國　立　交　通　大　學

運輸科技與管理學系

Master Thesis

碩 士 論 文

Short-term traffic flow forecasting for urban roads

市區路段短期交通量預測

研　究　生：謝亞蓁

指導教授：黃家耀　助理教授

中華民國九十九年六月

Short-term traffic flow forecasting for urban roads

市區路段短期交通量預測

研　究　生：謝亞蓁　　　Student：Ya-Chen Hsieh

指導教授：黃家耀　　　Advisor：Ka-Io Wong

國立交通大學

運輸科技與管理學系

碩士論文

A Thesis

Submitted to Department of Transportation Technology and Management

College of Management

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Transportation Technology and Management

June 2010

Hsinchu, Taiwan, Republic of China

中華民國九十九年六月

市區路段短期交通量預測

學生：謝亞蓁　　　　　　　　　　　　　指導教授：黃家耀 助理教授

國立交通大學運輸科技與管理學系碩士班

摘　要

近年來短期交通量預測的應用日益受到重視，許多先進交通旅行者資訊系統(ATIS)及先進交通管理系統(ATMS)的應用都需要估計及預測路網之交通狀況，其目的在於提供有用之旅行資訊給旅行者及提升整體路網的效率。透過道路上各種不同偵測器所蒐集到之交通狀況歷史資料，我們可以掌握道路上之即時資訊並用來估計目前的交通狀況及預測短期內可能發生的交通狀況。目前大部分交通流量預測的文獻都著重在高速公路流量的預測，事實上市區道路的交通流量預測由於還要考慮機車、紅綠燈的影響，且市區路網也較高速公路路網複雜許多，因此有必要深入研究。

本研究建立一個都市地區短期交通量的預測模式，所採用之方法為整合自我迴歸移動平均模式(ARIMA)及時空自我迴歸移動平均模式(STARMA)，其中時空自我迴歸移動平均模式為將時間序列之空間分布關係考慮進模式當中之自我迴歸移動平均模式。我們除了利用台北市 24 個偵測器所蒐集之交通量資料來做實例測試，也測試了可獲得即時資訊不同的情況下，時空自我迴歸移動平均模式預測能力的差異。

研究結果顯示：兩種模式之模式估計誤差與預測誤差都很低且非常接近，顯示兩者都適合用來預估市區路段的交通量。然而，整合自我迴歸移動平均模式每個偵測器最多要校估 5 個參數，而時空自我迴歸移動平均模式卻只需要 6 個參數，因此當路網中的偵測器數量增加時，較簡單之時空自我迴歸移動平均模式較適合用來預估整體路網之交通量。另外，交通量並非一個獨立的系統，而是會受到附近地區交通狀況之影響，因此時空自我迴歸移動平均模式將時間序列之空間關係考慮進模式中，的確會提升模式之預測能力。測試可獲得即時資訊不同的情況下模式之預測能力，結果顯示利用即時資訊來預測較利用歷史資料預測來得準確。

關鍵字：市區交通量、短期預測、整合自我迴歸移動平均模式(ARIMA)、時空自我迴歸移動平均模式(STARMA)

Short-term traffic flow forecasting for urban roads

Student: Ya-Chen Hsieh Advisor: Dr. Ka-Io Wong

Department of Transportation Technology and Management
National Chiao Tung University

Abstract

The interests and applications of short-term traffic forecasting have been growing in the recent years. Many of the applications in Advance Traveler Information System (ATIS) and Advance Traffic Management Systems (ATMS) , which aim at providing useful information to travelers and improving the overall efficiency of road network, require an estimation and forecasting of the traffic conditions of the network. With a historical database of past traffic data from various types of vehicle detectors, real-time traffic information is collected which will be used to estimate the current traffic conditions and predict the condition in near future. Whereas most of the literature focused on the traffic flow prediction on the freeways, modeling traffic flow in urban arterials is more challenging as there are disturbances such as motorcycles and traffic signals in urban area.

In this study, traffic flow forecasting models for urban arterials are proposed. Seasonal autoregressive integrated moving average (ARIMA) and space-time autoregressive moving average (STARMA) model, which incorporates the spatial correlations of the time series, are investigated. A case study using the traffic data from 24 vehicle detectors in Taipei city, Taiwan are performed. The forecasting performance of STARMA model are also examined by static, 1-step ahead rolling and 2-step ahead rolling strategies when real-time information can be obtained.

The findings of this thesis are as follows. The estimated results reveal that both ARIMA and STARMA model are suitable for traffic flows forecasting in urban area. However, in the ARIMA model, there are up to five parameters for each detector, whereas there are only 6 parameters in the STARMA model. With a large number of detector locations in the system to be forecasted, the STARMA model shows a relative simple structure as compared to the ARIMA model which is univariate in nature. Traffic flows of urban area are not an isolated system and will be influenced by the flows from other adjacent locations, consequently, STARMA model considering the spatial relationship between each time series can improve the forecasting accuracy. Finally, the results of forecasting performance tests prove that using real-time information to forecast is better than merely using historical information to forecast.

Keywords：Urban traffic flow, short-term forecasting, ARIMA, STARMA.

Acknowledgement

本篇論文得以順利完成，要感謝每一個曾經幫助我與鼓勵我的人，首先要感謝的是指導教授黃家耀老師，交大六年的求學生涯中，有一半以上的時間都受到老師的照顧與指導，在我徬徨無助或遇到瓶頸時為我指引方向，使我豁然開朗，在撰寫論文期間耐心教導與修正，使我受益良多，雖然是我在撰寫論文，實際上老師付出的更多；學業以外，不論是讓我們去香港、澳門增廣見聞，或是時不時的聚餐、打保齡球，都讓我們的研究生活豐富了許多，在此向老師致上最深的謝意。口試期間，承蒙論文口試委員韓復華教授及黃寬丞副教授撥冗細閱，不吝於給予寶貴意見及指正，使本論文能夠更加完善，學生由衷感謝。

回憶研究所兩年的時光，感謝實驗室的學長姐智偉、總務、伶潔及學弟妹小 P、冠豪，在我遇到困難時給予我幫助與鼓勵、在我忙不過來時替我分憂解勞，尤其感謝同窗兩年的夥伴修安，除了和我一起念書、一起研究，還要負責處理實驗室各種大大小小的問題及添購各種設備，讓我完全不需要操心，能在一個麻雀雖小，但五臟俱全的實驗室裡安心研究。擔任經濟助教的日子裡，感謝永遠年輕的高凱老師及家銘學長，讓我除了學習如何擔任一個好助教，也帶著我們吃吃喝喝及看電影，讓我的研究生活增添了許多樂趣。

感謝正妹佩君、很有原則的阿僑、瞇瞇眼的駿哥、每次當我司機的達達、貼心的怡安、未來的名教授惟茵、可愛的怡文及陪伴我度過無數個挑燈夜戰日子的室友姝婷，除了一起上課、寫作業及打報告，在我心情低落時給予我安慰及堅定的力量，在我心情煩悶時聽我抱怨及發牢騷，在我無聊時陪我玩樂陪我瘋，能夠認識你們這群好朋友，陪伴我一起歡笑與成長，是我生活在新竹苦悶日子裡最幸福的一件事。另外也要感謝其他 Lab 的程式達人丞博與佳儒，在我跑程式時給予我莫大的幫助，筱葳、采蓁及吳 lab 的學弟妹，陪伴我一起度過了快樂的出遊時光等等，與你們分享的點點滴滴，我也永遠不會忘記。

最後要感謝我最愛的父母、儒儒、蓓蓓、芳芳、小阿姨等家人，謝謝你們永遠在我身邊支持、關心及包容我，提供我最安全、最舒適及最溫暖的避風港，讓我能無後顧之憂的完成學業，未來我也會帶著你們的支持與鼓勵繼續前進，希望也能夠成為你們的依靠。最後謹以此篇論文獻給我所有的親朋好友們，謝謝你們大家。

Table of Contents

Chinese Abstract i

English Abstract ii

Acknowledgement iii

Table of Contents iv

List of Tables vi

List of Figures vii

Chapter 1 Introduction 1

1.1 Background 1
1.2 Problem Definition 1
1.3 Objectives 2
1.4 Thesis Organization 3

Chapter 2 Literature Review 5

2.1 The Forecasting Problem 5
2.2 Methodology for Traffic Forecasting 6
2.2.1 Univariate Models 6
2.2.2 Multivariate Models 9

Chapter 3 Model Building 13

3.1 ARIMA Model 13
3.1.1 Model Assumptions 14
3.1.2 Model Identification 17
3.1.3 Model Estimation 21
3.1.4 Diagnostic Checking 22
3.1.5 Model Forecasting 24
3.2 STARMA Model 24
3.2.1 Weight Matrices in STARMA 25
3.2.2 STARMA Modeling Procedure 28
3.3 Concluding Remark 32

Chapter 4 Case Study 33

4.1 Data Collection from Vehicle Detectors 33
4.2 Seasonal ARIMA 34
4.3 STARMA Model 48
4.3.1 Weight Matrices in STARMA 48

4.3.2 Numerical analysis 49
4.4 Forecasting and Comparisons 56

Chapter 5 Strategies of Dynamic Forecasting in ATMS 58

5.1 Forecasting Procedure and Performance Criteria 58
5.1.1 Look back and Look Ahead Procedures 58
5.1.2 Performance Criteria 61
5.2 Comparison of Forecasting Performance 62

Chapter 6 Conclusions and Recommendations 66

6.1 Conclusions 66
6.2 Recommendations for Future Research 67

References 68

Appendix 70

List of Tables

TABLE 3.1 GUIDELINES FOR CHOOSING NONSEASONAL OPERATORS....20
TABLE 3.2 GUIDELINES FOR CHOOSING SEASONAL OPERATORS....21
TABLE 4.1 ADF TEST FOR FIRST SEASONAL DIFFERENCED TRAFFIC DATA FROM VD NO.2 37
TABLE 4.2 PARAMETERS ESTIMATED BEFORE DIAGNOSTIC CHECK FOR VD NO.2 (ARIMA)....39
TABLE 4.3 AUTOCORRELATION CHECK FOR THE RESIDUALS FOR VD NO.2 (1ST MODEL)....39
TABLE 4.4 PARAMETERS ESTIMATED AFTER DIAGNOSTIC CHECK FOR VD NO.2 (ARIMA) 40
TABLE 4.5 AUTOCORRELATION CHECK FOR THE RESIDUALS FOR VD NO.2 (FINAL MODEL)....41
TABLE 4.6 ARIMA PARAMETERS ESTIMATED FOR ALL 24 VEHICLE DETECTORS 42
TABLE 4.7 ESTIMATION AND FORECASTING ERRORS BY RMSE AND MAPE (ARIMA)....44
TABLE 4.8 SPATIAL NEIGHBORING RELATIONSHIP OF THE 24 VEHICLE DETECTORS....49
TABLE 4.9 SPACE-TIME AUTOCORRELATION FUNCTIONS OF THE DIFFERENCED SERIES....51
TABLE 4.10 SPACE-TIME PARTIAL AUTOCORRELATION FUNCTIONS OF THE DIFFERENCED SERIES51
TABLE 4.11 PARAMETERS ESTIMATED BEFORE DIAGNOSTIC CHECK (STARMA) 52
TABLE 4.12 PARAMETERS ESTIMATED AFTER DIAGNOSTIC CHECK (STARMA)....52
TABLE 4.13 SPACE-TIME AUTOCORRELATIONS OF THE RESIDUALS 53
TABLE 4.14 SPACE-TIME PARTIAL AUTOCORRELATIONS OF THE RESIDUALS 53
TABLE 4.15 ESTIMATION AND FORECASTING ERRORS BY RMSE AND MAPE (STARMA) 55
TABLE 4.16 COMPARISON OF FORECASTING ERRORS OF ARIMA AND STARMA MODELS....57
TABLE 5.1 OBSERVATIONS AND PREDICTION EQUATIONS USED IN THE STATIC SCHEME....61
TABLE 5.2 OBSERVATIONS AND PREDICTION EQUATIONS USED IN THE STEP AHEAD ROLLING SCHEME 61
TABLE 5.3 COMPARISON OF STARMA MODEL BY STATIC, 1-STEP AND 2-STEP AHEAD FORECAST 63

List of Figures

Figure 1.1 Flow chart of this thesis 4
Figure 2.1 Forecasting with time series model in temporal and spatial dimensions 9
Figure 3.1 Flow chart of the model building procedure 14
Figure 3.2 Spatial order in one-dimensional and two-dimensional systems 26
Figure 3.3 A typical road network with tree structure (with no cycles) 27
Figure 4.1 Locations of the vehicle detectors 33
Figure 4.2 Time series plot of the traffic flow data of VD No.2 35
Figure 4.3 SAC for the traffic flow data of VD No.2 36
Figure 4.4 SPAC for the traffic flow data of VD No.2 36
Figure 4.5 SAC for the differenced traffic data of VD No.2 38
Figure 4.6 SPAC for the differenced traffic data of VD No.2 38
Figure 4.7 Time series plot of traffic flow data of VD No.4 45
Figure 4.8 Comparison of predictions with observations from VD No.2 46
Figure 4.9 Comparison of predictions with observations from VD No.10 47
Figure 4.10 Time series plots of the traffic data of selected VDs 50
Figure 5.1 Estimation and forecasting data used in rolling scheme 59
Figure 5.2 Comparison of STARMA by static and rolling forecast (VD No.9) 65
Figure 5.3 Comparison of STARMA by static and rolling forecast (VD No.24) 65

Chapter 1 Introduction

1.1 Background

Providing a convenient and comfortable transportation environment without congesting and delay has long been one of the most important transport authorities' objectives. With more and more advanced technologies and information systems applying to transportation field in recent years, making transportation systems becoming more intelligent, this problem can be solved. The term intelligent transportation systems (ITS) means the application of information technologies such as computer hardware and software, sensing, telecommunications and control techniques to improve transportation system operations, including improvements in efficiency, capacity, safety and environmental impacts. With intelligent equipments being installed as a part of transportation infrastructure, traveler information can be collected as well as disseminated, traffic signals and variable message signs can be controlled and toll can be electronically collected, which provide vital supports in traffic management, pavement monitoring and overall system maintenance.

Two of the most important areas of ITS are Advanced Traveler Information Systems (ATIS) and Advanced Traffic Management Systems (ATMS). ATIS aims at providing road users and travelers accurate and useful travel information such as travel time, delay and real time route guidance, which can help road users and travelers make better travel decisions. ATMS controls the traffic in the network in real time not only alleviate traffic congestion but also improve the use of system's capacity and thus improve overall network efficiency. All of these applications require the forecasting of traffic to the near future. This study aims at finding a flow forecasting model which can estimate and predict the traffic flows based on the traffic data measurement from the past.

1.2 Problem Definition

There are basically two distinct categories of predictive traffic conditions: strategic and short-term. The difference between strategic and short-term traffic information is the length of the forecasting horizon. Traditional strategic predictive models usually utilize a large amount of data and predict over a period of month or year which is needed for making decisions on long-term transportation planning. On the other hand, short-term predictive models often forecast traffic conditions within a day or an hour that captures the dynamics of traffic, therefore, short-term forecasting models are more suitable for traffic management and information systems.

On the area of implementation, whereas most of the literature focused on the traffic flow prediction on the freeways, modeling traffic flow in urban arterials is more challenging as there are disturbances such as motorcycles and traffic signals in urban area. Furthermore, the freeway network topology is simple, whereas the networks of urban arterials are more complicated, and the two problems are fundamentally different. An accurate model for urban area should involve the spatial relationships of traffic flow at different adjacent locations. In this study, the traffic flow forecasting for urban arterials is investigated. Seasonal autoregressive integrated moving average (ARIMA) and space-time autoregressive moving average (STARMA) are considered as our algorithms. Traffic flows data collecting from vehicle detectors installed on arterials in the Taipei city, Taiwan, were used to illustrate the modeling procedure in our numerical examples.

1.3 Objectives

This study aims at finding a short-term traffic flow forecasting model for urban roads. While most of the literature focus on traffic flow forecasting in the freeway, our study intend to apply our model to forecast urban traffic flow.

Traffic flows in the urban arterials are more complicated since other disturbances such as traffic signals, motorcycles or pedestrians may exist. Furthermore, as the main objective of freeway is to connect major cities, the freeway networks are usually simple and straightforward. However, urban arterials are built to connect all places in the city, so the networks of urban arterials are usually more complicated. Forecasting traffic flow at a particular location will not be sufficient, so integrating spatial factors in the forecasting model will consider in our following model.

One of the basic advanced traffic management systems (ATMS) is the traffic signal system which intends to improve the safety and capacity used of the transportation system. Once we have the past information describing the traffic conditions such us traffic flow or occupancies within an urban area, we can use it to forecast short-term traffic flow and further connect them to the urban traffic control systems. With the real-time and predicted traffic conditions in the near future, traffic authorities may be able to set a more suitable traffic signal plan that relieves traffic congestion and reduce the travel time and delay of travelers.

Since most of the literatures have demonstrated the superiority of univariate ARIMA model over other forecasting approaches in traffic flow prediction. In attempt to compare the forecasting abilities of the univariate and multivariate models, the univariate ARIMA and its extended form of space-time ARMA model which integrates the spatial-temporal dependencies of each time series are considered to be the forecasting approaches in our study.

The classic seasonal autoregressive integrated moving average (ARIMA), which is widely used in literature for forecasting problems, is tested for its validity in urban traffic models. A space-time autoregressive moving average (STARMA) model, which is new to the traffic forecasting literature, is also presented, and it is considered to be suitable for the urban traffic models with its spatial-temporal dimension. With the calibrated models, we would also demonstrate a dynamic execution of the models for the ATMS purposes, in which several forecasting strategies using real-time database are described.

1.4 Thesis Organization

The study is organized as follow. In this chapter, we have depicted the background of the short-term forecasting problem, problem definition and the objectives of this thesis. Chapter 2 reviews some previous literature concerning short-term traffic forecasting. Univariate predicting models are useful to forecast traffic conditions at a particular location. However, for the purpose of ATMS which requires a macroscopic view of the traffic pattern, multivariate forecasting models that consider observations collecting at more than one locations representing information from a wide area of the network are more appropriate. Chapter 3 presents the model building procedures of the proposed model as well as the model assumptions. Seasonal ARIMA and space-time ARIMA are considered as our algorithms. The identification, estimation and diagnostic checking methods of these models are introduced. Chapter 4 illustrates the use of those models by forecasting traffic volume on arterials; their forecasting performances are also compared and discussed. Then, in attempt to apply the forecasting model to ATMS, we will perform different forecasting strategies using real-time database to exhibit the forecasting performance of the STARMA model in Chapter 5. Finally, the conclusions of this study and the recommendations of future research will be presented in Chapter 6. Figure 1.1 shows the flow chart of this thesis.

Figure 1.1 Flow chart of this thesis

Chapter 2 Literature Review

2.1 The Forecasting Problem

Bowerman and O'Connell (1993) stated that the term forecast implies prediction of future events and conditions, and the act of making such predictions is called forecasting. Forecast must rely on information that has occurred in the past, so a forecaster needs to analyze past data and predict future condition according the result of the analysis. To begin a forecast, the first step is to analyze the data so as to identify a pattern or trend that could be used to represent it. Makridakis and Wheelwright (1978) stated that such patterns are generally assumed to be two different forms, which are univariate and multivariate methods. Generally univariate methods assumed a pattern that is solely determined as a function of time, and utilizing the historical data (in the form of time series) of the variable is sufficient to detect the basic pattern for forecasting. On the other hand, multivariate methods assumed there exist some relationships between two or more variables, and this pattern or trend can be extrapolated or extended into the future. A forecasting technique assumes that the pattern or trend that has been identified will continue in the future, so a forecasting technique will result in a good prediction only if the assumption is valid. If the identified pattern does not remain the same in the future, a forecaster should try to anticipate the changing time of the pattern so that the model would be adjusted to predict the data accurately.

Short-term traffic forecasting necessitates the collection of historical time series data. Time series data are a set of particular variable observations, each one being recorded at a specific time, arranging in sequence of time they occur; in other words, they are ordinal observation set that observed in accordance of time of a dynamic system. A time series is said to be deterministic if future values of the time series can be exactly determined by past values. On the other hand, if the future value of a time series is only partly determined by past values, it is a stochastic or random series. A stochastic time series model consists of trend, seasonal and error terms in the basic model, plus other relevant components, and the mathematical form is,

$$y_t = \mu_t + \gamma_t + \varepsilon_t \qquad t = 1,2,\ldots,T \tag{2.1}$$

where y_t is the observation at time *t*, μ_t is a slowly varying trend component, γ_t is a periodic seasonal component and ε_t is an irregular error or disturbance component.

Such data measuring the traffic stream characteristics such as traffic flow, speed and lane occupancy are recorded by various types of vehicle detector and are widely available on freeway and arterials for the operation and incident detection purposes. With a historical database of past traffic data, collected real-time traffic information can be used to estimate the current traffic conditions and predict the conditions in near future.

2.2 Methodology for Traffic Forecasting

As ITS implementations becoming more and more important throughout the world, attentions have been paid on constructing a short-term traffic forecasting algorithms that could forecast future traffic conditions accurately. Vlahogianni et al. (2004) provided an extensive review on the short-term traffic forecasting problem. It was concluded that although a wide variety of studies were devoted to develop algorithms in the fields of short-term traffic forecasting during the past decades, there was still not a solid framework for modeling traffic forecasting algorithms. The objectives of traffic forecasting in different studies are diverse, such as travel time forecasting, traffic flow forecasts and traffic state forecasting problems exist in the literature, with various methodologies making use of combinations of traffic data like speed, flow, occupancy or vehicle trajectories. The review explored a vast amount of research in the perspective of objectives and methods in this field and classified the modeling process into three essential stages, i.e. the determination of the scope, the conceptual output specification and the process of modeling, which involves several decisive issues such as the choices of appropriate methodology, the type of input and output data and the resolution of the data. Integrating the abovementioned process can result in a systematic logical flow that can be seen as a framework for developing short-term traffic forecasting models. Their logical flow also showed that when constructing short-term traffic forecasting algorithms, in addition to the basic three stages which are concept, realization and evaluation, the accuracy and representational power of a model also serve as important factors in determining the implementation effectiveness.

In the following section, we will review some of the univariate and multivariate time series models used in the literature to predict traffic flow.

2.2.1 Univariate Models

Determining an appropriate methodology is one of the major issues in all forecasting problems. In general, the widely used prediction models can be broadly classified into two techniques: parametric and non-parametric. Parametric techniques assume a form of probability distribution with a finite number of parameters which usually have physical interpretations, and they are most useful when conditions are expected to remain the same. Examples of parametric techniques are linear and non-linear regressions, historical average algorithms, exponential smoothing techniques, autoregressive integrated moving average models (ARIMA) and Kalman filtering. On the other hand, non-parametric techniques do not assume any specific functional form for the dependent and independent variables, as we only have little prior knowledge about the true function form to be estimated. When applying the non-parametric modeling technique, in addition to estimating the parameters of the functions, the primary object of nonparametric models is to estimate the underlying functional forms. The most popular non-parametric techniques include non-parametric regression and neural networks, in which the functional forms are not explicit.

When the physical meaning and relationship of the variables are known, parametric modeling is commonly used. Among the parametric models, the ARIMA model is a common technique used in many areas and has been proved for its advantages over some other forecasting methods. For some previous researches focused on predicting traffic flow with ARIMA models, Williams et al. (1998) first applied seasonal ARIMA models to urban freeway traffic flow prediction problems. They compared several seasonal ARIMA models with Winters exponential smoothing techniques for single-interval forecasting. Using the same data set as Smith (1995), which employed nearest-neighbor, neural networks and historical average models, it is found that the seasonal ARIMA models performs better as compared to the results of Smith for a similar problem.

Ghosh et al. (2005) used three different time series models, which are random walk model, Holt Winters' exponential smoothing technique and seasonal ARIMA model to forecast traffic flow of one junction in Dublin. Comparing the error estimated from these three models they concluded that both exponential smoothing technique and seasonal ARIMA performed much better than random walk model.

Other approaches treat time series data as nonparametric models with the spatial interest in neural networks. Dougherty (1995) gave a comprehensive overview on the topics of transport applications of neural networks. He suggested that the three most common used paradigms were backpropagation, learning vector quantisation and adaptive resonance theory, which were typical of supervised, reinforcement and self-organising learning. These three paradigms could be regarded as a standard tool kit for neural networks applied to transportation problems. It is also emphasized that a particular drawback in previous studies is the lack of standard procedure to analyze and compare the result of different researches, and therefore future research could aim at establishing a more specific approach to issues such as comparison between different methodologies with neural networks structures.

Ledoux (1997) considered a cooperation based neural network traffic flow model that could be treat as part of a real time adaptive urban traffic control system (UTC). In his model, a single local neural network was used to model a single signalized link at the beginning. Then, those local neural networks are connected so as to model the traffic flow over a wide network of junctions. While first and second generation of the UTC system utilized the past historical data and modified average historical pattern to predict traffic flow, to design a more adaptive UTC system, this paper proposed a third generation UTC system, i.e. predicted conditions according to current traffic measurements only. Demonstrated by simulation data, the paper concluded that neural networks could be used to model the traffic conditions in the near future.

Yin et al. (2002) proposed a fuzzy-neural network model (FNM) to predict the traffic flows in urban area. Their model composes two modules: a gate network (GN) and an expert network (EN). Whereas the GN classified the input data into a set of fuzzy clusters by fuzzy approach, the EN functioned as a traditional neural network model (NNM) aiming at finding out the relationship between input and output data. Their research found that FNM not only outperformed the traditional NNM in its predicting power, but also with a smaller computing time requirement. Furthermore, FNM can adaptively adjust the coefficients of the model to enhance its predicting power, which is more suitable for predicting real-time traffic conditions.

In general, seasonal ARIMA and neural networks are two different kinds of techniques, not only because of their parametric/non-parametric natures, but also their inherent data structure and dependence to the past data. Seasonal ARIMA model is a time series function of past information which can be expressed as,

$$y_{lt} = f(y_{lt-1}, y_{lt-2}, \ldots, y_{l1}) \tag{2.2}$$

where y_{lt} is traffic volume data at location *l* and time *t*, it only requires the information in the past to forecast future conditions of the same detector location.

On the other hand, with the neural networks model, the traffic data at location *l* and time *t* can be expressed as,

$$y_{lt} = g(y_{lt-1}, y_{l-1t-1}, \ldots, y_{1t-1}) \tag{2.3}$$

which is a function of data from other locations (e.g. upstream detectors) at time *t*-1. This is a common data structure assumed in studies like Yin et al. (2002), assuming the traffic volume of downstream depends on the volumes on its upstream locations some while ago, capturing the spatial effects. The two approaches are displayed in Figure 2.1, showing that seasonal ARIMA is a temporal model whereas neural network model is a spatial one.

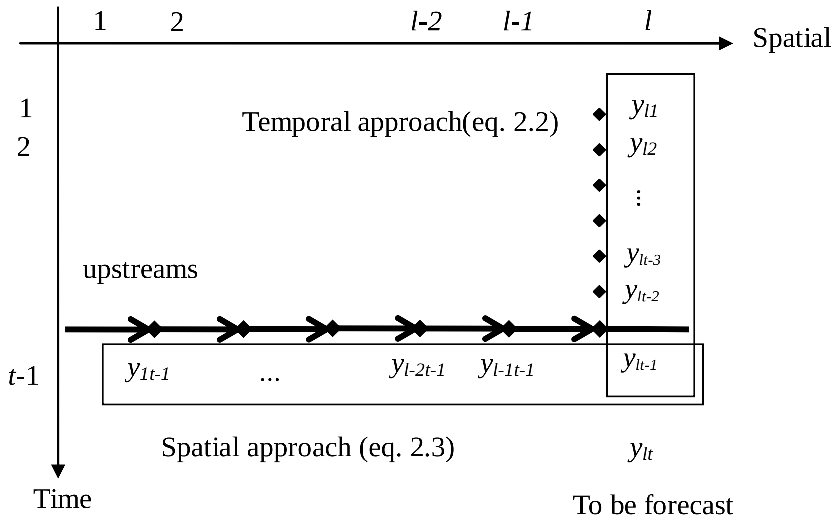


Figure 2.1 Forecasting with time series model in temporal and spatial dimensions

2.2.2 Multivariate Models

Forecasting of traffic volume at a particular location using univariate model is useful for the traffic operation at a local level. However, the deficiency of the univariate model is that they do not consider other factors or relationship (if available) during forecasting, as compared to the multivariate model which considers more than one time series data. For the purpose of ATMS which requires a macroscopic view of the traffic pattern, traffic forecasting for a wide area of the network is necessary, so a multivariate forecasting model is required.

Most of the previous literature concerning traffic forecasting tends to predict traffic data with univariate models since multivariate models are more complicated in the identification of the model and estimation of the parameters. Fortunately, with the great improvement in the computational science in recent years, we are capable of using computers to deal with even more difficult problems within shorter time, so the difficulties of identification and estimation of multivariate model could be solved.

Also, there could be some relationship between traffic data at a particular location and traffic data from its upstream or downstream (in the cases that traffic congestion occurs) detectors, knowing the information from other locations within the network may increase the accuracy of the forecasting. Therefore, with the time series data collected at different locations through the network and the ease of utilizing multivariate models, the possibility of multivariate model being used to predict traffic conditions in near future have increased.

Models that aim at depicting the relationship between time series observations at a particular location and its neighbor locations are referred to as space-time models. A class of space-time models that are characterized by the autoregressive and moving average forms of univariate time series lagged in both space and time are referred to as space-time autoregressive moving average (STARMA) models. STARMA is a special case and restricted form of vector ARMA, and the first complete adaption and modeling process of STARMA was proposed by Pfeifer and Deutsch (1980), who suggested a three-stage iterative procedure for space time modeling. STARMA was an extension of the three-stage iterative model building procedure developed by Box and Jenkins (1976), including identification of the model, estimation of parameters of the tentative model and diagnostic checking (the details will be elaborated in Chapter 3 of the thesis). At the end of their paper, an assault arrest example for the Boston area was employed to demonstrate the modeling procedure.

Pfeifer and Bodily (1990) applied the space-time autoregressive moving average (STARMA) model to fit demand-related data from eight hotels from a single hotel chain in a large U.S. city. Previous researches using STARMA models were all aiming at illustrating the modeling procedure or simply testing the hypotheses about the spatial-temporal structure of the observed data, none of them is used for forecasting purpose, therefore, it is the first paper utilizing STARMA model to forecast economic data and furthermore, comparing the forecasting performance of a single multivariate STARMA model with several univariate ARMA models. Though the after-the-fact analysis of the forecasting test did not conclude that adding the spatial term in the model or simultaneously estimating the parameters for all of the N time series can increase the forecasting accuracy, the analysis suggested that the relative simplicity of the STARMA model explained its superior performance in the test. As the number of observation sites becoming increases, so does the time series needed to be forecast. In this case, the simplicity of the model form will be an advantage in empirical test of forecasting performance.

Kamarianakis and Prastacos (2003) and Kamarianakis and Prastacos (2005) used a space-time autoregressive integrated moving average (STARIMA) model to represent traffic flow speed in urban area. In their model, they assumed that the distances between observed locations are sufficiently long so as to ignore the congestion effects; consequently, downstream locations only depend on upstream locations but not vice versa. Weighting matrices representing the distances between the observation locations were incorporated into the STARIMA model as the spatial characteristics of the space-time modeling process. The modeling procedure was illustrated by utilizing two months traffic flow data with different traffic-flow characteristics from 25 loop detectors in the Athens City, Greece to examine the stability of the estimation of the parameters. In their first paper, in addition to STARIMA, two univariate (historical average and ARIMA) and one multivariate VARMA models were also considered. The results revealed that though the forecasting performance of ARIMA were slightly better than VARMA and STARIMA models, their differences were not significant. The error estimated for historical average technique were more than eighty percent, indicating the inadequacy of historical average technique being used to forecast traffic data in an urban area. In their second paper, only ARIMA and STARMA were considered. They compared the root mean square error of STARIMA model with the average standard errors of ARIMA models and found that these models were quite close; nevertheless, the total number of parameters used for STARIMA model was 7 whereas there were total of 75 parameters in the ARIMA models. They concluded that while univariate techniques can accurately estimate flow speed, as the number of detectors growing larger, using univariate techniques may have some computational problems that it is not capable to forecast traffic data in real world urban network. Therefore, multivariate STARMA was recommended.

Lin et al. (2009) also applied STARMA to forecast short-term urban traffic flow, similar to the one proposed by Kamarianakis and Prastacos (2005). In their study, downstream traffic volume depends only on upstream traffic volume as well as the relative distances. Furthermore, they took seasonal difference to exclude seasonal term in their model. Their study demonstrated the potential of using STARMA model to improve accuracy of urban traffic flow prediction.

There are other techniques incorporating the spatial-temporal dimensions in the traffic forecasting. Yang (2006) studied the spatial-temporal dependencies of traffic flow and developed a spatial-temporal Kalman filter (STKF) forecasting model to compare with ARIMA and neural network (NN). In their study, rather than using a single model, an adaptive forecasting model selection strategy is proposed, which could select spatial-temporal Kalman filter to forecast with real-time data, but switch to use historical average method if real-time data was not available. The results showed that the superiority of spatial-temporal Kalman filter to ARIMA and NN models when real-time information was available for the forecast. Whereas the historical average method outperformed ARIMA and NN when there is no real-time information in the forecast.

In summary, there were a wide variety of univariate as well as multivariate forecasting approaches being used to predict short-term traffic flow. Since most of the literatures have demonstrated the superiority of univariate ARIMA model over other forecasting approaches in traffic flow prediction, we selected the univariate ARIMA model as one of our forecasting technique. Furthermore, it is stated in Yang (2006) that traffic flow is not an isolated phenomenon and should be presented in a multivariate form. One would expect the multivariate model that simultaneously takes all of the possible factors into account when forecasting would give more details of variable relationships and traffic pattern, and have a better forecasting performance. In attempt to compare the forecasting abilities of the univariate and multivariate models, the univariate ARIMA and its extended form of space-time ARMA model, which integrates the spatial-temporal dependencies of each time series, are considered to be the forecasting approaches in our study.

Chapter 3 Model Building

In this chapter, we will discuss the methodology and mathematical formulations of our algorithms, i.e. univariate ARIMA model and multivariate space-time ARMA model (STARMA). ARIMA model is a temporal dimension forecasting model whereas STARMA model extends ARIMA to a spatial-temporal dimension model.

3.1 ARIMA Model

Autoregressive integrated moving average (ARIMA) model, also known as Box-Jenkins method, is an iterative modeling procedure proposed by Box and Jenkins (1976). Since the class ARIMA models consist of simple autoregressive process (AR), simple moving average process (MA) or autoregressive integrated moving average process, the first step of the modeling procedure is to identify a tentative model from postulate general class of models. After determining an appropriate model to be applied, the next step of the procedure is using historical data to estimate the parameters of the model. Calculating the residuals of the fitted model for checking adequacy of the model is the stage called diagnostic checking. If the fitted model is adequate, then it can be used to forecast in the last stage. Otherwise, the process iterates and returns the step to the first stage, i.e. determining another tentative model to be applied and then go through the modeling process again. Figure 3.1 shows the flow chart of the model building procedure.

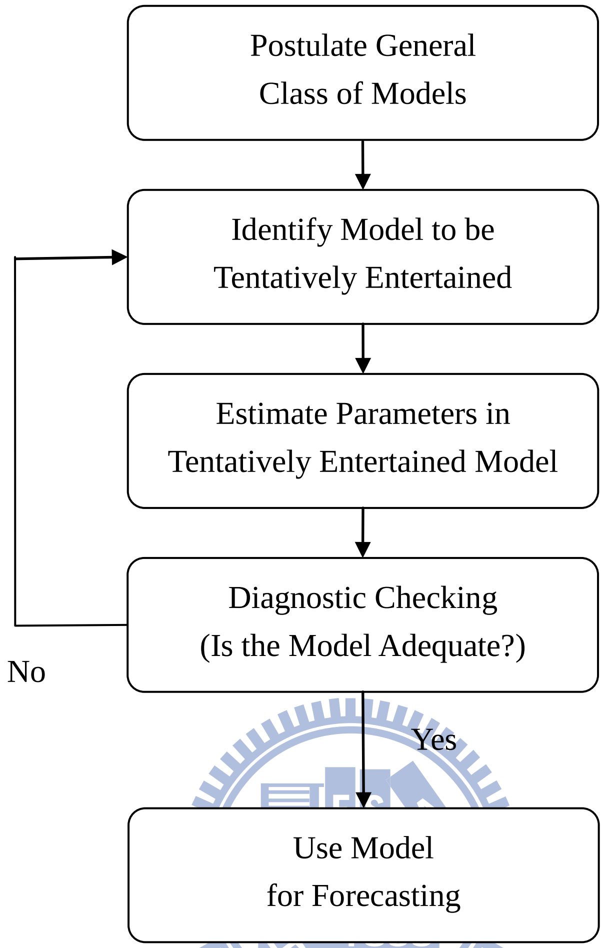


Figure 3.1 Flow chart of the model building procedure

3.1.1 Model Assumptions

The ARIMA modeling procedure requires the time series data process the some properties. Before applying the ARIMA modeling procedure, we should note that the data has to be checked if they satisfy the stationary and invertible assumptions. Here we give a brief introduction for those who are not familiar with the stationary and invertibility condition, and further details may be found in some well-known books such as Box and Jenkins (1976) and Bowerman and O'Connell (1993).

Classical ARIMA models depict stationary time series. A time series is said to be stationary if the joint distribution of any set of n observations $y_1, y_2, \ldots, y_n$ is the same as the joint distribution of $y_{1+k}, y_{2+k}, \ldots, y_{n+k}$ for all n and k. In other words, the statistical properties such as mean and variance of a stationary time series are practically constant through the time. On the other hand, if the mean of a time series is a function of time such as a linear or quadratic function of *t*, it is said to be a nonstationary time series. If the time series used to forecast is nonstationary, it may be converted into a stationary one with transformation techniques. The process of transformation is as follow.

Assuming a nonstationary time series y_t, where $t = 1,2,...,T$ and the first-differenced transformed values of the time series z_t can be defined as,

$$z_t = y_t - y_{t-1} \quad \text{for} \quad t = 2,...,T \tag{3.1}$$

where $y_1, y_2, \ldots, y_t$ are the original nonstationary time series values and $z_1, z_2, \ldots, z_t$ are the first differences time series values.

Taking first differences may be a good way to transform nonstationary time series in to stationary forms; however, sometimes the first differences of the raw data are still nonstationary. In these cases, the second differences of the time series y_t, where $t = 1,2,...,T$ are,

$$z_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2} \quad \text{for} \quad t = 3,...,T \tag{3.2}$$

Previous experiments have revealed that most of the nonstationary and nonseasonal time series can be converted into stationary time series by first difference or second difference. And for those who are nonstationary and seasonal, seasonal differenced transformation ($z_t = y_t - y_{t-L}$, where L denotes the seasonality) would be more accurate.

In this study we applied augmented Dickey-Fuller (ADF) test to examine whether the time series is stationary, Dickey-Fuller test is proposed by Dickey and Fuller (1979) to examine whether a time series has a unit root. When a time series has a unit root, the series is not stationary, and thus testing whether a time series has a unit root is similar to testing whether a time series is stationary. Consider the simple first order autoregressive (AR1) model,

$$y_t = \alpha y_{t-1} + \varepsilon_t \tag{3.3}$$

where ϕ is the coefficient and ε_t is the white noise series error term.

A series is said to be white noise if it satisfies the zero mean and constant variance σ^2 conditions. A series is said to have a unit root if the sum of the autoregressive parameters is equal to 1. As a result, testing whether the sum of the autoregressive parameters is 1 or not can help determine whether a time series has unit root, and hence whether it is stationary or not. The null and alternative hypotheses are,

$H_0 : \alpha = 1$(unit root exist, y_t is nonstationary)

$H_1 : |\alpha| < 1$(y_t is stationary)

The t statistics is,

$$t_{\alpha=1} = \frac{\hat{\alpha}-1}{SE(\hat{\alpha})} \tag{3.4}$$

where $\hat{\alpha}$ is the least squares estimate of α from equation (3.3), $SE(\hat{\alpha})$ is the standard error of $\hat{\alpha}$. If the DF test returns a significant probability value, we should reject the null hypothesis that the time series has a unit root, and hence, the time series is stationary.

Said and Dickey (1984) augmented the basic autoregressive unit root test to accommodate general ARMA(p, q) models with unknown orders when ε_t is not white noise, and proposed a test referring to as the augmented Dickey-Fuller (ADF) test. To discuss augmented Dickey-Fuller test, consider the $(p+1)^{th}$ order autoregressive time series,

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_{p+1} y_{t-p-1} + \varepsilon_t \tag{3.5}$$

ADF test is different form DF test in that ADF includes the term ∇y_{t-p} to allow for ARMA error process, and there are three variation types of model, i.e. no-intercept, nonzero mean and time trend.

1. The no-intercept model is parameterized as,

$$\nabla y_t = y_t - y_{t-1} = \delta y_{t-1} + \theta_1 \nabla y_{t-1} + \ldots + \theta_p \nabla y_{t-p} + \varepsilon_t \tag{3.6}$$

2. The model with nonzero mean term α_0 is,

$$\nabla y_t = \alpha_0 + \delta y_{t-1} + \theta_1 \nabla y_{t-1} + \ldots + \theta_p \nabla y_{t-p} + \varepsilon_t \tag{3.7}$$

3. The model with nonzero mean and a time trend term t can be expressed as,

$$\nabla y_t = \alpha_0 + \gamma t + \delta y_{t-1} + \theta_1 \nabla y_{t-1} + \ldots + \theta_p \nabla y_{t-p} + \varepsilon_t \tag{3.8}$$

where $\delta = \alpha_1 + \ldots + \alpha_{p+1} - 1$ and $\theta_k = -\alpha_{k+1} - \ldots - \alpha_{p+1}$

The null hypothesis of the DF test is that $\delta = 0$ which represents a unit root exists in equations (3.6), (3.7) and (3.8), respectively. These ADF tests statistics have been shown to have asymptotically the same distribution as the Dickey-Fuller test statistics, and thus if the ADF test returns a significant probability value, we should reject the null hypothesis that the time series has a unit root, and hence, the time series is stationary.

Another assumption of ARIMA model is the invertible assumption. Considering a linear process that an observation z_t at time t can be expressed as a weighted sum of observations in the past,

$$z_t = \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_k z_{t-k} + \varepsilon_t \tag{3.9}$$

where δ is a constant mean that time series values fluctuate, ε_t is the white noise series error term, and $\phi_1, \phi_2, \phi_3, \ldots, \phi_k$ are the coefficients of the white noise series of order *k*.

When z_t is expressed as a function of past observations, a linear process is said to be not invertible if the coefficients relating to the past observations do not decline as we move further into the past. On the contrary, an invertible linear process suggested that these coefficients do decline, which indicate that the recent observations have more influence on the value than the distant observations. While we can intuitively expect that future traffic conditions affect more by recent traffic conditions than past traffic conditions; therefore, we could say that the time series of traffic observations satisfied the invertible condition.

3.1.2 Model Identification

Having tentatively determined the difference order to transform a nonstationary time series to a stationary time series, we are then able to examine the behavior of the sample autocorrelation function (SAC) and the sample partial autocorrelation function (SPAC) for the stationary time series to identify the ARMA models.

Considering stationary time series $z_b, z_{b+1}, \ldots, z_n$, the sample autocorrelation function (SAC) at lag *k*, denote by r_k, which measures the linear relationship between time series observations separated by *k*-lag time units, can be expressed as

$$r_k = \frac{\sum_{t=b}^{n-k} (z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=b}^{n} (z_t - \bar{z})^2} \tag{3.10}$$

where $$\bar{z} = \frac{\sum_{t=b}^{n} z_t}{(n-b+1)} \tag{3.11}$$

The value of r_k will always lie between -1 and 1, when the observations separated by *k* time units tend to move together in a linear fashion with a positive slope, i.e. when both of $(z_t - \bar{z})$ and $(z_{t+k} - \bar{z})$ are positives or both of them are negatives, then their product will be positive. On the contrary, when the observations separated by *k* time units tend to move together in a linear fashion with a negative slope, i.e. when $(z_t - \bar{z})$ is positive, $(z_{t+k} - \bar{z})$ is negative or vice versa, then the value of their product will be negative. The more the value of r_k closer to 1 or -1, the more the observations tend to move together with the linear fashion.

The behavior of SAC is said to be cut off at *k* when a spike exists at lag *k* which means that the sample autocorrelation at lag *k* is statistically large. And the behavior of SAC dies down when the function, instead of cutting off, decreases in a steady fashion such as a damped exponential (with or without oscillation), damped sine-wave fashion, or a fashion combined both damped exponential and sine-wave.

Similarly, the sample partial autocorrelation (SPAC) at lag *k* r_{kk} is

$$r_{kk} = \begin{cases} r_1 & \text{if } k = 1 \\ \dfrac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} & \text{if } k = 2,3,\ldots \end{cases} \tag{3.12}$$

where $r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j}$ for $j = 1,2,\ldots,k-1$.

Just like SAC, the behavior of SPAC is said to be cut off after lag *k* if no spikes exist at lag greater than *k*. And SPAC dies down whenever the function decreases in a steady fashion such as a damped exponential (with or without oscillation), damped sine-wave fashion, or a fashion combined both damped exponential and sine-wave.

If the SAC and SPAC of the time series data either cut off or die down fairly quickly, then the time series should be considered stationary. On the contrary, if the SAC and SPAC of the time series data die down quite quickly or extremely slowly, then the time series should be considered nonstationary. Later we will present some guidelines for determining ARIMA process based on the behavior of SAC and SPAC.

The ARIMA model composed of three essential elements: AR stands for autoregressive part; I stands for differencing, a technique that transfers time series data to a stationary form; and MA stands for moving average part. A seasonal ARIMA model is the model which accounts for the seasonal variations, such as months or seasons in a year, in which the pattern of the time series repeats itself in each cycle by the seasonal effect. The construction of the model is described as follows.

A time series with autoregressive process of order p can be expressed as follow,

$$z_t = \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + \varepsilon_t \tag{3.13}$$

where δ is a constant mean that time series values fluctuate, $\phi_1, \phi_2, \phi_3, \ldots, \phi_p$ are the coefficients of the autoregressive process of order p, and ε_t is the white noise error components in the time series.

A moving average model of order q is defined by the equation,

$$z_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \tag{3.14}$$

where δ is a constant mean that time series values fluctuate, $\theta_1, \theta_2, \theta_3, \ldots, \theta_q$ are the coefficients of the moving average process of order q, and ε_t is the white noise error components in the time series.

Combining the above three elements and making use of the backshift operator B^k, which shifts the subscript of a time series observation or error term backward in time by k periods, we then consider the mixed autoregressive moving average process of order *(p, d, q)*, which is,

$$\phi_p(B)(1-B)^d y_t = \delta + \theta_q(B)\varepsilon_t \tag{3.15}$$

where $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)$ is the nonseasonal autoregressive operator of order p, $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)$ is the nonseasonal moving average operator of order q and B is the backshift operator.

If the time series has seasonal variation, a simple ARIMA model no longer fits, and a seasonal ARIMA is required. A general multiplicative seasonal model of order $(p, d, q)(P, D, Q)_L$ which is analogous to the mixed autoregressive moving average process of order *(p, d, q)* can be expressed as follow,

$$\phi_p(B)\phi_P(B^L)(1-B)^d(1-B^L)^D y_t = \delta + \theta_q(B)\theta_Q(B^L)\varepsilon_t \tag{3.16}$$

where ϕ_P, θ_Q, *P, D, Q* are the seasonal counterparts of ϕ_p, θ_q, *p, d, q* respectively and L denotes the seasonality.

The purpose of identifying a seasonal ARIMA model is to determine the values of p, d, q and P, D, Q by examining the functions of SAC and SPAC, while the cycle length L is usually given explicitly. Table 3.1 presents the guidelines for determining nonseasonal operators and Table 3.2 for seasonal operators.

Table 3.1 Guidelines for choosing nonseasonal operators

Guideline	Behavior of SAC and SPAC	Nonseasonal Operators
1	SAC has spikes at lags 1,2,…,q and cuts off after lag q, and SPAC dies down	Nonseasonal moving average of order q $\theta_q(B) = (1-\theta_1 B - \ldots - \theta_q B^q)$
2	SAC dies down, and SPAC has spikes at lags 1,2,…,p and cuts off after lag p	Nonseasonal autoregressive of order p $\phi_p(B) = (1-\phi_1 B - \ldots - \phi_p B^p)$
3	SAC has spikes at lags 1,2,…,q and cuts off after lag q, and SPAC has spikes at lags 1,2,…,p and cuts off after lag p	$\theta_q(B)$ or $\phi_p(B)$ If SAC cuts off more abruptly than SPAC, use $\theta_q(B)$. If SPAC cuts off more abruptly than SAC, use $\phi_p(B)$. If both SAC and SPAC cut off equally abruptly, use $\theta_q(B)$ and $\phi_p(B)$, then choose the one that yields the best model.
4	Both SAC and SPAC have no spikes at all lags	No nonseasonal operator
5	Both SAC and SPAC die down	Both $\theta_q(B)$ and $\phi_p(B)$

Reference：Adapted from Bowerman and O'Connell (1993)

Table 3.2 Guidelines for choosing seasonal operators

Guideline	Seasonal Behavior of SAC and SPAC	Seasonal Operators
6	SAC has spikes at lags *L*, 2*L*, …, *QL* and cuts off after lag *QL*, and SPAC dies down	Seasonal moving average of order *Q* $\theta_Q(B^L)=(1-\theta_{1,L}B^L-\ldots-\theta_{Q,L}B^{QL})$
7	SAC dies down, and SPAC has spikes at lags *L*, 2*L*, …, *PL* and cuts off after lag *PL*	Seasonal autoregressive of order *P* $\phi_P(B^L)=(1-\phi_{1,L}B^L-\ldots-\phi_{P,L}B^{PL})$
8	SAC has spikes at lags *L*, 2*L*, …, *QL* and cuts off after lag *QL*, and SPAC has spikes at lags *L*, 2*L*, …, *PL* and cuts off after lag *PL*	$\theta_Q(B^L)$ or $\phi_P(B^L)$ If SAC cuts off more abruptly at seasonal level than SPAC, use $\theta_Q(B^L)$. If SPAC cuts off more abruptly at seasonal level than SAC, use $\phi_P(B^L)$. If both SAC and SPAC cut off equally abruptly at seasonal level, use $\theta_Q(B^L)$ and $\phi_P(B^L)$, then choose the one that yields the best model.
9	Both SAC and SPAC have no spikes at all seasonal lags	No seasonal operator
10	Both SAC and SPAC die down fairly quickly at seasonal level	Both $\theta_Q(B^L)$ and $\phi_P(B^L)$

Reference：Adapted from Bowerman and O'Connell (1993)

Once we the values of *p, d, q* and *P, D, Q*, a tentative model have been determined and the modeling procedure can led to next stage. In our study, we utilized the SAS statistic software (SAS, 2008) to compute the SAC and SPAC so as to identify these values.

3.1.3 Model Estimation

After selecting a tentative model form through the identification process, the next stage is using the collected historical data to estimate the parameters of the model. Estimating an adequate model that contains the characteristics of the data is necessary without question. As Fisher (1956) stated that, it is necessary that efficient use of the data should have been made in the fitting process. Without doing this, inadequacy caused by inappropriate fitting rather than inadequate model form may arise.

The most widely used method for estimating parameters is the maximum likelihood (ML) method. Box and Jenkins (1976) also pointed out that many examples have revealed that when the sample size is moderate and large enough, the log-likelihood function will be unimodal and can be approximated by a quadratic function over a sufficiently extensive region near the maximum, and under such situations, the log-likelihood function can be described by its maximum. Maximum likelihood estimate is a procedure which its estimated values of parameters maximizing the log-likelihood function.

As the procedure of maximum likelihood estimation is complicated and difficult to implement, most of the statistics software provide an alternative to calculate by the least squares approach. It is proved that when the random errors are normally distributed, the maximum likelihood estimates can be approximate by least squares point estimates. The least squares approach begins with preliminary point estimates of the parameters and then modify them according to the mean squared error (MSE). At each iteration, the changes of the parameters are in the direction of minimizing the MSE, and this iterative searching technique will continue until the parameters corresponding to the minimum mean square error are found. At the end of the estimation process, the final least squares point estimates of the parameters are obtained.

3.1.4 Diagnostic Checking

Evaluating the adequacy of the fitted model is the purpose of diagnostic checking. If a fitted model is found to be inadequate at this stage, the modeling procedure should return to the identification stage and go through the process again, and therefore, ARIMA model is an iterative modeling procedure. Two kinds of observations in a model can be determined as inadequate. The first one is unable to depict the observed correlation of the process, and there exists significant correlation in the residuals over time. The second one may be the model is too complex, and here we introduce the concept of parsimony proposed by Box and Jenkins (1976) to define this kind of adequate fitted model. A model is said to be adequate if it satisfies the principle of parsimony, which means that to fit the model that can describe the pattern of historical data with minimum number of parameters. In other words, if any estimated parameter in the fitted model is proved to be statistically insignificant, it is not a good fitted model.

The first step of the diagnostic checking stage is to analyze the residuals from the fitted model. The observed data can be adequately represented if the residuals of the fitted model is white noise, that is, they are approximately normal distribution with mean zero and constant variance. By calculating the sample autocorrelation function of the residuals and sample partial autocorrelation function of the residuals, we can examine whether the residuals is white noise. If the residuals are white noise, the sample autocorrelation functions should be zero. Otherwise, they may present a pattern that should be included in the ARIMA model, identifying this pattern and integrating into the fitted model will usually get a better fitted model.

In this study, we applied the SAS software to model the ARIMA procedures. When testing whether the residuals is unrelated or contain additional information that should be included into the model, SAS software tests the null hypotheses that the set of autocorrelations is white noise using the Ljung-Box statistic calculated by,

$$\chi_m^2 = n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k} \tag{3.17}$$

where $r_k = \frac{\sum_{t=1}^{n-k}\varepsilon_t\varepsilon_{t+k}}{\sum_{t=1}^{n}\varepsilon_t^2}$ is the sample autocorrelation of residuals separated by a lag of k lags, ε_t is the residual series, n is number of observations.

If the residuals are unrelated, the autocorrelations of the residuals should be small, and hence χ_m^2 should be small. The smaller the χ_m^2 is, the smaller are the autocorrelations of the residuals and the more unrelated are the residuals. Consequently, a large χ_m^2 indicates that the model is inadequate. Setting the probability of a Type I error equal to α and apply it to Ljung-Box statistic we can reject the adequacy of the model if the p-value is less than α, where p-value is the area under the curve of the chi-square distribution having $k-n_p$ degrees of freedom to the right of χ_m^2, and n_p is the number of parameters that must be estimated in the model. The value of α indicates the confidence level of the test statistics. If we set α to be 0.05, the p-value less than 0.05 represents that the χ_m^2 of the tentative selected model has very little chance to be smaller than $\chi_{m[\alpha]}^2$, that is, $\chi_m^2 > \chi_{m[\alpha]}^2$, and the model is strongly inadequate. On the other hand, if the p-value is greater than 0.05 represents that χ_m^2 is less than $\chi_{m[\alpha]}^2$, we can say that the tentative model is suitable for the data.

The second step of the diagnostic checking stage is to test the statistical significance of the estimated parameters, and the testing procedure is as follow. Let θ be any particular parameter in the model, $\hat{\theta}$ be the point estimate of θ, and $s_{\hat{\theta}}$ be the standard of the point estimate $\hat{\theta}$. Then the t-value is defined as,

$$t = \frac{\hat{\theta}}{s_{\hat{\theta}}} \tag{3.18}$$

When the absolute value of t is large, then $\hat{\theta}$ is large, which indicates that parameter θ is not equal to zero and should be included in the ARIMA model. Experiences have found that it is reasonable to include the parameter in the model if the absolute t-value of that parameter is greater than 2, in other words, if any parameter whose t statistic is smaller than 2, it should be removed from the model.

3.1.5 Model Forecasting

If an appropriate model has been identified and its parameters have been fitted without violating the significant test, then with the historical traffic data collected in the past, it can be used to estimate current traffic conditions as well as forecast traffic conditions in the future.

3.2 STARMA Model

The space-time autoregressive moving average (STARMA) model can be viewed as a special case of the general vector ARMA (p, q) model. As Pfeifer and Deutsch (1980) stated that a process considering $N \times N$ autoregressive and moving average parameter matrices to represent the autocorrelations and cross-correlations of the N time series is referred to as the vector ARMA model. The general vector ARMA models will collapse to a STARMA models if the N time series appear to be a single random process operating at different sites and the dependencies between the N series is a function of their relative positions which can be captured by a weight matrix $W^{(l)}$; that is, when the diagonal elements in those $N \times N$ matrices are assumed to be equal and the off-diagonal elements are assumed to be a linear combination of the weight matrices $W^{(l)}$.

In addition, STARMA model is an extension of the ARIMA model to incorporate the spatial correlations of the time series data, incorporating spatial factors in ARIMA model is specifically useful in modeling traffic conditions since traffic measurements collected at a particular location may be influenced by upstream as well as downstream traffic conditions. We can expect that if an urban arterial is congested, its neighboring arterials will also be congested. Therefore, it is reasonable to speculate that extending ARIMA to STARMA would improve the prediction performance of the STARMA model.

Similar to the ARIMA model, the STARMA model is also an iterative modeling procedure, the first stage is to identify a tentative STARMA class of model to be employed, and the space-time sample autocorrelation and space-time sample partial autocorrelation functions of the data are examined. As the STARMA considers the relationships of the time series at the detectors locations, we have to assign a weight matrix representing the network structure and capturing the spatial characteristics of the data collecting points in advance to incorporate those dependencies into the model. The weight matrix used in STARMA will be described in the following section.

3.2.1 Weight Matrices in STARMA

Introducing the spatial lag operator can assist the understanding of STARMA. Assume that $z_i(t), i = 1,2,...,N$ are the N time series observations at N locations in space, let $L^{(l)}$ be the spatial operator of spatial order l, such that,

$$L^{(0)} z_i(t) = z_i(t) \tag{3.19}$$

$$L^{(l)} z_i(t) = \sum_{j=1}^{N} w_{ij}^{(l)} z_j(t) \tag{3.20}$$

where $w_{ij}^{(l)}$ are a set of weights used to represent the configuration of the N locations, $w_{ij}^{(l)}$ is set to nonzero if locations i and j are l^{th} order neighbors for all i.

Formulating the above equations in vector representation we get,

$$L^{(0)} z(t) = W^{(0)} z(t) = \mathrm{I}_N z(t) \tag{3.21}$$

$$L^{(l)} z(t) = W^{(l)} z(t) \quad \text{for } l > 0 \tag{3.22}$$

where I_N is an $N \times N$ identity matrix, $W^{(l)}$ is an $N \times N$ square matrix with weight $w_{ij}^{(l)}$ and each row summing to one,

$$\sum_{j=1}^{N} w_{ij}^{(l)} = 1 \quad \text{for all } i \tag{3.23}$$

The specific setting rule of these weights for different spatial order l can be determined by the model builder to indicate the relationship between their locations. For example, the physical properties of the system such as the length of the common boundary between contiguous countries i and j, or the distance between the centers of countries, can be considered when setting weights $w_{ij}^{(l)}$. Furthermore, as Pfeifer and Deutsch (1980) stated that, these weights must reflect a hierarchical ordering of spatial neighbors, i.e. first order neighbors are those nearest to the location of interest; second order neighbors are farther away than first order neighbors, but closer than third order neighbors. An example of the first four spatial order neighbors of a particular location for both one-dimensional line of locations and two-dimensional grid system are shown in Figure 3.2.

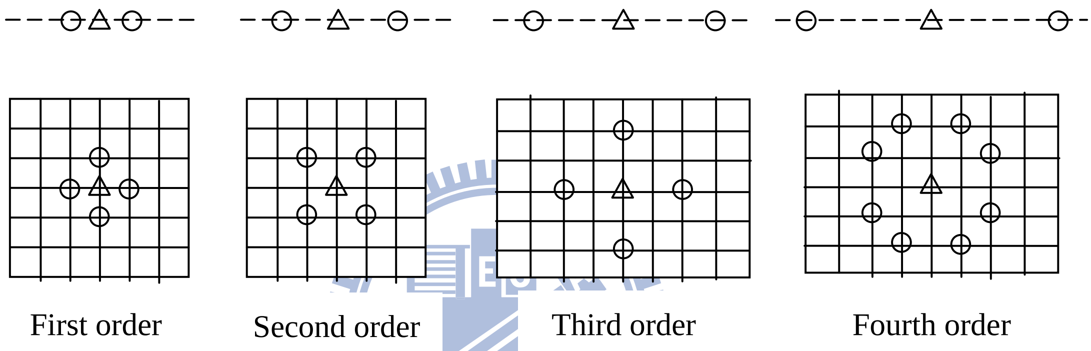


Figure 3.2 Spatial order in one-dimensional and two-dimensional systems
(Source：Pfeifer and Deutsch, 1980)

By the definition of the spatial operators, we can define a $N \times N$ square matrix $W^{(l)}$ with elements $w_{ij}^{(l)}$ which is nonzero if the locations of vehicle detectors i and j are l^{th} order neighbors for all i. $W^{(l)}$ is referred to as the l^{th} order weight matrix which satisfies a hierarchical order of spatial relationship, whereas first order neighbors are those nearest to the measuring locations, second order neighbors are those farther than first orders but closer than third order ones. The zero order neighbor of each location is itself, and therefore $W^{(0)}$ is an $N \times N$ identity matrix.

As the main purpose of the weight matrix is to incorporate the physical characteristics of the traffic flow network into the STARMA model, we should first analyze it. Here we used the definition the same as Kamarianakis and Prastacos (2005). As shown in Figure 3.3, typically, traffic flow network can be expressed as a tree structure where the directions of the tree vectors indicate the directions of the traffic flow and the nodes of the network represent the measuring locations where vehicle detectors located. Note that in the literature the tree network (and its sub-network) is suggested to process the many-to-one property, meaning that the links from upstream are merging to the downstream but no splitting up. However, this is rarely the case in a real transportation network where roads can be parallel or in grid forms; therefore, we used less restricted rules in setting our weight matrices. If we consider the distance between the locations of vehicle detectors is long enough so that the end of queue do not spill back to the upstream. As a result, only downstream locations will be influenced by flow at the upstream locations, and the reverse effect is much weaker. Therefore, the first spatial order includes its direct upstream, whereas the second spatial order includes its further upstream as well as the links in parallel, where merges to the same downstream.

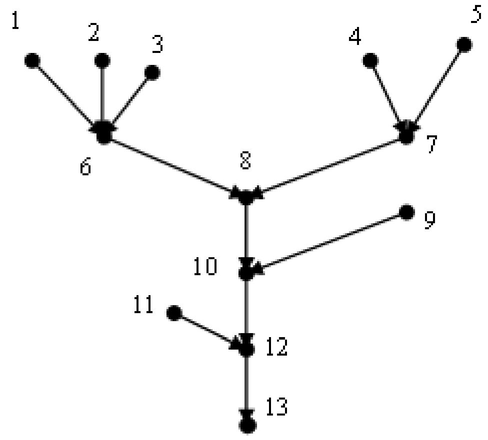


Figure 3.3 A typical road network with tree structure (with no cycles)
(Source：Kamarianakis and Prastacos, 2005)

Considering the characteristics of traffic flow network and assuming that the l^{th} order neighbors of each location i have equal weights and $\sum_{j=1}^{N} w_{ij}^{(l)} = 1$ for all i, the network topology of Figure 3.3 can be expressed as the following weight matrices for first and second spatial order shown in equation (3.24) and (3.25). Noted that these rules of weight matrix settings are user defined, as long as the row sum is equal to 1 (if not zero). In our study, we assumed that the contributions from all upstream are identical, without considering the ratios if the traffic volumes are turning from the left or right at the intersections. This is because it is difficult to differentiate the turning proportions from the vehicle detectors. However, one may assume the weighting factors to be proportional to average flow volume of the upstream. Other rules such as setting weights according to different turning movement at intersections or the effects due to weather (e.g. rainy day) may also be a possible alternative.

$$W^{(1)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.33 & 0.33 & 0.33 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{3.24}$$

$$W^{(2)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 \end{bmatrix} \tag{3.25}$$

3.2.2 STARMA Modeling Procedure

Since we have defined the meaning of spatial operator, we are now able to introduce the model classes of STARMA. Similar to the time series $z_i(t)$ in univariate ARIMA model, $z_i(t)$ in STARMA model depends on the past observations and prediction errors at the same location *i*, as well as the past observations of its neighboring locations at different spatial order. Therefore, $z_i(t)$ can be expressed as a linear combination of past observations and prediction errors at location *i* as well as its neighbor locations,

$$z_i(t)=\sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}L^{(l)}z_i(t-k)-\sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}L^{(l)}\varepsilon_i(t-k)+\varepsilon_i(t) \tag{3.26}$$

where p is the autoregressive order, q is the moving average order, λ_k is the spatial order of the k^{th} autoregressive term, m_k is the spatial order of k^{th} moving average term, ϕ_{kl} and θ_{kl} are parameters to be estimated, and $\varepsilon_i(t)$ are the random normal errors with

$$E[\varepsilon_i(t)]=0 \tag{3.27}$$

$$E[\varepsilon_i(t)\varepsilon_j(t+s)]=\begin{cases}\sigma^2 & i=j, s=0\\ 0 & \text{otherwise}\end{cases} \tag{3.28}$$

Any model that can be formulated as equation (3.29) is called a STARMA ($p_{\lambda_1,\lambda_2,\ldots,\lambda_p}, q_{m_1,m_2,\ldots,m_q}$) model, and the matrix representation of the model is,

$$z(t)=\sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}W^{(l)}z(t-k)-\sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}W^{(l)}\varepsilon(t-k)+\varepsilon(t) \tag{3.29}$$

with $\varepsilon(t)$ normal with mean zero and,

$$E[\varepsilon(t)\varepsilon(t+s)']=\begin{cases}\sigma^2\mathrm{I}_N & s=0\\ 0 & \text{otherwise}\end{cases} \tag{3.30}$$

Two of the special cases of the STARMA model classes are STAR model and STMA model. If the moving average terms q equal to zero and the model remains autoregressive term in the STARMA model, it collapse to a space-time autoregressive (STAR) model, and STAR model with order $p_{\lambda_1,\lambda_2,\ldots,\lambda_p}$ is referred to as a STAR($p_{\lambda_1,\lambda_2,\ldots,\lambda_p}$) model,

$$z(t)=\sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}W^{(l)}z(t-k)+\varepsilon(t) \tag{3.31}$$

On the other hand, a model with only moving average terms of order $q_{m_1,m_2,\ldots,m_q}$ is referred to as a STMA ($q_{m_1,m_2,\ldots,m_q}$) model,

$$z(t)=\varepsilon(t)-\sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}W^{(l)}\varepsilon(t-k) \tag{3.32}$$

Examining the space-time sample autocorrelation functions and space-time sample partial autocorrelation functions of the time series will help determine which of the STARMA, STAR, and STMA model classes are the time series belong to.

With the definition of the spatial operator, the space-time covariance between l^{th} and k^{th} order neighbors at time lag *s*, $\gamma_{lk}(s)$, which is the average covariance between the weighted l^{th} order neighbors of any location and the weighted k^{th} order neighbors of the same location at *s* time lags in the future can be defined as,

$$\gamma_{lk}(s) = E\left\{\sum_{i=1}^{N}\frac{L^{(l)}z_i(t)L^{(k)}z_i(t+s)}{N}\right\} \tag{3.33}$$

And the vector form of the space-time covariance function is,

$$\gamma_{lk}(s) = E\left\{\sum_{i=1}^{N}\frac{[\mathrm{W}^{(l)}z(t)]'[\mathrm{W}^{(k)}z(t+s)]}{N}\right\} \tag{3.34}$$

Once we have the space-time covariance function, we are interested in the space-time autocorrelation functions. As Martin and Oeppen (1975) noted that due to several possible scalings that may be used, the definition of space-time autocorrelation functions is not as simple as the one in the univariate domain. Since we hope that the sample autocorrelations to have constant variance at all spatial lags, an appropriate definition for the space-time autocorrelation between l^{th} and k^{th} order neighbors at time lag *s* is,

$$\rho_{lk}(s) = \frac{\gamma_{lk}(s)}{[\gamma_{ll}(0)\gamma_{kk}(0)]^{1/2}} \tag{3.35}$$

In particular, $\rho_{lk}(s)$ will not equals to $\rho_{kl}(s)$, but it will be the same as $\rho_{kl}(-s)$, i.e. $\rho_{lk}(s) = \rho_{kl}(-s)$, and the sample estimate of the space-time autocorrelation coefficients can be expressed as,

$$\begin{aligned}\hat{\rho}_{lk}(s) &= \frac{\hat{\gamma}_{lk}(s)}{[\hat{\gamma}_{ll}(0)\hat{\gamma}_{kk}(0)]^{1/2}} \\ &= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T-S}L^{(l)}z_i(t)L^{(k)}z_i(t+s)}{[\sum_{i=1}^{N}\sum_{t=1}^{T}(L^{(l)}z_i(t))^2\sum_{i=1}^{N}\sum_{t=1}^{T}(L^{(k)}z_i(t))^2]^{1/2}}\end{aligned} \tag{3.36}$$

Knowing the space-time autocorrelations of the time series is not enough to select a tentative model from STARMA model classes, the space-time sample partial autocorrelation functions is also necessary. The space-time sample partial autocorrelations can be derived from the general form of STAR model, let the STAR model with order λ in space and k in time be,

$$z(t) = \sum_{j=0}^{k}\sum_{l=1}^{\lambda}\phi_{jl}W^{(l)}z(t-j)+\varepsilon(t) \tag{3.37}$$

Multiplying equation (3.37) by the lag term $[W^{(h)}z(t-s)]'$ we can obtain,

$$z(t-s)'W^{(h)}{}'z(t) = \sum_{j=1}^{k}\sum_{l=1}^{\lambda}\phi_{jl}z(t-s)'W^{(h)}{}'W^{(l)}z(t-j)+z(t-s)'W^{(h)}{}'\varepsilon(t) \tag{3.38}$$

Since $E[z(t-s)'\varepsilon(t)]=0$ for all $s>0$, taking the expected values and dividing both sides by N in equation (3.38) we get,

$$\gamma_{ho}(s) = \sum_{j=1}^{k}\sum_{l=0}^{\lambda}\phi_{jl}\gamma_{hl}(s-j) \tag{3.39}$$

which are the space-time auto-covariance in terms of Yule-Walker type equations. If we substitute $s=1,2,\ldots,k$ and $h=0,1,\ldots,\lambda$ in these equations yields a set of linear equations, and we can obtain the space-time partial autocorrelations by solving the last coefficient, ϕ'_{kl} as $l=0,1,\ldots,\lambda$ for $k=1,2,\ldots$, therefore the space-time partial autocorrelations of spatial order λ could be approximate as the last coefficient of each successive STAR process fitted.

With the space-time autocorrelations and space-time spatial partial autocorrelations, one may examine whether the space-time autocorrelations and partial autocorrelations are tail off, cut off, or both tail off to tentative select a STARMA, STAR, or STMA classes of model to be fitted in the following procedure.

Then, since efficient estimation of the parameters is necessary, fitting the tentative selected model is the second stage of the procedure. And the next stage is the process called diagnostic checking which evaluates the adequacy of the fitted model, if the sample space-time autocorrelation function of the residuals from the tentative selected model is white noise and all of the parameters in the fitted model are proven statistically significant, then the fitted model is adequate for describing the data and hence can be applied to forecast future in the last stage. Otherwise, the procedure has to return to the first stage, i.e. fine tuning for another tentative model to be applied and then go through the modeling process again.

3.3 Concluding Remark

In this chapter we have briefly introduced the modeling procedure of the univariate ARIMA model and multivariate STARMA model, the most essential and significant difference between univariate and multivariate model is the input data of the model. In general, univariate models can only fit the model with a particular variable, which means that even though there is a historical database collecting information from different locations in the system, univariate models remain forecast future conditions by past data at a particular location separately without utilizing information collecting from other locations and thus neglect the possible influences from different locations. And since unvariate models operate individually, the main advantages are that they are easy to employ and different variables can be estimated with different parameters thus being tightly parameterized.

On the other hand, the multivariate model considers past information not only from an isolate data collecting point but also past information from other nearby locations. The modeling procedure of multivariate models such as finding the relationship between different variables or parameters estimating are often more complicated than model univariate models several times. Since multivariate models can simultaneously consider the influences across locations in the system, whenever there are some independencies between the forecasting targets, we can expect that the STARMA model can reveal more information as compared to estimating several univariate models, one for each detector location.

Chapter 4 Case Study

In this chapter, we will illustrate the modeling procedure of the forecasting techniques presented in Chapter 3. In section 4.1 we will give a brief analysis of the traffic data being used, the applications of ARIMA and STARMA models to fit those traffic data will be depicted in section 4.2 and 4.3 respectively. At the end of this chapter, section 4.4 will show the comparison of forecasting performance of these models.

4.1 Data Collection from Vehicle Detectors

Taipei City, the capital city of Taiwan, has an area of 270 km^2 and there are over 700 vehicle detectors collecting traffic data in the area. To illustrate the ARIMA and STARMA procedure by modeling traffic flow in the urban area, we select a study area of about 5km by 4 km as displayed in Figure 4.1. In order to show the relationship of the traffic at different locations, 24 vehicle detectors for the west bound traffic are collected, with their locations also shown in Figure 4.1, and the exact locations of vehicle detectors under study are shown in Appendix 1. Only flow data of same direction would be used in our study. It is noted that while there are other vehicle detectors in the study area, they are collecting traffic at different directions. And we assumed that there are no dependencies between vehicle detectors of different directions; those detectors are excluded in this study.

Figure 4.1 Locations of the vehicle detectors

To compare the applicability of the models on urban arterial roads, traffic data, including traffic volume, speed and lane occupancy, were collected every 5 minutes from these detectors. Since the maximum cycle length of traffic signal in Taipei City is 200 seconds, if we apply 5-min data as our model input, these data may be strongly affected by traffic signal leading a wrong model. Vythoulkas (1993) stated that using intervals shorter than 10 minutes declines the quality of information that one could obtain in the prediction systems. Highway Capacity Manual (2000) also indicated the 15-min interval as the best interval for traffic flow prediction, therefore, those 5-min data are aggregated into 15-min intervals for better forecasting accuracy, and hence there are 96 observations in each day in the analysis.

Data sets of four weeks were recorded from 29th June 2009 to 24th July 2009. The data of the first three weeks were used for calibrating the model, and the data of last week were used for validation purpose. We only consider data from Tuesday to Thursday in the analysis, as these days are having similar daily traffic pattern. Monday, Friday and weekends are ignored here to avoid the large deviations compared the mid-week traffic conditions. The data of twelve days, with a total of 1152 observations, for each detector were analyzed here. The data of nine days, i.e. 864 observations were used to calibrate the model parameters, and the data of last three days with 288 observations were use to check the performance of the model.

Last but not least, ARIMA and STARMA models are time series models that require continuous data without missing. However, due to some reasons such as signal failure or communication problems, sometimes vehicle detectors will be unable to collect traffic data and hence missing data appears. In these cases, we will use the average of its former and later observation to interpolate or replacing the zeros so as to get a continuous database.

4.2 Seasonal ARIMA

In this section we will fit the seasonal ARIMA model for all of the 24 detectors, since the modeling procedure of the ARIMA model is analogous; we will only consider the traffic volume data from the downstream detector VD No.2 on Minsheng W. Road to illustrate the modeling procedure of seasonal ARIMA model but show all the model results of 24 vehicle detectors at the end of this section.

To fit the seasonal ARIMA model, Figure 4.2 shows the time series plot of the traffic data of VD No.2, the vertical axis represents the volume data within 15 minute interval and the horizontal axis is the time index. The time series plots reveal that the data is nonstationary with seasonality every 96 observations, i.e. data of one day.

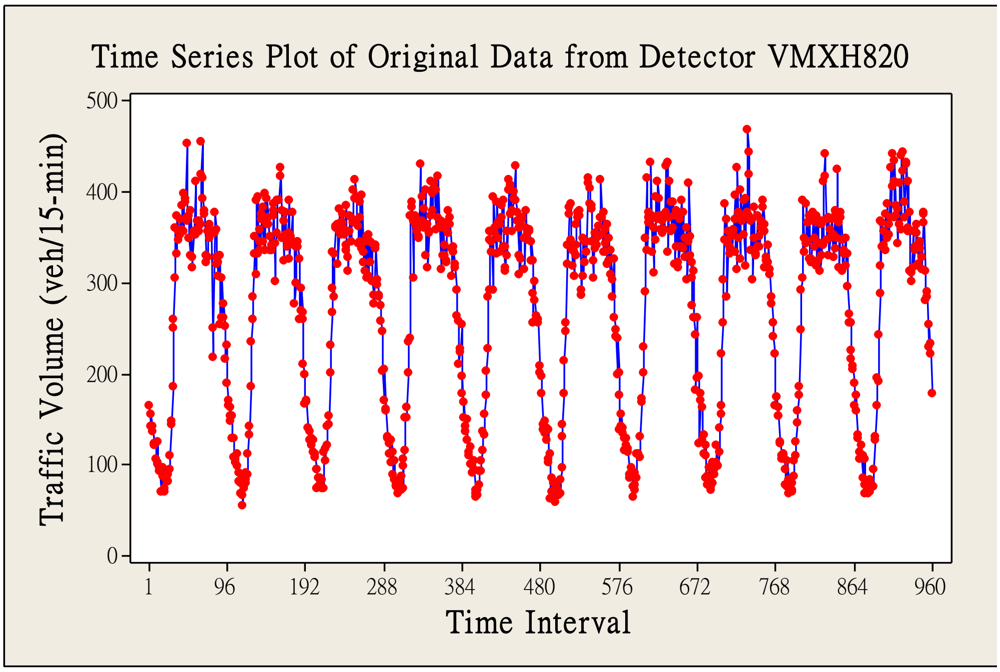


Figure 4.2 Time series plot of the traffic flow data of VD No.2

We used the ARIMA procedure in statistical software SAS to calculate the sample autocorrelation function (SAC) and sample partial autocorrelation function (SPAC) so as to fit the ARIMA models, and the corresponding SAC and SPAC plots are shown in Figure 4.3 and Figure 4.4. As the SAC of original time series appear to be nonstationary with seasonality every 96 observations, according to the assumption of ARIMA model that the time series to be fitted is stationary, we then transformed them by first seasonal differenced i.e. taking difference of 96 time lags to get a stationary time series.

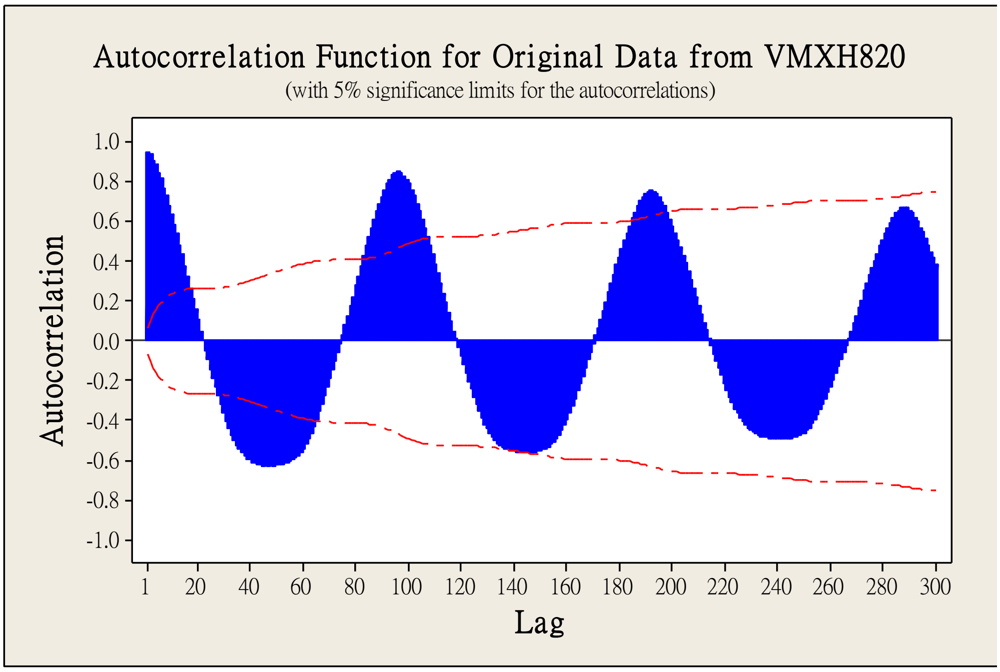


Figure 4.3 SAC for the traffic flow data of VD No.2

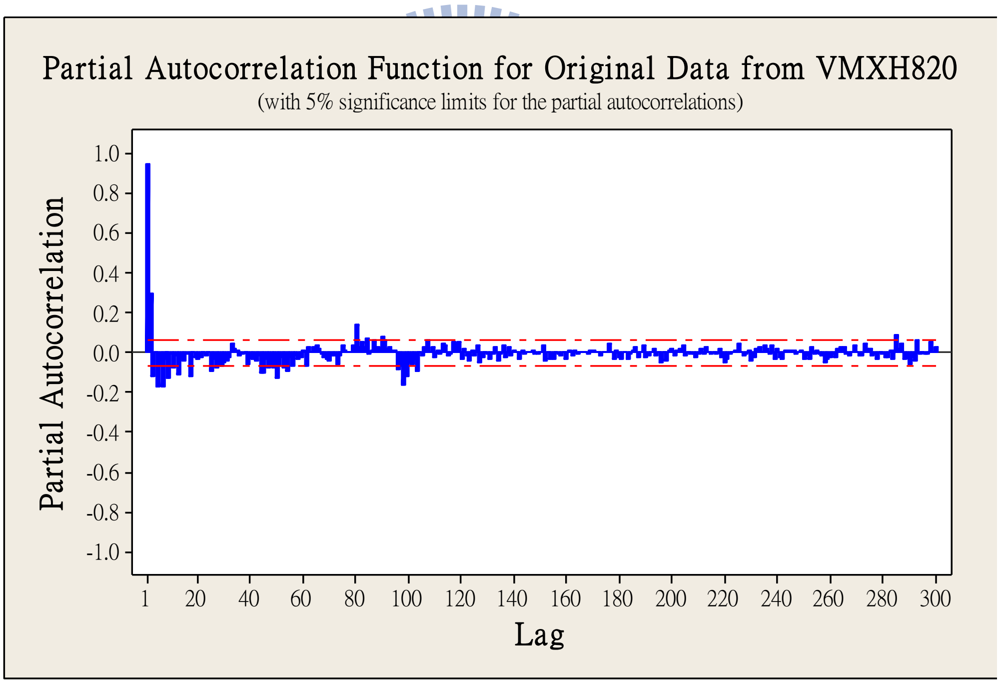


Figure 4.4 SPAC for the traffic flow data of VD No.2

While examining the SAC and SPAC plot can give a roughly identification of whether the time series is stationary, to be more precisely about the time series to be fitted is stationary, augmented Dickey-Fuller (ADF) test is employed. Rather than testing the original time series, we tested whether the first seasonal differenced time series is stationary or not. The null hypothesis is that the time series has a unit root at lag d, which means that the time series is nonstationary, Table 4.1 shows transformed traffic data of VD No.2 and three variations of the test statistics, which investigates whether the model includes a constant mean or a time trend when fitting the model are employed, and different lags were also examined.

We can see from Table 4.1 that the first column specifies three types of models, which are zero mean, single mean, or trend. The third column (Rho) and the fifth column (Tau) are the test statistics for unit root testing. Other columns are their p-values. Since the test results show that all of the p-values are small enough to reject the null hypothesis that the series has a unit root, which means that the seasonal differenced time series is stationary, we can then proceed to the next step to identify which *p* and *q* should be selected as a tentative model.

Table 4.1 ADF test for first seasonal differenced traffic data from VD No.2

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-648.993	0.0001	-23.67	<.0001		
	1	-490.312	0.0001	-15.63	<.0001		
	2	-600.500	0.0001	-14.65	<.0001		
Single Mean	0	-649.057	0.0001	-23.66	<.0001	279.84	0.0010
	1	-490.435	0.0001	-15.62	<.0001	122.06	0.0010
	2	-600.749	0.0001	-14.64	<.0001	107.18	0.0010
Trend	0	-649.116	0.0001	-23.64	<.0001	279.50	0.0010
	1	-490.491	0.0001	-15.61	<.0001	121.90	0.0010

Figure 4.5 and Figure 4.6 present the SAC and SPAC plots of the first seasonal differenced values. As they both die down quickly at seasonal level, implies that we should consider both autoregressive and moving average process when fitting the ARIMA model. Furthermore, as shown in Figure 4.5, the first seasonal differenced plot of ACF of VD No.2 cuts off at lag 2 and has a spike at lag 96, implies that the second-order autoregressive (AR2) process of nonseasonal and seasonal autoregressive order of lag 96 (AR96) must be included in the process. Similarly, Figure 4.6 shows that the first seasonal differenced plot of PACF of VD No.2 has spikes at lag 1, lag 2 and lag96, therefore, MA1 or MA2 of nonseasonal and MA96 of seasonal moving average process also seemed to be a possible ARIMA component.

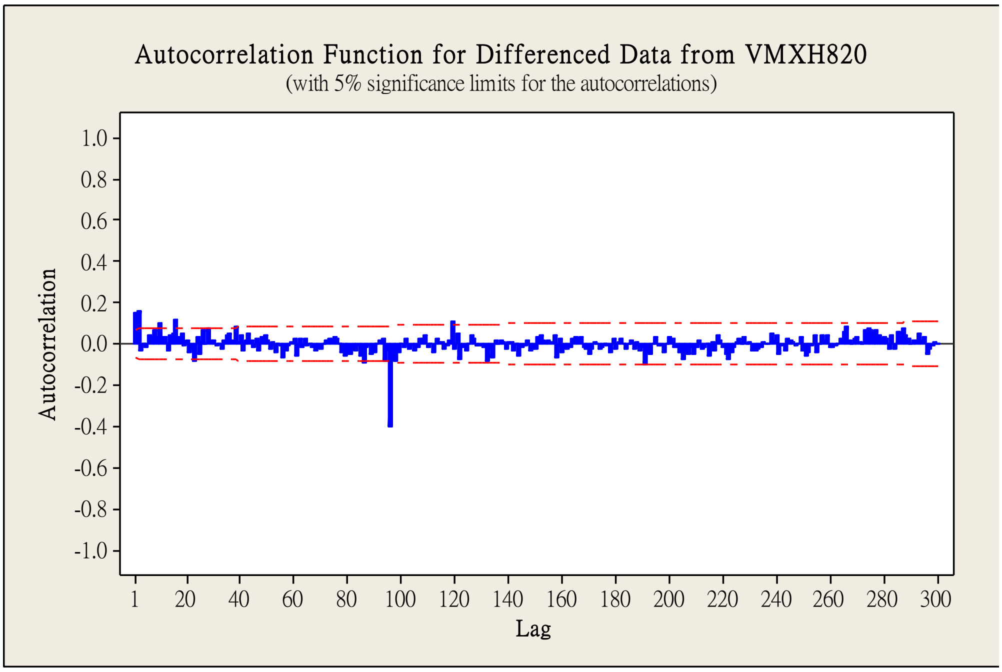


Figure 4.5 SAC for the differenced traffic data of VD No.2

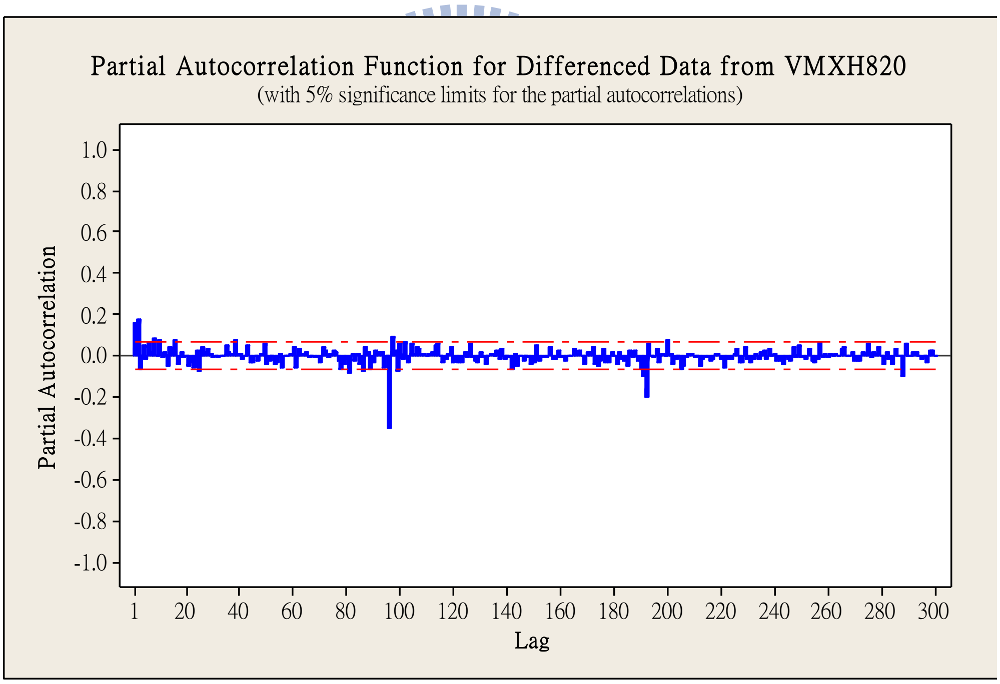


Figure 4.6 SPAC for the differenced traffic data of VD No.2

Combining the plots of ACF and PACF and following the guidelines in Table 3.2, the model of reasonable combinations of parameters for VD No.2 is of the form,

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \varepsilon_t - \theta_{96}\varepsilon_{t-96} \tag{4.1}$$

where ϕ_1 and ϕ_2 are the coefficients of the autoregressive process of order 1 and 2, ε_t is the random error components in the time series, θ_{96} is the coefficient of the moving average process of order 96.

The results of the coefficient estimations using maximum likelihood estimate with reasonable combinations of parameters and their corresponding *t*-statistics before diagnostic check for the seasonal ARIMA model are shown in Table 4.2. While all of the *t*-statistics for the parameters were found to be significance greater than 2, meaning that these parameters should be included in the model. Table 4.3 reveals the autocorrelations of the residuals, while all of the parameters tentative selected in the model are all significant, since the p-value in the autocorrelation check for the residuals were found to be smaller than 0.05, indicated that there are still some parameters need to be included in the model.

Table 4.2 Parameters estimated before diagnostic check for VD No.2 (ARIMA)

Parameter	Estimated value	Standard Error	t Value	Approx Pr > |t|
MA96	0.60929	0.03030	20.11	<.0001
AR1	0.14836	0.03599	4.12	<.0001
AR2	0.09899	0.03599	2.75	0.0061

Table 4.3 Autocorrelation check for the residuals for VD No.2 (1^{st} model)

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	9.21	3	0.0267	0.009	0.013	-0.089	0.008	-0.043	0.043
12	21.51	9	0.0106	-0.027	0.088	-0.007	0.074	0.013	0.042
18	38.77	15	0.0007	-0.084	0.018	0.003	0.113	-0.011	0.039
24	47.72	21	0.0008	0.000	0.039	-0.050	-0.007	-0.053	0.066
30	56.41	27	0.0008	-0.046	0.044	-0.021	0.065	-0.017	0.043
36	61.72	33	0.0018	-0.026	-0.017	-0.049	0.035	0.009	0.043
42	68.21	39	0.0026	-0.017	0.071	-0.016	0.003	-0.046	0.014
48	73.59	45	0.0045	0.018	0.040	-0.043	0.022	-0.044	0.017
54	78.31	51	0.0083	0.017	0.005	-0.044	0.031	-0.050	0.002
60	85.43	57	0.0087	-0.001	-0.060	-0.025	0.029	-0.007	0.058
66	90.33	63	0.0136	-0.059	0.017	-0.033	-0.004	-0.021	0.023
72	92.64	69	0.0304	-0.028	0.005	-0.006	-0.003	0.001	0.043
78	96.73	75	0.0465	-0.014	-0.003	0.006	0.063	0.010	-0.022
84	103.82	81	0.0446	-0.060	0.029	-0.041	-0.011	-0.040	0.021
90	111.77	87	0.0380	-0.043	-0.026	-0.002	0.039	-0.069	-0.015
96	114.42	93	0.0653	-0.023	0.034	-0.000	-0.035	0.011	-0.003

Repeating the estimation and diagnostic checking procedure for all of the statistics indicate that the tentative selected model is suitable for the data and all of the p-value of autocorrelations of the residuals are greater than 0.05. Therefore, the model suitable for describing the data of VD No.2 is,

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_{96} \varepsilon_{t-96} - \theta_{97} \varepsilon_{t-97} \tag{4.2}$$

The corresponding parameter estimations and autocorrelation check of residuals are shown in Table 4.4 and Table 4.5. Since all of the *t*-statistics of estimated parameters are greater than 2, representing that all of these parameters should be included in the model. Furthermore, since almost the p-value in the autocorrelation check for the residuals were found to be greater than 0.05, indicating that the residuals are white noise at the 95% confidence level, this model could be used to forecast traffic conditions for VD No.2. As a result, each observation of VD No.2 at time *t* can be expressed as a linear combination of previous observations at time *t*-1, *t*-2 and the prediction error made at time *t*-1, *t*-96 and *t*-97, plus a random error.

Table 4.4 Parameters estimated after diagnostic check for VD No.2 (ARIMA)

Parameter	Estimated value	Standard Error	t Value	Approx Pr > |t|
MA1	-0.9426	0.0254	-37.05	<.0001
MA96	0.6242	0.0300	20.84	<.0001
MA97	0.5802	0.0341	17.00	<.0001
AR1	-0.7576	0.0420	-18.05	<.0001
AR2	0.2193	0.0361	6.08	<.0001

Table 4.5 Autocorrelation check for the residuals for VD No.2 (Final model)

To Lag	Chi-Square	DF	Pr >ChiSq	Autocorrelations					
6	2.71	1	0.0998	-0.007	0.038	-0.020	-0.037	0.012	-0.006
12	9.15	7	0.2421	0.022	0.043	0.042	0.027	0.059	0.000
18	16.35	13	0.2306	-0.033	-0.017	0.041	0.072	0.029	0.003
24	20.35	19	0.3736	0.042	-0.001	-0.011	-0.038	-0.021	0.036
30	22.22	25	0.6229	-0.018	0.018	0.008	0.037	0.011	0.012
36	25.48	31	0.7458	0.006	-0.046	-0.015	0.005	0.036	0.018
42	28.00	37	0.8569	0.007	0.047	0.008	-0.020	-0.015	-0.013
48	30.30	43	0.9277	0.040	0.014	-0.021	0.002	-0.020	-0.011
54	33.93	49	0.9500	0.039	-0.021	-0.024	0.011	-0.031	-0.027
60	42.69	55	0.8868	0.026	-0.089	0.003	0.006	0.004	0.044
66	45.26	61	0.9343	-0.046	0.005	-0.019	-0.022	-0.006	0.005
72	46.92	67	0.9704	-0.012	-0.012	0.013	-0.021	0.018	0.027
78	50.51	73	0.9793	0.003	-0.015	0.021	0.043	0.025	-0.033
84	54.95	79	0.9820	-0.046	0.017	-0.034	-0.022	-0.033	0.007
90	61.44	85	0.9747	-0.031	-0.038	0.004	0.028	-0.063	-0.019
96	65.41	91	0.9803	-0.017	0.017	0.014	-0.051	0.032	-0.014

Repeating the modeling procedure of seasonal ARIMA model to fit the model of remaining 23 vehicle detectors, we can obtain the best fitted model for all 24 vehicle detectors and their corresponding parameters estimated shown in Table 4.6. As previous stated, fitting the model separately for data from different vehicle detectors may lead to inconsistent of the number of parameters across different models, hence some of the best fitted model can forecast future using only two parameters (VD No.6, VD No.12 and VD No.13), but some of the observation data need up to six parameters (VD No.4 and VD No.11) to obtain the best fitted model.

Table 4.6 ARIMA Parameters estimated for all 24 vehicle detectors

VD No.	Parameter estimated						
	AR1	AR2	AR3	MA1	MA2	MA96	MA97
1	-0.6732	0.2896		-0.8849		0.6737	0.6028
2	-0.7576	0.2193		-0.9426		0.6242	0.5802
3	-0.7071	0.2504		-0.8901		0.6692	0.6352
4	0.2269	0.9375	-0.2208		0.8457	0.6732	-0.6512
5	-0.7245	0.2339		-0.8916		0.6109	0.5686
6	0.2990					0.7022	
7	0.2347	0.0731				0.7000	
8	-0.6538	0.3068		-0.9182		0.6318	0.5685
9		0.1961		-0.3789		0.6878	0.2509
10		0.2182	0.1238	-0.4229		0.7251	0.3003
11	1.0351	0.1214	-0.1877	0.9249		0.6255	-0.5946
12	0.2727					0.7145	
13	0.3013					0.7073	
14		0.1592		-0.2073		0.6629	0.1837
15	-0.6342	0.3253		-0.9167		0.7280	0.6753
16	0.3963	0.0703				0.5667	
17	0.1564	0.1018				0.5994	
18	1.3212	-0.3438		0.9347		0.5913	-0.5618
19	1.1801	-0.2078		0.9061		0.5503	-0.5119
20	0.2873				-0.1077	0.6098	-0.0514
21	0.2520	0.1644				0.5342	
22	0.2095				-0.1153	0.5774	
23		0.2989		-0.2369		0.6566	0.2131
24	0.8273			0.5910		0.7041	-0.4687

To evaluate the performance of the fitted model, since there are various criteria being used in the literature and no one has concluded that which of them is better, so we selected the root mean squared error (RMSE) and the mean absolute percentage error (MAPE) as criteria in our study. The definition of RMSE and MAPE are,

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{4.3}$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100\% \tag{4.4}$$

where y_t is the observation at time *t*, $\hat{y}_t$ is the estimated or predicted value at time *t*, and *n* is the sample size over the study period.

With the definition of RMSE and MAPE, Table 4.7 present the model residuals and forecasting errors of ARIMA models for all 24 detectors estimated by RMSE and MAPE. The calibrating data are traffic flows from Tuesday to Thursday of the first three weeks, and the forecasting data are traffic flows on Tuesday of last week, i.e. data on 21st July. As expected, most of the model residuals appear to be smaller than the forecasting errors since the parameters were estimated based on calibrating data. Hence, when we used them to forecast, the forecasting errors would be greater than model residuals, but their differences were not too much. Lewis (1982) stated that MAPE lower than 10% represents the forecasting is highly accurate, MAPE between 10% and 20% represents the forecasting is quite well, and MAPE between 20% and 50% represents the forecasting is reasonable. From Table 4.7 we can find that most of the model residuals and forecasting errors estimated by MAPE are smaller than 20%. Furthermore, if we use the average volume of each detector as the weight, then the weighted average of 24 detectors also revealed that both estimating errors and forecasting errors were smaller than 20%, so we can conclude that the forecasting abilities of ARIMA models we fitted are quite well.

Table 4.7 Estimation and forecasting errors by RMSE and MAPE (ARIMA)

VD No.	Estimation errors		Forecasting errors	
	RMSE	MAPE	RMSE	MAPE
1	23.71	10.15%	28.26	11.47%
2	26.96	8.18%	31.82	9.27%
3	47.71	12.38%	52.29	13.13%
4	45.83	11.48%	115.92	45.93%
5	47.23	11.29%	56.45	14.41%
6	20.58	10.05%	18.45	10.36%
7	19.04	9.91%	16.68	10.18%
8	21.51	11.22%	21.84	14.50%
9	19.25	9.64%	17.33	8.51%
10	12.50	24.54%	10.35	22.33%
11	18.79	15.09%	21.69	18.43%
12	18.89	13.71%	17.77	13.47%
13	17.58	15.49%	19.61	16.81%
14	24.93	11.21%	22.32	11.20%
15	23.45	11.86%	27.48	13.77%
16	24.14	9.92%	24.74	9.28%
17	46.47	10.07%	48.55	10.81%
18	50.17	8.59%	49.61	10.61%
19	30.66	19.37%	30.37	23.58%
20	38.35	16.53%	35.11	18.35%
21	40.72	15.34%	42.10	18.25%
22	42.84	15.66%	43.95	14.14%
23	32.26	19.59%	36.71	20.40%
24	30.19	20.65%	33.22	28.46%
Weighted Average	36.86	12.70%	46.83	16.18%

In particular, the forecasting of VD No.4 has a large error, with model RMSE=45.83, MAPE=11.48% and forecasting RMSE=115.92, MAPE=45.93%; whereas the errors estimated from most of the other detectors by RMSE and MSE are beneath 50 and 30%, respectively. Possible reason is that, if we referred to the original time series plot of VD No.4 as shown in Figure 4.7, the traffic flow from VD No.4 from interval 864 to interval 960, i.e. observations being forecasted, are averagely lower than observations used to calibrate. That's why the forecasting error is extremely large compare to other vehicle detectors. And this special case also revealed the drawback of ARIMA model that it is weak to forecast time series with sudden change in the pattern.

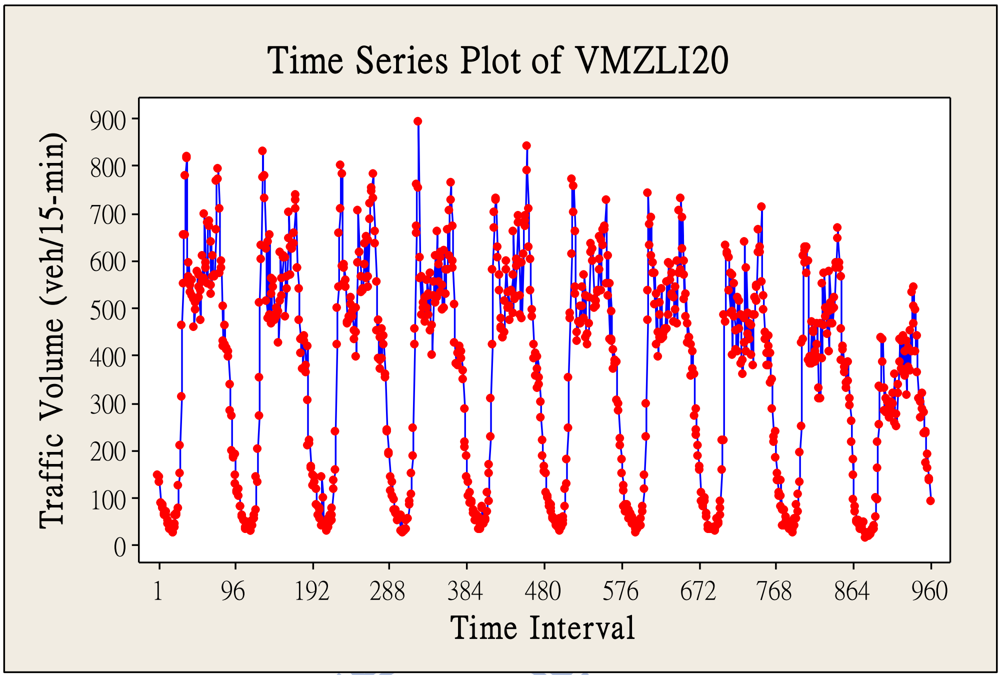
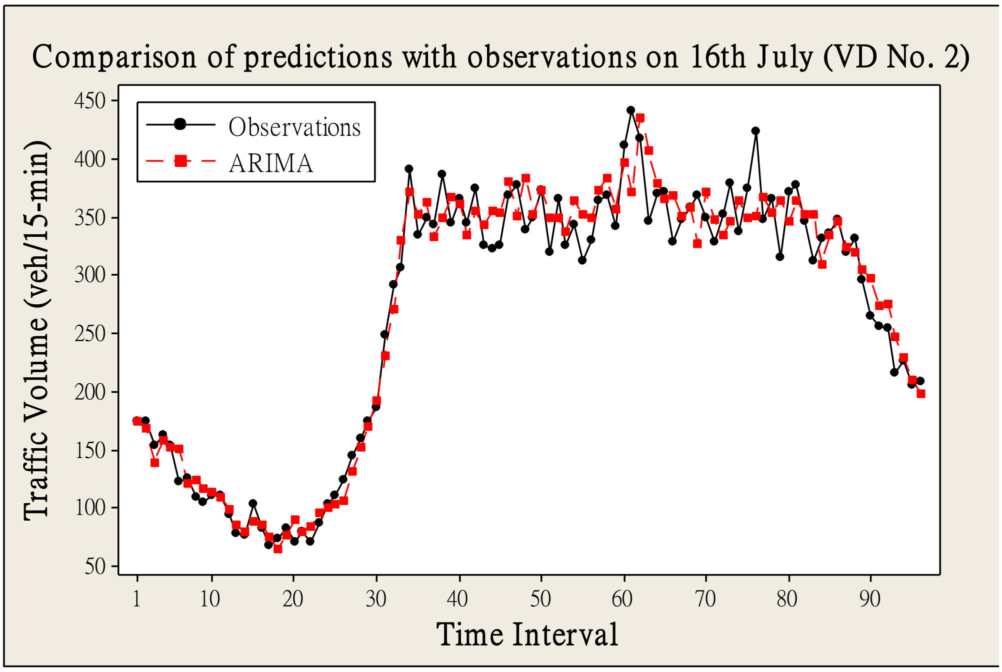
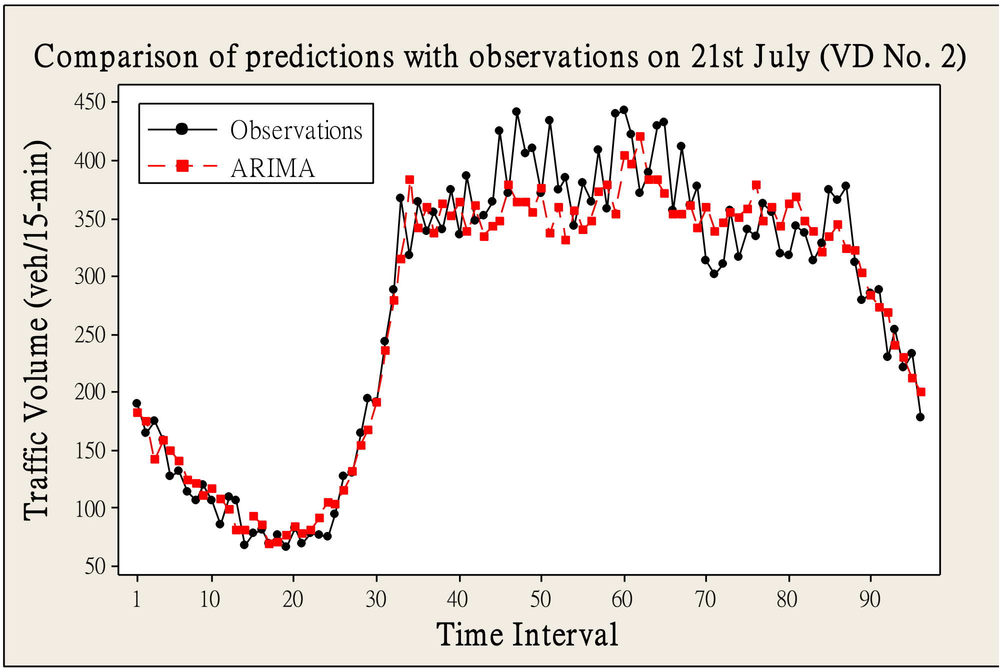


Figure 4.7 Time series plot of traffic flow data of VD No.4

If we consider VD No.4 as a special case thus not in our consideration, among 23 detectors, we selected the one with minimum estimating error with RMSE=26.96 and MAPE=8.18% (VD No.2) and the one with maximum error with RMSE=12.50 and MAPE=24.54% (VD No.10) and compared their estimated volume with the observations on 16th (estimating) and 21st (forecasting) July shown in Figure 4.8 (a), 4.8 (b), 4.9 (a) and 4.9 (b), which present typical 24-hr flow variation pattern of VD No.2 and VD No.10. As each time interval represents 15 minutes, traffic flows increase rapidly from time interval 28 to 34, indicate that a large amount of traffic flows appear from 7:00 am to 8:30 am. Since then they fluctuate at a higher level before 76th interval, which means that traffic flows continue fluctuating at a high level during daytime, and then start to decrease around 7:00 pm. We can also observe from Figure 4.8 (a), 4.8 (b), 4.9 (a) and 4.9 (b) that the model fits the observations quite well during off-peak when the flows are low in values and the observations were not fluctuating. But the model may not capture the extreme peak variations when it comes to rapid fluctuations in the peak hour. Furthermore, if we compare the traffic flow of VD No.2 and VD No.10, we can find that data from VD No.10 fluctuate more than data from VD No.2; as a result, it is not surprising that the estimated error from VD No.2 is smaller than that from VD No.10.

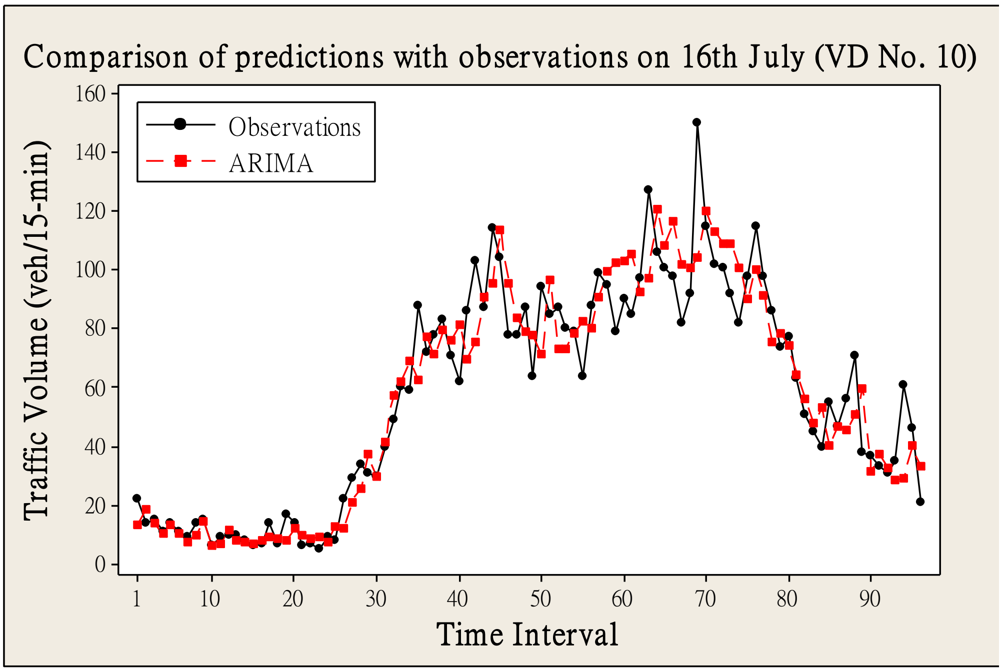


(a) Comparison of 7/16 data from VD No.2

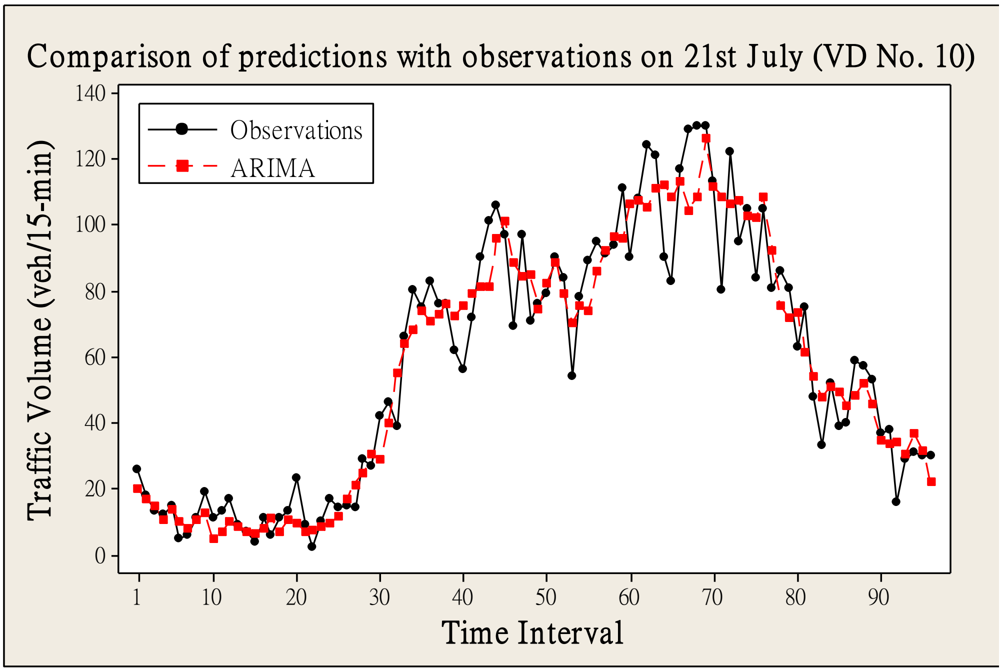


(b) Comparison of 7/21 data from VD No.2

Figure 4.8 Comparison of predictions with observations from VD No.2



(a) Comparison of 7/16 data from VD No.10



(b) Comparison of 7/21 data from VD No.10

Figure 4.9 Comparison of predictions with observations from VD No.10

4.3 STARMA Model

The modeling approaches of multivariate space-time autoregressive moving average (STARMA) will be examined in this section. As previously stated, the univariate ARIMA models used to forecast each of the time series separately, ignoring the dependencies between each other. Nevertheless, traffic flow is not isolated from nearby locations and instead, can be strongly influenced by other traffic flows within the system. Hence the multivariate STARMA model which integrates weight matrices into the model on account of the spatial dependencies between time series would be a more adequate model to forecast traffic flow. As the STARMA considers the relationships of the time series at the detectors locations, we have to assign a weight matrix representing the network structure and capturing the spatial characteristics of the data collecting points in advance to incorporate those dependencies into the model.

4.3.1 Weight Matrices in STARMA

In order to incorporate the physical characteristics of the traffic flow network into STARMA model, consider the locations of vehicle detectors that were shown in Figure 4.1. We define a hierarchical system showing the spatial neighboring relationship between the locations of the 24 detectors, with details shown in Table 4.8. The first order neighbors are those nearest to the location of interest, and the second order neighbors are those closer than third order but farther than first order ones.

As we have mentioned in section 3.2, we assumed that, in moderate traffic congestion, upstream traffic conditions are not influenced by the downstream conditions. Some of the upstream detectors such as VD No.5 (VMYN820) and VD No.16 (VKWNV20) are on the boundary of the study area and not influenced by any other detectors. Hence, they may have no neighboring detectors. Furthermore, we only consider the first and second order neighbors in our study. As the third order implies a 30-45 minutes traveling time (for our case with a 15 minutes time internal), we can expect that the traffic flows from third-order neighbors would have little influences and are neglectable.

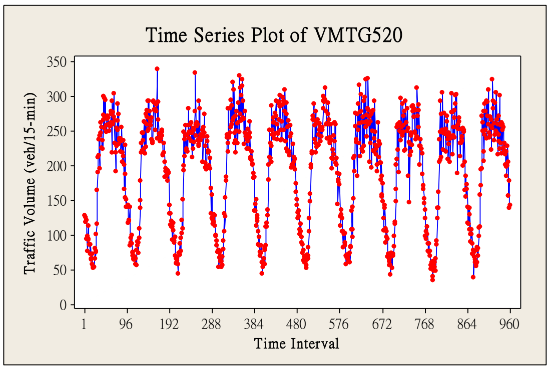
Table 4.8 Spatial neighboring relationship of the 24 vehicle detectors

VD No.	Detector ID	1st order	2nd order	VD No.	Detector ID	1st order	2nd order
1	VMTG520	2	6	13	VJTJ960	14,15	16
2	VMXH820	3	6	14	VKLLH20	15,16	9
3	VMZL960	4,5	9	15	VKRM820	16	-
4	VMZLI20	5	-	16	VKWNV20	-	-
5	VMYN820	-	-	17	VKLGD20	18	11
6	VMFIG20	7,8	9	18	VKAHN20	-	13,19
7	VMDL820	8,9	3	19	VIRHZ20	-	22
8	VMEKD00	9	3,14	20	VIPIZ60	20,21	23
9	VMDL800	-	4,15	21	VIPJA20	21,22	24
10	VM7FI60	-	1,11	22	VINKW20	22,23	-
11	VLKGF40	12	6	23	VINLD61	23,24	-
12	VLGGY60	-	6	24	VINM760	-	-

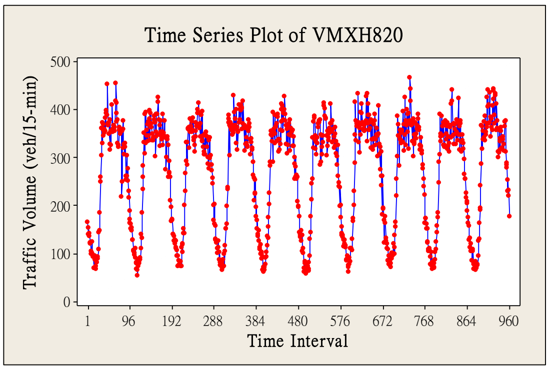
With the above chosen spatial neighboring relationship of the vehicle detectors, we are able to apply the definition of weight matrix presented in section 3.2. The l^{th} order neighbors of each location i have equal weights and the row sum equals to one. As a result, the above spatial neighboring relationship can be transferred into weight matrices of l^{th} order which will be used in the STARMA model in the next section.

4.3.2 Numerical analysis

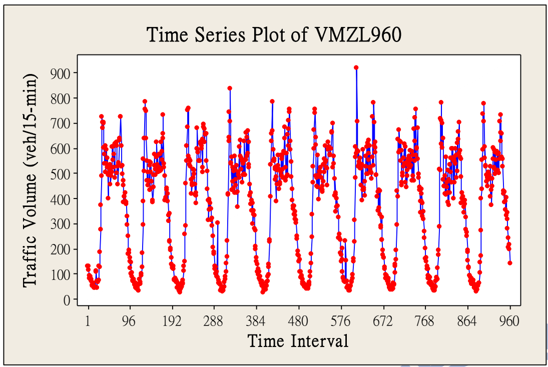
To fit STARMA model, Figure 4.10 (a)-(d) show some of the time series plots of traffic flow observations for the 24 vehicle detectors from 30th June 2009 to 16th July 2009. Only data from Tuesday to Thursday are used here as it is known that weekdays at the middle of week have a better consistent pattern as compared to Monday/Friday and weekends. In the figure, the vertical axis represents the volume data within 15 minute interval and the horizontal axis is the time index. As seen in the figures, the traffic volume at different detector locations varies. For example, the traffic volume of VD No.1 is less than 350 (veh/15 min), but the volume of VD No.4 is almost 900 (veh/15 min). They all reveal that a pattern that is nonstationary with seasonality at every 96 observations, i.e. a cycle of one day. Therefore, similar to ARIMA model, we need to take seasonal difference of 96 time lag for all 24 time series to transfer the original nonstationary series into stationary time series.



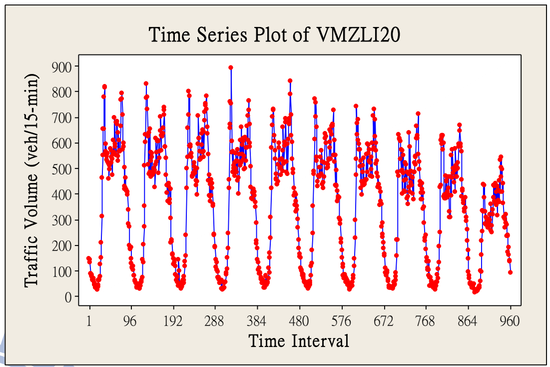
(a) VD No.1



(b) VD No.2



(c) VD No.3



(d) VD No.4

Figure 4.10 Time series plots of the traffic data of selected VDs

The Augmented Dickey-Fuller (ADF) test is then employed to all of the differenced time series to check whether they are stationary. Since the testing procedure repeats the one depicted in ARIMA modeling, we will not show all of the testing results of the 24 detectors respectively. The test results, which are not shown here, indicate that all of the p-values are small enough to reject the null hypothesis that the series has a unit root, and it is confirmed that the seasonal differenced time series are stationary. We can then proceed to the next step to identify which p and q should be selected as a tentative model.

To fit the STARMA model, similar to the modeling procedure of ARIMA, first we should examine the space-time autocorrelation and space-time partial autocorrelation functions of the time series to identify the autoregressive and moving average term for fitting the model. Table 4.9 and Table 4.10 display the space-time autocorrelation and space-time partial autocorrelation function for the differenced series. The space-time autocorrelations appear to have a spike at temporal lag 2 and 96 of zero spatial lag at nonseasonal and seasonal level, indicating that the moving average term of order 2 and order 96 should be included in the STARMA model. The space-time partial autocorrelations indicate that we should include the autoregressive term of order 96 into the STARMA model. Therefore, the tentative selected model to fit the data of the 24 vehicle detectors will be,

$$
\begin{aligned}
z_t &= \phi_{10} z_{t-1} + \phi_{11} W^{(1)} z_{t-1} + \phi_{12} W^{(2)} z_{t-1} + \phi_{20} z_{t-2} + \phi_{21} W^{(1)} z_{t-2} + \phi_{22} W^{(2)} z_{t-2} \\
&\quad + \phi_{30} z_{t-3} + \phi_{(96)} z_{t-96} - \theta_{10} \varepsilon_{t-1} - \theta_{20} \varepsilon_{t-2} - \theta_{(96)} \varepsilon_{t-96} + \varepsilon_t
\end{aligned} \tag{4.5}
$$

Table 4.9 Space-time autocorrelation functions of the differenced series

Spatial lag (l)	0	1	2
Time lag (s)			
1	0.2403	0.0714	0.0306
2	0.1418	0.0796	0.0109
3	0.0271	0.0095	0.0039
4	0.0666	0.0559	0.0060
5	0.0127	0.0095	0.0014
6	0.0793	0.0458	0.0089
7	0.0419	0.0489	0.0048
…			
95	-0.0765	-0.0266	-0.0177
96	-0.4097	-0.0975	-0.0088
97	-0.0924	-0.0217	0.0081

Table 4.10 Space-time partial autocorrelation functions of the differenced series

Spatial lag (l)	0	1	2
Time lag (s)			
1	0.2349	0.0068	0.0118
2	0.0808	0.0434	-0.0098
3	-0.0244	-0.0197	-0.0003
4	0.0522	0.0366	-0.0018
5	-0.0120	-0.0079	-0.0043
6	0.0657	0.0157	0.0036
7	0.0071	0.0352	-0.0055
…			
95	-0.0665	-0.0060	-0.0189
96	-0.3900	0.0149	0.0179
97	0.1001	0.0025	0.0159

Next we need to estimate the coefficients of the parameters tentatively selected in the identification stage. We employed the SAS statistic software to estimate the model for the coefficients of parameters by maximum likelihood approach. The estimated values and the corresponding *t*-statistics are shown in Table 4.11. As presented in the table, the t-statistics of parameters ϕ_{12}, ϕ_{20}, ϕ_{21} and θ_{10} appear to be insignificant, these parameters should be removed from the model.

Table 4.11 Parameters estimated before diagnostic check (STARMA)

Parameter	ϕ_{10}	ϕ_{11}	ϕ_{12}	ϕ_{20}	ϕ_{21}	ϕ_{22}
Estimated value	0.2529	0.0263	0.0128	0.0111	0.0073	0.0194
t value	25.95	3.54	1.84	1.01	0.98	2.8
Parameter	ϕ_{30}	ϕ_{96}	θ_{10}	θ_{20}	θ_{96}	RMSE
Estimated value	0.0351	0.0176	0.0042	0.0414	-0.666	71.75
t value	4.87	2.26	0.5	5.07	-143	

Reformulating the STARMA model step-by-step, we have the final model as,

$$\begin{aligned} z_t &= \phi_{10} z_{t-1} + \phi_{11} W^{(1)} z_{t-1} + \phi_{12} W^{(2)} z_{t-1} + \phi_{22} W^{(2)} z_{t-2} + \phi_{30} z_{t-3} + \phi_{(96)} z_{t-96} \\ &\quad - \theta_{20} \varepsilon_{t-2} - \theta_{(96)} \varepsilon_{t-96} + \varepsilon_t \end{aligned} \tag{4.6}$$

The parameters estimated for equation (4.6) and their corresponding *t*-statistics are displayed in Table 4.12. Since all of the *t*-statistics indicate that the tentative selected model is suitable for the data, therefore, we can proceed to check the space-time autocorrelations and partial autocorrelations of the residuals, shown in Table 4.13 and 4.14. Further checking confirms that except for the lags that are already included in the model, the autocorrelation and the partial correlations of the residuals of other lags appear to be insignificant, and we can conclude that the tentative selected STARMA model is suitable for describing the data.

Table 4.12 Parameters estimated after diagnostic check (STARMA)

Parameter	ϕ_{10}	ϕ_{11}	ϕ_{22}	ϕ_{30}	θ_{20}	θ_{96}	RMSE
Estimated value	0.2585	0.0292	0.0191	0.0372	-0.0479	0.6651	37.96
t value	46.19	4.01	2.84	6.02	-9.77	143.32	

Table 4.13 Space-time autocorrelations of the residuals

Spatial lag (l)	0	1	2
Time lag (s)			
1	-0.0396	-0.0214	0.0244
2	0.0779	0.0750	0.0088
3	-0.0644	-0.0420	0.0023
4	0.0814	0.0601	0.0023
5	-0.0470	-0.0387	-0.0082
6	0.1000	0.0425	0.0081
7	-0.0117	-0.0134	-0.0011
…			
95	0.0231	0.0179	0.0037
96	0.0472	0.0453	0.0335
97	-0.0005	-0.0099	0.0231

Table 4.14 Space-time partial autocorrelations of the residuals

Spatial lag (l)	0	1	2
Time lag (s)			
1	-0.0393	-0.0141	0.0246
2	0.0627	0.0540	-0.0003
3	-0.0531	-0.0210	0.0021
4	0.0640	0.0302	-0.0036
5	-0.0277	-0.0151	-0.0111
6	0.0807	0.0013	0.0028
7	0.0095	0.0035	-0.0057
…			
95	0.0252	0.0207	-0.0028
96	0.0208	0.0202	0.0208
97	-0.0001	-0.0066	0.0200

The result revealed that the parameters corresponding to the first-order neighbors appear to be more significant than second-order neighbors as expected. However, a surprising result is that while the parameter representing previous two interval (ϕ_{20}) is not significant enough to be included in the model, the t-statistics of parameter representing previous three interval (ϕ_{30}) is found to be greater than 2, implying that some dependencies exist between current traffic flow and traffic flow three intervals ago. And finally, with these parameters, each observation at a particular site i collecting at period t can be expressed as a combination of previous three and the 96th observations of its own, observations of prior two intervals of its second neighbors as well as the prediction error made at two periods ago, prediction error made yesterday at the same time of its own and a random error, which can be used to forecast future traffic flow later.

Table 4.15 shows the model residuals and forecasting errors of the fitted STARMA model estimated by RMSE and MAPE. The results reveal that except for VD No.10 and VD No.24, all of the model residuals estimated by MAPE are smaller than 20%, with a good estimation for fitting these data. Although the forecasting errors are slightly larger than model residuals, most of the errors are still below 20% and furthermore, some of them even below 10%. Therefore, we can conclude that the forecasting ability of STARMA model we fitted is quite well. Furthermore, the weighted average value, which is averaging the RMSE and MAPE values of the 24 VDs weighted by their corresponding average flow values, is also shown in the table. The weighted average MAPE is 12.71% for estimation and 15.87% for forecasting. The slight increase in the average forecasting errors is mostly contributed by the VD No.4, which has a high error in its forecasting.

Table 4.15 Estimation and forecasting errors by RMSE and MAPE (STARMA)

VD No.	Estimation errors		Forecasting error	
	RMSE	MAPE	RMSE	MAPE
1	25.11	10.50%	27.53	11.26%
2	27.72	8.42%	31.04	9.14%
3	50.08	12.25%	46.02	11.97%
4	49.20	11.92%	125.92	50.93%
5	49.54	11.46%	51.29	13.31%
6	20.69	10.08%	18.45	10.35%
7	19.01	9.89%	16.54	10.25%
8	21.91	11.21%	21.48	13.96%
9	19.45	9.54%	17.66	8.71%
10	12.84	24.85%	10.59	22.93%
11	19.41	15.31%	20.43	14.52%
12	18.99	13.82%	17.99	13.80%
13	17.58	15.48%	19.91	16.82%
14	25.04	11.24%	21.88	11.06%
15	23.88	11.63%	27.91	13.98%
16	24.67	9.98%	23.19	8.83%
17	47.05	10.09%	46.65	10.46%
18	51.59	8.61%	48.99	11.21%
19	31.31	19.40%	29.24	16.19%
20	38.57	16.41%	34.06	17.06%
21	41.31	15.16%	40.58	17.45%
22	43.22	15.53%	41.50	12.96%
23	32.73	18.92%	36.66	21.84%
24	30.69	20.57%	33.41	29.14%
Weighted Average	37.96	12.71%	47.08	15.87%

4.4 Forecasting and Comparisons

With the best fitted ARIMA and STARMA models obtained in section 4.2 and 4.3, we are now capable to compare the forecasting performance of these models. The volume data of Tuesday to Thursday from 30th June to 16th July 2009 are used for examine the fitness of the model and data of 21st July are used for the validation purpose. With the above model parameters, we forecasted 96 future observations and compared them with real observations. Then we calculated the forecasting errors by RMSE and MAPE for all of the 24 vehicle detectors as shown in Table 4.16, and if the STARMA model outperforms ARIMA model, it is written in boldface.

From Table 4.16 we can find that multivariate STARMA is superior to univariate ARIMA for 14 of the 24 detectors on both RMSE and MAPE basis. ARIMA performs better than STARMA for 7 of the 24 detectors on both RMSE and MAPE basis, whereas 3 out of 24 detectors the RMSE of STARMA is smaller but MAPE is larger than ARIMA models. If we use the average volume of each detector as the weight, then the weighted average of 24 detectors revealed that both ARIMA and STARMA are suitable for forecasting urban traffic flow, we can say that ARIMA models and STARMA model perform equally well. However, the number of parameters used in these two models is largely different. To forecast traffic flows of 24 vehicle detectors, there are four to five parameters for each detectors in the ARIMA model (noted that parameters used for each detector is different), but there are only 6 parameters in STARMA model. The number of parameters used in the model is quite important since when it comes to forecast traffic flows of real world network where the number of vehicle detectors within the system is much greater, the simpler STARMA model would be more suitable than ARIMA models in forecast traffic flow for the whole network. Therefore, the STARMA is recommended.

In overall, the forecasting ability of STARMA model is comparialble to the ARIMA models, and with the introductin of spatial parameters ϕ_{11} and ϕ_{22}. As we have stated that traffic flow is not an isolated system and will be influenced by traffic flows from other locations nearby, it is reasonable that including the spatial term in the model will enhance the forecasting ability of the fitted model. Therefore, when estimating parameters, ARIMA models operate separately whereas STARMA can simultaneously consider data from all of the 24 detectors, so their correlation can be examined from the space-time autocorrelation and space-time partial autocorrelation functions.

Table 4.16 Comparison of forecasting errors of ARIMA and STARMA models

VD No.	ARIMA		STARMA	
	RMSE	MAPE	RMSE	MAPE
1	28.26	11.47%	**27.53**	**11.26%**
2	31.82	9.27%	**31.04**	**9.14%**
3	52.29	13.13%	**46.02**	**11.97%**
4	**115.92**	**45.93%**	125.92	50.93%
5	56.45	14.41%	**51.29**	**13.31%**
6	18.45	10.36%	**18.45**	**10.35%**
7	16.68	**10.18%**	**16.54**	10.25%
8	21.84	14.50%	**21.48**	**13.96%**
9	**17.33**	**8.51%**	17.66	8.71%
10	**10.35**	**22.33%**	10.59	22.93%
11	21.69	18.43%	**20.43**	**14.52%**
12	**17.77**	**13.47%**	17.99	13.80%
13	**19.61**	**16.81%**	19.91	16.82%
14	22.32	11.20%	**21.88**	**11.06%**
15	**27.48**	**13.77%**	27.91	13.98%
16	24.74	9.28%	**23.19**	**8.83%**
17	48.55	10.81%	**46.65**	**10.46%**
18	49.61	**10.61%**	**48.99**	11.21%
19	30.37	23.58%	**29.24**	**16.19%**
20	35.11	18.35%	**34.06**	**17.06%**
21	42.10	18.25%	**40.58**	**17.45%**
22	43.95	14.14%	**41.50**	**12.96%**
23	36.71	**20.40%**	**36.66**	21.84%
24	**33.22**	**28.46%**	33.41	29.14%
Weighted Average	**46.83**	16.18%	47.08	**15.87%**

Chapter 5 Strategies of Dynamic Forecasting in ATMS

5.1 Forecasting Procedure and Performance Criteria

The estimation and forecasting techniques described in Chapter 4 assumed the application for a static system, i.e., forecasting a future condition (like a day) with historical data without using the updated data collected with time. Due to the progressing in the transportation and communication technologies in Intelligent Transportation Systems, such as vehicle detectors and global position systems, can collect traffic data in the real-time and send information from all data collecting points to the control center. The availability of these real-time data enables us to update our forecasting of traffic conditions. In this chapter, we will consider different forecasting strategies using real-time database to exhibit the forecasting performance of the STARMA model.

5.1.1 Look back and Look Ahead Procedures

Look back and look ahead span size are two issues need to be explicated when executing a forecasting. Look back span size is similar to our estimation process, and is related to data for the parameter estimation. As West and Mccracken (1998) mentioned, there are several strategies for selecting the size of database in the parameter estimation. Recursive scheme, rolling scheme and fixed scheme are primary used on the forecasting literature, whereas the results of these three strategies may act diversely. Recursive scheme estimates parameters with all available data at hand every time, so as time goes by, the look back span size, i.e. the number of observations used to forecast, will increase. Rolling scheme, on the other hand, estimates parameters based on fixed look back span size, say R, so the oldest observation will be dropped as new data adds in. The third fixed scheme is different from those two that it merely estimates the parameter using data of first R periods and does not update the parameters as new data adds in, and it is also the one we called static forecasting.

Another issue is look ahead span size, that is, how far you look forward. The most often used maximum likelihood estimation estimates parameters by minimizing the sum squares of 1-step ahead forecast errors, whereas adaptive forecasting estimates parameters by minimizing the sum squares of l-step ahead forecast errors, and $l \geq 1$ is the prediction period ahead of interest.

To illustrate how these three strategies work, consider a time series Y_T and T is the time we started to forecast, and $l \geq 1$ is the prediction period ahead of interest. First, assume we are now at time T, and we wish to look ahead l periods to forecast observation at time $T+l$. Then at time $T+1$, we wish to forecast observations at time $T+l+1$ and so on. Totally, the number of forecasting values we wish to obtain is M observations. If we label the real observation at time $T+l$ as y_{T+l}, and the predicted value at time $T+l$ as $\hat{y}_T(l)$, then the prediction period and data being used for each scheme are as follow.

1. Recursive scheme：

Step 1：Using observations y_1 to y_T to estimate parameters and predict the observation at time $T+l$, i.e. $\hat{y}_T(l) = f(y_T, y_{T-1},..., y_1)$.

Step 2：Using observations y_1 to y_{T+1} to estimate parameters and predict the observation at time y_{T+l+1}, i.e. $\hat{y}_{T+1}(l) = f(y_{T+1}, y_T,..., y_1)$.

…

Step M：Using observations y_1 to y_{T+M} to estimate parameters and predict observation at time $T+l+M-1$, i.e. $\hat{y}_{T+M-1}(l) = f(y_{T+M}, y_{T+M-1},..., y_1)$.

2. Rolling scheme：(Shown in Figure 5.1)

Step 1：Using observations y_1 to y_T to estimate parameters and predict the observation at time $T+l$, i.e. $\hat{y}_T(l) = f(y_T, y_{T-1},..., y_1)$.

Step 2：Using observations y_2 to y_{T+1} to estimate parameters and predict the observation at time y_{T+l+1}, i.e. $\hat{y}_{T+1}(l) = f(y_{T+1}, y_T,..., y_2)$.

…

Step M：Using observations y_M to y_{T+M-1} to estimate parameters and predict observation at time $T+l+M-1$, i.e. $\hat{y}_{T+M-1}(l) = f(y_{T+M-1}, y_{T+M-2},..., y_M)$.

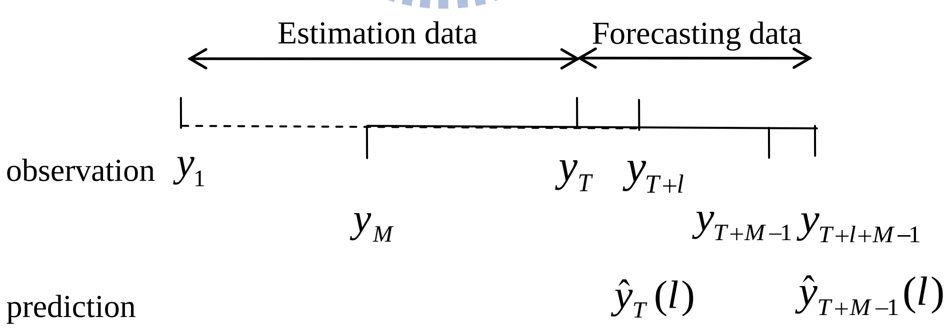


Figure 5.1 Estimation and forecasting data used in rolling scheme

3. Fixed/Static scheme：

Step 1：Using observations y_1 to y_T to estimate parameters and predict the observation at time $T+l$, i.e. $\hat{y}_T(l) = f(y_T, y_{T-1},..., y_1)$.

Step 2：Using observations y_1 to y_T to estimate parameters and predict the observation at time y_{T+l+1}, i.e. $\hat{y}_{T+1}(l) = f(y_T, y_{T-1},..., y_1)$.

…

Step M：Using observations y_1 to y_T to estimate parameters and predict the observation at time $T+l+M-1$, i.e. $\hat{y}_{T+M-1}(l) = f(y_T, y_{T-1},..., y_1)$.

Since the estimation process is a tedious work, it is reasonable to assume that parameter estimation will be updated with the database once a day or once a week, which is different from adaptive forecasting that updates parameters every time whenever new data adds in. Therefore in the following the parameters estimated for the recursive, rolling and static schemes will remain the same for a short span. In other words, we assumed the parameters used for all three schemes be the same when we forecast observations within a day. The only differences between the three schemes are the data used to fit into the model, which could be actual observations or immediate predicted data. And since the STARMA model we fitted in previous chapter revealed that each observation is related to previous 96 periods of its own and neighbor observations, recursive and rolling scheme in our study can be treated as the same scheme, thus we will only compare the forecasting performance of rolling and static schemes in our following studies.

When using rolling scheme, a new observation will add in at each iteration, hence, when $l=1$, we predict $\hat{y}_T(1)$ based on observations y_1 to y_T, i.e. $\hat{y}_T(1) = f(y_T, y_{T-1},..., y_1)$, then $\hat{y}_{T+1}(1)$ with observations y_2 to y_{T+1}, i.e. $\hat{y}_{T+1}(1) = f(y_{T+1}, y_T,..., y_2)$,…, and finally, $\hat{y}_{T+M}$ with y_M to y_{T+M-1}, i.e. $\hat{y}_{T+M-1}(1) = f(y_{T+M-1}, y_{T+M-2},..., y_M)$. This kind of forecasting is so-called 1-step ahead forecasting.

When $l=2$, a 2-step ahead forecasting is conducted. We made prediction $\hat{y}_{T+1}$ based on observations y_1 to y_T, i.e. $\hat{y}_T(1) = f(y_T, y_{T-1},..., y_1)$. However, when proceeding to predict observation at time $T+2$, since the observed value at time $T+1$ is not known yet, we based on observations y_2 to y_T as well as the prediction $\hat{y}_T(1)$, i.e. $\hat{y}_T(2) = f(\hat{y}_T(1), y_T,..., y_2)$,then $\hat{y}_{T+1}(2) = f(\hat{y}_{T+1}(1), y_{T+1},..., y_3)$ …, and finally, predict $\hat{y}_{T+M-2}(2)$ with observations y_M to y_{T+M-2} and prediction $\hat{y}_{T+M-2}(1)$, i.e. $\hat{y}_{T+M-2}(2) = f(\hat{y}_{T+M-2}(1), y_{T+M-2},..., y_M)$.

In general, the data used to predict the same forecasting period is not only determined by which scheme you select to estimate parameters, but also related to the look ahead lead time you select when forecasting, and thus influencing the forecasting performance of the model. We compared the forecasting performance of rolling scheme with 1-step and 2-step ahead and static scheme. The above discussions of observations and predictions comparisons of static and step ahead rolling scheme are summarized in Table 5.1 and Table 5.2, where the iteration number of prediction is 96 as we used in Chapter 4.

Table 5.1 Observations and prediction equations used in the static scheme

Observation	Static scheme
y_{T+1}	$\hat{y}_T(1) = f(y_T, y_{T-1} \cdots, y_{T-95})$
y_{T+2}	$\hat{y}_T(2) = f(\hat{y}_T(1), y_T, \ldots, y_{T-94})$
y_{T+3}	$\hat{y}_T(3) = f(\hat{y}_T(2), \hat{y}_T(1), y_T \ldots, y_{T-93})$
…	
y_{T+95}	$\hat{y}_T(95) = f(\hat{y}_T(94), \ldots, \hat{y}_T(1), y_T, y_{T-1})$
y_{T+96}	$\hat{y}_T(96) = f(\hat{y}_T(95), \ldots, \hat{y}_T(1), y_T)$

Table 5.2 Observations and prediction equations used in the step ahead rolling scheme

Observation	2-step ahead rolling	1-step ahead rolling
y_{T+1}	$\hat{y}_T(1) = f(y_T, y_{T-1} \cdots, y_{T-95})$	$\hat{y}_T(1) = f(y_T, y_{T-1} \cdots, y_{T-95})$
y_{T+2}	$\hat{y}_T(2) = f(\hat{y}_T(1), y_T, \ldots, y_{T-94})$	$\hat{y}_{T+1}(1) = f(y_{T+1}, y_T, \ldots, y_{T-94})$
y_{T+3}	$\hat{y}_{T+1}(2) = f(\hat{y}_{T+1}(1), y_{T+1}, \ldots, y_{T-93})$	$\hat{y}_{T+2}(1) = f(y_{T+2}, y_{T+1}, \ldots, y_{T-93})$
…		
y_{T+95}	$\hat{y}_{T+93}(2) = f(\hat{y}_{T+93}(1), y_{T+93}, \ldots, y_{T-1})$	$\hat{y}_{T+94}(1) = f(y_{T+94}, y_{T+93}, \ldots, y_{T-1})$
y_{T+96}	$\hat{y}_{T+94}(2) = f(\hat{y}_{T+94}(1), y_{T+94}, \ldots, y_T)$	$\hat{y}_{T+95}(1) = f(y_{T+95}, y_{T+94}, \ldots, y_T)$

5.1.2 Performance Criteria

To measure the performance of the fitted model, likewise, we utilized root mean squared error (RMSE) and mean absolute percentage error (MAPE) as our criteria. Since the data used in static and rolling strategies are different in that some of them are observations while some of them are predictions, the definition of MSE and MAPE are slightly different from the one presented in section 4.2. With the above definition of y_{T+s} and $\hat{y}_T(s)$ represent the real observation and predicted value at time $T+s$, respectively, the RMSE (root mean squared error) for static scheme can be defined as follow,

$$\mathrm{RMSE}_{static} = \sqrt{\frac{1}{M}\sum_{s=1}^{M}(y_{T+s} - \hat{y}_T(s))^2} \tag{5.1}$$

where $\hat{y}_T(s) = f(\hat{y}_T(s-1), \ldots, \hat{y}_T(1), y_T, \ldots, y_1)$, s is the prediction of interest and M is the iteration number of prediction.

Whereas the RMSE (root mean squared error) for rolling scheme is,

$$\mathrm{RMSE}_{rolling}(l) = \sqrt{\frac{1}{M}\sum_{s=1}^{M}(y_{T+s} - \hat{y}_{T+s-l}(l))^2} \tag{5.2}$$

where s is the prediction of interest, l is the look ahead span size in rolling scheme and M is the iteration number of prediction.

Likewise, the MAPE (mean absolute percentage error) we applied in our study for static scheme is,

$$\mathrm{MAPE}_{static} = \frac{1}{M}\sum_{s=1}^{M}\left|\frac{y_{T+s} - \hat{y}_T(s)}{y_{T+s}}\right| \times 100\% \tag{5.3}$$

Whereas MAPE for rolling scheme being used is,

$$\mathrm{MAPE}_{rolling}(l) = \frac{1}{M}\sum_{s=1}^{M}\left|\frac{y_{T+s} - \hat{y}_{T+s-l}(l)}{y_{T+s}}\right| \times 100\% \tag{5.4}$$

5.2 Comparison of Forecasting Performance

In this section, we will exhibit how STARMA model perform in forecasting traffic flow based on static and rolling strategies. Using RMSE and MAPE as criteria, Table 5.3 exhibits the forecasting errors of the STARMA models for each of the 24 vehicle detectors and their average performances for predicting with static scheme, 1-step ahead rolling scheme and 2-step ahead rolling scheme. Since the predictions of 1-step ahead rolling are derived from latest information as well as previous information, i.e. information before one period; predictions of 2-step ahead rolling are derived from information two periods ago; static strategy, on the contrary, do not update information as time goes by. So intuitively, it is anticipated that the forecasting errors estimated by 1-step ahead rolling are the smallest, and errors estimated by 2-step ahead rolling are better than static strategy.

For the forecasting errors in Table 5.3 satisfying this expectation, the estimated RMSE and MAPE are indicated in boldface. Using the average volume of each detector as the weight, a weighted average of RMSE and MAPE is also calculated, which designates the expected trend, with RMSE=47.08 and MAPE=15.87% for static strategy, RMSE=46.37 and MAPE=15.45% for 2-step ahead rolling and RMSE=44.29 and MAPE=14.80% for 1-step ahead rolling. Therefore, it proves that, 1-step ahead outperforms both 2-step ahead and static strategies, implying that using real-time information to forecast is better than merely using historical information for forecasting.

Table 5.3 Comparison of STARMA model by static, 1-step and 2-step ahead forecast

VD No.	Static		2-step ahead		1-step ahead	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
1	27.53	11.26%	27.89	11.37%	31.18	12.41%
2	31.04	9.14%	30.96	9.14%	32.25	9.66%
3	46.02	11.97%	46.71	12.03%	54.89	13.27%
4	**125.92**	**50.93%**	**120.89**	**48.59%**	**96.56**	**37.14%**
5	51.29	13.31%	51.71	13.35%	57.31	14.06%
6	**18.45**	10.35%	**18.26**	10.37%	**17.90**	9.95%
7	16.54	**10.25%**	16.35	**10.08%**	16.91	**10.03%**
8	21.48	**13.96%**	21.45	**13.81%**	22.41	**13.37%**
9	**17.66**	8.71%	**17.52**	8.60%	**16.79**	8.78%
10	10.59	22.93%	10.73	23.16%	10.91	23.57%
11	20.43	14.52%	20.80	14.72%	23.68	16.78%
12	17.99	13.80%	18.00	13.81%	18.90	13.72%
13	19.91	16.82%	20.02	16.65%	21.65	17.86%
14	21.88	11.06%	22.02	11.16%	24.10	11.88%
15	27.91	13.98%	28.07	13.99%	28.41	14.13%
16	**23.19**	**8.83%**	**22.95**	**8.73%**	**21.78**	**8.50%**
17	46.65	10.46%	46.67	10.50%	47.63	10.50%
18	**48.99**	**11.21%**	**48.88**	**11.03%**	**46.45**	**9.54%**
19	29.24	**16.19%**	29.38	**16.05%**	29.62	**15.00%**
20	34.06	17.06%	34.22	17.04%	36.09	17.60%
21	40.58	**17.45%**	41.01	**17.26%**	43.09	**16.80%**
22	41.50	**12.96%**	41.79	**12.77%**	45.69	**12.65%**
23	36.66	21.84%	36.32	18.93%	38.73	19.20%
24	33.41	**29.14%**	33.38	**25.01%**	33.50	**24.37%**
Weighted Average	**47.08**	**15.87%**	**46.37**	**15.45%**	**44.29**	**14.80%**

In particular, the error reduction of the detector No.4 is of the largest, with RMSE=125.92 and MAPE=50.93% for static strategy. RMSE=120.89 and MAPE=48.59% for 2-step ahead strategy and RMSE=96.56 and MAPE=37.14% for 1-step ahead strategy. As we have explained that the extraordinarily high forecasting error comes from the inconsistent pattern on 21st July, whose traffic flow is averagely lower than other days. Hence, the results of detector No.4 demonstrated that the superiority of rolling forecast over static forecast in cases of the estimation data set has a different pattern than the forecasting data set. In this study, to avoid large deviations of the data, we only used the traffic flows from Tuesday to Thursday to fit our model, but if we extend our model to fit not only traffic flow from Tuesday to Thursday but also Monday and Friday, or even holidays, it seems reasonable that the evidence of rolling forecast being superior to static strategy would be more clear. Further research can apply this kind of data to investigate it.

We selected the one with minimum error and the one with maximum and compare their predictions with observations on 21st July shown in Figure 5.2 and Figure 5.3. The forecasting error of detector No.9 is of the smallest, with static RMSE=17.66 and MAPE=8.71%, 2-step ahead RMSE=17.52 and MAPE=8.60% and 1-step ahead RMSE=16.79 and MAPE=8.78%, respectively. On the other hand, the predicting performance of detector No.24 is the worst among 24 detectors, with static RMSE=33.41 and MAPE=29.14%, 2-step ahead RMSE=33.38 and MAPE=25.01% and 1-step ahead RMSE=33.50 and MAPE=24.37%. Figure 5.2 and Figure 5.3 reveal that the model fits the observations quite well during off-peak when the flows are low in values and the observations were not fluctuating. But when traffic flow comes to higher level, the model may not capture the extreme peak variations in the peak hour. And since the variation of traffic flow from detector No.24 is more severe than from detector No.9, it is not surprising that the forecasting error of detector No.24 would be greater than detector No.9.

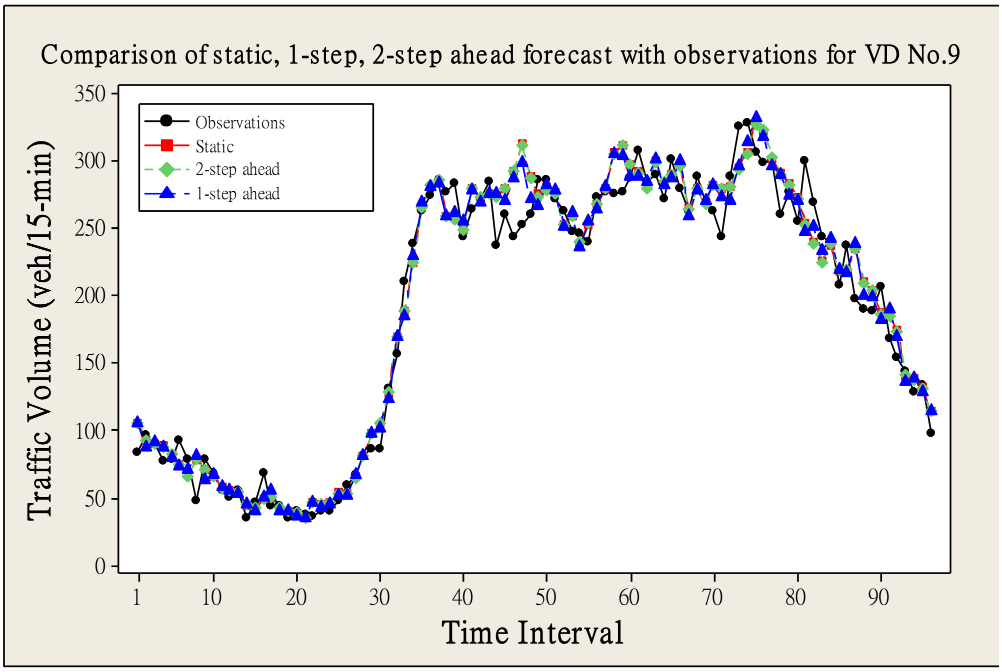


Figure 5.2 Comparison of STARMA by static and rolling forecast (VD No.9)

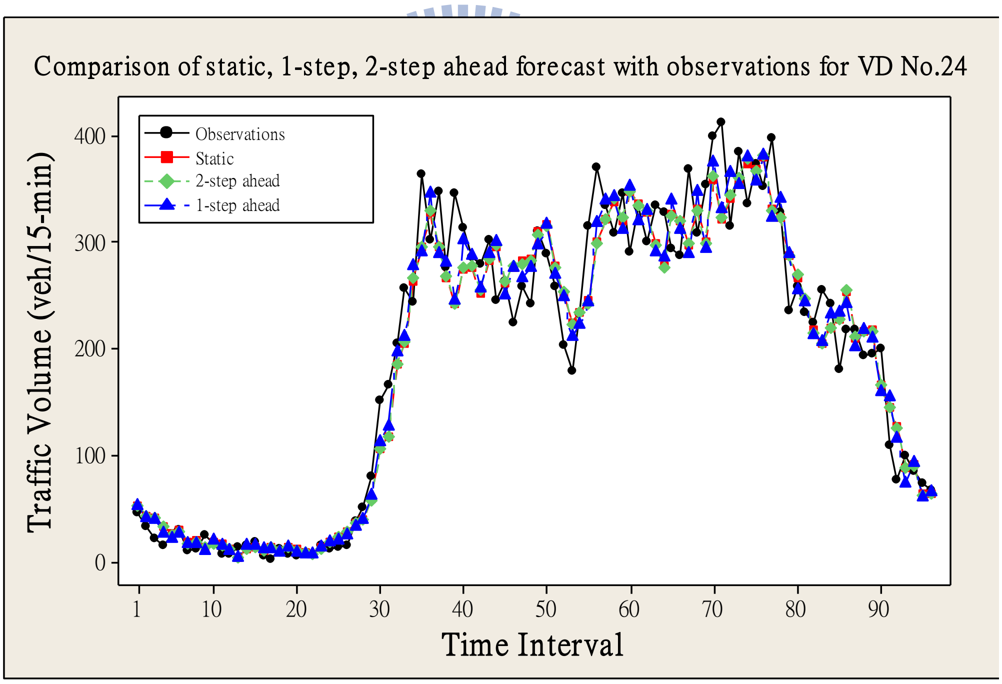


Figure 5.3 Comparison of STARMA by static and rolling forecast (VD No.24)

Chapter 6 Conclusions and Recommendations

The short-term traffic forecasting is an important issue in ATIS and ATMS, which aim at providing useful information to travelers and improving the overall efficiency of road network. In this study, urban arterials flow prediction models which forecast the short-term traffic flows based on the past traffic data measurement from a set of 24 vehicle detectors in Taipei city, Taiwan are modeled by means of seasonal autoregressive integrated moving average (ARIMA) and space-time autoregressive moving average (STARMA) model. The forecasting performance of STARMA model are examined by static, 1-step ahead rolling and 2-step ahead rolling strategies when real-time information can be obtained. The summary of the findings of this thesis and recommendations for future research directions are summarized as follows.

6.1 Conclusions

Comparing the predicted and forecasting traffic volume of the fitted ARIMA models and STARMA model with actual observations, we can find that most of the model residuals appear to be smaller than forecasting errors, but their differences were not too much. In addition, since most of the model residuals and forecasting errors estimated by MAPE are smaller than 20% for both ARIMA models and STARMA model, implying that both ARIMA and STARMA models are suitable for predicting traffic flows in an urban area.

Our results also show that, while both seasonal ARIMA models and STARMA model are able to forecast volume data at a high degree of accuracy, the number of parameters used in the ARIMA and STARMA models are largely different. In the ARIMA model, there are up to five parameters for each detector which is estimated standalone, whereas there are only 6 parameters in the STARMA model. With a large number of detector locations in the system to be forecasted, the STARMA model shows a relative simple structure as compared to the ARIMA model which is univariate in nature.

Furthermore, this result seems reasonable as traffic flows of urban area are not an isolated system and will be influenced by the flows from other adjacent locations, consequently, STARMA model considering the spatial relationship between each time series can explain the nature of the problem in a better way, and therefore improve the forecasting accuracy.

Finally we performed static and rolling strategies with the 1-step ahead and 2-step ahead span size to forecast traffic flow of all 24 vehicle detectors so as to compare the forecast performance of STARMA model when the data being used are based on real-time data or predicted values. The results revealed that weighted forecasting performance of all 24 detectors satisfy the expectation that the performance of 1-step rolling is the best, and the 2-step ahead rolling outperforms the static strategy. Therefore, it proves that using real-time information to forecast is better than merely using historical information to forecast.

6.2 Recommendations for Future Research

In this study, the traffic flow data used to fit the ARIMA models and STARMA model were from Tuesday to Thursday in order to avoid large deviations. ARIMA is suitable for forecasting time series with small variation. Therefore, if traffic volume data with larger variability such as data from Monday to Friday or even data from Monday to Sunday are employed, the forecasting ability of the STARMA model would be more impressive.

Also, the setting of the weight matrices is user-defined in the STARMA model. A common intuition is to assume that downstream flows only depend on upstream flows for uncongested conditions but not vice versa. However, the upstream traffic could be influenced by downstream congestion if long queues exist. How to define different settings of the weight matrices to incorporate this effect can be further investigated.

Furthermore, due to some limitations of the database, only traffic flow data is used to fit the prediction model. One may consider using not only the traffic flow data, but also other traffic measurements such as speed and occupancy, as the other branch of multivariate approaches.

References

BOWERMAN, B. L. & O'CONNELL, R. T. (1993) *Forecasting and Time Series: An Applied Approach*. Duxbury Press, Belmont, Callifornia.

BOX, G. E. P. & JENKINS, G. M. (1976) *Time Series Analysis: Forecasting & Control*, Prentice Hall.

DICKEY, D. A. & FULLER, W. A. (1979) Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association,* 74**,** 427-431.

DOUGHERTY, M. (1995) A review of neural networks applied to transport. *Transportation Research Part C: Emerging Technologies,* 3**,** 247-260.

FISHER, R. A. (1956) *Statistical Methods and Scientific Inference*, Macmillan Pub Co.

GHOSH, B., BASU, B. & O'MAHONY, M. (2005) Time-series modeling for forecasting vehicular traffic flow in Dublin. *85th Annual Meeting of Transportation Research Board.* Washington,D.C.

KAMARIANAKIS, Y. & PRASTACOS, P. (2003) Forecasting Traffic Flow Conditions in an Urban Network: Comparison of Multivariate and Univariate Approaches. *Transportation Research Record: Journal of the Transportation Research Board,* 1857**,** 74-84.

KAMARIANAKIS, Y. & PRASTACOS, P. (2005) Space-time modeling of traffic flow. *Computers & Geosciences,* 31**,** 119-133.

LEDOUX, C. (1997) An urban traffic flow model integrating neural networks. *Transportation Research Part C: Emerging Technologies,* 5**,** 287-300.

LEWIS, C. D. (1982) *Industrial and business forecasting methods : a practical guide to exponential smoothing and curve fitting*, London, Butterworth Scientific.

LIN, S.-L., HUANG, H.-Q., ZHU, D.-Q. & WANG, T.-Z. (2009) The application of space-time ARIMA model on traffic flow forecasting. *Machine Learning and Cybernetics, 2009 International Conference on.*

MAKRIDAKIS, S. & WHEELWRIGHT, S. C. (1978) *Interactive forecasting:univariate and multivariate methods*. San Francisco, CA：Holden-Day.

MARTIN, R. L. & OEPPEN, J. E. (1975) The Identification of Regional Forecasting Models Using Space: Time Correlation Functions. *Transactions of the Institute of British Geographers,* 66**,** 95-118.

PFEIFER, P. E. & BODILY, S. E. (1990) A test of space-time arma modelling and forecasting of hotel data. *Journal of Forecasting,* 9**,** 255-272.

PFEIFER, P. E. & DEUTSCH, S. J. (1980) A Three-Stage Iterative Procedure for Space-Time Modeling. *Technometrics,* 22**,** 35-47.

SAID, S. E. & DICKEY, D. A. (1984) Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika*, 71**,** 599-607.

SAS (2008) *SAS9.2*. SAS Institute Inc. www.sas.com

SMITH, B. L. (1995) Forecasting freeway traffic flow for intelligent transportation systems applications. Ph.D. Thesis. Department of Civil Engineering University of Virginia, Charlottesville.

TRANSPORTATION RESEARCH BOARD, N. R. C., WASHINGTON DC (2000) *Highway Capacity Manual*, Transportation Research Board, National Research Council, Washington DC.

VLAHOGIANNI, E., GOLIAS, J. C. & KARLAFTIS, M. G. (2004) Short-term traffic forecasting: Overview of objectives and methods. *Transport Reviews,* 24**,** 533-557.

VYTHOULKAS, P. C. (1993) Alternative approaches to short-term traffic forecasting for use in driver information systems. *The 12th International Symposium on Traffic Flow Theory and Transportation.* Berkley, CA.

WEST, K. D. & MCCRACKEN, M. W. (1998) Regression-Based Tests of Predictive Ability. *International Economic Review,* 39**,** 817-840.

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: Journal of the Transportation Research Board,* 1644 132-141.

YANG, Y. (2006) Spatial-Temporal Dependency of Traffic Flow and its Implications for Short - term Traffic Forecasting. PhD Thesis, Department of Urban Planning & Environmental Management University of Hong Kong.

YIN, H., WONG, S. C., XU, J. & WONG, C. K. (2002) Urban traffic flow prediction using a fuzzy-neural approach. *Transportation Research Part C: Emerging Technologies,* 10**,** 85-98.

Appendix

Appendix 1 Locations of vehicle detectors under study

VD No.	Vehicle detector	No. of Lanes	Location	GisX	GisY
1	VMTG520	2	民生西路 290 號前	300902	2772388
2	VMXH820	3	民生西路 77 號前	301579	2772475
3	VMZL960	6	民生東路三段 19 號前	303466	2772532
4	VMZLI20	6	民生東路三段 65 巷口	303828	2772526
5	VMYN820	6	民生東路三段 138 號前	304354	2772466
6	VMFIG20	2	長春路 34 號對面人行道上	302155	2772178
7	VMDL820	2	長春路 138 號前	302775	2772150
8	VMEKD00	2	長春路 198 號對面	303209	2772145
9	VMDL800	2	長春路 251 號前	303641	2772147
10	VM7FI60	2	南京西路 205 號前	300730	2772052
11	VLKGF40	2	長安西路 158 號前	301239	2771742
12	VLGGY60	2	長安西路 88 號對面人行道上	301552	2771688
13	VJTJ960	2	八德路一段 43 巷口	302797	2770994
14	VKLLH20	3	八德路二段 249 號前	303709	2771298
15	VKRM820	3	八德路二段 346 巷口	304122	2771375
16	VKWNV20	3	八德路三段 32 號對面(社教館)	304851	2771457
17	VKLGD20	4	忠孝西路一段 70 號前(消防隊)	301085	2771256
18	VKAHN20	6	忠孝東路一段 12 號對面人行道上	301861	2771100
19	VIRHZ20	4	仁愛路一段 29 號前	301927	2770381
20	VIPIZ60	4	仁愛路二段 27 號前	302385	2770358
21	VIPJA20	4	仁愛路二段 71 號旁	302752	2770339
22	VINKW20	4	仁愛路三段 47 號前槽化島上	303228	2770326
23	VINLD61	4	仁愛路三段 123 巷前槽化島上	303828	2770328
24	VINM760	4	仁愛路四段 27 巷前槽化島上	304169	2770317

簡　歷

姓名：謝亞蓁

生日：民國 74 年 12 月 6 日

地址：桃園市同安街 496 巷 1 號

電話：(03)355-4016

電子郵件：avahsieh.tem97g@g2.nctu.edu.tw

學歷：

民國 99 年 6 月　國立交通大學運輸科技與管理學系碩士班畢業

民國 97 年 6 月　國立交通大學運輸科技與管理學系畢業