



UNIVERSITY OF
MARYLAND

National Transportation Center

Project ID: NTC2015-SU-R-06

ADVANCED VOLATILITY MODELS FOR IMPROVING TRAVEL TIME PREDICTION

Final Report

By

Ali Haghani

Department of Civil and Environmental Engineering

University of Maryland
College Park, MD 20742

phone: (301) 405-1963

Fax: (301) 405-2585

E-mail: haghani@umd.edu

Yanru Zhang

University of Maryland

for

National Transportation Center at Maryland (NTC@Maryland)

1124 Glenn Martin Hall

University of Maryland

College Park, MD 20742

February, 2015

ACKNOWLEDGEMENTS

This project was funded by the National Transportation Center @ Maryland (NTC@Maryland), one of the five National Centers that were selected in this nationwide competition, by the Office of the Assistant Secretary for Research and Technology (OST-R), U.S. Department of Transportation (US DOT).

DISCLAIMER

The contents of this report reflect the views of the authors, who are solely responsible for the facts and the accuracy of the material and information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation University Transportation Centers Program in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof. The contents do not necessarily reflect the official views of the U.S. Government. This report does not constitute a standard, specification, or regulation.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	1
1.0 INTRODUCTION.....	2
1.1 PROBLEM STATEMENT.....	2
1.2 RESEARCH OBJECTIVES	3
1.3 RESEARCH CONTRIBUTIONS	3
1.4 REPORT OUTLINE.....	4
2.0 LITERATURE REVIEW	5
2.1 PARAMETRIC APPROACHES.....	5
2.1.1 Naïve Methods	5
2.1.2 Autoregressive Linear Processes	6
2.1.3 State Space Models	6
2.2 NON-PARAMETRIC APPROACHES.....	6
2.2.1 Non-parametric Regression	6
2.2.2 Neural Network.....	7
2.2.3 Other Artificial Intelligence Methods.....	8
2.3 HYBRID APPROACHES	8
2.3.1 Classification Based Approach.....	8
2.3.2 Classification Based Approach.....	8
2.3.3 Decomposition Technique	9
2.3.4 Ensemble Trees.....	9
2.3.5 Other Combination Approaches	10
2.4 PREDICTION INTERVAL BASED APPROACHES.....	10
2.4.1 Ensemble Methods.....	10
2.4.2 Statistical Volatility Based Approach.....	11
2.5 SUMMARY	12
3.0 STATISTICAL VOLATILITY MODELS FOR RELIABLE TRAVEL TIME PREDICTION.....	13
3.1 MEAN PREDICTION MODELS.....	14
3.1.1 Theoretical Background of ARIMA Models	14
3.1.2 ARIMA Model Optimization.....	16
3.1.3 Model Evaluation Criteria.....	16
3.2 VOLATILITY MODELS	17
3.2.1 GARCH-type Models	17
3.2.2 Component GARCH Models.....	18
3.2.3 Stochastic Volatility Model	20
3.2.4 Prediction Interval Estimation	22
3.3 APPLICATION OF COMPONENT GARCH MODELS IN TRAVEL TIME PREDICTION.....	23
3.3.1 Modeling Conditional Mean.....	23
3.3.2 Testing the ARCH Effect.....	25
3.3.3 Estimating the Volatility Model.....	25
3.3.4 Construct the Mean and Prediction Intervals.....	26
3.3.5 Results and Discussion	28

3.3.6	Summary	32
3.4	APPLICATION OF STOCHASTIC VOLATILITY MODEL	33
3.4.1	Model Fitting	37
3.4.2	Results and Analysis	38
3.4.3	Summary	42
4.0	CONCLUSION AND RECOMMENDATIONS.....	45
4.1	SUMMARY	45
4.2	CONCLUSION.....	46
4.3	FUTURE RECOMMENDATION.....	47
5.0	REFERENCES.....	48

LIST OF TABLES

Table 1: Estimated MPIL and PICP values for GARCH, C-GARCH and MC-GARCH models.	29
Table 2: Selected segments for this study.....	35

LIST OF FIGURES

Figure 1: Concept of prediction interval (Shrestha and Solomatine 2006).	14
Figure 2: Box Plot of Absolute Deviation from Predicted Mean.	24
Figure 3: Multiplicative component GARCH forecasting results: decomposition of the volatility into its various components (32-33).	27
Figure 4: Predicted mean and PI for multiplicative component GARCH model.	28
Figure 5: Comparing performance of GARCH, C-GARCH and MC-GARCH models during peak hours.....	30
Figure 6: Comparing performance of GARCH, C-GARCH and MC-GARCH during non-peak hours.....	31
Figure 7: Comparison of prediction intervals constructed by GARCH, C-GARCH and MC-GARCH models.....	32
Figure 8: Bluetooth sensor location of the study.	34
Figure 9: A scatter plot of travel times on four paths.	36
Figure 10: Prediction results for peak hour travel time at four segments.	39
Figure 11: Comparison of performance measures for six dataset by using ARIMA-GARCH and ARIMA-SV model with (a) 5 minute time interval (b) 15 minute time interval.....	42

EXECUTIVE SUMMARY

Travel time effectively measures freeway traffic conditions. Easy access to this information provides the potential to alleviate traffic congestion and to increase the reliability in road networks. Accurate travel time information through Advanced Traveler Information Systems (ATIS) can provide guidance for travelers' decisions on departure time, route, and mode choice, and reduce travelers' stress and anxiety. In addition, travel time information can be used to present the current or future traffic state in a network and provide assistance for transportation agencies in proactively developing Advanced Traffic Management System (ATMS) strategies. Despite its importance, it is still a challenging task to model and estimate travel time, as traffic often has irregular fluctuations. These fluctuations result from the interactions among different vehicle-driver combinations and exogenous factors such as traffic incidents, weather, demand, and roadway conditions. Travel time is especially sensitive to the exogenous factors when operating at or near the roadway's capacity, where congestion occurs. Small changes in traffic demand or the occurrence of an incident can greatly affect the travel time. As it is impossible to take into consideration every impact of these unpredictable exogenous factors in the modeling process, travel time prediction problem is often associated with uncertainty. This research uses innovative data mining approaches such as advanced statistical and machine learning algorithms to study uncertainty associated with travel time prediction.

1.0 INTRODUCTION

Travel time is widely acknowledged as an effective measure of highway traffic conditions, which can be easily understood by both travelers and transportation agencies (Yeon *et al.* 2008). Access to accurate travel time information has the potential to alleviate traffic congestion, to minimize its negative environmental and societal side effects, and to increase road network reliability (Choe *et al.* 2002). Advanced Traveler Information System (ATIS) provides travelers with accurate and timely traffic information via dynamic message signs, radio, and internet. The pre-trip travel time information gives guidance for travelers' decisions for departure time, route, and mode choice, and reduces travelers' stress and anxiety. In addition, travel time information can also be used to present the current or future traffic state in a network and to provide assistance for transportation agencies in proactively developing Advanced Traffic Management System (ATMS) strategies. For example, travel time is one of the performance measures in the Freeway Performance Measurement System (PeMS) developed by the California Department of Transportation (Choe *et al.* 2002). The Split Cycle Offset Optimization Technique system and the Sydney Coordinated Adaptive Traffic system are two successful traffic operation systems that use travel time information as their module input (Yang *et al.* 2010). Therefore, travel time information is a critical input and output for intelligent transportation system.

1.1 PROBLEM STATEMENT

The success of ATIS and ATMS relies not only on the availability and accuracy of historical and real time traffic information, but also on future traffic information. A wide range of methodologies in travel time forecasting has been proposed to model traffic characteristics and to produce short term forecasts. Most of these methods are based on the historical travel time data concurrently collected from various detection systems, such as vehicles with GPS or Bluetooth devices, electronic toll system, and video detection. Especially with the recent technology advances in vehicle tracking, direct and accurate travel time information can be easily obtained. The technology improvements make the development of an online travel time prediction algorithm more meaningful.

Despite the proliferation of traffic prediction methodologies in the existing literature, modeling and estimating travel time is still a challenging task (Yildirimoglu and Geroliminis 2013). Travel time experiences strong fluctuations across different periods and traffic conditions. These fluctuations result from the interactions among different vehicle-driver combinations, and exogenous factors such as traffic incidents, weather, demand, and roadway conditions. Travel time is especially sensitive to the exogenous factors when operating at or near the roadway's capacity, where congestion occurs. Small changes in traffic demand or the occurrence of an incident can greatly affect the travel time. As the impact of these unpredictable exogenous factors is impossible to be considered fully in the modeling process, travel time prediction problem is often associated with uncertainty.

Addressing uncertainties associated with travel time, and therefore, travel time reliability has become a topic of interest in recent years. FHWA defines travel time reliability as "the consistency or dependability in travel times, as measured from day-to-day and/or across different times of the day" (Ahmed and Cook 1979), and proposes several travel time reliability measures: 90th or 95th

percentile travel time, buffer index, and planning time index. For prediction purpose, another efficient measure is prediction intervals (PIs) (Levin and Tsao 1980), based on which one can assess the reliability of the travel time forecasting results. Prediction interval is an estimated interval, which covers the expected travel time value with a predetermined probability (Davis *et al.* 1990). In other words, PIs give a likely range of the predicting results to represent the uncertainties associated with travel time. Availability of PIs allows the travelers and traffic managers to quantify the level of uncertainty associated with predicted travel time and thus to make multiple strategies on route and departure time choice to deal with the worst and best conditions. Wide PIs indicate a higher uncertainty of the future traffic conditions and travelers should expect extreme delays, while narrow PIs mean the traffic condition are relatively stable (Hamed *et al.* 1995).

In brief, due to the dynamic and stochastic nature of traffic, travel time prediction is one of the most challenging tasks in ATIS and ATMS. In order to provide meaningful traffic information to travelers and traffic managers, it is critical to develop an accurate and reliable traffic prediction algorithm that not only reduces the absolute value of prediction error but also takes into consideration the uncertainty associated with travel time prediction.

1.2 RESEARCH OBJECTIVES

The primary goal of this research is to identify uncertainties associated with travel time prediction. Both accuracy and reliability issues are addressed in terms of freeway travel time prediction. Prediction accuracy emphasizes the difference between the predicted and the actual value, or in other words, prediction error. Most existing travel time prediction methods in the literature focus on improving travel time prediction accuracy without considering the uncertainty issue. On the other hand, reliability puts more emphasis on uncertainty associated with prediction. Instead of providing a point value (an average of travel time during a certain time interval), a prediction interval is proposed to represent how likely it will capture the observed value. To achieve both objectives, this research introduces and implements a data driven approach to predict travel time and to assess uncertainty associated with predictions.

A statistical volatility model is promising in terms of modeling uncertainty, as it not only provides the opportunity to develop a more accurate mean model but also produce an effective and efficient prediction interval. This research proposes and compares different types of statistical volatility-based travel time prediction models. The preliminary study results indicate that a statistical volatility model can be a promising approach to account for uncertainties associated with prediction.

1.3 RESEARCH CONTRIBUTIONS

This section lists the main contributions to the state-of-the-art research offered in this report.

We have developed two innovative component volatility-based travel time prediction models to better characterize long-term and short-term volatility and cyclical patterns in travel time data. Because of the daily, weekly or even monthly recurrent traffic congestion, travel time data often show strong cyclical patterns. In the traffic prediction field, existing volatility models do not consider the possible cyclical patterns in the residual series, often referred to as a seasonal

component. Conventional generalized autoregressive conditional heteroskedasticity (GARCH) models are often criticized as unsatisfactorily modeling data series that show pronounced seasonal patterns (Engle and Sokalska 2012). The decomposition technologies provide the potential to deal with trend and seasonal components in the data. Driven by the successful application of statistical volatility models in transportation analyses, the component GARCH models are proposed. The component GARCH models are similar to the structure of the GARCH model but include trend and seasonal elements. The component GARCH models allow for a more versatile structure with the potential to provide more accurate traffic volatility forecasting along freeway corridors.

We have introduced advanced solution technologies, namely Bayesian inference using the Markov Chain Monte Carlo (MCMC) based on ancillarity-sufficiency interweaving strategy (ASIS) proposed by Kastner and S. Frühwirth-Schnatter (2013), for stochastic volatility (SV) model in travel time volatility forecasting. The proposed method greatly improves the efficiency and robustness of the SV model. We have compared the proposed SV model with the GARCH model by using freeway travel time data and have demonstrated the advanced SV model as a competitive alternative in modeling the volatility nature in traffic.

1.4 REPORT OUTLINE

The rest of this report is organized as follows. The next chapter reviews previous works in freeway travel time prediction. The existing literature is categorized as parametric, non-parametric, hybrid and prediction interval based approaches. Chapter 3 describes statistical volatility models in travel time reliability prediction. The model structure and the concept of prediction interval are introduced. The statistical volatility models are composed of two parts: mean and prediction interval. The mean prediction models and the volatility models (to construct prediction intervals) are discussed in Chapter 3.1 and in Chapter 3.2. Chapter 3.3 performs a case study using volatility models in travel time prediction. Chapter 4 concludes the report and also provides further recommendations for future research in this topic.

2.0 LITERATURE REVIEW

Short-term traffic prediction as a critical component in a real-time ITS environment has seen an explosion of interest since the 1980s. A large number of forecasting algorithms have been proposed in the literature. Because of the complexity of the traffic prediction problem, existing traffic prediction approaches are different from one another in different aspects. Vlahogianni *et al.* (2004) suggested that the process of developing short-term traffic prediction can be divided into three essential clusters (scope, conceptual output and modeling) that involves two important issues (design and modeling parameters). The designing process mainly focuses on the objective of forecasting that includes what types of application, where to implement, and the desired output of the model. The modeling parameters procedure is the way to achieve the goal determined during the design process. This study belongs to the modeling parameters procedure.

Since this study primarily focuses on freeway travel time prediction, we will emphasize the literature on different modeling techniques in this area. There are several review or comparative studies in existing short-term traffic prediction methods (Vlahogianni *et al.* 2004; Arem *et al.* 1997; Ishak and Al-Deek 2002; Van Lint and Van Hinsbergen 2012) and interested readers can refer to these studies for further references. Vlahogianni *et al.* (2004) classified traffic prediction models as consisting of three modeling approaches: parametric, non-parametric and hybrid methods. The parametric approach usually assumes a specific form for the dependent and independent variables. The modeling process involves model identification, parameter estimation, model diagnostic checking and prediction. The parametric models often have more assumptions than non-parametric methods. If these assumptions are satisfied by the data, the parametric approach can produce accurate estimation. Otherwise, the parametric approach can be misleading. The non-parametric models are data driven approaches that usually do not assume a specific structure of the data. These algorithms heavily depend on the quality of the available data. Another traffic forecasting approach is the hybrid method that combines different models or the same type of models with different initial values or parameters to obtain better prediction performances. In addition, as uncertainties are often involved in traffic prediction, prediction interval-based approaches consider the future traffic parameter (such as volume, speed or travel time) as a distribution instead of a point value. Mostly existing prediction-interval based approaches belong to the hybrid method. Because of its importance, we discuss this approach as a separate category. The following sections will summarize the major research findings of existing literature based on the taxonomy proposed by Vlahogianni *et al.* (2004).

2.1 PARAMETRIC APPROACHES

2.1.1 Naïve Methods

The naïve method can be interpreted as a simple and easy implementation method without many model assumptions. Historical average and smoothing (Farokhi Sadabadi *et al.* 2010) techniques received extensive attention in practical applications (Smith and Demetsky 1997; Williams 1998). The historical average model simply averages historical traffic data on either a certain time of the day, day of the week or other time periods based on the assumption that traffic shows similar patterns throughout the day, week or year, e.g. traffic patterns day to day often show remarkable similarities and these patterns are useful for prediction. The historical average method has already

been applied to the urban traffic control systems (UTCS) (Stephanedes et al. 1981) and other various traveler information systems (Jeffery *et al.* 1987; Kaysi *et al.* 1993). However, as traffic conditions are highly dynamic, the naïve method is often a poor predictor.

2.1.2 Autoregressive Linear Processes

In the early 1990s, transportation researchers developed an alternative approach—autoregressive linear process such as the autoregressive integrated moving average (ARIMA) type models in predicting traffic. The ARIMA-type models were first introduced by Ahmed and Cook (Ahmed and Cook 1979) and Levin and Tsao (1980) in freeway traffic flow and occupancy prediction. Their studies indicated that the ARIMA-type models provide better forecasting accuracy compared with historical average and smoothing techniques. Applications of the ARIMA model in traffic parameters forecasting was also discussed in later studies (Davis *et al.* 1990; Hamed *et al.* 1995; Kamarianakis and Prastacos, 2005; Williams *et al.* 1998; Cetin and Gomert 2006). As traffic parameters show spatial-temporal correlations, both Kamarianakis and Prastacos (2005) and Min and Wynter (2011) adopted a multivariate spatial-temporal autoregressive moving average model to predict traffic flow. Because of its well defined theoretical foundation and effectiveness in prediction (2009), the ARIMA-type models gradually become standard methods to compare with newly developed forecasting models. However, the ARIMA-type models are sensitive to extreme values. This makes the ARIMA-type model less efficient when modeling data with large variations.

2.1.3 State Space Models

State space models belong to the multivariate forecasting category, which can be applied to multiple inputs – multiple outputs systems. It is worth noting that the ‘state space model’ and the more widely known ‘Kalman filter model’ refer to the same model structure. The term ‘state space’ refers to the model, the term ‘Kalman filter’ refers to the process of estimating and updating model parameters. Stathopoulos and Karlaftis (2003) applied a multivariate state space model to predict flow at an urban signalized arterial. Ghosh *et al.* (2009) applied a structural time series model to forecast traffic flow in a congested urban transportation network. The structural time series model is a special form of the state-space model that represents the observed time series as a sum of different components. Their study results indicate that the proposed model is computationally more efficient and can trace the evolution of each individual component separately. The Kalman filtering algorithm allows the parameters of the model to be updated with new data available. Therefore it enables dynamic traffic prediction (Whittaker *et al.* 1997; Okutani and Stephanedes 1984; Chien and Kuchipudi 2003; Nanthawichit *et al.* 2003; van Lint 2008; Wang *et al.* 2008; Yang *et al.* 2004)

2.2 NON-PARAMETRIC APPROACHES

2.2.1 Non-parametric Regression

Nonparametric regression is a form of regression analysis that does not predetermine a specific form for the predictor. It relies on data to describe the relationship between dependent and independent variables and is based on the principle of pattern recognition and chaotic systems (Smith *et al.* 2002). Smith and Demetsky (1996) demonstrated the advantages of the nonparametric regression approach when compared with the neural network. Another study by Smith *et al.* (2002)

suggested that the heuristic forecast generation method improves the performance of the nonparametric regression but does not necessarily perform better than the seasonal ARIMA model. Clark (2003) proposed a nonparametric regression technique described as a k nearest neighbor (k-NN) model to predict traffic state variables. Davis and Nihan (1991) applied the k-NN model in predicting freeway traffic. They suggested that the k-NN method is comparable to the linear time-series approach. Robinson and Polak (2005) proposed the use of the k-NN technique to model urban link travel time. They discussed in details the selection of a distance metric, local estimation measure and value of k. Myung *et al.* (2011) applied the k-NN to predict travel time with data from a vehicle detector system and an automatic toll collection system. Zou *et al.* (2009) utilized both k-NN and a multi-topology neural network model in predicting freeway travel time. Their proposed model provides reliable travel time predictions in uncongested, congested, and transition traffic conditions.

2.2.2 Neural Network

Van lint *et al.* (2002) applied a state space neural network that is capable of dealing with the spatio-temporal relationships of traffic. Yin *et al.* (2002) developed a fuzzy-neural model (FNM) to predict traffic flow in an urban network. Their model applied a gate network (GN) that divides input data into several clusters using a fuzzy approach, and then applied the expert network (EN) to specify the input-output relationship. Ishak *et al.* (2003, 2004) compared three different neural networks: simple recurrent networks (Jordan–Elman), partial recurrent networks (PRNs), and time-lagged feed forward networks (TLFN), with different input parameters for traffic prediction. Jiang and Adeli (2005) proposed the dynamic time-delay wavelet neural network model in freeway traffic flow forecasting. The proposed model considers both the time of the day and the day of the week when predicting traffic flow. Liu *et al.* (2005) applied the state-space neural network model to predict travel time when there is missing data. Their proposed method is insensitive to missing data and can provide accurate forecasting. Quek *et al.* (20026) applied a specific class of fuzzy neural network models in short term traffic flow prediction. The proposed model known as a pseudo outer-product fuzzy neural network using the truth-value-restriction method (POPFNN-TVR) was shown to outperform the conventional feed forward neural network using back propagation (BP) learning.

Zheng *et al.* (2006) suggested that a certain model has superior performance for a particular time period and combining single neural network predictors may improve forecasting accuracy. They developed a Bayesian combined neural network model that combines the back propagation and the radial basis function neural networks in traffic flow forecasting. The credit of each individual model is estimated based on the theory of conditional probability and Bayes' rule and largely depends on the accumulative prediction performance of previous time step. Zeng & Zhang (2013) applied four different neural network models and proposed one model in freeway travel time forecasting. These models include multilayer feed forward neural network, time-delay neural network, state-space neural network, and nonlinear autoregressive with exogenous inputs. In their study, they analyzed the effect of different input variables and the temporal-spatial inputs on model performance. Their study results indicate that the temporal-spatial inputs can greatly improve the model performance; the state-space neural network and the time delayed state-space neural network outperform the other models.

2.2.3 Other Artificial Intelligence Methods

Sun *et al.* (2006) considered spatial and temporal correlations of traffic flow among adjacent road links and developed a Bayesian network approach to forecast traffic flow. In their paper, the joint probability distribution between the upstream and downstream locations is described as a Gaussian mixture model (GMM) with parameters estimated via the competitive expectation maximization (CEM) algorithm. Hong (2011) proposed a traffic flow forecasting model that combines the seasonal support vector regression model with chaotic simulated annealing algorithm (SSVRCSA). In his paper, the chaotic simulated annealing algorithm is proposed to determine the value of three parameters in a SVR model; seasonal adjustment factors are then applied to deal with the cyclic trend.

2.3 HYBRID APPROACHES

2.3.1 Classification Based Approach

The classification-based hybrid approach classifies traffic data into different groups first and then assigns different models according to the characteristics of data in different groups. Danech-Pajouh and Aron (1991) proposed an ATHENA model that groups data according to their similarities and then assigns a different linear model to each cluster. Antoniou *et al.* (2013) developed a dynamic data-driven framework for traffic state estimation and prediction. Their model first clusters existing observations into several groups, and then predicts the future traffic state by modeling the evolution of traffic history states as a Markov process. By estimating a flexible model for each cluster, future traffic speed can be obtained. Van Der Voort *et al.* (1996) combined Kohonen maps with ARIMA time series. The study results were promising compared with the ATHENA model. Later, Chen *et al.* (2001) used a self-organizing map (SOM) to initially classify traffic data into different groups and then applied the ARIMA and the multi-layer perception (MLP) model as two prediction methods. Their study results suggested that the SOM/ARIMA hybrid approach is more sensitive to missing data than the SOM/MLP hybrid approach.

2.3.2 Classification Based Approach

The Kalman filtering method is a promising method to train and update model parameters and has been applied to different kinds of forecasting models to enable continuous parameter updating. Yu *et al.* (2010) applied the support vector machine (SVM) method to predict baseline travel time and used the Kalman filtering technique to adjust the prediction results with updated information. Stathopoulos and Dimitriou (2008) proposed a forecasting approach that utilizes a fuzzy rule based system (FRBS) that nonlinearly combines traffic flow forecasting results from an online adaptive Kalman filter (KF) and an artificial neural network (ANN) model. Their study results indicate that the combined approach improves the forecasting accuracy compared with each individual model. Liu *et al.* (2006) proposed the extended Kalman filter (EKF) method to train the parameters of the state-space neural networks (SSNN). The proposed algorithm was tested in an urban network and the study results indicate that the proposed method is 20 times faster than the SSNNLM model with a slightly worse forecasting accuracy.

2.3.3 Decomposition Technique

Decomposition techniques decompose a complicated data set into small elements. In terms of prediction, decomposition techniques can be utilized to reduce noisy information in traffic data to improve their prediction performance (2010); they could also be used as the basis for combining different models (2012). Some popular decomposition techniques include: Fourier methods, discrete wavelet transform (DWT) and empirical mode decomposition (EMD).

Hamad *et al.* (2009) suggested that more accurate prediction of speed data can be obtained through decomposing the time series into its basic components. They utilized the empirical model decomposition to filter out unimportant elements and applied a multilayer, feed forward neural network with BP to predict freeway travel speed. Later, Chen and Wu (2012) applied empirical model decomposition and gray theory (1989) in predicting bus travel time. Wei and Chen (2012) forecasted metro passenger flow through empirical model decomposition and neural networks. They applied the EMD method to decompose traffic flow as several intrinsic mode function (IMF) components and selected the important information as input for back-propagation neural networks (BPNN). Their study results indicate that treating the important and non-important IMF as different inputs of the BPNN would improve the forecasting accuracy. Jiang and Adeli (2004) proposed a hybrid wavelet packet-ACF method to analyze traffic flow time series and concluded that the discrete wavelet packet transform method de-noises the signal even more effectively than the conventional wavelet transform. Xie *et al.* (2007) applied a discrete wavelet decomposition method to remove noise in original traffic data and utilized the Kalman filter prediction model to the modified data to predict future traffic. Their study results indicate that removing noise in the original traffic data has the potential to improve the performance of a Kalman filter model in traffic volume forecasting. Daniel *et al.* (2010) applied a wavelet based method to remove noise in the original traffic data and applied self-organizing neural networks as the prediction method. Wang and Shi (2012) developed a hybrid traffic speed forecasting model based on support vector machine (SVM) regression theory. In their study, they constructed a new kernel function using a wavelet function to capture the non-stationary characteristics of the data and then used the Phase Space Reconstruction theory to identify the input space dimension. They assumed that the collected data are often accompanied with measurement errors; therefore they applied wavelet de-noising method to remove the noise in the traffic speed data.

2.3.4 Ensemble Trees

Leshem and Ritov (2007) proposed a traffic flow prediction algorithm by combining Random Forests algorithm into an Adaboost algorithm as a weak learner. The proposed algorithm is proved to be able to deal with missing data and is effective in predicting multiclass classification problems. Hamner (2010) applied a random forest method in travel time prediction and their method is able to provide accurate travel time prediction. Wang (2011) applied an ensemble bagging decision tree (ensemble BDT) to predict weather impact on airport capacity and demonstrated the superior performance of ensemble BDT compared with single SVM classifier. Ahmed and Abdel-Aty (2013) utilized a stochastic gradient boosting method in identifying hazardous conditions based on traffic data collected from different sensors. Their study results suggested that the proposed stochastic gradient boosting method has considerable advantages over classical statistical approaches. Similarly, Chung (2013) applied boosted regression trees to study crash occurrence. Both studies utilized the boosting method to study classification problems.

2.3.5 Other Combination Approaches

Zheng *et al.* (2006) proposed a freeway traffic flow prediction method that combined a back propagation neural network and a radial basis function neural network based on Bayesian model combination approach. The output of the proposed model is a weighted combination of the output of the two neural networks and the weight is estimated based on conditional probability and Bayes' rule. Hinsbergen *et al.* (2009a, 2009b) trained feed-forward neural networks and state-space neural networks using Bayesian inference theory. A simple average over all group members was used to combine neural networks into a group. Their study results indicate that the proposed framework provides a more accurate forecasting of both the mean and the prediction intervals. Zhang and Liu (2011) predicted travel time index by utilizing six baseline individual predictors as basic combination components and combined them through four combined predictors including equal weights (EW), optimal weights (OW), minimum error (ME) and minimum variance (MV) methods. Here, travel time index is the ratio of average travel time and free flow travel time. Vlahogianni *et al.* (2007) proposed a modular neural network prediction model that considers both spatial and temporal correlations of traffic data. In their proposed model, the spatial representation of traffic information collected from individual location is addressed through a system's modularity. Each module consists of a time delayed feed-forward neural network (TDNN) that represents the time evolution of traffic for a corresponding location. The temporal optimization of the input windows in each TDNN is through genetic algorithms (GAs).

2.4 PREDICTION INTERVAL BASED APPROACHES

Although a wide range of approaches has been applied to the traffic prediction field and has shown promising predicting abilities, some of them have limited abilities to capture the uncertainty and variability of traffic, as they only provide a point estimate to represent future traffic conditions. Traffic condition is a complex phenomenon, as it is often affected by the interactions among different vehicles and exogenous factors such as incident, weather, demand, and roadway conditions. Small changes in current traffic conditions may greatly affect future travel time. For example, an incident during peak hour may result in extreme delays in the near future. Due to the highly dynamic nature of traffic, predicting travel time is often associated with uncertainty, especially during non-recurrent congestion when incident or bad weather occurs. A point estimate provides limited information regarding the uncertainty and unreliability of travel time. On the other hand, prediction intervals (PIs) have the potential to provide more reliable forecasting results by providing a confidence band to indicate how reliable the forecasting results are. There are few studies using prediction intervals to model uncertainties associated with travel time prediction.

2.4.1 Ensemble Methods

Khosravi *et al.* (2011a, 2011b) developed different neural network based approaches to provide PIs to capture uncertainties in travel time. As there is always a mismatch between the predicted and actual values, PIs provide a range that can capture the uncertainty. Van Lint (2006) proposed an ensemble of state-space neural network (SSNN) models to predict prediction intervals and mean travel time. The constructed prediction interval captures the uncertainty associated with travel time prediction. Zeng and Zhang (2013) employed an ensemble method (using multiple instances of the same neural network model with different initial conditions) to derive a prediction band, but the method can be computationally intense. Fei *et al.* (2011) proposed a Bayesian inference-based

dynamic linear model that considers freeway travel time as the sum of the median of historical travel times, time varying random variations in travel time, and a model evolution error. Their proposed model prediction result is a travel time distribution that can generate a mean and a prediction interval representing uncertainty associated with travel time prediction. Van Hinsbergen (2009) proposed an approach that combines neural networks in a group using Bayesian inference theory to predict travel time with prediction intervals. Li and Rose (2011) developed a model that models average travel time and travel time variability separately to incorporate uncertainty in travel time prediction. All these studies indicate that travel prediction is a complex problem often associated with uncertainty. Prediction interval based approaches provide the potential to capture the dynamic changes of traffic.

2.4.2 Statistical Volatility Based Approach

Besides the ensemble approaches, another popular method that was able to capture the uncertainty and variations of data is the statistical volatility approach (Engle 1982). This approach relaxes the constant variance assumption and models time changing variance as a function of its past values. As a result, the statistical volatility modeling approach can capture the dynamic changes of travel time variations and can provide more accurate PIs. The first volatility model, the ARCH model, was proposed by Engle in 1982 for financial analysis purposes (1982). Later, different variations of the ARCH model were formulated and the GARCH (1986) model is one of the most widely used models. Driven by its successful applications in financial and other areas (2011), transportation professionals began to apply the family of GARCH models to predict traffic volatilities. Kamarianakis *et al.* (2005) suggested that traffic conditions are much more volatile during heavy traffic or congestion periods than at other times and effective modeling of variance can produce more accurate confidence intervals. They tested the performance of the ARIMA-GARCH model by using traffic flow data in an urban network and demonstrated that traffic flow data displayed time dependent volatilities. They also suggested that further studies should consider the asymmetric effects of positive and negative shocks.

Zhang *et al.* (2013) considered the asymmetric effects of positive and negative shocks and studied two asymmetric GARCH models: EGARCH and GJR-GARCH in travel time forecasting. Their study result indicated that the GJR-GARCH model performs better. Tsekeris and Stathopoulos (2006) incorporated fractionally integrated components in both the conditional mean and the conditional variance equations and proposed the ARFIMA-FIAPARCH model. They found that the proposed model improves the accuracy of predicted volatility. Similarly, Karlaftis and Vlahogianni (2009) suggested over-differentiation leads to over-inflated MA terms and applied the ARFIMA-FIGARCH model for traffic flow prediction. Tsekeris and Stathopoulos (2009) predicted urban traffic variability through a stochastic volatility modeling approach. Their study results demonstrated that the stochastic volatility model outperforms the GARCH model as a latent stochastic process and can better represent the speed variability dynamics than a stochastic process with a predetermined structure concerning the decaying impact of shocks. Yang *et al.* (2010) applied seasonal ARIMA, ANN and historical mean methods as the mean equation and the GARCH model to predict urban vehicle travel time. Their study results indicate that the proper selection of the mean equation can lead to excellent results. Xia *et al.* (2013) considered the relationship between flow and speed and proposed a VAR-MGARCH model to predict traffic flow and speed for urban roads. Guo (2010) proposed an online algorithm of the autoregressive moving average (ARMA)-GARCH model through Kalman filters in predicting traffic speed.

2.5 SUMMARY

In summary, traffic prediction algorithms can be categorized into parametric, non-parametric and hybrid approaches. The parametric approaches usually have a clear model structure and well established theoretical foundations. Compared with nonparametric approaches, they are easier to interpret. Some of them, for example, the historical average, have been implemented in existing traffic control devices for providing traffic information. However, this type of method sets a series of strict model assumptions. Misusing the model with the wrong data (data that does not meet certain model assumptions) will lead to inaccurate prediction. As traffic in different locations shows different characteristics, it is necessary to understand both the model and the data to select the appropriate model.

The non-parametric approach usually has fewer modeling assumptions compared with the parametric approach. Some popular methods include non-parametric regression, neural networks, SVM and other artificial intelligent methods. The structure of this type of model is usually developed based on the data. Especially, the neural network methods are analogous to a ‘black-box’. The users give some simple inputs to the model and get decent predictions, but they usually are unaware of the structure of the model. During the past few decades, the non-parametric methods attracted significant attention in the traffic prediction field because of their ability in modeling complex data.

The hybrid methods consider traffic as a complex phenomenon that cannot be represented by a single model. By taking advantage of different models, it is aimed at improving model prediction performance. However, because it involves several different models, the hybrid methods are often complex.

Besides prediction accuracy, it is also critical to develop a reliable prediction model that considers uncertainty associated with travel time prediction. Therefore, the prediction interval based approach is proposed to take into consideration this uncertainty. There are generally two categories: ensemble and statistical volatility methods. The ensemble method constructs the prediction interval through developing different base models while the statistical volatility method models the evolution of the changing behavior of the variance part of traffic data.

Although a large number of traffic prediction algorithms are proposed in the literature, prediction accuracy and reliability are still two challenging issues. In terms of prediction accuracy, most existing models can be highly accurate during non-peak hours as variations of travel time are not significant from day to day. However, prediction accuracy deteriorates during peak hour. Improving travel time prediction, especially during congested periods, is critical. To address this issue, it is important to develop an advanced prediction method that is able to model the complex relation of traffic data. As mentioned in the previous section, the ensemble methods have shown promising prediction results. This research studies and proposes a novel ensemble method for predicting travel time. At the same time, reliability is also an essential issue in travel time prediction. Since there is always a mismatch between the predicted and the actual value, there is a need to measure or predict this “mismatch” or uncertainty. Prediction interval based approaches are able to model this uncertainty. However, it is a relatively new concept and there is limited literature in this field. This research will further explore more advanced prediction interval based approaches to better model this uncertainty.

3.0 STATISTICAL VOLATILITY MODELS FOR RELIABLE TRAVEL TIME PREDICTION

The observed travel time can be decomposed into a conditional mean (u_t) and a residual (r_t) component. The traditional time series based travel time prediction methods study correlations between travel time at different time lags or at different locations, assuming constant variance of the data (r_t is constant across different time intervals). Therefore, they only model the time variation of the data for the first order moment and predict the mean part u_t . However, uncertainty often exists in the data, especially for travel time, which can be dramatically affected by some unexpected external factors, such as bottlenecks, traffic incidents, work zones, weather and special events. The point prediction results become less reliable because of the presence of these unexpected factors. As travelers are used to the daily congestion due to regular traffic demand, it is the unexpected delays that generate the most dissatisfaction. Modeling the uncertainties, referred to as the conditional standard deviation (or residual r_t), would improve forecasting reliability.

Equation (1) is the basic structure of a travel time prediction model.

$$x_t = u_t + r_t \quad (1)$$

where x_t is the observed travel time at time t , u_t represents the estimated conditional mean, and r_t is the residual part.

Traditional prediction methods only focus on the estimated conditional mean (u_t) component and treat the residual (r_t) part as having a constant variance. However, in real situations, the variations in traffic and travel times can be different during different time periods. Therefore, prediction interval based approaches are proposed to model this uncertainty (the residual part r_t). By providing a prediction interval, we have an idea of how likely this estimated range would capture future travel time. In other words, a prediction interval is an estimated range that captures the future observation, with a prescribed probability, given the current available observations. As illustrated in Figure 1, a prediction interval is comprised of an upper and lower prediction limit that indicates the accuracy of the model output with respect to the observed value. The traffic is a complex and uncertain system. Due to the uncertainty related with the data and the estimated model, there is often a mismatch between the model output and observed value. Models that only provide a point value (the predicted value) have limited abilities to capture the uncertainty and variability of travel time especially during congested situations. PIs provide a range to indicate how likely the travel time is during the next time interval. Therefore, PIs have the potential to capture the fluctuations and the stochastic traffic phenomena. Usually, a wider prediction interval is associated with larger variation in travel time.

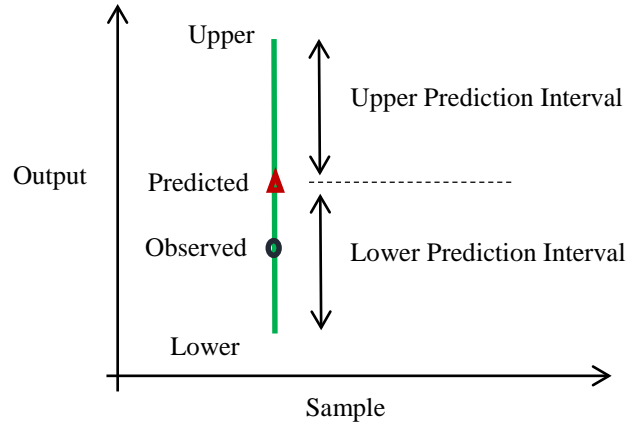


Figure 1: Concept of prediction interval (Shrestha and Solomatine 2006).

The statistical volatility-based travel time prediction models include two parts: predicted mean value (the red triangle in Figure 1) and prediction interval (the green vertical lines). This section will first study efficient mean prediction models for short-term travel time prediction purposes. The second part studies the application of statistical volatility model in predicting the variance part of the travel time data, and the construction of travel time prediction intervals to account for uncertainties associated with prediction.

3.1 MEAN PREDICTION MODELS

The first step of the modeling stage was to estimate the mean of the data. In the literature of traffic parameters forecasting, different mean equation models have been tested. For example, Kamarianakis *et al.* (2005) applied the ARIMA model as the mean equation of the volatility model. Karlaftis and Vlahogianni (2009) proposed using the ARFIMA model to capture the long memory in the conditional mean. Yang *et al.* (2010) adopted three different mean equations for volatility models: SARIMA, ANN, and historical average methods. Among these existing methods that have been proposed in the literature, the ARIMA-type model becomes one of the most widely used methods due to its ease of implementation and its well-known ability in traffic parameters modeling and forecasting (Karlaftis and Vlahogianni 2009). Therefore, for the purpose of studying the performance of different volatility models, this section applies the ARIMA model as the mean equation. However, it is worth noting that a proper mean equation model should not be restricted to the ARMA-type model. Properly choosing a mean equation regarding the structure of the data can lead to better model performance (Yang *et al.* 2010).

3.1.1 Theoretical Background of ARIMA Models

This section introduces the ARIMA model to capture the mean of this variation. ARIMA models, as one of the most general classes of time series models, predict the current value based on its past values. The $ARIMA(p, d, q)$ model is comprised of three parts: autoregressive $AR(p)$, integrated $I(d)$, and moving average $MA(q)$. If we define B as the backshift operator with

$B^k x_t = x_{t-k}$, The corresponding part of the ARIMA model for a given time series $\{x_1, x_2 \dots x_n\}$ consists of the following parts:

The autoregressive part of order p is denoted as $AR(p)$ and is formulated as:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + z_t \quad (1)$$

where $\phi_1 \dots \phi_p$ are parameters, z_t is a white noise process with zero mean and variance σ_z^2 . The equation can be rewritten in the following form by using the backshift operator:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)x_t = z_t \quad (2)$$

Or more concisely as:

$$\phi(B)x_t = z_t \quad (3)$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$.

As many time series are non-stationary, it is necessary to transform the original data to a stationary series. There are several different ways to transfer the data to stationary, such as difference the data, remove the trend of the data (if the data contain a trend), taking the logarithm or square root of the series for data with non-constant variance. In travel time prediction, differencing the original data is most often used technique to achieve stationary. Here differencing means taking the difference of two observations that are d periods apart. The integrated part with order d , denoted as $I(d)$, means the d th difference of the original data:

$$(1 - B)^d x_t \quad (4)$$

The moving average part with order q , denoted as $MA(q)$, is the process that the current value of x_t is a linear combination of a white noise series:

$$x_t = z_t + \theta_1 z_{t-1} + \theta_2 z_{t-2} + \dots + \theta_q z_{t-q} \quad (5)$$

where $\theta_1 \dots \theta_q$ are parameters, $z_t \dots z_{t-q}$ are white noise processes with zero mean and variance σ_z^2 . Equation (5) can be rewritten in a more concise form by using the backshift operator:

$$x_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)z_t \quad (6)$$

$$x_t = \theta(B)z_t \quad (7)$$

where $\theta(B) = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)$.

The $ARIMA(p, d, q)$ model is a generalized version of the autoregressive integrated moving average process with p as the number of autoregressive terms, d as the number of differences, and q as the number of lagged forecast errors:

$$\begin{aligned} (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d x_t \\ = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)z_t \end{aligned} \quad (8)$$

This can be written in a concise form as:

$$\phi(B)(1 - B)^d x_t = \theta(B)z_t \quad (9)$$

$$\{z_t\} \sim WN(0, \sigma^2)$$

Readers who are interested in theoretical foundations of the ARIMA model could refer to the book “Time series analysis and its applications: with R examples” (Shumway and Stoffer 2006) for details.

3.1.2 ARIMA Model Optimization

Order selection and parameter estimation are two major steps for the ARIMA model forecasting. Selecting a proper order of the ARIMA model is important for producing accurate forecasting results. From a prediction point of view, it is not a wise choice to select p and q arbitrarily large. Large p or q values potentially lead to over-fitting issues. To avoid over-fitting problems, a penalty factor is introduced to discourage a model with too many parameters. Some widely used criteria for model selection are the Akaike’s information criterion (AIC), the corrected Akaike information criterion, (AICc) and the Bayesian information criterion (BIC). The preferred model is the one with the minimum value of one for these criteria. This study selects the order of an appropriate ARIMA model based on the Akaike’s information criterion (AIC). The best model should be the one that has the smallest value of AIC:

$$AIC = -2 \log(L) + 2m \quad (10)$$

where L is the likelihood of the data for the specific model and m is the number of parameters selected for this model.

This research utilizes the method proposed by Hyndman and Khandakar (2007) to select the orders of the appropriate ARIMA model automatically.

After determining the best orders of the ARIMA model, the parameters of the model are estimated through the maximum likelihood method. Detailed information on theoretical background and steps in fitting an ARIMA model can be found in (Tsay 2010).

3.1.3 Model Evaluation Criteria

There are several well established methods to evaluate model performance of the mean equation model. This research applies two measures of effectiveness to test the mean part: the root mean squared error (RMSE) and the mean absolute percentage error (MAPE). The RMSE is a frequently used measure of the difference between values predicted by a model and the actual observation. It is measured in the same unit as the original data. The MAPE is another commonly used measure of effectiveness. Different from the RMSE measures, the MAPE is expressed in percentage terms. Therefore, it provides us a general sense of the error even without knowledge of what constitutes a “big” error for the data set.

The equations for the RMSE and MAPE are as follows:

$$\text{RMSE} = \left(\sum_{i=1}^n \frac{(t(i) - a(i))^2}{n} \right)^{1/2} \quad (11)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{t(i) - a(i)}{t(i)} \right| \quad (12)$$

where $t(i)$ is the actual value, $a(i)$ is the forecast value, and n is the total number of time periods.

3.2 VOLATILITY MODELS

As mentioned in Equation (1), the structure of the proposed method is the sum of the mean and the variance. Most traditional models only concentrate on the mean part and assume that the variance part simply satisfies the white noise properties. Volatility models relax the assumption and characterize the changing variance of the data through time. Volatility models aim at specifying how the conditional variance r_t evolves over time. Different ways to address the variance part r_t lead to different kinds of volatility models. In general, there are two different categories: the GARCH-type model and the stochastic volatility model.

3.2.1 GARCH-type Models

The GARCH-type models aim at capturing the changes of the variance part r_t . The first volatility model was proposed by Engle (1982), termed as the Autoregressive Conditional Heteroskedasticity (ARCH) model. This model was originally used to capture the uncertainty of financial data. This type of uncertainty refers to the variances and covariance that change over time. In his research, the discrete time stochastic process r_t is expressed as Bollerslev *et al.* (1992):

$$r_t = \sigma_t \epsilon_t \quad (13)$$

$$\sigma_t^2 = \text{Var}(r_t | F_{t-1}) = \text{Var}(x_t | F_{t-1}) \quad (14)$$

Here ϵ_t is an independent and identically distributed (i.i.d) process with zero mean and standard deviation one and F_{t-1} denotes information available through $t - 1$. The above equation forms the foundation of the volatility model. Its various extensions are all based on this equation.

Different ways of modeling σ_t lead to a wide variety of volatility models. Engle suggested in his paper that σ_t^2 can be a linear function of past squared values of the r_t process:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i r_{t-i}^2 \quad (15)$$

Note that α_0 is the intercept term with $\alpha_0 > 0$, α_i represents the unknown coefficient of r_{t-i}^2 that satisfy $\alpha_i \geq 0$ to ensure the conditional variance as positive, and m denotes the number of lags selected for the model. This structure clearly captures the cluster of volatilities. The current magnitude of r_t is based on past values of r_{t-i}^2 . Therefore, a sudden large change in the data would more likely lead to another large change. In other words, the probability of obtaining a large variance is greater than that of obtaining a small variance if the past value of the variance is large. This structure also works in traffic forecasting. For example, congestions would create unexpected delays, which lead to dramatic increase in travel time. This phenomenon would

usually last for considerable time periods. Although the ARCH process has been proven useful in modeling the uncertainty in data, it often requires a relatively long lag. In order to allow both a longer memory and a more flexible lag structure, a generalized version of the ARCH model – Generalized Autoregressive Conditional Heteroskedasticity (GARCH) – was proposed by Bollerslev (1986), which included the past value of σ_t^2 in the model structure:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i r_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 \quad (16)$$

in which $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \dots, m$, $\beta_j \geq 0$, $j = 1, \dots, s$, $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_j) < 1$. The appealing feature of the GARCH model is that it takes into account that σ_t^2 not only depends on past values of the Variance part r_{t-1} , but also on its own past values σ_{t-j}^2 . The GARCH model can be interpreted as an ARMA process. Applications of the GARCH model in many different fields demonstrate its ability in modeling uncertainty in data. Lots of variations of the volatility model inherit this unique structure. One successful type of variant is termed as component GARCH model.

3.2.2 Component GARCH Models

Component GARCH models aim at capturing the trend and seasonal (cyclical) components in data. In the traffic prediction field, no research has yet been conducted to study the seasonality and trend in data through the component GARCH model. One innovation of this research is introducing the component GARCH model in travel time reliability prediction. Two different component GARCH models are proposed in this research: the component GARCH (C-GARCH) and the multiplicative component GARCH (MC-GARCH) models.

As trend and seasonal components are often observed in data, it was argued that the conventional GARCH model is unable to provide adequate performance (Andersen and Bollerslev 1997) if trend or seasonal components exist. Several variations of the GARCH model were developed to take into account the trend and seasonal volatility patterns. We term these models as component models that decompose the data as trend, seasonal, and random elements. The trend component represents long term changes in the level of the data series while the seasonal factor is the periodic fluctuations within the data series. Two different structures can be considered as basic component models:

The additive model, where

$$x_t = s_t + t_t + e_t \quad (17)$$

The multiplicative model, where

$$x_t = s_t \times t_t \times e_t \quad (18)$$

where s_t represents the seasonal effect, t_t represents the trend, and e_t represents the errors.

The additive model is applied to data series where the magnitude of the seasonal fluctuation does not change regardless of the level of the data series. The multiplicative model applies to situations in which the seasonal variation increases/decreases with the level of the series.

Based on the additive model structure, the component GARCH (C-GARCH) model proposed by Engle and Lee (1993) decomposes the conditional variance into a long term and a transitory component. The equation of the component GARCH model is as follows.

$$\sigma_t^2 = q_t + \sum_{i=1}^m \alpha_i (r_{t-i}^2 - q_{t-i}) + \sum_{j=1}^s \beta_j (\sigma_{t-j}^2 - q_{t-j}) \quad (19)$$

$$q_t = \alpha_0 + \rho q_{t-1} + \varphi (r_{t-1}^2 - \sigma_{t-1}^2) \quad (20)$$

where α_i , β_j , ρ , φ are unknown parameters, as is α_0 .

In this model, the intercept term q_t is regarded as a time-varying process, which represents the long term component of the conditional variance. The difference of the conditional variance and the long term component $\sigma_{t-j}^2 - q_{t-j}$ is the transitory component that models the short-term volatilities.

The multiplicative component GARCH model (Engle and Sokalska 2012) assumes the variations increase with the level of data and decomposes the variance part into three multiplicative components: daily component d_t , deterministic diurnal pattern s_i , and stochastic intraday component $q_{t,i}$.

$$x_{t,i} = u_{t,i} + r_{t,i} \quad (21)$$

$$r_{t,i} = \sqrt{d_t s_i q_{t,i}} \varepsilon_{t,i} \quad (22)$$

where the travel time at time index i in day t consists of the conditional mean $u_{t,i}$ and variance $r_{t,i}$ equations. $\varepsilon_{t,i}$ is the i.i.d (0,1) standardized innovation which can follow a normal, a student-t distribution, etc. The daily part d_t models the variance of the data across different days. It can be derived from a multifactor risk model, a daily GARCH model or a multiple indicators model (Engle and Sokalska 2012). The deterministic diurnal component for each time index is estimated as:

$$\hat{s}_i = \frac{1}{T} \sum_1^T \frac{r_{t,i}^2}{d_t} \quad (23)$$

where T is the total number of days, t denotes the day, i denotes the regular spaced time intervals.

As indicated in the above equation, the diurnal component at time index i is the average of variance at time index i scaled by its corresponding variance for each day. Therefore, the diurnal component represents the regular intraday variations. After estimating the daily and deterministic intraday components, the rest component in the variance part is regarded as stochastic, and can be regarded as a GARCH (p,q) process. The normalized residual is:

$$z_{t,i} = r_{t,i} / \sqrt{d_t s_i} = \sqrt{q_{t,i}} \varepsilon_{t,i} \quad (24)$$

where the stochastic intraday component $q_{t,i}$ is assumed to follow the GARCH process:

$$q_{t,i} = \alpha_0 + \sum_{j=1}^p \alpha_j z_{t,i-j}^2 + \sum_{j=1}^q \beta_j q_{t,i-j} \quad (25)$$

From the perspective of travel time prediction, travel time exhibits both regular cyclical patterns (seasonal component) and stochastic patterns. Daily cyclical patterns distinguish travel time as peak hour and non-peak hour traffic. Stochastic patterns are the results of unexpected influential events, like bad weather conditions and traffic incidents. Capturing the time-varying features of traffic behavior is critical for travel time forecasting. In addition, decomposing data into cyclical and stochastic patterns provides a better understanding of the underlying structure of the data. The component GARCH model, a generalization of the traditional GARCH model, considers the seasonal and trend components in the variance part through decomposition.

3.2.3 Stochastic Volatility Model

Most of the existing studies focus on the GARCH-type model. There are limited studies applying a stochastic volatility type model in traffic prediction. Part of the reason is that the estimating process for the stochastic volatility model is much more complex compared with the GARCH-type model since a new parameter estimation method is introduced to improve the estimating performance of the stochastic volatility model.

The conditional volatility of the GARCH model in Equation (13) is a deterministic function of past quantities. Provided that all relevant information is available, the model could be specified at the present time period. The stochastic volatility model is a competitive alternative to the GARCH-type model by modeling volatilities non-deterministically, which treats the volatility as a random process and evolves stochastically over time. Since traffic involves interactions among different factors such as demand, incident, and weather, this often makes future traffic conditions uncertain. Modeling the conditional variance as an unobserved stochastic process allows for a more flexible applications of the SV model and can account for the uncertainty inherent in traffic phenomena (Tsekeris and Stathopoulos 2009).

Based on the canonical model (Kim *et al.* 1998) of the stochastic volatility class, the volatility part r_t can be expressed as follows:

$$r_t = \exp(h_t/2)\varepsilon_t \quad (26)$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma v_t \quad (27)$$

$$h_1 \sim N\left(\mu, \frac{\sigma^2}{1 - \phi^2}\right) \quad (28)$$

where r_t represents volatility of travel time during time interval t . ε_t is a Gaussian white noise sequence with mean 0 and variance 1, and v_t is also a Gaussian white noise sequence with mean 0 and variance 1 which is independent of ε_t . The unobserved process h_t is interpreted as a stochastic volatility process with parameters μ , ϕ and σ to be estimated. To setup the model, a prior distribution for parameters μ , ϕ and σ should be specified. According to Kastner and Frühwirth-Schnatter (2013), the level μ follows a normal distribution with mean m_μ and variance M_μ , or $\mu \sim N(m_\mu, M_\mu)$. To guarantee the persistence parameter $\phi \in (-1, 1)$, $(\phi + 1)/2$ follows

Beta distribution with $(\phi + 1)/2 \sim B(a_0, b_0)$, where a_0 and b_0 are positive parameters; the volatility of log variance $\sigma^2 \sim B_\sigma \cdot \chi_1^2 = \text{gamma}(1/2, 1/2B_\sigma)$, where B_σ is a single positive value that stands for the scaling of the transformed parameter σ^2 , χ_1^2 denotes a chi-squared distribution with one degree of freedom. The posterior distributions of the desired variables are estimated through Bayesian inference via the Markov Chain Monte Carlo (MCMC) method.

It should be noticed that by taking logarithms of the squared r_t in Equation (26), the SV model is transformed into a linear equation:

$$\log(r_t^2) = h_t + \log(\varepsilon_t^2) \quad (29)$$

$$\varepsilon_t \sim N(0, 1)$$

If $\log(\varepsilon_t^2)$ is approximated by a mixture of a normal distribution with m_{r_t} and $s_{r_t}^2$ being the mean and the variance of the r_t th mixture component respectively (Kastner and Frühwirth-Schnatter 2013), the above equation reduces to the form of a conditionally Gaussian state space model:

$$\hat{r}_t = m_{r_t} + h_t + \varepsilon \quad (30)$$

where $\hat{r}_t = \log(r_t^2)$, and $\varepsilon \sim N(0, s_{r_t}^2)$. This linearization makes efficient MCMC sampling possible.

Model reparameterization could potentially improve simulation efficiency in the volatility model. Denoting Equation (26) to (28) as the centered parameterization models, another version of the SV model is non-centered (NC) parameterization, where parameter μ is shifted from the state Equation (27) to the observation Equation (26) by setting $\hat{h}_t = (h_t - \mu)/\sigma$. The non-centered (NC) parameterization form is given as:

$$r_t \sim N(0, \exp(\mu + \sigma \hat{h}_t)) \quad (31)$$

$$\hat{h}_t = \phi \hat{h}_{t-1} + v_t \quad (32)$$

where $\hat{h}_t = (h_t - \mu)/\sigma$, v_t is a Gaussian white noise sequence with mean 0 and variance

1.

The choice of centered parameterization (C) or non-centered parameterization (NC) would dramatically affect the sampling efficiency. Both parameterizations have their advantages and disadvantages, which heavily depends on the true parameter values of the data generating process. There exists no ‘ultimate’ parameterization. Based on the ancillarity-sufficiency interweaving strategy (ASIS) introduced by Yu and Meng (2011), Kastner and Frühwirth-Schnatter (2013) proposed a strategy by interweaving C and NC to overcome this deficiency. Their study results show that interweaving C and NC leads to a robustly efficient sampler that always outperform either parameterization (C or NC) with respect to parameter costs in terms of design and computation. Their intuitive and efficient algorithm can be briefly summarized into six steps:

Choose appropriate starting values and repeat the following steps:

- (1) Draw h from parameterization C.
- (2) Draw μ , ϕ , σ from parameterization C.
- (3) Move to non-centered parameterization NC by the simple deterministic

transformation $\hat{h}_t = \frac{h_t - \mu}{\sigma}$.

- (4) Redraw μ, ϕ, σ from parameterization NC.
- (5) Move back to C by calculating $h_t = \mu + \sigma \hat{h}_t$ for all t.
- (6) Draw the indicators from parameterization C.

The detailed sampling steps in the ASIS involve extensive Bayesian approach with MCMC. In the following, a brief summary of the key ideas on how to estimate the parameters of the SV model is introduced. Readers are referred to Gelman *et al.* (2003) for further details. The following briefly explains the concept of Bayesian approach with MCMC.

Consider a problem that given a set of data, the posterior distribution of $P(\theta|x)$ is desired in Bayesian analysis in order to estimate parameter θ . Bayesian analysis seeks to estimate parameter θ by combining the prior knowledge about the parameters with the data. Denote $P(\theta)$ as the specified prior distribution of θ . The posterior distribution $P(\theta|x)$ is calculated through the conditional probability:

$$P(\theta|x) = \frac{P(\theta, x)}{P(x)} = \frac{P(x|\theta)P(\theta)}{P(x)} \quad (33)$$

where $P(x|\theta)$ denotes the likelihood function of the data for a given model, $P(x)$ denotes the marginal distribution of x . According to Bayes's rule:

$$P(\theta|x) \propto P(x|\theta)P(\theta) \quad (34)$$

The goal is to estimate θ through the posterior distribution $P(\theta|x)$. The Markov Chain Monte Carlo (MCMC) approach draws a sample from the posterior distribution and then calculates the estimator of θ . The Markov process is a stochastic process $\{x_t\}$ that the value of x_h does not depend on the value of x_s if the value of x_t is given and $s < t < h$. $\{x_t\}$ is a Markov process if its conditional distribution function satisfies the following criterion:

$$P(x_h|x_s, s \leq t) = P(x_h|x_t), h > t. \quad (35)$$

The basic idea of MCMC method is to simulate a Markov chain that has the desired probability distribution which is $P(\theta|x)$. The ASIS algorithm estimates the model parameters in steps two and four through Bayesian inference using MCMC.

3.2.4 Prediction Interval Estimation

Prediction intervals for the volatility model are estimated based on the idea of prediction intervals for regression models. To construct a prediction interval with $100(1 - \alpha)\%$ confidence, we assume that the error follows Gaussian distribution with zero mean and variance σ_t^2 . The prediction interval can be calculated as:

$$(u_t - z_{\alpha/2}\sigma_t, u_t + z_{\alpha/2}\sigma_t) \quad (36)$$

where u_t is the predicted mean, $z_{\alpha/2}$ denotes the standard score corresponding to the cumulative probability level of $\alpha/2$, and σ_t is the prediction variance from a volatility model.

As the concept of uncertainty or reliability is a relatively new area in traffic forecasting, there are few studies that provide criteria for PI assessment. One study by Khosravi *et al.* (2011)

suggested that two important aspects of PIs assessment should be considered: coverage probability and length. Coverage probability measures the percentage of the targets that lie within the predicted PIs. It measures how effective the constructed prediction intervals are. The mathematical representation of PI coverage probability (PICP) is as follows:

$$\text{PICP} = \frac{1}{n} \sum_{i=1}^n c_i \quad (37)$$

where $c_i = 1$ if $y_i \in [L(x_i), U(x_i)]$; otherwise, $c_i = 0$. $L(x_i)$ and $U(x_i)$ represent upper and lower bounds of the prediction interval of x_i , n is the total number of constructed PIs.

On the other hand, another criterion called mean PI length (MPIL) measures the average length of the PIs. It measures how efficient the constructed prediction intervals are. Assume we have two models that provide PIs with the same coverage probability; the one that gives a narrower prediction band is more efficient. The following equation gives the definition of the MPIL:

$$\text{MPIL} = \frac{1}{n} \sum_{i=1}^n (U(x_i) - L(x_i)) \quad (38)$$

Therefore, both criteria should be considered when evaluating the volatility models.

3.3 APPLICATION OF COMPONENT GARCH MODELS IN TRAVEL TIME PREDICTION

The performance of the GARCH, the C-GARCH, and the MC-GARCH models are investigated here by using data collected from Automatic Vehicle Identification (AVI) stations located along U.S. Highway 290 (or U.S. 290) in Houston, Texas. The entire study corridor is about five miles long and covers the Northwest Freeway in the westbound direction between I-610 and the junction of Farm to Market Road 1960 (FM1960). The IDs of the selected AVI stations are 29, 30, 31, 32, 33, and 34. The travel time between each pair of consecutive detectors were collected and aggregated into five-minute time intervals. Individual segment travel time at free-flow conditions is less than four minutes. The total period of the sample was the entire year of 2008 with missing data replaced by annual medians of the missing intervals. Since travel time patterns during weekdays and weekend are quite different (congestions are more likely to occur during weekdays), weekend data were excluded from the sample. As a result, 262 weekdays of travel time data that contain 75,456 five-minute observations are selected for this study.

3.3.1 Modeling Conditional Mean

Traffic data often shows periodic patterns. Travel time increases and varies significantly during peak hours compared with travel time during non-peak hours. It is difficult to precisely predict traffic when congestion occurs. Point prediction methods are often unable to capture traffic variation in congested situations, therefore providing less reliable or accurate prediction. As the performance of the point prediction methods often decreases when congestion occurs, it is expected that the residual series (after removing the predicted mean by ARIMA model) show higher variations during peak hours.

Figure 2 provides a boxplot of absolute deviation from the predicted mean for each 20-minute time interval (outliers have been removed from this plot). Each box statistically represents interval-to-interval and day-to-day variations of the residual series. The green line indicates the mean of the residual and the lower and upper boundary of each box is the 25th and 75th percentiles of the data for corresponding time intervals. As observed from this plot, statistics of each interval are different from each other. Both the mean and the percentiles of the data are different at different time intervals. This indicates that the residual series vary over time, and the constant variance assumption of the traditional time series models is violated. This further proves that a volatility model, which relaxes the constant variance assumption, is necessary. In addition, there is a pronounced increased variation at the beginning of 15:00 hour; subsequently, the variation decreases at 19:00 hour. Comparatively, variations during other time periods (non-peak hours) are less significant.

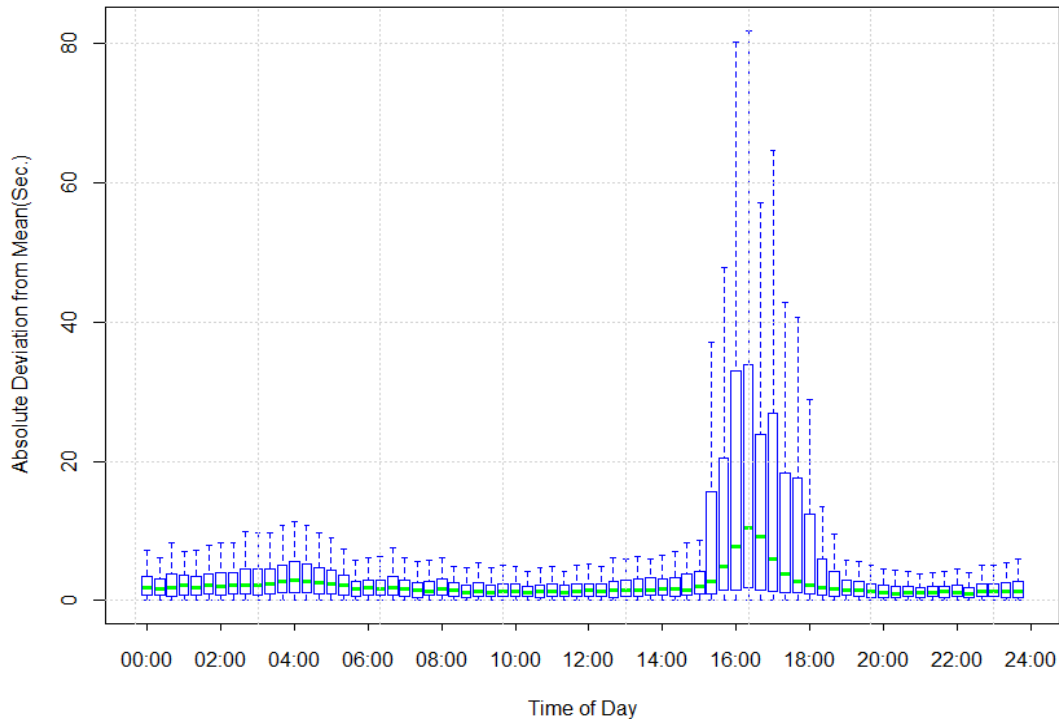


Figure 2: Box Plot of Absolute Deviation from Predicted Mean.

The other four studied segments also depicted similar diurnal patterns. It is clear that seasonal components exist in the residual time series. In addition, the mean of the absolute deviations during non-peak hours are close to zero, which means that the ARIMA model provides adequate prediction during non-peak hours. On the other hand, the prediction performance of the ARIMA model decreases during peak hours, as both the mean and the 75 percent statistics increase.

3.3.2 Testing the ARCH Effect

The basic assumption of the GARCH-type model is that square values of the residuals are correlated. Therefore, before applying different GARCH-type models to the data, there is a need to test if the data meet this assumption. Two tests are available: Ljung–Box statistics and Lagrange multiplier test (Tsay 2010). This study chose the Ljung-Box statistics to test if the first lags of the squared residuals are uncorrelated. The Ljung-Box test is as follows:

$$H_0: \rho_1 = \rho_2 = \dots \rho_m = 0$$
$$Q(m) = N(N + 2) \sum_{h=1}^m \frac{\rho_h^2}{N-h} \quad (39)$$

where N is the number of data points under study, ρ_h is the sample autocorrelation at lag h , and m is the number of lags being tested.

The critical region for rejecting the null hypothesis at significance level α is:

$$Q > \chi_{1-\alpha, m}^2$$

In terms of p values, the null hypothesis will be rejected if the p value is less than α . In our study, the Ljung-Box test is applied to the residual data of all five segments. P values of the test for all studied segments are significantly less than 0.01. Therefore, the null hypothesis is rejected at the significance level of 0.01. That is to say, correlations exist between squared values of residuals. The GARCH-type models are necessary.

3.3.3 Estimating the Volatility Model

Similar to the ARIMA type model, estimating the volatility model also involves order selection and parameter estimation. Several studies indicate that GARCH family model with order of (1, 1) was found adequate in representing the volatility dynamics. Therefore, GARCH, C-GARCH and MC-GARCH with order of (1, 1) were adopted for ease of implementation and comparison by using the R package ‘rugarch’ (Ghalanos 2013).

Since the multiplicative component GARCH model decomposes the data into a daily component, a deterministic diurnal pattern and a stochastic intraday component, the first step is to model the daily component. This study terms the average volatility for each day as the daily component. The daily component is estimated through the standard GARCH (1, 1) process. There are 262 daily data in total; the first 242 data points are used as the training data. After removing the daily component, the deterministic diurnal part is estimated as the annual average of the residual data at each time interval. The normalized residuals (22) are then used to produce the stochastic intraday component. The volatility components estimated by the MC-GARCH model are displayed in five panels as shown in Figure 3. The top panel shows the observed values of the residual data series. The second panel gives the estimated conditional variance, being the product of the following three components: deterministic intraday (panel three), daily (panel four), and stochastic intraday (panel five) components. As indicated in this figure, the MC-GARCH model is able to model the trend,

seasonal, and stochastic components of the data. This feature provides better understanding of the basic structure of the data and is easy to interpret. For example, the intraday [Deterministic] components indicate the regular cyclical patterns of travel time volatility, the intraday [Stochastic] components specify daily variations due to demand variation, incident or other abnormal traffic phenomenon.

3.3.4 Construct the Mean and Prediction Intervals

The final output of the proposed volatility models includes two measures: the predicted mean and the predicted PIs. The predicted mean part generally tells the expected value of travel time in the future, whereas the PIs tell how likely the observed value will lie within a certain range. In other words, wider PIs often indicate unreliable travel time and prediction. Thus, based on the combined information of predicted mean and PIs, travelers and operators would have a better sense of future traffic conditions. In this study, the ARIMA model provides mean values, and the prediction intervals are constructed according to Equation (36). Figure 4 plots some sample prediction results of the MC-GARCH model. The blue dot stands for observed travel-time data obtained at corresponding time intervals, and the red triangle stands for the predicted mean. The green lines represent the PIs constructed by the MC-GARCH model. It is obvious in this figure that there is always a mismatch between the predicted mean and the observed value. This partly results from the dynamic nature of traffic: travel time varies from time to time. Prediction intervals, on the other hand, are able to adequately capture this variation by covering most of the observed values. Therefore, this model provides an effective and efficient way to measure uncertainty associated with future travel time.

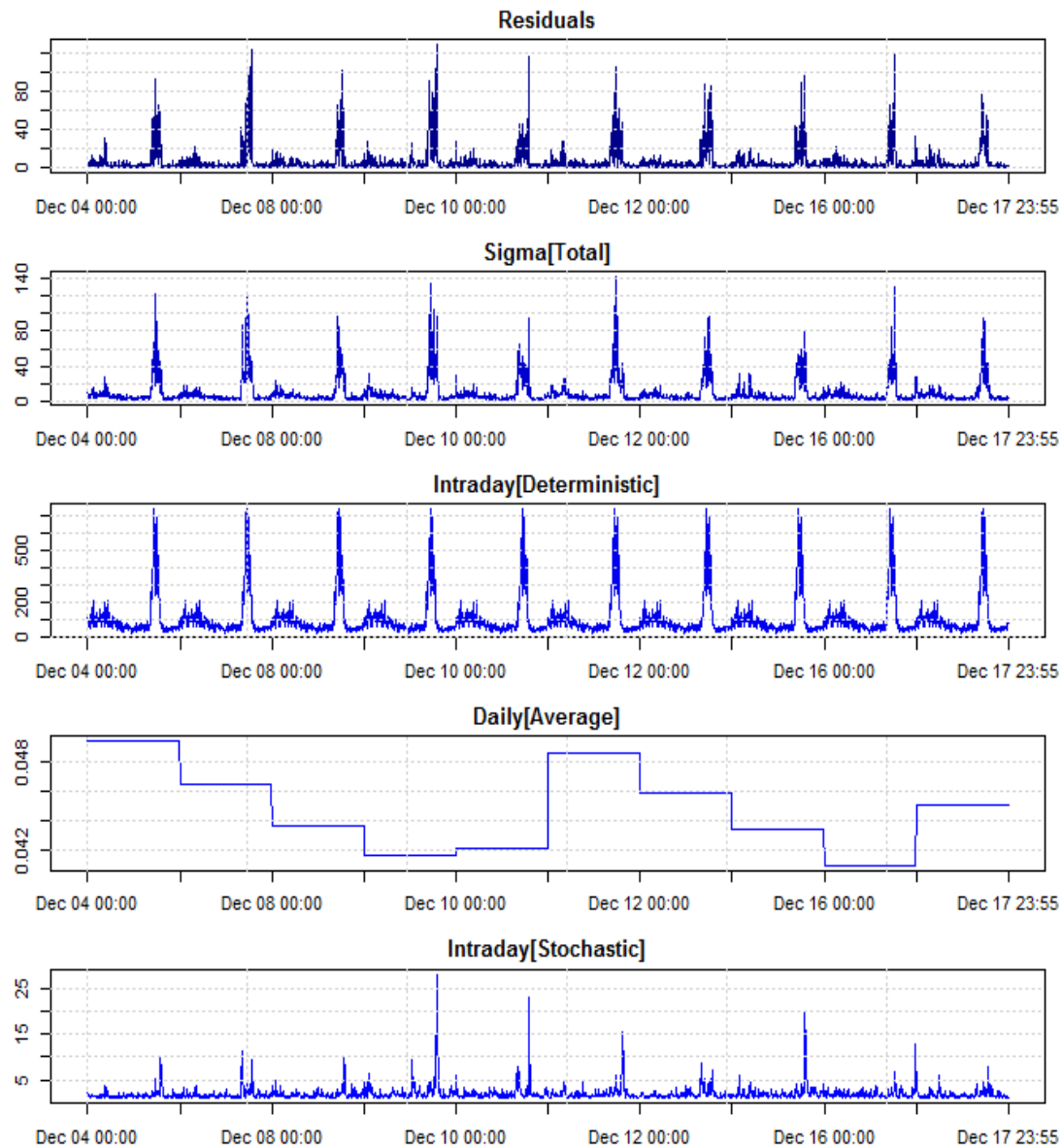


Figure 3: Multiplicative component GARCH forecasting results: decomposition of the volatility into its various components (32-33).

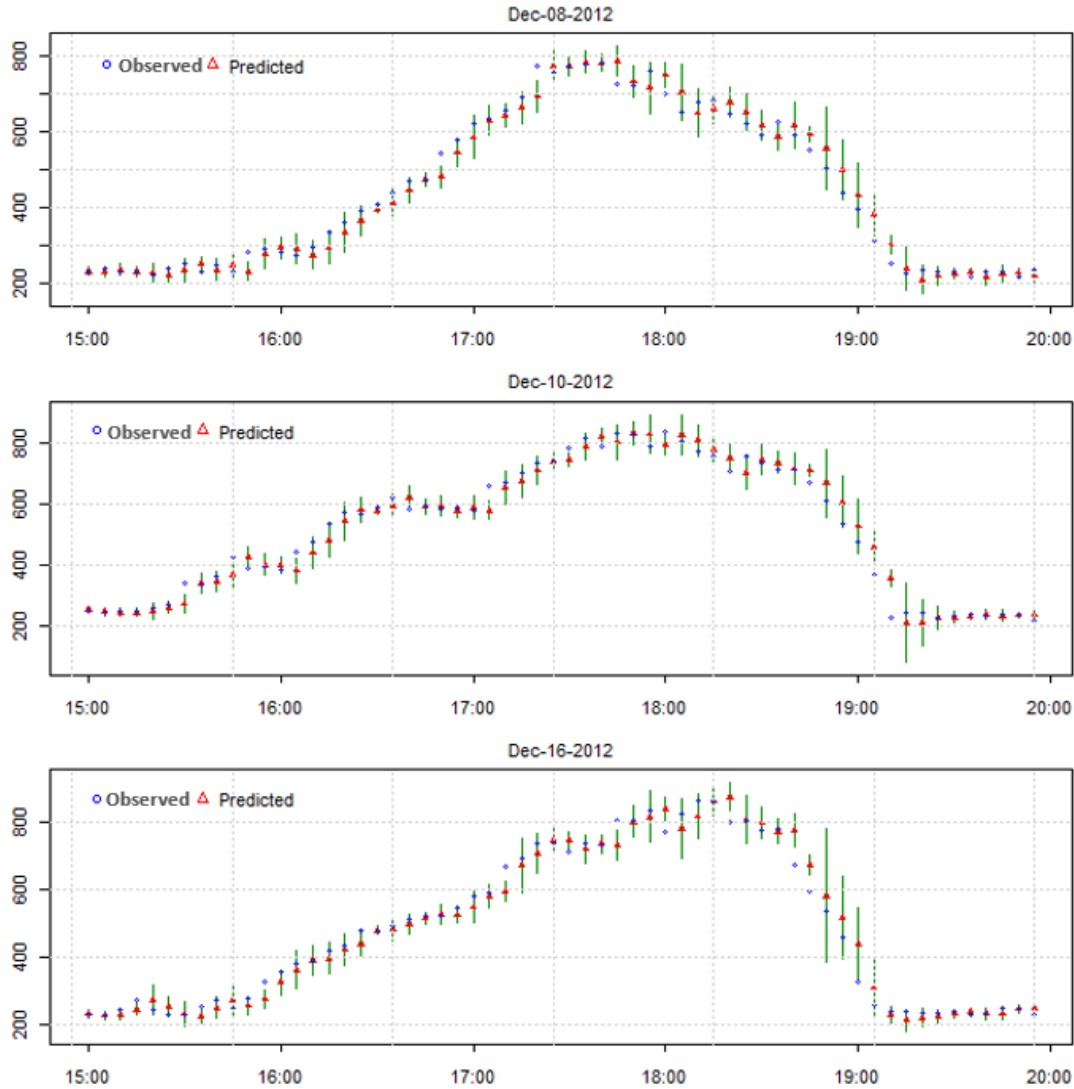


Figure 4: Predicted mean and PI for multiplicative component GARCH model.

3.3.5 Results and Discussion

PIs with different confidence levels are constructed through the GARCH, the C-GARCH and the MC-GARCH models for the five studied segments. The effectiveness and efficiency of the constructed PIs is evaluated based on the criteria of coverage probability and PI length. For each of the five studied segments, thirty days of five-minute travel-time data with 8,640 observations are used as the training data set and ten days travel time data with 2880 observations are used as the comparison (testing) dataset. We estimate individual models for each of the five segments. PIs are constructed with 95 percent, 90 percent, and 85 percent confidence levels, respectively.

Table 1 provides average MPIL and PICP values of the three models with different confidence levels during peak hours, nonpeak hours, and all day. During peak hours, the prediction interval coverage rates of the MC-GARCH are the highest compared to the C-GARCH and the GARCH model. For the 95 percent confidence level, the coverage rate of the MC-GARCH model (90.86

percent) is 2.86 percent higher than the second highest model (GARCH 88.00 percent). For the 90 percent confidence level, the coverage rate of the MC-GARCH model (87.35 percent) is 4.33 percent higher than the second highest model (GARCH 83.02 percent). For the 85 percent confidence level, the coverage rate of the MC-GARCH model (83.96 percent) is 5.18 percent higher than the second highest model (GARCH 78.78 percent). As the confidence level decreases, the advantage of the MC-GARCH model becomes obvious, proving that the MC-GARCH model is able to capture the volatility of traffic data during peak hours. In terms of mean prediction interval length, the C-GARCH is the smallest. However, on average, the C-GARCH model only reduced the length by 0.75 compared with the GARCH model.

During nonpeak hours, the MC-GARCH model also provides the highest coverage. But the advantage of MC-GARCH model is not obvious compared with the GARCH and the C-GARCH models in terms of either MPIL or PICP during non-peak hours. This is expected as travel time is relatively stable with small variations and the trend and seasonal patterns are not obvious during this period. Therefore, performance of these three models should be similar during non-peak hours.

Investigating the all-day performance of these three models indicates that the MC-GARCH model provides the highest PICP value, whereas both the C-GARCH and the GARCH models give lower MPIL values compared with the MC-GARCH model.

Table 1: Estimated MPIL and PICP values for GARCH, C-GARCH and MC-GARCH models.

Confidence Level	Model	Peak Hours		Nonpeak Hours		All Day	
		MPIL	PICP	MPIL	PICP	MPIL	PICP
95%	GARCH	115.98	88.00%	44.53	95.72%	56.69	94.41%
	C-GARCH	115.1	87.80%	45.13	95.10%	57.04	93.86%
	MC-GARCH	139.74	90.86%	45.09	95.87%	61.2	95.01%
90%	GARCH	97.04	83.02%	37.26	92.30%	47.43	90.72%
	C-GARCH	96.31	82.57%	37.77	91.42%	47.73	89.92%
	MC-GARCH	116.93	87.35%	37.73	93.08%	51.21	92.10%
85%	GARCH	85.21	78.61%	32.72	89.07%	41.65	87.29%
	C-GARCH	84.56	78.78%	33.16	87.79%	41.91	86.26%
	MC-GARCH	102.67	83.96%	33.13	90.47%	44.96	89.36%

In general, based on the estimation results, we can conclude that the MC-GARCH model tends to cover more targets compared with the C-GARCH and the GARCH model, especially during peak hours. The C-GARCH and GARCH models give a lower prediction band compared with MC-GARCH with the compromise of lower coverage rate. Since coverage rate of the C-GARCH and the GARCH model are much lower than corresponding confidence level during peak hours (for example, PIs of both the GARCH and the C-GARCH model cover around 78 percent of the targets for 85 percent confidence level), the MC-GARCH model generates more effective PIs, although a little bit wider than others.

To check the consistency of each model’s performance, Figure 5 and Figure 6 compare PICP values of the GARCH, the C-GARCH, and the MC-GARCH models at different confidence levels for individual segments. As shown in these figures, the orange, blue and green columns represent the GARCH, the C-GARCH and the MC-GARCH models; columns with different patterns represent different confidence levels. During peak hours (Figure 5), the MC-GARCH model generates PIs with the highest coverage for all segments. The advantage of the MC-GARCH model is significant with the highest difference of 6.94 percent compared with PIs (at the 85 percent confidence level) provided by the C-GARCH model for segment 33-34. On the other hand, the PICP values of the GARCH and the C-GARCH models are similar. The largest difference of PICP values between the GARCH and the C-GARCH is 1.63 percent. During non-peak hours (Figure 6), all three models provide high coverage rate at corresponding confidence levels. Differences among individual models during non-peak hours are not as significant as during peak hours. Comparing all three models’ performance between peak and non-peak hours suggests that peak hour coverage is relatively low, as traffic variations increase during peak hours. In addition, PI lengths during peak hours are also longer than PI lengths during non-peak hours. This is because the uncertainties during peak hours are more evident compared with non-peak hours.

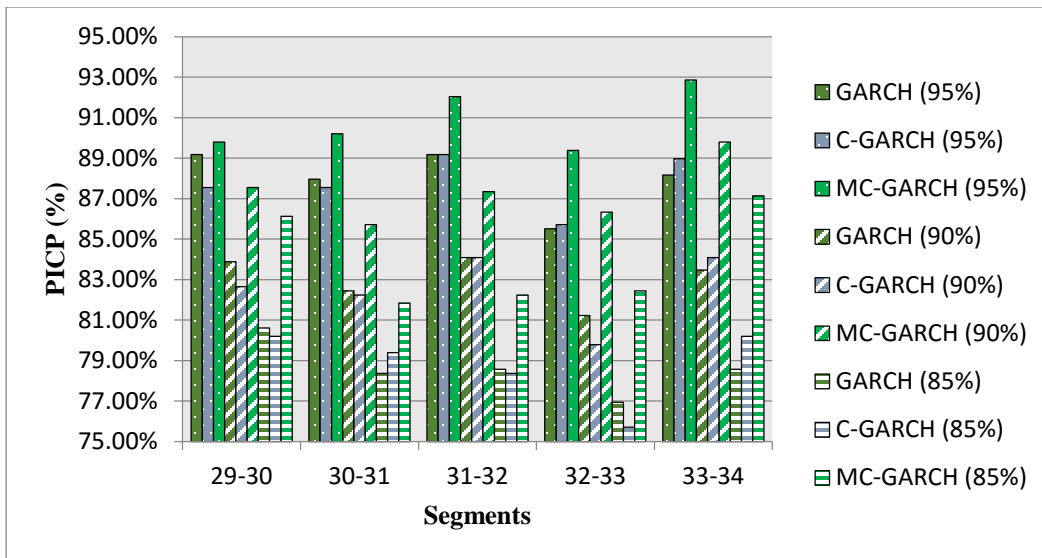


Figure 5: Comparing performance of GARCH, C-GARCH and MC-GARCH models during peak hours.

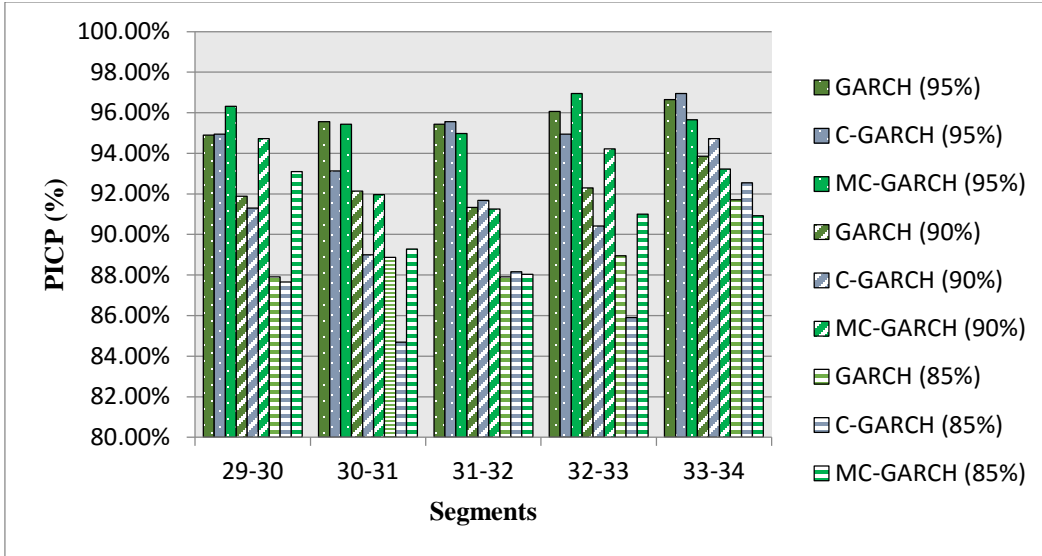


Figure 6: Comparing performance of GARCH, C-GARCH and MC-GARCH during non-peak hours.

Figure 7 provides an intuitive comparison of one-day peak hour PIs at the 95 percent confidence level constructed by the MC-GARCH, the C-GARCH and the GARCH models. The green dash lines stand for PIs of the MC-GARCH model, the yellow dash lines stand for PIs of the C-GARCH model, and the pink dash lines stand for PIs of the GARCH model. As shown in this figure, the PIs constructed by the C-GARCH and the GARCH almost overlap. It has also been depicted in Table 1 that there is no significant difference between MPIL and PICP values of the C-GARCH and the GARCH models. It seems the effect of the long-term component in the C-GARCH model is limited in this case. On the other hand, PIs of the MC-GARCH model are different from both the C-GARCH and the GARCH models. The MC-GARCH model tends to cover more targets by increasing the width of the prediction intervals at certain time intervals (Points identified by blue arrows). Overall, the C-GARCH and the GARCH models create PIs similar to each other. Compared with the C-GARCH and the GARCH models, the MC-GARCH model tends to cover more targets by increasing the length of its PIs during certain time intervals.

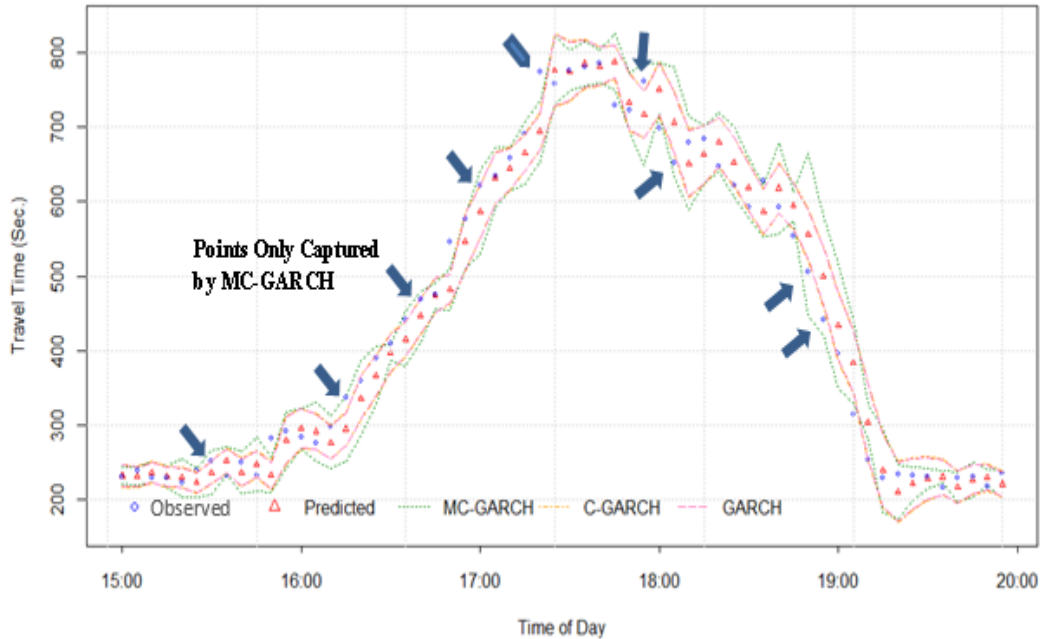


Figure 7: Comparison of prediction intervals constructed by GARCH, C-GARCH and MC-GARCH models.

3.3.6 Summary

As uncertainty associated with travel time prediction becomes an important topic for implementing an intelligent transportation system, statistical volatility models provide a promising way to generate more accurate PIs that account for variability in travel time prediction. The traditional GARCH model is argued to be inadequate when modeling data that show pronounced seasonal patterns. This study developed the C-GARCH and the MC-GARCH models in travel time prediction. To empirically evaluate performances of the proposed models, this study tested the GARCH, the C-GARCH and the MC-GARCH models by using freeway travel time data collected from Automatic Vehicle Identification (AVI) stations located along U.S. Highway 290 (or U.S. 290) in Houston, Texas. The forecasting results of the proposed models are attractive, especially during peak hours. The findings of this study include:

The proposed MC-GARCH model outperforms the GARCH and the C-GARCH models during peak hour prediction. A case study of the five selected segments highlighted the strength of the MC-GARCH model in providing more effective PIs in terms of coverage rate. The idea of decomposing travel time volatility is promising when data show cyclic patterns. By decomposing travel time volatility into daily, diurnal and stochastic components, the MC-GARCH model is able to capture uniqueness of each component and captures the seasonal effect of data.

The C-GARCH model treats travel time as a long term and transitory components. It works best if there is a trend. Based on the case study, the performance of PIs constructed by the C-GARCH model and the GARCH model are similar to each other. The effect of the long term volatility component in the C-GARCH model is not significant in this case.

During non-peak hours, there is no obvious advantage of all three models in terms of MPIL and PICP. This is partly due to the fact that travel time during nonpeak hours is relatively stable, with small variations around the mean. The trend and seasonal patterns are not obvious during this period.

Component GARCH models decompose travel time data into long term, short term and cyclical components. If there are cyclical components in the data, the MC-GARCH model has the potential to better capture uncertainties associated with travel time. In addition, the MC-GARCH model decomposes traffic volatility into several different components that can be easily interpreted and estimated. In this study, the daily component and the normalized residuals are modeled as a simple GARCH model. Besides the GARCH model, the daily component can also be estimated through multifactor risk model as suggested by Engle and Sokalska (2012). It is also worth trying different variations of GARCH models to estimate the normalized residuals. In addition, this study treats the intraday component as an average term. Further study could also explore different ways in defining the intraday component.

3.4 APPLICATION OF STOCHASTIC VOLATILITY MODEL

The stochastic volatility based method is investigated here by using travel time data collected with Bluetooth sensors along an 18-mile long corridor in Connecticut. The Bluetooth sensors (Figure 8) were temporarily installed by the University of Maryland team at Interstate 95 (I-95) to collect travel time information between October 19, 2012 and October 28, 2012. Bluetooth technique enables digital devices interconnect with each other using short-range wireless communications. Many mobile phones, car radios or other personal devices come equipped with Bluetooth wireless capability to communicate with other Bluetooth-enabled devices anywhere from 1 m to about 100 m (300 ft). In the context of travel time collection, the Bluetooth detector captures the electronic identifier, or tag, called Machine Access Control (MAC) address, in each Bluetooth enabled device and places a timestamp when the vehicle enters the detection range of the sensor. As the same vehicle passes subsequent detectors, the detected MAC can be matched allowing the calculation of travel time between these two locations. The Bluetooth detectors require at least two detectors to obtain travel time information. The Bluetooth detector has the advantage of providing more accurate traffic data with relatively low-cost installation (2010).

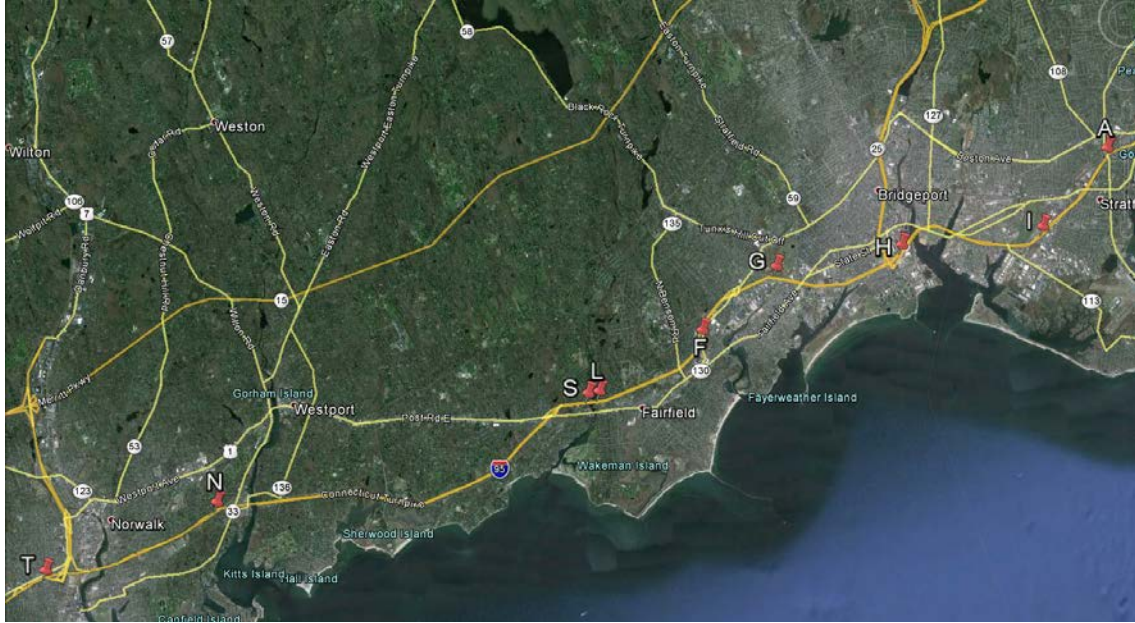


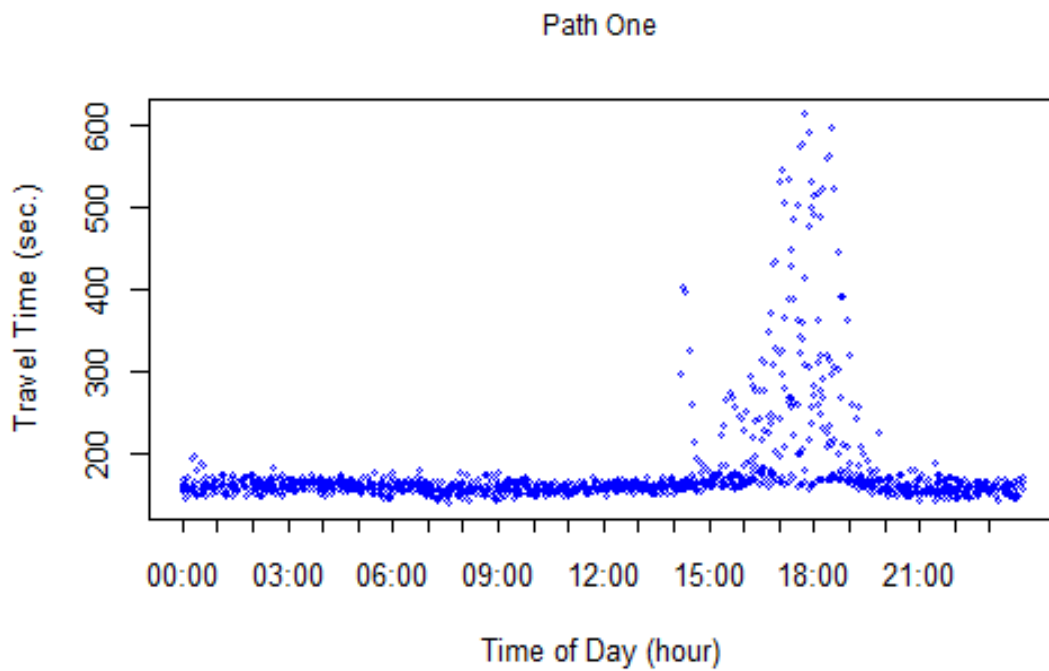
Figure 8: Bluetooth sensor location of the study.

Raw data from Bluetooth sensors are MAC IDs of the detected Bluetooth devices along with their detection time stored in a removable memory card. Sample travel time for a particular freeway stretch is obtained by matching the MAC ID between two Bluetooth sensors located at the endpoint of the freeway stretch. Figure 8 depicts the location of the selected Bluetooth sensors named as “T, N, S, L, F, G, H, I, A”. This study selects six segments as indicated in Table 2. Each path is comprised of a head and a rear sensor. Individual vehicle travel time for each path is the time it takes for the same vehicle to be detected by both the head and rear sensors. Raw data are filtered through a four-step offline filtering algorithm proposed by Haghani *et al.* (2010) to extract every five-minute ground truth travel time. Because traffic patterns during weekdays are significantly different from weekends, this study only focused on weekday travel time patterns. We separate travel time data for each pair of detectors under study as training dataset and testing dataset: the training dataset is every five minutes travel time from October 22, 2012 to October 25, 2012 with 1152 observations; the testing dataset is data obtained on October 26, 2012 with 288 observations.

Figure plots one week (excludes weekend) aggregated travel time over every five-minute time interval on four segments. The scatter plots (path one, two, five, six) illustrate considerable variations in travel time at each time interval over different days, especially during peak hours. Rush-hour traffic normally occurs between 3 pm and 8 pm on segments one and two, and between 6 am and 10 am on segments five and six. Travel time variations during non-peak and peak hours show different patterns. Variations of travel time during non-peak hours are much smaller than that during peak hours. The considerable variations of travel time across different times of the day can be attributed to several factors: demand variations over different times and days, variations of driver behavior under different kind of weather conditions, and incidents that disrupt normal traffic. These exogenous factors are often unpredictable, which makes travel time prediction a complex problem. Therefore, it is critical to treat traffic phenomena as a stochastic process.

Table 2: Selected segments for this study.

Path ID	Head Sensor	Rear Sensor	Distance (mile)		Starting at	Ending at
			Standard	Measured		
1	T	N	2.7	2.69	Fairfield Ave/Exit 14	CT-33/CT-136/Exit 17
2	N	S	5.8	5.94	CT-33/CT-136/Exit 17	Bronson Rd/Exit 20
3	A	I	1.5	1.51	Broad St/Exit 32	Surf Ave/Exit 30
4	I	H	2.4	2.36	Surf Ave/Exit 30	CT-25/CT-8/Exit 27
5	H	G	1.9	1.97	CT-25/CT-8/Exit 27	Fairfield Ave/State St/Exit 25
6	F	L	1.7	1.82	US-1/Exit 23	Bronson Rd/Exit 20



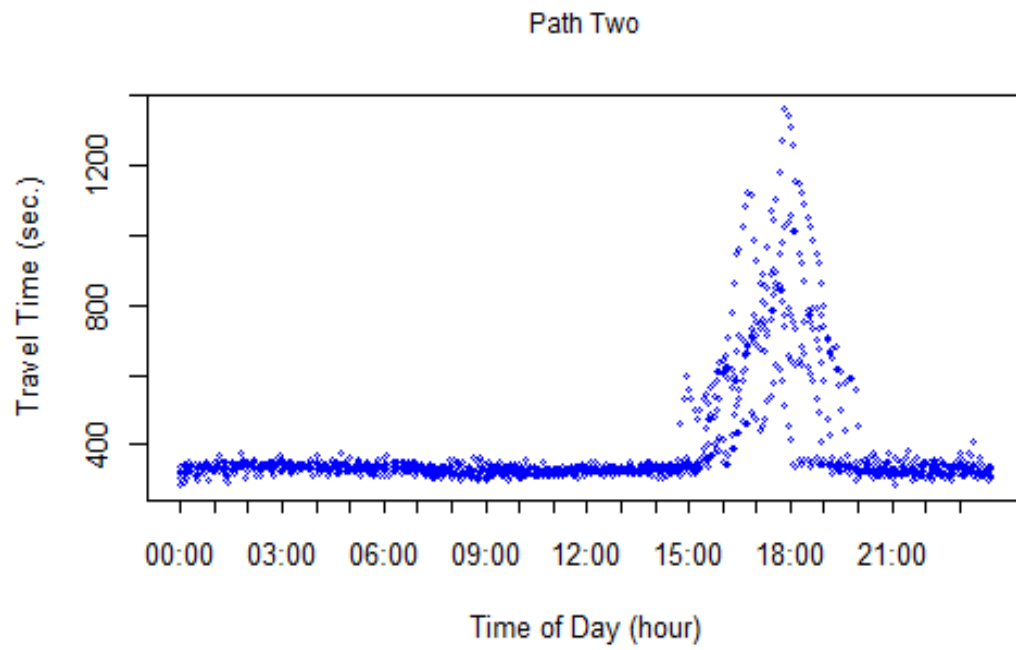
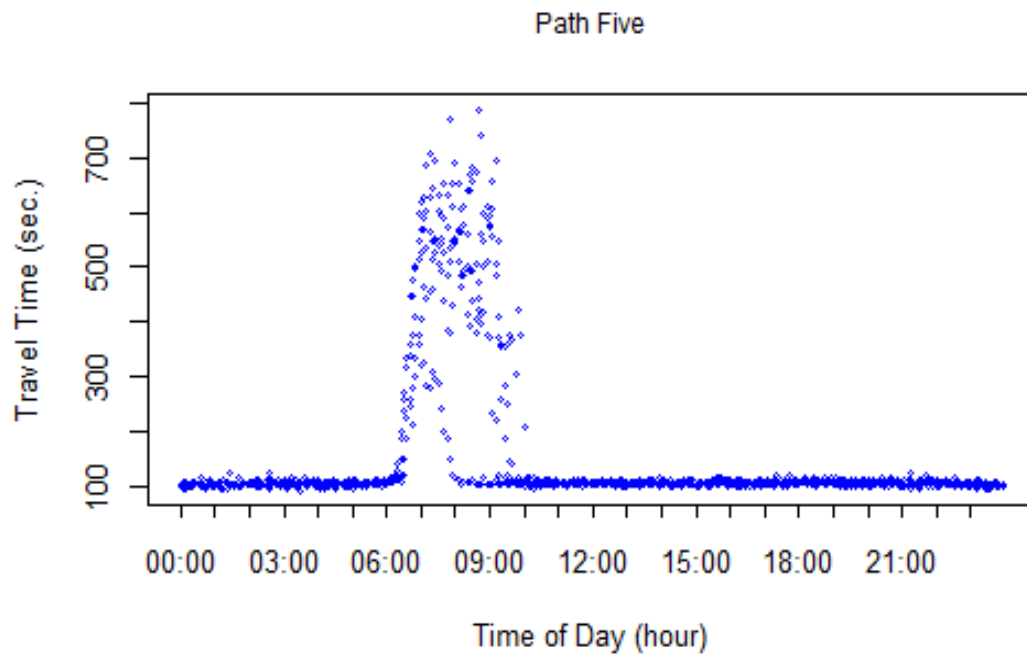


Figure 9: A scatter plot of travel times on four paths.



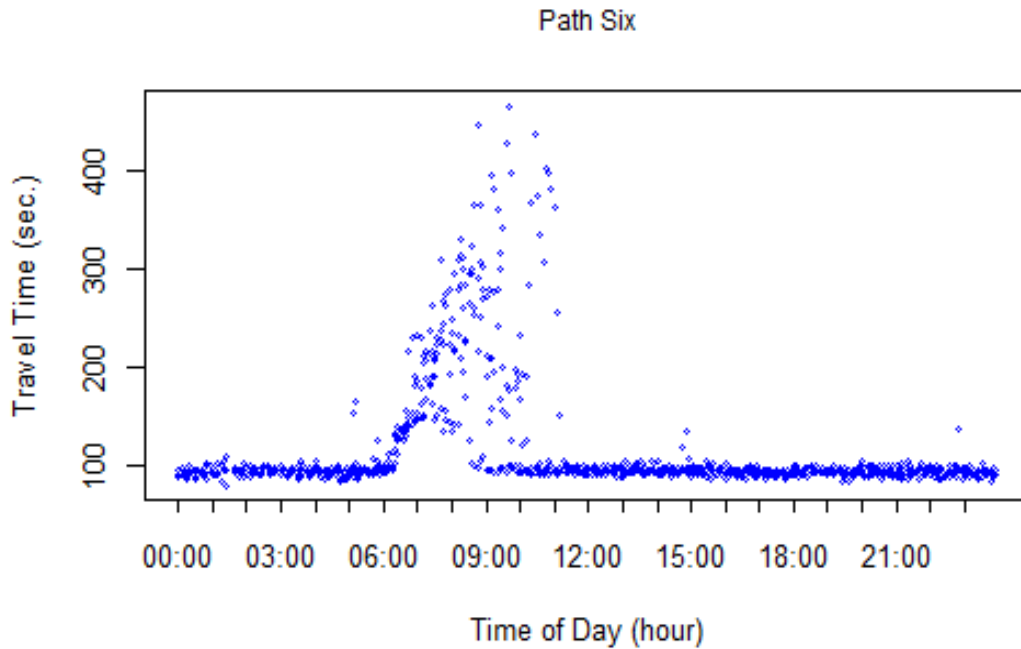


Figure 9 (Cont'd): A scatter plot of travel times on four paths.

3.4.1 Model Fitting

Both the GARCH and the SV models are tested to fit the second part of Equation 1. As the process to estimate the GARCH model has been already described in our previous study (2013), the following will focus on the estimation of the SV model.

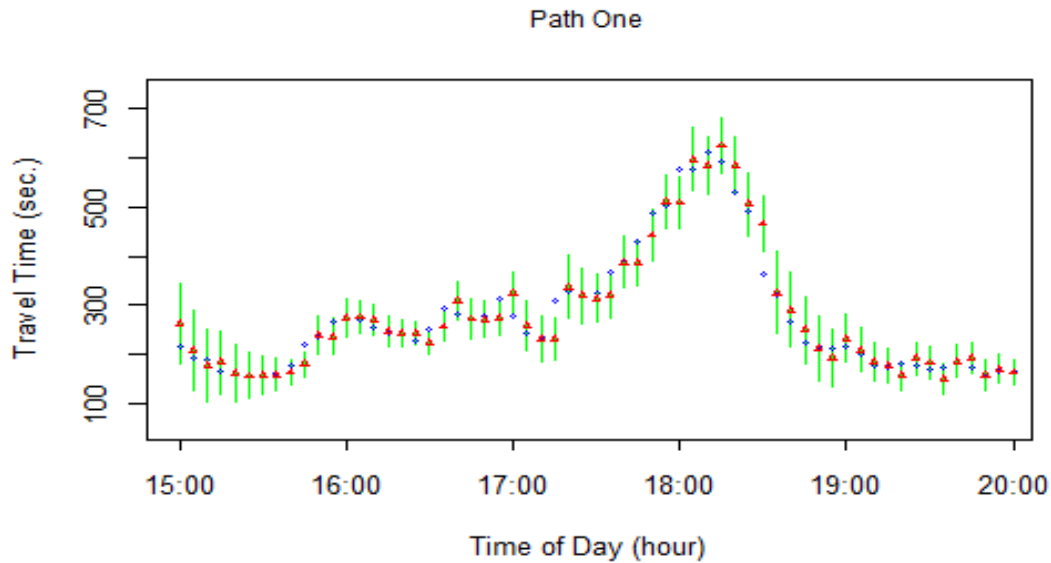
To perform Bayesian inference in the SV model, there is a need to specify parameters of the prior distribution of μ , ϕ and σ . As it already mentioned that μ follows Gaussian distribution, ϕ follows Beta distribution and σ^2 follows $B_\sigma \cdot \chi_1^2$, we need to specify five parameters: mean m_μ and standard deviation M_μ of the normal distribution for μ ; a_0 and b_0 of the beta distribution for $(\phi + 1)/2$; and B_σ for the scaling of the transformed parameter σ^2 . After specifying parameters of the prior distributions, the SV model is ready to be estimated by applying the MCMC method. The estimation is performed by using 5000 MCMC draws after a burn-in of 100 for each data set (burn-in means throwing away some iteration at the beginning of an MCMC run). The Bayesian inference using MCMC based on ancillarity-sufficiency interweaving strategy (ASIS) is implemented through the statistical software R package developed by Gregor & Sylvia (2013). To obtain the estimated prediction intervals, we sample 5000 random variables from a normal distribution with mean zero and variance one to obtain ε_t in Equation 4. According to Equation 4, 5000 samplings of r_t could be obtained. PIs with confidence level of $(1 - \alpha)100\%$ can be derived by taking $\alpha/2$ and $(1 - \alpha/2)$ percentile of r_t .

Figure shows travel time prediction results of the ARIMA-SV model for four segments during peak hours. Both predicted mean and PIs with a confidence level of 95% have been provided. The ARIMA predicted value is marked as a red triangle lying at the center of each prediction interval

(green vertical line segment). This plot indicates that there is always a mismatch between the observed and the predicted value, which is denoted as the prediction error. On the other hand, PIs constructed by the ARIMA-SV model cover most of the observed values. As indicated in this figure, most of the blue dots (the observed values) lie within the green vertical line segments (the PIs). Depending on the level of uncertainty in the data, the length of the PI for each time interval varies. Wider PI indicates higher uncertainty about the predicted travel time, while narrower PI indicates lower uncertainty. Providing PIs to travelers assist them to schedule their trips with more confidence.

3.4.2 Results and Analysis

To assess the effectiveness of the ARIMA-GARCH and the ARIMA-SV model, we compare these models for the six studied segments at two different time intervals: five and fifteen minute aggregation time intervals. Performance measures in terms of MAPE, RMSE, PICP and MPIL are summarized. Table 3 provides performance measure of the mean equation for the six studied segments. Since both models use the same ARIMA model to predict the mean value of travel time for each segment, each segment has only one MAPE and one RMSE value at each aggregation level. The tabulated results show that ARIMA model provides adequate prediction of future travel time in turns of MAPE, as it ranges from 3.45% to 4.47% at five minute time intervals, and ranges from 1.97% to 6.68% at fifteen minute time intervals. On the other hand, the RMSE value ranges from 3.79 up to 22.94 indicating that the absolute error is proportional to the variation of the data.



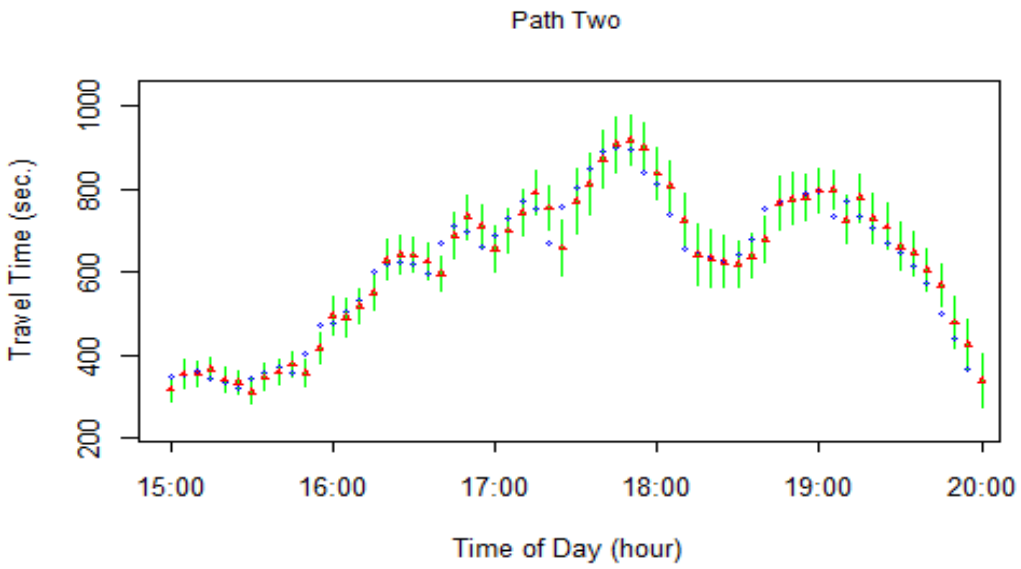
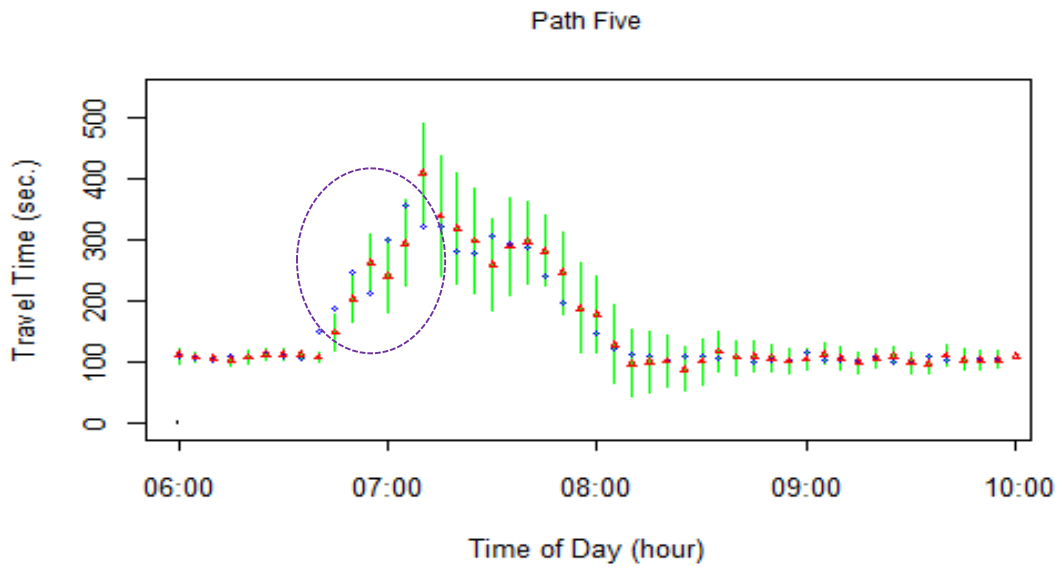


Figure 10: Prediction results for peak hour travel time at four segments.



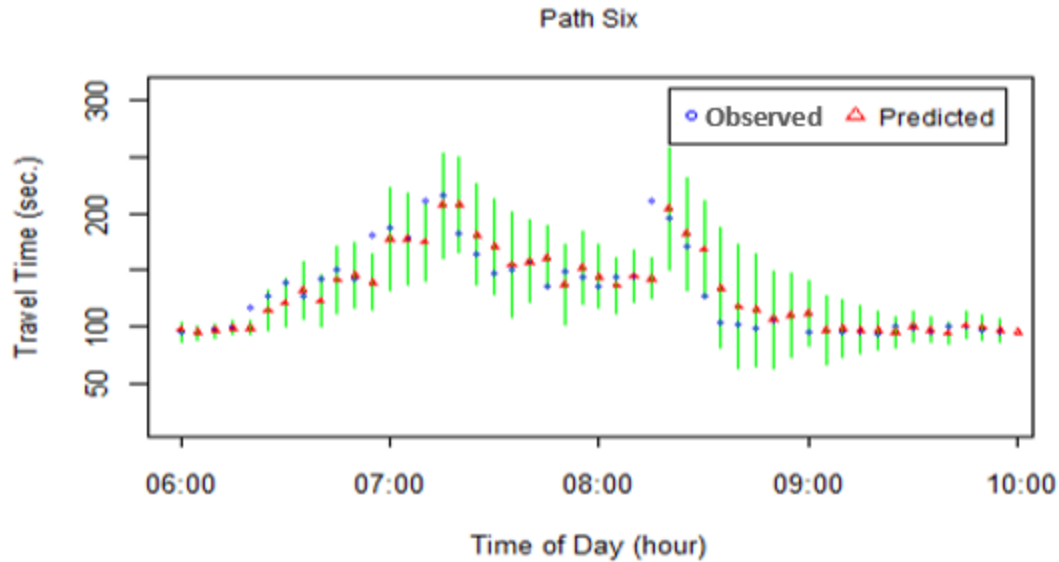


Figure 10 (Cont'd): Prediction results for peak hour travel time at four segments.

In Table 3, RMSE values for path six (8.41 at five minute time interval and 11.93 at fifteen minute time interval) are the smallest among these four segments, while RMSE values for path two (22.94 at five minute time interval and 34.42 at fifteen minute time interval) are among the largest. The ARIMA estimations for path three and four are the best in terms of both MAPE and RMSE values. This is because travel times on these two segments remain at a certain level with slight variations during the studied time period, which proves that ARIMA model performs better when the data are relatively stable. In contrast, the forecasting accuracy of ARIMA model will decrease as the variation of the data increases.

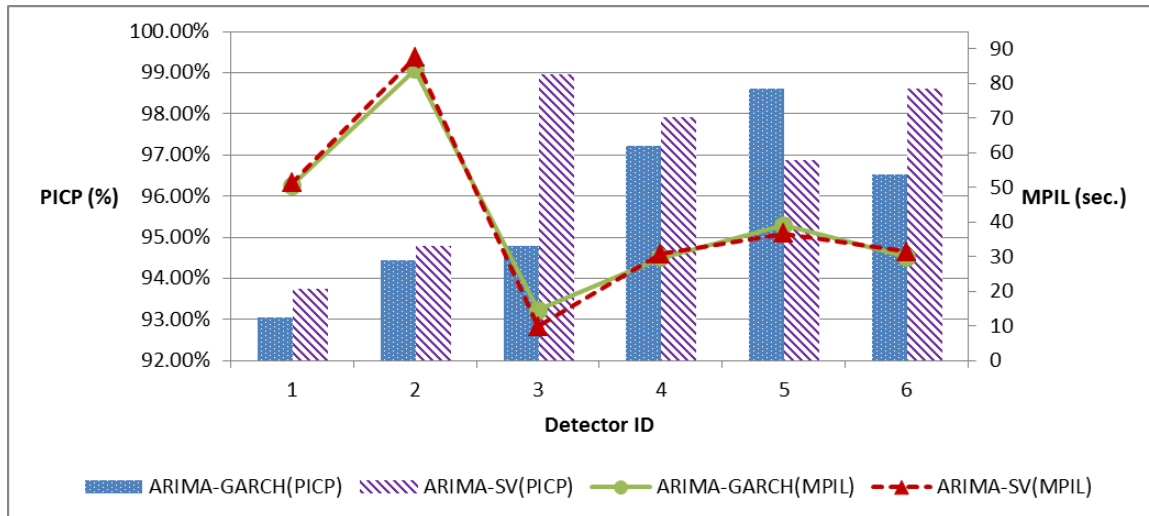
Table 3: Performance measures of the mean equation.

Path ID	Time Interval	1	2	3	4	5	6	Average Value
MAPE	5 min	4.47%	4.10%	3.45%	3.77%	4.72%	4.54%	4.18%
	15 min	6.68%	4.55%	1.97%	2.75%	4.79%	4.11%	4.14%
RMSE	5 min	19.67	22.94	3.79	6.82	11.82	8.41	12.24
	15 min	41.34	34.42	2.14	5.01	18.45	11.93	18.88

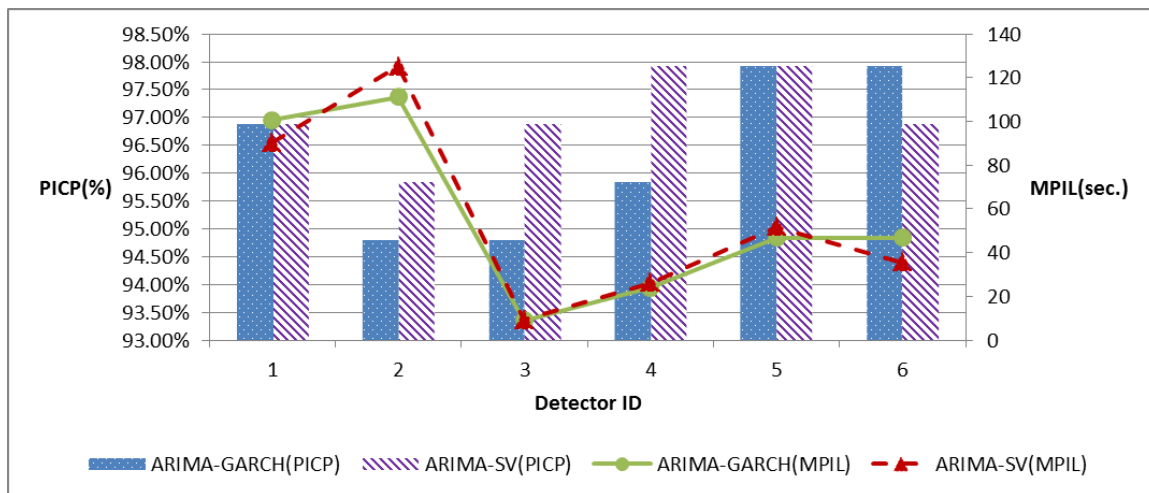
PICP and MPIL measure the prediction accuracy of the variance part. PICP measures the coverage probabilities of the prediction intervals. In our study, the prediction intervals for both ARIMA-GARCH and ARIMA-SV models are set with 95% confidence. Figure 11 compares PICP and MPIL values of both models at two different aggregation levels: five minute and fifteen minute. ARIMA-SV model provides higher coverage in most cases. At both five minute and fifteen minute aggregation levels, only in one out of five cases, the PICP value of the ARIMA-SV model is less than that of the ARIMA-GARCH model. The PICP value of the ARIMA-SV model ranges from 93.75% to 98.96% at five minute aggregation time interval, and ranges from 95.83% to 97.92% at

fifteen minute aggregation time interval. The ARIMA-SV model is capable to construct accurate PIs with predefined confidence level (95% in our study). Compared with the ARIMA-GARCH model, the ARIMA-SV model, in most cases, outperforms the ARIMA-GARCH model in terms of PCIP measure.

Regarding the MPIL value, which measures the width of the prediction interval, both models give similar performance. Comparing the MPIL values of both models for different segments indicates MPIL for segments 3 and 4 have the smallest value. This result demonstrates that the width of the prediction interval depends on the variation of traffic. As travel time for path three and four are relatively stable with minor variations, the MPIL for both models is low. Taking five minute aggregation time interval for example, the MPIL values for path three are 14.62 (ARIMA-GARCH) and 9.85 (ARIMA-SV) respectively, and 29.75 (ARIMA-GARCH) and 30.76 (ARIMA-SV) for path four. The MPIL value for segment two is the highest with 87.55 for the ARIMA-SV model. This is because the travel time varies the most on this path. It can be concluded that length of the constructed PI is sensitive to the variation of travel time on the studied segments, with higher variation leading to wider prediction interval, and vice versa. Since the width of prediction intervals provided by the ARIMA-SV and the ARIMA-GARCH models are similar and the ARIMA-SV model provides relatively higher coverage, the ARIMA-SV model outperforms the ARIMA-GARCH model in general.



(a)



(b)

Figure 11: Comparison of performance measures for six dataset by using ARIMA-GARCH and ARIMA-SV model with (a) 5 minute time interval (b) 15 minute time interval.

3.4.3 Summary

This study introduced an advanced stochastic volatility model to construct a prediction interval for each prediction point to capture this uncertainty. The proposed method was tested by using travel time data collected from Bluetooth sensors located along a freeway corridor in Connecticut. The proposed ARIMA-SV model was compared with the more widely used ARIMA-GARCH model. Different from the GARCH type models that assume deterministic nature of traffic volatility, the SV model considers this volatility as a non-deterministic process by specifying the variance follows some latent stochastic process. An advanced Monte Carlo Markov Chain estimation method for stochastic volatility model was applied to travel time reliability forecasting. The empirical experiment showed that the ARIMA-SV model outperforms the ARIMA-GARCH

model in terms of coverage probability and the length of PIs, as both models construct PIs that cover most of the observed value and the ARIMA-SV model tends to provide narrower PIs. In addition, by comparing the constructed PIs across different segments and at different types of aggregation time interval, it is revealed that the length of the constructed PI is sensitive to the variation of travel time. Higher variation leads to wider prediction interval, and lower variation is associated with narrower prediction interval. The width of the prediction interval indicates the variations of the prediction results and therefore provides a measure for the reliability of predicted travel time.

In summary, the proposed ARIMA-SV method shows its advantage in constructing more accurate prediction intervals. It accurately and effectively covers most of the observed values of travel time compared with the ARIMA-GARCH model. From a practical point of view, the proposed ARIMA-SV model provides not only a mean but also an upper and lower bound of future travel time, therefore capturing the uncertainty associated with prediction. As the mean value is unable to provide information regarding the variability of future traffic, the prediction interval can provide an upper and lower bounds that capture future travel time with predetermined confidence level. Therefore, the proposed model can be used in real time to provide more reliable and informative future traffic information for travelers. The ARIMA-SV model can be regarded as a promising algorithm to disseminate traffic information to travelers through traveler information systems to provide guidance for pre-trip planning as well as en-route navigation. Further research includes studying trend and seasonal patterns in the residuals series and to comprehensively evaluate different volatility model applications in travel time prediction field

4.0 CONCLUSION AND RECOMMENDATIONS

4.1 SUMMARY

Travel time prediction is a critical topic in the development of ITS systems. Especially with the rapid development of the Advanced Traveler Information Systems and Advanced Traffic Management Systems, more accurate and reliable travel time information is needed to enable the success of these systems' development. Apart from its importance, travel time estimation and prediction are complex and challenging tasks. Resulting from the interactions among different vehicle-driver combinations, and exogenous factors such as weather, demand, and roadway conditions, travel time often experiences strong fluctuations across different periods and traffic conditions. These rapid fluctuations are often complex and difficult to predict. Fully understanding these fluctuations and developing accurate travel time prediction algorithms is critical.

Inspired by the need of travel time predictions, a wide range of methodologies have been proposed in the literature. As discussed in Chapter 2, existing travel time prediction algorithms can be divided into four major categories: parametric, non-parametric, hybrid and prediction interval based approaches. The parametric methods usually have a well-established theoretical foundation but with lots of strict model assumptions. Comparatively, the non-parametric methods require less model assumptions but some of them may be difficult to interpret. The hybrid methods take advantage of different prediction models but some models may be too complex when making predictions. The travel time interval based algorithms belongs to the category of hybrid methods and it provides not only the mean but also a prediction bound to capture both prediction accuracy and reliability of the model. As it is a relatively new area in travel time prediction, there are limited studies in the literature.

In this research, both prediction accuracy and reliability issues have been addressed in freeway travel time prediction. Although most existing travel time prediction models are able to provide accurate predictions during non-peak hours, peak hour travel time prediction is still a challenging topic. Investigating travel time patterns during both non-peak and peak hours and developing a more accurate travel time prediction algorithm is critical. On the other hand, because of the difficulties in predicting travel time, especially during peak hours, another issue that needs to be considered is the reliability issues of the model. The model should consider situations when traffic is highly volatile when a point prediction becomes 'less accurate'. In this case, the prediction interval based approach provides a prediction bound to indicate how likely it will capture the observed travel time value and therefore is able to indicate how reliable the prediction is.

To capture the uncertainty and variations of travel time data, this study proposed two different statistical volatility models: component GARCH and stochastic volatility models. In general, the statistical volatility model predicts future traffic volatility based on its previous

volatility values. In a transportation system, travelers respond differently to unexpected changes in travel time. The presence of this volatility in traffic may lead to changes in driving behavior in order to compensate for the resulting changes in expected arrival time. These changes lead to increased traffic volatility with a decreasing rate over time in order to restore the past stability of the system. The volatility models capture this changed traffic patterns over time and make further prediction.

The component GARCH models consider situations when seasonal (cyclical) patterns or trends exist in data. For some road segments, when commuters account for a large percentage of the total traffic volume, travel time of these types of segments may show strong cyclical patterns. In this case, the seasonal component should be considered when modeling the data. Through decomposition, the component GARCH models potentially improve the prediction accuracy. Another type of the volatility model, the stochastic volatility model considers the conditional variance of travel time data as an unobserved stochastic process therefore allows for a more flexible application and can account for uncertainties inherent in traffic phenomena.

4.2 CONCLUSION

The following list provides the conclusions/findings of this research:

- Due to the complex nature of travel time prediction problem, the traditional point based prediction approach is unable to perfectly account for uncertainties in traffic. There is often a mismatch between the predicted mean and the observed value. A prediction interval based approach as an alternative way to represent uncertainties associated with travel time prediction has the potential to provide more reliable prediction information.
- Volatility-based travel time prediction models relax the constant variation assumption. This kind of method treats the current volatility as a function of its past values and can be used to construct more accurate PIs to capture travel time uncertainty.
- The component GARCH models are able to capture the seasonal patterns in travel time volatility. When seasonal (cyclical) patterns exist, the component GARCH model could be a better choice compared with the traditional GARCH models.
- The stochastic volatility models consider part of the change in travel time volatility are due to random shocks, while the GARCH type model treat the volatility as time changing but not stochastic process. Through using advanced Monte Carlo Markov Chain estimation method to fit the stochastic volatility model, the model is able to provide more accurate PIs.

4.3 FUTURE RECOMMENDATION

Although this research provided contributions to the existing literature in the area of freeway travel time prediction, there are some other research avenues that can be pursued. Future directions of the research are provided below:

- In this research, we only utilize travel time information. But with the advanced technology development, more and more data are available to use, such as incident, weather, work zone and so on. Since these events may have significant influence on travel time, utilizing this information could help improve prediction accuracy. For example, traffic congestion is more likely to occur due to inclement weather conditions as the freeway capacity drops while demand does not drop. In this case, if we can include weather conditions as explanatory variables, then the model would potentially capture the weather impact on travel time. When using weather information to predict travel time, the weather forecast information will be used, therefore, how to utilize the weather forecast information and considering its reliability could be another research topic. To sum up, future research should use the information of external impact factors when predicting travel time.
- Uncertainties associated with travel time prediction is relatively a new area in travel time prediction, few literature focuses on travel time uncertainty prediction. As indicated in the literature review section, there are generally two types of PIs based approach: ensemble methods and statistical volatility based approach. This research mainly focuses on using statistical volatility methods to model the uncertainty associated with travel time prediction. Two different types of volatility models have been proposed. The study results show that the PIs based approach shows its promising abilities in indicating the uncertainty associated with prediction. While in the future, we can also use the ensemble based algorithms to construct PIs when predicting travel time. It is also beneficial to have a comparison of the PIs constructed by volatility based and ensemble based method and discuss the advantages and disadvantages of each model in addressing uncertainty associated with travel time prediction.

5.0 REFERENCES

- Ahmed, M.S. and Cook, A.R. (1979) Analysis of freeway traffic time-series data by using Box-Jenkins techniques. *Transportation Research Record*.
- Ahmed, M.M. and Abdel-Aty, M. (2013) Application of Stochastic Gradient Boosting Technique to Enhance Reliability of Real-Time Risk Assessment. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2386, pp. 26-34,.
- Amit, Y. and Geman, D. (1997) Shape quantization and recognition with randomized trees. *Neural computation*, vol. 9, pp. 1545-1588.
- Antoniou, C., Koutsopoulos, H.N., and Yannis, G. (2013) Dynamic data-driven local traffic state estimation and prediction. *Transportation Research Part C: Emerging Technologies*, vol. 34, pp. 89-107.
- Andersen, T.G. and Bollerslev, T. (1997) Intraday periodicity and volatility persistence in financial markets. *Journal of empirical finance*, vol. 4, pp. 115-158.
- Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, vol. 31, pp. 307-327.
- Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992) ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics*, vol. 52, pp. 5-59.
- Breiman, L. (1984) *Classification and regression trees*. Belmont, Calif.: Wadsworth International Group.
- Breiman, L. (1996) Bagging predictors. *Machine learning*, vol. 24, pp. 123-140.
- Breiman, L. (2001) Random forests. *Machine learning*, vol. 45, pp. 5-32.
- Boto-Giralda, D., Díaz-Pernas, F.J. González-Ortega, D. Díez-Higuera, J. F. Antón-Rodríguez, M. Martínez-Zarzuela, M. and Torre-Díez, I. (2010) Wavelet-Based Denoising for Traffic Volume Time Series Forecasting with Self-Organizing Neural Networks. *Computer-Aided Civil and Infrastructure Engineering*, vol. 25, pp. 530-545.
- Cetin, M. and Comert, G. (2006) Short-term traffic flow prediction with regime switching models. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1965, pp. 23-31.
- Chen, C., Hu, J. Meng, Q. and Zhang, Y. (2011) Short-time traffic flow prediction with ARIMA-GARCH model. in *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pp. 607-612.
- Chen, H., Grant-Muller, S., Mussone, L., and Montgomery, F. (2001) A Study of Hybrid Neural Network Approaches and the Effects of Missing Data on Traffic Forecasting. *Neural computing & applications*, vol. 10, pp. 277-286, 2001/12/01.
- Chen, H.K., and Wu, C.J. (2012) Travel Time Prediction Using Empirical Mode Decomposition and Gray Theory Example of National Central University Bus in Taiwan. *Transportation Research Record*, pp. 11-19.
- Chien, S. and Kuchipudi, C. (2003) Dynamic Travel Time Prediction with Real-Time and Historic Data. *Journal of Transportation Engineering*, vol. 129, pp. 608-616.
- Choe, T., Skabardonis, A., and Varaiya, P. (2002) Freeway performance measurement system: operational analysis tool. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1811, pp. 67-75.
- Chung, Y.S. (2013) Factor complexity of crash occurrence: An empirical demonstration using boosted regression trees. *Accident Analysis & Prevention*, vol. 61, pp. 107-118.

- Clark, S. (2003) Traffic Prediction Using Multivariate Nonparametric Regression. *Journal of Transportation Engineering*, vol. 129, pp. 161-168.
- Danech-Pajouh, M. and Aron, M. (1991) ATHENA: a method for short-term inter-urban motorway traffic forecasting. *Recherche Transports Sécurité*.
- Davis, G. and Nihan, N. (1991) Nonparametric Regression and Short-Term Freeway Traffic Forecasting. *Journal of Transportation Engineering*, vol. 117, pp. 178-188.
- Davis, G.A., Nihan, N.L., Hamed, M.M. and Jacobson L.N. (1990) Adaptive forecasting of freeway traffic congestion. *Transportation Research Record*.
- Deng J.L. (1989) Introduction to grey system theory. *The Journal of grey system*, vol. 1, pp. 1-24.
- Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, pp. 987-1007.
- Engle R.F. and Sokalska M.E. (2012) Forecasting intraday volatility in the us equity market. multiplicative component garch. *Journal of Financial Econometrics*, vol. 10, pp. 54-83.
- Elith, J., Leathwick, J. R. and Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, vol. 77, pp. 802-813.
- Farokhi Sadabadi, K., Hamed, M., and Haghani, A. (2010) Evaluating moving average techniques in short-term travel time prediction using an AVI data set. in *Transportation Research Board 89th Annual Meeting*.
- Fei, X., Lu, C.-C. and Liu, K. (2011) A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction. *Transportation Research Part C: Emerging Technologies*, vol. 19, pp. 1306-1318.
- Friedman, J.H. (2001) Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pp. 1189-1232.
- Friedman, J.H. (2002) Stochastic gradient boosting. *Computational Statistics & Data Analysis*, vol. 38, pp. 367-378.
- Ghosh, B., Basu, B. and O'Mahony, M. (2009) Multivariate Short-Term Traffic Flow Forecasting Using Time-Series Analysis. *Intelligent Transportation Systems, IEEE Transactions on*, vol. 10, pp. 246-254.
- Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2003) *Bayesian data analysis*: CRC press.
- Ghalanos, A. (2013) rugarch: Univariate GARCH models. in *R package version 1.2-7 ed*.
- Guo, J. and Williams, B.M. (2010) Real-Time Short-Term Traffic Speed Level Forecasting and Uncertainty Quantification Using Layered Kalman Filters. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2175, pp. 28-37.
- Hamed, M., Al-Masaeid, H. and Said, Z. (1995) Short-Term Prediction of Traffic Volume in Urban Arterials. *Journal of Transportation Engineering*, vol. 121, pp. 249-254.
- Hamner, B. (2010) Predicting travel times with context-dependent random forests by modeling local and aggregate traffic flow. in *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*, pp. 1357-1359.
- Haghani, A., Hamed, M. Sadabadi, K. F. Young, S. and Tarnoff, P. (2010) Data collection of freeway travel time ground truth with bluetooth sensors. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2160, pp. 60-68.
- Hamad, K. Shourijeh, M. T. Lee, E. and Faghri, A. (2009) Near-Term Travel Speed Prediction Utilizing Hilbert-Huang Transform. *Computer-Aided Civil and Infrastructure Engineering*, vol. 24, pp. 551-576.
- Hastie, T., Tibshirani, R. Friedman, J. Hastie, T. Friedman, J. and Tibshirani, R. (2009) *The elements of statistical learning* vol. 2: Springer.

- Ho, T. K. (1995) Random decision forests. in *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, pp. 278-282.
- Ho, T. K. (1998) The random subspace method for constructing decision forests. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, pp. 832-844.
- Hong, W.C. (2011) Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm. *Neurocomputing*, vol. 74, pp. 2096-2107.
- Hyndman, R.J. and Khandakar, Y. (2007) Automatic time series for forecasting: the forecast package for R.
- Ishak S. and Al-Deek H. (2002) Performance evaluation of short-term time-series traffic prediction model. *Journal of Transportation Engineering*, vol. 128, pp. 490-498.
- Ishak, S. and Alecsandru, C. (2004) Optimizing traffic prediction performance of neural networks under various topological, input, and traffic condition settings. *Journal of Transportation Engineering*, vol. 130, pp. 452-465.
- Ishak, S., Kotha, P. and Alecsandru, C. (2003) Optimization of dynamic neural network performance for short-term traffic prediction. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1836, pp. 45-56.
- Jeffery, D., Russam, K., and Robertson, D. Electronic route guidance by AUTOGUIDE: the research background. *Traffic engineering & control*, vol. 28, pp. 525-529, 1987.
- Jiang, X. and Adeli, H. (2004) Wavelet Packet-Autocorrelation Function Method for Traffic Flow Pattern Analysis. *Computer-Aided Civil and Infrastructure Engineering*, vol. 19, pp. 324-337.
- Jiang, X. and Adeli, H. (2005) Dynamic wavelet neural network model for traffic flow forecasting. *Journal of Transportation Engineering*, vol. 131, pp. 771-779.
- Kamarianakis, Y. Kanas, A. and Prastacos, P. (2005) Modeling traffic volatility dynamics in an urban network. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1923, pp. 18-27.
- Kamarianakis, Y. and Prastacos, P. (2005) Space-time modeling of traffic flow. *Computers & Geosciences*, vol. 31, pp. 119-133.
- Karlaftis, M.G. and Vlahogianni, E.I. (2009) Memory properties and fractional integration in transportation time-series. *Transportation Research Part C: Emerging Technologies*, vol. 17, pp. 444-453.
- Kastner G. and Frühwirth-Schnatter S. (2013) Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis*.
- Kaysi, I., Ben-Akiva, M., and Koutsopoulos, H. (1993) *Integrated approach to vehicle routing and congestion prediction for real-time driver guidance*.
- Kearns, M. (1988) Thoughts on hypothesis boosting. *Unpublished manuscript, December*.
- Khosravi, A., Mazloumi, E. Nahavandi, S. Creighton, D. and Van Lint, J. (2011) Prediction intervals to account for uncertainties in travel time prediction. *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, pp. 537-547.
- Khosravi, A., Mazloumi, E. Nahavandi, S. Creighton, D. and Van Lint, J. (2011) A genetic algorithm-based method for improving quality of travel time prediction intervals. *Transportation Research Part C: Emerging Technologies*, vol. 19, pp. 1364-1376.
- Kim, S., Shephard, N. and Chib, S. (1998) Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, vol. 65, pp. 361-393.
- Koren, Y. (2009) The bellkor solution to the netflix grand prize. *Netflix prize documentation*.

- Lee, G. and Engle, R. (1993) A Permanent and Transitory Component Model of Stock Return Volatility. Available at SSRN 5848, 1993.
- Levin M. and Tsao Y.-D. (1980) On Forecasting Freeway Occupancies and Volumes (Abridgment). *Transportation Research Record*.
- Li, R. and Rose, G. (2011) Incorporating uncertainty into short-term travel time predictions. *Transportation Research Part C: Emerging Technologies*, vol. 19, pp. 1006-1018.
- Liu, H., van Zuylen, H. van Lint, H. and Salomons, M. (2006) Predicting urban arterial travel time with state-space neural networks and Kalman filters. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1968, pp. 99-108.
- Leshem, G. and Ritov, Y. a. (2007) Traffic Flow Prediction using Adaboost Algorithm with Random Forests as a Weak Learner. *International Journal of Intelligent Technology*, vol. 2.
- Min, W. and Wynter, L. (2011) Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, vol. 19, pp. 606-616.
- Myung, J., Kim, D.K., Kho, S.Y. and Park, C.H. (2011) Travel Time Prediction Using k Nearest Neighbor Method with Combined Data from Vehicle Detector System and Automatic Toll Collection System. *Transportation Research Record*, pp. 51-59.
- Nanthawichit, C., Nakatsuji, T. and Suzuki, H. (2003) Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway. *Transportation Research Record: Journal of the Transportation Research Board*, vol.
- Okutani, I. and Stephanedes, Y.J. (1984) Dynamic prediction of traffic volume through Kalman filtering theory. *Transportation Research Part B: Methodological*, vol. 18, pp. 1-11.
- Peng, Y., Lei, M. Li, J.-B. and Peng, X.-Y. (2012) A novel hybridization of echo state networks and multiplicative seasonal ARIMA model for mobile communication traffic series forecasting. *Neural Computing and Applications*, pp. 1-8, 2012/12/01.
- Quek, C., Pasquier, M. and Lim, B.B.S. (2006) POP-TRAFFIC: A novel fuzzy neural approach to road traffic analysis and prediction. *Intelligent Transportation Systems, IEEE Transactions on*, vol. 7, pp. 133-146.
- Robinson, S. and Polak, J. (2005) Modeling Urban Link Travel Time with Inductive Loop Detector Data by Using the k-NN Method. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1935, pp. 47-56.
- Schapire, R.E. (1990) The strength of weak learnability. *Machine learning*, vol. 5, pp. 197-227.
- Shrestha, D.L. and Solomatine, D.P. (2006) Machine learning approaches for estimation of prediction interval for the model output. *Neural Networks*, vol. 19, pp. 225-235.
- Shumway, R.H. and Stoffer, D.S. (2006) *Time series analysis and its applications : with R examples*, 2nd [updated] ed. New York: Springer.
- Smith, B.L., Williams, B.M. and Keith, R. (2002) Oswald Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, vol. 10, pp. 303-321.
- Smith, B. and Demetsky, M. (1996) Multiple-Interval Freeway Traffic Flow Forecasting. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1554, pp. 136-141.
- Smith B. and Demetsky M. (1997) Traffic Flow Forecasting: Comparison of Modeling Approaches. *Journal of Transportation Engineering*, vol. 123, pp. 261-266.
- Stathopoulos, A., Dimitriou, L., and Tsekeris T. (2008) Fuzzy modeling approach for combined forecasting of urban traffic flow. *Computer-Aided Civil and Infrastructure Engineering*, vol. 23, pp. 521-535.

- Stephanedes, V., Michalopoulos, P.G., and Plum, R.A. (1981) Improved estimation of traffic flow for Real-Time control (Discussion and closure). *Transportation Research Record*.
- Stathopoulos, A. and Karlaftis, M.G. (2003) A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies*, vol. 11, pp. 121-135.
- Sun, S., Zhang, C., and Yu, G. (2006) A Bayesian network approach to traffic flow forecasting. *Intelligent Transportation Systems, IEEE Transactions on*, vol. 7, pp. 124-132.
- Tsay, R.S. (2010) *Analysis of financial time series*, 3rd ed. Cambridge, Mass.: Wiley.
- Tsekeris, T. and Stathopoulos, A. (2006) Real-time traffic volatility forecasting in urban arterial networks. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1964, pp. 146-156.
- Tsekeris, T. and Stathopoulos, A. (2009) Short-term prediction of urban traffic variability: Stochastic volatility modeling approach. *Journal of Transportation Engineering*, vol. 136, pp. 606-613.
- Van Areem, B., Kirby, H.R., Van Der Vlist, M.J., and Whittaker, J.C. (1997) Recent advances and applications in the field of short-term traffic forecasting. *International Journal of Forecasting*, vol. 13, pp. 1-12.
- Van Der Voort, M., Dougherty, M., and Watson, S. (1996) Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies*, vol. 4, pp. 307-318.
- Van Hinsbergen, C.P.I., van Lint, J.W. and Van Zuylen, H. (2009) Bayesian training and committees of state-space neural networks for online travel time prediction. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2105, pp. 118-126.
- Van Hinsbergen, C., Van Lint, J., and Van Zuylen, H. (2009) Bayesian committee of neural networks to predict travel times with confidence intervals. *Transportation Research Part C: Emerging Technologies*, vol. 17, pp. 498-509.
- Van Lint, J.W.C. (2008) Online learning solutions for freeway travel time prediction. *Ieee Transactions on Intelligent Transportation Systems*, vol. 9, pp. 38-47.
- Van Lint J.W.C. and Van Hinsbergen C.P.I.J. (2012) Short-Term Traffic and Travel Time Prediction Models. *Artificial Intelligence Applications to Critical Transportation Issues*, p. 22.
- Vlahogianni, E.I., Karlaftis, M.G. and Golias J.C. (2007) Spatio-Temporal Short-Term Urban Traffic Volume Forecasting Using Genetically Optimized Modular Networks. *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, pp. 317-325.
- Van Lint, J.W., Hoogendoorn, S., and van Zuylen, H. J. (2002) Freeway travel time prediction with state-space neural networks: Modeling state-space dynamics with recurrent neural networks. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1811, pp. 30-39.
- Van Lint, J., Hoogendoorn, S. and van Zuylen, H.J. (2005) Accurate freeway travel time prediction with state-space neural networks under missing data. *Transportation Research Part C: Emerging Technologies*, vol. 13, pp. 347-369.
- Van Lint, J. (2006) Reliable real-time framework for short-term freeway travel time prediction. *Journal of Transportation Engineering*, vol. 132, pp. 921-932.
- Vlahogianni, E.I., Golias, J.C., and Karlaftis, M.G. (2004) Short-term traffic forecasting: Overview of objectives and methods. *Transport reviews*, vol. 24, pp. 533-557.

- Wang, Y., Papageorgiou, M., and Messmer, A. (2008) Real-time freeway traffic state estimation based on extended Kalman filter: Adaptive capabilities and real data testing. *Transportation Research Part A: Policy and Practice*, vol. 42, pp. 1340-1358.
- Wang, Y. (2011) Prediction of weather impacted airport capacity using ensemble learning. in *Digital Avionics Systems Conference (DASC), 2011 IEEE/AIAA 30th*, pp. 2D6-1-2D6-11.
- Wang, J. and Shi, Q. (2012) Short-term traffic speed forecasting hybrid model based on Chaos–Wavelet Analysis-Support Vector Machine theory. *Transportation Research Part C: Emerging Technologies*.
- Wei, Y. and Chen, M.-C. (2012) Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C: Emerging Technologies*, vol. 21, pp. 148-162.
- Williams, B., Durvasula, P., and Brown, D. (1998) Urban Freeway Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1644, pp. 132-141.
- Whittaker, J., Garside, S. and Lindveld, K. (1997) Tracking and predicting a network traffic process. *International Journal of Forecasting*, vol. 13, pp. 51-61.
- Xia, J., Nie, Q. Huang, W. and Qian, Z. (2013) Reliable Short-Term Traffic Flow Forecasting for Urban Roads Using a Multivariate GARCH Model. in *Transportation Research Board 92nd Annual Meeting*.
- Xie, Y., Zhang, Y. and Ye, Z. (2007) Short-Term Traffic Volume Forecasting Using Kalman Filter with Discrete Wavelet Decomposition. *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, pp. 326-334.
- Yang, M., Liu, Y., and You, Z. (2010) The reliability of travel time forecasting. *Intelligent Transportation Systems, IEEE Transactions on*, vol. 11, pp. 162-171.
- Yang, F., Yin, Z., Liu, H.X. and Ran, B. (2004) Online recursive algorithm for short-term traffic prediction. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1879, pp. 1-8.
- Yeon, J., Elefteriadou, L., and Lawphongpanich, S. (2008) Travel time estimation on a freeway using Discrete Time Markov Chains. *Transportation Research Part B: Methodological*, vol. 42, pp. 325-338.
- Yin, H., Wong, S., Xu, J., and Wong, C. (2002) Urban traffic flow prediction using a fuzzy-neural approach. *Transportation Research Part C: Emerging Technologies*, vol. 10, pp. 85-98.
- Yildirimoglu M. and Geroliminis N. (2013) Experienced travel time prediction for congested freeways. *Transportation Research Part B: Methodological*, vol. 53, pp. 45-63.
- Yu, Y. and Meng, X.L. (2011) To center or not to center: That is not the question—an Ancillarity–Sufficiency Interweaving Strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics*, vol. 20, pp. 531-570.
- Yu, B., Yang, Z.Z., and Chen, K. (2010) Hybrid model for prediction of bus arrival times at next station. *Journal of Advanced Transportation*, vol. 44, pp. 193-204.
- Zeng, X. and Zhang, Y. (2013) Development of Recurrent Neural Network Considering Temporal-Spatial Input Dynamics for Freeway Travel Time Modeling. *Computer-Aided Civil and Infrastructure Engineering*.
- Zhang, Y. Sun, R. Haghani, A. and Zeng, X. (2013) Univariate Volatility-Based Models for Improving Quality of Travel Time Reliability Forecasting. in *Transportation Research Board 92nd Annual Meeting*.

- Zhang, Y. and Liu, Y.C. (2011) Analysis of peak and non-peak traffic forecasts using combined models. *Journal of Advanced Transportation*, vol. 45, pp. 21-37.
- Zheng, W., Lee, D.H. and Shi, Q. (2006) Short-term freeway traffic flow prediction: Bayesian combined neural network approach. *Journal of Transportation Engineering*, vol. 132, pp. 114-121.
- Zhou, Z.H. (2012) *Ensemble methods: foundations and algorithms*. Boca Raton, FL: Taylor & Francis.
- Zou, N., Wang, J., Chang, G.-L., and Paracha, J. (2009) Application of Advanced Traffic Information Systems. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2129, pp. 62-72.