Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

APPLICATION OF TIME-SERIES AND WAVELET ANALYSIS FOR TRAFFIC FLOW MODELLING AND INCIDENT DETECTION

by

BIDISHA GHOSH

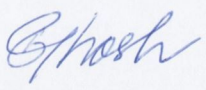
A dissertation submitted to the University of Dublin in the partial fulfilment of the requirements for the Degree of Doctor of Philosophy

Department of Civil, Structural and Environmental Engineering

October 2006

DECLARATION

I hereby declare that this dissertation, in whole or in part has not been submitted as an exercise for a degree at this or any other University. I further declare except where reference is given in the text, the work is entirely my own. I also give the library permission to lend or copy the thesis, upon request, for academic purposes.



Bidisha Ghosh

October 2006

DEDICATION

The work is dedicated to my parents, Gopa and Siddhartha, my brother Rahul and my fiancé Vikram for helping me to survive through the roller coaster ride of postgraduate research.

SUMMARY

Efficient and sustainable use of existing transportation network has become a major issue in socio-economic growth of the developed parts of today's world. Intelligent Transportation System (ITS) is an important tool to increase the sustainability and efficiency of an existing traffic management system. Implementation of ITS in an existing transport network requires the use of efficient short-term traffic condition (such as, traffic flow, density or speed) forecasting and incident detection algorithms.

In the present study, univariate and multivariate traffic flow modelling and incident detection techniques applicable to ITS equipped urban traffic control (UTC) systems are proposed. The proposed models are validated at single or multiple junctions within the congested urban transport network at the city-centre of Dublin.

In developing efficient short-term traffic flow simulation and forecasting model, initially univariate time-series techniques are applied to model the traffic volume observations at a single junction. Seasonal autoregressive integrated moving average (SARIMA) model is found to be the most rational and efficient time-series analysis technique to simulate and forecast 15 minute aggregate traffic volume at a single junction at the city-centre of Dublin. Instead of using the common maximum likelihood and/or least square method to estimate the parameters of the SARIMA model, a Bayesian estimation technique is proposed and used for the same. The forecasts from the Bayesian SARIMA model can better match the rapid variability of univariate traffic flow at an urban signal controlled arterial. To increase the flexibility and efficiency of univariate short–term the traffic flow simulation model, traffic flow theory based models are integrated with the statistical time-series traffic flow forecasting algorithms.

The short-term traffic flow simulation and forecasting algorithms are extended to a network context by using multivariate modelling techniques. Multi-input multi-output short-term traffic flow forecasting models are developed using multivariate structural time-series modelling methodology. The direction of traffic flow within a network is not considered in this model. To include the directions of traffic, the concept of merging traffic flow theory based models with statistical traffic flow forecasting algorithms is extended to a multivariate regime. The proposed multivariate short-term traffic flow forecasting

methodology can be used in simulating traffic volumes at locations on the transport network where no data is available.

Another transformation technique to analyse time series data, called the wavelet analysis technique is used to model traffic flow observations by individually focussing on traffic volume components of different time-scales. An efficient non-functional trend model and a Bayesian hierarchical model are proposed for simulating univariate traffic volume data. The univariate modelling methodology is further applied in developing an automatic incident detection technique for ITS equipped urban congested arterials.

One of the major contributions of this thesis is in introducing the concept of merging statistical time-series traffic volume forecasting models with the 'traffic flow theory' models. This concept is applied and validated comprehensively in both univariate and multivariate regime. Another significant contribution is in the introduction of a Bayesian framework in this thesis to address and alleviate the drawbacks of the conventional time-series and wavelet based forecasting models. Further, in the time-series model using wavelet analysis, a technique to capture the trend by a non-functional form has been has formulated. The Bayesian updating models (both for SARIMA and wavelet based) are able to track the changes in the traffic flow both during the peak and off-peak hours accurately, thus exhibiting promising prospects in the integration of the real time traffic modelling and management schemes for urban traffic control systems. In addition to the above mentioned contributions, a new concept of time varying variance of the traffic volume has been proposed in developing wavelet analysis based traffic flow simulation and automatic incident detection algorithms. The incident detection algorithm has been shown to perform with high accuracy and a low false alarm rate.

ACKNOWLEDGEMENT

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Biswajit Basu for his invaluable advice, encouragement and support throughout the research study. My co-supervisor Prof. Margaret O'Mahony should be thanked for introducing me to the world of transport engineering.

I would like to thank Dr. John Haslett and Dr. Sourabh Bhattacharya from the Statistics Department and Dr. W.Y. Szeto for their advice in developing certain parts of the thesis. I appreciate the help of the staff of the traffic control centre of Dublin City Council, especially Dave Traynor for allowing me to use the SCATS database for the research. Thanks to Clare, Brian, Kaushali and Vikram for helping me with the data collection. I would also like to acknowledge the help of the faculty, staff and the postgraduate students of the Department of Civil, Structural and Environmental Engineering.

Thanks to my family and friends in India and in Ireland for their constant support and encouragement. Finally thanks to Vikram for being a true companion in every sense.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ACF	Autocorrelation Function
AIC	Akaike's Information Criterion
AIDA	Automatic Incident Detection Algorithms
ANN	Artificial Neural Networks
APE	Absolute Percentage Error
ARIMA	Autoregressive Integrated Moving Average
ATIS	Advanced Traveller Information Systems
ATMS	Advanced Traffic Management Systems
CTM	Cell Transmission Model
DS	Degree of Saturation
DTA	Dynamic Traffic Assignment
DWT	Discrete Wavelet Transform
ES	Exponential Smoothing
FIFO	First-In-First-Out
FTT	Friesz Travel Time
HWES	Holt-Winters' exponential smoothing
ITS	Intelligent Transportation Systems
LWR	Lighthill, Whitham and Richards Model
MAPE	Maximum Absolute Percentage Error
MCMC	Markov Chain Monte Carlo
MRA	Multi-Resolution Analysis
MST	Multivariate Structural Time-Series
NID	Normal Independent Distribution
OD	Origin-Destination
PACF	Partial Autocorrelation Function
RBFNN	Radial Basis Function Neural Network
RMSE	Root Mean Square Error
SARIMA	Seasonal Autoregressive Integrated Moving Average
SCATS	Sydney Co-ordinated Adaptive Traffic System
SCOOT	Split Cycle and Offset Optimisation Technique
STAMP	Structural Time-Series Analyser, Modeller and Predictor
STM	Structural Time-Series Models
SUTSE	Seemingly Unrelated Time-Series Equations

UTC	Urban Traffic Control
VARMA	Vector Autoregressive Moving Average
vph	Vehicle Per Hour

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 BACKGROUND

Transportation or transit of vehicle, resources and human beings from one place to other using different modes through a network of road, air or waterways form one of the most important aspects of society. From time immemorial, there has always been a need to expand the existing transportation network to cater for the ever increasing population of the world. In the developed urban parts of today's world physical expansion of the existing transportation system can often prove difficult as it may incur several detrimental side-effects. Increased environmental pollution, fossil fuel (or other forms of energy) consumption, land take, traffic congestion and casualties are a few of the major detrimental side-effects of expansion, which can harm the social and environmental well-being of the human population. As an effect, the emphasis today is focussed towards developing balanced and sustainable transport solutions as opposed to infrastructure-intensive and capital-intensive transport strategies (McQueen and McQueen, 1999).

Intelligent Transportation Systems (ITS) is a step towards attaining sustainability by increasing the efficiency of an existing transport system. ITS aims at efficient traffic management and increased capacity within an existing network by introducing extensive and multipurpose use of advanced technologies and telecommunication systems to transport infrastructure. The two main aspects of ITS are: Advanced Traffic Management Systems (ATMS) and Advanced Traveller Information Systems (ATIS). ATMS involves network management (including incident management), second and third generation urban traffic signal control and congestion management strategies. ATIS helps to provide real-time network information (obtained from ATMS) to travellers pre-trip and during the trip in the vehicle.

Implementation of ATMS and ATIS in a network includes a cycle of processes: automatic collection of travel information from the transport network; management of the collected information using new control strategies; and finally dissemination of travel information to

the users or the travellers of the network. For proactive traffic control required in ATMS and for real-time route guidance required in ATIS, there is a need for a continuous flow of information on traffic conditions in real-time and in near (short-term) future. Short-term traffic forecasting is the process of estimating traffic conditions in future time, given current and recent past information on the same (Kaysi et al., 1993).

The work in this thesis is an effort to improve the existing short-term traffic flow forecasting and incident management methodologies using advanced time-series and wavelet analysis based techniques.

1.2 REVIEW OF LITERATURE

This section presents a brief review of the existing and past research done in the field of ATMS related to short-term traffic forecasting and incident detection. As the scope of the thesis is specific to the time-series techniques and wavelet analysis, the review focuses mostly on the literatures involving the same. In the review, literatures on short-term traffic flow forecasting and incident detection are presented in two different subsections.

1.2.1 Short-term Traffic Flow Forecasting

Short-term traffic forecasting is an important tool to follow evolution of traffic conditions over time in a transport network. This type of advanced forecasting methodologies having a time horizon of 15 minute or less (Smith et al., 2002) can provide information to support short-range operational modifications to improve the efficiency of the network at a finer scale. With the increasing need to develop more adaptive (site and time specific) traffic management systems, considerable research attention has been focussed on short-term traffic forecasting. Extensive reviews of this subject have been given by Van Arem et al. (1997) and Vlahogianni et al. (2004).

Classification of Methodologies

The well-known short-term forecasting methods can broadly be classified into univariate and multivariate approaches. The univariate approach is based on modelling traffic condition related variables (such as speed, flow or occupancy etc.) utilising observations

from any single detector, whereas developing a single model considering several sites for input and output is termed as multivariate approach. Unlike univariate models, these models are capable of capturing the temporal as well as the spatial evolution of traffic conditions over time in a transportation network. But due to ease of computation, univariate models are more common in short-term traffic forecasting literature (Kamarianakis and Prastacos, 2003).

Both the multivariate and univariate models can be technically classified into two paradigms: the empirical approaches employing fairly standard statistical methodology and/or heuristic methods for traffic flow forecasting without referring to the actual traffic dynamics and theoretical models based on traffic process theory (Van Arem et al., 1997).

In the existing literatures on empirical modelling, two separate modelling approaches are evident: parametric techniques and non-parametric techniques. The non-parametric techniques do not consider any definite functional form of the dependent and independent variables. The two main non-parametric techniques are non-parametric regression and neural networks (Vlahogianni et al., 2004). The non-parametric techniques are data driven in nature and accuracy depends partially on the quality of available data. Non-parametric regression is based on concepts of pattern recognition and chaotic systems (Davis and Nihan, 1991; Smith et al., 2002). Neural network techniques (e.g. Vythoulkas, 1993; Smith and Demetsky, 1994; Kirby et al., 1997; Lingras and Mountford, 2001; Yin et al., 2002; Lee et al., 2004; Vlahogianni et al., 2005) apply a non-linear regression framework to a potentially non-linear, non-Gaussian and non-stationary traffic condition observations (Van Arem, 1997) and give good predictions. Due to an intrinsic multi-input nature neural network models are often favoured in the space-time or multivariate models (Zhang et al., 1998). The main disadvantage of the neural network methods is their complex internal structure. The parametric techniques used in the short-term traffic forecasting literatures are mainly time-series models.

- **Time-Series Models**

The popular parametric (time-series models) techniques in short-term traffic forecasting literatures are random walk models, linear and non-linear regression, historical average algorithms, smoothing techniques (e.g. Smith and Demetsky, 1997; Williams et al., 1998) and autoregressive linear processes. Random walk or naïve prediction models are

extensively applied to present day urban traffic control systems (e.g. SCATS and SCOOT). They do not need any site specific fitting and provide sufficient accuracy for specific applications (Williams, 1999). In historical average methods (e.g. Smith and Demetsky, 1997) the single step predictions are weighted average of the past observations. The predicted traffic volume lacks any connection to the current conditions and the technique is not suitable for dynamic traffic management. The smoothing techniques (e.g. Smith and Demetsky, 1997; Williams et al., 1998) generally involve single, double or Holt-Winter's exponential smoothing techniques and these methods perform much better than the previous two models (Ghosh et al., 2005).

Of all autoregressive linear processes, the Autoregressive Integrated Moving Average (ARIMA) family of models are most successful as short-term forecasting techniques. They were first introduced by Ahmed and Cook (1979). An ARIMA (0,1,3) model was compared with the double exponential smoothing model, simple moving average model (with orders of 5, 10, and 20), and exponential smoothing model with adaptive response in the paper. The ARIMA model had better forecasting accuracy than the other three. Levin and Tsao (1980) compared the performance of two ARIMA models, where they showed ARIMA (3,1,0) performs better than ARIMA (0,1,0). Hamed et al. (1995) first applied an ARIMA model to forecast urban traffic volume. The other different variations of ARIMA models popular in the short-term traffic forecasting literatures are subset ARIMA (Lee and Fambro, 1999), ATHENA and seasonal ARIMA (SARIMA) models. ATHENA is a type of hybrid model which use a mixture of methods to construct a smaller and more efficient network (Vlahogianni et al., 2004). Van der Voort et al. (1996) developed an ATHENA model combining Kohonen self-organizing maps with ARIMA models.

Williams first introduced SARIMA models to short-term traffic flow forecasting literature (Williams et al., 1998; Williams and Hoel, 2003). The seasonal variations of the ARIMA model (e.g. Williams and Hoel, 2003 and Ghosh et al., 2005) perform better than linear regression, historical average and simple ARIMA (Chung and Rosalion, 2001) techniques. Moreover, Smith et al. (2002) has shown that the SARIMA model in situations where applicable, performs better than non-parametric regression. The major criticism towards using the ARIMA class of models is regarding their tendency to concentrate on the mean values and inability to predict the extremes (Vlahogianni et al., 2004). But the SARIMA models are multi-step in nature and once fitted to a particular site, have high accuracy, are computationally cheap and easy to implement in real scenarios.

The standard SARIMA model is parametric in nature and the estimation of the parameters can be done using classical (maximum likelihood estimates or least squares estimates) or Bayesian methods (Box and Jenkins, 1976). Box and Jenkins (1976) proposed Bayesian estimation while developing the general SARIMA model along with other classical methods. Monahan (1983) did a full Bayesian analysis of the ARIMA models and concluded that the Bayesian technique holds a distinct advantage over the classical approach as statistical analysis of ARIMA models is a certain class of non-standard problem where no classical/frequentist approach is widely accepted (Monahan, 1983). Use of *prior* (initially assumed distribution) in Bayesian estimation helps to include more information in the model, which could not be included otherwise. Unlike classical, Bayesian inference reduces statistical inference to probabilistic inference by defining a joint distribution for both the parameters and the observable data (Neal, 1993).

The use of Bayesian statistics is quite recent in the field of traffic flow modelling and forecasting. Some work has been done using Bayesian networks (using the concept of neural networks) in short-term traffic flow forecasting (Zhang et al., 2004). Hierarchical regression models with Bayesian inference are used in modelling freeway traffic flows considering the variability of parameter values throughout a day due to day-specific idiosyncrasies (Tebaldi et al., 2002). No literature is available in applying Bayesian parameter estimation methods to SARIMA models.

- **Theoretical Short-term Traffic Forecasting Models**

The short-term traffic forecasting techniques based on the traffic flow theory of the demand side or of the supply side are categorised as theoretical approaches. "Theoretical modelling can be both physical modelling of the vehicle passage variables on the supply side; and the behavioural modelling (e.g. Yang et al., 1998; Ashok and Ben-Akiva, 2000) of the trip and OD flows for the demand side" (Van Arem, 1997).

Physical models are generally based on state-space methodology considering unobserved traffic condition variables as hidden states (Whittaker et al., 1997). Physical models existing in short-term traffic forecasting literature are discussed in more detail in the next subsection. In the demand side of theoretical approaches, the behavioural models are based on different types of assignment models (like Dynamic Traffic Assignment model) which help to dynamically assign traffic flows to the paths and links of the transport network

Dynamic traffic assignment (DTA) is a well-researched field of transport modelling which has evolved extensively after the seminal work by Merchant and Nemhauser (1978a, 1978b). Existing literature in this field can be divided into two general categories: simulation based formulations (microscopic) and analytical formulations (macroscopic or mesoscopic). Due to the limited scope of discussion, the review here is focussed only on the analytical approach. The analytical approach is actually the formulation of the transport network models as optimisation models, non-linear complimentarity models or variational inequality models (Carey and Ge, 2004).

According to Heydecker and Addison (1998), the analytical macroscopic models/formulations considering instantaneous link travel time as a function of the link-flow variables (viz. link-inflow, link-volume or link-exit flow), for the whole link, are considered as the 'whole-link models'. A certain type of whole-link model is termed as linear delay function model by Nie and Zhang (2002a) in their work. This variational inequality based model, introduced by Friesz (1993), is first-in-first-out (FIFO) consistent and considers that the link travel time at any instant is a function of the link volume at that instant (linear delay function). Extensive studies (Friesz et al. 1993; Astarita 1996; Wu et al. 1998; Xu et al. 1999; Friesz et al. 2001; Carey and McCartney 2002) have been performed on this model to study the FIFO conditions and the potential applications. A comparatively different macroscopic approach is the hydrodynamic model of traffic flow on a freeway. Seminal work in this approach was done by Lighthill, Whitham (1955) and Richards (1956). A theory and formulation (LWR model) applicable to traffic behaviour on a long crowded road was presented in these papers. A first order finite difference based numerical approximation of the LWR model called the cell transmission model (CTM) was introduced by Daganzo (Daganzo, 1994 & 1995b).

As several forms of analytical macroscopic models exist in DTA literature it is important for the transport planners to find out the suitability of these models when implemented in a real transport network. Very few research work, except those by Nie and Zhang (2002b) and Carey and Ge (2003), have been carried out by researchers on the comparison of different macroscopic analytical link models. There is a need to focus on the performances of the existing analytical macroscopic DTA models when subjected to different types of variability in short-term traffic demand in a real transport network.

➢ Multivariate Short-term Traffic Forecasting Models

Most of the researches mentioned and discussed in the previous sections involve short-term univariate modelling of traffic flow observations. These models involve a single input and single output variable. Cases with multiple inputs and single or multiple outputs are termed as multivariate models in short-term traffic forecasting literatures. As in the case of univariate models, existing literature on multivariate models shows mostly two types of plausible models, i.e. the empirical (statistical parametric and neural network) models and the physical models.

ARIMAX (Williams, 2003) model is an example of multiple input single output multivariate statistical parametric models. In the paper by Williams, (2003) three upstream sites are used to predict the traffic volume at a downstream junction. This type of technique helps to model traffic flows at intersections which have multiple merges and diverges upstream of the site. The multi-input multi-output statistical parametric models are mainly multivariate variation of the existing univariate parametric statistical models, for e.g. the multivariate ARIMA model (Kamarianakis and Prastacos, 2003), space-time ARIMA model (Kamarianakis and Prastacos, 2002). These models can account for the dimension of space in a transport network. But the models are computationally demanding as the multivariate nature involves estimation of a large number of parameters. In the space-time ARIMA modelling, use of weighting matrices estimated on the basis of the causal distances among the various data collection points to introduce spatial dimension reduced the computational load to some extent. The main difficulty in the extensive use of the multivariate ARIMA models in short-term traffic forecasting technology is the lack of availability of software for estimating the parameters.

Multivariate structural time-series models based on state-space methodology were introduced as a short-term traffic forecasting technique by Stathopoulos and Karlaftis (2003). These multiple input–multiple output models do not require the stationary data unlike the other time series techniques. The multivariate structural time-series models are generally straightforward extensions of the univariate theory unlike ARIMA models. Missing observations and inclusion of exogenous variables like weather conditions or traffic flow observations of other upstream junctions can be handled comparatively easily (Harvey, 1989; West and Harrison, 1997; Durbin and Koopman, 2001). Use of the state-

space methodology in statistics being very recent there is a wide scope of application of these models in multivariate short-term traffic flow forecasting.

However, the multivariate empirical forecasting approaches ignore 1) the choice behaviour of travellers such as route and departure time choices, 2) traffic dynamics such as shockwaves, queue formation, and queue dissipation, as well as 3) dynamic traffic interactions across multiple links such as queue spillback. The other type of multivariate models, i.e. the physical models can capture traffic flow dynamics, but mainly at free flow state and cannot capture queue formation or queue spillback scenarios. These models are much inferior to empirical models in terms of capturing the cyclic nature of traffic demand. Forecasting is based on the Markov process and mostly one step ahead in nature (Whittaker et al., 1997). Hence there is a need to develop multivariate multi-step traffic forecasting model which can capture the traffic flow dynamics as well as the temporal variation of traffic conditions. This can be achieved only by combining the complex parametric empirical traffic forecasting models with theoretical traffic dynamics models.

Scope of Application

Literature on short-term traffic flow forecasting mainly concern freeways, expressways and highways, as there is a greater available variety and concentration of traffic conditions data from those roadways (Kirby et al., 1997). In the case of urban signalized arterials, the forecasting application becomes more complex and concentrates more on dynamic traffic control part of ATMS (Vlahogianni et al., 2004). Only a few literatures (Vythoulkas, 1993; Head, 1995; Stathopoulos and Karlaftis, 2003; Vlahogianni et al., 2005) are available in this field. There is a potential scope of research in applying existing freeway-based forecasting algorithms to congested metropolitan areas along with developing separate forecasting techniques for the same.

1.2.2 Traffic Flow Modelling Using Wavelet Analysis

As mentioned in the previous section, the time-series techniques are popular in traffic flow modelling and forecasting. Traffic flow data is rich in non-stationary events such as incidents/accidents, congestion are similar other phenomena which occur at different time-scales (durations). But time-series techniques are unable to perform the analysis of traffic

flow data at different time-scales. In this context, the concept of wavelet transform has been used in this thesis which can efficiently approximate a given set of traffic flow observations at different time-scales of interest. Wavelet analysis (Daubechies, 1992; Mallat, 1998) has been widely used in fields like digital signal processing, structural health monitoring, image processing etc. It is only recently that researchers in ITS have shown interest in using this technique for extracting useful information from the existing archived data related to traffic conditions.

The traffic flow time-series observations (non-stationary data) are required to be transformed to a stationary process for being modelled using conventional time-series techniques (Chatfield, 2004). The advantage of the application of wavelet analysis is in the inherent capability of accounting for non-stationary time-series.

In wavelet analysis based traffic flow modelling, multi-resolution analysis (MRA) technique (Mallat, 1989) has the potential to be used. This technique with the advantage of a fast computational algorithm has been used in feature or pattern extraction at different scales in data/image analysis and can be used as well for multi-scale forecasting of traffic flow. Incident detection is another area in ITS where the MRA techniques can be used to detect patterns corresponding to an incident from the traffic flow data.

The technique of de-noising using wavelet pre-processing has been used by a few researchers for increasing the efficiency of existing short-term traffic flow forecasting models. Sun et al. (2004) used wavelet pre-processing in combination with the local linear predictor for short-term traffic forecasting. They concluded that wavelet pre-processing increases the efficiency of the prediction method and the multi-resolution approach provides an intuitive visualization tool to better understand the ITS data. Chen et al. (2004) combined wavelet transform with Markov model to forecast traffic volume. In this paper, discrete wavelet transform was used to decompose the traffic volume observations at different resolutions and then using Markov model forecast for each level of decomposition. For the final prediction, the forecasts at all levels are combined.

In traffic pattern modelling, Qiao et al.(2003) developed an optimization process that can provide the optimized aggregation level and sampling frames of real-time data in ITS using continuous and discrete wavelet transform (Mathworks, 2000). Venkatanarayana et al. (2006) carried out another study on traffic pattern recognition using the concept of data

reduction with discrete wavelet transform. Ghosh-dastidar and Adeli (2003) developed a mesoscopic-wavelet model for simulating traffic flow patterns and extracting congestion characteristics in freeway work zone.

Due to very limited research on wavelet based traffic flow modelling in highly congested metropolitan areas, there is a scope of further exploration in this field. The model in this thesis proposes a non-analytic form of the regular traffic trend in an urban arterial which is possible to be captured using wavelet analysis of the traffic flow data (Ghosh et al., 2006). Wavelet analysis technique is comparatively more popular in internet network traffic predictions. Similar to internet network traffic, vehicular traffic also shows a complex scaling behaviour. A time-scale decomposition technique as in the paper by Mao (2005) is applied to the vehicular traffic as variations of traffic flow at different time scales are generated due to different factors.

Further applications of wavelet analysis in this thesis are in the area of incident detection and literatures in this field are described in the next section.

1.2.3 Incident Detection

The implementation of incident management systems in an urban transport network ensures efficient management of traffic by minimizing the effect of operational problems like non-recurrent congestion. Practical and reliable automatic incident detection algorithms (AIDA) are important to reduce and localize the effect of incidents. With the advent of ATMS (one important aspect of ITS), development of faster and reliable AIDA using several advanced telecommunication based techniques has received wide research attention. In the last three decades, a significant amount of research has been done on developing efficient incident detection algorithms using different approaches. Depending on the technical basis of the AID algorithms, they can be classified into a few main categories.

- **Comparative Algorithms**

Comparative algorithms compare a pair of traffic parameters from two contiguous upstream and downstream locations or it compares the traffic measurements against a

certain threshold. The most common traffic parameter used in these algorithms is the traffic occupancy. The simplest way to find out the accident occupancy pattern is to find the increases in upstream occupancy and the decreases in downstream occupancy. The most popular comparative algorithms are the California algorithm (Payne and Tignor, 1978) and the low-pass filtering algorithm (Stephanedes and Chassiakos, 1993).

- **Statistical Algorithms**

Statistical algorithms use standard statistical techniques to identify abnormal behaviour in tracking variables, such as average speed or lane occupancy. Common statistical algorithms include standard normal deviate algorithm (Dudek et al., 1974) and Bayesian algorithm (Levin and Krause, 1978). The Standard normal deviate algorithm has a high false alarm rate due to identification of recurrent congestion as incident in some cases. Bayesian algorithm has a low false alarm rate, but comparatively high detection time.

Time-series based algorithms are also a type of statistical algorithm. Time series algorithms employ time-series analysis to predict the traffic parameters. The moving average (MA) algorithm (Whitson et al., 1969), double exponential MA algorithm (Cook and Cleveland, 1974) and ARIMA algorithm (Ahmed, 1982) are examples of time-series algorithms. In the ARIMA algorithm, a 95% confidence interval for the forecasts of traffic occupancy is calculated and observations outside this prediction interval are considered as incidents. The detection rate for this algorithm is 100% in freeway conditions. The main disadvantage of this algorithm is the presence of random noise in the models which may interfere with the changes incurred by the occurrence of incidents.

- **Traffic Theory Based Algorithms**

Traffic theory based algorithms uses a variety of traffic flow theories in modelling and estimating the traffic states. McMaster algorithm (Persaud and Hall, 1989; Persaud et al., 1990; Hall et al., 1993) which depends on catastrophe theory is the most popular traffic theory based algorithm. In the McMaster algorithm the state of traffic is determined based on its location in the flow-occupancy diagram and incidents are detected based on the transition of the point from one state to another. The Discrete

State Propagation Algorithm (Guin, 2004) is another traffic theory based AIDA which uses the discrete form of macroscopic hydrodynamic model of traffic flow.

- **Advanced Computational Methodology Based Algorithms**

With the development of ITS related advanced data collection systems, certain information technology based techniques are employed in developing advanced computational methodology based AIDA. Research on three important types of advanced computational intelligence based AIDA are mentioned here.

Image Processing Based Algorithms

In ITS, video recording using cameras is often used in collecting traffic condition data. Image processing techniques are used to extract information regarding traffic parameters from these captured images in video-traffic surveillance systems. These image processing techniques can also be used to identify anomalies in traffic parameters due to incidents (Michalopoulos, 1993; Ikeda et al. 1999; Trivedi et al., 2000; Hu et al. 2004). Image processing based AIDA has a high detection rate and a low false alarm rate.

Artificial Intelligence Algorithms

Artificial intelligence incident detection algorithms use techniques like neural network, fuzzy logic etc which are essentially developed in the field of artificial intelligence. Artificial neural networks (ANN) (Ritchie and Cheu, 1993; Cheu and Ritchie, 1995; Stephenedes and Liu, 1995; Dia and Rose, 1997; Abdulhai and Ritchie, 1999; Ishak and Al-Deek, 1998, 1999), fuzzy logic (Han and May, 1990; Lin and Chang, 1998; Chang and Wang, 1995) and fuzzy neural networks (Hsiao, 1994) are the some of the artificial intelligence based techniques which have been applied to incident detection.

Wavelet Based Algorithms

Wavelet analysis based techniques have been introduced to the AIDA paradigm only at the beginning of 21st century. Samant and Adeli (2000), Adeli and Samant (2000), Adeli and Karim (2000) and Karim and Adeli (2002) used wavelet based methods for incident detection.

The first two papers jointly propose an AIDA based on discrete wavelet transform, linear discriminant analysis and neural networks. This algorithm mainly utilises the concept of recognizing incident patterns from incident-free patterns. Discrete wavelet transform and linear discriminant analysis are used for pattern extraction, de-noising and pre-processing of data (Samant and Adeli, 2000) for use in an adaptive conjugate gradient neural network model for traffic incident detection (Adeli and Samant, 2000). For incidents with duration of more than 5 minutes, this AIDA has 100% detection rate and about a 1% false alarm rate for freeways with two or three lanes. Samant and Adeli (2001) increased the efficiency of the fuzzy neural-network approach presented by Hsiao et al. (1994), by combining it with a discrete wavelet transform (Samant and Adeli, 2000). In Adeli and Karim (2000), a wavelet transform was used with a fuzzy data clustering method for feature extraction. This feature extraction was integrated with the radial basis function neural network (RBFNN) for incident detection. In these studies the wavelet transform is mainly used to filter out noise and short-term fluctuations removing the high-resolution components to reduce the high-rate of false alarm due to recurrent congestion from non-wavelet based AIDA. In the paper by Teng and Qi (2003), unlike the filter-based work, the high-resolution levels of traffic measurements (obtained from discrete wavelet transform) are employed in freeway incident detection along with the low-resolution levels.

All the wavelet based AIDA techniques are essentially applicable to freeway scenarios. To develop AIDA applicable to congested metropolitan scenario, an automatic incident detection technique applicable mainly to urban arterials is proposed in the thesis. In urban arterials, unlike the existing wavelet based freeway AID algorithms the high resolution components show more pronounced effects of non-recurrent congestion (considering, average incident time around 5 minutes) than the low-resolution components. In the algorithm proposed here a discrete wavelet filter is used to remove the coarse level from the traffic data (traffic flow and occupancy) instead of de-noising it (Samant and Adeli, 2000; Adeli and Samant, 2000; Adeli and Karim, 2000).

1.3 ORGANIZATION OF THESIS

The thesis is presented in seven chapters following this chapter.

A modelling of univariate traffic flow in a congested urban transportation network is presented in **Chapter 2**. Three different time-series models, viz. the random walk model, the Holt-Winters' exponential smoothing (HWES) technique and the seasonal autoregressive integrated moving average (SARIMA) model are used for modelling of traffic volume in the city of Dublin. Simulation and short-term forecasting of univariate traffic flow data is done using these models. The data used for modelling are obtained from inductive loop-detectors which are a part of the existing Urban Traffic Control system in Dublin. The results show that the SARIMA and the HWES technique give highly competitive forecasts and match considerably well with the traffic flow data obtained during rush hours.

The parameters of the SARIMA model are commonly estimated using classical (maximum likelihood estimate and/or least square estimate) methods. In **Chapter 3**, instead of using the classical inference a Bayesian method is employed to estimate the parameters of the SARIMA model considered for modelling. In Bayesian analysis, the Markov chain Monte Carlo simulation technique is used to solve the posterior integration problem in a high dimension. Each of the estimated parameters from the Bayesian method has a probability density function conditional on the observed traffic volumes. The forecasts from the Bayesian model can better match the traffic behaviour of extreme peaks and rapid fluctuation. Similar to the estimated parameters, each forecast has a probability density curve with the maximum probable value as the point forecast. Individual probability density curves provide a time-varying prediction interval unlike the constant prediction interval from the classical inference. The time-series data used for fitting the Bayesian SARIMA model and the classical SARIMA model are obtained from a certain intersection (TCS 183) in the city centre of Dublin.

For effective short-term simulation and forecasting of traffic flow for assignment to a real-life ITS equipped urban transport network it is necessary to integrate the traffic flow theory models with the statistical traffic flow forecasting algorithms. For this purpose, in **Chapter 4,** the behaviour of two analytical macroscopic link models; a whole-link model and a hydrodynamic model are compared in response to short-term variability of inflow loadings with time to find out a suitable theoretical traffic flow model. The outflow profiles from the two models are compared under conditions of rapidly or suddenly changing traffic inflow (signifying occurrence of recurrent or non-recurrent congestion) and inductive loop-detector observations (which were used in the two previous chapters for time-series

modelling). The solution algorithm used for the whole-link model is similar to the previous pertinent literature. But, in the case of using the analytical solution of the hydrodynamic model, instead of the regular shock wave solution, another simplistic assumption is considered which leads to the unique solution for the cases dealt with in this chapter and is also consistent with the solution using Newell's method. Under the realistic inflow conditions shown in **Chapter 4**, the two models perform differently unlike cases when the inflows are static or changing more slowly over time. Based on these inferences, the numerical approximation of the hydrodynamic model is chosen as the possible theoretical model which can be merged with the time-series forecasting algorithms to develop a new type of short-term simulation and forecasting methodology in a univariate scenario.

Chapter 5 extends the univariate modelling scenario to a multivariate regime. The time series models used for modelling are mostly univariate and hence, traffic volume forecasts can be made for only a single site at a time. There are a number of computational issues which arise in extending the univariate SARIMA time-series models to the multivariate SARIMA model with more than one site and these have been dealt with by introducing a new type of time-series models, called structural time-series models (in their multivariate form) in this chapter.

Chapter 6 extends the concept of merging the theoretical and empirical approaches of short-term traffic flow forecasting from a single link (univariate) to a transport network (multivariate). An efficient short-term space-time traffic flow forecasting strategy integrating the empirical-based SARIMA time-series forecasting technique with the theoretical-based first-order macroscopic traffic flow model - CTM is developed. In the transport network, traffic flows at the downstream intersections are simulated using CTM. The traffic flow demands used at the origin intersections of the network are the time series forecasts obtained from the univariate SARIMA time-series models trained on historical loop-detector observations at those upstream junctions. A case study in the Dublin city centre with serious congestion is performed to test the effectiveness of the forecasting strategy. The results show that the forecasts at the destination intersections only deviate around 10% from the original observations, and seem to indicate that the proposed forecasting strategy is one of the effective approaches to predict the real-time traffic flow level in a congested network. Other than that, this strategy can be utilised to simulate traffic volume at any non-critical intersections within the network where no continuous data collection is taking place.

In **Chapter 7**, a wavelet analysis method is used to model univariate traffic flow observations. The approach in this study first decomposes the historical traffic volume to an approximate part associated with low frequency fluctuations and several detail parts associated with high frequency fluctuations by means of discrete wavelet transform. The approximate part, averaged over a few days, represents the daily trend of traffic flow at the site of data collection by a non-functional form. A formulation is proposed where the residuals (after subtracting the trend from the traffic flow observations), are modelled as random Gaussian noise. But in this approach it is observed that the variance of the residual is not stable and varies with time. Hence, in a second approach the residual is modelled as a Bayesian hierarchical process, where the variability of variance is considered separately. In the next part of the chapter, the concept of time-varying variances of the residuals and trend modelling using wavelet analysis is applied in developing a practical and reliable automatic incident detection (AID) algorithm. The two-station AID methodology is developed for application to the urban arterials. Unlike the existing wavelet based freeway AID algorithms where the low resolution components show the maximum abrupt change due to incidents, in the case of urban arterials the high resolution components are more sensitive to such changes. The time-varying variance of the residuals of the past traffic measurements (traffic density and occupancy) are studied and compared with the current variance of the same in an online window. An incident is detected when the variance of the upstream occupancy and the downstream traffic flow falls outside the allowable range of variability. The traffic measurements and incidents are simulated from the same CTM based simulation strategy as described in chapter 6. Instead of using time-series forecasts (as described in chapter 6), real time traffic demand is used for the simulation. The proposed algorithm is demonstrated to be successful in detecting incidents in case of more than one lane blockings.

Chapter 8 is the concluding chapter of the thesis. A summary of the works presented in the thesis is given along with the conclusions drawn. Some recommendations about future work using the methodologies developed in the thesis are suggested in this chapter.

CHAPTER 2

UNIVARIATE SHORT-TERM TRAFFIC FLOW FORECASTING USING TIME-SERIES ANALYSIS: CLASSICAL APPROACH

2.1 INTRODUCTION

In this chapter, three conventional and non-conventional time-series techniques, the random walk model, Holt Winters' exponential smoothing model and the seasonal autoregressive integrated moving average model, are used for simulation and short-term forecasting of traffic flow observations obtained from congested urban arterials. The traffic flow data used in this purpose are collected from the embedded loop-detectors of the signalised junctions in the transportation network at the city-centre of Dublin. Suitability of the three time-series models in forecasting short-term traffic flow data is judged by comparing the error estimates like root mean square error (RMSE) or maximum absolute percentage error (MAPE) of the different models.

2.2 DESCRIPTION OF UNIVARIATE TRAFFIC FLOW DATA USED FOR MODELLING

The univariate traffic flow data used for time-series modelling are obtained from the inductive loop-detectors embedded in the streets of junction TCS 183 at the city-centre of Dublin, as a part of the urban traffic control (UTC) data collection system. A map of the junction is given in figure 2.1. It is a four-legged junction, with one-way traffic on two approaches. Tara Street (one-way) has four lanes, with traffic flowing from south to north. The traffic volume passing through Tara Street, measured on the loop-detectors numbered 1, 2, 3, 4 are continuously recorded.

The time interval of data collection is unique to the data collection system of the existing urban traffic control system of any city. The data interval can vary from few seconds to one hour. Short-term forecasting algorithms applicable to traffic management system should have a prediction horizon of 15 minute or less (Smith and Demetsky, 1997). The

univariate observations obtained collectively over each 15 minute from these four detectors are used for the modelling.

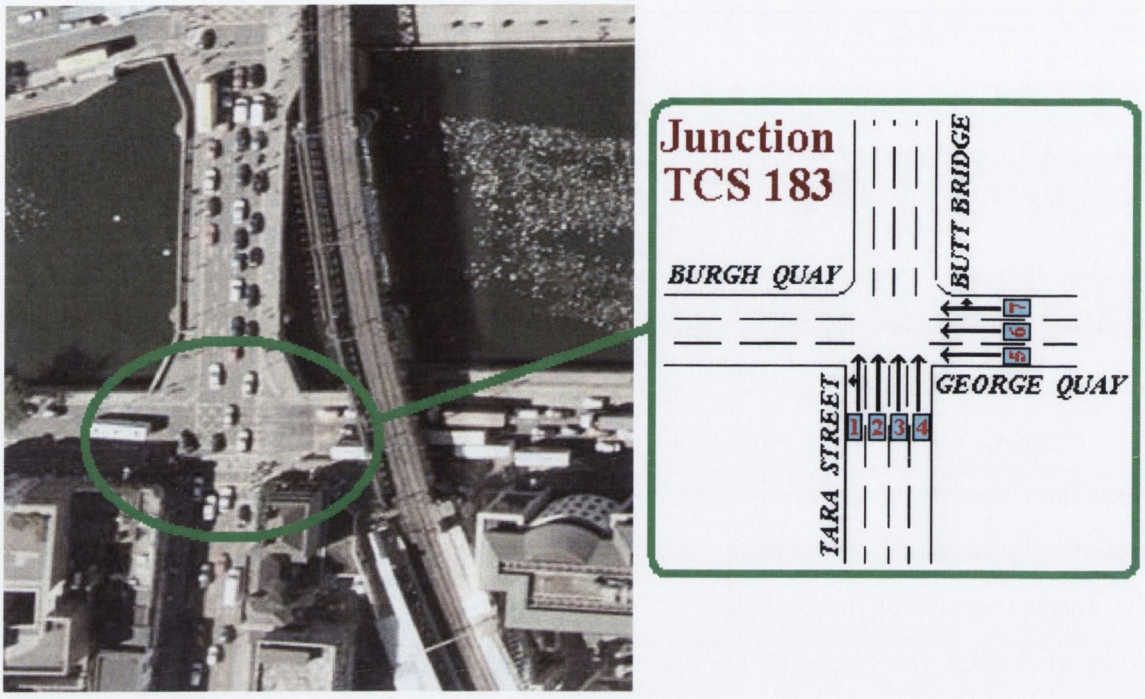


Figure 2.1 Diagram of the Junction TCS 183.

If the data collection period is over a month, two types of periodicity, viz. daily and weekly, can be observed. The weekend travel behaviour is much unlike the travel behaviour in the weekdays (figure 2.2). For developing a daily seasonal model, the traffic volume observations during the weekdays should only be considered. Williams (1999) showed that the daily models have a better accuracy than the weekly models. Hence, a daily seasonal model excluding weekend travel is opted for.

The data used for modelling was recorded from 3rd November 2003 midnight to 30th November 2003 midnight, excluding the weekends. The total number of observations is 1920, i.e., in total, 20 days of data are used. A plot of the data is shown in figure 2.3. Fifteen minutes are considered as unit time index in this plot and consequently 96 observations are plotted over each day.

The data show a bimodal nature, but the trough between the two peaks is not very distinct. The traffic volumes observed are considerably lesser in early morning hours than the

values obtained in rush hours. Abrupt changes are also quite frequent due to abundance of incidents in a busy city street.

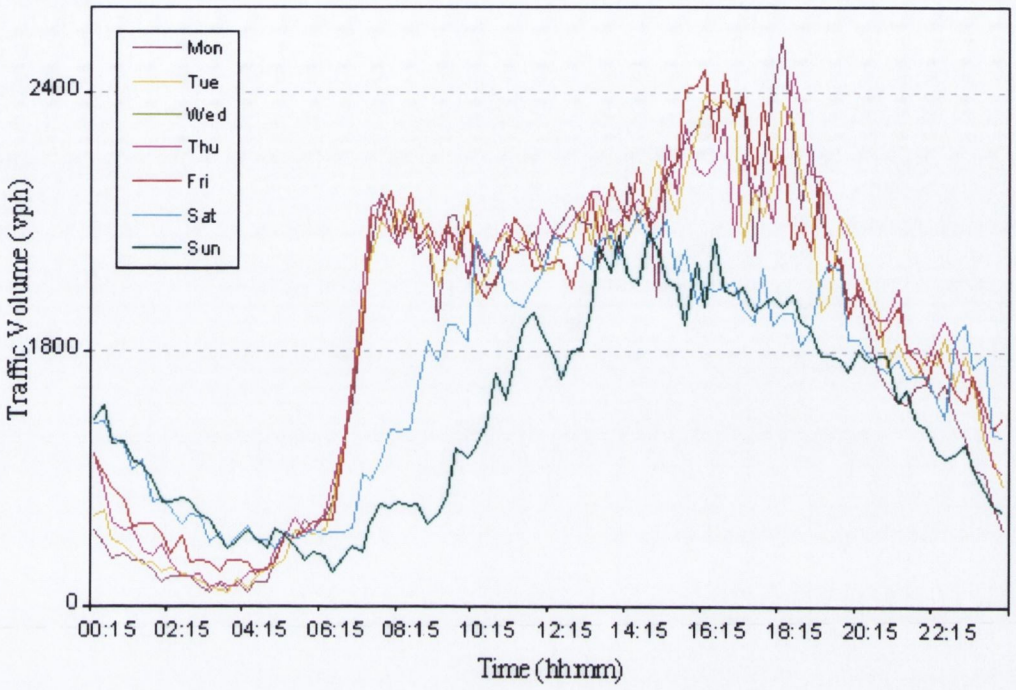


Figure 2.2 Traffic Volume at Junction TCS 183 Over Each Day in A Week.

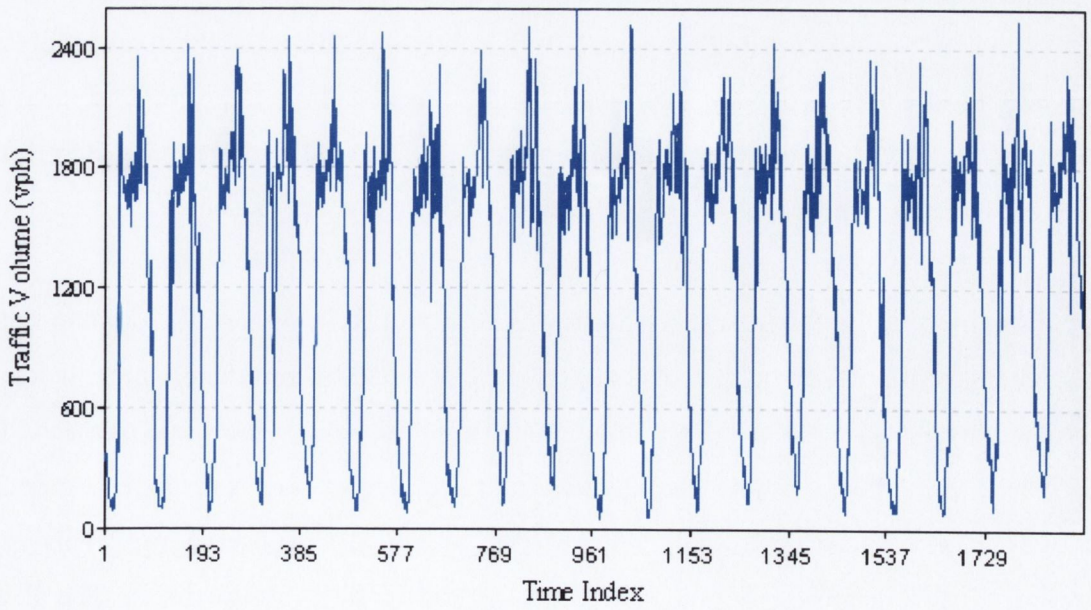


Figure 2.3 Time Series Plot of the Traffic Data.

Analysis of the Data for Time-Series Modelling

Any time-series data is required to be a *stationary* process in order to be analyzed by existing time-series models (Chatfield 2004). A time series is said to be *strictly stationary* if the joint distribution of the process $X(t_1),....X(t_k)$ is the same as the joint distribution of $X(t_{1+\tau}),....X(t_{k+\tau})$, for all $t_1,....t_k,\tau$. The joint distribution remains unaffected by shift of the time origin by an amount τ and the distribution depends only on the intervals between $t_1,....t_k$, for any value of *k*. Strict stationarity is a strong condition to be fulfilled. For weak stationarity or *second-order stationarity*, the time-series data set should have a constant mean and a constant variance. In other words, the mean *μ(t)* should be equal to a constant, μ, for all *t* and auto-covariance $\gamma(\tau)$

$$\begin{aligned}\gamma(\tau) &= E\{[X(t)-\mu][X(t+\tau)-\mu]\} \\ &= \text{Cov}\,[X(t), X(t+\tau)]\end{aligned} \tag{2.1}$$

should depend only on τ, the absolute difference between *t* and $t+\tau$ (τ is the lag at which auto-covariance is calculated). To ensure this condition, the mean is subtracted from each observation of the univariate traffic flow observation dataset over 20 days.

The graph in which sample autocorrelation coefficients [Appendix-A] are plotted against the lag *k* for *k* = 0, 1, ...P (P is generally less than one fourth of the total number of observations) is called a *correlogram*. The sample autocorrelation coefficients measures correlation between observations at different lags and provide useful descriptive information about the properties of the time-series data.

The correlogram of the zero-mean traffic flow data is plotted in figure 2.4. The two dotted lines, nearly parallel to the x-axis of the plot, denote the 95% confidence interval for the ACF (autocorrelation function) (Chatfield, 2004) [Appendix-A]. If the value of the ACF is within these lines, then it can be considered to be negligible or equal to zero. In figure 2.4, high ACF values come repetitively after a time period of 96 observations. This is an indication of the intrinsic periodic nature of the traffic flow observations. The PACF (partial autocorrelation function) [Appendix-A] of the centred traffic data (figure 2.4) is plotted as well. This plot confirms the seasonality i.e. the repetitive high correlation at lag 96 (i.e. comparison of the 1st observation with the 97th observation).

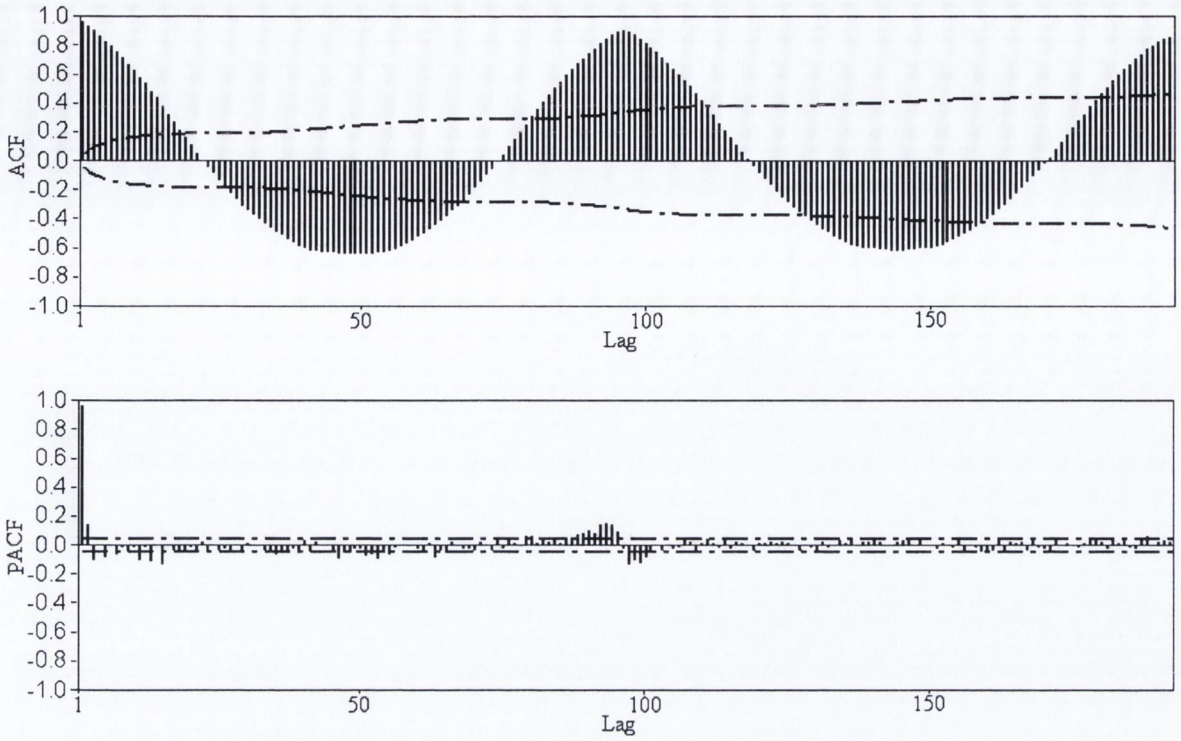


Figure 2.4 Correlogram and Partial Auto-Correlation Function Plot of the Centred Traffic Data.

To stabilise the ACF with time (required to ensure weak stationarity) a seasonal difference (over lag 96 i.e. 24 hours) is taken. The correlogram and PACF plot of the differenced and centred traffic data (figures 2.5a and 2.5b) show that the ACF or PACF values drop to zero (i.e. within the limiting lines) rapidly. Hence, the series is now stationary in nature.

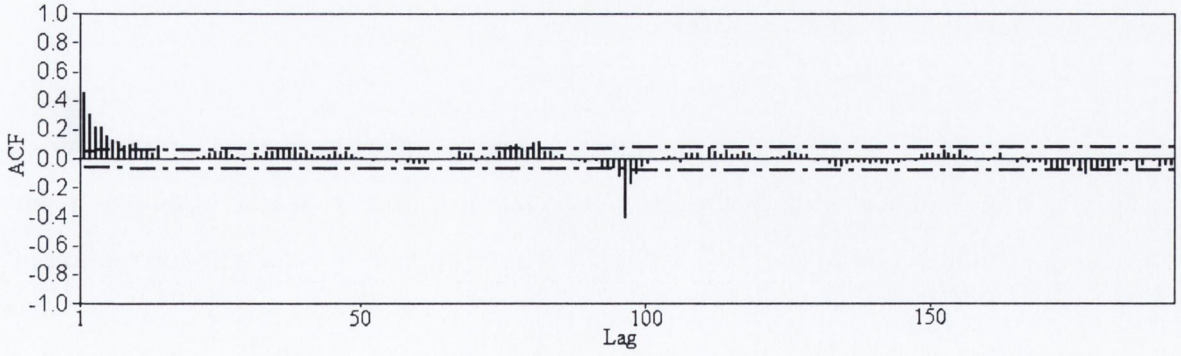


Figure 2.5a Correlogram Plot of the Centred Differenced Traffic Data.

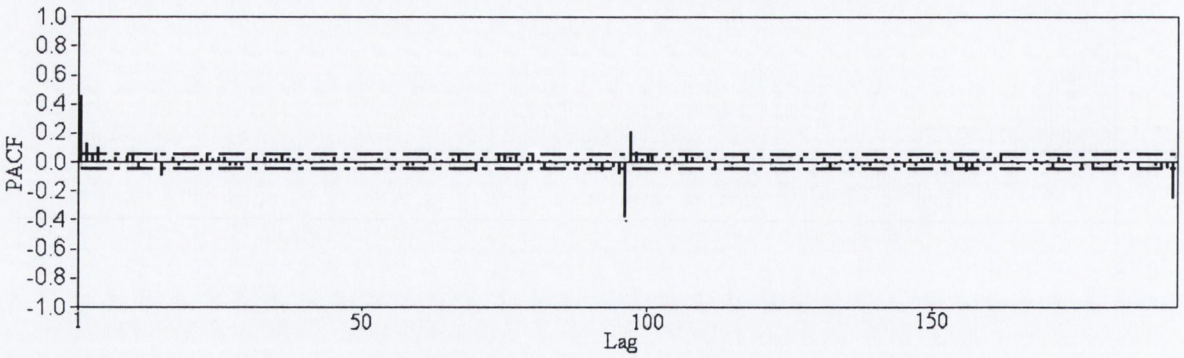


Figure 2.5b Partial Auto-Correlation Function Plot of the Centred Differenced Traffic Data.

The traffic flow data have been modelled using three approaches:

1. Random walk approach;
2. Holt-Winters' trend and seasonal smoothing technique;
3. Seasonal ARIMA model;

Out of these three approaches, the original data series is used only in Holt-Winters' trend and seasonal smoothing technique. The differenced and centred traffic flow data series is used for the other two techniques. All three models are used for forecasting and the forecasts and the relative errors from each of the models are calculated to find the most suitable modelling technique for this particular traffic data.

2.3 RANDOM WALK MODEL

The *random walk model* is one of the most widely used naïve method of time-series modelling. The random walk approach uses the most recent actual observation for forecasting the immediate future. This approach is quite popular in economic forecasting.

A linear stochastic process $\{y_t\}$ is considered as a random walk process if

$$y_t = y_{t-1} + \varepsilon_t \tag{2.2}$$

where, $\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$ and ε_t is a discrete time random process or white noise with mean zero and variance σ^2. This equation holds for a zero mean process. For observations with finite mean, the mean has to be subtracted from the data.

The random walk model does not take into account the effect of the historical observations. The model is non-stationary and can be used directly on non-stationary data. But, in this study the centred and the differenced traffic flow time-series data are used to take into account of the seasonal nature of the series. To include seasonality the following model is used.

$$y_t^* = y_t - y_{t-96} \tag{2.3}$$

Hence, $y_t^* = y_{t-1}^* + \varepsilon_t$ (2.4)

where, $\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$. The full equation of the model is,

$$y_t = y_{t-1} + y_{t-96} - y_{t-97} + \varepsilon_t \tag{2.5}$$

Here y_t is the observed data at an instant of time t; ε_t is the random error at time instant t. This model is good for one-step-ahead prediction. As the seasonal difference of the data is used for generating the forecasts, the forecasts for the next day (i.e. 96 points of future predictions) (figure 2.6) are essentially one-step ahead predictions.

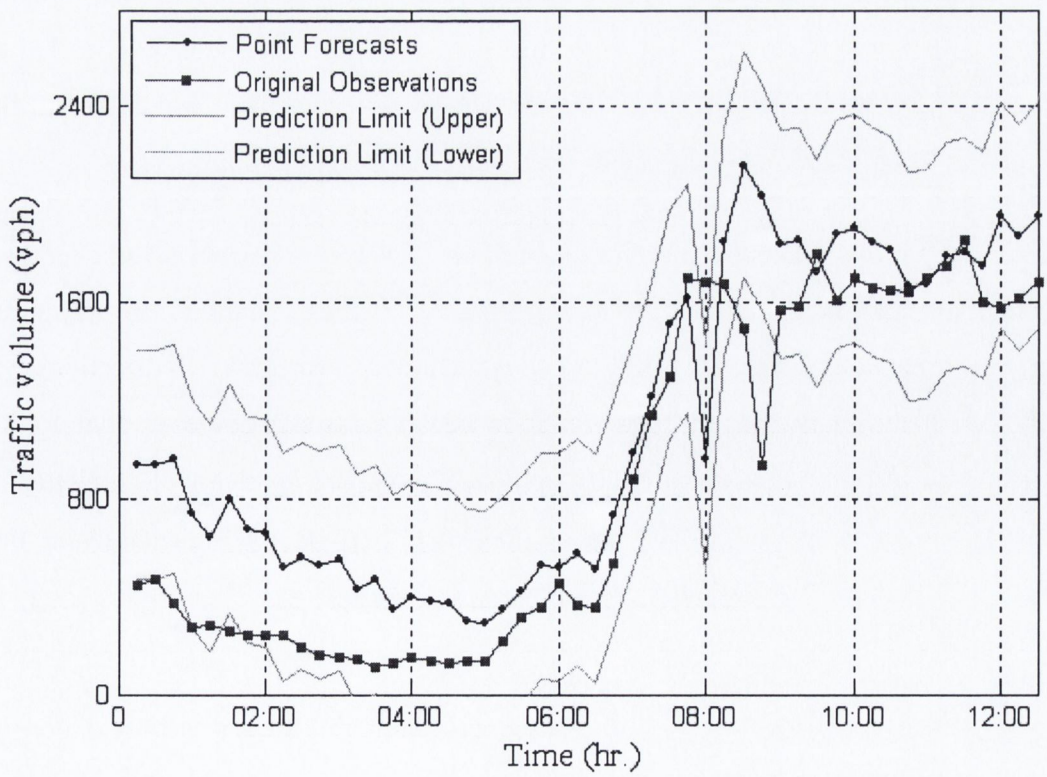


Figure 2.6 Random Walk Model Forecasts.

The point predictions along with 95% prediction interval are given in figure 2.6. Traffic volumes per 15 minute (with the time indexed, at an interval of 15 minute) are plotted here for the original and the forecasted cases. Some points of the lower limit of the confidence interval give negative values. As traffic flow observations can not be negative in reality, those negatives forecast points of the confidence interval are taken as zero. The forecasting precision from the random walk model, i.e. the RMSE and the MAPE values [Appendix-A] are given and compared with other models in table 2.2 and in section 2.6.

2.4 HOLT-WINTERS EXPONENTIAL SMOOTHING TIME-SERIES MODEL

Exponential smoothing (ES) is a weighted moving average technique, where the weights decrease exponentially as the observations get older. So, the maximum weight is given to the most recent observation. Forecasting using exponential smoothing is not based on a probability structure, and hence can be considered as an 'ad-hoc' method (Chatfield, 2001). The basic equation of simple exponential smoothing is as follows,

$$L_t = \alpha\, y_{t-1} + (1-\alpha)\, L_{t-1} \qquad 0 < \alpha \leq 1 \quad t \geq 3 \tag{2.6}$$

At any time period t, L_t stands for the smoothed observation or *exponentially weighted moving average*, and y_t stands for the original observation. The parameter α is called the *smoothing constant*. Nonlinear optimiser like the Marquardt procedure can be used to find out the value of α by minimising the mean square error (MSE) [Appendix-A].

Holt-Winters' trend and seasonal smoothing technique (HWES) is a generalized version of the exponential smoothing technique for dealing with the variations in trend and seasonality in time-series data. In HWES, the weights are reduced from two directions, viz. seasonally and historically. In this method, three equations are used to account for the changes in level, trend and seasonality. If α, β and γ are the level, trend and seasonal smoothing parameters respectively and s is the number of periods in a seasonal cycle, then the equations for the 'additive' HWES model are:

- For updating the level index L_t (Overall Smoothing),

$$L_t = \alpha \frac{y_t}{S_{t-s}} + (1-\alpha)\left(L_{t-1} + b_{t-1}\right) \tag{2.7}$$

- For updating the trend index b_t (Trend Smoothing),

$$b_t = \beta\left(L_t - L_{t-1}\right) + \left(1-\alpha\right)\left(L_{t-1} + b_{t-1}\right) \tag{2.8}$$

- For updating the seasonal index S_t (Seasonal Smoothing),

$$S_t = \gamma\frac{y_t}{L_t} + (1-\gamma)S_{t-S} \tag{2.9}$$

Using these three indices, the equation for forecast, F_{t+m} for the additive model is,

$$F_{t+m} = L_t + mb_t + S_{t-S+m} \tag{2.10}$$

where, m is number of periods ahead. The constants, α, β and γ are estimated by minimizing the MSE using any non-linear optimisation technique.

The centred and differenced traffic volume observations over a month follow similar values and patterns for each day in a repetitive fashion (figure 2.3). The data show definite additive seasonality and no discernable trend over time. Hence, HWES being particularly good in modelling data with seasonality is suitable for modelling the traffic flow data. As stationarity is not an essential requirement for analysis using HWES, original traffic flow observations over 15 minutes are used instead of the centred and differenced data.

The initial values of the smoothing parameters are taken as, $\alpha = 0.05$, $\beta = 0.02$, $\delta = 0.03$ for modelling the original traffic flow data. The optimum values of the smoothing parameters are found out by minimizing the mean absolute percentage error (MAPE). The forecasts from the model are discussed in section 2.6.

2.5 BOX-JENKINS SARIMA TIME-SERIES MODEL

Simple Autoregressive Integrated Moving Average (ARIMA) Process

In statistics, ARIMA is an important class of time series model for linear non-stationary stochastic processes. A simple ARIMA model constitutes of three parts, 'AR', i.e. the *autoregressive* part; 'I', i.e. the *differencing* part; 'MA', i.e. the *moving average* part;

In practice most time series are non-stationary in nature (e.g. traffic flow observations to be modelled in this study). *Differencing* is one of the filter/transformation techniques for removing non-stationarity from any time-series data by eliminating the trend in the dataset. It can be described as follows:

The 'first difference' $y_t^{'}$ of any time-series data is,

$$y_t^{'} = y_t - y_{t-1} \tag{2.11}$$

An AR(p) process, is a multiple regression process in which each time-series observation 'y_t' is regressed in terms of its predecessors, 'y_s', for $s < t$, by the equation,

$$y_t = \sum_{l=1}^{p} \alpha_l \, y_{t-1} + Z_t \tag{2.12}$$

where, $\alpha_1, \alpha_2, \alpha_3 ... \alpha_p$ are the coefficients of the auto regressive process of the order p.

To ensure stationarity to the AR process the parameters should satisfy the condition,

$$|\alpha_l| < 1 \tag{2.13}$$

No restrictions are required on the parameters to ensure invertibility (Box and Jenkins, 1976).

An MA(q) process is simply a finite linear filter applied to a white noise sequence $\{Z_t\}$, of the form

$$y_t = Z_t + \sum_{j=1}^{q} \beta_j \, Z_{t-j} \tag{2.14}$$

where, β_1, β_2, β_3, ... β_q are the coefficients of the moving average process of the order q.

An MA(q) process is invertible if,

$$|\beta_j| < 1 \tag{2.15}$$

No restrictions are required on the parameters to ensure stationarity (Box and Jenkins 1976).

The equation representing an ARIMA (p, d, q) model for a time-series sequence y_t (t = 1, 2...n ;) is

$$\phi(B)(1-B)^d y_t = \theta(B)Z_t \tag{2.16}$$

where, $\phi(B) = (1-\alpha_1 B - \alpha_2 B^2 - ....\alpha_p B^p)$ (2.17)

and $\theta(B) = (1-\beta_1 B - \beta_2 B^2 - ....\beta_q B^q)$ (2.18)

Z_t is a white noise sequence; B is the 'backshift operator'; In ARIMA (p,d,q), p denotes the order of the AR process, d denotes the order of differencing and q denotes the order of MA process. The importance of the ARIMA process lies in the *principle of parsimony*. Any time-series can be modelled by an ARIMA model involving fewer parameters than a pure AR or MA process (Chatfield, 2004).

***SARIMA* Process** (Box and Jenkins, 1976)

Instead of simple ARIMA, a seasonal ARIMA (SARIMA) model is used for a time series with intrinsic seasonality. There are two types of seasonal models, *additive* and *multiplicative*. Here the *multiplicative* SARIMA model $(p, d, q)(P, D, Q)_s$ is used. In the *multiplicative* model the non-seasonal part (p, d, q) and the seasonal part $(P, D, Q)_s$ are multiplied together. The equation used for the *multiplicative seasonal ARIMA model* is as follows:

$$\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^D y_t = \theta(B)\Theta(B^S)Z_t \tag{2.19}$$

where, ϕ, θ have the same significance as described in the earlier section and Φ, Θ are their seasonal counterparts, S denotes the time period of seasonality.

SARIMA Model Building (Box and Jenkins, 1976)

The centred traffic volume data are modelled using the using Box and Jenkins methodology for SARIMA class of time-series models. There are three steps in fitting a SARIMA model to any time-series data.

1. Model Identification
2. Model Estimation
3. Model Diagnostic Checking

Model Identification: The objective of model identification is to obtain some idea about the values of p, P, d, D, q and Q. ACF and PACF [Appendix-A] are two principal tools of identifying a SARIMA model.

The centred traffic flow observations are non-stationary and seasonal in nature (figure 2.4) A seasonal difference of the data is taken to account for the non-stationarity of the data, as described in section 2.2. The correlogram of the differenced data (figure 2.5) shows no significant non-stationarity. The ACF has larger values at lag 1 and at lag 96 and an exponential decay from lag 1 is very distinct. Hence, the stochastic process must be seasonal at lag 96 and AR(1) or AR(2) can be a non-seasonal component of the model. Apart from that, MA(1) seems to be a plausible seasonal component. Therefore, some variations of the SARIMA model of the nature of $(p,0,q)(P,1,Q)_{96}$ should give a suitable model for representing the traffic flow data.

Model Estimation: To estimate the parameters of the suitable SARIMA model either a frequentist (classical) or a Bayesian technique is to be used. In frequentist or classical technique maximum likelihood estimate or least square method is used (Box and Jenkins, 1976). Exact maximum likelihood estimates are often preferred to the least square method (Chatfield, 2004). The estimation procedure using a Bayesian framework is described in details in chapter 3.

A classical estimation procedure is used in estimating the parameters of the suitable SARIMA model. In table 2.1, some of the different possible variations of the SARIMA model and the error estimates for each of the models are shown. The error estimates used here are the MSE and the AIC (Akaike's Information Criterion) values. AIC is one of the most common penalized likelihood procedures (Akaike, 1974).

$$AIC = -2\log L + 2m \tag{2.20}$$

where, $m = (p+q+P+Q)$ and L denotes likelihood. For difficulties often encountered in calculating the likelihood a useful approximation is used.

$$-2\log L \approx n(1+\log(2\pi)) + n\log \sigma^2 \tag{2.21}$$

where, n is the number of observations and σ^2 is the variance of the residuals. For all the SARIMA models discussed in this study, the total number of observations (n) is the same

and hence while calculating the AIC value for the models, the first part of the likelihood (i.e. $n(1+\log(2\pi))$ from equation 2.21) is not considered. AIC is not a relative error and does not have much significance when used for comparing models from different contexts. Only when the AIC values of the different models from the same context are compared, the best fit model is the one with minimum AIC value.

SARIMA MODEL	COEFFICIENTS						MSE	AIC
	AR1	AR2	MA1	MA2	AR96	MA96		
$(1,0,1)(1,1,1)_{96}$	0.7500	0	0.3491	0	0.8348	0.0928	1441	13965.6
$(2,0,1)(0,1,1)_{96}$	1.2048	0.2649	0.7905	0	0	0.8513	1434	13963.49
$(1,0,2)(0,1,1)_{96}$	0.8667	0	0.4625	0.1186	0	0.8499	1440	13964.61

Table 2.1 Error Estimates from Different SARIMA Models

From the error estimates in table 2.1, SARIMA $(2,0,1)(0,1,1)_{96}$ seems to be the most suitable model. Both the AIC and MSE values are minimum in case of this model. However, the other SARIMA models also have quite competitive error estimates.

Model Diagnostic Checking: The residuals or errors after fitting a suitable SARIMA model to a time-series dataset are calculated by subtracting the fitted values from the observation (in this case, the centred traffic volume data). The ACF and the PACF are important tools to identify any remaining structure in the residuals. The time plot of the 'standardized residuals' (mean zero and variance one) is helpful in identifying outliers. A *portmanteau lack-of-fit test* of the residuals is an additional test of fit for the SARIMA model (Makridakis, 1998).

As less than 5% of the ACF or PACF values fall distinctly outside the limiting lines, the ACF and PACF plots (figure 2.7) show that the residuals are essentially white noise. Standardised residuals from the SARIMA model are plotted to identify the outliers (figure 2.8). The data ('standardized residuals') will be treated as outlier if it falls outside ±3. In the plot less than 2% of the data are identified as outliers. So, the SARIMA $(2,0,1)(0,1,1)_{96}$ model can be considered as a reasonably good fit. The forecasts from the SARIMA model

are discussed and compared with the forecasts from the other two models in the next section.

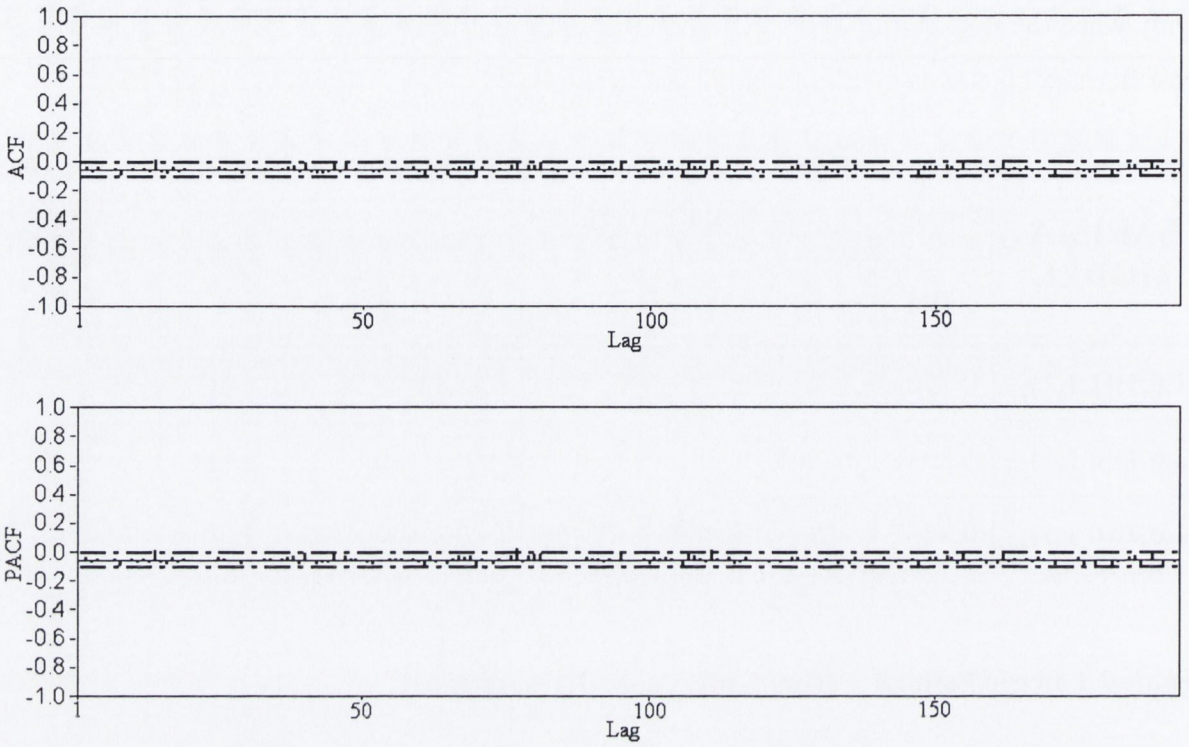


Figure 2.7 Correlogram and Partial Auto-Correlation Function Plot of the Residuals.

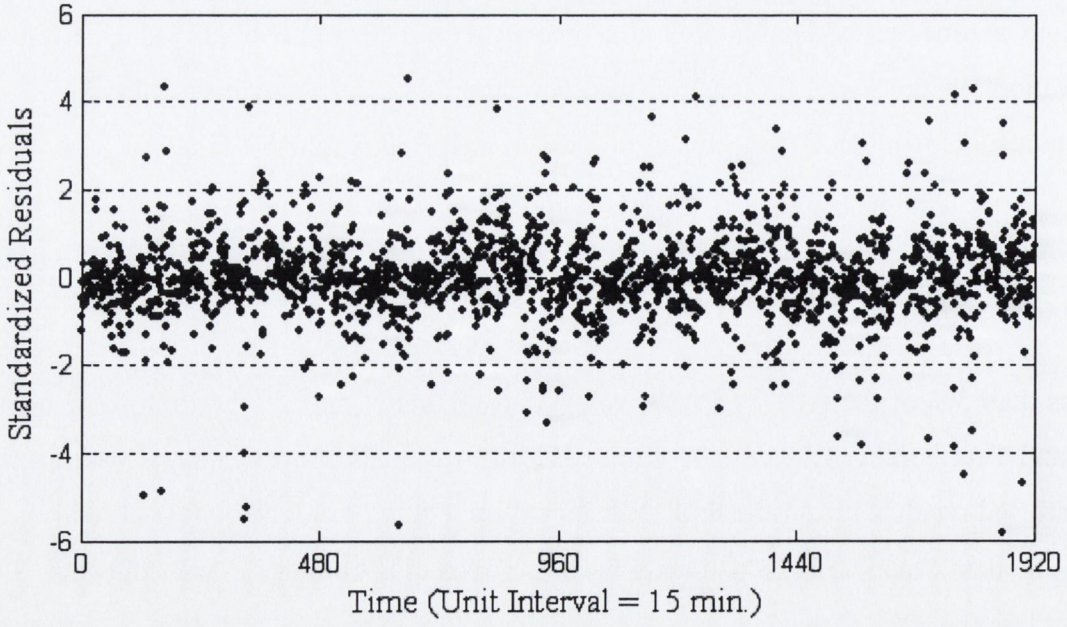


Figure 2.8 Standardized Residuals.

2.6 FORECASTS

Univariate traffic flow predictions involve forecasting future traffic flow based only on the previous observations at the same site. Forecasts from univariate traffic flow time-series models can be of two types,

1. Single interval or one-step ahead forecasts
2. Multiple interval or *k*-step ahead forecasts

Single Interval Prediction: The univariate single interval prediction (y_{t+1}) for a traffic flow time-series dataset y_t is,

$$y_{t+1} = f(y_t, y_{t-1}, y_{t-2}.....) \tag{2.22}$$

Single interval predictions are generally in a Markov chain.

Multiple Interval Prediction: The univariate multi-step prediction is,

$$y_{t+k} = f(y_t, y_{t-1}, y_{t-2}.....) \qquad k = 2,3... \tag{2.23}$$

All three methods described here generate multiple interval or multi-step predictions.

All three of the methods mentioned for time series modelling in the earlier sections in this chapter are used to model the traffic flow observations available from the four loop-detectors on the Tara Street, junction TCS 183 and 50 points in the future are predicted. The traffic flow data obtained on the 1st of December 2003, or data collected in next 12.5 hours (50x15 = 750minute = 12.5hours) are compared with these forecasts. As the random walk model is not comparable to the other two from the point of forecasting precision, hence this model is not used in the plot for comparing the forecasts. The forecasts from the HWES and the SARIMA models along with the original observations are plotted in figure 2.9.

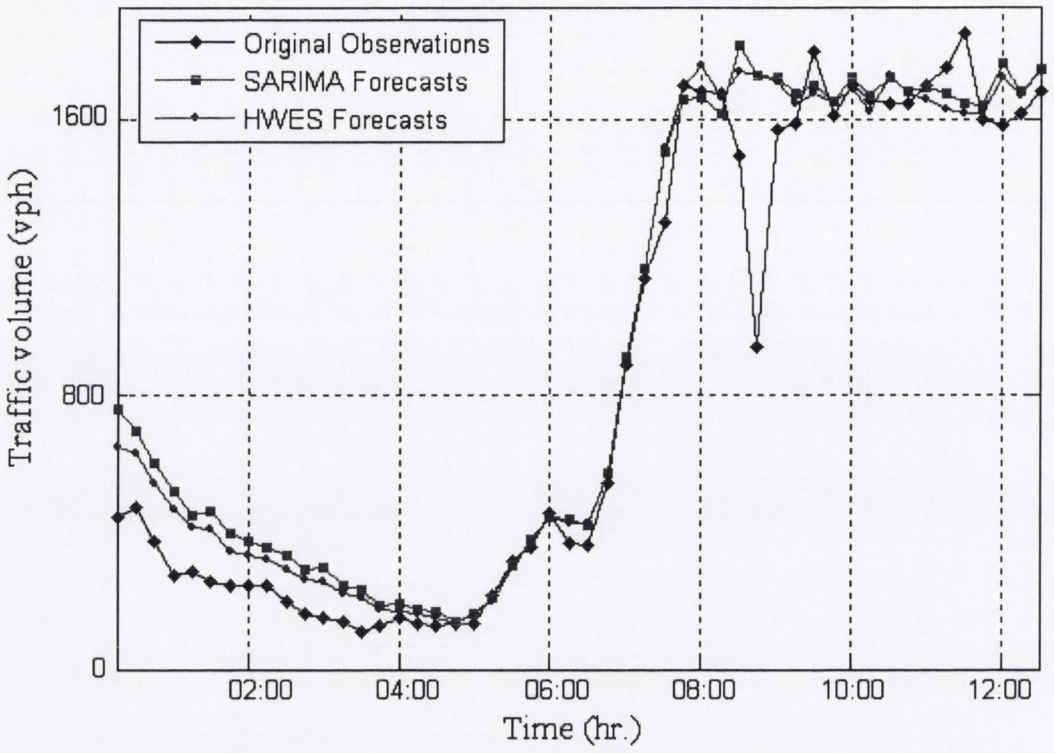


Figure 2.9 Forecasts from HWES and SARIMA Models.

In figure 2.9 only the point forecasts are shown in each case and the prediction intervals are not employed. The point forecasts from the HWES and the SARIMA models are considerably close to the original observations, whereas the forecasts from the random walk model does not match so well (figure 2.6).

The SARIMA model is used to model the centred traffic flow data. So, for comparing with the original non-centred observations the mean of the traffic volume dataset is added to all the forecasts from the model. The errors from the three different time-series models are compared in table 2.2.

MODEL	RMSE	MAPE
Random Walk Model	85.82	85.2%
Exponential Smoothing	37.8	22.11%
Seasonal ARIMA $(2,0,1)(0,1,1)_{96}$	41.89	28.55%

Table 2.2 Comparison of the Time Series Models

The RMSE and MAPE values show that the HWES model is slightly better than the SARIMA model. This inference is unlike the results when the same techniques are applied for modelling the traffic flow in urban freeways (Williams, 1999). The error estimates obtained, show quite high values of MAPE and RMSE. It is quite evident from figure 2.9, that the first few points of forecasts from both the HWES and the SARIMA model do not match well with the observed traffic flow data.

If the data series is so adjusted that the forecasts are made from 6 a.m. instead of 12 a.m., then the error estimates from both the HWES and the SARIMA model decrease considerably. The forecasts from the HWES and the SARIMA model along with the original observations, for 6 a.m. to 6 p.m. on 1st of December 2003, are plotted in figure 2.10. The error estimates for 48 points (i.e. 12 hours) of forecasts, starting from 6 a.m. on 1st of December 2003, are given in table 2.3.

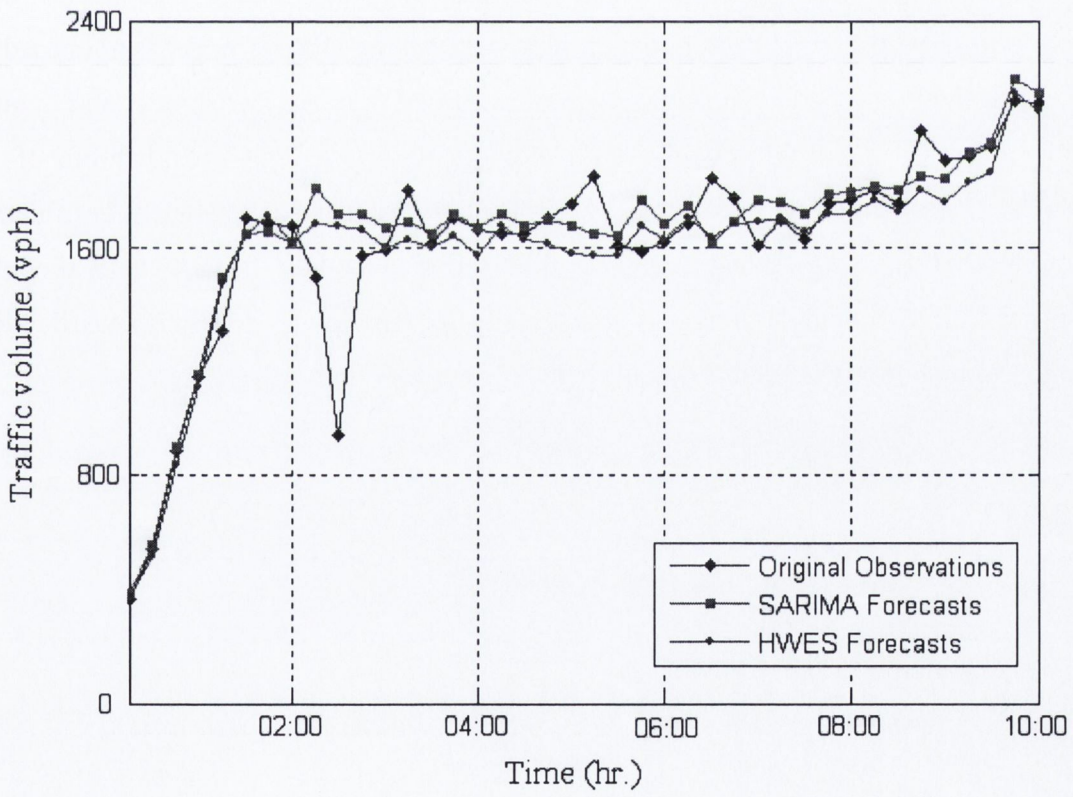


Figure 2.10 Forecasts from Both Methods at Rush Hours.

MODEL	RMSE	MAPE
Exponential Smoothing	61.47	11.22%
Seasonal ARIMA $(2,0,1)(0,1,1)_{96}$	61.93	11.28%

Table 2.3 Comparison of Two Time Series Models in the Rush Hours

Another reason for the high values of RMSE and MAPE is sudden changes in the observed traffic flow data at some points of prediction. None of the models perform well in these regions. On eliminating those points the error estimates from both the models, decrease further (MAPE approximately equal to 8%).

Based on the forecasting precision, both the HWES and the SARIMA model can be considered as suitable time-series techniques for univariate short-term traffic flow prediction for an urban signalized transport network. But the HWES model is not based on the classical theory of statistical stochastic processes like the SARIMA model and is often considered as an *ad-hoc* (Chatfield, 2001) method of forecasting. Consideration of the probability theory gives SARIMA model a robust theoretical basis. Hence, considering both the theoretical and practical aspects of time-series modelling, the SARIMA model is chosen as the most suitable time-series technique for univariate short-term traffic flow prediction in favour of the HWES model.

CHAPTER 3

UNIVARIATE SHORT-TERM TRAFFIC FLOW FORECASTING USING TIME-SERIES ANALYSIS: BAYESIAN APPROACH

3.1 INTRODUCTION

In chapter 2, the SARIMA model was chosen among three time-series models used for simulation and short-term forecasting of univariate traffic flow observations from congested urban arterials. The standard SARIMA model is parametric in nature and the estimation of the parameters can be done using classical (maximum likelihood estimates or least squares estimates) or Bayesian methods (Box and Jenkins 1976). In this chapter, the SARIMA model parameters are estimated using both the classical (maximum likelihood estimate) and the Bayesian methods. The Bayesian estimation of the model parameters is performed using an innovative simulation procedure. Qualitative and quantitative comparisons of the Bayesian methods with the classical methods are shown. A comparison between the forecasts from the two approaches is presented and the additional information obtained from the Bayesian models in the context of traffic management is discussed (Ghosh et al., 2006a).

3.2 FITTING SARIMA MODEL TO TRAFFIC FLOW DATA

To develop a Bayesian method of estimation of the parameters of a traffic flow time-series model, univariate traffic flow data from the loop-detectors 1, 2, 3, 4 of approach 1 in junction TCS 183 (figure 2.1) in the city-centre of Dublin are fitted to a suitable SARIMA time-series model. The data used for the modelling were recorded from 6 a.m. on 2nd September 2004 to 6 a.m. on 30th September 2004. A plot of the data over the first week of September 2004 is shown in figure 3.1. Fifteen minutes are considered as unit time index in this plot and consequently 96 observations are plotted over each day. The traffic flow observations follow a regular pattern in general. But, certain high peaks (e.g. evening peak on 5/09/04) and unusual drops (e.g. late evening hours of 8/9/05 and hours just before evening peak of 9/09/04) introduces rapid variability which is difficult to model using the classical SARIMA model described in chapter 2. The weekend travel behaviour is much unlike the travel behaviour in the weekdays (for details see section 2.2). So, for developing

a daily seasonal model the traffic flow observations only during the weekdays are considered. The total number of observations excluding weekends is 1920, i.e., in total, 20 days of data are used.

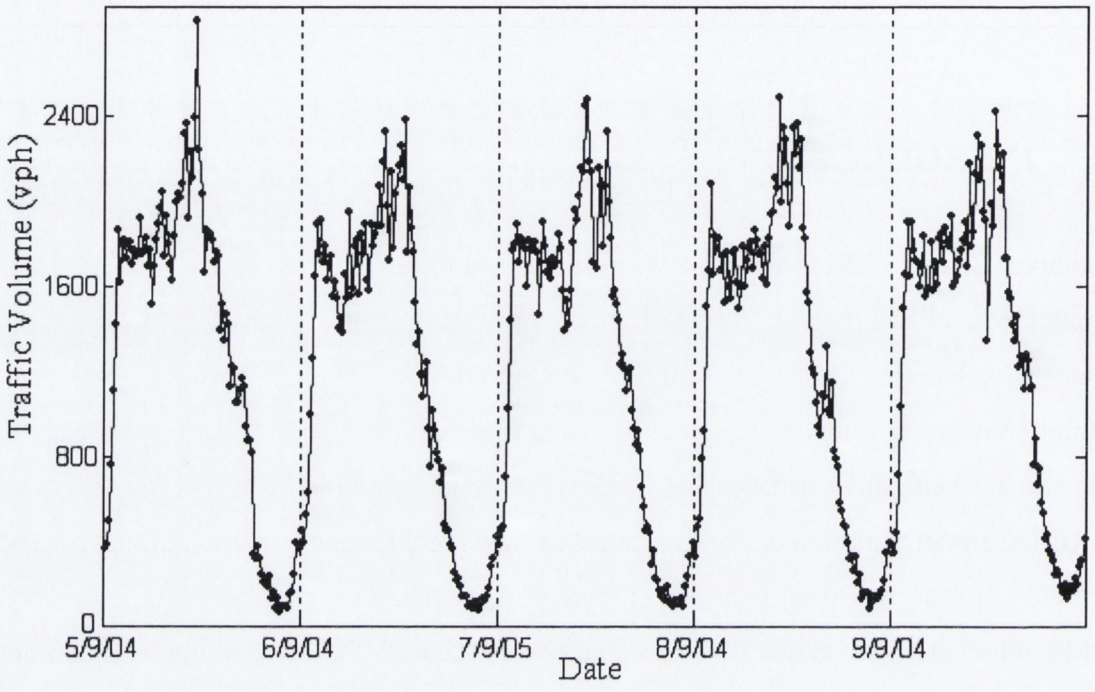


Figure 3.1 Time Series Plot of the Traffic Data over a Week.

For weak stationarity, the mean of the traffic flow dataset is subtracted from each observation and a seasonal difference (over lag 96 i.e. 24 hours) is taken. A correlogram of the differenced data is plotted in figure 3.2a. The correlogram shows that the ACF of the differenced dataset has larger values at lag 1 and at lag 96.

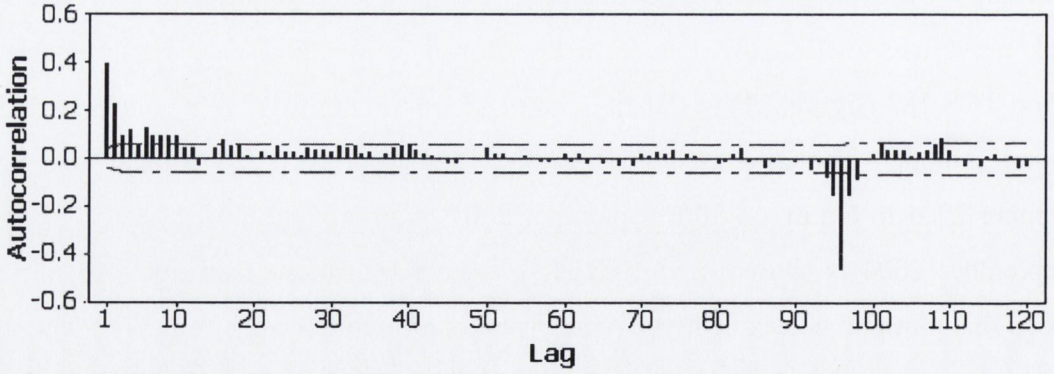


Figure 3.2a Correlogram Plot of the Centered Differenced Traffic Data.

The PACF (partial autocorrelation function) of the centred traffic data (figure 3.2b) is plotted as well. This plot shows the possibility of presence of a seasonal moving average component in the model.

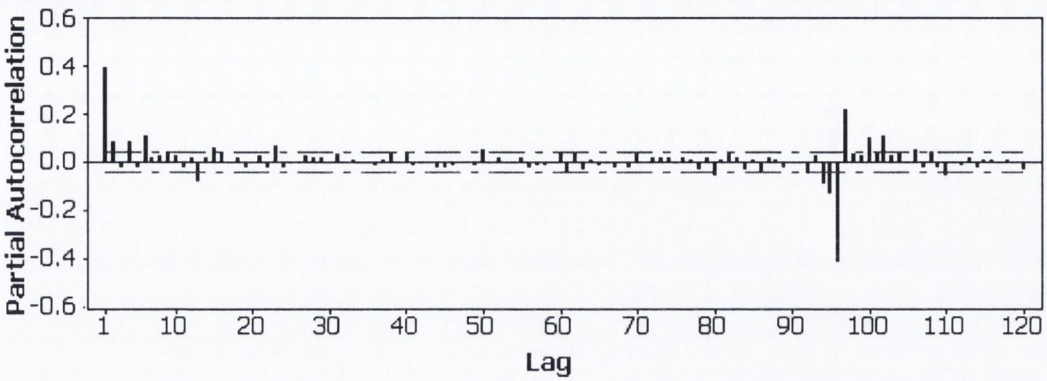


Figure 3.2b Partial Auto-Correlation Function Plot of the Centered Differenced Traffic Data.

On calculating the error estimates (Table 2.1), it can be concluded that a SARIMA $(2,0,0)(0,1,1)_{96}$ (Ghosh et al., 2005) is the best fit model for junction TCS 183. But, there are some other SARIMA models also, which though not the best fit, perform very well compared with the best fit. All of these SARIMA models can capture the daily repetitive nature of traffic flow and the dependence of present traffic conditions on immediate past. SARIMA $(1,0,0)(0,1,1)_{96}$ model, (which has error estimates varying less than 1% from the best fit model) is fitted to the traffic flow observations data obtained from junction TCS 183. This model is chosen instead of the best fit model to be computationally parsimonious. However, it may be vital that if required for specific situation, this model can be easily extended by addition of a non-seasonal MA part to be more versatile without much additional computation.

The equation of SARIMA $(1,0,0)(0,1,1)_{96}$ model used for modeling the centered traffic flow observations in this chapter is,

$$\left(1-\phi_1 B\right)\left(1-B^{96}\right)Y_t = \left(1-\Theta_1 B^{96}\right)\varepsilon_t \tag{3.1}$$

which leads to $Y_t = \phi_1 Y_{t-1} - Y_{t-96} - \phi_1 Y_{t-97} + \varepsilon_t - \Theta_1 \varepsilon_{t-96}$ (3.2)

where, $\varepsilon_t \sim N(0,\sigma^2)$. The unknown parameters to be estimated are ϕ_1, Θ_1 and σ and is represented by a vector $\xi^T = (\phi_1, \Theta_1, \sigma^2)$. To estimate ξ, either a frequentist (classical) or a Bayesian technique is to be used. In frequentist or classical technique maximum likelihood estimate or least square method is used to estimate the vector ξ (Box and Jenkins 1976). Exact maximum likelihood estimates are often preferred to least square method (Chatfield 2004).

3.3 PARAMETER ESTIMATION OF SARIMA MODEL USING BAYESIAN METHOD

In the Bayesian approach, all unknown quantities are considered as random variables and uncertainties over those quantities are represented using probability distributions conditional to the available data. While estimating any parameter using classical/frequentist methods, sampling distribution of the parameter is mostly assumed as normal or Gaussian. This approach is quite crude in the sense that in real situations the sampling distributions of the parameters can be different from normal. With Bayesian analysis, realistic approximations to the sampling distribution are considered and the inferences are reached using generic techniques and the observed data.

General Example of Bayesian Inference

The basic principle behind Bayesian statistics is as follows. Some prior ideas about any parameter or data set can be obtained from prolonged and detailed observations or by comparing with similar conditions. If these prior ideas are considered as 'prior beliefs' in any hypotheses regarding those situations or events, then conditional to the observed data (likelihood) an exact solution or 'posterior' distribution of the parameter can be obtained (Lee 1997).

The general idea can be explained with the help of a small example. The unknown parameter, the vector θ, is to be estimated using a set of n observations of any event y, which depend on the value of θ in a known way.

From Bayes' theorem it follows that-

$$\Pi(y \mid \theta) = \frac{\Pi(y,\theta)}{\Pi(\theta)} = \frac{\Pi(y)\Pi(\theta \mid y)}{\Pi(\theta)} \tag{3.3a}$$

or, $\Pi(y \mid \theta) \propto \Pi(y)\Pi(\theta \mid y)$ (3.3b)

where, the proportionality constant is,

$$1/\Pi(\theta) = 1/\left[\int \Pi(y)\Pi(\theta \mid y)dy\right] \text{ (in continuous case)} \tag{3.3c}$$

This leads to the basis of Bayesian analysis,

$$(\textbf{\textit{posterior}})\Pi(\theta \mid y) \propto \Pi(\theta)(\textbf{\textit{prior}}).L(\theta \mid y)(\textbf{\textit{likelihood}}) \tag{3.3d}$$

The likelihood is the part which changes the 'prior' (*a priori* belief) in the light of the actual data to the real distribution i.e. the 'posterior'. Depending on the data diagnostics, the likelihood function is chosen. More explicitly, depending on data diagnostics a distribution is assumed or chosen for the available observations. This assumption may not always match with the real distribution.

From the posteriors, the marginal distributions of the different parameters are determined. Using the posteriors, a 'predictive distribution' (distribution of the predictions of y, i.e. y_{t+1}, y_{t+2}..... etc. in future) for y can be found out. The predictive distribution takes into account the uncertainty in the values of the unknown parameters and the residual uncertainty about y when these parameters are known. The predictive distribution is,

$$\Pi\left(y_{n+1} \mid y_{1,2,3.....n}\right) \propto \int \Pi\left(y_{n+1} \mid \theta, y_{1,2,3.....n}\right)\Pi\left(\theta \mid y_{1,2,3.....n}\right)d\theta \tag{3.3e}$$

Therefore, $\Pi\left(y_{n+1} \mid y_{1,2,3.....n}\right) = \int L\left(\theta \mid y_{1,2,3.....n+1}\right)\Pi\left(\theta \mid y_{1,2,3.....n}\right)d\theta$ (3.3f)

Using the values of y, obtained from the predictive distribution, the suitability of the likelihood and priors chosen is decided.

The Likelihood Function (Box and Jenkins 1976)

For both the frequentist (maximum likelihood estimation) and the Bayesian methods for the estimation of the vector ξ, the likelihood function is required to be calculated. Considering the centred traffic flow data diagnostics (mean and quantile-quantile plot) the demand data for each day when considered over a few weeks seem to be normally distributed. If $\boldsymbol{Y}$ is the vector of n observations of the centred traffic flow time-series data (Y_t), then the normal probability distribution $p(\boldsymbol{Y} \mid \xi)$ of the dataset, depends on the vector ξ. The appropriate function to contemplate the various values of the vector ξ which has given rise to the fixed set of observations of vector $\boldsymbol{Y}$ is the likelihood function $L(\xi \mid \boldsymbol{Y})$. The likelihood function $L(\xi \mid \boldsymbol{Y})$ is of the same form as $p(\boldsymbol{Y} \mid \xi)$, only $\boldsymbol{Y}$ is fixed ξ is the variable.

$$L(\xi \mid \boldsymbol{Y}) = \sigma^{-n} f(\xi) \exp\left(-\frac{S(\xi)}{2\sigma^2}\right) \tag{3.3}$$

where, $S(\xi) = \sum_{t=-\infty}^{n} [\varepsilon_t \mid \boldsymbol{Y}, \xi]^2$

In equation 3.3, $f(\xi)$ is a function of the vector ξ which is to be calculated based on inverse of the $n \times n$ covariance matrix of the Y_t's. To ease the computational complexity for calculating $f(\xi)$ the history of the observed process and of the errors are incorporated as in Ravishankar and Ray (1996). In this study, the history of observations ($Y_{t-1}, Y_{t-2,} \ldots\ldots\ldots Y_{t-97}$) are available from the SCATS database of past observations and the history of the unknown errors ($\varepsilon_{t-1}, \varepsilon_{t-2,} \ldots\ldots\ldots \varepsilon_{t-96}$) (since the model parameters are unknown) is assumed to follow a similar normal distribution as the error ε_t. The multivariate (n variate) normal distribution of the SARIMA process can be represented as

$$\prod(Y_1, Y_2, Y_3 \ldots\ldots\ldots Y_n \mid \xi) = \prod(Y_1 \mid Y_2, Y_3 \ldots\ldots\ldots Y_n, \xi) \prod(Y_2 \mid Y_3 \ldots\ldots\ldots Y_n, \xi) \ldots\ldots\ldots \prod(Y_n \mid \xi) \tag{3.4}$$

In equation 3.2, the moving average part is assumed to be represented by,

$$e_t = \varepsilon_t - \Theta_1 \varepsilon_{t-96} \tag{3.5}$$

The observations are assumed to have the following multivariate normal distribution

$$\prod(Y_1, Y_2, Y_3 \ldots\ldots Y_n) \sim \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{3.6}$$

where, the mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2 \ldots\ldots \mu_n)^T$ and $\Sigma_{n\times n}$ is the covariance matrix of Y_t.

The mean values of the distribution are

$$\mu_t = \phi_1 Y_{t-1} + Y_{t-96} - \phi_1 Y_{t-97} \qquad t = 1, 2, \ldots\ldots n; \tag{3.6a}$$

From Ravishankar and Ray (1996), the likelihood function is

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid Y_{1,2,3\ldots\ldots n}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{t=1}^{n}(Y_t - \mu_t)^T \boldsymbol{\Sigma}^{-1}(Y_t - \mu_t)\right\} \tag{3.7}$$

When the number of observations is greater than 50 the matrix $\Sigma_{n\times n}$ can be simplified to a scalar form (Box, Jenkins, 1976). Equation 3.7 when applied to the SARIMA model for the present study approximates to (considering the simplified form of the matrix $\Sigma_{n\times n}$ as $\tau = \sigma^2 + \Theta_1{}^2\sigma^2$),

$$L(\mu, \tau \mid Y_{1,2,3\ldots\ldots n}) = (2\pi\tau)^{-n/2} \exp\left(-\frac{1}{2}\tau^{-1}\sum_{t=1}^{n}(Y_t - \mu_t)^2\right) \tag{3.8}$$

$$L(\phi_1, \Theta_1, \sigma^2 \mid Y_{1,2,3\ldots\ldots n}) \propto \sigma^{-n}\left(1 + \Theta_1{}^2\right)^{-n/2} \exp\left\{-0.5\sigma^{-2}\left(1 + \Theta_1{}^2\right)^{-1}\left(\mathrm{A} + \phi_1 \mathrm{B} + \phi_1^2 \mathrm{C}\right)\right\} \tag{3.9}$$

where, $\mathrm{A} = \sum_{t=1}^{n}(Y_t - Y_{t-96})^2$ (3.9a)

$$\mathrm{B} = -2\sum_{t=1}^{n}(Y_t - Y_{t-96})(Y_{t-1} - Y_{t-97}) \tag{3.9b}$$

$$\mathrm{C} = \sum_{t=1}^{n}(Y_{t-1} - Y_{t-97})^2 \tag{3.9c}$$

For the priors, there are some restrictions on the range of ϕ_1, Θ_1. The invertibility region for the model, required by the condition that the roots of $(1 - \phi_1 B)$=0 should lie outside the unit circle is defined by the inequality, -1< ϕ_1< 1. The condition, $|\Theta_1|$ <1 is required to ensure stationarity (Box and Jenkins 1976). The priors considered here for ϕ_1, Θ_1 are uniform priors which give

$$\Pi(\phi_1) = 1/2 \text{ (considering uniform prior from -1 to 1)} \tag{3.10}$$

$$\Pi(\Theta_1) = 1/2 \text{ (considering uniform prior from -1 to 1)} \tag{3.11}$$

and, considering the non-informative prior for σ^2, gives

$$\Pi(\sigma^2) \propto 1/\sigma^2 \ (\sigma^2 \text{ is always positive}). \tag{3.12}$$

Hence, the posterior density is

$$\Pi(\phi_1, \Theta_1, \sigma^2 \mid Y_{1,2,3.....n}) \propto \Pi(\phi_1)\Pi(\Theta_1)\Pi(\sigma^2) L(\phi_1, \Theta_1, \sigma^2 \mid Y_{1,2,3.....n}) \tag{3.13}$$

which yields,

$$\Pi(\xi \mid Y_{1,2,3.....n}) \propto \sigma^{-n-2}\left(1+\Theta_1^2\right)^{-n/2} \exp\left\{-0.5\sigma^{-2}\left(1+\Theta_1^2\right)^{-1}\left(A+\phi_1 B+\phi_1^2 C\right)\right\} \tag{3.14}$$

By integrating the other unknown parameters except for the one whose distribution is to be estimated, the 'marginal (posterior) distributions' of the each of the unknown parameters

$[\Pi(\sigma|\boldsymbol{Y}, \Theta_1, \phi_1), \Pi(\Theta_1|\boldsymbol{Y}, \sigma, \phi_1), \Pi(\phi_1|\boldsymbol{Y}, , \sigma, \Theta_1)]$ can be found out.

To obtain the marginal distributions of each unknown parameter from the posterior distribution, the other two unknowns are to be integrated out. Hence, to find the marginal distribution of the three unknown parameters the following integrations are to be performed.

$$\Pi(\phi_1 \mid \boldsymbol{Y}) \propto \int_{-1}^{1}\int_{0}^{\infty} \sigma^{-n-2}\left(1+\Theta_1^2\right)^{-n/2} \exp\left\{-0.5\sigma^{-2}\left(1+\Theta_1^2\right)^{-1}\left(A+\phi_1 B+\phi_1^2 C\right)\right\} d\sigma d\Theta_1 \tag{3.15a}$$

$$\Pi(\Theta_1 \mid \boldsymbol{Y}) \propto \int_0^{\infty}\int_{-1}^{1} \sigma^{-n-2}\left(1+\Theta_1^{\,2}\right)^{-n/2} \exp\left\{-0.5\sigma^{-2}\left(1+\Theta_1^{\,2}\right)^{-1}\left(\mathrm{A}+\phi_1\mathrm{B}+\phi_1^2\mathrm{C}\right)\right\} d\phi_1 d\sigma^2 \qquad (3.15b)$$

$$\Pi(\sigma^2 \mid \boldsymbol{Y}) \propto \int_{-1}^{1}\int_{-1}^{1} \sigma^{-n-2}\left(1+\Theta_1^{\,2}\right)^{-n/2} \exp\left\{-0.5\sigma^{-2}\left(1+\Theta_1^{\,2}\right)^{-1}\left(\mathrm{A}+\phi_1\mathrm{B}+\phi_1^2\mathrm{C}\right)\right\} d\phi_1 d\Theta_1 \qquad (3.15c)$$

From the posterior distribution again the 'predictive distribution' can be obtained using the relationship,

$$\Pi(Y_{n+1} \mid Y_1, Y_2, Y_3 \ldots\ldots Y_n) = \int L\left(\xi \mid Y_{1,2,3\ldots\ldots n}\right)\Pi\left(\xi \mid Y_{1,2,3\ldots\ldots n}\right) d\xi \qquad (3.15d)$$

Simulation of Distributions

The determination of the posterior distributions often involves integration of complex integrals (mostly integrations of equation set 3.15) with high dimensionality. Numerical integrations are often performed to compute the distributions for which the analytical solution is intractable. However, for large models numerical integration may lead to too many approximations and may even become intractable. In many high dimensional cases of Bayesian analysis, certain refinements of Monte Carlo integration methods are often used (Carlin,1996).

There are different non-iterative and iterative variations of these refinements. *Markov Chain Monte Carlo* (MCMC) is the particular iterative variation of Monte Carlo method in which the simulated values are not in *iid* but are in a Markov chain. In summary, the goal of the MCMC is, given a target distribution $\Pi(x)$, a Markov chain $\{x_n\}$ is required to be constructed whose limiting distribution is $\Pi(x)$.There are two popular MCMC algorithms,

(i) Gibbs Sampler (Geman and Geman, 1984) and,

(ii) Metropolis-Hastings Algorithm (Metropolis et al., 1953 and Hastings, 1970)

Gibbs Sampler

The Gibbs sampler is a multivariate extension of chained data augmentation algorithm (Tanner, 1996). To simulate the probability distributions for each of the parameters θ_1, θ_2..... θ_d in an unknown vector $\boldsymbol{\theta}$ (θ_1, θ_2..... θ_d), given an initial condition ($\theta_1^{(0)}$, $\theta_2^{(0)}$..... $\theta_d^{(0)}$)the algorithm iterates the following steps:

1. Sample $\theta_1^{(i+1)}$ from $p\left(\theta_1 \middle| \theta_2^{(i)}, \ldots\ldots \theta_d^{(i)}, \boldsymbol{Y}\right)$ (3.16a)

2. Sample $\theta_2^{(i+1)}$ from $p\left(\theta_2 \middle| \theta_1^{(i)}, \theta_3^{(i)}, \ldots\ldots \theta_d^{(i)}, \boldsymbol{Y}\right)$ (3.16b)

   .
   .
   .
   .

3. Sample $\theta_d^{(i+1)}$ from $p\left(\theta_d \middle| \theta_1^{(i)}, \theta_2^{(i)}, \ldots\ldots \theta_{d-1}^{(i)}, Y\right)$ (3.16c)

The vectors ($\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\theta}^{(1)}$..... $\boldsymbol{\theta}^{(t)}$) are a realization of a Markov chain. The joint distribution of ($\theta_1^{(i)}$, $\theta_2^{(i)}$..... $\theta_d^{(i)}$) converges to $p\left(\theta_1^{(i)}, \theta_2^{(i)}, \ldots\ldots \theta_d^{(i)} \middle| \boldsymbol{Y}\right)$.

Unlike Metropolis-Hastings algorithm, the Gibbs Sampler requires the simulation from the full conditional distributions (equations 3.15 a, b & c in this chapter). This kind of sampling is possible only in the case of standard distributions. As the full conditionals in equations 16a, b and c are non-standard in nature, the Metropolis-Hastings algorithm technique is chosen in this paper to simulate the marginal distributions from the posterior distributions.

Metropolis-Hastings Algorithm

Being an MCMC technique Metropolis-Hastings algorithm (Metropolis et al., 1953, Hastings, 1970) tries to generate a Markov chain $x^{(0)}$, $x^{(1)}$,...having given the 'target distribution', $\Pi(x)$, as stationary distribution. Assume, the target distribution $\Pi(x)$ has a density function $\pi(x)$ known up to the normalizing constant, denoted by κ. Considering some initial value $x^{(t)}$, a proposed simulation x^{sim} is generated from a candidate generating density $q(x^{(t)}, x^{sim})$ and then accepted with a probability $\alpha(x^{(t)}, x^{sim})$, where,

$$\alpha(u,v)=\begin{cases}\min\left\{\dfrac{\pi(v)q(v,u)}{\pi(u)q(u,v)},1\right\} & \text{if } \pi(u)q(u,v)>0\\ 1 & \text{if } \pi(u)q(u,v)=0\end{cases} \qquad (3.17a)$$

The criteria for acceptance is that the value of $\alpha\,(x^{(t)}, x^{sim})$ should be greater than a random number generated from standard normal distribution. If the proposed value is accepted then, $x^{(t+1)} = x^{sim}$; if not accepted then, $x^{(t+1)} = x^{(t)}$. By continuous iterations (until equilibrium or convergence is reached) a series of dependent realizations, forming a Markov chain, are simulated with $\Pi(x)$ as the equilibrium distribution (Tanner, 1996). The normalizing constant, κ is cancelled in the ratio (eq. 3.17a); hence knowledge of κ is not needed to implement the algorithm. This makes the algorithm generic in approach. The choice of candidate generating density influences the simulation performance greatly. Depending on the performance of the Metropolis-Hastings algorithm, the stationary/the equilibrium density are accepted as approximation for the real marginal densities.

The candidate values of the parameter ϕ_1 is to follow a normal random walk as the proposal distribution is taken as a normal centered at the current value with a standard deviation equal to 0.05. Considering this marginal distribution 'ϕ_1' as the 'target distribution' and using an arbitrary starting value of ϕ_1 (say, $\phi_{1\text{-}in}$), a simulated value of ϕ_1 (say, $\phi_{1\text{-}sim}$) is obtained with $\phi_{1\text{-}sim} \sim N(\phi_{1\text{-}in}, \tau_1)$ .The simulated value of $\phi_{1\text{-}sim}$ is accepted following the Metropolis-Hastings' algorithm. As a symmetric transition function (normal distribution) is assumed, only the Metropolis algorithm is sufficient. According to this algorithm, each simulated value of ϕ_1 is accepted with a probability

$$\alpha(\phi_1)=\frac{\Pi\left(\phi_{1\_sim}\mid\Theta_1,\sigma^2,y_{1,2,3.....n}\right)}{\Pi\left(\phi_{1\_in}\mid\Theta_1,\sigma^2,y_{1,2,3.....n}\right)} \text{ or 1 which is minimum} \qquad (3.17b)$$

with the same acceptance criteria as previously mentioned. For the next simulated value of ϕ_1 instead of the arbitrarily assumed value of $\phi_{1\text{-}in}$, $\phi_{1\text{-}sim}$ is used and a new $\phi_{1\text{-}sim}$ is simulated using the Markov chain (normal random walk). The new $\phi_{1\text{-}sim}$ is accepted using the same method as before. In this way, 10000 values of the parameter ϕ_1 are simulated (figure 3.3). Two parallel chains of simulation are shown in the figure. The chains start from different initial values, one from 0.1 and another from 0.8. Both the chains seem to converge towards the same value.

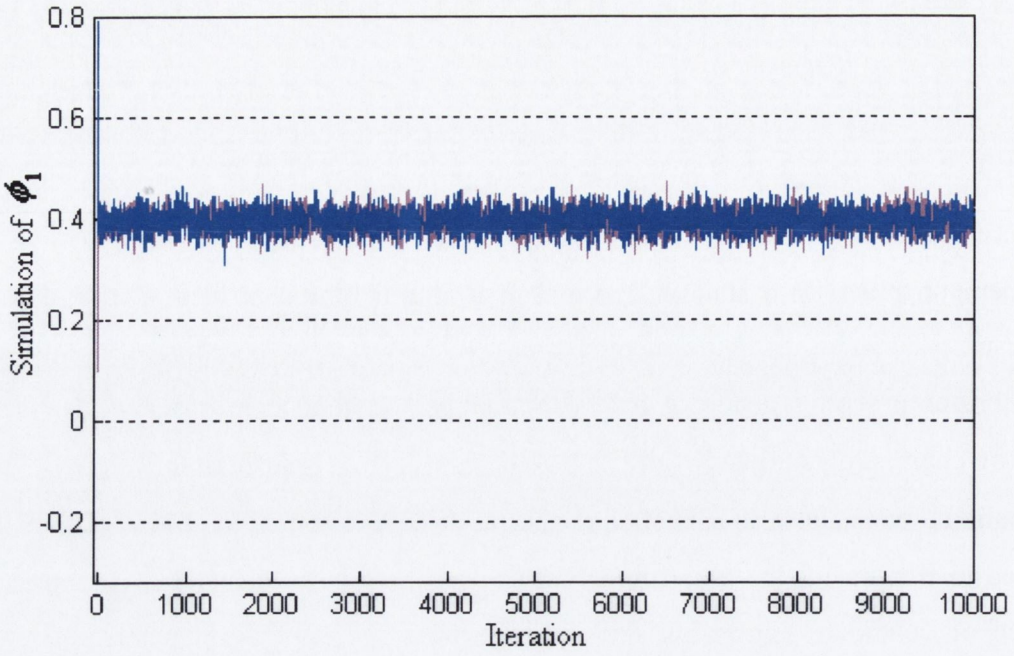


Figure 3.3 Simulation of Values of ϕ_1.

The same type of technique is applied to the other two parameters Θ_1 and σ^2. Each of them is separately simulated from a normal 'proposal' distribution centred at the current value and with standard deviations equal to 0.01 and 0.5 respectively for Θ_1 and σ^2 (figures 3.4a and 3.4b).

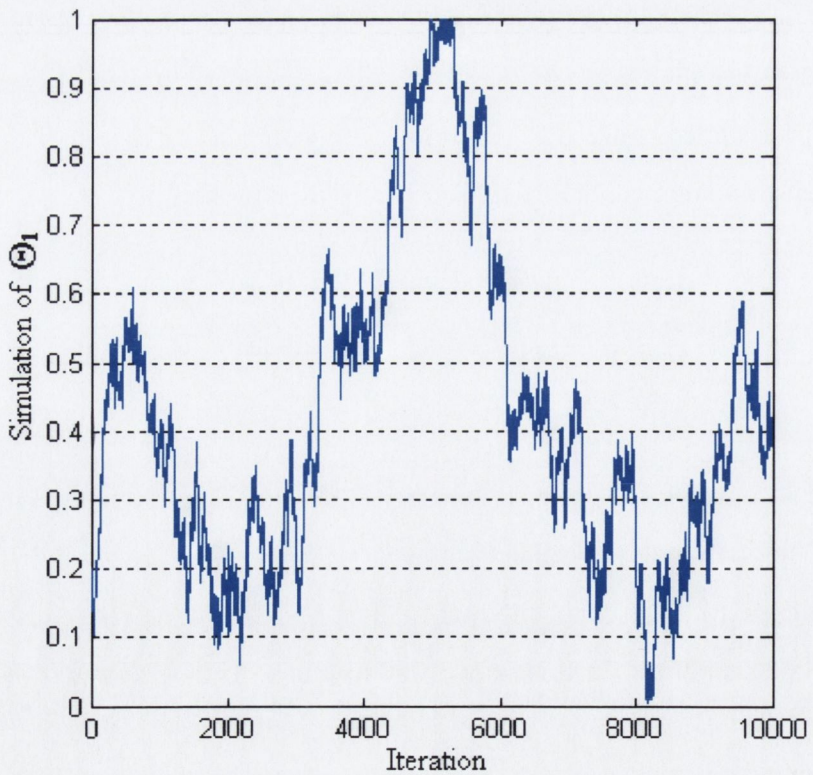


Figure 3.4a Simulation of the Values of Θ_1 from Individual Normal Distribution.

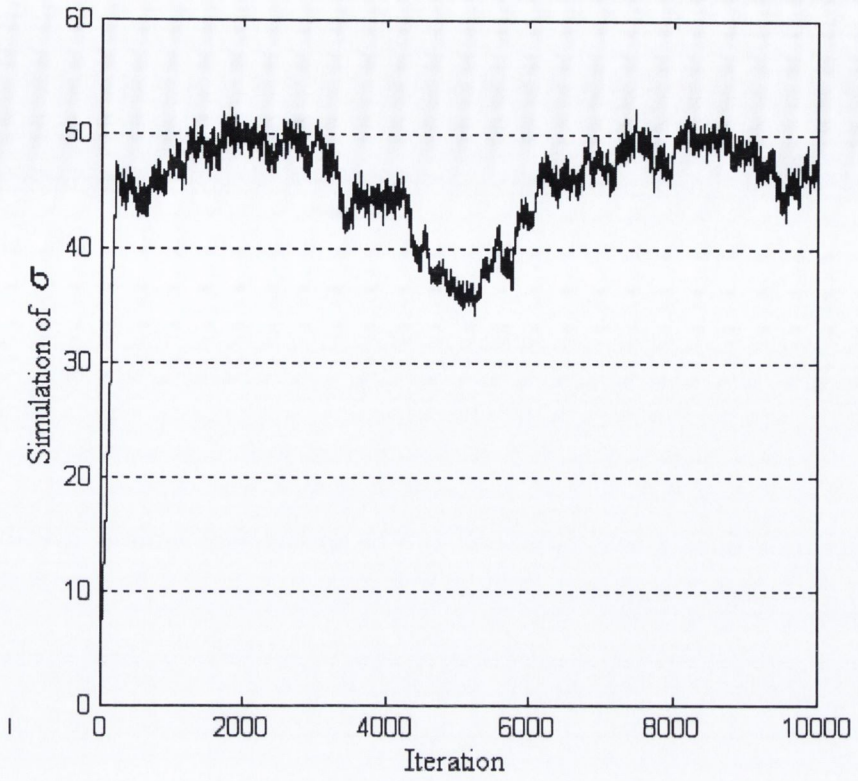


Figure 3.4b Simulation of the Values of σ^2 from Individual Normal Distribution.

But the results obtained due to this assumption show poor convergence. The reason behind the slow convergence of these two parameters is quite high correlation (correlation coefficient -0.7) between them and they cannot really be simulated using separate normal random walk proposal distributions.

Blocking

A general approach towards this problem is using the concept of *blocking* (Carlin, 1996) or updating multivariate blocks of parameters is employed leading to faster convergence. From Roberts and Sahu (1997), blocking of components leads to faster convergence rates, e.g. the larger the blocks are updated simultaneously the faster the convergence. Though blocking is more computationally demanding than component wise updating, it moves the high correlation to the random vector generator (Roberts & Sahu, 1997). To perform blocking, certain methods are used to get an idea of the probable joint density function of the other two parameters. Both the parameters are to be updated using a bi-variate distribution approach.

The original expression of the posterior was given in equation 3.15. The parameter ϕ_1 has to be integrated out from the expression of the equation 3.14, to get a tentative idea of the bi-variate proposal kernel. However, since analytical integration is not possible within the bounded intervals of [-1,1], the integral is approximated by changing the limits from [-1,1] to (-∞,∞) for the variables. The joint distribution has a form (integrating out the parameter ϕ_1)

$$p^*(\Theta_1,\sigma^2) \propto \sigma^{-T+1}(1+\Theta_1^2)^{\frac{1-T}{2}} \exp\left[-\left(\frac{AC-B^2}{2C}\right)\sigma^2(1+\Theta_1^2)^{-1}\right] \tag{3.18}$$

Given Θ_1, the conditional distribution of σ^2 is an inverse gamma distribution with parameters,

$$\alpha = T/2 \tag{3.19a}$$

$$\beta = \left[(AC-B^2)/2C\right] \Big/ \left(1+\Theta_1^2\right) \tag{3.19b}$$

Thus, simulation from the bi-variate proposal kernel (equation 3.18) is possible if a suitable approximation of the marginal of Θ_1 can be obtained. The marginal of Θ_1 (obtained by integrating out ϕ_1 and σ^2) is given by,

$$F(\Theta_1) = \int_{-\infty}^{\Theta_1} \left(\sqrt{1+u^2}\right) du \Big/ 2.2956 \tag{3.20}$$

Simulation from this density is performed by using the inverse transform method. Given that $F(\Theta_1) \sim$ U(0,1), and solving equation 3.20 for Θ_1, values of Θ_1 are simulated. As, Θ_1 and σ^2 are correlated, they are to be simulated from a joint distribution $p^*(\Theta_1,\sigma^2)$ to facilitate 'block updating'. Using the simulated value of Θ_1 from equation 3.22, a value of σ^2 is simulated from the inverse gamma distribution described in equation 3.19. Both the values are accepted with a probability,

$$p(\Theta_1,\sigma^2)=\frac{\Pi\left(\Theta_{1\_sim},\sigma^2_{sim}\mid\phi_1,y_{1,2,3.....n}\right)}{\Pi\left(\Theta_{1\_in},\sigma^2_{in}\mid\phi_1,y_{1,2,3.....n}\right)}\frac{q\left(\Theta_{1\_in},\sigma^2_{in}\right)}{q\left(\Theta_{1\_sim},\sigma^2_{sim}\right)} \tag{3.21}$$

where,

$$q\left(\Theta_1,\sigma^2\right)=\frac{\exp\left(-\left(\frac{\text{AC-B}^2}{2\text{C}}\right)\sigma^{-2}(1+\Theta_1{}^2)^{-1}\right)}{\sigma^{T+1}\left(1+\Theta_1{}^2\right)^{\frac{T-1}{2}}} \tag{3.22}$$

in accordance with the same acceptance criteria as previously mentioned. Following the same procedure as described for ϕ_1, 10000 values of both the parameters are simulated (figures 3.5 and 3.6).

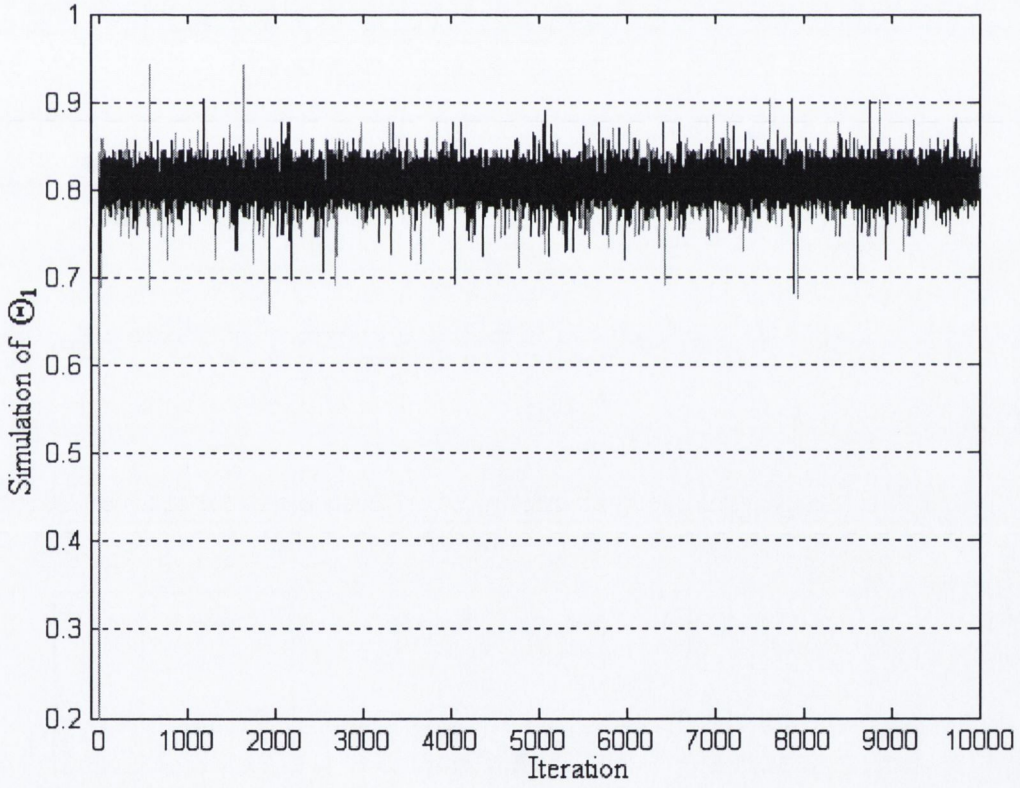


Figure 3.5 Simulations for the Values of Θ_1.

The values for Θ_1 and σ^2, are simulated from two sets of initial values as performed earlier in the case of simulation of ϕ_1. The chains are seen to converge satisfactorily for both the parameters. Using the simulated values of ξ, 10000 values of Y_{n+1} are simulated from $\prod(Y_{n+1}\mid Y_1,Y_2,Y_3.....Y_n)$ as described in equation 3.16d.

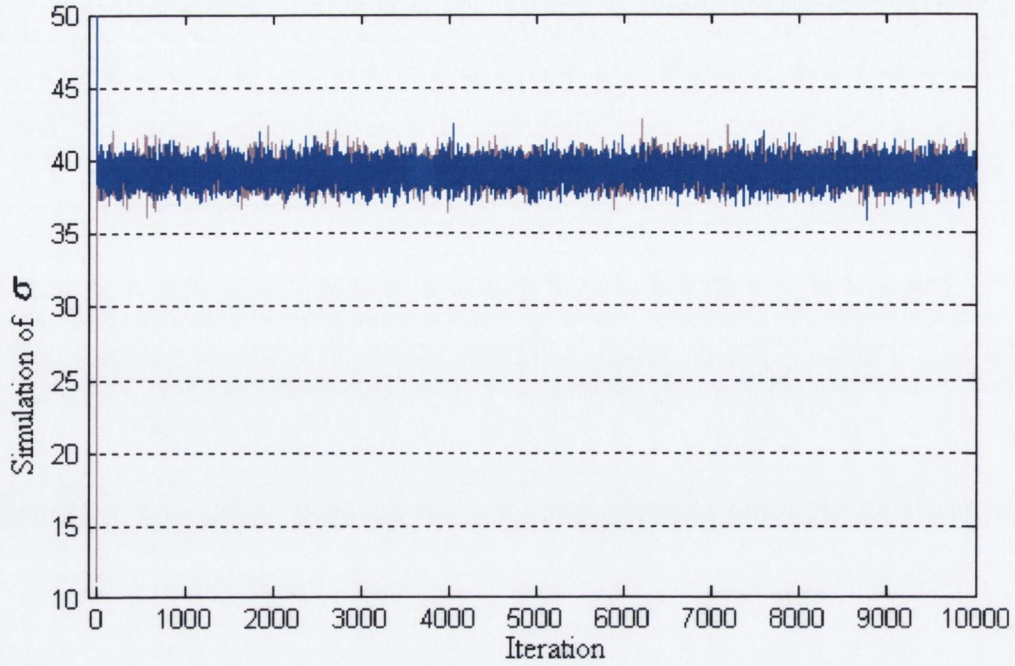


Figure 3.6 Simulations for the Values of σ.

Traffic volumes are predicted from 6:30 a.m. to 12 p.m. on 30th September 2004 with the proposed Bayesian SARIMA model with the estimated parameters (figure 3.7). The period with twenty-three data points covers the morning peak hours of traffic.

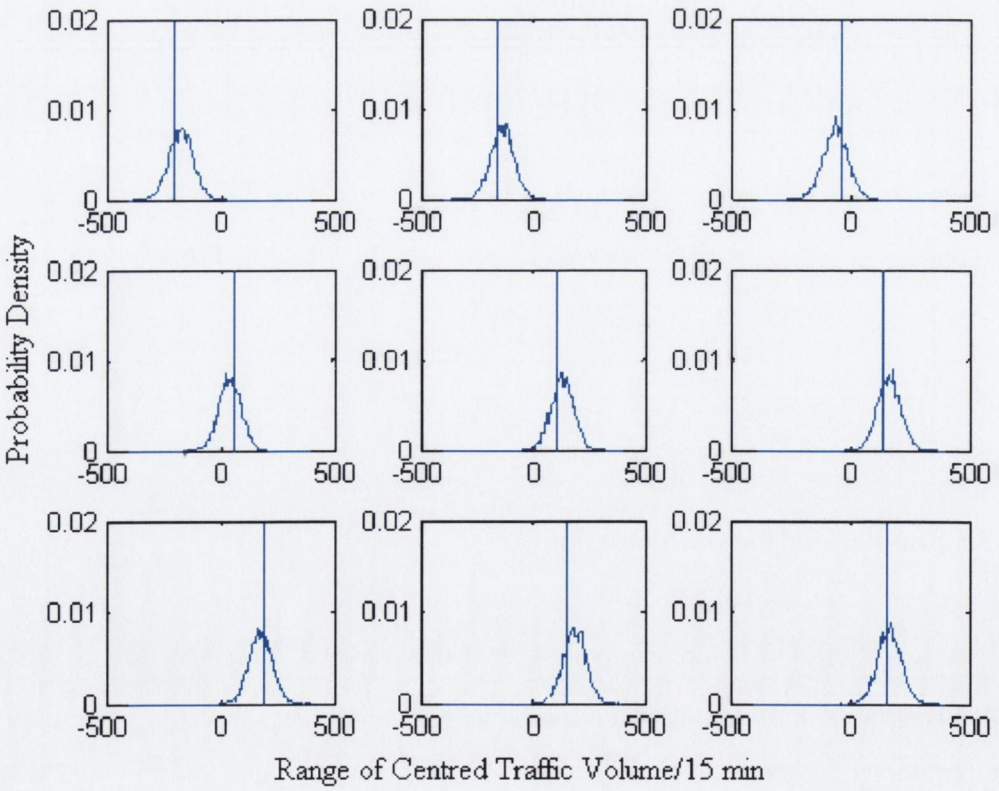


Figure 3.7 Predictive Distributions for 9 Future Data Points/Forecasts.

The densities of the first nine points of forecasts along with the observed data are plotted in nine subplots of figure 3.7. In the subplots, the vertical axis shows the probability density of each point in the range. The horizontal axis specifies the range of the forecast point which ideally varies from $-\infty$ to $+\infty$. The density functions in the subplots vary considerably with different mean and variance values which can only be obtained using Bayesian inference. As in the case of the original observations no uncertainty is involved, they are represented by a vertical line in each subplot.

3.4 PARAMETER ESTIMATION OF SARIMA MODEL USING CLASSICAL METHOD

To compare the Bayesian estimates with the results from classical method, the estimation of parameters using classical technique is carried out. The exact maximum likelihood method (Fuller, 1996) is used for estimating ξ of the SARIMA $(1,0,0)(0,1,1)_{96}$ model. The non-seasonal AR1 parameter ϕ_1 has a value 0.4215 and the seasonal MA (SMA) parameter Θ_1 has a value of 0.8215. According to the classical theories the parameters are supposed to have normal densities. The normal density plots of ϕ_1 and Θ_1 are shown in figure 3.8(A) and (B). The standard deviation σ of the SARIMA process is 43.5.

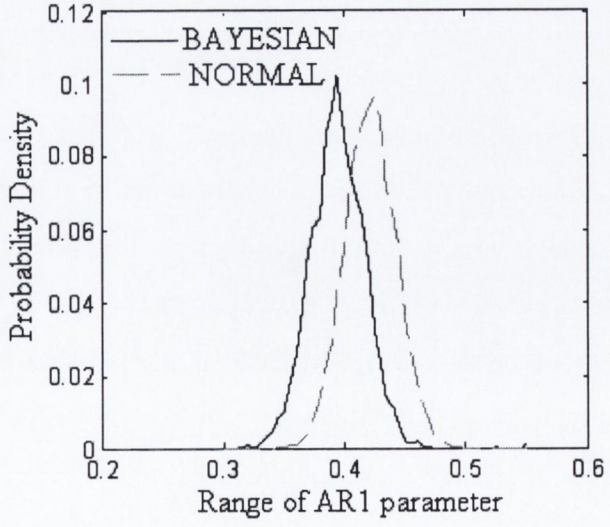


Figure 3.8(A) Bayesian and Normal Density (from Classical Estimate) Plots for ϕ_1.

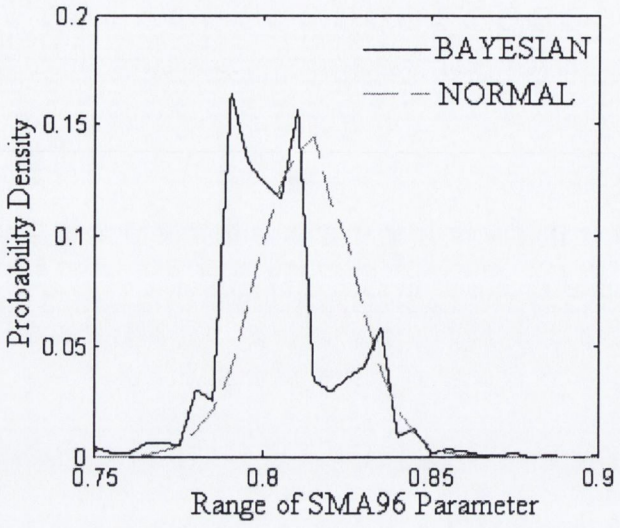


Figure 3.8 (B) Bayesian and Normal Density (from Classical Estimate) Plots for Θ_1.

3.5 COMPARISON OF BAYESIAN AND CLASSICAL INFERENCES

In figures 3.8 (A) and 3.8(B), the density estimates of the SARIMA parameters from Bayesian inference are plotted along with the normal density plots from classical inference. In both the subplots, the parameter density obtained from Bayesian estimate is not exactly normal. The variability of parameters is considered in Bayesian inference while forecasting. But, in the case of classical forecasting the standard error associated with the parameter is ignored and the mean of the density plot is used only as a constant.

The five and half hour traffic volume predications from the Bayesian and classical inference are plotted along with the original observations in figure 3.9 and figure 3.10 respectively. The absolute percentage error (APE) of the predictions from Bayesian inference is 5.4%, where as the APE for classical forecasts is 5.1%. Though the overall error estimates are very comparable, only the Bayesian inference can capture the extreme morning peak. The point forecasts in figure 3.10 vary within a range of 50 cars/15min, whereas in figure 10 the same varies within a range of 100cars/15min. Hence, the Bayesian forecasts have a bigger range of variability compared to the classical forecasts which are nearly flat in nature. The Bayesian predictions can more closely match the rapid variability of the real observations unlike the classical forecasts.

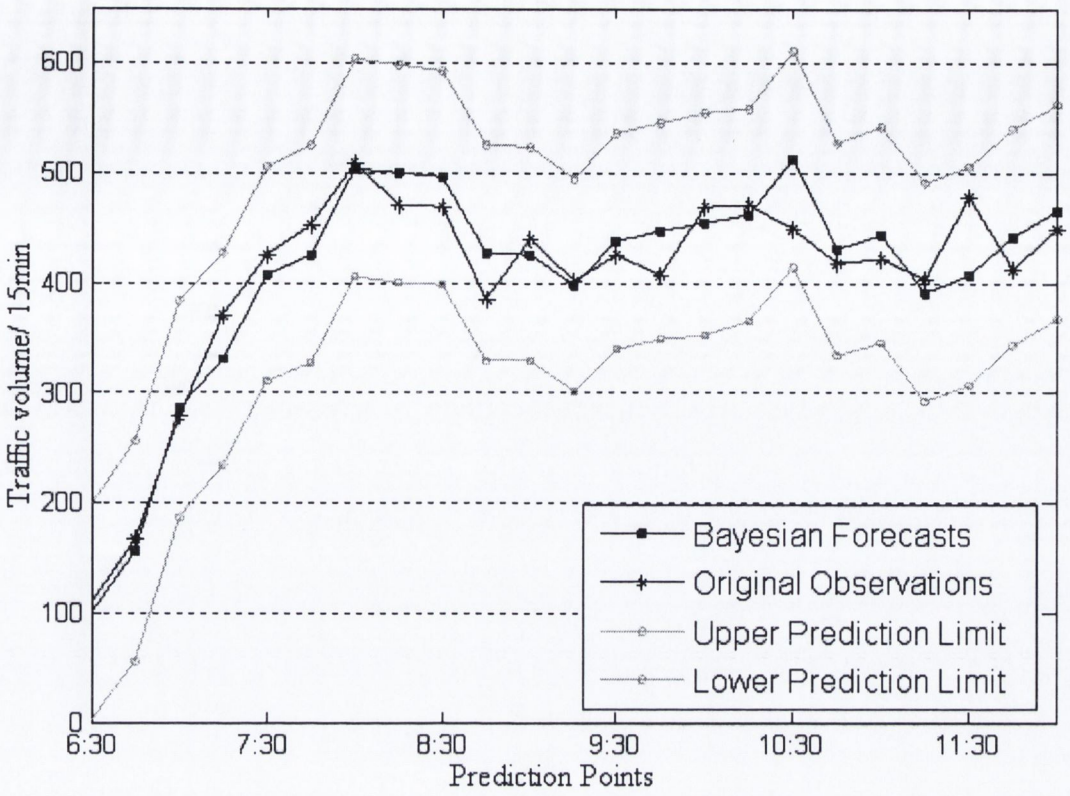


Figure 3.9 Forecasts from Bayesian Inference.

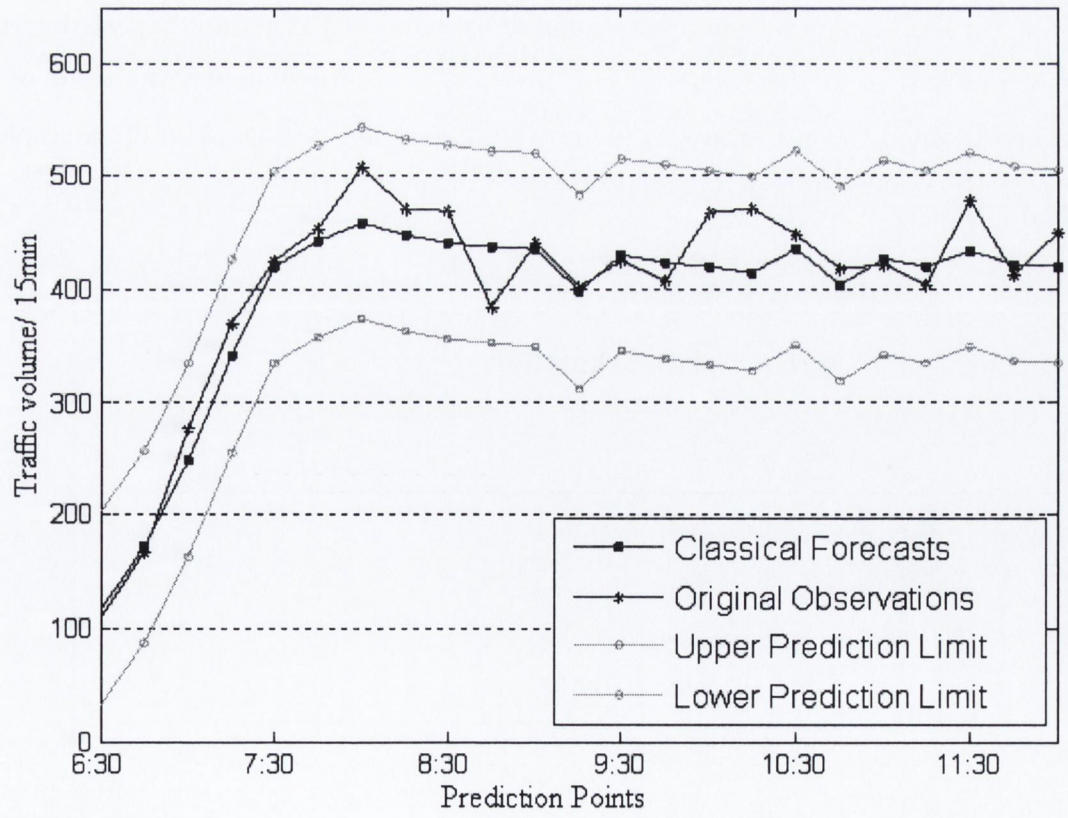


Figure 3.10 Forecasts from Classical/Frequentist Approach.

Both the plots in figure 3.9 and figure 3.10 include two other lines specifying the upper and lower limit of prediction. In figure 3.9, this prediction interval is the 95% confidence interval. A C % confidence interval for a parameter χ is obtained by finding L and U such that,

$$p(L < \chi < U) = C \tag{3.23a}$$

The Bayesian analogue of classical/frequentist confidence interval is a credible set (Carlin 1996). In general, a $100C$% credibility interval for a parameter χ given a random sample $X_1, ..., X_n$ is an interval ($L(X_1, ..., X_n)$, $U(X_1, ..., X_n)$) such that

$$p(L(X_1, X_2 ..... X_n) < \chi < U(X_1, X_2 ..... X_n)) = \int_{L(X_1, X_2 ..... X_n)}^{U(X_1, X_2 ..... X_n)} f(\chi \mid X_1, X_2 ..... X_n) d\chi = C \tag{3.23b}$$

Hence, the credibility interval is conditional on the dataset. A 95% credibility interval means that there is a probability of 0.95 that the forecast/parameter lies within that particular interval at that point in time. Whereas 95% confidence interval in a classical case means that if the same procedure as used for constructing the interval is repeated several times then in 95% of the cases the range of the intervals will include the true value of the forecast/parameter. The credibility set is more rational as it is not based on the asymptotic theory.

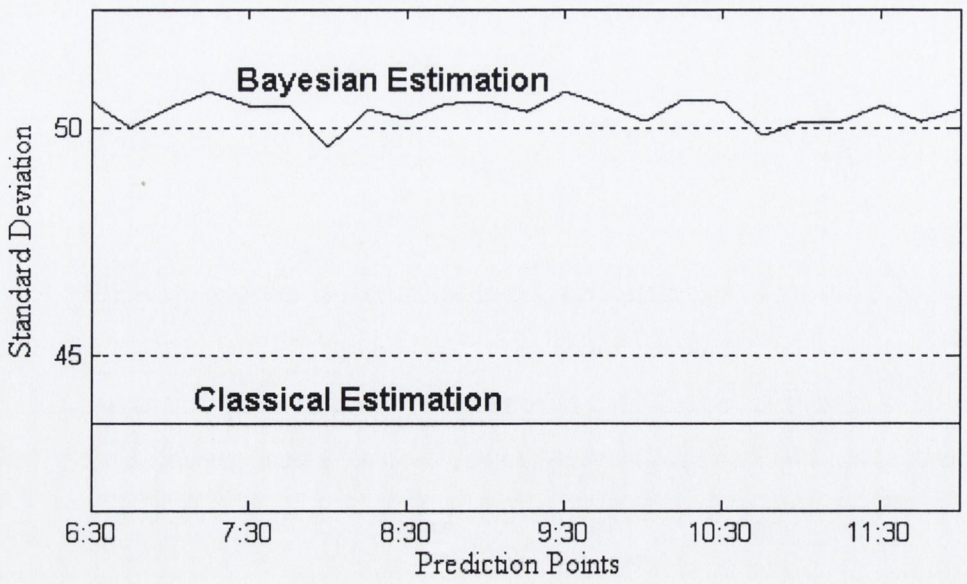


Figure 3.11 Variance from Bayesian and Classical Inference.

As seen in figure 3.7, each prediction of Bayesian inference has a distribution of its own. So, each future point has a different variance. In figure 3.11, the plot of the standard deviation from Bayesian inference varies continuously with time accounting for the happening of the events (Bayesian update) whereas the same from classical inference is constant over time failing to account for the change in the conditional probability of a forecast.

The main strength of the Bayesian method of parameter estimation lies in capturing the rapid variability of the traffic estimates which can not be captured by classical estimation due to application of methods like least squares which try to minimise the overall error rather than capturing the local variability of the observation points. To emphasise the potential of the Bayesian time-series model, another set of traffic flow data from TCS 183 are analysed using the Bayesian SARIMA $(1,0,0)(0,1,1)_{96}$ model developed in this chapter. The traffic flow observations from 3rd November 2004 to 30th November 2004 data (excluding weekends) are modelled using both the Bayesian and Classical methods of parameter estimation. The forecasts from both the models along with original observations during the morning peak hours of 1st December 2004 are plotted in figure 3.12. Similar to the previous case (figures 3.9 and 3.10) the Bayesian forecasts seem to better predict the peak.

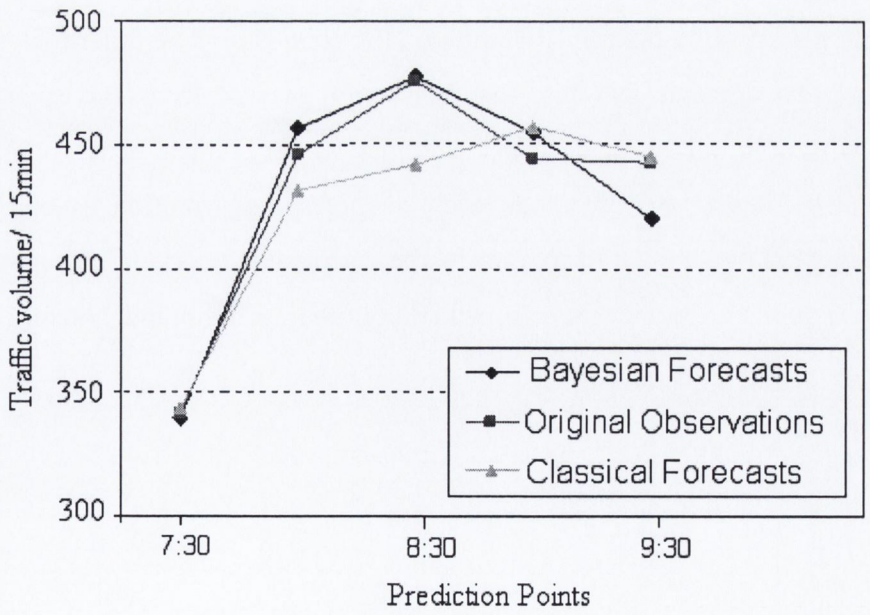


Figure 3.12 Bayesian and Classical Forecasts on 1-12-2004.

3.6 ADVANTAGES OF BAYESIAN SARIMA MODEL

The study in this chapter shows that the Bayesian inference of SARIMA models provides a more rational technique towards short-term traffic flow prediction compared to the commonly applied classical inference. The advantages of using a Bayesian SARIMA model instead of its classical counterpart are as follows:

- Bayesian inference is conditional on the observed traffic volumes and generates individual probability density curve (may not be normal as assumed in case of maximum likelihood estimates) for each of the SARIMA model parameters as well as for each point of prediction. Instead of using a constant parameter value estimated by classical methods, variability of the parameters conditional on the traffic flow observations can be considered in Bayesian inference.

- Forecasts from the Bayesian approach can better model the traffic behaviour in reality with rapid fluctuations and extreme peaks. The same from the classical SARIMA model concentrate more towards the mean and cannot capture everyday features like the morning and evening peak effectively.

- In a real traffic flow, uncertainties associated with traffic forecasts over different times of the day should vary depending on the congestion and other factors (such as weather). To account for this, the prediction interval from Bayesian inference varies with time unlike the classical inference where point forecasts and a constant confidence interval are obtained based on normal distribution assumption. This provides more realistic information to the traffic planners and transport network managers and is important for efficient and optimal traffic management.

CHAPTER 4

BEHAVIOUR OF DIFFERENT 'TRAFFIC FLOW THEORY' MODELS UNDER SHORT-TERM TRAFFIC INFLOW VARIABILITY

4.1 INTRODUCTION

The univariate short-term traffic flow forecasting algorithms developed in chapters 2 and 3 are statistical techniques for modelling traffic flow. But these models cannot capture the physical basis of traffic flow movement which the traffic flow theory models can capture effectively. To efficiently simulate and forecast univariate traffic volume for assignment of traffic on a real-life ITS equipped urban transport network, it is necessary to integrate the traffic flow theory models with the statistical algorithms. To find out a suitable theoretical traffic flow model for this purpose, the behaviour of different traffic flow theory models under the application of different types of traffic inflow with short-term variability are required to be studied.

Two types of well established analytical macroscopic link models are considered in this chapter to investigate their suitability of application when subjected to different realistic traffic loading scenarios in a real transport network. In this study, a certain type of whole-link model termed as linear delay function model (Nie and Zhang, 2002b) or the Friesz Travel time model (Carey and Ge, 2003) is chosen as the first type of macroscopic analytical link model for investigation. A comparatively different macroscopic approach, the LWR model or hydrodynamic model, is considered as the other type of macroscopic analytical link model for this study. Both the analytical macroscopic link models considered here for comparison are solved using solution algorithms, which include certain variations in the techniques from the generally used ones. The models are compared on the basis of their outflow profiles. Different types of traffic inflow profiles with occurrence of incidents or with the effects of signal control are applied as inflow on a single urban arterial.

4.2 FRIESZ TRAVEL TIME MODEL

This model is quite popular (Astarita, 1996; Wu et al., 1998 and Xu et al., 1999) with DTA researchers using optimal control, variational inequality or the complementarity approach. Friesz (1993) solved the dynamic user equilibrium problem assuming a linear travel time function. In the FTT (Friesz travel time), linear travel time function is the only simple travel time function observing first-in-first-out (FIFO) condition as indicated by Nie and Zhang (2002b). To ensure FIFO under all inflow conditions, a linear travel time function is used in this study for the FTT model. The model consists of the following equations,

1. Travel time equation (flow behaviour):

$$\tau(t) = f(x(t))$$
$$= \alpha + \beta.x(t)^n \qquad (4.1)$$

[while considering a linear travel time function, n =1] where, $\tau(t)$ is the link travel time and $x(t)$ is the link volume at time instant t. The free flow travel time for the link is α and β is the reciprocal of maximum feasible constant outflow from the link. The travel time function $f(x(t))$ should be continuous and non-decreasing.

With a constant flow U and total link travel time T, the total link volume x will be UT. Substituting that in equation 4.1 (with $n = 1$),

$$T = \frac{\alpha}{1 - \beta U} \qquad (4.1a)$$

Considering the fact that link travel time should be positive and finite, for $U < 1/\beta$ a solution exists. When $U \to 1/\beta$, the link travel time approaches infinity (Carey and McCartney, 2002). For inflows which exceed maximum stable outflow, traffic continues to exit from the link in a stable way but the travel time increases manifolds (figure 4.1).

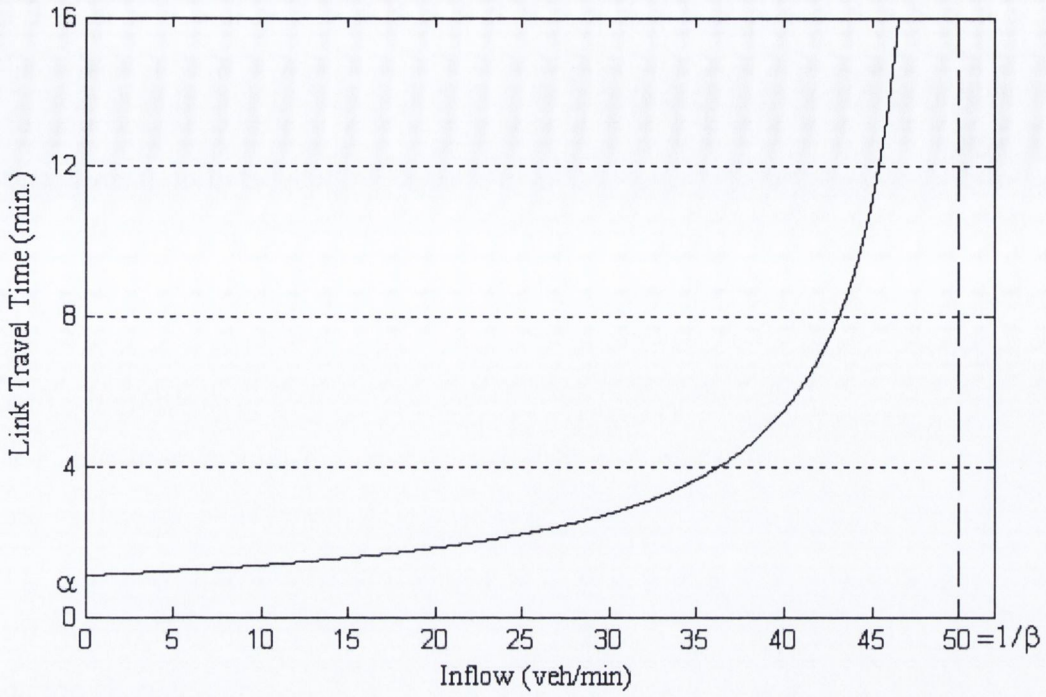


Figure 4.1 Stable Travel Time vs. Constant Inflow.

2. Equation of state:

$$x(t) = x(0) + \int_0^t (u(s) - v(s))ds \; ; \tag{4.2}$$

where, $u(s)$ and $v(s)$ are inflow and exit-flow rate to the link at time s and $x(0)$ is the initial existing link-volume and $x(t)$ is the instantaneous link volume at time t.

3. Flow conservation equation:

$$x(0) + \int_0^t u(s)ds = \int_0^{t+\tau(t)} v(s)ds$$

Taking the first derivative with respect to t and rearranging gives,

$$v(t+\tau(t)) = \frac{u(t)}{1+\tau'(t)} = \frac{u(t)}{1+f'(x)(u(t)-v(t))} \tag{4.3}$$

***Solution algorithm used*:** The solution can be derived analytically up to a few steps for very simple inflow patterns like step-function inflows or linear inflows. But the analytical solution becomes intractable in the cases of more general inflow patterns. To account for this intractability a numerical technique has to be used for the solution (Carey and McCartney, 2002). The solution algorithm used here is similar to the algorithm used by Ran et al. (1993) and Carey and McCartney (2002).

At the start, $v(t)$ is zero until time $t = \alpha$, as the link is assumed to be empty before the first vehicle enters the link. Under this assumption, $u(t)$, $v(t)$, $x(t)$ are known until time t = α. Uniform and fine discretization of time is used up to this limit. $v(t+\tau(t))$ values are calculated from equation (4.3) using the values of $u(t)$ at discretised time intervals till t. For the next phase, the time intervals are those calculated values of link travel time, $(t+\tau(t))$ for which corresponding exit flow rate, $v(t+\tau(t))$ values are known. The time intervals are no longer uniform but the technique involving the use of known values of $u(t)$, $v(t)$ and $x(t)$ is applied as before to calculate the instantaneous link travel time for each instant. The outflow calculated using this algorithm, generally contains a pseudo-periodicity in the profile which is explained and dealt with extensively in a paper by Carey (2003). This pseudo-periodic nature though inherent to the model, is less predominant in the case of a linear (travel time) link performance function than the quadratic or the higher order ones.

4.3 HYDRODYNAMIC MODEL

The kinematic wave model of freeway traffic flow and its finite difference approximations (Daganzo, 1995b) (cell-transmission model) (Daganzo, 1994, 1995a) are considered as the benchmark analytical macroscopic model in DTA. The kinematic wave model or the LWR model is considered as the simplest and most commonly used continuum model. This model includes,

1. The quasi-linear hyperbolic conservation equation (*Equation of continuity*)

$$\frac{\partial q}{\partial x} + \frac{\partial k}{\partial t} = 0 \tag{4.4}$$

2. The relationship between flow and density (*Equation of state*)

$$q = S(k(x,t),x,t) \tag{4.5}$$

In equations 4.4 and 4.5, q is the flow, k is the density and $S(k(x,t),x,t)$ is a function of k, x and t. Lighthill and Whitham (1955) have analyzed in detail the case when the flow does not depend directly on changes over time, independent of other variables. When a homogeneous highway is assumed, i.e. when the variability of the spatial and the temporal characteristics along the highway is not significantly affecting the flow, the equation can be considered as follows,

$$q = S(k(x,t)) \tag{4.6}$$

To make the model consistent to the whole-link model considered before, $S(k(x,t))$ is obtained by rearranging the travel time function used for the whole-link model (Carey and Ge, 2003).

***Solution algorithm used*:** Unlike other studies using numerical methods to solve the LWR model for comparison with whole-link travel time model (Carey and Ge, 2003; Nie and Zhang, 2002a) this solution is based on an analytical solution. But due to some intractability encountered in the full analytic solution, certain assumptions are used (in the case of shock waves or accelerating fans).

Equation 4.4, with the variable *'q'* replaced using the relation of equation 4.6, gives the following first-order quasi-linear hyperbolic conservation equation.

$$\frac{\partial S(k)}{\partial k}\frac{\partial k}{\partial x} + \frac{\partial k}{\partial t} = 0 \tag{4.7}$$

where, $0 \le x \le \infty$; $0 \le t \le \infty$ and $\frac{\partial S(k)}{\partial k}$ is called the *wave speed.*

With the help of the equation of state (equation 4.5) and a set of well-posed initial and boundary conditions, the conservation equation 4.7 can be solved by the *method of characteristics*.

A *characteristic* is a trajectory in time-space along which the density remains constant even if there are small perturbations elsewhere in the system. In this case the *characteristic equation* is,

$$\frac{dt}{dr} = \frac{\partial k}{\partial S(k)} \tag{4.8}$$

Using the equations $\frac{dx}{dr} = 1$ (4.9)

and, $\frac{dk}{dr} = 0$ (as k is constant on the characteristics) (4.10)

where, r is an independent variable assumed to satisfy the equations 4.8, 4.9 and 4.10.

The *characteristic equation* becomes, $t = \int_{x_0}^{x} \frac{\partial k}{\partial S(k)} dx + t_o$ (4.11)

The initial conditions are, $q(x,t)$ = inflow rate to the link = $u(t)$, when $x = 0$. Using the relation from equation 4.6, initial conditions $k(0,t)$ and $k(x,0)$ are calculated. The characteristics curves initiate from the point $(0,t_o)$. Each curve has a constant density $k(x,t)$ along its spread and that value of the density function is calculated from the coordinates of the initial point (x,t). So, when x equals the link length, the value of $q(x,t)$ gives the outflow rate from the link which is varying over time.

This apparently simple and continuous solution method becomes very complicated and computationally intensive if the characteristic lines intersect with each other within the time and space (link-length) of interest over which the solution is sought. When the density or flow concentration increases and the characteristic lines converge, a shock wave is formed. With the decrease in density the characteristic curves diverge and acceleration fans are formed. According to Newell (1993), it is necessary to remove the formal solutions of the resulting equations, which are inconsistent with a preferred direction of time and space. Hence, the results obtained from the shocks, in negative space and time, are not acceptable. Using the concept of cumulative flow or traffic volume as described by Newell, the laborious calculations for shock paths can be avoided. Here a particular assumption

(explained later below) has been used to reach the unique solutions when the characteristics diverge. This assumption is validated using Newell's approach.

In this solution algorithm, while dealing with the acceleration fans, i.e. where the characteristics diverge, it is assumed that concentrations obtained at the boundaries of the acceleration fans are the unique values of the solution. The movement in negative time (going back in time) is not considered, as it is not possible in reality. The assumption is explained explicitly while solving the problems described here.

4.4 COMPARISON OF BEHAVIOUR UNDER DIFFERENT TRAFFIC DEMANDS

The two models discussed in this chapter are compared regarding their performance under different types of inflow or traffic demands.

4.4.1 Comparability of the Models

The whole-link model is based on a travel time function and the LWR model is based on a flow-density function. If the flows are constant over time, the travel time model is equivalent to the flow density model, i.e., given either of these two functions one can simply be rearranged to obtain the other and hence the parameters of the other. In all the cases of different types of traffic inflow patterns studied in this chapter, this concept is used to ensure consistency of the two models, i.e. starting from a numerical example of one model, rearranging it (temporarily assuming constant inflows) the other can be obtained.

In FTT model a linear travel time function is used, hence $\tau(t) = \alpha + \beta x(t)$. Since $x(t)$ is the instantaneous link-volume (here, instantaneous mean over a minute), if L is the link length then considering a uniform density all along the link and constant flow rate, the *flow-density relation* [$q = Q(k)$] becomes,

$$q = Lk/(\alpha + \beta Lk) \tag{4.12}$$

Figure 4.2, shows that the flow-rate, q, from equation 4.12 is a concave function of k, passing through the origin and ending in an asymptote, $q = 1/\beta$.

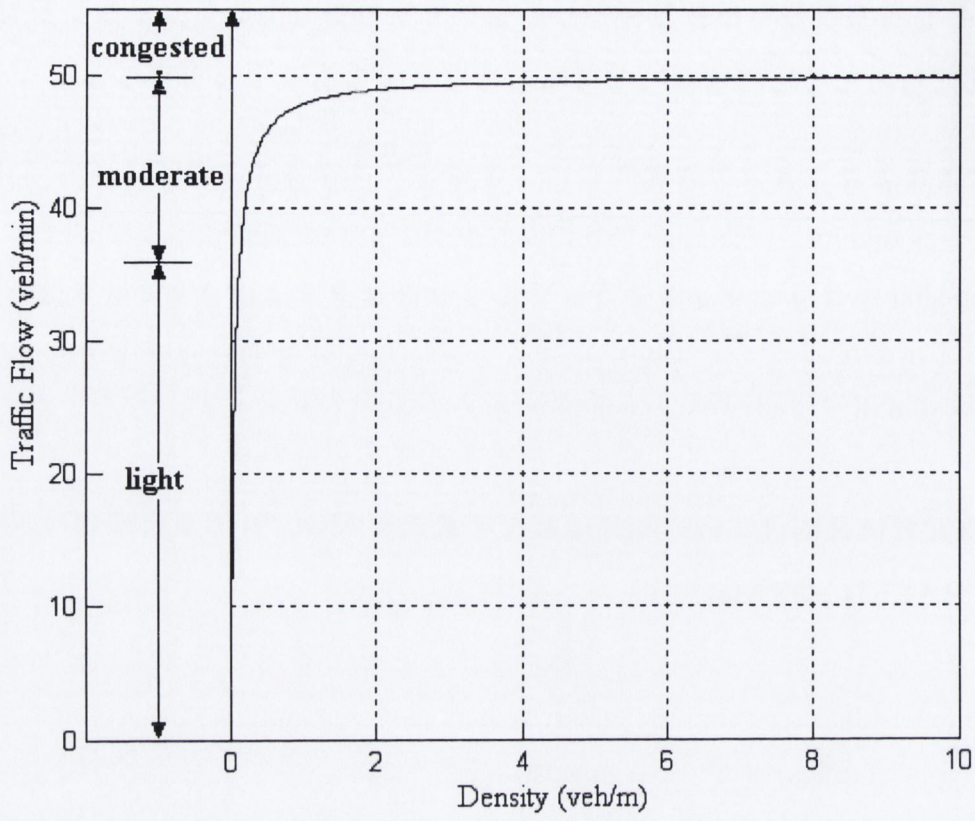


Figure 4.2 Flow-Density Curve.

In the LWR model, the equation 4.6 becomes, $q = Lk/(\alpha + \beta Lk)$. Hence, the partial differential equation 4.7 becomes,

$$\frac{L\alpha}{(\alpha + \beta Lk)^2}\frac{\partial k}{\partial x} + \frac{\partial k}{\partial t} = 0 \quad \text{where } 0 \le x \le \infty,\ 0 \le t \le \infty. \tag{4.13}$$

The initial condition $q(0,t) = u(t)$ remains the same.

The characteristics equation 4.8 becomes, $\dfrac{\partial t}{\partial r} = \dfrac{(\alpha + \beta Lk)^2}{L\alpha}$ (4.14)

and,
$$\begin{cases} k(0,t) = \dfrac{\alpha\, u(t)}{(L - \beta L u(t))} & \text{for } t \ge 0 \\ k(x,0) = 0 & \text{for } x \ge 0 \end{cases} \tag{4.15}$$

According to equation 4.15, if $u(t) > 1/\beta$, then the initial density becomes negative, i.e. the solution does not exist. Practically, as average density k is calculated over a minute for an inflow more than $1/\beta$ for less than 1 minute, the solution exists.

To compare the two models under different traffic conditions, all the inflows described in the following subsections are applied on a homogeneous link length L = 1200 m. In the linear travel time function considered for the FTT model, a free flow travel time of $\alpha = 1.1$ minute and $\beta = 0.02$ min/veh. are used in all the cases studied. This implies that the peak capacity flow is 50 vehicles/minute (= $1/\beta$). The outflows are calculated using the solution algorithms described for the two methods in sections 4.2 and 4.3.

4.4.2 Comparison Using a Gradually Changing Inflow

The first comparison is carried out to investigate the behaviour of the FTT and LWR models in light traffic conditions with inflow slowly changing over time. For an inflow function, $u(t) = 15\left(\frac{t}{t+1}\right)^2$, the outflow rate is calculated using both models.

From figure 4.3, it can be seen that the outflows calculated from both models are almost identical. In light traffic conditions, the inflows are always in the initial upward sloping part of the flow-density function (figure 4.2). Even though the flow-density function is nonlinear, the initial upward sloping part is almost linear which implies a constant travel time. In that scenario both the models would always give same results.

Hence, another comparison for slowly changing inflow in the case of moderately congested traffic condition (the inflow reach the curved section above the upward linear part of the flow density curve in figure 4.2) is shown in figure 4.4.

To simulate this situation, an inflow function $u(t) = 45\left(\frac{t}{t+1}\right)^2$ is applied on both the models. The outflow rates obtained are also very similar but not identical. The profiles are of same nature but outflow from LWR is marginally different from the outflow obtained from the FTT model. The results from the two models are similar enough to approximate the LWR model by the whole link model.

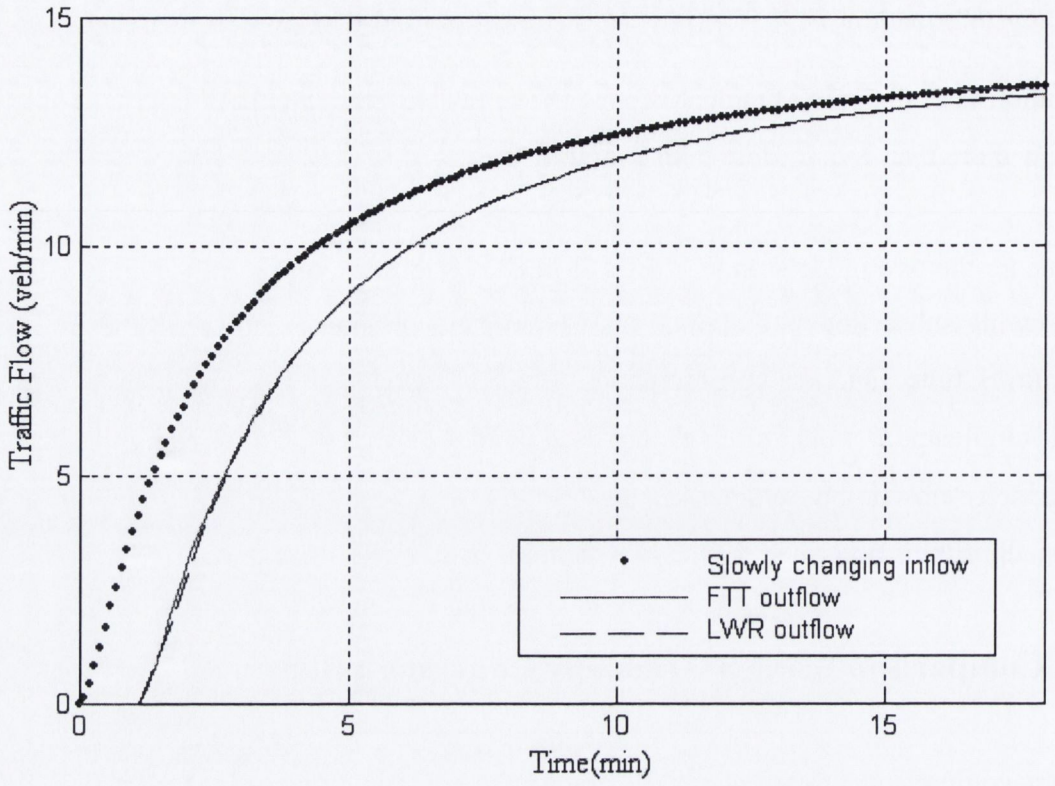


Figure 4.3 Inflow and Resulting Out-Flow Profiles for Both the Models under Light Traffic Condition.

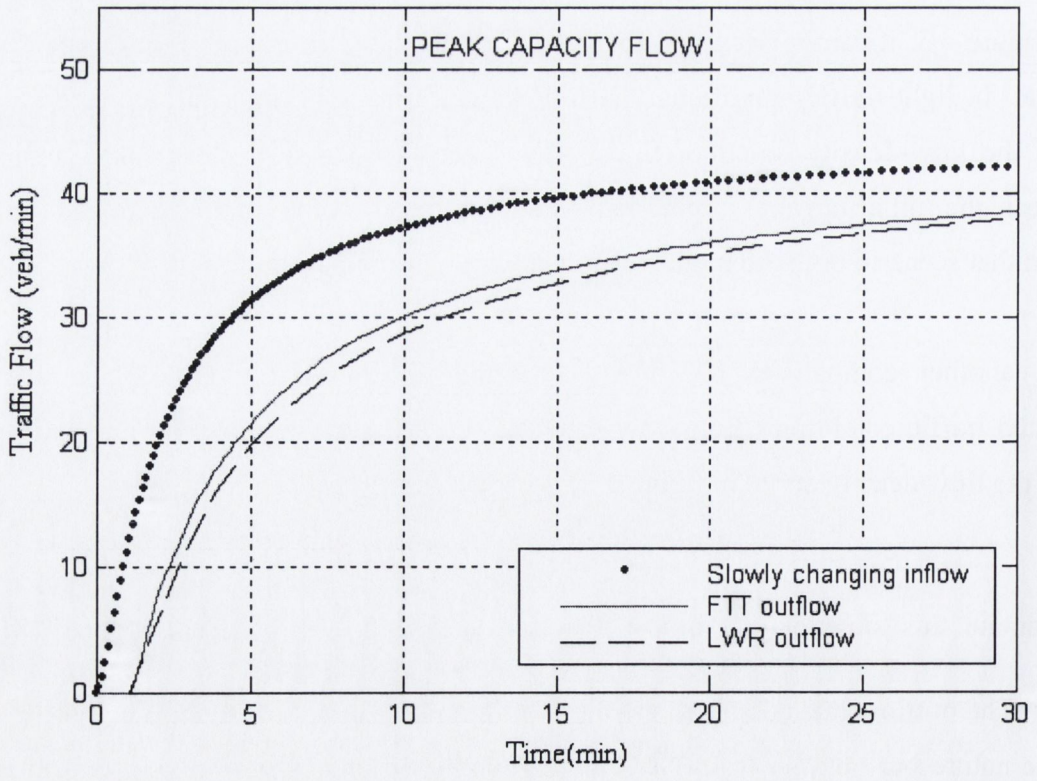


Figure 4.4 Inflow and Resulting Out-Flow Profiles for Both the Models under Moderate Traffic Condition.

An interesting fact to consider in the outflow profile from FTT model, in figures 4.3 and 4.4, is that the pseudo-periodic nature of the outflow is not distinguishable at all.

4.4.3 Comparison Using Inflow Loadings with Impulse

In all observed cases of suddenly or rapidly changing flow-rate, a linear inflow is considered as the underlying existing loading. An impulsive load is introduced for a small interval of time in addition to the loading already used.

The impulse loading used is of two types, continuous and discontinuous. Continuous loading means that the impulsive load is introduced in such a way that at the points of superposition (of the impulsive flow to the existing flow) the magnitude and the slope of both curves are the same. In the case of the discontinuous flow an abrupt change in magnitude and slope occurs.

Rapidly Changing Inflow (Continuous Impulse) Under a Moderately Congested Traffic Condition

To investigate the effect of introduction of sudden or rapid changes in inflow, first a moderately congested traffic condition is considered. The existing inflow to the link is a linear inflow function of the form $u(t) = 1.5\ t$. The impulse load superposed on this existing loading is a fifth order polynomial. It is assumed that at the points of superposition the impulse function is continuous and differentiable and a given maximum value of the flow rate occurs at a given time instant, at $t = 7$. Under these assumptions, a fifth order polynomial representing the impulsive inflow loading function is constructed. Hence, the inflow loading is given as,

$$u(t) = -3t^5 + 12.5t^4 - 193t^3 + 1454.7t^2 - 5320.3t + 7557.9 \quad \text{when, } 5 \le t \le 10. \qquad (4.16)$$

Impulsive loading is introduced at the fifth minute and continued for a time interval of five minutes. After the tenth minute, the inflow returns to normal. Given this nature of inflow to

the link, the exit flow/ outflow from the link is calculated using both the FTT and LWR model.

A diagram of the inflow and the outflows from both the LWR model and the FTT model are given in figure 4.5. The piecewise linear nature of the outflow from FTT model is due to the numerical discretization of the solution algorithm used. The time intervals used in FTT algorithm become non-uniform from the second iteration /step onwards. The exit flow calculated using the FTT model does not show any pseudo-periodicity in the region studied.

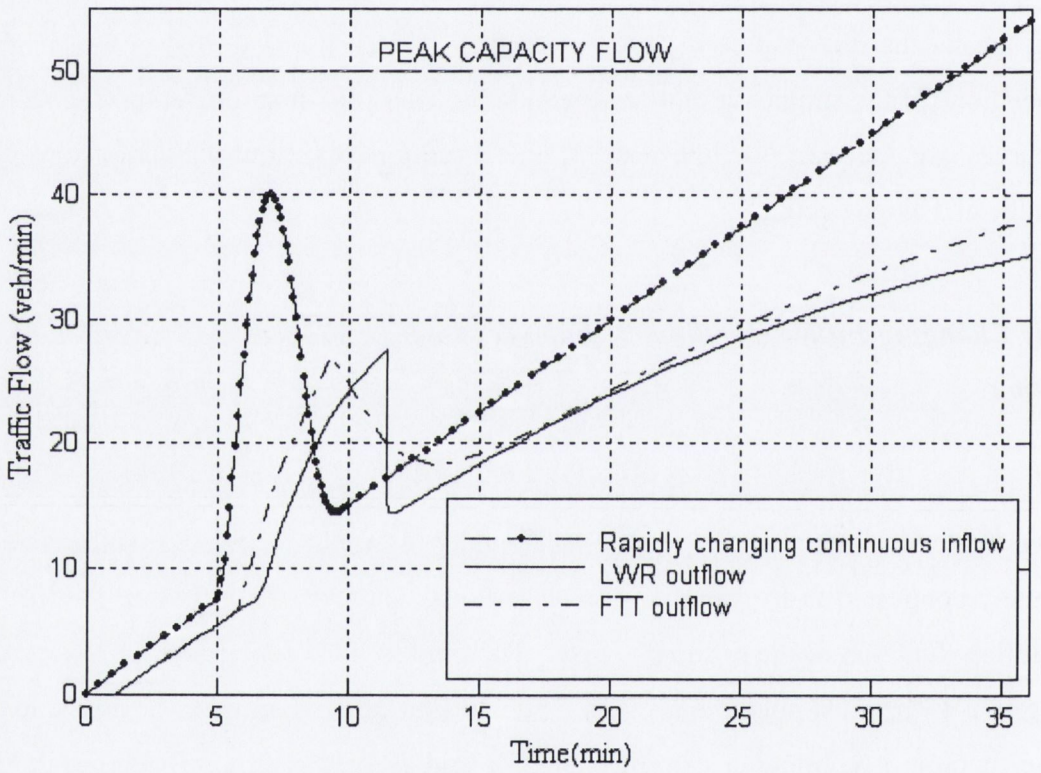


Figure 4.5 Inflow and Resulting Out-Flow Profiles for Both the Models for Rapidly Changing Continuous Inflow under a Moderate Traffic Condition.

In figure 4.6 the map of the characteristic curves is shown to explain the method of solution adopted for calculating the outflow from LWR model. The characteristic lines intersect and generate an acceleration fan at the point of sudden change in concentration of the traffic due to impulsive loading. The solution algorithm described in section 4.3, gives a "multiple solution" zone at the point of the link exit, which is encircled in figure 4.6.

(The characteristic lines cannot intersect each other. In case of the shock wave solution, they terminate at the point of intersection.) This multiple solution zone actually gives different values of density for the same space-time co-ordinate, as seen by the dotted line outflow profile in the figure 4.6.

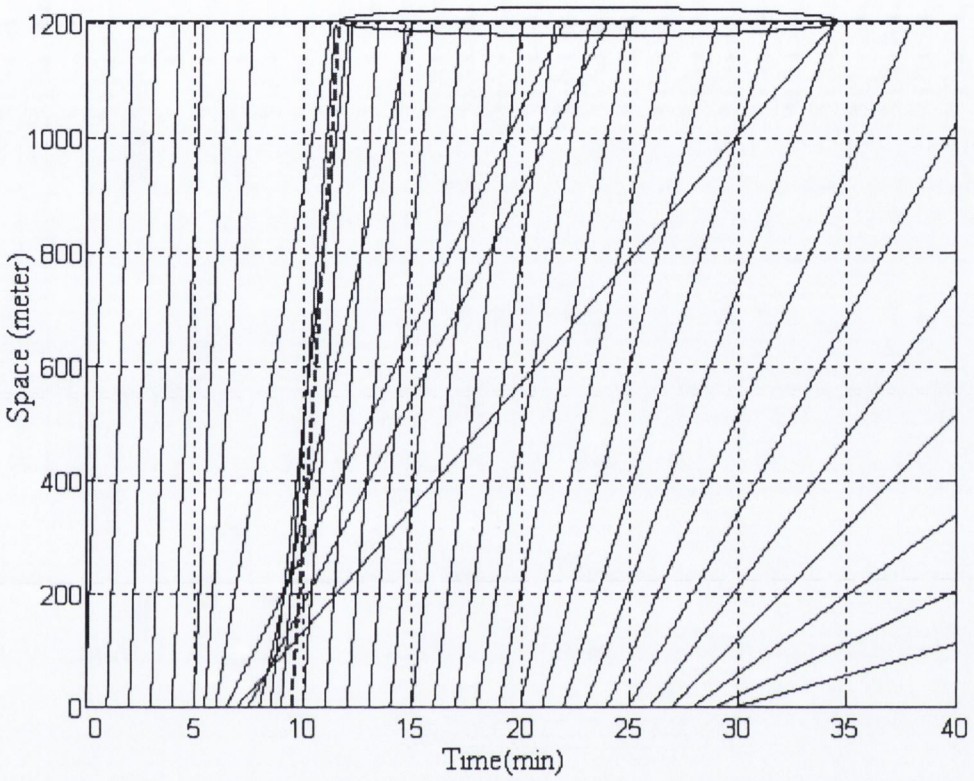


Figure 4.6 Characteristics Map For LWR Model.

According to the proposed solution in this study, the characteristic lines in forward time with lesser densities are considered as a feasible solution. More precisely the characteristic lines with maximum positive slope are considered to give unique solutions in the region of acceleration fans (i.e. the encircled region in figure 4.6). The characteristic line, which is followed to reach the unique solution, is shown in bold dashed.

The solution, when using this assumption is shown with dash-dot line, while the multiple solutions are shown as dotted line (figure 4.5). This feasible solution obtained using this assumption is verified using Newell's solution techniques as described in section 4.3. The lower envelope of the cumulative outflow (figure 4.7) is chosen in this case. Due to this, a small jump in the region of end of the impulsive loading can be seen in the outflow profile. The unique solution obtained using Newell's technique is same as the one obtained using the assumption proposed in this study.

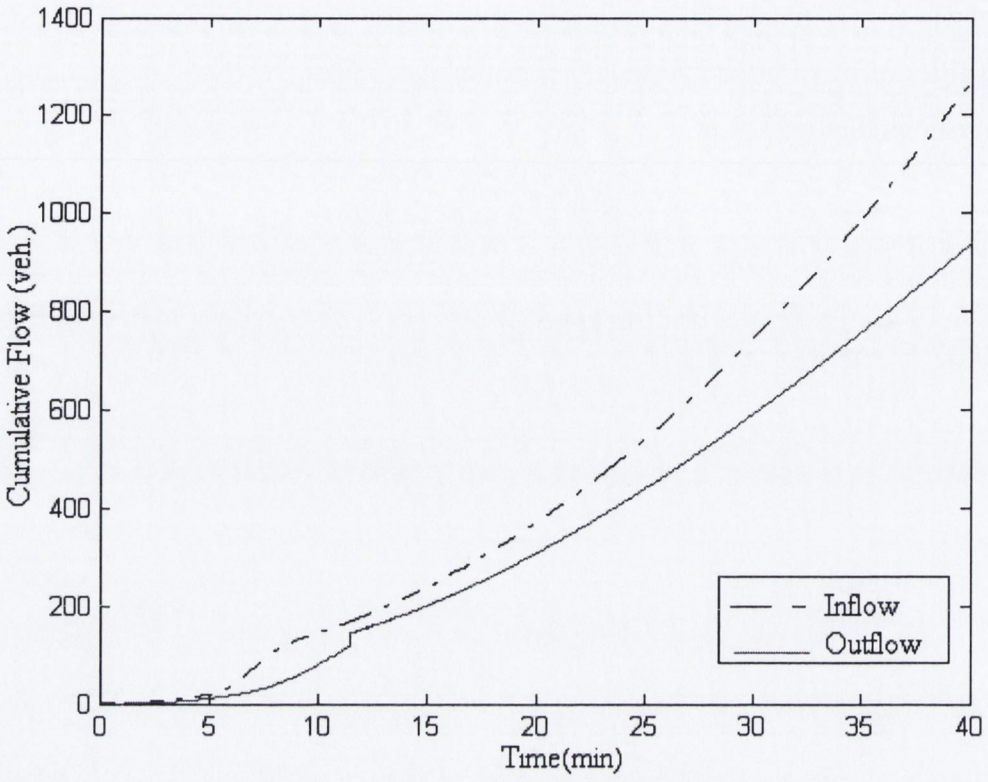


Figure 4.7 Cumulative Inflow and Outflow Profiles Using Newell's Method.

The two outflows from the two models reach almost the same maximum outflow rate. Only the travel time (the horizontal distance between the outflow and inflow profile in a flow vs. time graph) in LWR model increases as the inflow approaches the peak capacity flow for reasons discussed in section 4.3.

In the region of impulse the outflow profiles are both quantitatively and qualitatively different from each other. To judge whether this difference is due to a rapid change in inflow, a light traffic condition (where the FTT model is a good approximate of LWR model) is considered next.

Rapidly Changing Inflow (Continuous Impulse) Under a Light Traffic Condition

The initial inflow applied to the link is $u(t) = 1.5\ t$. The impulse load superposed is a fifth order polynomial. The polynomial is constructed using the same assumptions as described in the previous case of rapidly changing inflow. Thus the impulsive function is,

$$u(t) = -0.2t^5 + 9.9t^4 - 152.5t^3 + 1146.9t^2 - 4186t + 5937.5 + t \quad \text{when, } 6 \le t \le 8. \qquad (4.17)$$

The rapid change is introduced on the sixth minute and removed at the eighth minute. The linear loading following the rapid change is $u(t) = t$. The inflow profile and the two exit flow profiles obtained from the FTT model and the LWR model are shown in figure 4.8. The exit flows are calculated as in the previous section. No significant pseudo periodicity was observed in case of the exit flow calculated from the FTT model. The solution for the LWR model is plotted considering the feasible solution from the multiple-solution zone.

Except for the zone of rapid change in outflow due to corresponding change in inflow, the outflow profiles from the two models match well. The outflow from the LWR model shows a higher maximum outflow than the FTT model. To examine the effect of rapid changes in inflow under a light traffic conditions in further detail, a discontinuous inflow function is used next to introduce impulsive loading.

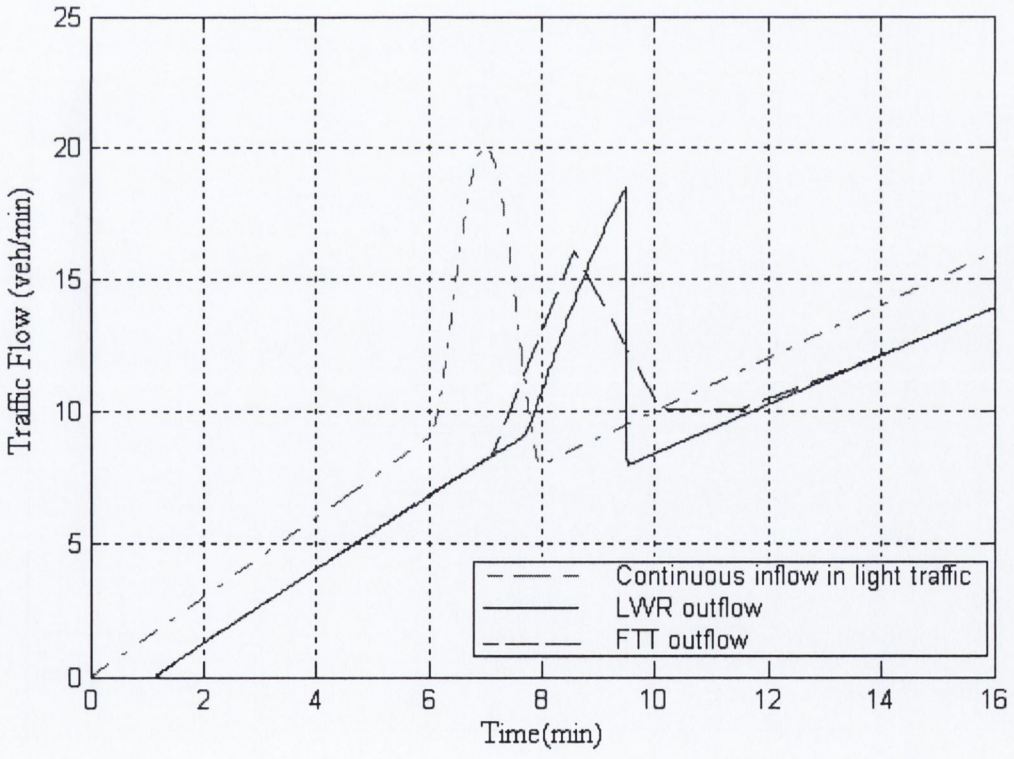


Figure 4.8 Inflow and Resulting Out-Flow Profiles for Both the Models for Rapidly Changing Continuous Inflow under a Light Traffic Condition.

Rapidly Changing Inflow (Discontinuous Impulse) Under a Light Traffic Condition

An existing linear inflow loading of $u(t) = t$ is considered for application to the link. At the fifth minute from the time of observation, a rectangular impulsive loading of the form $u(t) = t+5$ is introduced for a duration of one minute. After the sixth minute the extra load is removed. This sudden increase in load or the impulse is plotted in figure 4.9. For this inflow to the link, the exit flow/ outflow from the link is calculated using both the FTT and LWR models based on the solution algorithms described in this chapter.

In figure 4.9, the exit flow calculated using the FTT model shows pseudo-periodicity. This pseudo-periodicity has been reported for FTT model by Carey (2003). However this was unobservable with the cases of continuous impulses studied in this paper. The sudden and large discontinuity/change in the inflow is possibly the prime reason for pseudo-periodicity for the FTT solution as seen in figure 4.9.

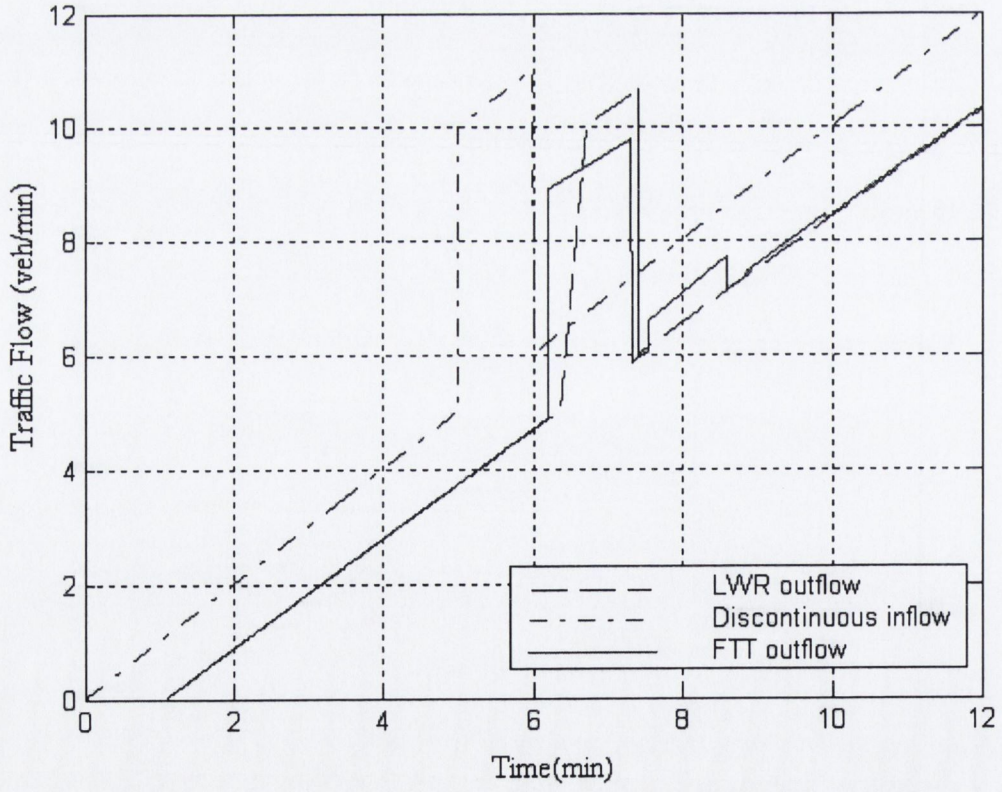


Figure 4.9 Inflow and Resulting Out-Flow Profiles for Both the Models for Rapidly Changing Discontinuous Inflow under a Light Traffic Condition.

As seen in the previous cases under light traffic condition, except for the zone of rapid change in outflow corresponding to the inflow, the outflow profiles from two models match well and the outflow from the LWR model show a distinctive higher maximum than the outflow from the FTT model. Due to the presence of pseudo periodicity the nature of the two outflow profiles become more dissimilar than the previous case with light traffic condition.

Slowly Changing and Rapidly Changing Inflows (Continuous Impulse) In a Congested/Heavy Traffic Condition

To study the effect of rapidly changing inflow and compare it with the effect of a slowly changing inflow would have on the FTT and LWR outflows under congested traffic conditions the following investigations are carried out. A rapidly changing inflow and a slowly varying inflow, rising to the same peak flow value as the rapidly changing one, are applied to the link in congested traffic conditions.

The slow varying non-linear inflow function is, $u(t) = 55\left(\frac{t}{t+1}\right)^2$. The rapid changing continuous inflow function is a fourth order polynomial, derived using the same assumptions described in previous cases of continuous impulse. The function is,

$$u(t) = 18t^4 - 280t^3 + 2114t^2 - 7751t + 11030 \quad \text{when, } 5 \le t \le 10. \tag{4.18}$$

Over an existing inflow loading of $u(t) = 1.5\ t$, the impulsive loading is applied at the fifth minute and removed at the tenth minute. The two inflow profiles are shown in figure 4.10. Exit flows from the LWR model and the FTT model are calculated under both slowly varying and rapid changing inflows and are plotted in figure 4.11.

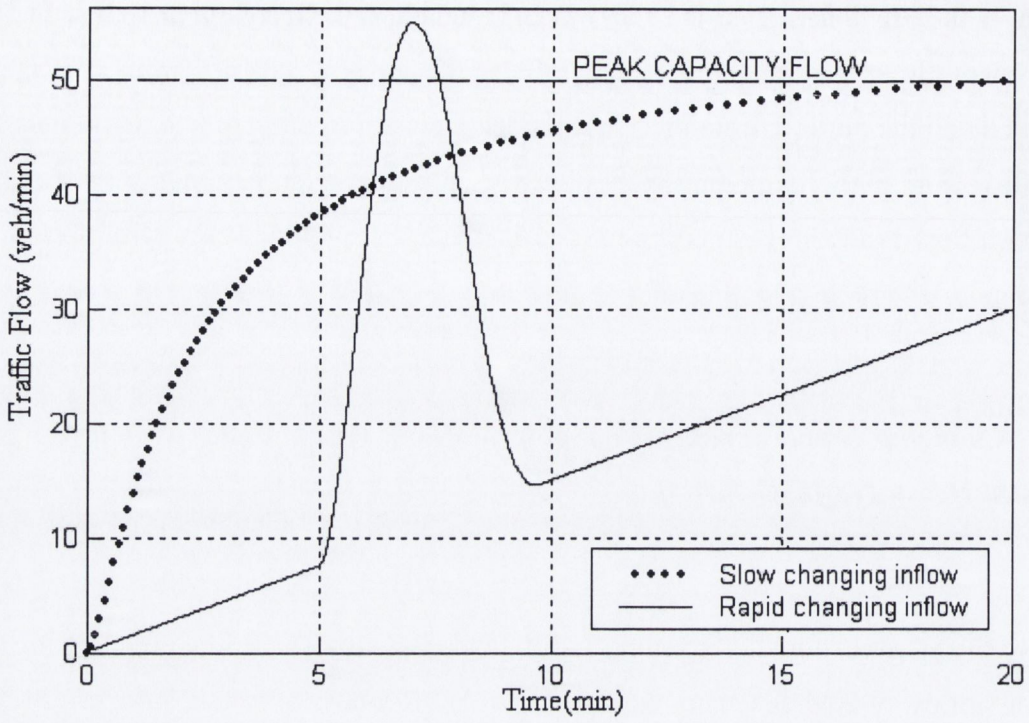


Figure 4.10 Slowly Varying Nonlinear Inflow and Rapidly Changing Inflow in Congested Traffic Condition.

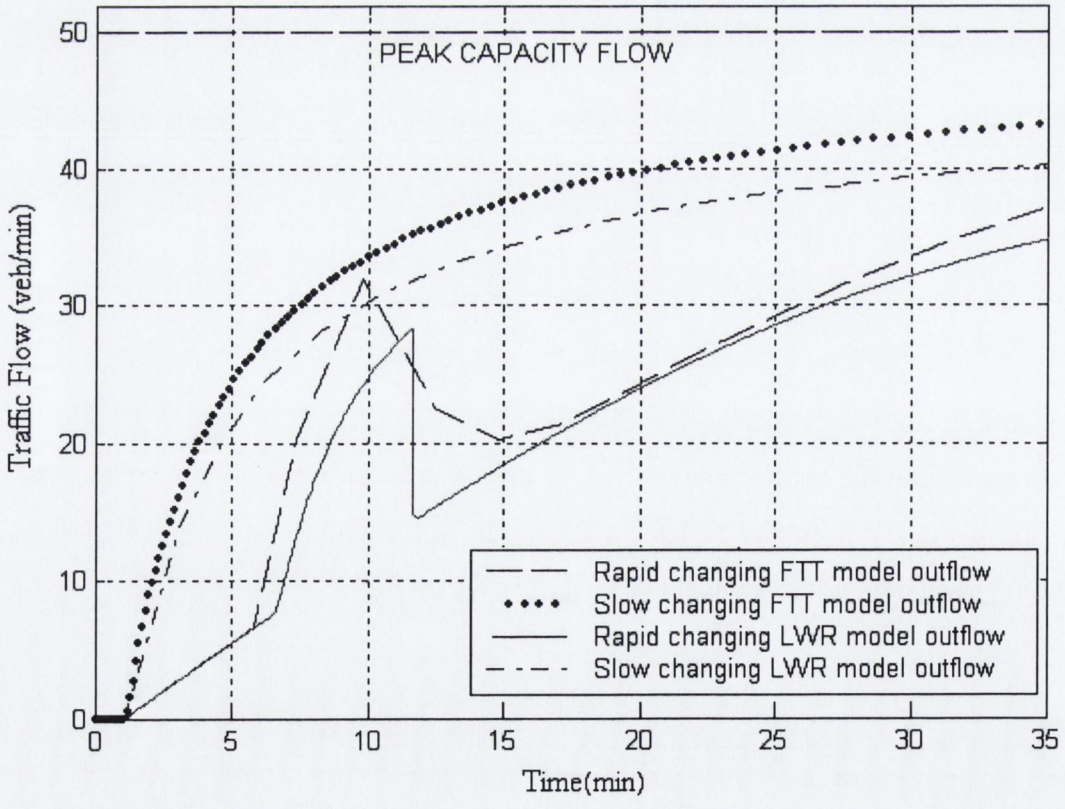


Figure 4.11 Out-Flow Profiles for Both the Models for Inflows Shown In Figure 4.10.

The exit flows calculated from the two models under the slowly varying inflow loading, show gradual increase in difference of magnitude with increase in inflow. But the nature of the outflow profiles remains the same and hence the FTT model could still very well be considered as an approximation to the LWR model. However, for a rapidly changing inflow the outflows from the FTT and the LWR models are qualitatively different. The FTT model gives quantitatively higher estimates than the LWR model for most of the time.

4.4.4 Comparison Using Loop-Detector Observations as Traffic Demand

To study the applicability of the FTT and LWR models in a real scenario, both the models are tested using the real time inductive loop-detector traffic flow observations. The traffic flow observations are obtained from the intersection TCS183 (figure 2.1) at the city-centre of Dublin. The traffic flow counts obtained from the SCATS database (UTC of Dublin) are during the green time of the signal. It is assumed that the traffic flow during the green time maintains a constant rate. A plot of the traffic flow observations on 22nd June 2005 from 7.00 a.m. to 7:10 a.m. is given in figure 4.12.

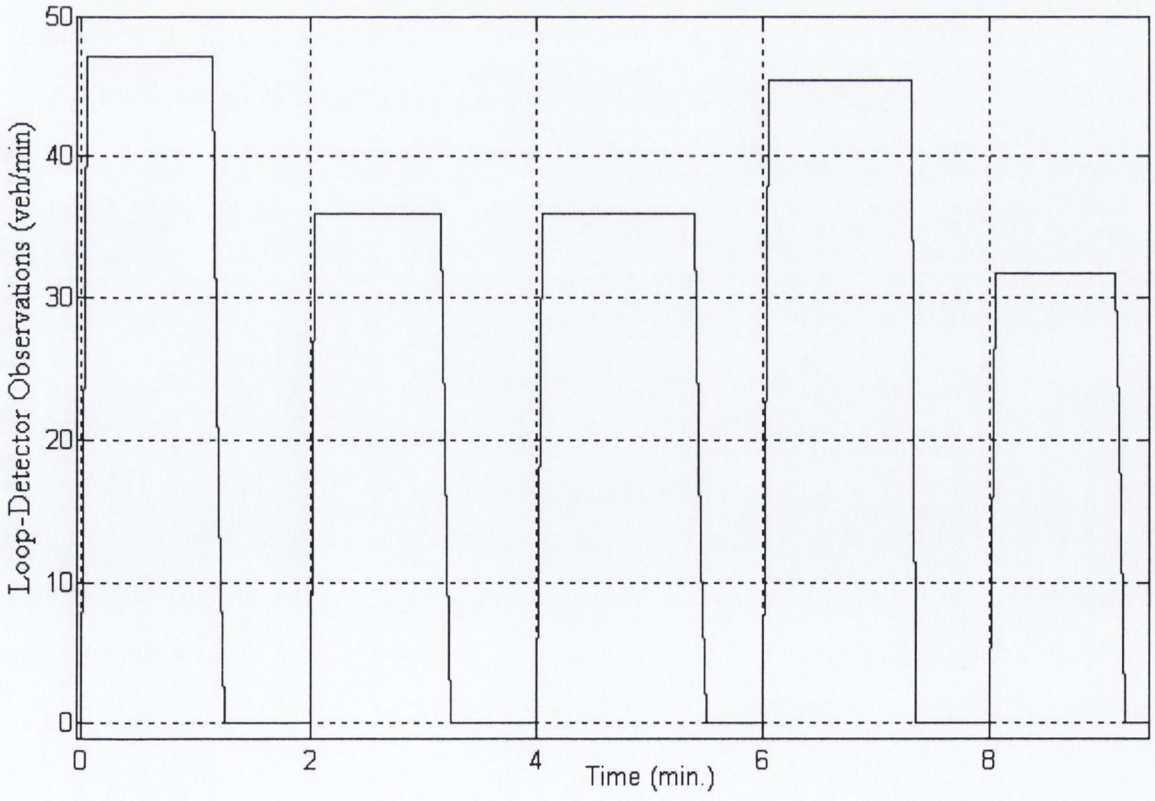


Figure 4.12 Loop-Detector Observations on 22-06-2005 at 7 a.m. to 7:10 a.m.

During 7 a.m. to 7:10 a.m. each cycle of traffic signal at intersection TCS183 is of 120 seconds. In figure 4.12, the traffic count during the red phase is taken as zero and during the green phase is taken as a constant rate.

The traffic flow plot shown in figure 4.12 is used as the inflow [$u(t)$] to both the models. The link length and the saturation flow parameters to be used in the FTT and LWR model are measured from the site. The link length, L, is approximately 120 metre and the saturation flow is 1744 veh/hr/lane. Considering 4 lanes of the junction, the peak capacity outflow from the link is taken as 130 vehicles/minute ($= 1/\beta$). The same free-flow speed as used in earlier comparisons is used here. In the linear travel time function considered for the FTT model, a free flow travel time of $\alpha = 0.3$ minute and $\beta = 0.0086$ min/veh. are used. The same conditions of comparison as described in section 4.4.1 holds for this case.

Similar solution algorithms as described in sections 4.2 and 4.3 are used in calculating the outflows. But as the inflows can not represented as a function of time (t), certain minor changes are required for the FTT solution algorithm. In travel time ($t+\tau(t)$) calculated using the previously described solution algorithm is approximated to the nearest half of a minute, so that the corresponding inflow rate, $u(t+\tau(t))$ values are known. Using these known values of $u(t+\tau(t))$, the corresponding exit flows are calculated using equation 4.3.

Exit flows from the LWR model and the FTT model are calculated using the loop-detector observations as inflows and are plotted in figure 4.13. The exit flows calculated from the FTT and LWR models are quite similar.

But the exit flow calculated using the FTT model shows pseudo-periodicity. This pseudo-periodicity is seen at the beginning of each green phase at the exit of the link. The travel time in LWR model increases continuously with queue formation and is different from the step increase of travel time in the FTT model at the beginning of each green phase at the exit. The travel time from LWR model is nearly constant during the end of each green phase. But it increases slightly during the same phase in the FTT model.

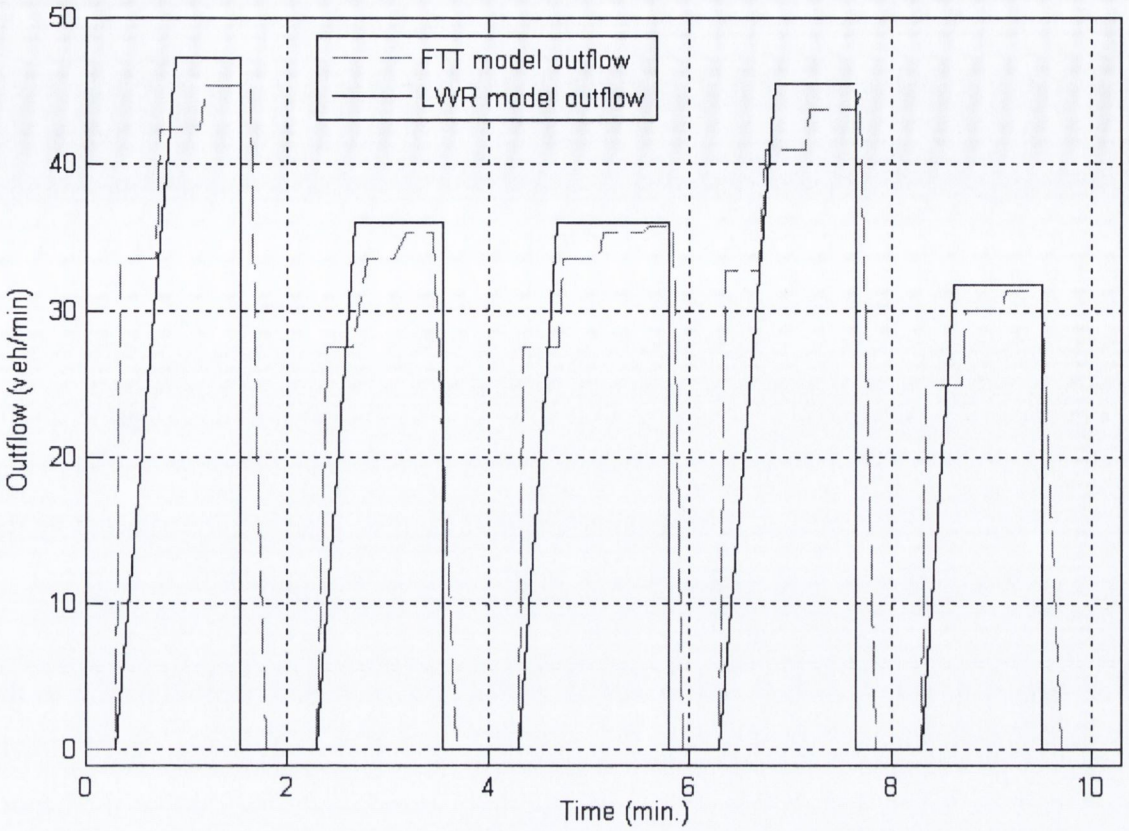


Figure 4.13 Out-Flow Profiles for Both the Models Using Loop-detector Observations Shown in figure 4.12.

4.5 COMMENTS ON THE MODELS

A few observations from the comparative study of the two models under slow and rapidly varying inflow are as follows.

- The FTT and LWR models studied in this chapter results in similar outflows under continuous and slowly changing inflow under light to moderate traffic conditions. The two models actually converge if the inflow is in the initial linear part of the flow density curve, i.e. under light to moderate traffic conditions. However, even under light traffic condition when the inflows change rapidly, the models yield exit flow profiles of different nature and magnitude in a certain region following the rapid changes in inflows.

- Under light traffic condition, the LWR model gives a higher maximum value of the outflow profile than that obtained from the FTT model. With the increase in traffic inflow, the LWR model starts giving lower estimates and in case of rapidly changing congested traffic conditions the FTT model overestimates the outflow (considering the LWR model as the bench mark as previous researchers comparing different macroscopic models).

- The effect of rapidly changing inflow persists for a longer duration of time on the outflow from the FTT model as compared to the LWR model. This effect including pseudo periodicity (though inherent to the FTT model), is seen to be affected by the nature of network loading (continuous or discontinuous impulse).

- The 'whole-link model' (FTT) can be considered as a good approximation to the hydrodynamic model (LWR) in case of light/moderate or even congested slowly varying traffic inflow. However, this approximation would have to be used more carefully, in case of sudden changes in the inflow. These sudden changes may be caused due to signal control of the traffic or due to occurrence of some incident on the link in a realistic traffic scenario.

Considering these observations and the comparison of both the models in a real time scenario, it can be concluded that for a signal controlled traffic, the LWR model gives continuous and more realistic exit flows from a link. But the analytical solution shown in this chapter may become very tedious if applied to a transport network consisting of several links and nodes or junctions. Hence, the Cell Transmission Model (Daganzo, 1994, 1995b) which is the first order finite difference numerical approximation of the LWR model is used in the subsequent chapters while working with loop-detector observations in a transport network.

CHAPTER 5

MULTIVARIATE SHORT-TERM TRAFFIC FLOW FORECASTING USING TIME-SERIES ANALYSIS

5.1 INTRODUCTION

The short-term traffic flow forecasting and simulation models described in chapters 2, 3 and 4 are all based on information obtained from a single site. Traffic management related to ITS involves simultaneous monitoring and control of multiple stations in a transport network. Hence, extension of the univariate traffic models to the multivariate regime is the next logical step towards developing efficient traffic flow models for ITS. Direct extension of the univariate SARIMA models to multivariate regime (Space-Time ARIMA, Vector ARMA) involves problems regarding number of parameters to be estimated and identifiablity of the estimated parameter values. A different class of time-series models called structural time-series models (STM) which can be easily extended to a multivariate regime are proposed in this chapter to model traffic flow observations from multiple sources within an urban transport network. The multi-input multi-output traffic flow forecasting model developed in this chapter can be applied to multiple junctions within a transport network which are not all interconnected as origin-destination pairs.

5.2 THEORETICAL BACKGROUND

The STM methodology is a particular time-series analysis technique which is set up in terms of components which have a direct physical interpretation (Harvey, 1989). The different components of STM are the trend, seasonal, cyclical and calendar variation together with the effect of explanatory variables and interventions (outlier and structural breaks). The basic principle behind a STM is similar to that of the HWES model, but more complex. Multivariate STMs are straightforward extension of the univariate STMs and involve less computational complexities than the other existing multivariate time-series techniques. An overview of the univariate and multivariate STM model definitions is given in this section. A detailed discussion on this subject is available in Harvey (1989) and Durbin and Koopman (2001). A software package called STAMP 6.0 (Structural Time-

Series Analyser, Modeller and Predictor) is used in this study for modelling traffic flow observations using STM.

Univariate STM Methodology

A univariate structural time-series model is formulated based on the unobserved components which have a direct interpretation in terms of the temporal variability of a time series dataset. Consequently the evolution of the components such as trend or seasonality over time and their contribution to the final predictions can be observed clearly. A univariate STM for a time series dataset y can be described by the following general equation involving all possible types of temporal components in its form:

$$\begin{cases} y_t = \mu_t + \gamma_t + \psi_t + \nu_t + \varepsilon_t, \\ \\ \varepsilon_t \sim \text{NID}\ (0, \sigma_\varepsilon^2) \end{cases} \qquad t = 1, \ldots, T \qquad (5.1)$$

where, μ_t is the trend, γ_t is the seasonal, ψ_t is the cycle, ν_t is the first-order AR component and ε_t is the irregular or the random error. For the purpose of traffic flow modelling, the univariate and multivariate STM are considered to be comprised of three components; stochastic trend, seasonality and irregular. Hence, the equation 5.1 reduces to the following form:

$$\begin{cases} y_t = \mu_t + \gamma_t + \varepsilon_t, \\ \\ \varepsilon_t \sim \text{NID}\ (0, \sigma_\varepsilon^2) \end{cases} \qquad (5.2)$$

The stochastic trend component (μ_t) represents the long-term movement in a time-series which can be extrapolated into the future. In the case of traffic flow observations over a few weeks from a developed urban transport network, this long-term movement does not show any significant gradient and should be modelled for the local fluctuations. A Markov model of the stochastic trend can be considered in this purpose.

$$\begin{cases} \mu_t = \mu_{t-1} + \eta_t \\ \\ \eta_t \sim \text{NID}\ (0, \sigma_\eta^2) \end{cases} \qquad \text{(Change of slope is not considered)} \qquad (5.3)$$

The irregular disturbance/variance σ_ε^2 and the stochastic trend (level) variance σ_η^2 are mutually uncorrelated.

The periodic nature of a time-series dataset can be modelled in a trigonometric form in STM.

$$\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{j,t} \tag{5.4}$$

where each $\gamma_{j,t}$ is generated by

$$\begin{bmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{bmatrix} = \begin{bmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{bmatrix} + \begin{bmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{bmatrix}, \quad \begin{matrix} j = 1,\ldots,\left[s/2\right], \\ t = 1,\ldots,T, \end{matrix} \tag{5.5}$$

where $\lambda_j = 2\pi j / s$ is the frequency, in radians, and the seasonal disturbances ω_t and ω_t^* are mutually uncorrelated random normal disturbances with zero mean and common variance σ_ω^2. When s is even, the equation 5.5 at $j = s/2$ collapses to

$$\gamma_{j,t} = \cos\lambda_j \gamma_{j,t-1} + \omega_{j,t} \tag{5.6}$$

The equations 5.1 to 5.5 define the STM used in this study. The disturbances of the individual components of the STM ($\sigma_\varepsilon^2, \sigma_\eta^2$ and σ_ω^2) mentioned in these equations are all mutually uncorrelated. These equations are generally solved in state-space form using Kalman filter based algorithms (Kalman, 1960; Harvey, 1989).

In some cases, the time series observations to be modelled have dynamic relationships with some other independent variables. They are called explanatory or exogenous variables and inclusion of these variables in STM, may improve the forecasting precision of the models. The inclusions of the explanatory variables change the first part of equation 5.1 to the following form:

$$y_t = \mu_t + \gamma_t + \psi_t + \nu_t + \sum_{i=1}^{k} \sum_{\tau=0}^{q} \Delta_{i\tau} x_{i,t-\tau} + \varepsilon_t \tag{5.7}$$

where, $x_{i,t-\tau}$ is an exogenous variable, k is the total number of exogenous variables, τ is the time lag and $\Delta_{i\tau}$ is a set of unknown constants. The significance of τ is that in some cases, the lagged value of the dependent variable can be considered as an exogenous variable in STM.

Multivariate STM Methodology

Multivariate time-series data can be chiefly classified into two distinct types, Panel data and interactive data (Harvey, 1989). In the case of panel data, the time-series variables are subjected to the same or similar influences but the individual elements do not interact with each other. As the variables follow similar temporal nature, they can be modelled jointly. In contrast, multivariate interactive time-series data consists of a set of variables which have some behavioural relationships among themselves and interact dynamically with each other. This distinction is important to find out the suitable type of multivariate structural time-series analysis technique. The multivariate traffic flow observations (i.e. observations from different stations in the same transport network) modelled in this chapter are considered to be subjected to similar influences but not to have any dynamic interaction (detailed explanation in section 5.3) and are modelled as panel data.

The panel data can be modelled by a multivariate structural time-series (MST) technique where the various components of the different time-series variables are allowed to be contemporaneously correlated. Such a type of MST model is referred to as, *seemingly unrelated time-series equations* (SUTSE) (Harvey, 1989). The univariate equations described in the previous subsections can easily be extended to SUTSE model.

$$\begin{cases} y_t = \mu_t + \gamma_t + \varepsilon_t, \\ \\ \varepsilon_t \sim \text{NID}\,(0, \Sigma_\varepsilon) \end{cases} \qquad t = 1, \ldots, T \tag{5.8}$$

where, y_t is a vector of $N \times 1$ time-series observations which depends on the unobserved trend component , μ_t seasonal component γ_t and irregular component ε_t which are also vectors. Σ_ε is the $N \times N$ variance matrix of the irregular disturbances. In multivariate regime, the equations 5.3, 5.4, 5.5 and 5.6 change in a similar manner as in the case of

equation 5.2. The various unobserved components of the univariate STM now become vectors in the MST model and the disturbances of these components become $N \times N$ variance matrices.

The inclusion of explanatory variables in the MST model is simple and similar to the univariate approach described in equation 5.7.

$$y_t = \mu_t + \gamma_t + \sum_{\tau=0}^{q} \delta_\tau x_{t-\tau} + \varepsilon_t \tag{5.9}$$

where, x_t is a vector of $K \times 1$ explanatory variables. Elements of the unknown parameter matrix δ_τ can be specified to be zero, thereby excluding certain explanatory variables from particular equations.

A vector process y_t is said to be homogenous if all linear combinations of its N elements have the same stochastic properties. In a multivariate homogenous system, the disturbance matrices are required to satisfy the following equation:

$$\Sigma_k = q_k \Sigma_* \quad \text{and} \quad \Sigma_\varepsilon = h\Sigma_* \tag{5.10}$$

where, q_k, $k = 1,\ldots\ldots, g$ and h are non-negative scalars and Σ_* is an $N \times N$ matrix. In the cases possible assumption of homogeneity in a SUTSE model decreases the computational complexity to a great extent.

Comparison of STM and SARIMA Methodologies

A comparative discussion on the advantages and disadvantages of the newly introduced STM methodology and previously studied SARIMA techniques (in chapters 2 and 3) is presented in this section. The key points to mention in this context are as follows:

- One of the major advantages of STM methodology is its transparency (Durbin and Koopman, 2001). In STM, the different components which make up the time-series are

modelled separately unlike the SARIMA methodology where the trend and seasonal components are eliminated by differencing. Hence, in STM it is easy to understand the evolution and contribution of each of the components in the final results. Also, due to the recursive nature of the STM models any known structural change over time is easy to implement. On the other hand, SARIMA models cannot include such structural changes since they are homogenous and stationary in form. Treatment for missing values in a time-series is also very simple in STM equations. Explanatory variables, outliers, structural breaks etc. can also be easily modelled in a STM framework as described in the previous sub-sections. Introduction of the same variables in SARIMA models require tedious computational efforts.

- In a multivariate regime, MST models are straightforward vector extensions of the univariate STM. The SUTSE models do not involve estimation of huge co-variance matrices as Vector ARMA (VARMA) models. Apart from that, similar to the univariate STM, explanatory variables can be included easily to the MST model equations.

- Classical time-series analysis using the ARMA (SARIMA, ARIMAX) class of models is based on the theory of stochastic statistical processes. Hence the time-series dataset to be modelled using SARIMA technique is always required to be checked for its stationarity. If the dataset is not stationary, then transformations are required to be performed to achieve weak stationarity (ref. chapter 2). As STM is not based on this theory, no such transformations or check are required for the application of this methodology. This can be considered as an advantage considering the aspect of application to a real time traffic data modelling scenario. However, the STM methodology being dependent on the State Space formulation does not have a strong base in classical statistical theories. Another point in favour of SARIMA model is that, this class of models being hugely popular in the field of time-series analysis, numerous user-friendly application based commercial software are available for univariate time-series analysis. The only software available for univariate and multivariate STM are STAMP (Koopman et al., 1999a) and SsfPack (Koopman et al., 1999b).

To sum up, structural time-series models are more general, more flexible and can be easily transferred from a univariate to multivariate regime than the ARIMA class of models.

Hence, in this study for developing a parsimonious and efficient multivariate traffic flow model, STM methodology is favoured to SARIMA technique.

5.3 PROPOSED MULTIVARIATE TRAFFIC FLOW TIME-SERIES MODEL

A multi-input multi-output (where the number of input intersections are more than number of output intersections) short-term traffic flow simulation and forecasting model is proposed in this study for efficient modelling of traffic in a congested urban transport network. The proposed model is developed for a set of 15 minute aggregate traffic volume observations from different approaches and different intersections of the transport network.

Unlike the previous multivariate traffic flow models developed for urban transport networks (Stathopoulos and Karlaftis, 2003; Kamarianakis and Prastacos, 2002), the locations of the sites of data collection within the transport network are not required to be considered in the proposed methodology. The aim of this approach is to develop a multivariate traffic flow simulation and forecasting model for multiple intersections within a transport network which may not be situated on the same route. As the intersections or stations of observations are not situated on the same route it is highly unlikely that the same platoon of vehicles will pass through different intersections at different time instants. Hence, information about the directions of traffic flow is not essential. A SUTSE model is ideal for modelling such multivariate time-series observations as the behavioural relationship among the variables are not considered.

As an improvement to the panel data modelling methodology a spatial dimension is introduced to the proposed multivariate traffic flow model. The traffic flow observations from the nearest available upstream intersection of each of the modelled data collection stations are included as explanatory variables to the MST model equations. This ensures that the effect of any abrupt change occurring in the upstream junction can be accounted for in the model.

Case Study

The proposed multivariate traffic flow forecasting methodology is applied to a congested urban transportation network at the city centre of Dublin to test the effectiveness of the

forecasting strategy. A small network of ten intersections within the transport network is chosen in this purpose. 15 minute univariate aggregate traffic flow observations from each of the ten intersections and their nearest available upstream junctions are modelled using the proposed multivariate traffic flow model. There are two important points which are required to be checked before applying the proposed methodology.

1. The location and the distance of the sites chosen should be arbitrary. The average travel time between two sites on the same route within the network should not be more than 15 minutes. Considering a 30 km/hr free flow speed within a congested urban transport network, the radius of the simulation network should not be more than 7.5 km.

2. The second and the most important point is that the univariate traffic flow observations from different data-collection site should not have behavioural or physical relationship among themselves. If and only if the multivariate traffic flow observations behave as a panel data set, the SUTSE model can be applied.

In the figure 5.1 a map of the chosen urban transport network at the city-centre of Dublin is given. The ten junctions where the multivariate time-series model is applied for short-term traffic volume simulation and prediction are shown with numbered yellow squares in the map. In the figure, the direction of the univariate traffic movement at each intersection is shown with a pink arrow. The origin of the pink arrow is marked with a numbered dark brown circle which signifies the nearest upstream junction to each intersection from which traffic volume data can be obtained. The length of the pink arrow signifies the distance between an intersection and its nearest available upstream junctions. If this distance is considerably large then it is possible that the changes in traffic flow at the upstream junction may not influence the traffic flow at the downstream intersection.

It is evident from the figure, that the choice of the site locations at which 15 minute traffic volume is modelled is arbitrary. The chosen intersections are not on the same route within the transport network and none of the two stations have a distance of more than 7.5 km. Hence, the chosen network of ten intersections conforms to the conditions mentioned in point one. The ten intersections at which the proposed multivariate short-term traffic flow model is applied for simulation and forecasting are termed as output intersections in the rest of the text.

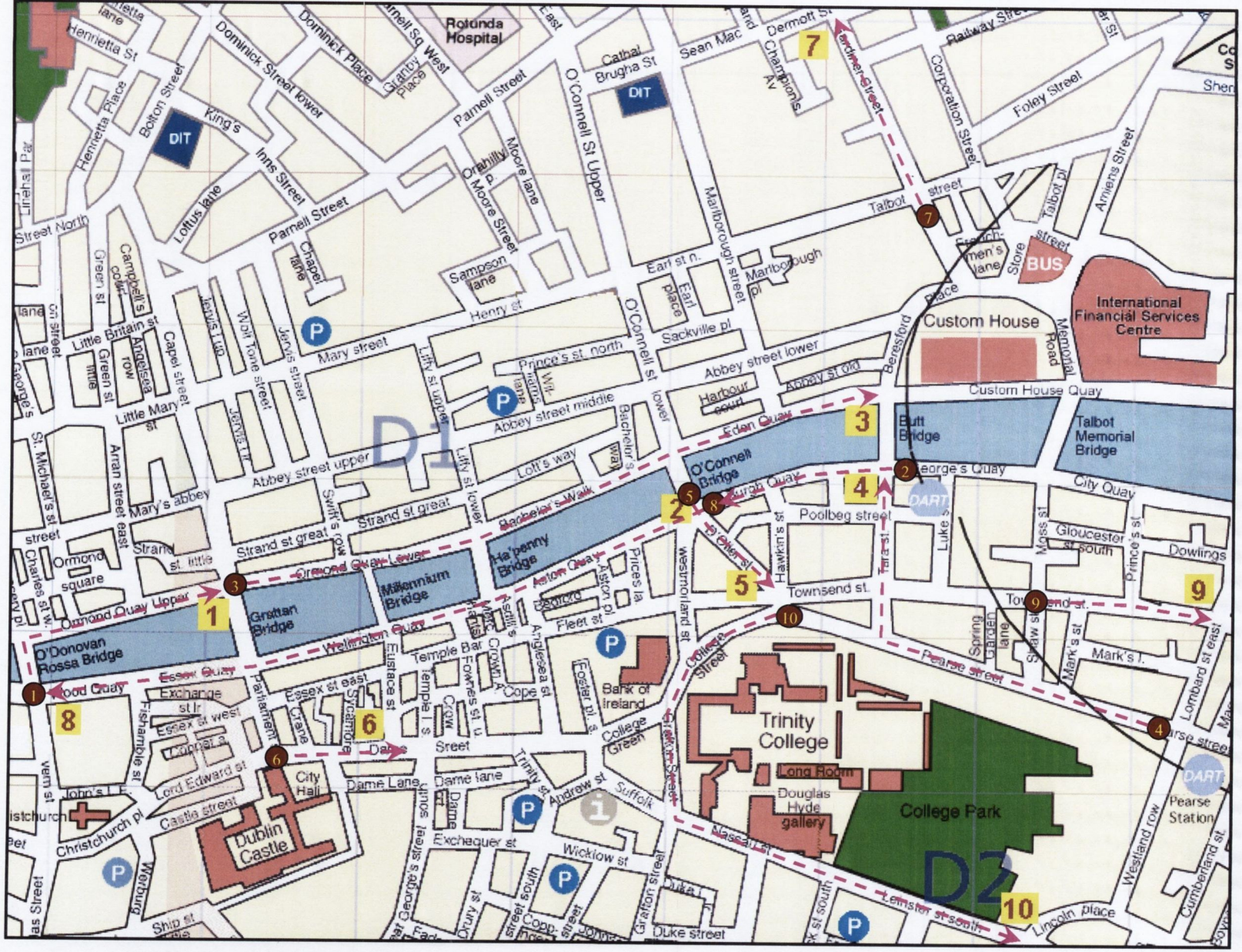


Figure 5.1 Map of the chosen Transport Network.

Station in Map	Intersection Name	Data Collecting Loop-Detectors	Upstream Junction & Loop-Detectors
1	TCS 26	4, 5, 6	TCS 182 (1, 2)
2	TCS 196	10, 11, 12, 13	TCS 193 (5, 6, 7)
3	TCS 17	5, 6, 7, 8	TCS 183 (5, 6, 7)
4	TCS 183	1, 2, 3, 4	TCS 26 (4, 5, 6)
5	TCS 232	1, 2, 3, 4	TCS 146 (1, 2, 3, 4)
6	TCS 49	1, 2,	TCS 48 (1, 2)
7	TCS 166	1, 2, 3	TCS 188 (3, 4)
8	TCS 193	1, 2, 3, 4	TCS 232 (6, 7)
9	TCS 439	1, 2, 3	TCS 196 (6, 7, 8, 9)
10	TCS 269	1, 2, 3	TCS 196 (10, 11, 12, 13)

Table 5.1 Details of the Ten Output Sites

In table 5.1 further details about the ten output intersections are given along with the name of the nearest upstream intersection at which traffic flow observations are available. The 15 minute aggregate univariate traffic volumes from the mentioned loop-detectors in the upstream junctions are used as explanatory variables in the SUTSE model equations. In equation 5.9, the elements of the matrix δ_τ are so chosen that the forecasts from each intersection is affected only by the changes at its upstream junction and not by the changes at other upstream intersections.

The traffic flow observations used for modelling from all the chosen intersections were recorded from 3rd November 2003 6:30 a.m. to 26th November 2003 6:30a.m., excluding the weekends. A cross-section time-series plot of the traffic flow observations from the ten output intersections during 4th and 5th of November in 2003 is given in figure 5.2. The plot shows that there is a definite temporal similarity among the curves. The output junctions at which the direction of the univariate traffic flow fall on the routes towards the city-centre have high traffic volumes during the morning peak hours whereas the junctions for which the same fall on the routes away from the city-centre have higher traffic volumes during evening peak hours than the morning peak. Consequently, the peak hourly volumes from all these ten output junctions may not have high positive correlations, but the time of occurrence of the maximum traffic volumes passing through the junctions are very similar. Considering this contemporaneous correlation among the ten output traffic volume time-series datasets, they can be modelled as panel data using SUTSE models.

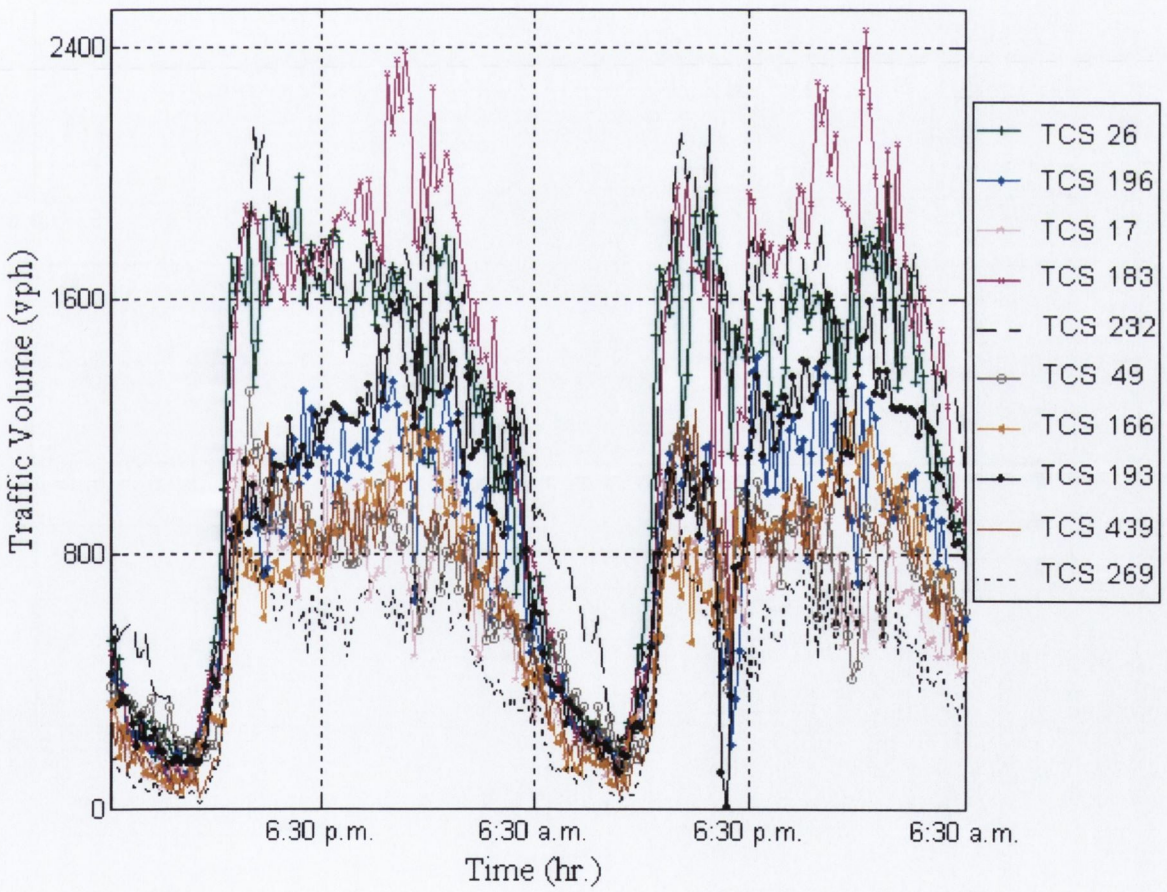


Figure 5.2 Plot of Two day Traffic volumes from Ten Output Intersections.

All the ten series of traffic flow observations are modelled using SUTSE models with equations 5.9 and equations 5.3, 5.4, 5.5 and 5.6 in their vector forms. For all the ten

output junctions, 50 points in the future are forecasted (figure 5.3). The traffic flow data obtained on the 26-11-2003, or data collected in the next 12.5 hours (50x15 = 750minute = 12.5hours) i.e. 6:30 a.m. to 8:00 p.m. are compared with these forecasts.

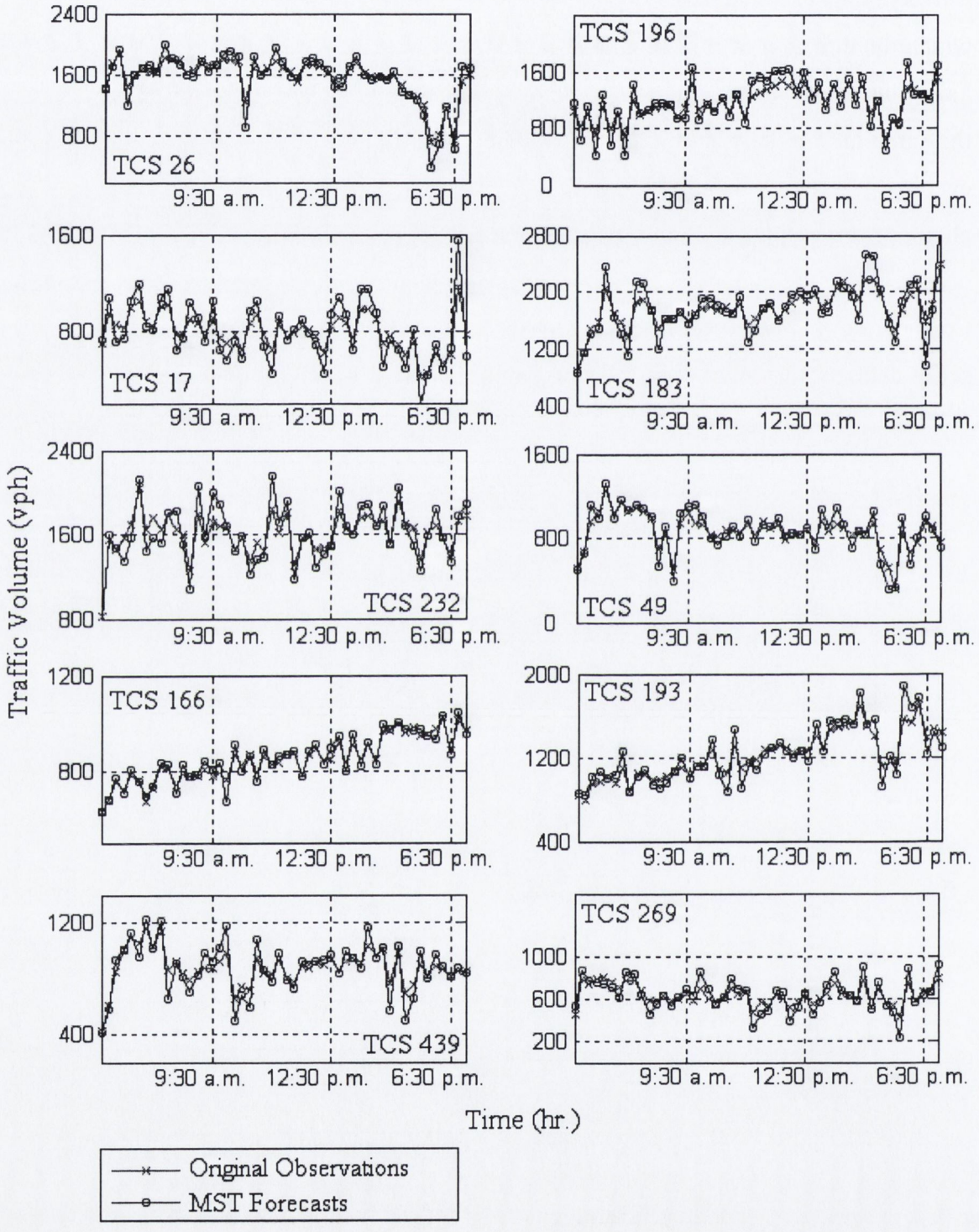


Figure 5.3 Forecasts from Ten Output Intersections from the SUTSE model.

The forecasting precision (MAPE) from the SUTSE model for each of the ten output junctions with and without considering the influence of upstream junctions are given in table 5.2.

Intersection	**Distance from Upstream junction**	**MAPE** (with upstream data)	**MAPE** (without upstream data)
TCS 26	Close	5.89%	11.47%
TCS 196	Close	10.9%	12.57%
TCS 17	Far	12.66%	12.12%
TCS 183	Far	7.4%	7.19%
TCS 232	Close	6.52%	9.24%
TCS 49	Close	7.96%	12.88%
TCS 166	Close	4.94%	8.28%
TCS 193	Far	6.2%	6.7%
TCS 439	Close	7.4%	7.6%
TCS 269	Far	10.1%	9.8%

Table 5.2 Forecasting Errors from Ten Chosen Intersections

From figure 5.3 and table 5.2, it can be seen that the MAPE value for the junctions improve significantly in case of most of the junctions when the traffic flow observations from the nearest available upstream junctions are incorporated in the SUTSE model as explanatory variables. The four junctions for which the MAPE value do not improve or deteriorate do not have any upstream junction nearby from where loop-detector observations can be available. In cases like this where the nearest upstream junction is situated at a considerable distance from the output intersection, it is wise to ignore the influence of the traffic volume changes at the upstream junction.

The MAPE value from the univariate SARIMA model for the traffic flow observations at the intersection TCS 183 during the month of November in 2003 was around 11% (Chapter 2, Table 2.3) in comparison to the MAPE value from the SUTSE model of 7.19 % for the same junction. Hence, the proposed multivariate traffic flow time-series model is a more efficient technique than the ordinary univariate SARIMA model for short-term simulation and forecasting of traffic volume in a congested urban network.

5.4 DISCUSSION ON MST TRAFFIC FLOW MODEL

The multivariate structural time-series model developed in this chapter has two distinct advantages over the univariate time-series models;

- Simultaneous simulation and modelling of traffic conditions at multiple intersections.

- Consideration of the effect of changes in traffic conditions at the immediate upstream junction to improve the predictions at the downstream output junction.

The forecasting precision of the SUTSE methodology based multivariate short-term traffic flow model is better than the same from the univariate SARIMA models. But, the distance of the nearest available upstream junction from the output intersection influences the forecasting precision to a great deal. Hence, for developing efficient and robust multivariate short-term traffic flow forecasting algorithm it is necessary to understand the movement of traffic between the upstream junctions and the forecasting sites. To address this issue, the concept of merging the 'traffic flow theory' models with the statistical time-series models, introduced in chapter 4 in a univariate scenario should be extended to a multivariate regime. Due to the availability of software and ease of modelling, the SARIMA class of univariate time-series models are used in this merging scheme instead of the univariate structural time-series models discussed in this chapter.

CHAPTER 6

MULTIVARIATE SHORT-TERM TRAFFIC FLOW FORECASTING MERGING C.T.M. AND UNIVARIATE TIME-SERIES ANALYSIS

6.1 INTRODUCTION

A multivariate traffic flow time-series model was developed in chapter 5 for efficient short-term simulation and forecasting of traffic volume data in an ITS equipped urban congested transport network. The application of the model to a real-life transport network illustrated the importance of merging the 'traffic flow theory' models with the statistical time-series models. To develop a multivariate multi-step traffic forecasting model which can capture the traffic flow dynamics as well as the temporal variation of traffic conditions, the SARIMA time-series forecasting technique is integrated with the Cell-Transmission Model (CTM) in this chapter.

CTM has excellent capability to capture traffic dynamics and dynamic traffic interactions across multiple links. On the other hand, the SARIMA model can capture the within day, daily, weekly, monthly and other seasonal effects in traffic flows. Hence, the SARIMA model is used to predict future traffic volume counts at sites defined as the origins or entry points of a transport network. Using the predictions as traffic demand input, CTM model is used to simulate the traffic flows at junctions in the network where no continuous data collection is taking place. The proposed space-time approach is particularly suitable wherein loop-detectors are often not used or not installed at non-critical junctions inside a congested traffic network but can be affected due to queue spillback which may often occur at these locations. To test the effectiveness of the forecasting strategy, a case study in the Dublin city-centre with severe congestion is performed.

6.2 OVERVIEW OF CELL TRANSMISSION MODEL

The cell transmission model (CTM) (Daganzo, 1994 & 1995b) is a first order finite difference based numerical approximation of the Lighthill-Whitham-Richards (LWR) model. The LWR model (Lighthill & Whitham, 1955; Richards, 1956) or the

hydrodynamic theory of the traffic flow underlies most of the present day macroscopic traffic operation models. It captures the full range of the fundamental flow–density-speed relationships including shockwaves, queue formation and queue dissipation features both in congested and un-congested regime. The model and its solution, based on the *method of characteristics*, are described in chapter 4. The model is applied to a single link in the chapter. To implement the model in a transport network solutions based on numerical methods are more appropriate. CTM is the finite difference solution developed by Daganzo (1994, 1995b) adopting a simplified flow-density relationship to the LWR problem.

6.2.1 Basic Principles on A Simple Highway

According to Daganzo (1994), if the flow (q)-density (k) relationship on a simple highway can be expressed in the form,

$$q = \min\left\{vk, Q_{\max}, w(k_j - k)\right\}, \qquad \text{for } 0 \le k \le k_j \tag{6.1}$$

then the LWR equations can be approximated by a set of difference equations where state of the system can be updated with the tick of a clock. In the equation 6.1, v, $Q_{\max}$, w, k_j are free-flow speed, the maximum flow (or capacity), the shockwave speed and the jam-density respectively.

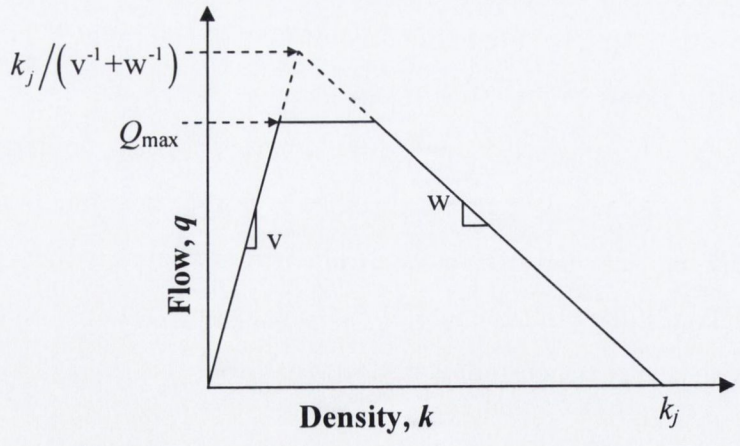


Figure 6.1 The Trapezoidal and Triangular flow-Density Relation.

Though CTM can apply the any flow-density relationship, CTM either uses triangular or trapezoidal flow-density curve for depicting flow propagation (figure 6.1) which simplifies the computational effort to a great extent. According to the standard trapezoidal flow-

density curve used for CTM, there is a constant free-flow speed (higher speed) at low densities and a constant shockwave speed (always less than the free-flow speed) at high densities. Empirically it can be shown that the free-flow speed gradually decreases mainly while the density approaches the flow capacity and otherwise it is fairly constant over a wide range of low densities.

In CTM, the road or the link is divided into homogenous sections (or cells) and the time is divided into intervals such that each cell length is equal to the distance travelled by free-flowing traffic in one time interval. The figure 6.2 shows the basic building block of CTM depicting the flow propagation relationship between cells i and $i+1$.

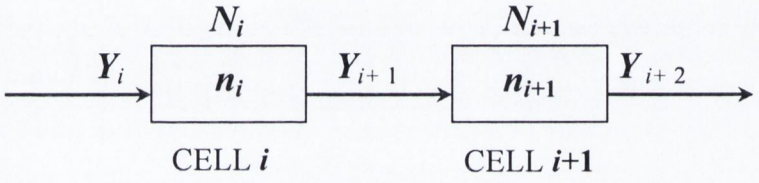


Figure 6.2: The Basic CTM Model for a Single Link with 2 Cells at Any Given Time Step. Ref: Lo et al (2001)

$Y_i(t)$ is the inflow to cell i and $Y_{i+1}(t)$ is the outflow from cell i or inflow to cell $i+1$ at the same time instant t. The total number of vehicles in the cell is $n_i(t)$ and $N_i(t)$ is the holding capacity of cell i . $N_i(t)$ and $Q_{\max}$ can vary with time, but this variability is not considered here.

$$N_i = k_j n_1 L \tag{6.2}$$

where, n_1 is the number of lanes and L is the length of each cell. $Q_{\max}$ is a product of time interval or time step and capacity of each cell. In CTM, the LWR model is approximated by the following set of recursive equations (Daganzo, 1994, 1995b):

1. The equation of state:

$$Y_{i+1}(t) = \min\{ n_i(t), Q_{\max}, \delta [N_{i+1}(t) - n_{i+1}(t)]\} \tag{6.3}$$

where, Q_{max} is the maximum number of vehicles that can enter cell $i+1$ at any single tick of clock; $[N_{i+1}(t) - n_{i+1}(t)]$ is the available space in cell $i+1$ and δ is the ratio of shockwave

speed to free-flow speed (w/v). This equation covers both congested and un-congested scenarios. In light traffic conditions, the number of vehicles in the upstream cell, $n_i(t)$, constrains flow. In congested traffic, the last term depicting available space in the downstream cell constraints flow. In a bottleneck scenario, the middle term decides the inflow to the downstream cell (Lo, 2001).

2. The conservation equation:

$$n_i(t+1) = n_i(t) + Y_i(t) - Y_{i+1}(t), \tag{6.4}$$

This equation updates the flow in consecutive cells at each time step. Equations 6.3 and 6.4 provide the basic principle of modelling traffic flow on a series of straight cells, representing a simple highway or link. On a finite road, the first and the last cell of the link should ascertain the boundary conditions. The first cell should act as a gate cell. An origin cell is to be introduced before the gate cell which will allow only the link travel demand as inflow to the gate cell. The origin cell should act like a parking lot. The destination cell or the last cell of the link should act as a sink and should have infinite holding capacity.

If the cells are signalized, then the maximum holding capacity can be represented as a binary variable. For any signalized cell i, $Q_i(t)$ varies as follows (Lo, 1999, 2001)-

$$Q_i(t) = \begin{cases} Q_{\max} & \text{if } t \in \text{ green phase,} \\ 0 & \text{if } t \in \text{ red phase.} \end{cases} \tag{6.5}$$

6.2.2 Extension to Networks

A network consists of an ensemble of directed links and nodes. Daganzo (1995b) suggested that if a maximum three links are attached to a node, then three types of scenario may arise. Daganzo (1995b) suggested the CTM representation of the first two network topologies (*merges* and *diverges*) in his paper. Any other types of junctions can be broken down to an ensemble of merges, diverges and ordinary links. Chang (1998) further developed these representations for '*signalized merges*' and '*turning lane diverges*' case. Here the three basic building blocks for representing a traffic network in CTM framework are discussed.

1. *ordinary* if one link enters and one leaves the node (figure 6.2)
2. *merge* if two links enter a node but one link leaves it (figure 6.3)
3. *diverge* if two links leave and one link enters (figure 6.4)

Ordinary Links

Ordinary links can be modelled for CTM using the equations described in the last subsection. The only improvements are required while modelling two consecutive cells from two different links with different free-flow speeds. Then the variables $Q_{\max}$ and δ for the sending cell and the receiving cell can be different. Hence, equation 6.3 will change to,

$$Y_{i+1}(t) = \min\{ n_i(t), \min\{Q_{\max}(i), Q_{\max}(i+1)\}, \delta_i [N_{i+1}(t) - n_{i+1}(t)]\} \quad (6.6)$$

The time variability of $Q_{\max}$ is not considered in equation 6.6 as in equation 6.3. To simplify equation 6.6, two different variables can be considered.

- Maximum flow that can be sent,

$$S_i(t) = \min(n_i, Q_{\max}(i)); \quad (6.7a)$$

- Maximum flow that can be received,

$$R_{i+1}(t) = \min\{Q_{\max}(i+1), \delta_i[N_{i+1}(t) - n_{i+1}(t)]\}; \quad (6.7b)$$

Hence, the new form of equation 6.6 is,

$$Y_{i+1}(t) = \min\{S_i(t), R_{(i+1)}(t)\} \quad (6.8)$$

The equations 6.4 and 6.5 will be applicable in their stated form in modelling the ordinary links.

Unsignalised and Signalised Merges (Daganzo, 1995b)

Merges are one of the most important movements to be modelled while representing a network in CTM framework. Figure 6.3 shows a 'merge' manoeuvre.

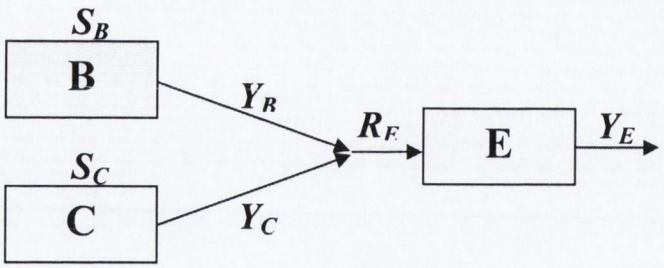


Figure 6.3 A Merge Manoeuvre.

In figure 6.3, S_B and S_C are the maximum possible outflows from the two sending cells **B** and **C** respectively while R_E is the maximum possible inflow to cell **E;** Y_B and Y_C are the outflows from cells **B** and **C.** The index t for these notations is dropped for simplicity. S_B, S_C and R_E are to be calculated using equations 6.7a and b. p_B and p_C are the merge ratios (not shown in the figure), representing the proportions of R_E coming from cells **B** and **C** respectively with their sum equal to 1. p_B and p_C are determined by the design of any intersection with merging manoeuvre. These parameters can vary with time, but the time dependence is not considered here.

The flows shown in figure 6.3 must satisfy the following conditions,

$$Y_B(t) \le S_B(t); \quad Y_C(t) \le S_C(t); \tag{6.9a}$$

and, $$Y_B(t) + Y_C(t) \le R_E(t) \tag{6.9b}$$

Three types of causality scenarios may arise in case of an unsignalised merge.

1. Free flow condition downstream and no queuing upstream to the node;
2. Queuing in both the approaches due to downstream congestion;
3. Flow in one approach (or the priority approach) blocking the flow of the other (or the complementary approach).

In case 1, when outflows from cells **B** and **C** are controlled by the conditions downstream (cell **E**), the following equations must govern the state of traffic.

$$\text{if } R_E \geq S_B + S_C, \quad \text{then} \begin{cases} Y_B(t) = S_B \\ \\ Y_C(t) = S_C \end{cases} \tag{6.10}$$

In case 2, the conditions in both the sending cells are controlled by the conditions downstream and in case 3, the priority approach is controlled by the conditions downstream whereas the flows in the other sending cell is controlled by the conditions upstream. In both these cases, equation 6.10 is not applicable. The following equations will govern the state of traffic in those cases,

$$\text{if } R_E < S_B + S_C, \quad \text{then} \begin{cases} Y_B(t) = \text{mid}\{S_B, (R_E - S_C), p_B R_E\} \\ \\ Y_C(t) = \text{mid}\{S_C, (R_E - S_B), p_C R_E\} \end{cases} \tag{6.11}$$

In equations 6.10 and 6.11 the time index t is dropped from the notations S_B, S_C and R_E for ease of expression.

In case of a signalized junction, flows from cell **B** and cell **C** will not flow at the same time to cell **E**. In case of a signalised intersection, the equations 6.10 and 6.11 will become simpler. The merge ratios p_B and p_C will be either 1 or 0. If cell **B** is considered as the active approach (approach in green phase), then the equations 6.10 and 6.11 will simplify to:

$$p_B(t) = 1; \qquad Y_B(t) = min\{S_B, R_E\} \tag{6.12a}$$

and for the approach in red phase, $p_C(t) = 1; \qquad Y_C(t) = 0$ (6.12b)

Equations 6.10 to 6.12 are the generalization of equation 6.3 and 6.5 to uniquely define flows in a merge manoeuvre. To update the cell occupancies thereafter equation 6.4 has to be used.

Diverges

To accurately model a network, Daganzo (1995b) also introduced the idea of modelling diverging manoeuvres (figure 6.4). In case of the one-way street system modelled here,

diverging applies to the through and turning movement of vehicles from any one-way street, sharing the same green time. Same as Lo (2001), the exogenous turning proportions are considered as time invariant. As shown in figure 6.4 vehicles from cell **B** can flow to two different destination cells **E** and **C** in two different directions.

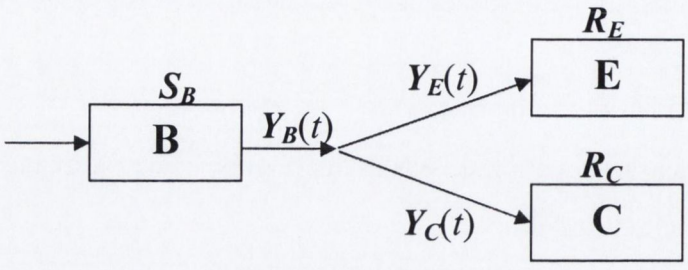


Figure 6.4 A Diverge Manoeuvre at a Time Instant.

In figure 6.4, $Y_C(t)$, $Y_E(t)$ and $Y_B(t)$ are the inflows to cells **C** and **E** and the outflow from cell **B** respectively at time instant t; R_E and R_C are the maximum possible inflows to the two receiving cells **E** and **C** respectively while S_B is the maximum possible outflow from the cell **B.** In the figure 6.4, the time index t is dropped from the notations S_B, R_C and R_E for ease of expression. β_C and β_E are the proportions of S_B going to cells **C** and **E** respectively.

Considering that the turning proportions are exogenously known, the inflow to each cell is,

$$Y_C(t) = \beta_C Y_B(t) \quad \text{and,} \quad Y_E(t) = \beta_E Y_B(t), \tag{6.13}$$

As both cells **C** and **E** are considered as destination of the outflow from cell **B,** if any of them is unable to accommodate the allocated inflow then the entire outflow is restricted. This is to maintain FIFO (first in first out) principle. This assumes that vehicles unable to exit freezes further movement in the cell. These conditions can be maintained by the following equations,

$$Y_B(t) = \min\left\{S_B(t), {R_E}/{\beta_E}, {R_C}/{\beta_C}\right\} \tag{6.14}$$

As in cases of ordinary links and merges, the state of the system can be updated further by using equation 6.4. No protected turning is considered in this formulation.

6.3 FORECASTING METHODOLOGY

Figure 6.5 gives a schematic diagram of a hypothetical traffic management system based on the proposed multivariate forecasting methodology.

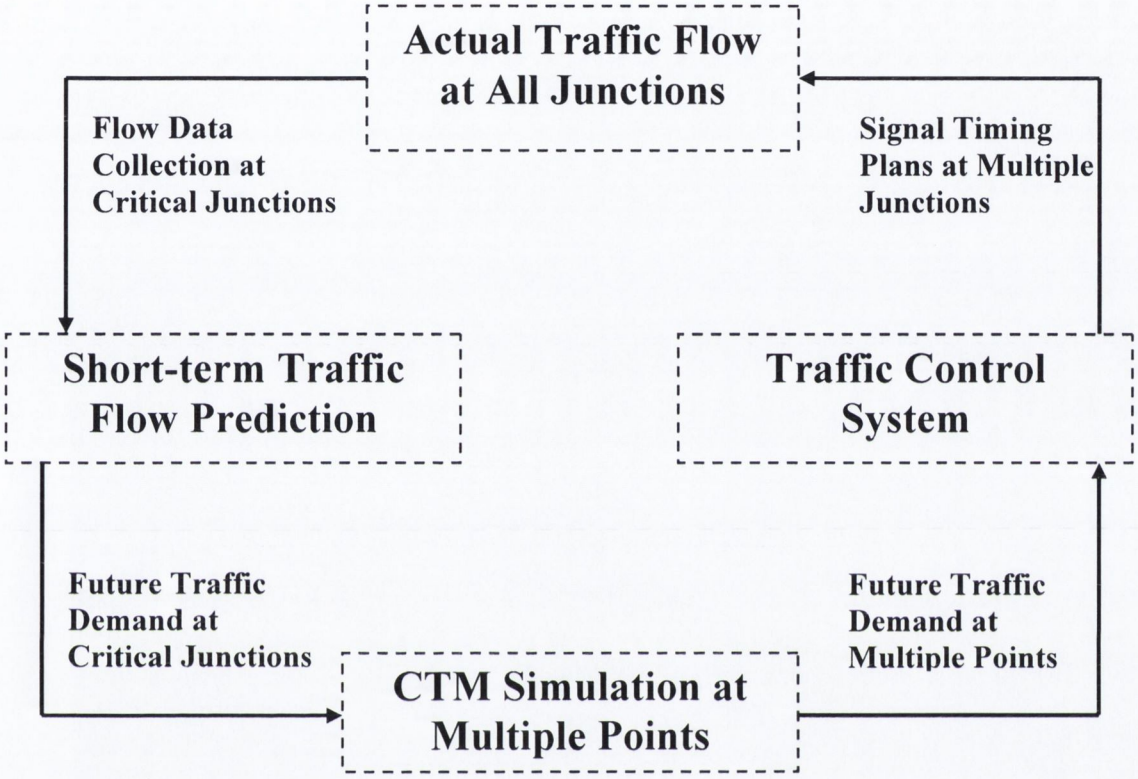


Figure 6.5 Multivariate Forecasting Scheme.

To start from the top, in the proposed system, actual traffic flow information will be obtained from inductive loop-detectors (data collection system) at certain critical junctions within a transportation network. At those critical junctions, future traffic demand is predicted using short-term traffic forecasting techniques. In this chapter, the SARIMA time-series model is used as a short-term traffic forecasting algorithm to predict future traffic demand based on the current and the historical traffic flow observations. This future traffic demand data are now used as input to a CTM formulation of the transportation network. Hence, future traffic flows and densities can be simulated at all points within the network. These simulations can further be used in a traffic control system to generate signal timing at all junctions within the transport network.

6.4 CASE STUDY

The proposed forecasting methodology is applied to a transportation network at Dublin city-centre to test the effectiveness of the forecasting strategy.

6.4.1 Site Description

The transportation network used here for modelling, using combined CTM and time series forecasting approach, is a part of the busy city centre of Dublin (figure 6.6A and 6.6 B).

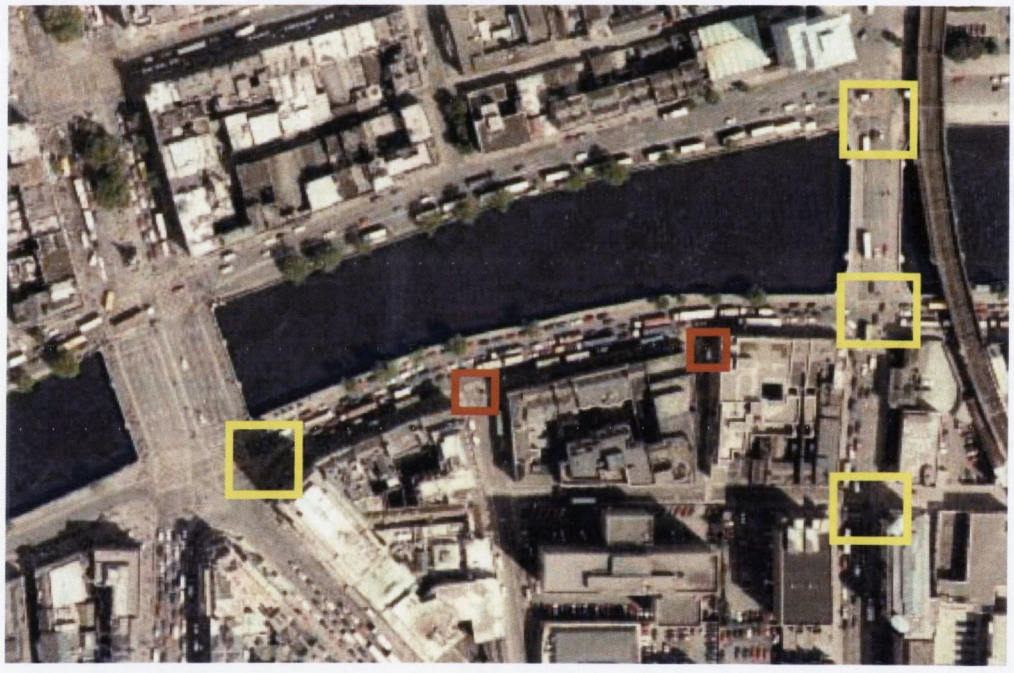


Figure 6.6A Satellite Image of the Six Junctions of the Transportation Network.

The four signalised junctions of the network are marked with yellow squares and the two un-signalised junctions are marked with orange squares in figure 6.6A. The number of lanes, the position of bus-lanes and the direction of traffic flow at all six junctions are shown in the schematic diagram of the transport network. Considerable queue formation and congestion can be encountered in this site during the peak hours (figure 6.6A). The main thoroughfares are **Tara Street** and the **Quays**, which carry one-way traffic all through. Two crossings of **Tara Street** with **Poolbeg Street** and the **Quays** are considered here. Two un-signalised side streets, **Hawkins' Street** and **Corn Exchange Place** joining the quays from the south side are also considered within the network. All of the crossroads carry one-way traffic as well as

carry one-way traffic as well as the main streets. Ireland has a left-hand drive system with no protected turning movements in this site.

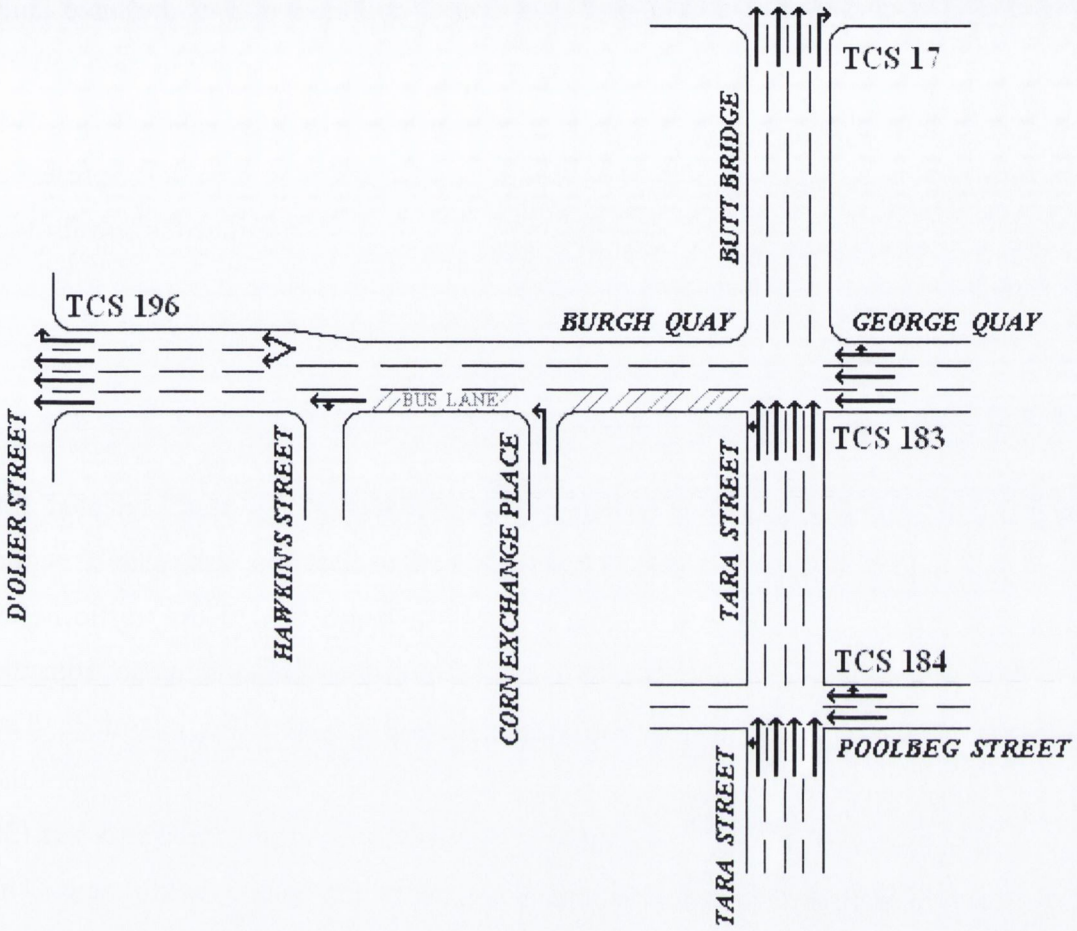


Figure 6.6B Schematic Diagram of the Transportation Network.

6.4.2 Field Data and Collection Procedures

For developing the cell representation of the chosen transportation network certain site specific details, such as free flow speed, saturation flow etc. are required. The following measurements were taken for calibration and application of CTM to the transportation network. The measurements were taken following the directions of the Highway Capacity Manual (1997).

1. Free Flow Speed

The free flow speed was measured manually from the site. Two reference points along the link were selected as the starting point and the ending point. The time interval of a

free-flow vehicle travelling from the starting point to the ending point was taken while the distance between the reference points was measured using a measuring tape. The free flow speed is the distance travelled by a free-flow vehicle in unit time interval. The free flow speed changes with the width of the lane. The lane width of the three lane arterials and four lane arterials being different, two sets (60 observations in each set) of observations were taken. The free flow speed of each lane in a three lane arterial was 27.3 km/hr and for each lane in a four lane arterial was 35.6 km/hr. Due to the presence of a turning manoeuvre just before junction TCS 184, the free flow speed is less for the three lane arterial.

2. Shockwave Speed

Similar to free flow speed measurement, two reference points were selected for measuring shock wave speed. The stop line before the traffic light was selected as the starting point and another reference point a certain distance behind the traffic lights was selected as the end point. The time taken to form a continuous queue from the starting point to the end point after the traffic lights turned red was recorded. The distance is measured using a measuring tape. The shockwave speed is the length of the continuous queue formed in unit time interval. Similar to free flow speed two sets (30 observations in each set) of observations were taken for each type of lane. The shockwave speed of each lane in a three lane arterial was 14.6 km/hr and for each lane in a four lane arterial was 13.34 km/hr.

3. Saturation Flow

At a signalised intersection, in a given lane, if every vehicle consumes an average of h seconds of green time, and if the signal continues to be uninterruptedly green, then S vph could enter the intersection where S is the saturation flow rate (vehicles per hour of green time per lane) given by

$$S = \frac{3600}{h} \tag{6.15}$$

Here h is the saturation 'headway' or the average discharge 'headway' at the junction measured in seconds. Successive discharge headways after the green indication is received are conceptually plotted in figure 6.7.

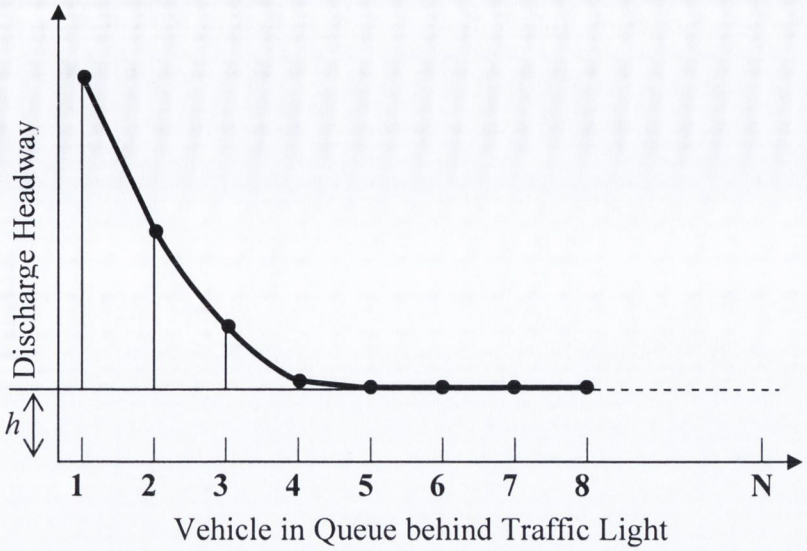


Figure 6.7 Discharge Headways of Departing Signals.

The first 'headway' is relatively long as it includes drivers' reaction and acceleration time. Each successive 'headway's get slightly smaller with overlapping of the reaction time. The level headway or the saturation headway starts from the fourth or the fifth vehicle. Occasionally when a queue of eight vehicles formed behind the traffic lights, average discharge headway is recorded over five vehicles after the third vehicle for a single observation of saturation headway. Two sets of such 40 observations are taken for two types of lanes in the network. The saturation flow for each lane in three lane arterial was 1722 veh/hr/lane and for each lane in four lane arterial was 1744 veh/hr/lane.

4. Turning Movement Percentages

In several points within the chosen transportation network, 'turning lane diverges' and 'unsignalised merge' manoeuvres take place. For a CTM framework representation of the network, the turning and merging percentages (β_C , β_E , p_B and p_C) are required to be exogenously determined. These proportions are obtained for this study by manually collecting data over four average weekdays (14th, 15th, 22nd and 23rd June 2005). Traffic counts from the SCATS database were also used for this purpose. The merging and diverging percentages calculated from these data are given in table 6.1. The merge percentages are not required to be exogenously calculated for signalized merges (equation 6.12) and are not explicitly given in the table.

Cell No.	Percentage from cell no.(1)	Percentage from cell no. (2)	Percentage to cell no.(1)	Percentage to cell no. (2)
3, Quays to Butt Bridge and through			12% to cell **6**	88% to cell **5**
10, merging traffic	80.1% from cell **9**	19.9% from cell **8**		
17, merging from Tara Street and Poolbeg Street	66.6% from cell **16**	33.4% from cell **15**		
20, Tara street to Poolbeg street and through			13.8% to cell **22**	86.2% to cell **23**
27, merging traffic	21.4% from cell **25**	78.6% from cell **26**		
29,Tara Street to 2 left turning lanes and through			53.2% to cell **30**	46.8% to cell **31**
30, 2 Left turning Tara Street lanes to Quays and through			47.8% to cell **32**	52.2% to cell **33**
37, merging from Tara Street and quays	37.1% from cell **35**	62.9% from cell **36**		
37, left turning and through traffic in Butt bridge			86.6% to cell **38**	13.4% to cell **93**
49, C.R. Place and Quays merging traffic	5.2% from cell **90**	94.8% from cell **48**		
57, left turning and through traffic in quays			18.9% to cell **58**	81.1% to cell **59**
58, through and left turn to Hawkins' Street			3.0 % to cell **91**	97% to cell **60**
61, right turn and through traffic (2 lanes)			81.9 % to cell **71**	18.1% to cell **80**

Table 6.1 Turning and Merge Percentages

5. Traffic Signal Timing Plan (SCATS Database)

There are four signalized intersections in the network (figure 6.6A). The traffic signals in these intersections are controlled by the existing urban or area traffic control (UTC) system called SCATS.

An UTC system consists of a number of interlinked and interdependent traffic signals. There can be two main types of UTC (Luk, 1984),

- Fixed-time Control
- Traffic responsive Control

In fixed time control three or four fixed signal timing plans for each junction are employed depending on the time of the day. The fixed signal timing plans are prepared based on the historical data. Vehicle detectors are not required in this case. In traffic responsive method the traffic control parameters are optimized according to the prevailing traffic conditions.

SCATS is a traffic responsive urban or area traffic control system, where the area under control is divided into a set of subsystems of one to ten intersections which share a common cycle time. Four phase split plans which express green plus inter-green time as percentages of the cycle time are available for each intersection with the subsystem. Each subsystem has five internal offset plans between adjacent intersections within the subsystem. The elements of the subsystem, i.e. the common cycle time, the phase split plan number and the offset plan number are based on the degree of saturation (DS). The degree of saturation is measured from the inductive loop-detectors embedded in the streets. The elements of a subsystem can vary slightly from the fixed plan, depending on the DS from the last cycle. Two adjacent subsystems can be 'married' or interlinked by external offsets of a subsystem. Married subsystems share a common cycle time.

In the current transport network, there are three subsystems. TCS 183 and TCS 184 belong to subsystem SS-6, TCS 196 to subsystem SS-27 and TCS 17 to subsystem SS-

22. SS-22 is married to SS-6 by an external offset. SS-27 is not linked. Cycle time within this network varies from 40 seconds to 60 seconds.

The signal timing plans at these four junctions from over 4:00 p.m. to 5:00 p.m. of 30th June 2005 are recorded from the SCATS database. The phase splits vary by 2 to 3 seconds for each cycle. Only the fixed phase split plans (signal timings) are shown in table 6.2. The cycle lengths for junctions TCS 17 and TCS 196 were not available and were assumed to be 120 seconds based on the observations over the hour. The yellow and all red times (inter-green times) are 1 and 2 seconds respectively, for all four signalised junctions.

Junction	Street Name	Phase	Offset	Green Time	Red Time
TCS 183	*Tara Street*	A	0	77sec	43sec
	Burgh Quay & George Quay	B	77sec	43sec	77sec
TCS 184	*Tara Street*	A	0	100sec	20sec
	Poolbeg Street	B	100sec	20sec	100sec
TCS 17	*Butt Bridge*(through and left turn)	AC	112 sec	72sec	48sec
	Butt Bridge(right turn)	A	130 sec	50sec	70sec
TCS 196	*Burgh Quay*	E	54sec	46sec	74sec

Table 6.2: The Signal Timing Plans

6. Actual Traffic Count (SCATS database)

From figure 6.5, the actual traffic flow observations are one of the main data input to the proposed multivariate forecasting methodology. The traffic flow data collection is a part of SCATS traffic control system.

Inductive loop-detectors are embedded near the stop line of each of the signalized intersection in the network (figure 6.6A). They record the traffic volume during each green phase. The counts during the green phase are aggregated over 5 minutes or 15

minutes to store in the SCATS database. In this study, 15 minute traffic flow observations obtained from the database from 16th May 2005 to 30th June 2005 are used in the time-series modelling.

For the unsignalised junctions, no time-series modelling is done. For traffic demand input during 4:00 p.m. to 5:00 p.m. of 30th June 2005 to the CTM framework at Hawkins' street and Corn Exchange Place crossing (figure 6.6B), actual traffic counts aggregated over each 15 minute are recorded manually.

6.4.3 Cell Representation of the Site

The chosen transport network (figure 6.6A and 6.6B) is presented in a CTM framework in figure 6.8. Depending on the free flow speed and space discretization criteria, each arterial or link is divided into equal length cells. The possible movements, in and out of a cell, are represented by directional arrows.

All the cells in figure 6.8 are drawn in the same dimension (except for the origin cells). But the lane numbers of all of the cells are not the same and may vary from 1 lane to 4 lanes. The numbers of lanes in each of the cells are given in table 6.3.

Number of Lanes	Cell Numbers
1 LANE	1, 3, 5, 6, 8, 11,12, 13, 14, 15, 16, 18, 20, 22, 23, 25, 32, 33, 58, 60, 62, 63, 64, 65, 66, 67, 68, 69, 70, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92
2 LANES	2, 4, 7, 9, 17, 30, 31, 34, 71, 72, 73, 74, 75, 76, 77, 78, 79,
3 LANES	10, 19, 21, 24, 26, 36, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 59, 61,
4 LANES	27, 28, 29, 37

Table 6.3: Number of Lanes in Each Cell

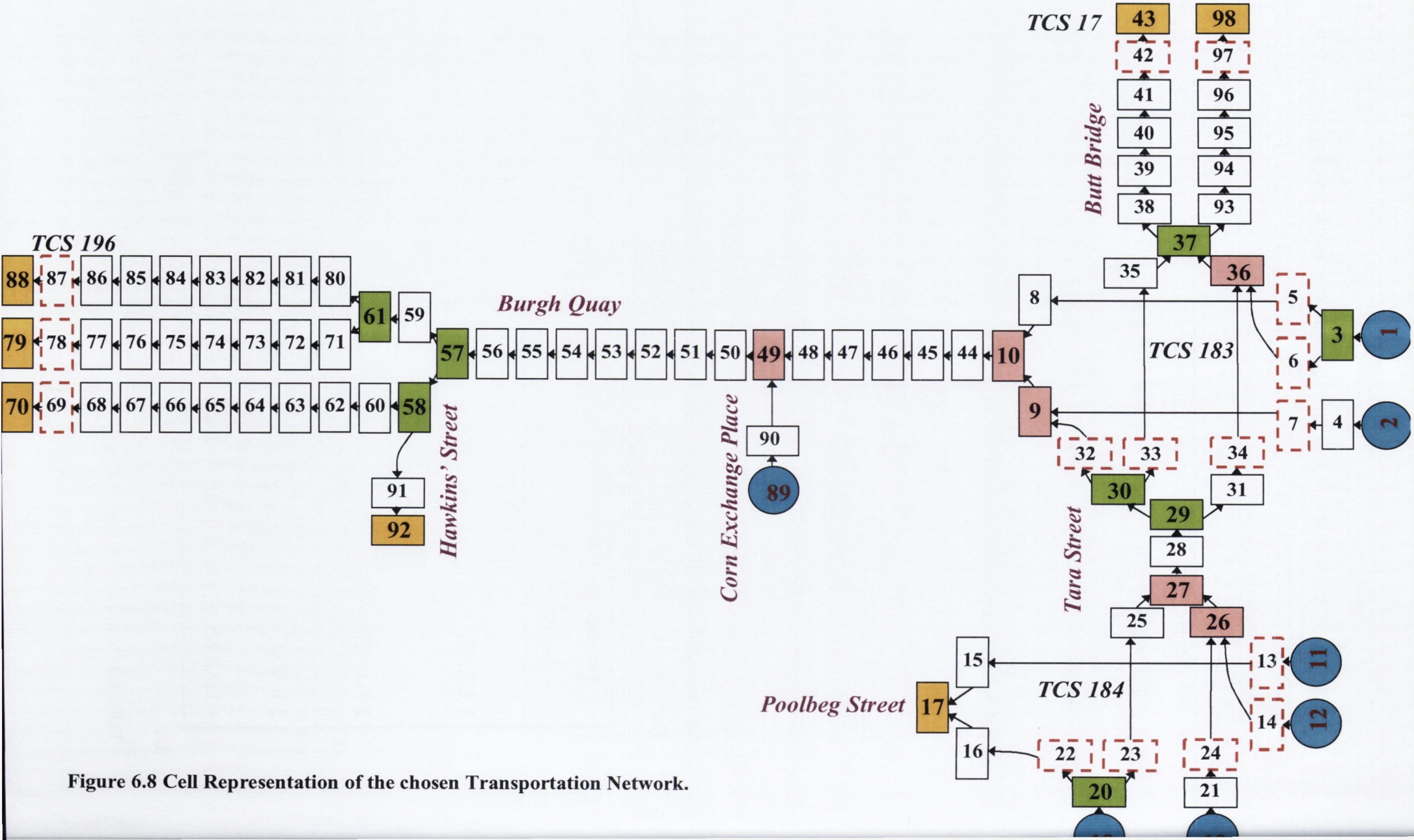


Figure 6.8 Cell Representation of the chosen Transportation Network.

The origin cells are shown in blue circles. There are seven origin cells in four different approaches. The number of the detectors recording traffic counts at each origin cell is given in table 6.2. Six origins are mentioned in the table. In the seventh origin, cell **89**, no inductive loop-detector is embedded at street level for measuring the traffic volume. The Corn Exchange place being a relatively quiet street, the traffic volume observations over each 15 minute are collected manually as described earlier.

The destination cells are shown in golden rectangles. There are 7 destination cells in the network in four different approaches. The merging cells are coloured in tan and the diverging cells are coloured in lime green. The cells in which outflow is controlled by traffic lights are shown with dotted red outline. The merges in which the two merging flows come from signalized cells are signalized merges and flows in those cells are governed by equations 6.12 and 6.4. In other tan coloured cells, the traffic flows are determined by equations 6.10, 6.11 and 6.4.

6.4.4 Time-Series Modelling

The input traffic demands used in the CTM model at six origin points (excluding the origin point at cell **89**) are the predicted demands calculated from the univariate short-term traffic forecasting model of the existing and past traffic demand data at those sites.

The SARIMA time series model is used as a forecasting technique to model the traffic flow observations recorded from 16th May 2005 early morning to 30th June 2005 early morning. The weekends and the bank holidays are not considered in the model for reasons described in chapter 1. In each day, 96 observations are obtained. The data show definite seasonality in pattern over a period of 24 hours at all six points. This leads to the idea of fitting a seasonal time-series model. As the site and situation is the same (or very similar) to that of the model described in chapter 1, similar SARIMA models are used in modelling the observations. The most suitable SARIMA models fitted to each of the origin points are given in table 6.4.

The traffic demand during 4:00 p.m. to 5:00 p.m. of 30th June 2005 at the six origins of the chosen transport network are predicted using the individual SARIMA time series models for those origin points.

Origin	Loop-detector	SARIMA Model
OR1	Loop 7, Junction TCS 183	$(2,0,1)(0,1,1)_{96}$
OR2	Loops 5 & 6, Junction TCS 183	$(2,0,1)(0,1,1)_{96}$
OR3	Loop 6, Junction TCS 184	$(2,0,1)(0,1,1)_{96}$
OR4	Loop 5, Junction TCS 184	$(1,0,1)(1,1,1)_{96}$
OR5	Loop 1, Junction TCS 184	$(2,0,1)(0,1,1)_{96}$
OR6	Loops 2,3 & 4, Junction TCS 184	$(2,0,1)(0,1,1)_{96}$

Table 6.4 The Traffic Demand Models

6.5 RESULTS

The traffic flows at the seven downstream destination points (figure 6.8) during 4:00 p.m. to 5:00 p.m. of 30th June can be obtained through CTM simulation using predicted traffic demand values (from the SARIMA models) at the seven input points. The simulated future traffic flows during 4:00 p.m. to 5:00 p.m. of 30th June 2005 at the seven downstream destination points are compared with real time traffic flow data collected by the embedded inductive loop-detectors at those cells (figure 6.9). Instead of showing the comparison of traffic counts for each destination cell of the same intersection separately, the comparison is shown for the intersection as a whole.

According to table 6.5, most of the predictions are within 10% of the actual observations, which are quite acceptable. The MAPE [Appendix-A] from the 1 hour prediction data set is only 4.4% in junction TCS 17 and 10.6% in junction TCS 196.

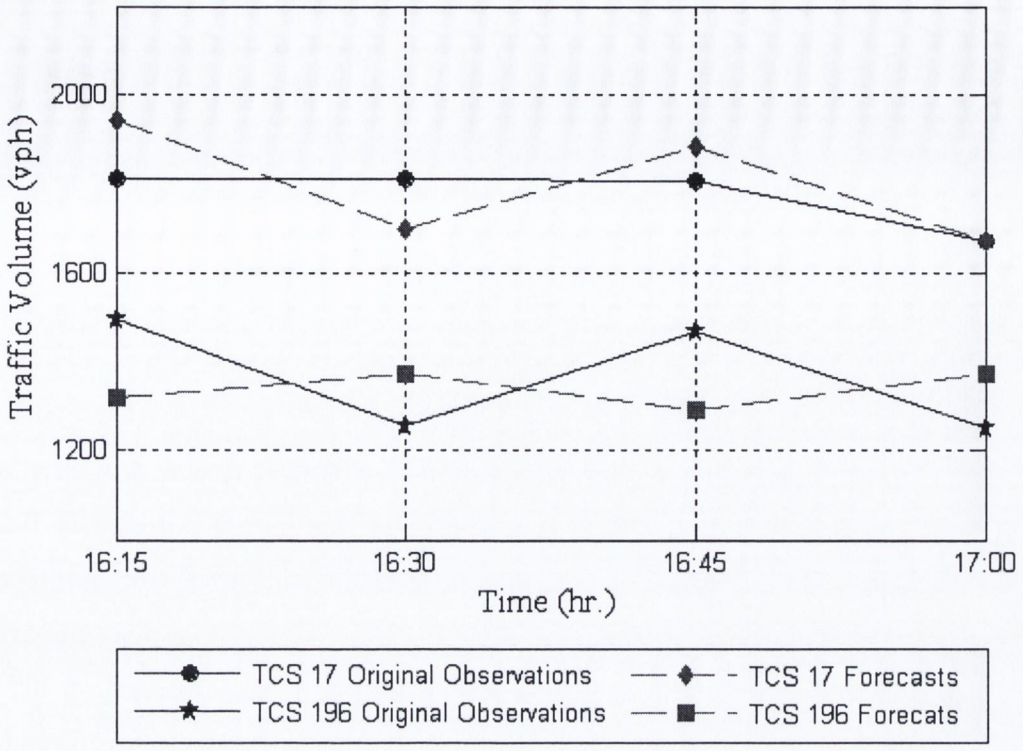


Figure 6.9 Original Observations and Model Forecasts From Two Destination Intersections.

Time Period (hr.)	TCS17 Real Counts	TCS17 Model Counts	APE (%)	TCS196 Real Counts	TCS196 Model Counts	APE (%)
16:00-16:15	453	485.14	7.10	374	329.33	11.94
16:15-16:30	452	424.05	6.18	314	342.83	9.18
16:30-16:45	451	470.30	4.28	367	322.98	12.0
16:45-17:00	418	416.95	0.25	313	342.83	9.53

Table 6.5 The Error Estimates

The possible reasons for the marginal error or discrepancy could be due to the following:

- There may be manual errors during data collection.
- The loop-detector data can be inaccurate owing to the fact that some of the inductive loops in any big network can be faulty.
- Lane-changing behaviour cannot be captured by CTM model, but happens in reality.
- As mentioned before, the average fixed cycle lengths were assumed in intersection TCS 17 and intersection TCS 196, but the cycle length are not exactly fixed in reality, and could not be calculated as the pedestrian phase times were not recorded in the SCATS database. Finally, instead of an exact signal time plan, the average cycle times of 120 seconds were used over 4:00 p.m. to 5:00 p.m. of 30th June 2005. This error was in the order of 1 to 2 seconds over each cycle and may attribute a considerable percentage to the final error values.

In addition to the benefit of forecasting jointly at multiple network points, another major contribution of this model is that the proposed methodology is effective to predict the real-time traffic flow level at the junctions where no continuous data collection takes place in a congested network with frequent queue spillback occurrence. The 15 minute aggregate traffic volumes are simulated at cell **10** and cell **57** of the network (figure 6.8) as example, as there are no inductive loop-detectors present at these two sites in the real-life transport network. During 4:00 p.m. to 5:00 p.m. of 30th June, the simulated traffic volumes of these two cells obtained from the proposed multivariate traffic flow model are given in table 6.6.

	4:00-4:15 p.m.	**4:15-4:30 p.m.**	**4:30-4:45 p.m.**	**4:45-5:00 p.m.**
Cell **10**	1296 vph	1388 vph	1288 vph	1388 vph
Cell **57**	1236 vph	1352 vph	1312 vph	1360 vph

Table 6.6 Simulation of 15 Minute Traffic Volumes

The results show that the forecasts at junctions TCS 17 and TCS 196 only deviate around 10% from the original observations. It is expected, that the accuracy of the results can

improve if the demand data can be obtained over a smaller prediction time interval. Apart from that the proposed methodology is effective in simulating traffic volumes at points on the transport network where no data is available. Hence, this model can be effectively used for simulation and for forecasting of multivariate traffic volume data in an urban congested transport network.

CHAPTER 7

APPLICATION OF WAVELET ANALYSIS FOR TRAFFIC FLOW FORECASTING AND INCIDENT DETECTION

7.1 INTRODUCTION

In the previous five chapters of this thesis, studies focussing on the time-series analysis based univariate and multivariate traffic flow modelling techniques for application to different ITS based scenarios have been discussed. In a set of traffic flow observations, there can be events occurring which may influence short-term changes at different time-scales. Wavelet analysis is a transformation technique to analyse time series data like traffic flow observations by individually focussing on traffic volume components of different time-scales. In this chapter, the technique of wavelet analysis is applied in developing two main aspects of ITS, i.e. traffic flow modelling and incident detection. In traffic flow modelling, a non-functional trend model and a Bayesian hierarchical model are proposed for simulating univariate traffic volume data at an intersection where there is no regular data collection taking place. The wavelet based trend modelling methodology is further applied in developing an automatic incident detection technique in the remaining part of the chapter based on the variability of the statistical variance of the traffic flow parameters.

7.2 MULTI-RESOLUTION DISCRETE WAVELET ANALYSIS

Wavelet transform provides a time-frequency representation of any signal/time-series data. The basis of wavelet analysis is decomposing a signal into shifted and scaled versions of the original (or *mother*) wavelet. Multi-resolution wavelet analysis uses techniques by which different frequencies are analyzed with different resolutions by using an efficient numerical algorithm.

Discrete wavelet transform (DWT) (Mallat, 1989) consists of the collection of coefficients

$$\left.\begin{cases} c_J(k) = < X, \varphi_{Jk}(t) > \\ d_j(k) = < X, \psi_{jk}(t) > \end{cases}\right\} j,k \in Z \tag{7.1}$$

where < *, * > denotes inner product, $\{d_j(k)\}$ are the detail coefficients at level j (j = 1,2,...J) and $\{c_J(k)\}$ are the approximate coefficients at level J. The signal X to be analyzed is integrally transformed with a set of basis functions

$$\psi_{jk}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right) \tag{7.2}$$

which is constructed from the mother-wavelet $\psi(t)$ by a time-shift operation and a dilation operation. The function $\varphi_{Jk}(t)$ is a time shifted version of the mother-wavelet scaling function $\varphi_J(t)$: $\varphi_{Jk}(t) = \varphi_J(t-k)$. $\varphi_J(t)$ is a low-pass function which can separate the low frequency component of the signal. Thus DWT decomposes a signal into a large timescale (low frequency) approximation and a collection of details at different smaller timescales (higher frequencies).

The original signal can be reconstructed back from the decomposed approximation and the detail components. Thus, the original signal can be represented as,

$$X(t) = A_J(t) + \sum_{j \le J} D_j(t) \tag{7.3}$$

where, $A_J(t)$ is the reconstruction of the approximation coefficients c_J at level J and $D_j(t)$ is the reconstruction obtained from the detail coefficients d_j at level j. In the reconstructed approximation (A_J) and in the reconstructed details obtained at each stage/level (D_1, D_2, D_3 ... D_J) the numbers of data points remain the same as the original dataset.

7.3 TREND MODELLING

The 'trend' of a time-series data can loosely be defined as the 'long-term' change of the mean level of the data (Chatfield, 2001). In the daily trend modelling of a traffic flow time-series data, the word 'long-term' indicates stability over time on a daily basis. To develop a

background model for the daily traffic flow observations at an intersection, the representative trend underlying the traffic flow over a day is required to be modelled. In this study, a multi-resolution analysis of the traffic flow observations is performed to find the representative daily trend (Ghosh et al., 2006).

DWT associated with the basis *Daubechies' 4* (db4) is used to decompose the signal (time-series traffic flow observations) into different time scales. The scaling function and the basis function of the wavelet basis *Daubechies' 4* are shown in figure 7.1.

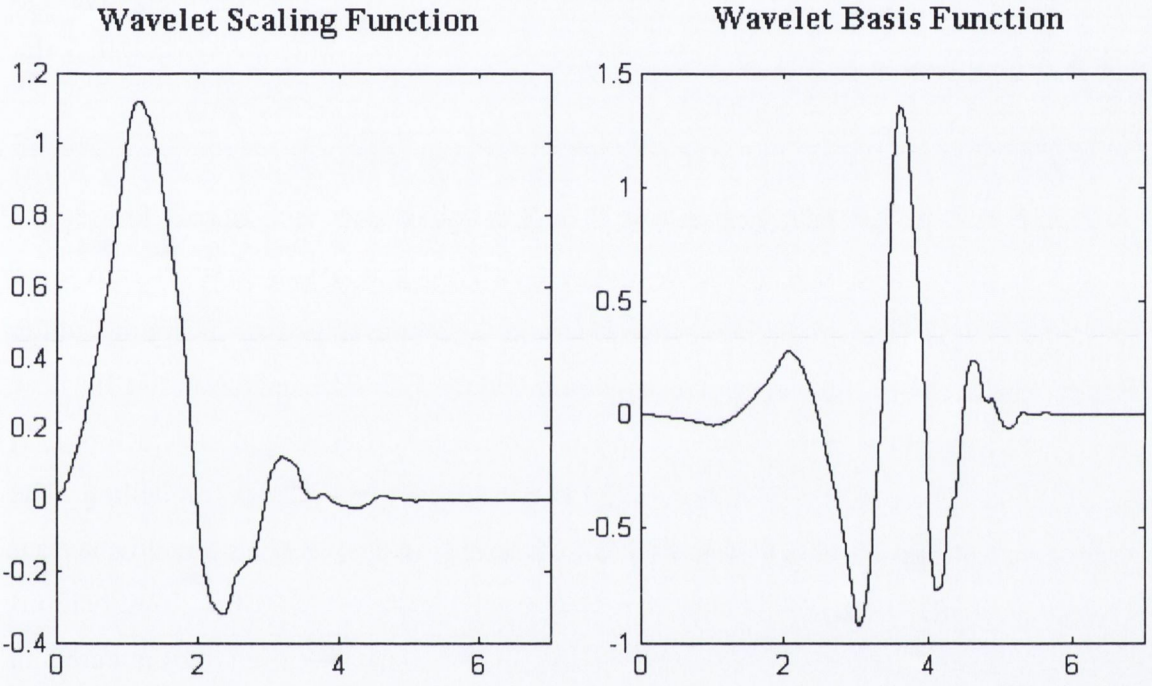


Figure 7.1 The Scaling Function and Basis Function of DWT *Daubechies' 4*.

Initially, the original signal is decomposed into approximation coefficients c_1 (low frequency/fluctuations or variability) and detail coefficients d_1 (high frequency/fluctuations or variability). The approximation coefficients c_1 (relative low frequency components) are again decomposed to approximation coefficients c_2 (low fluctuations) and detail coefficients d_2 (high fluctuations) at the next level. This procedure is repeated for further decomposition. The aim of repeating the decomposition procedure is to find an optimum approximation level for extracting the trend in the data. The optimum approximation level is the one in which the reconstructed approximation coefficients, A_m (m is the optimum approximation level), are the optimal smoothed estimate of the traffic flow data which can truly represent the traffic flow pattern on an average day. This is essentially a de-noising

technique in signal processing. The local variability in traffic flow observations due to signal control in the urban arterials is considered as the noise (for the mathematical treatment) in this methodology.

The traffic flow pattern at any particular approach at any intersection is similar for the weekdays (section 2.2). But there can be some day-to-day variability due to other factors, like day of the week, accidents or recurrent congestion in some other part of the transport network etc. These factors are uncontrollable and can not be modelled as such. So, to obtain a 'regular trend over an average day', the A_m values over some regular days (approximately, 20 days in this study) are to be averaged for a single day. The average trend has a non-analytical functional form and has better flexibility in representing the mean traffic flow over an average day.

Case Study

The described methodology of finding 'regular trend over an average day' is applied to the intersection TCS 183 at the city-center of Dublin (figure 2.2). The univariate traffic flow observations obtained collectively over each 5 minute interval from the loop detectors 1, 2, 3 and 4 of the approach one at intersection TCS 183 are used for the modelling. The representative trend is modelled for an average weekday. As the weekend travel behaviour is very much unlike the travel behaviour in the weekdays, the traffic flow observations only during the weekdays are included the model. Since, the data set used does not contain any missing data, no special treatment for missing data is required to be utilised here.

15th June 2005 is chosen arbitrarily as the regular weekday for the modelling. A plot of the traffic flow observations from the chosen site on that day is given in figure 7.2. The traffic flow data series shows a non-stationary (Section 2.2) nature. Wavelet analysis can handle non-stationarity unlike other time-series modeling techniques where this issue is required to be separately dealt with and transformation of the time-series data is required to be performed (Ghosh et al., 2005).

The traffic flow observations over twenty days i.e. four weeks (as weekends are not included), in the month of June-July from the chosen site are then decomposed into three levels of resolution using MRA with *Daubechies' 4* wavelet basis function. The three different levels represent the three different time scales. The wavelet coefficients for

approximation and detail are then reconstructed at all three levels. For any modelling purpose the reconstructed values of the approximation and detail coefficients are always used in this study.

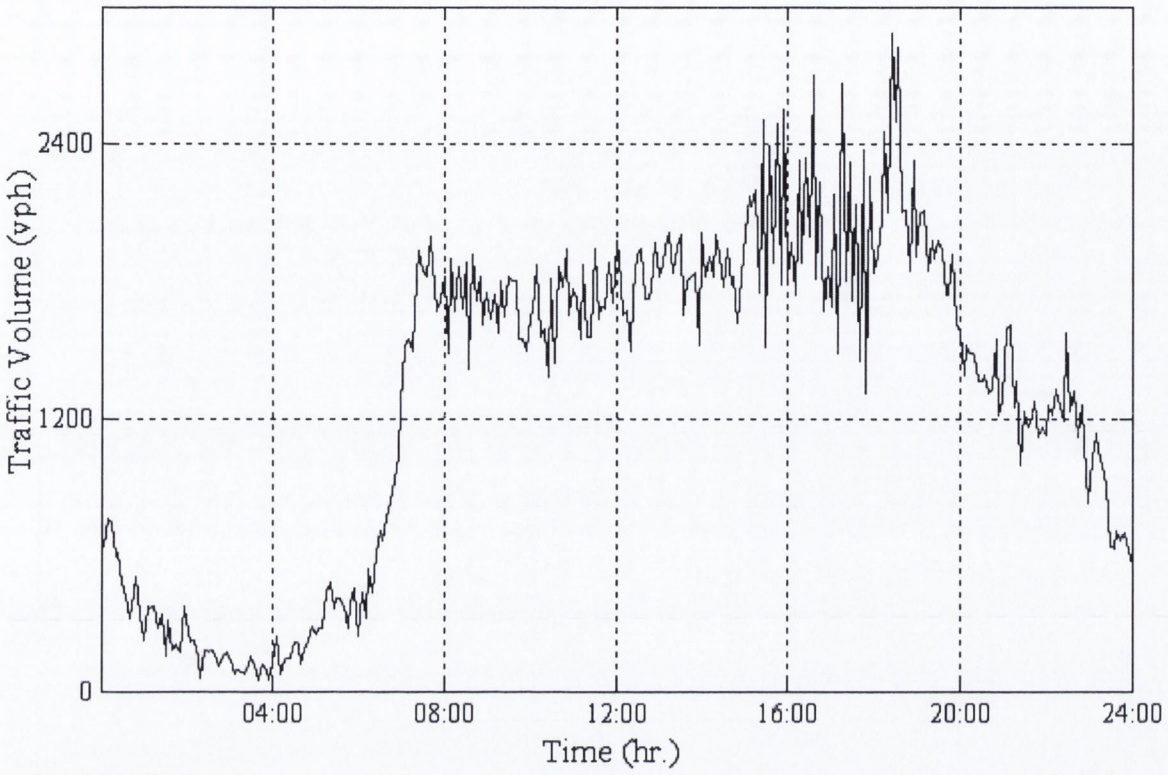


Figure 7.2 Traffic Flow Observations on 15-06-2005 at Junction TCS 183.

At each level, during decomposition the high frequency part of the data is separated from the low resolution or the low frequency part. The low frequency part at level three is quite smooth and can be used as a representative of the overall trend over a day in the traffic data. Hence, the third level is considered as the optimum level of decomposition to obtain optimum smoothed estimates of the times-series data. The approximations at level three, and the details at level one, two and three on 15th June 2005 are plotted in figure 7.3.

To model a representative trend, level three approximations of the traffic flow time-series observations over 20 days are taken. An average over 20 days of the level 3 reconstructed approximation coefficients for a single day is used to model the representative daily trend for approach one of intersection TCS 183. Taking the average coefficients helps to reduce the effect of certain uncontrollable conditions as described before.

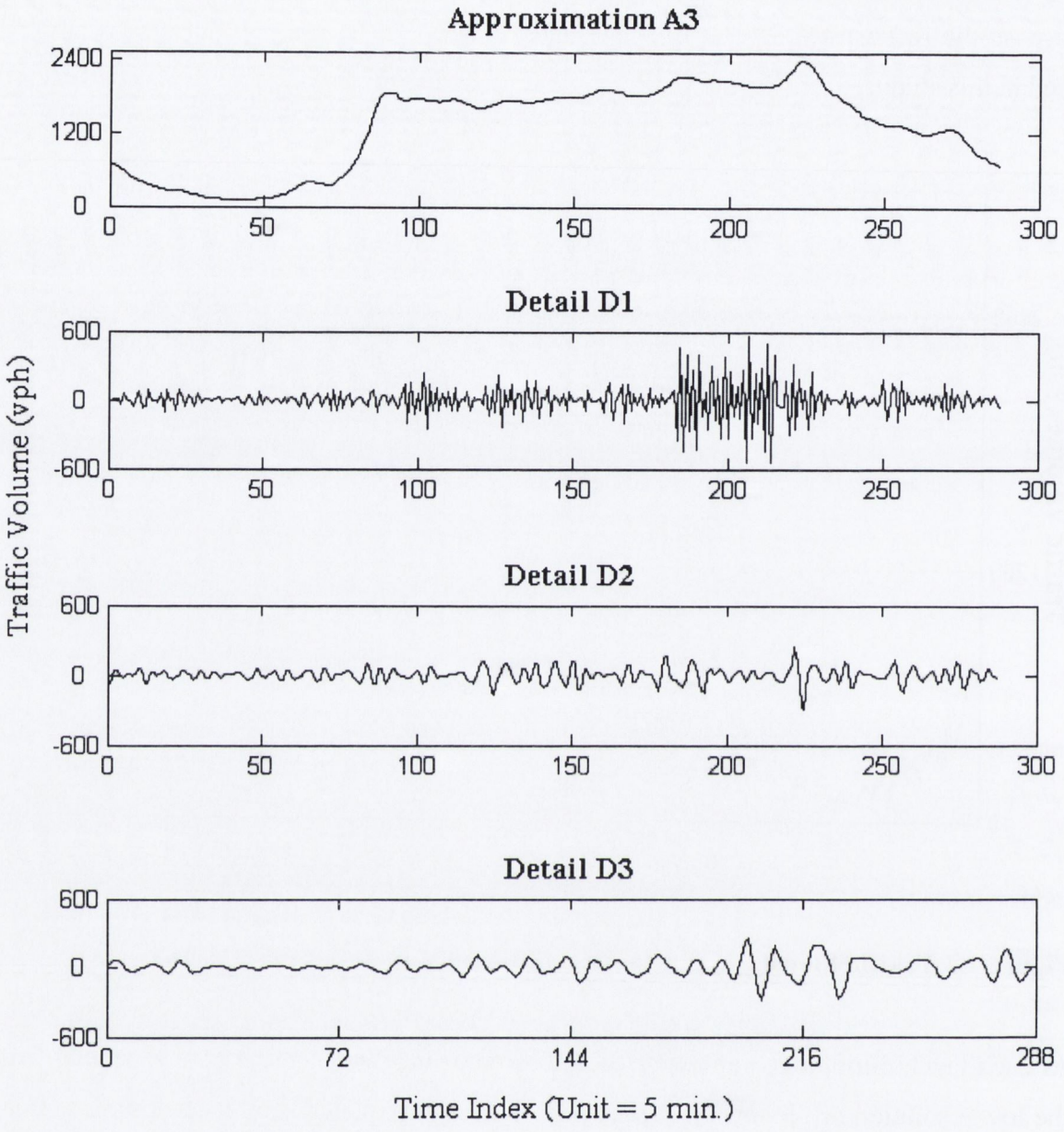


Figure 7.3 Reconstructed Approximation Coefficients at Level 3 and Reconstructed Details Coefficients at Level 1, 2, 3 on 15th June 2005.

In figure 7.4 the 'regular trend over an average day' is plotted over the traffic flow observations over 15th June 2005. From the graph it can be seen that the simple trend gives a very good approximation of the traffic volume on any arbitrary day.

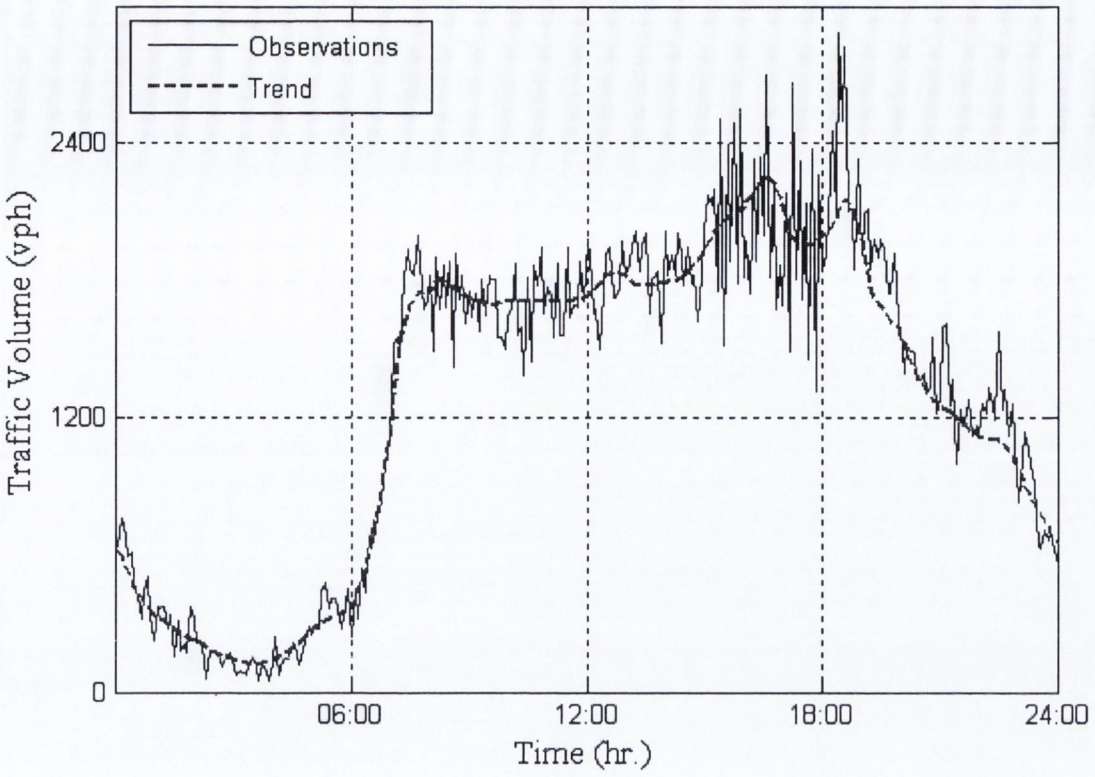


Figure 7.4 Trend and Original Observations on an Arbitrary Day (15-06-2005).

7.4 RESIDUAL MODELLING

The residuals of the model are obtained by subtracting the average trend from the traffic flow observations on 15th June 2005. A dot plot of the residuals on that day is given in figure 7.5. The residuals obtained have a standard deviation of 188.28 vph. The mean of the original observations over on 15th of June is around 1301.75 vph. The MAPE value obtained after fitting the trend model to the 5 minute traffic flow observations on 15th June 2005 is nearly 15%. Hence, it can be concluded that the background trend modelling using a MRA technique gives reasonably accurate information about the traffic flow observations at any intersection in a transport network without depending on the current information from the site.

Though obtaining a 'regular trend over an average day' is the main feature of the background model for traffic flow at an intersection, for better accuracy the residual obtained after fitting the trend are also required to be modelled. Generally these high frequency components are random in nature. But with an increase in decomposition level, self-similar nature can be observed in them.

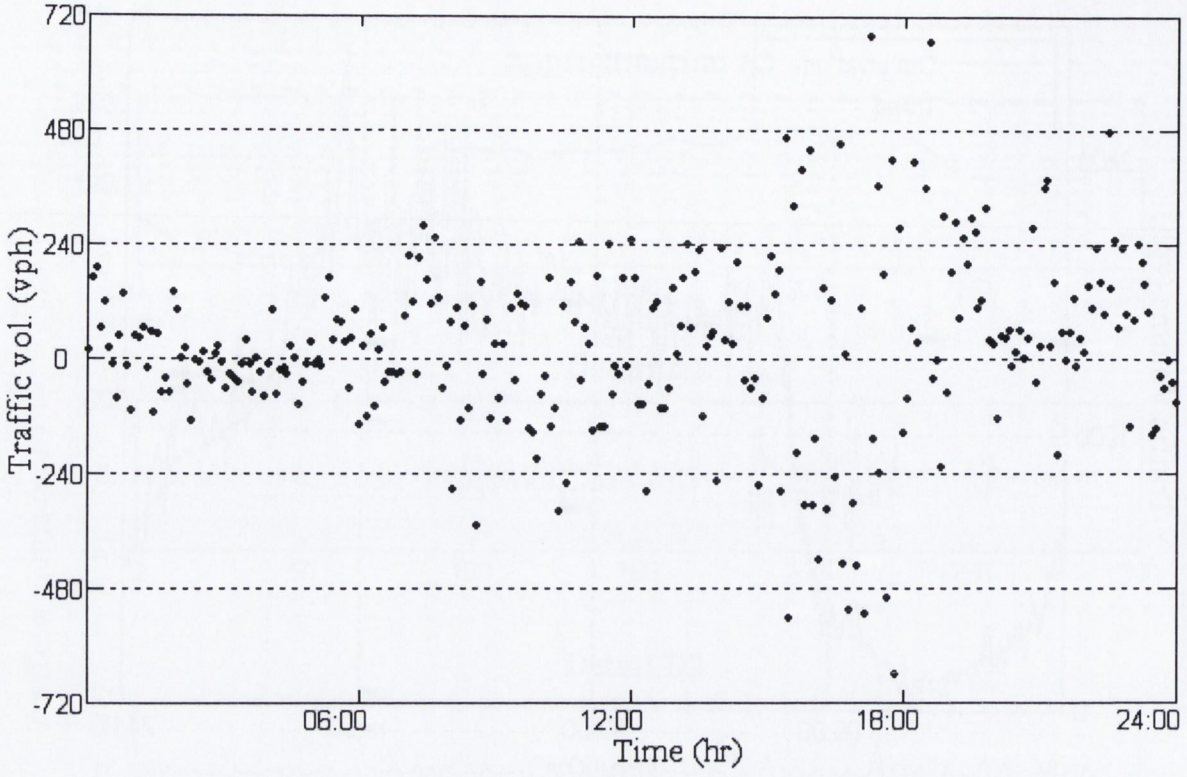


Figure 7.5 Dot Plot of Residual on 15-06-2005.

For developing a crude model of the residuals, a histogram of the residual data on 15th June 2005 is plotted in figure 7.6. From the histogram of the residuals it is evident that the residuals can be approximately modelled as a normal distribution. The normal probability density fit of the residuals is plotted on the histogram in figure 7.6. The normal density fit matches the histogram only crudely. A confidence interval on the 'regular trend over an average day' can be provided based on this residual model. The following equation is used in simulating the confidence limit of this background model based on 'regular trend over an average day'.

$$y_{sim} = A_{trnd} + \mu_{res} + \varepsilon_{res} \qquad \text{where, } \varepsilon_{res} \sim \mathrm{N}(0,\sigma) \tag{7.4}$$

where, y_{sim} is the simulated traffic volume from the background model; A_{trnd} is the average value of level three approximation obtained from the trend model; μ_{res} is the mean of the residual dataset and ε_{res} is the random part of the residuals. ε_{res} values are randomly generated from a normal distribution with mean zero and standard deviation σ which is the sample standard deviation of the residuals. A 95% confidence interval of the residuals is used to form the confidence limit of the background model.

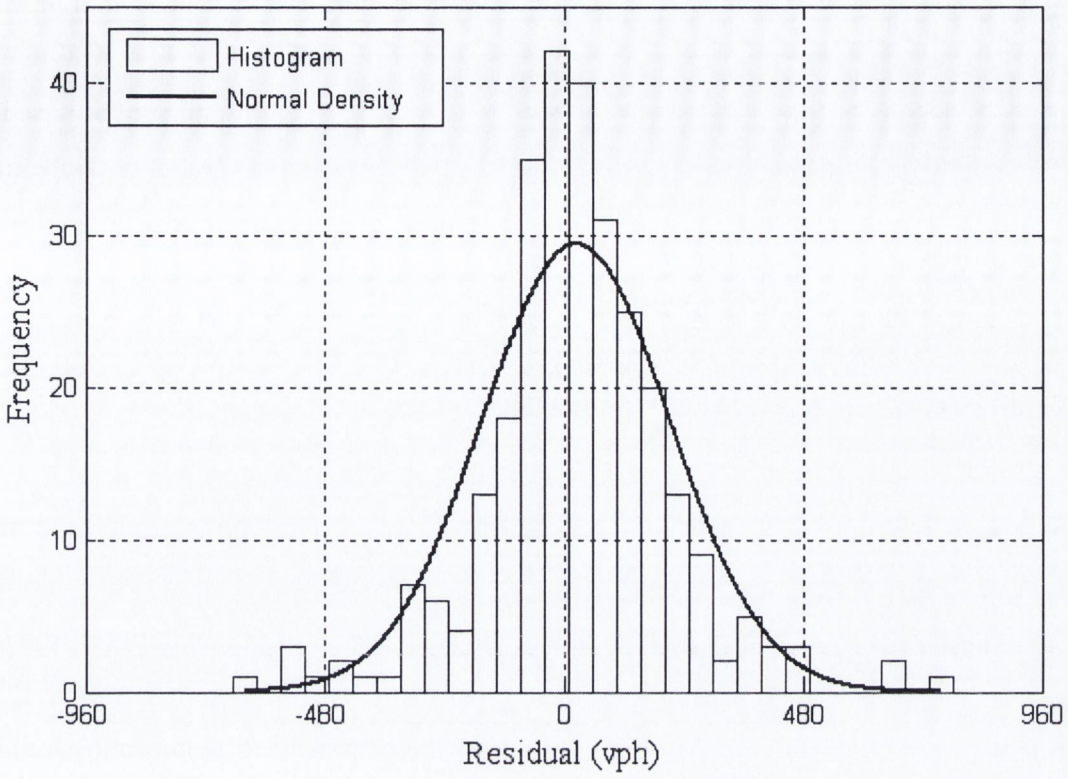


Figure 7.6 Histogram and Normal Probability Density Plot of the Residual on 15-06-2005.

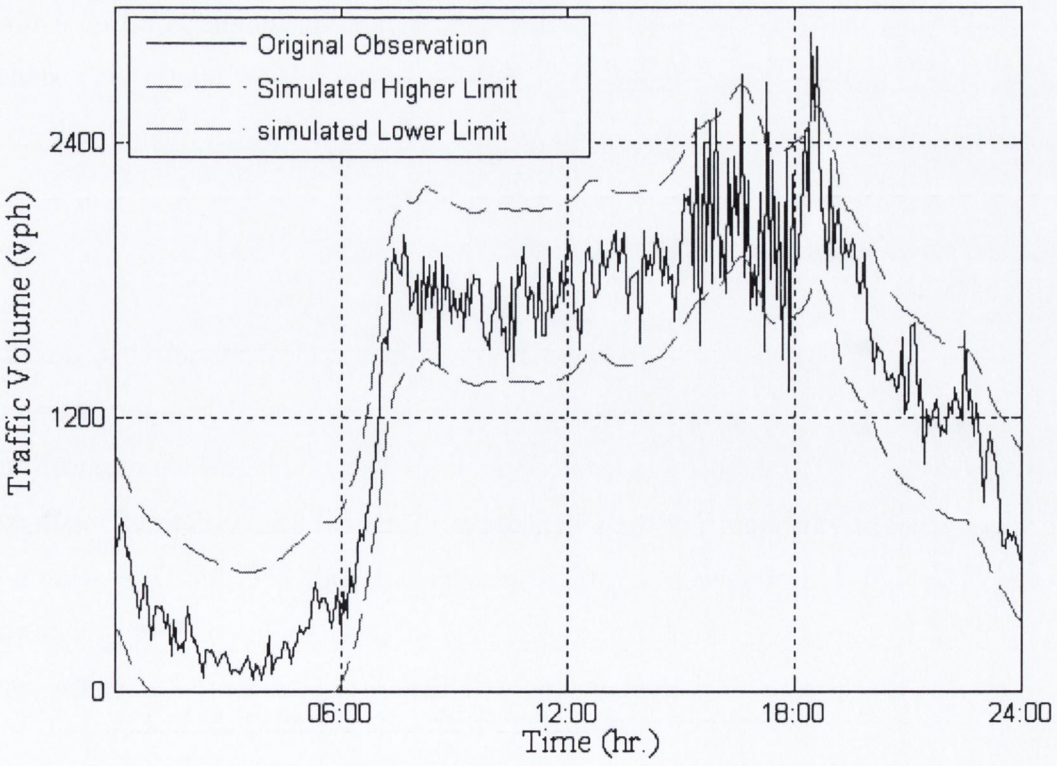


Figure 7.7 Simulated and Original Traffic Volumes on 15-06-2005.

In figure 7.7 the original traffic flow observations on 15th June 2005 are plotted along with simulated confidence intervals from the background model. Most of the observations fall within the confidence limits. This proves that the simplistic background model based on 'regular trend over an average day' can very well be used as an approximate traffic volume simulation model for the intersections where continuous information on traffic conditions is not available.

7.4.1 Time-Variant Variances of the Residual

It is observed from figure 7.5 that the spread of the residual data points around the mean value is not uniform. The variability of the residual data in off-peak hours of early morning and late night is much less than the variability of the same during peak hours of the day. The variability is the highest during the evening peak hours. This non-uniform variability signifies that the variance of the residual should not be estimated as a constant parameter for an entire day, but as a variable varying with the time of the day. In a heuristic approach, the residual can be modelled with variance values changing after each half-hour of the day. The half-hourly variance can be calculated for each half-hour of the day from the residual data (obtained from 5 minute traffic flow observations on 15-06-2005) during that half-hour. But for 5 minute aggregate traffic volume data, the sample size for calculating variance over each half-hour is very small and can not be considered as a true estimate. Hence, instead of this heuristic method Bayesian hierarchical model is used as a standard statistical method to model the residues with its time varying variance.

Bayesian Hierarchical Model of the Residues

In simple words, the basic idea behind the Bayesian hierarchical model is to develop a parametric statistical model with parameters which are represented by other parametric statistical models (i.e. it is doubly stochastic in nature). The essence of hierarchical model is that the dependencies among variables in a statistical model can be defined much easily with a tree-like structure. In the case of a Bayesian (concept introduced in section 3.2) hierarchical model, while calculating *posterior* density from equation 3.3d, *priors* which themselves depend on other parameters not included in the *likelihood* function can be accounted for.

In this study, the variance of the residual is dependent on time and has to be modelled accordingly using another parametric statistical model. Hence, the residual obtained from the 5 minute aggregate traffic flow observations on 15-06-2005, after fitting the trend model are further modelled using a Bayesian hierarchical model to account for the time-varying nature of the variance of the residual dataset. If the $\boldsymbol{R}$ is the vector of the residual, then in a normal hierarchical model,

$$\boldsymbol{R}_t \sim \mathrm{N}\left(m, \sigma_t^2\right) \qquad t = 1,2....T \tag{7.5}$$

where, m is the sample mean of the residual on 15-06-2005 and σ_t is the standard deviation of the residual for each time instant denoted by a subscript t. As 5 minute aggregate traffic volume is modelled, the vectors $\boldsymbol{\sigma}$ and $\boldsymbol{R}$ are both of dimension $\{T\mathrm{x}1\}$ where T is the number of time intervals or time instants in a day (for 5 minute aggregate traffic flow observations, $T = 288$) .

The variance σ^2, of the residual dataset $\boldsymbol{R}$ changes with the time of the day. To model this time varying variability of the variance, the following parametric distribution is proposed.

$$\log(\boldsymbol{\sigma}) \sim \mathrm{N}\left(\log(\boldsymbol{y}), \tau^2\right) \tag{7.6a}$$

$$\text{which leads to } \boldsymbol{\sigma} \sim \mathrm{LN}\left(\log(\boldsymbol{y}), \tau^2\right) \tag{7.6b}$$

As $\boldsymbol{\sigma}$ is always positive, a lognormal distribution is taken in equation 7.6 to ensure that all σ_t lie within $(0,\infty)$. The lognormal distribution for each σ_t is centred at y_t with standard deviation of τ. The variances of the high resolution components (sum of level 1, 2 and 3 reconstructed detail coefficients) from the 20 day traffic flow observations calculated over each hour of a day are considered as the initial estimates of the standard deviation of the residual (y_t) for that hour of the day. The elements of the vector $\boldsymbol{y}$ are same for all time instants within the same hour of the day. The values of the vector $\boldsymbol{y}$ of dimension $\{T\mathrm{x}1\}$ are calculated from the 20 day data set and these values are constant for traffic flow observations on an average day at the particular intersection TCS 183 considered in this study. Hence, the Bayesian hierarchical model developed in this study can be considered as applicable to any arbitrary day of the year.

The variance τ^2 of the lognormal distribution of vector σ is assumed to follow a uniform distribution, within a range $(0,k)$

$$\tau \sim \mathrm{U}(0,k) \qquad k<\infty \tag{7.7}$$

where, k is an arbitrary constant signifying the maximum limit of the values of τ. The exact value of k does not influence the estimation process. In this study the equations 7.5, 7.6b and 7.7 define the Bayesian hierarchical model for the residuals obtained after subtracting the regular average trend from the traffic flow observations on 15th June 2005.

In the Bayesian hierarchical model the unknown parameters to be estimated are $\sigma(\sigma_1,\sigma_2.....\sigma_{288})$ and τ. These unknown parameters are represented by a vector $\xi=(\tau,\sigma_1,\sigma_2...\sigma_{288})^T$. To estimate the vector ξ, the Bayesian estimation technique described in chapter 3 is to be used.

For the Bayesian inference, the posterior density of the normal hierarchical model is

$$p(\xi|\boldsymbol{R},\boldsymbol{t})=p(\tau)L(\sigma|\boldsymbol{R},\boldsymbol{t})L(\tau|\sigma,\boldsymbol{t}) \tag{7.8}$$

where, $p(\xi|\boldsymbol{R},\boldsymbol{t})$ is the posterior density of ξ; $L(\sigma|\boldsymbol{R},\boldsymbol{t})$ is the likelihood function of σ and $L(\tau|\sigma,\boldsymbol{t})$ is the likelihood function of τ; $p(\tau)$ is the prior density of parameter τ.

According to equation 7.5, $\boldsymbol{R}$ is assumed to follow a normal distribution. Hence, the likelihood function of σ given the vector $\boldsymbol{R}$ and the time instant vector $\boldsymbol{t}$ (unit time interval = 5 minute) is

$$L(\sigma|\boldsymbol{R},\boldsymbol{t})=\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi\sigma_t^{\,2}}}\exp\left(-\frac{R_t^2}{2\sigma_t^{\,2}}\right) \tag{7.9}$$

Similarly according to equation 7.6b where σ is assumed to follow a likelihood function, the likelihood function of τ given σ, $\boldsymbol{y}$ and $\boldsymbol{t}$,

$$L(\tau|\boldsymbol{\sigma},\boldsymbol{t}) = \prod_{t=1}^{T} \frac{1}{\sigma_t \tau\sqrt{2\pi}} \exp\left[-\frac{(\log\sigma_t - \log y_t)^2}{2\tau^2}\right] \tag{7.10}$$

$p(\tau)$ is equal to a constant as the prior density of τ is assumed as flat on the range $(0,\infty)$. Hence, the posterior density from equation 7.8 is

$$p(\xi|\boldsymbol{R},\boldsymbol{t}) \propto \prod_{t=1}^{T} \frac{1}{\sigma_t} \exp\left(-\frac{R_t^2}{2\sigma_t^2}\right) \prod_{t=1}^{T} \frac{1}{\sigma_t \tau} \exp\left[-\frac{(\log\sigma_t - \log y_t)^2}{2\tau^2}\right] \tag{7.11a}$$

which yields,

$$p(\xi|\boldsymbol{R},\boldsymbol{t}) \propto \left(\frac{1}{\tau^T}\right) \prod_{t=1}^{T} \frac{1}{\sigma_t^2} \exp\left(-\frac{R_t^2}{2\sigma_t^2} - \frac{(\log\sigma_t - \log y_t)^2}{2\tau^2}\right) \tag{7.11}$$

$$\int_0^\infty \ldots \int_0^\infty \int_0^\infty \left(\frac{1}{\tau^T}\right) \prod_{t=1}^{T} \frac{1}{\sigma_t^2} \exp\left(-\frac{R_t^2}{2\sigma_t^2} - \frac{(\log\sigma_t - \log y_t)^2}{2\tau^2}\right) d\tau d\sigma_1 \ldots d\sigma_T \tag{7.12}$$

Similar to the case discussed in chapter 3, by integrating out the other unknown parameters except for the one whose distribution is to be estimated, the 'marginal distributions' of the each of the unknown parameters can be found out from the integral in equation 7.12. The computation of the marginal distributions of the unknown parameters in ξ involves evaluation of a complex integral with problems of high dimensionality.

The MCMC method (described in section 3.2) is used here to simulate the distributions of the unknown parameters. To simulate the marginal probability distributions for the unknown parameters in the vector $\xi(\tau,\sigma_1,\sigma_2.....\sigma_T)$, given an initial condition $(\tau^{(0)},\sigma_1^{(0)},\sigma_2^{(0)}.....\sigma_T^{(0)})$ the following 289 steps are to be iterated (i denotes the number of iteration):

1. Sample τ^{i+1} from $p(\tau^{i+1}|\sigma_1^{(i)},\sigma_2^{(i)}.....\sigma_T^{(i)},\boldsymbol{y}_t,\boldsymbol{t})$ using Gibbs sampler technique

2. Sample σ_1^{i+1} using Metropolis Hastings technique

.
.

289. Sample σ_T^{i+1} using Metropolis Hastings technique

The initial conditions $\left(\tau^{(0)}, \sigma_1^{(0)}, \sigma_2^{(0)} \ldots\ldots \sigma_T^{(0)}\right)$ are as follows,

$$\begin{cases} \boldsymbol{\sigma}^{(0)} = |\boldsymbol{y} - 0.5| \\ \tau^{(0)} \sim \text{Inverse-gamma}\left[\dfrac{T-1}{2}, \dfrac{\sum_{t=1}^{T} (\log \sigma_t^{(0)} - \log y_t)^2}{2}\right]^{-1} \end{cases} \tag{7.13}$$

In step 1, the Gibbs sampler (section 3.2) technique is used to simulate the distribution of τ. From the posterior density in equation 7.12, a full conditional distribution for τ is as follows,

$$p(\tau|\boldsymbol{\sigma}, \boldsymbol{R}, \boldsymbol{t}) \propto \left(\frac{1}{\tau^T}\right) \exp\left(-\frac{\sum_{t=1}^{T} (\log \sigma_t - \log y_t)^2}{2\tau^2}\right) \tag{7.14}$$

The full conditional distribution of τ can be observed as an inverse gamma distribution with parameters,

$$\alpha = T/2 - 0.5 \tag{7.15a}$$

$$\beta = 0.5 \sum_{t=1}^{T} (\log \sigma_t - \log y_t)^2 \tag{7.15b}$$

where the density function of the inverse gamma distribution is as follows,

$$p(\tau) = \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{-(\alpha+1)} \exp(-\beta/\tau) \tag{7.15c}$$

For simplicity the superscript denoting the number of iteration is not used in the equations 7.14 and 7.15.

Steps 2 to 289 are similar in nature and are used to simulate the values of $\sigma_1, \sigma_2 \ldots\ldots \sigma_T$ in each iteration. The simulation of $\boldsymbol{\sigma}$ is done using Metropolis-Hastings technique. The candidate values of each of the elements of the vector $\boldsymbol{\sigma}$ are simulated from following proposal distribution,

$$\boldsymbol{\sigma}^i \sim \mathrm{LN}\left[\log(\boldsymbol{y}), \left(\tau^i\right)^2\right] \tag{7.16}$$

The simulated value of $\boldsymbol{\sigma}^i$ at each iteration is accepted following the Metropolis algorithm. According to this algorithm, each simulated value of the elements of the vector $\boldsymbol{\sigma}^i$ in each iteration is accepted with a probability

$$\hat{p}(\sigma_t) = \frac{p\left(\sigma_t^i \middle| \tau^i, \sigma_1^{i-1}, \sigma_2^{i-1}, \ldots . \sigma_{t-1}^{i-1}, \sigma_{t+1}^{i-1}, \ldots\ldots\ldots\ldots \sigma_T^{i-1}, \boldsymbol{R}, \boldsymbol{t}\right)}{p\left(\sigma_t^{i-1} \middle| \tau^i, \sigma_1^{i-1}, \sigma_2^{i-1}, \ldots . \sigma_{t-1}^{i-1}, \sigma_{t+1}^{i-1}, \ldots\ldots\ldots\ldots \sigma_T^{i-1}, \boldsymbol{R}, \boldsymbol{t}\right)} \tag{7.17}$$

or 1 whichever is minimum, with the same acceptance criteria as previously mentioned in section 3.2.

The steps 1 to 289 are repeated for 10000 times to simulate 10000 values for all the unknown parameters. The simulated values of τ are plotted in a graph in figure 7.8. The simulation show high convergence towards a constant value of about 0.4.

In case of the vector of time varying standard deviation $\boldsymbol{\sigma}$, all the two eighty-eight elements of the vector are simulated separately. Instead of showing a plot of 10000 simulated values of each σ_t, the mean of the simulated values of each σ_t are shown in figure 7.9. The sample standard deviation obtained from the previous Gaussian noise model of the residual is shown as a horizontal line at 188.28 vph in the same figure. The estimates of $\boldsymbol{\sigma}$ obtained from the Bayesian hierarchical model change with the time of the day. The estimates during the peak hours are much more than the estimated values σ_t during the rest of the day and the estimates during the early hours in the morning are the lowest of all. The nature of the variability of the variance of the residual conforms to the spread of the residual data points in figure 7.5.

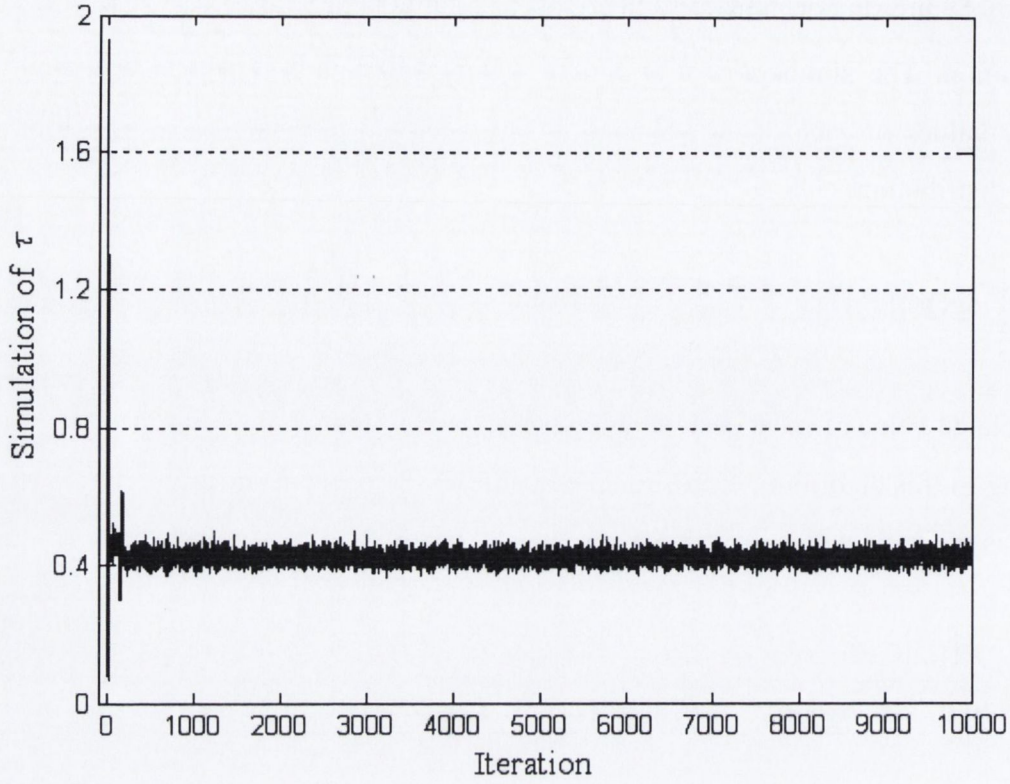


Figure 7.8 Simulations of Values of τ.

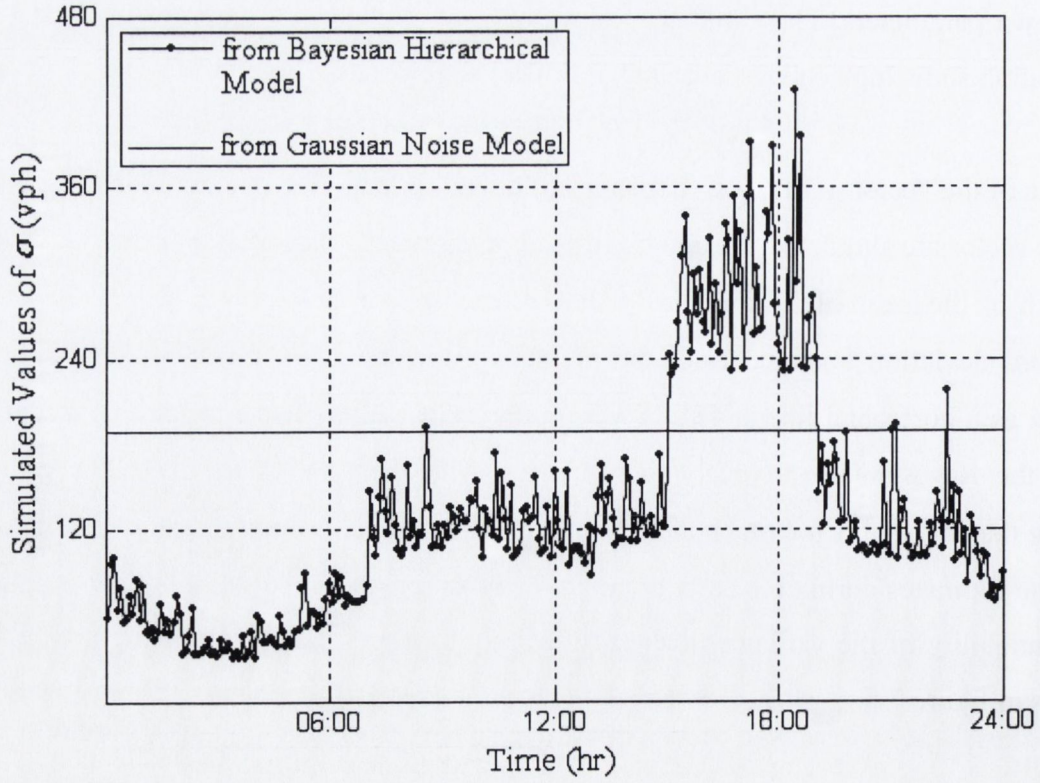


Figure 7.9 Simulations of Values of σ.

To illustrate the effectiveness of the Bayesian hierarchical model of the residual a 95% confidence interval similar to the Gaussian noise model is constructed on the regular average trend. The original 5 minute traffic flow observations from intersection TCS 183 on 15th June 2005 are plotted along with the simulated 95% confidence limit in figure 7.10. Unlike the previous model, all the traffic flow observations on 15-06-2005 fall within the simulated limits. Being a Bayesian method, the confidence limits adapt according to the variability of the residual data. This adaptation proves most effective during the evening peak hours where most of the observations fall outside the simulated limits from the Gaussian noise model.

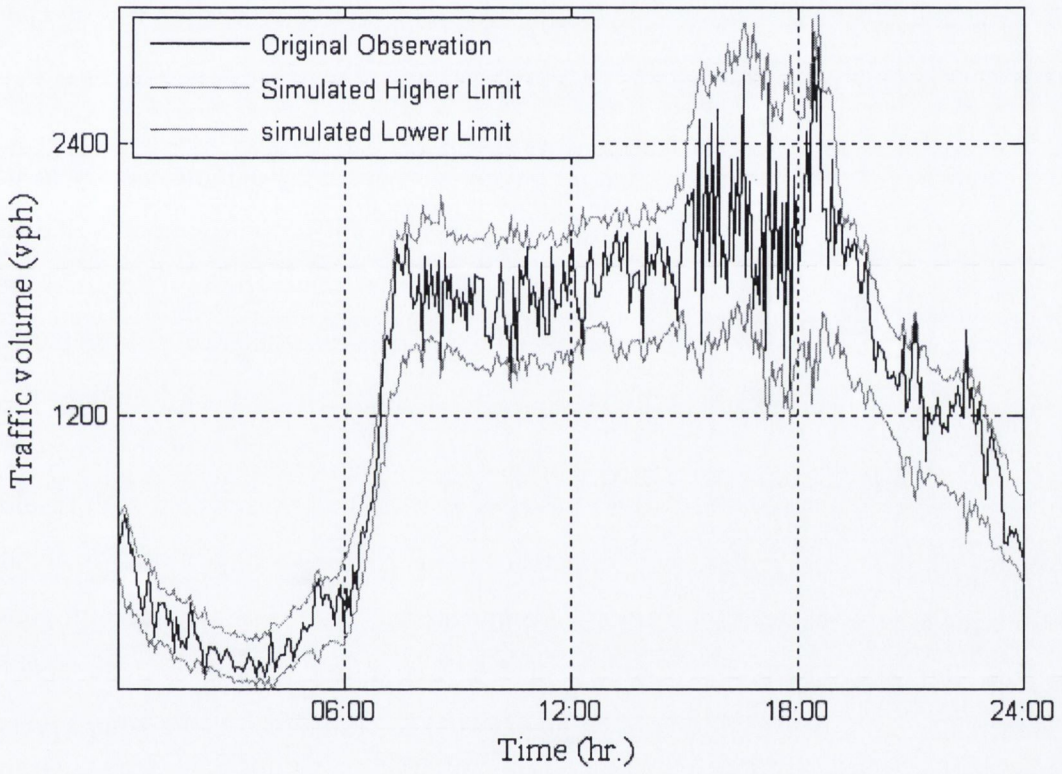


Figure 7.10 Simulated and Original Traffic Volumes on 15-06-2005.

The regular average trend plus residual hierarchical model can be considered as an effective traffic flow simulation model for a particular intersection in a congested urban network.

7.5 APPLICATION OF TIME-VARIANT VARIANCES IN INCIDENT DETECTION

Travel-time delays, reduction in arterial capacity and air-pollution are some of the main detrimental effects of non-recurrent congestion or incidents on busy urban networks. The implementation of incident management systems in an urban transport network ensures efficient management of traffic by minimizing the effect of the operational problems like non-recurrent congestion. Practical and reliable automatic incident detection algorithms (AIDA) are important to reduce and localize the effect of incidents. The concept of time-varying variances of the residuals and trend modelling using wavelet analysis developed in the previous sections is applied in developing a practical and reliable AIDA in this part of the study.

The automatic incident detection technique developed here is applicable mainly to urban arterials where the stop and go effect is quite discrete in traffic measurements due to the close spacing of traffic signals/traffic lights. MRA can decompose individual traffic parameters as signals in different resolutions, which allows to concentrate in the particular scale and frequency level in which the effect of an anomaly is the most pronounced. This feature of wavelet analysis is exploited in this algorithm in an MRA of the main traffic parameters. Unlike the existing wavelet based freeway AID algorithms the high resolution components show more pronounced effects of non-recurrent congestions or incidents (considering, average incident time around 5 minutes) than the low-resolution components. In the algorithm proposed here a discrete wavelet filter is used to remove the coarse level from the traffic data (traffic flow and occupancy). The residuals i.e. the finer levels are analysed based on the variability of their variance over time within a day. Statistical and deterministic comparison of variance with and without incident is applied for detection.

7.5.1 Wavelet-based Automatic Incident Detection Methodology

The study develops an AIDA which combines a wavelet-based MRA technique and a statistical methodology. The multi resolution involves working with discrete wavelet transform and the statistical part involves hypothesis testing. The combined methodology is described in this section.

Traffic incidents are exceptions to everyday traffic behaviour in any urban network. The extracted features from traffic measurement during incidents are significantly different from the same during regular/non-incident conditions. The regular traffic behaviour is modelled using the background model described in sections 7.3 and 7.4 (Trend and Residual modelling). The traffic flow and cell-occupancy data aggregated at 20 or 30 seconds interval and simulated over several days (say, *n* days) can be decomposed into 3 levels at 3 different time scales using Discrete Wavelet Transform (DWT) with *Daubechies' 4* wavelet basis function. The wavelet coefficients for approximation and detail can then be reconstructed. As described in section 7.3 the approximation at level 3 can be taken as the trend of the traffic data over these *n* days and sum of the detail reconstructions at levels 1, 2 and 3 is considered as the residual of the fit. The variability of the traffic flow observations from the trend is measured by the time varying variances of the residual traffic data (traffic flow and cell-occupancy). The time varying variances are calculated over each half of an hour from the mid-night over 24 hours.

Online Sliding Window

To detect the occurrences of incident within a short interval of time in a congested urban arterial, the current and recent past loop-detector observations from different stations aggregated over 20 sec or 30 sec are compared with the regular observations from those sites. A wavelet based MRA (similar to the one used for building the background model) is to be used on simulated traffic flow and cell-occupancy data to remove some low-resolution variability (mainly due to the fluctuation of mean with time) from the dataset. Discrete wavelet *Daubechies' 4* at level 3 is used on each data-point (traffic flow or cell-occupancy) with an online sliding window of 3 simulations (i.e. 1.5 minute with 30 sec aggregate simulations) before and after the point. Variance of the residuals (obtained using wavelet decomposition, as described in section 7.4) from the 7 points over 3.5 minutes is compared with the variance of that half-hour or half-hourly variance as obtained from the historical observations of *n* days. Occurrence of any incident is detected from violation of the condition of comparison of the two variances.

Comparison of Variance

The AIDA described in this study is based on online comparison of the current variance of different traffic parameters windowed over 3.5 minutes (i.e. 7 points) with the variance of those parameters obtained as per historical data. Two types of comparison criteria are used.

1. Confidence interval of the hourly variance based on t-distribution
2. Limit on variability of variance

Confidence Interval of the Hourly Variance Based on t-Distribution

Hourly variances of traffic flow and cell-occupancy for every half-hour within a day are calculated based on the past n day's traffic flow observations and cell-occupancy simulations (Ghosh et al., 2005a). Using the mean and standard deviation for each half-hourly variance, based on a t-distribution (as sample size < 30) a 99.9% confidence interval is constructed over each half-hourly variance of traffic flow and cell-occupancy. As the online sliding window used for calculating current online variance considers only 7 information points, the number of degrees of freedom is chosen as 6.

Limit on Variability of Variance

In this comparison a limit is set to the variability of current online variance of flow observations and cell-occupancy simulations. It is assumed that the value of the current online variance should not be less than the minimum or more than the maximum, of the hourly variances of 5 hours centred at the current hour.

Using both these criteria, if the current online variance of two consecutive points of flow observations and cell-occupancy simulations fall outside this interval, occurrence of an incident is detected.

7.5.2 Simulation of Incidents

The proposed AIDA is tested on the inductive loop-detector observations from the UTC system employed in the city of Dublin, Ireland. A macroscopic model based simulation strategy (Ghosh et al., 2005a) is applied on a real-life transportation network to simulate traffic flow dynamics including physical queue formation and delay due to the occurrence of incidents. The multivariate CTM based forecasting and simulation strategy described in chapter 6 is used as the simulation model for this study.

The CTM model involves three main traffic parameters; traffic flow, measured in vehicles per hour (vph), occupancy measured in percentage of capacity (%) and speed measured in km/hr and is dependent on the flow-density relationship. Using historical loop-detector observations as traffic demands, traffic flow and cell-occupancy (occupancy percentage calculated for a single cell) aggregated at 20 or 30 seconds interval at different points within a real transport network can be simulated using the CTM based simulation strategy.

The UTC used in Dublin is SCATS and it mainly collects the traffic flow and saturation data at different critical junctions within the transportation network. The transport network described in chapter 6 (figure 6.6A and 6.6B) is used here for simulation of traffic flow and cell-occupancy data. Considerable queue formation and congestion can be encountered in this site during the peak hours. The cell representation of the network is given and described in section 6.4.3 of chapter 6 (figure 6.8).

Using historical loop-detector observations over 24 days (13-06-2005 to 14-07-2005, excluding the weekends), traffic flow and cell-occupancy aggregated at 30 seconds interval are simulated at different points within the network.

7.5.3 Case Study

Wavelet Decomposition

The 30 seconds aggregate traffic flow and cell-occupancy data simulated over 24 days at different points within the network shown in figure 6.8, are decomposed using the discrete wavelet *Daubechies' 4* till level 3. The residuals or the sum of the finer level details is then calculated by de-trending the original traffic conditions data, as described in section 7.5.1. In figure 7.11, the traffic volume simulated over a day is plotted along with its trend (i.e. level 3 DWT approximations). The approximations approximately follow the mean of the data. The half-hourly variances of the residual traffic flow data show considerable variability over time (figure 7.12).

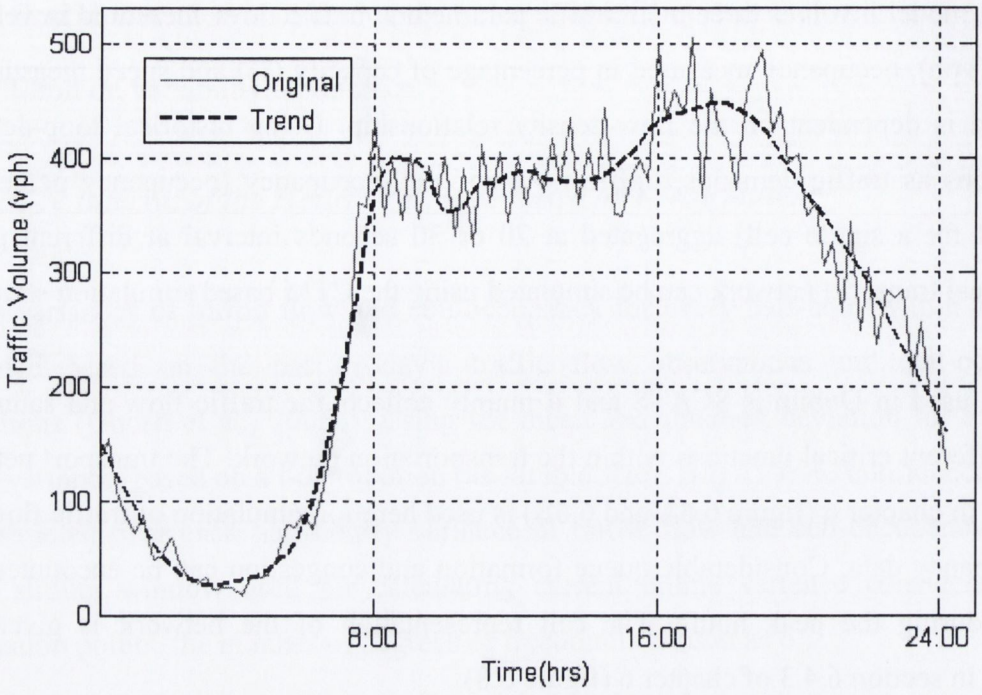


Figure 7.11 Trend and Original Observation over a Day.

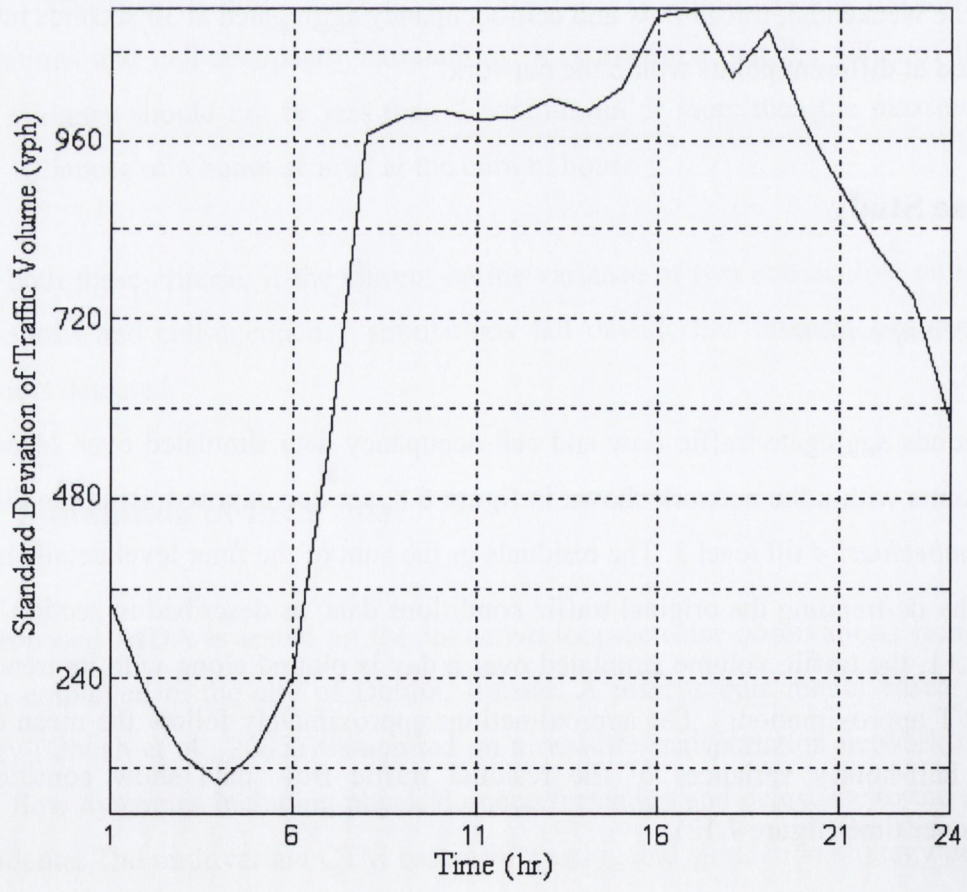


Figure 7.12 Variability Of Variance Plot.

As the flow-density relationship used in the simulation model is linear in nature, the variability of variance for cell-occupancy is similar to the variability of variance due to flow. Hence, it is not reasonable to model the residual as a white noise sequence with a constant variance.

Simulation of Incident

To simulate the incidents in the CTM framework described in chapter 6, it is assumed that the incidents occur in a particular cell within the network. The capacity of the cell in which the incident occurs becomes zero during the occurrence of the incident. Hence during that time, the outflow from the upstream cell is blocked or reduced depending on the severity of the incidence. This will result in increased upstream cell-occupancy and decreased downstream traffic inflow similar to a real life incident in any transport network. As CTM framework includes queuing in its formulation, the effect of upstream queuing is simulated very realistically.

Incidents are simulated at cell **18**, cell **40** and cell **54** in the network shown in figure 6.8. Traffic flow and cell-occupancy data at each 30 sec interval is collected at the upstream and the downstream of the cells at which incident is simulated. Out of the several incidents at the three cells, one simulated in cell **40**, at 3:40 p.m. on 15th June 2005 spanning over nearly five minutes is shown as a case study for analysis in this study. The simulated incident caused blockage of two lanes. The upstream and downstream 30 second aggregate traffic flow and cell-occupancy simulations during 3:35 p.m. and 3:50 p.m. with and without occurrence of an incident is plotted in figure 7.13, 7.14, 7.15 and 7.16.

It can be seen from the simulations, the traffic volume counts 'with incident' are less than that 'without incident' (at about 3:40 p.m.) at the upstream (figure 7.13) and downstream (figure 7.14) of cell **40,** which is the point of occurrence of the incident. In case of cell-occupancy data, the simulations decreases at downstream (figure 7.16), but increases at the upstream point (figure 7.15) during the occurrence of the incident.

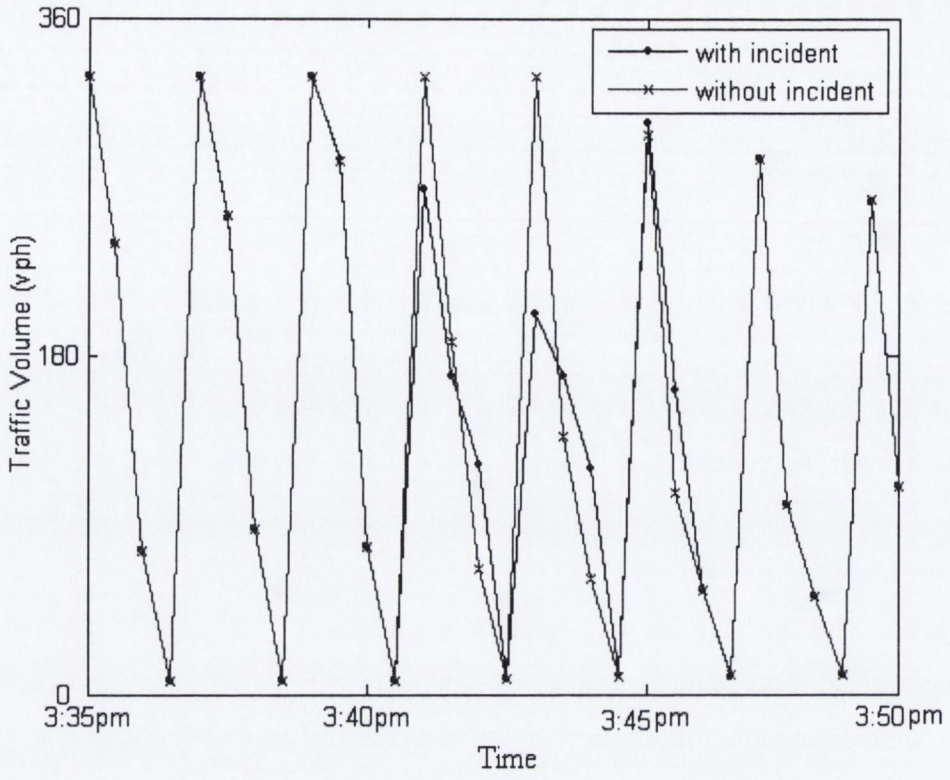


Figure 7.13 Plot of Upstream Traffic Flow.

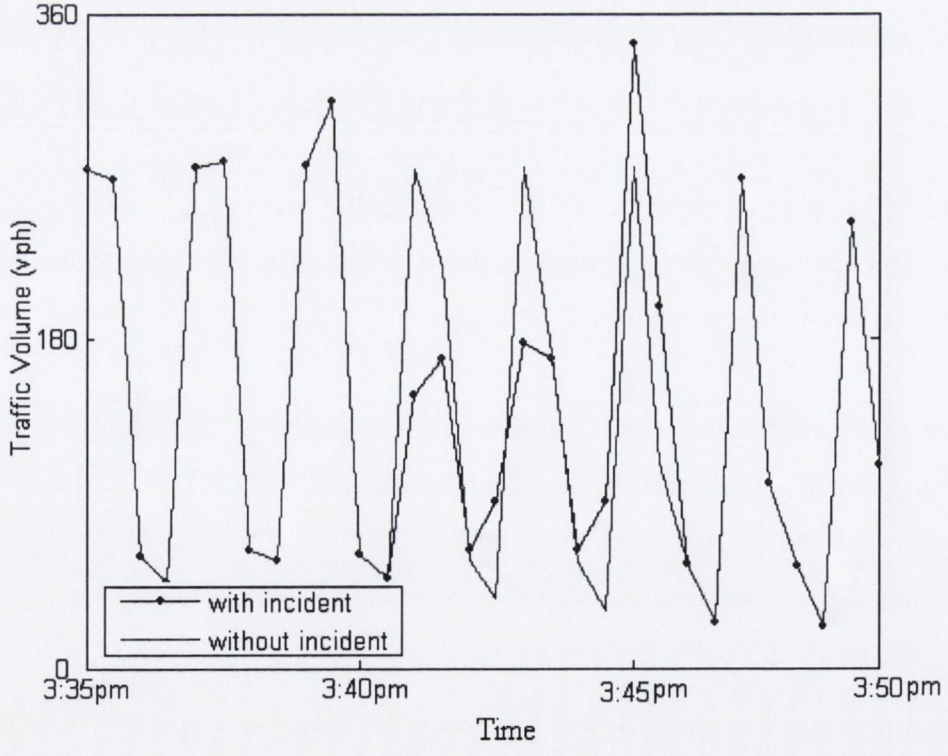


Figure 7.14 Plot of Downstream Traffic Flow.

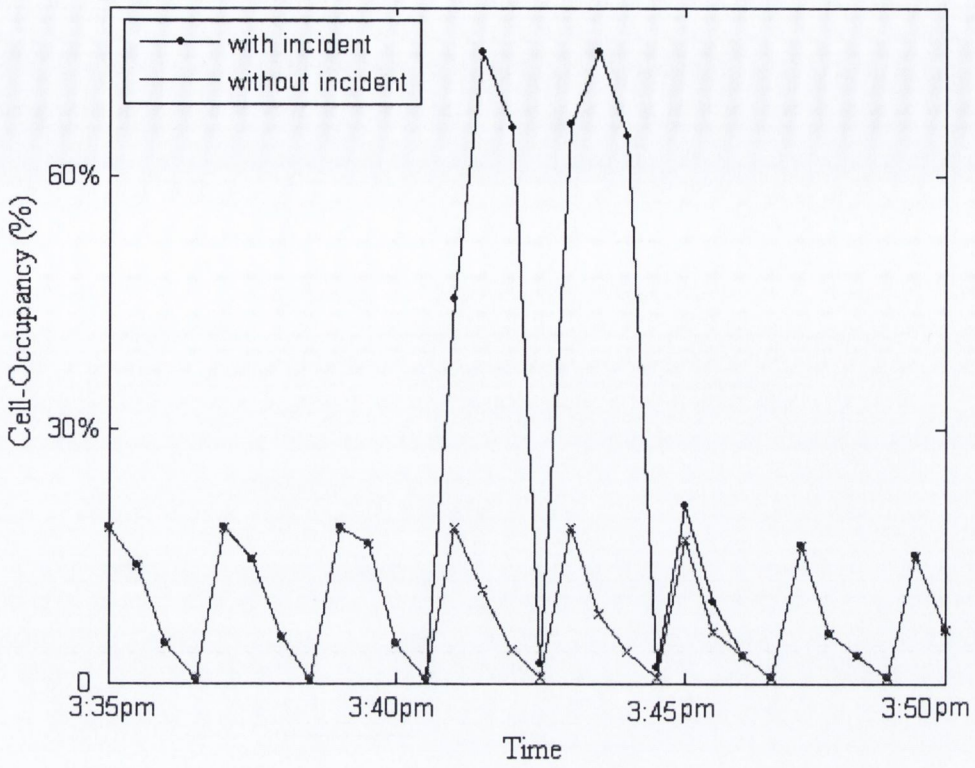


Figure 7.15 Plot of Upstream Cell-Occupancy.

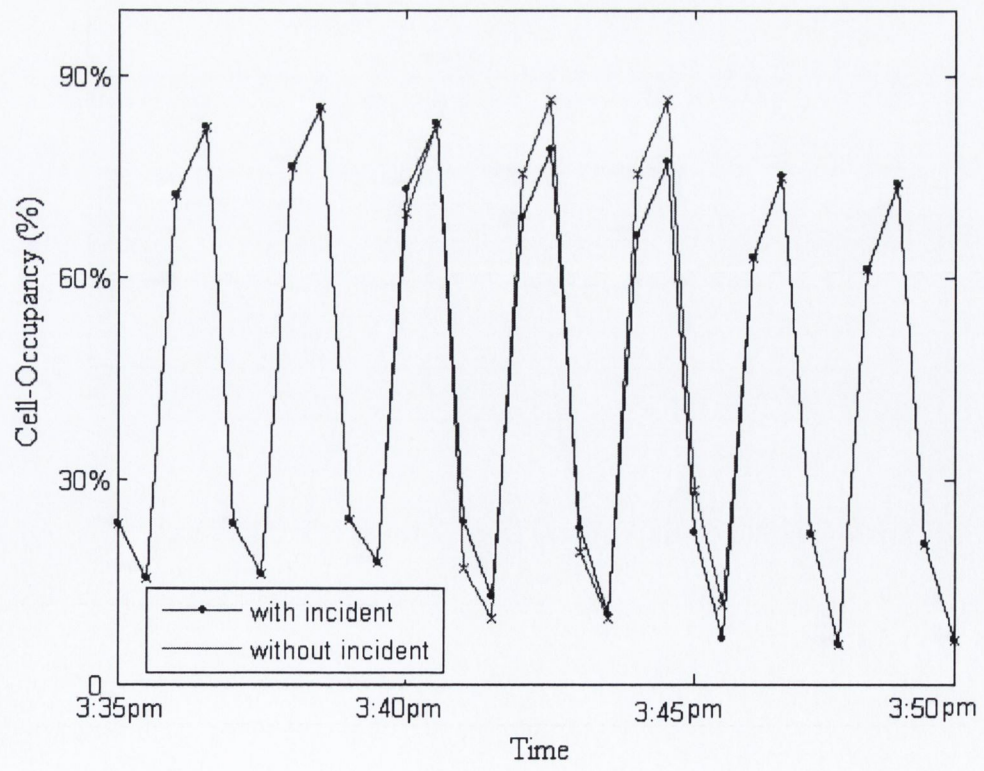


Figure 7.16 Plot of Downstream Cell-Occupancy.

The decomposition of downstream traffic volume at all the three levels with and without incident are shown in figure 7.17. The effect of incident is the most evident in details at level 2 and at level 1. By effective removing of the underlying low-resolution part of the signal, the changes occurring due an incident is made more evident.

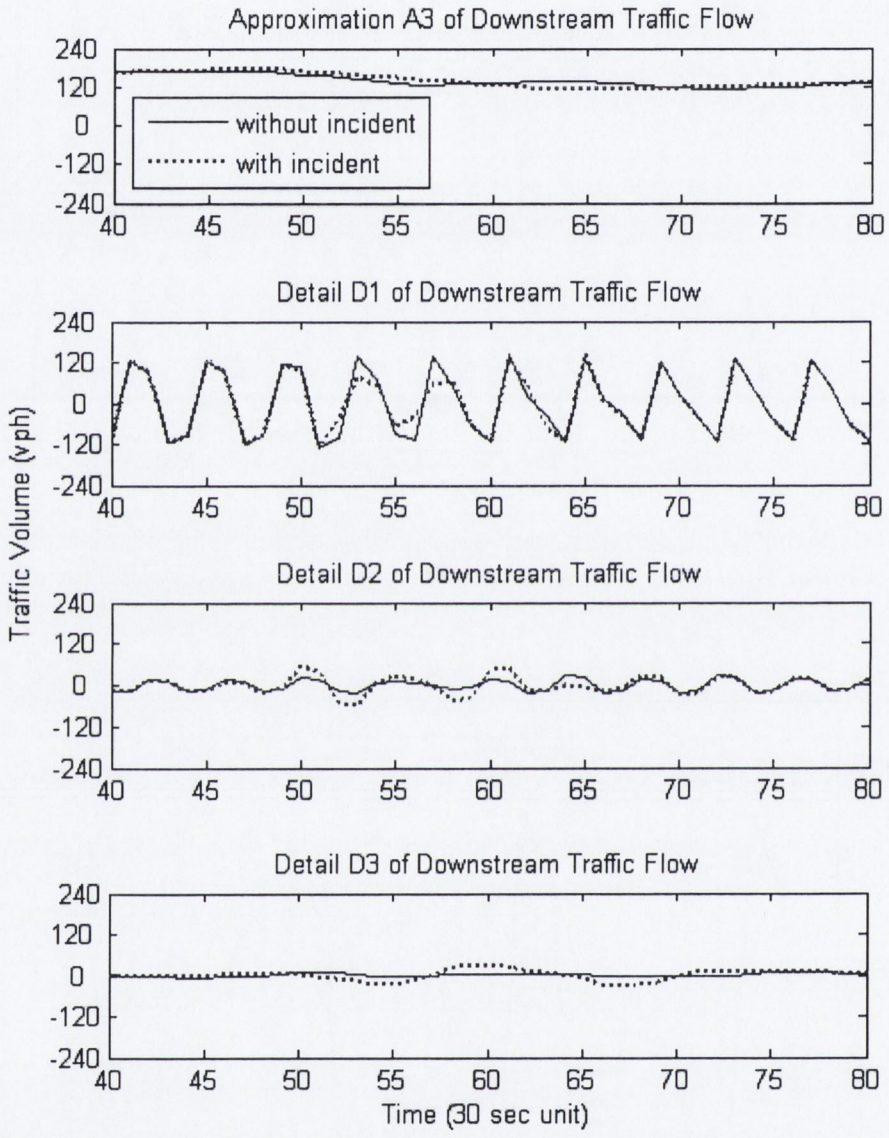


Figure 7.17 Downstream Traffic Volume Decomposed at All Levels.

Comparison of Variance

The current online variances over 3.5 minutes are calculated using a sliding window as described in section 7.5.1. A plot of the current online variances for downstream traffic flow and upstream cell occupancy during the time of incident is shown in figure 7.18. The confidence interval of half-hourly variance for upstream occupancy and downstream traffic

flow are shown as two parallel solid lines with no marker in both the cases. The interval obtained from the deterministic criteria is shown as two parallel dotted lines with no marker in the graphs.

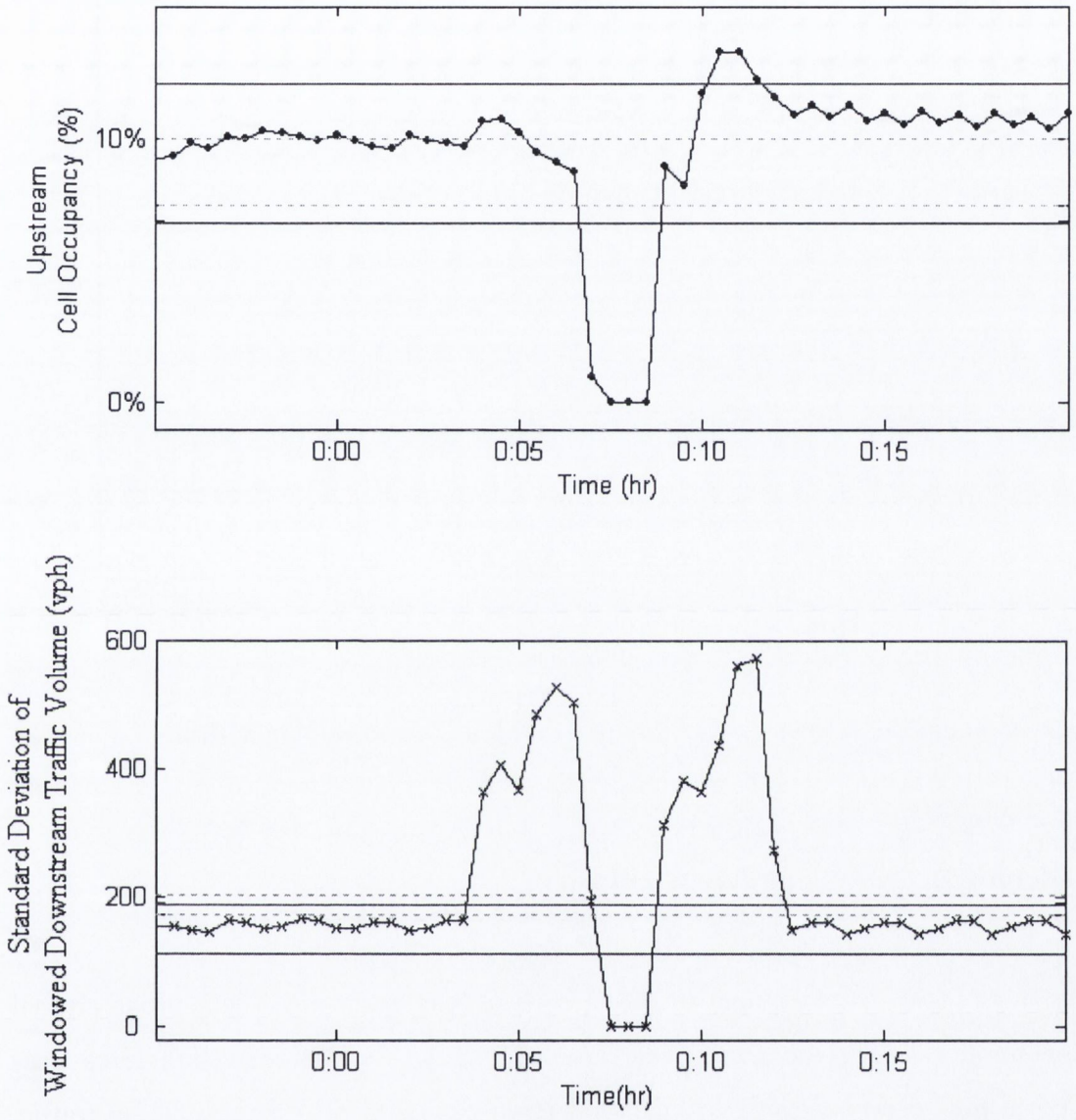


Figure 7.18 Comparison of Variance.

Several points (> 2) on both the curves fall outside the allowable range satisfying either criteria 1 or criteria 2. The diagram of the online graphic window showing incident detection is presented in figure 7.19. The two states are denoted by 0 (no violation of confidence interval) and 1(violation of confidence interval) along the Y-axis in the figure. The downstream traffic flow parameter is represented by a red dot and the upstream cell-occupancy is represented by a blue square. When both the parameters are in state 1 for

more than two consecutive time points the alarm for detecting incidents goes off. The time points in which a incident is detected is shown in a circle in the figure.

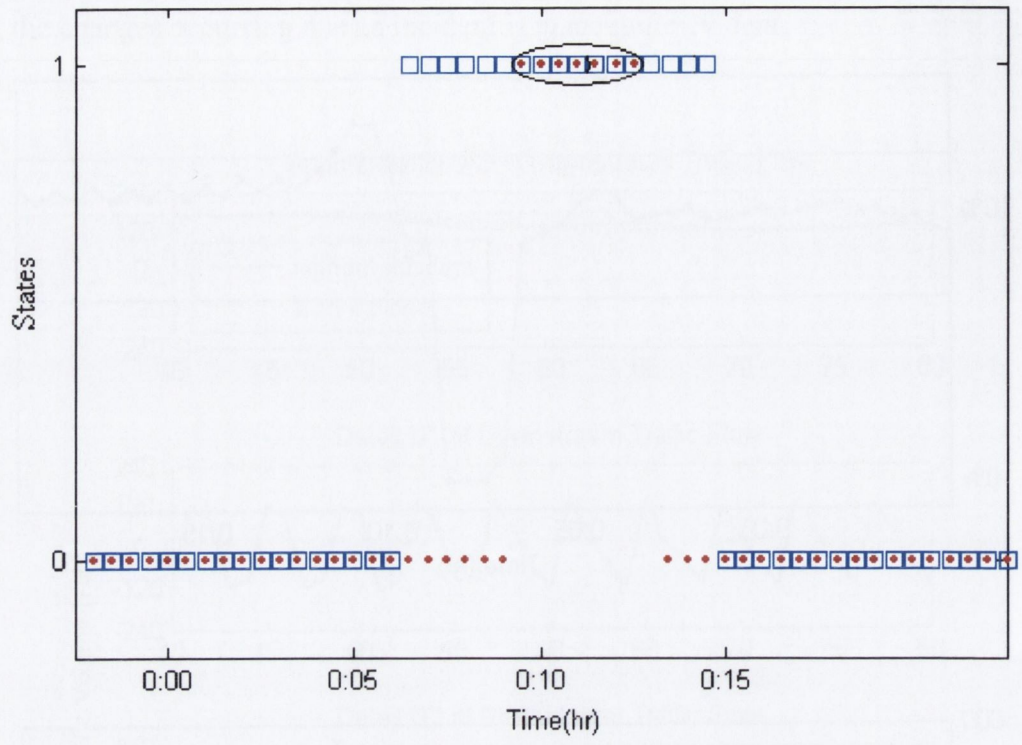


Figure 7.19 Graphic Window of Automatic Incident Detection Algorithm.

7.5.4 Performance of the Algorithm

Three types of incidents blocking 1 or 2 or all 3 lanes can occur in a 3 lane urban arterial chosen here to estimate the performance of the AIDA. Incident detection rates vary according to the severity of an incident and its consequent effect on surrounding traffic. Each lane type and corresponding detection rate from the proposed AIDA are given in table 7.1.

Incident Severity	Number Of Lanes Blocked	Detection Rate	False Alarm Rate
Type I	3	100%	0.83%
Type II	2	98%	0.83%

Table 7.1 Lane Type and Performance

The algorithm can not detect 1 lane blocking incident. All the false alarms are due to 1 lane blocking incidents which last for less than 3 minutes. In a congested urban network, single lane blockage for a few minutes primarily occurs due to congestion. Hence, in the city-centre of Dublin, the changes due to this kind of single lane blockage should be included in regular hourly variability and can not be detected using this algorithm. This particular flaw of the algorithm can be overlooked based on the fact that minor incidents blocking one lane for a few minutes are not required to be detected as their effects are extremely localized. It performs very well for 3 lane and 2 lane blocking incidents with a 0.83 % overall false alarm rate. The low false alarm rate and high detection rate provides the proposed AIDA with excellent potential for being used in a congested urban scenario.

Hence, the proposed technique can successfully detect incidents with 2/3 lane blockings in a 3 lane urban arterial network with insignificant false alarm. The simulation procedure mainly uses loop-detector data obtained from the SCATS system employed in traffic control of the city-centre of Dublin on a previously calibrated network. As SCATS does not include any intrinsic incident detection system, this methodology described here can be incorporated for incident detection in urban networks employing SCATS for traffic control.

CHAPTER 8

CONCLUSIONS

8.1 RESEARCH SUMMARY

Efficient traffic flow modelling and incident detection techniques applicable to ITS equipped urban transportation networks are developed in this thesis. The various models developed for this purpose are mainly using time-series analysis or wavelet analysis based techniques. The models developed have been validated on single or multiple junctions in the transport network at the city-centre of Dublin to test the effectiveness of the proposed methodologies.

In the view of developing an efficient yet simplistic univariate short-term traffic forecasting model for the junctions in an urban transport network (like Dublin city-centre), three different time-series techniques viz. the random walk model, the HWES model and the SARIMA model are used for simulation and forecasting of the 15 minute aggregate traffic volume observations obtained from those junctions. The periodic nature of the traffic flow data is confirmed from the correlogram and the partial ACF plot of the traffic volume observations. According to the forecasting performance, the HWES model and the SARIMA model are found to be suitable candidates for short-term univariate traffic flow forecasting in an urban signalised transport network. Unlike the *ad-hoc* (Chatfield, 2001) HWES model, the SARIMA model is based on probability theory and is consequentially chosen for further investigation and improvement in relation with traffic flow modelling.

In the existing traffic flow modelling studies using the SARIMA model, the parameters of the model are estimated using classical time-series analysis techniques (maximum likelihood estimates and/or least square estimates). A Bayesian inference technique has been employed in the thesis to estimate the parameters of the SARIMA model instead of using the classical approach. Each of the estimated parameters from the Bayesian inference has a probability density function conditional to the observed traffic volumes. In formulation of the Bayesian analysis, MCMC simulation algorithm is proposed and is used to solve the posterior integration problem in high dimensions. Superiority of the Bayesian

SARIMA model to the classical SARIMA model is presented in a qualitative and quantitative comparison between the forecasts from both the approaches.

Univariate statistical traffic flow models lack the physical basis of traffic movement. The concept of inclusion of the time-series observations to the traffic flow theory models is considered to improve the robustness of the univariate time-series models used for traffic flow modelling. In this context, the responses from two traffic flow theory models, viz. the hydrodynamic model and the whole-link model are compared under the application of various types of traffic demands such as the inductive loop-detector observations from an urban arterial or suddenly changing continuous or discontinuous inflow. The solution algorithms used for the models are similar to the existing studies in this field. However, while considering the analytical solution procedure of the hydrodynamic model instead of the regular shock wave analysis, a simplistic assumption has been introduced to reach a unique solution. From this comparative study, the numerical approximation of the hydrodynamic model i.e. the CTM model is chosen to provide a theoretical basis to the univariate short-term time-series traffic flow forecasting models.

Next, the research on developing an efficient traffic flow model for an urban transport network is extended to a multivariate regime. Two types of multivariate traffic flow models are proposed in this thesis. The first model is a multi-input multi-output (number of input sites greater than the number of output sites) traffic flow model based on an empirical methodology of the SUTSE model. The relative distances among the data-collection sites are not required to be considered in developing this model. The effects of the changes in the traffic volume in the immediate upstream junction to any data-collection site are included as explanatory or exogenous variables in the SUTSE model. The second multivariate traffic flow model is essentially an extension of the concept of merging the theoretical and the empirical traffic flow modelling techniques to a multivariate frame work. A multivariate multi-step traffic forecasting model is developed which can capture the traffic flow dynamics as well as the temporal variation of the traffic conditions by integrating the SARIMA time-series forecasting technique to the CTM model. A case study using the 15 minute aggregate traffic volume observations from multiple sites in the congested transport network in the city-centre of Dublin is presented for both the multivariate traffic flow models.

The applicability of the wavelet analysis methods in traffic flow modelling is studied in the last part of the thesis. Wavelet analysis technique is utilised in developing a background simulation model for the traffic flow observations from a signalised urban arterial. First of all, a 'regular average daily trend' is modelled by isolating the low resolution component from the high resolution components of the univariate 5 minute aggregate traffic volume observations. Later the high resolution components are modelled using a Bayesian hierarchical model. The time-varying variability of the high resolution components is taken into account in the model. An application of the simulation methodology is proposed by developing an AIDA for the signalised urban arterials. A case study on the developed AIDA has been illustrated using a CTM based simulation strategy.

8.2 RESEARCH FINDINGS AND CRITICAL ASSESSMENTS

This section presents the conclusions that can be drawn on the applicability and effectiveness of the models developed in this thesis in relation to an ITS equipped congested urban transport network.

The models were tested on the traffic volume observations and other traffic conditions obtained from the different parts and signalised intersections of the transport network at the city centre of Dublin. The traffic flow observations were collected as a part of the SCATS system (the ITS equipped UTC of Dublin) and hence the models tested can be incorporated to the system. The 15 minute/5 minute/2 minute/30 seconds aggregate traffic volume observations used for this purpose of were observed to posses certain properties which are worthwhile to discuss.

- The travel behaviour and the traffic flow pattern at the city-centre during the weekends are distinctively different from that of the weekdays. The models developed in this thesis focuses on modelling the weekday traffic conditions.

- According to the correlogram of the original traffic volume observations, the data are non-stationary in nature and a seasonal difference is required to be applied in certain time-series techniques to model the observations.

- The traffic flow observations in weekdays show a bimodal nature with two discernable crests signifying the morning and evening peak hours. The traffic flow observations over a few weeks (except the weekends) show a definitive 'daily' periodicity and a stable and nearly constant trend (which may change for greater time intervals with changes in prevailing social and economic conditions).

To find the suitability of the existing freeway based univariate short-term traffic flow time-series models (the random walk model, HWES model and the SARIMA model) when applied to the intersections at the city-centre of the city of Dublin, the parameters of the three univariate short-term time-series models were estimated only for the 15 minute aggregate traffic volume observations from a single junction (TCS 183) at the city-centre of Dublin. Being a part of the same congested urban network the other intersections at the city-centre can be modelled using the same techniques. The HWES technique was found to be the most suitable for the intersection TCS 183. However, the forecasting performance of the SARIMA model is only 0.06% less than the HWES model. The error estimates obtained using all these models are considerably large compared to the results when the same models are applied on a freeway (Williams, 1999). Both the forecasting models performed better for peak hour traffic volume observations than for the off-peak hour data. If the changes in traffic volume due to occurrence of incidents can be treated as outliers, then it is possible to get better forecasting precision from these models. Both the HWES and the SARIMA models cannot predict early morning observations very well, however, if the starting point of the traffic flow observations are moved to 6 a.m. instead of 12 p.m. the forecasting performances for both the models increase considerably. As the SARIMA model is based on probability theory, it was chosen in favour of HWES model as the most suitable time-series technique to model the traffic volume observations obtained from an urban signalised junction.

A Bayesian estimation procedure was employed for estimating the parameters of a suitable SARIMA model used for short-term traffic flow forecasting. In Bayesian inference, the model parameters and the points of forecasts are estimated and expressed as probability density functions conditional on the observed traffic volumes unlike the constant values obtained from the classical inference. Hence, the variability of the parameters over time according to the variable traffic flow pattern can be considered in the Bayesian SARIMA model. Due to this additional information, the forecasts from the Bayesian SARIMA model can better match the urban traffic flow pattern with rapid fluctuations and extreme peaks.

Computational complexity has been one of the main limitations for not applying Bayesian techniques more widely in the field of statistical inference. The use of practical assumptions with respect to the available traffic information in the Bayesian estimation technique proposed in this study reduces the complexity of the model.

As the traffic inflow has short-term variability, for the purpose of calculation of traffic outflows, two well-known traffic flow theory models (whole-link model and hydrodynamic model) were compared under the application of different inflow conditions on a single link or arterial. The suddenly changing inflows (continuous and discontinuous) represented the real life situations of recurrent and non-recurrent congestions. The inductive loop-detector traffic volume observations were used for checking the applicability of these theoretical models in a real life ITS equipped urban signalised arterial. The study showed that the nature of outflows from both the models were partially dependent on the solution algorithms used. From the point of merging these theoretical models with time-series forecasting techniques, the numerical solution of the hydrodynamic model is the most suitable solution strategy.

Multivariate models for traffic forecasting were also proposed in this thesis and it had been concluded that these models can increase the efficiency of the traffic flow simulation and forecasting algorithms used in the ITS system of any urban transport network. The SUTSE based multivariate time-series traffic forecasting models can consider the effect of traffic flow variations at upstream junctions (as explanatory variables), data outliers and missing values more simplistically than the VARMA or ARIMAX models. Unlike VARMA or ARIMAX models, the computational complexity involved in the proposed multivariate traffic flow model is comparatively less since the estimation of the covariance matrix is not necessary in this procedure. Consequently, the computational burden increases linearly with the size of the network in contrast to a quadratic increase for the VARMA and ARIMAX models. As the location of the data-collection sites are not required to be included in the model, such information is not required to be collected from the real-life transport network on which the multivariate traffic flow model is applied.

In this study, the concept of merging the theoretical and the statistical traffic flow modelling techniques was extended to a network context, i.e. to a multivariate regime. The developed multivariate multi-step traffic flow forecasting methodology can capture the effect of queue spillback and the seasonal nature of the traffic flow pattern in a congested

transport network. The model performed satisfactorily when applied to a network of multiple intersections and signalized arterials at the city-centre of Dublin. Another major contribution of this methodology is that the real-time traffic volume can also be predicted at the junctions where traffic counts are not collected regularly. The requirement of the model is that the application of this forecasting algorithm in an ITS equipped transport network necessitates accurate calibration of the simulation network by collecting information from the real-life system, however this is not a major limitation.

Traffic flow models and incident detection techniques suitable for an ITS equipped UTC were developed using wavelet analysis. The main advantage of using wavelet analysis is that the low resolution components of a traffic flow observation dataset can be isolated from its high resolution components with the help of the technique. This gives the flexibility to zoom into a particular frequency which is the most important for a certain analysis. A novel and important concept of time-varying variance of the high resolution components from the traffic flow observations was introduced. This concept was exploited in developing a Bayesian hierarchical model of the residuals (which is also the summation of the all the high resolution components) obtained after subtracting the 'regular average daily trend' from a set of traffic flow observations. The 'trend' was possible to be modelled in a non-functional form by virtue of the application of MRA using wavelets. In addition to the traffic flow simulation technique, the concept was also utilised while developing a wavelet based AIDA for urban signalized arterials. Using the MRA methodology, it was possible to isolate and model the high resolution components where the changes due to occurrence of the incidents were the most prominent. The high detection rates for two and three lane blocking incidents on urban arterials have proved that the proposed AIDA is efficient and can be incorporated in ITS equipped UTC systems like SCATS where no separate incident detection technique exists currently.

8.3 RECOMMENDATIONS FOR FURTHER RESEARCH

The studies in this thesis are aimed at developing traffic flow models and suitable AIDA for ITS equipped UTC systems. Apart from this thesis and a handful of available information from other studies all other existing research in this field focus on the application of ITS in a highway and freeway based transport network. Hence, it is felt that

there are definite scopes and requirements for further studies in the development and application of ITS in an urban scenario.

A straightforward extension from the studies on the Bayesian SARIMA model developed for short-term univariate traffic flow simulation and forecasting is to introduce an online Bayesian updation scheme for estimating the unknown parameters of the model. This study will enable the traffic management authorities in real-time implementation of the more efficient Bayesian univariate traffic forecasting algorithm in ITS based UTC systems like SCATS which currently use an inefficient random walk model for the same purpose.

The traffic flow observations collected from the inductive loop-detectors of congested urban arterial show the presence of outlier observations due to the occurrence of recurrent and non-recurrent congestions (incidents). Apart from that, due to serviceability problems, a certain percentage of the loop-detectors in an urban transport network often remain defunct and consequently the presence of 'missing values' in the traffic flow observations is a regular phenomenon. The STM methodology has been observed to be the most suitable time-series analysis technique for the purpose of accounting for these problems while modelling traffic flow observations. STM methodology was employed in developing a multivariate traffic flow model in this thesis. Further research on this methodology in univariate and multivariate regime for modelling traffic flow data containing outliers and missing values should be performed to exploit the full potential of the methodology as a continuation of this thesis.

Apart from the other influencing factors, traffic volume in a network is affected largely by some factors like weather conditions which are beyond human control. There is no significant existing research in studying this aspect and studies on the effect of weather conditions included in both univariate and multivariate STM as explanatory variables is a potential area of research for investigation.

The concept of 'time-varying variance' developed in chapter 7 has immense potential in improving the efficiency of the traffic flow simulation and forecasting strategies through the development of new algorithms. It is suggested that the 'time-varying variance' of the residuals used in the Bayesian hierarchical technique be modelled using a regression equation with time as the independent variable. The parameters of the regression model can be estimated using Bayesian inference.

The CTM based multivariate multi-step traffic flow simulation and forecasting model developed in Chapter 6 has a huge potential for future research. The model can be utilised to develop a traffic signal control and management scheme by using a multi-objective optimisation algorithm on maximising the outflow from the exit points in the network and by minimising the total travel time delay of the network. In this model, further work is required for including the lane changing behaviour to the CTM model along with modelling the phenomenon of queue spillback. The applicability of the multivariate multi-step traffic flow model developed in this thesis has a promising possibility of application on an ITS equipped freeway transport network.

REFERENCES

Abdulhai, B. and S. G. Ritchie. (1999) Enhancing the Universality and Transferability of Freeway Incident Detection Using a Bayesian-Based Neural Network. *Transportation Research Part C: Emerging Technologies*, Vol.7, pp. 261-280.

Adeli, H. and A. Samant. (2000) An Adaptive Conjugate Gradient Neural Network-Wavelet Model For Traffic Incident Detection. *Computer-Aided Civil Infrastructure Engineering*, Vol.15, pp. 251–260.

Adeli, H., and A. Karim. (2000) Fuzzy-Wavelet RBFNN Model for Freeway Incident Detection. *Journal of Transportation Engineering*, ASCE, Vol. 126, pp. 464–471.

Ahmed, M. S. and Cook, A. R. (1979) Analysis of freeway traffic time-series data by using Box–Jenkins techniques. *Transportation Research Record: Journal of the Transportation Research Board, No. 722*, pp.1–9.

Ahmed, M. S. and Cook, A. R. (1982) Application of Time-Series Analysis Techniques to Freeway Incident Detection. *Transportation Research Record: Journal of the Transportation Research Board, No. 841*, pp. 19-21.

Ashok, K. and Ben-Akiva, M. E. (2000) Alternative Approaches For Real-Time Estimation and Prediction of Time-Dependent Origin–Destination Flows. *Transportation Science*, Vol. 34, pp. 21-36.

Akaike, H. (1974). A New Look at Statistical Model Identification. *IEEE Transactions on Automatic Control*, Vol. AC-19, pp. 716-723.

Astarita, V. (1996) A Continuous Time Link Model For Dynamic Network Loading Based On Travel Time Function. *13th International Symposium on Theory of Traffic Flow*, pp. 79-103.

Box, G. E. P. and Jenkins, G. M. (1976) *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Carey, M. and McCartney, M. (2002) Behaviour of a Whole-Link Travel Time Model Used in Dynamic Traffic Assignment. *Transportation Science*, Vol.36 (1), pp. 83-95.

Carey, M. (2003) Pseudo Periodicity in Travel-time Model Used in Dynamic Traffic Assignment. *Transportation Research Part B: Methodological*, Vol.37, pp. 769-792.

Carey, M. and Ge, Y. E. (2003) Comparing Whole-Link Travel Time models. *Transportation Research Part B: Methodological*, Vol.37, pp. 905-926.

Carey, M. and Ge, Y. E. (2003) Efficient Discretization of Link Travel Time Models Used in DTA. *Networks and Spatial Economics*, Vol. 4(3), pp. 269-290.

Carlin, B. P. and Louis, T. A. (1996) *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall, New York.

Chang E. (1998) Implementing a cell-based signal control model in Hong Kong. *M. Phil. Thesis*, Hong Kong University of Science & Technology, China.

Chang, E. P. and. Wang, S. H. (1995) Improved Freeway Incident Detection Using Fuzzy Set Theory. *Transportation Research Record: Journal of the Transportation Research Board, No. 1453*, pp. 75-82.

Chatfield, C. (2001) *Time-Series Forecasting*. Chapman and Hall/CRC, London.

Chatfield, C. (2004) *The Analysis of Time Series: An Introduction, Sixth Edition*. Chapman and Hall/CRC, London.

Chen, S., Wang, W. and Qu, G.(2004) Combining Wavelet Transform and Markov Model to Forecast Traffic Volume, Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai.

Cheu, R. L. and Ritchie, S. G. (1995) Automated Detection of Lane-Blocking Freeway Incidents Using Artificial Neural Networks. *Transportation Research Part C: Emerging Technologies*, Vol. 3, pp. 371-388.

Chung, E. and Rosalion, N. (2001) Short Term Traffic Flow Prediction. *Proceedings of the 24th Australian Transportation Research Forum*, Hobart, Tasmania.

Cook, A. R. and Cleveland, D. E. (1974) Detection of Freeway Capacity-Reducing Incidents by Traffic-Stream Measurements. *Transportation Research Record: Journal of the Transportation Research Board, No.495*, pp. 1-11.

Daubechies, I. (1992) *Ten Lectures on Wavelets*. SIAM, Philadelphia.

Daganzo, C. F. (1994) The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. *Transportation Research. Part B: Methodological*, Vol.28, pp. 269-287.

Daganzo, C. F. (1995a) The Cell Transmission Model, Part II: Network Traffic. *Transportation Research Part B: Methodological*, Vol.29, pp. 79-93.

Daganzo, C. F. (1995b) Finite Difference Approximation of the Kinematic Wave Model of Traffic Flow. *Transportation Research Part B: Methodological*, Vol.29, pp. 261-276.

Davis, G. and Nihan, N. (1991) Nonparametric Regression and Short-Term Freeway Traffic Forecasting. *Journal of Transportation Engineering*, ASCE, Vol. 117, pp. 178-188.

Dia, H. and Rose, G. (1997) Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data. *Transportation Research. Part C: Emerging Technologies*, Vol. 5, pp. 313-331.

Dudek, C. L., Messer, C. J. and Nuckles, N. B. (1974) Incident Detection on Urban Freeway. *Transportation Research Record: Journal of the Transportation Research Board, No. 495*, pp. 12-24.

Durbin, J. and Koopman S. J. (2001) *Time Series Analysis by State Space Methods*. Oxford Statistical Science Series, Oxford University Press.

Friesz, T. L., Bernstein, D., Smith, T. E., Tobin, R. L. and Wie, B. W. (1993) A Variational Inequality Formulation of the Dynamic Network User Equilibrium Problem. *Operations Research*, Vol.41 (1), pp. 179-191.

Friesz, T. L., Bernstein, D., Suo, Z. and Tobin, R. L. (2001) Dynamic Network User Equilibrium with State-Dependent Time Lags. *Networks and Spatial Economics*, Vol. 1(3/4), pp. 319-347.

Fuller, W.A. (1996) *Introduction to Statistical Time Series (Second Edition)*. John Wiley and Sons, Inc., New York.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Trans. On Pattern Analysis and Machine Intelligence,* Vol. 6, pp. 721-741.

Ghosh, B., Basu, B. and O'Mahony, M. M. (2005) Time-Series Modelling for Forecasting Vehicular Traffic Flow in Dublin. *84th Annual Meeting of Transportation Research Board (CD-ROM), TRB, Washington, D. C.*

Ghosh, B., Szeto, W. Y., Basu, B. and O'Mahony, M. M. (2005a) Cell-Based Short-Term Traffic Flow Forecasting Using Time Series Modelling. *The Proceedings of the 10th International Conference Of Hong Kong Society For Transportation Studies.*

Ghosh, B., Basu, B. and O'Mahony, M. M. (2006) Analysis of Trend in Traffic Flow. *Proceedings of Institute of Engineering and Technology: Irish Signals and Systems Conference*, pp. 415-419.

Ghosh, B., Basu, B. and O'Mahony, M. M. (2006a) A Bayesian Time-Series Model For Short-Term Traffic Flow Forecasting. In press, *Journal of Transportation Engineering*, ASCE.

Ghosh-Dastidar, S. and Adeli, H. (2003) Wavelet-Clustering-Neural Network Model for Freeway Incident Detection. *Computer-Aided Civil and Infrastructure Engineering*, Vol.18 No.5, pp. 325-338.

Guin, A. (2004) An Incident Detection Algorithm Based on a Discrete State Propagation Model of Traffic Flow. *Ph. D. Dissertation*, Georgia Institute of Technology.

Hall, F. L., Shi, Y. and Atala, G. (1993). On-Line Testing of the McMaster Incident Detection Algorithm under Recurrent Congestion. *Transportation Research Record: Journal of the Transportation Research Board, No.1394*, pp. 1-7.

Hamed, M. M., Al-Masaeid, H. R. and Bani Said, Z.M. (1995) Short-Term Prediction of Traffic Volume in Urban Arterials. *Journal of Transportation Engineering*, ASCE, Vol. 121(3), pp. 249–254.

Han, L. D. and May, A. D. (1990) Artificial Intelligence Approaches for Urban Network Incident Detection and Control. *Traffic Control Methods: Proceedings of the Engineering Foundation Conference*, New York, pp. 159–176.

Harvey, A. C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press.

Hastings, W. K. (1970) Monte Carlo Sampling Methods Using Markov Chains and Their Application. *Biometrika*, Vol. 57, pp. 97-100.

Head, L. K. (1995) Event-Based Short-Term Traffic Flow Prediction Model. *Transportation Research Record: Journal of the Transportation Research Board, No. 1510*, pp. 45–52.

Heydecker, B. G. and Addison, J. D. (1998) Analysis of Traffic Models for Dynamic Equilibrium Traffic Assignment. *Transportation Networks: Dynamic Flow Modelling and Control*, Springer, London, pp. 213-231.

Highway Capacity Manual, Special Report 209 (1997) Transportation Research Board, Washington, D.C.

Hsiao, C. H., Lin, C. T. and Cassidy, M. (1994) Application of Fuzzy Logic and Neural Networks to Automatically Detect Freeway Traffic Incidents. *Journal of Transportation Engineering,* ASCE, Vol. 120(5), pp. 753-772.

Hu, W., Xiao, X., Xie, D., Tan, T., and Maybank, S. (2004) Traffic Accident Prediction Using 3-D Model-Based Vehicle Tracking. *IEEE Transaction on Vehicle Technology*, Vol. 53(3), pp. 677–694.

Ikeda, H., Matsuo, T., Kaneko, Y., and Tsuji, K. (1999) Abnormal Incident Detection System Employing Image Processing Technology. Proceedings of IEEE International Conference of Intelligent Transportation Systems (ITS'99), pp.748–752.

Ishak, S. S. and Al-Deek, H. M. (1998) Freeway Incident Detection Using Fuzzy Art (Adaptive Resonance Theory). *Proceedings of Fifth International Conference on Applications of Advanced Technologies in Transportation Engineering*, pp. 59-66.

Ishak, S. S. and Al-Deek, H. M. (1999) Fuzzy Art Neural Network Model for Automated Detection of Freeway Incidents. *Transportation Research Record: Journal of the Transportation Research Board, No.1634,* pp. 56-63.

Kalman, R. E. (1960) A New Approach to Linear Prediction And Filtering Problems. *Transactions of the ASME - Journal of Basic Engineering*, Vol. 82, pp. 35-45.

Kamarianakis, Y. and Prastacos, P. (2002) Space-Time Modelling Of Traffic Flow. *European Regional Science Association Conference*. Available from www.ersa2002.org.

Kamarianakis, Y. and Prastakos, P. (2003) Forecasting traffic flow conditions in an urban network: comparison of multivariate and univariate approaches. *82nd Annual Meeting of Transportation Research Board, (CD-ROM), TRB, Washington, D. C.*

Karim, A. and Adeli, H. (2002) Incident Detection Algorithm Using Wavelet Energy Representation of Traffic Patterns. *Journal of Transportation Engineering*, ASCE, Vol. 128, pp. 232-242.

Kaysi, I., Ben-Akiva, M. and Koutsopoulos, H. (1993) An integrated approach to vehicle routing and congestion prediction for real-time driver guidance. *Transportation Research Record: Journal of the Transportation Research Board, No.1408*, pp. 66–74.

Kirby, H. R., Watson, S. M. and Dougherty, M. S. (1997) Should We Use Neural Network Or Statistical Models For Short-Term Motorway Traffic Forecasting? *International Journal of Forecasting*, Vol.13, pp. 43-50.

Koopman, S. J., Harvey, A.C., Doornik, J.A. and Shephard, N. (1999a) S*tructural Time Series Analysis, Modelling and Prediction Using STAMP.* London: Timberlake consultants Press.

Koopman, S. J., Shephard, N. and Doornik, J.A. (1999b) Statistical Algorithms for Models in State Spaceusing SsfPack 2.2 (with discussion). *Econometrics Journal,* Vol. 2, pp.113-166.

Lee, D. H., Zheng, W. Z. and Shi, Q. X. (2004) Short-Term Freeway Traffic Flow Prediction Using a Combined Neural Network Model. *84th Annual Meeting of Transportation Research Board (CD-ROM), TRB, Washington, D. C.*

Lee, P. M. (1997) *Bayesian Statistics: An Introduction (Second edition).* Arnold, A member of the Hodder Headline group, London.

Lee, S. and Fambro, D. B. (1999) Application of Subset Autoregressive Moving Average Model For Short-Term Freeway Traffic Volume Forecasting. *Transportation Research Record: Journal of the Transportation Research Board, No.*1678, pp. 179–188.

Levin, M. and Krause, G. M. (1978) Incident Detection: A Bayesian Approach. *Transportation Research Record: Journal of the Transportation Research Board, No. 682,* pp. 52-58.

Levin, M., Tsao and Y. D., (1980) On Forecasting Freeway Occupancies and Volumes. *Transportation Research Record: Journal of the Transportation Research Board, No. 773*, pp. 47–49.

Lighthill, M. J. and Whitham, J. B. (1955) On Kinematic Waves. I. Flow Movement in Long Rivers. II. A theory Of Traffic Flow on Long Crowded Roads. *Proceedings of Royal Society London, Part A,* Vol. 229, pp. 281-345.

Lin, C. K. and G. L. Chang. (1998) Development of a Fuzzy-Expert System for Incident Detection and Classification. *Mathematical and Computer Modelling*, Vol.27 (9-11), pp.9-25.

Lingras, P., Mountford, P. (2001) Time Delay Neural Networks Designed Using Genetic Algorithms For Short-Term Inter-City Traffic Forecasting. IEA/AIE 2001, LNAI 2070, pp. 290–299.

Lo, H., (1999) A Novel Traffic Signal Control Formulation. Transportation Research, Part A 33, 433-448.

Lo, H., (2001) A cell-based traffic control formulation: strategies and benefits of dynamic timing plans. Transportation Science 35 (2), pp. 148-164.

Lo, H., Chang, E. and Chan, Y.C., 2001. Dynamic network traffic control. Transportation Research A 35, pp. 721-744.

Luk, J.Y. K. (1984) Two Traffic-Responsive Area Traffic Control Methods: SCATS and SCOOTS. *Traffic Engineering & Control*, Vol. 25(1), pp. 14-22.

Makridakis, S., Wheelwright, S. C. and Hyndman, R. J. (1998) *Forecasting: Methods and Applications (Third Edition)*. John Willey & Sons, Inc., New York.

Mallat, S. (1989) A Theory For Multi-Resolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11(7), pp. 674–693.

Mallat, S. (1998) *A Wavelet Tour of Signal Processing (2nd Edition).* Academic Press, London.

Mao, G. (2005) A Timescale Decomposition Approach to Network Traffic Prediction. IEICE Transactions on Communication, Vol. E88–B(10), pp. 3974-3981.

Mathworks. (2000) *Wavelet toolbox for use with MATLAB: User's Guide Version 2.* Mathworks Inc. (http://www.mathworks.com)

McQueen, B. and McQueen, J. (1999) *Intelligent Transportation Systems Architecture.* Artech House Publishers, Inc., U.S.A.

Merchant, D. K. and Nemhauser, G. L. (1978a) A Model and an Algorithm for the Dynamic Traffic Assignment Problem. *Transportation Science*, Vol.12 (3), pp. 183-199.

Merchant, D. K. and Nemhauser, G. L. (1978b) Optimality Conditions For A Dynamic Traffic Assignment Model. *Transportation Science*, Vol.12(3), pp. 200-207.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953) Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, Vol. 21, pp. 1087-1092.

Michalopoulos, P. G., Anderson, C. A. and Jacobson, R. D. (1994) Development and Field Deployment of a Machine Vision Based Incident Detection System. *Proceeding of Moving Toward Deployment, IVHS AMERICA 1994 Annual Meeting.*

Monahan, J. F. (1983) Fully Bayesian analysis of ARMA time series models. *Journal of Econometrics*, Vol. 21, pp. 307-331.

Neal, R. M. (1993) Probabilistic inference using Markov Chain Monte Carlo methods. *Technical Report* CRG-TR-93-1.

Newell, G. F. (1993) A Simplified Theory Of Kinematic Waves In Highway Traffic, Part II: Queuing At Freeway Bottlenecks. *Transportation Research B*, Vol.27, pp. 289-303.

Nie, X. J. and Zhang, H. M. (2002a) The Delay-Function-Based Link Models: Their Properties and Computational Issues. *Technical Report UCD - ITS - Zhang -2002-2.*

Nie, X. J. and Zhang, H. M. (2002b) A Comparative Study Of Some Macroscopic Link Models Used In Dynamic Traffic Assignment. *Technical Report UCD - ITS - Zhang - 2002-3.*

Payne, H. J. and Tignor. S. C. (1978) Freeway Incident-Detection Algorithms Based on Decision Trees with States. *Transportation Research Record: Journal of the Transportation Research Board, No. 682*, pp. 30-37.

Persaud, B. N., Hall, F. L. and Hall, L. M. (1990) Congestion Identification Aspects of the Mcmaster Incident Detection Algorithm. *Transportation Research Record: Journal of the Transportation Research Board, No.1287*, pp. 167-175.

Persaud, B. N. and Hall, F. L. (1989) Catastrophe Theory and Patterns in 30-Second Freeway Traffic Data--Implications for Incident Detection. *Transportation Research. Part A: General*, Vol. 23, pp. 103-113.

Qiao, F., Wang, X. and Yu, L. (2003) Optimizing Aggregation Level for ITS Data Based On Wavelet Decomposition. *82th Annual Meeting of Transportation Research Board (CD-ROM), TRB, Washington, D. C.*

Ran, B., Boyce, D. E. and Leblanc, L. J. (1993) A New Class of Instantaneous Dynamic User-Optimal Traffic Assignment Models. *Operations Research*, Vol. 41(1), pp. 192-202.

Ravishanker, N. and Ray, B. K. (1997) Bayesian Analysis Of Vector ARMA Models Using Gibbs Sampling. *Journal of Forecasting*, Vol. 16, pp.177-194.

Richards, P.I. (1956) Shockwaves on the Highway. *Operations Research*, Vol.4, pp. 42-51.

Ritchie, S. G. and Cheu, R. L. (1993) Simulation of Freeway Incident Detection Using Artificial Neural Networks. *Transportation Research. Part C: Emerging Technologies*, Vol.1, pp. 203-217.

Roberts, G. O. and Sahu, S. K. (1997) Updating Schemes, Correlation Structure, Blocking and Parameterization of Gibbs Sampler. *Journal of Royal Society, Series B (methodological)*, Vol. 59 (2), pp. 291-317.

Samant, A. and Adeli, H. (2000) Feature Extraction for Traffic Incident Detection Using Wavelet Transform and Linear Discriminant Analysis. *Computer-Aided Civil Infrastructure Engineering* Vol.15, pp. 241–250.

Smith, B. L., Demetsky, M. J. (1994) Short-Term Traffic Flow Prediction: Neural Network Approach. *Transportation Research Record: Journal of the Transportation Research Board, No. 1453*, pp. 98–104.

Smith, B. L. and Demetsky, M. J. (1997) Traffic Flow Forecasting: Comparison of Modelling Approaches. *Journal of Transportation Engineering*, Vol. 123(4), pp. 261–266.

Smith, B. L., Williams, B. M., and Oswald, R. K. (2002) Comparison of Parametric And Nonparametric Models for Traffic Flow Forecasting. *Transportation Research, Part C: Emerging Technologies*, Vol. 10 (4), pp. 257–321.

Stathopoulos, A., Karlaftis, M. G. (2003) A Multivariate State-Space Approach for Urban Traffic Flow Modeling and Prediction. *Transportation Research Part C: Emerging Technologies*, Vol. 11(2), pp. 121–135.

Stephanedes, Y. J., and Chassiakos, A. P. (1993) Freeway Incident Detection through Filtering. *Transportation Research Part C: Emerging Technologies,* Vol.1, pp. 219-233.

Stephanedes, Y. J. and Liu, X. (1995) Artificial Neural Networks for Freeway Incident Detection. *Transportation Research Record: Journal of the Transportation Research Board, No. 1494*, pp. 91-97.

Sun, H., Xiao, H., Yang, F., Ran, B., Tao, Y. and Oh, Y. (2005) Wavelet Preprocessing for Local Linear Traffic Prediction. *85th Annual Meeting of Transportation Research Board, (CD-ROM), TRB, Washington, D. C.*

Tanner, M. A. (1996) *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions* (*Third Edition*), Springer-Verlag, New York.

Tebaldi, C., West, M. and Karr, A. F. (2002) Statistical Analysis of Freeway Traffic Flows. *Journal of Forecasting*, Vol. 21, pp. 39-68.

Teng, H. and Y. Qi. (2003) Application of Wavelet Technique to Freeway Incident Detection. *Transportation Research Part C: Emerging Technologies,* Vol. 11, pp. 289-308.

Trivedi, M. M., I. Mikic, and G. Kogut. Distributed Video Networks for Incident Detection and Management. *Proc. 2000 IEEE Intelligent Transportation Systems*, Detroit, Michigan, 2000.

Van Arem, B., Kirby, H. R., Van Der Vlist, M. J. M. and Whittaker, J. C. (1997) Recent Advances and Applications in the Field of Short-Term Traffic Forecasting. *International Journal of Forecasting*, Vol.13, pp.1-12.

Van der Voort, M., Dougherty, M. and Watson, S. (1996) Combining Kohonen Maps with ARIMA Time Series Models to Forecast Traffic Flow. *Transportation Research Part C: Emerging Technologies*, Vol. 4(5), pp. 307–318.

Venkatanarayana, R., Smith, B. L. and Demetsky, M. J. (2006) Traffic Pattern Identification using Wavelet Transforms. *84th Annual Meeting of Transportation Research Board (CD-ROM), TRB, Washington, D. C.*

Vlahogianni, E. I., Golias, J. C. and Karlaftis, M. G. (2004) Short-Term Forecasting: Overview of Objectives and Methods. *Transport Reviews*, Vol. 24 (5), pp. 533-557.

Vlahogianni, E. I., Karlaftis, M. G. and Golias, J. C. (2005) Optimized and Meta-Optimized Neural Networks for Short-Term Traffic Flow Prediction: A Genetic Approach, *Transportation Research Part C: Emerging Technologies*, Vol. 13(2), pp. 211–234.

Vythoulkas, P. C. (1993) Alternative Approaches to Short-Term Traffic Forecasting for Use in Driver Information Systems. *Transportation and Traffic Theory, Proceedings of the 12th International Symposium on Traffic Flow Theory and Transportation.*

West, M. and Harrison, P. J. (1997) *Bayesian forecasting and dynamic models*. Srpinger, New York.

Whitson, R. H., Burr, J. H., Drew, D. R. and McCasland, W. R. (1969) Real-Time Evaluation of Freeway Quality of Traffic Service. *Highway Research Record*, Vol. 289, pp. 38–50.

Whittaker, J., Garside, S., and Lindveldm, K. (1997) Tracking And Predicting A Network Traffic Process. *International Journal of Forecasting*, Vol. 13, pp. 51-61.

Williams, B. M., Durvasula, P. K. and Brown, D. E. (1998) Urban Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models. *Transportation Research Record: Journal of the Transportation Research Board, No. 1644*, pp.132–144.

Williams, B. M. (1999) Modelling and Forecasting Vehicular Traffic Flow as a Seasonal Stochastic Time Series Process. *Ph. D. Dissertation.* University of Virginia, Charlottesville.

Williams, B. M. and Hoel, L. A. (2003) Modelling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. *Journal of Transportation Engineering*, ASCE, Vol. 129(6), pp. 664-672.

Wu, J. H., Chen, Y. and Florian, M. (1998) The Continuous Dynamic Network Loading Problem: A Mathematical Formulation And Solution Method. *Transportation Research B*, Vol. 32,173-187.

Xu, Y. W., Wu, J. H., Florian, M., Marcotte, P. and Zhu, D. L. (1999) Advances in the Continuous Dynamic Network Loading Problem. *Transportation Science*, Vol. 33, pp. 341-353.

Yang, H., Akiyama, T. and Sasaki, T. (1998) Estimation of Time-Varying Origin-Destination Flows from Traffic Counts: A Neural Network Approach. *Mathematical and Computer Modelling*, Vol. 27 (9-11), pp. 323-334.

Yin, H. B, Wong, S. C., Xu, J. M. and Wong, C. K. (2002) Urban Traffic Flow Prediction Using a Fuzzy-Neural Approach. *Transportation Research Part C: Emerging Technologies*, Vol. 10, pp. 85-98.

Zhang, C., Sun, S. and Yu, G. (2004) A Bayesian Network Approach to Time Series Forecasting of Short-Term Traffic Flows. *Proceedings, the 7th IEEE International Conference on Intelligent Transportation Systems* (*ITSC2004*), Washington, D.C., pp. 216-221.

Zhang, G., Patuwo, B. E., Hu, M. Y. (1998). Forecasting with Artificial Neural Networks: the State of Art. *International Journal of Forecasting*, Vol. 14, pp. 35–62.

APPENDIX A

MSE

Mean square error (MSE) is the expected value of the square of the error or the difference between the observed value and the model output. MSE is calculated using the following equation:

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - x_t)^2$$

where, y_t and x_t are the simulated and real values of a statistical process at the time instant t.

RMSE

Root mean square error (RMSE) is the square root of MSE. It measures the magnitude of the error.

$$\text{RMSE}=\sqrt{\text{MSE}}$$

MAPE

Mean absolute percentage error (MAPE) is a relative measure of the absolute prediction or simulation error.

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\left(y_t - x_t\right)\Big/ x_t\right|\%$$

ACF

Autocorrelation is the correlation between values of the same time-series at different time periods. For a given time-series $\{x_1, x_2 \ldots x_n\}$, the sample ACF at lag j is,

$$r_j = \frac{\sum_{t=j+1}^{n}\left[\left(x_t - \bar{x}\right)\left(x_{t-j} - \bar{x}\right)\right]}{\sum_{t=j+1}^{n}\left(x_t - \bar{x}\right)^2}$$

where, $\bar{x} = \frac{\sum_{t=1}^{n} x_t}{n}$, is the sample mean.

The pattern of autocorrelation for different values of lags is known as the autocorrelation function or ACF. A plot of the ACF against the lag is known as the correlogram (Makridakis et al. 1998).

PACF

Partial autocorrelation is a measure of the correlation used to identify the extent of relationship between the current values of a variable with the earlier values of that (lagged values) while holding the effects of all other time lags constant (Makridakis et al. 1998).

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots + \alpha_j x_{t-j}$$

The PACF of the order j is denoted by the constant α_j which can be calculated by regressing x_t against $x_{t-1,}\ldots x_{t-j}$ from the regression equation described above.