# FORECASTING TRAVEL TIME AND VARIATIONS IN TRAVEL TIME DUE TO VEHICLE ACCIDENTS IN SPATIO-TEMPORAL CONTEXT ALONG FREEWAY


by

Rashid Mohammad Zahid Reza


A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Infrastructure and Environmental Systems

Charlotte

2015

Approved by:

______________________________
Dr. Srinivas S. Pulugurtha

______________________________
Dr. Rajaram Janardhanam

______________________________
Dr. Martin R. Kane

______________________________
Dr. Wei Fan

______________________________
Dr. Barry Sherlock

# ABSTRACT

RASHID MOHAMMAD ZAHID REZA. Forecasting travel time and variations in travel time due to vehicle accidents in spatio-temporal context along freeway. (Under the direction of DR. SRINIVAS S. PULUGURTHA)

The growth in population, technological advancements, increase in average life expectancy, and newer and fuel-efficient vehicles has led to a meteoric rise in travel demand over the past few decades. However, the road network capacity has not increased at the same progressive rate, resulting in congestion and associated traffic problems. Traffic incidents are major contributors of non-recurring congestion in most of the urban areas in the United States.

Travel time is an effective parameter to quantify the congestion at segment- or corridor-level. Short-term traffic and travel time prediction plays a vital role in advanced traveler information systems (ATIS) and assists in proactive management of transportation network. However, forecasting travel time is complex under over-saturated conditions and in the presence of an incident. Besides, travel time itself cannot explain the exact impact of the incident as it varies with respect to time and over space. Incorporating all factors that affect traffic and travel time increases the magnitude of the complexity. Therefore, this research focuses on an application of Autoregressive Integrated Moving Average (ARIMA) model, incorporating travel time information over space (from neighboring segments) and time, to forecast travel time and relative variations in travel time (RVTT) along a freeway corridor in spatio-temporal context. The RVTT was considered instead of variation in travel time to negate the effect of difference in segment lengths and other geometric characteristics for a meaningful comparison. Two

types of RVTT were considered: travel time / expected travel time and travel time / minimum travel time.

Travel time data was collected from INRIX and incident data was gathered from Traveler Information Management System (TIMS) from 2010 to 2012. Databases were developed using data, for 150 “vehicle accident” affected days and 100 sample days of data when there were no incidents, along a ~19-mile freeway corridor of I-77 S in the city of Charlotte, North Carolina.

Four categories of Cronbach’s α were computed at 10-minute intervals for each segment. The higher value was selected as the corresponding Cronbach’s α to capture the expected travel time of that segment. Minimum travel time of the segment was estimated as the minimum of all the travel time samples in a 10-min interval for each day-of-the-week. However, when the minimum travel time was zero then the second minimum travel time was taken as it was assumed that no vehicle passed that segment or data was recorded during the specific time interval.

Results obtained indicate that the difference between the observed and the expected travel time is less than 10% for almost 85% samples and less than 15% for 90% of the samples considered from 2010 to 2012. This indicates the effectiveness of Cronbach’s α in capturing the expected travel time for a certain time interval of a segment.

Lagged regression model was then developed using data for 18 segments. Three scenarios (1) travel time (TT), (2) travel time/expected travel time (TT/ExpTT), and (3) travel time/ minimum travel time (TT/MinTT) were considered for both without incident and under “vehicle accident” conditions. Developed models from all three scenarios showed that the travel times and RVTT for consecutive segments are highly correlated, and both upstream and downstream segments have an influence on the current state of the target segment. Moreover, all the significant predictor variables had a time lag of 10 minutes. In other words, the prediction horizon is 10 minutes.

The Mean Absolute Percent Error (MAPE) and Mean Absolute Deviation (MAD) of the developed lagged regression model are estimated for every segment, irrespective of the incident condition. MAPE and MAD values of all segments of TT are less than 10% for all but one segment, for which MAPE value was marginally greater than 10%. For both TT/ExpTT and TT/MinTT scenario, MAPE and MAD value for all segments was less than 10%, except for one segment.

For model validation, a total of 80 days of data was considered (45 without incident days and 35 "vehicle accident" affected days). Results showed that MAPE and MAD values are less than 15% for TT and TT/MinTT scenarios of almost all the segments. However, for TT/ExpTT scenario, all of them are less than 15%. Moreover, both the calibrated and validated models demonstrated that modeling using TT/ExpTT would yield accurate results than TT/MinTT. Except for four segments, forecasting accuracy of TT/ExpTT was higher than TT/MinTT for all other segments. Overall, the adopted methodology successfully forecasted travel time and variations in travel time for both without incident and under "vehicle accident" condition.

# DEDICATION

This dissertation is dedicated to few persons whose contribution in my life is limitless. I like to thank my father, Dr. M. A. Rashid, my mother, Khadija Akhter, and my sister, Ishrat Mizushima, who have been constantly supporting and motivating me to accomplish this honor from my home country Bangladesh. I also like to thank the Bangladeshi community at Charlotte for letting me have a wonderful life. Lastly, I want to thank my wife, Naima Khan, who has been assisting me for the last three years without a single hint of pain and with whom I have passed few hard times of my life. I want to thank her again for being part of my life.

# ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Srinivas S. Pulugurtha, Professor in the Department of Civil and Environmental Engineering for his valuable guidance and support throughout my Ph.D. study at The University of North Carolina at Charlotte. I would like to thank Dr. Rajaram Janardhanam, Dr. Martin R. Kane, Dr. Wei Fan and Dr. Barry Sherlock for serving on my dissertation committee. I would also like to thank my colleagues in the transportation lab for helping me with perspectives and discussions about the study.

This dissertation is prepared based on information collected for a research project funded by the United States Department of Transportation - Office of the Assistant Secretary for Research and Technology (USDOT/OST-R) under Cooperative Agreement Number RITARS-12-H-UNCC. The support and assistance by the project program manager and the members of Technical Advisory Committee in providing comments on the methodology are greatly appreciated. The contributions from other team members at the University of the North Carolina at Charlotte are also gratefully acknowledged. Special thanks are extended to Ms. Kelly Wells of North Carolina Department of Transportation (NCDOT) and Mr. Michael VanDaniker of the University of Maryland (UMD) for providing access and help with INRIX data.

# DISCLAIMER

The views, opinions, findings, and conclusions reflected in this dissertation are the responsibility of the researchers of the subject research project only and do not represent the official policy or position of the USDOT/OST-R, NCDOT, UMD, INRIX, or any other State, or the University of North Carolina at Charlotte or other entity. The researchers are responsible for the facts and the accuracy of the data presented herein. This dissertation does not constitute a standard, specification, or regulation.

TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1: INTRODUCTION

Travel time is one of the most important link-, segment- or corridor-level performance measures. It can provide more meaningful information of a traffic network than other traffic variables of the network (Zeng et al. 2009). Travel time is easy to comprehend and visualize by everyone. Regular transportation systems users, businesses as well as transportation planners and engineers rely on travel time as a measure to make appropriate decisions.  Travel time also helps evaluate the performance of transportation system and serves as a criterion to determine the optimal route from an origin to a destination.

## 1.1 Background

The growth in population, technological advancements, increase in average life expectancy, and newer and fuel-efficient vehicles has led to a meteoric rise in travel demand over the past few decades. However, the road network capacity has not increased at the same progressive rate. Therefore, congestion as well as associated traffic problems has been increasing steadily. An increase in congestion means driver has to spend more time in traffic on their way to the destination. Traffic congestion affects in many ways. The Federal Highway Administration (FHWA) classified congestion into three broad categories: traffic influencing events, traffic demand, and physical highway features.

Based on these categories, FHWA has identified seven root causes for congestion. They are listed as follows (Cambridge Systematics 2005).

1. Physical bottlenecks (“Capacity”),
2. Traffic incidents,
3. Work zones, Weather,
4. Traffic control devices,
5. Special events, and,
6. Fluctuations in normal traffic flow.

One or more than one source listed above can cause congestion. In general, traffic congestion occurs when the volume of traffic is greater than the road capacity. This is also termed as over-saturated condition. Due to congestion, transportation system users spend extra time in traffic when traveling from their origin to destination.

The 2012 Urban Mobility Report summarizes the impact of congestion on travel time. The report is based on an analysis of data considering over 498 urban areas in the United States from 1982 to 2011 (Schrank et al. 2012). A few highlighted from the aforementioned report are summarized next.

1. Total travel delay per year increased from 1.1 billion hours to 5.5 billion hours (total percent increase is 400%) in the last 30 years.
2. Delay on any weekday is higher than weekend and around 63% of the total delay occurs during peak hours and the remaining 37% occurs during off-peak hours.

3. Local roads observe more delay (peak hour delay accounts for 34% and off-peak hour delay accounts for 26%) than freeway (peak hour delay accounts for 29% and off-peak hour delay accounts for 11%).
4. Passenger vehicles contribute to 73% of the vehicle miles traveled in urban areas, which imparted 78% of the congestion cost whereas the remaining congestion is imparted by heavy vehicles.
5. About 13% of vehicle miles traveled occurred at extreme congestion level, which caused 64% of travel delay in 2011.
6. Commuters suffered an average of 6 hours of delay per year during weekdays (~52 hours of delay per year in areas over 3 million population) in 2011. Throughout the 498 urban areas in the United States, yearly per commuter delay had increased to 38 hours from 16 hours (percent increase is 137.5%) in between 2011-1982.
7. Travel Time Index (TTI) has increased from 1.07 to 1.18, which revealed that a 20-minute free-flow trip takes 23.6 minutes instead of 21.4 minutes during the peak period.
8. Planning Time Index (PTI) for freeways is 3.09, which indicated that a transportation system user should plan for a 46.35-minute trip to ensure on time arrival at the destination 95% of the time during congested hours when compared to complete a 15-minute trip in light traffic conditions.

In addition to the increased travel time, congestion has a negative impact on environment and economy as it causes excess fuel consumption and gas emissions. The Centre for Economics and Business Research (CEBR) forecasted that by 2030, the $CO_2$

equivalent emissions in monetary terms would be $538.2 million for the United States including both direct and indirect cost of households. Key parameters that have an effect on emissions include economic growth, gross domestic product (GDP) per capita, and growth of population, total vehicle miles travelled on roads by commuters, business and freight transportation users (CEBR 2014). The present condition and trends pertaining to fuel consumption and air quality are summarized next based on the 2012 Urban Mobility Report (Schrank et al. 2012).

1. Consumption of extra fuel during congestion or wasted fuel has increased to 2.9 billion gallons in 2011 (which can fill four New Orleans Superdomes) from 0.5 billion gallons in 1982.
2. Congestion cost, which is a combination of yearly delay time and wasted fuel, increased almost 5 times from $24 billion to $121 billion in the last 30 years. With this $121 billion, it is possible to cover the lost productivity and direct medical expenses of 12 average flu seasons (time span characterized by the prevalence of outbreaks of influenza).
3. 56 billion pounds of additional $CO_2$ greenhouse gas is released into the atmosphere (which is same as total liftoff weight of 12,400 full fuel loaded Space Shuttles) during urban congested conditions compared to 10 billion pounds of additional $CO_2$ in 1982.
4. Yearly congestion cost per auto commuter was $818 in 2011 compared to an inflation-adjusted cost of $342 in 1982.

5. $73 billion (78%) of the delay cost was from passenger car and the remaining is from truck operation excluding any value for the goods being transported in the trucks.

Overall, the impact of congestion can be summarized as longer travel times i.e., loss of productivity and negative impact on economy, excess fuel consumption, and increase in emissions (harmful effect on the environment).

Congestion can be subdivided into two parts: recurring congestion and non-recurring congestion. Hallenbeck et al. (2003) defined non-recurring congestion as unusual congestion caused by an unexpected or transient event compared to other days. It can be caused by a variety of factors like lane blocking due to crash, disabled vehicles, debris on the roadway; lane closure due to construction; inclement weather; roadside distractions due to roadside construction, electronic signs, a fire beside the freeway etc. Alike recurring congestion, non-recurring congestion caused by incidents may lead to sudden changes in traffic condition resulting in significant delays on urban roads. Along with congestion, incidents have many effects such as secondary incidents resulting in increased delays, bottlenecking, and rubbernecking (Tennessee Department of Safety 2003).

FHWA estimated that 25% of the congestion is caused by traffic incidents (Cambridge Systematics 2005). It is one of the major contributors of non-recurring congestion in most of the urban areas in the United States. Among the different types of traffic incidents, vehicle crash is the most severe one. From 2004 to 2013, a total of 339,833 fatal crashes and 57,372,000 non-fatal crashes occurred throughout the United States. Due to fatal crashes, a total 373,598 people were killed while 24,247,000 were injured due to both fatal and non-fatal crashes within this timeframe. Progress has been made to reduce yearly crash rate as well as fatality rate and injury rate nationwide. In 2013, the fatality rate per 100 million VMT decreased to 1.09, the lowest point historically, compared to 1.44 per 100 million VMT in 2004. A similar scenario was observed for the fatality rates per population, licensed drivers, and registered vehicles. In 2004, the injury rate per 100 million VMT was 94 but after 2008 the injury rate per 100 million VMT did not exceed 80. It was estimated as equal to 77 in 2014 (NHTSA 2015).

## 1.2 Problem Statement

One of the most important parameters to estimate the effect of traffic and incidents such as "vehicle accident" on congestion is travel time (Cambridge Systematics et al. 2008). Travel time information is an essential part of Advanced Traveler Information Systems (ATIS) and Advanced Traffic Management Systems (ATMS). Travel time provides significant information pertaining to trip assignment. It influences start time of the trip, mode of travel, routing decision, or even cancel their trip. To planners, all these details are vital to make better and informed decisions on strategies for the management and the optimal guidance of traffic. Moreover, it is important to forecast how transportation network performance would vary under a given set of conditions in the near future. Therefore, short-term prediction is necessary for improved traveler information or traffic management (Booz Allen & Hamilton Inc. 1998).

Under free-flow conditions, travel time can be easily predicted based on the free flow speed. Even if the traffic state remains stable over time as well as across space and the traffic volume starts increasing, travel time is still very predictable as a function of relevant traffic characteristics such as travel time, speed, volume and occupancy.

However, the traffic condition becomes complex and is non-linear in nature when traffic volume levels are close to the over-saturated condition. Furthermore, incidents induce more complexities to the traffic system, which reduces the preciseness of travel time prediction. Therefore, the prediction of travel time during incident and immediately after an incident is a challenge for planners and engineers (Zeng et al. 2009).

Appropriate and reliable data collection is another important issue. Various traffic sensing technologies have been used to collect traffic data for travel time estimation, including point to point travel time collection (e.g., manual probe vehicle, license plate recognition systems, automatic vehicle identification systems, Bluetooth detectors, etc.) and station based traffic state measuring devices (e.g., loop detector, video camera, and remote traffic microwave sensor) (Chen et al. 2013). Private companies such as INRIX, Tom Tom and HERE integrate different sources of data to provide section-based traffic state data (e.g. speed, average travel time, etc.).

Short-term traffic prediction plays a vital role in ATIS and proactive management of transportation network. It has already been mentioned that as the traffic and congestion has been increasing steadily, an accurate prediction of future travel time is becoming more important in trip planning for commuters. The core of the short-term forecasting is to identify the fundamental pattern in the traffic data and utilize that information to forecast future traffic pattern such as its speed, flow, and travel time. In short-term traffic forecasting, there are two groups of models: univariate and multivariate. Univariate models use past internal patterns of traffic variables at one specific site to forecast future values. However, multivariate forecasting methods can capture both temporal and spatial evolvement of traffic (Yang et al. 2014).

The effectiveness of multivariate method enhances with the incorporation of surrounding traffic condition. The accuracy of traffic forecasting is closely related to the use of neighboring segments' traffic information. The spatial-temporal relationships between traffic variables of consecutive segments are highly correlated under different traffic conditions (Ahmed et al. 1979, Yang et al. 2014).

However, travel time itself cannot provide the exact impact of the incident as it varies with respect to time (peak period compared to off-peak period). Travel pattern changes even in the presence of recreational events like game day, festival, long-weekend. Incidents can add new dimension to those scenarios based on the numbers of blocked lanes, damaged vehicles clearance time, arrival pattern of the vehicles, accident occurrence time-of-the-day, etc. Variation in travel due to "vehicle accident" will help better understanding the effect of the accident and will help the planners to come up with an effective alternative route plan. For example, if regular travel time along a freeway segment is 4 minutes but increases to 8 minutes due to a "vehicle accident", expressing that the vehicle travel time has increased by 100% or two times will be a more effective way to explain the scenario.

The length of all the segments is not same in the transportation network. Travel time variation can eliminate the impact of the difference in lengths. For instance, it is assumed that two segments of 3 miles and 5 miles shows regular travel time of 5 minutes and 3 minutes, but due to incident, travel time changes to 8 minutes and 7 minutes, respectively. Segment 1 still shows higher travel time than segment 2 but travel time variation on segment 2 is higher than segment 1. Considering these facts travel time variation simply outperforms travel time as a performance measure. Therefore, the variation in travel time also needs attention in addition to travel time.

The variation in travel time requires identification of incident free travel time. Obviously, it is hard to say what could be the travel time if there is no accident occurred at that time. Most probable or expected travel time can provide some insights in this regard. Therefore, identifying or computing the expected travel time is the first step to determine the approximate travel time variation.

Various methods were devised in the past to forecast different traffic parameters. However, very little is done on developing models to forecast travel time and examine spatio-temporal variations in travel time due to incidents such as "vehicle accident".

This Dissertation, therefore, addresses this need and focuses on adopting multivariate time series Autoregressive Integrated Moving Average (ARIMA) model in forecasting the effect of incidents on travel time and variations in travel time as well as to check the effectiveness Cronbach's α to estimate the expected travel time.

## 1.3 Goal and Objectives

The goal of the proposed research is to improve operational performance on urban transportation networks through dissemination of information, by transportation system managers, that can help make transportation system users make better mode, route choice and departure time decisions. The following objectives are identified and defined to achieve the research goal.

1. To define the variation in travel time,
2. To research and develop a method to estimate the expected travel time for segments of a freeway for each time interval,

3. To forecast travel time for segments of a freeway when there is no “vehicle accident” and when there is a “vehicle accident”,
4. To research and evaluate the spatial-temporal relationships in travel time and travel time variations,
5. To model and assess the effect of “vehicle accident” on system performance so that planners can provide proper route guidance and individuals can make informed trip planning, and,
6. To calibrate the developed model for enhanced prediction and validate the application of the methodology.

1.4 Research Significance

The traffic condition of neighboring segments over time play a significant role in short-term traffic condition forecasting of the target segment. For instance, it is expected that traffic condition on the target segment will be highly correlated with its immediate downstream and / or upstream segment traffic condition. This relationship could vary over time after a “vehicle accident”. Forecasting and disseminating this information will help transportation system users make improved trip planning decisions.

Different technologies have been adopted to forecast travel time in the past. However, not many studies have focused on travel time forecasting under “vehicle accident” condition. Moreover, travel time itself cannot precisely assess the effect of incident as the length of the segment varies and travel time of a segment is not same throughout the whole day.

Generally, delay is used as a means to estimate the effect of an incident on the transportation network. However, it would be very hard to measure delay in the absence of traffic volume data. This dissertation uses variations in travel time and relative variation in travel time to capture the effect of an incident.

## 1.5 Organization of the Dissertation

The reminder of this Dissertation is comprised of six chapters. Chapter 2 provides a literature review of previous works on forecasting different traffic parameters, both under incident and without incident condition and on capturing expected travel time. Chapter 3 introduces Cronbach's α for estimating expected travel time and ARIMA model and its different parameters (Auto-correlation function, Partial Auto-correlation function, etc.) for time series forecasting. Chapter 4 discusses about the study area and data required for this research. Chapter 5 presents the definition of variation in travel time, study corridor, and database development. It also covers the systematic approach for finding maximum corresponding Cronbach's α for each segment by time-of-the-day as well as day-of-the-week and the systematic method of building ARIMA model from travel time data for each segment. Chapter 6 shows the effectiveness of the adopted methods in estimating the expected travel time, presents results from ARIMA model to forecast travel time and variation in travel time for both with "vehicle accident" and without "vehicle accident" incident condition, and model validation. Finally, chapter 7 presents the conclusions and recommendations for future research.

# CHAPTER 2: LITERATURE REVIEW

This chapter presents a review of different methods to capture the expected travel time, to forecast travel time and its variation under "vehicle accident" and without vehicle accident. It includes a review of literature on the support vector regression, the time series model, the neural network, combination of the time series and the neural network, and other methods for forecasting of travel time, different empirical methods on expected travel time computation, and previous studies on capturing the effect of incidents.

## 2.1 Performance Evaluation of Machine Learning Models in Forecasting Traffic Parameters

Machine learning, a subfield of computer science, is a method that uses past available data to improve computational performance. This process can deal with different types of data varying from digitized human-labeled training sets data to data obtained from human environment interaction.  Machine learning algorithms have been used in the field of computer science as well as biological science. They have been successfully implemented to solve problems that involve or related to classification, clustering, regression, anomaly detection, recognition, online learning, and grammar induction (Bishop 2006).

Machine learning tasks are generally classified into three sections depending on the nature of the learning system: supervised learning process, unsupervised learning process, and reinforcement learning process. Supervised learning is a type of machine learning task that tries to find a function from labeled training data. Unsupervised learning tries to figure out the hidden structure in unlabeled data, while reinforcement learning concerns with how software agents deals with environment to maximize cumulative performance. Selected prominent machine learning algorithms are listed next (Bishop 2006).

1. Support vector machine (SVM)
2. Support vector regression (SVR)
3. Decision trees
4. *k*-NN
5. Linear regression
6. Naïve Bayes
7. Neural networks
8. Logistic regression

SVM theory was invented by V. Vapnik in 1995 at the AT&T Bell Laboratories (Vapnik 1995). Since then it has been successfully applied for forecasting different traffic parameters. SVM can generalize the training data as well as can reach the global minima. Wu et al. (2003) applied SVM to forecast travel time along a Taiwan freeway. Results showed that SVM outperformed current travel time prediction method and historical mean prediction method (average of all historical travel time samples for a specific time-of-the-day and day-of-the-week). Konkaew et al. (2013) adopted SVM for prediction of short-term travel time along an expressway in Bangkok, Thailand. Data was obtained from video image processing cameras. Results revealed that SVM approach significantly outperformed historic approach for all prediction horizons from 0 minute to 30 minutes in 10 minutes increment. Chen et al. (2012) conducted a study on predicting travel time of Bus Rapid Transit (BRT) using SVM and Global Positioning System (GPS) data. They also applied Kalman filter algorithm to adjust the results of the predicted travel time. Results proved that the proposed model could predict BRT vehicle travel time with a high prediction accuracy for both off-peak hour and peak hours.

Not only travel time, SVM can successfully forecast traffic volume (traffic flow). Gao et al. (2010) used SVM for forecasting traffic volume using three months traffic volume data in China. Relative forecasting error of SVM was less than both Grey Predicting (GP) model and individual ANN model. The SVM combined forecasting method showed higher accuracy than linear combined forecasting method and ANN combined forecasting method. Sun et al. (2012) forecasted traffic flow using SVM. Gao et al. (2010) also incorporated feature extraction methods such as Principal Component Analysis (PCA), Self-Organization Map (SOM) network, and Multidimensional Scaling (MDS). Among these three, MDS increased the computation speed the most.

Traffic flow varies with respect to time. Therefore, Li (2009) added a time dependent reconstruction process to SVM structure. This process outperformed SVM without a time-dependent structure. Absolute mean error (AME) and mean squared error (MSE) of the prediction model were 5.1 veh/5 min and 6.0 veh/5 min, respectively. Casto-Neto et al. (2008) applied Online SVR (OL-SVR) method for forecasting short-term traffic flow under atypical conditions like vehicular crashes, hazardous weather, work zone, holidays, etc. The OL-SVR model was compared with Gaussian maximum likelihood (GML), Holt exponential smoothing, and ANN. Results showed that OL-SVR performed better than any other considered model.

A significant number of studies focused on forecasting speed profile along both arterial and freeway. Yildrim et al. (2008) predicted speed using SVM. Data was collected from remote traffic microwave sensors. Speed in one sensor was predicted either with respect to historical speed data or based on the speed correlations with other sensors. Predictor horizon varied from 5 minutes to 60 minutes. Results showed that SVM performed very accurately and provided better results than *k* Nearest Neighbor (*k*NN) method.

Generally, forecasting is made for a segment or corridor. But Asif et al. (2012) used SVR to predict traffic speed in a large and heterogeneous road network, which could provide network wide scenario of speed profile. Moreover, k-means clustering, PCA, and SOM were adopted to capture the spatial and temporal performance for both network level and for individual links for multiple prediction horizons. SVR was also successfully implemented for incident duration forecasting and incident detection.

Besides SVM, KNN was also used to forecast different traffic parameters. The basic principle of KNN is that it tries to match current traffic scenario with historic traffic pattern. Zhang et al. (2013) used KNN method for forecasting short-term traffic flow along expressway. Results showed that the forecasting accuracy of the proposed method was more than 90%. Chen et al. (2013) demonstrated the application of KNN for prediction of travel time. They used spatio-temporal traffic state as control variable instead of travel time. Results showed that this method reduced the prediction error by 7.5 minutes to 3 minutes on a 50-minute trip. Naïve clustering approach was also used for forecasting travel time in the past. Deb Nath et al. (2011) applied Naïve clustering approach for forecasting travel time and found that it provided better results than Rule Based method, Naïve Bayesian Classification (NBC) method, Successive Moving Average (SMA) and Chain Average (CA).

Forecasting is generally performed for individual road segments and prediction horizons. However, Dauwels et al. (2014) developed a collective prediction for multiple road segments and prediction-horizons. They developed different models through partial least squares (PLS), higher order partial least squares (HO-PLS) and N-way partial least squares (N-PLS). These models successfully forecasted traffic conditions for multiple road segments as well as prediction-horizons.

Overall, different machine learning methods were applied successfully to forecast different traffic parameters. ANN, which is a part of machine learning technique and used to forecast traffic parameters, is discussed in the next section.

## 2.2 Performance Evaluation of Different Types of Artificial Neural Network (ANN) Models in Forecasting Traffic parameters

Artificial Neural Networks (ANN) is a large-scale parallel-linked system, which simulates the structure of the human brain with self-adapting modeling and studying functions (Angeli 2010). Generally, ANN consists of a number of structural constituents that helps in non-linear computations. The objective function of neurons is then computed through a two way process: at first, by summing up the product of the input signals and synaptic weights, and secondly, by transforming the summation of the inputs through a transfer function to produce an output (Hassoun 1996). An example ANN model is shown in Figure 1.

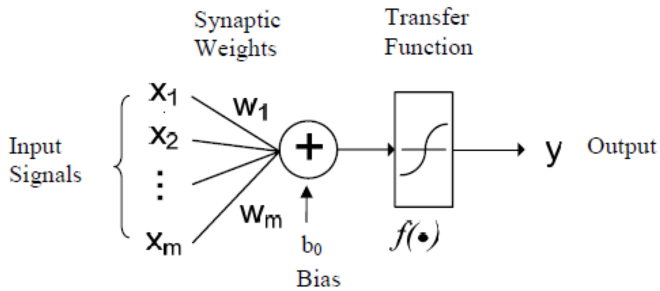



FIGURE 1: Model of neuron (Zeng 2009)

The framework of ANN can be applied to a wide range of forecasting problems with a high degree of accuracy. However, sometimes ANN yields mixed results in solving linear problems, and hence not applicable for any type of data (Zeng 2009).

ANN has been widely used to solve the travel time forecasting problem. Many ANN approaches had resulted in satisfactory prediction performance under certain conditions. Lint et al. (2002) applied state-space neural network (SSNN) for travel time prediction of freeway. Output travel time showed zero mean normally distributed residuals were within a range of 10% of the real expected travel times most of the cases. Furthermore, analyses of the internal states and the weight configurations revealed that the SSNN based internal models are closely related to the underlying traffic processes. Lint et al. (2005) also proposed a framework for travel time prediction along freeway, which exhibits capability in dealing with missing or corrupt input data. This neural network framework also consisted of SSNN topology. Robustness of missing data was tackled by means of simple imputation (data replacement) schemes, such as exponential forecasts and spatial interpolation. Results showed that this framework of SSNN travel time prediction yields accurate and robust travel time predictions on both synthetic and real data. Liu et al. (2009) applied SSNN to forecast travel time on urban arterial streets. Modeling for a long urban arterial street in the Netherlands was done by assembling multiple segments. The results indicated that this proposed model is capable of forecasting travel time with satisfying accuracy.

For better training of ANN topology, different methods have been adopted in the past. Li et al. (2009) used SSNN with adaptive filters to predict the urban arterial travel time. Adaptive filters were proposed to improve the effectiveness through training the SSNN instead of conventional approaches. Model performance was tested with urban arterial data. The performance of the proposed model was compared with that of conventional SSNN and Back Propagation Neural Network (BPNN). Results showed that the proposed method successfully forecasted urban arterial travel time.

Not only SSNN, other ANN techniques have been applied to forecast travel time and different other traffic characteristics. For example, Park et al. (1998) applied Modular ANN to forecast travel time for every unsupervised cluster based on historical data of each segment. This method was then compared with other ANN methods. It was observed that the modular ANN (MNN) outperformed the conventional singular ANN, Kalman filtering model, exponential smoothing model, historical profile, and real-time profile. Cho et al. (2003) used MNN to forecast the arrival times of trains at highway-railroad grade crossings. Independent variable was train speed profile and the dependent variable was arrival time at highway-railroad grade crossing. Four models were developed based on the data input and prediction interval. Approximately, 499 trains were used for training and 183 trains were used for testing the MNN. It was found that modular architecture gave better results than simple ANN model and standard regression techniques.

Another ANN model is Spectral basis ANN that has been successfully implemented to forecast different traffic characteristics. Park et al. (1999) employed a spectral basis artificial neural network (SNN) which adopted sinusoidal transformation technique to increase the linear separability of the input features to forecast link travel time using real-information. SNN outperformed Kalman filtering model, exponential smoothing model, historical profile, and real time profile and gave similar results to that of modular ANN. However, SNN took less effort than MNN. Rilett et al. (2001) also adopted SNN to forecast freeway corridor-level travel time following a two-step approach and using link-level travel time. They reported that results were better than that of other model based on real-time or historical corridor profile.

BPNN has also been adopted to forecast different traffic characteristics. Hu et al. (2009) worked on a predicting model and data acquisition plan based on BPNN. The drawback of BPNN is that it easily falls into the local convergence. Therefore, they introduced Genetic Algorithms to improve the traditional BPNN using its global search capability. Results showed that GA-BPNN showed better convergence speed and predicting accuracy. Lee et al. (2009) developed a novel travel time forecasting model using ANN with cluster method. The cluster method was employed to reduce the number of input variables. Results showed that the mean absolute percentage error (MAPE) of the predicted travel time was less than 22%, indicating a good forecasting performance.

ANN was also applied to predict travel time in the presence of an incident. Ran et al. (2006) applied neural network model for corridor-level travel time prediction in the presence of traffic incidents. Three different scenarios were considered based on the types of the input variables: (1) incident related information only; (2) current traffic condition information only, (3) both the incident related information and the current traffic condition information. Results demonstrate that third scenario delivered a successful model. Zeng et al. (2009) predicted corridor travel time under incident conditions by adopting two steps: firstly, predicting of the segment travel time, and, secondly, aggregating corridor-level travel time of the predicted segmental results with and without incident. To address the dynamic nature of traffic system, Time Delay Neural Network (TDNN), SSNN, an Extended SSNN (ExtSSNN), and a traditional BPNN were tested alongside the influence of incident to predict corridor-level travel time. The empirical results showed that the SSNN and ExtSSNN outperformed other models. Likewise, He et al. (2010) implemented BPNN to predict travel time during incident to assist with incident management.

Overall, past research shows that ANN has been applied successfully through different network architectures (ANN, SSNN, SNN, TDNN, MNN, etc.), data sources (GPS, simulated travel time, etc.), different types of inputs (speed, travel time, real-time data, historical data, occupancy, upstream and downstream traffic data, incident information, etc.) for both static and dynamic network conditions.

## 2.3 Performance Evaluation of Time Series Methods in Forecasting Traffic Parameters

A time series is defined as the sequence of observations or points over time. The trend of a time series is easily understood through delineation of a scatter plot. Time series can be either continuous or discrete. Observations are available at every instant in case of continuous time series, while discrete time series is the one where there are some intervals among the observations (Easton et al. 1997).

Different time series methods were developed and adopted by researchers in the past. Billings et al. (2006) applied ARIMA model to predict arterial travel time using GPS probe vehicle data. The Yule-Walker method, the Levison-Durbin algorithm, the Burg's algorithm, the Innovations algorithm, and the Hannan-Rissanen procedure were used to estimate the ARIMA parameters. Models for each segment were validated using both the residual analysis and Portmanteau lack-of-fit test. The findings indicate reasonable accuracy in predicting section-level travel time. Hu et al. (2010) also used time series model for prediction of travel time for urban network beside of a simulation-based model. The simulation-based model, DynaTAIWAN was used to estimate travel time using traffic flow patterns, O-D flows, network, alternative routes and other traffic conditions. ARIMA was integrated with signal delay to predict travel time for the arterial street. Numerical results indicated that both proposed simulation-based travel time prediction model and the ARIMA model provided reasonable travel time information.

Yang et al. (2014) used ARIMA model alongside pre-whitening cross-correlation function (CCF) and observed spatio-temporal relationships between data of consecutive segments. They found that traffic condition is highly correlated over space and time.

Traffic characteristics other than travel time had been successfully forecasted using ARIMA model in the past. Lee et al. (1999) implemented ARIMA model for forecasting traffic volume and found that a subset ARIMA model outperformed full ARIMA model. Rashidi et al. (2013) conducted study on short-term prediction of bus dwell time using ARIMA based on historical automatic vehicle location (AVL) data and found that Wakeby distribution outperformed the lognormal distribution for both peak and off-peak periods.

Time series model was also developed to forecast arrival time of buses. Suwardo et al. (2009) applied ARIMA model to predict bus travel time based on historical travel time series data. Historical data set was collected from the bus service operated on a divided 4-lane 2-way highway in Malaysia. The analysis of both directions was performed separately. Their results showed that the developed models clearly fit with the observed values for both directions. Like regression model, ARIMA does not require the factors affecting bus travel time such as delay at link, bus stop, and intersection for forecasting.

Not only ARIMA, other time series models have been successfully used to forecast different traffic characteristics alongside travel time. Guin et al. (2006) presented a Seasonal Autoregressive Integrated Moving Average (SARIMA) time series based approach for predicting future travel times using historical point detection travel time data. Empirical testing of the model was performed using data obtained from video detection systems in Atlanta, Georgia. Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were used for selecting the orders of SARIMA models. However, the quality of the predictions for the random walk model degrades rapidly for a 4-step ahead forecast. Marković et al (2007) developed a SARIMA model to predict the travel time on an urban road in Zagreb, Croatia using GPS data. Khoei et al. (2013) used Bluetooth data for short-term travel time prediction. The SARMIA model accounted the seasonality coefficient considering previous lags of the same day and the values from the same time of previous days. The results successfully validated different prediction horizons (5 min to 90 minutes). However, the random walk model and seasonality may not be applicable to all contexts.

Not only travel time, SARIMA was also applied for traffic flow forecasting in the past. Xia et al. (2006) developed a methodology for dynamic travel time prediction based on continuous data generated by single-loop detectors (and similar devices) and incident reports generated by the traffic monitoring system. Their method involved multiple-step-ahead prediction for flow rate and occupancy in real time. SARIMA model was developed with an embedded adaptive predictor. This predictor adjusted the prediction error based on traffic data that became available every five minutes at each station. The impact of incidents was evaluated based on estimates of incident duration and the queue incurred. Tests and comparative analyses showed that this method was able to capture the real-time characteristics of the traffic and provide more accurate travel time estimates particularly when incidents occur.

Williams et al. (2003) compared SARIMA, random walk, historical average, and combination of random walk and historical average for traffic flow prediction. Their results indicate that one-step SARIMA predictions consistently outperformed other heuristic forecast benchmarks. Ghosh et al. (2004) compared three different time-series models, viz. random walk model, Holt-Winters' exponential smoothing technique and SARIMA model for predicting traffic flow at an arterial intersection in a congested urban transportation network in the city of Dublin, Ireland. The data used for modeling was obtained from loop-detectors at a certain junction in the city center of Dublin. SARIMA and Holt-Winters' exponential smoothing technique gave highly competitive forecasts and matched the traffic flow data during rush hours.

Other time series model such as Space-Time Autoregressive Integrated Moving Average (STARIMA), Autoregressive Integrated Moving Average with Generalized Autoregressive Conditional Heteroscedasticity (ARIMA-GARCH), Kohonen-enhanced ARIMA (KARIMA), etc. were also used to model traffic parameters in the past. Khan et al. (2012) conducted a study on predicting travel time in an urban traffic scenario for Sydney region, Australia as well as large section of Sydney's urban landscape. The model was built using Quadstone Paramics simulator. The baseline demand model was calibrated based on load shift parameter in the pre-defined demand matrix. Analysis was performed for six level of service (LOS) settings and accuracy was reported for prediction times of several tens of minutes. Though the model performed well in the steady state case, the basic approach is not suitable for modeling the urban traffic setting. Chen et al. (2011) proposed ARIMA-GARCH model for short-term traffic flow prediction. The model combined linear ARIMA model with non-linear GARCH model, to capture both the conditional mean and conditional heteroscedasticity of traffic flow series. The model was developed based on PeMS single loop detector data. The performance of the hybrid model was compared with that of standard ARIMA model. The results showed that the conditional heteroscedasticity could not bring satisfactory improvement to prediction accuracy and in some cases; GARCH (1, 1) model even showed deteriorated performance.

Van Der Voort et al. (1995) introduced KARIMA method for short-term traffic flow forecasting. The technique used a Kohonen self-organizing map as an initial classifier, which was associated with individual ARIMA model. The model successfully demonstrated traffic flow forecasting of French motorway for prediction horizons of half an hour and an hour.

Time series model was also integrated with other models such as ANN and fuzzy logic. Through hybridization, the time series was decomposed into linear and non-linear components, which can deal with single model as well as multiple models. Therefore, novel hybridization of ANN and ARIMA model can overcome limitation of ANN and yield a more accurate forecasting model than traditional ARIMA models. Moreover, theoretical and empirical findings have indicated that integration of different models can be an effective way of improving upon their predictive performance, especially when the models in combination are quite different. Ho et al. (2002) investigated a comparative study of the Box-Jenkins ARIMA model and the ANN model in time series prediction. The ANN architecture comprised of Multi-Layer Feed-Forward Network and the Recurrent Neural Network (RNN). Simulation results showed that both the ARIMA and the RNN model outperformed the feed-forward model in short-term forecasting. Zeng et al. (2008) also developed a hybrid-predicting model for short-term traffic flow that combined both ARIMA and multilayer artificial neural network (MLANN). ARIMA is suitable for linear prediction and MLFNN is suitable for non-linear prediction. Their experimental results with real data sets indicated that the combined model could outperform the other models in forecasting accuracy. Tan et al. (2008) proposed an aggregation approach for traffic flow prediction that was based on the moving average, exponential smoothing, ARIMA, and neural network models. The data source was the weekly traffic flow series, a daily traffic flow, and an hourly traffic flow series. The predictions that resulted from the moving average, exponential smoothing, and ARIMA models were used as the input for ANN. The output traffic flow of the aggregated model then compared with the naïve, ARIMA, non-parametric regression, and ANN model. The outcome revealed that the data aggregation model could forecast more accurately than any individual model considered in their research. He-Jiuran et al. (2013) proposed a method to predict passenger flow using ARIMA model and Radial Basis Function (RBF) neural network, which is suitable for processing non-linear problem. The ARIMA - RBF model analyzed passenger flow's temporal characteristics. The proposed model was used to forecast passenger flow of rail transit in Beijing, China. Results showed that the ARIMA - RBF model's average daily forecast error was 2%, which was less than ARIMA and RBF itself. Overall, past research showed that time series model could forecast accurately the traffic parameters when applied alone or along with other models.

## 2.4 Performance Evaluation of Different Other Methodologies in Forecasting Traffic Parameters

Besides the methodologies discussed in the previous sections, other methodologies were used in forecasting traffic characteristics. For instance, Messer et al. (1973) developed a method for predicting the travel time required to traverse a freeway segment that is experiencing incident congestion. The model was developed based on kinematic wave theory of Lighthill and Whitham. In their research, the model was deterministic. The study presented speeds of the various shock waves and travel-time. This model could also be used to predict queue backups and delays due to lane closures caused by scheduled maintenance operations. Hobeika et al. (1994) predicted short-term traffic volume based on current traffic, historical average, and upstream traffic. Three models were developed for prediction: a combination of historical average and upstream traffic, a combination of current traffic and upstream traffic, and a combination of all three variables. The three models were evaluated using 15-minute freeway data from induction loop detectors. The third model surpassed other models in producing accurate forecasts under congested traffic condition in ranges of 30- to 45-minutes. Roozemond et al. (1997) also used historic data and current data to develop a dynamic model for forecasting travel times of links. They reported that future forecasting could be done using both current data as well as travel times for some time ahead based on current and historic data. However, historic data based prediction models give the current and future travel time based on historical travel time of an assumed stationary condition. Therefore, historical mean based prediction model is more reliable where the traffic pattern in that area of interest is relatively stable with not much disruption from congestion. Sun et al. (2003) applied local linear regression model for short-term prediction of traffic flow using historical and current traffic flow data as input variable. The performance of the proposed model was compared with the previous results of non-parametric approaches such as KNN and kernel methods and results showed that proposed method outdone both KNN and the kernel smoothing methods. Kalman Filtering was also used to forecast traffic characteristics. Barcelo et al. (2010) adopted ad hoc procedures based on Kalman Filtering and reported that it could predict travel time successfully. Zhang et al. (2011) applied adaptive model based on the Kalman Filter Model (AKFM) and Classical Kalman Filter Model (CKFM) for short-term traffic flow forecasting. Results obtained showed that the stability and prediction capability of AKFM is better than CKFM.

## 2.5 Previous Studies on Quantifying the Effect of Incident

Researchers in the past have worked on quantifying the effect of incident on the transportation system. Koutsopoulos et al. (1991) presented a theoretical link travel-time estimation model in which incident delay was estimated using a deterministic model. However, in their study the incident and its attributes (reduced capacity and incident duration) are randomly generated. Al-Deek et al. (1991) evaluated the benefits of a route guidance system specifically in the case of incident congestion. In their model, the incident situation also was assumed deterministic, and a deterministic queuing model was then used to estimate the queuing delay.

Fu et al. (1995) developed a model to predict travel time in an urban traffic environment considering dynamic and stochastic nature of the traffic during incident conditions. In contrast to traditional deterministic delay models of incident, the model presented explicitly considered the stochastic attributes of incident duration. This model predicted the delay that a vehicle would experience as it traveled through non-recurring congestion brought about by an incident. A mixed discrete and continuous vehicle-delay model was first derived and estimators of the mean and variance of vehicle delay were identified. Results showed that the deterministic model might over-estimate or under-estimate the expected incident delay, depending on the arrival of the vehicle to the incident location, and under certain circumstances the incident delay showed a high variance even when the expected delay was low.

Wang et al. (2008) developed a new algorithm, based on a modified deterministic queuing theory, for quantifying incident-induced delays (IID) on freeways. The algorithm used incident log data and loop detector data as inputs. The incident log data was obtained from the Washington Incident Tracking System (WITS) database and the archived loop data used in their research was downloaded from the Traffic Data Acquisition and Distribution (TDAD) website. Identifiable errors in both incident and loop data were either eliminated or corrected before the data was transferred to the Incident Study (IS) database. The proposed algorithm was implemented in the Advanced Roadway Incident Analyzer (ARIA) system. To verify the accuracy and validity of the algorithm, a microscopic simulation model for the Evergreen Point Bridge on State Route (SR) 520 in Washington State was developed using VISSIM traffic simulation software. Corthout et al. (2010) used Marginal Incident Computation (MIC) method to quantify the congestion spillback. This model was combined with Dynamic Network Loading (DNL) of Dynamic Traffic Assignment (DTA) to capture the variability of capacity and demand fluctuations that lead to a variation of travel time. MIC was applied to each node for tracing the congestion spillback as well as to reduce computational speed in case of a large network.

## 2.6 Literature Review on Factors Affecting Traffic Characteristics

Travel time is the outcome of dynamic and non-linear traffic system, which varies with respect to space and time. The complicated relationship between travel time and other traffic variables are difficult to quantify mathematically and very much depended on the assumptions. Therefore, past researchers worked on defining and examining the relationships among various variables affecting travel time. The journey speed along an arterial street depends not only on the arterial road geometry but also on the traffic flow characteristics and traffic signal coordination (Lum et al. 1998). Likewise, a few other factors that have an effect on travel time are listed below.

1. Holiday and special incidents (Karl et al., 1999); that contribute to an increase in travel time.
2. Signal delay (Wu, 2001); which affects the flow on arterial and local streets.
3. Weather conditions (Chien et al. 2003); traffic flow can be severely disrupted under inclement weather conditions.
4. Congestion level (Richardson 2004); which varies by time-of-the-day.
5. Forecasting period (Kisgyorgy et al. 2002); higher the forecasting horizon, higher the prediction error.

Vythoulkas (1993) mentioned that traffic forecasting accuracy is related to the use of neighboring location traffic information. Chandra et al. (2009) assumed that the effect of upstream and downstream traffic is symmetrical. However, Yang et al. (2014) showed that upstream and downstream segment does not have equal impact on a target segment.

## 2.7 Performance Evaluation of Different Empirical Methods to Capture the Expected Travel Time

This part of the literature review focuses on past research to estimate the expected travel time. As the expected travel time gives an idea about the probable time to reach the destination, it plays a significant role for travelers in planning for their trip. Moreover, different studies have defined the expected travel time in different ways. Therefore, this definition needs special context to avoid possible misunderstanding.

Florida's Reliability Method (FDOT 2000) defined the expected travel time as the median of all the travel time samples of a study period rather than the mean. This was done to avoid the influence of a major incident. FDOT defined reliability as the percentage of trips that does not cross the summation of expected travel time and percentage of the expected travel time (acceptable additional time) of that study period. A study was conducted along I-40 in the Orlando, FL area for 6 months and 5-20% was considered as acceptable additional time. Results showed that when acceptable additional time was accounted as 20%, the expected time reliability is around 85% i.e., for 85% of the trips, time travel is within the range.

The expected travel time was estimated for a given route for various time intervals during which the demand is relatively constant for a given route. Van Amelsfort et al. (2005) considered the average of minimum and maximum travel time as the expected travel time. Yeon et al. (2008) defined the expected travel time for durations during which the demand is relatively constant along a corridor. Sun et al. (2010) defined ensemble mean travel time over a number of days as expected travel time. Spot speed data was collected from point detectors that were used to calculate ensemble mean speed and ensemble variance averaging all the spot speed data from a point detector for a specific time interval and for multiple days. Ensemble mean speed and moment approximation method was then used to plot piecewise constant speed trajectory, from which ensemble mean travel time was calculated from a specific departure time.

The expected travel time was also defined as travel time on a degraded link to be lesser than the free flow travel time for the link with a specific tolerance level to estimate travel time reliability and capacity reliability (Al-Deek et al. 2006). Puvvala et al. (2015) measured the most expected travel time based on Cronbach's α or Cronbach's coefficient. Their basic objective was to find the most reliable travel time for a time period based on the internal consistency of the travel time. Eight different scenarios were considered. Factors considered to examine the consistency included day-of-the-week, weekend/weekday, time-of-the-day, 85$^{th}$ percentile of travel time, and average travel time. Observations showed that a majority of the trips have a higher value of Cronbach's α when the average travel time was taken into consideration instead of the 85th percentile travel time.

## 2.8 Summary of Findings from Literature Review

Different methodologies and their effectiveness in forecasting traffic parameters including travel time both for with "vehicle accident" and without incident scenario was discussed in the previous sections. Previously, to evaluate the effect of the incident, most of the studies used delay as a measure of effectiveness. It was calculated using either queuing theory or shock wave theory or simulation software (Al-Deek et al. 1991, Fu et al. 1995). Besides, most of these studies were deterministic in nature. Moreover, it requires having vehicle arrival information to calculate delay. This study aims to define the effect of an incident as a function of travel time instead of number of vehicle arrivals and delay.

This chapter also shows that, not many studies were conducted on forecasting traffic characteristics in the presence of accident over time and space and using time series model. Moreover, to compute the expected travel time, this study focuses on the similarity of time series pattern or consistency in the travel times.

# CHAPTER 3: BACKGROUND ON CRONBACH'S α AND ARIMA MODELS

Developing a process and identifying an effective method is vital to estimate the expected travel time and to forecast travel time and travel time variation with or without "vehicle accident". This chapter introduces Cronbach's α to estimate the expected travel time and ARIMA model to forecast travel time and its variation. The background discussion related to Cronbach's α and ARIMA are explained based on content presented by Puvvala (2014), Puvvala et al. (2015), Box et al. (2008), and Montgomery et al. (2008).

## 3.1 Cronbach's $\alpha$

Cronbach's $\alpha$ is used as a measure of internal consistency or reliability of a test/ survey/ questionnaire. It was named after Lee Cronbach in 1951 (Cronbach 1951). The principle of Cronbach's $\alpha$ is defined from the basic classical test theory where the reliability of test scores is expressed as the ratio of the true-score variance and the observed-score variance. The observed-score variance is the summation of the true-score variance and the measurement error (Harvill 1991). For example, if a student scores 86% in a test where he knows 80% of the questions of that test, then 86 is the observed score and true score is 80. The additional 6% is because of conjecture, which is defined as the measurement error or the unreliability of the test.

For the aforementioned problem, Cronbach's α is computed using the following expression (Cronbach 1951).

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K} \sigma_{Y_i}^2}{\sigma_X^2}\right) or \ \frac{K}{K-1}\left(1 - \frac{V1}{V2}\right) \quad (3.13)$$

$$V1 = \sum_{i=1}^{K} \sigma_{Y_i}^2 \, ; \, V2 = \sigma_X^2$$

Here, K is the number of questions,

$\sigma_X^2$ is the variance of the observed total test scores of a person, and,

$\sigma_{Y_i}^2$ is the variance of the sums of scores of a question for all the five persons.

If the reliability score is high, the measurement error would be low as well as it would be less correlated with the true score. In other words, the correlation between true score and observed score would be high if test reliability is high. Several assumptions are made in estimating Cronbach's $\alpha$.

Questionnaire in table 1 can provide an idea about Cronbach's $\alpha$. Consider a case where one needs to determine the reliability of questionnaire, which consists of three questions to measure the analytical ability of five persons with various educational levels. The test is prepared to rate the person's ability to analyze a given dataset and to test the validity of the questionnaire. It is assumed that ability correlated with education level. The test results are also shown in Table 1, as binary variables (Puvvala, 2014).

TABLE 1: Summary of results from test scores

| Students | Questions | | | Total |
|---|---|---|---|---|
| | Q1 | Q2 | Q3 | |
| 1 | 1 | 1 | 1 | 3 |
| 2 | 0 | 0 | 1 | 1 |
| 3 | 0 | 0 | 0 | 0 |
| 4 | 1 | 0 | 0 | 1 |
| 5 | 1 | 0 | 0 | 1 |
| Item Variances | 0.24 | 0.16 | 0.24 | |
| V2 | | | | 0.96 |
| V1 | | | | 0.64 |

From Table 1,

Sum of individual variances (V1) = 0.24 + 0.16 + 0.24 = 0.64

Variance of the total scores (V2) = 0.96

Number of questions (items) = 3

Based on K and computed V1 and V2 from Table 2,

$$\alpha = \frac{3}{3-1}\left(1 - \frac{0.64}{0.96}\right)$$

$$=> \ \alpha = 0.50$$

A ‘zero’ value of Cronbach’s α indicates that the questions does not measure the analytical ability of a person or questionnaire is not capable to capture the ability of a person. On the other hand, if Cronbach’s α is ‘one’, it indicates that all the questions designed did a perfect job for make questionnaire valid. It only happens when the scores of a student remain same for all questions making him score three. The computed Cronbach’s α in the above example is 0.5, indicating that the questions are reliable in estimating the analytical ability of the person.

In the above example, the persons are the primary source of variance while questions are the secondary source of variance. In this research, time-of-the-day, day-of-the-week and week-of-the-year are considered as sources of variance for both primary and secondary factor. Considering one factor at a time, multiple Cronbach's α are evaluated once with time-of-the-day as primary factor and next with day-of-the-week or week-of-the-year as primary factor (Puvvala et al., 2015). In general, the primary factor causes the changes in the observations and correlation is evaluated through the secondary factor.

In summary, Cronbach's α measures the correlation or similarity of entities from various items i.e., the correlation among the columns of dataset.

3.2 Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF)

The use of ACF and PACF is an imperative part of the Box–Jenkins approach for modeling time series through identifying appropriate parameters of ARIMA (*p, d, q*) model. ACF defines the linear dependency between variables with respect to two points of time (Box et al. 2008). ACF actually tells how much correlation there is between two neighboring data points in a time series or between two time series. Theoretically, the ACF of lag *k* is expressed as follows (Box et al. 2008)

$$\rho_y(k) = \frac{E[(y_t - \mu_y)(y_{t+k} - \mu_y)]}{\sqrt{E[(y_t - \mu_y)^2]E[(y_{t+k} - \mu_y)^2]}}$$
$$= \frac{Cov(y_t, y_{t+k})}{\sigma_t \sigma_{t+k}}$$

For a stationary time series, the variance at time $t$ is the same as the variance at time $t + k$. Therefore, the denominator is just the variance of $y_t$ (Box et al. 2008)

$$\rho_y(k) = \frac{Cov(y_t, y_{t-k})}{Var(y_t)} \qquad (3.1)$$

ACF is very useful as it provides a partial description of the ARIMA (*p, d, q*) model. This function identifies the appropriate moving average (MA) model and its order as it cuts off after lag *q* (Box et al. 2008).

The PACF also plays an important role in time series data analyses by identifying the appropriate lag in an autoregressive (AR) model as it cuts off after lag *p* (Box et al. 2008). In general, PACF is a conditional correlation, which defines the correlation between one response variable and one of the predictor variables considering other predictor variables. For instance, partial auto-correlation between $y_t$ and $x_{t-k}$ is defined as the conditional correlation between $y_t$ and $x_{t-k}$, accounting other predictor variables in between two time points *t* and *t-k* (Box et al. 2008). PACF correlates the residuals from two regressions: $y_t$ is predicted from $x_{t-k+1}$ to $x_{t-1}$ in the first regression, while $x_{t-k}$ is predicted from $x_{t-k+1}$ up to $x_{t-1}$ in the second regression (Box et al. 2008). The $k^{th}$ order (lag) PACF is expressed as follows (Box et al. 2008).

$$\rho_y(k) = \frac{Cov(y_t, x_{t-k} | x_{t-k+1}, x_{t-k+2}, \cdots x_{t-1})}{\sqrt{\left(Var\,(y_t | x_{t-k+1}) \ldots\ldots Var\,(y_t | x_{t-1})\right)\left(Var(x_{t-k} | x_{t-k+1}) \ldots\ldots Var\,(x_{t-k} | x_{t-1})\right)}} \qquad (3.2)$$

3.3 Autoregressive (AR) Model

In an AR model, the variable of interest is a linear combination of past values of the variable of the same time series or different time series (Box et al. 2008). The AR model of order *p,* AR *(p)* is expressed as follows (Box et al. 2008).

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + e_t \qquad (3.3)$$

The above equation is like multiple regressions with $y_t$ as response variable and lagged values of $y_t$ as predictor variables (Box et al. 2008). Here, $\delta$ is a constant and $\varepsilon_t$ is white noise. Using the backward shift operator, the AR (*p*) is expressed as follows (Box et al. 2008).

$$y_t = \delta + (1 + \phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p) y_t + e_t$$

Where, $B^n y_t = y_{t-n}$. *B* operator shifts the data backward to period *n*. For example, $B^2$ operator shifts the data backward two periods. *B* operator can be replaced by *L* (lag) operator. Using eq$^n$. 3.3, the first order AR, AR (1), is expressed as follows (Box et al. 2008).

$$y_t = \delta + \phi_1 y_{t-1} + e_t$$

Likewise, the second order AR model, AR (2), is expressed as follows (Box et al. 2008).

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$$

3.4 Moving Average (MA) Model

Alike AR, MA model is also a linear regression model, which uses past errors as predictor variables (Box et al. 2008). In this process of order *q*, each observation ($y_t$) is formed by a weighted average of random disturbances going back by a number of periods equal to the order *q*. It is denoted as MA (*q)* and expressed as follows (Box et al. 2008).

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \qquad (3.4)$$

Where, $\varepsilon_t \approx iid\ \mathrm{N}(0, \sigma_\varepsilon^2)$ implying that $\varepsilon_t$ are identically and independently distributed as well as normally distributed with “0” mean and constant variance. All the variables are uncorrelated e.g., *Cov* ($\varepsilon_t$, $\varepsilon_{t-1}$) = 0 and so as with response variable i.e., *Cov*

($\varepsilon_t$, $y_t$) = 0. Therefore, $\varepsilon_t$ is defined as white noise. Parameters $\theta_1$, $\theta_2$… $\theta_k$ can be positive or negative. The change in parameters results different patterns of time series. The first order MA model, MA (*1*), is expressed as follows (Box et al. 2008).

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

The second order MA model, MA (*2*), is expressed as follows (Box et al. 2008).

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

Using the backward shift operator (*B*), the MA (*q*) process is expressed as follows (Box et al. 2008).

$$y_t = \mu + (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t$$

3.5 Theoretical Auto-correlation Function (ACF) for Autoregressive (AR) Model

The generalized equation of ACF of AR model for different lag *k* is presented in this section. The $p^{\text{th}}$ order AR model, AR (*p*), is shown in equation 4.3. The covariance for lag *k* is expressed as follows (Box et al. 2008).

$$\begin{aligned}
&\gamma(k) \\
&= Cov(y_t, y_{t-k}) \\
&= Cov(\delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, y_{t-k}) \\
&= \sum_{i=1}^{p} \phi_i Cov(y_{t-i}, y_{t-k}) + Cov(\varepsilon_t, y_{t-k}) \\
&= \sum_{i=1}^{p} \phi_i \gamma(k-i) + E(\varepsilon_t(\mu + \varepsilon_{t-k} + \theta_1 \varepsilon_{t-k-1} + \theta_2 \varepsilon_{t-k-2} + \cdots)) \\
&= \sum_{i=1}^{p} \phi_i \gamma(k-i) + \begin{cases} \sigma_\varepsilon^2 & \text{when } k = 0 \\ 0 & \text{when } k > 0 \end{cases}
\end{aligned}$$

When $k = 0$,

$$\gamma(0)$$
$$= \sum_{i=1}^{p} \phi_i \gamma(-i) + \sigma_\varepsilon^2$$
$$= \gamma(0)\sum_{i=1}^{p} \phi_i \gamma(i)/\gamma(0) + \sigma_\varepsilon^2$$
$$= \gamma(0)\sum_{i=1}^{p} \phi_i \rho(i) + \sigma_\varepsilon^2$$
$$= \sigma_\varepsilon^2 \Big/ \left[1 - \sum_{i=1}^{p} \phi_i \rho(i)\right]$$

When $k > 0$,

$$\gamma(k) = \sum_{i=1}^{p} \phi_i \gamma(k-i)$$

Adopting eq$^n$. 3.1, ACF of lag $k$ for AR ($p$) is expressed as follows (Montgomery et al. 2008).

$$\gamma(k)/\gamma(0) = \sum_{i=1}^{p} \phi_i \gamma(k-i)/\gamma(0)$$
$$\rho(k) = \sum_{i=1}^{p} \phi_i \rho(k-i) \qquad (3.5)$$

The ACF for the first order AR model, denoted as AR (1), for different lags is expressed as follows (Montgomery et al. 2008).

For $k = 0$,

$$\rho(1) = \gamma(0)\big/\gamma(0) = 1 = \phi_1^0$$

For $k = 1$,

$$\rho(1) = \sum_{i=1}^{1} \phi_i \rho(1-i) = \sum_{i=1}^{1} \phi_1 \rho(1-1) = \phi_1 \rho(0) = \phi_1 * 1 = \phi_1^1$$

For $k = 2$,

$$\rho(1)=\sum_{i=1}^{1}\phi_1\rho(2-i)=\phi_1\rho(2-1)=\phi_1\rho(1)=\phi_1*\phi_1=\phi_1^2$$

$\vdots$

$\vdots$

For $k=k$,

$$\rho(k)=\sum_{i=1}^{k}\phi_1\rho(k-i)=\phi_1^2\sum_{i=2}^{k}\rho(k-i)=\phi_1^3\sum_{i=3}^{k}\rho(k-i)=\cdots=\phi_1^k\sum_{i=k}^{k}\rho(k-k)=\phi_1^k\rho(0)=\phi_1^k$$

Therefore, the generalized equation of ACF for AR (1) for any lag $k$ = 0, 1, 2, 3… $k$ is expressed as (Montgomery et al. 2008):

$$\rho(k)=\phi_1^k \qquad (3.6)$$

From eq[n]. 3.6, it is evident that the ACF decays exponentially with an increase of lag $k$. To be stationary, the roots of $\phi$ ($B$) = (1- $\phi_1$B) = 0 must lie outside the unit circle i.e., $|\phi_1| < 1$. If $0 < \phi_1 < 1$, all auto-correlations are positive and if $-1 < \phi_1 < 0$, the sign of auto-correlations will show alternating pattern with a negative value at start (Box et al. 2008) . AR (1) process is always invertible.

The ACF for the second order AR model, denoted as AR (2), for different lags is expressed as follows (Montgomery et al 2008).

For $k$ = 1,

$$\rho(1)=\sum_{i=1}^{2}\phi_i\rho(1-i)=\phi_1\rho(1-1)+\phi_2\rho(1-2)=\phi_1\rho(0)+\phi_2\rho(-1)=\phi_1*1+\phi_2\rho(1)$$
$$=\phi_1/(1-\phi_2)$$

For $k$ = 2,

$$\rho(2)=\sum_{i=1}^{2}\phi_i\rho(2-i)=\phi_1\rho(2-1)+\phi_2\rho(2-2)=\phi_1\rho(1)+\phi_2\rho(0)=\phi_1^2/(1-\phi_2)+\phi_2$$

For $k$ = 3,

$$\rho(3) = \sum_{i=1}^{2} \phi_i \rho(3-i) = \phi_1 \rho(3-1) + \phi_2 \rho(3-2) = \phi_1 \rho(2) + \phi_2 \rho(1)$$
$$= \phi_1^3 / (1-\phi_2) + \phi_1 \phi_2 + \phi_1 \phi_2 / (1-\phi_2)$$

⋮

⋮

For $k = k$,

$$\rho(k) = \sum_{i=1}^{2} \phi_i \rho(k-i) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2)$$

Therefore, the generalized equation of ACF of AR (2) is as expressed as follows (Montgomery et al. 2008).

$$\rho(k) = \begin{cases} \phi_1/(1-\phi_2) & \text{When } k = 1 \\ \phi_1 \rho(k-1) + \phi_2 \rho(k-2) & \text{When } k = 2, 3, ..., k \end{cases} \qquad (3.7)$$

ACF tail for AR (2) also decays exponentially (when $0 < \phi_1$ and $0 < \phi_2$ or when $0 > \phi_1$ and $0 < \phi_2$) alike AR (1) for the real roots or shows damped sign wave for complex roots. For $0 > \phi_1$ and $0 < \phi_2$, ACF shows exponential decay with alternating sign starting with a negative value. AR (2) process is always in the inverted form and to be stationary the roots of $\phi$ ($B$) = (1- $\phi_1 B$- $\phi_2^2 B$) = 0 must lie outside the unit circle (Box et al. 2008). To be stationary, the conditions of AR (2) are shown as follows (Montgomery et al. 2008).

$$\phi_1 + \phi_2 < 1$$
$$\phi_2 - \phi_1 < 1$$
$$-1 < \phi_2 < 1$$

3.6 Theoretical Auto-correlation Function (ACF) of Moving Average (MA) Model

The generalized equation of ACF of MA model for different lag $k$ is presented in this section. The equation of the ACF for $q^{th}$ order MA model, MA ($q$), is shown in section 3.3. If lag $k$ is less than or equal to order of the MA model (i.e., $k \leq q$) then the ACF model is expressed as follows based on equation 3.1 (Montgomery et al. 2008).

$$\rho_y(k)$$
$$= \frac{\gamma_y(k)}{\gamma_y(0)}$$
$$= \frac{Cov(y_t, y_{t-k})}{Var(y_t)}$$
$$= \frac{Cov(\mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_k\varepsilon_{t-k} - \theta_{k+1}\varepsilon_{t-k-1} - \cdots - \theta_q\varepsilon_{t-q}, \mu + \varepsilon_{t-k} - \theta_1\varepsilon_{t-k-1} - \cdots - \theta_q\varepsilon_{t-k-q})}{Var(\mu - \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q})}$$
$$= \frac{E[(\varepsilon_t - \cdots - \theta_k\varepsilon_{t-k} - \theta_{k+1}\varepsilon_{t-k-1} - \cdots - \theta_q\varepsilon_{t-q})(\varepsilon_{t-k} - \theta_1\varepsilon_{t-k-1} - \cdots - \theta_{q-k}\varepsilon_{t-q} - \cdots - \theta_q\varepsilon_{t-k-q})]}{Var(\varepsilon_t) - Var(\theta_1\varepsilon_{t-1}) - \cdots - Var(\theta_q\varepsilon_{t-q})}$$
$$[\text{As } Cov(\varepsilon_t, \varepsilon_{t-n}) = E(\varepsilon_t\varepsilon_{t-n})]$$
$$= \frac{(-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q)\sigma_\varepsilon^2}{(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma_\varepsilon^2}$$
$$[\text{As } E(\varepsilon_t\varepsilon_{t-n}) = 0 \text{ and } E(\varepsilon_{t-n}\varepsilon_{t-n}) = \sigma_\varepsilon^2]$$
$$= \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2} \qquad (3.8)$$

If lag $k$ is greater than the order of the MA model (i.e., $k > q$) then last weighted error or (t-q)$^{th}$ error term of the $y_t$ is $\varepsilon_{t-q}$, which will not even match with the first weighted error of $y_{t-k}$ (Montgomery et al. 2008). Therefore, the numerator of equation is zero implying that ACF is zero. For example, ACF for first order MA of different lag is expressed as follows.

$$\rho_y(1) = \begin{cases} \dfrac{-\theta_1}{1+\theta_1^2} & \text{When } k = 1 \\ 0 & \text{When } k > 1 \end{cases}$$

From the values of $\rho_y$ (1), it is evident that auto-correlation cuts off after lag 1 for MA (1). If $\theta_1$ is negative then $\rho_y$ (1) is positive and if $\theta_1$ is positive then $\rho_y$ (1) is negative. MA (1) process is always stationary and invertible for $|\theta_i| < 1$ (Box et al. 2008). For second order, MA, denoted as MA (2), for different lags is as expressed as follows.

$$\rho_y(2) = \begin{cases} \dfrac{-\theta_1 + \theta_1\theta_2}{1+\theta_1^2+\theta_2^2} & \text{when } k = 1 \\ \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2} & \text{when } k = 2 \\ 0 & \text{when } k > 2 \end{cases}$$

From the values of $\rho_y$ (2), it can be said that auto-correlation cuts off after lag 2 for MA (2). Like MA (1), MA (2) is always stationary. If $\theta_1 > 0$ and $\theta_2 > 0$, the value of lag 1 is either positive or negative while the value of lag 2 is negative. If $\theta_1 < 0$ and $\theta_2 > 0$, the value of lag 1 is positive but the value of lag 2 is negative (Box et al. 2008). MA (2) process is invertible for the following conditions.

$$\theta_1 + \theta_2 < 1$$
$$\theta_2 - \theta_1 < 1$$
$$-1 < \theta_2 < 1$$

3.7 Theoretical Partial Auto-correlation Function (PACF) of the Autoregressive (AR) Model

Following eq$^n$. 4.3, the new regression equation for $y_{t-k}$ is expressed as follows.

$$y_{t-k} = \phi_{k1} y_{t-k+1} + \phi_{k2} y_{t-k+2} + \cdots + \phi_{kk} y_t + e_{t-k} \qquad (3.9)$$

Multiplying both sides with $y_{t-k+j}$, the above eq$^n$. 3.9 is converted to the following equation.

$$y_{t-k} y_{t-k+j} = \phi_{k1} y_{t-k+1} y_{t-k+j} + \cdots + \phi_{kk} y_t y_{t-k+j} + e_{t-k} y_{t-k+j}$$

Incorporating covariance to both sides in the above equation will lead to the following equation (Box et al. 2008)).

$$Cov(y_{t-k} y_{t-k+j}) = \phi_{k1} Cov(y_{t-k+1}, y_{t-k+j}) + \cdots + \phi_{kk} Cov(y_t, y_{t-k+j}) + Cov(e_{t-k}, y_{t-k+j})$$
$$\gamma_j = \phi_{k1}\gamma_{j-1} + \phi_{k2}\gamma_{j-2} + \cdots + \phi_{kk}\gamma_{j-k}$$

Dividing both sides with $\gamma_0$ leads to:

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \cdots + \phi_{kk}\rho_{j-k}$$

For, $j = 1, 2, \ldots, k$, the following system of equations are derived:

$$\begin{aligned} \rho_1 &= \phi_{k1} + \phi_{k2}\rho_1 + \cdots + \phi_{kk}\rho_{k-1} \\ \rho_2 &= \phi_{k1}\rho_1 + \phi_{k2} + \cdots + \phi_{kk}\rho_{k-2} \\ &\vdots \\ \rho_k &= \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \cdots + \phi_{kk} \end{aligned}$$

Using Cramer’s rule successively for $k = 1, 2 \ldots k$

$$\phi_{11} = \rho_1$$
$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$
$$\vdots$$

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

The generalized equation from Levinson and Durbin's recursive formula is then defined as follows (Box et al. 2008).

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}} \qquad (3.10)$$

where, $\phi_{kj} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j}$, $j = 1,2,\cdots,k-1.$

Based on the above formula, PACF is obtained for different lags *k =1*, *2*... *k* for AR (1). They are presented as follows.

$$\phi_{11} = \frac{\rho_1 - \sum_{j=1}^{1-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{1-1} \phi_{k-1,j}\rho_{k-j}} = \rho_1 = \phi_1$$

$$\phi_{22} = \frac{\rho_2 - \sum_{j=1}^{2-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{2-1} \phi_{k-1,j}\rho_{k-j}} = \frac{\rho_2 - \phi_{11}\rho_1}{1 - \phi_{11}\rho_1} = \frac{\phi_1^2 - \phi_1\phi_1}{1 - \phi_1\phi_1} = 0$$

$$\phi_{22} = \frac{\rho_2 - \sum_{j=1}^{2-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{2-1} \phi_{k-1,j}\rho_{k-j}} = \frac{\rho_2 - \phi_{11}\rho_1}{1 - \phi_{11}\rho_1} = \frac{\phi_1^2 - \phi_1\phi_1}{1 - \phi_1\phi_1} = 0$$

$$\phi_{33} = \frac{\rho_3 - \sum_{j=1}^{3-1} \phi_{3-1,j}\rho_{3-j}}{1 - \sum_{j=1}^{3-1} \phi_{3-1,j}\rho_{3-j}} = \frac{\rho_3 - \phi_{21}\rho_2 - \phi_{22}\rho_1}{1 - \phi_{21}\rho_2 - \phi_{22}\rho_1} = \frac{\rho_3 - (\phi_{11} - \phi_{22}\phi_{11})\rho_2 - \phi_{22}\rho_1}{1 - (\phi_{11} - \phi_{22}\phi_{11})\rho_2 - \phi_{22}\rho_1}$$

$$= \frac{\rho_3 - \phi_{11}\rho_2 + \phi_{22}\phi_{11}\rho_2 - \phi_{22}\rho_1}{1 - \phi_{11}\rho_2 + \phi_{22}\phi_{11}\rho_2 - \phi_{22}\rho_1} = \frac{\phi_1^3 - \phi_1\phi_1^2 + 0 * \phi_1\rho_2 - 0 * \rho_1}{1 - \phi_1\phi_1^2 + 0 * \phi_1\rho_2 - 0 * \rho_1} = 0$$

Therefore, generally PACF for AR (1) for different lag is expressed as follows.

$$\phi_{kk} = \begin{cases} \phi_1 & \text{when } k = 1 \\ 0 & \text{when } k = 2, 3, \ldots \end{cases}$$

From the PACF value, it seems that it cuts off after lag 1. Moreover, the value can be either positive or negative based on the value of $\phi_1$. The PACF value would be positive for $0 < \phi_1 < 1$, while the PACF value would be negative for $-1 < \phi_1 < 0$. AR (1) is always in the invertible condition. To be stationary, $|\phi_1|$ needs to be less than 1 (Box et al. 2008). Same as AR (1), the PACF for AR (2) for different lags from the eq$^n$. 3.9 is expressed as follows (Montgomery et al. 2008).

When $k$ = 1,

$$\phi_{11} = \frac{\rho_1 - \sum_{j=1}^{1-1} \phi_{1-1,j}\rho_{1-j}}{1 - \sum_{j=1}^{1-1} \phi_{1-1,j}\rho_{1-j}} = \rho_1 = \frac{\phi_1}{1 - \phi_2}$$

When $k$ = 2,

$$\phi_{22}$$

$$= \frac{\rho_2 - \sum_{j=1}^{2-1} \phi_{2-1,j}\rho_{2-j}}{1 - \sum_{j=1}^{2-1} \phi_{2-1,j}\rho_{2-j}} = \frac{\rho_2 - \phi_{1,1}\rho_1}{1-\phi_{1,1}\rho_1} = \frac{\phi_1\rho_1 + \phi_2\rho_0 - \phi_{1,1}\rho_1}{1-\phi_{1,1}\rho_1} = \frac{\frac{\phi_1^2}{1-\phi_2} + \phi_2 - \frac{\phi_1^2}{(1-\phi_2)^2}}{1-\phi_{1,1}\rho_1}$$

$$= \frac{\frac{\phi_1^2(1-\phi_2) + \phi_2(1-\phi_2)^2 - \phi_1^2}{(1-\phi_2)^2}}{1 - \frac{\phi_1^2}{(1-\phi_2)^2}} = \frac{\frac{\phi_1^2(1-\phi_2) + \phi_2(1-\phi_2)^2 - \phi_1^2}{(1-\phi_2)^2}}{\frac{(1-\phi_2)^2 - \phi_1^2}{(1-\phi_2)^2}} = \frac{\frac{\phi_1^2 - \phi_1^2\phi_2 + \phi_2 - 2\phi_2^2 + \phi_2^3 - \phi_1^2}{(1-\phi_2)^2}}{\frac{1-2\phi_2+\phi_2^2-\phi_1^2}{(1-\phi_2)^2}}$$

$$= \frac{\frac{\phi_2 - 2\phi_2^2 + \phi_2^3 - \phi_1^2\phi_2}{(1-\phi_2)^2}}{\frac{1-2\phi_2+\phi_2^2-\phi_1^2}{(1-\phi_2)^2}} = \frac{\frac{\phi_2(1-2\phi_2+\phi_2^2-\phi_1^2)}{(1-\phi_2)^2}}{\frac{1-2\phi_2+\phi_2^2-\phi_1^2}{(1-\phi_2)^2}} = \phi_2$$

When $k = 3$

$$\phi_{33}$$

$$= \frac{\rho_3 - \sum_{j=1}^{3-1} \phi_{3-1,j}\rho_{3-j}}{1 - \sum_{j=1}^{2-1} \phi_{2-1,j}\rho_{2-j}} = \frac{\rho_3 - \phi_{2,1}\rho_2 - \phi_{2,2}\rho_1}{1-\phi_{2,1}\rho_2 - \phi_{2,2}\rho_1}$$

$$= \frac{\phi_1\rho_2 + \phi_2\rho_1 - (\phi_{1,1} - \phi_{2,2}\phi_{1,1})\rho_2 - \phi_{2,2}\rho_1}{1-(\phi_{1,1} - \phi_{2,2}\phi_{1,1})\rho_2 - \phi_{2,2}\rho_1}$$

$$= \frac{\phi_1\rho_2 + \phi_2\rho_1 - (\phi_{1,1} - \phi_{2,2}\phi_{1,1})\rho_2 - \phi_{2,2}\rho_1}{1-\phi_{1,1}\rho_2 + \phi_{2,2}\phi_{1,1}\rho_2 - \phi_{2,2}\rho_1}$$

$$= \frac{\phi_1(\phi_1\rho_1 + \phi_2) + \phi_2\rho_1 - (\phi_{1,1} - \phi_{2,2}\phi_{1,1})(\phi_1\rho_1 + \phi_2) - \phi_{2,2}\rho_1}{1-(\phi_{1,1} - \phi_{2,2}\phi_{1,1})(\phi_1\rho_1 + \phi_2) - \phi_{2,2}\rho_1}$$

$$= \frac{\frac{\phi_1^3}{1-\phi_2} + \phi_1\phi_2 + \frac{\phi_1\phi_2}{1-\phi_2} - (\frac{\phi_1^3}{(1-\phi_2)^2} - \frac{\phi_1^3\phi_2}{(1-\phi_2)^2} + \frac{\phi_1\phi_2}{1-\phi_2} - \frac{\phi_1\phi_2^2}{1-\phi_2}) - \frac{\phi_1\phi_2}{1-\phi_2}}{1 - (\frac{\phi_1^3}{(1-\phi_2)^2} - \frac{\phi_1^3\phi_2}{(1-\phi_2)^2} + \frac{\phi_1\phi_2}{1-\phi_2} - \frac{\phi_1\phi_2^2}{1-\phi_2}) - \frac{\phi_1\phi_2}{1-\phi_2}}$$

$$= \frac{\frac{\phi_1^3}{1-\phi_2} + \phi_1\phi_2 - \frac{\phi_1^3}{(1-\phi_2)^2} + \frac{\phi_1^3\phi_2}{(1-\phi_2)^2} + \frac{\phi_1\phi_2^2}{1-\phi_2} - \frac{\phi_1\phi_2}{1-\phi_2}}{1 - \frac{\phi_1^3}{(1-\phi_2)^2} + \frac{\phi_1^3\phi_2}{(1-\phi_2)^2} + \frac{\phi_1\phi_2^2}{1-\phi_2} - \frac{\phi_1\phi_2}{1-\phi_2}}$$

$$= \frac{\dfrac{\phi_1^3(1-\phi_2)+\phi_1\phi_2(1-\phi_2)^2-\phi_1^3+\phi_1^3\phi_2+\phi_1\phi_2^2(1-\phi_2)-\phi_1\phi_2(1-\phi_2)}{(1-\phi_2)^2}}{1-\dfrac{\phi_1^3}{(1-\phi_2)^2}+\dfrac{\phi_1^3\phi_2}{(1-\phi_2)^2}+\dfrac{\phi_1\phi_2^2}{1-\phi_2}-\dfrac{\phi_1\phi_2}{1-\phi_2}}$$

$$= \frac{\dfrac{\phi_1^3-\phi_1^3\phi_2+\phi_1\phi_2-2\phi_1\phi_2^2+\phi_1\phi_2^3-\phi_1^3+\phi_1^3\phi_2+\phi_1\phi_2^2-\phi_1\phi_2^3-\phi_1\phi_2+\phi_1\phi_2^2}{(1-\phi_2)^2}}{1-\dfrac{\phi_1^3}{(1-\phi_2)^2}+\dfrac{\phi_1^3\phi_2}{(1-\phi_2)^2}+\dfrac{\phi_1\phi_2^2}{1-\phi_2}-\dfrac{\phi_1\phi_2}{1-\phi_2}}$$

$$= \frac{\dfrac{0}{(1-\phi_2)^2}}{1-\dfrac{\phi_1^3}{(1-\phi_2)^2}+\dfrac{\phi_1^3\phi_2}{(1-\phi_2)^2}+\dfrac{\phi_1\phi_2^2}{1-\phi_2}-\dfrac{\phi_1\phi_2}{1-\phi_2}} = 0$$

For $k = k$,

$$\phi_{kk} = 0$$

Therefore, the PACF for AR (2) for different lags is expressed as follows (Montgomery et al. 2008).

$$\phi_{kk} = \begin{cases} \dfrac{\phi_1}{1-\phi_2} & \text{when } k = 1 \\ \phi_2 & \text{when } k = 2 \\ 0 & \text{when } k = 3, 4, \ldots \end{cases}$$

From above equation, it seems that PACF cuts off after lag 2. Lags are all positive for real roots and start with positive for complex roots when $0 < \phi_1$ and $0 < \phi_2$. On the other hand, alternating in sign starts with negative for real roots and all negative for complex roots when $0 > \phi_1$ and $0 < \phi_2$. AR (2) process is always in the inverted form and to be stationary the roots of $\phi(B) = (1- \phi_1 B- \phi_2^2 B) = 0$ must lie outside the unit circle

(Box et al. 2008). To be stationary, AR (2) must satisfy the following conditions (Montgomery et al. 2008):

$$\phi_1 + \phi_2 < 1$$
$$\phi_2 - \phi_1 < 1$$
$$-1 < \phi_2 < 1$$

3.8 Theoretical Partial Auto-correlation Function (PACF) of the Moving Average (MA) Model

Following eq$^n$. 3.10, the PACF value can be obtained for different lags of MA (1) and is expressed as follows (Box et al. 2008).

For $k$ = 1,

$$\theta_{11} = \frac{\rho_1 - \sum_{j=1}^{1-1} \theta_{1-1,j}\rho_{1-j}}{1 - \sum_{j=1}^{1-1} \theta_{1-1,j}\rho_{1-j}} = \rho_1 = \frac{-\theta_1}{1+\theta_1^2}$$

$$= \frac{-\theta_1(1-\theta_1^2)}{(1+\theta_1^2)(1-\theta_1^2)} = \frac{-\phi_1(1-\theta_1^2)}{(1-\theta_1^4)} = \frac{-\phi_1^1(1-\theta_1^2)}{(1-\theta_1^{2(1+1)})}$$

For $k$ = 2,

$$\theta_{22}=\frac{\rho_2-\sum_{j=1}^{2-1}\theta_{2-1,j}\rho_{2-j}}{1-\sum_{j=1}^{2-1}\theta_{2-1,j}\rho_{2-j}}=\frac{\rho_2-\theta_{1,1}\rho_1}{1-\theta_{1,1}\rho_1}=\frac{-\theta_{1,1}\rho_1}{1-\theta_{1,1}\rho_1}=\frac{-\frac{\theta_1^2}{(1+\theta_1^2)^2}}{1-\theta_{1,1}\rho_1}$$

$$=\frac{-\frac{\theta_1^2}{(1+\theta_1^2)^2}}{1-\frac{\theta_1^2}{(1+\theta_1^2)^2}}=\frac{-\frac{\theta_1^2}{(1+\theta_1^2)^2}}{\frac{(1+\theta_1^2)^2-\theta_1^2}{(1+\theta_1^2)^2}}=\frac{-\frac{\theta_1^2}{(1+\theta_1^2)^2}}{\frac{1+2\theta_2^2+\theta_2^4-\theta_1^2}{(1-\theta_2)^2}}$$

$$=\frac{-\frac{\theta_1^2}{(1+\theta_1^2)^2}}{\frac{1+\theta_2^2+\theta_2^4}{(1-\theta_2)^2}}=\frac{-\theta_1^2}{1+\theta_2^2+\theta_2^4}=\frac{-\theta_1^2(1-\theta_1^2)}{(1-\theta_1^2)(1+\theta_2^2+\theta_2^4)}=\frac{-\theta_1^2(1-\theta_1^2)}{(1-\theta_1^6)}=\frac{-\theta_1^2(1-\theta_1^2)}{(1-\theta_1^{2(2+1)})}$$

$\vdots$

$\vdots$

For $k = k$,

$$\theta_{kk}=\frac{-\theta_1^k(1-\theta_1^2)}{(1-\theta_1^{2(k+1)})} \qquad (3.11)$$

From eq$^n$. 3.11, it is evident that the PACF decays exponentially with an increase of lag $k$. To be invertible, the roots of $\theta(B) = (1- \theta_1 B) = 0$ must lie outside the unit circle i.e., $|\theta_1| < 1$. If $0 < \theta_1 < 1$, all auto-correlations are negative and if $-1 < \theta_1 < 0$, the sign of auto-correlations will show alternating pattern with a positive value at start (Box et al. 2008). MA (1) process is always stationary.

The deduction of the theoretical PACF of MA (2) is very complex but can be shown as the sum of two exponentials if the roots are real and as decreasing sine waves if the roots are complex.

3.9 Stationary Time Series and Differencing

A stationary time series is one whose properties do not vary with respect to time whenever it is observed. Therefore, statistical properties such as mean, variance, and correlation of a stationary time series are all constant over time (Box et al. 2008). In other words, the time series without trends or without seasonality can be regarded as stationary time series.

The reason for trying to have a stationary time series is to obtain meaningful sample statistics, which might be helpful in providing description of the future characteristics. For instance, if the series is consistently increasing or decreasing over time, the sample mean and variance will grow up or down with respect to time and with the size of the sample. It may under-estimate the mean and variance in future periods. If the mean and variance of a series are not well defined, then the correlations among variables and errors are also not meaningful. (Box et al. 2008)

A time series can be made stationary through transformations such as differencing and logging. The term differencing is actually the change between consecutive observations in the original time series and can be applied consecutively more than once to make the time series stationary. It can be named as first difference, second difference, third difference, and so on. Differencing removes the change in the level of time series. It stabilizes the mean eliminating trends and seasonality (Chase Jr. 2009).

The first difference $y_t^{/}$ of time series $y_t$ is expressed as follows.

$$y_t^{/} = y_t - y_{t-1}$$

The differenced series will have $t-1$ values for a time series since it is not possible to calculate a difference $y_1^{/}$ for the first observation. Occasionally, the first order-differenced data $y_1^{/}$ will not appear stationary.

Therefore, it may be necessary to difference the data multiple times to obtain a stationary series. The degree of differencing determines the *d* value of AIRMA *(p, d, q)* model. One way to determine more objectively if differencing is required is the use of a unit root test. A number of unit root tests are available. Based on the assumptions, the tests may lead to conflicting answers. One of the most popular tests is the Augmented Dickey-Fuller (ADF) test. The null hypothesis of this test is that the model is not stationary. Using the usual 5% threshold, differencing is required if the p-value is greater than 0.05. Another popular unit root test is the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. However, the hypotheses are just opposite i.e., the null-hypothesis is that the data are stationary. In this case, small p-values (e.g., less than 0.05) suggest that differencing is required (Chase Jr. 2009).

3.10 Cross-Correlation Function (CCF)

In the relationship between two time series ($y_t$ and $x_t$), $y_t$ may be related to the past lags of the x-series. The CCF can identify the lags of the x-variable that might be useful predictors of $y_t$. CCF is defined as the product-moment correlation between two time series of different time lags (Box et al. 2008). The same principle was applied in this research to identify the correlations of travel time series for different road segments. The general form of CCF for upstream and downstream travel times is shown next (Box et al. 2008).

$$\rho_{s,s-1}^{k}(k) = \frac{E[(T_{s,t+k} - \mu_s)(T_{s-1,t} - \mu_{s-1})]}{\sigma_s \sigma_{s-1}}$$

Here, $\mu_s$ and $\mu_{s-1}$ are the means of $T_s$ and $T_{s-1}$ and $\sigma_s$ and $\sigma_{s-1}$ are the standard deviation of $T_s$ and $T_{s-1}$ and $k$ is the time lag between two series. In general, the CCF can be described as "lead" and "lag" relationship. When $k > 0$, $T_{s-1,t}$ "leads" $T_{s,t}$ and when $k < 0$, $T_{s-1,t}$ "lags" $T_{s,t}$. For instance, consider $k = -$ h where $h$ is a positive integer, CCF measures the relationship between upstream travel time at $h$ minutes before $t$ and the downstream travel time at time $t$.

3.11 Autoregressive Integrated Moving Average (ARIMA) Model

ARIMA models were first introduced in the early 1900s and were popularized by George Box and Gwilym Jenkins in the early 1970s. They developed a comprehensive approach that integrates the relevant information required to understand and use ARIMA models. They formalized their theory and methodology by developing a process to select the best ARIMA model from a group of candidate models. As a result, ARIMA models are often referred to as Box-Jenkins models (Box et al. 2008).

Although the theoretical notation is quite sophisticated, applying ARIMA models is not that difficult, particularly with the advances in automating the Box-Jenkins procedure using forecasting software packages (Chase Jr. et al. 2009). The Box-Jenkins approach incorporates key elements from both time series and regression methods for forecasting. As a result, practitioners must have a solid understanding of regression before attempting to apply the Box-Jenkins approach to create an ARIMA model. When applying ARIMA models, two basic steps are required: (1) analysis of the data series and (2) selection of a forecasting model (from several candidate models) that best fits the data series (Montgomery et al 2009). In general, ARIMA is a combination of differencing with AR and a MA model. It is, hence, called as ARIMA ($p, d, q$) model where (Box et al. 2008):

$p$ = order of the AR part

$d$ = degree of first differencing involved

$q$ = order of the MA part

The generalized equation of ARIMA ($p$, $d$, $q$) model is presented next (Box et al. 2008).

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} + e_t \quad (3.12)$$

Where, $c$ is a constant. Based on equation 3.12, the transfer function-noise model can be expressed as follows (Montgomery et al. 2008).

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots + \phi_p B^p) y_t = c + (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)\, \varepsilon_t$$

If the function is transformed and there is white noise then the above formula is expressed as follows.

$$=> \phi_y(B)(1-B)^d y_t = \theta_y(B)\varepsilon_t$$

$$=> \varphi_y(B) y_t = \theta_x(B)\varepsilon_t \qquad \text{where, } \varphi_y(B) = \phi_y(B)(1-B)^d$$

$$=> \varepsilon_t = \theta_y(B)^{-1} \varphi_y(B) y_t$$

In this notation, $\theta_y(B)^{-1}\varphi_y(B)$ acts as a filter and when applied to $y_t$ generates a white noise time series. Therefore, it is referred as "pre-whitening" (Montgomery et al. 2008).

3.12 Summary

Overall, in this chapter, background information pertaining to Cronbach’s α and ARIMA is provided. The study area and data used in this research is explained in the next chapter.

# CHAPTER 4: STUDY AREA AND DATA COLLECTION

Travel time and incident are required to forecast travel time and its variation under incident condition and without incident condition. For this research, only “vehicle accident” incident type and travel time data were collected for a freeway and considered for modeling. Incident data should include its severity, location and date/time of occurrence to identity their position along the study corridor. This chapter consists of selection of study area and collection of data.

## 4.1 Selection of Study Area

For this research, Mecklenburg County was selected as the study area. This county is located in the state of North Carolina, United States. The population of Mecklenburg County was 919,628 in 2012, making it the most populated as well as the most densely populated county in North Carolina ($49^{th}$ among 3,143 counties in the United States). Its largest city is Charlotte, which is also its county seat. In 2012, the estimated population of Charlotte, according to the United States Census Bureau, was 774,442 making it the $27^{th}$ largest city in the United States (US Census Bureau 2012).

Charlotte is the home of the corporate headquarters of Bank of America and the east coast operations of Wells Fargo. Some notable attractions of Charlotte are the Bank of America Stadium (Home of the Carolina Panthers of the National Football League - NFL), the Charlotte Hornets of the National Basketball Association (NBA), NASCAR

Sprint Cup race and the NASCAR All-Star Race, the Wells Fargo Championship, the NASCAR Hall of Fame, Carowinds amusement park, and the United States National Whitewater Center (O' Daniel 2012). The area is also served by the Charlotte Douglas International Airport, a major international hub, which was ranked as the 23rd busiest airport in the world in 2013 (anna.aero 2008).

4.2 Selection of the Study Corridor

Along southbound direction of I-77 in the Mecklenburg County of North Carolina, a 19.13-mile long segment was considered as the study corridor for this research. A part of the study corridor is shown in Figure 2. Pink triangles in Figure 2 define the starting point of each Traffic Message Channel (TMC) code (could simply be referred to as a "segment") that extends in the southbound direction.

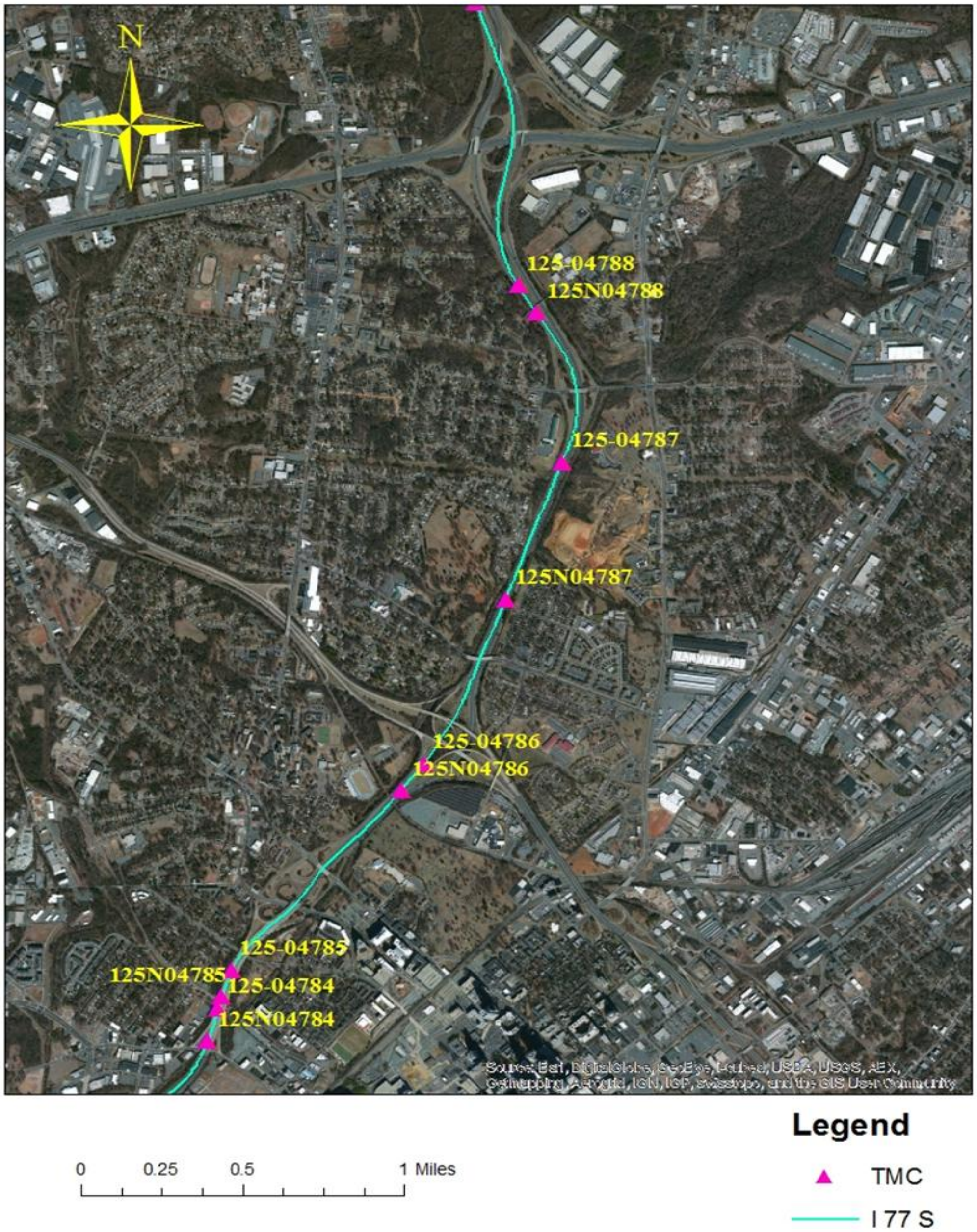



FIGURE 2: Part of selected study corridor along I-77 S

The study corridor extends from TMC codes 125-04777 to 125N04792, consisting of 28 TMC codes. From downstream to upstream, road segments were designated as X1 to X28 based on the sequence along the selected corridor of I-77 S i.e., downstream segment was designated as X1 and upstream segment is designated as X28. A description of each study segment along I-77 S is provided in Table 2.

TABLE 2: Description of the segments along the study corridor

| Segment No | TMC Code | Length (miles) | Direction | Location | Segment Type |
|---|---|---|---|---|---|
| X1 | 125N04778 | 0.48 | Southbound | Nations Ford Rd/Exit 4 | Internal |
| X2 | 125-04778 | 0.56 | Southbound | Nations Ford Rd/Exit 4 | External |
| X3 | 125N04779 | 0.57 | Southbound | Tyvola Rd/Exit 5 | Internal |
| X4 | 125-04779 | 0.67 | Southbound | Tyvola Rd/Exit 5 | External |
| X5 | 125N04780 | 0.28 | Southbound | Woodlawn Rd/Exit 6 | Internal |
| X6 | 125-04780 | 0.14 | Southbound | Woodlawn Rd/Exit 6 | External |
| X7 | 125N04781 | 0.21 | Southbound | NC-49/Tryon St/Exit 6 | Internal |
| X8 | 125-04781 | 0.74 | Southbound | NC-49/Tryon St/Exit 6 | External |
| X9 | 125N04782 | 0.39 | Southbound | Clanton Rd/Exit 7 | Internal |
| X10 | 125-04782 | 0.61 | Southbound | Clanton Rd/Exit 7 | External |
| X11 | 125N04783 | 0.22 | Southbound | Remount Rd/Exit 8 | Internal |
| X12 | 125-04783 | 0.57 | Southbound | Remount Rd/Exit 8 | External |
| X13 | 125N04784 | 0.96 | Southbound | I-277/US-74/Exit 9 | Internal |
| X14 | 125-04784 | 0.11 | Southbound | I-277/US-74/Exit 9 | External |
| X15 | 125N04785 | 0.19 | Southbound | Morehead St/Exit 10 | Internal |
| X16 | 125-04785 | 0.09 | Southbound | Morehead St/Exit 10 | External |
| X17 | 125N04786 | 0.80 | Southbound | Trade St/5th St/Exit 10 | Internal |
| X18 | 125-04786 | 0.11 | Southbound | Trade St/5th St/Exit 10 | External |
| X19 | 125N04787 | 0.95 | Southbound | Brookshire Fwy/Exit 11 | Internal |
| X20 | 125-04787 | 0.49 | Southbound | Brookshire Fwy/Exit 11 | External |
| X21 | 125N04788 | 0.53 | Southbound | Lasalle St/Exit 12 | Internal |
| X22 | 125-04788 | 0.11 | Southbound | Lasalle St/Exit 12 | External |
| X23 | 125N04789 | 0.97 | Southbound | Statesville Ave/Exit 13 | Internal |
| X24 | 125-04789 | 1.65 | Southbound | Statesville Ave/Exit 13 | External |
| X25 | 125N04790 | 0.58 | Southbound | US-21/Exit 16 | Internal |
| X26 | 125-04790 | 2.20 | Southbound | US-21/Exit 16 | External |
| X27 | 125N04791 | 0.66 | Southbound | Reames Rd/Exit 18 | Internal |
| X28 | 125-04791 | 3.78 | Southbound | Reames Rd/Exit 18 | External |

Each segment shown in Table 1 is categorized as either internal (designated by N or P) or external path (designated by + or -). An internal path is the segment which starts at an off-ramp and ends at an on-ramp, while an external path is the road segment that leads up to the point of interchange or intersection. Each segment (TMC code) will have a +/-/N/P symbol; "+" sign indicates northbound or westbound direction and "-" sign indicates southbound or eastbound direction along the external path. "P" sign indicates northbound or westbound direction and "N" sign indicates southbound or eastbound direction along the internal path. Moreover, each TMC code is a combination of country code, location code, internal/external path and its direction, and TMC code itself. For instance, in a TMC code like 125-04791, "1" stands for country code USA, "25" is the location code, "-" defines the external path in southbound or eastbound direction and the last four digit is actual TMC code. A schematic diagram of TMC code is shown in Figure 3 (INRIX 2013).

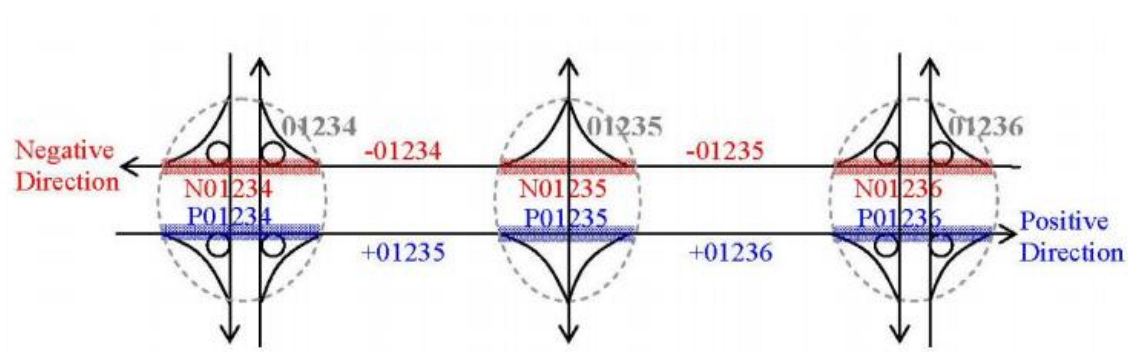



FIGURE 3: Schematic diagram of internal and external path (INRIX 2013)

### 4.3 Collection of Travel Time and Incident Data

The database needs to be developed in a way that it would have both travel time and incident information for analysis and modeling. Incident data was obtained from the North Carolina Department of Transportation’s (NCDOT) Traveler Information Management System (TIMS) website, which was created to provide accurate, timely and easy accessible information to the travelling public. TIMS recorded 172,350 incidents for North Carolina from 2010 to 2012. This database contained different information about incidents such as unique incident ID, incident type, start and end time, location (mile marker and coordinates), number of lanes closed, reason of incident, direction of traffic, road name, and severity (Table 3).

TABLE 3: TIMS incident database

| Incident ID | Road Name | Start Time | End Time | MM | Incident Type | Closed Lanes | Severity |
|---|---|---|---|---|---|---|---|
| 344518 | I-77 | 1/1/2011 11:13:00 | 1/1/2011 12:13:00 | 11 | Vehicle Accident | 0 | 2 |
| 344844 | I-77 | 1/3/2011 10:43:00 | 1/3/2011 11:00:48 | 45 | Vehicle Accident | 1 | 3 |
| 344867 | I-77 | 1/3/2011 18:11:00 | 1/3/2011 19:11:00 | 9 | Vehicle Accident | 1 | 2 |
| 344987 | I-77 | 1/5/2011 12:16:00 | 1/5/2011 12:40:44 | 1 | Vehicle Accident | 1 | 3 |
| 344997 | I-77 | 1/6/2011 1:00:00 | 1/5/2011 23:27:59 | 30 | Night Time Maintenance | 0 | 1 |
| 344999 | I-77 | 1/6/2011 20:30:00 | 1/7/2011 1:00:00 | 10 | Night Time Maintenance | 2 | 2 |
| 345000 | I-77 | 1/7/2011 0:00:00 | 1/7/2011 3:00:00 | 10 | Night Time Maintenance | 1 | 2 |
| 345283 | I-77 | 1/6/2011 13:38:00 | 1/6/2011 14:12:00 | 31 | Vehicle Accident | 1 | 3 |
| 345288 | I-77 | 1/6/2011 14:22:00 | 1/6/2011 14:29:03 | 4 | Congestion | 0 | 2 |
| 345348 | I-77 | 1/8/2011 6:00:00 | 1/8/2011 8:03:30 | 8 | Maintenance | 1 | 2 |
| 345509 | I-77 | 1/10/2011 16:27:00 | 1/10/2011 17:41:12 | 43 | Vehicle Accident | 1 | 3 |
| 345542 | I-77 | 1/10/2011 19:02:00 | 1/10/2011 20:02:00 | 1 | Vehicle Accident | 1 | 2 |

The classification of incident severity in the TIMS database is not same as KABCO groupings (National Safety Council 1990). Instead, it is classified into three simple groups: Type 1, Type 2, and Type 3. According to TIMS, the type “1” severity incident includes that the incident caused blockage of lane/lanes. Incidents that caused shoulder blockage is defined as type “2” severity, while type “3” severity incidents are incidents with no lane blockage but congestion. Total records of incidents in this database for I-77 were 3,975. TIMS reported 14 types of incidents along I-77. Incident type and number of incidents by incident type are summarized in Table 4.

TABLE 4: Incident type and number of incidents

| No | Incident Type | No. of Incidents | Cumulative Percentage |
|---|---|---|---|
| 1 | Night Time Maintenance | 1549 | 38.97 |
| 2 | Vehicle Accident | 857 | 21.56 |
| 3 | Congestion | 611 | 15.37 |
| 4 | Night Time Construction | 327 | 8.23 |
| 5 | Maintenance | 307 | 7.72 |
| 6 | Disabled Vehicles | 115 | 2.89 |
| 7 | Construction | 61 | 1.53 |
| 8 | Other | 50 | 1.26 |
| 9 | Fire | 38 | 0.96 |
| 10 | Road Obstruction | 17 | 0.43 |
| 11 | Special Event | 17 | 0.43 |
| 12 | Weekend Construction | 17 | 0.43 |
| 13 | Fog | 4 | 0.10 |
| 14 | Weather Event | 4 | 0.10 |
| 15 | Signal Problem | 1 | 0.03 |

Statistics showed that along I-77, 857 “vehicle accident” type incidents occurred from 2010 to 2012. This is 21.6% of the total incidents recorded along I-77 corridor in the Charlotte region. This is 2$^{nd}$ in the ranking; just behind “Night Time Maintenance” which accounts for 39.0% of the total incidents. For this study, only “vehicle accident” type incidents were chosen for analysis and modeling. Databases were developed using data, for 150 “vehicle accident” affected days and 100 sample days of data when there were no incidents, along a ~19-mile freeway corridor. For validation, 45 “vehicle accident” affected days and 35 incident free days were selected and set aside. The selected incidents (pink dots) along I-77 S are shown in Figure 4.

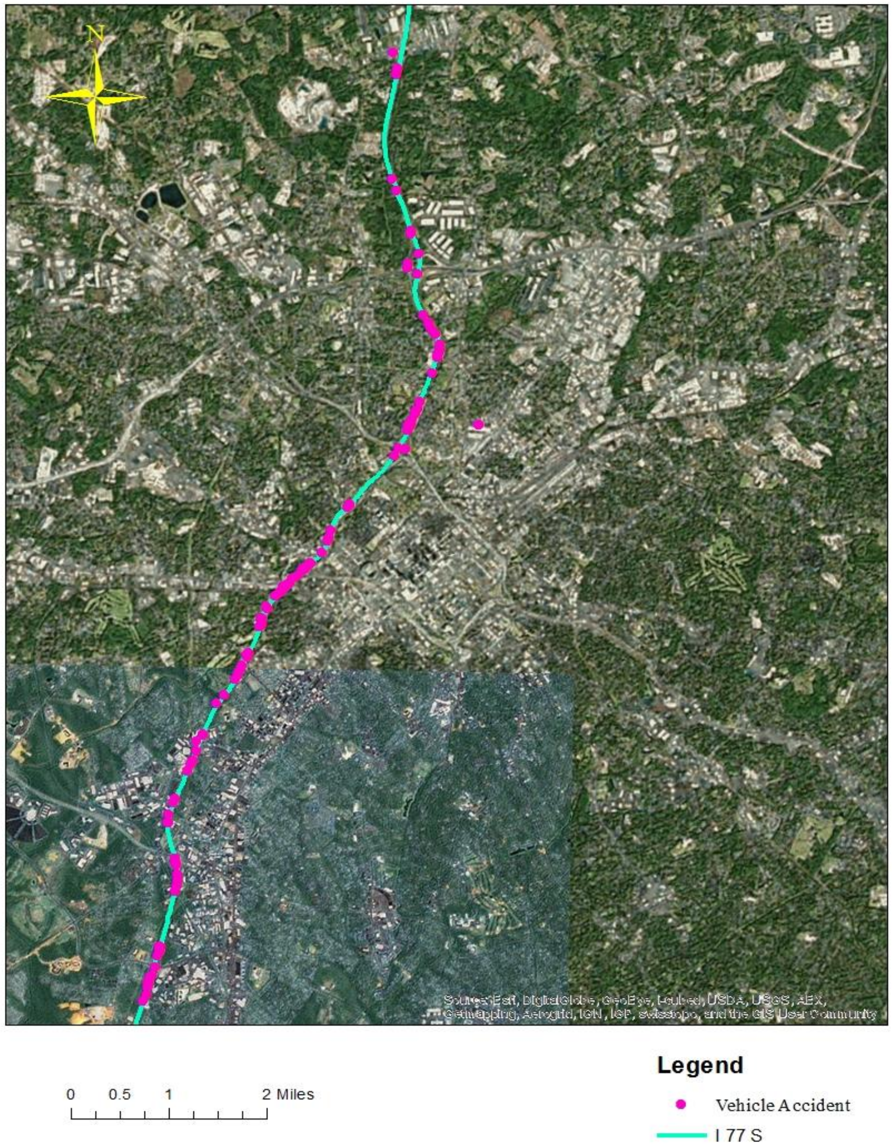



FIGURE 4: Incident locations along I-77 S

Travel time data was obtained from INRIX from 2010 to 2012. INRIX collects real-time (24×7) speed data using numerous probes for more than 260,000 miles of roads, including interstates and major roads, all over the United States. In the INRIX database, each segment is defined by a unique nine digit code known as TMC code created by Tele Atlas and NAVTEQ. For every TMC code, corresponding average travel times computed from the all vehicles observed at one-minute interval is reported. Besides travel time information, INRIX also provides speed, average speed, reference speed, and confidence score. Speed is the current estimated time to traverse the roadway segment in miles per hour; average speed is the historical average mean speed for the roadway segment for that hour-of-the-day and day-of-the-week in miles per hour; and reference speed is the mean “free flow” speed for the roadway segment in miles per hour. As the data type indicator, score has three discrete values: “30” (real time data), “20” (combination of real time and historical data), and “10” (historical data). Travel time information for the study corridor was obtained from this database. Sample INRIX database is shown in Table 5 (INRIX 2013).

## 4.4 Summary

This chapter introduces the two databases, INRIX and TIMS, required for this research. INRIX provides travel time information, which is collected at 1-min intervals, and TIMS provides information pertaining to incidents. From both the databases, information was collected from 2010 to 2012.

TABLE 5: INRIX database

| TMC Code | Time Stamp | Speed (mph) | Average Speed (mph) | Reference Speed (mph) | Travel Time (min) | Confidence score |
|---|---|---|---|---|---|---|
| 125+04639 | 1/3/2011 12:33:56.56 | 63 | 56 | 61 | 0.037 | 30 |
| 125+04640 | 1/3/2011 12:33:56.56 | 57 | 59 | 62 | 0.028 | 30 |
| 125+04641 | 1/3/2011 12:33:56.56 | 63 | 59 | 61 | 1.254 | 30 |
| 125+04642 | 1/3/2011 12:33:56.56 | 56 | 40 | 57 | 0.452 | 30 |
| 125+04643 | 1/3/2011 12:33:56.56 | 66 | 61 | 62 | 0.255 | 30 |
| 125+04644 | 1/3/2011 12:33:56.56 | 66 | 63 | 64 | 0.478 | 30 |
| 125+04645 | 1/3/2011 12:33:56.56 | 73 | 63 | 63 | 0.677 | 30 |
| 125+04646 | 1/3/2011 12:33:56.56 | 74 | 61 | 61 | 0.46 | 30 |
| 125+04647 | 1/3/2011 12:33:56.56 | 70 | 63 | 63 | 0.407 | 30 |
| 125+04648 | 1/3/2011 12:33:56.56 | 61 | 63 | 62 | 0.772 | 30 |
| 125+04649 | 1/3/2011 12:33:56.56 | 56 | 63 | 63 | 0.411 | 30 |
| 125+04650 | 1/3/2011 12:33:56.56 | 60 | 63 | 63 | 0.543 | 30 |
| 125+04651 | 1/3/2011 12:33:56.56 | 64 | 61 | 61 | 0.248 | 30 |
| 125+04652 | 1/3/2011 12:33:56.56 | 62 | 62 | 63 | 0.456 | 30 |

# CHAPTER 5: ESTIMATION OF CRONBACHs' α AND DEVELOPING ARIMA MODEL

This chapter describes the overall methodology in a systematic way. It includes discussion on 1) the estimation of Cronbach's α to estimate the expected travel time; 2) the database development; 3) the adoption of appropriate ARIMA model; and, 4) how the CCF and lagged regression were used to build the prediction model.

## 5.1 Estimation of Maximum Corresponding Cronbach's $\alpha$ and Expected Travel Time

As discussed in Chapter 3, Cronbach's α captures the similarity among the data series. The same principle was adopted to the travel time series. If the similarity is high then a similar pattern of travel time series is observed throughout the year for each time interval. Similarity in the travel time series of a segment is estimated based on two categories (day-of the-week and weekday/ weekend) and two factors (time-of-the-day and week-of the-year). The average travel times were considered as the travel time measure. All these factors yielded four categories of Cronbach's α value in total (Table 6).

TABLE 6: Characteristics of each category of Cronbach's 'α'

| No. | Category | Primary factor | Secondary factor | Travel Time Measure Used |
|---|---|---|---|---|
| α1 | Day-of-the-week | Time-of-the-day | Week-of-the-year | Average |
| α2 | Weekday/Weekend | Time-of-the-day | Week-of-the-year | Average |
| α3 | Day-of-the-week | Week-of-the-year | Time-of-the-day | Average |
| α4 | Weekday/Weekend | Week-of-the-year | Time-of-the-day | Average |

Among the four values, the higher value was selected to compute the expected travel time of a segment for a specific time-period. The computation of Cronbach's α is explained using a simple example considering data shown in Table 7 for TMC Code 125N04781(X7).

TABLE 7: Sample travel time data of a segment and 'day-of-the-week' used for computing 'α1'

| TOD (Primary factor) | Week-of-the-year (secondary factor) | | | | | | | Sum of travel times |
|---|---|---|---|---|---|---|---|---|
| | Week 1 | Week 2 | Week 3 | . | . | . | Week 49 | |
| 12:00 AM-12:10 AM | 0.67 | 1.94 | 0.59 | . | . | . | . | 22.32 |
| 12:10 AM-12:20 AM | 0.65 | 1.94 | 0.60 | . | . | . | . | 22.51 |
| 12:20 AM-12:30 AM | 0.62 | 2.46 | 0.57 | . | . | . | . | 22.68 |
| 12:30 AM-12:40 AM | 0.59 | 1.39 | 0.54 | . | . | . | . | 22.51 |
| 12:40 AM-12:50 AM | 0.60 | 0.68 | 0.59 | . | . | . | . | 22.56 |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| 11:50 PM-12:00 AM | 0.65 | 2.00 | 0.60 | . | . | . | . | 25.91 |
| Item Variance | 0.15 | 0.14 | 0.0004 | . | . | . | . | |

Instead of all the weeks, only 49 weeks are considered as data was missing for the three "Mondays". To calculate α1, variance 1 (V1) is the defined as the variance of summation of travel times (rightmost column) of all 144-time intervals. In this case, V1 is 14.92. Variance 2 (V2) is the summation of all the item variances (bottom row). Value of V2 for this case is 1.98. Therefore, Cronbach's α1 is =(49/48*(1-(1.988/14.92))) = 0.88.

5.2 Development of Databases

As travel time information, alone does not explicitly explain the sole effect of an incident, this study focused on a measure based on variation in travel time (VTT). VTT due to an incident is defined as the difference between travel time under incident condition and travel time when there is no incident during the same period. However, instead of VTT, the Relative Variations in Travel Time (RVTT) was used to capture the effect of incident in this research. The purpose of using RVTT is two-fold. Firstly, the length and other geometric characteristics of every segment along a corridor are not equal. RVTT can eliminate the effect of the difference in lengths and other geometric characteristics when assessing the effect. Secondly, VTT can be a positive value, a negative value or "0". On the other hand, RVTT is always positive. For instance, if the observed travel time and forecasted travel time were same, then the ratio would be "1".

In this research, two different types of RVTT were considered. The first one is expressed as the ratio of travel time to the expected travel time of the same time interval and same segment, whereas the second one is expressed as the ratio of travel time to the minimum travel time. The former captures the sole effect of the incident, while the later captures the effect of incident along with possible effect due to fluctuations in traffic volume.

Databases were developed which consisted of both without incident and “vehicle accident” affected days of travel time information. From TIMS, databases were developed using data, for 150 “vehicle accident” affected days and 100 sample days of data where there were no incidents, along a ~19-mile freeway corridor. The data was gathered from 2010 to 2012. The selected “vehicle accident” type crash had different severity type and the number of lanes blocked due to the incident. In total, four databases were developed for this study.

1. The first database was named as “Travel Time” database. This had travel time information for all 316 days and for all 28 segments. Each day was split into 10-minute intervals – example, 2011-01-01 10:00:00 AM to 2011-01-01 10:10:00 AM, 2011-01-01 10:10:00 AM to 2011-01-01 10:20:00 AM, and so on. Travel time for each interval was computed by averaging all the travel time samples within that specific time interval for that segment. As INRIX data was collected and summarized for every 1-minute interval, each time interval could have a maximum 10 travel time data samples. Travel time in this database was denoted as “TT”.
2. The second database was named as “Expected Travel Time” database. This had average travel time of all 28 segments, for each year from 2010 to 2012, for each day-of-the-week (e.g., Sunday, Monday, etc.). To develop this database, all seven days of the week were split into 10-minute intervals - 10:00:00 AM to 10:10:00 AM, 10:10:00 AM to 10:20:00 AM, and so on. Unlike “Travel Time” database, “”Expected Travel Time” database did not have date part in the timestamp. The average travel time of any time interval for a specific

segment was computed by averaging all the samples for that specific day-of-the-week of that TMC code. Travel time in this database was titled as "ExpTT". This database was used to compute the expected travel time for each segment during each time interval.

3. The third database was named as "Minimum Travel Time" database. In this, each day-of-the-week (e.g., Sunday, Monday, etc.) was split into 10-minute intervals. This database contained the yearly minimum travel time of a TMC code for each day-of-the-week and time interval. Instead of daily minimum travel time of a segment, yearly minimum travel time of a segment for a specific time split was estimated. Daily minimum travel time can be during any 10-min time interval, most probably at night. However, the objective is to compare travel time during a specific time interval with minimum travel time of the same time interval (account for traffic). For example if an incident occurs around 10:00 AM, then comparing travel time at that period with minimum travel time of 1:00 AM (assuming that minimum travel time is observed between 1:00 AM-1:10 AM ) is seemingly inappropriate considering the study objective. However, when the minimum travel time was zero then the second minimum travel time was taken as it was assumed that no vehicle passed that segment during the specific time interval. The travel time in this database was titled as "MinTT".
4. Finally, all the three databases were connected using the segment ID (TMC code), day-of-the-week, year, and time interval. Two more columns were generated in the final table beside of TT. They were "TT/ExpTT" (which is

the ratio of TT to ExpTT) and “TT/MinTT” (which is the ratio of TT to MinTT).

## 5.3 Development of ARIMA model

To forecast the travel time, ARIMA model was developed based on TT. To forecast variations in travel time, ARIMA models were also developed based on TT/ExpTT and TT/MinTT. The process adopted to develop the ARIMA models is discussed in this section. It is illustrated using data for segment X7. The same process was applied for TT/ExpTT and TT/MinTT for all 28 segments.

### 5.3.1 Pattern Recognition to Check the Stationary Condition and Necessity of Differencing of a Time series

Before developing a statistical model, the preliminary task is to observe the data trend through simple time series scatter diagram. This can be done by plotting a set of N observations with respect to time. This simple time series plot is used to assess and make a decision about whether a plot is stationary or not.

There is no growth or decline in the data in case of stationary data. If there is a trend, positive or negative auto-correlations will dominate the ACF or PACF plots, which will make impossible to uncover the other pattern in the data series. Therefore, it is necessary to remove the trend to allow other correlation structures to be seen before choosing the appropriate model (AR, MA, or ARIMA).

The trend of the plot was observed through visual inspection. Figure 5 shows time series plot of travel times along X7. From visual inspection, it is clear that the trend does not show the variation of the mean of the time series plot. This indicates that the time series plot is stationary.

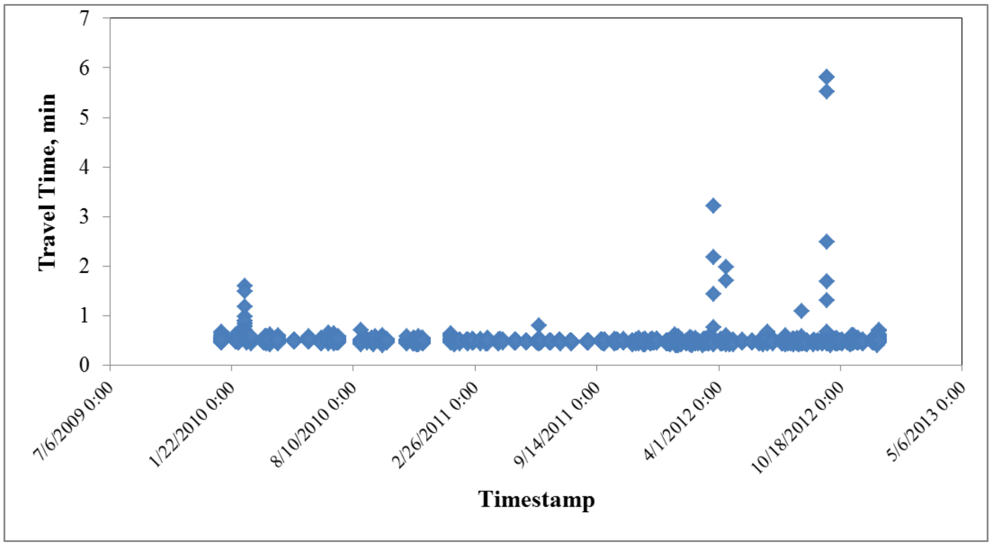



FIGURE 5: Time series plot of travel times along X7

If a plot is stationary then there is no need for differencing. If the plot is non-stationary, the time series needs differencing. Sometimes a time series needs differencing more than once. The time series plot of $n^{th}$ difference should be considered in this case. Normally, the correct amount of differencing is the lowest order of differencing that yields a time series which has a well-defined mean and whose ACF plot decays rapidly to zero, either from above or below (Chase Jr. 2009).

Differencing tends to introduce “negative” correlation. Therefore, if the series has positive auto-correlations and a high number of lags (e.g., 10 or more), then it probably needs a higher order of differencing as differencing reduces the auto-correlation or add negative auto-correlation. However, excess differencing can even drive the lag - 1 auto-correlation to a negative value (Chase Jr. 2009).

If the lag - 1 auto-correlation is zero or negative, or, the auto-correlations are small and pattern less, then the series does not need further differencing (Figure 6). In this case, even if the graph shows any pattern in the auto-correlations, it would be recommended not to perform any more differencing. One of the most common errors in ARIMA modeling is to "over-difference" the time series, which may lead to adding an extra AR or MA term to undo the damage (Chase Jr. 2009).

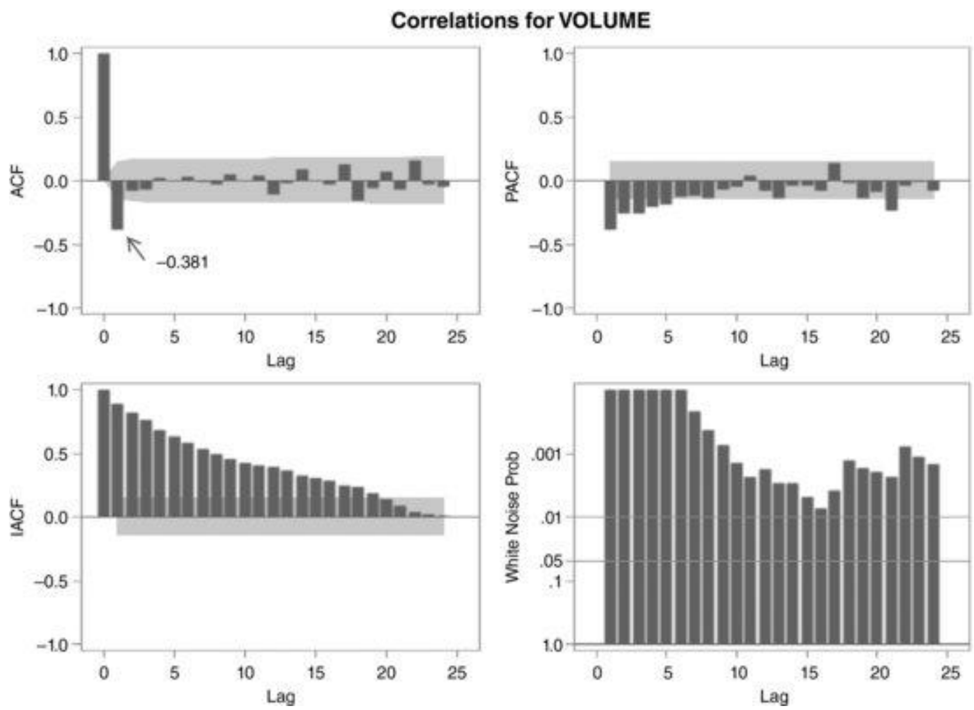



FIGURE 6: Data plots with first differencing (Chase Jr. 2009)

If the lag - 1 auto-correlation is less than - 0.5 (and theoretically a negative lag - 1 auto-correlation should never be greater than 0.5 in magnitude), the series may be over-differenced. The time series plot of an over-differenced series may look quite random in pattern i.e., a pattern of excessive changes in sign from one observation to the next observation or up-down-up-down and beyond from observations to observations. From

Figure 6, after the first non-seasonal differencing, there is no need to conduct a second differencing as lag - 1 is negative and less than - 0.5 (exactly -0.381) (Chase Jr. 2009).

However, if the second non-seasonal differencing is done (Figure 7), lag - 1 comes close to zero. The signs of over-differencing is evident i.e., a pattern of changes of sign from one observation to the next which is confirmed from the ACF plot and shows a positive spike close to zero (0.1393). In addition, the ACF plot is decaying slowly with positive and negative inverse lags. This is a sure sign of over-differencing (Chase Jr. 2009).

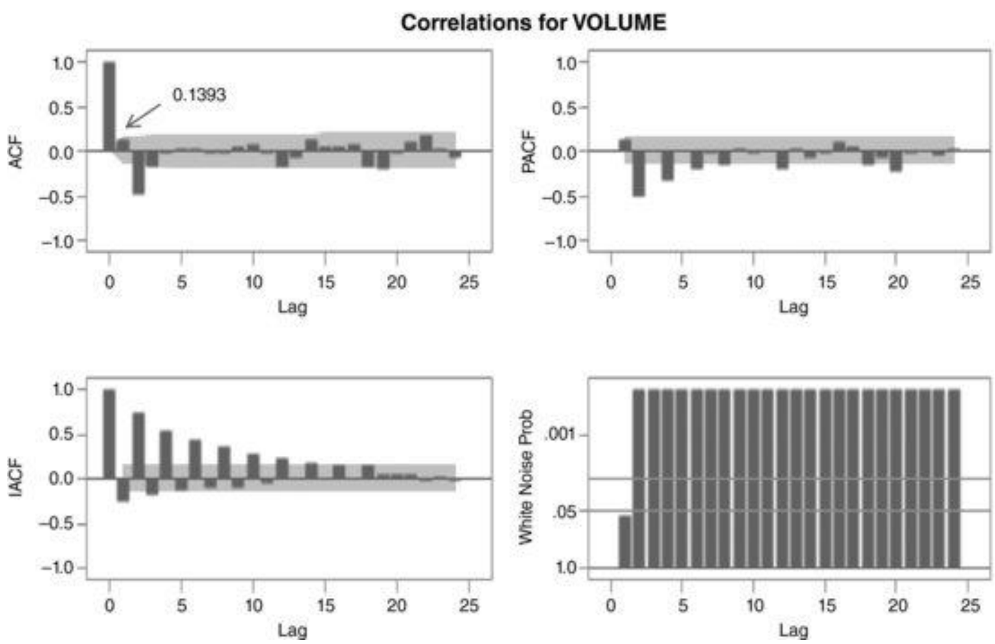



FIGURE 7: Data plots with second differencing (Chase Jr. 2009)

To verify the aforementioned discussion, ACF and PACF plot of the travel-time data series along X7 is shown in Figure 8. ACF plot does not show positive auto-correlations with higher number of lags (exactly 6, which is less than 10. This indicates that the plot is not under-differenced. Moreover, the lag - 1 auto-correlation is not even negative (close to +0.85). The pattern does not show randomness. Therefore, this series need not be over-differenced.

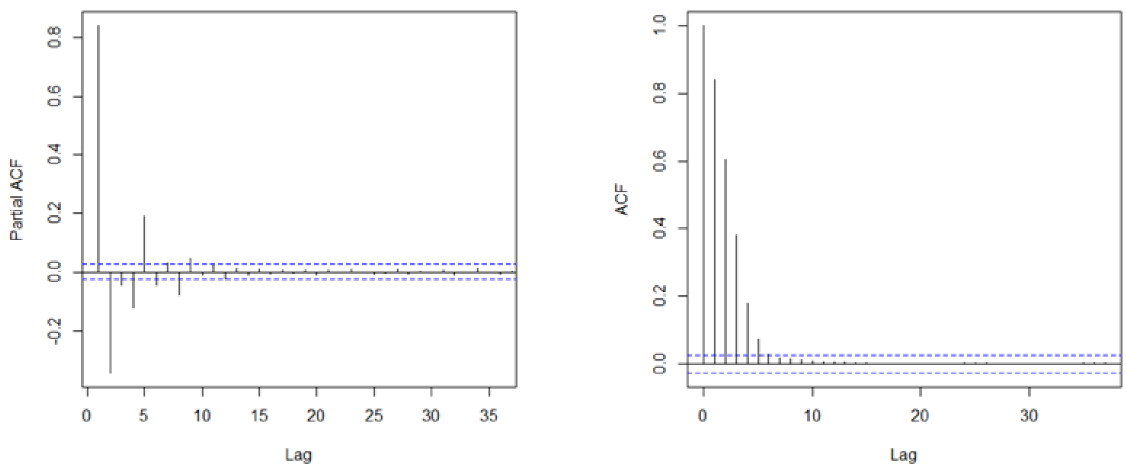



FIGURE 8: ACF and PACF plot of travel time data along X7 without Differencing

Moreover, in Figure 9, after applying the second non-seasonal differencing of the data, lag-1 comes close 0.25. This is close to zero and shows a change in the sign of the pattern, indicating the sign of over-differencing.

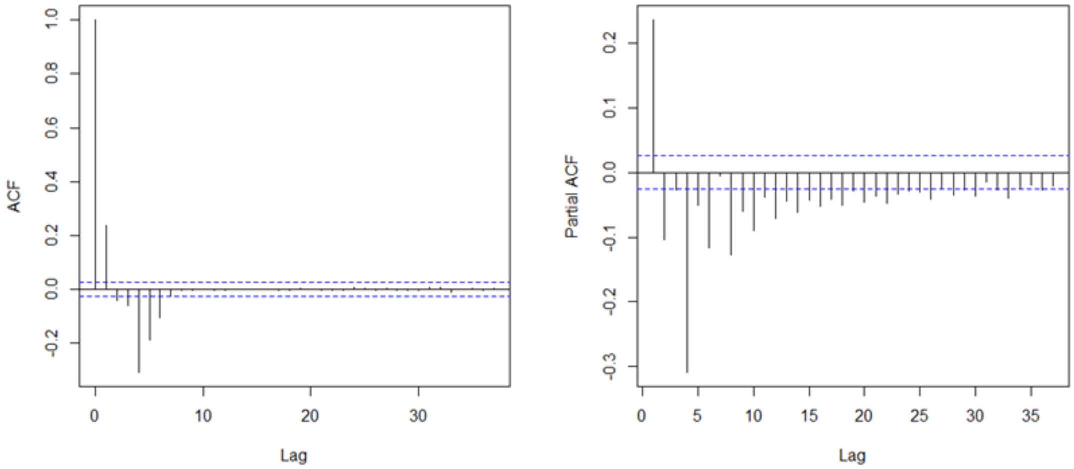



FIGURE 9: ACF and PACF plot of travel time data along X7 with first differencing

From figures 8 and 9, the travel time series of X7 does not need differencing. Performing the unit root test is another way to determine the necessity of differencing. Augmented Dickey - Fuller (ADF) test was implemented in R to confirm that differencing is not needed. Results obtained indicate that Dickey-Fuller = -15.9553, Lag order = 17, and p-value < 0.01. The null hypothesis does not satisfy, as the p-value is less than 0.01. Therefore, the unit root test confirmed that differencing is not needed. A model with no orders of differencing assumes that the original series is stationary itself.

When a time series requires first order of differencing, it is assumed that the original series is either random walk or exponentially smooth with a constant average trend. On the other hand, the series is assumed to follow a random walk model with a time varying trend when a time series model requires second order of differencing (Montgomery et al. 2008).

5.3.2 Identifying the Numbers of AR or MA Terms

After a time series has been made stationary by differencing, the next step in developing an ARIMA model is to determine whether AR or MA terms are needed to correct any auto-correlation that remains in the differenced series. By looking at the ACF and PACF plots of the differenced series, one can tentatively identify the numbers of AR and/or MA terms that are needed.

The stationary series displays an “AR signature” if the PACF displays a sharp cutoff while the ACF decays more slowly (i.e., has significant spikes at higher lags). This implies that the auto-correlation pattern can be explained by adding AR terms instead of adding MA terms. The number of AR terms added to the ARIMA model is same as the lag at which the PACF cuts off. AR signature is commonly associated with positive auto-correlation at lag - 1 i.e., it tends to arise in series which are slightly under-differenced. For example, in an AR (1) model for a series, *y* is expressed as follows (Chase Jr. 2009).

$$\begin{aligned} &(1-\phi_1 B)y_t = \mu \\ &\Rightarrow y_t - \phi_1 y_{t-1} = \mu \qquad (5.1) \end{aligned}$$

The AR term acts like a first difference if the AR coefficient is equal to 1. It does nothing if the AR coefficient is zero and acts like a partial difference if the coefficient is between 0 and 1. So, if the series is slightly under-differenced i.e., if the non-stationary pattern of positive auto-correlation has not been eliminated, it would require a partial difference by displaying an AR signature (Chase Jr. 2009).

However, this is not always the simplest way to explain a given pattern of auto-correlation. For example, sometimes it is more efficient to add MA terms instead of AR terms. Therefore, when the ACF displays a sharp cutoff while the PACF decays more slowly, the stationary series displays an "MA signature". This implies that the auto-correlation pattern can be explained by adding MA terms (lags of the forecast errors) instead of adding AR terms. It means the ACF plays the same role for MA terms that the PACF plays for AR terms i.e., the ACF tells how many MA terms are likely to be needed to remove the remaining auto-correlation from the differenced series. A MA signature is commonly associated with negative auto-correlation at lag - 1 i.e., it tends to arise in the series and is slightly over-differenced. The reason for this is that an MA term can "partially cancel" an order of differencing in the forecasting equation. To observe this, an ARIMA (0, 1, 1) model without constant is equivalent to a Simple Exponential Smoothing model as shown in the following equation (Chase Jr. 2009).

$$\begin{aligned} (1-B)^1 \, y_t &= \theta_1 B \varepsilon_t \\ \Rightarrow \; y_t - y_{t-1} &= \theta_1 \varepsilon_{t-1} \end{aligned} \qquad (5.2)$$

To evaluate the parameters for ARIMA model, the initial model was assumed as ARIMA (0, 0, 0). ACF and PACF plots of residuals of ARIMA (0, 0, 0) are shown in Figure 4. It is evident that the ACF shows slight tail off but PACF shows a sharper cutoff than the ACF. In particular, the PACF has four significant spikes, while the ACF has six. Moreover, auto-correlation of lag -1 is positive. Therefore, the differenced series is suggestive to an AR signature as it shows of slight under-differenced scenario. Four different AR models were considered for analysis: ARIMA (1, 0, 0), ARIMA (2, 0, 0), ARIMA (4, 0, 0), and ARIMA (5, 0, 0).

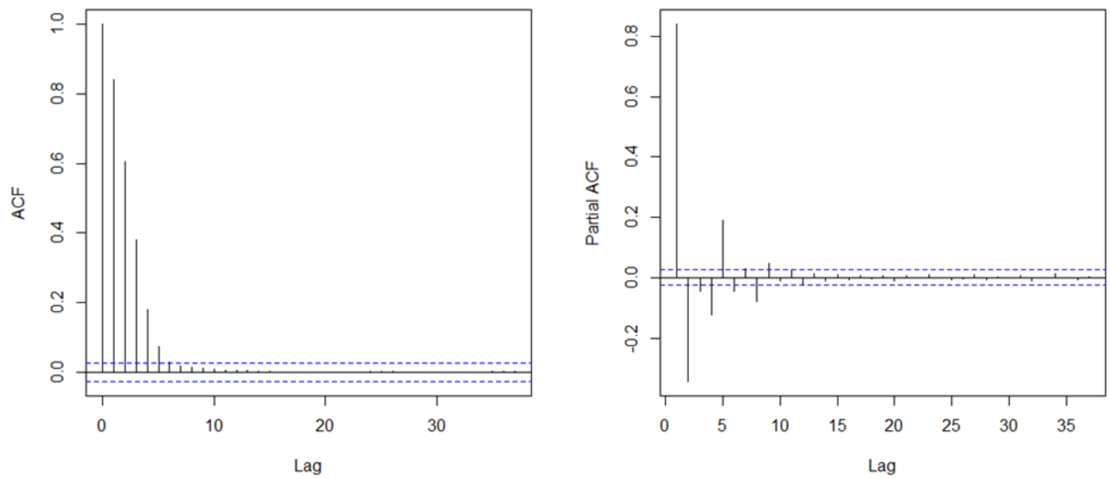



FIGURE 10: ACF and PACF plots of the residuals of the ARIMA (0, 0, 0) model

All four models were fit and $AIC_c$ values were computed (Table 8). Of these four models, the best is the ARIMA (2, 0, 0) as it has the smallest $AIC_c$ value (equal to -12673.3).

TABLE 8: Comparison of Different ARIMA Model

| Model | $AIC_c$ |
|---|---|
| ARIMA (1, 0, 0) | -11673.4 |
| ARIMA (2, 0, 0) | -12673.3 |
| ARIMA (4, 0, 0) | -12467.3 |
| ARIMA (5, 0, 0) | -12376.6 |

For next trial, residuals of ARIMA (2, 0, 0) model was fit. ACF and PACF plot of the residuals of the model is shown in Figure 11. The procedure is similar to the previous one with the exception that ACF shows lower number of significant spikes than PACF. Therefore, the differenced series is suggestive to an MA signature, as it shows slight over-differenced scenario. Two different ARIMA models were considered for analysis:

ARIMA (2, 0, 3) and ARIMA (2, 0, 4). ARIMA (2, 0, 3) ($AIC_c$ = -12740.48) shows lower $AIC_c$ value than ARIMA (2, 0, 4) ($AIC_c$ = -12739.96).

The presence of the unit root in the AR part of the model was checked i.e., whether the sum of the AR coefficients is close to one or not. If it is close to or equal to one, the number of AR terms needs to be reduced by one. Similar condition is applicable for MA signature when it is suggestive i.e., the sum of the MA coefficients is close to one or not.

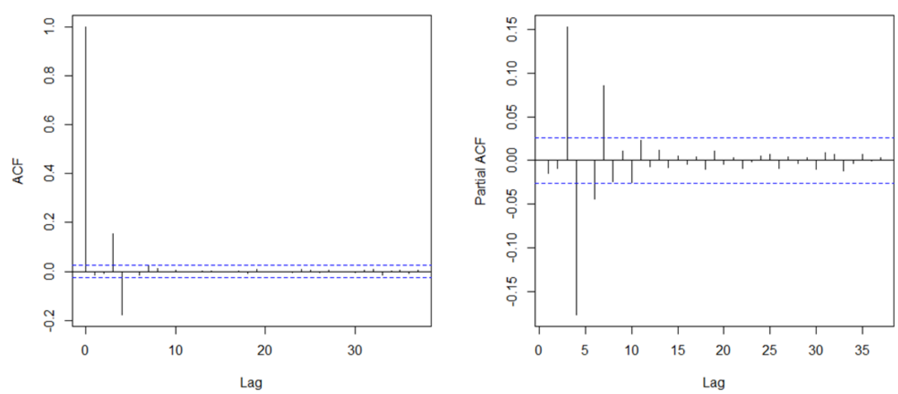



FIGURE 11: ACF and PACF plots of the residuals of the ARIMA (2, 0, 0) model

Following the aforementioned procedure, ACF and PACF of the residuals of ARIMA (2, 0, 3) were plotted (Figure 12). All the spikes in PACF are now within the significance limits and the residuals appear to be white noise (as explained in Chapter 3).

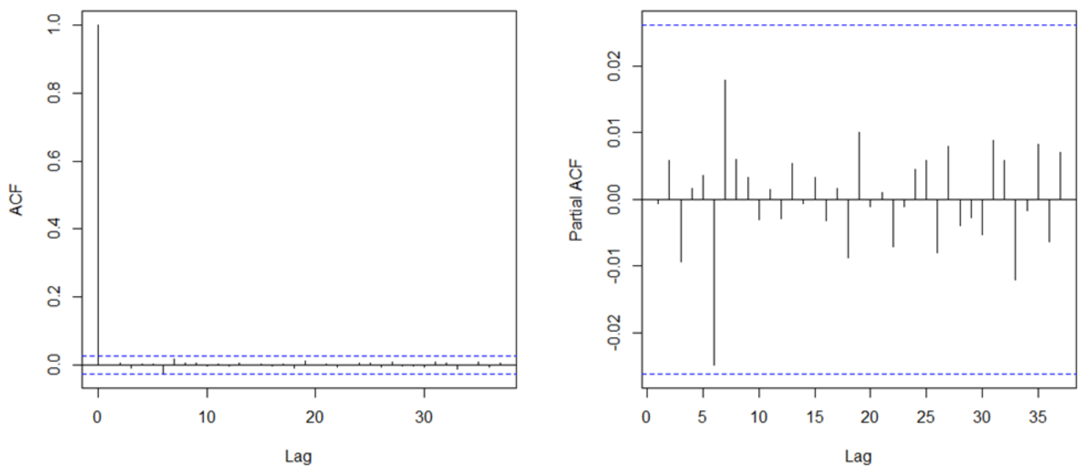



FIGURE 12: ACF and PACF plots of the residuals of the ARIMA (2, 0, 3) model

A Ljung - Box test was applied to test whether the residuals have any remaining auto-correlations or not. The Ljung - Box test is named after Greta M. Ljung and George E. P. Box. Instead of testing randomness at each distinct lag, it tests the "overall" randomness based on a number of lags instead of randomness at each distinct lag. Therefore, it is known as Portmanteau test. The hypothesis of the Ljung–Box test is expressed as follows (Box et al. 2008).

$H_0$: The data are independently distributed (i.e., the correlations in the population from which the sample is taken are zero, so that any observed correlations in the data result from randomness of the sampling process).

$H_a$: The data are not independently distributed.

A Ljung-Box test of ARIMA (2, 0, 3) is shown in Figure 13. The figure reveals that the residuals have no remaining auto-correlations as all the spikes are within the significant limit $\pm 1.96/\sqrt{T}$ (Box et al. 2008, Yang et al. 2014).

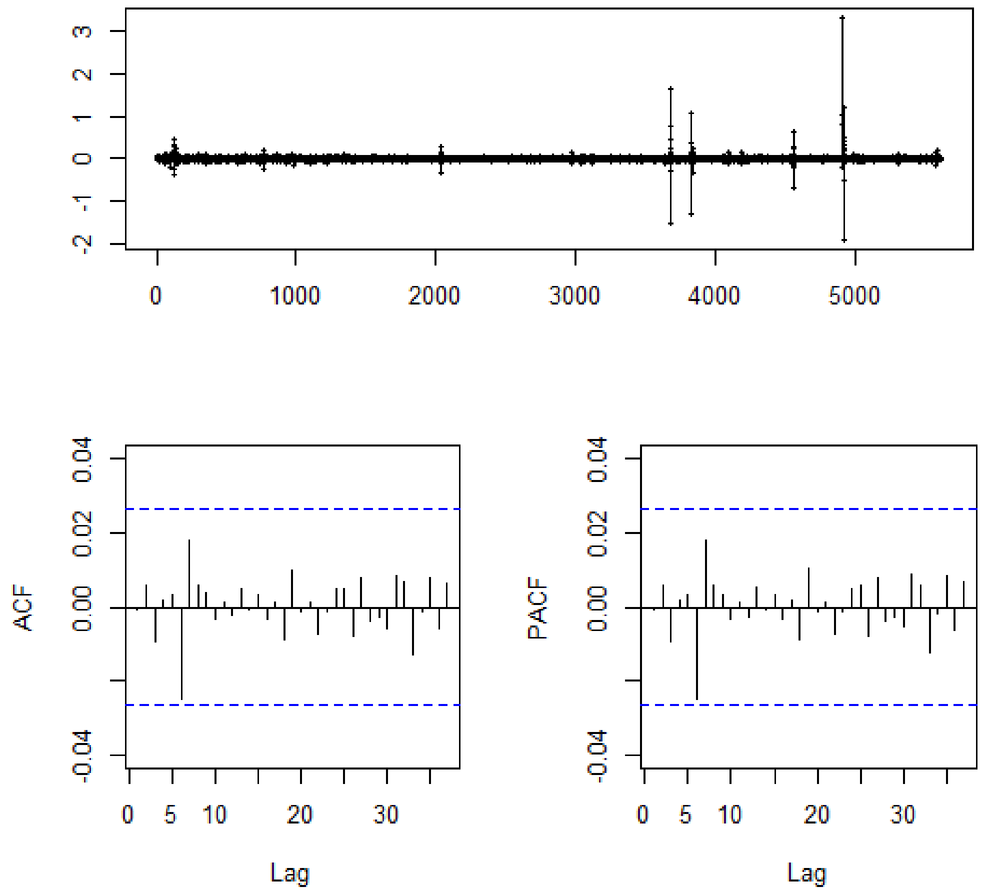



FIGURE 13: Ljung-Box test results

5.3.3 Identifying the Seasonal Part of the Model

The seasonal part of an ARIMA model has the same structure as the non-seasonal part. It can have an AR factor, a MA factor, and/or an order of differencing. However, in the seasonal part of the model, all of these factors operate across multiple lags (the number of periods in a season). A seasonal ARIMA model is classified as an ARIMA (p, d, q) (P, D, Q) model, where P = number of seasonal autoregressive (SAR) terms, D = number of seasonal differences, and Q = number of seasonal moving average (SMA) terms. Two rules need to be considered for finding appropriate seasonal ARIMA model (Chase Jr. 2009). They are listed next.

1. If the series has a strong and consistent seasonal pattern, then one single order of seasonal differencing needs to be used instead of multiple order of seasonal differencing or more than 2 orders of total differencing (seasonal + non-seasonal).
2. If the auto-correlation at the seasonal period is positive, SAR term could be added to the model. If the auto-correlation at the seasonal period is negative, SMA term could be added to the model. It would be better not to mix SAR and SMA terms in the same model, and to avoid using more than one of either kind.

5.4 Implementation of the Pre-whitening Cross-Correlation Function (CCF) and Lagged Regression for Forecasting

The CCF can be used to help identify the form of the transfer function appropriate for an input series. Cross-correlations are the correlations across time between pre-whitened values of Y and X. Pre-whitening is the process of making a series into white noise before in putting it into the model. This is required as the correlation structure of the X variable could hamper the estimates of the cross-correlation between X and Y. The pre-whitening technique assumes that the input variables do not depend on past values of the response variable. If there is feedback from the response variable to an input variable, as evidenced by significant cross-correlation at negative lags, both the input and the response variables need to be pre-whitened before meaningful cross-correlations can be computed. However, a few CCFs reveal so many significant spikes, which make it very difficult to find the most significant one to predict the variable. Therefore, to simplify this complicated auto-correlation process and to identify accurate patterns, the series must be pre-whitened first. The pre-whitening process involves following steps (Box et al. 2009).

1. An AR model with minimum AIC is fitted to $x_t$.
2. The $p$ and $q$ are used to pre-whiten $x_t$ and $y_t$ by multiplying with $\theta_x^{-1}(B)\Phi_x(B)(1-B)^d$ and generate two pre-whitened series $\alpha_t$ and $\beta_t$.
3. CCF is computed for the pre-whitened series $\alpha_t$ and $\beta_t$.

Figure 14 shows the difference between regular CCF with many significant spikes and pre-whitened CCF with few significant number of spikes such as k = 0, 1, -1 which can be used to predict $y_t$.

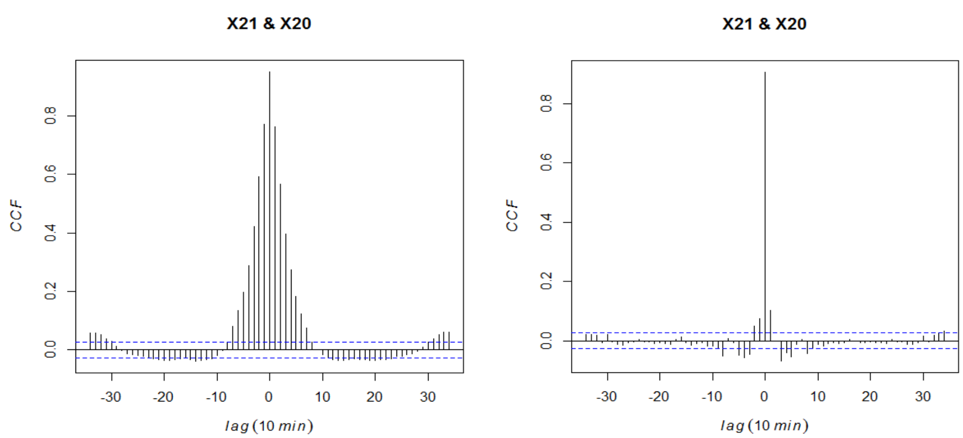



FIGURE 14: CCF and pre-whitened CCF

In forecasting, the leading factors for one segment should be identified first from pre-whitening cross-correlation factors. From these leading factors, lagged regression identifies which variables are statistically significant to predict the models. To assess spatial affects, different numbers of upstream and downstream segments were fitted and evaluated as predictor variables. The lagged regression model for segment X7 can be expressed as follows.

$$V_{X7,t} = \sum \phi_{s,k} V_{s,t-k} + \cdots + \varepsilon_k$$

Where, *k* is the time lag, s is the segment ID, and $\varepsilon_t$ is zero-mean uncorrelated error term. Since R-squared for the regression model always increases, as variables are added, R-squared is not an effective measurement for model comparison. Therefore, MAPE and mean absolute deviation (MAD) for holdout performance was used to compare different models (Yang et al. 2014).

5.5 Overall Methodology for ARIMA Model Building

The methodology discussed in this chapter is summarized using Figure 15.

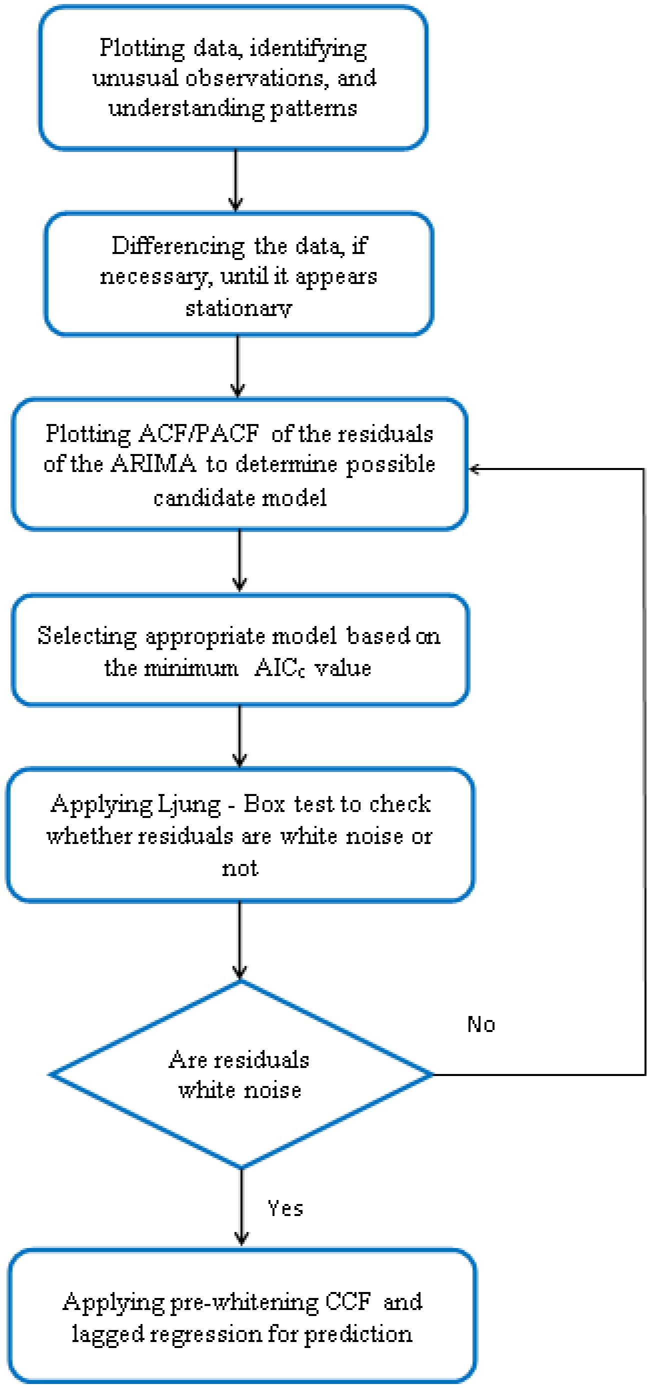



FIGURE 15: Overall methodology for ARIMA model building

# CHAPTER 6: ANALYSIS AND RESULTS

The first part of this chapter evaluates the effectiveness of Cronbach's α in estimating the expected travel time. This includes the reliability of the Cronbach's α value and comparison of expected travel time and observed travel time using real-world data. The second part of this chapter presents the results and analysis of the ARIMA model in forecasting travel time and relative variations in travel time. Besides, a comparative analysis was performed to assess if expected travel time or minimum travel time would better forecast the outputs. At the end, results for both with 'vehicle accident' and without incident condition are shown in a graphical format.

## 6.1 Identifying Corresponding Cronbach's α of a TMC Code

Four categories of Cronbach's α (from Table 9) were considered to compute the expected travel time of each segment. It was computed for each year as well as each day-of-the-week (i.e., Monday, Tuesday, etc.) separately. Sample output from analysis is shown in Table 8 for segment, X6 i.e., TMC code 125-04780. Cronbach's coefficient, α1 and α3, for each day-of-the-week from Saturday to Friday were computed for 49, 50, 50, 49, 51, 50, and 52 days, respectively. For weekday, α2 and α4, were computed for total 252 (=50+ 49+ 51+ 50+ 52) days and, for weekend, α2 and α4 were computed for total 99 (=49+50) days.

TABLE 9: Cronbach's α associated with varying categories, primary, and secondary factors for TMC code 125-04780

| TMC Code | DOW | WD | α1 | α2 | α3 | α4 | Max(α) |
|---|---|---|---|---|---|---|---|
| 125-04780 | 1 | 0 | 0.081 | 0.136 | 0.702 | **0.844** | 0.844 |
| 125-04780 | 2 | 1 | 0.895 | **0.982** | 0.847 | 0.922 | 0.982 |
| 125-04780 | 3 | 1 | 0.968 | **0.982** | 0.838 | 0.922 | 0.982 |
| 125-04780 | 4 | 1 | 0.863 | **0.982** | 0.887 | 0.922 | 0.982 |
| 125-04780 | 5 | 1 | 0.973 | **0.982** | 0.847 | 0.922 | 0.982 |
| 125-04780 | 6 | 1 | 0.938 | **0.982** | 0.874 | 0.922 | 0.982 |
| 125-04780 | 7 | 0 | 0.167 | 0.136 | **0.902** | 0.844 | 0.902 |

The Cronbach's α value selected for a TMC code is the maximum value. The corresponding travel time for this maximum value is selected as the expected travel time. For example, α2 (in bold font) for Monday is higher than any other computed Cronbach's α for the TMC code. Therefore, the travel time corresponding to this Cronbach's α is considered as the expected travel time on Monday for segment X6.

The corresponding Cronbach's α values of different segments based on 2010 data for different days of the week is shown in Table 10. The coefficients α1, α2, α3, and α4 are highlighted in green, red, yellow, and blue color, respectively. For most of the segments, α2 is higher than other coefficients on weekdays. This implies that it is beneficial to categorize by week-of-the year for the majority of the weekday trips rather than considering single day of different weeks. In other words, travel time of a specific time interval shows more similarity in a week rather than specific day-of-the-week of different weeks. Besides, time-of-the-day plays a vital role in planning a trip for weekdays as it varies throughout the day. It is comprehensible that morning and evening peak-hour traffic characteristics are different on a weekday when compared to the off-peak period. Travel time during weekends showed different trends than the weekday trips for all the segments. As α3 and α4 are higher during weekends, time-of-the-day does not play a significant role during weekends. Travel time patterns are consistent throughout the day on weekends.

TABLE 10: Cronbach's 'α for different segments - 2010 data

| Segment | Corresponding Cronbach's α in 2010 | | | | | | |
|---|---|---|---|---|---|---|---|
| | Saturday | Sunday | Monday | Tuesday | Wednesday | Thursday | Friday |
| X1 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X2 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X3 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X4 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X5 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X6 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X7 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X8 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X9 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X10 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X11 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X12 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X13 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X14 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X15 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X16 | α3 | α4 | α4 | α4 | α4 | α4 | α4 |
| X17 | α3 | α4 | α4 | α4 | α4 | α4 | α4 |
| X18 | α3 | α4 | α4 | α4 | α4 | α4 | α4 |
| X19 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X20 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X21 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X22 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X23 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X24 | α3 | α4 | α4 | α4 | α3 | α4 | α4 |
| X25 | α3 | α4 | α4 | α3 | α4 | α4 | α4 |
| X26 | α3 | α4 | α4 | α3 | α4 | α4 | α4 |
| X27 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X28 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |

Table 11 summarizes computed Cronbach's 'α's based on 2011 data. Observations indicate that trends in travel time on weekdays are different from weekends.

Time-of-the-day plays a major role in trip planning. During weekends, travel times vary and can be categorized as day-of-the-week or week-of-the-year dependent.

TABLE 11: Corresponding Cronbach's 'α for different segments in 2011

| Segment | Corresponding Cronbach's α in 2011 | | | | | | |
|---|---|---|---|---|---|---|---|
| | Saturday | Sunday | Monday | Tuesday | Wednesday | Thursday | Friday |
| X1 | α4 | α3 | α3 | α4 | α4 | α4 | α4 |
| X2 | α4 | α3 | α2 | α2 | α2 | α2 | α2 |
| X3 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X4 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X5 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X6 | α4 | α4 | α2 | α2 | α2 | α2 | α2 |
| X7 | α4 | α4 | α2 | α2 | α2 | α2 | α2 |
| X8 | α4 | α3 | α2 | α2 | α2 | α2 | α2 |
| X9 | α4 | α3 | α2 | α2 | α2 | α2 | α2 |
| X10 | α4 | α3 | α2 | α2 | α2 | α2 | α2 |
| X11 | α4 | α4 | α2 | α2 | α2 | α2 | α2 |
| X12 | α4 | α3 | α2 | α2 | α2 | α2 | α2 |
| X13 | α4 | α3 | α2 | α2 | α2 | α2 | α2 |
| X14 | α4 | α4 | α2 | α2 | α2 | α2 | α2 |
| X15 | α4 | α4 | α3 | α2 | α3 | α3 | α3 |
| X16 | α3 | α4 | α2 | α3 | α2 | α2 | α2 |
| X17 | α4 | α4 | α3 | α3 | α3 | α3 | α4 |
| X18 | α4 | α4 | α3 | α3 | α3 | α3 | α4 |
| X19 | α4 | α4 | α3 | α3 | α3 | α4 | α4 |
| X20 | α4 | α4 | α2 | α2 | α2 | α2 | α2 |
| X21 | α4 | α3 | α2 | α2 | α2 | α2 | α2 |
| X22 | α4 | α3 | α2 | α2 | α2 | α2 | α2 |
| X23 | α4 | α4 | α2 | α2 | α2 | α2 | α2 |
| X24 | α4 | α4 | α3 | α3 | α4 | α4 | α4 |
| X25 | α4 | α3 | α3 | α4 | α4 | α4 | α3 |
| X26 | α4 | α3 | α3 | α4 | α4 | α4 | α4 |
| X27 | α4 | α3 | α3 | α4 | α4 | α4 | α4 |
| X28 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |

Results from Table 12, also bolster the claim from the results of tables 10 and 11 that the fundamental difference between weekday and weekend trip planning is time-of-the-day. However, a striking observation from Table 12 is that several segments show a higher α2 value during weekends. This could imply that weekend travel patterns are mimicking weekday travel patterns over time.

TABLE 12: Corresponding Cronbach's 'α for different segments in 2012

| Segment | Corresponding Cronbach's α in 2012 | | | | | | |
|---|---|---|---|---|---|---|---|
| | Saturday | Sunday | Monday | Tuesday | Wednesday | Thursday | Friday |
| X1 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X2 | α2 | α2 | α2 | α2 | α2 | α2 | α2 |
| X3 | α2 | α2 | α2 | α2 | α2 | α2 | α2 |
| X4 | α2 | α2 | α2 | α2 | α2 | α2 | α2 |
| X5 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X6 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X7 | α4 | α3 | α2 | α2 | α2 | α2 | α2 |
| X8 | α3 | α3 | α2 | α2 | α2 | α2 | α2 |
| X9 | α4 | α3 | α2 | α2 | α2 | α2 | α2 |
| X10 | α4 | α3 | α2 | α2 | α2 | α2 | α2 |
| X11 | α4 | α3 | α2 | α2 | α2 | α2 | α2 |
| X12 | α4 | α3 | α2 | α2 | α2 | α2 | α2 |
| X13 | α4 | α4 | α2 | α2 | α2 | α2 | α2 |
| X14 | α4 | α4 | α2 | α2 | α2 | α2 | α2 |
| X15 | α2 | α2 | α2 | α2 | α2 | α2 | α2 |
| X16 | α2 | α2 | α2 | α2 | α2 | α2 | α2 |
| X17 | α2 | α2 | α2 | α2 | α2 | α2 | α2 |
| X18 | α2 | α2 | α2 | α2 | α2 | α2 | α3 |
| X19 | α2 | α2 | α2 | α2 | α2 | α2 | α2 |
| X20 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X21 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X22 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X23 | α3 | α4 | α2 | α2 | α2 | α2 | α2 |
| X24 | α2 | α2 | α4 | α3 | α4 | α4 | α4 |
| X25 | α1 | α2 | α4 | α3 | α4 | α1 | α1 |
| X26 | α2 | α2 | α1 | α2 | α1 | α1 | α1 |
| X27 | α2 | α2 | α2 | α2 | α2 | α2 | α2 |
| X28 | α2 | α3 | α2 | α2 | α2 | α2 | α2 |

6.2 Level of Reliability of Cronbach's α

The Cronbach's coefficient indicates the primary and secondary factors for similarity or consistency in travel time. However, it does not indicate the level of reliability. Corresponding Cronbach's α provides information about which coefficient to select but the level of reliability provides the extent of similarity among the travel-time series pattern. If the level of reliability is high, similarity among the travel time series are also high. In this case, the expected travel time would be close to the observed travel time. When the Cronbach's α value is greater than or equal to 0.9, then the level of reliability is "A", Cronbach's α value is between 0.7 to 0.9, then the level of reliability is "B"; Cronbach's α value is between 0.5 to 0.7, then the level of reliability is "C"; Cronbach's α value is between 0.4 to 0.5, then the level of reliability is "D"; and for value of less than 0.4, the level of reliability is "E". Generally, the level of reliability A is denoted as "Excellently Reliable", B as "Highly Reliable", C as "Reliable", D as "Poorly Reliable", and E as "Unreliable". The computed level of reliability categories A, B, C, D, and E are colored as red, blue, yellow, green, and orange and presented in tables 13-15 for years 2010, 2011 and 2012.

From visual inspection, it is clear that red is dominant in Table 13 for year 2010. Out of 196 values presented in Table 13, 136 samples (69.38%) values are greater than or equal to 0.9 ( "Excellently Reliable"). About 25% of the values are categorized as "Highly Reliable". None of the values indicates poor or unreliable. The level of reliability on Sunday is relatively lower than other day-of-the-week.

Similar to Table 13, Table 14 shows that 126 (64.28%) of the values are greater than or equal to 0.9 and classified as "Excellently Reliable", while 68 samples (34.69%)

are classified as “Highly Reliable”. None of the values indicates poor or unreliable. Unlike in 2010, weekend indicates a “Highly Reliable” condition.

TABLE 13: Maximum Cronbach’s α associated with different TMC codes and day-of-the-week in 2010

| Segment | Maximum Cronbach's α in 2010 | | | | | | |
|---|---|---|---|---|---|---|---|
| | Saturday | Sunday | Monday | Tuesday | Wednesday | Thursday | Friday |
| X1 | 0.93 | 0.81 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 |
| X2 | 0.94 | 0.82 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X3 | 0.79 | 0.75 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| X4 | 0.90 | 0.79 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X5 | 0.84 | 0.76 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X6 | 0.90 | 0.84 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X7 | 0.94 | 0.89 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X8 | 0.89 | 0.81 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| X9 | 0.85 | 0.82 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| X10 | 0.96 | 0.77 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X11 | 0.94 | 0.78 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| X12 | 0.94 | 0.78 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| X13 | 0.97 | 0.76 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X14 | 0.97 | 0.79 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 |
| X15 | 0.96 | 0.78 | 0.93 | 0.93 | 0.93 | 0.93 | 0.93 |
| X16 | 0.96 | 0.75 | 0.90 | 0.90 | 0.90 | 0.90 | 0.90 |
| X17 | 0.93 | 0.77 | 0.87 | 0.87 | 0.87 | 0.87 | 0.87 |
| X18 | 0.93 | 0.84 | 0.90 | 0.90 | 0.90 | 0.90 | 0.90 |
| X19 | 0.95 | 0.82 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 |
| X20 | 0.95 | 0.73 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X21 | 0.94 | 0.64 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X22 | 0.93 | 0.76 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| X23 | 0.89 | 0.84 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| X24 | 0.93 | 0.91 | 0.84 | 0.84 | 0.86 | 0.84 | 0.84 |
| X25 | 0.94 | 0.93 | 0.90 | 0.92 | 0.90 | 0.90 | 0.90 |
| X26 | 0.84 | 0.78 | 0.87 | 0.89 | 0.87 | 0.87 | 0.87 |
| X27 | 0.92 | 0.91 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| X28 | 0.92 | 0.91 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |

TABLE 14: Maximum Cronbach's α associated with different TMC codes and day-of-the-week in 2011

| Segment | Maximum Cronbach's α in 2011 | | | | | | |
|---|---|---|---|---|---|---|---|
| | Saturday | Sunday | Monday | Tuesday | Wednesday | Thursday | Friday |
| X1 | 0.86 | 0.90 | 0.96 | 0.92 | 0.92 | 0.92 | 0.92 |
| X2 | 0.89 | 0.93 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X3 | 0.77 | 0.68 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X4 | 0.85 | 0.79 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X5 | 0.87 | 0.83 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X6 | 0.90 | 0.90 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X7 | 0.80 | 0.80 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X8 | 0.78 | 0.80 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| X9 | 0.72 | 0.93 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| X10 | 0.78 | 0.86 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| X11 | 0.84 | 0.84 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| X12 | 0.85 | 0.88 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X13 | 0.87 | 0.92 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| X14 | 0.89 | 0.89 | 0.90 | 0.90 | 0.90 | 0.90 | 0.90 |
| X15 | 0.91 | 0.91 | 0.82 | 0.70 | 0.88 | 0.80 | 0.89 |
| X16 | 0.88 | 0.88 | 0.87 | 0.87 | 0.87 | 0.87 | 0.87 |
| X17 | 0.92 | 0.92 | 0.96 | 0.85 | 0.87 | 0.85 | 0.84 |
| X18 | 0.92 | 0.92 | 0.96 | 0.85 | 0.87 | 0.85 | 0.84 |
| X19 | 0.93 | 0.93 | 0.97 | 0.85 | 0.86 | 0.85 | 0.85 |
| X20 | 0.92 | 0.92 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| X21 | 0.84 | 0.87 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| X22 | 0.89 | 0.90 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| X23 | 0.92 | 0.92 | 0.93 | 0.93 | 0.93 | 0.93 | 0.93 |
| X24 | 0.89 | 0.89 | 0.88 | 0.88 | 0.87 | 0.87 | 0.87 |
| X25 | 0.92 | 0.92 | 0.87 | 0.86 | 0.86 | 0.86 | 0.89 |
| X26 | 0.77 | 0.88 | 0.97 | 0.94 | 0.94 | 0.94 | 0.94 |
| X27 | 0.73 | 0.82 | 0.95 | 0.92 | 0.92 | 0.92 | 0.92 |
| X28 | 0.87 | 0.85 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |

Table 15 reveals that 57.14% of the values are greater than or equal to 0.9, indicating "Excellent Reliability". About 38.78% of the values indicate "Highly Reliable" condition. The number of values between 0.5 to 0.7 is 4.59%), highest compared to other study years. Saturday and Sunday either are with level of reliability B or lower.

TABLE 15: Maximum Cronbach's α associated with different TMC codes and day-of-the-week in 2012

| Segment | Maximum Cronbach's α in 2012 | | | | | | |
|---|---|---|---|---|---|---|---|
| | Saturday | Sunday | Monday | Tuesday | Wednesday | Thursday | Friday |
| X1 | 0.72 | 0.68 | 0.85 | 0.85 | 0.85 | 0.85 | 0.85 |
| X2 | 0.79 | 0.79 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X3 | 0.82 | 0.82 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X4 | 0.72 | 0.72 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X5 | 0.81 | 0.61 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X6 | 0.83 | 0.82 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X7 | 0.78 | 0.79 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| X8 | 0.85 | 0.84 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| X9 | 0.81 | 0.82 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| X10 | 0.87 | 0.88 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| X11 | 0.80 | 0.81 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| X12 | 0.86 | 0.87 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| X13 | 0.69 | 0.69 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 |
| X14 | 0.69 | 0.69 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 |
| X15 | 0.74 | 0.74 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 |
| X16 | 0.68 | 0.68 | 0.93 | 0.93 | 0.93 | 0.93 | 0.93 |
| X17 | 0.92 | 0.92 | 0.90 | 0.90 | 0.90 | 0.90 | 0.90 |
| X18 | 0.85 | 0.85 | 0.83 | 0.83 | 0.83 | 0.83 | 0.84 |
| X19 | 0.89 | 0.89 | 0.86 | 0.86 | 0.86 | 0.86 | 0.86 |
| X20 | 0.89 | 0.86 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| X21 | 0.90 | 0.89 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| X22 | 0.89 | 0.89 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| X23 | 0.84 | 0.82 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 |
| X24 | 0.92 | 0.92 | 0.88 | 0.89 | 0.88 | 0.88 | 0.88 |
| X25 | 0.90 | 0.88 | 0.86 | 0.88 | 0.86 | 0.91 | 0.87 |
| X26 | 0.94 | 0.94 | 0.93 | 0.88 | 0.92 | 0.93 | 0.92 |
| X27 | 0.89 | 0.89 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| X28 | 0.83 | 0.87 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |

From Table 13 to Table 15, it is observed that Cronbach's α value is gradually decreasing for weekend and so is the level of reliability. The characteristics of weekend travel time pattern is shifing towards time-of-the-day dependent pattern, indicating the fact that Charlotte is growing with increasing weekend activity. However, weekend trip has yet not become totally time-of-the-day dependent like weekday. Therefore similarity is revelaing mixed vibe of both weekday and weekend and similarity in travel-time series pattern has been decreased.

6.3 Comparison of Expected Travel Time and Observed Travel Time

The average of all the samples in a year for each time interval at segment-level was used to estimate the expected travel time based on the category of the corresponding Cronbach's α. To check the validity, the expected travel time was then compared with the observed travel time for each time interval. The percent difference in the estimated and observed value is computed using the following equation.

$$\%\,\text{Difference} = \frac{\left|\text{Observed Travel Time - Expected Travel Time}\right|}{\text{Observed Travel Time}}$$

Moreover, the number of samples by percent difference between the expected travel time and observed travel time was computed for each segment. The maximum number of 10-minute intervals during a year is = 6 / hour × 24 hours × 365 days = 52,560. This was used to compute the percent of samples based on the percent difference in expected travel time and observed travel time.

Table 16 shows the percent of samples by percent difference between the estimated expected travel time and the observed travel time for each selected segment during 2010. The percent difference is less than or equal to 10% for most of the segments except X6, X20, and X21 in 2010. From segment length perspective, it is hard to explain as length is both below and above 0.5 miles. Therefore, this needs to be researched to find out the exact reason of this difference. More than 30% difference is observed for less than 5% of the samples for almost all the segments (except X20 and X21).

Table 17 shows the percent of samples by percent difference between the estimated expected travel time and the observed travel time for each selected segment during 2011. The percent difference is less than or equal to 10% for all the segments except X12 and X13 in 2011. More than 30% difference is observed for segment X12 and X13. For all other segments, the percent difference is below 5%.

Table 18 shows the percent of samples by percent difference between the estimated expected travel time and the observed travel time for each selected segment during 2012. Except X11, X12, X13, X14, and X15, the percent difference is less than or equal to 10% for all other segments in 2012.

Overall, the comparison shows that travel time estimated based on Cronbach's α can be an effective means to measure the expected travel time.

TABLE 16: Percentage of sample with percentage difference between expected and observed travel time in 2010

| Segment | % Difference Between Expected and Observed Travel Time in 2010 | | | | | | |
|---|---|---|---|---|---|---|---|
| | ≤5 | 5<x≤10 | 10<x≤15 | 15<x≤20 | 20<x≤25 | 25<x≤30 | >30 |
| X1 | 70.3 | 18.6 | 5.1 | 2.4 | 1.2 | 0.8 | 1.6 |
| X2 | 68.4 | 20.4 | 5.1 | 2.1 | 1.2 | 0.9 | 2.3 |
| X3 | 62.6 | 22.5 | 6.8 | 2.6 | 1.3 | 1.0 | 3.4 |
| X4 | 70.2 | 18.3 | 5.0 | 2.1 | 1.1 | 0.8 | 2.6 |
| X5 | 70.5 | 17.8 | 5.0 | 2.1 | 1.2 | 0.9 | 2.5 |
| X6 | 61.2 | 18.5 | 6.2 | 3.1 | 1.9 | 1.7 | 7.4 |
| X7 | 65.3 | 20.9 | 5.5 | 2.4 | 1.6 | 1.0 | 3.3 |
| X8 | 60.4 | 23.0 | 10.2 | 3.2 | 1.2 | 0.6 | 1.4 |
| X9 | 64.3 | 27.3 | 5.9 | 1.5 | 0.3 | 0.1 | 0.5 |
| X10 | 71.3 | 18.4 | 4.9 | 1.9 | 0.9 | 0.6 | 2.0 |
| X11 | 68.6 | 20.1 | 5.5 | 1.6 | 1.1 | 0.5 | 2.6 |
| X12 | 80.8 | 15.3 | 2.5 | 0.6 | 0.2 | 0.2 | 0.4 |
| X13 | 81.0 | 15.3 | 2.5 | 0.5 | 0.2 | 0.2 | 0.4 |
| X14 | 77.0 | 15.6 | 3.7 | 1.7 | 0.8 | 0.5 | 0.7 |
| X15 | 77.6 | 16.8 | 4.1 | 0.9 | 0.3 | 0.1 | 0.3 |
| X16 | 68.0 | 19.2 | 5.5 | 2.3 | 1.5 | 0.9 | 2.7 |
| X17 | 65.3 | 20.2 | 6.0 | 2.4 | 1.3 | 0.9 | 3.6 |
| X18 | 67.8 | 18.7 | 5.7 | 2.2 | 1.3 | 0.9 | 3.4 |
| X19 | 68.7 | 17.9 | 5.5 | 2.5 | 1.4 | 1.0 | 3.0 |
| X20 | 61.2 | 18.5 | 6.2 | 3.1 | 1.9 | 1.7 | 7.4 |
| X21 | 66.1 | 16.3 | 5.0 | 2.4 | 1.9 | 1.6 | 6.7 |
| X22 | 65.4 | 19.5 | 6.1 | 2.8 | 1.8 | 1.2 | 3.2 |
| X23 | 68.2 | 23.1 | 6.2 | 1.3 | 0.3 | 0.2 | 0.6 |
| X24 | 79.4 | 15.8 | 3.5 | 0.7 | 0.2 | 0.1 | 0.4 |
| X25 | 71.3 | 18.3 | 4.4 | 1.6 | 0.9 | 0.6 | 2.9 |
| X26 | 72.0 | 17.9 | 5.5 | 2.0 | 1.2 | 0.6 | 0.8 |
| X27 | 77.2 | 18.1 | 3.5 | 0.7 | 0.2 | 0.1 | 0.3 |
| X28 | 77.0 | 15.6 | 3.7 | 1.7 | 0.8 | 0.5 | 0.7 |

TABLE 17: Percentage of sample with percentage difference between expected and observed travel time in 2011

| Segment | % Difference Between Expected and Observed Travel Time in 2011 | | | | | | |
|---|---|---|---|---|---|---|---|
| | ≤5 | 5<x≤10 | 10<x≤15 | 15<x≤20 | 20<x≤25 | 25<x≤30 | >30 |
| X1 | 82.4 | 13.6 | 2.6 | 0.7 | 0.2 | 0.2 | 0.3 |
| X2 | 77.3 | 13.9 | 3.6 | 1.8 | 1.1 | 0.7 | 1.2 |
| X3 | 76.8 | 13.1 | 3.9 | 2.0 | 1.3 | 0.9 | 2.5 |
| X4 | 73.8 | 15.3 | 4.1 | 2.0 | 1.3 | 0.9 | 2.5 |
| X5 | 68.6 | 18.9 | 4.5 | 2.0 | 1.4 | 1.1 | 4.0 |
| X6 | 63.6 | 23.9 | 5.0 | 2.4 | 1.3 | 0.9 | 3.4 |
| X7 | 70.8 | 17.8 | 4.3 | 1.8 | 1.3 | 1.0 | 3.4 |
| X8 | 70.8 | 16.5 | 4.9 | 2.3 | 1.4 | 0.9 | 3.1 |
| X9 | 71.4 | 17.0 | 4.4 | 2.1 | 1.4 | 1.0 | 3.4 |
| X10 | 70.1 | 19.0 | 4.6 | 1.8 | 1.1 | 0.7 | 2.6 |
| X11 | 66.8 | 18.8 | 6.1 | 2.6 | 1.3 | 0.9 | 3.9 |
| X12 | 57.6 | 20.4 | 7.5 | 3.6 | 2.1 | 1.7 | 7.4 |
| X13 | 65.7 | 16.1 | 5.5 | 2.6 | 1.8 | 1.5 | 7.3 |
| X14 | 67.6 | 20.7 | 4.3 | 2.6 | 1.5 | 1.2 | 2.6 |
| X15 | 62.4 | 21.2 | 6.4 | 3.0 | 1.8 | 1.6 | 4.2 |
| X16 | 66.3 | 21.8 | 5.9 | 2.1 | 1.5 | 0.8 | 2.1 |
| X17 | 76.3 | 16.6 | 3.8 | 1.3 | 0.8 | 0.4 | 1.4 |
| X18 | 76.4 | 16.5 | 3.8 | 1.3 | 0.8 | 0.4 | 1.4 |
| X19 | 82.1 | 14.0 | 2.6 | 0.6 | 0.3 | 0.2 | 0.7 |
| X20 | 73.1 | 18.1 | 4.1 | 1.6 | 0.8 | 0.6 | 2.2 |
| X21 | 76.1 | 14.9 | 3.6 | 1.4 | 0.8 | 0.5 | 3.3 |
| X22 | 72.7 | 19.7 | 2.7 | 1.0 | 0.9 | 0.5 | 3.0 |
| X23 | 71.3 | 20.0 | 4.1 | 1.8 | 1.4 | 1.0 | 1.0 |
| X24 | 81.9 | 14.3 | 3.0 | 0.5 | 0.3 | 0.2 | 0.3 |
| X25 | 82.6 | 14.4 | 3.0 | 0.3 | 0.1 | 0.0 | 0.2 |
| X26 | 86.7 | 11.3 | 2.3 | 0.2 | 0.0 | 0.0 | 0.1 |
| X27 | 81.8 | 15.0 | 3.2 | 0.3 | 0.1 | 0.0 | 0.1 |
| X28 | 81.3 | 12.8 | 3.6 | 1.5 | 0.7 | 0.3 | 0.4 |

TABLE 18: Percentage of sample with percentage difference between expected and observed travel time in 2012

| Segment | % Difference Between Expected and Observed Travel Time in 2012 | | | | | | |
|---|---|---|---|---|---|---|---|
| | ≤5 | 5<x≤10 | 10<x≤15 | 15<x≤20 | 20<x≤25 | 25<x≤30 | >30 |
| X1 | 85.3 | 11.9 | 1.9 | 0.4 | 0.1 | 0.1 | 0.3 |
| X2 | 77.1 | 14.1 | 3.8 | 1.7 | 1.1 | 0.8 | 1.5 |
| X3 | 76.5 | 12.8 | 3.8 | 1.9 | 1.1 | 0.9 | 3.0 |
| X4 | 73.7 | 14.3 | 4.5 | 2.1 | 1.2 | 1.0 | 3.0 |
| X5 | 71.3 | 15.3 | 4.3 | 2.0 | 1.3 | 1.1 | 4.6 |
| X6 | 66.5 | 19.3 | 5.7 | 2.2 | 1.4 | 1.0 | 3.8 |
| X7 | 72.5 | 14.8 | 4.1 | 2.1 | 1.4 | 1.1 | 4.0 |
| X8 | 72.9 | 15.1 | 4.0 | 2.2 | 1.5 | 1.0 | 3.4 |
| X9 | 70.6 | 16.2 | 4.3 | 2.3 | 1.6 | 1.1 | 3.9 |
| X10 | 68.9 | 16.9 | 5.8 | 2.6 | 1.5 | 0.9 | 3.4 |
| X11 | 65.3 | 16.9 | 7.2 | 3.5 | 1.7 | 1.1 | 4.3 |
| X12 | 59.2 | 16.7 | 6.8 | 3.8 | 3.3 | 2.1 | 8.2 |
| X13 | 66.6 | 17.4 | 5.0 | 3.0 | 1.1 | 1.4 | 5.5 |
| X14 | 66.6 | 17.4 | 5.0 | 3.0 | 1.1 | 1.4 | 5.5 |
| X15 | 66.6 | 17.0 | 5.4 | 2.7 | 1.3 | 1.1 | 5.9 |
| X16 | 64.1 | 21.9 | 5.2 | 2.0 | 1.8 | 0.9 | 4.1 |
| X17 | 75.9 | 14.0 | 3.4 | 1.9 | 1.2 | 0.9 | 2.6 |
| X18 | 76.6 | 15.1 | 4.2 | 1.9 | 0.8 | 0.4 | 1.0 |
| X19 | 81.6 | 12.0 | 3.4 | 1.3 | 0.7 | 0.3 | 0.7 |
| X20 | 74.4 | 15.7 | 4.1 | 1.7 | 1.1 | 0.7 | 2.4 |
| X21 | 79.7 | 11.9 | 2.3 | 1.2 | 0.8 | 0.5 | 3.5 |
| X22 | 73.3 | 19.0 | 2.3 | 1.1 | 0.7 | 0.7 | 2.9 |
| X23 | 78.4 | 14.0 | 3.1 | 1.4 | 1.0 | 0.8 | 1.2 |
| X24 | 82.4 | 14.4 | 2.2 | 0.4 | 0.2 | 0.1 | 0.5 |
| X25 | 84.5 | 13.5 | 1.1 | 0.2 | 0.2 | 0.1 | 0.4 |
| X26 | 89.2 | 10.0 | 0.6 | 0.1 | 0.0 | 0.0 | 0.1 |
| X27 | 82.5 | 14.7 | 2.1 | 0.4 | 0.1 | 0.1 | 0.1 |
| X28 | 83.0 | 11.9 | 2.9 | 1.2 | 0.5 | 0.2 | 0.3 |

## 6.4 Model Building for Travel Time Forecasting

For model building, a total 250 days of data was considered. These include 100 days without incident and 150 days with “vehicle accident” data. Pre-whitened CCF narrowed down the search horizon for predictor variables, and lagged regression model helped to identity the significant predictor variables for building the model for each segment. To develop the lagged regression model, different leading and lagging time periods such as *t*+10, *t*+20, *t*+ 30… and *t*-10, *t*-20, *t*-30…, respectively were considered. Symbol $t \pm$ n defines the n minutes before or after time *t*. To evaluate the model for each segment, MAPE and MAD were used instead of $R^2$ as the value of $R^2$ value increases with the addition of new or dummy variables. The target segment and its lagged regression model with its appropriate predictor variables to forecast travel time is shown in Table 19. Each fitted model consisted of both upstream and downstream segments travel time. However, a few segments along this study corridor does not have sufficient upstream or downstream segments for lagged regression analysis. For example, X1 does not have any downstream segment and X28 does not have upstream segment. Therefore, for analysis, only segments X6 to X24 were considered.

MAPE is less than 10% for almost all the segments except segment X19. However, MAD values are less than 10% for all the segments. Therefore, the overall prediction errors are reasonable and the forecasted value is close to the observed value. From Table 19, it is evident that the time lag for all significant predictor variables is 10 minutes.  Moreover, each target segment has predictor variables representing both upstream and downstream segments.

TABLE 19: Performance comparison of fitted models for travel time forecasting

| Target Segment | Lagged Regression Model with Predictor Variables | Holdout Performance | |
|---|---|---|---|
| | | MAPE (%) | MAD |
| $X6_t$ | $0.01+0.02X4_{t-10}+0.22X5_{t-10}+0.18X7_{t-10}+0.02X8_{t-10}+0.06X14_{t-10}-0.56X15_{t-10}+0.25X16_{t-10}$ | 8.82 | 0.14 |
| $X7_t$ | $0.02+0.19X5_{t-10}+0.37X6_{t-10}+0.09X8_{t-10}+0.02X10_{t-10}+0.01X11_{t-10}-0.01X13_{t-10}+0.06X14_{t-10}$ | 8.63 | 0.14 |
| $X8_t$ | $-0.12+0.19X1_{t-10}-0.08X4_{t-10}+0.39X5_{t-10}+0.77X6_{t-10}+1.41X7_{t-10}+0.25X9_{t-10}+0.31X10_{t-10}+0.05X11_{t-10}-0.03X12_{t-10}-0.02X13_{t-10}$ | 9.55 | 0.15 |
| $x9_t$ | $0.03+0.11X1_{t-10}-0.08X2_{t-10}+0.04X3_{t-10}+0.17X7_{t-10}+0.3X8_{t-10}+0.29X10_{t-10}-0.05X11_{t-10}-0.06X18_{t-10}$ | 9.83 | 0.16 |
| $X10_t$ | $0.09+0.18X1_{t-10}+0.13X2_{t-10}-0.09X3_{t-10}+0.11X4_{t-10}+0.21X8_{t-10}+0.15X9_{t-10}+0.13X11_{t-10}+0.05X12_{t-10}+0.09X13_{t-10}$ | 9.93 | 0.16 |
| $X11_t$ | $-0.02-0.18X1_{t-10}+0.14X4_{t-10}-0.24X5_{t-10}+0.09X8_{t-10}-0.07X9_{t-10}+0.15X10_{t-10}+0.27X12_{t-10}+0.21X13_{t-10}+0.14X17_{t-10}-0.45X18_{t-10}$ | 9.87 | 0.16 |
| $X12_t$ | $0.02+0.2X3_{t-10}-0.22X4_{t-10}+0.28X5_{t-10}-0.09X8_{t-10}+0.27X10_{t-10}+1.03X11_{t-10}-0.09X17_{t-10}+0.17X18_{t-10}+0.51X19_{t-10}-0.13X20_{t-10}$ | 8.76 | 0.15 |
| $X13_t$ | $-0.2-0.29X2_{t-10}+0.48X5_{t-10}-0.21X8_{t-10}+0.53X10_{t-10}+1.05X11_{t-10}-0.11X12_{t-10}+12.4X15_{t-10}-0.35X17_{t-10}+1.05X18_{t-10}-0.33X19_{t-10}$ | 6.11 | 0.11 |
| $X14_t$ | $0.01+0.03X1_{t-10}+0.01X2_{t-10}-0.01X3_{t-10}+0.01X4_{t-10}+0.02X6_{t-10}+0X13_{t-10}+2.58X15_{t-10}-0.03X17_{t-10}+0.02X18_{t-10}+0.01X19_{t-10}$ | 9.15 | 0.14 |
| $X15_t$ | $0-0.01X1_{t-10}+0X4_{t-10}+0X13_{t-10}+0.26X14_{t-10}+0.08X16_{t-10}+0.01X17_{t-10}-0.01X18_{t-10}+0X20_{t-10}$ | 8.45 | 0.18 |
| $X16_t$ | $0-0.06X5_{t-10}+0.17X6_{t-10}+0X13_{t-10}-0.07X14_{t-10}+1.39X15_{t-10}+0.05X17_{t-10}-0.01X19_{t-10}+0.01X20_{t-10}-0.01X21_{t-10}$ | 9.04 | 0.17 |
| $X17_t$ | $0.08-0.13X6_{t-10}+0.19X11_{t-10}-0.05X12_{t-10}-0.1X13_{t-10}-2.47X14_{t-10}+9.97X15_{t-10}+2.46X16_{t-10}+0.39X18_{t-10}+0.15X19_{t-10}+0.22X20_{t-10}-0.08X21_{t-10}+0.08X26_{t-10}$ | 9.96 | 0.20 |
| $X18_t$ | $-0.13-0.22X6_{t-10}+0.06X8_{t-10}-0.09X9_{t-10}-0.58X11_{t-10}+0.07X12_{t-10}+0.31X13_{t-10}+1.78X14_{t-10}-7X15_{t-10}+0.44X17_{t-10}+0.18X19_{t-10}+0.04X23_{t-10}$ | 9.11 | 0.16 |
| $X19_t$ | $0.06-0.12X11_{t-10}+0.22X12_{t-10}-0.13X13_{t-10}+0.5X14_{t-10}-0.53X16_{t-10}+0.17X17_{t-10}+0.27X18_{t-10}+0.24X20_{t-10}+0.18X25_{t-10}+0.05X27_{t-10}$ | 14.68 | 0.25 |
| $X20_t$ | $0.01-0.44X15_{t-10}+0.24X16_{t-10}+0.12X17_{t-10}+0.09X19_{t-10}+0.2X21_{t-10}+1.39X22_{t-10}+0.06X23_{t-10}$ | 9.28 | 0.18 |
| $X21_t$ | $-0.01+1.1X15_{t-10}-0.46X16_{t-10}-0.09X17_{t-10}+0.03X18_{t-10}+0.5X20_{t-10}+0.99X22_{t-10}+0.37X23_{t-10}-0.04X24_{t-10}$ | 7.47 | 0.15 |
| $X22_t$ | $0-0.01X19_{t-10}+0.08X20_{t-10}+0.02X21_{t-10}+0.07X23_{t-10}+0.01X24_{t-10}-0.01X26_{t-10}+0X27_{t-10}$ | 8.46 | 0.13 |
| $X23_t$ | $0.08+0.06X19_{t-10}+0.09X20_{t-10}+0.28X21_{t-10}+2.06X22_{t-10}+0.29X24_{t-10}-0.01X27_{t-10}$ | 9.34 | 0.13 |
| $X24_t$ | $0.17-0.02X19_{t-10}-0.03X21_{t-10}+0.64X22_{t-10}+0.5X23_{t-10}+0.83X25_{t-10}+0.11X26_{t-10}+0.05X28_{t-10}$ | 6.40 | 0.13 |

## 6.5 Model Building for Forecasting Travel Time / Expected Travel Time and Travel Time / Minimum Travel Time

The computed MAPE and MAD from the observed and forecasted travel time / expected travel time (TT/ExpTT) for each segment are shown in Table 20. Alike travel time model, MAPE values are less than 10% for almost all the segments except X11 (but less than 15%). The MAD values are also low. This indicates that the difference between the forecasted and the observed travel time estimates is also very low. The predictor variables are from both upstream and downstream of the target segment. Moreover, for all the models to forecast TT/ExpTT, all significant predictor variables are in time lag of 10 minutes.

The computed MAPE and MAD from the observed and forecasted travel time / minimum travel time (TT/MinTT) for each segment is shown in Table 21. MAPE value is less than 10% for all the segments except X11 and X13 (but less than 15%). The MAD values are also low. Like in the case of travel time and TT/ExpTT, all significant predictor variables are in time lag of 10 minutes for TT/MinTT models.

A comparison of results between the fitted models from Table 22 shows that TT/ExpTT performs better and leads to relatively accurate estimates than TT/MinTT, except for segments X12, X19, and X24.

TABLE 20: Fitted model for forecasting travel time / expected travel time

| Target Segment | Lagged Regression Model with Predictor Variables | Holdout Performance | |
|---|---|---|---|
| | | MAPE (%) | MAD |
| $X6_t$ | $0.09+0.52X5_{t-10}+0.29X7_{t-10}+0.07X8_{t-10}+0.03X13_{t-10}-0.1X15_{t-10}+0.14X16_{t-10}-0.04X17_{t-10}$ | 9.21 | 0.15 |
| $X7_t$ | $0.08+0.24X5_{t-10}+0.28X6_{t-10}+0.32X8_{t-10}+0.02X9_{t-10}+0.07X10_{t-10}-0.03X13_{t-10}+0.05X14_{t-10}-0.03X16_{t-10}$ | 9.05 | 0.14 |
| $X8_t$ | $0+0X1_{t-10}+0X6_{t-10}+0X7_{t-10}+0X9_{t-10}+0X10_{t-10}+0X17_{t-10}+0X18_{t-10}$ | 9.56 | 0.16 |
| $X9_t$ | $0+0.05X5_{t-10}+0.12X7_{t-10}+0.48X8_{t-10}+0.41X10_{t-10}-0.04X12_{t-10}+0.05X13_{t-10}$ | 9.03 | 0.17 |
| $X10_t$ | $0.15+0.13X2_{t-10}-0.09X3_{t-10}+0.15X4_{t-10}-0.07X5_{t-10}+0.1X7_{t-10}+0.24X8_{t-10}+0.09X9_{t-10}+0.17X11_{t-10}-0.05X12_{t-10}+0.3X13_{t-10}-0.08X14_{t-10}-0.08X15_{t-10}+0.05X16_{t-10}$ | 8.25 | 0.16 |
| $X11_t$ | $-0.02+0.15X1_{t-10}-0.08X2_{t-10}+0.06X3_{t-10}+0.19X10_{t-10}+0.81X12_{t-10}-0.08X13_{t-10}+0.08X17_{t-10}-0.12X18_{t-10}$ | 11.27 | 0.21 |
| X12 | $-0.03-0.14X1_{t-10}+0.04X2_{t-10}-0.07X10_{t-10}+0.95X11_{t-10}+0.18X13_{t-10}+0.09X19_{t-10}$ | 9.95 | 0.19 |
| $X13_t$ | $-0.13-0.09X2_{t-10}+0.1X5_{t-10}-0.11X8_{t-10}+0.43X10_{t-10}-0.1X11_{t-10}+0.24X12_{t-10}+0.25X14_{t-10}+0.63X15_{t-10}-0.11X16_{t-10}-0.11X18_{t-10}$ | 8.02 | 0.15 |
| $X14_t$ | $0.08+0.03X7_{t-10}+0.07X13_{t-10}+0.79X15_{t-10}-0.31X17_{t-10}+0.25X18_{t-10}+0.04X19_{t-10}+0.05X23_{t-10}$ | 8.57 | 0.16 |
| $X15_t$ | $-0.05-0.05X6_{t-10}-0.03X12_{t-10}+0.17X13_{t-10}+0.64X14_{t-10}+0.21X16_{t-10}+0.23X17_{t-10}-0.13X18_{t-10}$ | 9.26 | 0.19 |
| $X16_t$ | $-0.02-0.19X5_{t-10}+0.23X6_{t-10}+0.05X10_{t-10}+0.03X12_{t-10}-0.12X13_{t-10}+0.63X15_{t-10}+0.46X17_{t-10}-0.07X19_{t-10}$ | 8.55 | 0.13 |
| $X17_t$ | $0.05-0.04X8_{t-10}+0.02X11_{t-10}-0.33X14_{t-10}+0.34X15_{t-10}+0.17X16_{t-10}+0.74X18_{t-10}+0.03X19_{t-10}+0.03X20_{t-10}$ | 8.38 | 0.12 |
| $X18_t$ | $0.02+0.04X8_{t-10}-0.02X11_{t-10}-0.04X13_{t-10}+0.27X14_{t-10}-0.13X15_{t-10}+0.72X17_{t-10}+0.07X19_{t-10}+0.07X20_{t-10}$ | 9.69 | 0.13 |
| $X19_t$ | $0.12+0.15X12_{t-10}-0.12X13_{t-10}+0.12X14_{t-10}-0.18X16_{t-10}+0.15X17_{t-10}+0.35X18_{t-10}+0.21X20_{t-10}-0.07X22_{t-10}+0.12X23_{t-10}+0.17X25_{t-10}$ | 9.19 | 0.14 |
| $X20_t$ | $-0.01+0.04X10_{t-10}-0.04X15_{t-10}+0.08X17_{t-10}+0.19X18_{t-10}+0.09X19_{t-10}+0.25X21_{t-10}+0.29X22_{t-10}+0.11X24_{t-10}$ | 8.21 | 0.18 |
| $X21_t$ | $0.03-0.15X18_{t-10}+0.49X20_{t-10}+0.18X22_{t-10}+0.62X23_{t-10}-0.11X24_{t-10}-0.05X25_{t-10}$ | 8.86 | 0.16 |
| $X22_t$ | $-0.02-0.04X19_{t-10}+0.33X20_{t-10}+0.11X21_{t-10}+0.6X23_{t-10}+0.1X24_{t-10}-0.05X25_{t-10}$ | 8.92 | 0.11 |
| $X23_t$ | $0.03+0.04X20_{t-10}+0.15X21_{t-10}+0.24X22_{t-10}+0.39X24_{t-10}+0.07X25_{t-10}+0.07X28_{t-10}$ | 9.19 | 0.21 |
| $X24_t$ | $0.09-0.01X21_{t-10}+0.04X22_{t-10}+0.33X23_{t-10}+0.33X25_{t-10}+0.5X27_{t-10}+0.06X28_{t-10}$ | 6.95 | 0.18 |

TABLE 21: Fitted model for forecasting travel time / minimum travel time

| Target Segment | Lagged Regression Model with Predictor Variables | Holdout Performance | |
|---|---|---|---|
| | | MAPE (%) | MAD |
| $X6_t$ | $0.07-0.04y2+0.05X3_{t-10}+0.44X5_{t-10}+0.28X7_{t-10}+0.15X8_{t-10}+0.03X11_{t-10}-0.04X12_{t-10}+0.03X13_{t-10}+0.05X14_{t-10}-0.23X15_{t-10}+0.24X16_{t-10}$ | 9.76 | 0.23 |
| $X7_t$ | $0.14-0.05X4_{t-10}+0.27X5_{t-10}+0.23X6_{t-10}+0.35X8_{t-10}+0.01X9_{t-10}+0.1X10_{t-10}-0.04X13_{t-10}+0.03X14_{t-10}+0.05X15_{t-10}-0.05X16_{t-10}$ | 9.61 | 0.22 |
| $X8_t$ | $-0.19+0.11X1_{t-10}-0.11X4_{t-10}+0.16X5_{t-10}+0.14X6_{t-10}+0.4X7_{t-10}+0.13X9_{t-10}+0.31X10_{t-10}-0.09X13_{t-10}+0.09X14_{t-10}$ | 9.71 | 0.19 |
| $X9_t$ | $0.12-0.1X2_{t-10}+0.08X3_{t-10}+0.08X7_{t-10}+0.57X8_{t-10}+0.36X10_{t-10}+0.03X13_{t-10}$ | 9.85 | 0.19 |
| $X10_t$ | $0.22+0.14X2_{t-10}-0.1X3_{t-10}+0.14X4_{t-10}-0.06X5_{t-10}+0.09X7_{t-10}+0.26X8_{t-10}+0.07X9_{t-10}+0.2X11_{t-10}-0.08X12_{t-10}+0.26X13_{t-10}-0.14X14_{t-10}$ | 8.89 | 0.16 |
| $X11_t$ | $0.08+0.07X4_{t-10}-0.06X5_{t-10}+0.04X6_{t-10}+0.28X10_{t-10}+0.75X12_{t-10}-0.08X13_{t-10}+0.05X15_{t-10}-0.04X16_{t-10}+0.12X17_{t-10}-0.12X18_{t-10}-0.11X19_{t-10}$ | 13.76 | 0.39 |
| X12 | $-0.1+0.04X5_{t-10}-0.06X6_{t-10}-0.16X10_{t-10}+1.05X11_{t-10}+0.19X13_{t-10}-0.07X15_{t-10}+0.08X16_{t-10}-0.08X17_{t-10}+0.25X19_{t-10}-0.06X20_{t-10}$ | 8.97 | 0.18 |
| $X13_t$ | $-0.3+0.09X6_{t-10}-0.04X7_{t-10}-0.13X8_{t-10}+0.03X9_{t-10}+0.47X10_{t-10}-0.12X11_{t-10}+0.21X12_{t-10}+0.51X14_{t-10}+0.47X15_{t-10}-0.07X16_{t-10}+0.11X17_{t-10}-0.17X18_{t-10}-0.1X19_{t-10}$ | 10.02 | 0.21 |
| $X14_t$ | $0.23+0.06X8_{t-10}-0.11X10_{t-10}-0.02X12_{t-10}+0.21X13_{t-10}+0.67X15_{t-10}-0.04X16_{t-10}-0.27X17_{t-10}+0.25X18_{t-10}+0.06X19_{t-10}$ | 9.18 | 0.16 |
| $X15_t$ | $-0.06-0.02X7_{t-10}+0.01X11_{t-10}+0.15X13_{t-10}+0.55X14_{t-10}+0.26X16_{t-10}+0.2X17_{t-10}-0.04X19_{t-10}-0.03X20_{t-10}+0.02X21_{t-10}-0.04X23_{t-10}$ | 9.83 | 0.19 |
| $X16_t$ | $0.05+0.02X7_{t-10}-0.03X11_{t-10}+0.05X12_{t-10}-0.03X13_{t-10}-0.07X14_{t-10}+0.61X15_{t-10}+0.33X17_{t-10}+0.13X18_{t-10}-0.07X19_{t-10}+0.03X20_{t-10}$ | 8.83 | 0.13 |
| $X17_t$ | $0.04+0.03X11_{t-10}-0.03X12_{t-10}+0.02X13_{t-10}-0.22X14_{t-10}+0.22X15_{t-10}+0.14X16_{t-10}+0.77X18_{t-10}+0.04X19_{t-10}+0.01X20_{t-10}$ | 9.78 | 0.12 |
| $X18_t$ | $0.07-0.02X11_{t-10}-0.04X13_{t-10}+0.14X14_{t-10}+0.04X16_{t-10}+0.63X17_{t-10}+0.09X19_{t-10}+0.1X20_{t-10}-0.02X21_{t-10}$ | 9.98 | 0.13 |
| $X19_t$ | $0.04-0.23X11_{t-10}+0.31X12_{t-10}-0.11X13_{t-10}+0.13X14_{t-10}-0.09X15_{t-10}-0.11X16_{t-10}+0.14X17_{t-10}+0.37X18_{t-10}+0.18X20_{t-10}-0.06X22_{t-10}+0.1X23_{t-10}+0.12X25_{t-10}+0.2X27_{t-10}$ | 8.67 | 0.14 |
| $X20_t$ | $0.01+0.03X13_{t-10}-0.05X14_{t-10}-0.04X15_{t-10}+0.04X16_{t-10}+0.26X18_{t-10}+0.08X19_{t-10}+0.22X21_{t-10}+0.29X22_{t-10}+0.09X23_{t-10}+0.08X24_{t-10}$ | 9.75 | 0.18 |
| $X21_t$ | $-0.01-0.06X7_{t-10}+0.07X8_{t-10}-0.07X10_{t-10}+0.1X15_{t-10}-0.08X16_{t-10}-0.14X18_{t-10}+0.48X20_{t-10}+0.2X22_{t-10}+0.69X23_{t-10}-0.12X24_{t-10}$ | 9.96 | 0.16 |
| $X22_t$ | $0+0X7_{t-10}+0X8_{t-10}+0X19_{t-10}+0X20_{t-10}+0X21_{t-10}+0X23_{t-10}+0X24_{t-10}+0y26$ | 9.12 | 0.11 |
| $X23_t$ | $0.02+0.03X19_{t-10}+0.04X20_{t-10}+0.15X21_{t-10}+0.22X22_{t-10}+0.43X24_{t-10}+0.07y26+0.04X28_{t-10}$ | 9.35 | 0.14 |
| $X24_t$ | $0.1-0.01X13_{t-10}+0.02X14_{t-10}-0.01X19_{t-10}+0.02X20_{t-10}-0.01X21_{t-10}+0.04X22_{t-10}+0.33X23_{t-10}+0.24X25_{t-10}+0.14y26+0.12X27_{t-10}+0.03X28_{t-10}$ | 6.75 | 0.11 |

## 6.6 Model Validation for Travel Time and Variations in Travel Time Forecasting

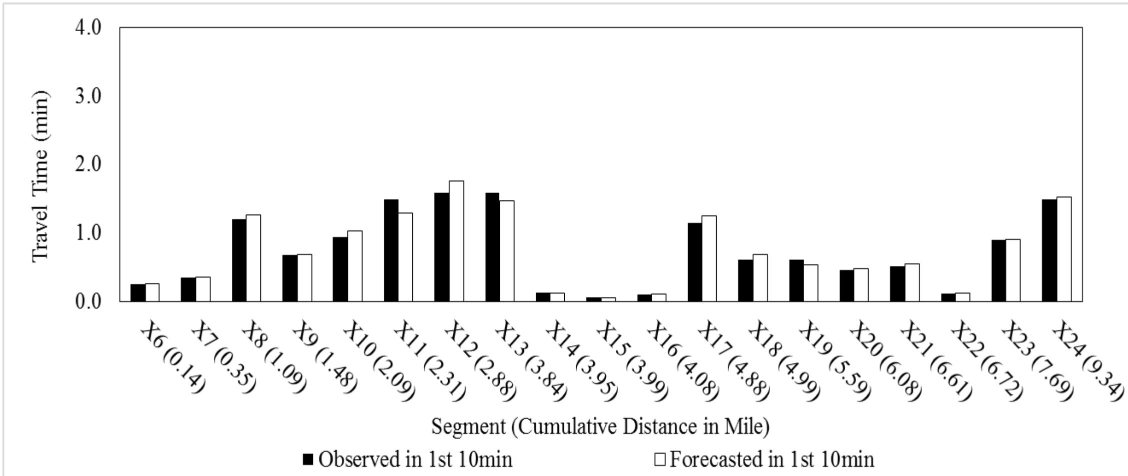
For model validation, a total of 80 days of data was considered from 2010 to 2012. Among these, 45 days are “vehicle accident” affected days with different severity types, different number of lanes blocked and crashes during different time-periods. The remaining 35 days had no reported incidents along the study corridor. MAPE and MAD were computed for all the segments using the fitted lagged regression model. The results from model validation of travel time, TT/ExpTT and TT/MinTT are shown in Table 22. The model validation results using the fitted model based on TT/ExpTT led to MAPE values less than 15% for all segments. For fitted models for travel time and TT/ MinTT, the computed MAPE is less than 15% for almost all the segments. X19 in travel time and X22 in TT/MinTT have MAPE values higher than 15%. However, the model validation results using the fitted model based on TT/MinTT showed comparatively higher MAPE values than TT/ExpTT for most of the segments. Therefore, forecasting RVTT might be more accurate considering ExpTT than MinTT.

## 6.7 Sample Spatio-temporal Variation of Travel Time without Incident Condition

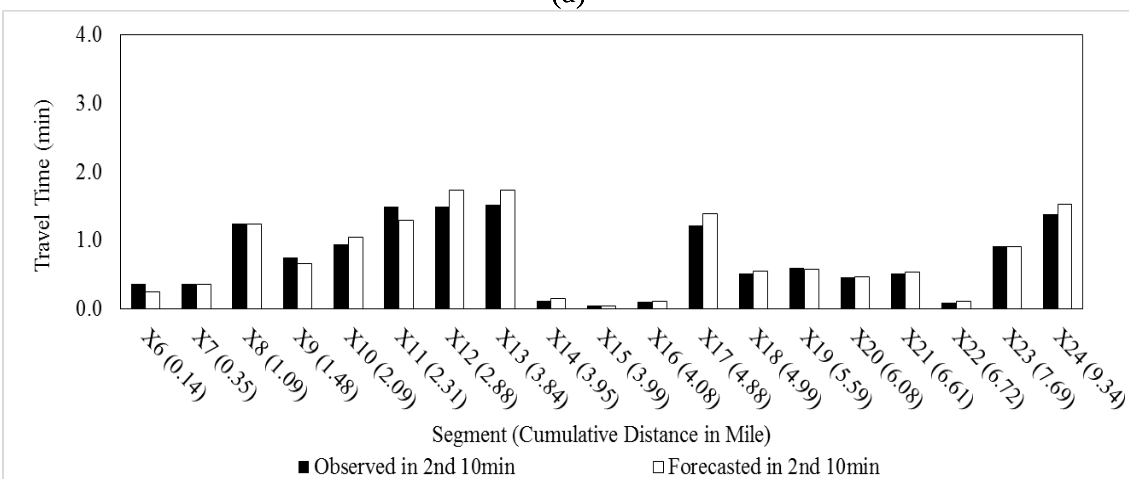
A graphical representation of spatio-temporal variation of travel time, when there is no incident, using the fitted lagged regression model of each segment is shown in Figure 16. The study period is selected from 2012-06-1817:00:00 to 2012-06-18 17:30:00. During this period, there was no incident along I-77 S. In Figure 16, the observed and forecasted travel times are very close to each other for all three 10 minute intervals, which demonstrates that the model successfully replicates the real-world scenario without incident condition.

TABLE 22: Model validation of travel time, travel time / expected travel time and travel time/ minimum travel time

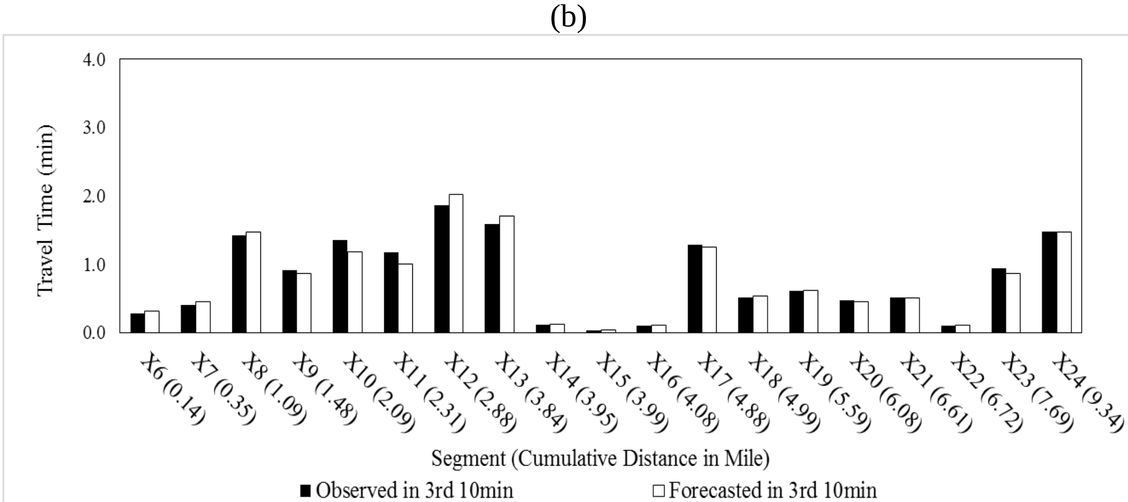
| Target Segment | Model Validation | | | | | |
|---|---|---|---|---|---|---|
| | Travel Time | | Travel Time/ Expected Travel Time | | Travel Time/ Minimum Travel Time | |
| | MAPE (%) | MAD | MAPE (%) | MAD | MAPE (%) | MAD |
| $X6_t$ | 11.84 | 0.31 | 10.42 | 0.18 | 13.03 | 0.25 |
| $X7_t$ | 10.63 | 0.19 | 10.35 | 0.18 | 12.61 | 0.22 |
| $X8_t$ | 12.66 | 0.15 | 10.66 | 0.19 | 12.77 | 0.23 |
| $x9_t$ | 13.51 | 0.27 | 10.09 | 0.20 | 15.77 | 0.40 |
| $X10_t$ | 13.57 | 0.30 | 11.25 | 0.28 | 11.79 | 0.21 |
| $X11_t$ | 9.89 | 0.21 | 13.33 | 0.21 | 13.76 | 0.39 |
| $X12_t$ | 9.86 | 0.18 | 10.78 | 0.19 | 11.57 | 0.25 |
| $X13_t$ | 9.32 | 0.15 | 12.22 | 0.18 | 13.02 | 0.31 |
| $X14_t$ | 9.15 | 0.18 | 10.67 | 0.21 | 11.75 | 0.22 |
| $X15_t$ | 10.11 | 0.18 | 10.16 | 0.26 | 11.21 | 0.21 |
| $X16_t$ | 12.31 | 0.22 | 9.85 | 0.19 | 11.54 | 0.22 |
| $X17_t$ | 10.36 | 0.25 | 10.01 | 0.14 | 9.96 | 0.17 |
| $X18_t$ | 13.21 | 0.41 | 11.69 | 0.20 | 9.98 | 0.13 |
| $X19_t$ | 18.32 | 0.56 | 9.33 | 0.17 | 10.13 | 0.17 |
| $X20_t$ | 9.82 | 0.18 | 9.29 | 0.18 | 13.50 | 0.22 |
| $X21_t$ | 9.41 | 0.15 | 12.67 | 0.19 | 10.43 | 0.19 |
| $X22_t$ | 10.61 | 0.16 | 11.21 | 0.18 | 15.42 | 0.33 |
| $X23_t$ | 10.35 | 0.18 | 10.41 | 0.27 | 10.32 | 0.18 |
| $X24_t$ | 7.03 | 0.14 | 7.32 | 0.19 | 8.81 | 0.12 |




(a)




(b)



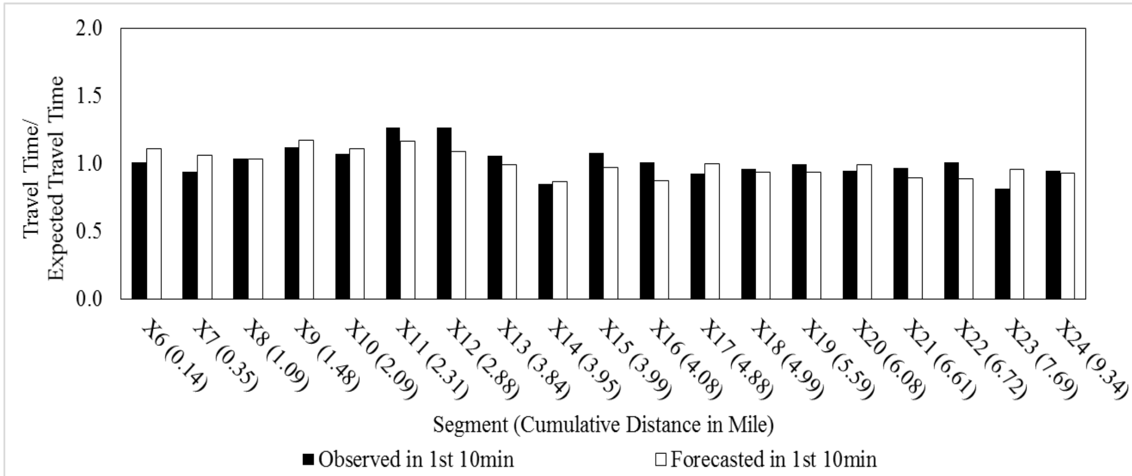

(c)

FIGURE 16: Sample spatio-temporal variation of travel time without incident condition for (a) 1st 10 minute, (b) 2nd 10 minute, and (c) 3rd 10 minute intervals

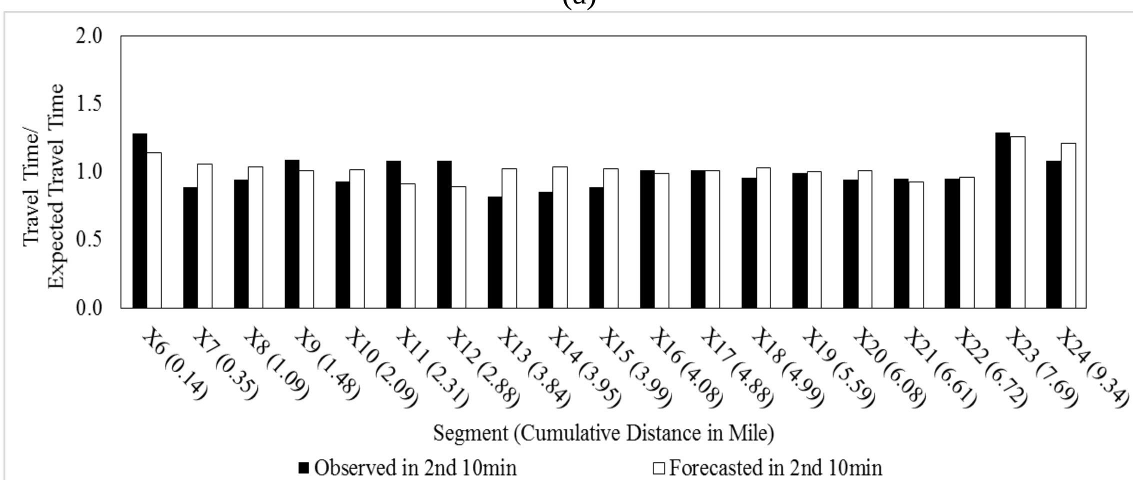
### 6.8 Sample Spatio-temporal Variation of Travel Time/ Expected Travel Time and Travel Time/ Minimum Travel Time without Incident Condition

A graphical representation of RVTT using the fitted lagged regression model of each segment is shown in figures 17 and 18. The time period is same the one mentioned in the previous section. The results for the three time intervals based on TT/ExpTT are shown in Figure 17. The figure shows that the model successfully forecasts the TT/ExpTT. In most of the cases, the ratio is close to “1”. This implies that the expected travel time is marginally higher or lower than the observed travel time of the segment during the corresponding time interval.

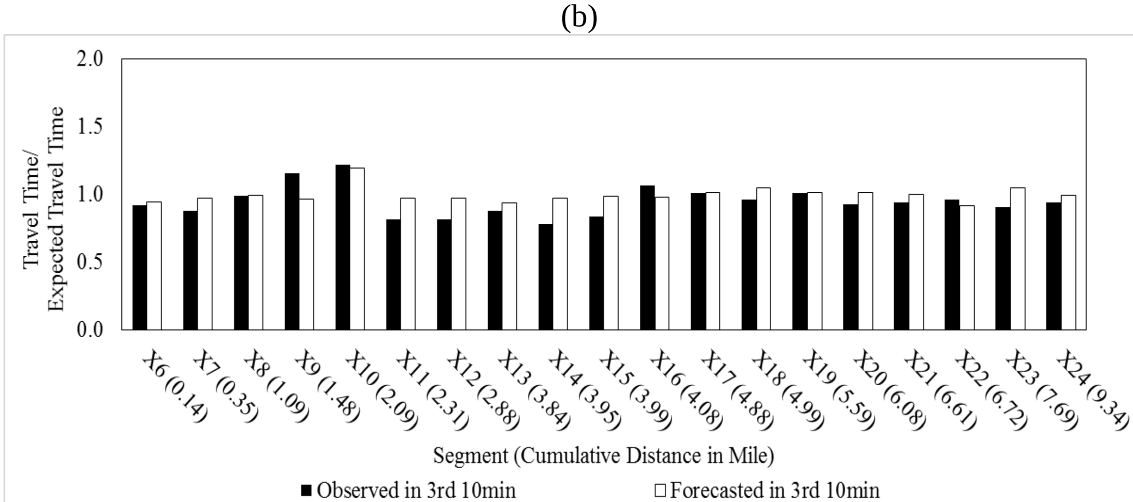
Figure 18 shows RVTT based on TT/MinTT. The forecasted TT/MinTT model is close to the observed TT/MinTT. However, from X6 to x16, this ratio is mostly greater than “1”. This is because the minimum observed travel time is supposed to be less than the observed travel time of the segment during the corresponding time interval.




(a)




(b)




(c)

FIGURE 17: Sample spatio-temporal variation of travel time/ expected travel time for without incident condition for (a) 1st 10 minute, (b) 2nd 10 minute, and (c) 3rd 10 minute intervals

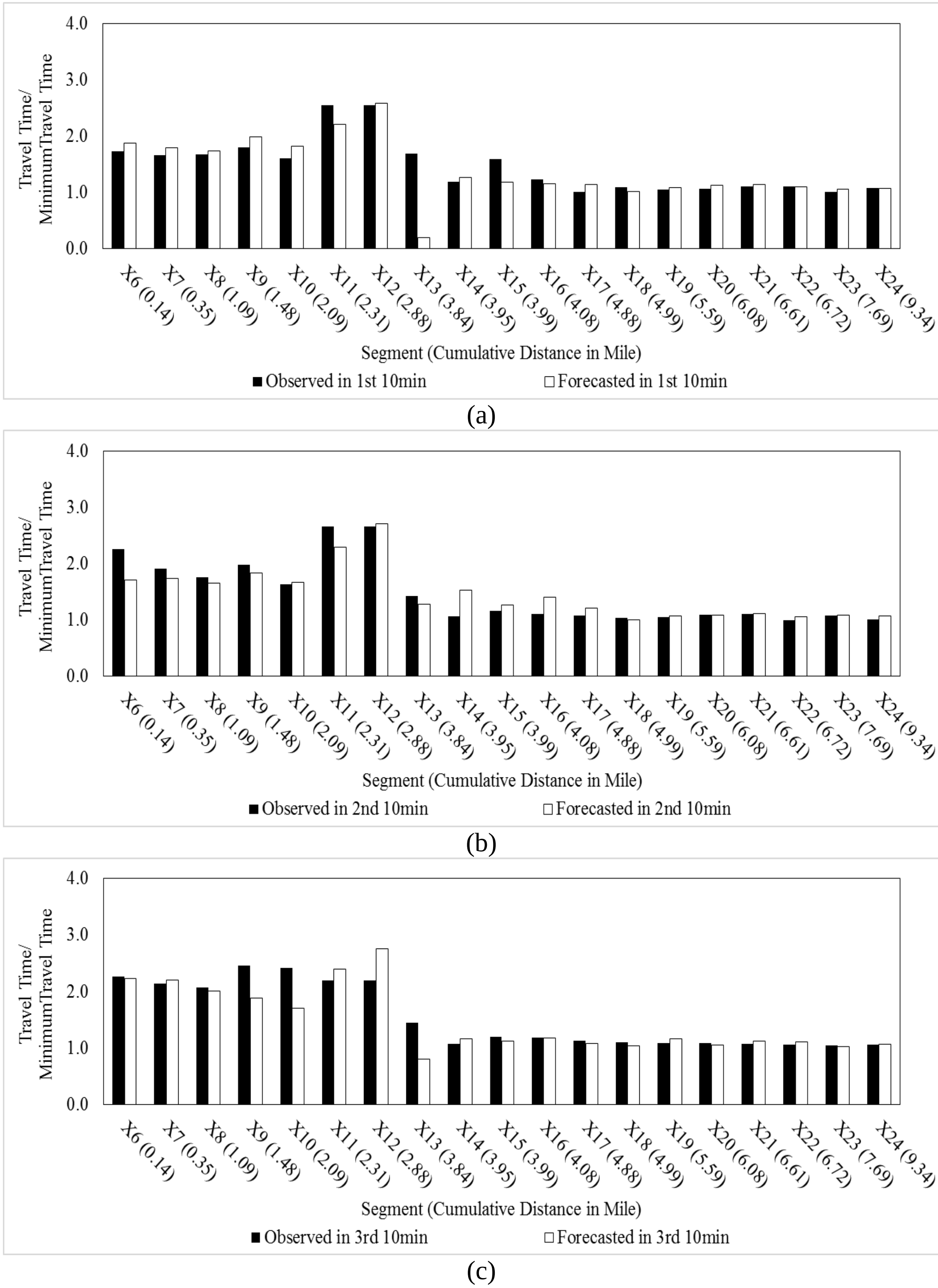



FIGURE 18: Sample spatio-temporal variation of travel time/ minimum travel time for without incident condition for (a) 1st 10 minute, (b) 2nd 10 minute, and (c) 3rd 10 minute intervals

## 6.9 Sample Spatio-temporal Variation of Travel Time with "Vehicle Accident" Condition

This part focuses on the comparison of forecasted and observed travel time due to a "vehicle accident" occurred at 2010-06-22 15:00:00 and the crash injury type were "B". The number of lanes blocked was only one. Figure 19 shows that the forecasted and observed travel time is very close to each other for almost all the segments. The queue kept building on and propagating in the upstream direction up to 30 minutes after the incident occurrence (Figure 19).

## 6.10 Sample Spatio-temporal Variation of Travel Time/ Expected Travel Time and Travel Time/ Minimum Travel Time with "Vehicle Accident" Condition

This part focuses on the implementation of the lagged regression model in capturing RVTT due to a "vehicle accident". The selected "vehicle accident" of crash injury type "B" occurred at 2010-06-22 15:00:00 and the. The number of lanes blocked was only one. Figure 20 shows the RVTT based on TT/ExpTT. The model successfully forecasted the RVTT for most of the segments as the observed and forecasted TT/ExpTT are very close to each other. The ratios also reveal that the effect of the incident starts at X8, 10 minutes after the incident occurrence. The effect starts building and propagating in the upstream direction. An observation of performance on surrounding segments shows that the "vehicle accident" had an effect at least up to 5:20:00 PM.

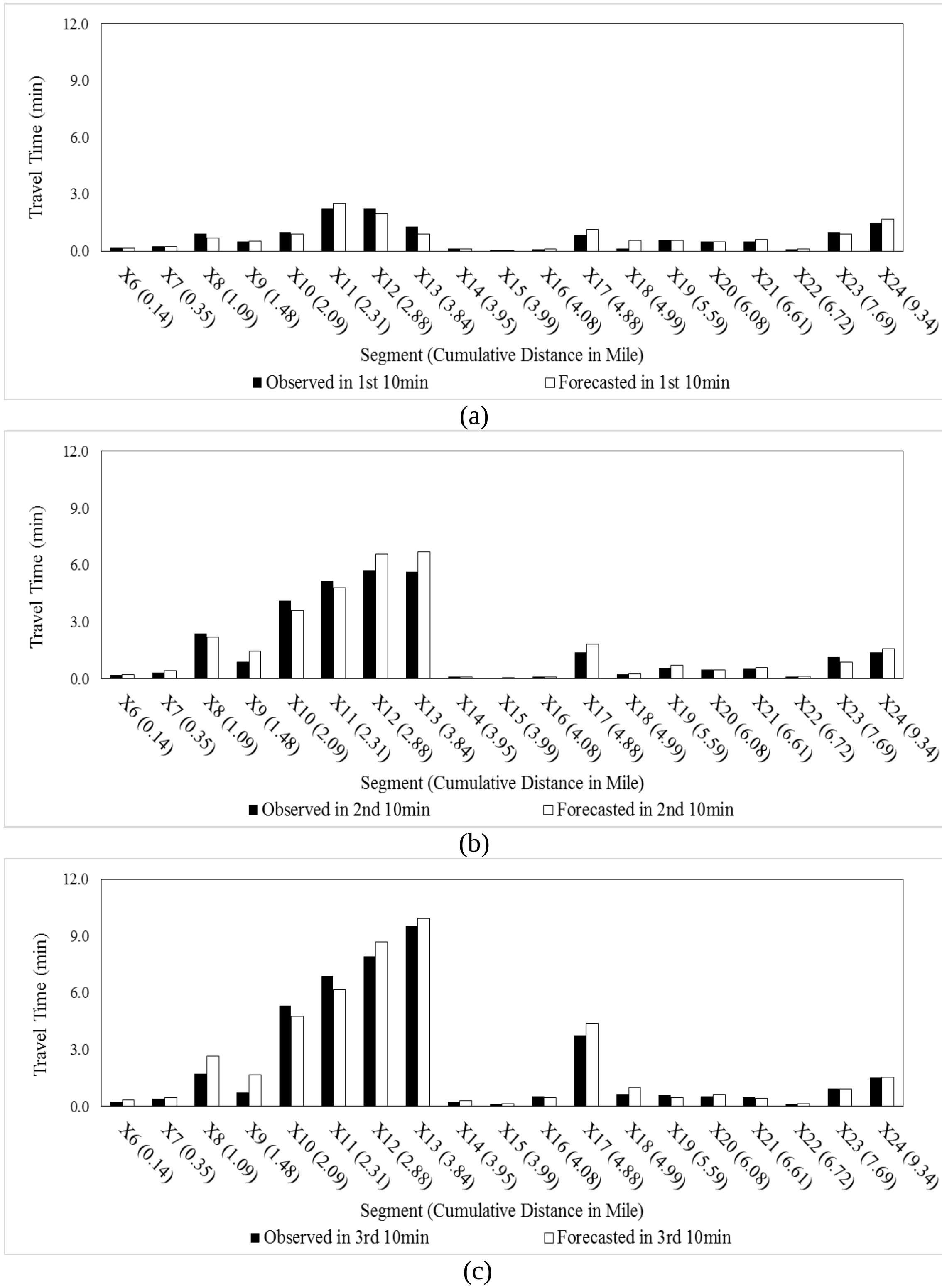
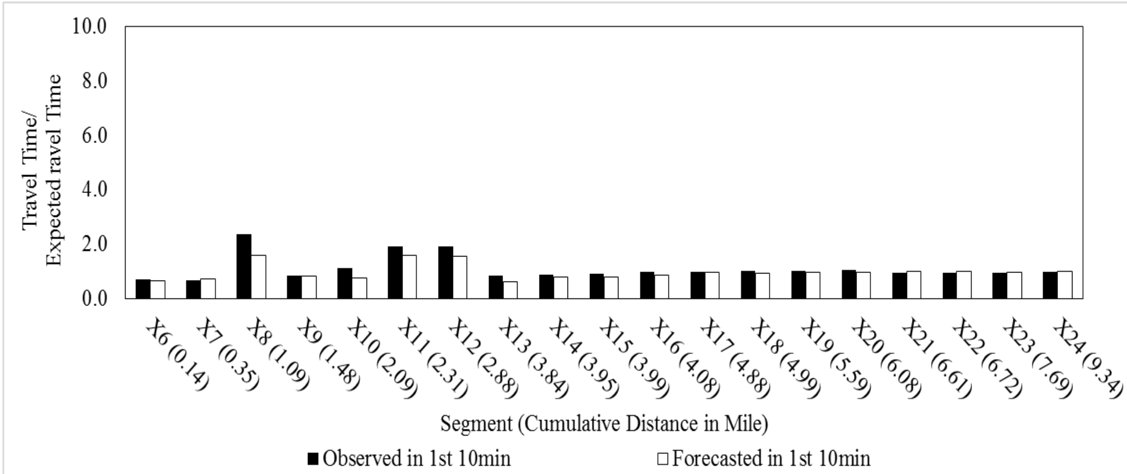
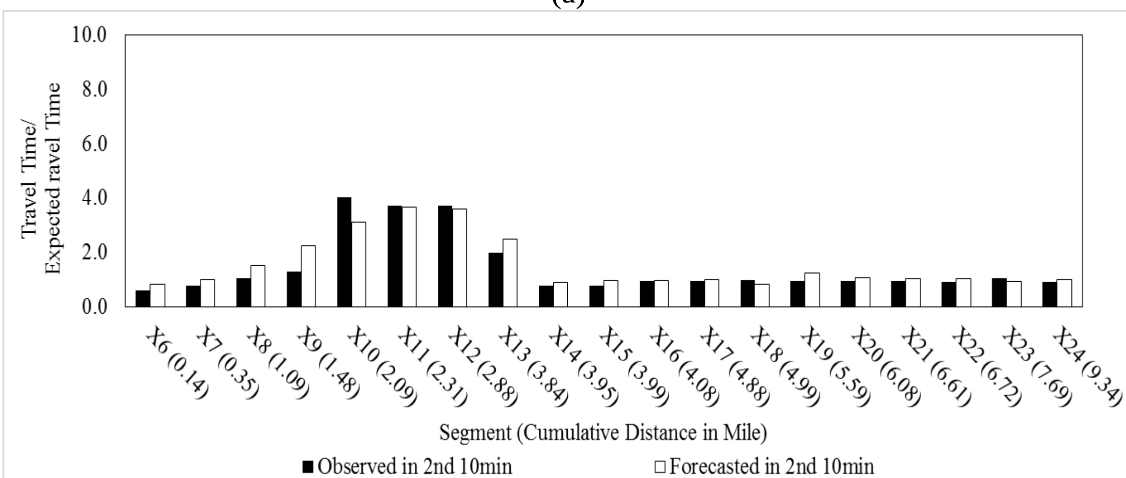


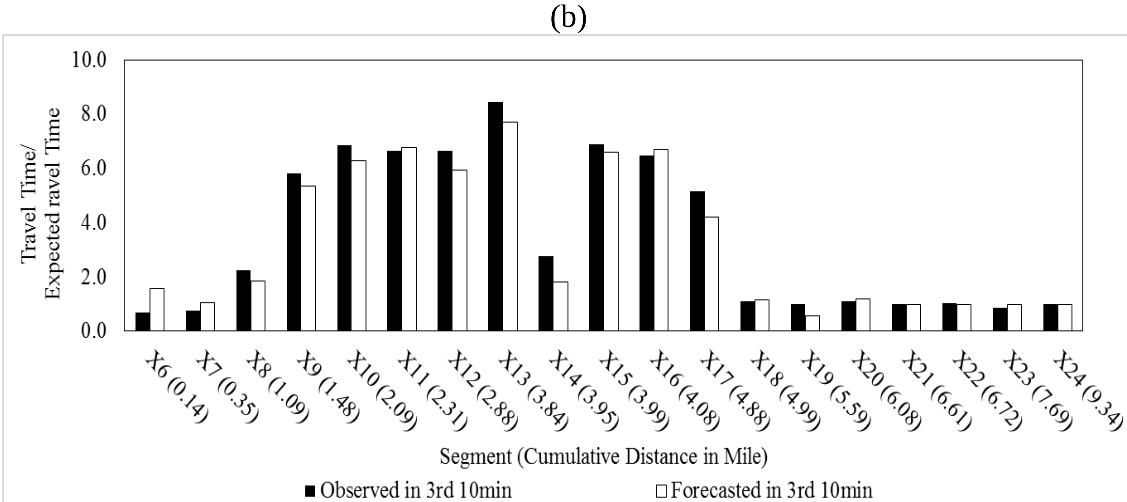

FIGURE 19: Sample spatio-temporal variation of travel time with “vehicle accident” condition for (a) 1st 10 minute, (b) 2nd 10 minute, and (c) 3rd 10 minute intervals




(a)




(b)



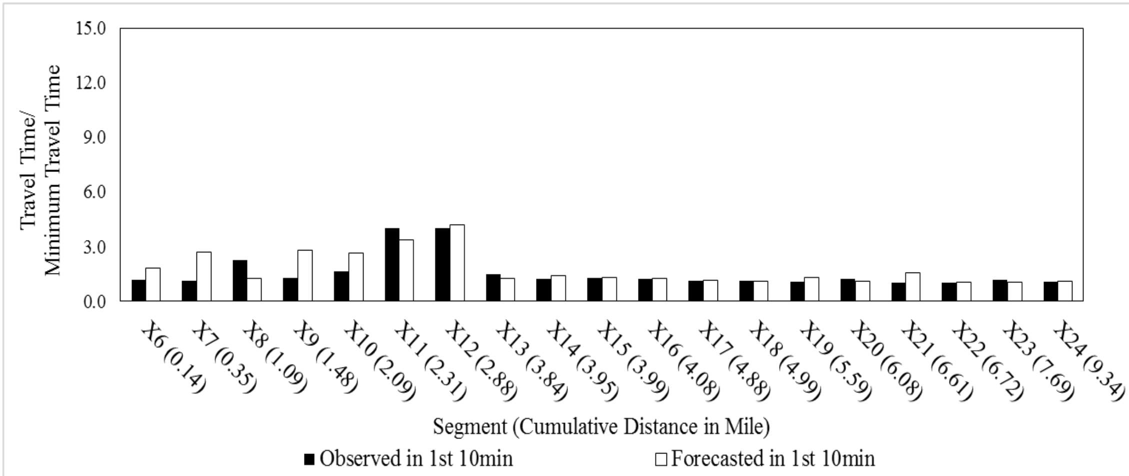

(c)

FIGURE 20: Sample spatio-temporal variation of travel time/expected travel time in "vehicle accident" condition for (a) 1st 10 minute, (b) 2nd 10 minute, and (c) 3rd 10 minute intervals

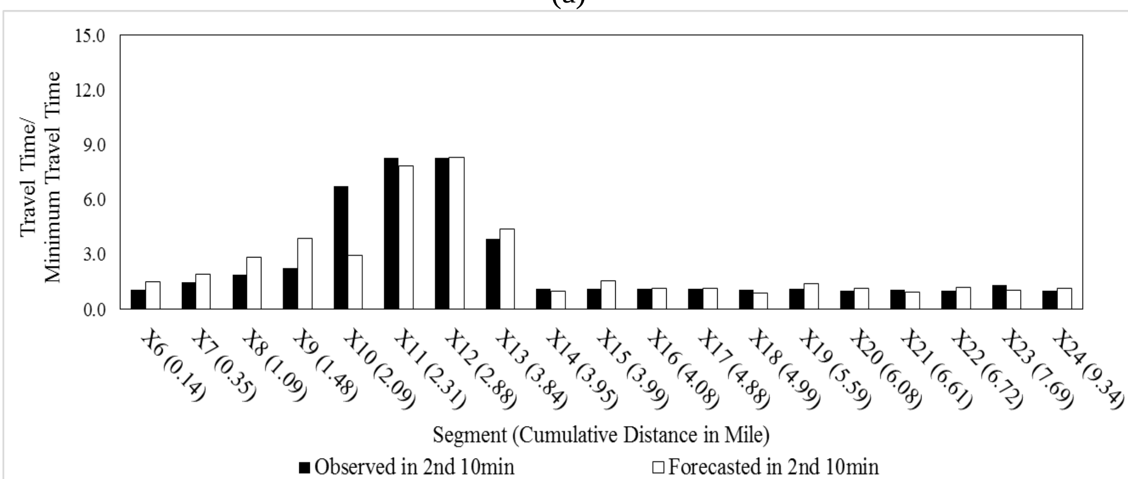
Figure 21 depicts the RVTT due to the same "vehicle accident" mentioned previously based on TT/MinTT. It shows that the model successfully forecasted the RVTT. The estimates are close to the observed values. Comparing figures 20 and 21, it can be said that the propagation nature of the effect of incident over time and space is same for both TT/ExpTT and TT/MinTT. Both figures 20 and 21 show a fine distinction between travel time and TT/ExpTT or TT/MinTT. As the length of X14, X15, and X16 is low, the travel time is very low for all three 10-min time intervals (Figure 19). However, from figure 20 and figure 21, it seems that the "vehicle accident" also had an effect on other segments and reaches a ratio equal to 4.0. This implies that the travel time observed is 3 times more than the expected travel time or minimum travel time.

6.11 Summary

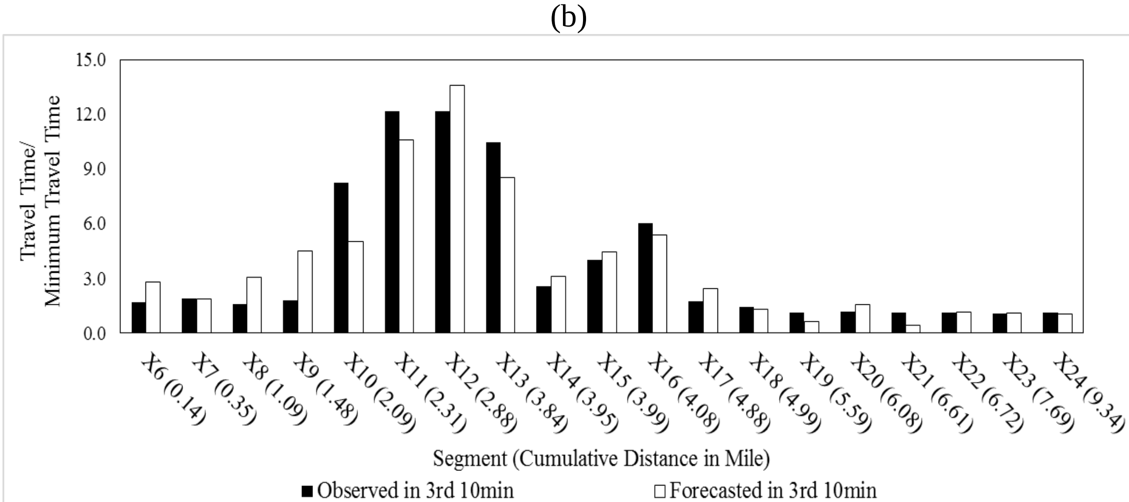
Overall, the results obtained show that the ARIMA model alongside pre-whitening CCF could be an effective method to forecast both travel time and relative variations in travel time. Travel time forecasting cannot exactly mimic the sole effect of a "vehicle accident", which can be overcome by estimating or modeling Travel Time / Expected Travel Time and Travel Time / Minimum Travel Time. Both the fitted model and validated model shows that MAPE and MAD value less than 15% for 99% of estimations.




(a)




(b)




(c)

FIGURE 21: sample spatio-temporal variation of travel time/expected travel time in "vehicle accident" condition for (a) 1st 10 minute, (b) 2nd 10 minute, and (c) 3rd 10 minute intervals

# CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS

Travel time and variations in travel time was forecasted in this study. In the first step, Cronbach's α was used to obtain the expected travel time. ARIMA model with pre-whitening CCF was adopted for forecasting as the second step. A summary of findings from this study and potential scope for future research is discussed next.

## 7.1 Summary of Findings

When a non-recurring incident such as "vehicle accident" occurs, traffic condition changes over time and space. Travel time of the segment or corridor is affected due to the incident. Accurately forecasting conditions without incident is also equally important to assess the effect of incident. To solve this problem, the "expected travel time" or most probable travel time based on the similarity in travel-time series pattern needs to be analyzed. For this purpose, Cronbach's α by time-of-the-day and week-of-the-year was computed. Findings from Cronbach's α computation are summarized next.

1. Four different categories of Cronbach's α: α1, α2, α3, and α4 were estimated based on two categories and two factors: day-of the-week, weekday/ weekend, time-of-the-day, and week-of the-year. The average travel was used as the travel time measure.
2. Corresponding Cronbach's α of a TMC code was estimated separately for each day-of-the-week (i.e., Monday, Tuesday etc.) and each year for 2010 to 2012.

3. Level of reliability shows 63.6% (374 out of 588 samples) of the segments are in “Excellently Reliable” condition. Moreover, reliability does not go below “C” or “Reliable” condition.
4. Results show that the expected travel time is very close to the observed travel time. In 2010, more than 85% of samples show that the percentage difference between observed and expected travel is equal or less than 10% for 23 segments out of 28 segments. Results are observed to be fairly consistent and similar during 2011 and 2012.
5. From 2010 to 2012, more than 75% of the segments have shown less than or equal to 10% difference between observed and expected travel time.
6. Thirty percent (30%) difference is observed no more than for 5% of the segments.

ARIMA model alongside pre-whitened CCF was applied to model travel time and relative variations in travel time due to a “vehicle accident” as well as when there is no incident. Results showed that the adopted methodology could be successfully implemented to forecast travel time as well as relative variations in travel time, irrespective of the incident condition. Major findings are mentioned below.

1. Two types of relative variations in travel time were considered: travel time/expected travel time and travel time/minimum travel time.
2. Fitted models were developed not only based on the past travel time of target segment but also with respect to the travel time of different lags of the surrounding segments.

3. All the fitted models consisted of the predictor variables from both upstream and downstream segments; all the significant variables were in time lag of 10 minutes.
4. MAPE and MAD values of the developed lagged regression model are estimated for every segment, irrespective of the incident condition. MAPE and MAD values for all segments based on travel time model are less than 10% except for one segment, which has a MAPE value greater than 10% but less than 15%.
5. In TT/ExpTT, MAPE and MAD value for all segments was less than 10%, except for one segment for which the value was marginally greater than 10%.
6. Alike TT and TT/ExpTT scenario, MAPE and MAD value of TT/MinTT for all segments was less than 10%, except for one segment for which the value was between 10-15%.
7. Validation, based on a total of 80 days of data (45 incident free days and 35 "vehicle accident" affected days), showed that MAPE and MAD values are less than 15% for all segments in travel time model scenario except for one segment for which MAPE value was around 18%.
8. For TT/ExpTT and TT/MinTT scenario, all the segments had MAPE value less than 15%.
9. Both the calibrated and validated models demonstrate that forecasting accuracy is relatively higher for TT/ExpTT than TT/MinTT.

## 7.2 Recommendations for Future Research

ARIMA models were developed for only 'vehicle accident' type incidents.

However, there are several other incident types. Models similar to those developed in this research could be generated to estimate their effect on travel time.

The effect of one vehicle accident was modeled in this research. Congestion and driver behavior following a vehicle accident could lead to secondary vehicle accidents. These secondary vehicle accidents further aggravate congestion and travel time variation. Modeling the effect of secondary vehicle accidents warrants research.

In this study, 85th percentile travel time was disregarded to forecast the relative variation in travel time. This as well as other travel time percentiles could be considered for investigation and for better understanding of the effect of an incident. Time aggregation of this study is 10 minutes. The applicability of 5 minutes and 15 minutes for time aggregation could be explored and compared to aggregation at 10-minute intervals. These merit an investigation.

# REFERENCES


Ahmed, M. S., and A. R. Cook. Analysis of Freeway Traffic Time-series Data by Using Box Jenkins Techniques. *In Transportation Research Record: Journal of the Transportation Research Board, No. 722*, Transportation Research Board of the National Academies, Washington, D.C., Vol.2., 1979, pp. 62–72.

Al-Deek, H. M., and E. B. Emam. Computing Travel Time Reliability in Transportation Networks with Multistates and Dependent Link Failures. In *Journal of Computing in Civil Engineering*, Vol. 20, No.5, 2006, pp. 317-327.

Al-Deek, H., and A. Kanafani. *Incident Management with Advanced Traveler Information System*. Proceedings of the 2nd Vehicle Navigation and Information System Conference, Dearborn, Michigan, 1991.

Angeli, C. *Diagnostic Expert Systems: From Expert's Knowledge to Real-Time Systems*. Advanced Knowledge Based Systems: Model, Applications & Research, Vol. 1, 2010, pp. 50 - 73.

Asif, M. T., J. Dauwels, C. Y. Goh, A. Oran, E. Fathi, M. Xu, M. M. Dhanya, N. Mitrovic, and P. Jaillet. Spatiotemporal Patterns in Large-Scale Traffic Speed Prediction. In *IEEE Transactions on Intelligent Transportation Systems,* Vol.15, 2014, pp.794-804.

Billings, D., and J. S. Yang. *Application of the ARIMA Models to Urban Roadway Travel Time Prediction - A Case Study*. Presented at IEEE International Conference on Systems, Man, Cybernatics, Taipei, Taiwan, Vol.3, 2006, pp. 2529–2534.

Bishop, C. M. Pattern Recognition and Machine Learning. Springer, 2006.

Booz Allen & Hamilton Inc. *Intelligent Transportation Systems Field Operational Test Cross Cutting Study Advance Traveler Information Systems*. Publication FHWA-JPO-99-038,Federal Highway Administration (FHWA), United States Department of Transportation, 1998.

Cambridge Systematics, Inc., Dowling Associates, Inc., System Metrics Group, Inc., and Texas Transportation Institute (TTI). *NCHRP Report 618: Cost-Effective Performance Measures for Travel Time Delay, Variation, and Reliability*. Transportation Research Board of the National Academies, Washington D.C., 2008.

Cambridge Systematics. *Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation*. 2005, http://www.ops.fhwa.dot.gov/congestion_report/executive_summary.htm. Accessed July 15, 2015.

Castro-Neto M., Y. S. Jeong, M. K. Jeong, and H. Lee. Online-SVR for Short-Term Traffic Flow Prediction under Typical and Atypical Traffic Conditions. In *Expert Systems with Applications: An International Journal*, Vol.36, 2009, pp.6164-6173.

Centre for Economics and Business Research (CEBR). *The Future Economic and Environmental Costs of Gridlock in 2030*. London, 2014. http://inrix.com/wp-content/uploads/2015/07/Whitepaper_Cebr-Cost-of-Congestion.pdf. Accessed August 29, 2015.

Chandra, S. R, and H. Al-Deek. Predictions of Freeway Traffic Speeds and Volumes Using Vector Autoregressive Models. In *Journal of Intelligent Transportation Systems*, Vol. 13, No. 2, 2009, pp.53 -72.

Chen C., J. Hu, Q. Meng, and Y. Zhang. *Short-Time Traffic Flow Prediction with ARIMA-GARCH Model*. Presented at Intelligent Vehicles Symposium (IV), Baden-Baden, 2011, pp.607-612.

Chen H., H. A. Rakha, and C. C. McGhee. *Dynamic Travel Time Prediction using Pattern Recognition*. Presented at 20th ITS World Congress, 2013.

Chen, H., and H. Rakha. *Data-Driven Particle Filter for Multi-step Look-Ahead Travel Time Prediction*. Presented at 93rd Annual Meeting of Transportation Research, Washington D.C., 2014.

Chen, X. M., H. B. Chen, H. B. Gong, and J. N. Wang. BRT Vehicle Travel Time Prediction Based On SVM and Kalman Filter. In *Journal of Transportation Systems Engineering and Information Technology,* Vol. 12, 2012, pp. 29–34.

Chien, S. I. J., and C. M. Kuchipudi. Dynamic Travel Time Prediction with Real-Time and Historic Data. In *Journal of Transportation Engineering*, Vol. 129, No. 6, 2003, 608-616.

Cho, H. and L. R. Rilett. Forecasting Train Arrival Times at At-Grade Crossings. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1844*, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 94-102.

Corthout, R., C. M. J. Tampère, and L. H. Immers. Marginal Incident Computation: An Efficient Algorithm to Determine Congestion Spillback due to Incidents. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2099*, Transportation Research Board of the National Academies, Washington, D. C., 2009, pp. 22-29.

Cortina, J. M. (1993). What is Coefficient Alpha? An Examination of Theory and Applications. Journal of Applied psychology, Vol. 78, pp. 98-104.

Dauwels, J., A. Aslam, M. T. Asif, X. Zhao, N. M. Vie, A. Cichocki, and P. Jaillet. *Predicting Traffic Speed in Urban Transportation Subnetworks for Multiple Horizons*. Presented at13th International Conference on Control Automation Robotics & Vision (ICARCV), Singapore, 2014, pp.547-552.

Deb Nat, R. P., N. K. Chowdhury, and M. Aono. A Naïve Clustering Approach in Travel Time Prediction. In *International Journal of Data Engineering*, Vol. 2, 2011, pp. 62-74.

Easton, V. J., and J. H. McColl. Statistics Glossary v1.1.: The STEPS Project. 1997. http://www.stats.gla.ac.uk/steps/glossary/time_series.html. Accessed by September 22, 2015.

Florida Department of Transportation (FDOT). *Florida's Mobility Performance Measures Program Summary Report*. Office of State Transportation Planner, Tallahassee, FL, August 2000.

Fu, L., and L. R. Rilett. *Estimation of Expected Minimum Paths in Dynamic and Stochastic Traffic Networks*. Proceedings of the 6$^{th}$ International Conference on Vehicular Navigation and Information Systems (VNIS), Seattle, Washington, August 1995.

Gao S., Z. Zhang, and C. Cao. Road Traffic Freight Volume Forecast Using Support Vector Machine Combining Forecasting. In *Journal of Software*, Vol. 6, 2011, pp. 1680-1687.

Ghosh B., B. Basu, and M. O'Mahony. *Time-Series Modeling For Forecasting Vehicular Traffic Flow in Dublin*. Presented at the 84th Annual Meeting of Transportation Research Board, Washington D.C., 2005.

Guin, A. *Travel Time Prediction Using a Seasonal Autoregressive Integrated Moving Average Time Series Model*. Presented at Intelligent Transportation Systems Conference, Toronto, 2006, pp. 493-498.

Hallenbeck, M. E., J. M. Ishimaru, and J. Nee. *Measurement of Recurring Versus Non-Recurring Congestion*. WA-RD 568.2. Washington State Transportation Center, October 2003.

Harvill, L.M. NCME Instructional Module (standard error of measurement). *Educational Measurement*. Vol. 10, pp. 33–41.

Hassoun, M. H. *Fundamentals of Artificial Neural Networks*. In Proceedings of the IEEE, Vol.84, 1996, pp.906-910.

Hawkins, R. and P. Stopher. Collecting Data with GPS: Those Who Reject, and Those Who Receive. In Proceedings of 27th Australian Transport Research Forum, TSC, Adelaide, Australia, 2004, pp. 29-31.

He, D., F. Ni, and S. Yang. Prediction of Freeway Travel Time in Incident Management Evaluation Based on Genetic Neural Network. In *Traffic and Transportation Studies*, 2010, pp. 778-786.

Ho, S. L., M. Xie, and T. N. Goh. A Comparative Study of Neural Network and Box-Jenkins ARIMA Modeling in Time Series Forecasting. In *Computers and Industrial Engineering*, Vol. 42, 2002, pp. 371-375.

Hobeika, A. G., and C. K. Kim. *Traffic-Flow-Prediction Systems Based on Upstream Traffic*. In Proceedings of Vehicle Navigation and Information Systems Conference, 1994, pp. 345-350.

Hu K., A. Song, M. Xia, X. Ye, and Y. Dou. An Adaptive Filtering Algorithm Based On Genetic Algorithm-Back Propagation Network. In *Mathematical Problems in Engineering*, 2013.

Hu, T. Y., and W. M. Ho. *Travel Time Prediction for Urban Networks: The Comparisons of Simulation-Based and Time-Series Models*. Presented at 17th ITS World Congress, Busan, 2010.

INRIX. *I-95 Interface Guide Version 3.3*. INRIX, June 2010.

Jiuran H, and S. Bingfeng. *The Application of ARIMA-RBF Model in Urban Rail Traffic Volume Forecast*. In Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering, Paris, France, 2013.

Karl, C.A., S. Charles, and R. Trayford. *Delivery of Real-Time and Predictive Travel Time Information: Experiences from a Melbourne Trial*. In Proceedings of 6th World Congress on Intelligent Transport Systems, 1999, Toronto, Canada.

Khan R. I, B. Landfeldt, and A. Dhamdhere. *Predicting Travel Times in Dense and Highly Varying Road Traffic Networks Using STARIMA Models*. School of Information Technologies, University of Sydney, Australia, 2012.

Khoei, A. M., A. Bhaskar, and E. Chung. *Travel Time Prediction on Signalized Urban Arterials by Applying SARIMA Modelling on Bluetooth data*. Presented at 36th Australasian Transport Research Forum, Queensland, Australia, 2013.

Kisgyorgy, L. and L. R Rilett. Travel Time Prediction by Advanced Neural Network. In *Periodica Polytechnica Civil Engineering*, Vol. 46, No. 1, 2002, pp. 15-32.

Konkaew T., S. Pan-ngum, and W. Pattara-atikom. Travel Time Prediction on Freeway by Support Vector Machines. In *Journal of Society for Transportation and Traffic Studies*, Vol. 4, 2013, pp. 34-44.

Koutsopoulos, H. N., and H. Xu. An Information Discounting Routing Strategy for Advanced Traveler Information Systems. In *Transportation Research Part C: Emerging Technologies*, Vol. 3, 1991, pp. 249-264.

Lee, S., and D. B. Fambro. Application of Subset Autoregressive Moving Average Model for Short-term Freeway Traffic Volume Forecasting. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1678,* Transportation Research Board of the National Academies, Washington, D.C., 1999, pp. 179–188.

Lee, W. H., S. S. Tseng and S. H. Tsai. A Knowledge-Based Real-Time Travel Time

Prediction System for Urban Network. In *Expert Systems with Applications*, Vol. 36, 2009, pp.4239-4247.

Li Q. Short-Time Traffic Flow Volume *Prediction Based on Support Vector Machine with Time-Dependent Structure*. In Instrumentation and Measurement Technology Conference (I2MTC '09), Singapore, 2009, pp.1730-1733.

Li, Y., A. Montasir, and X. Guo. *Urban Arterial Travel Time Prediction Using State Space Neural Network with Adaptive Filters*. Presented at International Conference on Transportation Engineering, 2009, pp. 2925-2930.

Liu H., H. van Zuylen, H. van Lint, M. Salomons. Predicting Urban Arterial Travel Time with State-Space Neural Networks and Kalman Filters. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1968*, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 99–108.

Lum, K.M., H. S. L. Fan, S. H. Lam, and P. Olszewski. Speed-Flow Modeling of Arterial Roads in Singapore. In *Journal of Transportation Engineering*, Vol. 124, No. 6, 1998, pp. 213-222.

Ma, X., Y. Wu, and Y. Wang. DRIVE Net: An E-Science of Transportation Platform for Data Sharing, Visualization, Modeling, and Analysis. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2215*, TRB, National Research Council, Washington, D.C., 2011, pp.37-49.

Marković H. Mining Spatio-Temporal Data for Travel Time Estimation in Urban Traffic Networks. Master Thesis, the Department of Electronics, Microelectronics, Computer and Intelligent Systems, Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia, 2007.

Messer, C. J., C. L. Dudek, and J. D. Friebele. Method for Predicting Travel Time and Other Operational Measures in Real-Time During Freeway Incident Conditions. Highway Research Record 461, TRB, National Research Council, Washington, D.C., 1973, pp. 1-11.

National Highway Traffic Safety Administration (NHTSA). *Traffic Safety Fact Sheet: 2013 Data*. Washington D.C., 2015. http://www-nrd.nhtsa.dot.gov/Pubs/812169.pdf. Accessed September 20, 2015.

Park, D., and L. Rilett. Forecasting Multiple-Period Freeway Link Travel Times Using Modular Neural Networks. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1617*, Transportation Research Board of the National Academies, Washington, D.C., 1998, pp. 163–170.

Park, D., and L. Rilett. Spectral Basis Neural Networks for Real-Time Travel Time Forecasting. In *Journal of Transport Engineering*, Vol. 125, 1999, pp. 515–523.

Puvvala, R. Cronbach's Coefficient as a Performance Measure to Assess Link-Level Reliability. The University of North Carolina at Charlotte, Master's Thesis, Charlotte, NC, 2014.

Puvvala, R., S. S. Pulugurtha and V. R. Duddu. *Cronbach's Alpha as a Performance Measure to Assess Link-Level Reliability*. Presented at Transportation Research Board 94th Annual meeting, Transportation Research Board of the National Academies, Washington, D.C., 2015.

Ran, B., F. Yang, Y. Tao, and Z. Qiu. *Travel Time Prediction in Presence of Traffic Incidents Using Different Types of Neural Networks*. Proceedings of Transportation Research Board 85th Annual Meeting, Washington, D.C., 2006.

Rashidi S, and P. Ranjitkar. Approximation and Short-Term Prediction of Bus Dwell Time Using AVL Data. In *Journal of Eastern Asia Society for Transportation Studies*, Vol. 10, 2013, pp. 1281–1291.

Reza, R. M. Z., S. S. Pulugurtha, and V. R. Duddu. *ARIMA Model for Forecasting Short-Term Travel Time due to Incidents in Spatio-Temporal Context*. Presented at 94th Annual Meeting of Transportation Research, Washington D.C., 2015.

Rilett, L., and D. Park. (2001) Direct Forecasting of Freeway Corridor Travel Times Using Spectral Basis Neural Networks. In *Transportation Research Record: Journal of Transportation Research Board, No. 1752*, TRB, National Research Council, Washington, D.C., 2001, pp. 140-147.

Roozemond, D. Forecasting Travel Times Based on Actuated And Historic Data. Presented at the 3rd International Conference on Urban Transport and the Environment, 1997, Acquasparta, Italy, pp. 23-25.

Schrank D., B. Eisele, and T. Lomax. *2012 Urban Mobility Report*. Texas A&M Transportation Institute, College Station, TX, 2012. tti.tamu.edu/documents/mobility-report-2012.pdf. Accessed July 15, 2015.

Sun L., J. Yang, and H. Mahmassani. Travel Time Estimation Based on Piecewise Truncated Quadratic Speed Trajectory. In *Transportation Research Part A: Policy and Practice*, Vol. 42, No. 1, pp. 173–186, 2008.

Sun, Z., Y. Wang, and J. Pan. *Short-Term Traffic Flow Forecasting Based on Clustering and Feature Selection*. IEEE World Congress on Computational Intelligence, Hong Kong, 2008, pp. 577–583.

Sun, H., H. X. Liu, H. Xiao, R. R. He, and B. Ran. Use of Local Linear Regression Model for Short-Term Traffic Forecasting. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1836*, TRB, National Research Council, Washington, D.C., 2003, pp. 143-150.

Suwardo, W., M. Napiah, and I. Kamaruddin. ARIMA Models for Bus Travel Time Prediction. In *Journal of the Institute of Engineers*, Vol. 71, 2010, pp. 49-58.

Tan M., S. Wong, J. Xu, Z. Guan, and P. Zhang. An Aggregation Approach to Short-

Term Traffic Flow Prediction. In *IEEE Transactions on Intelligent Transportation Systems*, Vol. 10, 2009, pp. 60-69.

Tennessee Department of Safety. *Strategic Plan for Highway Incident Management in Tennessee*. Office of Incident Management - Transportation Management Center and Vanderbilt University, Nashville, Tennessee, 2003.

Van Amelsfort, D. H. Valuation of Uncertainty in Travel Time and Arrival Time: Some Findings from a Choice Experiment. In *European Transport Conference*, Strasbourg, France, 2005.

Van der Voort, M., M. Dougherty, and S. Watson. Combining Kohonen Maps with ARIMA Time Series Models to Forecast Traffic Flow. In *Transportation Research Part C: Emerging Technologies*, Vol. 4, 1996, pp.307-318.

Van Lint, J. W. C., and S. P. Hoogendoorn. *Robust Travel Time Prediction with State-Space Neural Networks*. Presented at 13$^{th}$ Mini-EURO Conference, Handling Uncertainty in the Analysis of Traffic and Transportation Systems, Bari, Italy, 2002.

Van Lint, J. W. C., S. P. Hoogendoorn, H. J. van Zuylen. Accurate Freeway Travel Time Prediction with State-Space Neural Networks under Missing Data. In *Transportation Research Part C: Emerging Technologies*, Vol. 13, 2005, pp. 347–369.

Vapnik, V. The Nature of Statistical Learning Theory. Springer-Verlag, New York, 1995.

Vythoulkas, P. C. *Alternative Approaches to Short Term Traffic Forecasting for Use in Driver Information Systems*. Presented at 12$^{th}$ International Symposium on the Theory of Traffic Flow and Transportation, California, 1993.

Williams, B. M., and L. A. Hoel. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. In *Journal of Transportation Engineering*, Vol. 129, No. 6, 2003, pp. 664–672.

Wu, C. F. The Study of Vehicle Travel Time Estimation using GPS, Department of Transportation Technology & Management, Master Thesis at National Chiao Tung University, Taipei, Taiwan, 2001.

Wu, C. H., C. C. Wei, D. C. Su, M. H. Chang, and J. M. Ho. Travel Time Prediction with Support Vector Regression. Intelligent Transportation Systems. In *Proceedings of IEEE*, Vol. 2, 2003, pp. 1438 – 1442.

Xia J. Dynamic Freeway Travel Time Prediction Using Single Loop Detector and Incident Data. University of Kentucky Doctoral Dissertations, Tennessee, 2006.

Y. Wang, M. Hallenbeck, P. Cheevarunothai, and T. Northwest. *Quantifying Incident-Induced Travel Delays on Freeways Using Traffic Sensor Data*. Technical report, Transportation Northwest, University of Washington, 2008.

Yang J., L. D. Haan, and P. B. Freeze. Short-Term Freeway Speed Profiling Based on Longitudinal Spatial-Temporal Dynamics. *In Transportation Research Record: Journal of the Transportation Research Board, No. 2467*, Transportation Research Board of the National Academies, Washington, D.C., Vol. 2, 2014, pp. 62–72.

Yeon J., L. Elefteriadou, and S. Lawphongpanich. Travel Time Estimation on a Freeway Using Discrete Time Markov Chains. In *Transportation Research Part B: Methodological*, Vol. 42, No. 4, 2008, pp. 325–338.

Yildirim, U., and Z. Cataltepe. *Short Time Traffic Speed Prediction Using Data from a Number of Different Sensor Locations*. Presented at 23rd International Symposium on Computer and Information Sciences, Istanbul, Turkey, 2008, pp. 1 – 6.

Yu, C. H. An Introduction to Computing and Interpreting Cronbach Coefficient Alpha in SAS. Proceedings of 26th SAS User Group International Conference, 2001.

Zeng, D. H., J. M. Xu, J. W. Gu, L. Y. Liu, and G. Xu. Short-Term Traffic Flow Prediction Using Hybrid ARIMA and Ann Models. Presented at Power Electronics and Intelligent Transportation System, Guangzhou, China, 2008, pp.621-625.

Zeng, X. Dynamically Predicting Corridor Travel Time under Incident Conditions Using a Neural Network Approach. MS Thesis Submitted to the Office of Graduate Studies of Texas A&M University, Texas, 2009.

Zhang L., Q. C. Liu, W. C. Yang. An Improved K-Nearest Neighbor Model for Short-Term Traffic Flow Prediction. In *Procedia – Social and Behavioral Sciences*, Vol. 96, 2013, pp. 653–66.