Engineering*, Vol. 129, No. 5, 2003, pp. 484-493.

Adeli, H., and H. S. Park. *Neurocomputing for Design Automation*. Boca Raton, FL: CRC Press, 1998.

Al-Deek, H.M., Garib, A., and A.E. Radwan. New Method for Estimating Freeway Incident Congestion. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1494*, Transportation Research Board of the National Academies, Washington, D.C., 1995, pp. 30-39.

Bai, Y., and Y. Li. Determining Major Causes of Highway Work Zone Accidents in Kansas (Phase 2) (No. K-TRAN: KU-06-1). Kansas Department of Transportation, 2007.

Bartin, B., K. Ozbay, and S. Mudigonda. Interactive Lane Closure and Traffic Information Tool Based on a Geographic Information System. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2272*, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 44-55.

Benekohal, R., H. Ramezani, and K.A. Avrenli. WorkZoneQ User Guide for Two-Lane Freeway Work Zones. FHWA-ICT-13-019, 2013.

Bluetooth Travel Time Technology Evaluation - Using the BlueTOAD. Trafficcast Website: http://trafficcast.com/docs/PennDOT_BlueTOAD_final_report_incl_charts_4_Jan_2010.pdf, accessed on February 21, 2015.

Boto-Giralda, D., F. J. Diaz-Pernas, D. Gonzalez-Ortega, J. F. Diez-Higuera, M. Anton-Rodriguez, M. Martinez-Zarzuela, and I. Torre-Diez. Wavelet-Based Denoising for Traffic Volume Time Series Forecasting with Self-Organizing Neural Networks. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 25(7), 2010, pp. 530-545.

Bureau of Public Roads. Traffic Assignment Manual. U.S. Department of Commerce, Urban Planning Division, Washington D.C., 1964.

Cambridge Systematics, Inc., Travel time data collection - White Paper. http://www.camsys.com/pubs/WhitePaper_OD_TTData_Collection.pdf, accessed February 18, 2016.

Chatterjee, I., P. Edara, S. Menneni, and C. Sun. Replication of Work Zone Capacity Values in a Simulation Model. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2130*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 138-148.

Chen, H. and H. Rakha. Agent-Based Modeling Approach to Predict Experienced Travel Times. Presented at 93rd Transportation Research Board Annual Meeting, Washington D.C., 2014.

Chien, S., D. G. Goulias, S. Yahalom, and S. M. Chowdhury. Simulation-based Estimates of Delays at Freeway Work Zones. *Journal of Advanced Transportation*, Vol. 36, 2002, pp. 131-156.

Chien, S., and K. Ozbay. Development of New Jersey Rates for the NJCMS Incident Delay Model. Publication FHWA-NJ-2008-001. New Jersey Department of Transportation, 2012.

Chien, S., and P. Schonfeld. Optimal Work Zone Length for Four-lane Highways. *Journal of Transportation Engineering*, Vol. 127, No.2, 2001, pp. 124-131.

Chien, S., L. Spasovic, J. Lee, K. Mouskos, and B. Du. Feasibility of Lane Closures Using Probe Data. Draft Report. New Jersey Department of Transportation, 2016.

Chien, S., and Y. Tang. Scheduling Highway Work Zones with Genetic Algorithm Considering the Impact of Traffic Diversion. *Journal of Advanced Transportation*, Vol. 44, 2014, pp. 287-303.

Chitturi, M., and R. Benekohal. Comparison of QUEWZ, FRESIM and QuickZone with Field Data for Work Zones. Presented at 83rd Transportation Research Board Annual Meeting, Washington, D.C., 2004.

Choe, T., A. Skabardonis, and P. Varaiya. Freeway Performance Measurement System: Operational Analysis Tool. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1811*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 67-75.

Chung, Y. Quantification of Non-recurrent Congestion Delay Caused by Freeway Accidents and Analysis of Causal Factors. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2229*, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 8-18.

Chung, Y., H. Kim, and M. Park. Quantifying Non-Recurrent Traffic Congestion Caused by Freeway Work Zones Using Archived Work Zone and ITS Traffic Data. Transportmetrica, Vol. 8, No. 4, 2012, pp. 307-320.

Colombaroni, C. and G. Fusco. Artificial Neural Network Models for Car Following: Experimental Analysis and Calibration Issues. *Journal of Intelligent Transportation Systems*, Vol. 18, No. 1, 2014, pp. 5-16.

Copeland, L. User's Manual for QUEWZ-98. Texas Transportation Institute, FHWA/TX-98/1745-2, 1998.

CORSIM Website: http://www-mctrans.ce.ufl.edu/featured/TSIS/Version5/corsim.htm, accessed on February 2, 2016.

CoVal Systems. Introduction to OpenReach: http://www.covalsystems.com/latest/openreach/openreach.html, accessed on July 10, 2016.

Du, B., and S. Chien. Feasibility of Shoulder Use for Highway Work Zone Optimization. *Journal of Traffic and Transportation Engineering*, Vol. 1, No.4, 2014, pp. 235-246.

Du, B., S. Chien, J. Lee, L. Spasovic, and K. Mouskos. An Artificial Neural Network Model for Estimating Temporal and Spatial Freeway Work Zone Delay Using Probe-vehicle Data. *Transportation Research Record: Journal of the Transportation Research Board, No. 2573*, Transportation Research Board of the National Academies, Washington, D.C., 2016, pp. 164-171.

Du, B., J. Lee, S. Chien, B. Dimitrijevic, and K. Kim. Short-term Freeway Work Zone Capacity Estimation Using Support Vector Machine Incorporated with Probe-vehicle Data. Presented at 94th Transportation Research Board Annual Meeting, Washington, D.C., 2015.

Dudek, C. L. and S. H. Richards. Traffic Capacity through Urban Freeway Work Zones in Texas. In *Transportation Research Record: Journal of the Transportation Research Board, No. 869*, Transportation Research Board of the National Academies, Washington, D.C., 1982, pp. 14-18.

Edara, P. K. and B. H. Cottrell. Estimation of Traffic Mobility Impacts at Work Zones: State of the Practice. Presented at 86th Transportation Research Board Annual Meeting, Washington, DC, 2007.

Edara, P., C. Sun, and Z. Zhu. Calibration of Work Zone Impact Analysis Software for Missouri. Missouri Department of Transportation, 2013.

Elefteriadou, L., D. Arguea, A. Kondyli, and K. Heaslip. *Impacts of Trucks on Arterials LOS and Freeway Work Zone Capacity (Part B)*. Publication TRC-FDOT-54954-b-2007, Florida Department of Transportation, 2007.

Elhenawy, M., H. Chen, and H. Rakha. Dynamic Travel Time Prediction Using Genetic Programming. Presented at 93rd Transportation Research Board Annual Meeting, Washington, D.C., 2014.

Fact and Statistics - Work Zone Delay: http://www.ops.fhwa.dot.gov/wz/resources/facts_stats/delay.htm, accessed on July 2, 2016.

Federal Highway Administration. Work Zone Impacts Assessment - An Approach to Assess and Manage Work Zone Safety and Mobility Impacts of Road Projects, FHWA-HOP-05-068, 2006.

Federal Highway Administration. Work Zone Road User Cost - Concepts and Applications, FHWA-HOP-12-005, 2011.

Ferman, M., D. Blumenfeld, and X. Dai. An Analytical Evaluation of a Real-Time Traffic Information System Using Probe Vehicles. *Journal of Intelligent Transportation Systems*, Vol. 9, No. 1, 2005, pp. 23-34.

Fudala, N. and M. Fontaine. Work Zone Variable Speed Limit Systems: Effectiveness and System Design Issues. VTRC 10-R20, Virginia Transportation Research Council, 2010.

Gambatese, J. A., and M. Johnson. Work Zone Design and Operation Enhancements. Research Final Report SPR 669. Oregon Department of Transportation, 2010.

Ghosh-Dastidar, S., and H. Adeli. Neural Network-Wavelet Microsimulation Model for Delay and Queue Length Estimation at Freeway Work Zones. *Journal of Transportation Engineering*, Vol. 132, No. 4, 2006, pp. 331-341.

Habtemichael, F. G., M. Cetin, and K. A. Anuar. Methodology for Quantifying Incident-Induced Delays on Freeways by Grouping Similar Traffic Patterns. Presented at 94th Transportation Research Board Annual Meeting, Washington, D.C., 2015.

Haghani, A., M. Hamedi, and K. F. Sadabadi. I-95 Corridor Coalition Vehicle Probe Project. University of Maryland, College Park, 2009.

Heaslip, K., Kondyli, A., Arguea, D., Elefteriadou, L., and F. Sullivan. Estimation of Freeway Work Zone Capacity through Simulation and Field Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2130*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 40-50.

HERE Website: http://www.here.com/, accessed on June 14, 2015.

Transportation Research Board. *Highway Capacity Manual*. Washington, D.C.: National Academy of Sciences, 2010.

Highway Statistics 2013. U.S. Department of Transportation, Federal Highway Administration, 2013.

Hou, Y., P. Edara, and C. Sun. Traffic Flow Forecasting for Urban Work Zones. IEEE Transactions on Intelligent Transportation Systems, Vol. 16(4), pp. 1761-1770, 2015.

I-95 Corridor Coalition, "Vehicle Probe Project General Benefits White Paper", 2010.

Image Sensing Systems Website: http://www.imagesensing.com/products/radar-detection/autoscope-g4.html, accessed on February 18, 2016.

INRIX Website: http://www.inrix.com/. accessed on June 14, 2015.

Jiang, Y. Estimation of Traffic Delays and Vehicle Queues at Freeway Work Zones. Presented at 80th Transportation Research Board Annual Meeting, Washington D.C., 2001.

Jiang, X., and H. Adeli. Freeway Work Zone Traffic Delay and Cost Optimization Model. *Journal of Transportation Engineering*, Vol. 129, No. 3, 2003, pp. 230-241.

Joseph, C. T., E. Radwan, and N. M. Rouphail. Work Zone Analysis Model for the Signalized Arterial. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1194*, Transportation Research Board of the National Academies, Washington, D.C., 1988, pp. 112-119.

Karim, A., and H. Adeli. Radial Basis Function Neural Network for Work Zone Capacity and Queue Estimation. *Journal of Transportation Engineering*, Vol. 129, No. 5, 2003, pp. 494-503.

Krammes, R. A. and G. O. Lopez. Updated Capacity Values for Short-term Freeway Work Zone Lane Closures. Presented at 73rd Transportation Research Board Annual Meeting, Washington, D.C, 1994.

Kim, T., D. J. Lovell, and J. Paracha. A New Methodology to Estimate Capacity for Freeway Work Zones. Presented at 80th Transportation Research Board Annual Meeting, Washington D.C., 2001.

Kumar, K., M. Parida, and V. K. Katiyar. Short-term Traffic Flow Prediction in Heterogeneous Condition Using Artificial Neural Network. Transport, Vol. 30(4), 2015, pp. 397-405.

Lighthill, M. J. and G. B. Whitham. On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads. Proc. R. Soc. Lond. Ser. Math. Phys. Sci., Vol. 229, No. 1178, 1955, pp. 317–345.

Lin, W. H. and R. W. Hall. BTS (Verison 1.1) - Bottleneck Traffic Simulator User's Manual. Advanced Technology for the Highway Institute of Transportation Studies, University of California at Berkeley, 1991.

Lord, D. and F. Mannering. The Statistical Analysis of Crash-frequency Data: A Review and Assessment of Methodological Alternatives. *Transportation Research Part A: Policy and Practice*, Vol. 44, No. 5, 2010, pp. 291-305.

Maher, A., M. Jafari, E. Bossett, M. O'Connell, and J. Buison. Plan4Safety Website: http://cait.rutgers.edu/tsrc/plan4safety, accessed on July 10, 2016.

MATLAB Neural Network Toolbox. The MathWorks, Inc., Natick, Mass., 2014.

McCoy, P. T., M. Pang, and E. R. Post. Optimum Length of Two-lane Two-way No-passing Traffic Operation in Construction and Maintenance Zones on Rural Four-lane Divided Highways. In *Transportation Research Record: Journal of the Transportation Research Board, No. 773*, Transportation Research Board of the National Academies, Washington, D.C., 1980, pp. 20–24.

Memmott, J. L., and C. L. Dudek. Queue and User Cost Evaluation of Work Zones (QUEWZ). In *Transportation Research Record: Journal of the Transportation Research Board, No. 979*, Transportation Research Board of the National Academies, Washington, D.C., 1985, pp. 12-19.

Mendez, M. Work Zone Performance Data Collection – Way to blue. Thinking Highway Website: http://www.bitcarrier.com/sites/default/files/ThinkingHighways.pdf, accessed on February 20, 2016.

Meng, Q., and J. Weng. Cellular Automata Model for Work Zone Traffic. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2188*, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 131-139.

Middleton, D., R. Rajbhandari, R. Brydia, P. Songchitruksa, E. Kraus, S. Hernandez, K. Cheu, V. Iragavarapu, and S. Turner. Synthesis of TXDOT Uses of Real-Time Commercial Traffic Data. FHWA/TX-12/0-6659-1, 2011.

Mimbela, L. and L. Klein. A Summary of Vehicle Detection and Surveillance Technologies use in Intelligent Transportation Systems. FHWA Website: http://www.fhwa.dot.gov/policyinformation/pubs/vdstits2007/05.cfm, accessed on February 18, 2016.

Minitab Support Website: http://support.minitab.com/en-us/minitab-express/1/help-and-how-to/modeling-statistics/regression/supporting-topics/basics/a-comparison-of-the-pearson-and-spearman-correlation-methods/, accessed on July 2, 2016.

Mudge, R., H. Mahmassani, R. Haas, A. Talebpour, L. Carroll. Work Zone Performance Measurement Using Probe Data. FHWA-HOP-13-043, 2013.

Neubert, L., L. Santen, A. Schadschneider, and M. Schreckenberg. Statistical Analysis of Freeway Traffic. *Traffic and Granular Flow*, Springer, New York, 1999, pp. 307-314.

New Jersey Department of Transportation. 2015 Straight Line Diagrams Website: http://www.state.nj.us/transportation/refdata/sldiag/, accessed on Jul. 10, 2016.

New Jersey Department of Transportation Road User Cost Manual Website: http://www.state.nj.us/transportation/eng/documents/RUCM/, accessed on October 10, 2016.

New Jersey Turnpike Authority Website: http://www.state.nj.us/turnpike/ez-pass.html, accessed on February 21, 2016.

New Zealand Transport Agency. Transport Model Development Guidelines, 2014.

Oregon Department of Transportation. Web-Based Work Zone Traffic Analysis Tool Users' Guide, 2010.

Pan, B., U. Demiryurek, C. Gupta, and C. Shahabi. Forecasting Spatiotemporal Impact of Traffic Incidents for Next-generation Navigation Systems. *Knowledge and Information Systems*, Vol. 45, No. 1, 2015, pp. 75-104.

PARAMICS Website: http://www.paramics-online.com/, accessed on February 2, 2016.

Puckett, D. and M. Vickich. Bluetooth-Based Travel Time/Speed Measuring Systems Development. University Transportation Center for Mobility Project Number 09-00-17, 2010.

QuickZone Website: http://www.ops.fhwa.dot.gov/wz/traffic_analysis/quickzone/index.htm, accessed on February 2, 2016.

Request for Proposal for Traffic Data and Associated Services along the I-95 Corridor. RFP No. 82085N, issued by the University of Maryland College Park, 2007.

Richards, P. I. Shock Waves on the Highway. Operations Research, Vol. 4, No. 1, 1956, pp. 42–51.

Sbayti, H., Y. Chiu, and H. S. Mahmassani. Calibration of Dynamic Traffic Assignment Models for Operational Planning: Methodology and Results using DYNASMART-P. *INFORMS*, San Jose, 2002.

Scholkopf, B., A. J. Smola, R. C. Williamson, and P. L. Bartlett. New Support Vector Algorithms. *Neural computation*, Vol. 12, No. 5, 2000, pp. 1207-1245.

Schrank, D., T. J. Lomax, and S. Turner. TTI's 2015 Urban Mobility Report. Texas Transportation Institute, Texas A&M University System, College Station, 2015.

Schroeder, B. J., N. M. Rouphail, B. M. Williams, A. Hajbabaie, B. Aghdashi, S. Kim, K. Tabrizi, and B. Narron. Delay and User Cost Estimation for Work Zones on Urban Arterials. Institute for Transportation Research and Education, North Carolina State University, 2015.

Seymour, E., N. Chaudhary, D. Middleton, R. Brydia, and L. Miller. White Paper: State of ITS Industry and Assessment of Project Types. FHWA/TX -11/0-6672-1, 2011.

Smith, S., R. Worrall, D. Roden, R.A. Pfefer, and M. Hankey. Application of Freeway Simulation Models to Urban Corridors, Volume 1: Final Report, Publication No. FHWA-RD-92-103, FHWA, 1992.

Suykens, J., T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific Pub. Co., Singapore, 2002.

Tang, Y., and S. Chien. Scheduling Work Zones for Highway Maintenance Project Considering Variable Time-cost Relation. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2055*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 21-30.

Tarnoff, P., J. Wasson, S. Young, N. Ganig, D. Bullock, and J. Sturdevant. The Continuing Evolution of Travel Time Data Information Collection and Processing. Presented at 88th Transportation Research Board Annual Meeting, Washington, D.C., 2009.

Taylor N. B. The CONTRAM Dynamic Traffic Assignment Model. *Networks and Spatial Economics*, Vol. 3, No. 3, pp. 297-322, 2003.

Taylor, D. R., S. Muthiah, B.T. Kulakowski, K. M. Mahoney, and R. J. Porter. Artificial Neural Network Speed Profile Model for Construction Work Zones on High-speed Highways. *Journal of Transportation Engineering*, Vol. 133, No. 3, 2007, pp. 198-204.

TomTom Website: http://www.tomtom.com/en_us/. accessed on June 14, 2016.

Turner, D., P. Ray, M. Wang, M. Wang, M. Fincher, M. Lanctot, and Q. Cui. Work Zone Lane Closure Analysis Model (No. ALDOT 930-721). University Transportation Center for Alabama, 2009.

Ullman, G. L., J. L. Timothy, and T. Scriba. A Primer on Work Zone Safety and Mobility Performance Measurement. US Department of Transportation, Federal Highway Administration, 2011.

Vapnik V. *The Nature of Statistical Learning Theory*. New York, NY: Springer, 1995.

Vapnik V. *Statistical Learning Theory*. New York, NY: John Wiley and Sons, 1998.

Vemuri, A. T., M. M. Polycarpou, and P. D. Pant. Short-term Forecasting of Traffic Delays in Highway Construction Zones Using On-line Approximators. *Mathematical and computer modelling*, Vol. 27, No. 9, 1998, pp. 311-322.

VISSIM Website: http://vision-traffic.ptvgroup.com/en-us/products/ptv-vissim/, accessed on February 2, 2016.

Vo, T. An Investigation of Bluetooth Technology for Measuring Travel Times on Arterial Roads: A Case Study on Spring Street. MS thesis. Georgia Institute of Technology, Atlanta, 2011.

Wang, Y., R.L. Cheu, and T.F. Fwa. Highway Maintenance Scheduling Using Genetic Algorithm with Microscopic Traffic Simulation. Presented at 81st Transportation Research Board Annual Meeting, Washington, D.C., 2002.

Washburn, S., T. Hiles, and K. Heaslip. Impact of Lane Closure on Roadway Capacity: Development of a Two-Lane Work Zone Lane Closure Analysis Procedure (Part A), 2008.

Weng, J., and Q. Meng. Effects of Environment, Vehicle and Driver Characteristics on Risky Driving Behavior at Work Zones. *Safety Science*, Vol. 50, No. 4, 2012, pp. 1034-1042.

Weng, J., and Q. Meng. Estimating Capacity and Traffic Delay in Work Zones: An Overview. Transportation Research Part C: Emerging Technologies, Vol. 35, 2013, pp. 34-45.

White, J. and M. Porter. El Segundo Area ITS Project: Arterial Traveler Information – Is it Possible? Presented at the ITE Western District 2010 Annual Meeting, San Francisco, California, 2010.

Wirasinghe, S. Determination of Traffic Delays from Shock-wave Analysis. *Transportation Research*, No. 12, 1978, pp. 343-348.

Wright, J. and J. Dahlgren. Using Vehicles Equipped with Toll Tags as Probes for Providing Travel Times. California Partners for Advanced Transit and Highways, 2000.

Wu, C., J. Ho, and D.T. Lee. Travel-time Prediction with Support Vector Regression. IEEE Transactions on Intelligent Transportation Systems, Vol. 5, No. 4, 2004, pp. 276-281.

Wardrop, J. G. Some Theoretical Aspects of Road Traffic Research. Proceedings of Institute of Civil Engineers, Part II, Vol. 1, No. 2, 1952, pp. 325-378.

Xiao, J. and Y. Liu. Traffic Incident Detection Using Multiple-Kernel Support Vector Machine. In *Transportation Research Record: Journal of the Transportation*

Research Board, No. 2324, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 44-52.

Xie, Y., K. Zhao, Y. Sun, and D. Chen. Gaussian Processes for Short-term Traffic Volume Forecasting. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2165*, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 69-78.

Yang, N., Schonfeld, P., and M.W. Kang. A Hybrid Methodology for Freeway Work-Zone Optimization with Time Constraints. *Public Works Management & Policy*, Vol. 13, No. 3, 2008, pp. 1-12.

Yu, R., and M. Abdel-Aty. Utilizing Support Vector Machine in Real-time Crash Risk Evaluation. *Accident Analysis & Prevention*, Vol. 51, 2013, pp. 252-259.

Zhang, Y. and Y. Xie. Forecasting of Short-term Freeway Volume with V-support Vector Machines. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2024*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 92-99.