

 Open access • Journal Article • DOI:10.1016/J.TRB.2015.02.008

Nonlinear multivariate time–space threshold vector error correction model for short term traffic state prediction — Source link

Tao Ma, Zhou Zhou, Baher Abdulhai

Institutions: University of Toronto

Published on: 01 Jun 2015 - Transportation Research Part B-methodological (Pergamon)

Topics: Traffic flow, Statistical model, Error detection and correction, Piecewise linear function and Error correction model

Related papers:

- [Real-time road traffic prediction with spatio-temporal correlations](#)
- [Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results](#)
- [Traffic volume forecasting based on radial basis function neural network with the consideration of traffic flows at the adjacent intersections](#)
- [Comparison of parametric and nonparametric models for traffic flow forecasting](#)
- [A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/nonlinear-multivariate-time-space-threshold-vector-error-4iemge5y76>

Nonlinear Multivariate Time-Space Threshold Vector Error Correction Model for Short Term Traffic State Prediction

by

Tao Ma

A thesis submitted in conformity with the requirements
for the degree of the Doctor of Philosophy
Graduate Department of Civil Engineering
University of Toronto

© Copyright by Tao Ma 2016

Nonlinear Multivariate Time-Space Threshold Vector Error Correction Model for Short Term Traffic State Prediction

Tao Ma

Doctor of Philosophy

Graduate Department of Civil Engineering
University of Toronto

2016

Abstract

The time-space threshold vector error correction (TS-TVEC) model is proposed and developed in this research for short term (hourly) traffic state prediction. The theory and method of cointegration with error correction mechanism is employed in the general design of the new statistical model TS-TVEC. An inherent connection is revealed between the error correction model and the transformed fundamental diagrams in macroscopic traffic flow theory. Error correction model is a linear model established on difference space, whereas, the fundamental relations between traffic variables exhibit piecewise linearity when they are transformed to difference space. Meanwhile, dynamics of traffic variables can be reflected in difference space. A threshold regime switching framework is implemented to deal with unknown structural breaks in non-stationary traffic time series to capture multiple traffic states. Spatial cross-correlated information is incorporated with a piecewise linear vector error correction model. As the TS-TVEC model is designed for nonlinear multivariate forecasting, it is compared to Neural Network, and Support Vector Regression model that are commonly used in this regard to test the effectiveness and robustness of the new statistical model. The TS-TVEC model was experimented on a small scale case in our empirical study and applied on a larger scale case thereafter. Both cases show that the TS-TVEC model is an effective tool that is capable of modelling the complexity of stochastic traffic processes and potentially applicable to real time traffic state prediction.

Key words: cointegration, vector error correction, threshold regime switching, short term traffic state prediction, neural network, support vector regression

Acknowledgments

The author is sincerely grateful to those who provided help during the course of this research including Professor Baher Abdulhai, Professor Zhou Zhou, Professor Amer Shalaby, Professor Fang Yao, Professor Matthew Roorda, Matthieu Stigler, Professor Roberto Baragona, Professor Domenico Cucina, Marcus Beck, Rebecca Strung, Professor Markus Bussmann, Professor Brent Sleep, Mohamed Elshenawy, Ming Jie Ma, and Stephen Keen.

The new methodology in chapter 3 and empirical study in chapter 4 in this thesis were published in the Journal of Transportation Research Part B. Special thanks go to the editor-in-chief and two reviewers for their many helpful insights and valuable comments that are also beneficial to this thesis.

The author would also like to thank Professor Roy Kwon for chairing the Final Oral Examination, and the external examiner Professor Pengfei Li from University of Waterloo for his review and appraisal of the thesis.

My deepest gratitude goes to my wife and kids, my parents, and my brother Professor Ma for years of their support.

To those

who persevere in adversity,

who help others turn dreams to reality,

whom I love

Table of Contents

Acknowledgments.....	iii
Table of Contents.....	v
List of Tables.....	viii
List of Figures.....	ix
Chapter 1 Introduction.....	1
1.1 Introduction.....	1
1.2 Challenges of traffic modelling and forecasting.....	2
1.2.1 Data source.....	2
1.2.2 Stochastic and dynamic traffic variables.....	3
1.2.3 Statistical characteristics.....	9
1.3 Contributions of the research.....	17
1.4 Scope of the research.....	18
Chapter 2 Literature review.....	20
2.1 Traffic state estimation.....	21
2.1.1 Second order macroscopic traffic flow models.....	21
2.1.2 Cell transmission model for traffic state estimation.....	24
2.1.3 Critiques and limitation.....	25
2.2 Traffic state prediction.....	26
2.2.1 Neural Network.....	27
2.2.2 Kalman recursion.....	29
2.2.3 Support vector regression.....	31
2.2.4 Nonparametric regression.....	32
2.2.5 Time series.....	34
2.3 Summary.....	38
Chapter 3 Methodology.....	39
3.1 General design of the proposed statistical model.....	39
3.2 Theoretical rationale.....	40
3.2.1 Spurious regression and cointegration.....	40
3.2.2 Error correction representation.....	41
3.3 Inherent connection.....	42

3.3.1	Cointegration between traffic variables	42
3.3.2	Transformed fundamental diagrams and error correction model	43
3.4	Time-space vector error correction model	44
3.5	Threshold regime switching framework	46
3.5.1	Prototype of threshold model.....	46
3.5.2	Threshold cointegration vector error correction model	48
3.5.3	Capacity of threshold structure	49
3.6	Model estimation.....	49
Chapter 4	Testing and evaluation	53
4.1	TS-TVEC model	53
4.1.1	Test cointegration and threshold effect.....	53
4.1.2	Order selection for the TS-TVEC model	55
4.1.3	Differencing between spatial time series	58
4.1.4	Model specification.....	58
4.1.5	Check residuals	59
4.2	MLP Neural Network.....	64
4.2.1	Input layer and output layer	66
4.2.2	The number of hidden layers	66
4.2.3	Hidden layer neurons and training iterations	67
4.3	Support vector regression.....	69
4.3.1	Kernel method.....	69
4.3.2	Support vector regression	72
4.3.3	Parameter specification.....	74
4.4	Comparison of predictions	74
Chapter 5	Large scale application.....	80
5.1	Highway network	80
5.2	Time series data.....	84
5.3	Model determination	88
5.4	Forecasting	92
5.5	Discussion	92
Chapter 6	Summary, discussion and conclusion.....	102
6.1	Summary	102

6.2 Discussion	103
6.3 Conclusion.....	104
6.4 Future research	106
References.....	107
Appendix A Exhibitions of model fitness and hourly rolling predictions for 35 locations	121
Appendix B Fundamental Diagrams from location 1 and 2	157
Copyright Acknowledgements.....	163

List of Tables

Table 1.1 Unit Root Test.....	12
Table 2.1 Selected K -nearest neighbors	33
Table 4.1 Phillips-Ouliaris cointegration test	54
Table 4.2 Johansen cointegration test (eigenvalue)	54
Table 4.3 Johansen cointegration test (trace).....	54
Table 4.4 Hansen threshold effect test (univariate)	56
Table 4.5 Hansen threshold effect test (multivariate).....	56
Table 4.6 Hansen and Seo threshold cointegration test	56
Table 4.7 Model order selection based on AIC, BIC, and MSE criterion	57
Table 4.8 Legend used in Tables 4.9, 4.10, 4.11, 4.12 and 4.13.....	59
Table 4.9 Augmented Dickey-Fuller (ADF) unit root test	60
Table 4.10 DF-GLS unit root test	60
Table 4.11 KPSS unit root test.....	61
Table 4.12 Zivot-Andrews unit root test.....	61
Table 4.13 Comparison of variance	62
Table 4.14 Specification of estimated TS-TVEC Models (2 examples)	63
Table 4.15 Ljung-Box Test for Residual series	64
Table 4.16 Comparison of 90-minute interval one-step-ahead rolling prediction for 2 days.....	75
Table 4.17 Comparison of 60-minute interval one-step-ahead rolling prediction for 48 hours ...	76
Table 5.1 Locations of prediction	83
Table 5.2 Phillips-Ouliaris cointegration test	86
Table 5.3 Johansen cointegration test (eigenvalue)	86
Table 5.4 Hansen and Seo threshold cointegration test	86
Table 5.5 Model selection, specification and prediction Highway 401 Col WB	89
Table 5.6 Model selection, specification and prediction Highway 401 Exp EB	90
Table 5.7 Model selection, specification and prediction Highway 401 Exp WB.....	90
Table 5.8 Model selection, specification and prediction Highway 401 Col EB.....	91
Table 5.9 Model selection, specification and prediction Highway 400 and 404 NB & SB	91
Table 5.10 Summary of model prediction accuracy at 35 locations.....	92

List of Figures

Fig. 1.1 Data map for empirical study	3
Fig. 1.2 Profile of daily traffic volume, speed, and occupancy	5
Fig. 1.3 Time series plot of traffic volume, speed, and occupancy	6
Fig. 1.4 3D scatter plot of traffic volume, speed and occupancy, location 3.....	6
Fig. 1.5 Fundamental Diagrams, traffic data from location 3.....	7
Fig. 1.6 Kerner’s three-phase traffic flow theory	8
Fig. 1.7 Four traffic states in the flow-occupancy diagram	9
Fig. 1.8 Plot of traffic volume, ACF, and spectrum	10
Fig. 1.9 Differenced time series of traffic volume, and speed.....	12
Fig. 1.10 Zivot test for unknown structural change in traffic speed time series.....	14
Fig. 1.11 Autocorrelogram (ACF) of traffic volume	15
Fig. 1.12 Cross-correlograms of multivariate traffic time series	16
Fig. 2.1 Illustration of cell transmission model	24
Fig. 3.1 3D scatter plot of differenced volume, speed, and occupancy, location 3	43
Fig. 3.2 Fundamental relationships between differenced traffic variables, location 3	44
Fig. 4.1 Asymptotic threshold distribution estimated by bootstrap method	55
Fig. 4.2 Residuals, autocorrelogram, and p -values of Ljung-Box, Location 3.....	64
Fig. 4.3 MLP Feed Forward Neural Network Architecture (Beck, 2014).....	65
Fig. 4.4 Choose hidden layer neurons and training iterations, $q - v$, location 2.....	69
Fig. 4.5 Architecture of SVR and related kernel method (Scholkopf and Smola, 2002)	73
Fig. 4.6 TS-TVEC model fit and one-step-ahead rolling prediction for 48 hours, location 1	77
Fig. 4.7 MLP model fit and one-step-ahead rolling prediction for 48 hours, location 1	78
Fig. 4.8 SVR model fit and one-step-ahead rolling prediction for 48 hours, location 1.....	79
Fig. 5.1 the GTA 400 series of Highways and data collecting stations (1)	81
Fig. 5.2 the GTA 400 series of Highways and data collecting stations (2)	82
Fig. 5.3 typical traffic time series from the data set (401DE0100DEC, 02-03/2014)	85
Fig. 5.4 Bootstrap density distribution of Hansen and Seo (2002) test	87
Fig. 5.5 Structural break test with 200 points of speed series.....	88
Fig. 5.6 Structural break test with 200 points of occupancy series.....	88
Fig. 5.7 401.EB.Exp.1, Hwy400 Basket Wave, 401DW0030DEE	94

Fig. 5.8 401.EB.Col.7, Bayview Ave. Leslie St., 401DE0180DEC	95
Fig. 5.9 401.WB.Exp.3, Allen Rd. @ Hwy401, 401DE0060DWE.....	96
Fig. 5.10 401.WB.Col.3, Keele St. Dufferin St., 401DE0030DWC.....	97
Fig. 5.11 401.WB.Col.5, Bathurst St. Avenue Rd., 401DE0090DWC	98
Fig. 5.12 400.NB.2, Finch Ave. W. Steeles Ave. W., 400DN0080DNS.....	99
Fig. 5.13 404.NB.1, Sheppard Ave. E. Finch Ave. E., 404DN0010DNS	100
Fig. 5.14 404.SB.1, Sheppard Ave. E. Finch Ave. E., 404DN0010DSS.....	101

Chapter 1

Introduction

1.1 Introduction

Traffic state prediction is an essential component of Intelligent Transportation Systems (ITS) and applications. With an accurate prediction of traffic state, we are able to use Advanced Traffic Management System (ATMS) and Advanced Traveller Information System (ATIS) to effectively manage traffic flow with proactive dynamic traffic control and traveler's route guidance.

It is often debatable whether to increase capacity by construction of new roads with enormous investments or improve utilization of the existing road network with intelligent traffic control strategy. The latter is preferable when adding more roads as a measure of capacity expansion is infeasible in high density urban areas due to limited space and the prohibitive cost. However, the success of proactive dynamic traffic control depends on the premise that accurate prediction of traffic state is available and reliable.

Traffic state estimation or prediction is to estimate, or forecast, macroscopic traffic quantities including flow, density, and speed at critical locations on networks of transportation infrastructure. Many methods for traffic state estimation or prediction are investigated in the literature. Although these studies in the literature constitute a rich body of knowledge to build upon and are indeed useful, ample room exists for improvement. None of the existing methods and models comprehensively captures all at once the challenges exhibited in modelling and forecasting traffic flow due to its stochastic and dynamical properties, and some special statistical characteristics as will be shown in this thesis. In this research, we developed an alternative interdisciplinary methodology and applications that show both theoretical and practical significance in traffic state prediction.

Our ultimate goal is to contribute to bridging the gap between infrastructure supply and traffic demand, maximize capacity utilization of existing infrastructure, timely formulate traffic management solutions to mitigate traffic congestion, and reduce travel time in a complex urban environment.

1.2 Challenges of traffic modelling and forecasting

The evolution of highway traffic state is a dynamic and stochastic process (Jia et al., 2010) that is predictable. However, the challenges that arise in traffic state estimation or prediction primarily stem from the autonomous and interactive dynamics of stochastic traffic variables and their intricate statistical characteristics. Hence, the state prediction has to consider combinatorial factors including the fundamental relation of traffic variables, autocorrelation and multivariate cross-correlation over time and space, multiple traffic states, seasonality, non-stationarity, cointegration of multivariate traffic time series, and unknown structural change in the time series that leads to non-linearity, etc. Details are elaborated in the following sections.

1.2.1 Data source

In order to study the characteristics of traffic processes, the time series of traffic volume, speed, and occupancy are collected from various locations on the highway system in the Greater Toronto Area (GTA), Ontario, Canada. A data map in Fig.1.1 shows three groups of locations where traffic data are collected. These locations include Highway 400 northbound at north of Sheppard Avenue, Highway 401 eastbound expressway at east of Highway 400, and Highway 401 eastbound expressway at east of Yonge Street. The locations of prediction are indicated with red dots, and their adjacent upstream and downstream loop detectors are indicated with blue dots. The traffic data are collected by the Ministry of Transportation Ontario (MTO) and provided to researchers at the ITS (Intelligent Transportation Systems) Center of the University of Toronto. From those nine loop detectors in Fig.1.1, twenty-seven hourly time series, including traffic volume, speed, and occupancy, are collected and analyzed in our empirical study, and presented in chapter 1 and 4. These data sets are used for testing the effectiveness and robustness of the new statistical model. Later in chapter 5, we expand the scope to test the new models on a larger freeway network (Fig. 5.1 and 5.2).



Fig. 1.1 Data map for empirical study

1.2.2 Stochastic and dynamic traffic variables

- Stochasticity

Definition: A stochastic process y on a set X is a random quantity indexed by $x \in X$.

This means that for every x , we get a random quantity $y(x)$ taking values in \mathbf{R} , or more generally, in a set \mathfrak{R} . A stochastic process is characterized by the joint probability distributions of y on arbitrary finite subsets of X ; in other words, of $(y(x_1), \dots, y(x_m))$.

(Scholkopf and Smola, 2002)

A traffic process is a stochastic process that is predictable. It is a non-deterministic process that varies over time, and can be expressed as Eq. (1.1).

$$\{y(s,t), s \in \Omega, t \in T\} \quad (1.1)$$

where $y(\cdot, t)$ is a random variable on the sample space Ω for each time $t \in T$. s denotes a location of transportation network. A sequence of realization of this random process is given by

$y(s, \cdot)$ for each $s \in \Omega$ at each time $t \in T$. In contrast to a chaotic process such as Brownian motion or white noise, a stochastic process is predictable, because a probability distribution function with finite variance usually can be drawn from the successive random quantities generated from $y(s, t)$ where the random quantities are not independent. However, functions of a chaotic process are so erratic that no information is contained in $x(t)$ about the value $x(t + \delta)$, no matter how small δ is, therefore, a chaotic system is not predictable.

- Dynamics

Traffic dynamics includes autonomous and interactive dynamics. The autonomous dynamics reflect the change of an individual macroscopic traffic variable over time and space. The interactive dynamics indicate the fundamental relations between traffic variables and lead to statistical cross-correlation analysis of multivariate traffic time series.

Autonomous dynamics and stochastic characteristics of traffic variables can be easily observed by the mode of variations, their intensive oscillation, magnitude of fluctuation, extreme values, outliers, trend, and cyclical pattern from the plots of traffic flow, speed, and occupancy time series. These characteristics can be observed in both hourly and a finer resolution traffic profile as shown in Fig.1.2 and Fig.1.3. The interactive dynamics of traffic variables are reflected by the 3D scatter plots of flow, speed and occupancy in Fig.1.4 which implies the intrinsic connections between traffic variables.

- Fundamental relations between macroscopic traffic variables

Traffic flow Q (veh/hr), speed V (km/hr), and density ρ (veh/km) are three major variables used to describe the state of traffic flow at a macroscopic level. These traffic variables constitute three pairs of fundamental relationships which are revealed by the Fundamental diagram. Fig. 1.5 shows a plot of the Fundamental Diagram using the samples from location 3. Since the density of traffic flow is scarcely available from a direct survey, sensor occupancy o is usually used instead. In traffic theory, the Greenshields (1935) or Greenberg (1959) models are commonly used to describe these relationships that are subject to the hydrodynamic equation of volume equals multiplication of speed and density (i.e. $Q = V \cdot \rho$).

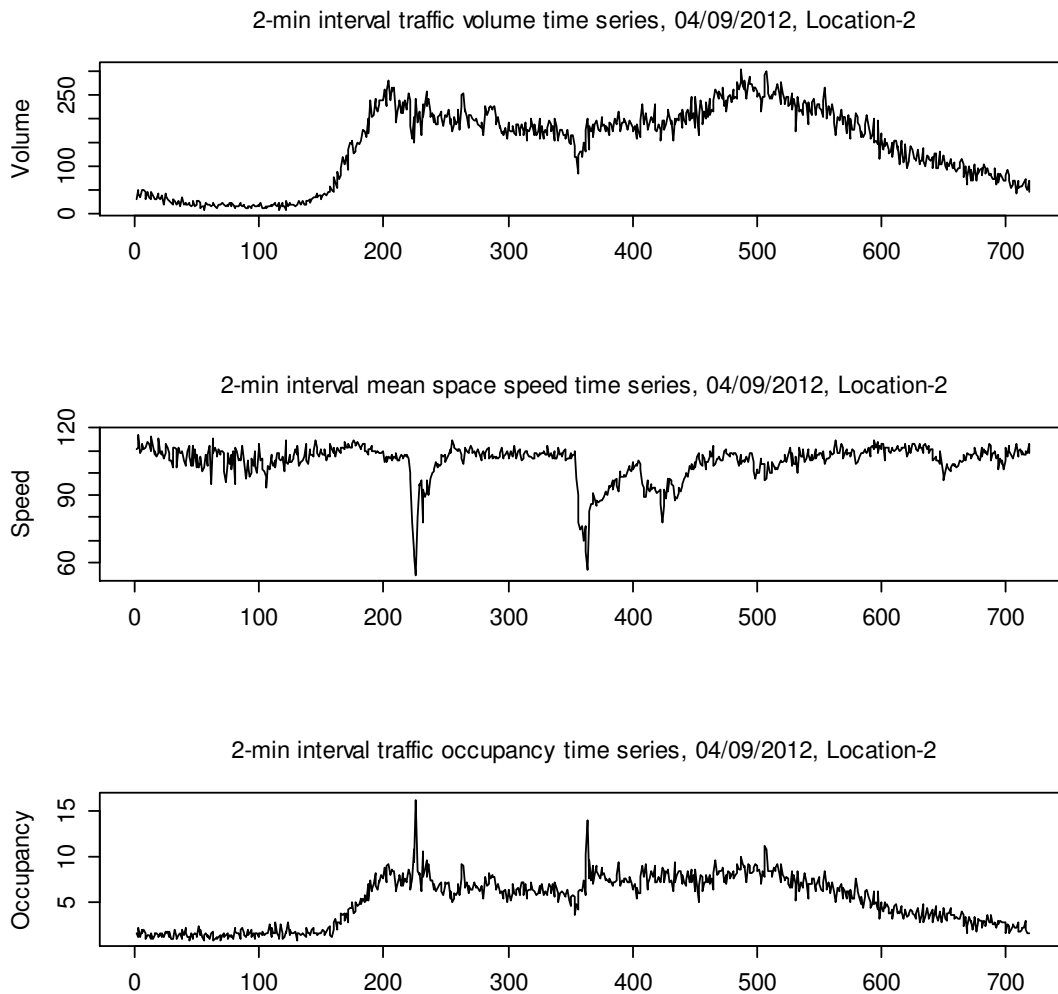


Fig. 1.2 Profile of daily traffic volume, speed, and occupancy

The Fundamental Diagram such as Fig.1.5 can interpret not only the long run relationship of traffic variables, but also show the traffic states. In Fig.1.5, 13% occupancy is a critical point that splits traffic flow between free flow and congested states. The diagram of speed vs. occupancy shows that traffic flow is able to maintain the posted speed or above when occupancy is less than 13%. In a free flow state, as density increases, speed slightly drops but not significantly. However, both diagrams of volume vs. occupancy and speed vs. volume show that volume would still increase even though the speed of the traffic flow drops slightly until occupancy reaches a critical point. If traffic demand increases further (e.g. during rush hours) such that the occupancy reaches a critical point (13%), the diagram of volume vs. occupancy shows that traffic volume approaches the capacity of the roadway, then becomes unstable. When the occupancy surpasses the critical point ($>13\%$), all diagrams show that both traffic volume and speed would

quickly decline. The traffic flow enters an unstable or mildly congested state. The congestion would deteriorate to a traffic jam state as soon as more vehicles join in the flow. When the occupancy approaches its maximum point, traffic speeds approach zero.

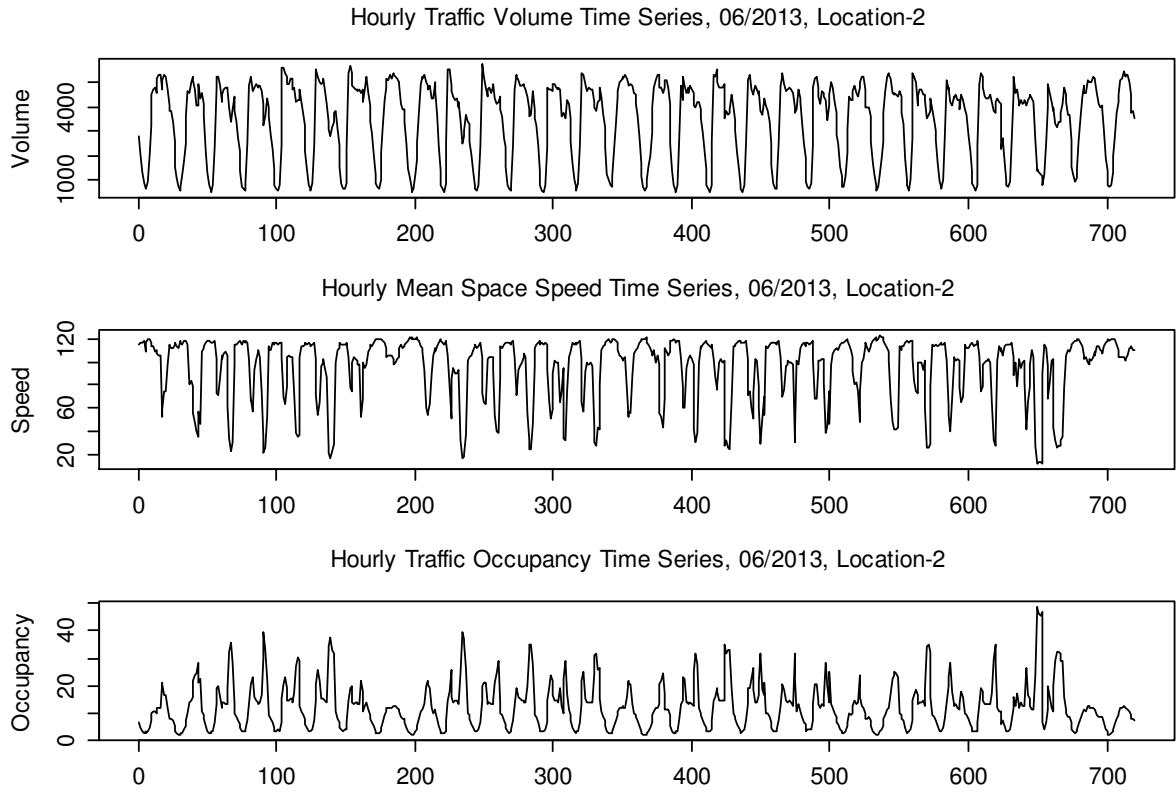


Fig. 1.3 Time series plot of traffic volume, speed, and occupancy

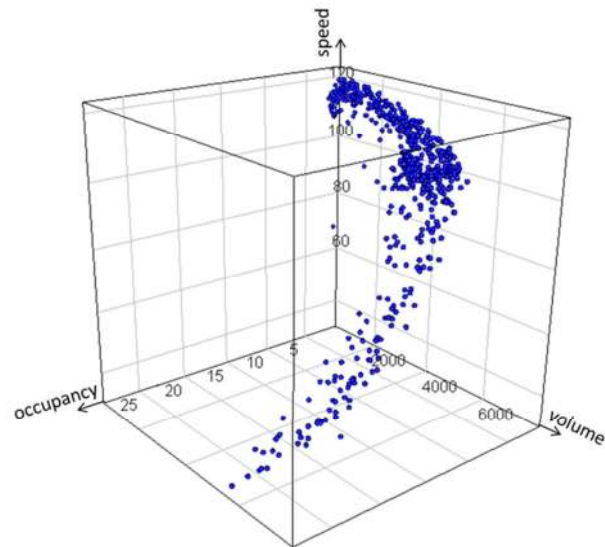


Fig. 1.4 3D scatter plot of traffic volume, speed and occupancy, location 3

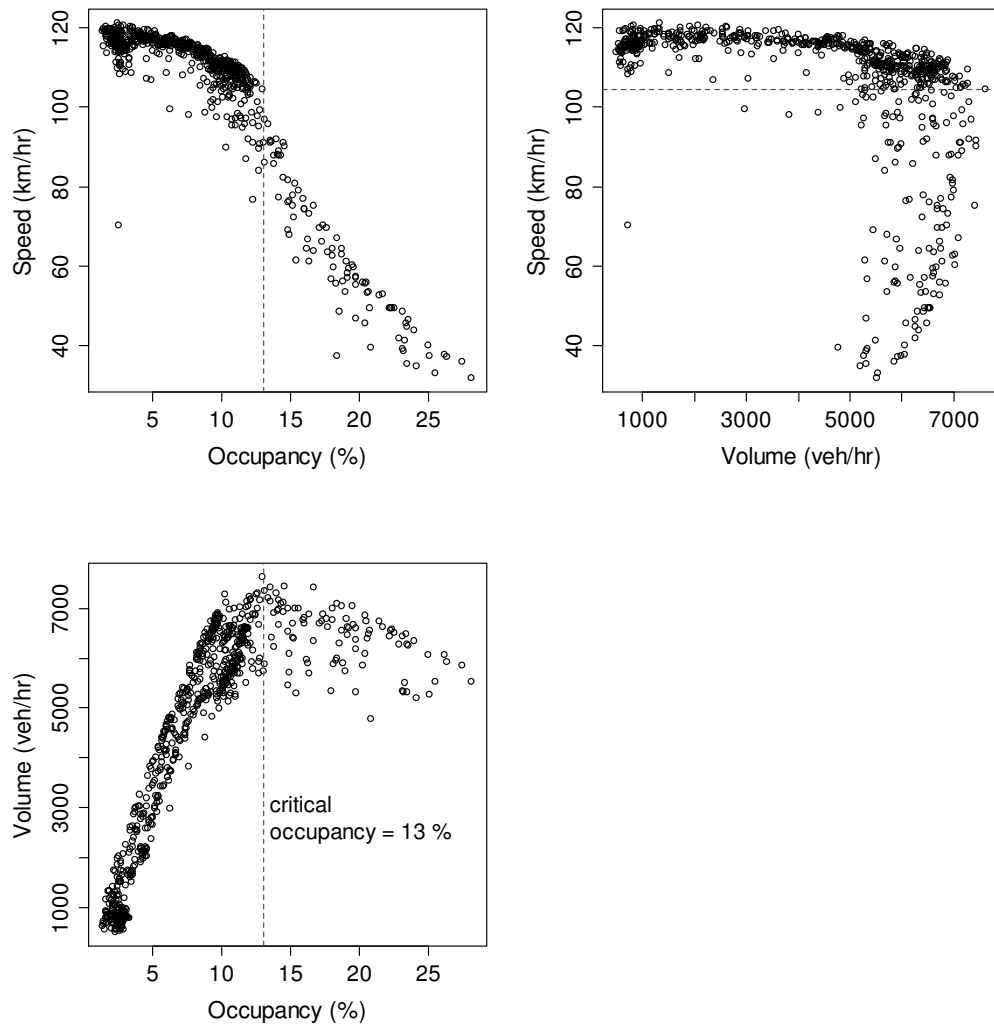


Fig. 1.5 Fundamental Diagrams, traffic data from location 3

- Traffic states

Traffic states are closely associated with the balance between traffic demand and roadway capacity. Traffic flow operates at a free flow state when the demand is lower than the roadway capacity, whereas, traffic flow operates at a congested state when the demand exceeds the roadway capacity. The two traffic states, free flow and congestion, can be easily identified based on the fundamental diagram of a flow-density relationship.

A recent study more comprehensively classifies traffic states into three phases: free traffic, wide moving jams, and synchronized flow (Kerner 2004, 2009, van Hinsbergen et al., 2007).

According to three-phase traffic theory, the congestion state is further divided into two states:

synchronized flow (S) and wide moving jams (J). In the synchronized flow, the speed of vehicles drop significantly, and the density of vehicles increases, therefore, there is no noticeable change in the flow rate observed. At wide moving jams, both flow rate and speed drop significantly, traffic flow becomes relatively uniform than synchronized flow. In order to show Kerner's three-phase traffic flow model, Fig. 1.6 directly cited Figure 12.1(c) on pp. 246 of Kerner's book (Kerner, 2009). Interested readers may also refer to transportation research Circular E-C149 on pp. 22-42 (Greenshields 75 Symposium, 2011) for this theory. However, Treiber, Kesting and Helbing (2010) argued that Kerner's three-phase traffic theory can be reproduced with simple two-phase models with a fundamental diagram if the model parameters are suitably chosen.

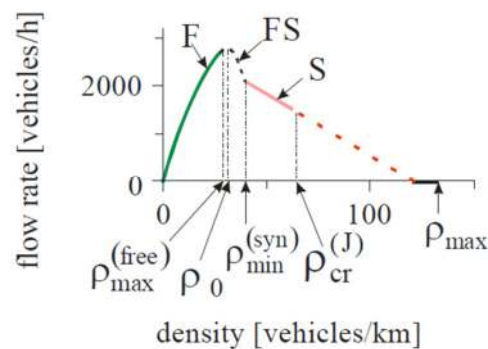


Fig. 1.6 Kerner's three-phase traffic flow theory
that is directly cited from Figure 12.1(c) on pp. 246 of (Kerner, 2009)

Yang and Qiao (1998) use two-layer self-organizing pattern recognition neural network to classify highway traffic states into distinctive cluster centers. The method is found to be more flexible than Highway capacity manual (HCM) and feasible for online application. In addition, based on the neural clustering approach to the analysis of high frequency traffic data collected from arterial roads in Athens, Vlahogianni et al. (2005, 2008a, 2008b) identified four traffic regimes by clustering traffic patterns from the flow-occupancy relationship. These traffic states include free-flow condition where no queue forms, two synchronized conditions that reflect traffic flow evolution near capacity, and congested condition. Two synchronized conditions are classified based on diverging statistical behavior including a stochastic structure where volume highly fluctuates with low occupancy, and a deterministic structure where volume highly fluctuates with sharp increase in occupancy. In order to show the four traffic states classified by

Vlahogianni et al., Fig.1.7 directly cited Figure 3 of the paper authored by Vlahogianni et al. (2005).

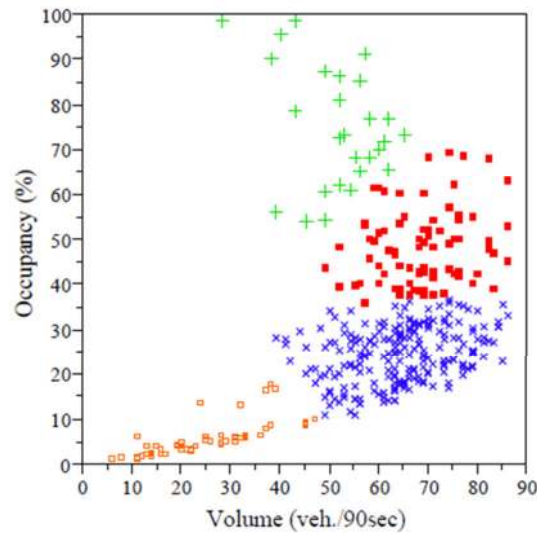


Fig. 1.7 Four traffic states in the flow-occupancy diagram that is directly cited from the paper authored by Vlahogianni et al. (2005).

1.2.3 Statistical characteristics

From a statistical viewpoint, the characteristics of traffic flow may exhibit seasonality, non-stationarity, conditional heteroscedasticity (Anand et al., 2008), unknown structure break, temporal and spatial correlations, and cointegration.

- Seasonality

Seasonality in a time series is a predictable pattern of change that recurs or repeats over a fixed and known period d , where d defines the length of seasonality. It can be expressed as $S_{t+d} = S_t$, where S_t is a function with known period d and referred to as a seasonal component in a time series. A traffic time series that is collected long enough likely shows multi-seasonality including yearly, weekly, and daily pattern. The seasonality is mainly reflected in traffic volume series rather than speed and occupancy time series. Fig.1.8 shows the plot of traffic volume series from location 3, its autocorrelogram (ACF), and spectral density. The definition of autocorrelation and spectral density can be referenced to (Brockwell and Davis, 2002). The repeated pattern in the plot of traffic volume series and the ACF indicate the existence of

seasonality in the time series. Seasonality can also be clarified by frequency analysis with spectral density.

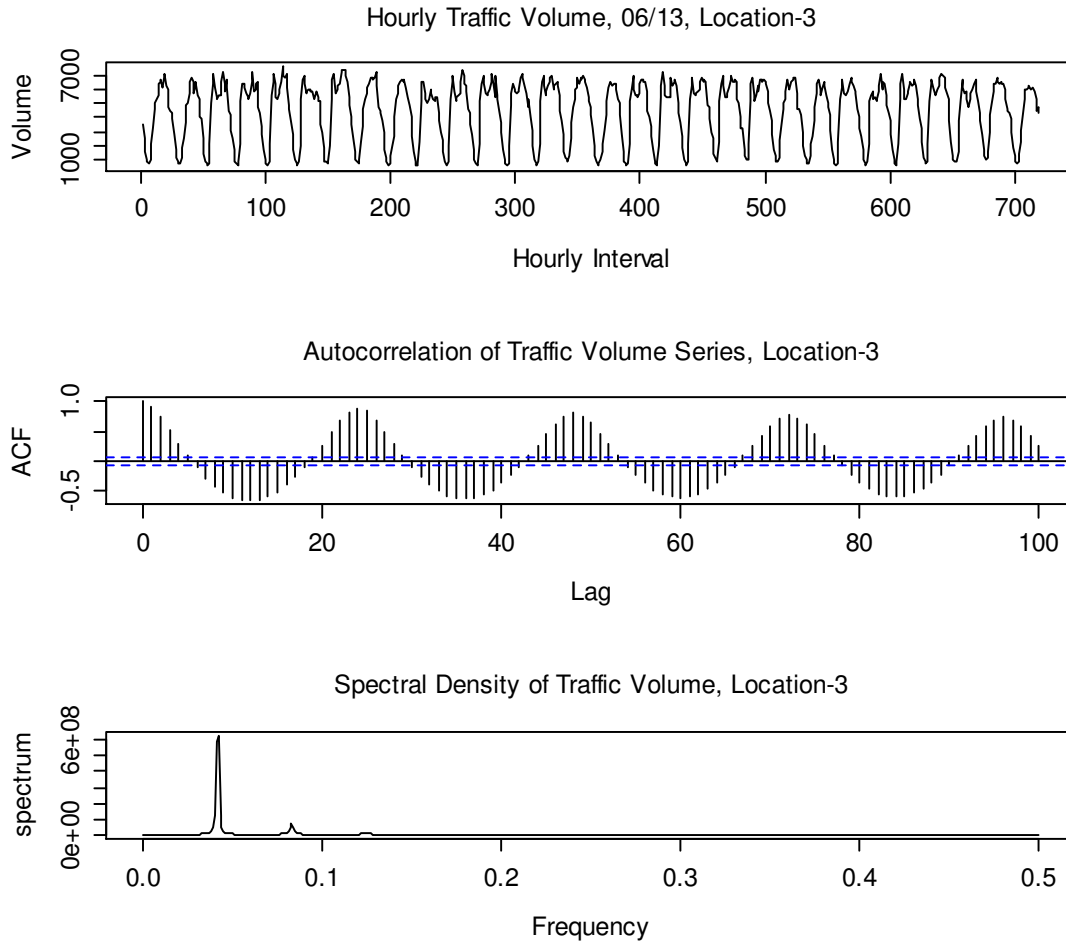


Fig. 1.8 Plot of traffic volume, ACF, and spectrum

- Stationary and non-stationary

A stochastic process y_t is said to be (weakly) stationary if both Eq. (1.2) and Eq. (1.3) hold. Namely, both mean function and covariance function are independent of time t and finite over time. The covariance γ_h in Eq. (1.3) is a function of time lag h which is an integer for short memory time series.

$$E(y_t) = \mu < \infty, \quad \forall t \in \Omega \quad (1.2)$$

$$E[(y_t - \mu)(y_{t-h} - \mu)] = \gamma_h, \quad \forall t, h \in \Omega \quad (1.3)$$

A stationary normal process has a property known as time reversibility, i.e. the joint probability distribution of the process does not change as the process is reversed in time. In general, a stationary process is a statistical equilibrium such that statistical inferences about the stochastic process structure can be easily made (Cryer, 2008).

However, the evolution of traffic state appears to be a non-stationary and time-irreversible process. A typical daily traffic volume profile is analogous to an asymmetric M shape as such shown in Fig.1.2 with a slow rise before the morning peak and gradual decay after the evening peak. Some irregular abrupt declinations with slow rises appear in the speed profile. From the plots, it can be easily identified that traffic process is non-stationary and time-irreversible because it has a non-constant mean function. As well, the process reversed in time does not have the same distribution as the original process (Kamarianakisa et al., 2012).

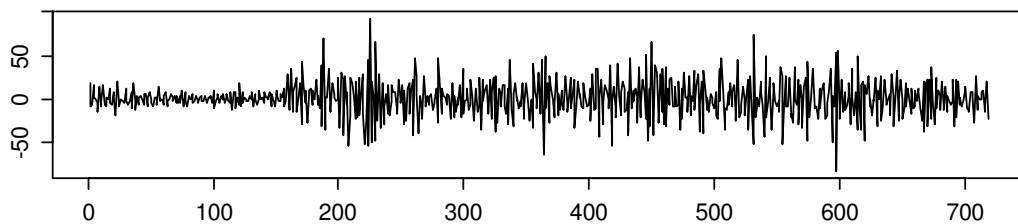
The degree of non-stationarity can be examined by unit root tests. These statistical tests include the Augmented Dickey-Fuller (ADF) (Dickey and Fuller 1979, 1981), DF-GLS (Elliott et al., 1996), and KPSS (Kwiatkowski et al., 1992) and Zivot-Andrews unit root test (Zivot and Andrews, 1992). Table 1 shows the results of four types of unit root tests to check the degree of non-stationarity in the traffic time series. These results indicate that the hypothesis of unit roots cannot be rejected such that the traffic time series from the three locations are non-stationary.

Non-constant variance, known as conditional heteroscedasticity, can be identified by the difference technique that is usually used to remove the trend and seasonal component from a time series. Fig.1.9 shows the differenced time series of traffic volume and speed respectively. It can be easily identified from the plots that the variance of differenced traffic volume and speed are not constant. The magnitude of fluctuation around the mean changes over time.

Table 1.1 Unit Root Test

Unit Root Test (Stationarity Test)						
Traffic Series	Location 1		Location 2		Location 3	
	Test Statistic	p-value	Test Statistic	p-value	Test Statistic	p-value
test 1:						
Augmented Dickey-Fuller (ADF) Test, H0: unit roots, non-stationary Critical values (cv) for a significance level of: 1% cv = -2.58, 5% cv = -1.95, 10% cv = -1.62						
Volume	-0.3495	p > 10%	-0.1751	p > 10%	-0.0985	p > 10%
Speed	-0.0962	p > 10%	0.1628	p > 10%	0.3826	p > 10%
Occupancy	-0.0434	p > 10%	-1.2046	p > 10%	-1.0247	p > 10%
test 2:						
Elliot, Rothenberg and Stock (DF-GLS) Test, H0: unit roots, non-stationary Critical values (cv) for a significance level of: 1% cv = -2.57, 5% cv = -1.94, 10% cv = -1.62						
Volume	-0.4936	p > 10%	-0.4241	p > 10%	-0.2335	p > 10%
Speed	-0.4892	p > 10%	-0.5413	p > 10%	-0.8017	p > 10%
Occupancy	-0.1867	p > 10%	-0.48	p > 10%	-0.6446	p > 10%
test 3:						
KPSS Unit Root Test, H0: stationary Critical values (cv) for a significance level of: 10% cv = 0.119, 5% cv = 0.146, 2.5% cv = 0.176, 1% cv = 0.216						
Volume	0.2436	p < 1%	0.1853	1% < p < 2.5%	0.2197	p < 1%
Speed	0.2718	p < 1%	0.175	2.5% < p < 5%	0.2267	p < 1%
Occupancy	0.2458	p < 1%	0.1772	1% < p < 2.5%	0.2605	p < 1%
test 4:						
Zivot-Andrews Unit Root Test, H0: unit roots in the case of a structural break Critical values (cv) for a significance level of: 1% cv = -5.57, 5% cv = -5.08, 10% cv = -4.82						
Volume	-2.8991	p > 10%	-3.6017	p > 10%	-3.58	p > 10%
Speed	-4.0613	p > 10%	-3.6495	p > 10%	-4.0012	p > 10%
Occupancy	-3.2978	p > 10%	-4.3414	p > 10%	-4.4776	p > 10%

Differenced traffic volume time series



Differenced mean space speed time series

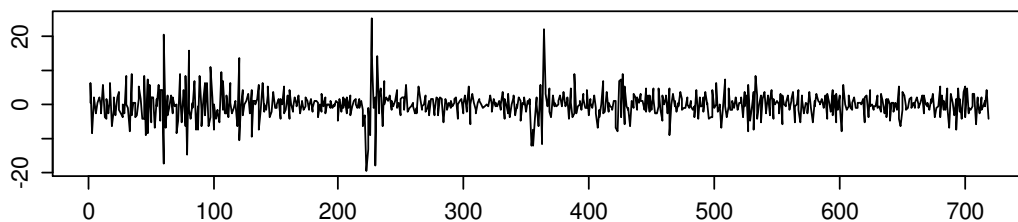


Fig. 1.9 Differenced time series of traffic volume, and speed

- Structural change

Unknown structural change may occur in the first, the second, or higher order moment of a time series. This phenomenon can be visualized by abrupt declination or inclination in the plot of the time series. The point where the structural shift takes place is usually unknown before the change occurs.

Traffic flow is generally classified as either a free flow state or congested state as discussed earlier. In the free flow state, traffic variables exhibit simple symmetric fluctuation around the steady state. In the congested state, asymmetric fluctuation deviates from the steady state. The abrupt declination, or inclination of traffic variables caused by unexpected disturbances to traffic flow, instability, or incidents can be defined as outliers in the context of conventional time series modelling. However, from a viewpoint of traffic flow theory, those outliers cannot be simply removed in traffic modelling because they imply traffic congestion state, which is of primary interest in traffic management. This instability of the data generating process often results in structural changes in most traffic time series. Abrupt structural changes can be easily observed in the plot of speed time series in Fig.1.2, Fig.1.3, and in the plot of the differenced volume and speed in Fig.1.9.

Unknown structural changes can be detected by the Zivot-Andrews unit root test. The null hypothesis H_0 is a unit root process with drift (random walk) that excludes exogenous structural change. The alternative hypothesis is a trend stationary process that allows for a one-time break in the level, the trend, or both. Fig.1.10 shows a Zivot-Andrews unit root test resulted from a daily traffic speed time series. The test statistic equals -5.9528 (critical values: 0.01= -5.57, 0.05= -5.08, 0.1= -4.82). The null hypothesis is rejected at 1% significance in favor of a one-time structural break at position 354 as shown in Fig.1.10. The Zivot-Andrew test estimates the breakpoint to be at a place where the ADF unit root t -test statistic is minimized.

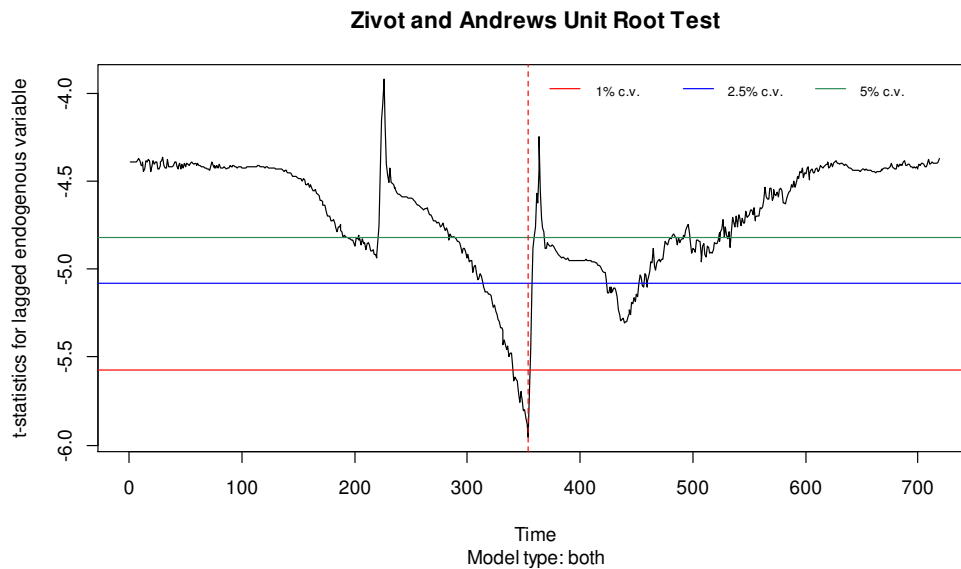


Fig. 1.10 Zivot test for unknown structural change in traffic speed time series

- Nonlinearity

Non-stationarity in the first two order moments and unknown structural change in traffic time series lead to nonlinear model representation because a global linear approximation is often inappropriate. The most common nonlinear models are the threshold autoregressive (TAR) models of Tong (1978, 1983), and the autoregressive conditional heteroscedastic (ARCH) models of Engle (1982) and the generalized ARCH (GARCH) models of Bollerslev (1986). The threshold model is suitable for conditional mean functions, whereas, the ARCH model for conditional variance functions.

- Temporal and spatial correlation

Temporally, traffic process exhibits correlation between time t and $t+h$. The current traffic state has immediate impact on the traffic state at the next time instant at the same location. The impact becomes weaker as the time lag gets further apart, also known as ergodicity in the context of time series analysis. In other words, the current traffic state is closely related to the traffic state in the most recent past.

Autocovariance and autocorrelation reflect a linear dependence between two random variables at different time points within a single time series. In the literature, there are about thirteen different ways to quantify the correlation coefficient (Rodgers and Nicewander, 1988). The formula

commonly adopted is based on Pearson's product moment correlation coefficient (PMCC) (Pearson, 1920). The autocovariance function (ACVF) and the autocorrelation function (ACF) of time series $\{X_t\}$ at lag h are shown in Eq. (1.4) and Eq. (1.5) respectively.

$$\gamma_x(h) = Cov(X_{t+h}, X_t) = E[(X_t - \mu_{x,t})(X_{t-h} - \mu_{x,t-h})] \quad (1.4)$$

$$\rho_x(h) \equiv \frac{\gamma_x(h)}{\gamma_x(0)} = Cor(X_{t+h}, X_t) \quad (1.5)$$

The sample autocorrelogram of traffic volume time series in Fig.1.11 provides great insight on temporal autocorrelations.

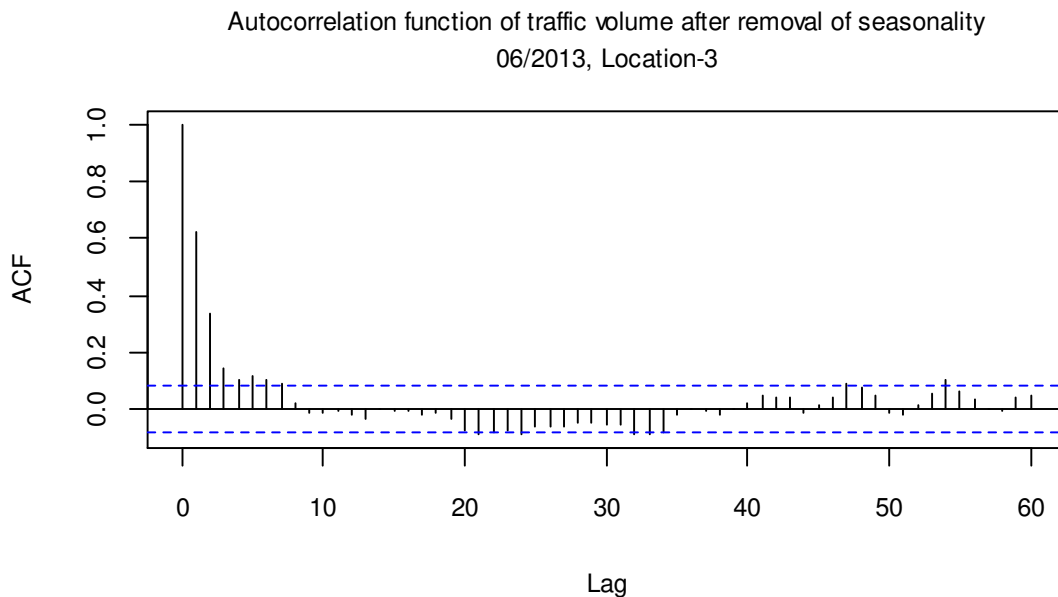


Fig. 1.11 Autocorrelogram (ACF) of traffic volume

Spatially, traffic processes exhibit correlation between the traffic flows at two adjacent locations. Traffic flow at any location of the roadway network may receive impact from its upstream and downstream. The impact gets weaker as the spatial distance between two locations gets greater. The correlation also exists between two traffic variables at the same location such as traffic volume series and speed series. These local and spatial linear dependences between traffic variables and between spatial neighborhoods can be verified by cross-covariance and cross-correlation of multivariate time series.

Given the m -dimension time series vector $\mathbf{X}_t = \{x_{1t}, x_{2t}, \dots, x_{mt}\}$, where each x_{it} ($i = 1, 2, \dots, m$) represents a single time series, the cross-autocovariance function (CACVF) and the cross-autocorrelation function (CACF) of vector \mathbf{X}_t at lag h are shown in Eq. (1.6) and Eq. (1.7) respectively.

$$\mathbf{R}_X(h) = Cov(\mathbf{X}_{t+h}, \mathbf{X}_t) = E[(\mathbf{X}_t - \boldsymbol{\mu}_{X,t})(\mathbf{X}_{t-h} - \boldsymbol{\mu}_{X,t-h})] \quad (1.6)$$

$$\boldsymbol{\rho}_X(h) \equiv \frac{\mathbf{R}_X(h)}{\mathbf{R}_X(0)} = Cor(\mathbf{X}_{t+h}, \mathbf{X}_t) \quad (1.7)$$

Fig.1.12 shows the sample cross-correlograms of two pairs of traffic time series. The top cross-correlogram reflects the cross-correlation between the traffic volume and speed at the same location; the bottom correlogram indicates the spatial cross-correlation between the upstream traffic volume and the volume at the site of prediction.

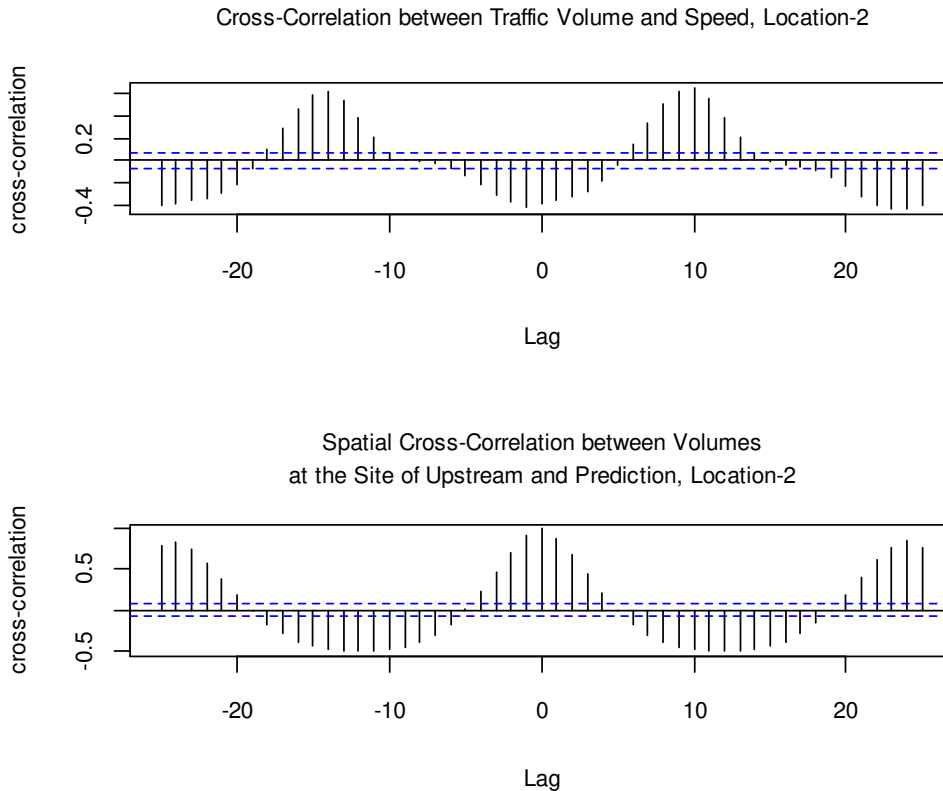


Fig. 1.12 Cross-correlograms of multivariate traffic time series

- Integration and cointegration

In this research, a key finding is that cointegration exists between macroscopic traffic variables. Loosely speaking here, a non-stationary time series variable x becomes stationary after taking d times difference, then x is said to be integrated order d , denoted as $x \sim I(d)$. Moreover, a linear combination of non-stationary time series variables becomes stationary or has their integration order reduced, then the variables are said to be cointegrated. Their precise definition, resulting model, and statistical test are presented in Chapter 3 and Chapter 4.

1.3 Contributions of the research

In order to model and forecast traffic state more accurately using advanced time series techniques, all the aforementioned challenges presented in section 1.2 including stochasticity and dynamics of traffic processes, the fundamental relation of traffic variables, multiple traffic states, seasonality, non-stationarity, serial and cross-section dependence over time and space, cointegration, and unknown structural change in traffic time series have to be taken into account and captured during the course of modelling and forecasting. While many modelling methods and models have been developed during the last two decades in this domain, there is no single time series model available in the literature that can incorporate these factors all at once. It is therefore natural and intuitive to seek and devise a model structure that is able to concurrently take care of all these challenges and factors for network-wide application. Following this train of thought, the time-space threshold vector error correction (TS-TVEC) model is proposed for short term (hourly) traffic state prediction. The details of TS-TVEC model and theoretical basis are elaborated in chapter 3. The experiments and practice are shown in chapters 4 and 5 respectively. The contributions and potential impact of the TS-TVEC model include:

1. As cointegration is found to exist between traffic variables, the theory and method of cointegration with the vector error correction mechanism are employed, for the first time, in the general design of the proposed statistical model TS-TVEC for short term traffic state prediction that deals with multivariate traffic time series.
2. A strong inherent connection between error correction model and traffic flow theory is revealed through the transformation of the well-known Fundamental Traffic Diagrams.

The piecewise linear relationship among the differenced traffic variables justifies the appropriateness of using an error correction model to reveal the dynamics between traffic variables.

3. A threshold regime switching framework with a threshold identification mechanism is implemented to overcome any unknown structural changes in non-stationary traffic time series. The regime switching structure enables TS-TVEC to capture symmetrical and asymmetrical deviation, multiple traffic states, and conditional heteroscedasticity of traffic variables.
4. Spatial and temporal cross-correlated information is incorporated with piecewise linear vector error correction model.
5. MSE (mean squared error) of prediction with cross validation method is employed as a criterion in the TS-TVEC model selection.
6. The empirical study in chapter 4 shows that the TS-TVEC model outperformed MLP Neural Network and Support Vector Regression in forecasting traffic volume with approximately 6% of improvement on average. In chapter 5, the larger scale applications are performed to 35 freeway locations with approximately 315 time series. The mean absolute percentage error (MAPE) in traffic volume prediction is between 4.66% and 23.03% with median 7.45%, the coefficient of variation is between 5.87% and 41.29% with median 9.29%, and the mean square error is between 18864 and 296926 with median 109632.

1.4 Scope of the research

The modelling and forecasting efforts in this research focus on macroscopic traffic quantities including flow, speed, and occupancy which are important indicators of traffic state on a roadway. There is no rigorous definition for the “short-term” prediction which depends on a specific application. Most short term traffic state predictions in literature are concerned with an hourly interval or less (van Hinsbergen et al., 2007). This research focuses on hourly interval one-step-ahead rolling prediction.

Advanced time series techniques are adopted for short term traffic state prediction at data collecting locations. The proposed nonlinear TS-TVEC model employs a threshold regime-switching structure to deal with multi-state and non-stationary multivariate traffic time series with unknown possible structural breaks. The vector error correction model as a piecewise linear model is embedded in the regime-switching structure to deal with cointegration of traffic variables. The proposed methodology is tested and applied to the 400 series of Highways in the Greater Toronto Area (GTA). The details of its experiment and large-scale application are presented in chapters 4 and 5.

The indicators used to measure the accuracy of predictions include MSE (mean squared error), MAPE (mean absolute percentage error) and Coefficient of Variation. MSE is a primary index assessing prediction capacity. It is the second moment of prediction error and thus incorporates both the variance of the predictive model and its bias. MAPE measures the average absolute error in percentage, and the Coefficient of Variation equals to ratio of the standard deviation of prediction to the mean. It measures the magnitude of prediction error per unit.

The next sections are organized in the following order: literature review in chapter 2, theoretical elaboration and model specification in chapter 3, empirical study in chapter 4, large scale implementation in chapter 5, and discussion and conclusion in chapter 6.

Chapter 2

Literature review

Literatures in the domain of short term traffic state estimation and prediction show that many methods have been attempted in the past few decades. There are different ways to categorize the methods of prediction. From a methodological perspective, this chapter reviews those methods based on two classes of research works: traffic state estimation and traffic state prediction.

In our review, traffic state estimation refers to methods that are mainly based on traffic flow theory to approximate the traffic situation at the middle section of a road stretch from boundary conditions. Traffic state prediction refers to methods that forecast the traffic situation at a given location of a highway stretch based on historical time series data collected from the loop detector at the location of prediction and its adjacent upstream and downstream locations.

A major part of traffic flow theory that associates with traffic state estimation is second order macroscopic traffic flow models. Those models are developed from kinematic physics, fluid or gas-kinetic, and wave theory among others to describe traffic phenomena. They are initially taken the form of partial differential equations (PDE) known as continuum second order models. The numerical solution to continuum second order models lies in the space and time discretization. Discrete second order models with a supply-demand method results in application of cell transmission model (CTM) to traffic state estimation (Daganzo 1993, 1994, 1995a).

On the other hand, many methodologies developed in the past can be aligned with the stream of traffic state prediction. Those methods include, but are not limited to, Neural Networks, Kalman Filter, Support Vector Regression, Nonparametric regression, and time series, etc. Those methods are developed on the basis of statistical error theory where an optimal solution is found by minimizing error between the model predictions and observed values.

This chapter reviews second order macroscopic traffic flow models for traffic state estimation, and various methodologies for traffic state prediction.

2.1 Traffic state estimation

Model-based traffic state estimation is to estimate macroscopic traffic quantities for any intermediate point on a freeway or arterial segment based on theoretical traffic flow models and boundary conditions. Many researchers use the continuum section-based or discrete cell-based modelling approach. Representative works include the Cell Transmission Model (CTM) (Daganzo 1993, 1994, 1995a), variational kinematic waves (Daganzo 2005), CTM-based traffic state estimation models (Muñoz et al., 2003), second-order traffic flow model with Kalman filter (Nanthawichit et al. 2003, Wang and Papageorgiou 2005, 2008), CTM-based second-order traffic flow model with particle filtering (Mihaylova et al., 2007), the LWR partial differential equation (Lighthill and Whitham 1955, Richards 1956) with the Lagrangian measurements (Herrera and Bayen 2010), Newell's simplified kinematic wave model (Newell, 1993), stochastic Newell's three-detector method (Laval et al., 2012, Deng et al., 2013), etc.

Second order macroscopic traffic flow models as the most relevant models and a major approach to traffic state estimation are briefly reviewed in this section. Their application to traffic state estimation and limitation are discussed as well. There is a great body of literature in traffic flow theory that is related to second order macroscopic traffic flow models. Interested readers may refer to relevant articles and books (TRB report 165-1975, Papageorgiou 1983, Leutzbach 1988, Hoogendoorn 1999, Garavello and Piccoli 2006, Palais 2008, and Treiber and Kesting 2013, etc.) for more details.

2.1.1 Second order macroscopic traffic flow models

The LWR model and Payne's model (Payne 1971) are established on the assumption of a simple homogeneous road section. Hence, their application is limited. In the context of inhomogeneous road section where merging or diverging zones of on-and-off ramps are present, or the number of lanes is varying due to lane closure or open, Treiber and Kesting (2013) derived more generic macroscopic traffic dynamic models including continuity equation, acceleration equation, and flow-conserving equation. The three equations are able to describe dynamics of traffic speed, density, and flow in a more general road situation where different type of bottleneck can be included.

Let us assume that the traffic flow, density and speed are differentiable functions of time and space denoted by $Q(x,t)$, $\rho(x,t)$ and $V(x,t)$ respectively on the (x,t) plane where x denotes the distance in the driving direction along road stretch and t denotes time. The continuity equation and flow-conserving equation can be written compactly in the vector form as Eq. (2.1).

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{U})}{\partial x} = \mathbf{S}(\mathbf{U}) \quad (2.1)$$

where

$$\mathbf{U} = \begin{pmatrix} \rho \\ Q \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} Q - D \frac{\partial \rho}{\partial x} \\ \frac{Q^2}{\rho} - \left(\frac{Q \cdot D}{\rho} \right) \frac{\partial \rho}{\partial x} + P - \eta \frac{\partial}{\partial x} \left(\frac{Q}{\rho} \right) \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} g_{ramp} - \frac{Q}{I} \frac{dI}{dx} \\ \frac{\rho \cdot V_e - Q}{\tau} + S_{inh} \end{pmatrix}$$

In Eq. (2.1), \mathbf{U} denotes the components of the traffic state, \mathbf{F} denotes the associated density and flow, and \mathbf{S} denotes the sources of road inhomogeneities. D and η are coefficients, $\frac{\partial}{\partial x} \left(D \frac{\partial \rho}{\partial x} \right)$ denotes a diffusion term, g_{ramp} is the generation (dissipation) rate in vehicles from on-ramps (off-ramps) per unit time per unit length, I denotes number of lanes at merging and diverging zone that is a continuous function of road spatial coordinate x , $\frac{\partial P}{\partial x}$ denotes traffic pressure,

$S_{inh} = \frac{Q \cdot g_{ramp}}{\rho} - \frac{Q^2}{\rho \cdot I} \frac{dI}{dx} + \rho \cdot A_{ramp}$ is the source term associated with road inhomogeneities,

$A_{ramp} = \frac{(V_{ramp} - V) \cdot Q_{ramp}}{\rho \cdot I \cdot L}$ denotes the rate of the mean speed change at the on-ramp merging and off-ramp diverging zone, L is the length of merging and diverging zone.

The acceleration equation takes the form of Eq. (2.2).

$$\frac{\partial V}{\partial t} + \overbrace{V \frac{\partial V}{\partial x}}^{convection} = \overbrace{\frac{V_e(\rho, V, \rho_a, V_a) - V}{\tau}}^{relaxation} - \underbrace{\frac{1}{\rho} \frac{\partial P}{\partial x}}_{pressure} + \overbrace{\frac{1}{\rho} \frac{\partial}{\partial x} \left(\eta \frac{\partial V}{\partial x} \right)}^{anticipation} + A_{ramp} \quad (2.2)$$

The acceleration equation consists of three components including convective rate of change, relaxation source, and the spatial anticipation where $\frac{\partial}{\partial x} \left(\eta \frac{\partial V}{\partial x} \right)$ denotes diffusion term.

If the model parameters are properly adjusted, the acceleration equation and continuity equation can be respectively reduced to Payne's model (Payne 1971) and LWR model (Lighthill and Whitham 1955, Richards 1956). Interested readers may refer to Treiber and Kesting (2013) for detailed derivation of those equations.

Mathematically speaking, dynamics of macroscopic traffic variables can be described with both differential equation and difference equation. Differential equation is to emphasize the trend and rate of change, whereas, difference equation is to focus on the relation of the current and past time. Therefore, a discrete version of second order models is suitable for traffic state forecasting. The continuum second-order macroscopic traffic models in the form of hyperbolic partial differential equations are barely useful for the traffic state estimation unless they are discretized in time and space.

The method of choice for time and space discretization is explicit finite difference that is a numerical integration scheme based on the Taylor series. Due to its asymmetric finite-difference approximation for gradients, the discretization is classified as forward finite difference in time and backward in space denoted as FTBS (also known as upwind method), and forward finite difference in time and forward in space denoted as FTFS (also known as downwind method). The discretization divides the highway stretch into cells of length Δx , and time into time steps with constant intervals Δt . The partial differential equations are approximately decomposed into groups of N difference equations. The discrete speed model is presented in Eq. (2.3). The discrete density and flow model are given in Eq. (2.4).

$$V_j^{n+1} = V_j^n + \frac{\Delta t}{\tau} (V_e(\rho_j^n) - V_j^n) - \frac{\Delta t}{\Delta x} V_j^n (V_j^n - V_{j-1}^n) - \left(\frac{\Delta t}{\rho_j^n + \chi} \right) \left(\frac{P_{j+1}^n - P_j^n}{\Delta x} + \frac{\eta(V_{j+1}^n - 2V_j^n + V_{j-1}^n)}{\Delta x^2} \right) + \Delta t \cdot A_{ramp} \quad (2.3)$$

$$U_j^{n+1} = \begin{cases} U_j^n - \frac{\Delta t}{\Delta x} (F_j^n - F_{j-1}^n) + \Delta t \cdot S_j \\ \text{FTBS (upwind method) if characteristic velocity } c > 0, \text{ free-flow} \\ U_j^n - \frac{\Delta t}{\Delta x} (F_{j+1}^n - F_j^n) + \Delta t \cdot S_j \\ \text{FTFS (downwind method) if characteristic velocity } c < 0, \text{ congested} \end{cases} \quad (2.4)$$

2.1.2 Cell transmission model for traffic state estimation

Daganzo (1993, 1994, and 1995a) among others proposed a major approach to traffic state estimation using discrete second order macroscopic traffic flow models with a supply-demand method that is known as cell transmission model (CTM).

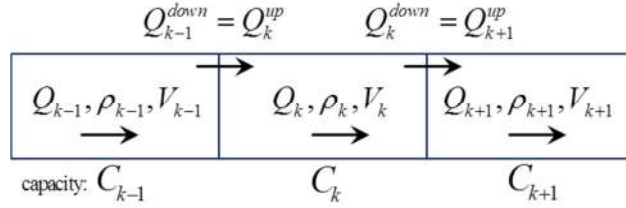


Fig. 2.1 Illustration of cell transmission model

Briefly, a cell transmission model is illustrated in Fig.2.1 where a unidirectional highway stretch is divided into iterated coupled cells of length Δx with the index k increasing in a downstream direction, and time is discretized into time interval Δt . The outflow from an upstream cell is equal to the inflow to the immediate neighbor cell in the downstream during each time interval Δt . The C_k denotes maximum capacity of cell k . The supply and demand method is to determine the value such as $Q_k^{down} = Q_{k+1}^{up}$ at cell boundaries, whereas, the discrete second order macroscopic traffic flow models in Eq. (2.3) and (2.4) are used to update the state variables such as Q_k, ρ_k and V_k in a cell.

The supply and demand method is to determine the flows transmitting cell boundaries from Eq.(2.5) by taking the minimum value of supply S_{k+1} provided by the $(k+1)^{th}$ cell or demand D_k required by the k^{th} cell. The supply provided by cell $(k+1)$ is either limited to its capacity C_{k+1} for free flow traffic situation or given by the actual flow of moving jam in this cell. The demand is either given by the actual outflow from cell k for free flow traffic situation or equal to the maximum value of its capacity for congested situations.

$$Q_k^{down} = Q_{k+1}^{up} = \min(S_{k+1}, D_k) \quad (2.5)$$

where

$$S_{k+1} = \begin{cases} Q_{k+1} & \text{cell (k+1) congested, } \rho_{k+1} > \rho_c \\ C_{k+1} & \text{otherwise} \end{cases}$$

$$D_k = \begin{cases} Q_k & \text{free traffic in cell k, } \rho_k \leq \rho_c \\ C_k & \text{otherwise} \end{cases}$$

Methods of asymmetric finite difference with a dynamically switching mechanism between upwind and downwind method depending on the local characteristic velocities are called Godunov schemes (Treiber and Kesting 2013). The state variables in each cell are updated by discrete second order macroscopic traffic flow models in Eq. (2.3) and (2.4) using Godunov schemes. The switching mechanism depends on the dominant factor among supply and demand. The upwind method is used if demand dominates (i.e. free traffic, characteristic velocity $c > 0$), while the downwind method is employed if supply dominates (i.e. congested traffic, characteristic velocity $c < 0$).

2.1.3 Critiques and limitation

In traffic state estimation with CTM, the forecasting interval is determined by discretized time interval Δt in discrete second order macroscopic traffic flow models. However, the second order models may exhibit instabilities if Δt is improperly chosen for discretization. Hence, Δt is usually restricted by the first and second Courant–Friedrichs–Lévy (CFL) conditions in Eq. (2.6) and Eq.(2.7) respectively (Treiber and Kesting 2013).

$$\Delta t < \frac{\Delta x}{\max|c|} \approx \frac{\Delta x}{V_f} \quad \text{first CFL condition} \quad (2.6)$$

$$\Delta t < \frac{(\Delta x)^2}{2D} \quad \text{second CFL condition} \quad (2.7)$$

where c denotes characteristic velocity; V_f denotes free flow speed, D is coefficient of diffusion term in the second order traffic models. Therefore, the choice of Δt according to Eq.(2.6) and Eq. (2.7) hardly meets the need of hourly interval forecasting in this research. The application of CTM is not only restricted by the first and second CFL conditions, but also suffers from discretization errors, convergence issues, and numerical instabilities (Treiber and Kesting 2013).

It is noteworthy that discrete second order traffic flow models use first order difference equations of density, speed, and flow to predict the traffic state. However, first order difference equations only consider the correlation between current time and one-step time lag, and ignore potential correlation with other time lags. This limitation impacts the accuracy of prediction.

In addition, physical boundary conditions have to be clearly specified when traffic state at a road stretch is simulated by second order traffic flow models. For instance, the traffic demand at the entry and exit of the road stretch, on-and-off ramps, as well as the velocity distribution and variance of the inflow vehicles must be known.

Many critiques arise when traffic flow operations cannot be accurately simulated by second order models. Papageorgious (1998) argued that any macroscopic traffic flow theory will unlikely reach the descriptive accuracy attained in other domains of science such as Newtonian physics. Daganzo (1995b) pointed out that the higher order models exhibit a fundamentally flawed model structure because they can produce negative flows and negative speeds at the tails of congested regions. Aw and Rascle (2000) attempted to remedy second order model by replacing the space derivative with a convective derivative to overcome Daganzo's observations. Helbing (1996, 1997) argues that most macroscopic traffic flow models neglect the finite space requirement of vehicles, the velocity variance, and the finite reaction and braking times of driver-vehicle units, thus, may result in unrealistic jam density.

It is worth mentioning that second order traffic flow models are not established on the basis of statistical error theory, hence, lack capability to reflect statistical characteristics of traffic processes. However, traffic processes are stochastic processes where statistical methods are necessary.

2.2 Traffic state prediction

Research works classified as traffic state prediction focus on forecasting traffic situation at selected locations with local time series data. The methodologies involve multi-disciplinary research such as control theory, artificial intelligence, applied mathematics, statistical science, etc. Over the past two decades, the attention and effort devoted to modelling and forecasting macroscopic traffic states from univariate temporal correlation to multivariate temporal-spatial correlation and from linear to nonlinear form. Those models may be loosely classified as

statistical and non-statistical methods. Some representatives include the class of Time Series models, e.g. seasonal ARIMA (Autoregressive Integrated Moving Average) (Williams et al., 2003) and STARIMA (Space-Time ARIMA) (Kamarianakis and Prastacos, 2003), Kalman Filter State-Space model (Antonioni et al., 2005, Garside 1997, Okutani and Stephanedes, 1984), Neural Network (Qiao et al., 2001, Abdulhai et al., 2002, Zheng et al., 2006), Nonparametric Regression (Chang et al., 2012, Smith et al., 2002), and Support Vector Regression (Wu et al., 2004), Bayesian Network (Castillo et al., 2008).

2.2.1 Neural Network

Neural Network models have found widespread applications in many fields such as data mining, pattern recognition, classification, and prediction when large data sets are available. They are frequently adopted for short-term traffic forecasting (Chan et al., 2012, Abdi et al., 2010, Vlahogianni et al., 2008). Neural Networks have been widely used as a prevailing alternative to time series modelling. The major advantage of the Neural Network model lies in its capacity to deal with multivariate nonlinear problems and its flexibility to easily accommodate any unknown relationship between variables. Neural Networks are good at recognizing the patterns out of training data and stores the patterns in its weight matrix, and relies on said patterns to forecast the future. However, its dilemma is difficulty with interpreting autocorrelations and cross-correlations between variables such that it cannot be used for statistical inference. In the literature, there is debate about its capacity and efficiency to deal with seasonality of time series. Some argue that Neural Networks are able to deal with seasonality of time series (Sharda et al. 1992), but others argue that Neural Networks are not able to deal with seasonality and pre-processing of data is required (Zhang and Qi, 2005).

Neural Network provides a mapping mechanism from the input to the output space with an input and output pairwise data set and a topology structure that consists of multiple layers of neurons and sets of weights indicating the strength of connections between neurons. From a mathematical viewpoint, Neural Network is a nonlinear function $f : R^N \rightarrow R^M$ mapping n dimension real space to m dimension real space. The weight vector is the coefficient that stores mapping information. From a statistical viewpoint, Neural Network is a class of nonlinear regression models without an explicit mathematical expression. Being analogous to coefficient estimation of a regression process, the network training process is to adjust weight vectors with an error

backward propagation method to minimize the error between actual model output and desired output. The Neural Network model is flexible and adaptive in the sense that it relaxes the assumption that the residuals need to be white noise for the standard regression model, and the explicit correlation between pairwise data is not required (Kutner, 2005).

There are three major elements to determine the properties of a Neural Network: (1) network architecture (neurons, layers, and connections), (2) activation function, (3) learning rules for the weight adjustment. Usually, different types of problems require different specifications of Neural Network. Neural Networks developed for the purpose of traffic prediction include BP Neural Networks (BPNN) (Liu et al., 2009, Zhu et al. 2010, Guo and Zhu, 2009, Li et al. 2010), Time Delay Neural Networks (TDNN) (Abdulhai et al., 2002), Elman Neural Network (ENN) (Gao et al., 2007) and Radial Basis Function (RBF) (Yang et al., 2010, Wang and Xiao, 2003). A multilayer perceptron (MLP) Neural Network is set up for traffic state prediction in this research. The MLP Neural Network architecture for traffic state prediction refers to Fig.4.3 in chapter 4.

However, some limitations exist that make it difficult to apply the approach to large-scale network-wide forecast. (1) The application of Neural Network to traffic state prediction is based on the assumption that the historical traffic pattern would repeatedly occur in the future such that the traffic pattern at a future time instance could be matched by a similar one in the historical data set which was used for network training. Hence, large and comprehensive data sets are required for training the network in order to cover all the patterns possible. (2) There lacks systematic ways and theoretical basis to determine optimal architecture of Neural Network including the number of neurons, layers and connections for a specific problem. In fact, no clear and consensus guidelines in the literature are available to determine the number of layers and neurons in the hidden layer. (3) Neural Network parameters are not interpretable and may not provide useful insights in the investigation of causal relationships in traffic dynamics (Kamarianakisa et al. 2012). (4) Statistical diagnostic tests are currently not available for Neural Network models such as the tests for goodness of fit, identification of influential observations, outliers, and significance effect of the various predictors. It is difficult to infer confidence interval. The overparameterization problem will appear if the number of neurons in the hidden layer is more than just a few. This is analogous to keeping insignificant predictors in a linear regression model. Neural Network does not produce unique solutions to the weight vector because there are infinitely many sets of parameters that can lead to the correct model, thus any

particular set of parameters may lose intrinsic meaning. In addition, the problem of being trapped in the local optimum when minimizing network training errors is a challenge of Neural Network (Kutner et al., 2005).

2.2.2 Kalman recursion

Kalman recursion model is a typical version of the Structural State Space model family. It may be classified as Kalman prediction, filtering and smoothing model (Brockwell and Davis, 2002). The original design of Kalman prediction model is intended to forecast unobservable (implicit) system variables by observable (explicit) exogenous variables. It is a one-step-ahead prediction. The Kalman prediction model uses the linear relationship between the exogenous variables and the system variables, and uses forecast errors (innovations) of the observable exogenous variables at the previous step to recursively correct the forecast of the system variables at the current step.

The essence of the Kalman prediction is the same as the vector ARMA (autoregressive moving-average) model (Brockwell and Davis, 2002). The algorithm is analogous to the innovation forecasting algorithm for a ARMA(1,1) process, but replaces innovations of an endogenous variable with an observable exogenous variable. The Kalman Gain coefficient matrix is calculated in a way analogous to the calculation of the innovation term coefficient by the innovation algorithm for the ARMA(1,1) process. Stathopoulos and Karlaftis (2003) reported that the state space model reduced into an ARIMA model in some cases of predictions.

Kalman recursion model consists of state equation (2.8) and observation equation (2.9) (Brockwell and Davis, 2002).

$$X_{t+1} = \Phi_t X_t + V_t \quad (2.8)$$

$$Y_t = H_t X_t + W_t \quad (2.9)$$

where X_t denotes an unobservable system variable, Y_t an observational variable, Φ_t , H_t are coefficient matrices, $V_t \sim WN(0, Q_t)$ is a system error vector with a zero mean and covariance Q_t , $W_t \sim WN(0, R_t)$ is a measurement error vector with a zero mean and covariance R_t , V_t and W_t are linearly independent. The system equation in the Kalman recursion is a vector form of the

AR(1) model and any AR(p) model can be transformed into a vector form of the AR(1) model. The observation equation in the Kalman recursion is equivalent to the innovation part of the ARMA model, which continuously measures model deviations and updates the system equation. Given Y_0, \dots, Y_t , the one-step-ahead prediction is recursively computed by Eq. (2.10), and then corrected by Eq. (2.11).

$$\hat{X}_{t+1/t} = \Phi_{t+1, t} \hat{X}_t \quad (2.10)$$

$$\hat{X}_{t+1} = \hat{X}_{t+1/t} + K_{t+1}(Y_t - H_t \hat{X}_t) \quad (2.11)$$

where $t+1/t$ are the intermediate steps, $\hat{X}_{t+1/t}$ denotes forecast value prior to correction, \hat{X}_{t+1} denotes forecast value posterior to correction, the Kalman Gain matrix K_{t+1} is computed by Eq.(2.12) as to minimize the error covariance P_{t+1} that is calculated by Eq. (2.13).

$$K_{t+1} = P_{t+1/t} H_{t+1}^T (H_{t+1} P_{t+1/t} H_{t+1}^T + R_{t+1})^{-1} \quad (2.12)$$

$$P_{t+1} = (I - K_{t+1} H_{t+1}) P_{t+1/t} \quad (2.13)$$

where

$$P_{t+1/t} = \Phi_{t+1} P_t \Phi_{t+1}^T + Q_{t+1}, \quad X_0 \sim N(\mu_0, P_0), \quad P_0 = E[(X_0 - \mu_0)(X_0 - \mu_0)']$$

X_0 denotes the initial information known about the system at $t = 0$. In the context of traffic state predictions, the Extended-Kalman-Filtering (EKF) method is recently combined with second order macroscopic traffic flow models (Wang et. al, 2008, Bellemans et al., 2006, Tampere and Immers, 2007). This approach requires a complicated transformation to fit traffic variables into the rigid Kalman recursion structure because observable (explicit) traffic variables are also system variables (implicit) themselves. The vector transformation of traffic variables from their explicit to implicit form is barely justifiable from the viewpoint of the time series-based model.

As both the Kalman recursion structure and the second order macroscopic traffic flow models are first order difference equations, it implies that the future traffic state is only correlated with the traffic state at one-step-past time instant and the current measurement. Other possible

correlations between the future state and the past and spatial correlations are not taken into account.

In addition, the Kalman recursion is a linear model. Its prediction performance is compromised from a steady traffic state to a state with significant fluctuations. The shorter the forecast horizon is the less accurate the prediction is, because the nonlinear characteristics of stochastic traffic process become more significant.

2.2.3 Support vector regression

Support Vector Regression (SVR) is the pith of Support Vector Machine (SVM) algorithms. SVM can efficiently perform both linear and nonlinear classification by mapping the input data points into high-dimensional feature spaces through a kernel function. SVM is to construct a set of hyperplanes and find the best hyperplane (called the maximum-margin hyperplane, also known as maximum margin classifiers) that separates the data points in high dimensional feature spaces. Hyperplane is defined by a linear vector function. During the process of searching an optimal hyperplane, the empirical classification error is minimized and the geometric margin of hyperplane is maximized (Moore, 2003).

Recently, researchers take advantage of the regression optimization process SVR out of SVM for traffic prediction. In essence, SVR is to determine a suitable predictive function through a nonlinear regression process with pairs of input and output data in a high dimensional feature space based on quadratic programming optimization techniques. Analogous to the perceptron Neural Network, SVR regression process is a training process where the parameter vector of an optimal hyperplane function are determined with a set of training data by minimizing an overall error objective function. The estimated predictive function is then used for traffic variable prediction (Yang and Lu, 2010). Mathematical details of SVR are presented in section 4.3 in chapter 4.

It is reported that SVR outperforms BPNN and SARIMA for traffic variable prediction (Hong et al., 2010, Theja and Vanajakshi 2010). Similar to Neural Network, one drawback of the SVR is that the model parameters are hardly interpretable.

2.2.4 Nonparametric regression

Parametric methods mean that the model structure is predefined, and only the parameters of the model need to be estimated from data. Nonparametric methods mean that both the model structure and the parameters are determined from data (van Hinsbergen et al., 2007).

Nonparametric regression, i.e. K-Nearest Neighborhood (KNN), is often adopted for clustering and classification. The K-Nearest Neighborhood nonparametric regression is developed based on chaotic system theory rather than stochastic system theory which are the basis of time series models. In their research, Disbro and Frame (1989) argue with some evidence that traffic flow exhibits chaotic behavior. Wang et al. (2005) also argue that traffic system is a nonlinear system and traffic flow data exhibits chaotic properties. Shang et al. (2005) argue that traffic time series has a strong chaotic signature due to the positive largest Lyapunov Exponent and the low correlation dimension estimated using phase space techniques. The KNN nonparametric regression consists of four elements including state space, distance metric, forecast generation method, and neighbor database management (Smith et al., 2002).

With the K-Nearest Neighborhood method for traffic prediction (Chang et al., 2012, Yuan and Wang, 2012), K observations are selected from a historical database based on their nearness to the current observation of traffic variable to form a nearest neighborhood. In order to determine k nearest neighborhood, a state vector has to be defined for the purpose of comparison between the current observation and each historical observation. A typical state vector $x_{j,t}$ can be constructed as Eq. (2.14).

$$x_{j,t} = [Q_{j,t}, Q_{j,t-1}, \dots, Q_{j,t-d}] \quad (2.14)$$

where $j=1, \dots, n$ is the index of historical cases in the state space, $Q_{j,t}$ denotes traffic measurement such as flow rate at current time t in historical case j , d is an integer number of time lags. However, there are no clear rules available to define state vector. The state vector may contain more information other than lagged observations. Nearness can be defined by a distance metric. Among various distance metrics, the Euclidean distance is commonly used to measure the similarity between current traffic observation and historical traffic data in KNN regression

based prediction. Eq. (2.15) shows a typical example to calculate the Euclidean distance based on the state vector defined in Eq. (2.14).

$$ED_j = \frac{1}{d+1} |x_{c,t} - x_{j,t}| = \frac{1}{d+1} \left(\sum_{i=0}^d (Q_{c,t-i} - Q_{j,t-i})^2 \right)^{1/2} \quad (2.15)$$

where c denotes index of the current case to be forecast, $x_{c,t}$ is the current state vector.

The nearest neighbor selection is a process of matching the historical state vector $x_{j,t}$ with the current state vector $x_{c,t}$. After the neighbor search procedure identifies k neighbors, the historical state vectors, their corresponding traffic observations $Q_{i,t+1}$ at time $(t+1)$ and Euclidean distance ED_i , $i=1, \dots, k$ are sorted in order according to their nearness. An example is shown in Table 2.1 based on the state vector defined in Eq. (2.14).

Table 2.1 Selected K -nearest neighbors

1	$Q_{1,t}, Q_{1,t-1}, \dots, Q_{1,t-d}$	$Q_{1,t+1}$	ED_1
\vdots	\vdots	\vdots	\vdots
k	$Q_{k,t}, Q_{k,t-1}, \dots, Q_{k,t-d}$	$Q_{k,t+1}$	ED_k

Given the $Q_{i,t+1}$ and ED_i , $i=1, \dots, k$, the forecast can be calculated for the current state. Many different forms of forecasting equation are attempted in the literature. A typical forecasting equation weighted by inverse of distance is shown as Eq. (2.16) (Smith et al., 2002, Chang et al., 2012).

$$Q_{c,t+1} = \sum_{i=1}^k \frac{Q_{i,t+1}}{ED_i} \bigg/ \sum_{i=1}^k \frac{1}{ED_i} \quad (2.16)$$

where $Q_{c,t+1}$ denotes the forecast value for the current state, $Q_{i,t+1}$, $i=1, \dots, k$ denotes the neighbor historical traffic observations at time $(t+1)$, ED_i , $i=1, \dots, k$ denotes the neighbor Euclidean distance.

Smith, William, and Oswald (2002) investigated the prediction performance of KNN nonparametric regression by comparing with refined heuristic methods and the seasonal ARIMA

model. They concluded that KNN nonparametric regression did not match the performance of the seasonal ARIMA forecasts, thus providing some evidence for the argument that traffic process is more stochastic than chaotic. They also suggested that defining different traffic states may lead to better prediction results.

The efficiency of nonparametric regression is directly dependent on the quality and size of the database. It is desirable to possess a large database including as many likely conditions as possible. However, execution time searching for the nearest neighbor will be compromised as the database expands. This is a significant issue for real-time applications of KNN to traffic flow prediction (Smith and Oswald, 2000).

2.2.5 Time series

Time series researchers believe that the future can be predicted from information hidden in the dust of history. The essence of time series modelling techniques is to perform autocorrelation analysis through covariance structure of time series to establish a finite difference equation where the value of a stochastic process at the current time can be expressed as a linear combination of past observations and unobservable white noise. Hence, given historical data, it is viable to extrapolate the future behavior of a stochastic process based on the relation between the current and past time (Brockwell and Davis, 2002, Cryer, 2008). The parameters of a time series model are interpretable and may reveal insightful causal relationships in the stochastic process (Kamarianakisa et al., 2012).

There are many classes of time series models that have been applied to traffic state predictions, such as ARIMA, SARIMA (Guo, 2005, Williams et al., 1998), ARIMAX (Williams, 1999), VARMA, STARMA (Kamarianakis and Prastacos, 2005, 2003), STAR (Sun and Liu, 2011), and SETAR (Liu et al., 2010, Al-Deek et al., 2001, D'Angelo et al., 1999, Ishak and Al-Deek, 2002). Among others, a few representative time series models are introduced in the next sections.

- ARIMA

ARIMA (Autoregressive Integrated Moving Average) (Box and Jenkins, 1976) was introduced to traffic prediction by Ahmed and Cook (1979) and Levin and Tsao (1980). It has been extensively applied to short term traffic state prediction as a benchmark of univariate time series

model. The specification of an ARIMA(p, d, q) model is shown in Eq. (2.17) (Brockwell and Davis, 2002).

$$\phi(B)(1-B)^d X_t = \theta(B)Z_t \quad (2.17)$$

where X_t is time series variable, $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ are p th and q th-degree polynomials, ϕ_1, \dots, ϕ_p and $\theta_1, \dots, \theta_q$ are coefficients, d denotes the order of difference, B is the backward shift operator, p and q represent the autoregressive and moving average order, respectively.

Prediction performance of ARIMA models for traffic prediction is mixed (Kamarianakis and Prastacos, 2005, Williams, 2001, Smith et al., 2002, Hamed et al., 1995, Lee and Fambro, 1999, Lee and Choi, 1998, Kirby et al., 1997, Nihan, 1980) due to its limitations: (1) it is a linear model, but the traffic evolution is a nonlinear stochastic process; (2) it is unable to reflect spatial correlation of topology structure of a road network because it only copes with a single time series of traffic data from the location of interest; (3) it is unable to capture rapid and sudden fluctuations caused by structural change in time series (Davis et al., 1991, Hamed et al., 1995).

- SARIMA

SARIMA (Seasonal ARIMA) is introduced by Williams et al. (1998) to model seasonality of traffic variables. The results indicated that SARIMA outperforms ARIMA and significantly improved prediction accuracy. The specification of a SARIMA(p, d, q)×(P, D, Q) model is shown in Eq. (2.18) (Brockwell and Davis, 2002).

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t \quad (2.18)$$

where $Y_t = (1-B)^d(1-B^s)^D X_t$, X_t is time series variable, Y_t is transitional variable, $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ are p th and q th-degree polynomials, $\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_P B^{s \cdot P}$ and $\Theta(B^s) = 1 + \Theta_1 B^s + \dots + \Theta_Q B^{s \cdot Q}$ are P th and Q th-degree polynomials, ϕ_1, \dots, ϕ_p , $\theta_1, \dots, \theta_q$, Φ_1, \dots, Φ_P , and $\Theta_1, \dots, \Theta_Q$ are

coefficients, d and D denote the order of ordinary and seasonal differencing, s denotes the length of seasonal period, B is the backward shift operator, p and q represent the autoregressive and moving average order, respectively. P and Q represent the autoregressive and moving average order at the length of seasonal period, respectively.

However, SARIMA is a linear model that deals with traffic time series with temporal correlation techniques. It lacks competence to incorporate spatial correlation and deal with any structural change in time series which leads to nonlinear problems. It does not consider the dynamics between traffic variables.

- STARIMA

STARIMA (Space-Time ARIMA) model is a recently developed time series model for traffic prediction. The advancement of STARIMA is to incorporate both spatial correlations and temporal autocorrelation in one model. It shows a tendency to reflect the spatial dynamics of traffic flow in time series models. It was first proposed by Pfeifer and Deutsch (1980). Spatial correlations are typically analyzed with the VARMA (vector autoregressive moving average) model, which was criticized for having too many parameters. STARIMA is a variant of VARMA such that it is refined to drastically reduce the number of parameters by grouping sets of time series according to their spatial order, i.e. a function of distance to the location under study. A linear combination of grouped time series is nested in the time lags of the ARIMA model.

STARIMA was introduced for traffic flow forecasting by Kamarianakis and Prastacos (2003), and was also investigated most recently by Cheng et al. (2012), Min and Wynter (2011), and Min et al. (2010). The specification of the STARIMA model is shown in Eq. (2.19).

$$(1 - B)^d Z_{i,t} = \sum_{k=1}^p \sum_{h=0}^{m_k} \phi_{k,h} W^{(h)} (1 - B)^d Z_{i,t-k} + \varepsilon_{i,t} + \sum_{k=1}^q \sum_{h=0}^{n_k} \theta_{k,h} W^{(h)} \varepsilon_{i,t-k} \quad (2.19)$$

where

$Z_{i,t}$ traffic variable at time t and location i
 p is autoregressive order
 q is moving average order
 k is time lag,
 h is spatial lag
 m is spatial order of k_{th} autoregressive term
 n is spatial order of k_{th} moving average term
 $\phi_{k,h}$ is the autoregressive parameter at time lag k and spatial lag h
 $\theta_{k,h}$ is the moving average parameter at time lag k and spatial lag h
 $W^{(h)}$ is the $N \times N$ matrix of weights for spatial order h ($W^{(h)} = I$)
 $\varepsilon_{i,t}$ is the white noise vector at time t location i

The success of the model is conditional on the construction of the spatial weight matrix. In a typical scenario, the spatial weight matrix is assumed to be a constant matrix for the same spatial order over time steps (Kamarianakis and Prastacos 2005).

Cheng et al. (2012) provide insight into determining a spatial weight matrix for its spatial order and weights. Cheng et al. (2012) argued that it is not always reliable to use a distance function to determine spatial order. An effective spatial order is dynamic such that it becomes larger under a free flow traffic state and smaller under a congested traffic state. Taking into account link speed, Min and Wynter (2011) construct a set of dynamic spatial weight matrices by using historical data where each spatial weight matrix is associated with a time period (time of day, day of week). A GSTARIMA (Generalized STARIMA) is a more flexible model proposed to capture the dynamics of network process allowing the AR (autoregressive) and MA (moving average) parameters to vary per location, but it has N multiple number of parameters of a regular STARIMA model for a traffic network with N nodes (Min et al., 2010). It is noteworthy that these models are multivariate linear models that cannot deal with nonlinearity in traffic time series.

- Regime Switching

A three-regime SETAR (self-exciting threshold autoregressive) model is employed for modelling and forecasting hourly traffic volume (Liu et al., 2010). The structural change of traffic time series is detected with CUSUM (cumulative sum), and the ARIMA-EM (expectation

maximization) model is built for forecasting by Cetin and Comert (2006). However, these studies focus on univariate traffic time series.

More recently, a temporal regime switching model is proposed for traffic volume prediction (Kamarianakis et al., 2012). Time-based thresholds are predefined and fixed to partition 24 hours into five time regimes: morning and afternoon peak periods, off-peak period, first-half and second-half night periods. A linear autoregressive model is constructed for each time-based regime. Regime switching occurs at fixed times of day. However, traffic demand fluctuation or incidents does not occur at precisely the same time each day of the week or at the same locations. Fixed temporal thresholds might yield significantly biased predictions. In addition, a Neural Network model is used with predetermined uncongested and congested regimes to perform short-term traffic flow and speed prediction (Dunne and Ghosh, 2012). However, these regime switching models for traffic flow prediction have no structural change detection and threshold identification mechanism involved.

2.3 Summary

In this chapter, we reviewed various methodologies for traffic state estimation and prediction. Every method has its own strengths as well as limitations. Comparatively speaking, time series models are promising candidates for short term traffic state prediction because they have more advantages than other methods. In general, time series models have a parsimonious model structure, are well interpretable, and may provide insightful interpretation of traffic dynamics. Time series models are able to capture spatial and temporal correlations of traffic flow evolution and provide statistical inference about the causal relations. Advanced time series analysis techniques can competently handle various challenges of modelling stochastic traffic processes.

Chapter 3 Methodology

3.1 General design of the proposed statistical model

In this research, a time-space threshold vector error correction (TS-TVEC) model is proposed for short term traffic state prediction to concurrently address multi-state, seasonality, non-stationarity, temporal and spatial correlations, and dynamics between traffic variables that lead to the interaction of multivariate traffic time series. In addition, it is found that an inherent connection exists between the error correction model and the transformed fundamental diagrams in macroscopic traffic flow theory. Error correction model is a linear model established on difference space, whereas, the fundamental relations between traffic variables exhibit piecewise linearity when they are transformed to difference space. Meanwhile, dynamics of traffic variables can be reflected in difference space.

The new statistical model (TS-TVEC) is designed based on comprehensive considerations of traffic flow characteristics, structure of traffic data, and integration with the fundamental traffic theory. The design principles have encompassed many factors including long run relationship of traffic variables, multiple traffic states, cointegration of traffic variables, short run dynamics between traffic variables, multivariate time series, seasonality, non-stationarity, unknown structural change in the time series, and temporal and spatial correlations. Combining all these considerations, the newly designed model TS-TVEC has the following capacities and features: (1) cointegration technique with an error correction mechanism is employed to account for long run equilibrium, short run dynamics, and temporal correlation between multivariate traffic series; (2) a threshold regime switching framework is employed to capture multiple traffic states and deal with unknown structural change in traffic time series conditional on the threshold vector detected; (3) the model provides different adjustment speeds to perform error correction within the regime switching framework, thus improving accuracy of prediction; (4) the model is capable of capturing both symmetrical deviation from long run traffic equilibrium under free flow traffic state and asymmetrical sudden shift under traffic congestion state; (5) spatial correlations are incorporated in the piecewise linear error correction model via exogenous terms which take the differenced spatial information between the location of prediction and neighborhood sites.

3.2 Theoretical rationale

The new statistical model TS-TVEC is designed to concurrently use the information of multiple traffic time series and their dynamics to improve the accuracy of traffic prediction. This section elaborates multivariate time series regression theory, method, and modelling technique. The relevant concepts include integration, spurious regression, cointegration of multivariate time series, error correction model, and threshold regime switching framework. The definition of integration is given by Engle and Granger (1987, pp 252) as follows:

Definition *A series with no deterministic component which has a stationary, invertible, ARMA representation after differencing d times, is said to be integrated of order d , denoted $x_t \sim I(d)$.*

3.2.1 Spurious regression and cointegration

Spurious regression is a pitfall that may appear in the multivariate time series regression when all the regressors are not cointegrated. The problem of spurious regression is first pointed out in a simulation study by Granger and Newbold (1974). The traditional statistical inferences do not hold in spurious regression according to the asymptotic distribution theory in Phillips (1986). For instance, let y_{1t} and y_{2t} be two independent $I(1)$ processes that are not cointegrated and $y_{it} = y_{it-1} + \varepsilon_{it}$, where $\varepsilon_{it} \sim WN(0,1)$, $i = 1, 2$. Regressing y_{1t} on y_{2t} gives the fitted model $y_{1t} = \hat{\beta}y_{2t} + \hat{u}_t$. This is a spurious regression and its statistical implications include (1) the residual series \hat{u}_t is usually autocorrelated and its order of integration is not reduced; (2) the true value of the coefficient β is zero, but the estimated coefficient $\hat{\beta}$ does not converge in probability to zero, instead it converges in distribution to random variables with a non-degenerated distribution; (3) let T denote the sample size. The OLS (ordinary least squares) t statistics diverge to $\pm\infty$ as $T \rightarrow \infty$ resulting in the null hypothesis of a zero coefficient wrongly rejected; (4) the coefficient of determination R^2 from the regression converges to unity as $T \rightarrow \infty$ so that the model appears to fit well even with wrong specifications (Zivot and Wang 2006).

Multivariate regression between $I(1)$ variables only makes sense when they are cointegrated. The concept of cointegration is introduced by Granger (1981) and Granger and Weiss (1983), and is precisely defined in Engle and Granger (1987, pp 253) as follows:

Definition *The components of the vector x_t are said to be cointegrated of order d , b , denoted $x_t \sim CI(d,b)$, if (i) all components of x_t are $I(d)$; (ii) there exists a vector $\alpha (\neq 0)$ so that $z_t = \alpha' x_t \sim I(d-b)$, $b > 0$. The vector α is called the cointegrating vector.*

Campbell and Perron (1991) added that if r linearly independent vectors $\alpha_i (i = 1, \dots, r)$ exist, x_t is said to be cointegrated with cointegrating rank r . In other words, the non-stationary time series in vector x_t are cointegrated if a linear combination of the components has its order of integration reduced. In case of $d = 1$, and $b = 1$, a linear combination $z_t \sim I(0)$ is stationary, which implies that the components of x_t do not drift too far apart.

3.2.2 Error correction representation

The relationship between error correction model and cointegration was first pointed out in Granger (1981). The theorem showing precisely that cointegrated time series can be represented by an error correction model is known as the Granger Representation Theorem and has been proven in Engle and Granger (1987), one of the essential papers for which Granger received the Nobel Prize in 2003. We only cite the most important part, statement (4), of the Granger Representation Theorem for the purpose of the research. A full version of the theorem can be found in Engle and Granger (1987, pp 255) for interested readers.

Granger Representation Theorem: *If the components of $N \times 1$ vector x_t are cointegrated with $d = 1$, $b = 1$ and with co-integrating rank r , then:*

Statement (4) There exists an error correction representation with $z_t = \alpha' x_t$ that is a $r \times 1$ vector of stationary random variables:

$$A(B)(1 - B)x_t = -\Phi z_{t-1} + \varepsilon_t \quad (3.1)$$

with $A(0) = I_N$

where $A(B)$ is p^{th} -degree polynomials and B is backshift operator. z_{t-1} are error correction term. ε_t are zero mean white noise with standard deviation σ . However, in order for the symbols to be consistent with the convention in our research, let $A(B) = 1 - \gamma_1 B - \dots - \gamma_p B^p$, $(1 - B)x_t = \nabla x_t$ and $z_{t-1} = ECT_{t-1}$, we rewrite Eq. (3.1) in the form of:

$$\nabla x_t = -\Phi ECT_{t-1} + \sum_{i=1}^p \gamma_i \nabla x_{t-i} + \varepsilon_t \quad (3.2)$$

Eq. (3.2) is a familiar form of the error correction model that is consistent with the convention used for our new model development. The Granger Representation Theorem shows that cointegration implies the existence of an error correction model (ECM), and vice versa, if $x_t \sim I(1)$ are generated by an error correction model, and then x_t is necessarily cointegrated.

Reiteration of these definitions and theorem can also be found for interested readers in the publications authored by Johansen (1995), Zivot and Wang (2006), Pfaff (2008), and Kirchgässner et al. (2013).

3.3 Inherent connection

In this research, it is found that (1) macroscopic traffic variables are cointegrated; (2) an inherent connection exists between error correction model in statistics and macroscopic traffic flow theory through the transformed fundamental diagrams.

3.3.1 Cointegration between traffic variables

Multivariate time series are usually cointegrated if they are generated by or share a common stochastic process. Although traffic volume, speed and occupancy time series collected from a loop detector are three individual time series, they are actually generated by the same stochastic process and describe the characteristics of the same traffic flow from three perspectives. Therefore, it is natural to see that they are cointegrated and share a common stochastic drift. Cointegration tests in our empirical study (Chapter 4) provide us statistical evidence to justify the existence of non-spurious regressive relations between traffic variables.

3.3.2 Transformed fundamental diagrams and error correction model

As a counterpart of the traditional Fundamental Diagrams in Fig.1.4 and 1.5, we construct the fundamental relationships between the differenced traffic variables as shown in Fig.3.1 and 3.2. The value of the difference represents the change of a traffic variable over a time interval. The relationships between the changes reflect the dynamic interaction of three traffic variables as traffic flow is temporally evolving. It is noteworthy that Fig.3.2 provides us clear evidence showing the existence of piecewise linear relationships among differenced traffic variables which coincide with the mathematical form of error correction model, because an error correction model is a linear regression model that works on differenced variables. Therefore, the evidence from Fig.3.2 justifies the appropriateness that we employ the family of error correction model to reveal the dynamics between the differenced traffic variables. From Fig.1.5 to Fig.3.2, the transformation of the fundamental diagrams from the level traffic variables to the differenced traffic variables has revealed a strong inherent connection between the macroscopic traffic flow theory and the family of error correction model in statistics. Similar to location 3, the fundamental diagrams that are constructed respectively with the level traffic variables and the differenced traffic variables for location 1 and 2 can be found in Appendix B. To the best of our knowledge, this is the first time that this finding is established and is one of the contributions of the research.

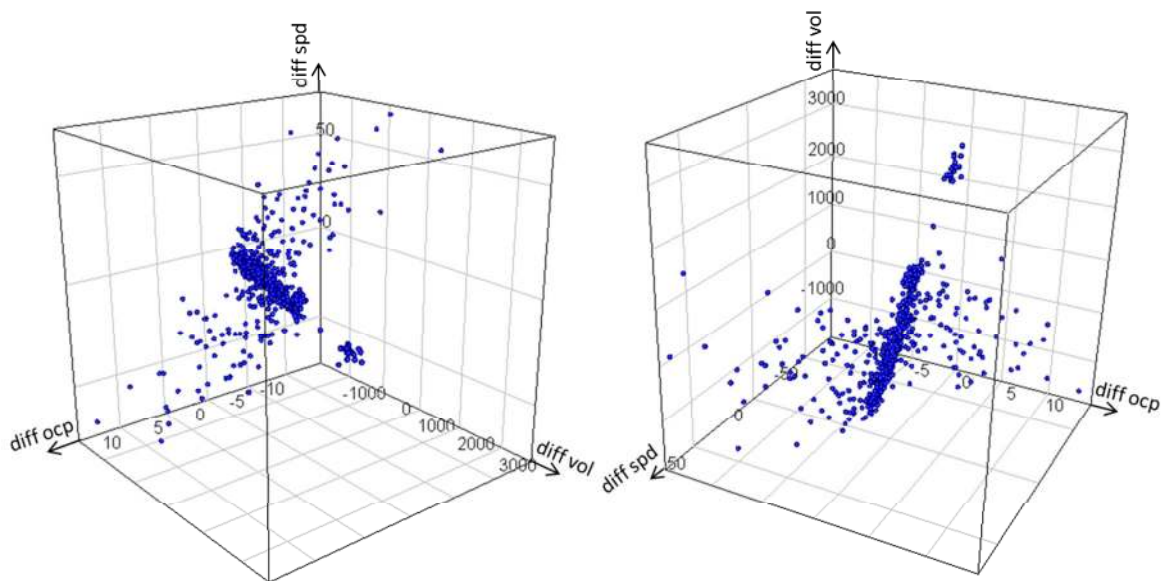


Fig. 3.1 3D scatter plot of differenced volume, speed, and occupancy, location 3

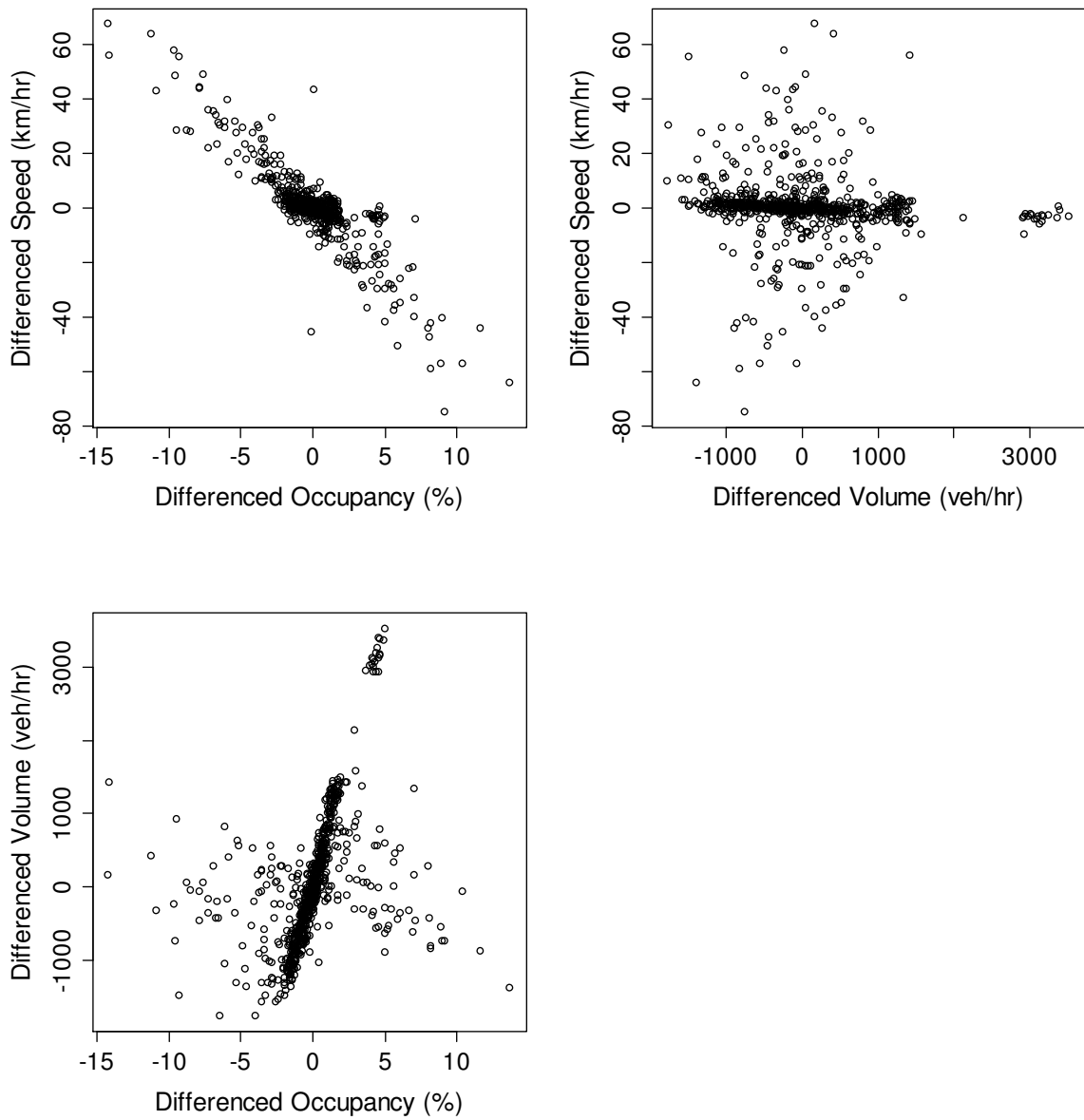


Fig. 3.2 Fundamental relationships between differenced traffic variables, location 3

3.4 Time-space vector error correction model

The error correction model in Eq. (3.2) is expanded to accommodate spatial correlations with exogenous terms. The resulting time-space vector error correction model is given by Eq. (3.3) which is referred as TS-VEC in this research. The error correction term is given by Eq. (3.4).

$$\begin{aligned}
\begin{bmatrix} \nabla x_{01,t} \\ \dots \\ \nabla x_{0n,t} \end{bmatrix} &= \begin{bmatrix} \phi_1 \\ \dots \\ \phi_n \end{bmatrix} + \begin{bmatrix} \phi_{01} \\ \dots \\ \phi_{0n} \end{bmatrix} ECT_{t-1} \\
&+ \sum_{i=1}^p \begin{bmatrix} \gamma_{11,i} & \dots & \gamma_{1n,i} \\ \dots & \dots & \dots \\ \gamma_{n1,i} & \dots & \gamma_{nn,i} \end{bmatrix} \begin{bmatrix} \nabla x_{01,t-i} \\ \dots \\ \nabla x_{0n,t-i} \end{bmatrix} \\
&+ \sum_{j=1}^l \sum_{s=1}^m \begin{bmatrix} \pi_{11,j}^s & \dots & \pi_{1n,j}^s \\ \dots & \dots & \dots \\ \pi_{n1,j}^s & \dots & \pi_{nn,j}^s \end{bmatrix} \begin{bmatrix} x_{01,t-j} - x_{s1,t-j} \\ \dots \\ x_{0n,t-j} - x_{sn,t-j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{01,t} \\ \dots \\ \varepsilon_{0n,t} \end{bmatrix}
\end{aligned} \tag{3.3}$$

$$\text{where } ECT_{t-1} = x_{01,t-1} - \sum_{k=2}^n \beta_{k-1} x_{0k,t-1} \tag{3.4}$$

and $x_{sk,t}$ denotes the k^{th} ($k = 0, 1, \dots, n$) traffic variable at the location s at time t . s denotes the location, $s = 0$ represents the location of prediction; $s \in [1, m]$ denotes the neighborhood site spatially correlated to the location of prediction; $\nabla x_{0k,t}$ denotes traffic variables in the 1st order difference. i and j denote time lag. $\phi_1 \dots \phi_n$ denote intercepts. ECT_{t-1} is the error correction term defined by Eq. (3.4) that accounts for the cointegration and long run equilibrium of traffic variables. $\phi_{01} \dots \phi_{0n}$ are the coefficients of the error correction term. $\nabla x_{01,t-i} \dots \nabla x_{0n,t-i}$ ($i = 1, 2, \dots, p$) are the p -order autoregressive lag terms which reflect short term dynamics and temporal evolution of traffic flow. γ measures the short-term effect and immediate response of $x_{0k,t}$ to a short term change in lag variable $x_{01,t-i} \dots x_{0n,t-i}$. The terms $(x_{01,t-j} - x_{s1,t-j}) \dots (x_{0n,t-j} - x_{sn,t-j})$ are defined as exogenous spatial terms which represent the difference between traffic variables at the location of prediction and its neighborhood site $s = 1, 2, \dots, m$ for time lag $j = 1, 2, \dots, l$. The coefficients π_j^s measure the spatial correlations between the location of prediction and neighborhood sites. ε_t represent white noise residuals with $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_s) = 0$, $\text{Var}(\varepsilon_t) = \sigma^2$.

The spatial correlations are taken into account via exogenous terms that contain the differenced traffic time series between the location of prediction and its neighborhood sites. The benefit is

that the differenced spatial information is more stationary with less noise. A numerical verification is provided in the empirical study, sub-section 4.1.3 differencing between spatial time series of chapter 4. In addition, the difference between traffic variables can reflect the change of traffic situation at neighborhood sites relative to the location under study. The exogenous terms in the model are flexible and elastic to accommodate any number of time lags and spatial orders without the need of constructing a spatial weight matrix that changes over time. It is much easier to construct spatial correlations in this way than constructing a spatial weight matrix in the STARIMA model.

3.5 Threshold regime switching framework

A linear model is sufficient to model stationary time series processes. However, non-stationary time series or structural breaks that may appear in any order moments of a time series leads to nonlinear problems. A threshold regime switching framework provides an approach to solving this type of nonlinear problem.

3.5.1 Prototype of threshold model

The original threshold model was first proposed by Tong (1978) and was further studied by Hansen (1999), and Bai and Perron (1998), among others. The time series between structural breaking points can be modeled by a piecewise linear model that is nested in a regime of the regime switching framework that is controlled by a threshold mechanism. When the transition variable exceeds a threshold value, the regime switching framework can use a different piecewise linear model by switching to a different regime. This mechanism involves structural change detection coupled with threshold identification. In his original threshold framework, Tong adopted the autoregressive model $AR(p)$ as a piecewise linear model in each regime known as SETAR (self-exciting threshold autoregressive) model that has been widely applied to a variety of areas (Tong, 1983, 1990, 2010). Let Y_t be a univariate time series. A

$SETAR(m, p, \dots, p)$ model takes the form

$$Y_t = \sum_{j=1}^m I_t^{(j)}(\gamma, d) \left(\beta_0^{(j)} + \sum_{i=1}^p \beta_i^{(j)} Y_{t-i} + e_t^{(j)} \right) \quad (3.5)$$

where m denotes the number of regimes, j denotes the j^{th} regime $j = (1, 2, \dots, m)$, the vector γ is called the threshold parameter, $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_{m-1}, \gamma_m)^T$ a linearly ordered subset of the real numbers, such that $-\infty = \gamma_0 < \gamma_1 < \dots < \gamma_{m-1} < \gamma_m = \infty$ which defines a partition of the real line \mathbb{R} . d is called the time delay parameter which may be any strictly positive integer with upper bound p , and $\beta_i^{(j)}$ denotes the coefficient in the j^{th} regime where $i = (0, 1, \dots, p)$, and

$$I_t^{(j)}(\gamma, d) = I(\gamma_{j-1} < Y_{t-d} \leq \gamma_j)$$

where $I(\bullet)$ is the indicator function

$$I(a) = \begin{cases} 1 & \text{if } a \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

The error $e_t^{(j)}$ in Eq. (3.5) is a uniformly square integrable martingale difference sequence, hence $E(e_t | Y_t, Y_{t-1}, Y_{t-2}, \dots) = 0$ and $E(e_t^2) = \sigma^2 < \infty$. A $SETAR(m, p, \dots, p)$ model has m regimes where the j^{th} regime occurs when $I_t^j(\gamma, d) = 1$. The class $SETAR(1, p)$ is the class of linear AR model of order p . For analytical reasoning, it is convenient to rewrite Eq. (3.5) in the following matrix form:

$$Y_t = \sum_{j=1}^m I_t^{(j)}(\gamma, d) (\mathbf{X}_{t-1}^T \boldsymbol{\beta}^{(j)} + e_t^{(j)}) \quad (3.6)$$

where $\mathbf{X}_{t-1} = (1 \ Y_{t-1} \ Y_{t-2} \ \dots \ Y_{t-p})^T$ a $k \times 1$ vector with $k = 1 + p$, and $\boldsymbol{\beta}^{(j)} = (\beta_0^{(j)} \ \beta_1^{(j)} \ \dots \ \beta_p^{(j)})^T$ a $k \times 1$ coefficient vector.

A more general class of threshold models may include SETARMA, TARSO, TARSC, and EAR, etc. The interested readers may refer to the books or articles authored by Tong (1983, 1980, and 1975), Tong and Lim (1980), Bartlett (1966), Lawrance and Lewis (1980), etc.

3.5.2 Threshold cointegration vector error correction model

Considering multiple states, non-stationarity, and structural break in traffic time series, we adopt threshold regime switching framework. However, in this research, a linear time-space vector error correction model is developed for each regime because cointegration exists in traffic time series. Hence, based on TS-VEC model and the concept of threshold cointegration, we propose the time-space threshold vector error correction model for short term traffic state prediction, named as TS-TVEC model. The multiple TS-VEC piecewise linear models are nested in a regime switching framework to formulate the TS-TVEC model. The resulting model specification is given by Eq. (3.7).

$$\nabla Y_{0,t} = \begin{cases} \Phi_0^1 + \Phi^1 ECT_{t-1} + \sum_{i=1}^p \Gamma_i^1 \nabla Y_{0,t-i} + \sum_{j=1}^l \sum_{s=1}^m \Pi_{j,s}^1 (Y_{0,t-j} - Y_{s,t-j}) + \varepsilon_t & \text{if } ECT_{t-1} \leq \theta_1 \\ \dots & \dots \\ \Phi_0^k + \Phi^k ECT_{t-1} + \sum_{i=1}^p \Gamma_i^k \nabla Y_{0,t-i} + \sum_{j=1}^l \sum_{s=1}^m \Pi_{j,s}^k (Y_{0,t-j} - Y_{s,t-j}) + \varepsilon_t & \text{if } \theta_{k-1} < ECT_{t-1} \leq \theta_k \\ \dots & \dots \\ \Phi_0^{r+1} + \Phi^{r+1} ECT_{t-1} + \sum_{i=1}^p \Gamma_i^{r+1} \nabla Y_{0,t-i} + \sum_{j=1}^l \sum_{s=1}^m \Pi_{j,s}^{r+1} (Y_{0,t-j} - Y_{s,t-j}) + \varepsilon_t & \text{if } \theta_r < ECT_{t-1} \end{cases} \quad (3.7)$$

where $Y_{s,t} = (x_{s1,t} \ \dots \ x_{sn,t})'$, s denotes the location, $s=0$ represents the location of prediction, and other values represent neighborhood sites. θ_k is the threshold parameter, the usual TS-VEC model parameters Φ_0^k , Φ^k , Γ_i^k and $\Pi_{j,s}^k$ have now the superscript k referring to the k^{th} regime when the ECT_{t-1} is between the threshold θ_{k-1} and θ_k . ECT_{t-1} as a transition variable indicates the deviation from the long run equilibrium. The deviation can be positive or negative where asymmetry may occur and jog among regimes. The coefficient Φ^k indicates how the error correction is made and the adjustment speed towards the equilibrium. Different regimes have different adjustment speeds for error correction. The sign of Φ^k is expected to be opposite to the sign of ECT_{t-1} . $\Theta = (\theta_1, \dots, \theta_r)'$ is the threshold vector that determines the number of regimes and tells when regime switching occurs.

3.5.3 Capacity of threshold structure

It is demonstrated in the literature that threshold models have very rich structure to capture many frequency-domain phenomena of nonlinear system including limit cycles, synchronization, sub-harmonics, higher harmonics, jump resonance, time-irreversibility and amplitude-frequency dependency (Tong, 1983). The concept of threshold cointegration was introduced to capture asymmetric deviation in the case of structural change in time series (Hansen and Seo 2002, Stigler and Tortora 2011).

In the context of traffic time series modelling and forecasting, the structural changes often appear in most traffic time series due to abrupt disturbances to traffic flow that may change the stability of the data generation process (Cetin and Comert, 2006). When traffic flow is in a free flow state, traffic variables fluctuate evenly around the long run equilibrium; when traffic flow is in a congestion state, the traffic variables exhibit asymmetrical deviation, such as volume and speed appear to abruptly decline from high to low values, whereas, occupancy appears to abruptly spike from low to high. Those deviated values may appear to jump from one regime to another regime.

The TS-VEC model is a linear model that performs a symmetrical error adjustment and is good for modelling the free flow situation, but lacks the capacity to capture asymmetric deviation in congested situation. However, the TS-TVEC model with threshold regime switching framework can provide a mechanism for correcting asymmetric deviation with different error adjustment speed in a different regime.

3.6 Model estimation

The estimation of a threshold regime switching error correction model is to determine the number of regimes, time delay parameter, threshold vector, order selection and coefficients. This task can be converted to a multilayer optimization problem. The determination of the number of regimes is a process of structural change detection coupled with threshold identification. A few conventional statistical tests can be used to detect structural change, i.e. the threshold effect. These statistical tests include the CUSUM test (Page, 1954) and the Likelihood Ratio (Hansen, 1999). However, their validity relies on the assumption of (weak) stationarity under the null

hypothesis. In the cases where the covariance structure of the time series is varying, i.e., the second- and higher-order non-stationarity appear in time series, these conventional tests of a null hypothesis may lead to inconsistent and biased testing results. In a recent research, a simple bootstrap procedure is proposed and shown to be consistent under no assumption of stationarity under the null hypothesis in the temporal dynamics of the time series (Zhou, 2013). The distribution of threshold variable is non-standard and complicated. The estimation of its distribution and determination of critical value p are challenging due to the presence of a nuisance parameter that is not identifiable under the null hypothesis. For testing threshold effects, the Bootstrap method is suggested for estimating its asymptotic distribution and critical value p (Hansen, 1999). A few methods are proposed for threshold identification such as Tong's method (Tong, 1983, 1990), Genetic Algorithm (Baragona et al., 2004), Least Squares (Hansen, 1999) and AIC criteria (Chen et al., 2012), etc.

Least Squares and grid search techniques (Narzo et al., 2014) are currently employed in our research to identify the threshold vector. To estimate a typical SETAR model as defined in Eq.(3.6) given (γ, d) , the conditional least square (CLS) estimator for $\beta^{(j)}$ is defined as

$$\hat{\beta}^{(j)}(\gamma, d) = \left(\mathbf{X}^T I^{(j)}(\gamma, d) \mathbf{X} \right)^{-1} \left(\mathbf{X}^T I^{(j)}(\gamma, d) \mathbf{Y} \right) \quad (3.8)$$

The residual sum of squared errors for given (γ, d) is

$$RSS_T(\gamma, d) = \sum_{t=1}^T \left(Y_t - \sum_{j=1}^m I_t^{(j)}(\gamma, d) \left(\mathbf{X}_{t-1}^T \hat{\beta}^{(j)} \right) \right)^2 \quad (3.9)$$

and we define the estimator of (γ, d) as the value that minimizes $RSS_T(\gamma, d)$:

$$\left(\hat{\gamma}, \hat{d} \right) = \arg \min_{\gamma, d} RSS_T(\gamma, d) \quad (3.10)$$

The structural estimators evaluated at the estimated threshold values are defined as:

$$\hat{\beta}^{(j)}(\hat{\gamma}, \hat{d}) = \left(\mathbf{X}^T I^{(j)}(\hat{\gamma}, \hat{d}) \mathbf{X} \right)^{-1} \left(\mathbf{X}^T I^{(j)}(\hat{\gamma}, \hat{d}) \mathbf{Y} \right) \quad (3.11)$$

While AIC (the information criterion of Akaike, 1973) or BIC (Akaike, 1978) criterion are usually used in most practice for model order selection, they are not designed to minimize the

prediction error. Instead, they are designed for selecting a parsimonious version of the model with a better goodness of fit. AIC or BIC criterion is robust for statistical inference. However, they are not necessarily a guarantee for the chosen model to have good performance of prediction, and vice versa.

As the goal of the research is to perform prediction and the mean squared error (MSE) of prediction is the most important indicator that is directly related to prediction, hence, MSE is employed as a criterion with cross validation method for order selection. Because the mean squared error of prediction depends not only on the white noise variance of the fitted model, but also on errors arising from the estimation of the parameters of the model, therefore, choosing a model with minimum MSE value of out-of-sample prediction accounts for both goodness of fit of the model and its predictive performance.

The data structure in Eq. (3.7) can be rearranged to be analogous to Eq. (3.6). Hence, the TS-TVEC model can be estimated in such a way that is similar to aforementioned SETAR estimation. The algorithm for TS-TVEC model estimation is summarized as follows. Let

$\mathbf{X}_{t-1} = \left(1, ECT_{t-1}, \nabla Y_{0,t-1} \cdots \nabla Y_{0,t-p}, (Y_{0,t-1} - Y_{1,t-1}) \cdots (Y_{0,t-l} - Y_{m,t-l})\right)^T$, where $ECT_{t-1}(\alpha) = \alpha^T Y_{0,t-1}$, and $\boldsymbol{\beta} = \left(\Phi_0, \Phi, \Gamma_1 \cdots \Gamma_p, \Pi_{1,l} \cdots \Pi_{l,m}\right)^T$, then the TS-TVEC model as defined in Eq. (3.7) may alternatively be written in a matrix form as Eq. (3.12).

$$\nabla Y_{0,t} = \sum_{j=1}^m I_t^{(j)}(\alpha, \theta) \left(\mathbf{X}_{t-1}^T \boldsymbol{\beta}^{(j)} + \varepsilon_t^{(j)} \right) \quad (3.12)$$

where $I_t^{(j)}(\alpha, \theta) = I(\theta_{j-1} < ECT_{t-1} \leq \theta_j)$ is the indicator function, α is cointegrating vector, θ denotes threshold variable, and $\boldsymbol{\beta}_i^{(j)}$ denotes the coefficient in the j^{th} regime.

1. Construct a grid search space on $[\alpha_{low}, \alpha_{up}]$ and $[\theta_{low}, \theta_{up}]$ based on the linear estimate α , and ECT_{t-1} .
2. Fix pair value (i, k) , where lag order $i = 1, 2, \dots, p$, and the number of threshold $k = 1, 2$.
3. For each pair value (α, θ) on this grid, calculate

$$\hat{\boldsymbol{\beta}}^{(j)}(\alpha, \theta) = \left(\mathbf{X}^T I^{(j)}(\alpha, \theta) \mathbf{X} \right)^{-1} \left(\mathbf{X}^T I^{(j)}(\alpha, \theta) Y \right) \quad \text{and}$$

$$RSS_T(\alpha, \theta) = \sum_{t=1}^T \left(\nabla Y_t - \sum_{j=1}^m I_t^{(j)}(\alpha, \theta) (X_{t-1}^T \hat{\beta}^{(j)}) \right)^2, \text{ respectively.}$$

4. Find $(\hat{\alpha}, \hat{\theta}) = \arg \min_{\alpha, \theta} RSS_T(\alpha, \theta)$ on this grid that yields the lowest value of RSS_T .

5. Determine model coefficient $\hat{\beta}^{(j)}(\hat{\alpha}, \hat{\theta})$ as defined in

$$\hat{\beta}^{(j)}(\hat{\alpha}, \hat{\theta}) = \left(X^T I^{(j)}(\hat{\alpha}, \hat{\theta}) X \right)^{-1} \left(X^T I^{(j)}(\hat{\alpha}, \hat{\theta}) Y \right).$$

6. For each pair value (i, k) , use the estimated model at step 5 to perform prediction using test data set, and calculate MSE.

7. Determine (i, k) for the best model that yields the lowest value of MSE.

The threshold regime switching model has not been systematically investigated and applied to traffic modelling and forecasting. Most of regime switching models in the literature do not involve the process of structural change detection and threshold identification. To the best of our knowledge, this is the first time that such a model structure as TS-TVEC is proposed and applied to traffic state prediction.

Chapter 4

Testing and evaluation

The empirical study with true time series data is presented in this chapter. The study includes (1) building the TS-TVEC model, a multilayer feed-forward Neural Network, and Support Vector Regression model; (2) performing one-step-ahead rolling prediction; (3) comparing the accuracy of their predictions to verify the effectiveness and robustness of the TS-TVEC model. The Neural Network and Support Vector Regression models are commonly used for nonlinear multivariate prediction in the literature. At each of three locations in Fig. 1.1, three pairs of traffic time series, i.e. volume vs. speed, volume vs. occupancy, and speed vs. occupancy, are used respectively to estimate the models. Those models are then used to perform one-step-ahead rolling predictions for the next 48 hours (2 days in a row).

4.1 TS-TVEC model

In this section, the statistical tests are performed to verify cointegration and threshold effect among traffic time series; TS-TVEC models are estimated and their specifications are presented; model order selection, residual check, and differencing technique applied to spatial time series are discussed.

4.1.1 Test cointegration and threshold effect

The application of TS-TVEC model is based on two hypotheses that need to hold in time series: (1) cointegration exists between time series; (2) threshold effect exists or a structural break is detected in the time series. Statistical tests are performed to verify the two hypotheses in traffic time series.

The hypothesis of cointegration is tested with the Phillips-Ouliaris cointegration test (Phillips and Ouliaris, 1990) and verified by the Johansen cointegration test (eigenvalue & trace) (Johansen, 1995). The null hypothesis of no cointegration is rejected at 5% significance by all tests. The results of the cointegration test for nine pairs of traffic time series from three locations are presented in Tables 4.1, 4.2, and 4.3 respectively.

Table 4.1 Phillips-Ouliaris cointegration test

Phillips-Ouliaris: Cointegration test with H0: no cointegration										
Pair of Traffic Variables	Type	Location 1		Location 2		Location 3		Critical Values		
		Test Statistic	p-value	Test Statistic	p-value	Test Statistic	p-value	10%	5%	1%
<i>Volume vs. Speed</i> (<i>q-v</i>)	Pu	272.37	p < 1%	279.84	p < 1%	285.58	p < 1%	27.85	33.71	48.00
	Pz	244.92	p < 1%	275.58	p < 1%	257.56	p < 1%	47.59	55.22	71.93
<i>Volume vs. Occupancy</i> (<i>q-o</i>)	Pu	312.86	p < 1%	284.88	p < 1%	318.84	p < 1%	27.85	33.71	48.00
	Pz	381.22	p < 1%	320.40	p < 1%	367.11	p < 1%	47.59	55.22	71.93
<i>Speed vs. Occupancy</i> (<i>v-o</i>)	Pu	477.53	p < 1%	372.99	p < 1%	345.13	p < 1%	27.85	33.71	48.00
	Pz	278.35	p < 1%	257.70	p < 1%	232.37	p < 1%	47.59	55.22	71.93

Table 4.2 Johansen cointegration test (eigenvalue)

Johansen cointegration test (eigenvalue) with H0: no cointegration										
Pair of Traffic Variables	Rank	Location 1		Location 2		Location 3		Critical Values		
		Test Statistic	p-value	Test Statistic	p-value	Test Statistic	p-value	10%	5%	1%
<i>Volume vs. Speed</i> (<i>q-v</i>)	r<=1	108.35	p < 1%	127.41	p < 1%	123.47	p < 1%	6.5	8.18	11.65
	r=0	229.27	p < 1%	185.94	p < 1%	189.99	p < 1%	12.91	14.9	19.19
<i>Volume vs. Occupancy</i> (<i>q-o</i>)	r<=1	116.66	p < 1%	119.84	p < 1%	121.56	p < 1%	6.5	8.18	11.65
	r=0	267.86	p < 1%	167.49	p < 1%	193.46	p < 1%	12.91	14.9	19.19
<i>Speed vs. Occupancy</i> (<i>v-o</i>)	r<=1	79.01	p < 1%	94.35	p < 1%	66.76	p < 1%	6.5	8.18	11.65
	r=0	209.11	p < 1%	212.22	p < 1%	190.08	p < 1%	12.91	14.9	19.19

Table 4.3 Johansen cointegration test (trace)

Johansen cointegration test (Trace) with H0: no cointegration										
Pair of Traffic Variables	Rank	Location 1		Location 2		Location 3		Critical Values		
		Test Statistic	p-value	Test Statistic	p-value	Test Statistic	p-value	10%	5%	1%
<i>Volume vs. Speed</i> (<i>q-v</i>)	r<=1	108.35	p < 1%	127.41	p < 1%	123.47	p < 1%	6.5	8.18	11.65
	r=0	337.62	p < 1%	313.35	p < 1%	313.47	p < 1%	15.66	17.95	23.52
<i>Volume vs. Occupancy</i> (<i>q-o</i>)	r<=1	116.66	p < 1%	119.84	p < 1%	121.56	p < 1%	6.5	8.18	11.65
	r=0	384.52	p < 1%	287.33	p < 1%	315.02	p < 1%	15.66	17.95	23.52
<i>Speed vs. Occupancy</i> (<i>v-o</i>)	r<=1	79.01	p < 1%	94.35	p < 1%	66.76	p < 1%	6.5	8.18	11.65
	r=0	288.13	p < 1%	306.58	p < 1%	256.84	p < 1%	15.66	17.95	23.52

The hypothesis of threshold effect is tested by the Hansen test and its extension (Hansen, 1999) for both univariate and multivariate time series. The threshold cointegration joint effect is tested by the Hansen and Seo test (Hansen and Seo, 2002). The results of testing linear vs. threshold effect for both univariate and multivariate time series are shown in Tables 4.4 and 4.5 respectively. The test results for threshold cointegration joint effect are presented in Table 4.6. The null hypothesis of linear cointegration is rejected at 5% significance. Asymptotic threshold

distribution is estimated by the bootstrap method and an example is shown in Fig.4.1. All these statistical tests are performed using the package of *urca* in the statistical software R (Pfaff and Stigler, 2013). Results of the threshold cointegration effect test indicate that threshold cointegration effect exists in the traffic time series under study. Therefore, the application of the TS-TVEC model is appropriate in this circumstance.

4.1.2 Order selection for the TS-TVEC model

Usually, AIC or BIC is commonly used to determine the order of a time series model. However, a TS-TVEC model is mainly used for prediction purpose. Therefore, MSE is proposed as a criterion for order selection. For comparison purpose, AIC, BIC, and MSE are computed respectively for each combination of lag and threshold. Table 4.7 shows the details of the AIC, BIC, and MSE for the three locations. The minimum values of the AIC, BIC and MSE are highlighted in yellow. The selected order p is highlighted in green. It is interesting to see that the choice of the number of lags given by the BIC and MSE criterion is very close most of the time in our experiments. It indicates that the MSE method is robust. The choice of order p from AIC criterion is usually too large for prediction.

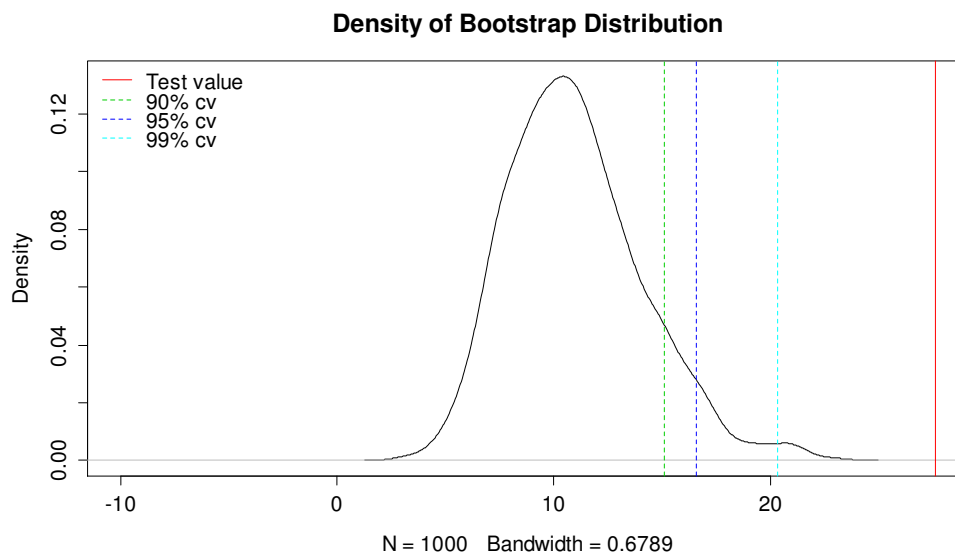


Fig. 4.1 Asymptotic threshold distribution estimated by bootstrap method

Table 4.4 Hansen threshold effect test (univariate)

Test Linearity against Threshold: Univariate, Hansen (1999), with Bootstrap Distribution																
Linearity against 2 or 3 Regimes																
Univariate	Regime Type	Location 1					Location 2					Location 3				
		Test Statistic	P-value	Bootstrap Critical Values			Test Statistic	P-value	Bootstrap Critical Values			Test Statistic	P-value	Bootstrap Critical Values		
				0.9	0.95	0.99			0.9	0.95	0.99			0.9	0.95	0.99
Volume	1vs2	706.79	0.00	28.09	29.12	29.94	523.01	0.00	30.74	31.21	31.60	529.34	0.00	23.92	24.35	24.70
	1vs3	1015.73	0.00	57.94	60.23	62.06	667.45	0.00	52.13	52.18	52.23	938.25	0.00	47.81	47.99	48.14
Speed	1vs2	311.25	0.00	31.38	33.77	35.68	191.97	0.00	28.69	29.00	29.25	165.22	0.00	23.73	24.25	24.67
	1vs3	339.73	0.00	63.72	64.13	64.47	332.01	0.00	55.82	56.73	57.46	204.24	0.00	53.36	54.89	56.12
Occupancy	1vs2	862.19	0.00	27.58	27.91	28.17	257.93	0.00	21.24	21.39	21.51	376.08	0.00	27.70	28.19	28.58
	1vs3	1075.85	0.00	52.90	55.04	56.76	338.27	0.00	54.31	54.91	55.39	456.62	0.00	52.51	52.72	52.90

Table 4.5 Hansen threshold effect test (multivariate)

Test Linearity against Threshold: Multivariate Extension, Hansen (1999), with Bootstrap Distribution																
Linear VAR against 2 or 3 Regime VAR																
Multivariate	Regime Type	Location 1					Location 2					Location 3				
		Test Statistic	P-Value	Bootstrap Critical Values			Test Statistic	P-Value	Bootstrap Critical Values			Test Statistic	P-Value	Bootstrap Critical Values		
				0.90	0.95	0.99			0.90	0.95	0.99			0.90	0.95	0.99
<i>Volume vs. Speed</i> (q-v)	1vs2	509.59	0.00	23.99	24.45	24.82	340.92	0.00	15.05	15.73	16.28	340.92	0.00	18.17	18.66	19.05
	1vs3	798.36	0.00	47.16	48.43	49.44	479.73	0.00	31.73	31.85	31.95	479.73	0.00	37.89	38.72	39.38
<i>Volume vs. Occupancy</i> (q-o)	1vs2	536.42	0.00	23.23	23.36	23.46	346.00	0.00	16.53	16.80	17.01	346.00	0.00	25.33	25.99	26.51
	1vs3	760.38	0.00	39.06	39.48	39.81	506.59	0.00	41.08	41.66	42.12	506.59	0.00	44.78	45.50	46.07
<i>Speed vs. Occupancy</i> (v-o)	1vs2	547.74	0.00	28.16	28.34	28.48	298.66	0.00	28.27	29.32	30.15	298.66	0.00	16.89	16.99	17.07
	1vs3	610.61	0.00	53.25	53.96	54.53	356.65	0.00	31.66	31.92	32.14	356.65	0.00	42.53	44.08	45.32

Table 4.6 Hansen and Seo threshold cointegration test

Test of Linear Cointegration versus Threshold Cointegration of Hansen and Seo (2002) with Bootstrap Distribution															
Number of Bootstrap Replications: 1000															
Multivariate	Location 1					Location 2					Location 3				
	Test Statistic	P-Value	Bootstrap Critical Values			Test Statistic	P-Value	Bootstrap Critical Values			Test Statistic	P-Value	Bootstrap Critical Values		
			0.90	0.95	0.99			0.90	0.95	0.99			0.90	0.95	0.99
<i>Volume vs. Speed</i> (q-v)	62.70	0.00	11.17	12.63	16.02	45.15	0.00	12.34	13.48	16.20	52.18	0.00	11.94	13.48	16.12
<i>Volume vs. Occupancy</i> (q-o)	74.59	0.00	9.02	10.64	12.98	50.42	0.00	9.53	10.72	14.13	37.71	0.00	9.28	10.69	13.70
<i>Speed vs. Occupancy</i> (v-o)	101.15	0.00	16.29	18.00	22.83	27.60	0.00	15.09	16.60	20.32	80.87	0.00	15.83	17.35	20.81

Table 4.7 Model order selection based on AIC, BIC, and MSE criterion

Model	Volume vs. Speed (q-v)						Volume vs. Occupancy (q-o)						Speed vs. Occupancy (v-o)						
	Threshold #	1			2			1			2			1			2		
Lag Order	AIC	BIC	MSE(p)	AIC	BIC	MSE(p)	AIC	BIC	MSE(p)	AIC	BIC	MSE(p)	AIC	BIC	MSE(p)	AIC	BIC	MSE(p)	
Location 1	p = 1	7539.47	7710.42	406929.40	7404.98	7663.50	878253.50	6490.51	6661.46	434432.50	6360.22	6618.73	1568832.00	2446.98	2617.93	873.73	2403.73	2662.25	857.79
	p = 2	7522.98	7727.19	462216.20	7383.40	7691.80	1220558.00	6446.78	6650.99	464982.50	6330.29	6638.69	1676715.00	2418.52	2622.73	950.28	2380.21	2688.61	1107.26
	p = 3	7504.04	7741.47	591770.20	7368.74	7726.96	1338586.00	6413.64	6651.07	906361.20	6316.84	6675.07	1355085.00	2392.24	2629.67	1079.76	2274.75	2632.98	1497.12
	p = 4	7479.27	7749.88	652591.90	7317.66	7725.67	2706680.00	6373.66	6644.27	910494.80	6308.50	6716.50	1655299.00	2362.45	2633.07	1065.09	2245.36	2653.36	2238.09
	p = 5	7464.80	7768.57	1228177.00	7304.38	7762.11	1989160.00	6355.12	6658.89	1731674.00	6291.28	6749.02	2774018.00	2349.84	2653.61	858.22	2232.98	2690.71	3110.55
	p = 6	7453.19	7790.08	1294465.00	7284.51	7791.92	2703999.00	6347.30	6684.19	2161066.00	6283.02	6790.43	1048825.00	2336.87	2673.76	6445.87	2194.79	2702.20	6085.33
	p = 7	7347.81	7717.78	1679181.00	7247.72	7804.76	1725764.00	6302.39	6672.36	1819287.00	6226.57	6783.60	1908345.00	2334.55	2704.53	2312.40	2185.38	2742.42	7921.58
	p = 8	7332.84	7735.86	2723634.00	7237.58	7844.19	2560302.00	6281.80	6684.82	2693122.00	6196.31	6802.92	2819964.00	2328.54	2731.57	2128.03	2117.49	2724.10	2688.68
	p = 9	7315.56	7751.60	6890541.00	7235.66	7891.79	4217225.00	6275.16	6711.19	545132979.00	6194.94	6851.07	2784259.00	2315.33	2751.37	1618.33	2091.59	2747.72	1470263.00
	p = 10	7375.83	7844.85	1153613.00	7199.70	7905.30	3915630.00	6292.50	6761.52	2506014.00	6168.16	6873.76	2505531.00	2310.37	2779.39	1448.73	2062.99	2768.59	50672.98
Location 2	Lag Order	AIC	BIC	MSE(p)	AIC	BIC	MSE(p)	AIC	BIC	MSE(p)	AIC	BIC	MSE(p)	AIC	BIC	MSE(p)	AIC	BIC	MSE(p)
	p = 1	8049.72	8220.67	546421.80	8035.74	8294.26	559438.10	6803.56	6974.51	458312.30	6784.59	7043.11	807054.90	3245.20	3416.15	1298.31	3211.84	3470.35	2373.15
	p = 2	8031.37	8235.58	656624.80	8024.27	8332.67	522058.00	6783.90	6988.11	815202.40	6755.54	7063.94	845185.80	3236.25	3440.46	1363.48	3198.38	3506.78	2499.75
	p = 3	8014.01	8251.44	641282.70	7996.43	8354.65	508418.80	6771.94	7009.37	754915.10	6753.39	7111.61	975177.10	3237.25	3474.68	1321.95	3207.95	3566.18	2391.65
	p = 4	8007.10	8277.72	623179.10	7984.97	8392.98	660015.00	6764.95	7035.57	954285.70	6746.24	7154.24	909110.50	3229.22	3499.84	1270.05	3199.61	3607.61	2708.79
	p = 5	8000.60	8304.36	671788.70	7967.57	8425.30	727666.90	6761.82	7065.59	915733.90	6764.60	7222.33	785773.10	3231.70	3535.47	2001.79	3202.71	3660.45	2782.66
	p = 6	7949.35	8286.24	451051.00	7938.33	8445.74	466458.60	6744.08	7080.96	484261.60	6731.14	7238.55	497553.70	3223.80	3560.68	1421.79	3196.09	3703.50	3060.82
	p = 7	7937.23	8307.20	451185.60	7932.07	8489.11	489677.80	6728.23	7098.20	521460.70	6707.60	7264.63	547806.80	3215.60	3585.57	1491.23	3171.43	3728.46	24770.29
	p = 8	7924.42	8327.44	462855.40	7914.66	8521.27	502433.40	6732.95	7135.97	1784199.00	6717.00	7323.61	1505483.00	3210.17	3613.19	1886.34	3154.71	3761.32	3983.89
	p = 9	7931.34	8367.38	899177.70	7881.76	8537.90	2051793.00	6722.22	7158.25	3355508.00	6691.49	7347.62	2496301.00	3213.29	3649.33	2094.87	3148.52	3804.65	49828.79
p = 10	7925.24	8394.26	747903.00	7891.26	8596.86	1051353.00	6719.04	7188.06	3279767.00	6693.83	7399.43	2535664.00	3219.08	3688.09	8330.12	3154.61	3860.22	20017.21	
Location 3	Lag Order	AIC	BIC	MSE(p)	AIC	BIC	MSE(p)	AIC	BIC	MSE(p)	AIC	BIC	MSE(p)	AIC	BIC	MSE(p)	AIC	BIC	MSE(p)
	p = 1	7819.04	7989.99	323904.10	7758.28	8016.80	346550.80	6259.64	6430.59	376989.20	6209.73	6468.25	391255.90	2466.48	2637.43	221.87	2457.19	2715.71	227.69
	p = 2	7812.34	8016.55	328810.80	7814.39	8122.78	609175.30	6243.67	6447.87	373933.20	6190.72	6499.12	386174.70	2442.84	2647.05	206.59	2427.65	2736.04	175.83
	p = 3	7789.25	8026.68	330624.50	7789.75	8147.98	721817.20	6223.81	6461.24	375019.50	6194.03	6552.25	546219.20	2406.05	2643.48	241.13	2382.56	2740.78	263.89
	p = 4	7771.88	8042.50	374564.10	7786.24	8194.24	658783.40	6174.51	6445.13	414492.90	6144.58	6552.58	434772.30	2401.94	2672.55	248.12	2379.86	2787.87	273.12
	p = 5	7755.76	8059.52	430052.70	7751.41	8209.14	808389.30	6188.14	6491.90	494755.60	6154.23	6611.96	471317.50	2400.72	2704.49	241.21	2373.76	2831.49	245.14
	p = 6	7740.49	8077.38	409979.80	7751.80	8259.21	566826.40	6172.94	6509.83	503652.50	6146.14	6653.55	2249988.00	2386.77	2723.66	239.62	2363.82	2871.23	246.57
	p = 7	7725.30	8095.27	536348.90	7715.85	8272.89	469323.30	6149.92	6519.89	569097.80	6144.22	6701.25	588963.20	2384.08	2754.06	199.06	2341.92	2898.96	200.94
	p = 8	7720.62	8123.64	553288.30	7714.19	8320.79	799360.10	6143.32	6546.34	658985.10	6073.21	6679.82	678934.90	2384.85	2787.88	197.39	2339.46	2946.07	236.02
	p = 9	7714.37	8150.41	569116.30	7708.71	8364.84	676420.60	6139.41	6575.45	761688.30	6074.58	6730.71	1087431.00	2378.13	2814.17	195.78	2324.82	2980.95	280.77
p = 10	7703.84	8172.86	554019.60	7702.46	8408.07	971224.40	6135.03	6604.05	727154.10	6082.87	6788.47	1001785.00	2380.05	2849.06	192.79	2320.34	3025.94	577.10	

With the cross validation method, a TS-TVEC model with minimum MSE value of out-of-sample prediction is chosen. According to the simplistic and parsimonious principle, the simpler model structure prevails if MSE values are the same.

4.1.3 Differencing between spatial time series

In the literature, differencing is a means of achieving stationarity for time series that is non-stationary. Similarly, this technique is applicable between two non-stationary time series that are spatially close enough to share the same stochastic drift. Our study shows that a time series resulted from the difference between two original time series has less degree of non-stationarity and smaller variance. It is equivalent to the case where a cointegrating vector is $\beta = (1, -1)'$. The essence of differencing between two time series is to cancel out their deterministic and/or stochastic trend.

Four statistical tests are employed to verify the degree of non-stationarity of the time series before and after differencing. These statistical tests include the Augmented Dickey-Fuller (ADF), DF-GLS, KPSS, and Zivot-Andrews unit root test. The difference is taken between the time series from the location of prediction and its adjacent upstream and downstream sites.

Tables 4.9, 4.10, 4.11, and 4.12 respectively show the results of four unit root tests. The results of all tests show that the degree of non-stationarity in the differenced time series is reduced according to the moving direction of p -value in contrast to the original time series. Table 4.13 shows that the variance of the differenced time series is much lower than the original time series. All these results indicate that the spatially differenced time series between the location of prediction and its neighborhood sites become more stationary with less noise. In this research, one-lag spatial time series from upstream and downstream sites are used for the exogenous term in the model, i.e. $l = 1$, and $m = 2$ are specified in Eq. (3.7). The legend used in Tables 4.9, 4.10, 4.11, 4.12, and 4.13 is defined in Table 4.8.

4.1.4 Model specification

Table 4.14 shows two model specification exemplars from nine estimated TS-TVEC models. The error correction term is denoted as ECT_{t-1} in the tables. Model specifications include the

coefficient estimate, its standard error, t -statistic, and p -value. The significance level of coefficients is indicated by the number of stars in the legend as follows: *** [0, 0.001], ** [0.001, 0.01], * [0.01, 0.05]. The mathematical forms for the estimated TS-TVEC models in Table 4.14 are presented in Eq. (4.1), (4.2), (4.3), and (4.4). The estimated threshold is $\theta = 50.78$ for volume vs. speed model, and $\theta = 498.9$ for volume vs. occupancy model. The non-spurious equilibrium, Eq. (3.4), can be verified by the results of cointegration tests in Tables 4.1, 4.2 and 4.3. It can also be justified by the negative sign of the error correction term (ECT_{t-1}) coefficient in model specifications in Table 4.14.

4.1.5 Check residuals

The residual series from each model is verified to be white noise by the Ljung-Box test available in the package of FitAR in the R software (McLeod and Zhang, 2008). Results of the Ljung-Box test for all 18 residual series from 9 TS-TVEC models are shown in Table 4.15. The p -values are all greater than 0.05. Therefore, the null hypothesis of independence cannot be rejected at 5% significance, which indicates that autocorrelations in the time series are well explained by adequate lag terms. Fig.4.2 shows an example of a standardized residual plot, autocorrelograms and p -values of the Ljung-Box statistics from $q \sim v$ model at location 3 to verify that residual series is a white noise.

Table 4.8 Legend used in Tables 4.9, 4.10, 4.11, 4.12 and 4.13

Legend:	
upvol	: upstream volume
vol	: volume at location of prediction
downvol	: downstream volume
vol - upvol	: differenced series between volume at location of prediction and upstream volume
downvol - vol	: differenced series between downstream volume and volume at location of prediction
upspd	: upstream speed
spd	: speed at location of prediction
downspd	: downstream speed
spd - upspd	: differenced series between speed at location of prediction and upstream speed
downspd - spd	: differenced series between downstream speed and speed at location of prediction
upocp	: upstream occupancy
ocp	: occupancy at location of prediction
downocp	: downstream occupancy
ocp - upocp	: differenced series between occupancy at location of prediction and upstream occupancy
downocp - ocp	: differenced series between downstream occupancy and occupancy at location of prediction

Table 4.9 Augmented Dickey-Fuller (ADF) unit root test

Augmented Dickey-Fuller (ADF) Test, H0: unit roots, non-stationary Critical values (cv) for a significance level of: 1% cv = -2.58, 5% cv = -1.95, 10% cv = -1.62						
Traffic Series	Location 1		Location 2		Location 3	
	ADF Test Statistic	p-value	ADF Test Statistic	p-value	ADF Test Statistic	p-value
upvol	-0.3409	p > 10%	-0.4244	p > 10%	-0.0599	p > 10%
vol	-0.3495	p > 10%	-0.1751	p > 10%	-0.0985	p > 10%
downvol	-0.2967	p > 10%	-0.0633	p > 10%	-0.2658	p > 10%
vol - upvol	-1.8341	5% < p < 10%	0.0498	p > 10%	-0.5549	p > 10%
downvol - vol	-0.5747	p > 10%	-2.3021	1% < p < 5%	-0.6237	p > 10%
upspd	0.2899	p > 10%	0.1038	p > 10%	0.3505	p > 10%
spd	-0.0962	p > 10%	0.1628	p > 10%	0.3826	p > 10%
downspd	0.2284	p > 10%	0.3569	p > 10%	0.3106	p > 10%
spd - upspd	-0.6371	p > 10%	-0.5366	p > 10%	-0.5800	p > 10%
downspd - spd	-1.6825	5% < p < 10%	-2.1077	1% < p < 5%	-3.1298	p < 1%
upocp	0.1860	p > 10%	-1.0786	p > 10%	-0.5572	p > 10%
ocp	-0.0434	p > 10%	-1.2046	p > 10%	-1.0247	p > 10%
downocp	-0.1329	p > 10%	-1.0579	p > 10%	-1.0675	p > 10%
ocp - upocp	-2.4576	1% < p < 5%	-12.8627	p < 1%	-1.3016	p > 10%
downocp - ocp	-0.5541	p > 10%	-7.6040	p < 1%	-1.2301	p > 10%

Table 4.10 DF-GLS unit root test

Elliot, Rothenberg and Stock (DF-GLS) Test, H0: unit roots, non-stationary Critical values (cv) for a significance level of: 1% cv = -2.57, 5% cv = -1.94, 10% cv = -1.62						
Traffic Series	Location 1		Location 2		Location 3	
	DF-GLS Test Statistic	p-value	DF-GLS Test Statistic	p-value	DF-GLS Test Statistic	p-value
upvol	-0.4774	p > 10%	-0.554	p > 10%	-0.201	p > 10%
vol	-0.4936	p > 10%	-0.4241	p > 10%	-0.2335	p > 10%
downvol	-0.4328	p > 10%	-0.2842	p > 10%	0.2573	p > 10%
vol - upvol	-0.4402	p > 10%	-0.2713	p > 10%	-0.2545	p > 10%
downvol - vol	-0.7054	p > 10%	-0.2121	p > 10%	-1.001	p > 10%
upspd	0.5864	p > 10%	-0.3908	p > 10%	-0.7209	p > 10%
spd	-0.4892	p > 10%	-0.5413	p > 10%	-0.8017	p > 10%
downspd	-0.4667	p > 10%	-0.6267	p > 10%	-0.471	p > 10%
spd - upspd	-1.8736	5% < p < 10%	0.475	p > 10%	-1.4004	p > 10%
downspd - spd	-0.9843	p > 10%	-0.2896	p > 10%	-1.6018	p > 10%
upocp	0.0669	p > 10%	-0.4362	p > 10%	-0.67	p > 10%
ocp	-0.1867	p > 10%	-0.48	p > 10%	-0.6446	p > 10%
downocp	-0.2874	p > 10%	-0.9892	p > 10%	-0.3093	p > 10%
ocp - upocp	-0.9441	p > 10%	-0.5029	p > 10%	-0.9098	p > 10%
downocp - ocp	-0.9569	p > 10%	-0.0832	p > 10%	-0.2539	p > 10%

Table 4.11 KPSS unit root test

KPSS Unit Root Test H0: stationary						
Critical values (cv) for a significance level of: 10% cv = 0.119, 5% cv = 0.146, 2.5% cv = 0.176, 1% cv = 0.216						
Traffic Series	Location 1		Location 2		Location 3	
	KPSS Test Statistic	p-value	KPSS Test Statistic	p-value	KPSS Test Statistic	p-value
upvol	0.2429	p < 1%	0.1977	1% < p < 2.5%	0.2189	p < 1%
vol	0.2436	p < 1%	0.1853	1% < p < 2.5%	0.2197	p < 1%
downvol	0.2402	p < 1%	0.1791	1% < p < 2.5%	0.1762	1% < p < 2.5%
vol - upvol	0.1484	2.5% < p < 5%	0.1752	2.5% < p < 5%	0.2	1% < p < 2.5%
downvol - vol	0.2461	p < 1%	0.1504	2.5% < p < 5%	0.1941	1% < p < 2.5%
upspd	0.2323	p < 1%	0.1615	2.5% < p < 5%	0.2509	p < 1%
spd	0.2718	p < 1%	0.175	2.5% < p < 5%	0.2267	p < 1%
downspd	0.2226	p < 1%	0.2199	p < 1%	0.1966	1% < p < 2.5%
spd - upspd	0.1157	p > 10%	0.1092	p > 10%	0.1809	1% < p < 2.5%
downspd - spd	0.0903	p < 1%	0.1018	p > 10%	0.0859	p > 10%
upocp	0.2772	p < 1%	0.1534	2.5% < p < 5%	0.2659	p < 1%
ocp	0.2458	p < 1%	0.1772	1% < p < 2.5%	0.2605	p < 1%
downocp	0.2277	p < 1%	0.2452	p < 1%	0.2337	p < 1%
ocp - upocp	0.1483	2.5% < p < 5%	0.1063	p > 10%	0.1404	5% < p < 10%
downocp - ocp	0.0941	p < 1%	0.1211	5% < p < 10%	0.0999	p > 10%

Table 4.12 Zivot-Andrews unit root test

Zivot-Andrews Unit Root Test, H0: unit roots in the case of a structural break						
Critical values (cv) for a significance level of: 1% cv = -5.57, 5% cv = -5.08, 10% cv = -4.82						
Traffic Series	Location 1		Location 2		Location 3	
	Zivot-Andrews Test Statistic	p-value	Zivot-Andrews Test Statistic	p-value	Zivot-Andrews Test Statistic	p-value
upvol	-2.7905	p >10%	-3.5443	p >10%	-3.5287	p >10%
vol	-2.8991	p >10%	-3.6017	p >10%	-3.58	p >10%
downvol	-2.8962	p >10%	-4.2777	p >10%	-2.8799	p >10%
vol - upvol	-3.5765	p >10%	-3.3989	p >10%	-2.9242	p >10%
downvol - vol	-2.8136	p >10%	-3.0726	p >10%	-2.936	p >10%
upspd	-4.1702	p >10%	-3.72	p >10%	-2.5345	p >10%
spd	-4.0613	p >10%	-3.6495	p >10%	-4.0012	p >10%
downspd	-3.9653	p >10%	-3.7282	p >10%	-4.2625	p >10%
spd - upspd	-5.4957	1% < p < 5%	-4.4323	p >10%	-5.8844	p < 1%
downspd - spd	-4.431	p >10%	-4.4719	p >10%	-2.8591	p >10%
upocp	-4.3618	p >10%	-3.8215	p >10%	-4.7792	p >10%
ocp	-3.2978	p >10%	-4.3414	p >10%	-4.4776	p >10%
downocp	-2.505	p >10%	-3.7395	p >10%	-4.6391	p >10%
ocp - upocp	-4.7359	p >10%	-4.4035	p >10%	-4.3937	p >10%
downocp - ocp	-5.7881	p < 1%	-5.5064	1% < p < 5%	-2.3707	p >10%

Table 4.13 Comparison of variance

Traffic Series	Variance								
	Volume			Speed			Occupancy		
	Location 1	Location 2	Location 3	Location 1	Location 2	Location 3	Location 1	Location 2	Location 3
<i>upstream</i>	6267967.00	1155107.00	2990877.00	204.21	722.81	345.48	27.64	76.87	31.14
<i>at loc of pred</i>	6320818.00	2427829.00	4708497.00	397.35	780.82	361.42	52.69	72.92	27.62
<i>downstream</i>	4795278.00	2352378.00	2477220.00	408.08	450.93	255.73	59.72	47.11	27.62
<i>loc of pred - upstream</i>	10212.91	240334.90	282030.00	85.30	26.31	35.69	8.52	2.56	1.68
<i>downstream - loc of pred</i>	148062.20	53790.95	437213.90	44.75	66.77	75.40	4.75	19.05	3.36

The mathematical form for volume vs. speed model in Table 4.14, the estimated threshold $\theta = 50.78$.

$$\nabla vol_t = \begin{cases} -0.287ECT_{t-1} - 9.188\nabla spd_{t-1} - 51.856i_{upocc_{t-1}} \\ + 0.656down_{ivol_{t-1}} - 43.323down_{iocc_{t-1}} + \varepsilon_{1t}, & ECT_{t-1} \leq 50.78 \\ -0.502ECT_{t-1} - 63.847down_{iocc_{t-1}} + \varepsilon_{1t}, & ECT_{t-1} > 50.78 \end{cases} \quad (4.1)$$

$$\nabla spd_t = \begin{cases} 0.007\nabla vol_{t-1} - 0.351\nabla spd_{t-1} - 0.798i_{upspd_{t-1}} \\ - 3.303i_{upocc_{t-1}} + \varepsilon_{2t}, & ECT_{t-1} \leq 50.78 \\ -0.014i_{upvol_{t-1}} - 0.768i_{upspd_{t-1}} \\ - 2.341i_{upocc_{t-1}} + \varepsilon_{2t}, & ECT_{t-1} > 50.78 \end{cases} \quad (4.2)$$

The mathematical form for volume vs. occupancy model in Table 4.14, the estimated threshold $\theta = 498.9$.

$$\nabla vol_t = \begin{cases} -0.364ECT_{t-1} + 25.478\nabla occ_{t-1} + 0.340i_{upvol_{t-1}} \\ + 0.394down_{ivol_{t-1}} - 33.923down_{iocc_{t-1}} + \varepsilon_{1t}, & ECT_{t-1} \leq 498.9 \\ -686.040 + 0.656ECT_{t-1} - 0.306\nabla vol_{t-1} \\ - 28.870down_{isped_{t-1}} - 98.077down_{iocc_{t-1}} + \varepsilon_{1t}, & ECT_{t-1} > 498.9 \end{cases} \quad (4.3)$$

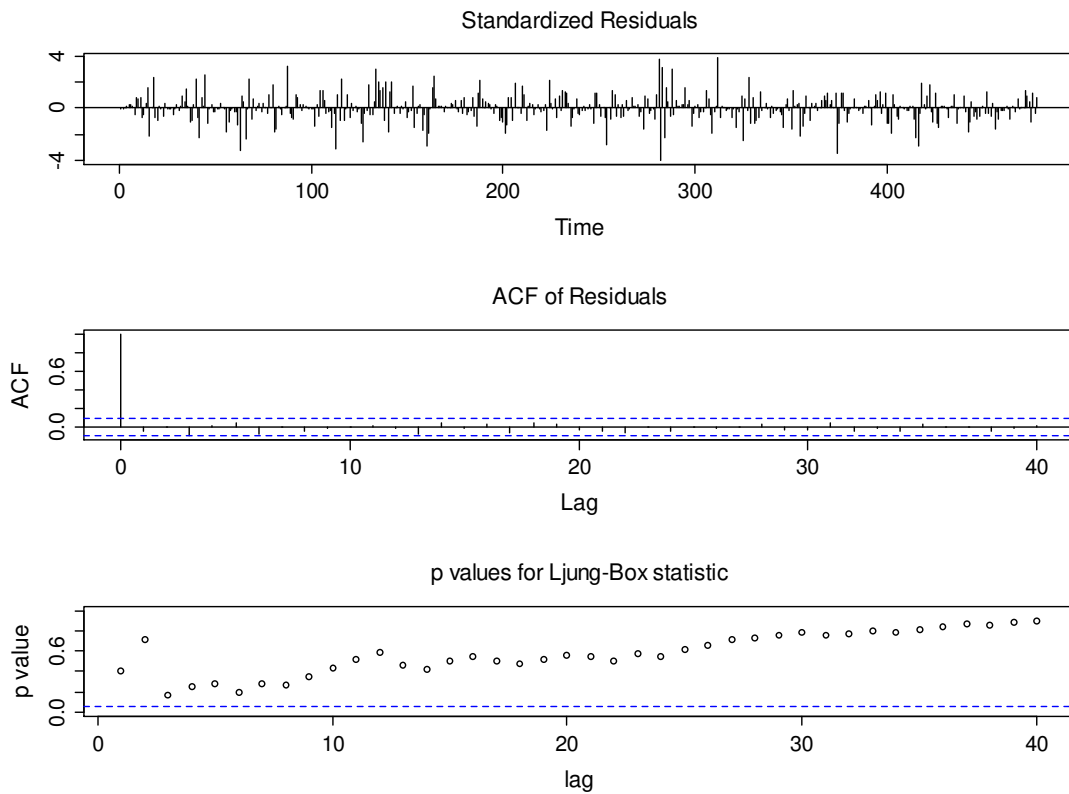
$$\nabla occ_t = \begin{cases} -0.002ECT_{t-1} - 0.001\nabla vol_{t-1} - 0.138\nabla occ_{t-1} - 0.203\nabla occ_{t-2} \\ + 0.003i_{upvol_{t-1}} + 0.065i_{upspd_{t-1}} + \varepsilon_{2t}, & ECT_{t-1} \leq 498.9 \\ 3.622 - 0.006ECT_{t-1} - 0.530\nabla occ_{t-1} + \varepsilon_{2t}, & ECT_{t-1} > 498.9 \end{cases} \quad (4.4)$$

Table 4.14 Specification of estimated TS-TVEC Models (2 examples)

Legend: significant level (p-value) of the coefficient										
***	[0, 0.001]									
**	[0.001, 0.01]									
*	[0.01, 0.05]									
Location 3										
<i>Volume vs. Speed</i>										
Regime-1	coefficients				Std. Error		t-value		Pr(> t)	
Variable	vol		spd		vol	spd	vol	spd	vol	spd
ECT	-0.287	*	0.006		0.11	0.00	-2.52	1.48	0.0121	0.1400
Const	-4.667		-0.928		25.42	0.91	-0.18	-1.02	0.8544	0.3097
vol(t-1)	0.096		0.007	**	0.07	0.00	1.40	2.84	0.1616	0.0048
spd(t-1)	-9.188	***	-0.351	***	1.97	0.07	-4.67	-4.97	0.0000	0.0000
i_upvol	0.289		-0.012		0.17	0.01	1.75	-1.94	0.0809	0.0527
i_upspd	-8.902		-0.798	***	4.82	0.17	-1.85	-4.61	0.0656	0.0000
i_upocc	-51.856	*	-3.303	***	26.35	0.95	-1.97	-3.49	0.0496	0.0005
down_ivol	0.656	***	0.009		0.16	0.01	4.19	1.55	0.0000	0.1227
down_ispd	-4.595		0.099		4.67	0.17	-0.98	0.59	0.3252	0.5527
down_iocc	-43.323	*	0.535		20.91	0.75	-2.07	0.71	0.0388	0.4765
Location 3										
<i>Volume vs. Occupancy</i>										
Regime-1	coefficients				Std. Error		t-value		Pr(> t)	
Variable	vol		occ		vol	occ	vol	occ	vol	occ
ECT	-0.364	***	-0.002	**	0.07	0.00	-5.22	-3.26	0.0000	0.0012
Const	12.262		0.003		15.87	0.10	0.77	0.03	0.4403	0.9743
vol(t-1)	0.095		-0.001	*	0.06	0.00	1.64	-2.48	0.1021	0.0136
occ(t-1)	25.478	**	-0.138	*	8.32	0.05	3.06	-2.52	0.0023	0.0122
vol(t-2)	0.016		0.000		0.05	0.00	0.31	-0.23	0.7574	0.8167
occ(t-2)	5.457		-0.203	***	7.71	0.05	0.71	-3.98	0.4792	0.0001
i_upvol	0.340	*	0.003	**	0.13	0.00	2.52	3.02	0.0120	0.0026
i_upspd	-4.716		0.065	*	3.93	0.03	-1.20	2.48	0.2307	0.0134
i_upocc	-29.424		0.234	.	20.75	0.14	-1.42	1.71	0.1568	0.0886
down_ivol	0.394	***	-0.001		0.12	0.00	3.41	-0.99	0.0007	0.3246
down_ispd	-1.425		0.032		3.40	0.02	-0.42	1.42	0.6758	0.1556
down_iocc	-33.923	*	0.168		16.48	0.11	-2.06	1.54	0.0402	0.1232
Location 3										
<i>Volume vs. Occupancy</i>										
Regime-2	coefficients				Std. Error		t-value		Pr(> t)	
Variable	vol		occ		vol	occ	vol	occ	vol	occ
ECT	0.656	*	-0.006	**	0.32	0.00	2.03	-2.93	0.0428	0.0036
Const	-686.040	**	3.622	*	231.49	1.53	-2.96	2.37	0.0032	0.0183
vol(t-1)	-0.306	*	0.001		0.16	0.00	-1.97	0.49	0.0494	0.6246
occ(t-1)	20.584		-0.530	***	22.90	0.15	0.90	-3.50	0.3692	0.0005
vol(t-2)	0.041		-0.001		0.16	0.00	0.26	-1.38	0.7985	0.1696
occ(t-2)	32.317		-0.341		34.65	0.23	0.93	-1.49	0.3514	0.1366
i_upvol	0.200		0.002		0.32	0.00	0.63	1.15	0.5321	0.2506
i_upspd	-6.163		-0.016		11.64	0.08	-0.53	-0.21	0.5968	0.8356
i_upocc	-121.500		-0.138		65.08	0.43	-1.87	-0.32	0.0625	0.7483
down_ivol	0.288		-0.001		0.32	0.00	0.91	-0.61	0.3614	0.5451
down_ispd	-28.870	**	0.098		10.79	0.07	-2.68	1.38	0.0077	0.1694
down_iocc	-98.077	*	0.462		44.39	0.29	-2.21	1.57	0.0276	0.1163

Table 4.15 Ljung-Box Test for Residual series

Ljung-Box test White Noise for Residuals				
H0: independence				
TVEC Model	Residuals from Equation	Location 1	Location 2	Location 3
		p-value	p-value	p-value
Volume vs. Speed	vol	0.06	0.82	0.33
	spd	0.08	0.52	0.05
Volume vs. Occupancy	vol	0.58	0.63	0.54
	occ	0.27	0.21	0.20
Speed vs. Occupancy	spd	0.05	0.56	0.93
	occ	0.13	0.14	0.39

Fig. 4.2 Residuals, autocorrelogram, and p -values of Ljung-Box, Location 3

4.2 MLP Neural Network

Multilayer Perceptron (MLP) Neural Network is a class of commonly used feed-forward Neural Network architecture with full connection between neurons. Fig.4.3 shows a typical MLP Neural Network architecture used in the study. The MLP Neural Network is chosen among others to

comparatively assess the performance of the TS-TVEC model because it has been proved that multilayer feed-forward networks are a class of universal approximators that are capable of approximating any measurable function to any desired degree of accuracy. Details will be discussed in the following subsection of determining the number of hidden layers.

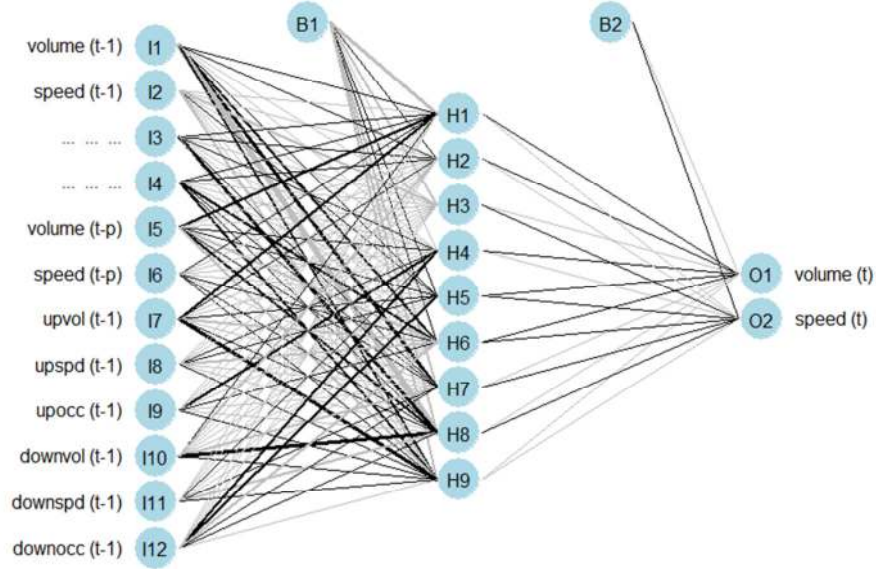


Fig. 4.3 MLP Feed Forward Neural Network Architecture (Beck, 2014)

The Neural Network architecture exhibited in Fig.4.3 approximates a time series model for prediction of traffic volume and speed using the sliding window technique for its input layer, where p denotes the number of lags of a sliding window, the nodes I7 – I12 represent exogenous spatial information from upstream and downstream neighborhood locations including volume, speed, and occupancy with one time lag. Nodes H1 – H9 represent neurons in the hidden layer. B1 and B2 represent the bias term for each neuron in the hidden and output layers. Functionally, a MLP model can be expressed by Eq. (4.1) which is equivalent to a nonlinear autoregressive model.

$$y_{k,t}^0 = \phi_{0,k} + \sum_{h=1}^n \phi_{h,k} f \left(\sum_{i=1}^p \sum_{j=1}^k \theta_{\Lambda 1, h} y_{j, t-i}^0 + \sum_{l=1}^r \sum_{s=1}^m \sum_{j=1}^q \theta_{\Lambda 2, h} (y_{j, t-l}^0 - y_{j, t-l}^s) + \theta_{0, h} \right) + \varepsilon_{k, t}$$

where,

$$k = 1, 2, \dots$$

$$\Lambda 1 = (i-1)k + j$$

$$\Lambda 2 = pk + (l-1)mk + (s-1)k + j$$

(4.5)

where $y_{k,t}^0$ represents a traffic variable in the output layer, k denotes the number of traffic variables in the vector for prediction, p and r denote the number of time lags, m denotes the number of neighborhood sites correlated to the site under study, q denotes the number of traffic variables considered in spatial correlation (commonly equals 3), s denotes the location, $s=0$ represents the location under study, the product of $(p*(k+r)*m*k)$ equals the number of input nodes, and n is the number of nodes in the hidden layer. f is a sigmoid transfer function such as the logistic $f(x) = \frac{1}{1 + \exp(-x)}$, $\phi_{h,k}$ is a vector of weights from the hidden to output nodes and $\theta_{\lambda_1,h}$ and $\theta_{\lambda_2,h}$ are weights from the input to the hidden nodes. $\phi_{0,k}$ and $\theta_{0,h}$ denote the bias terms equivalent to the intercept term in a linear model. Most forecasting problems usually employ a linear function between the hidden and output layers as shown in Eq. (4.1).

As there is no well-defined consensus guidance available in the literature for Neural Network specifications, we tried our best to follow some theoretical guidance in the literature as discussed next and use cross validation and early stopping techniques to construct the Neural Network.

4.2.1 Input layer and output layer

The sliding window technique is used to design the input layer of the MLP network in order to deal with time series data. The input layer includes both endogenous lag terms and exogenous terms. The lag terms are equivalent to the autoregressive part of the corresponding TS-TVEC model with order p , and the exogenous terms contain the same spatial information as the TS-TVEC model. The output layer contains the number of neurons equivalent to the number of traffic variables under prediction.

4.2.2 The number of hidden layers

In theory, the MLP feed-forward Neural Network only needs a single hidden layer that is sufficient to approximate a nonlinear autoregressive model such as Eq. (4.1) in the context of this research. Kolmogorov (1957) proved the following theorem that becomes a theoretical basis of Eq. (4.1) in this regard.

Theorem For any integer $n \geq 2$ there are continuous real functions $\psi^{p,q}(x)$ on the closed unit interval $E^1 = [0;1]$ such that each continuous real function $f(x_1, \dots, x_n)$ on the n -dimensional unit cube E^n is representable as

$$f(x_1, \dots, x_n) = \sum_{q=1}^{q=2n+1} \chi_q \left[\sum_{p=1}^n \psi^{pq}(x_p) \right] \quad (4.6)$$

where $\chi_q(y)$ are continuous real functions

Cybenko (1989) proved that a single hidden layer MLP feed-forward Neural Network can approximate any bounded continuous and multivariate function with arbitrary precision. Hornik et al. (1989) also rigorously proved that the standard multilayer feed-forward network with as few as one hidden layer is capable of approximating any measurable function to any desired degree of accuracy such that multilayer feed-forward networks are a class of universal approximators. Therefore, one hidden layer is employed for our final architecture of MLP in the sense that networks with more than one hidden layer can be converted to an equivalent network with just one hidden layer for the purpose of approximating a nonlinear Autoregressive model.

4.2.3 Hidden layer neurons and training iterations

The cross validation technique is commonly used to determine the number of neurons in the hidden layer. According to a popular forum of Neural Network, Neural Network FAQ, (Sarle, 1997), an intelligent choice of the number of hidden neurons depends on whether the analyst is using early stopping or some other form of regularization. In our experiments, we used the cross validation technique with an early stopping policy to find the near optimal number of neurons in the hidden layer and the maximum number of training iterations for weight matrix optimization.

The technique of cross validation and early stopping policy requires partitioning the data set into three complementary subsets including training, validation, and testing sets. The training set is used for performing model estimation with different numbers of neurons in the hidden layer. A validation set is used to estimate the generalization error as the iterative training proceeds. The generalization error is periodically measured with the mean squared error. Early stopping prevents the network from overfitting by monitoring the generalization error on a validation set during the course of the iterative training. If the generalization error shows no further

improvement in a certain number of iterations or even begins to increase with further optimization, then the iterative training stops from further optimization. The number of neurons in the hidden layer that yields the minimum generalization error is chosen as the final parameter of Neural Network architecture.

Early stopping uses a validation set to decide when to stop the training. It answers one of the difficult questions of optimization regarding stopping criterion, and is an efficient method in terms of computing time to decide the number of training iterations that affects both optimization and generalization. The test set of data is used for evaluation of prediction performance after the number of neurons in the hidden layer and the training iterations are determined. In our study, the grid search method is implemented for choosing the optimal number of neurons in the hidden layer and the maximum training iterations.

Fig.4.4 shows how to determine the number of neurons in the hidden layer and training iterations for a $q \sim v$ model from location 2 as an example using cross validation and early stopping techniques for its Neural Network architecture. A neuron is added to the hidden layer one at a time and the weight matrix is iteratively optimized until the generalization error of the validation set stops improving or begins to climb up. The optimal number of neurons in the hidden layer is identified as 17 with 27 training iterations to produce minimum error 0.1929.

In the empirical study, three nonlinear optimization routines are used to train the Neural Networks including BPNN (standard backpropagation), BFGS (Broyden-Fletcher-Goldfarb-Shannon), and SCG (scaled conjugate gradient). These algorithms are available in RSNNS and nnet packages in the open source software R.

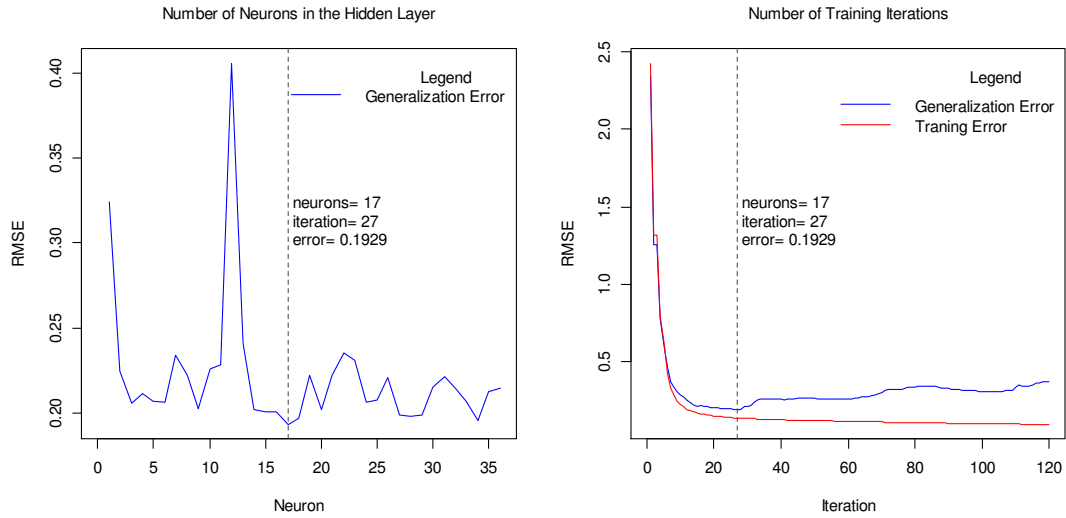


Fig. 4.4 Choose hidden layer neurons and training iterations, $q-v$, location 2

4.3 Support vector regression

Support vector regression (SVR) is selected to compare with TS-TVEC model because of its excellent performances in regression and time series prediction (Muller et al., 1997, Drucker et al., 1997, Stitson et al., 1999, Mattera and Haykin, 1999, Scholkopf and Smola, 2002, Cortes and Vapnik, 1995). The SVR is a machine learning algorithm that is closely analogical to nonparametric estimation. Developed from statistical learning theory that concentrates on kernel-based regression with quadratic optimization techniques, SVR is replacing Neural Network in a variety of fields.

4.3.1 Kernel method

In the context of support vector regression, kernels are introduced as similarity measures because they compute all geometric features including length of vector, angle and distance between vectors. Kernel is computed through an inner product function denoted by $\langle \dots \rangle$ which has the appealing property that its value is maximal whenever its arguments are equal. In order to employ inner products to measure similarity between objects, we need to represent them in a normed space ℓ_2^n which is defined next. As well, some important concepts are introduced as follows.

Definition (Features and feature space) (Herbrich, 2002) a function $\phi_i : \mathcal{X} \rightarrow R$ that maps each object $x \in \mathcal{X}$ to a real value $\phi_i(x)$ is called a feature. Combining n features ϕ_1, \dots, ϕ_n results in a feature mapping $\Phi : \mathcal{X} \rightarrow H \subseteq \ell_2^n$ and the space H is called a feature space, where, a normed space ℓ_2^n of sequences of length n is defined by $\ell_2^n \stackrel{\text{def}}{=} \left\{ (x_1, \dots, x_n) \in R^n \mid \sum_{i=1}^n |x_i|^2 < \infty \right\}$.

Definition (Kernel) (Herbrich, 2002) Suppose we are given a feature mapping $\Phi : \mathcal{X} \rightarrow H \subseteq \ell_2^n$. The kernel is the inner product function $k : \mathcal{X} \times \mathcal{X} \rightarrow R$ in H , i.e., for all $x_i, x_j \in \mathcal{X}$, $k(x_i, x_j) \stackrel{\text{def}}{=} \langle \Phi(x_i), \Phi(x_j) \rangle = \langle \mathbf{X}_i, \mathbf{X}_j \rangle$

The inner product function k is called the kernel. Any function involving the kernel complies with the Cauchy-Schwarz Inequality such that it is bounded and produces similar results.

Proposition (Cauchy-Schwarz inequality) (Scholkopf and Smola, 2002) If k is a positive definite kernel, and for all $x_1, x_2 \in \mathcal{X}$ input space, then

$$|k(x_1, x_2)|^2 \leq k(x_1, x_1) \cdot k(x_2, x_2). \text{ For all } \mathbf{X}, \mathbf{X}' \in H \text{ a feature space, then}$$

$$|\langle \mathbf{X}, \mathbf{X}' \rangle| \leq \|\mathbf{X}\| \cdot \|\mathbf{X}'\|.$$

By the Cauchy-Schwarz inequality, we know that the difference of kernel function evaluated at two image vectors is bounded. Hence, if we only consider parameter vectors with a constant norm, it follows that whenever two points are close to each other, any linear function would assign similar real values to them as well.

These two properties, linearity and bounded, make linear functions perfect candidates for designing the kernel classifier. Incorporating a kernel in the linear classifier, linear regression function is known as kernel classifier. The necessary and sufficient conditions of k to be a kernel are given by Mercer's theorem (Mercer, 1909).

Mercer's Theorem (1909): Suppose $k \in L_\infty(\mathcal{X} \times \mathcal{X})$ is a symmetric function, i.e., $k(x, x') = k(x', x)$, such that the integral operator $T_k : L_2(\mathcal{X}) \rightarrow L_2(\mathcal{X})$ given by

$$(T_k f)(x) := \int_{\mathcal{X}} k(x, x') f(x') dx'$$

is positive semidefinite, that is, for all $f \in L_2(\mathcal{X})$, we have

$$\int_{\mathcal{X}} \int_{\mathcal{X}} k(x, x') f(x) f(x') dx dx' \geq 0$$

Let $\psi_i \in L_2(\mathcal{X})$ be the eigenfunction of T_k associated with the eigenvalue $\lambda_i \geq 0$ and normalized such that

$$\|\psi_i\|^2 = \int_{\mathcal{X}} \psi_i^2(x) dx = 1, \text{ i.e. } \forall x \in \mathcal{X}: \int_{\mathcal{X}} k(x, x') \psi_i(x') dx' = \lambda_i \psi_i(x)$$

Then

1. $(\lambda_i)_{i \in \mathbb{N}} \in \ell_1$,
2. $\psi_i \in L_\infty(\mathcal{X})$,
3. $k(x, x') = \sum_{i=1}^{\infty} \lambda_i \psi_i(x) \psi_i(x')$ holds for all $(x, x') \in \mathcal{X}$ where the series converges absolutely and uniformly for all $(x, x') \in \mathcal{X}$.

Loosely speaking, Mercer's theorem means that if $\int_{\mathcal{X}} \int_{\mathcal{X}} k(x, x') f(x) f(x') dx dx' \geq 0$ for all $f \in L_2(\mathcal{X})$ holds, we can write $k(x, x')$ as a dot product or inner product in some feature space. Mercer's theorem not only gives necessary and sufficient conditions for k to be a kernel, but also suggests a constructive way of obtaining features Φ from a given kernel function k .

4.3.2 Support vector regression

The support vector regression is to estimate a linear function in Hilbert space through the kernel method and quadratic optimization technique. The linear function involving a kernel is known as kernel classifier. The kernel classifier takes the form

$$f(x) = \langle w, x \rangle + b \quad (4.7)$$

where $w \in \mathcal{X}$, $b \in R$. Regression of the kernel-based linear function (4.3) boils down to a quadratic optimization problem where the coefficients are estimated by maximizing soft margin as well as minimizing training error from the training samples. The primal objective function is

$$\begin{aligned} & \text{minimize} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) \\ & \text{subject to} \quad \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (4.8) \end{aligned}$$

Introducing Lagrange multipliers $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$, the Lagrange function can be constructed by the primal objective function subtracting the sum of all products of constraints and corresponding Lagrange multipliers (Fletcher, 1989, Bertsekas, 1995). Hence, the Lagrange function L takes the form

$$\begin{aligned} L : &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) - \sum_{i=1}^m (\eta_i \xi_i + \eta_i^* \xi_i^*) \\ & - \sum_{i=1}^m \alpha_i (\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) \\ & - \sum_{i=1}^m \alpha_i^* (\varepsilon + \xi_i^* - y_i + \langle w, x_i \rangle + b) \quad (4.9) \end{aligned}$$

The Lagrange function is a powerful theoretical mean that is employed to develop a dual objective function. Optimization can be achieved by minimizing the primal variables and simultaneously maximizing the Lagrange multipliers.

Let the partial derivatives of L in Eq. (4.5) with respect to the primal variables (w, b, ξ_i, ξ_i^*) equal to zero and substituting them into Eq. (4.4) yields the following dual optimization problem.

$$\begin{aligned} \text{maximize } W(\alpha, \alpha^*) &= -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) k(x_i, x_j) \\ &\quad - \varepsilon \sum_{i=1}^m (\alpha_i^* + \alpha_i) + \sum_{i=1}^m (\alpha_i^* - \alpha_i) y_i \quad (4.10) \\ \text{subject to } 0 \leq \alpha_i, \alpha_i^* \leq C &\text{ for all } i=1, \dots, m, \text{ and } \sum_{i=1}^m (\alpha_i - \alpha_i^*) = 0 \end{aligned}$$

where ε is the error, C is the upper bound, α_i, α_i^* are Lagrange multipliers, and $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$ is the kernel. Thus, the regression function turns to the form of Eq.(4.7) (Scholkopf and Smola, 2002).

$$f(x) = \sum_{i=1}^m (\alpha_i^* - \alpha_i) k(x_i, x) + b \quad (4.11)$$

In a summary, the SV regression procedures are shown in Fig.4.5. The input pattern is mapped into feature space by a mapping mechanism Φ . Then the kernel function $k(x, x')$ is evaluated by computing the dot product of the images of the input patterns in feature space. Finally a linear combination of the dot products weighted by $(\alpha_i - \alpha_i^*)$ plus the constant term b yields the prediction output. The similarity of outputs is usually measured in terms of a loss function.

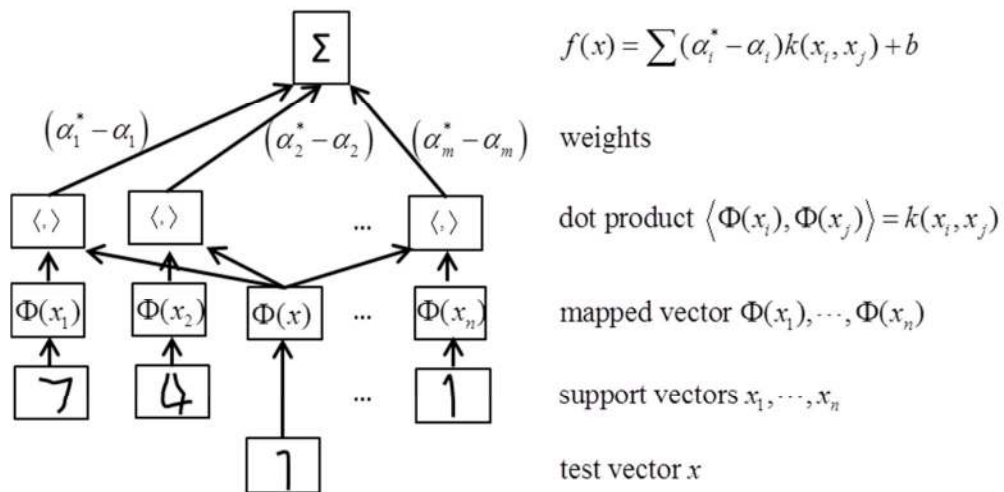


Fig. 4.5 Architecture of SVR and related kernel method (Scholkopf and Smola, 2002)

The process described here is very similar to regression in a neural network, with the difference, that in the SV case the weights in the input layer are a subset of the training patterns.

4.3.3 Parameter specification

In the study, the RBF (radial basis function) is chosen as the kernel function. It is highly effective in mapping nonlinear relationships. The γ is the key kernel parameter that needs to be chosen carefully, as it implicitly defines the structure of the high dimensional feature space and thus controls the complexity of the final solution. C is a general penalizing parameter for C -classification indicating the cost of constraints violation. It is referred to as the regularized constant that determines the tradeoff between the empirical error and the regularized term in the Lagrange formulation. The ε in the insensitive-loss function is called the tube size of SVR and is equivalent to the approximation accuracy placed on the training data.

Kernel parameter γ and cost parameter C were determined based on the cross validation method and grid search from the training data. The data is split into a training set and a test set. Hence, the `tune.svm(...)` function in the `e1071` package in R software is used for the purpose of parameter determination. The one-step-ahead rolling prediction is made based on the trained regression model.

4.4 Comparison of predictions

The effectiveness of the new TS-TVEC model is examined by comparing the accuracy of prediction performed by the TS-TVEC model, Neural Networks, and Support Vector Regression. Accuracy is measured in terms of MSE, Coefficient of Variation, and MAPE. The cross validation method with MSE criterion is used for the TS-TVEC model selection. Cross validation method with early stopping policy is used for Neural Network training and support vector regression.

Three sets of 60-minute interval data collected from three locations in Fig. 1.1 are used to test the effectiveness of the new TS-TVEC model. There are 27 time series from 3 locations involved in the modelling process. Three sets of 90-minute interval time series data are used to test the robustness of the TS-TVEC model. Table 4.16 shows the comparison of predictions between TS-TVEC and Neural Network using 90-minute interval traffic volume and speed time series. Table 4.17 shows the comparison of prediction performance between the new TS-TVEC models, Neural Network, and Support Vector Regression where one-step-ahead rolling predictions are

performed for 48-hour horizon based on 60-minute interval data. The comparison results indicate that the TS-TVEC model provides better prediction of traffic volume than Neural Network, and support vector regression. This observation holds in both 60-minute and 90-minute interval data. The average improvement in traffic volume prediction is approximately 6%. Numerically, the values of speed and occupancy are within a narrow range such that prediction error is more sensitive to a small change. The prediction coefficient of variation and MAPE of speed and occupancy by TS-TVEC model are close to the outcomes by Neural Network and support vector regression. Therefore, the cointegration model with regime switching mechanism provides a competitive approach to the non-linear forecasting problem for stochastic traffic processes in contrast to a gradient descent optimization method and kernel regression. Another observation is that the prediction errors by the 3 types of approaches are varying at different locations. It indicates that these models are data driven models. The selection of a forecasting method is associated with the data environment in question. As such, the method used in forecasting should be best suited for its corresponding data environment.

Table 4.16 Comparison of 90-minute interval one-step-ahead rolling prediction for 2 days

Model	Variable	Location 1					Location 2					Location 3					
		MSE	Coefficient of Variation	MAPE	Lag	Threshold	MSE	Coefficient of Variation	MAPE	Lag	Threshold	MSE	Coefficient of Variation	MAPE	Lag	Threshold	
TS-TVEC (q-v)	Volume	212668.30	0.07	0.07	3	1	195937.50	0.08	0.07	1	1	471574.30	0.11	0.07	2	1	
	Speed	30.63	0.05	0.04	3	1	217.52	0.13	0.08	1	1	298.15	0.15	0.07	2	1	
				Neuron	Iteration					Neuron	Iteration					Neuron	Iteration
MLP (BFGS) nnet pkg	Volume	459259.20	0.11	0.12	22	84	553020.30	0.14	0.17	12	15	678520.70	0.13	0.11	19	79	
	Speed	8.91	0.03	0.02			68.89	0.07	0.06			48.12	0.06	0.05			
MLP (BPNN) RSNNs pkg	Volume	3131926.00	0.29	0.49	7	9	712024.90	0.16	0.21	9	11	815408.20	0.14	0.24	7	58	
	Speed	85.91	0.08	0.08			353.34	0.17	0.15			38.05	0.05	0.05			

Fig.4.6 shows an example of overall model fitting and one-step-ahead rolling predictions of volume, speed and occupancy for a 48-hour horizon by the TS-TVEC model. It shows that the TS-TVEC model is able to capture both symmetrical deviation and asymmetrical sudden shifts. It is noteworthy that the predicted values oscillate closely around the observed values of three traffic variables. This justifies the error correction effect on deviations conditional on the threshold regime switching mechanism in the TS-TVEC model.

Table 4.17 Comparison of 60-minute interval one-step-ahead rolling prediction for 48 hours

Model	Variable	Location 1					Location 2					Location 3						
		MSE	Coefficient of Variation	MAPE	Lag	Threshold	MSE	Coefficient of Variation	MAPE	Lag	Threshold	MSE	Coefficient of Variation	MAPE	Lag	Threshold		
TS-TVEC (q-v)	Volume	63596.92	0.06	0.05	1	1	107671.10	0.09	0.08	6	1	114878.00	0.08	0.06	1	1		
	Speed	9.73	0.03	0.02	1	1	748.69	0.25	0.12	6	1	225.79	0.13	0.06	1	1		
TS-TVEC (q-o)	Volume	63955.81	0.06	0.05	1	1	89643.95	0.08	0.07	1	1	93322.65	0.07	0.06	2	1		
	Occupancy	0.58	0.18	0.17	1	1	7.94	0.36	0.26	1	1	4.38	0.35	0.16	2	1		
TS-TVEC (v-o)	Speed	11.27	0.03	0.02	5	1	201.01	0.13	0.08	4	1	263.72	0.14	0.06	2	2		
	Occupancy	0.50	0.17	0.17	5	1	7.35	0.34	0.19	4	1	5.64	0.40	0.17	2	2		
					Neuron	Iteration						Neuron	Iteration					
MLP (BFGS) nnet pkg	Volume	599755.20	0.19	0.17	12	103	114627.60	0.09	0.12	17	27	331446.20	0.14	0.18	11	23		
	Speed	9.64	0.03	0.02			148.87	0.11	0.09			23.99	0.04	0.03				
	Volume	584876.40	0.19	0.18	9	113	120539.60	0.10	0.11	12	20	127457.00	0.08	0.10	9	34		
	Occupancy	0.87	0.22	0.19			4.91	0.28	0.24			0.75	0.14	0.15				
	Speed	13.41	0.03	0.02	22	119	142.25	0.11	0.09	8	20	17.31	0.04	0.02	8	91		
	Occupancy	0.51	0.17	0.17			8.76	0.37	0.34			0.64	0.13	0.11				
MLP (BPNN) RSNNS pkg	Volume	467182.10	0.17	0.23	15	111	150218.50	0.11	0.10	18	42	493691.40	0.17	0.19	11	110		
	Speed	16.69	0.04	0.03			83.97	0.08	0.07			12.84	0.03	0.03				
	Volume	442952.80	0.16	0.26	13	113	157017.90	0.11	0.12	5	19	145323.10	0.09	0.11	19	116		
	Occupancy	0.91	0.23	0.16			8.98	0.38	0.32			1.23	0.19	0.23				
	Speed	18.79	0.04	0.04	4	16	56.72	0.07	0.05	9	18	22.91	0.04	0.03	6	120		
	Occupancy	3.88	0.47	0.71			3.44	0.23	0.27			1.69	0.22	0.29				
MLP (SCG) RSNNS pkg	Volume	438179.10	0.16	0.20	7	84	109982.10	0.09	0.09	17	97	293717.30	0.13	0.15	14	92		
	Speed	14.68	0.03	0.03			114.59	0.10	0.08			19.97	0.04	0.03				
	Volume	504681.40	0.17	0.25	16	32	193944.30	0.12	0.12	13	75	119951.80	0.08	0.08	16	118		
	Occupancy	1.60	0.30	0.21			2.91	0.22	0.17			0.68	0.14	0.15				
	Speed	7.01	0.02	0.02	14	71	122.24	0.10	0.07	9	37	14.99	0.03	0.02	14	110		
	Occupancy	1.38	0.28	0.34			9.08	0.38	0.39			0.82	0.15	0.19				
					C	Y						C	Y					
SVR e1071 pkg	Volume	545422.10	0.18	0.18	5	1.10	116394.70	0.10	0.09	2	0.10	227701.20	0.11	0.14	2	1.10		
	Speed	6.79	0.02	0.02	$\epsilon =$	0.01	22.41	0.04	0.03	$\epsilon =$	0.01	24.82	0.04	0.02	$\epsilon =$	0.01		
	Volume	606371.90	0.19	0.18	4	1.10	259361.50	0.14	0.18	20	0.10	146252.60	0.09	0.09	2	0.10		
	Occupancy	0.71	0.20	0.15	$\epsilon =$	0.01	3.45	0.24	0.20	$\epsilon =$	0.01	0.39	0.10	0.09	$\epsilon =$	0.01		
	Speed	1.35	0.01	0.01	20	0.10	18.76	0.04	0.03	4	0.10	13.03	0.03	0.02	1	0.10		
	Occupancy	0.08	0.07	0.07	$\epsilon =$	0.01	1.46	0.15	0.14	$\epsilon =$	0.01	0.58	0.13	0.11	$\epsilon =$	0.01		

In contrast, Fig.4.7 and Fig.4.8 shows overall model fitting and one-step-ahead rolling predictions of volume, speed, and occupancy for a 48-hour horizon by the MLP Neural Network and Support Vector Regression, respectively. It shows that both MLP network and SVR model are able to capture seasonal patterns of volume and occupancy without pre-processing the data. It appears that both MLP and SVR predictions lack to follow the overall curve of observed values.

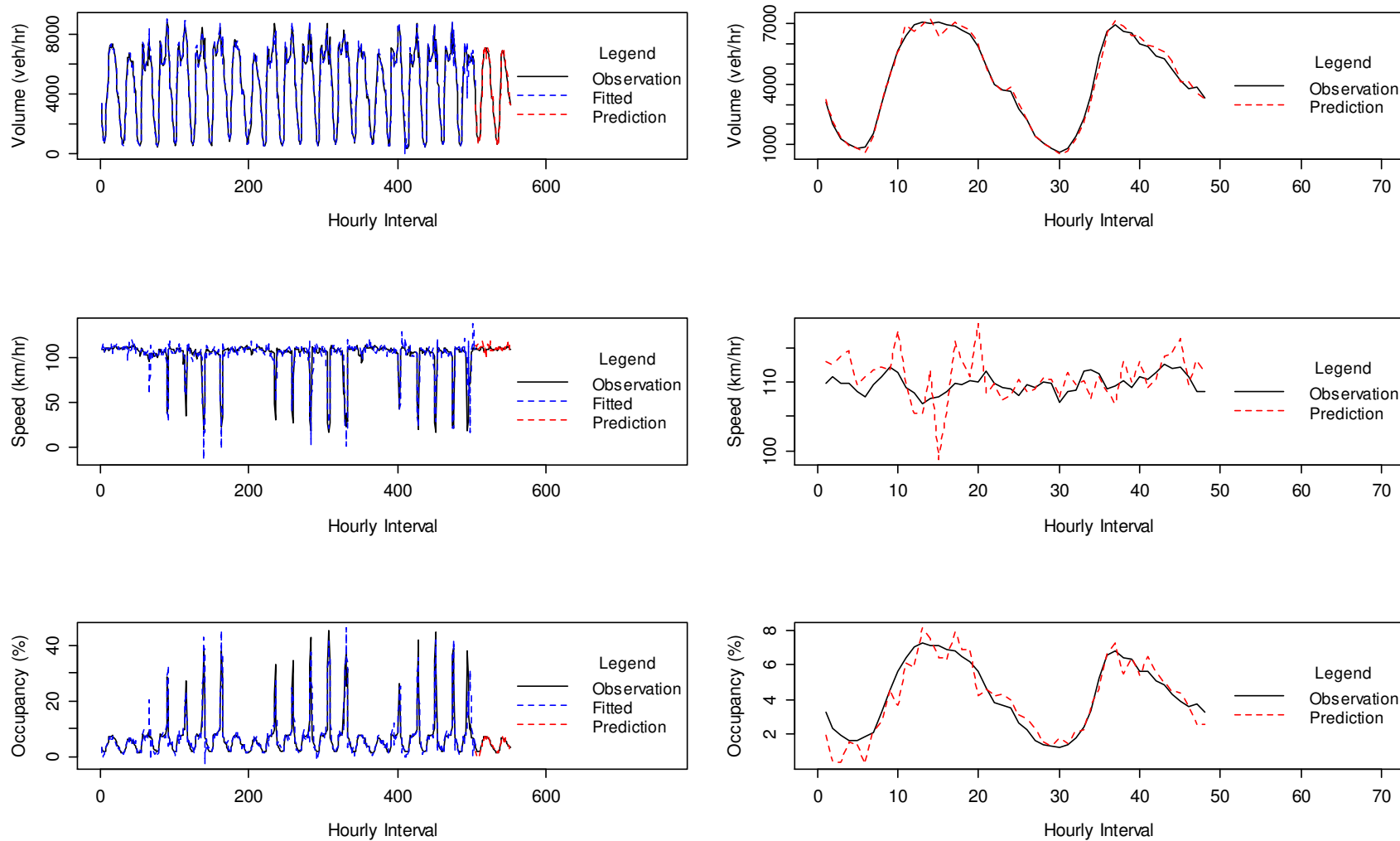


Fig. 4.6 TS-TVEC model fit and one-step-ahead rolling prediction for 48 hours, location 1

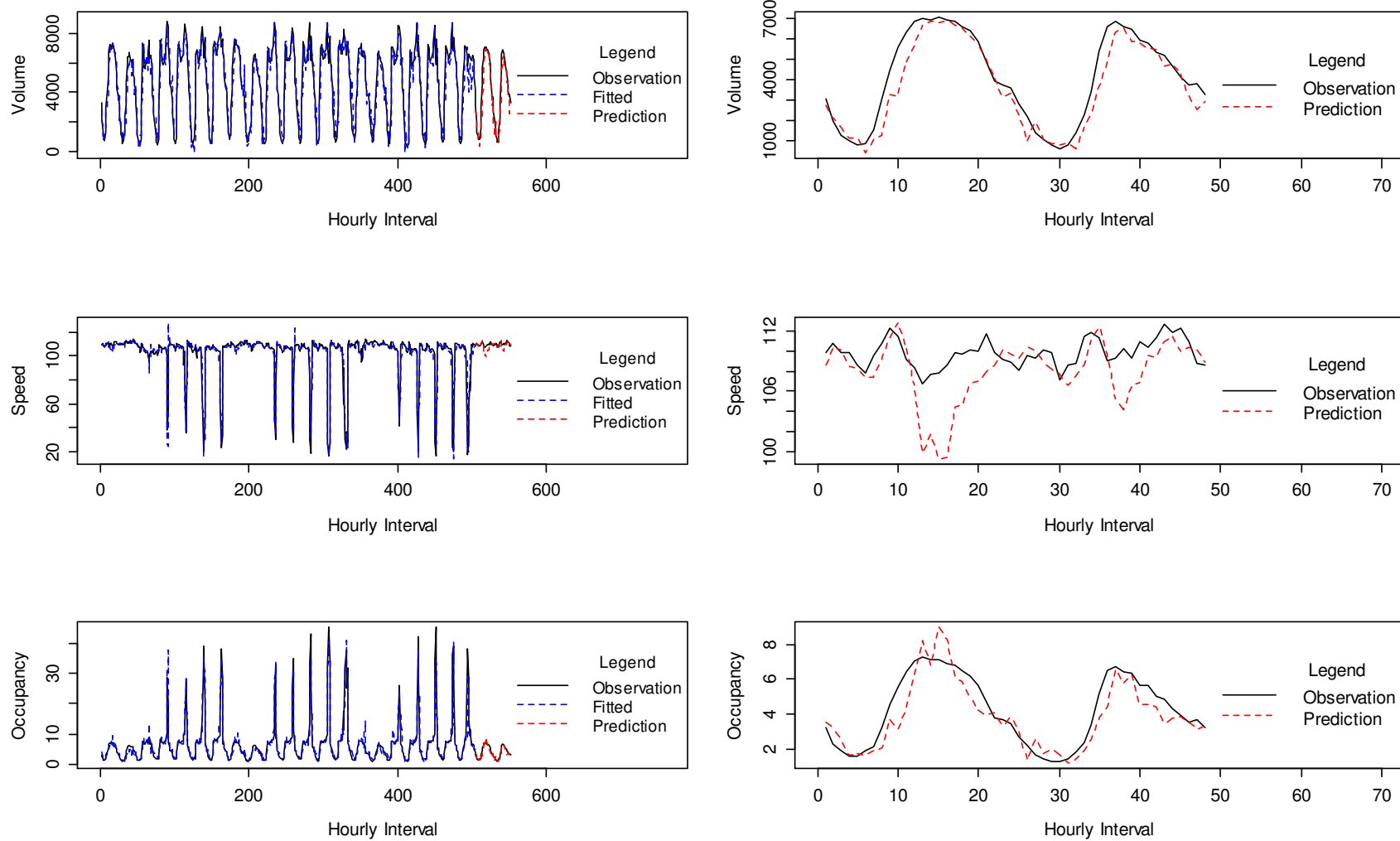


Fig. 4.7 MLP model fit and one-step-ahead rolling prediction for 48 hours, location 1

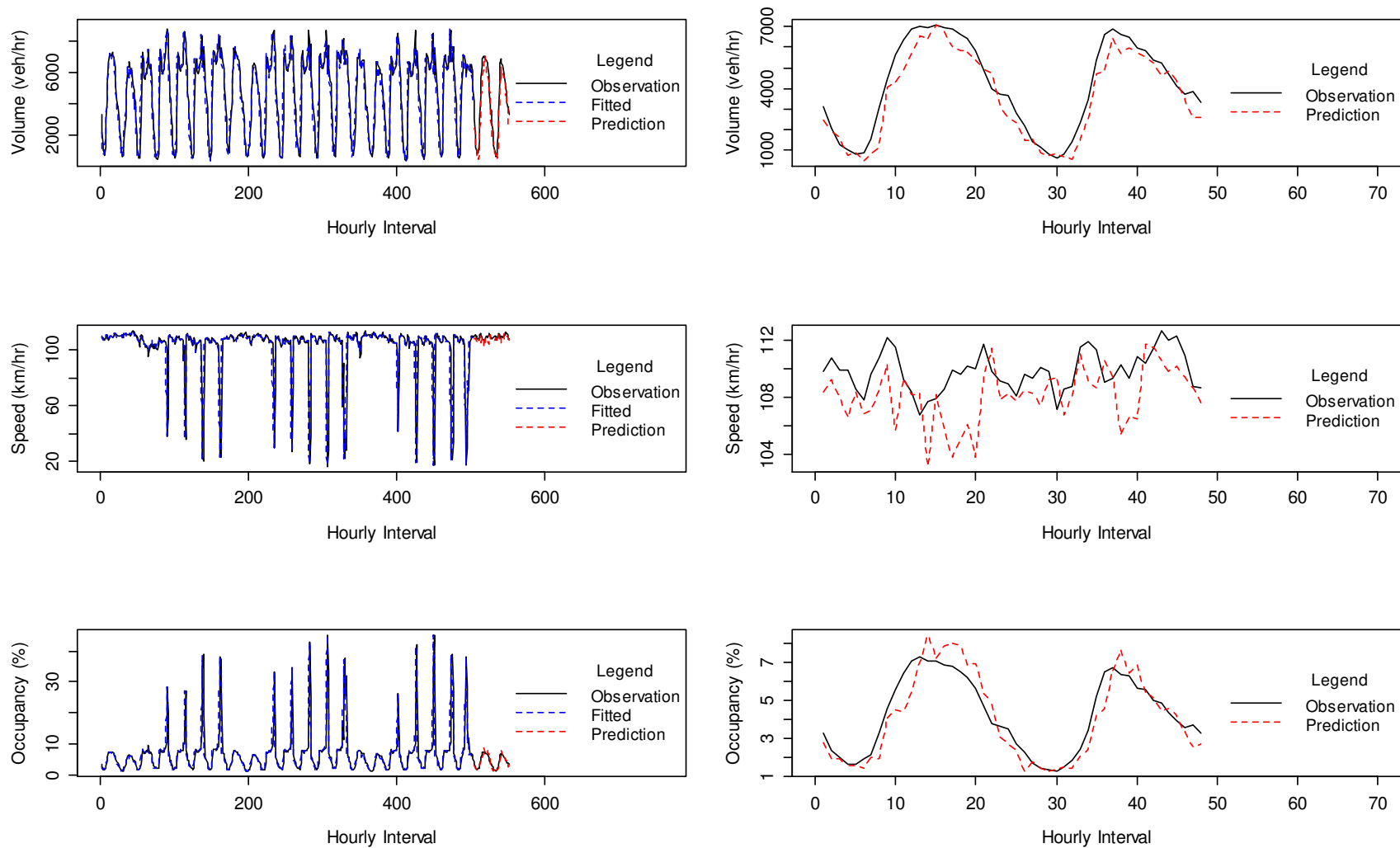


Fig. 4.8 SVR model fit and one-step-ahead rolling prediction for 48 hours, location 1

Chapter 5

Large scale application

In chapters 1 and 4, the TS-TVEC model is introduced and estimated with one-month long time series and its predictive performance is assessed with 48 hourly data from three locations in Fig.1.1. In order to further assess its predictive capability, effectiveness, robustness, and consistency, this chapter presents large scale applications of the methodology. The method is applied to a larger area that covers 35 sites of prediction on Highway 401, 400, and 404 in the Greater Toronto Area (GTA), Ontario, Canada, and exposed to more complex traffic situations with approximately 315 time series. Three TS-TVEC models are estimated for each site and one-step-ahead rolling prediction is performed for each traffic variable by a selected model for consecutive 168 hours, i.e. 7 days.

5.1 Highway network

Traffic time series are collected from loop detectors deployed on the GTA 400 series of freeways. The data collection stations on the freeways are managed and maintained by the Ministry of Transportation Ontario (MTO), and the traffic data are provided to the ITS lab at University of Toronto for purposes of research. Fig.5.1 shows the sites of prediction and related upstream and downstream sites on Highway 401 expressway east and west bound, and Highway 400 and 404 north and south bound. Fig.5.2 shows the sites of prediction and related upstream and downstream sites on Highway 401 collector east and west bound. In total, there are 35 sites of prediction denoted by red dots on the two maps that represent 35 highway segments. The locations of prediction selected on Highway 401 focus on core sections between Highway 400 and 404 that assume heavy traffic volumes in, out, and passing through the City of Toronto on a daily basis. These chosen locations cover 12 stretches on the Highway 401 expressway and 17 stretches on the collector road. Four locations of prediction are selected on Highway 400 north and south bound between Highway 401 and 407. Two locations are selected on Highway 404 north and south bound between Highway 401 and Finch Avenue. Table 5.1 lists the locations of prediction, loop detector ID, and road stretches.

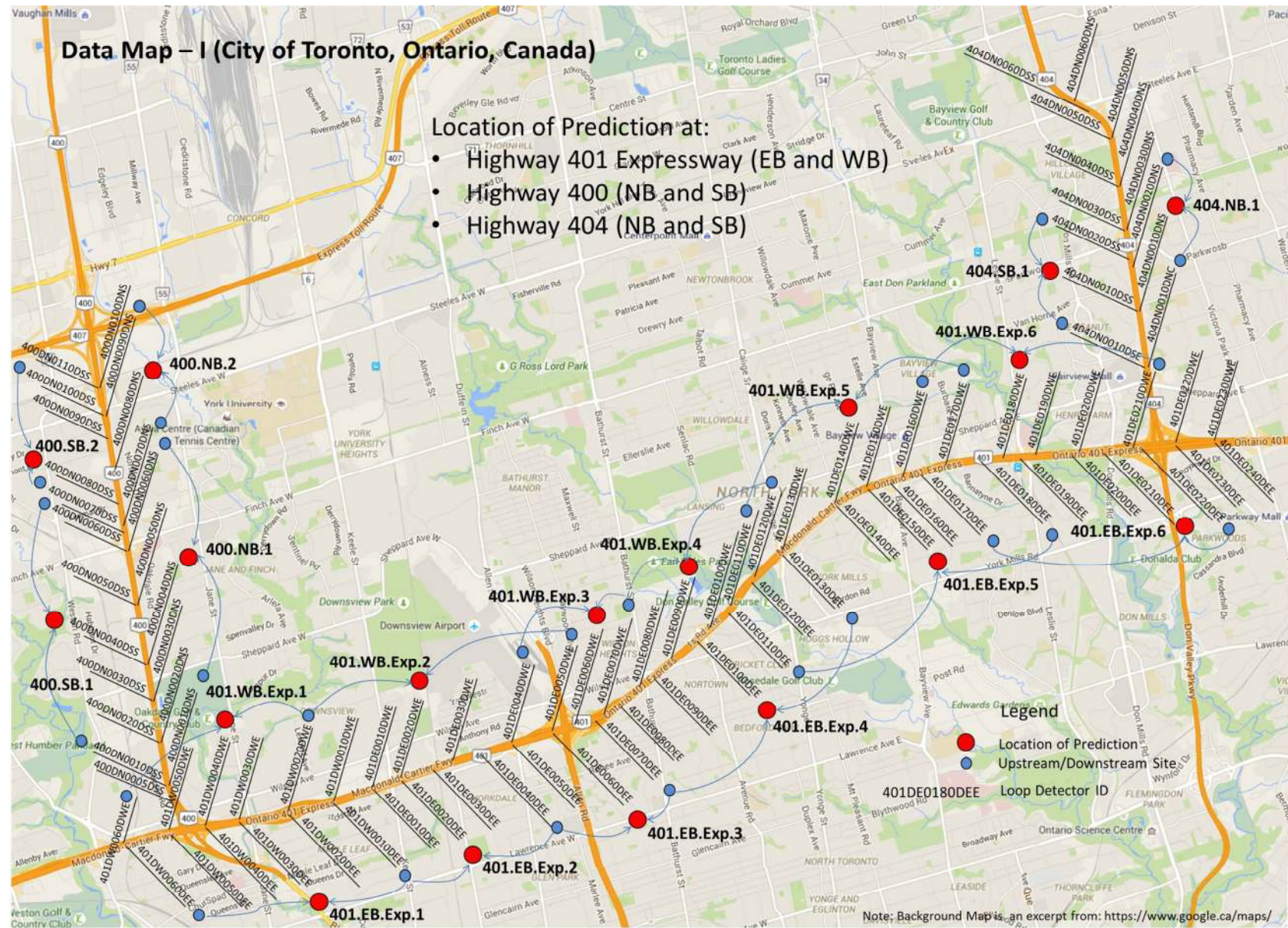


Fig. 5.1 the GTA 400 series of Highways and data collecting stations (1)

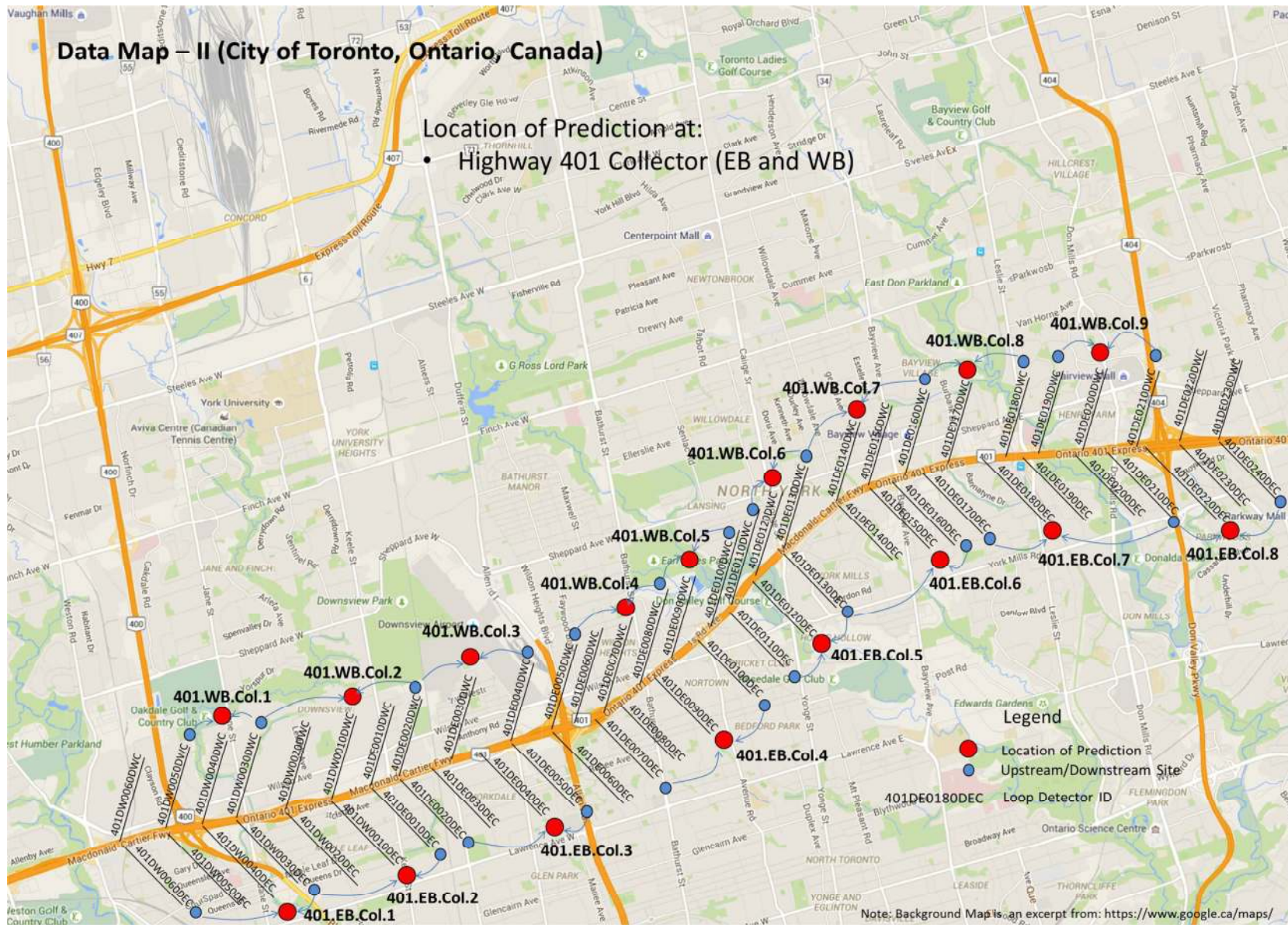


Fig. 5.2 the GTA 400 series of Highways and data collecting stations (2)

Table 5.1 Locations of prediction

Name	Location Index of Prediction	Detector ID	Road Section
HWY401 Expressway Eastbound	401.EB.Exp.1	401DW0030DEE	Hwy400 - Basket Wave.
	401.EB.Exp.2	401DE0020DEE	Basket Wave - Dufferin St.
	401.EB.Exp.3	401DE0060DEE	Allen Rd. @ Hwy401
	401.EB.Exp.4	401DE0100DEE	Allen Rd. - Yonge St.
	401.EB.Exp.5	401DE0150DEE	Yonge St. - Leslie St.
	401.EB.Exp.6	401DE0210DEE	Leslie St. - Hwy404
HWY401 Collector Eastbound	401.EB.Col.1	401DW0040DEC	Hwy400 - Jane St.
	401.EB.Col.2	401DW0010DEC	Jane St. - Keele St.
	401.EB.Col.3	401DE0040DEC	Keele St. - Dufferin St.
	401.EB.Col.4	401DE0090DEC	Bathurst St. - Avenue Rd.
	401.EB.Col.5	401DE0120DEC	Avenue Rd. - Yonge St.
	401.EB.Col.6	401DE0150DEC	Yonge St. - Bayview Ave.
	401.EB.Col.7	401DE0180DEC	Bayview Ave. - Leslie St.
	401.EB.Col.8	401DE0220DEC	Leslie St. - Hwy404
HWY401 Expressway Westbound	401.WB.Exp.1	401DW0040DWE	Hwy400 - Basket Wave.
	401.WB.Exp.2	401DE0020DWE	Basket Wave - Dufferin St.
	401.WB.Exp.3	401DE0060DWE	Allen Rd. @ Hwy401
	401.WB.Exp.4	401DE0090DWE	Bathurst St. - Yonge St.
	401.WB.Exp.5	401DE0140DWE	Yonge St. - Bayview Ave.
	401.WB.Exp.6	401DE0180DWE	Bayview Ave. - Hwy404
HWY401 Collector Westbound	401.WB.Col.1	401DW0040DWC	Hwy400 - Jane St.
	401.WB.Col.2	401DW0010DWC	Jane St. - Keele St.
	401.WB.Col.3	401DE0030DWC	Keele St. - Dufferin St.
	401.WB.Col.4	401DE0070DWC	Allen Rd. - Bathurst St.
	401.WB.Col.5	401DE0090DWC	Bathurst St. - Avenue Rd.
	401.WB.Col.6	401DE0120DWC	Avenue Rd. - Yonge St.
	401.WB.Col.7	401DE0140DWC	Yonge St. - Bayview Ave.
	401.WB.Col.8	401DE0170DWC	Bayview Ave. - Leslie St.
	401.WB.Col.9	401DE0200DWC	Leslie St. - Hwy404
HWY400 North & Southbound	400.NB.1	400DN0040DNS	Hwy401 - Finch Ave. W.
	400.NB.2	400DN0080DNS	Finch Ave. W. - Steeles Ave. W.
	400.SB.1	400DN0040DSS	Hwy401 - Finch Ave. W.
	400.SB.2	400DN0080DSS	Finch Ave. W. - Steeles Ave. W.
HWY404 North & Southbound	404.NB.1	404DN0010DNS	Sheppard Ave. E. - Finch Ave. E.
	404.SB.1	404DN0010DSS	Sheppard Ave. E. - Finch Ave. E.

5.2 Time series data

Fig. 5.3 shows typical two-month long hourly traffic time series including volume, speed, and occupancy from the Highway 401 collector road at Avenue Road. These plots represent most of traffic time series with some variations in the data set. Usually, a longer time series may present a larger variety of traffic situations including both recurring and non-recurring traffic incidents and congestion. Seasonality can be easily observed from the traffic volume and occupancy series. Non-stationary features and traffic state alternation can be identified from the traffic speed series.

Four statistical tests are performed to verify the existence of cointegration and threshold effect among traffic variables. Table 5.2 shows the results of the Phillips-Ouliaris cointegration test. The null hypothesis of no cointegration is rejected at 5% significance. The existence of cointegration between traffic variables is further verified by results of Johansen cointegration test (Eigen) shown in Table 5.3. Hansen and Seo test results in Table 5.4 reject the null hypothesis of linear cointegration at 5% significance and favor the threshold cointegration effect. Correspondently, Fig. 5.4 shows the density bootstrap distribution of the test statistics of three Hansen and Seo tests respectively. All p -values are zero. The Zivot-Andrew unit root test is also performed to verify the structural break in the speed and occupancy time series. Fig. 5.5 and 5.6 show that the null hypothesis of a unit root process with drift that excludes exogenous structural change is rejected at 5% significance.

The results of those statistical tests indicate that the threshold cointegration effect exists among traffic volume, speed and occupancy. The necessity of the TS-TVEC model is justified. In addition, the spatial time series that is used as the exogenous term of the model are chosen from the upstream and downstream site where the traffic flow converges or diverges from the traffic flow at the site of prediction. Spatial time series from the upstream and downstream site should be physically related but not identical to the traffic flow at the location of prediction.

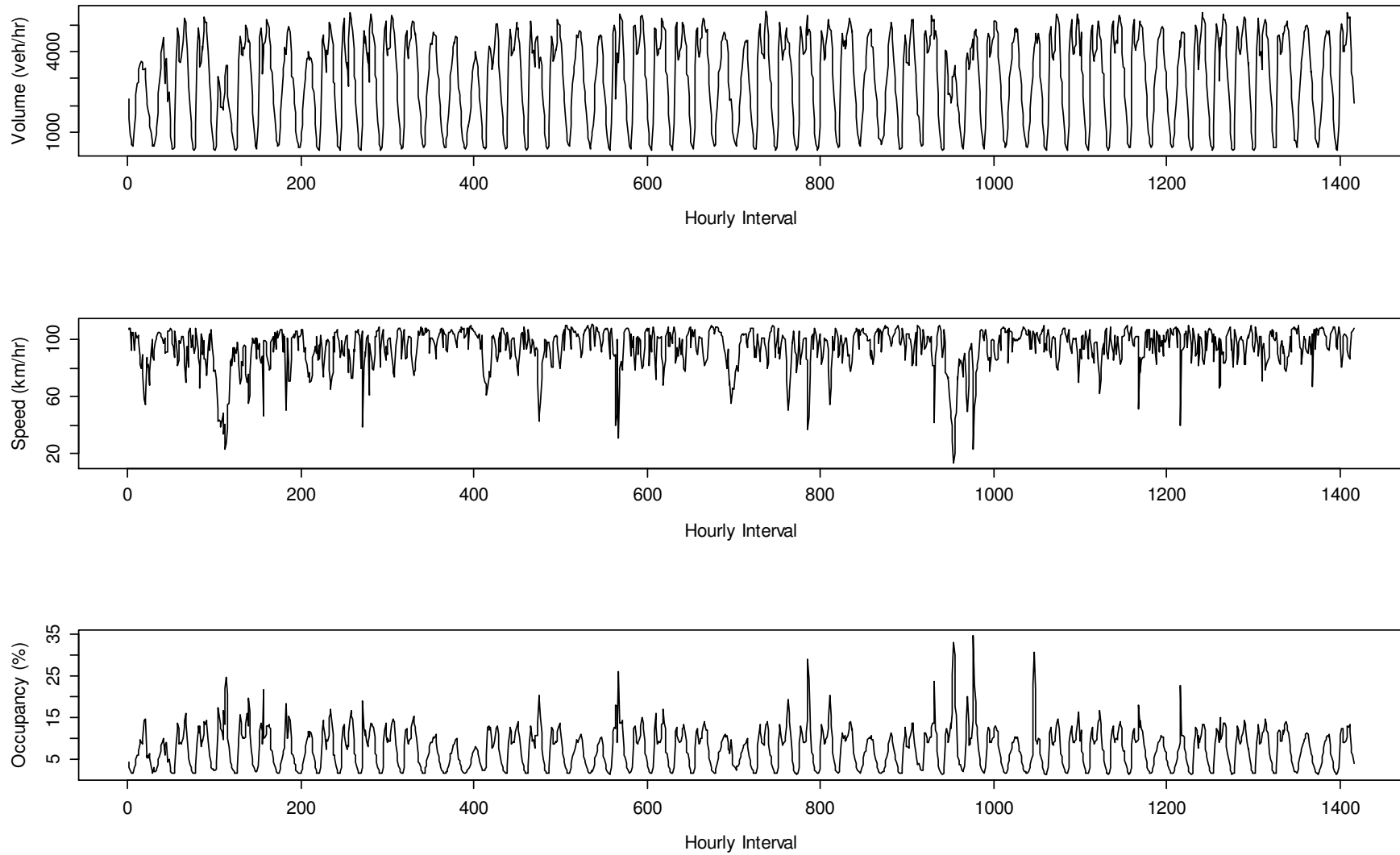


Fig. 5.3 typical traffic time series from the data set (401DE0100DEC, 02-03/2014)

Table 5.2 Phillips-Ouliaris cointegration test

Pair of Traffic Variables	Type	Test Statistic	p-value	Critical Values		
				10%	5%	1%
<i>Volume vs. Speed</i>	Pu	486.69	p < 1%	27.85	33.71	48.00
	Pz	515.47	p < 1%	47.59	55.22	71.93
<i>Volume vs. Occupancy</i>	Pu	660.25	p < 1%	27.85	33.71	48.00
	Pz	873.49	p < 1%	47.59	55.22	71.93
<i>Speed vs. Occupancy</i>	Pu	729.52	p < 1%	27.85	33.71	48.00
	Pz	568.67	p < 1%	47.59	55.22	71.93

Table 5.3 Johansen cointegration test (eigenvalue)

Pair of Traffic Variables	Rank	Test Statistic	p-value	Critical Values		
				10%	5%	1%
<i>Volume vs. Speed</i>	r<=1	157.53	p < 1%	6.5	8.18	11.65
	r=0	211.21	p < 1%	12.91	14.9	19.19
<i>Volume vs. Occupancy</i>	r<=1	201.43	p < 1%	6.5	8.18	11.65
	r=0	235.00	p < 1%	12.91	14.9	19.19
<i>Speed vs. Occupancy</i>	r<=1	155.51	p < 1%	6.5	8.18	11.65
	r=0	169.20	p < 1%	12.91	14.9	19.19

Table 5.4 Hansen and Seo threshold cointegration test

Multivariate	Test Statistic	P-Value	Bootstrap Critical Values		
			0.90	0.95	0.99
<i>Volume vs. Speed</i>	151.93	0.00	13.88	15.32	17.57
<i>Volume vs. Occupancy</i>	126.19	0.00	10.39	12.00	15.11
<i>Speed vs. Occupancy</i>	92.19	0.00	18.78	21.03	24.00

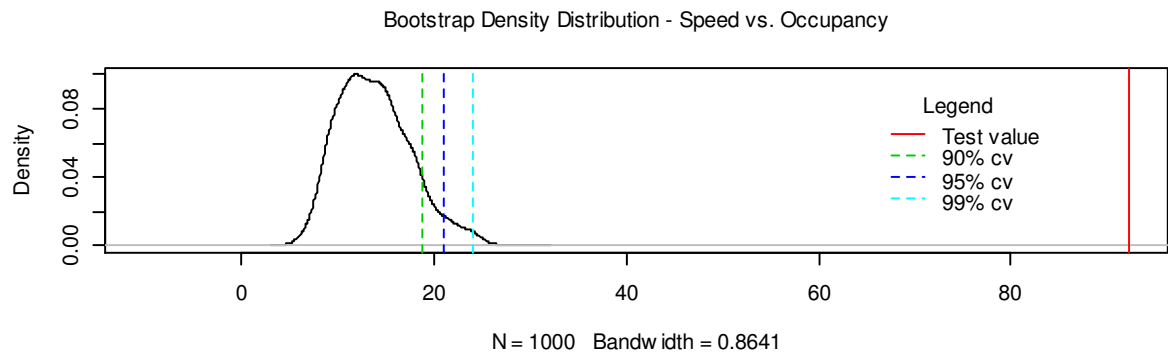
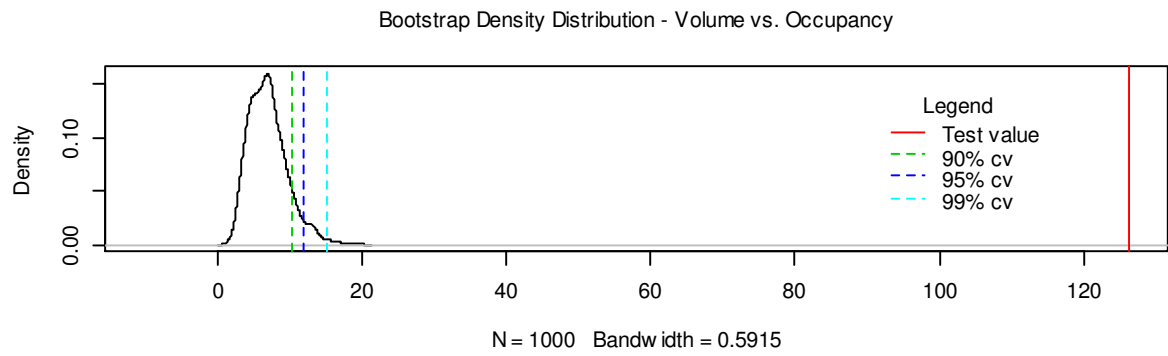
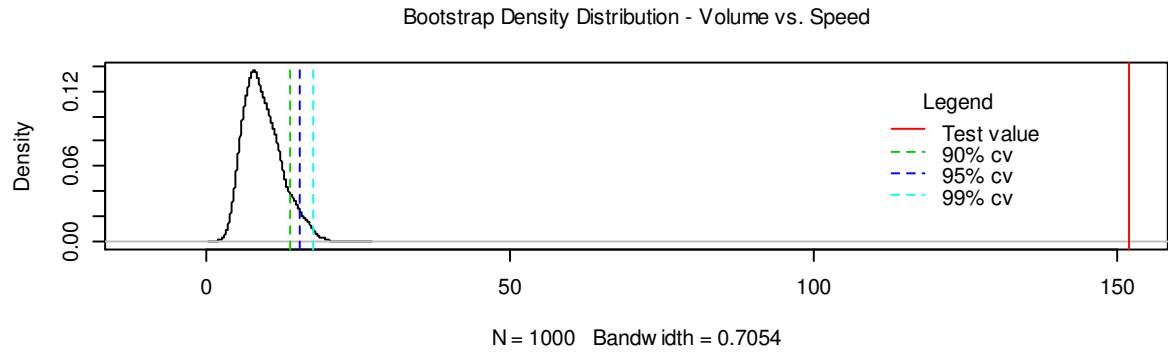


Fig. 5.4 Bootstrap density distribution of Hansen and Seo (2002) test

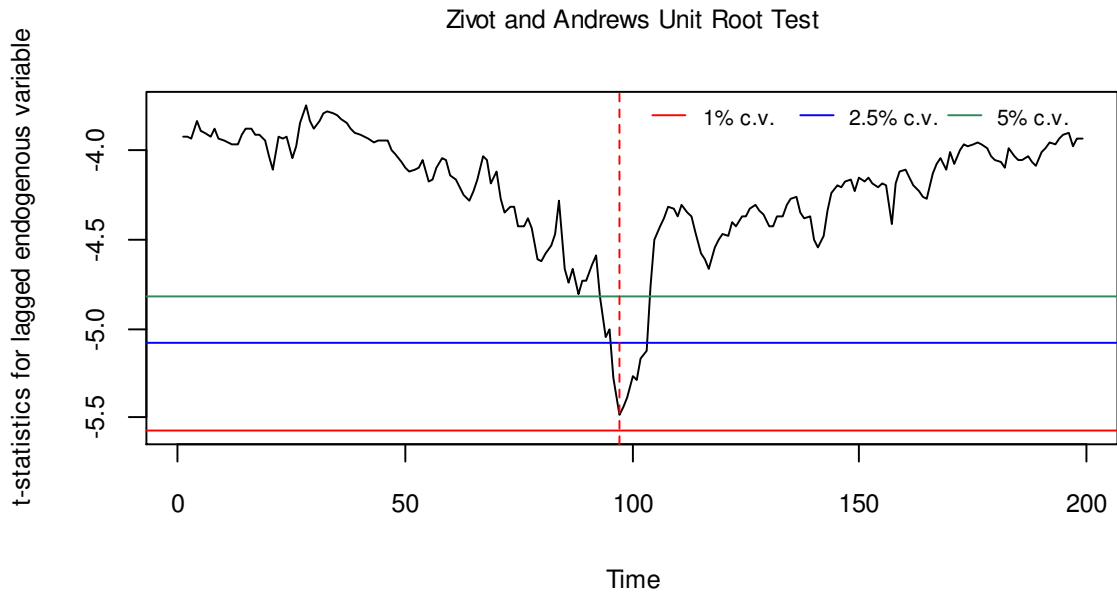


Fig. 5.5 Structural break test with 200 points of speed series

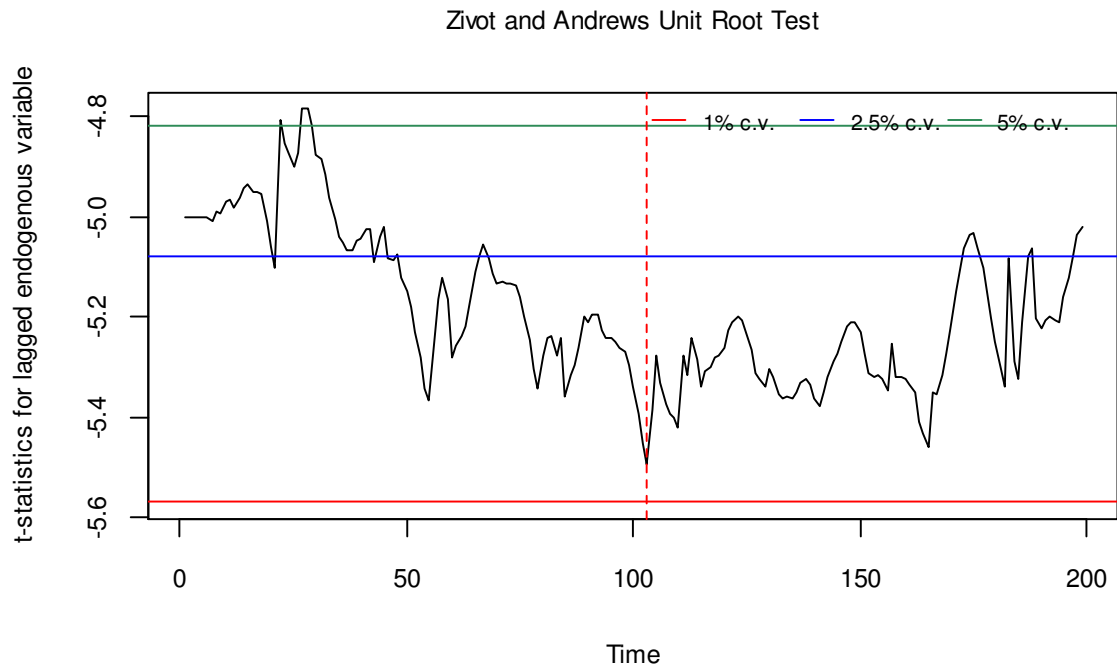


Fig. 5.6 Structural break test with 200 points of occupancy series

5.3 Model determination

For each site of prediction, three TS-TVEC models are estimated respectively based on three pairs of cointegrated traffic variables, i.e. volume vs. speed, volume vs. occupancy, and speed vs. occupancy. Traffic variable prediction is performed with the selected model. Model selection

concurrently takes into account the MSE, number of regimes, and lags. The cross validation method in conjunction with parsimonious principle is used for model selection. Statistical diagnostic tests are performed to examine the significance of the model coefficients, normality, and whiteness of noise in the process of model estimation. Results of model selection and specification for Highway 401 expressway and collector east and west bound, Highway 400 and 404 north and south bound are shown in Tables 5.5, 5.6, 5.7, 5.8, and 5.9 respectively.

Table 5.5 Model selection, specification and prediction Highway 401 Col WB

Location	Predictive Variable	Model Specification					Accuracy of Prediction		
		vol~spd	vol~occ	spd~occ	Lag	Threshold	MSE	Coefficient of Variation	MAPE
401.WB.Col.1 Hwy400 - Jane St. 401DW0040DWC	volume		x		6	1	121677.10	0.1944	0.1304
	speed			x	2	1	227.37	0.1492	0.1162
	occupancy			x	2	1	11.29	0.3787	0.1999
401.WB.Col.2 Jane St. - Keele St. 401DW0010DWC	volume	x			1	2	57355.63	0.0625	0.0567
	speed			x	2	2	65.91	0.0791	0.0528
	occupancy		x		1	1	3.60	0.2924	0.1790
401.WB.Col.3 Keele St. - Dufferin St. 401DE0030DWC	volume	x			1	1	18864.70	0.4129	0.2303
	speed			x	1	1	40.93	0.1293	0.1069
	occupancy			x	1	1	1.06	0.6925	0.5801
401.WB.Col.4 Allen Rd. - Bathurst St. 401DE0070DWC	volume	x			7	1	47082.57	0.0721	0.0726
	speed			x	1	1	142.46	0.1266	0.0785
	occupancy			x	1	1	4.62	0.3845	0.1529
401.WB.Col.5 Bathurst St. - Avenue Rd. 401DE0090DWC	volume	x			1	1	104457.10	0.0803	0.0750
	speed	x			1	1	16.14	0.0416	0.0310
	occupancy		x		4	1	0.40	0.1165	0.0924
401.WB.Col.6 Avenue Rd. - Yonge St. 401DE0120DWC	volume	x			10	1	57849.68	0.0771	0.0726
	speed	x			10	1	89.52	0.1113	0.0929
	occupancy		x		10	1	5.43	0.2480	0.2002
401.WB.Col.7 Yonge St. - Bayview Ave. 401DE0140DWC	volume		x		1	2	84243.81	0.1115	0.0722
	speed	x			1	2	284.64	0.1903	0.1597
	occupancy		x		1	2	6.49	0.4884	0.3897
401.WB.Col.8 Bayview Ave. - Leslie St. 401DE0170DWC	volume		x		2	1	128246.90	0.1257	0.0766
	speed			x	2	1	96.91	0.0990	0.0663
	occupancy			x	2	1	8.71	0.6790	0.2457
401.WB.Col.9 Leslie St. - Hwy404 401DE0200DWC	volume	x			2	1	182919.50	0.0925	0.0864
	speed			x	5	1	265.11	0.1993	0.1658
	occupancy			x	5	1	14.03	0.4097	0.3142

Table 5.6 Model selection, specification and prediction Highway 401 Exp EB

Location	Predictive Variable	Model Specification					Accuracy of Prediction		
		vol~spd	vol~occ	spd~occ	Lag	Threshold	MSE	Coefficient of Variation	MAPE
401.EB.Exp.1 Hwy400 - Basket Wave. 401DW0030DEE	volume		x		3	1	163350.70	0.1180	0.0764
	speed			x	10	1	143.81	0.1234	0.1004
	occupancy			x	10	1	9.61	0.3123	0.1964
401.EB.Exp.2 Basket Wave - Dufferin St. 401DE0020DEE	volume		x		6	2	296926.80	0.1394	0.0949
	speed			x	3	1	93.34	0.1078	0.0958
	occupancy			x	3	1	5.79	0.2150	0.1306
401.EB.Exp.3 Allen Rd. @ Hwy401 401DE0060DEE	volume	x			4	1	235112.10	0.1279	0.0759
	speed	x			4	1	162.86	0.1304	0.1192
	occupancy		x		9	1	9.19	0.2949	0.1707
401.EB.Exp.4 Allen Rd. - Yonge St. 401DE0100DEE	volume	x			3	1	79108.83	0.0828	0.0533
	speed			x	8	1	151.49	0.1205	0.0949
	occupancy		x		3	1	7.61	0.2537	0.1441
401.EB.Exp.5 Yonge St. - Leslie St. 401DE0150DEE	volume	x			1	1	193569.10	0.1045	0.0650
	speed			x	8	2	138.79	0.1139	0.0779
	occupancy			x	8	2	9.89	0.3978	0.1738
401.EB.Exp.6 Leslie St. - Hwy404 401DE0210DEE	volume		x		1	1	76487.54	0.1167	0.0667
	speed	x			1	1	89.54	0.1038	0.0852
	occupancy		x		1	1	6.05	0.3825	0.1563

Table 5.7 Model selection, specification and prediction Highway 401 Exp WB

Location	Predictive Variable	Model Specification					Accuracy of Prediction		
		vol~spd	vol~occ	spd~occ	Lag	Threshold	MSE	Coefficient of Variation	MAPE
401.WB.Exp.1 Hwy400 - Basket Wave. 401DW0040DWE	volume	x			2	1	177438.50	0.0907	0.0633
	speed	x			2	1	258.22	0.1638	0.1284
	occupancy		x		2	1	8.33	0.3604	0.2072
401.WB.Exp.2 Basket Wave - Dufferin St. 401DE0020DWE	volume		x		1	1	145905.50	0.0901	0.0785
	speed	x			3	1	232.74	0.1563	0.1105
	occupancy		x		1	1	6.85	0.2718	0.1469
401.WB.Exp.3 Allen Rd. @ Hwy401 401DE0060DWE	volume		x		5	1	148765.10	0.0971	0.0803
	speed			x	2	2	140.96	0.1106	0.0828
	occupancy		x		5	1	8.54	0.4377	0.2168
401.WB.Exp.4 Bathurst St. - Yonge St. 401DE0090DWE	volume	x			1	2	233762.40	0.1343	0.0948
	speed			x	7	1	230.57	0.2043	0.1527
	occupancy			x	7	1	9.02	0.3109	0.2268
401.WB.Exp.5 Yonge St. - Bayview Ave. 401DE0140DWE	volume		x		6	2	165100.50	0.0914	0.0706
	speed	x			5	2	213.77	0.1458	0.1077
	occupancy			x	2	2	12.73	0.3584	0.2133
401.WB.Exp.6 Bayview Ave. - Hwy404 401DE0180DWE	volume	x			3	1	107052.40	0.1312	0.0963
	speed			x	1	1	155.53	0.1417	0.1210
	occupancy			x	1	1	9.05	0.3434	0.2221

Table 5.8 Model selection, specification and prediction Highway 401 Col EB

Location	Predictive Variable	Model Specification					Accuracy of Prediction		
		vol~spd	vol~occ	spd~occ	Lag	Threshold	MSE	Coefficient of Variation	MAPE
401.EB.Col.1 Hwy400 - Jane St. 401DW0040DEC	volume	x			1	1	72905.69	0.1271	0.0787
	speed			x	1	1	232.01	0.1669	0.1687
	occupancy		x		2	1	40.56	0.8089	0.4016
401.EB.Col.2 Jane St. - Keele St. 401DW0010DEC	volume		x		8	1	103069.60	0.0951	0.0713
	speed			x	13	1	195.34	0.1614	0.1376
	occupancy			x	13	1	15.67	0.4293	0.3149
401.EB.Col.3 Keele St. - Dufferin St. 401DE0040DEC	volume	x			4	2	185171.80	0.1268	0.1017
	speed			x	1	2	141.58	0.1331	0.0952
	occupancy			x	1	2	12.38	0.4365	0.2319
401.EB.Col.4 Bathurst St. - Avenue Rd. 401DE0090DEC	volume	x			5	1	125111.40	0.1015	0.1413
	speed	x			5	1	155.59	0.1630	0.1219
	occupancy			x	7	1	41.89	0.5467	0.3162
401.EB.Col.5 Avenue Rd. - Yonge St. 401DE0120DEC	volume	x			3	2	146109.10	0.1343	0.0887
	speed	x			3	2	65.40	0.0832	0.0426
	occupancy			x	3	2	2.21	0.3391	0.2226
401.EB.Col.6 Yonge St. - Bayview Ave. 401DE0150DEC	volume		x		3	1	68916.99	0.0929	0.0966
	speed			x	1	1	261.25	0.1549	0.1058
	occupancy			x	1	1	20.98	0.9740	0.6263
401.EB.Col.7 Bayview Ave. - Leslie St. 401DE0180DEC	volume	x			1	1	51965.17	0.0745	0.0579
	speed			x	4	2	108.39	0.1154	0.0984
	occupancy			x	4	2	11.59	0.3540	0.2261
401.EB.Col.8 Leslie St. - Hwy404 401DE0220DEC	volume		x		5	1	21101.93	0.0877	0.0700
	speed			x	2	2	56.03	0.0720	0.0426
	occupancy			x	2	2	2.36	0.4742	0.1711

Table 5.9 Model selection, specification and prediction Highway 400 and 404 NB & SB

Location	Predictive Variable	Model Specification					Accuracy of Prediction		
		vol~spd	vol~occ	spd~occ	Lag	Threshold	MSE	Coefficient of Variation	MAPE
400.NB.1 Hwy401 - Finch Ave. W. 400DN0040DNS	volume	x			4	1	109632.30	0.0748	0.0671
	speed	x			4	1	106.54	0.0980	0.0763
	occupancy		x		2	1	7.04	0.4487	0.1963
400.NB.2 Finch Ave. W. - Steeles Ave. W. 400DN0080DNS	volume	x			4	1	127344.30	0.0783	0.0662
	speed			x	8	1	99.02	0.0946	0.0671
	occupancy			x	8	1	5.20	0.2988	0.1938
400.SB.1 Hwy401 - Finch Ave. W. 400DN0040DSS	volume	x			2	2	134720.30	0.0814	0.0612
	speed	x			2	2	204.12	0.1417	0.1259
	occupancy			x	1	1	21.46	0.5744	0.3001
400.SB.2 Finch Ave. W. - Steeles Ave. W. 400DN0080DSS	volume		x		3	1	109302.20	0.0765	0.0745
	speed			x	2	2	219.87	0.1431	0.0906
	occupancy		x		3	1	8.97	0.5002	0.2003
404.NB.1 Sheppard Ave. E. - Finch Ave. E. 404DN0010DNS	volume		x		6	1	68148.23	0.0587	0.0466
	speed			x	1	1	52.47	0.0693	0.0538
	occupancy			x	1	1	6.19	0.3778	0.1962
404.SB.1 Sheppard Ave. E. - Finch Ave. E. 404DN0010DSS	volume		x		3	1	79529.93	0.0743	0.0566
	speed			x	8	1	119.30	0.1430	0.1309
	occupancy			x	8	1	18.75	0.3632	0.4462

5.4 Forecasting

The hourly interval one-step-ahead rolling prediction is performed with the selected TS-TVEC model for seven days in a row for each traffic variable. The prediction includes 168 points of time. Prediction accuracy is assessed by the MSE, Coefficient of Variation, and MAPE that are shown in Tables 5.5, 5.6, 5.7, 5.8, and 5.9 respectively. Exhibitions of model fitness and 168 hourly predictions for each traffic variable from 8 representative sites are shown in Fig 5.7, 5.8, 5.9, 5.10, 5.11, 5.12, 5.13 and 5.14 respectively. The complete list of exhibitions of model fitness and hourly predictions from 35 sites can be found in the Appendix A.

5.5 Discussion

There are 105 models that are estimated at 35 freeway locations. As the TS-TVEC model is a data driven model, its lag order p is varying for different data sets at different locations. The value of p reflects the best model selection given the data set at each location of prediction. The overall model specifications show that the 2-regime model prevails among data sets in contrast to the linear or 3-regime models. It indicates that traffic at most locations is inclined to two states. The 3-regime model specification may imply more complicated traffic situations at corresponding locations. Table 5.10 summarizes prediction accuracy shown in Tables 5.5, 5.6, 5.7, 5.8, and 5.9.

Table 5.10 Summary of model prediction accuracy at 35 locations

Statistics	Volume				Speed				Occupancy			
	MSE	Coefficient of Variation	MAPE	Std Deviation	MSE	Coefficient of Variation	MAPE	Std Deviation	MSE	Coefficient of Variation	MAPE	Std Deviation
min	18864.7	5.87%	4.66%	137.35	16.14	4.16%	3.10%	4.02	0.4	11.65%	9.24%	0.63
max	296926.8	41.29%	23.03%	544.91	284.64	20.43%	16.87%	16.87	41.89	97.40%	62.63%	6.47
median	109632.3	9.29%	7.45%	331.11	142.46	12.93%	9.84%	11.94	8.71	37.87%	20.72%	2.95

The mean absolute percentage error (MAPE) in traffic volume prediction is between 4.66% and 23.03% with median 7.45%, the coefficient of variation is between 5.87% and 41.29% with median 9.29%, and the standard deviation is between 137.35 and 544.91 with median 331.11.

Similarly, the MAPE in traffic speed prediction is between 3.1% and 16.87% with median 9.84%, the coefficient of variation is between 4.16% and 20.43% with median 12.93%, and the standard deviation is between 4.02 and 16.87 with median 11.94.

The MAPE in traffic occupancy prediction is between 9.24% and 62.63% with median 20.72%, the coefficient of variation is between 11.65% and 97.40% with median 37.87%, and the standard deviation is between 0.63 and 6.47 with median 2.95.

In other words, prediction accuracy of TS-TVEC model is approximately 92.55% for traffic volume, 90.16% for speed, and 79.28% for occupancy at 35 locations of prediction.

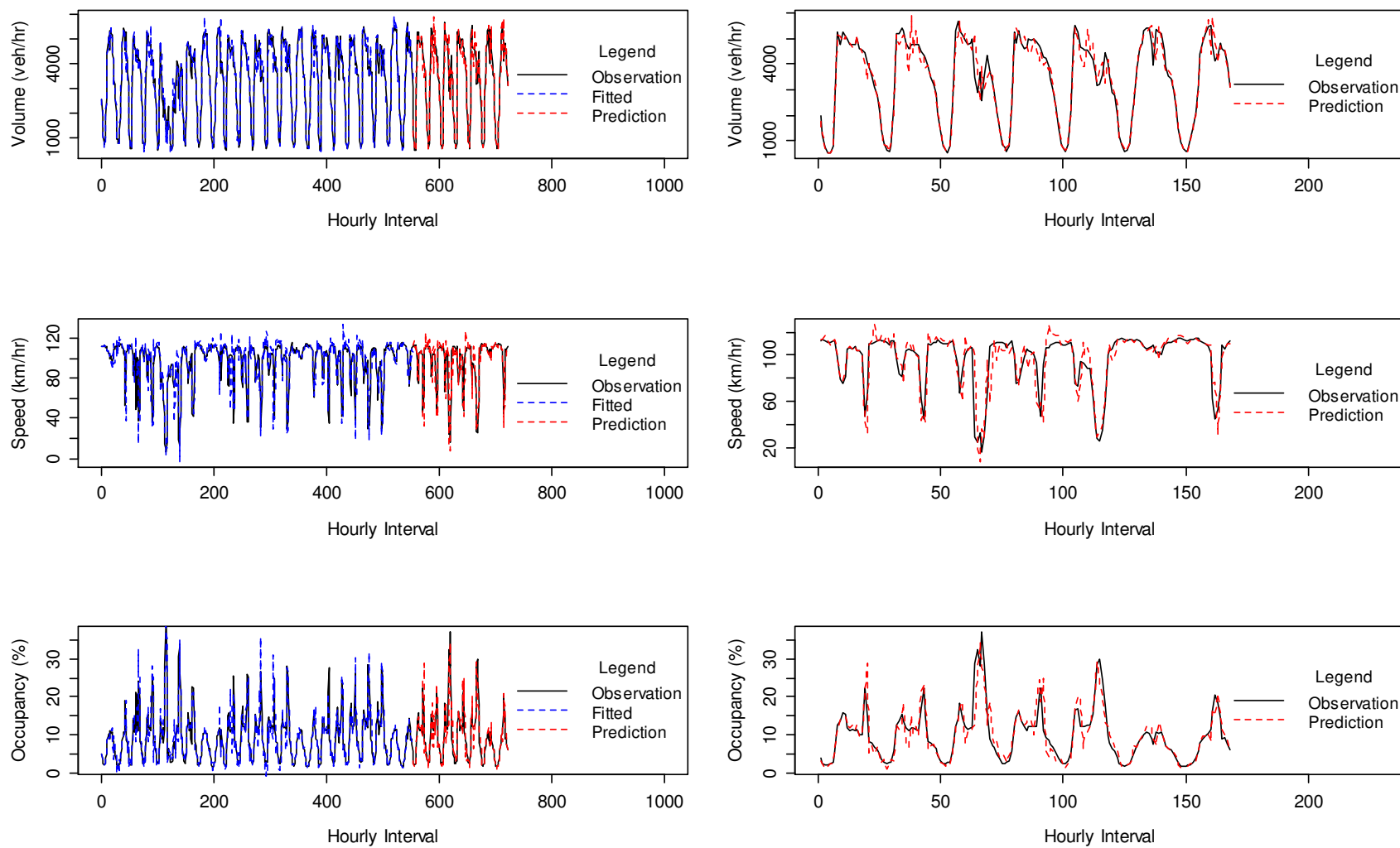


Fig. 5.7 401.EB.Exp.1, Hwy400 Basket Wave, 401DW0030DEE

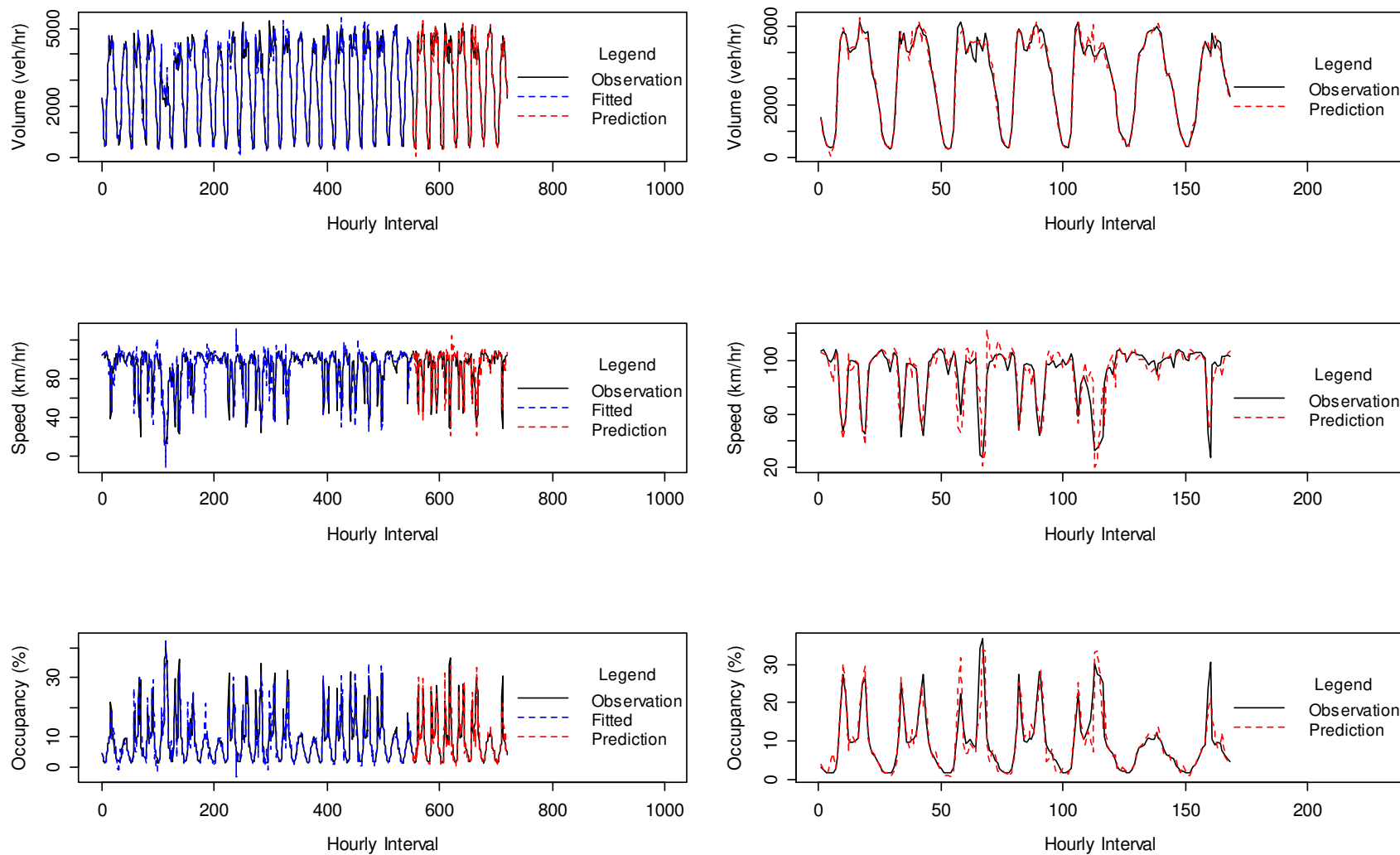


Fig. 5.8 401.EB.Col.7, Bayview Ave. Leslie St., 401DE0180DEC

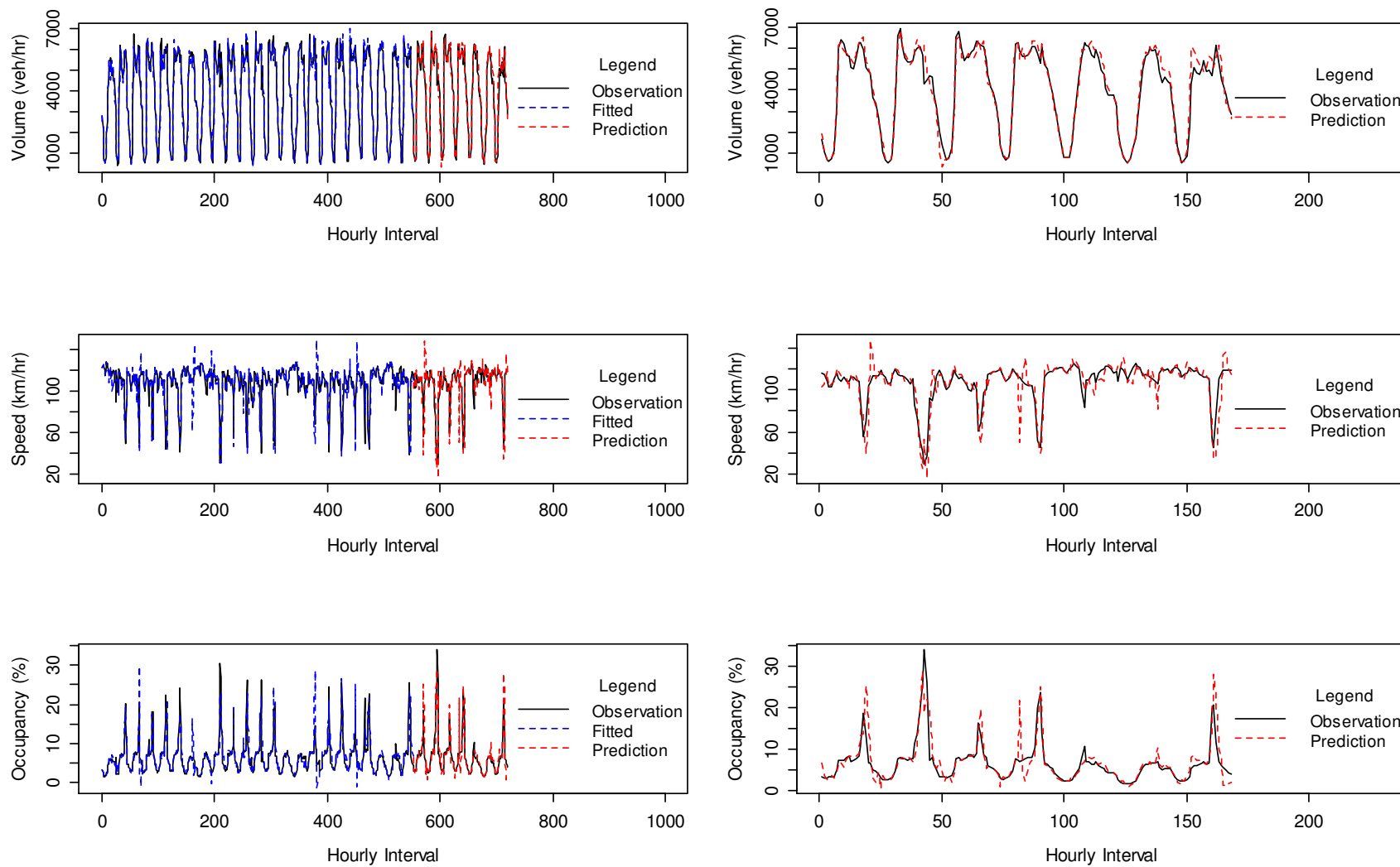


Fig. 5.9 401.WB.Exp.3, Allen Rd. @ Hwy401, 401DE0060DWE

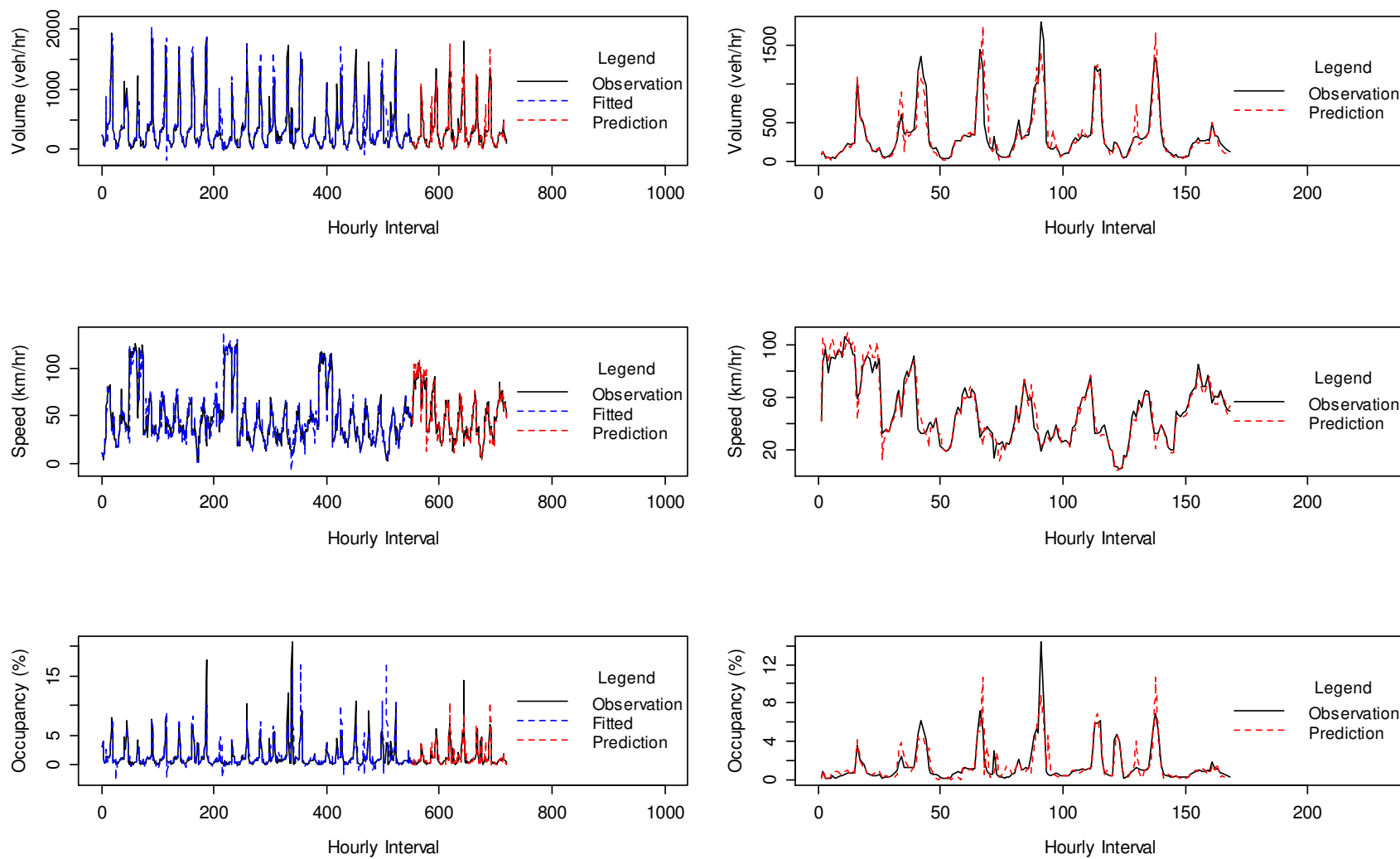


Fig. 5.10 401.WB.Col.3, Keele St. Dufferin St., 401DE0030DWC

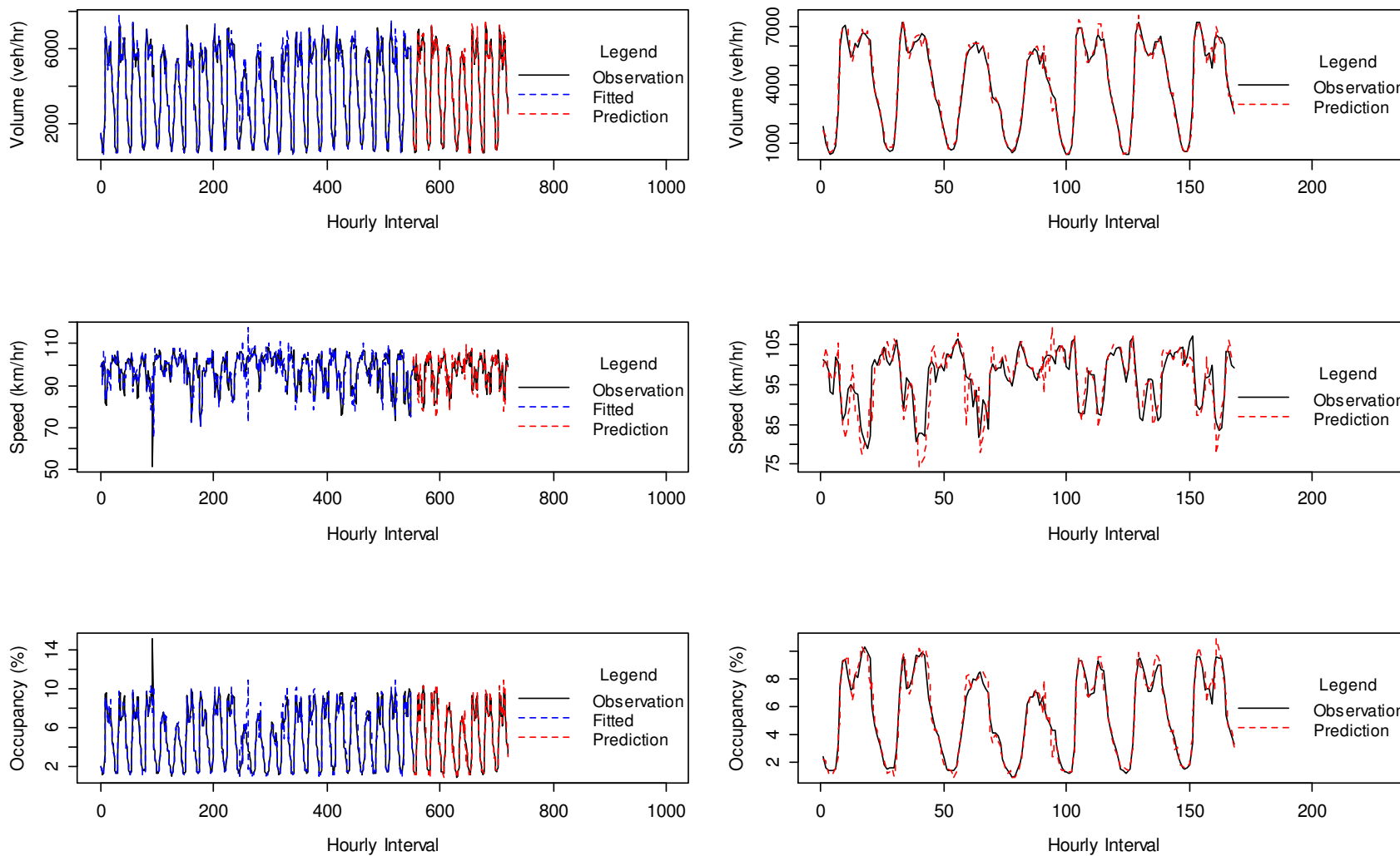


Fig. 5.11 401.WB.Col.5, Bathurst St. Avenue Rd., 401DE0090DWC

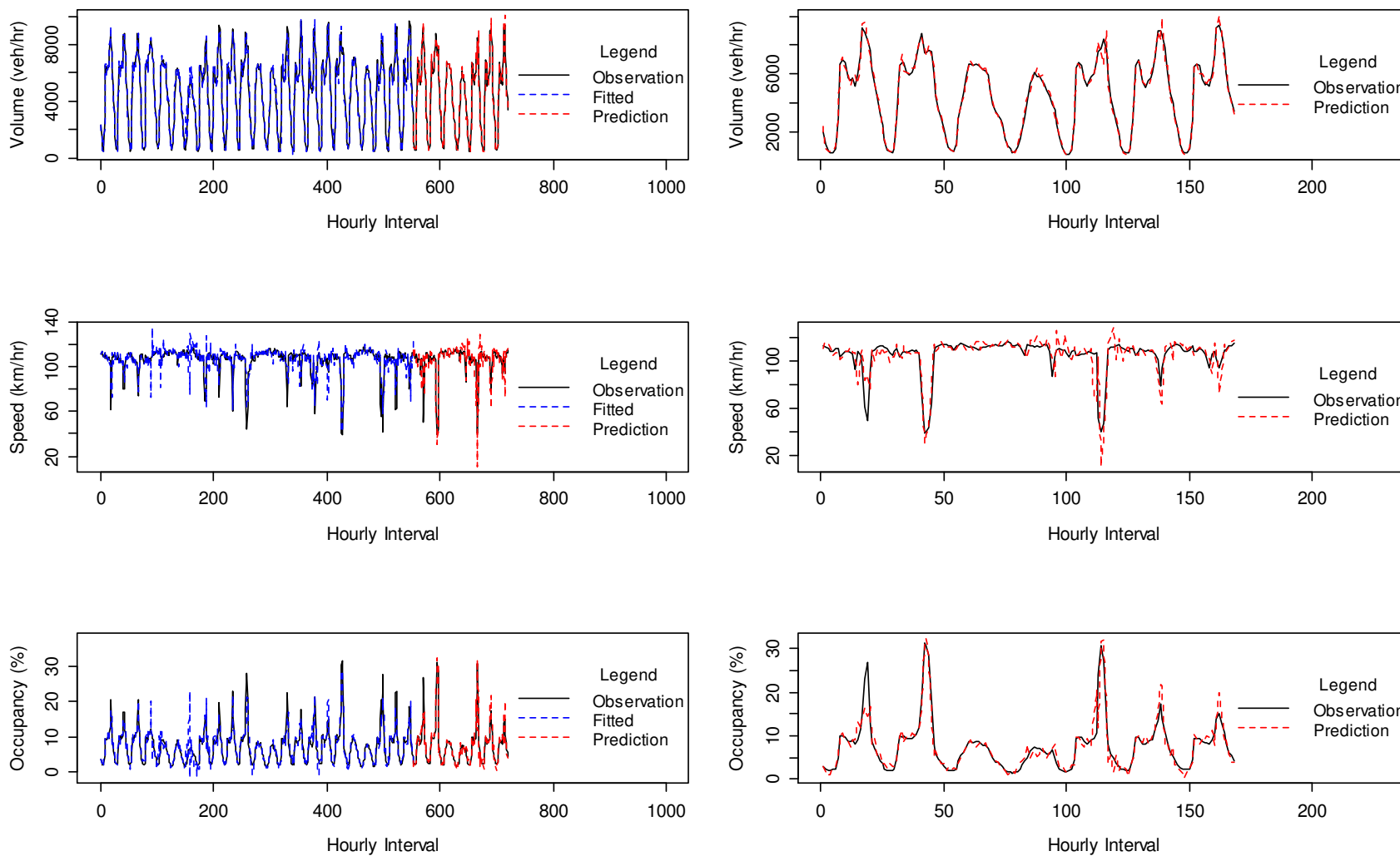


Fig. 5.12 400.NB.2, Finch Ave. W. Steeles Ave. W., 400DN0080DNS

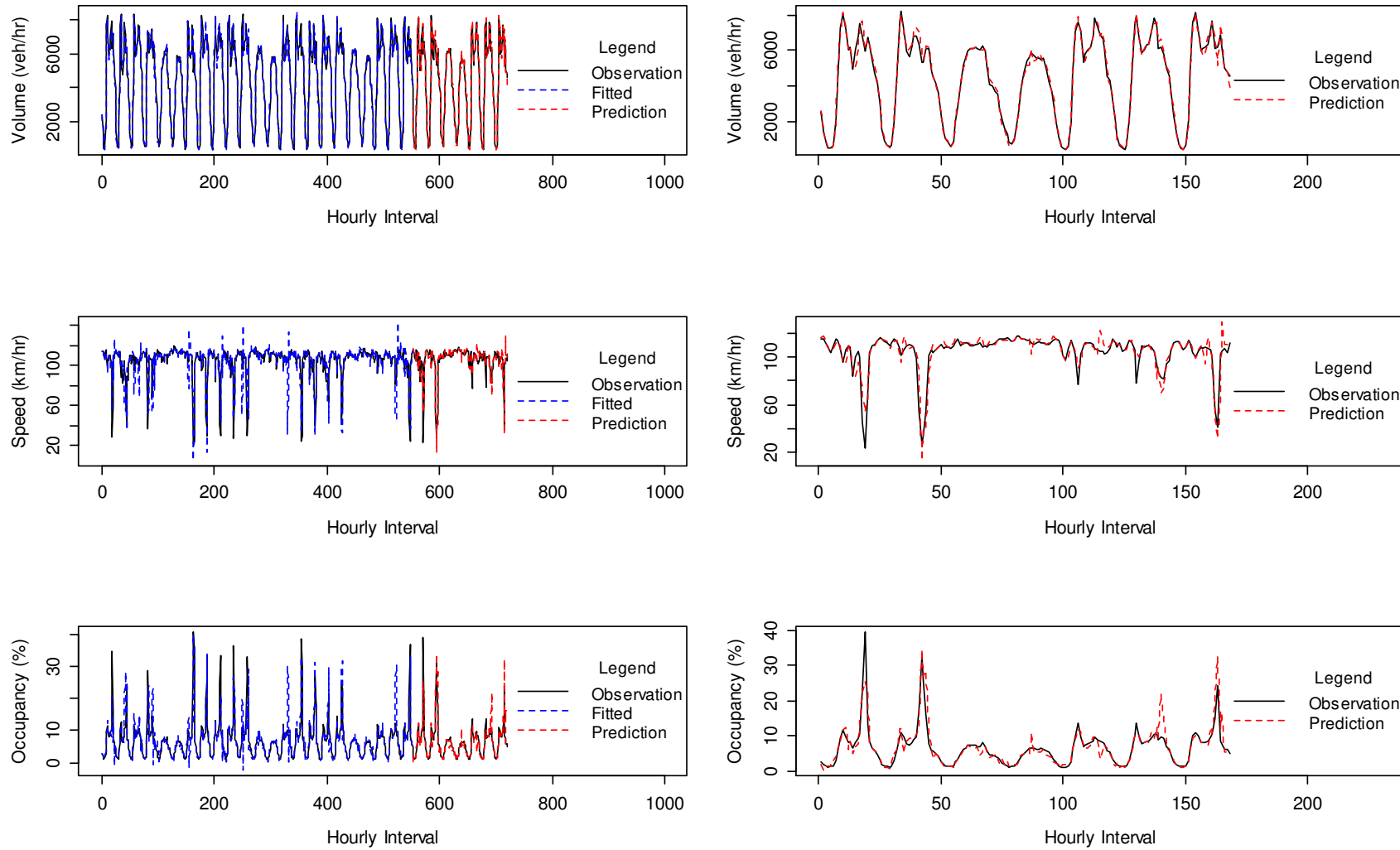


Fig. 5.13 404.NB.1, Sheppard Ave. E. Finch Ave. E., 404DN0010DNS

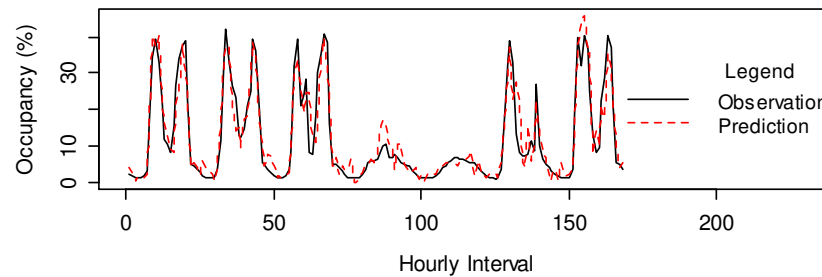
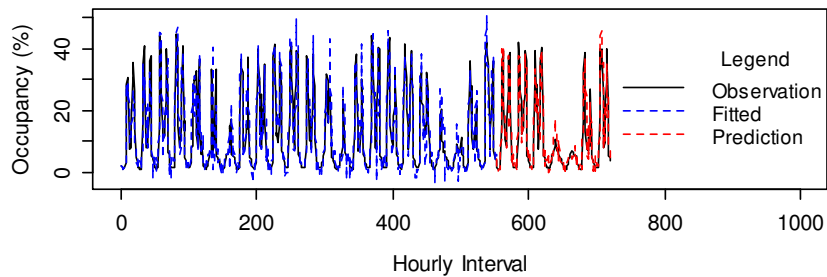
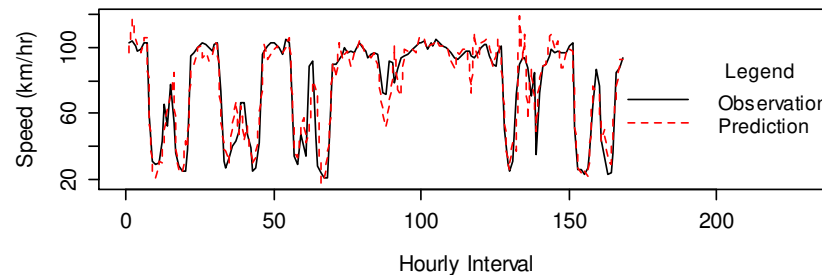
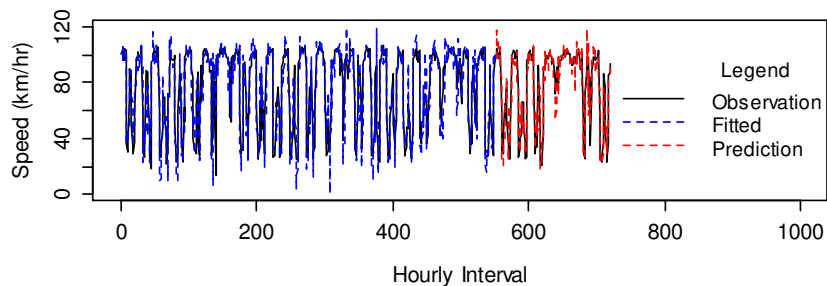
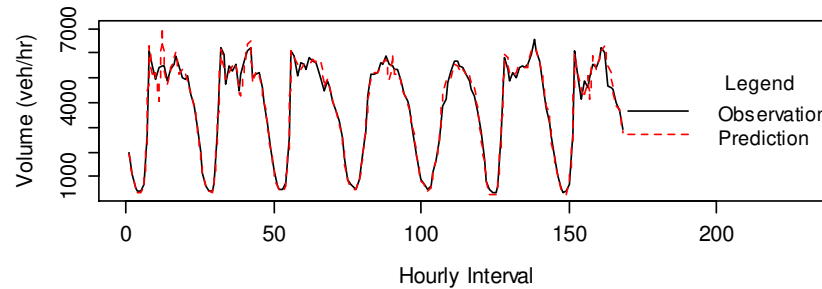
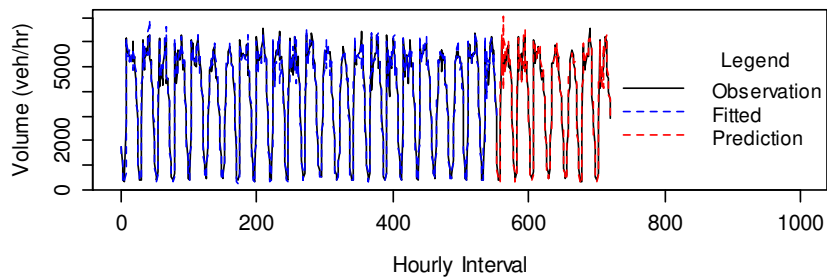


Fig. 5.14 404.SB.1, Sheppard Ave. E. Finch Ave. E., 404DN0010DSS

Chapter 6

Summary, discussion and conclusion

6.1 Summary

Short term traffic state prediction is an important component of Intelligent Transportation Systems. It is an essential instrument for the implementation of Advanced Traffic Management System (ATMS) and Advanced Traveller Information System (ATIS) to effectively manage traffic flow and provide traveler's route guidance.

In the past few decades, many methods for traffic state prediction were proposed as well as many attempts were made to improve the accuracy of prediction. Some representatives include, but not limited to, macroscopic traffic flow models, Neural Network, Kalman Filter, statistical time series models, Support Vector Regression, and nonparametric regression, etc. Although those methods made great contributions in this domain, there are still many improvements needed and gaps to fill in.

Based on our observation and study on a large amount of traffic time series data, the challenges that arise in traffic state prediction primarily are due to stochasticity and dynamics of traffic processes and their special statistical characteristics. Many of them have not been thoroughly and systematically taken into account in the existing approaches for traffic forecast.

In this research, we employ advanced time series modelling technique and developed time-space threshold vector error correction (TS-TVEC) model. The TS-TVEC model is able to concurrently take into account all those issues including autonomous and interactive dynamics of traffic variables, multiple traffic states, autocorrelation and cross-correlation of multivariate time series over time and space, seasonality, non-stationarity, cointegration of multivariate traffic time series, and unknown structural change in the time series. Both empirical study and large scale experiments show that TS-TVEC model is able to provide accurate predictions. The empirical study in chapter 4 shows that the TS-TVEC model outperformed MLP Neural Network and Support Vector Regression in forecasting traffic volume with approximately 6% of improvement on average. In chapter 5, the larger scale applications are performed to 35 freeway locations with approximately 315 time series. The mean absolute percentage error (MAPE) in traffic volume

prediction is between 4.66% and 23.03% with median 7.45%, the coefficient of variation is between 5.87% and 41.29% with median 9.29%, and the mean square error is between 18864 and 296926 with median 109632.

6.2 Discussion

In the context of data environment presented in Chapter 4, the results of comparison provide some evidence showing consistent effectiveness and robustness of the TS-TVEC model. However, we are not attempting to verify whether or not the TS-TVEC model will outperform and replace other families of models, such as Neural Network and Support Vector Regression, because these models are data oriented models such that their performance is closely associated with the data environment under study. Every family of models may exhibit their own advantages in specific circumstances. For instance, in an obvious linear data environment, according to the parsimony principle, a linear model will maintain its advantage. Nevertheless, it is our belief that TS-TVEC is a theoretically sound, powerful and competitive modelling method suitable for complex multivariate data environment where threshold cointegration effects are non-trivial. In the large scale application presented in Chapter 5, macroscopic traffic variables (i.e. volume, speed, and occupancy) are verified to be cointegrated with threshold effect. Therefore, the TS-TVEC model exhibits its advantages rendering its capacity to be fully used. It is potentially applicable to a wide variety of traffic circumstances and real time traffic state forecasting.

The TS-TVEC model also shows its advantage in need of a modest data size for model estimation. For instance, one-month hourly data is adequate for a TS-TVEC model to be estimated. One-month hourly data prior to the prediction time is able to provide sufficient information on the most recent dynamics of traffic flow including daily and weekly cyclic patterns.

In statistics, ergodicity refers to asymptotic independence and is a characteristic of a stochastic process. Loosely speaking, it means that the further apart two realizations of a time series are with respect to time, the closer to independence they become [Bernhard Pfaff 2008]. Hence, long time series is not necessarily helpful for improving model estimation and prediction.

The autocorrelograms and partial autocorrelograms show that traffic time series are short memory time series where observations separated by a long time span exhibit asymptotic independence. Therefore, one-month data is sufficient for the TS-TVEC model to identify the model structure. In practice, TS-TVEC model estimation should be a periodically rolling update process. The most recent one-month data prior to the prediction time should be always used for model estimation.

The TS-TVEC model has improved prediction performance due to the error correction mechanism, and the regime switching structure enables it to model multiple traffic states with various piecewise linear models.

In addition, in the study of numerous traffic time series, especially in traffic volume series, it is found that traffic seasonality is best represented in a weekly period. This finding is consistent with the specification of the seasonal ARIMA model proposed by Williams et al. (2003).

Three pairs of macroscopic traffic variables can be used respectively to construct the TS-TVEC model such that each traffic variable may have two parallel TS-TVEC models available for prediction.

6.3 Conclusion

According to the analysis to the hundreds of traffic time series, the special features that often exhibit in the traffic time series are summarized. From a statistical perspective, the features include seasonality, time and spatial correlation, non-stationary, and unknown structural change, whereas, from a traffic perspective, the features include the fundamental dynamics between traffic variables, and multiple traffic states. All these features are concurrently taken into account in the design of the TS-TVEC model such that the model has the capacity and flexibility to deal with the complex data environment in traffic prediction.

This research contributes in five aspects. It discovered the existence of cointegration effects among macroscopic traffic variables (i.e. volume, speed, and occupancy), introduced and adapted the error correction model to fit in traffic prediction, introduced the regime switching structure to reflect multi-states of traffic situation, incorporated spatially correlated information

into the TS-TVEC model, and revealed the inherent connection between error correction model and the transformed fundamental diagrams.

The discovery of the cointegration effects among macroscopic traffic variables is beneficial to better understanding the mechanism of the traffic data generating process, thus improving the prediction accuracy.

According to the Granger Representation Theorem, an error correction model must exist if the traffic variables are cointegrated, and vice versa. Hence, we are able to establish a vector error correction model between traffic variables due to their cointegration. Although the Granger Representation Theorem was proposed and proved since early 1980s in the economic sciences, to the best of our knowledge, this is the first time that the error correction model is introduced to and adapted for short term traffic prediction.

The vector error correction model essentially is a vector autoregressive model that is established on first order differenced variables with an error correction mechanism. The difference between two adjacent lags reflects the change of the variable over a time interval. Using the difference as a term in the model enables us to see the dynamics between variables, i.e. the amount of change in one variable causes the change in other variables.

Once traffic time series are mapped from the original level space to a difference space, the linear relationship between differenced traffic variables is revealed through transformed fundamental traffic diagrams. It justifies error correction model application for traffic prediction.

The linear relationship in the difference space may consist of multiple linear segments. The gap between two linear segments may be caused by the change of traffic state from free flow to congestion or vice versa. Hence, a regime switching structure with piecewise linear models embedded into different regimes is introduced to deal with structural breaks or traffic state alternation in traffic time series.

Spatially correlated information, i.e. traffic time series from upstream or/and downstream of the location of prediction, is incorporated into the model to enhance the accuracy of prediction. The TS-TVEC model provides flexible structure to enable any spatially correlated time series fused in the model. These spatial time series are presented in the form of difference in the model.

6.4 Future research

Currently, the TS-TVEC model focuses on hourly interval prediction. The prediction for other time intervals may be attempted in future research.

Missing data, or gaps, may appear in the traffic time series when a loop detector is malfunctioning for a certain period of time. Some demands exist for reconstructing the missing data or filling in the gaps. The methodology developed in this research may be adapted to serve this purpose.

The cointegration effect among traffic variables changes over time. In order to reflect the dynamics of cointegration, the model may be designed to adopt changeable coefficients.

References

- Abdi, J., Moshiri, B., Sedigh A. K., 2010. Comparison of RBF and MLP Neural Networks in Short-Term Traffic Flow Forecasting. IEEE. DOI: 10.1109/ICPCES.2010.5698623
- Abdulhai, B., Porwal, H., Recker, W., 2002. Short-term traffic flow prediction using neuro-genetic algorithms. *ITS J.* 7, 3–41.
- Ahmed, M. S., Cook, A. R., 1979. Analysis of freeway traffic time-series data by using Box–Jenkins techniques. *Transportation Research Board*, 722, pp. 1–9.
- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), *2nd International Symposium on Information Theory*. Akademiai Kiado, Budapest, pp. 267–281.
- Akaike, H., 1978. Time series analysis and control through parametric models. In: Findley, D.F. (Ed.), *Applied Time Series Analysis*. Academic Press, New York.
- Al-Deek, H., Ishak, S., Wang, M., 2001. A new short-term traffic prediction and incident detection system on I-4. The Florida Department of Transportation, University of Central Florida.
- Anand, N., Scoglio, C., Natarajan, B., 2008. GARCHNon-Linear Time Series Model for Traffic Modeling and Prediction. IEEE.
- Antoniou, C., Ben-Akiva, M., Koutsopoulos, H.N., 2005. Online calibration of traffic prediction models. *J. Transp. Res. Rec.* 1934, 235–245.
- Aw, A., Rasche, M., 2000. Resurrection of “second order” models of traffic flow. *SIAM J. Appl. Math.*, Vol. 60, No. 3, pp. 916–938.
- Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. *Econometrica* 66 (1), 47–78.
- Baragona, R., Battaglia, F., Cucina, D., 2004. Fitting piecewise linear threshold autoregressive models by means of genetic algorithms. *Comput. Stat. Data Anal.* 47, 277–295.

- Bartlett, M. S., 1966. *Stochastic Processes*. 2nd Ed. Cambridge Univ. Press.
- Beck, M. W. 2014. Package NeuralNetTools in R software, <<http://www.r-project.org/>>.
- Bellemans, T., De Schutter, B., Wets, G., De Moor, B., 2006. Model Predictive Control for Ramp Metering combined with Extended Kalman Filter-based Traffic State Estimation. Proceedings of the IEEE ITS Conference, Toronto, Canada.
- Bertsekas, D. P., 1995. *Nonlinear Programming*. Athena Scientific, Belmont, MA.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31, 307–327.
- Box, G.E.P., Jenkins, G.M., 1976. *Time Series Analysis: Forecasting and Control*. 2nd Ed. Holden-Day, San Francisco.
- Brockwell, P. J, Davis, R. A., 2002. *Introduction to Time Series and Forecasting*, second ed. Springer-Verlag.
- Campbell, J. Y. and Perron, P., 1991. Pitfalls and opportunities: What macroeconomists should know about unit roots, in ‘NBER Macroeconomic Annual 1991’, National Bureau of Economic Research, Cambridge, MA, pp. 141–218.
- Castillo, E., Menéndez, J. M., Sánchez-Cambronero, S., 2008. Predicting traffic flow using Bayesian networks. *Transp. Res. Part B* 42, 482–509.
- Cetin, M., Comert, G., 2006. Short-term traffic flow prediction with regime switching models. *Transp. Res. Rec.: J. Transp. Res. Board* 1965, 23–31.
- Chan, K. Y., Khadem, S., Dillon, T.S., 2012. Optimization of neural network configurations for short-term traffic flow forecasting using orthogonal design, IEEE World Congress on Computational Intelligence, Brisbane, Australia.
- Chang, H., Lee, Y., Yoon, B., Baek, S., 2012. Dynamic near-term traffic flow prediction: system oriented approach based on past experiences. *IET Intelligent Transport Systems*, 6 (3), 292–305.

- Chen, H., Chong, T. T.-L., Bai, J., 2012. Theory and applications of TAR model with two threshold variables. *Econometric Rev.* 31 (2), 142–170.
- Cheng, T., Haworth, J., Wang, J., 2012. Spatio-temporal autocorrelation of road network data. *J. Geogr. Syst.* 14, 389–413.
- Cortes, C., Vapnik, V., 1995. Support-vector network. *Machine Learning*, 20, 1-25.
- Cryer, J. D., Chan, K.-S., 2008. *Time Series Analysis with Applications in R*. 2nd Ed. Springer.
- Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Syst.* 2, 303–314.
- D'Angelo, M. P., Al-Deek, H. M., Wang, M. C., 1999. Travel-time prediction for freeway corridors. *Transportation Research Record*, 1676: p. 184-191.
- Daganzo, C.F., 1993. The Cell Transmission Model. Part I: A Simple Dynamic Representation Of Highway Traffic. Research Report in California PATH program, the University of California, Berkeley. UCB-ITS-PRR-93-7.
- Daganzo, C.F., 1994. The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transp. Res. Part B* 28 (4), 269–287.
- Daganzo, C.F., 1995a. The cell transmission model, Part II: Network traffic. *Transp. Res. Part B* 29 (2), 79–93.
- Daganzo, C.F., 1995b. Requiem for second-order fluid approximation of traffic flow, *Transportation Res. Part B*, 29 (4), pp. 277–286.
- Daganzo, C.F., 2005. A variational formulation of kinematic waves: basic theory and complex boundary conditions. *Transp. Res. Part B* 39 (2), 187–196.
- Davis, G. A., Niham, N. L., Hamed, M. M., Jacobson, L. N., 1991. Adaptive forecasting of freeway traffic congestion. *Transportation Research Record*, 1287, pp. 29–33.
- Deng, W., Lei, H., Zhou, X., 2013. Traffic state estimation and uncertainty quantification based on heterogeneous data sources: a three detector approach. *Transp. Res. Part B* 57, 132–157.

- Dickey, D.A., Fuller, W.A., 1979. Distributions of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* 75, 427–431.
- Dickey, D.A., Fuller, W.A., 1981. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49, 1057–1072.
- Disbro, J. E., Frame, M., 1989. Traffic flow theory and chaotic behavior. *Transportation Research Record* 1225, 109-115.
- Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., Vapnik, V., 1997. Support vector regression machines. In: *Advances in Neural Information Processing Systems*, 9, pages 155–161, MIT Press.
- Dunne, S., Ghosh, B., 2012. Regime-based short-term multivariate traffic condition forecasting algorithm. *J. Transp. Eng.* 138 (4).
- Elliott, G., Rothenberg, T. J., Stock, J. H., 1996. Efficient Tests for an Autoregressive Unit Root. *Econometrica*, Vol. 64, No. 4, 813–836.
- Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica*, 50, 987–1007.
- Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: representation, estimation, and testing. *Econometrica* 55 (2), 251–276.
- Fletcher, R., 1989. *Practical Methods of Optimization*. John Wiley and Sons, New York.
- Gao, H., Zhao, J., Jia, L., 2007. Short-term Traffic Flow Forecasting Model of Elman Neural Network Based on Dissimilation Particle Swarm Optimization. *IEEE*.
- Garavello, M., Piccoli, B., 2006. *Traffic Flow on Networks Conservation Laws Models*. AIMS Series on Applied Mathematics Volume 1. the American Institute of Mathematical Sciences.
- Garside, S., Lindveld, K., Whittaker, J., 1997. Tracking and predicting a network traffic process. *International Journal of Forecasting*, 13: pp. 51-61.

- Granger, C.W.J., Newbold, P., 1974. Spurious regressions in econometrics. *J. Econometrics* 2, 111–120.
- Granger, C.W.J., 1981. Some properties of time series data and their use in econometric model specification. *J. Econometrics* 16, 121–130.
- Granger, C.W.J., Weiss, A.A., 1983. Time series analysis of error-correcting models. In: *Studies in Econometrics, Time Series and Multivariate Statistics*. Academic Press, New York, pp. 255–278.
- Greenberg, H., 1959. An analysis of traffic flow. *Oper. Res.* 7.
- Greenshields, B.D., 1935. A study in highway capacity. *Highway Res. Board Proc.* 14.
- Greenshields 75 Symposium, 2011. *Transportation Research Circular E-C149, 75 Years of the Fundamental Diagram for Traffic Flow Theory*, TRB, Washington, DC, pp. 22-42
- Guo, J., 2005. Adaptive estimation and prediction of univariate vehicular traffic condition series. North Carolina State University, Civil Engineering.
- Guo, X., Zhu, Q., 2009. A Traffic Flow Forecasting Model Based on BP Neural Network. 2nd International Conference on Power Electronics and ITS. IEEE.
- Hamed, M. M., Al-Masaeid, H. R., Said, Z. M. B., 1995. Short-term prediction of traffic volume in urban arterials. *Journal of Transportation Engineering*, 121(3): p. 249-254.
- Hansen, B. E., 1999. Testing for linearity. *J. Econ. Surv.* 13 (5).
- Hansen, B. E., Seo, B., 2002. Testing for two-regime threshold cointegration in vector error correction models. *J. Econometrics* 110, 293–318.
- Helbing, D., 1996. Gas-kinetic derivation of Navier-Stokes-like traffic equations. *Physical Review E* 53(3), 2266-2381.
- Helbing, D., 1997. Modeling Multi-lane traffic flow with queuing effects. *Physica A* 242, 175-194.

- Herbrich, R., 2002. *Learning Kernel Classifiers, Theory and Algorithms*. The MIT Press.
- Herrera, J. C., Bayen, A. M., 2010. Incorporation of Lagrangian measurements in freeway traffic state estimation. *Transp. Res. Part B* 44, 460–481.
- Hong, W.-C., Dong, Y., Hung, W.-M., Wei, S.-Y., 2010. Seasonal adjustment in a SVR with chaotic simulated annealing algorithm traffic flow forecasting model. *Second World Congress on Nature and Biologically Inspired Computing*, Fukuoka, Japan.
- Hoogendoorn, S. P., 1999. *Multiclass Continuum Modelling of Multilane Traffic Flow*. Delft University Press.
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* 2, 359–366.
- Ishak, S., Al-Deek, H., 2002. Performance evaluation of short-term time-series traffic prediction model. *Journal of Transportation Engineering*, 128(6): p. 490-498.
- Jia, A, Roupail, N. M., Williams, B. M., 2010. Identification and calibration of site specific stochastic freeway breakdown and queue discharge. *Transportation Research Record* 2188:148–155.
- Johansen, S., 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.
- Kamarianakis, Y., Prastacos, P., 2003. Forecasting traffic flow conditions in an urban network: comparison of multivariate and univariate approaches. *Transp. Res. Rec.*, 1857 p. 74-84.
- Kamarianakis, Y., Prastacos, P., 2005. Space–time modeling of traffic flow. *Comput. Geosci.* 31, 119–133.
- Kamarianakis, Y., Shen, W., Wynter, L., 2012. Real-Time Road Traffic Forecasting Using Regime-Switching Space-Time Models and Adaptive LASSO. Published online in Wiley Online Library.
- Kerner, B. S., 2004. *The physics of traffic*. Springer-Verlag.

Kerner, B. S., 2009. Introduction to Modern Traffic Flow Theory and Control, The Long Road to Three-Phase Traffic Theory. Springer-Verlag.

Kirby, H. R., Watson, S. M., Dougherty, M. S., 1997. Should we use neural networks or statistical models for short-term motorway traffic forecasting? *International journal of forecasting*, 13: p. 43-50.

Kirchgässner, Gebhard, Wolters, Jürgen, Hassler, Uwe, 2013. Introduction to Modern Time Series Analysis, second ed. Springer-Verlag.

Kolmogorov, A. N., 1957. On the representation of continuous functions of several variables as superpositions of continuous functions of one variable and addition. In: Translated in Selected Works of A.N. Kolmogorov. *Mathematics and Mechanics*, vol. 1. Kluwer Academic Publishers., pp. 383–387 (1991).

Kutner, M., Nachtsheim, C., Neter, J., Li, W., 2005. Chapter 13, Applied Linear Statistical Models, 5th Ed. McGraw-Hill/Irwin.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *J. Econometrics* 54, 159–178.

Laval, J. A., He, Z., Castrillon, F., 2012. A stochastic extension of newell's three-detector method. *Transp. Res. Rec.: J. Transp. Res. Board* 12–2563, 73–80.

Lawrance, A. J., Lewis, P. A. W., 1980. The exponential autoregressive moving average EARMA(p, q) process. *J. Roy. Statist. Soc., B*, 42, 150-161.

Lee, S., Fambro, D.B., 1999. Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. *Transportation Research Record*, 1678: p. 179-188.

Lee, Y.-I., Choi, C.Y., 1998. Development of a link travel time prediction algorithm for urban expressway. *Proceedings of 5th World Congress on Intelligent Transport Systems*.

- Leutzbach, W., 1988. *Introduction to the Theory of Traffic Flow*. translated English version, Springer-Verlag.
- Levin, M., Tsao, Y.-D., 1980. On forecasting freeway occupancies and volumes. *Transportation Research Record*, 773, pp. 47–49.
- Li, C., Xu, S., Wen, X., 2010. Traffic Flow forecasting Algorithm Using Simulated Annealing Genetic BP Network. *International Conference on Measuring Technology and Mechatronics Automation*.
- Lighthill, M. J., Whitham, G. B., 1955. On kinematic waves: II a theory of traffic flow on long crowded roads. *Proc. R. Soc. A* 229 (1178), 317–345.
- Liu, T., Sun, X., Zhong, X., 2010. Short term traffic flow forecasting based on a three-regime SETAR model. *J. Highway Transp. Res. Dev. China* 27 (10), 122–127.
- Liu, Z., Yang, Z., Gao, P., 2009. Research on the Short-term Traffic Flow Prediction Method Based on BP Neural Networks. *IEEE*.
- Mattera, D., Haykin, S., 1999. Support vector machines for dynamic reconstruction of a chaotic system. In: *Advances in Kernel Methods—Support Vector Learning*, pages 211–242, Cambridge, MIT Press.
- McLeod, A. I., Zhang, Y., 2008. Improved subset autoregression: with R package. *J. Stat. Softw.* 28 (2), 1–28.
- Mercer, J., 1909. Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society, London*, A209: 415–446.
- Mihaylova, L., Boel, R., Hegyi, A., 2007. Freeway traffic estimation within particle filtering framework. *Automatica* 43 (2), 290–300.
- Min, W., Wynter, L., 2011. Real-time road traffic prediction with spatio-temporal correlations. *Transp. Res. Part C* 19, 606–616.

- Min, X., Hu, J., Zhang, Z., 2010. Urban traffic network modeling and short-term traffic flow forecasting based on GSTARIMA model. In: Proceedings of 13th International IEEE Conference 2010.
- Moore, A. W., 2003. Support Vector Machines, course notes. School of Computer Science, Carnegie Mellon University. www.cs.cmu.edu/~awm.
- Muller, K.-R., Smola, A., Ratsch, G., Scholkopf, B., Kohlmorgen, J., Vapnik, V., 1997. Predicting time series with support vector machines. In: Artificial Neural Networks ICANN'97, pages 999–1004, Berlin. Springer Lecture Notes in Computer Science, Vol. 1327.
- Muñoz, L., Sun, X., Horowitz, R., Alvarez, L. 2003. Traffic density estimation with the cell transmission model. In: Proceedings of the American Control Conference, Denver, Colorado, USA, pp. 3750–3755.
- Nanthawichit, C., Nakatsuji, T., Suzuki, H., 2003. Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway. *Transp. Res. Rec.* 1855, 49–59.
- Narzo, A. F. D., Aznarte, J. L., Stigler, M., 2014. Manual of tsDyn Package in R v0.9-41.
- Newell, G. F., 1993. A simplified theory of kinematic waves in highway traffic, Part I: General theory. *Transp. Res. Part B* 27 (4), 281–297.
- Nihan, N. L., 1980. Use of the Box and Jenkins time series technique in traffic forecasting. *Transportation (Netherlands)* 9: p. 125-143.
- Okutani, I., Stephanedes, Y. J., 1984. Dynamic prediction of traffic volume through Kalman filtering theory. *Transportation Research Part B*, 18(1): p. 1-11.
- Page, E. S., 1954. Continuous inspection scheme. *Biometrika* 41 (1/2), 100–115.
- Palais, R. S., 2008. An Introduction to Wave Equations and Solitons. In the Princeton Companion to Mathematics. T. Gower Ed., Princeton Univ. Press, 234-239.

- Papageorgiou, M., 1983. Applications of Automatic Control Concepts to Traffic Flow Modeling and Control. Lecture Notes in Control and Information Sciences. Springer-Verlag.
- Papageorgiou, M., 1998. Some remarks on macroscopic traffic flow modelling. *Transportation Res. Part A*. 32 (5), pp. 323-329.
- Payne, H. J., 1971. Models of freeway traffic and control. In: *Mathematical Models of Public Systems*, vol. 1, pp. 51–61. Simulation Council, La Jolla, CA
- Pearson, K., 1920. Notes on the History of Correlation. *Biometrika*, 13, 25-45.
- Pfaff, B., 2008. *Analysis of Integrated and Cointegrated Time Series with R*. Springer-Verlag.
- Pfaff, B., Stigler, M., 2013. Package *urca* in R software, <<http://www.r-project.org/>>.
- Pfeifer, P. E., Deutsch, S. J., 1980. A three-stage iterative procedure for space-time modeling. *Technometrics* 22 (1), 35–47.
- Phillips, P. C. B., 1986. Understanding spurious regression in econometrics. *J. Econometrics* 33, 311–340.
- Phillips, P.C.B., Ouliaris, S., 1990. Asymptotic properties of residual based tests for cointegration. *Econometrica* 58, 73–93.
- Qiao, F., Yang, H., Lam, W. H. K., 2001. Intelligent simulation and prediction of traffic flow dispersion. *Transp. Res. Part B* 35, 843–863.
- Richards, P. I., 1956. Shock waves on the highway. *Ope. Res.* 4 (1), 42–51.
- Rodgersa, J. L., Nicewandera, W. A., 1988. Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, 42(1), 59-66.
- Sarle, W. S. (Ed.). 1997. Neural Network FAQ, part 1 of 7: Introduction, Periodic Posting to the Usenet Newsgroup *comp.ai.neural-nets*, <<ftp://ftp.sas.com/pub/neural/FAQ.html>>.
- Scholkopf, B., Smola, A. J., 2002. *Learning with Kernels*, The MIT Press.

Shang, P., Li, X., Kamae, S., 2005. Chaotic analysis of traffic time series. *Chaos, Solitons and Fractals* 25 (1): 121–128.

Sharda, R., Patil, R. B., 1992. Connectionist approach to time series prediction: an empirical test. *J. Intell. Manuf.* 3, 317–323.

Smith, B. L., Oswald, R. K., 2000. Effects of Parameter Selection on Forecast Accuracy and Execution Time in Nonparametric Regression. 2000 IEEE intelligent Transportation Systems Conference Proceedings, Dearborn (MI), USA.

Smith, B. L., Williams, B. M., Oswald, R. K., 2002. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transp. Res. Part C* 10(4), 303–321.

Stathopoulos, A., Karlaftis, M. G., 2003. A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies*, 11(2):121 – 135.

Stigler, M., Tortora, P., 2011. Food crisis: did agricultural prices rise faster than they fell? *Res. Top. Agric. Appl. Econ.* 2, 19–30.

Stitson, M., Gammerman, A., Vapnik, V., Vovk, V., Watkins, C., Weston, J., 1999. Support vector regression with ANOVA decomposition kernels. In: *Advances in Kernel Methods—Support Vector Learning*, pages 285–292, MIT Press.

Sun, X., Liu, T., 2011. A STAR model for urban short-term traffic flow forecasting. 7th Advanced Forum on Transportation of China, Beijing. IET, pp. 185 – 190.

Tampere, C. M. J., Immers, L. H., 2007. An Extended Kalman Filter Application for Traffic State Estimation Using CTM with Implicit Mode Switching and Dynamic Parameters. *Proceedings of the ITS Conference*, Seattle, USA.

Theja, P.V.V. K., Lelitha V., 2010. Short Term Prediction of Traffic Parameters Using Support Vector Machines Technique. *Emerging Trends in Engineering & Technology*, International Conference on 2010, pp. 70-75, doi:10.1109/ICETET.2010.37.

- Tong, H., 1975. Determination of the order of a Markov chain by Akaike's information criteria. *J. Appl. Prob.* 12, 488-497.
- Tong, H., 1978. On a threshold model. In: Chen, C.H. (Ed.), *Pattern Recognition and Signal Processing*, NATO ASI Series E: Applied Sc., vol. 29. Sijthoff & Noordhoff, The Netherlands, pp. 575–586.
- Tong, H., 1980. A view on non-linear time series model building. *Time Series* (ed. O. D. Anderson). Amsterdam: North-Holland.
- Tong, H., Lim, K. S., 1980. Threshold autoregression, limit cycles and cyclical data (with discussion). *J. Roy. Statist. Soc., B*, 42, 245-292.
- Tong, H., 1983. Threshold models in non-linear time series analysis. *Lecture Notes in Statistics*, vol. 21. Springer-Verlag, New York, p. 323.
- Tong, H., 1990. *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press, p. 564.
- Tong, H., 2010. *Threshold Models in Time Series Analysis—30 Years On*. London School of Economics and Political Science.
- TRB Special Report 165, 1975. *Traffic Flow Theory*. Transportation Research Board, National Research Council.
- Treiber, M., Kesting, A., Helbing, D., 2010. Three-phase traffic theory and two-phase models with a fundamental diagram in the light of empirical stylized facts. *Transp. Res. Part B* 44, 983–1000.
- Treiber, M., Kesting, A., 2013. *Traffic Flow Dynamics, Data, Models and Simulation*. Springer-Verlag.
- van Hinsbergen, C.P.IJ., van Lint, J.W.C., Sanders, F.M., 2007. Short Term Traffic Prediction Models. ITS World Congress, Beijing, China

Vlahogianni, E., Karlaftis, M., Stathopoulos, A., 2005. An extreme value based neural clustering approach for identifying traffic states. Proceedings of the 8th International IEEE Conference on ITS, Vienna, Austria.

Vlahogianni, E., Geroliminis, N., Skabardonis, A., 2008a. Empirical and Analytical Investigation of Traffic Flow Regimes and Transitions in Signalized Arterials. *Journal of Transportation Engineering ASCE*, 134:512-522.

Vlahogianni, E., Karlaftis, M., Golias, J., 2008b. Temporal Evolution of Short-Term Urban Traffic Flow: A Nonlinear Dynamics Approach. *Computer-Aided Civil and Infrastructure Engineering* 23, 536–548, IEEE

Wang, J., Shi, Q., Lu, H., 2005. The Study of Short-Term Traffic Flow Forecasting Based on Theory of Chaos. Proceedings of Intelligent Vehicles Symposium, IEEE. pp. 869 – 874

Wang, X.-H., Xiao, J.-M., 2003. A Radial Basis Function Neural Network Approach to Traffic Flow Forecasting. IEEE.

Wang, Y., Papageorgiou, M., 2005. Real-time freeway traffic state estimation based on extended Kalman filter: a general approach. *Transp. Res. Part B* 39, 141–167.

Wang, Y., Papageorgiou, M., Messmer, A., 2008. Real-time freeway traffic state estimation based on extended Kalman filter: Adaptive capabilities and real data testing. *Transportation Research Part A* 42 (2008) 1340–1358

Williams, B. M., Durvasula, P. K., Brown, D. E., 1998. Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models. *Transportation Research Record*, 1644: pp. 132-141.

Williams, B. M., 1999. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal Stochastic Time Series Process (dissertation). Department of Civil Engineering, University of Virginia, Charlottesville.

Williams, B. M., 2001. Multivariate vehicular traffic flow prediction: An evaluation of ARIMAX modeling. *Transportation Research Record* 1776, 194-200.

- Williams, B. M., Hoel, L. A., 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: a theoretical basis and empirical results. *J. Transp. Eng. (ASCE)* 129, 664–672.
- Wu, C., Ho, J., Lee, D., 2004. Travel-time prediction with support vector regression. *IEEE Transp. Intell. Transp. Syst.* 5 (4), 276–281.
- Yang, H., Qiao, F., 1998. Neural network approach to classification of traffic flow states. *J. Transp. Eng. ASCE* 124 (6), 521–525.
- Yang, W., Yang, D., Zhao, Y., Gong, J., 2010. Traffic Flow Prediction based on Wavelet Transform and Radial Basis Function Network. *IEEE*.
- Yang, Y., Lu, H., 2010. Short-term Traffic Flow Combined Forecasting Model Based on SVM. *International Conference on Comp. & Info. Sci. IEEE*. pp. 262 - 265
- Yuan, Z.-W., Wang, Y.-H., 2012. Research on K Nearest Neighbor Non-parametric Regression Algorithm Based on KD-Tree and Clustering Analysis. *the 4th International Conference on Comp. & Info. Sci.* pp. 298–301.
- Zhang, P. G., Qi, M., 2005. Neural network forecasting for seasonal and trend time series. *Eur. J. Oper. Res.* 160, 501–514.
- Zheng, W.Z., Lee, D. H., Shi, Q. X., 2006. Short-term freeway traffic flow prediction: Bayesian combined neural network approach. *J. Transp. Eng. ASCE* 132 (2), 114–121.
- Zhou, Z., 2013. Heteroscedasticity and autocorrelation robust structural change detection. *J. Am. Stat. Assoc.* 108 (502), 726–740.
- Zhu, C., Xu, X., Yan, C., 2010. The Research of Method of Short-Term Traffic Flow Forecast Based on GA-BP Neural Network and Chaos Theory. *IEEE*.
- Zivot, E., Andrews, D. W.K., 1992. Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis. *Journal of Business & Economic Statistics*, 10(3), 251–270.
- Zivot, E., Wang, J., 2006. *Modeling Financial Time Series with S-PLUS*, second ed. Springer-Verlag.

Appendix A
Exhibitions of model fitness and hourly rolling predictions for 35
locations

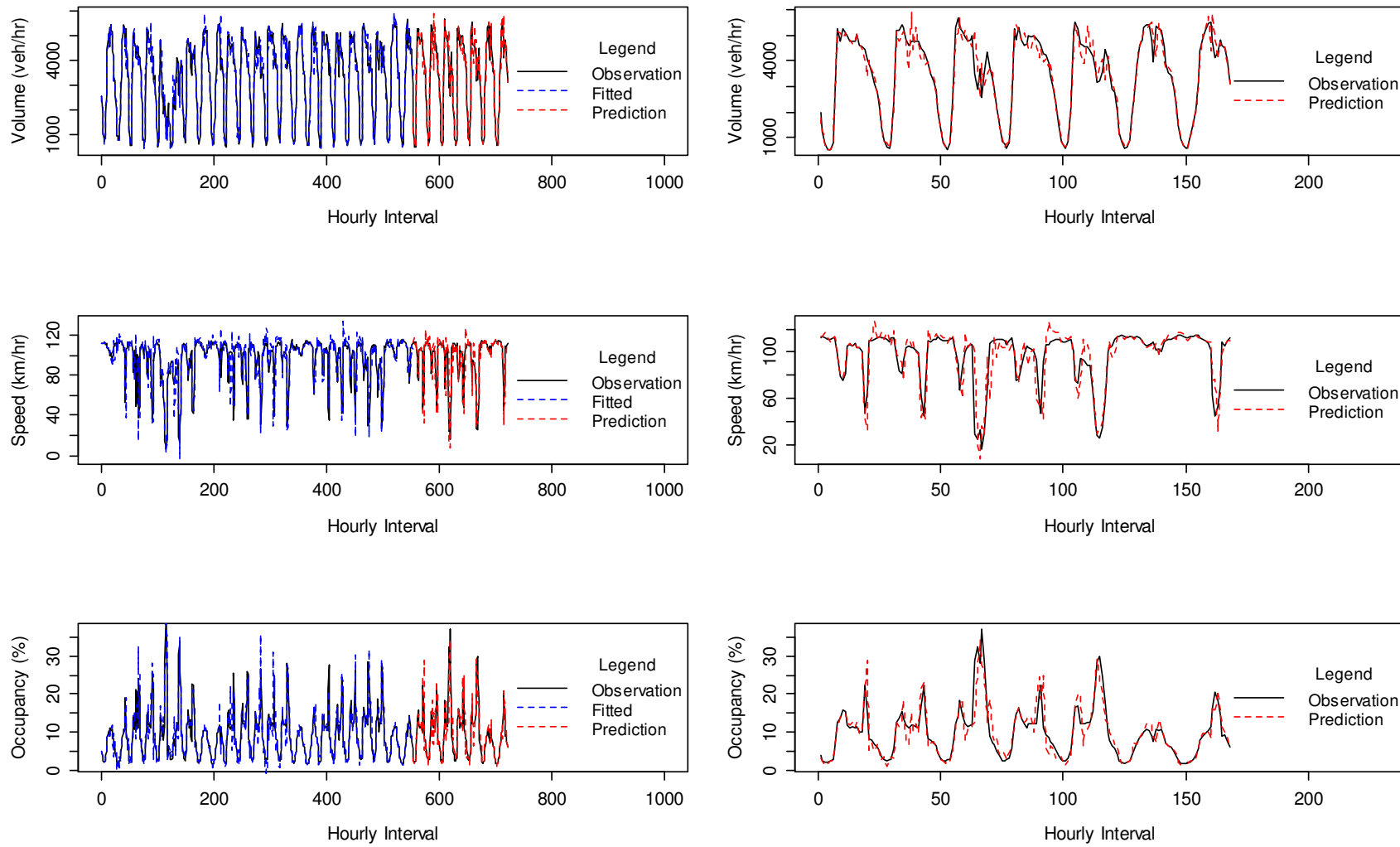


Fig. A. 1 401.EB.Exp.1, Hwy400 - Basket Wave, 401DW0030DEE

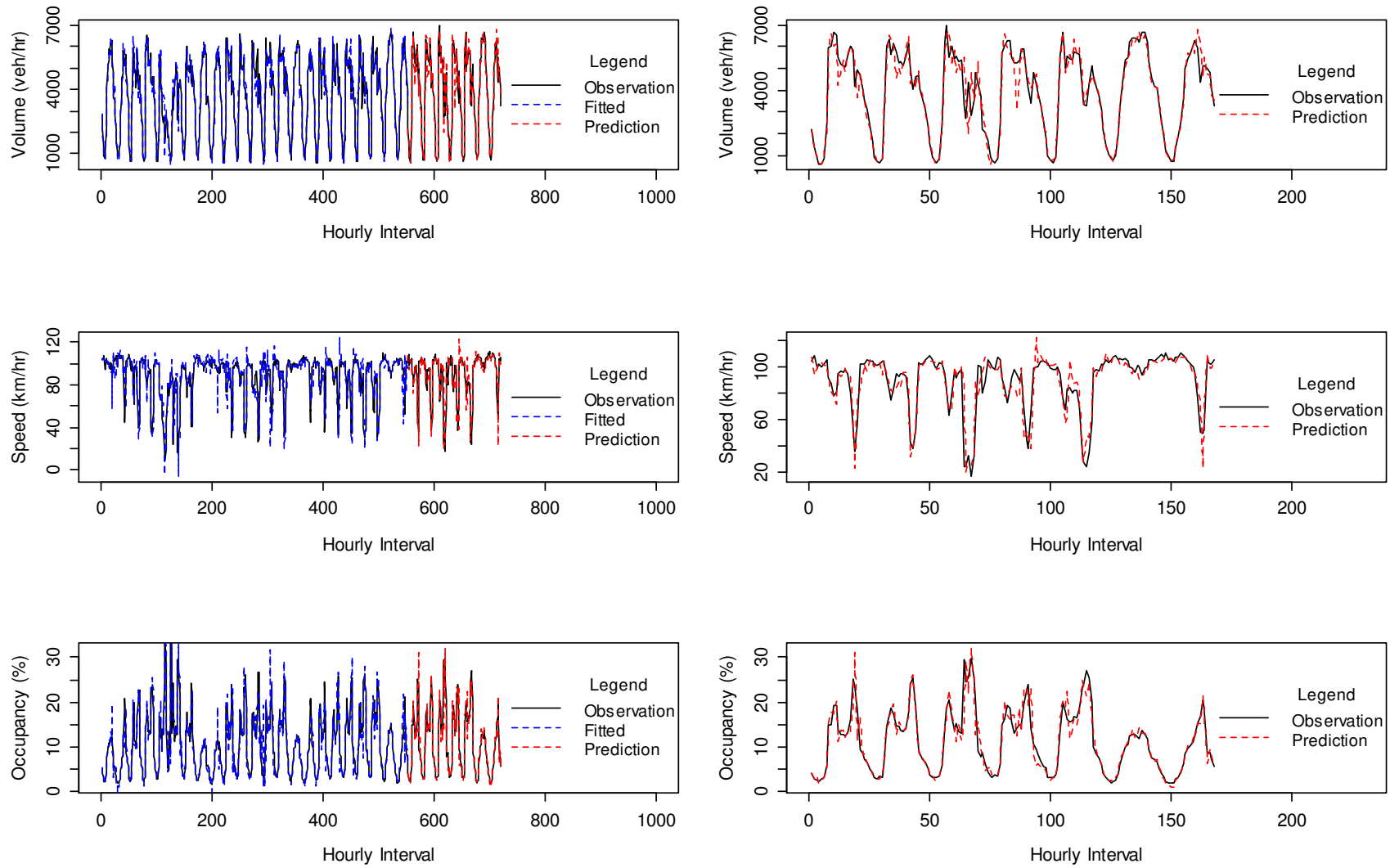


Fig. A. 2 401.EB.Exp.2, Basket Wave - Dufferin St., 401DE0020DEE

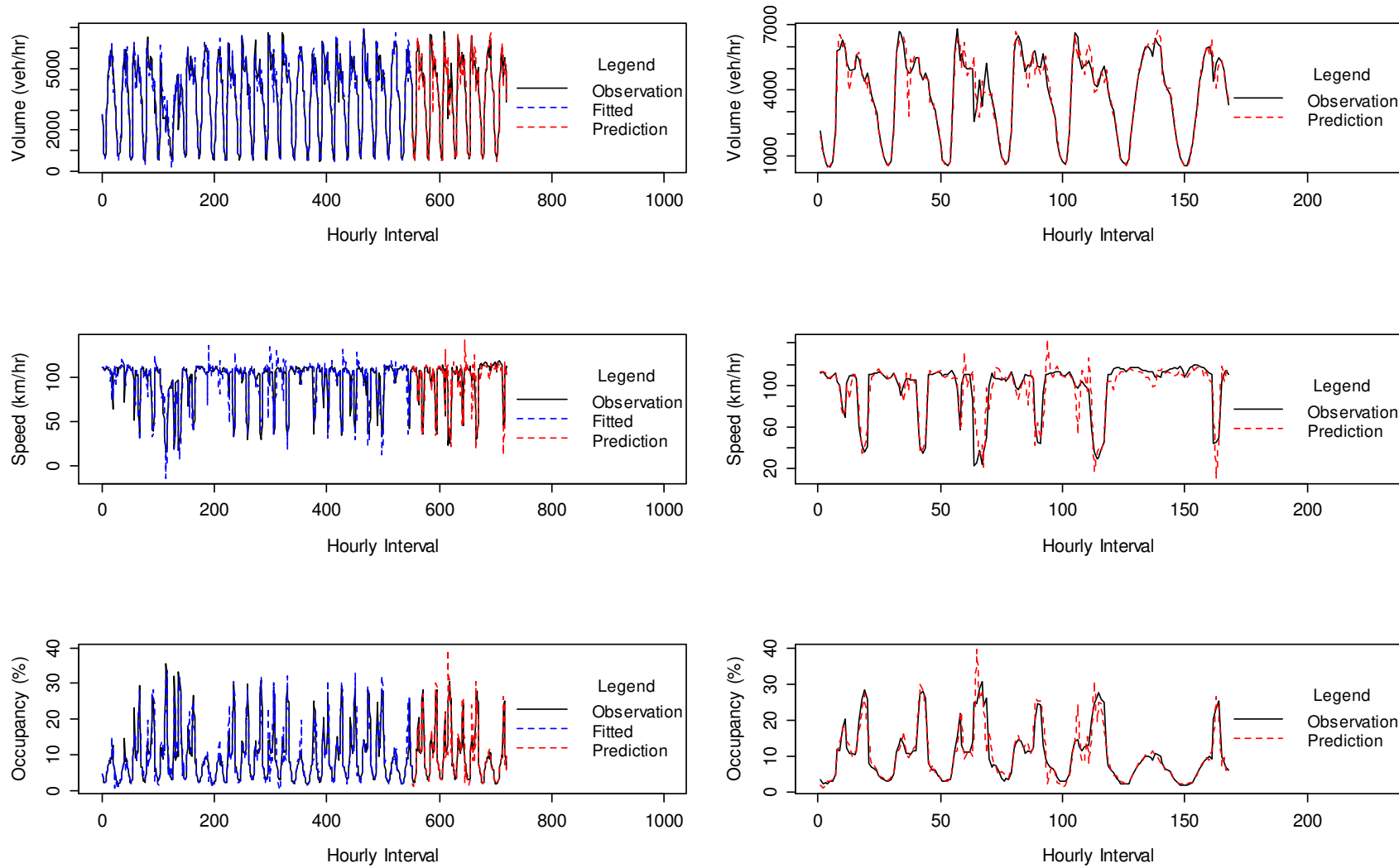


Fig. A. 3 401.EB.Exp.3, Allen Rd. @ Hwy401, 401DE0060DEE

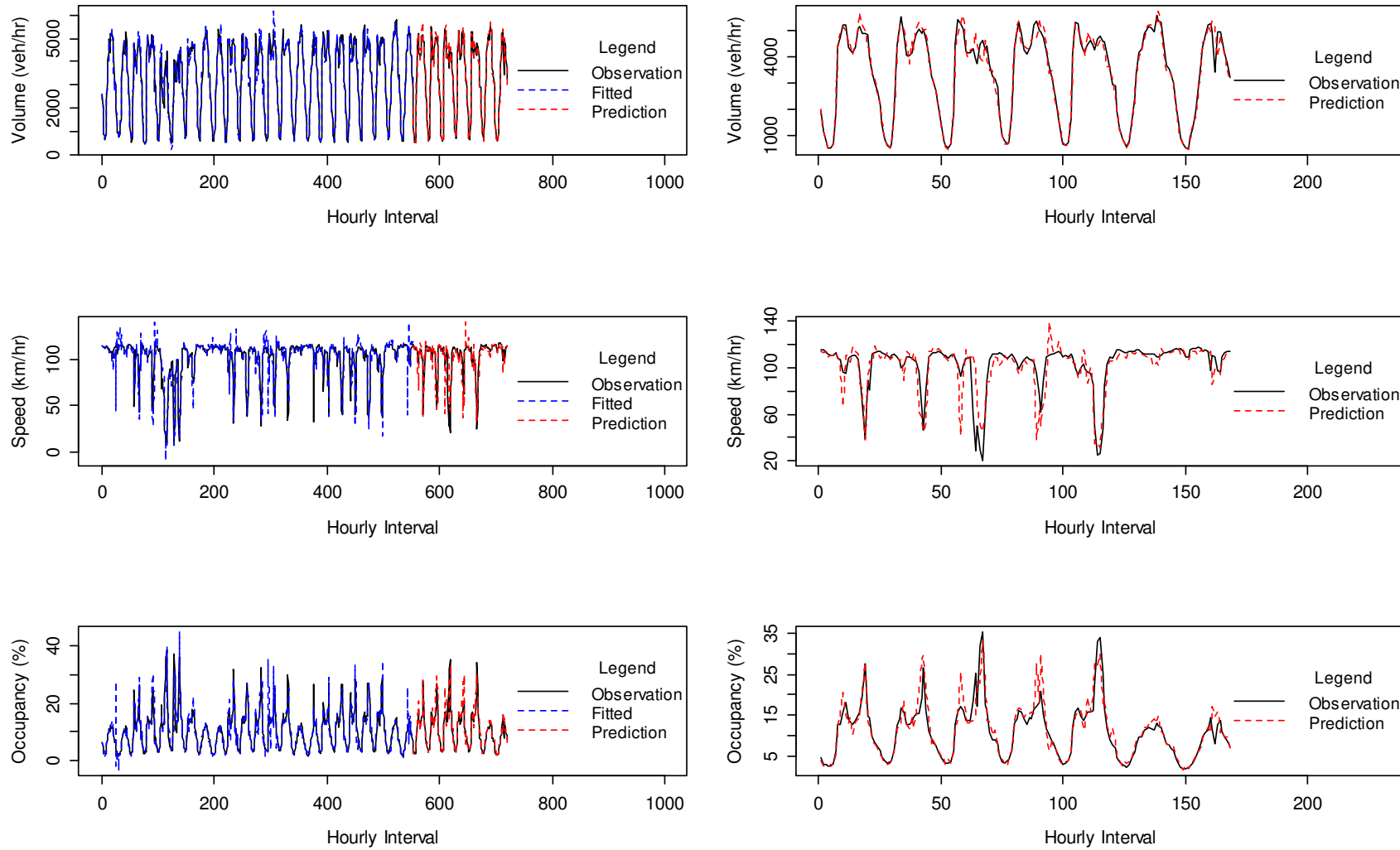


Fig. A. 4 401.EB.Exp.4, Allen Rd. - Yonge St., 401DE0100DEE

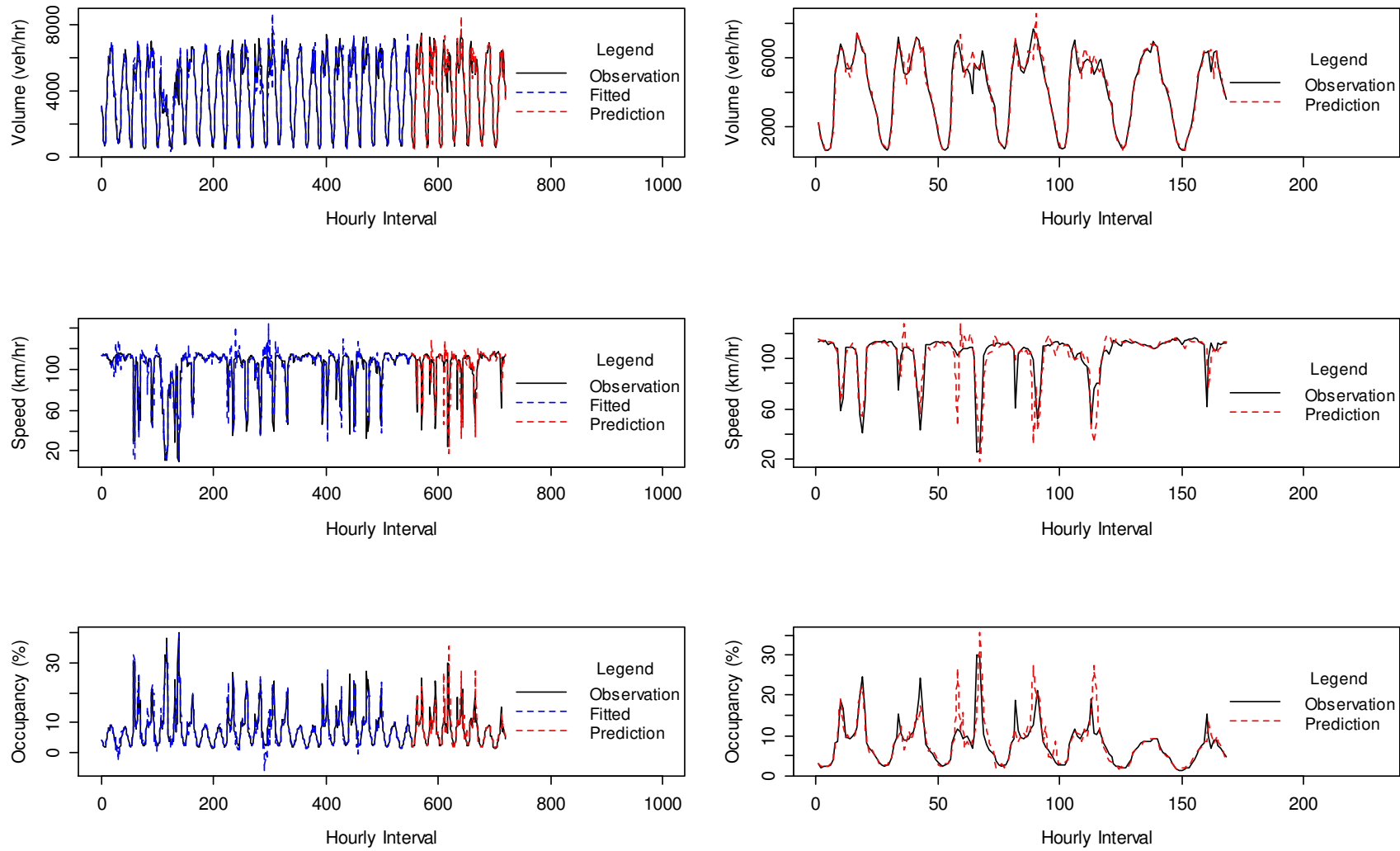


Fig. A. 5 401.EB.Exp.5, Yonge St. - Leslie St., 401DE0150DEE

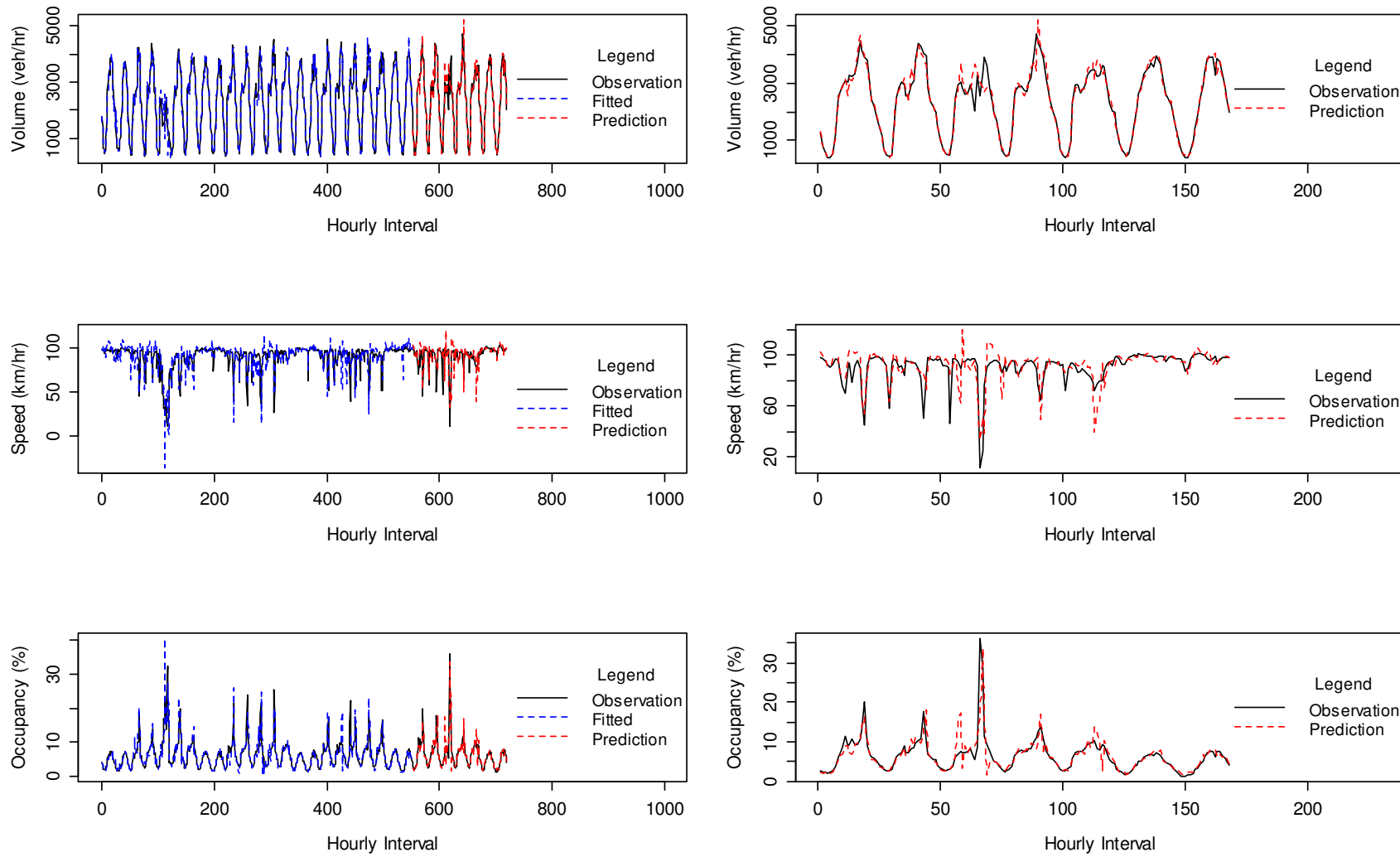


Fig. A. 6 401.EB.Exp.6, Leslie St. - Hwy404, 401DE0210DEE

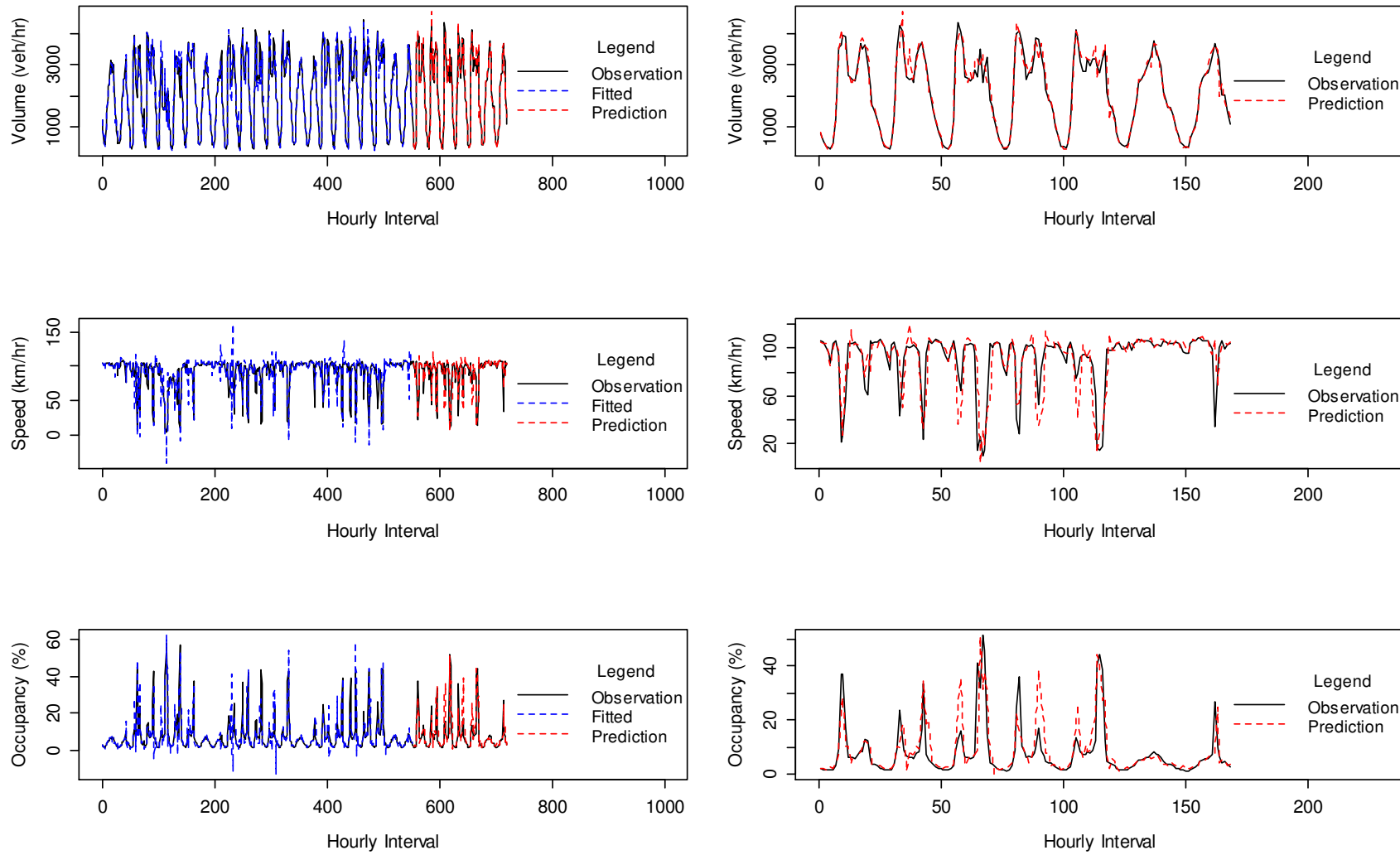


Fig. A. 7 401.EB.Col.1, Hwy400 - Jane St., 401DW0040DEC

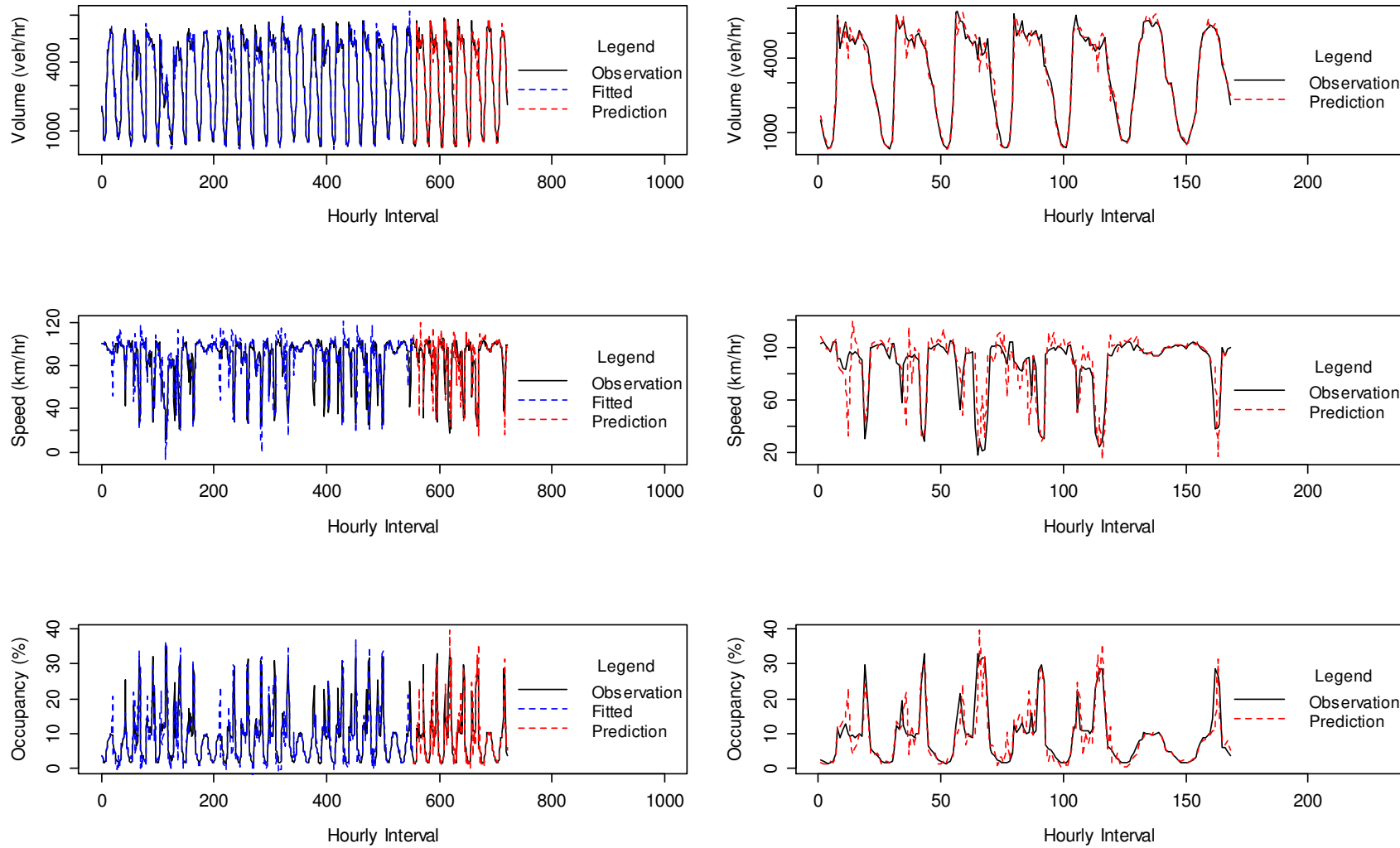


Fig. A. 8 401.EB.Col.2, Jane St. - Keele St., 401DW0010DEC

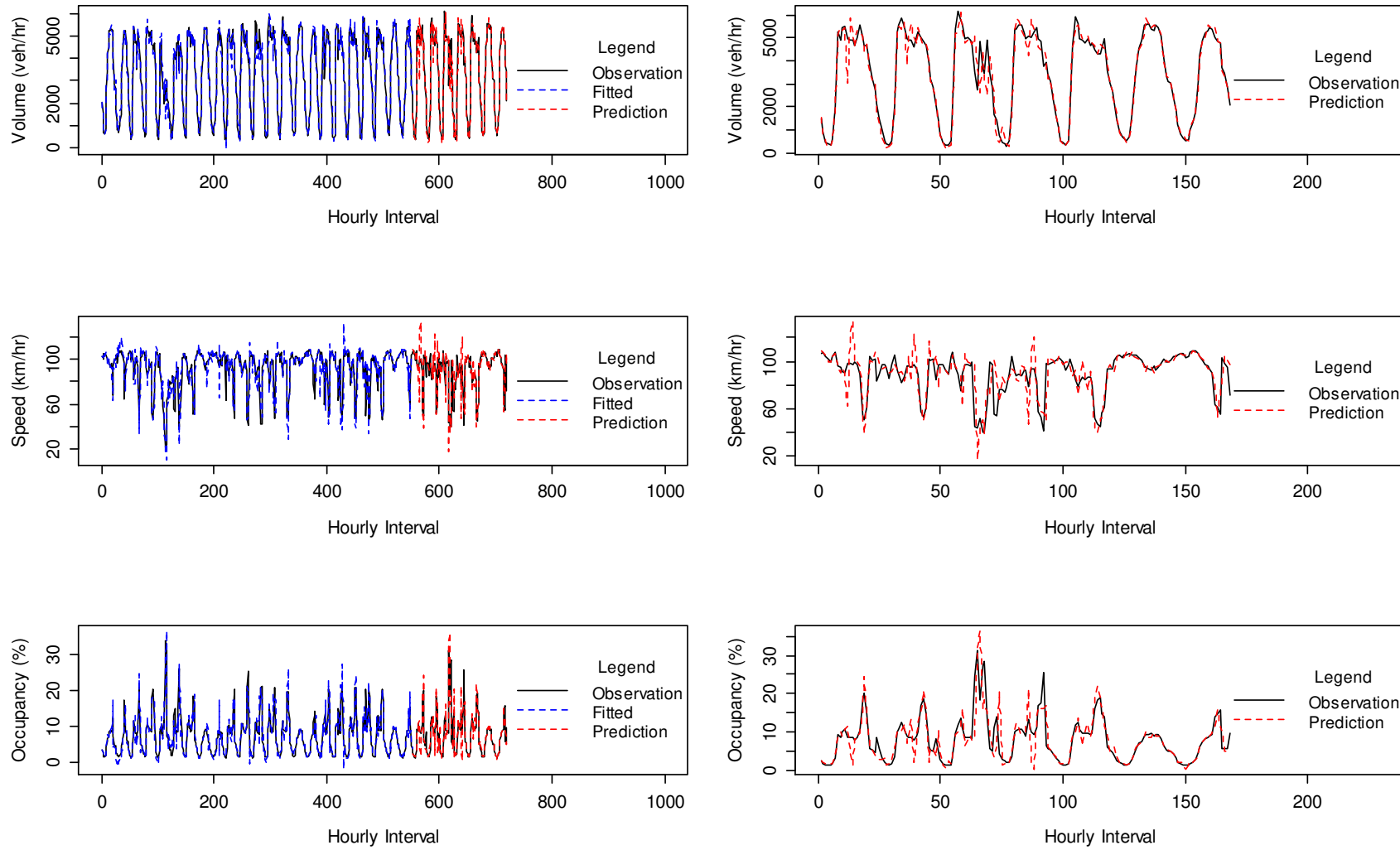


Fig. A. 9 401.EB.Col.3, Keele St. - Dufferin St., 401DE0040DEC

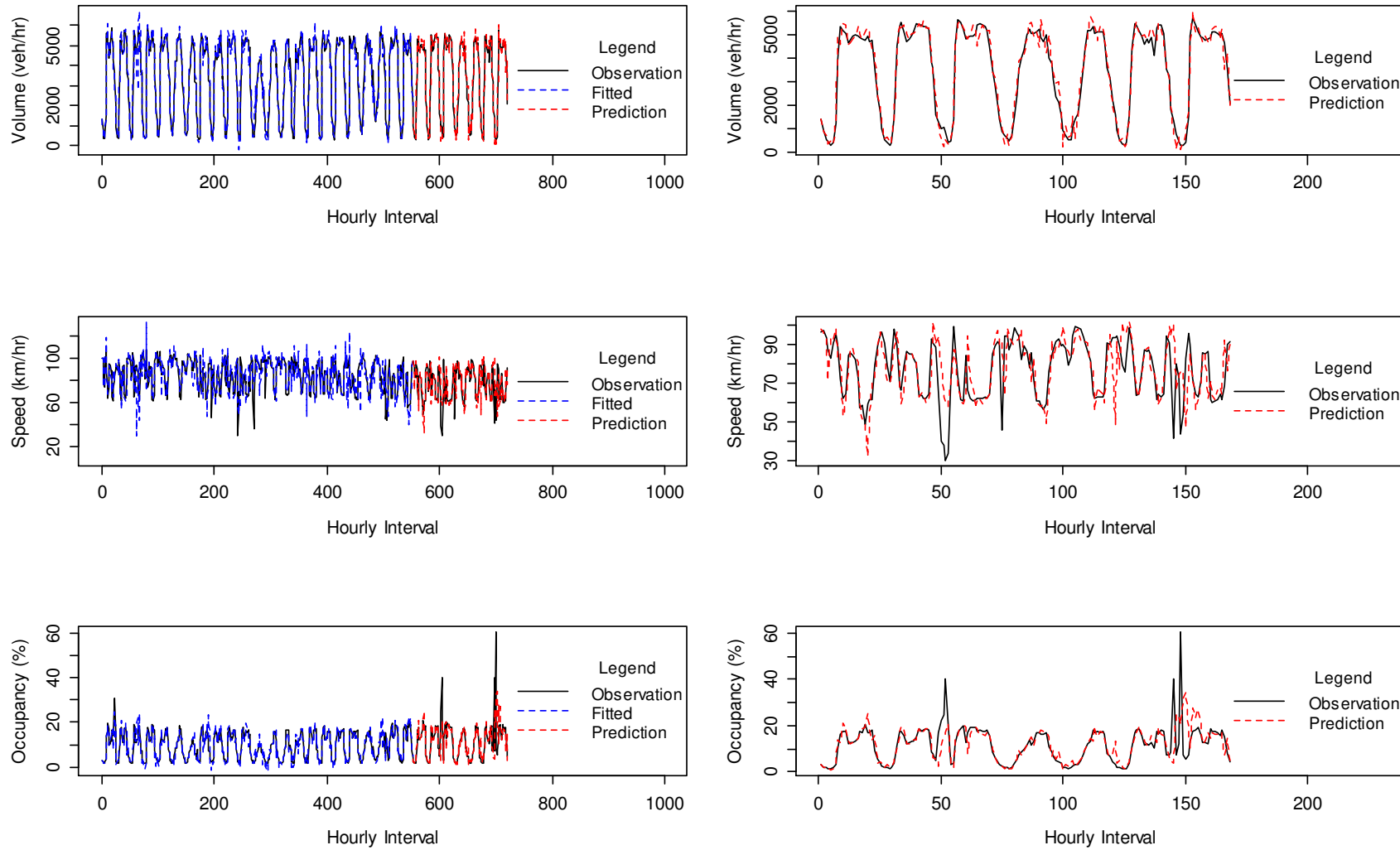


Fig. A. 10 401.EB.Col.4, Bathurst St. - Avenue Rd., 401DE0090DEC

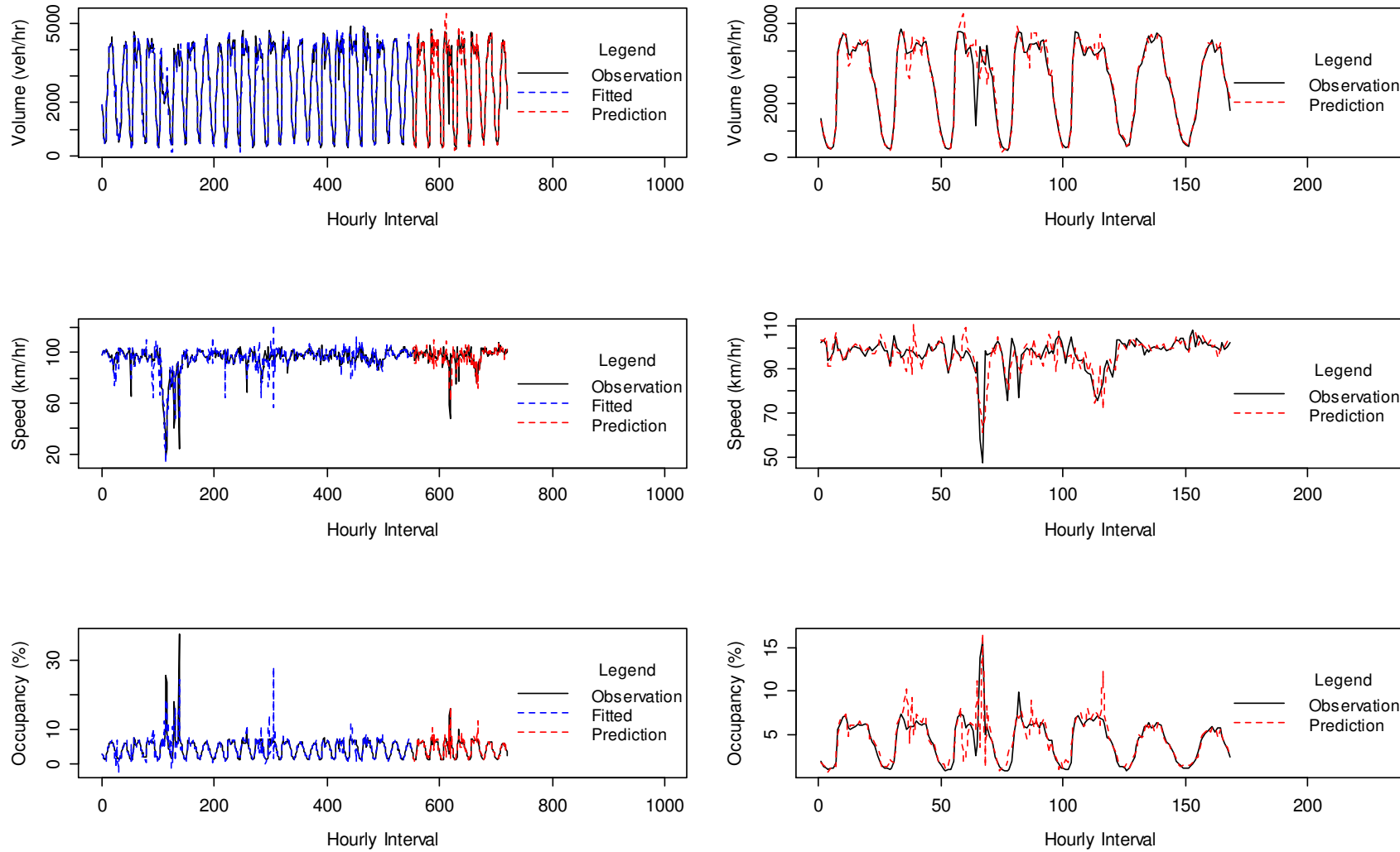


Fig. A. 11 401.EB.Col.5, Avenue Rd. - Yonge St., 401DE0120DEC

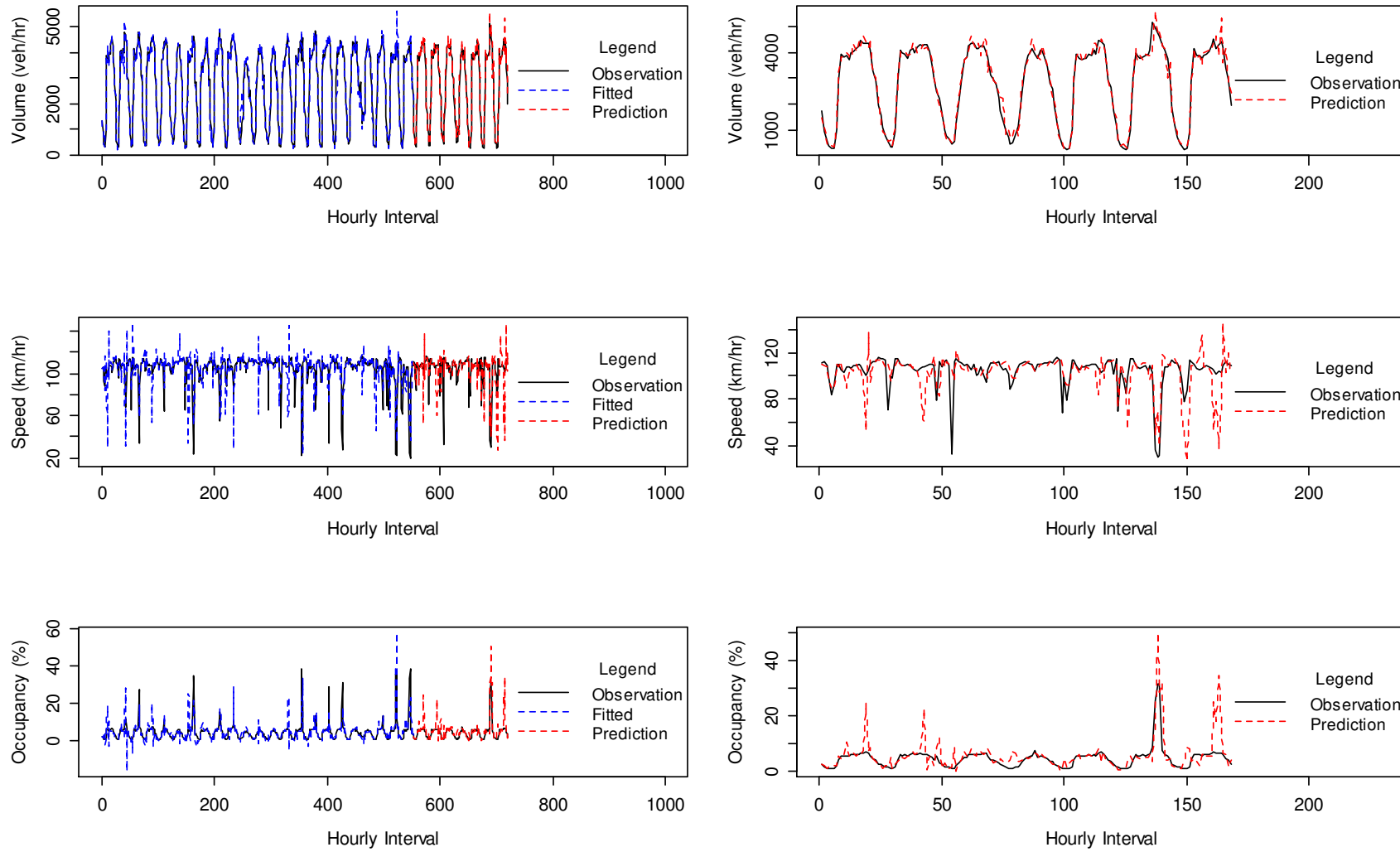


Fig. A. 12 401.EB.Col.6, Yonge St. - Bayview Ave., 401DE0150DEC

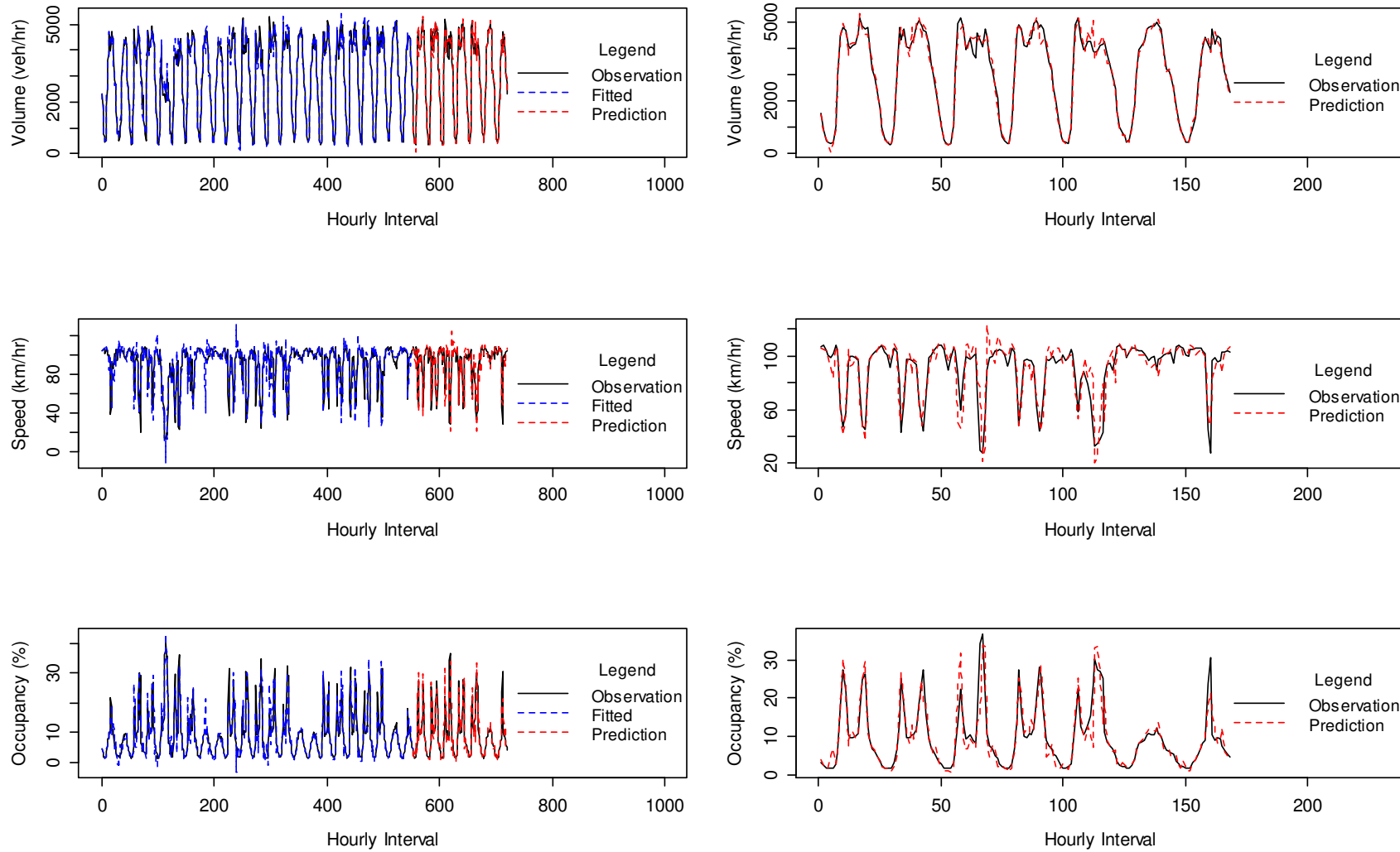


Fig. A. 13 401.EB.Col.7, Bayview Ave. - Leslie St., 401DE0180DEC

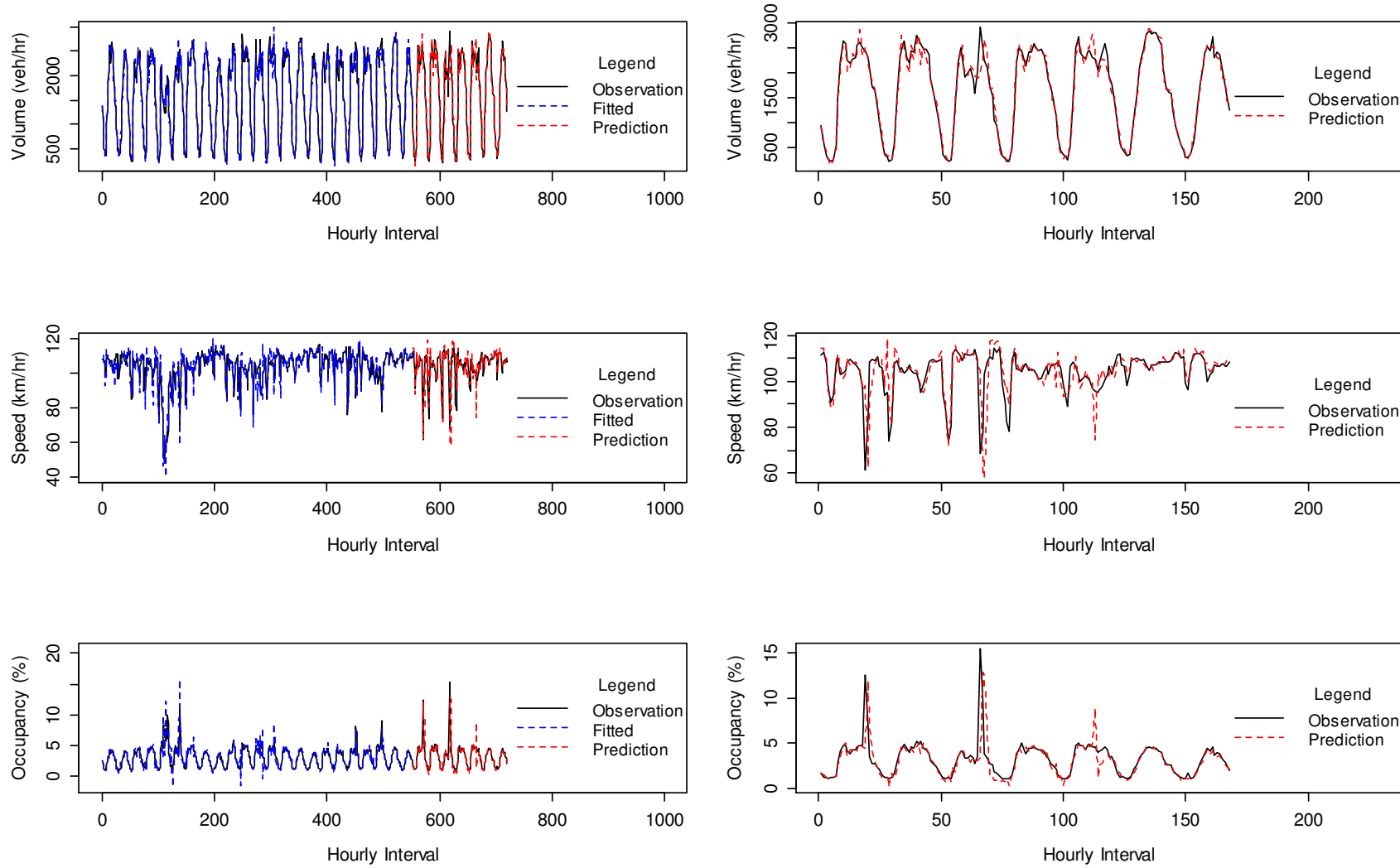


Fig. A. 14 401.EB.Col.8, Leslie St. - Hwy404, 401DE0220DEC

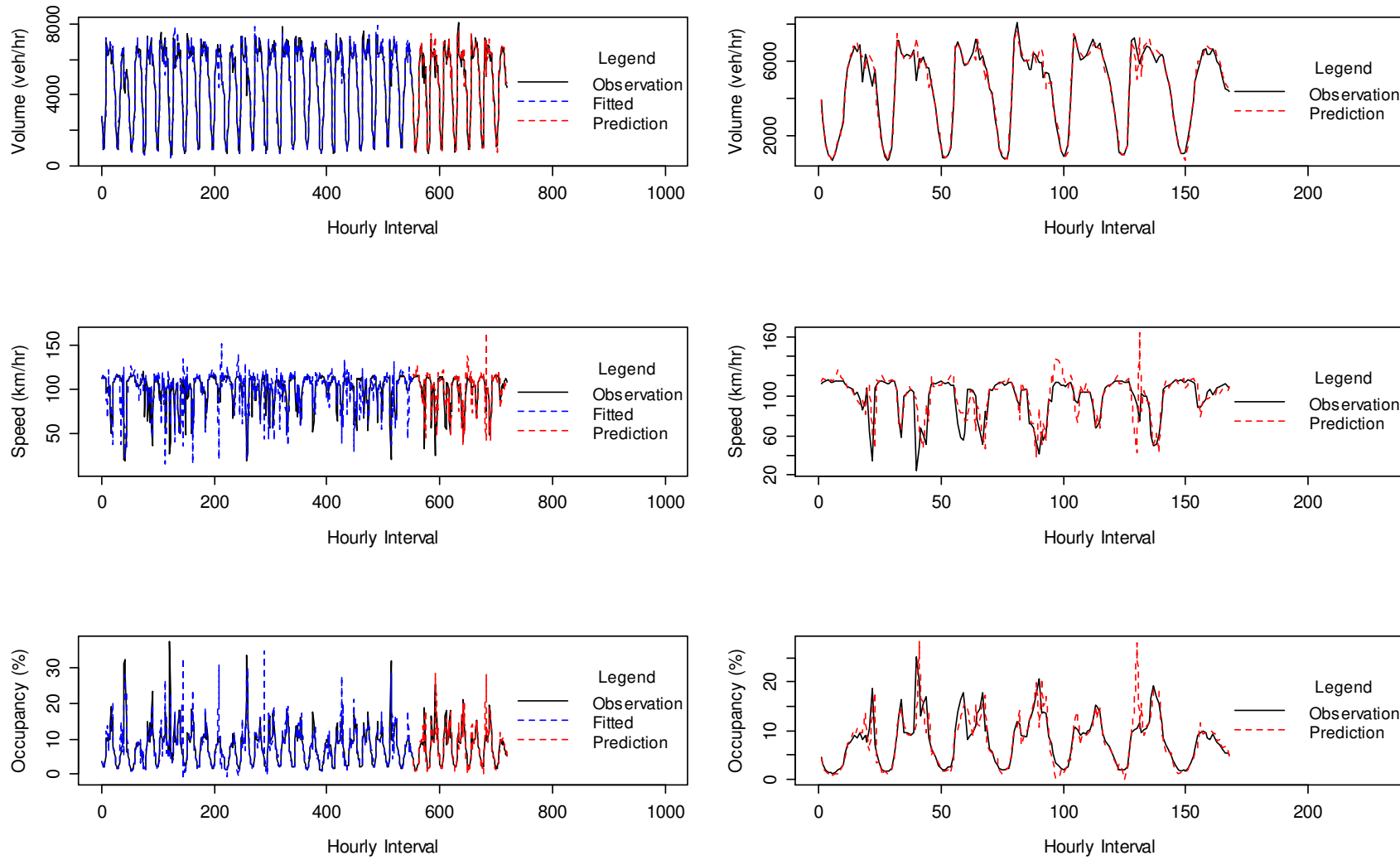


Fig. A. 15 401.WB.Exp.1, Hwy400 - Basket Wave, 401DW0040DWE

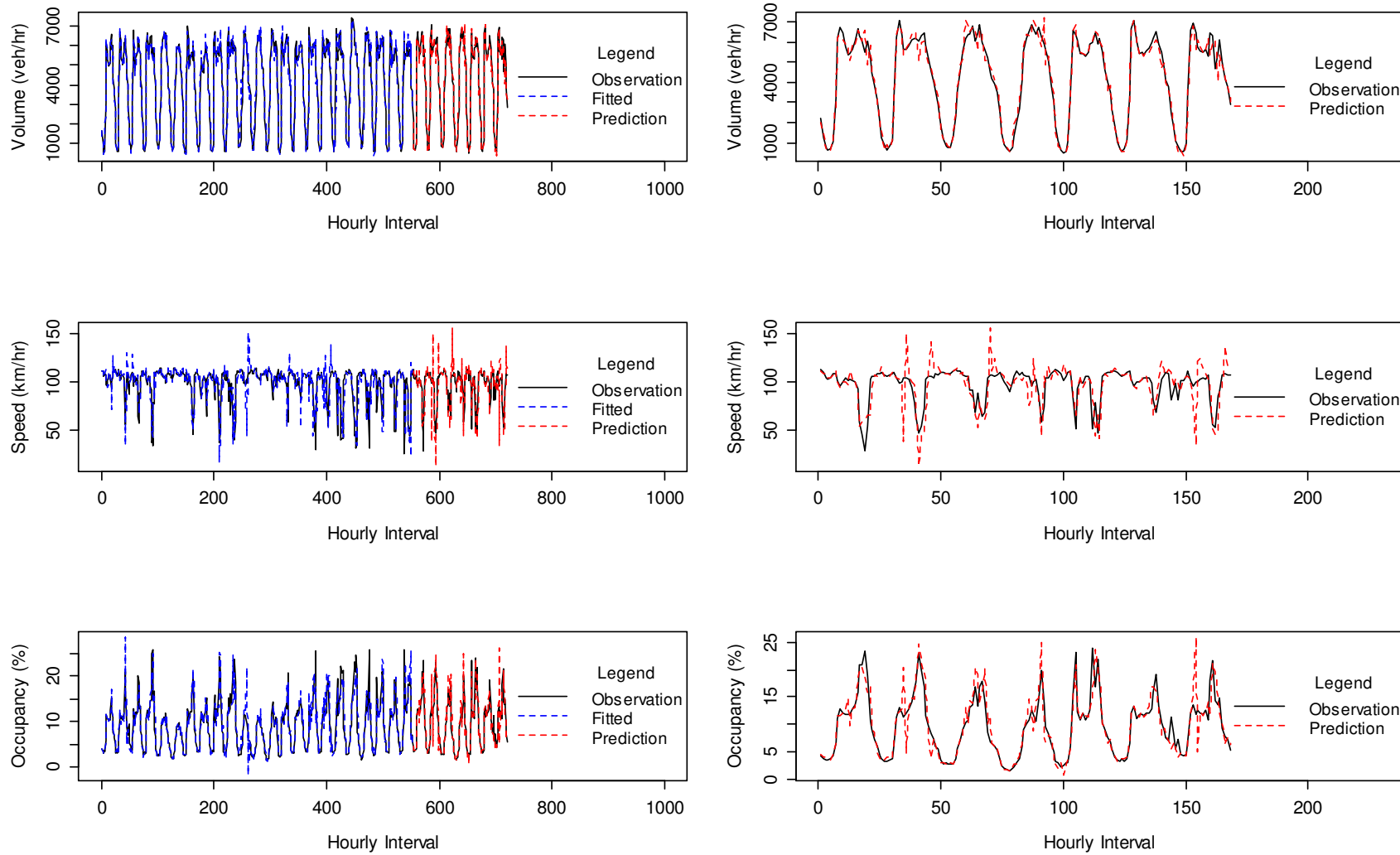


Fig. A. 16 401.WB.Exp.2, Basket Wave - Dufferin St., 401DE0020DWE

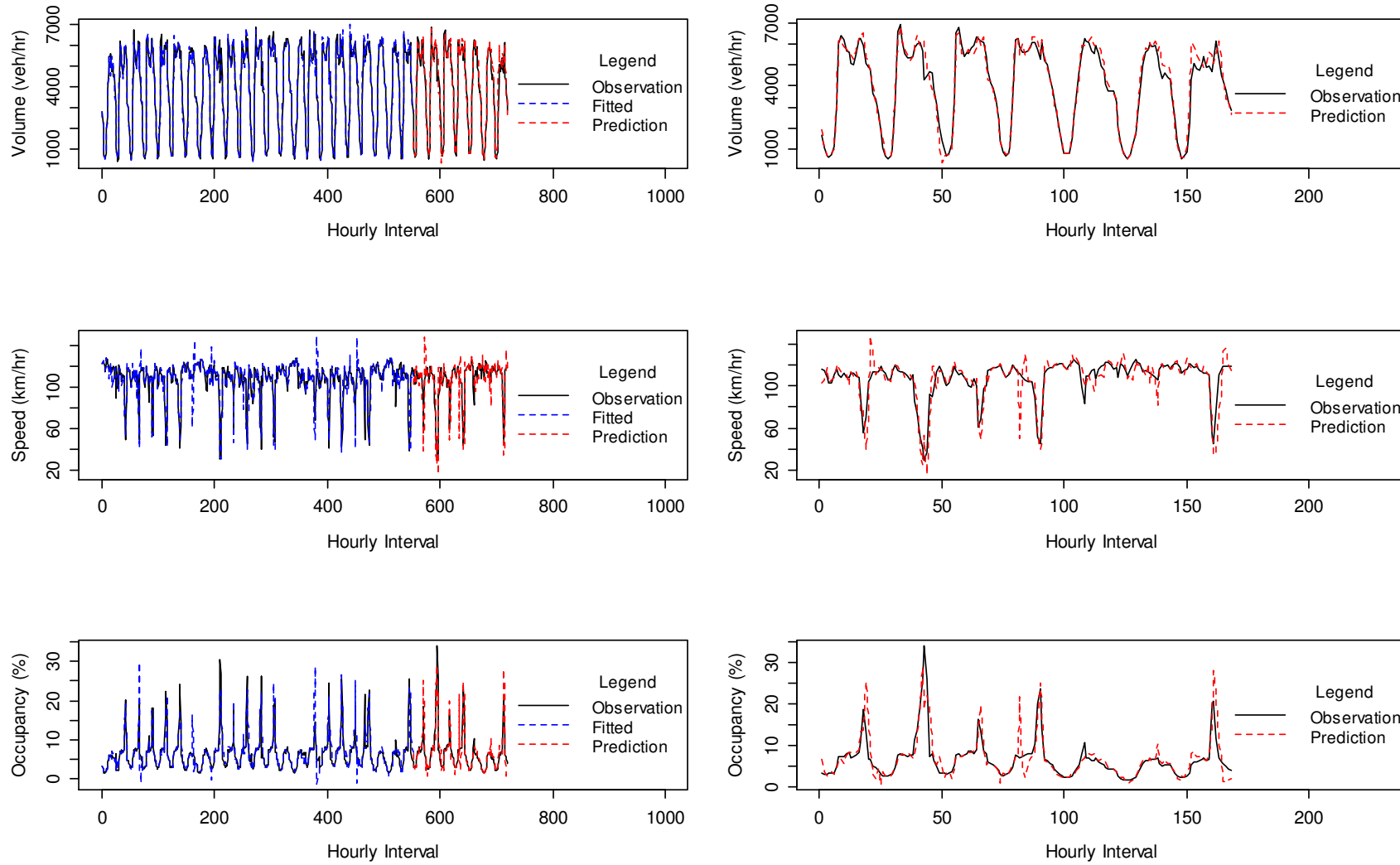


Fig. A. 17 401.WB.Exp.3, Allen Rd. @ Hwy401, 401DE0060DWE

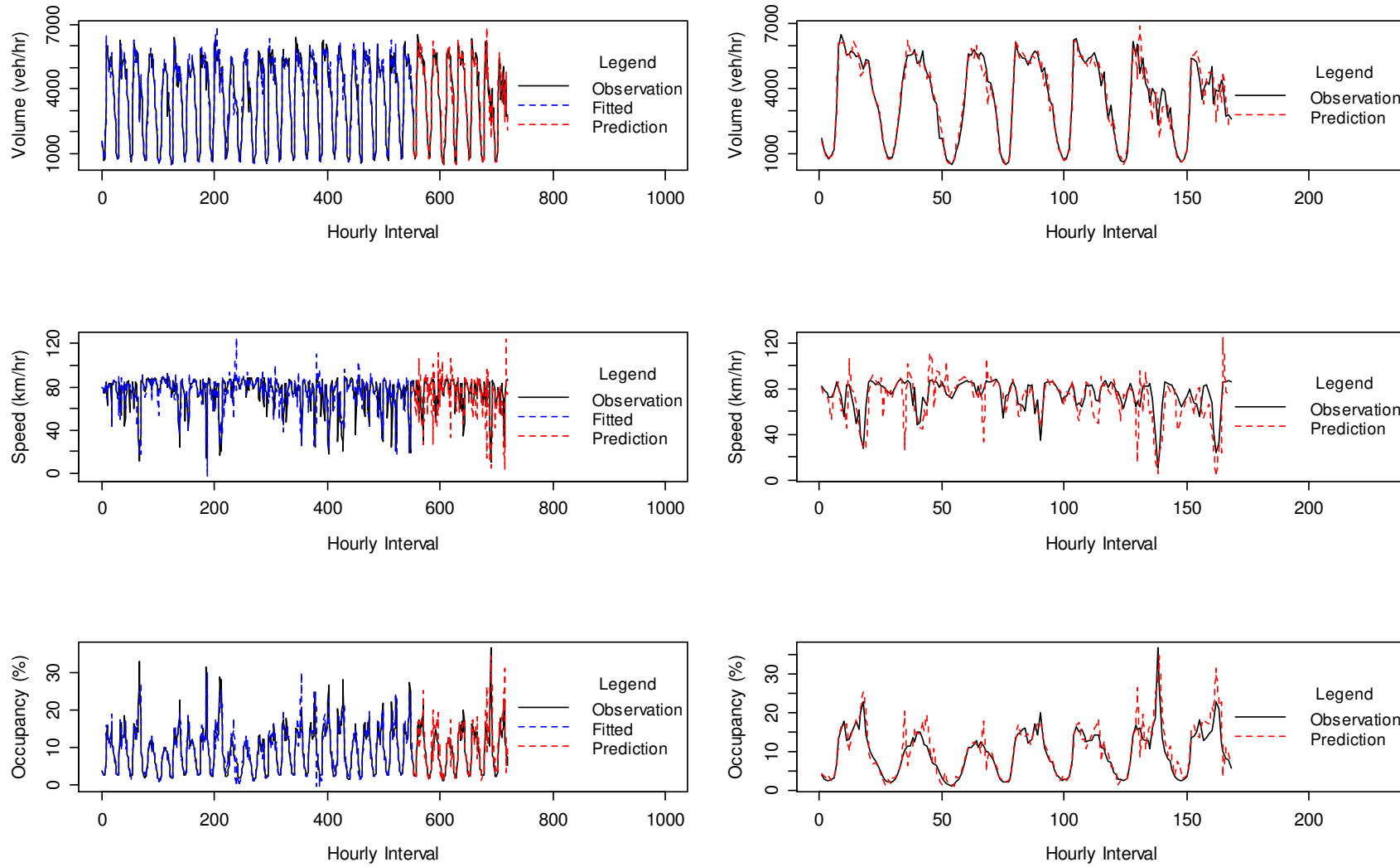


Fig. A. 18 401.WB.Exp.4, Bathurst St. - Yonge St., 401DE0090DWE

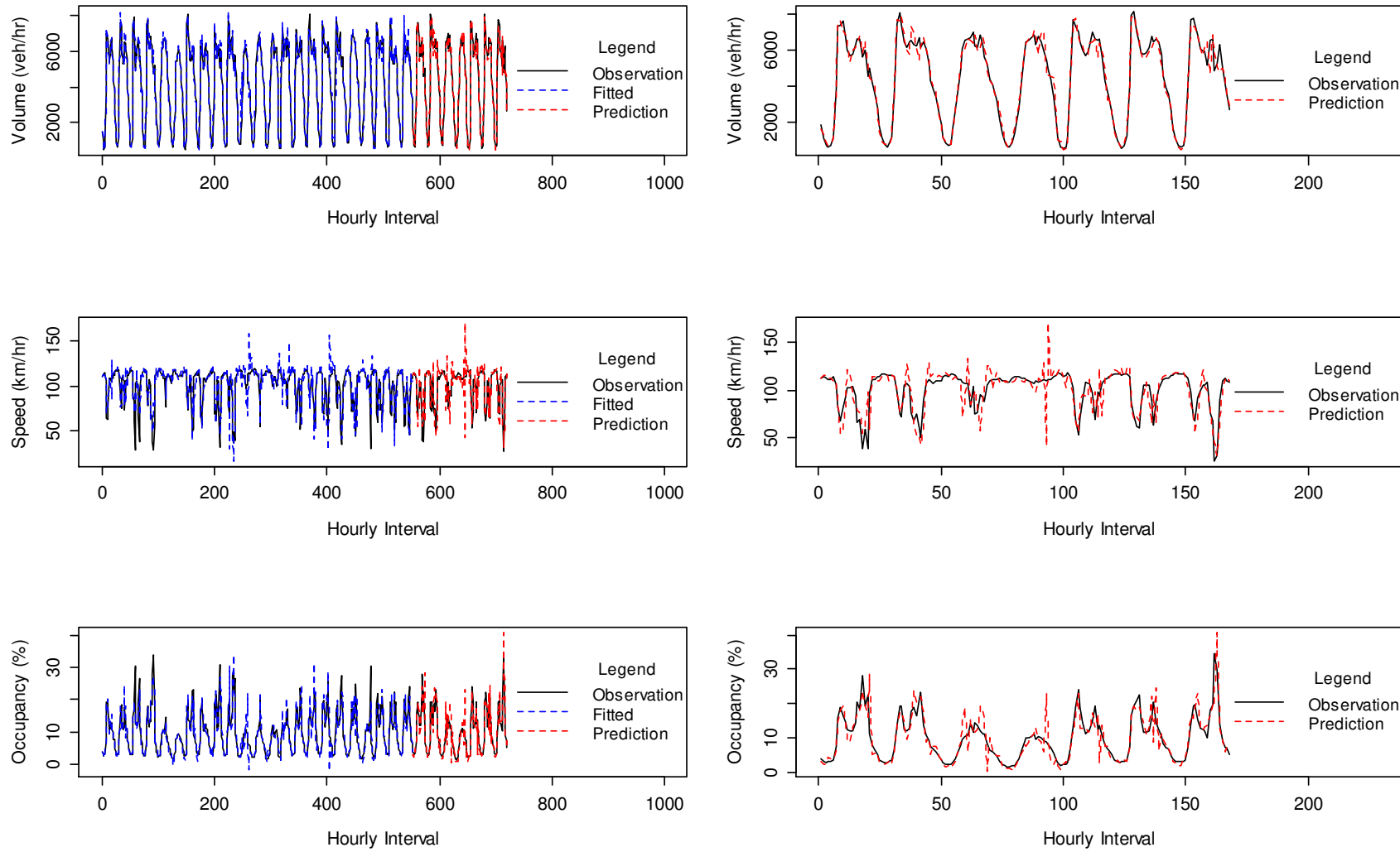


Fig. A. 19 401.WB.Exp.5, Yonge St. - Bayview Ave., 401DE0140DWE

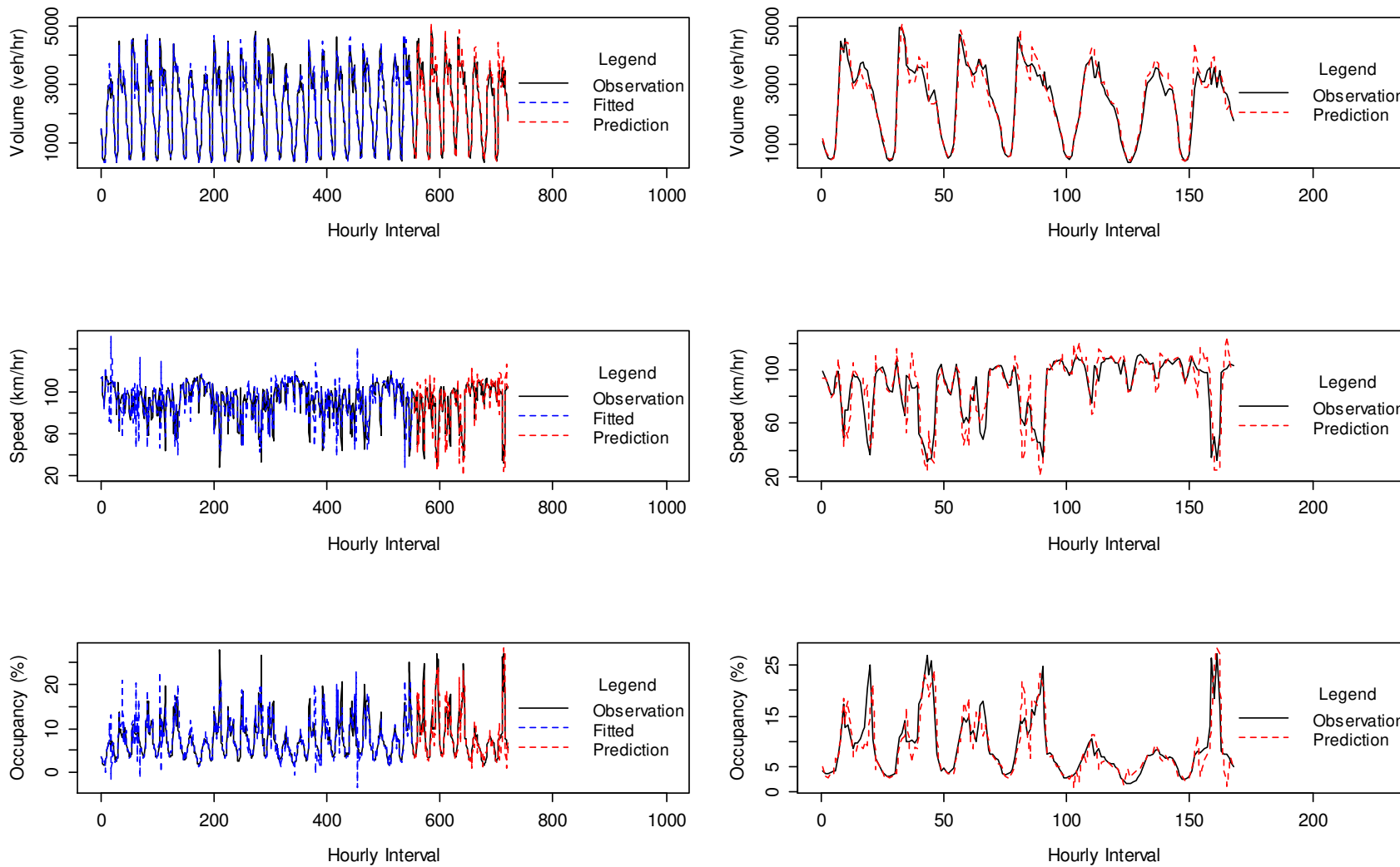


Fig. A. 20 401.WB.Exp.6, Bayview Ave. - Hwy404, 401DE0180DWE

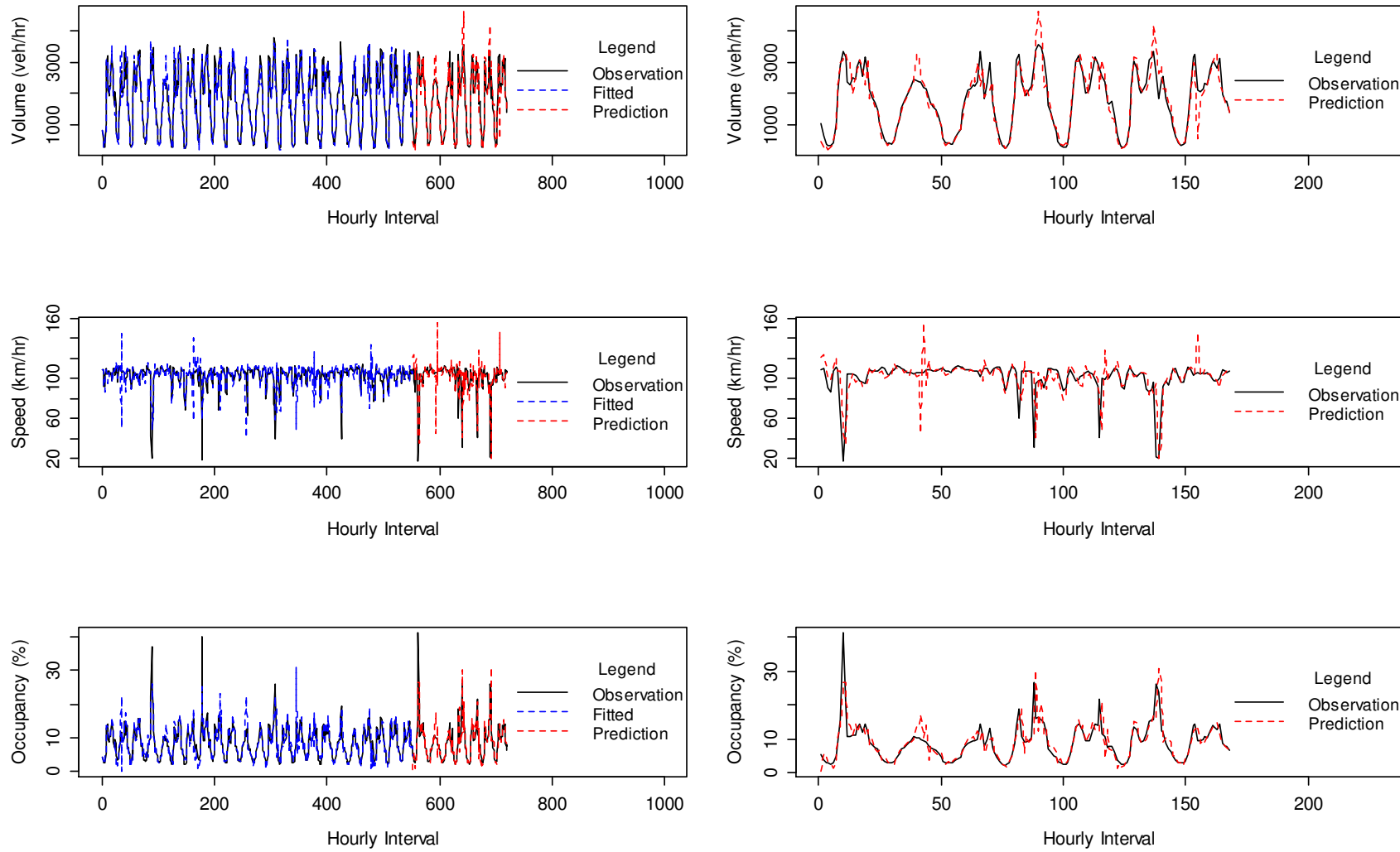


Fig. A. 21 401.WB.Col.1, Hwy400 - Jane St., 401DW0040DWC

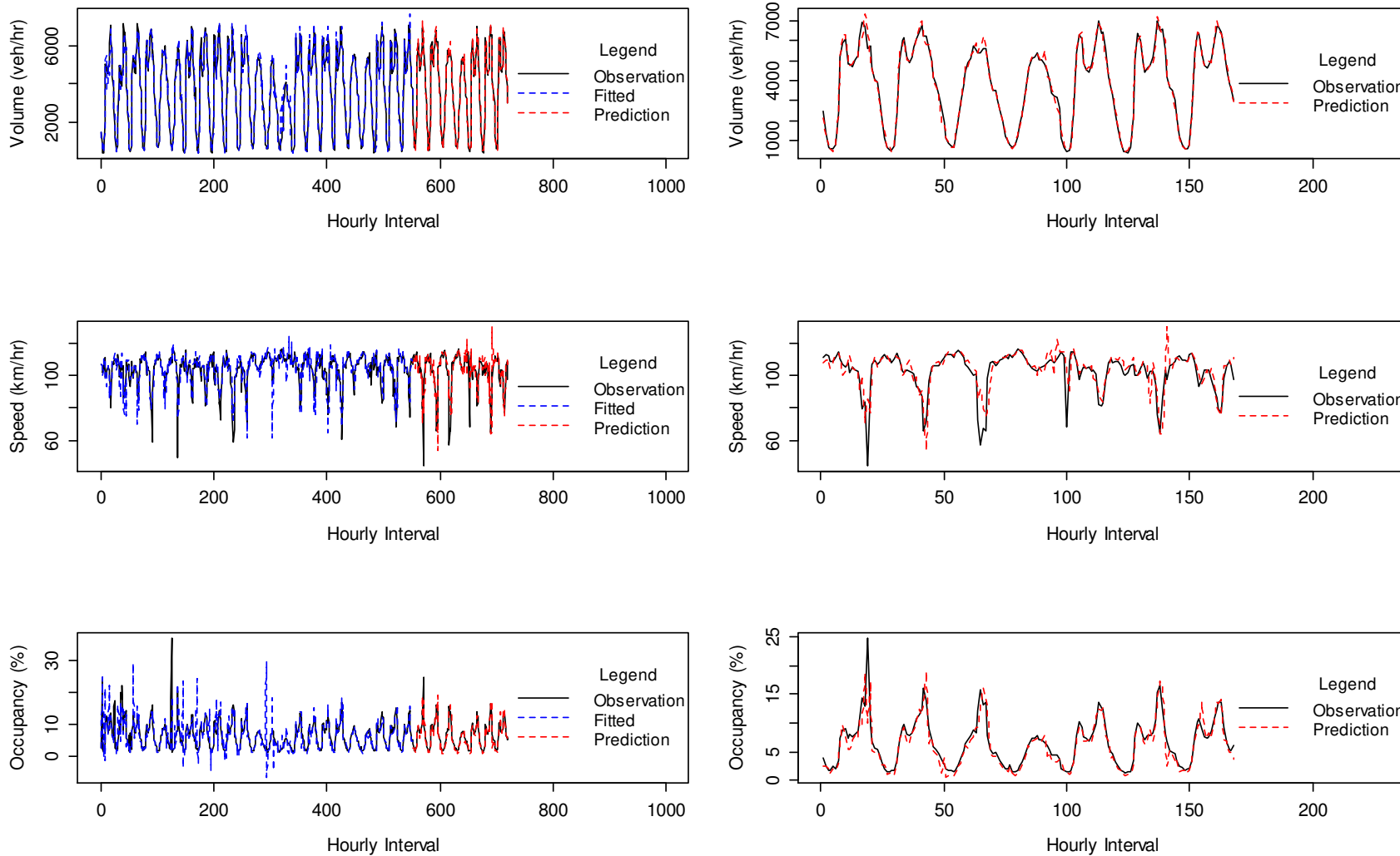


Fig. A. 22 401.WB.Col.2, Jane St. - Keele St., 401DW0010DWC

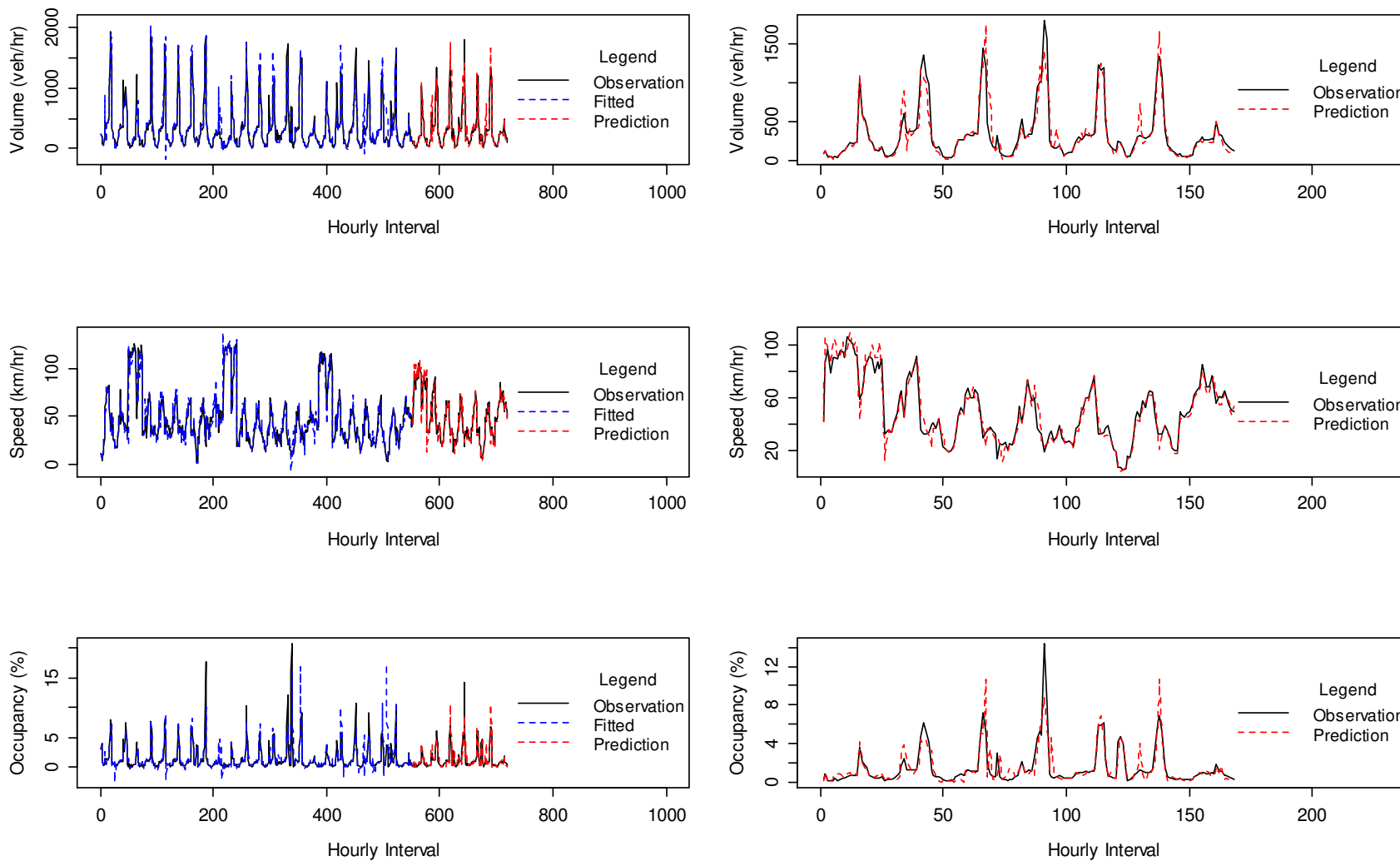


Fig. A. 23 401.WB.Col.3, Keele St. - Dufferin St., 401DE0030DWC

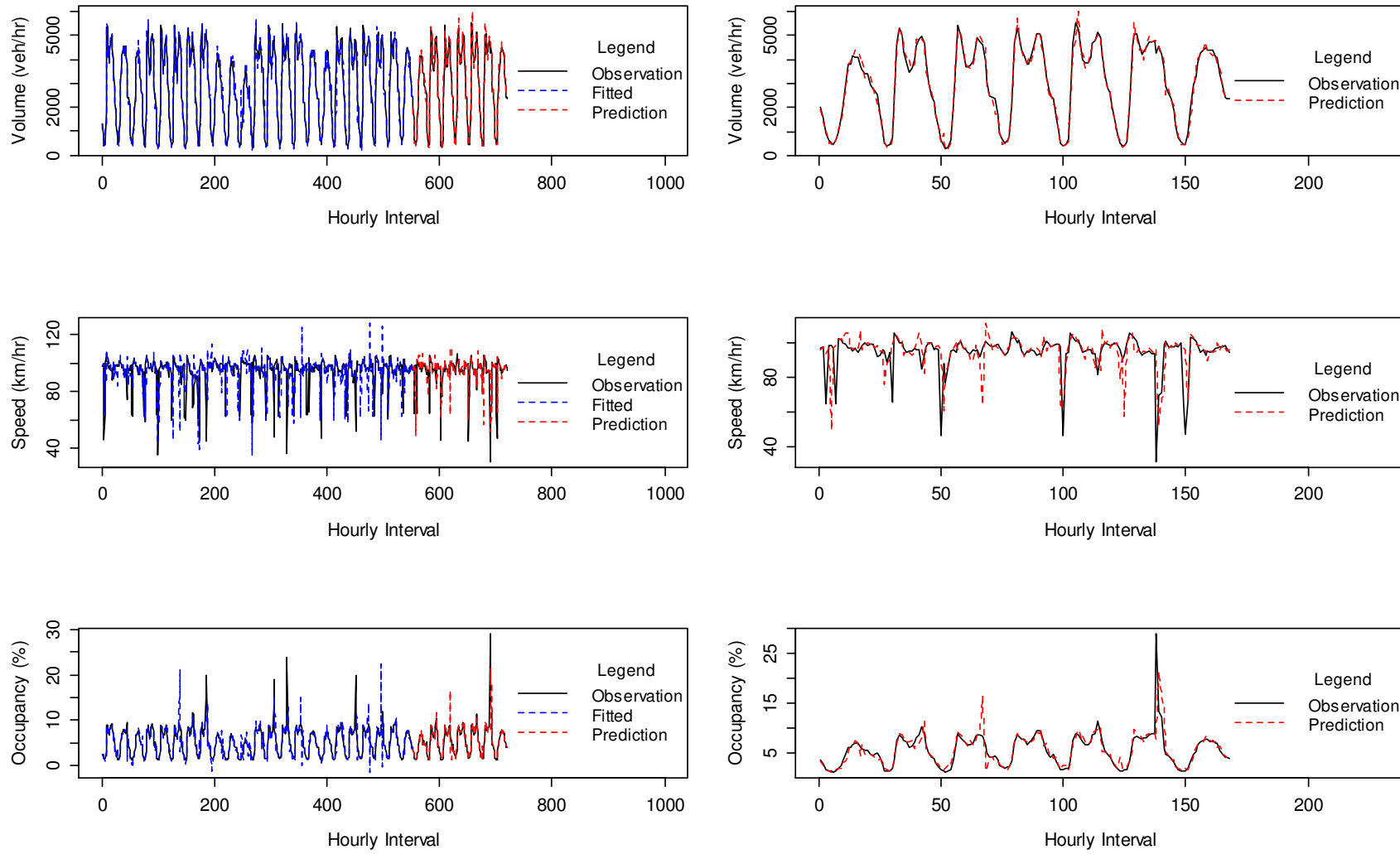


Fig. A. 24 401.WB.Col.4, Allen Rd. - Bathurst St., 401DE0070DWC

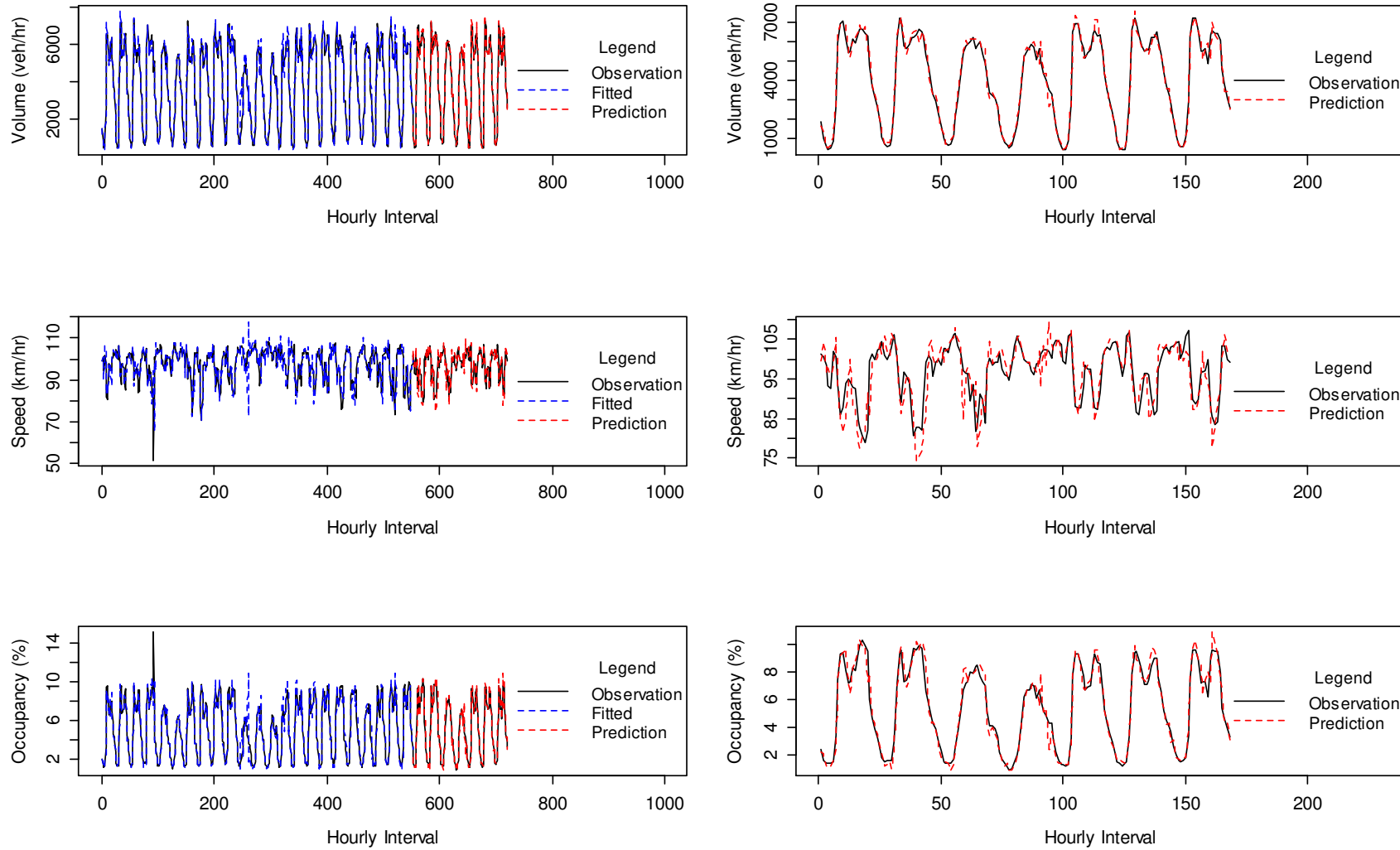


Fig. A. 25 401.WB.Col.5, Bathurst St. - Avenue Rd., 401DE0090DWC

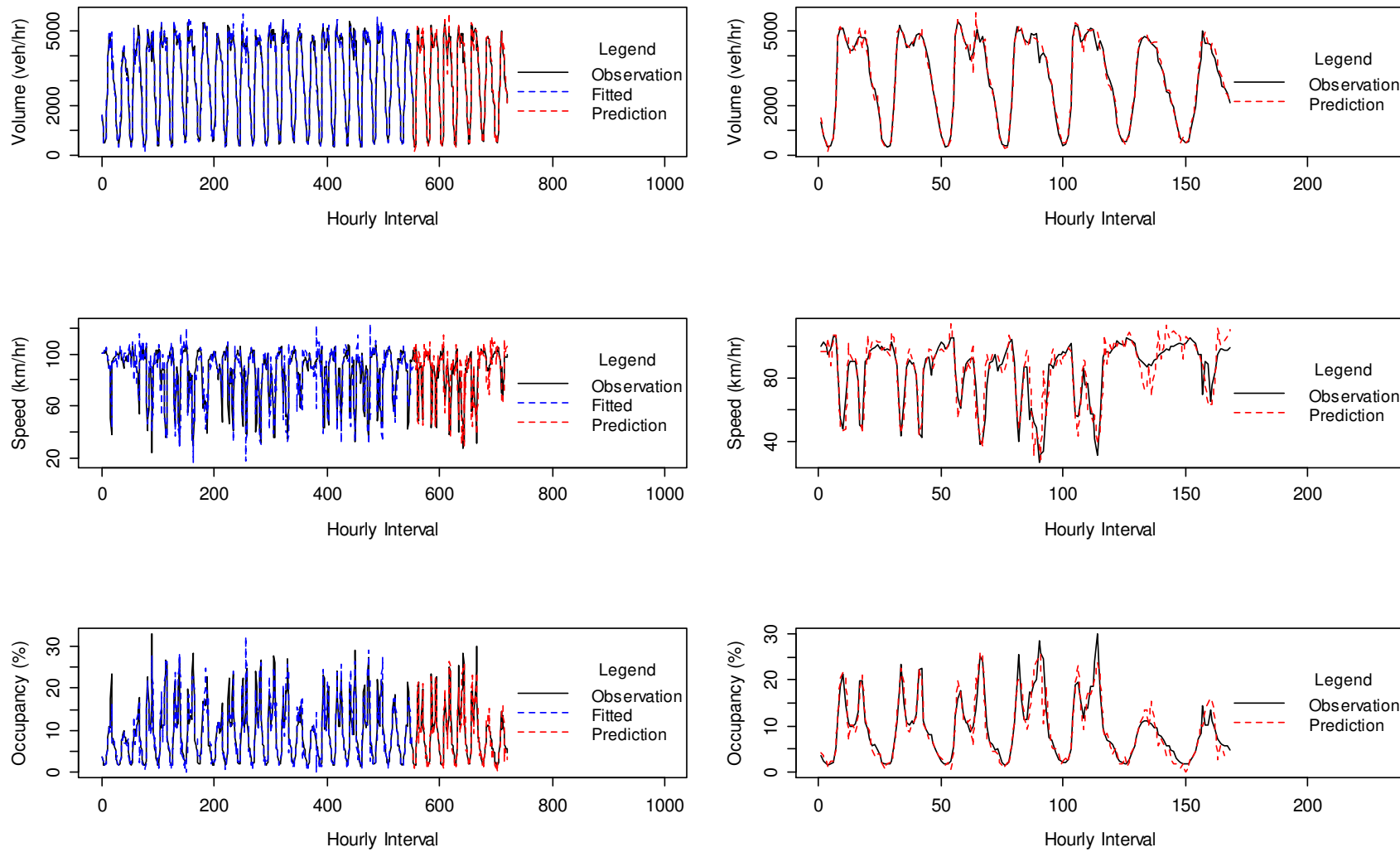


Fig. A. 26 401.WB.Col.6, Avenue Rd. - Yonge St., 401DE0120DWC

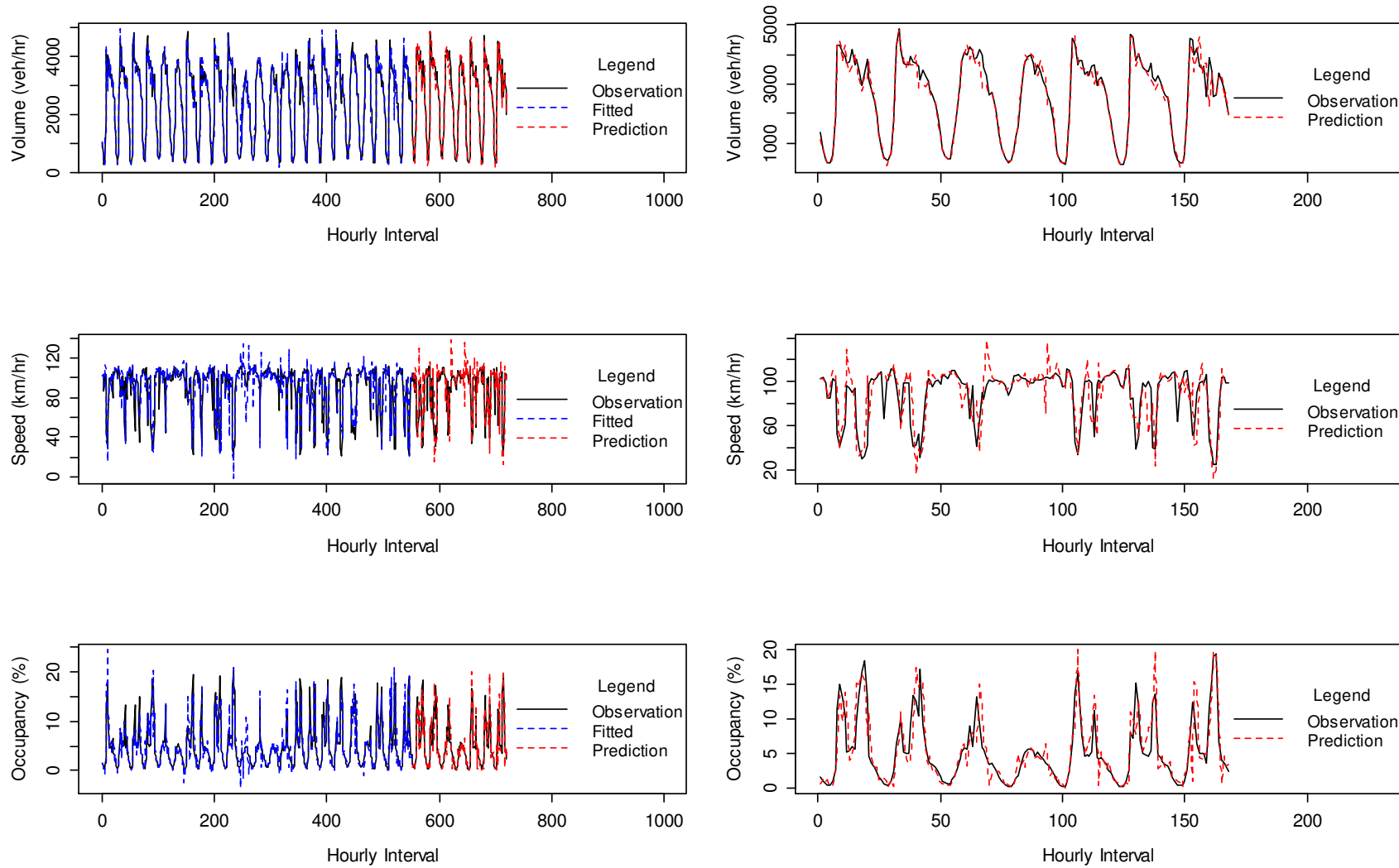


Fig. A. 27 401.WB.Col.7, Yonge St. - Bayview Ave., 401DE0140DWC

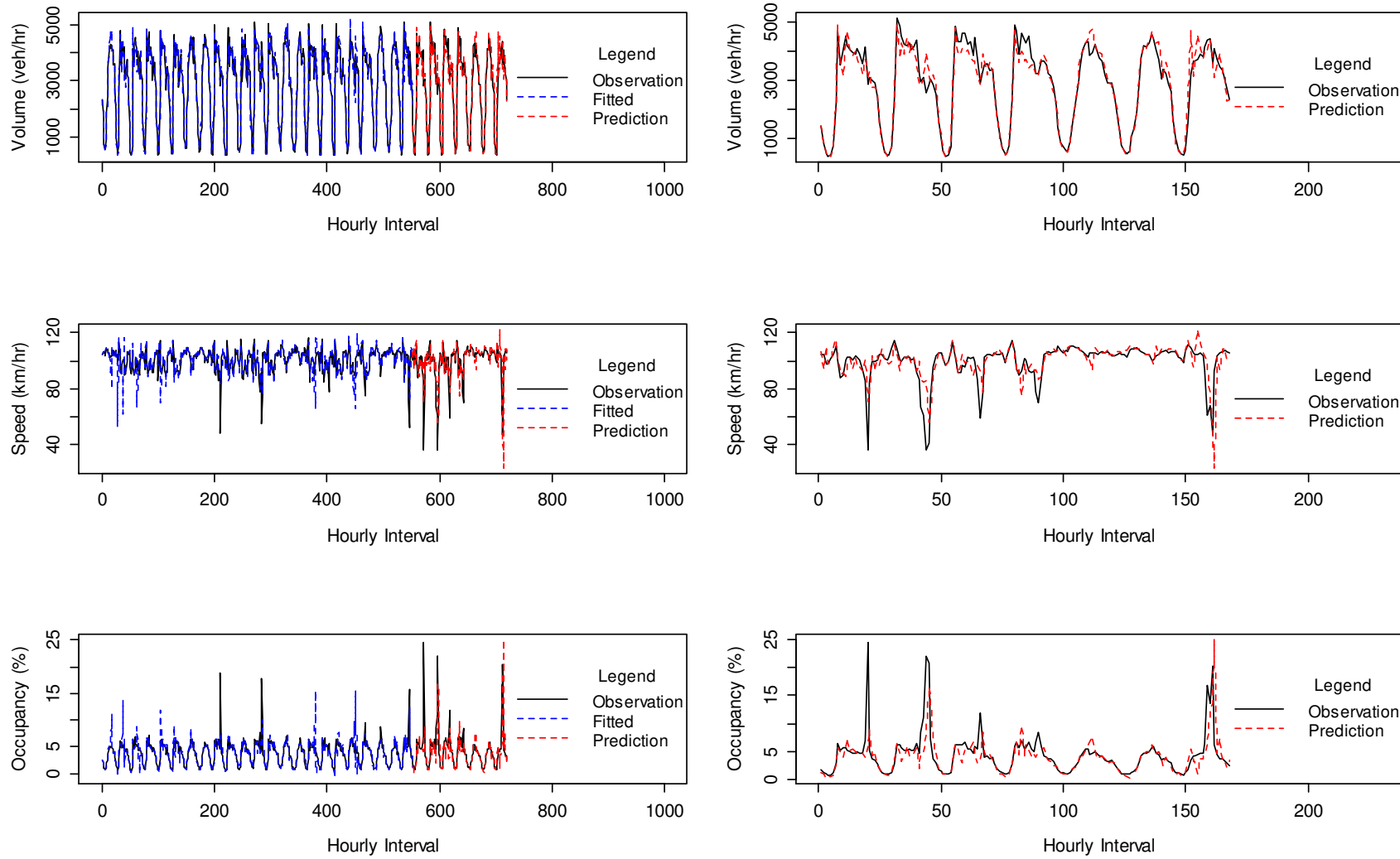


Fig. A. 28 401.WB.Col.8, Bayview Ave. - Leslie St., 401DE0170DWC

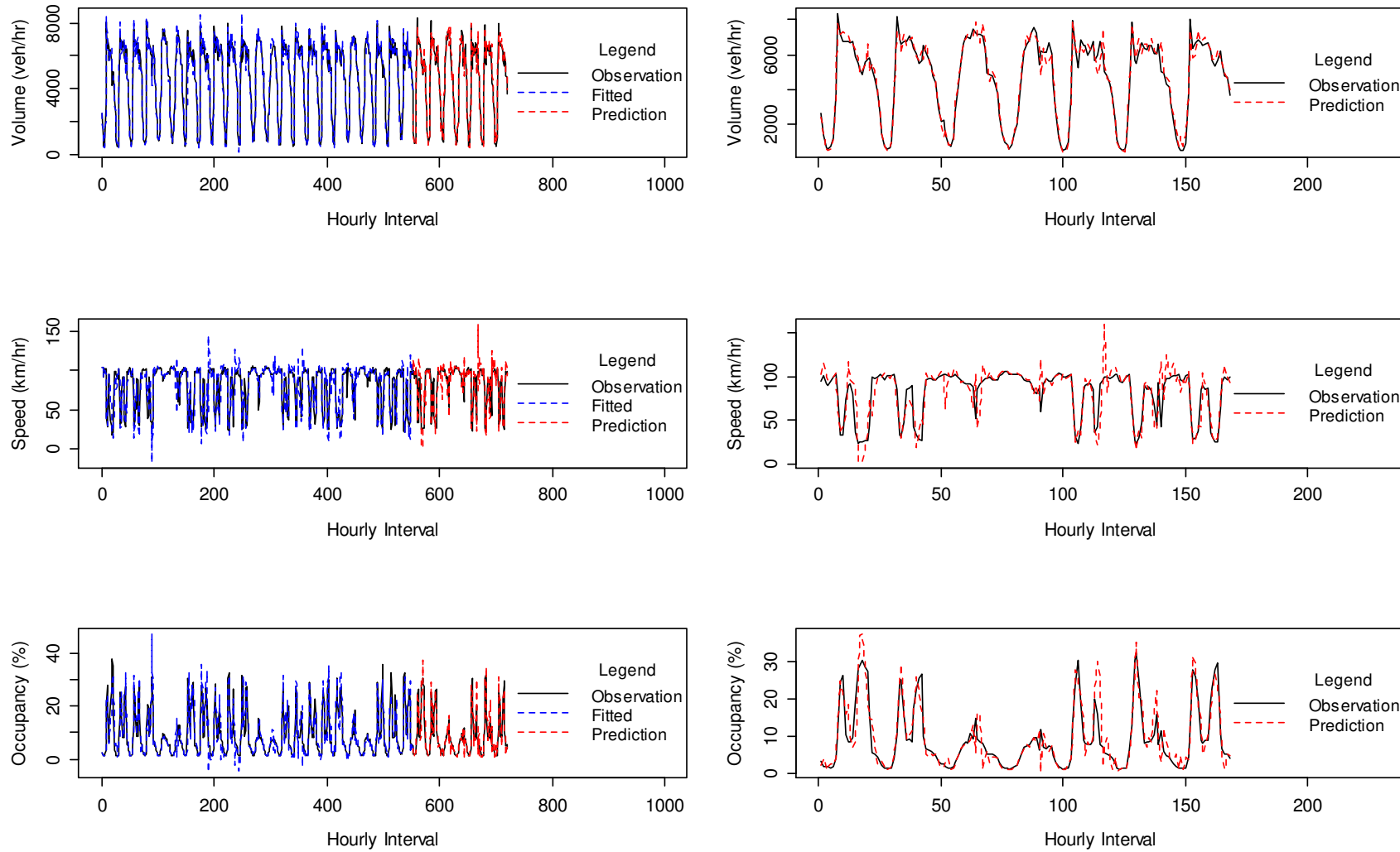


Fig. A. 29 401.WB.Col.9, Leslie St. - Hwy404, 401DE0200DWC

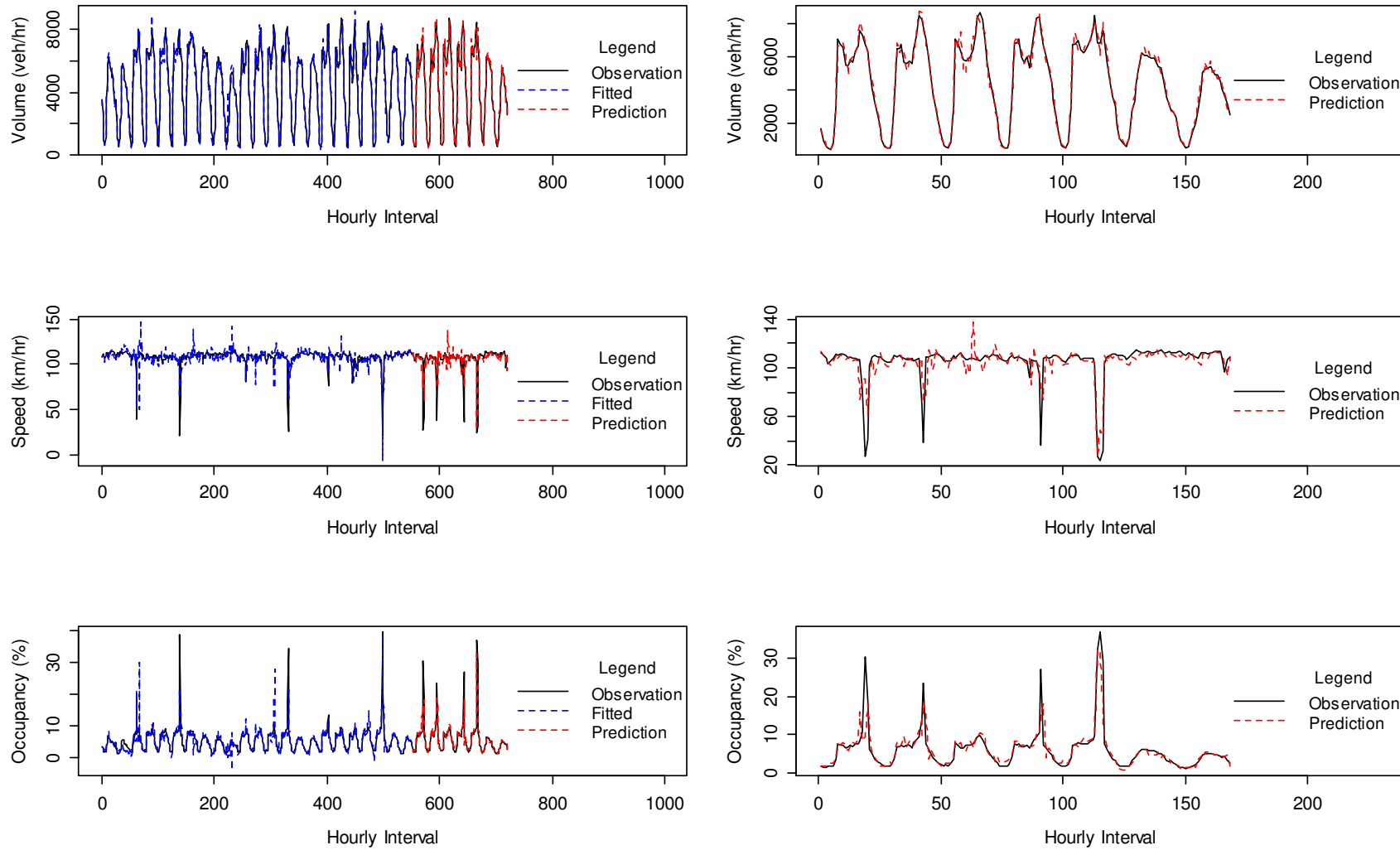


Fig. A. 30 400.NB.1, Hwy401 - Finch Ave. W., 400DN0040DNS

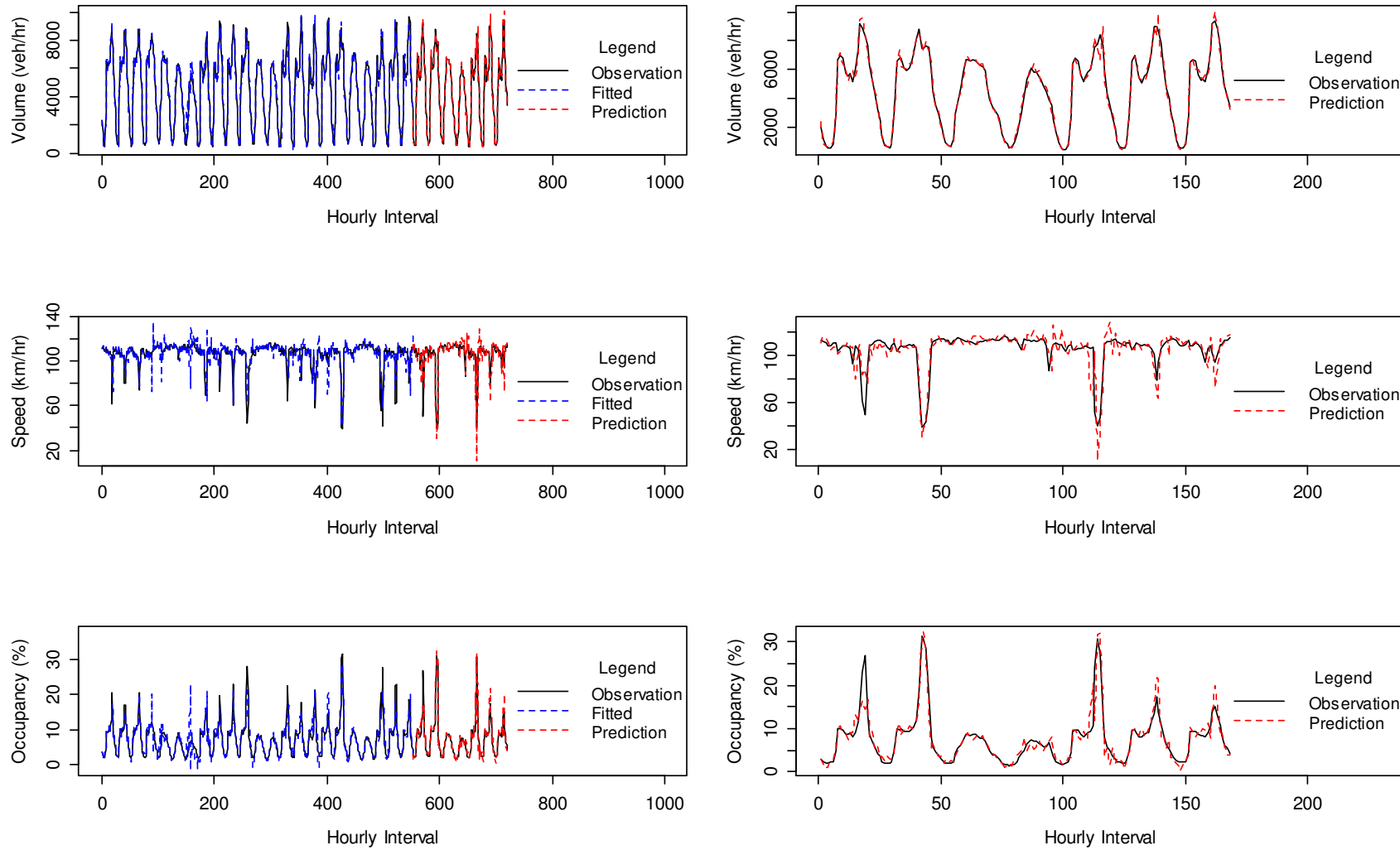


Fig. A. 31 400.NB.2, Finch Ave. W. - Steeles Ave. W., 400DN0080DNS

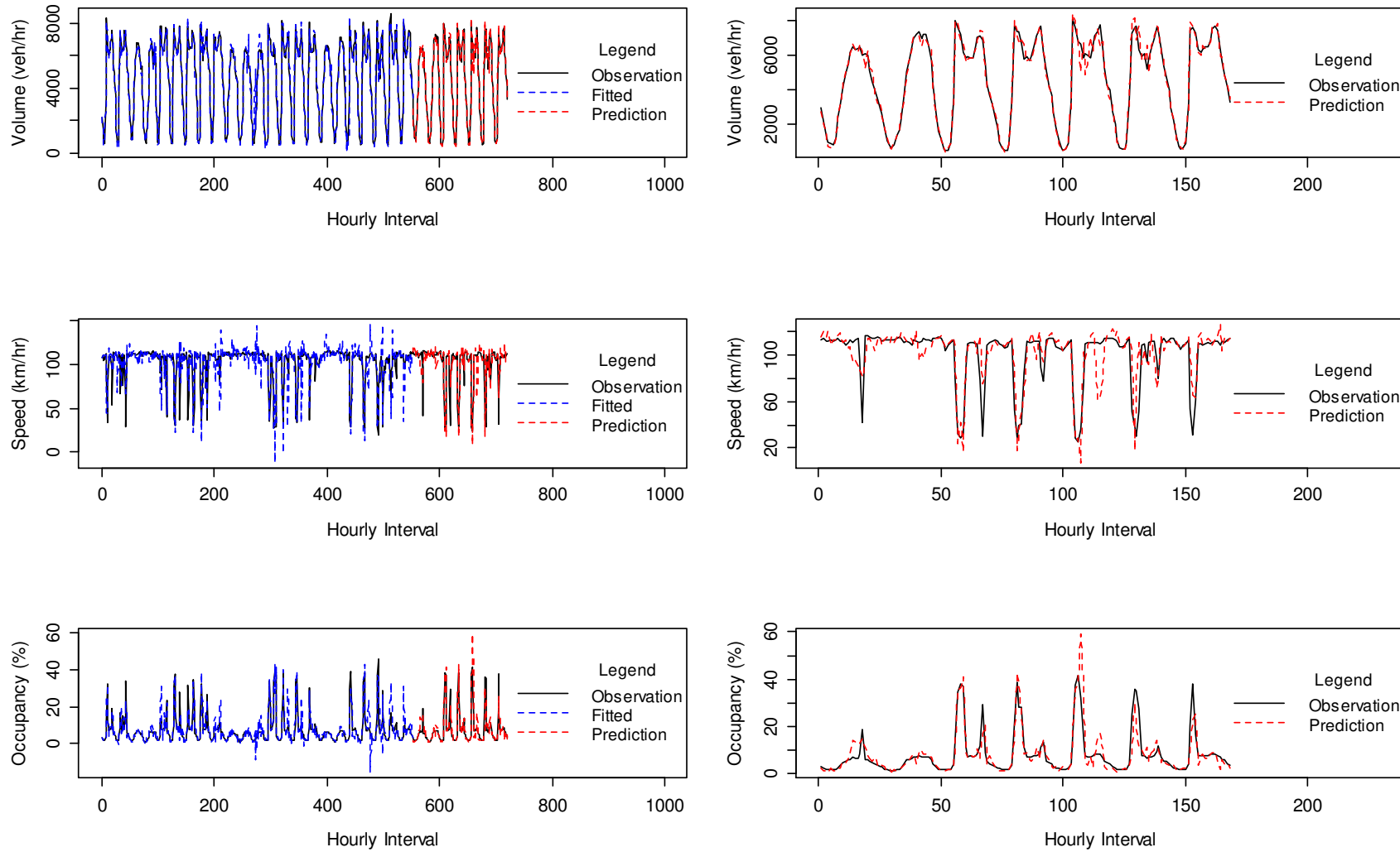


Fig. A. 32 400.SB.1, Hwy401 - Finch Ave. W., 400DN0040DSS

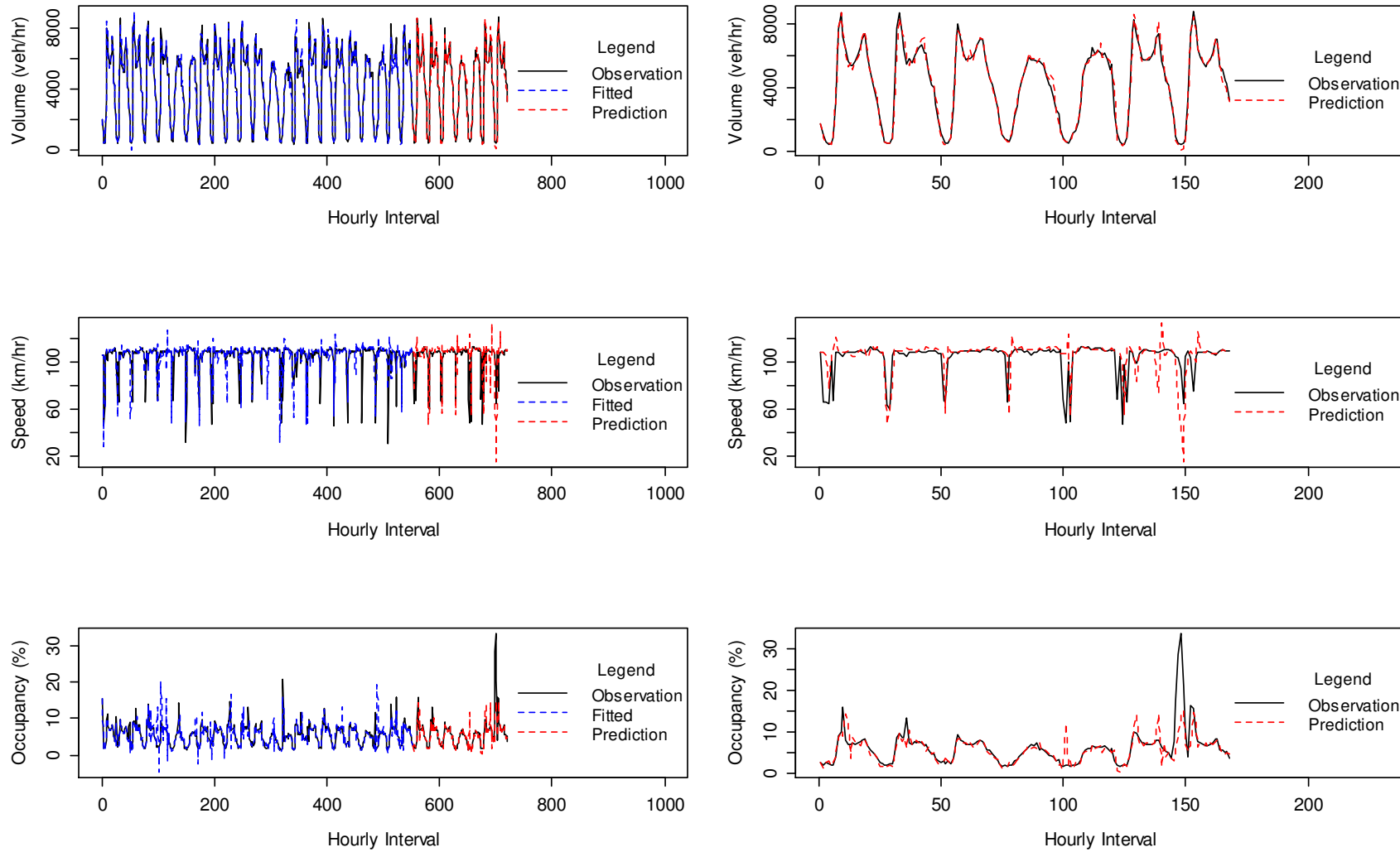


Fig. A. 33 400.SB.2, Finch Ave. W. - Steeles Ave. W., 400DN0080DSS

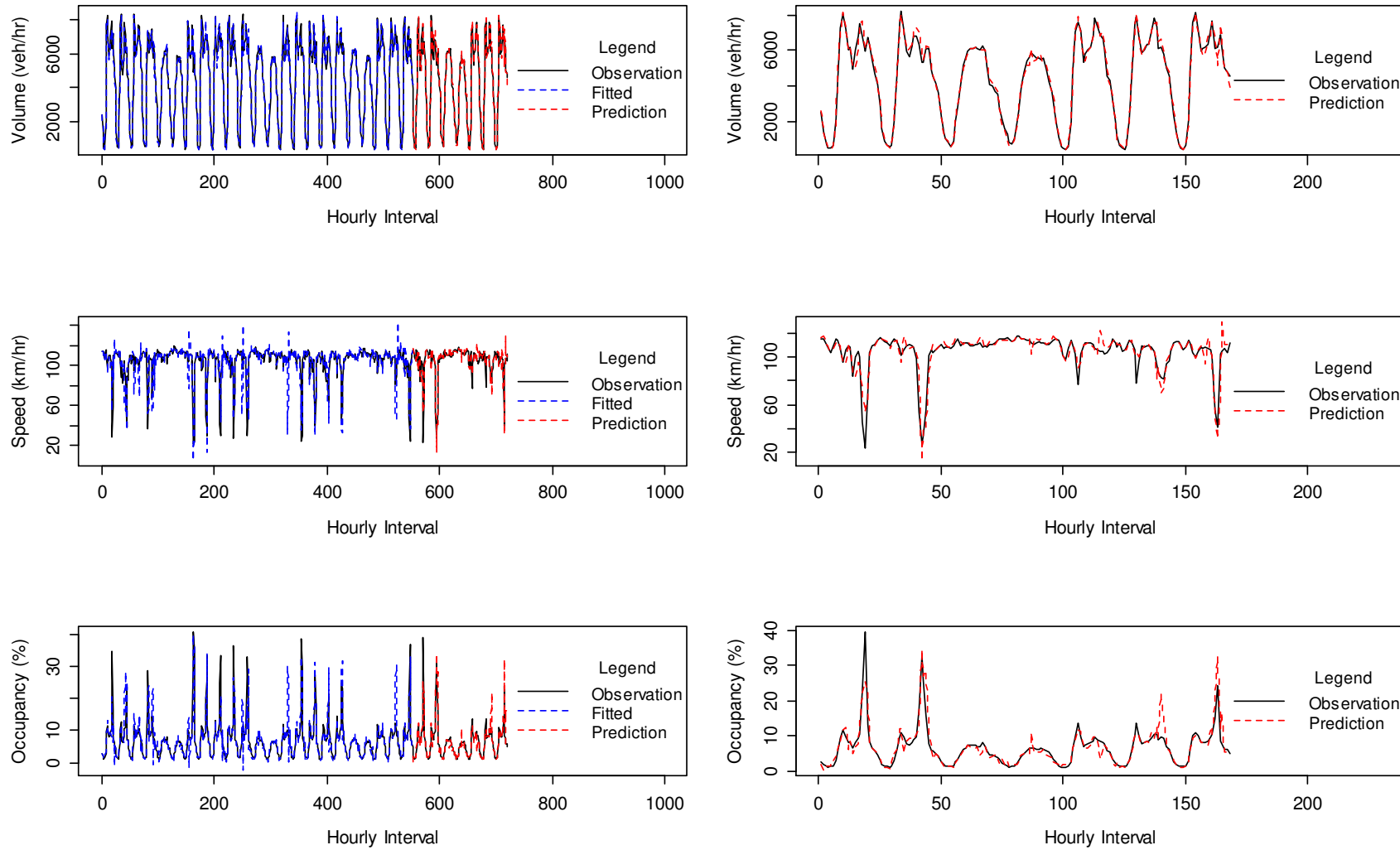


Fig. A. 34 404.NB.1, Sheppard Ave. E. - Finch Ave. E., 404DN0010DNS

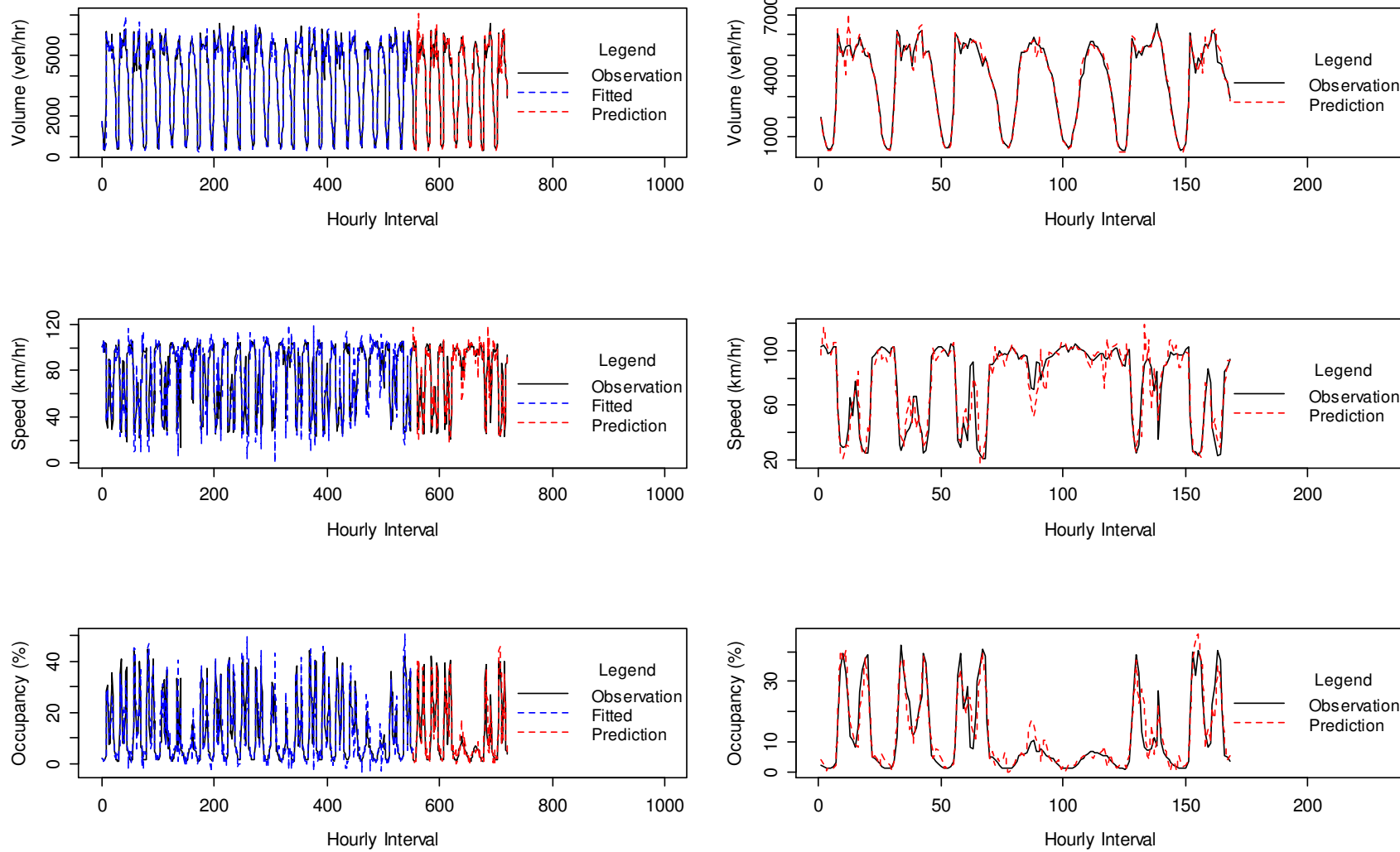


Fig. A. 35 404.SB.1, Sheppard Ave. E. - Finch Ave. E., 404DN0010DSS

Appendix B

Fundamental Diagrams from location 1 and 2

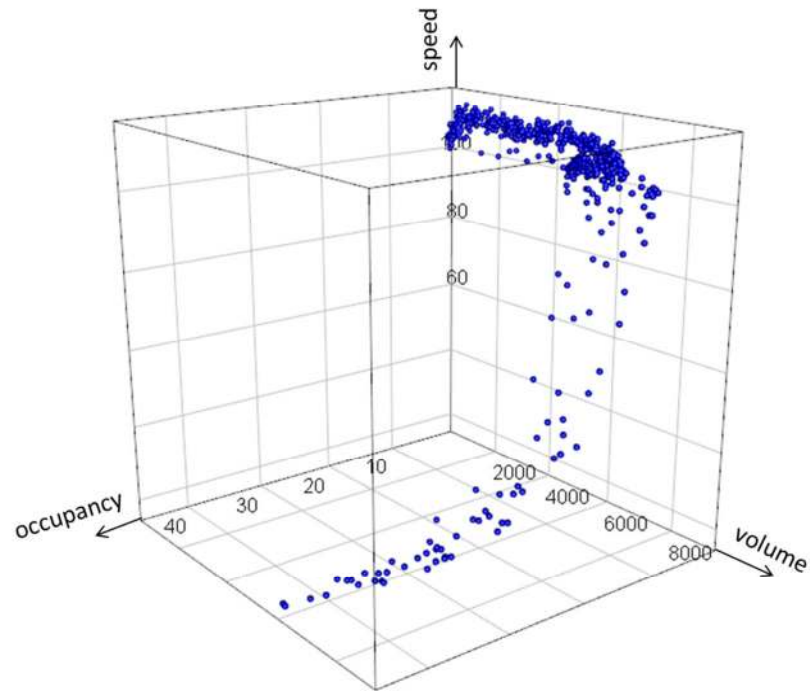


Fig. B. 1 3D scatter plot of traffic volume, speed and occupancy, location 1

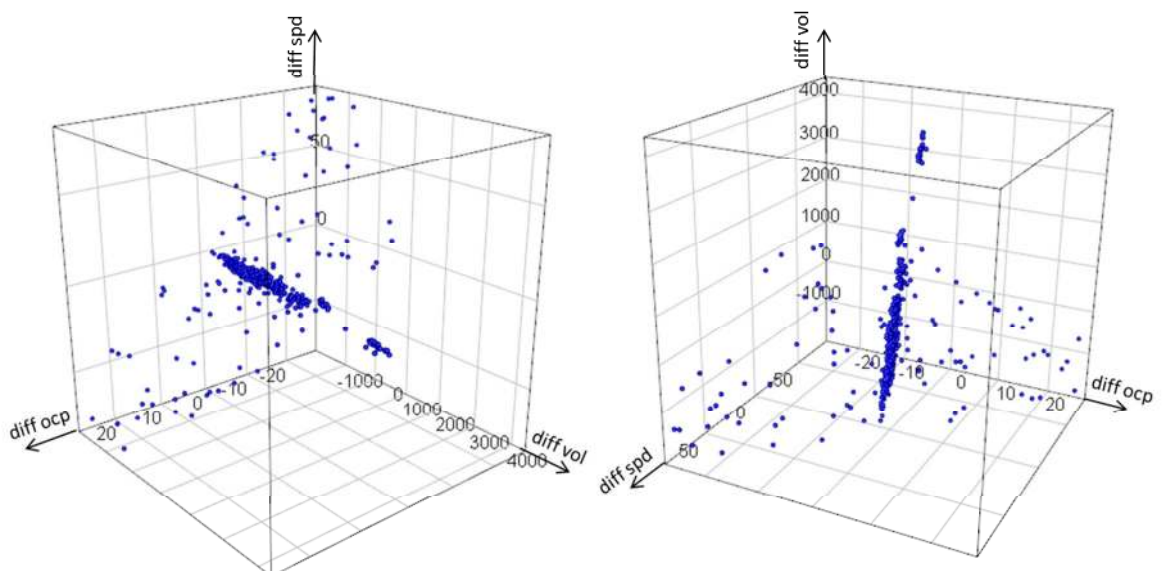


Fig. B. 2 3D scatter plot of differenced volume, speed, and occupancy, location 1

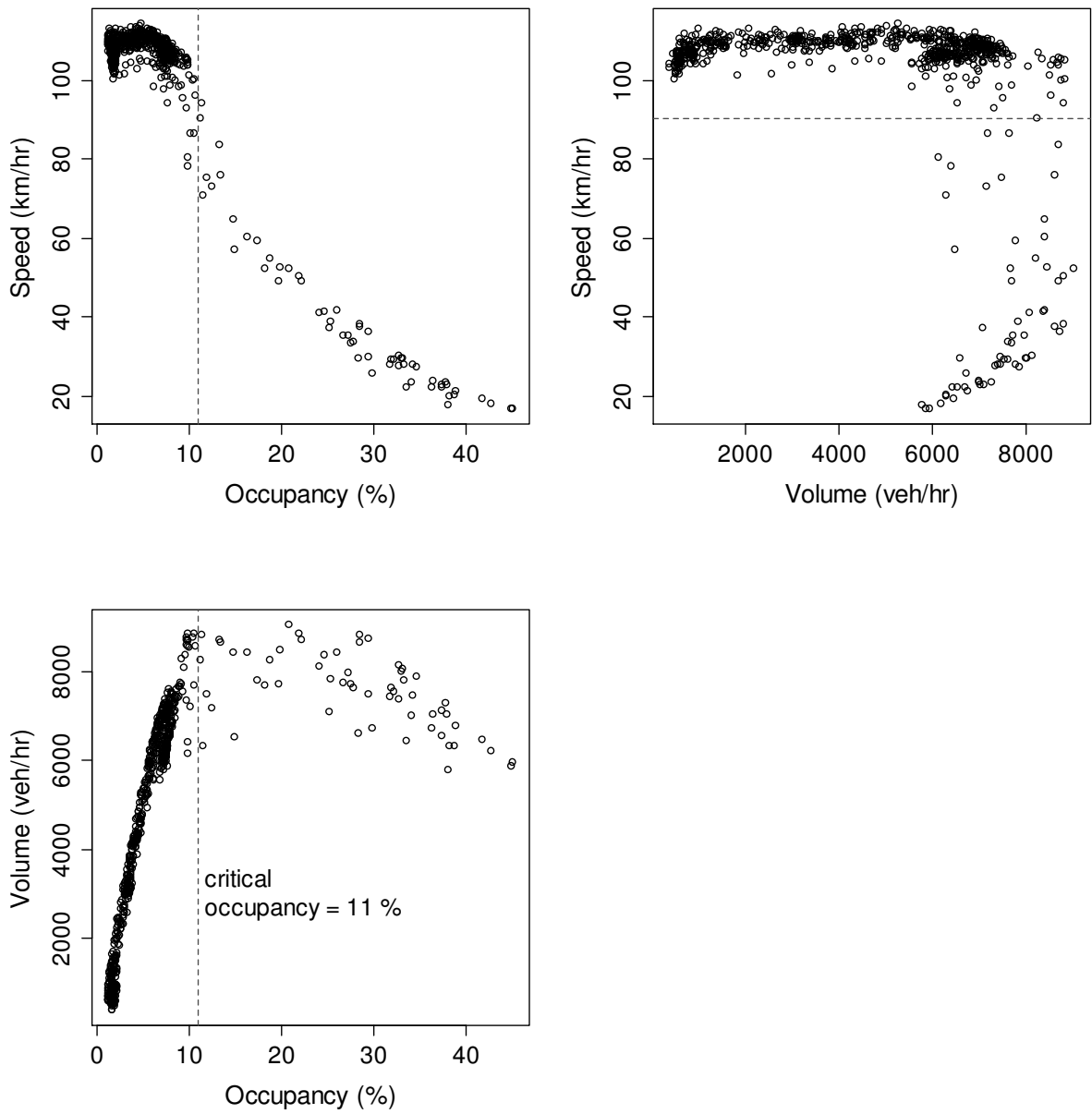


Fig. B. 3 Fundamental Diagrams, traffic data from location 1

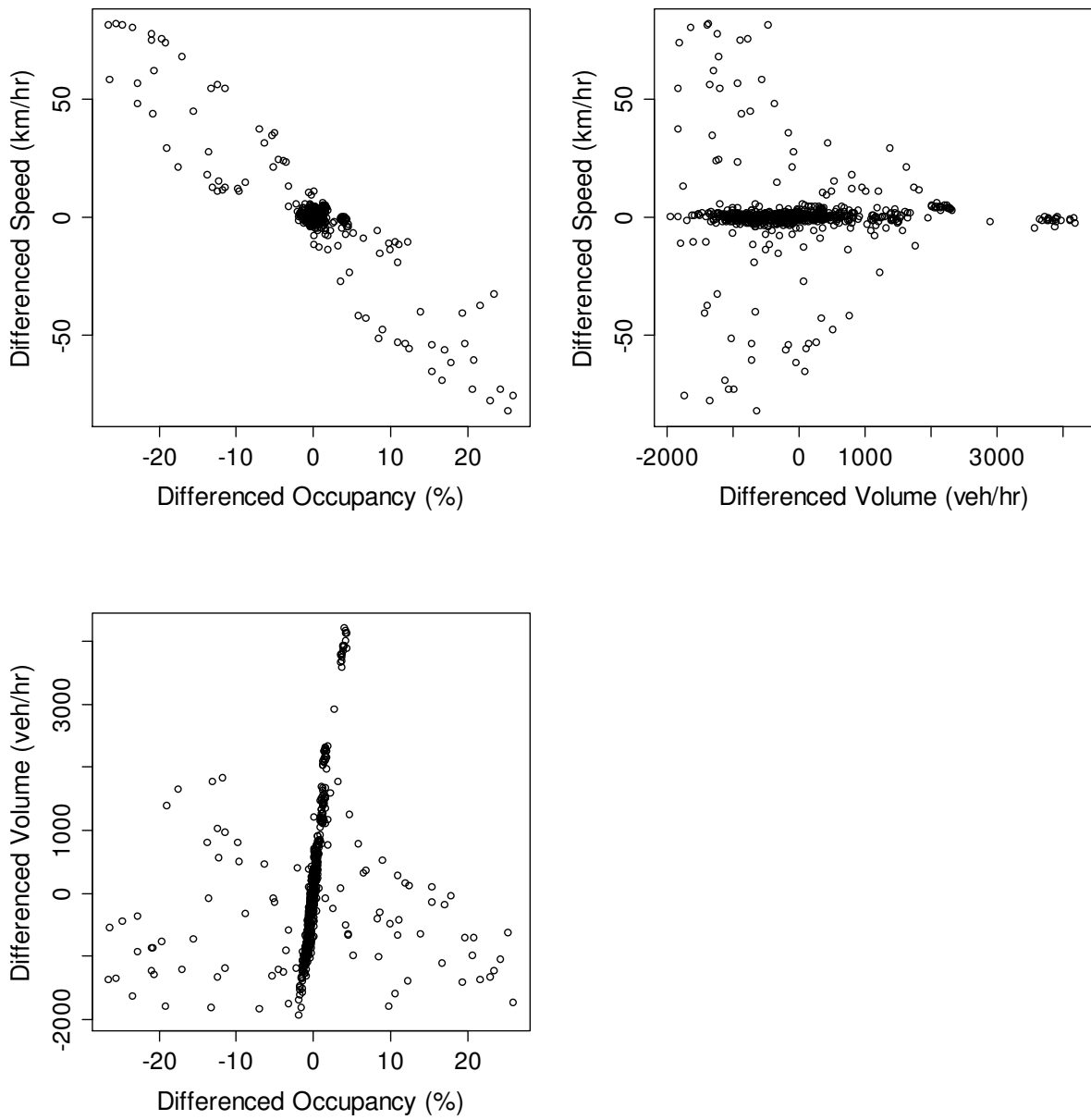


Fig. B. 4 Fundamental relationships between differenced traffic variables, location 1

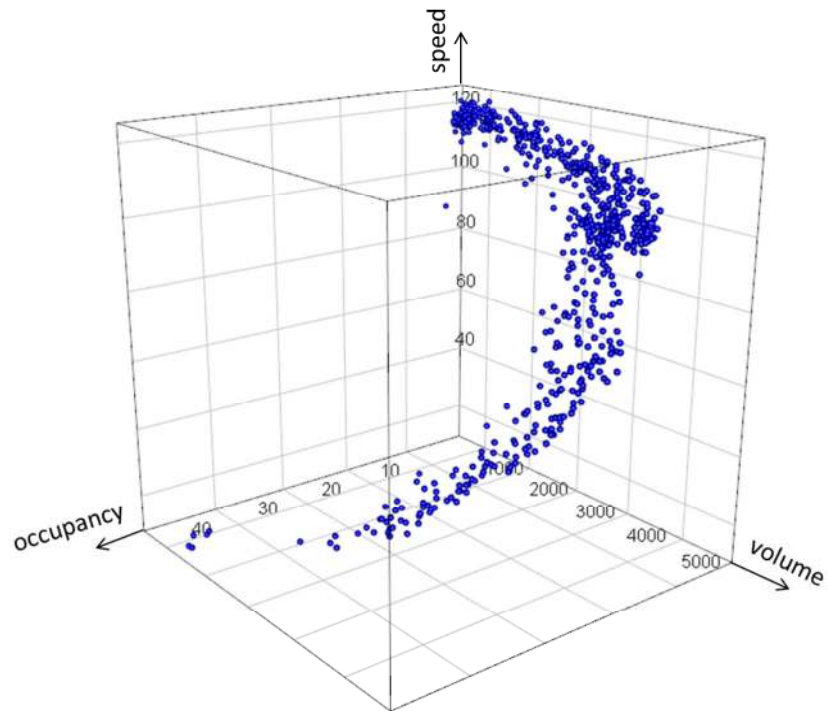


Fig. B. 5 3D scatter plot of traffic volume, speed and occupancy, location 2

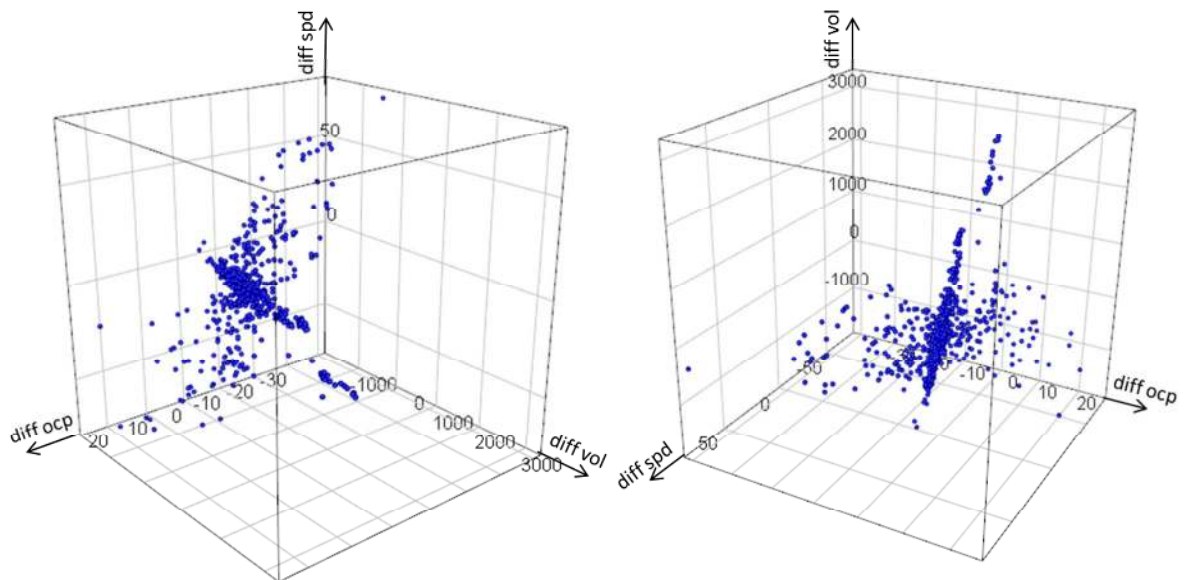


Fig. B. 6 3D scatter plot of differenced volume, speed, and occupancy, location 2

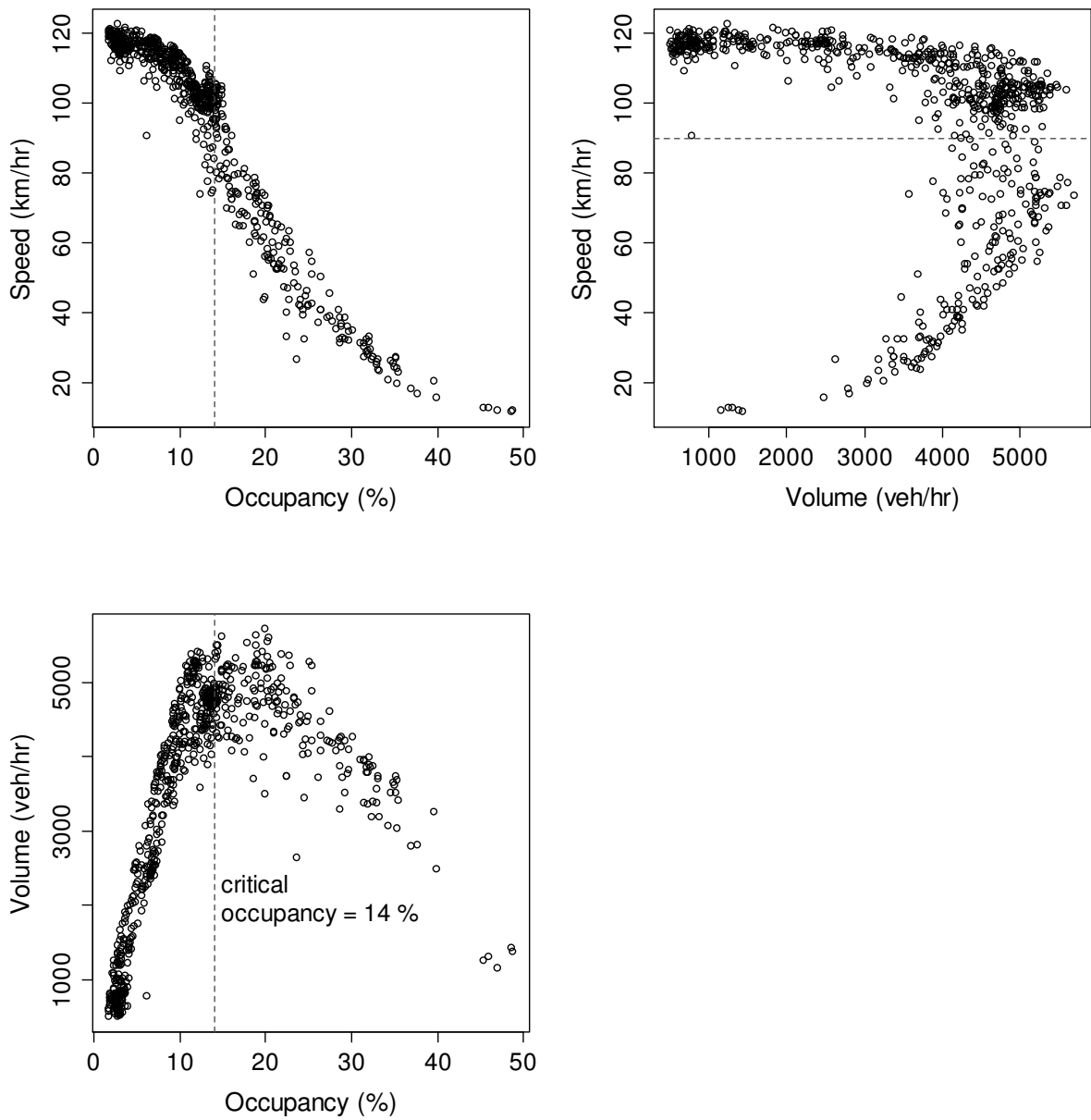


Fig. B. 7 Fundamental Diagrams, traffic data from location 2

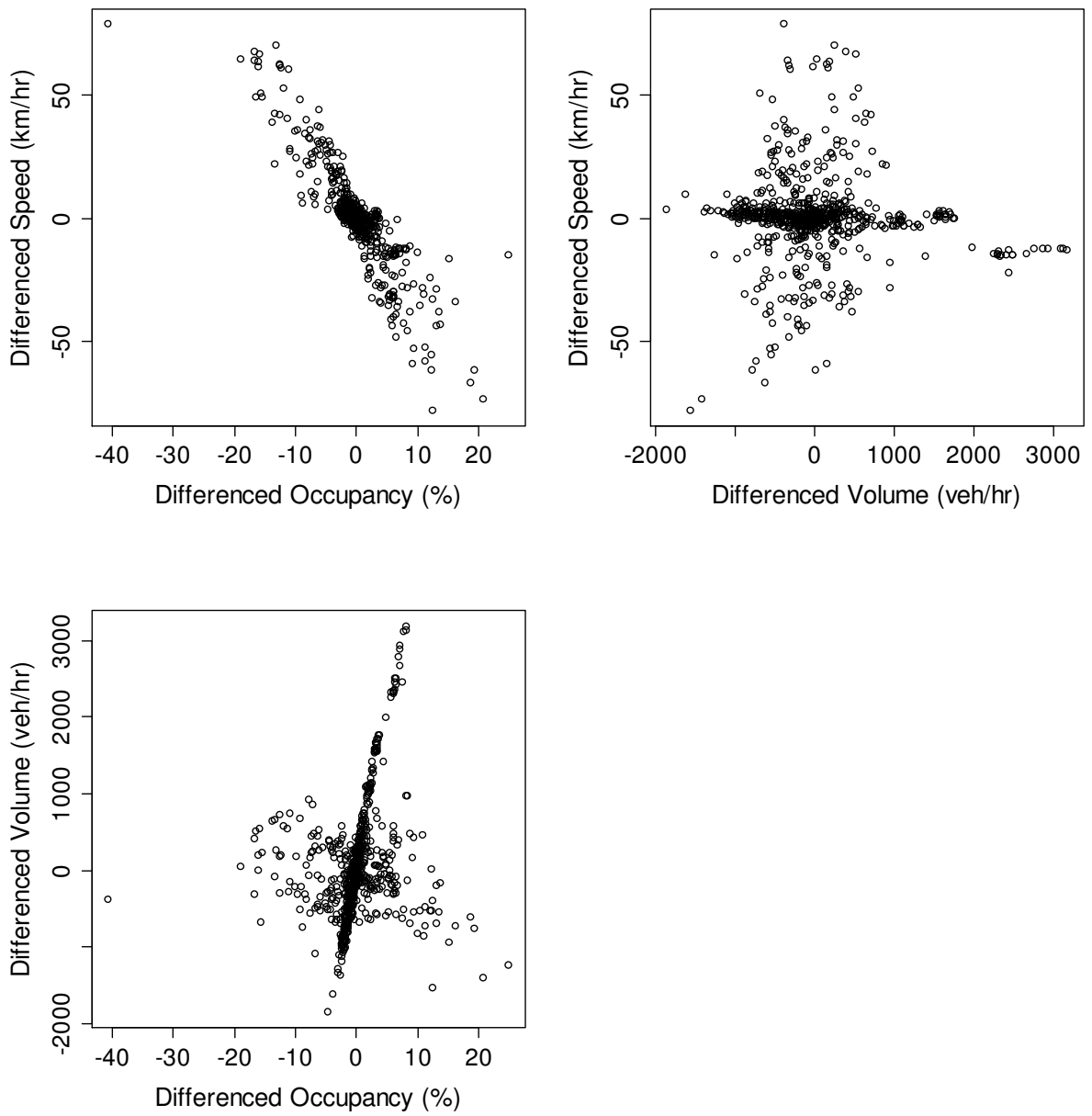


Fig. B. 8 Fundamental relationships between differenced traffic variables, location 2

Copyright Acknowledgements