Bayesian dynamic graphical models for high-dimensional flow forecasting in road traffic networks

Thesis

THE OPEN UNIVERSITY

Bayesian dynamic graphical models for high-dimensional flow forecasting in road traffic networks

by

Osvaldo Anacleto Junior, BSc, MSc

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

in the

Department of Mathematics and Statistics

The Open University

November 2012

Date of Submission: 19 October 2012

Date of Award: 5 December 2012

BRITISH LIBRARY

IMAGING SERVICES NORTH

Boston Spa, Wetherby
West Yorkshire, LS23 7BQ
www.bl.uk

THE FOLLOWING ITEMS HAVE BEEN EXCLUDED UNDER INSTRUCTION FROM THE UNIVERSITY

FIGURES

2.1 page 9

4.7 page 73

Abstract

Congestion on roads is a crucial problem which affects our lives in many ways. As a consequence, there is a strong effort to improve road networks in order to keep the traffic flowing. Flow forecasting models based on the large amount of traffic data, which are now available, can be a very useful tool to support decisions and actions when managing traffic networks. Although many forecasting models have been developed to this end, very few of them capture important features of high-dimensional traffic data and, moreover, operating most of these models is a hard task when considering on-line traffic management environments.

Dynamic graphical models can be a suitable choice to address the challenge of forecasting high-dimensional traffic flows in real-time. These models represent network flows by a graph, which not only is a useful pictorial representation of multivariate time series of traffic flow data, but it also ensures that model computation is always simple, even for very complex road networks. One example of such a model is the multiregression dynamic model (MDM).

This thesis focuses on the development of two classes of dynamic graphical models to forecast traffic flows. Firstly, the linear multiregression dynamic model (LMDM), which is an MDM particular case, is extended to allow important traffic characteristics in its structure, such as the heterocedasticity of daily traffic flows, measurement errors due to malfunctions in data collection devices, and the use of extra traffic variables as predictors to forecast flows. Due to its graphical structure, the MDM assumes independence of flows at the entrances of a road network. This thesis therefore introduces a new class of dynamic graphical models where the correlation across road network entrances is accommodated, resulting in better forecasts when compared to the LMDM.

All the methodology proposed in this thesis is illustrated using data collected at the intersection of three busy motorways near Manchester, UK.

Declaration of Authorship

Some of the work presented in this thesis has already been accepted for publication and was collaboratively done as follows.

Part of Chapter 4 and all the developments described in Chapter 5 have been accepted for publication under the title "Multivariate forecasting of road traffic flows in the presence of heteroscedasticity and measurement errors" and will appear in the *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, volume 2, in March, 2013. This paper was co-authored with Catriona Queen and Casper Albers, with me as the first author. While Catriona developed the first theoretical ideas for this paper and Casper did the preliminary analysis of the data, I developed further the theoretical ideas based on an extensive analysis of available data, also carried out by me. I also did all the practical work to implement the models.

The material in Chapter 6 has been accepted for publication in the *Australian and New Zealand Journal of Statistics* under the title "Forecasting multivariate road traffic flows using Bayesian dynamic graphical models, splines and other traffic variables". This paper was also co-authored with Catriona Queen and Casper Albers with me as the first author. Casper proposed the use of traffic variables in flow forecasting models using splines, and also carried out the preliminary data analysis. Based on these developments, I extended the data analysis of all the traffic variables considered in this work, proposed a different method to use traffic predictors in forecasting flow models still using splines, and I also extended the use of splines in the context of dynamic graphical models to parsimoniously accommodate high-frequency flow data and to remove previous unrealistic assumptions assumed by the model under study. Catriona supported the theoretical developments for this work and, again, I did all the practical work to implement the models.

A third paper co-authored with Catriona Queen, with me as the first author, is currently being finished, based on all the research presented in Chapter 7. Here, I originally proposed all the theory which included the definition of a new dynamic graphical model and the proof of its associated results. Catriona played a supporting role in these theoretical developments and gave several suggestions to refine and present them in a clearer manner. All the practical work involved, including data analysis, was also done by me.

Contents

List of Figures

List of Tables

Acknowledgements

I would firstly like to thank my supervisor, Dr. Catriona Queen. Her invaluable help, guidance, encouragement and several suggestions were crucial to write this thesis and to my development as a statistical researcher. Catriona not only was an excellent supervisor, but she also cared a lot about my wellbeing during these past years, for which I am very grateful. I will also always remember her numerous comments on my drafts when trying to clearly and accurately write from now on.

During my period as a PhD student, I had the privilege to interact with brilliant researchers from The Open University Statistics Group who are, above all, very kind people. I am very grateful for having this opportunity. I would particularly like to thank Prof. Paul Garthwaite for his suggestions, advice and fruitful discussions regarding the methodology I have developed during my research for this thesis. And also to the Faculty of Mathematics and Computing of The Open University for providing me with a PhD scholarship.

Thanks to the Highways Agency for providing the traffic data used in this thesis, and also to Les Lyman from Mott MacDonald for his very important help, patience and time when discussing about analysis of these data.

I strongly believe that one of the best things that happened to my life was to meet people who I can genuinely call friends. Some of them had a profound influence on me while I was doing the research to make this thesis possible, especially when I was facing personal problems which directly affected my academic activities. Here I mention my best friend André, who was a constant company during these past three years with his daily e-mails and frequent Skype calls. Many thanks also to my great friends Ivan and Ana for all the (long) conversations and for being the perfect company in all trips and concerts we have been together.

I must also mention some of the people who made my time in Milton Keynes much better: thank you Bethany, Dave, Andy, Mariano, Pete and Jay. And I am very grateful to my sister Fabiana, my brother-in-law Adilson and my nephews Diego and David for making me happy even being miles away from me.

Finally, I would like to thank Anna for her constant support, comprehension and patience, particularly when I ended up working most of the time while visiting her. And I extend my deepest gratitude to my mother, whom I will always be indebted. She has made unmeasurable efforts to provide the best she could give to me and my sister, and only the three of us know how extreme some of the circumstances she had to face were to do so. Her constant encouragement and comprehension not only helped making this thesis become a reality, but also gave me a strong support to achieve all my goals since my childhood. Mãe, muito obrigado por tudo!

Para minha mãe

"There will come a time in your life

when you will ask yourself a series of questions

am I happy with who I am?

am I happy with the people around me?

am I happy with what I'm doing?

am I happy with the way my life is going?

do I have a life or am I just living?

do not let these questions restrain or trouble you

just point yourself in the direction of your dreams

find your strength in the sound

and make your transition"

(Underground Resistance)

Chapter 1

Introduction

This thesis focuses on the problem of developing statistical models for road traffic flow forecasting. Congestion on roads has become a crucial problem of vital importance during the last decades, since its consequences can affect not only daily users of motorways or urban arterials, but also the environment, public health and the economy, for example. As a result, a strong effort has been made by government agencies to maximize efficiency of road networks to prevent congestion. These developments can be put into practise by defining decision rules or a set of actions to be taken given traffic conditions, and they form what is usually called *active traffic management systems*.

One important step for the development of these systems in England was the implementation of induction loops in some motorways to collect traffic information, resulting in a huge amount of data which can be updated on a minute-by-minute basis. The analysis of such data is crucial input to improve active traffic management systems, because it can give insights about current traffic conditions, therefore improving decisions and actions to improve road efficiency.

Since traffic information is generated over time, these data form a time series with possibly very high dimension. Therefore, short-term forecasting models can be useful to describe future traffic conditions. Additionally, since updated traffic conditions are obtained as data arrive, an *on-line* traffic management environment is required. In this context, traffic forecasting models must be also able to provide *real-time forecasts.*

This thesis will use what are known as dynamic graphical models to forecast traffic flows. These models represent the flows in the network by a graph. This graph is not only a useful pictorial representation of the network, but it also ensures that model computation is always simple, even for very complex road networks. Although a dynamic graphical model has already been shown to be extremely promising for short-term forecasting in a UK network, there is still work to be done in order to improve its forecast accuracy when considering real-time traffic data.

1.1 Research question

Traffic data have some characteristics that can be quite challenging to deal with from a statistical modelling perspective. To have a broad view of the traffic network under analysis, data have to be collected from a series of sites, which can generate a high-dimensional time series.

It is also important that a model captures some complex features of road networks. Given a network topology, traffic flows have a dynamic pattern based on possible driver routes, which defines a dependence structure between the time series: this can heavily affect data analysis and forecasts. In addition, events such as adverse traffic conditions or road accidents can cause sudden changes in traffic flows. There may also be changes in a network due to the development of new motorways or some temporary or permanent road blockages. At the same time, as decisions have

to be taken in real time, especially during busy periods, active traffic management systems require forecasts in an on-line environment.

In view of this, one question that arises is: can a statistical model be capable of generating accurate forecasts given the complexities from this field, and at the same time be simple enough in order to meet the requirements of an active traffic management system? Although it is well known that the model-building process is much more difficult for multivariate time series than univariate time series (Chatfield, 2003, West and Harrison, 1997), there are some alternative modelling approaches based on graphical representations of the time series that can address this question.

1.2 Thesis outline

This thesis begins with a description of an active traffic management system which has been implemented in England, as well as a description of the data used for the development of the proposed models here. These are followed by an analysis of these data, which come from a road intersection in Manchester, UK.

A critical review of the models that have been used so far to forecast traffic flows is presented in Chapter 3. This chapter also introduces the dynamic linear model, which has a crucial role in the models to be subsequently developed. As will be also discussed in Chapter 3, very few flow forecasting models take into account the multivariate nature of the data. In this context, the multiregression dynamic model (MDM), which uses a graph to represent multivariate time series, is presented in Chapter 4. A particular class of MDMs, namely the linear multiregression dynamic model (LMDM), has been applied to forecast multivariate flow time series. The LMDM is also presented in that chapter, followed by a procedure to elicit a graph for the traffic network from which the data are collected.

Both Chapters 5 and 6 show LMDM extensions considering traffic data features not previously considered in this model, and also taking into account the requirements of active traffic management systems. As will be seen in Chapter 2, it is not reasonable to assume that flow variability is constant over time, as has been assumed so far when using LMDMs to forecast traffic flows. Chapter 5 therefore shows how to deal with flow heterocedasticity in the LMDM. This chapter also describes how measurement errors, due to malfunctions in data collection devices, can be incorporated into this model.

Data currently being collected in English roads are not only flows, but also other traffic variables. There are currently few models which consider these extra traffic variables as predictors in forecasting flow models. Chapter 6 shows how these predictor variables can be easily included into the structure of the LMDM, resulting in more accurate flow forecasts. Since time series of flows can be built considering different time intervals for data aggregation, Chapter 6 discusses how different data aggregations show different traffic features. Additionally, Chapter 6 shows how the LMDM can accommodate traffic flows aggregated into time intervals suitable for real-time traffic forecasting.

When using LMDMs to forecast traffic flows, marginal independence is assumed among any time series that represent the entrances of the road network under study. However, this assumption may not be reasonable for traffic data. For example, all flows will increase during the build up of traffic in the morning rush hours, and, as another example, adverse weather conditions can affect the road network entrances equally. Motivated by the need of relaxing this restriction, Chapter 7 introduces a new class of dynamic graphical models where the data dependence structure among road network entrances is accommodated, thus resulting in better forecasts when compared to the LMDM.

Finally, possible future directions to be followed from the proposed statistical developments in this thesis are presented in Chapter 8.

Chapter 2

The Manchester network

2.1 Introduction

The traffic data which will be used to develop the models described in this thesis are introduced in this chapter, as well as the road intersection where they are collected. Time series of these traffic data and the relationships between flow and other available traffic variables are also analysed. We firstly begin with a description of an active traffic management system currently operating in England.

2.2 The Managed Motorways project

As described in Chapter 1, an active traffic management system comprises a set of decision rules or actions to be taken in order to keep the traffic flowing and, especially, to avoid congestion. An example of such system is the *Managed Motorways* project developed by the Highways Agency, which is responsible for managing and improving road networks in England. An overview of this project can be found in Highways Agency (2012b).

Managed Motorways aims to improve road capacity by controlling the flow using hard shoulders as additional lanes during congestion periods. The use of hard shoulders to reduce levels of congestion has been shown to be an useful action in some countries where it was adopted (Sultan *et al.*, 2008). It is also a cheaper and quicker alternative when compared to widening roads for example. Other actions considered in the Managed Motorways project to improve road efficiency relate to operating mandatory speed limits to controlling flow while traffic is in the network.

These actions depend on alerts which are triggered when certain values of traffic variables are exceeded. Particularly, mandatory speed limits are triggered when flow reaches specified (high) values. The data being used in this project are collected by induction loops which were installed in some roads in England as part of the *Motorway Incident Detection and Automatic Signalling* (MIDAS) system (Gibbens and Werft, 2005). These induction loops are installed at each lane of a road site, and they collect the following variables on a minute-by-minute basis:

- Flow: the number of vehicles passing over the induction loop per minute;
- Occupancy: the percentage of time per minute that vehicles are 'occupying' the inductive loop;
- Headway: the average time (per minute) between vehicles passing over the induction loop (in sec);
- (Time mean) speed: the average ratio of the distance between two (consecutive) induction loops in a road segment and the time taken by each vehicle to pass over these loops (in kph).

Although these variables are available on a minute-by-minute basis, data aggregation considering other time intervals may be more suitable, since different traffic features can be observed depending on the aggregation level considered. For the development

of flow forecasting models in this thesis, traffic variables will be averaged into 5-minute and 15-minute intervals. Traffic data aggregation will be further discussed in Chapter 6.

The Managed Motorways project began operating on the M42 motorway in Birmingham in 2006, and it is also being implemented in other motorways. See Highways Agency (2012a) for an implementation plan of this project in English road networks.

2.3 The Manchester network

Traffic data collected by induction loops are available for the M60/M62/M602 intersection in Manchester, hereafter called the *Manchester network*, where flow, occupancy, speed and headway information are collected at 32 sites. Figure 2.1 shows an aerial photograph of the network, and a schematic diagram of the Manchester network reflecting the layout of the data sites is given in Figure 2.2. In this diagram, the arrows show the direction of travel, and the data sites are labelled and indicated by yellow circles. This is one of the networks where Managed Motorways is being implemented (Highways Agency, 2012a).

Given the locations of the data collection sites, it is possible to have a description of how traffic flows through the network. One crucial characteristic to be considered when analysing data from this road intersection is that it takes only a few minutes to traverse the Manchester network. This means that vehicles can be counted at a number of different sites during the *same* time period, depending on the time interval used to aggregate the traffic variables. Information from different parts of the network can therefore be used as potential predictors when developing traffic models for a specific site.

FIGURE 2.1: Aerial photograph of the Manchester network (©2012 DigitalGlobe, GeoEye, Infoterra Ltd & Bluesky, The GeoInformation Group, Map data ©2012 Google)

2.4 Visualising the Manchester network time series

To visualise some traffic patterns which can be observed at the Manchester network, Figures 2.3 and 2.4 show time series plots of flow, occupancy, speed and headway at site 9200B collected between May 10^{th} and May 16^{th}, 2010. Data were aggregated into 5-min intervals to plot these series.

Figure 2.3(a) shows a time series plot of 9200B flows. It is clearly seen in this plot that flow patterns during the weekdays are different from flow patterns during weekends: while there are morning and afternoon peak periods around 07:00-08:00 and 16:00-17:00, respectively, on weekdays, a single peak period around 12:00 was observed on both Saturday and Sunday. Some flow outliers, probably caused by

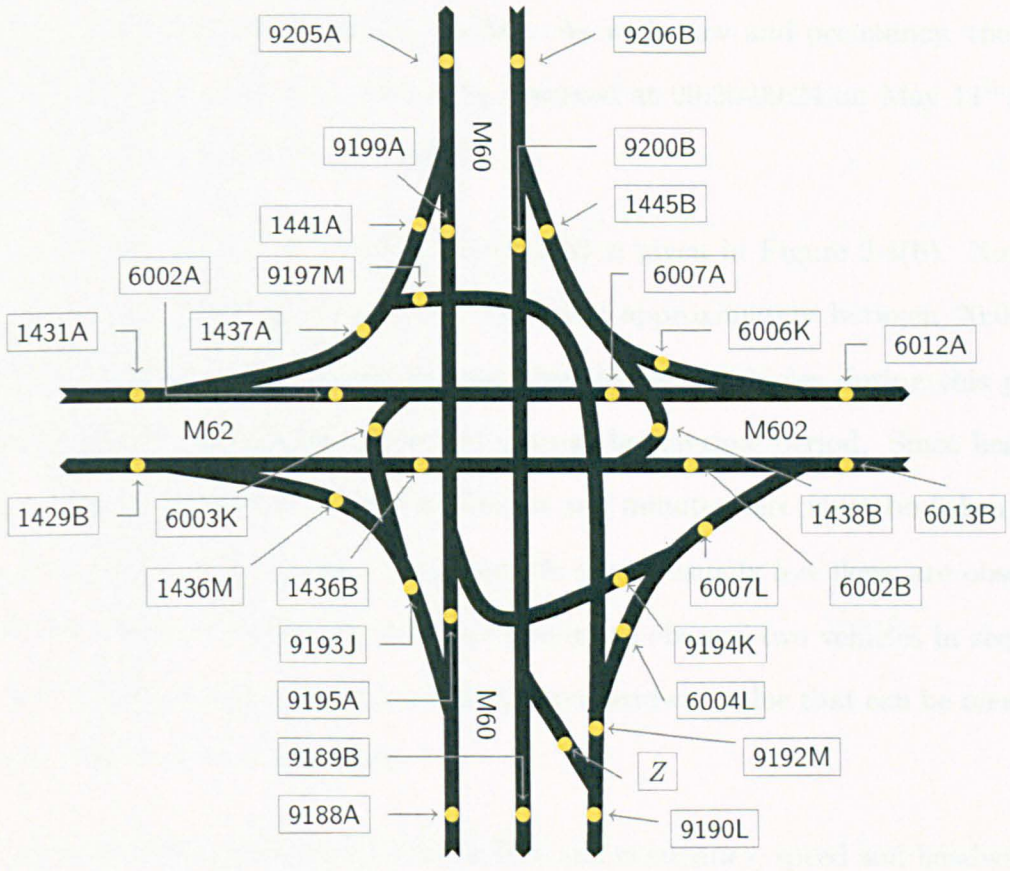

FIGURE 2.2: Schematic diagram of the Manchester network

congestion, can be noted between aproximately 08:30-09:30 on May 11th, 2010 and between 08:00-09:00 on May 14th, 2010, for example.

A similar pattern to that observed for flows is also seen when looking at occupancies from the same period, presented in Figure 2.3(b), where occupancy peaks are also around 07:00-08:00 and 16:00-17:00 on weekdays and weekend occupancy peaks are around 12:00. In particular, the 5-min intervals where the highest occupancy peaks were observed during this week (09:20-09:24 on May 11th, 2010 and 08:55-08:59 on May 14th, 2010) are the same periods where some of the highest flow peaks were observed in Figure 2.3(a).

Figure 2.4(a) shows 5-min speeds at site 9200B collected between May 10th and May 16th, 2010. There are apparently two speed regimes observed during weekdays, in

the sense that speeds approximately between 07:00 and 18:00 are lower than speeds observed at other periods during the day. As with flow and occupancy, the most extreme measurements that week were observed at 09:20-09:24 on May 11th, 2010 and at 08:55-08:59 on May 14th, 2010.

The headway plot for this week at site 9200B is given in Figure 2.4(b). Note the high level and also a high headway variability approximately between 20:00 and 07:00, which indicates a higher average time between vehicles during this period when compared to headways observed during the daytime period. Since headway is based on average times between vehicles per minute, care must be taken when analysing this traffic variable during periods where usually low flows are observed, as is quite likely to take more than one minute to observe two vehicles in sequence during low flow periods. Indeed, the maximum headway value that can be measured by loop detectors is 25.4 seconds.

An analysis of the relationship between flow and occupancy, speed and headway will be presented in Section 2.5.

2.4.1 Flow variability in the Manchester network

Figure 2.5 shows hourly box-plots of 15-min flows for each weekday from 07:00 to 18:59 at site 6013B of the Manchester network, using data observed from March to November 2010. These clearly show daily differences in level and variability of flows. There is a particularly high flow level observed between 12:00 and 14:59 during Fridays when compared to the same period during other weekdays. On the other hand, there is a lower flow level between 17:00 and 18:59 on Fridays in contrast to the same period during the other weekdays at site 6013B. It can be an indication that people usually leave work earlier on Fridays.

Differences in flow variability between weekdays during the same time period are also observed when looking at these boxplots. As an example, the flow variability observed on Fridays is lower than the flow variability on Mondays during 18:00-18:59 at site 6013B. Additionally, flow can also vary depending on the time of the day, as flow variability observed during 07:00-08:59 is higher than the flow variability observed between 09:00-13:59, for instance.

In Chapter 5, the variability of flows within days will be accommodated in a multivariate flow forecasting model.

2.5 Relationships between flows and other traffic variables

As will be discussed in Chapter 6, there is currently few flow forecasting models which consider extra traffic information in the form of predictor variables. Hence, it is important to analyse the relationship between flows and extra available variables to have some insights into how to include them as predictors when modelling flows.

The first row of Figure 2.6 shows scatterplots of flow at time t versus occupancy at previous time $t-1$ at site 9188A for February, June and October 2010. The plots indicate an increasing relationship between flow and occupancy until the latter reaches some value around 20, which is usually defined by traffic managers as the road capacity and varies from site to site. For occupancy values higher than this road capacity, the relationship then turns to be decreasing, which can lead to congestion. This relationship is similar to what is called *the fundamental diagram of traffic* (Ashton, 1966, Kühne, 2008), which will be discussed in Subsection 3.2.2 of Chapter 3.

Scatterplots of flow at time t versus headway at previous time $t-1$ at the same site for the same three months are shown in the second row of Figure 2.6. These plots confirm an intuitive relationship in the sense that the flow decreases as the average time between cars increases.

The last row of Figure 2.6 shows scatterplots of flow at time t versus speed at previous time $t-1$, again at site 9188A for the same three months. Most flow values are concentrated at speed values between 80 kph and 100 kph, with an apparently decreasing relationship in this region. There also seems to be an increasing relationship between flow at t and speed at $t-1$ for low speed values, although with a slightly higher level of variability: it is likely that many of these points are from situations where congestion occurred.

Plots of flow at t versus the other traffic variables at $t-1$ look broadly comparable at the other sites. Based in these observed relationships, it will be described in Chapter 6 how these extra traffic variables can be included as predictors in a multivariate flow forecasting model.

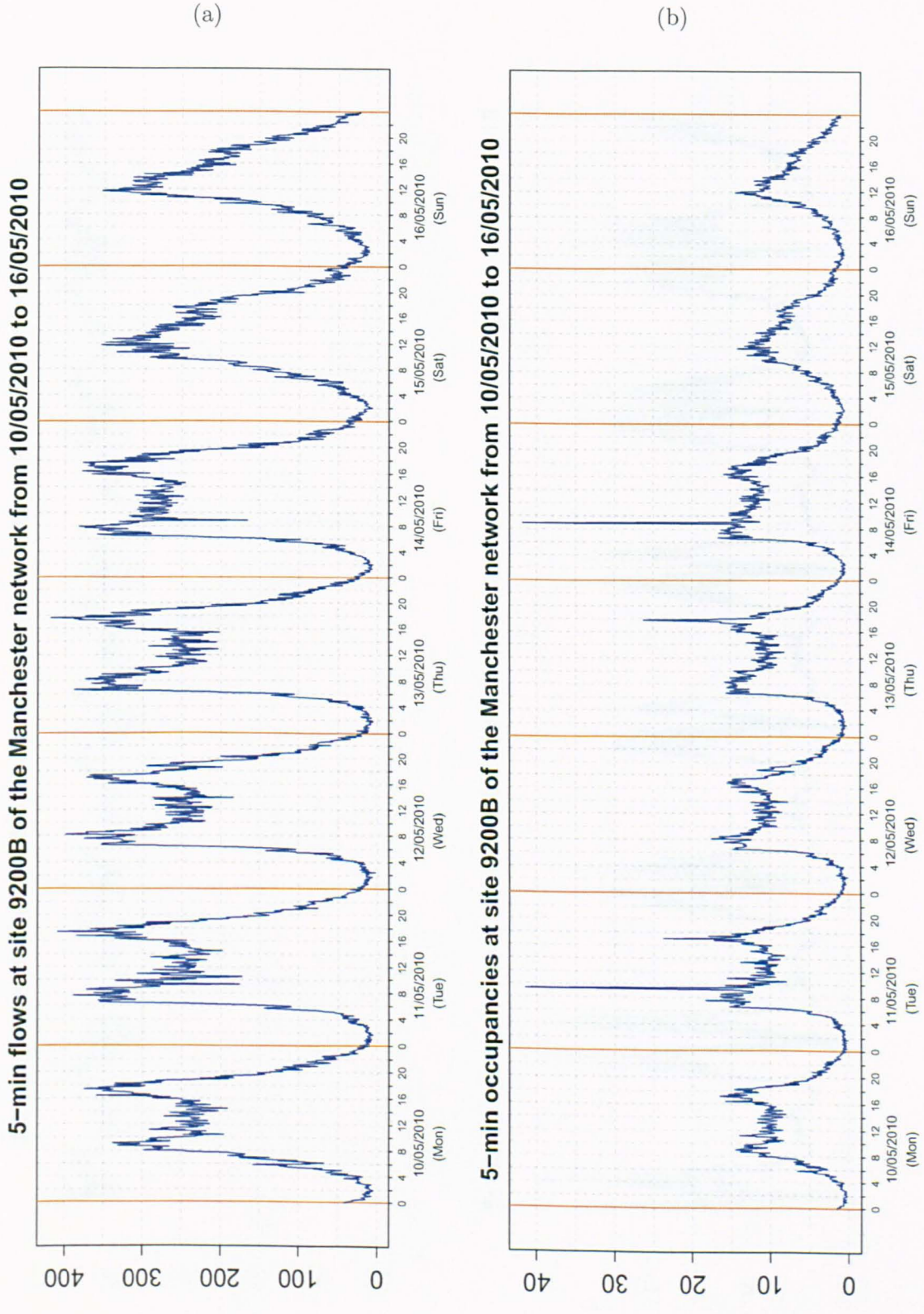

FIGURE 2.3: 5-min flows (a) and occupancies (b) at site 9200B for May 10^{th}-16^{th}, 2010.

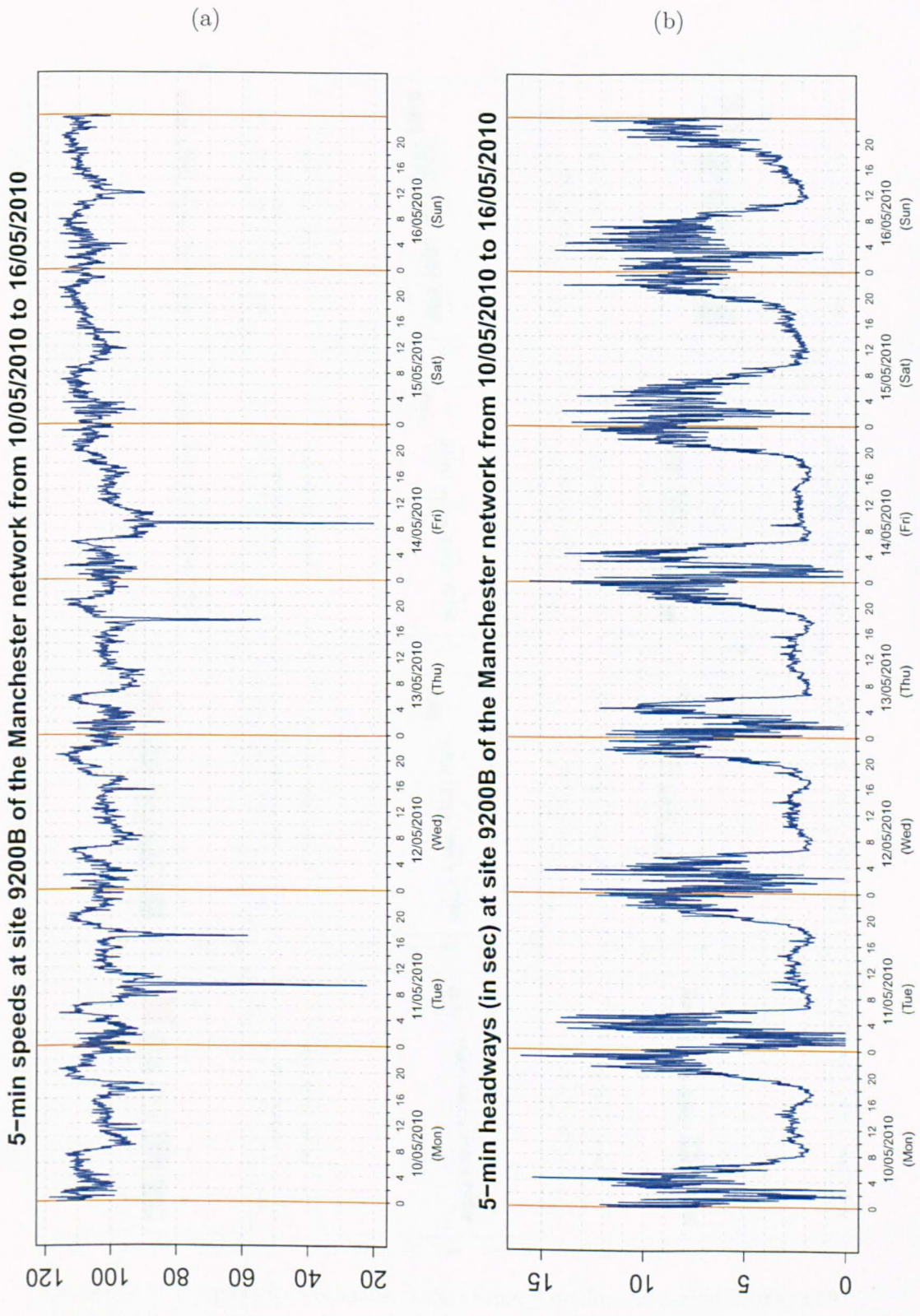

FIGURE 2.4: 5-min speeds (a) and headways (b) at site 9200B for May 10^{th}-16^{th}, 2010.

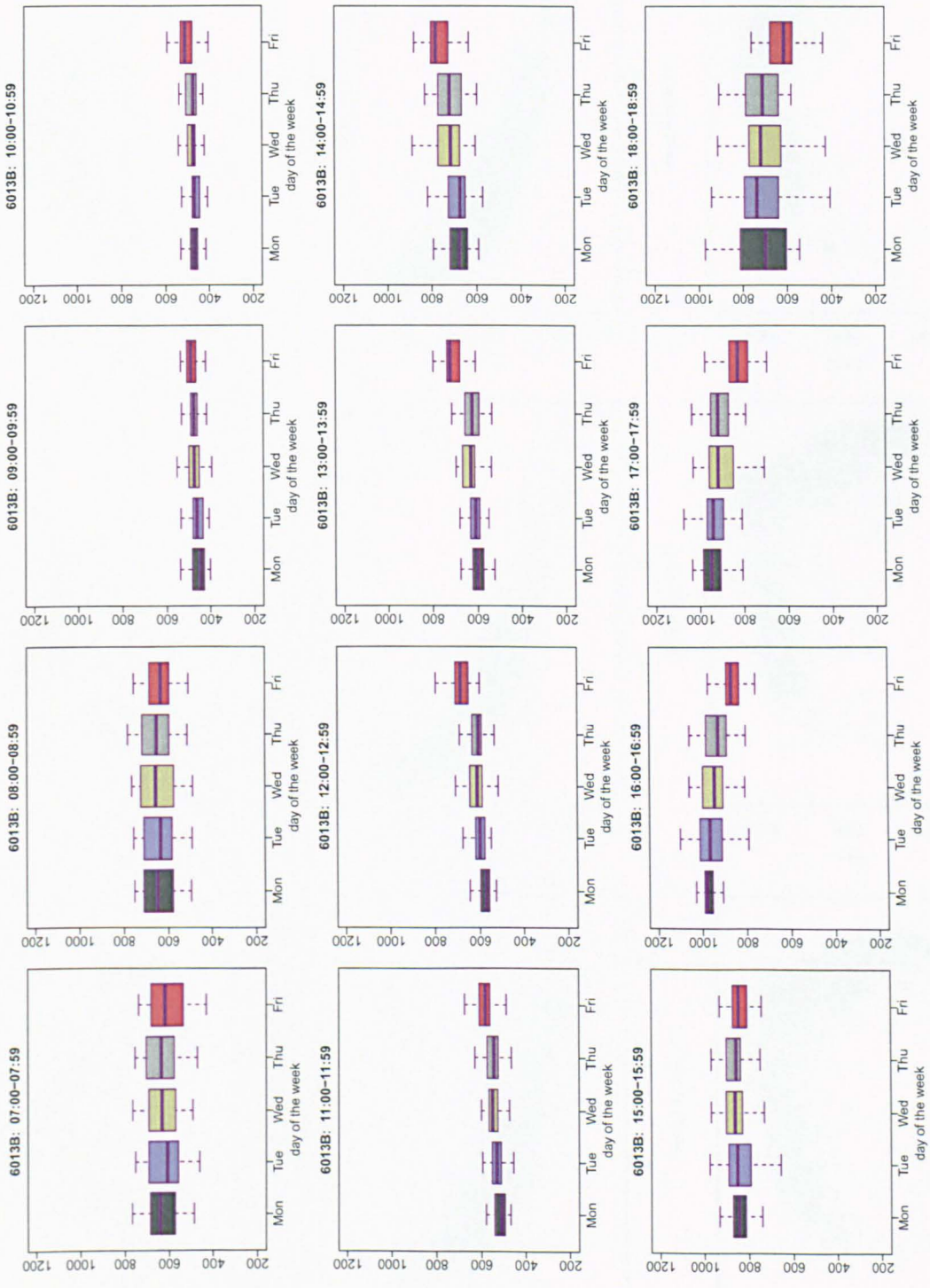

FIGURE 2.5: Boxplots by weekdays using 15-min flows for the period 07:00-18:59 (one boxplot for each hour) at site 6013B observed from March to November 2010.

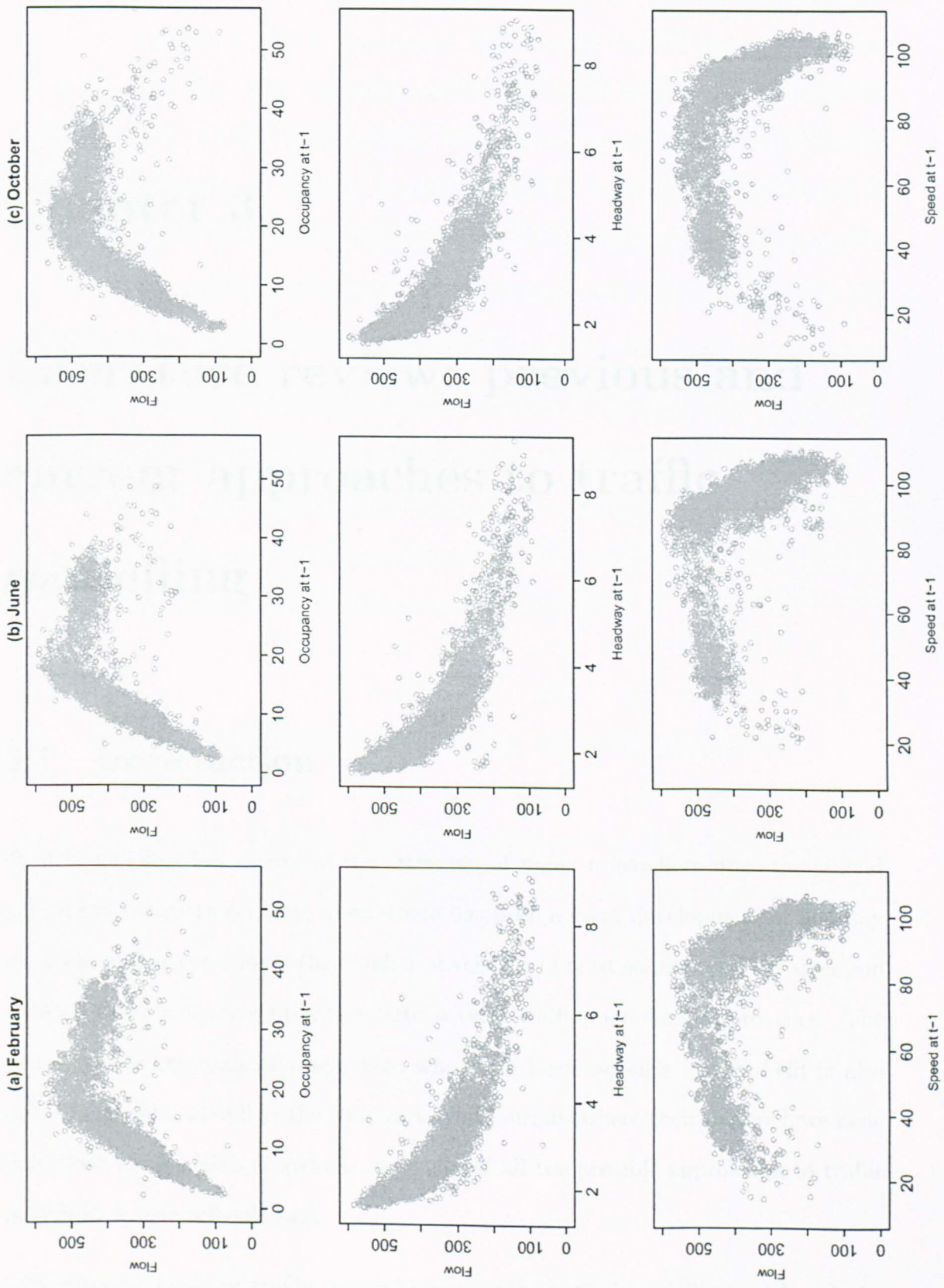

FIGURE 2.6: Scatterplots of 5-min flows at site 9188A at time t versus occupancy, headway and speed at $t-1$, in (a) February, (b) June and (c) October 2010.

Chapter 3

Literature review: previous and current approaches to traffic modelling

3.1 Introduction

Traffic modelling has attracted the attention of many researchers since the second half of the twentieth century, where there has been a great development of highway networks and an increase in the number of vehicles in most countries. This attention resulted in an extensive literature with several traffic modelling approaches. The variety of backgrounds of researchers who have been working in this field is also quite noticeable, as well as the wide variety of journals where their papers have been published. This makes an exhaustive study of all the possible approaches to traffic modelling a very difficult task.

Deterministic views of traffic dynamics use mathematical modelling, and such developments have provided important concepts about traffic behaviour. The first

statistical applications in the field were based on ARIMA models, which are established forecasting tools for time series. Multivariate ARIMA versions, namely VARMA models, have been recently considered for traffic forecasting. Due to their flexibility and the variety of commercial software available, neural network models have also been extensively applied in traffic modelling. Additionally, applications of state space models have been considered in this field, since their structure permits a sequential estimation of parameters for forecasting traffic flows. These are the approaches to be discussed in the next sections.

Some reviews covering the broad extent of mathematical traffic flow models can be found in Bellomo *et al.* (2002), Hoogendoorn and Bovy (2001) and Gartner *et al.* (2001). Van Arem *et al.* (1997), Vlahogianni *et al.* (2004) and Karlaftis and Vlahogianni (2011) review neural network techniques, statistical approaches and also some hybrid alternatives for short-term traffic forecasting.

3.2 Mathematical modelling of traffic flows

The initial approaches to model traffic had a mathematical basis, where deterministic relationships between traffic parameters in space and time were considered. Mathematical traffic flow models can be classified according to the level of detail defined to describe traffic behaviour. They are then basically divided into *microscopic* and *macroscopic* models. There is also a third type of models where analogies between flow of vehicles in a road and the flow of molecules in a gas are considered (Ashton, 1966).

3.2.1 Microscopic models

Microscopic models are used to have a description of the dynamics of individual drivers, taking into account the influence of surrounding vehicles (Bellomo *et al.*, 2002). Their main idea is to establish a mechanism describing aspects of the interaction between vehicles through mathematical formulations.

Most early attempts to develop microscopic models form what are called *car-following models*, which are based on descriptions of how one vehicle follows another (Chandler *et al.*, 1958). Given a sequence of vehicles following each other, the basic principle used in these models considers the response, which is the act of accelerating or braking related to the following vehicle, as a product of what are known as sensitivity and stimulus. Stimulus is defined as the speed difference between the leader and the follower in the sequence. Sensitivity can be viewed as a measure of interaction between vehicles, usually defined to be inversely proportional to the space between them (Edie, 1961, Herman and Potts, 1961). Mathematical expressions of these quantities can be found in Chandler *et al.* (1958) and Gazis *et al.* (1961). Some alternative principles and approaches to define interaction between individual vehicles and other types of microscopic models can be found in Hoogendoorn and Bovy (2001).

Car-following principles have been heavily applied in the development of simulation tools to have a computer-based view of traffic behaviour, resulting in several *microscopic simulation models*. Based on some initial conditions about driver behaviour and vehicle characteristics, simulation methods are used to determine how the traffic system evolves over time, given possible decisions each driver can make. A review of microscopic simulation models was made in the SMARTEST project (Algers *et al.*, 1997), where the main objective was to develop this class of models to solve specific traffic management problems.

Although these models can be quite useful to understand traffic behaviour, they demand a high level of detail about traffic data. Alternatively, traffic information on an aggregated level can be used to generate microscopic data (Hoogendoorn and Bovy, 2001), and some effort has been made to build suitable datasets for the development of such models (Wilson, 2010). However, given the structure of microscopic models, it can be quite difficult to update them in an on-line traffic environment, since the generation of real-time forecasts as new data arrive possibly requires a high number of simulations.

3.2.2 Macroscopic models

Different from car-following approaches, macroscopic models can be used to give a description of traffic on an aggregated level, not taking into account explicit relationships between different vehicles. Their basic principle is that driver behaviour depends on traffic characteristics that can be defined in terms of aggregated variables measured at a road of interest.

The first study concerning traffic flow behaviour adopted a macroscopic approach and was made by Greenshields (1935). In this work, the focus was on determining the road conditions in which congestion or slow traffic are likely to occur. Traffic measurements and relationships developed by Greenshields had a significant impact in the field during the following decades.

The model developed in this first study focused on the relationship between three traffic variables:

- flow (q): number of vehicles per unit time;
- density or concentration (k): number of vehicles per unit distance;
- mean speed (u): distance per unit of time.

The relation

$$q = u \times k. \tag{3.1}$$

can be defined among these variables. Greenshields used a 16 mm camera to measure these quantities in a road by taking pictures at regular time intervals (see Greenshields, 1934, for details). Motivated by his measurements, he realized that a linear relationship of the form

$$u = \alpha - \beta k, \tag{3.2}$$

for some α and β, could be assumed between speed and concentration.

Taking into account that concentration reflects the level of interaction between vehicles in a road, it can be noted that, as this variable (k) tends to zero, the speed (u) tends to reach its maximum. In equation (3.2), α is usually called the *free speed*, which is considered as the maximum speed a driver can reach if there are no interactions between vehicles in a road. On the other hand, the ratio α/β can be considered as the maximum concentration as speed tends to zero. Replacing (3.2) in (3.1) we have,

$$q = \alpha k - \beta k^2, \tag{3.3}$$

which represents a quadratic relationship between flow and concentration (as shown in Figure 3.1) and, as mentioned in Section 2.5 of Chapter 2, is called the fundamental diagram of traffic.

With this simple model and its empirical validation (Greenshields, 1935), it is possible to describe some intuitive traffic flow characteristics. For example, it is reasonable to assume that flow must be zero when concentration is zero, and also concentration reaches its maximum when traffic jams occur. It can be also thought that flow increases up to a given value and, after reaching it, interactions between vehicles are so high that flow tends to decrease, leading to a traffic jam. This value

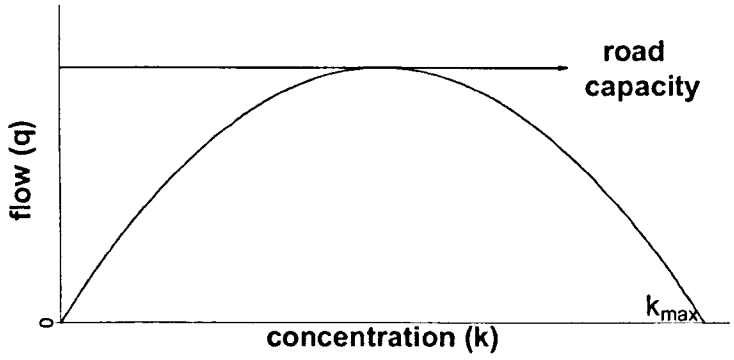

FIGURE 3.1: Fundamental diagram of traffic.

is called the *road capacity* and it is of great interest for traffic engineers in order to design new roads or to monitor traffic.

Although the equation proposed by Greenshields was widely accepted as the proper shape of the flow-concentration curve, some authors pointed out limitations of this model (see, for example, Ashton, 1966). Gartner *et al.* (2001) also noted that Greenshields' measurements were made during a holiday in a single lane road, where the level of interaction between vehicles is lower than levels observed in multi-lane freeways. There are indeed studies which propose different relationship descriptions of flow-concentration, speed-concentration and flow-speed. There has also been some effort to describe these three variables simultaneously (Gilchrist, 1988, Gilchrist and Hall, 1989). Equations describing these relationships are usually called *traffic stream models*. Additionally, there are some models (Edie, 1961, Greenberg, 1959) which consider discontinuites in the flow-concentration relationship, with different regimes during non congestion times (traffic flow values smaller than road capacity) and congestion times (traffic flows bigger than road capacity). It can reflect a limitation found in Ashton (1966) regarding Greenshields' model, where it was observed that the flow-concentration relationship cannot be considered to be the same for all values of k, since a driver probably only takes account of the vehicle ahead when driving in low concentrations, and several factors begin to influence him as concentration

increases.

3.2.2.1 Hydrodynamic approaches to traffic modelling

Macroscopic traffic models can be also developed by considering analogies between vehicular and fluid flows. These analogies motivate the formulation of partial differential equations which, together with their solutions, can describe the dynamics of macroscopic traffic characteristics (such as flow, concentration and speed) thus giving a picture of how traffic behaves over time. These models are usually called *continuum flow models* (Gartner *et al.*, 2001).

The key idea behind continuum flow models is what is called the *principle of flow conservation*, which can be stated as follows. Suppose a road section with two data collection sites, namely $S(1)$ and $S(2)$, such that they are separated by a very small distance Δx, there are not any intersections or junctions between them and the direction of traffic is from $S(1)$ to $S(2)$. The principle of flow conservation states that the increase in the flow at the downstream site $S(2)$ during a very small time Δt must be equal to the decrease of vehicles (concentration) in the section of the road between sites $S(1)$ and $S(2)$. This principle can be mathematically described as

$$\frac{\partial q}{\partial x} = -\frac{\partial k}{\partial t}. \tag{3.4}$$

where $\partial q/\partial x$ represents the increase in flow at $S(2)$ and $-\partial k/\partial t$ represents the decrease of vehicles in the section of the road between sites $S(1)$ and $S(2)$ (for details, see Ashton, 1966). If entrances or exits in a road section are considered, equation (3.4) can be written as (Gartner *et al.*, 2001),

$$\frac{\partial q}{\partial x} + \frac{\partial k}{\partial t} = g(x,t). \tag{3.5}$$

where $g(x,t)$ is the generation (or dissipation rate) of vehicles per time per unit length. Equation (3.5) is usually called *the conservation equation.*

Different solutions exist to the conservation equation, either by using additional equations or by considering the description of other macroscopic variables. Particularly, Payne (1971) considered a partial differential equation describing the dynamic behaviour of speed, and this idea was also used in Helbing (1996).

Lighthill and Whitham (1955) were the first to propose a solution to (3.5), considering a system with the conservation equation, the relationship given in (3.1) and assuming speed as a function of concentration (such as (3.2)). A similar solution was derived by Richards (1956), with this approach being usually called LWR (Lighthill, Whitham and Richards) models. As this class of models does not have a unique solution, some generalized solutions were proposed by Leutzbach (1988), and a description of numerical alternatives can be found in Gartner *et al.* (2001). Stephanopoulos and Michalopoulos (1979, 1981) describe some implementations of LWR models.

Since loop detectors in road networks collect traffic macroscopic data (as seen in Chapter 2), continuum flow models are presumably suitable for traffic forecasting. However, these data are used in continuum flow models to estimate additional equations to solve the conservation equation (3.5). As the resulting system of equations may not have closed-form solutions, it can then be infeasible to use them on a real-time basis. Some other limitations of macroscopic models can be found in Papageorgiou (1998).

3.3 ARIMA models in traffic forecasting

After the seminal work of Yule (1927), which introduced the notion of defining a time series as a realization of a stochastic process, there were several developments to describe dynamic systems using time series models with a random component, rather than considering strictly deterministic mathematical models only. In this context, given a sequence of random variables $\{X_t\}_{t\geq 1}$, which represents a time series of interest, a general class of models consists of describing the random variable at time t as a sum of a linear combination representing a function of previous values of the sequence, together with a linear combination of random variables with zero mean and constant variance, such that

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=0}^{q} \beta_j Z_{t-j} \tag{3.6}$$

where Z_{t-j}, $j = 0, \ldots, q$, are random variables with null mean and constant variance, and $\beta_0 = 1$.

The first sum in (3.6) represents the systematic term of what is called a *autoregressive process* of order p, denoted by AR(p). The idea is that the observed value of the random variable at time t depends on previous observed values of the sequence, and it has the same structure of the systematic term of the classic linear model, despite the fact that the systematic term in a autoregressive process contains realizations of the *same* random variable defined as the response (hence the name autoregressive). The second sum represents a moving average process of order q, denoted by MA(q), and it can be interpreted as being composed of the error term of the model at time t, represented by Z_t, as we also have in the classical linear model, plus the error terms related to differences between predicted and observed values at previous times, represented by Z_{t-j}, $i = 0, \ldots, q$.

The parameters to be estimated in this model from an observed time series are α_i and β_j, $i = 1, \ldots, p$, $j = 1, \ldots, q$. Using this terminology, the model described in (3.6) is usually called an *autoregressive moving average process* of order (p, q), or simply ARMA(p, q).

The key assumption in the estimation of ARMA models is that some statistical properties of the sequence of random variables, which represents the time series under study, must remain the same over time. Particularly, this means that $\{X_t\}_{t \geq 1}$ must satisfy two conditions:

i. $\mathrm{E}[X_t]$ and $\mathrm{Var}[X_t]$ are constant for every $t \geq 0$.

ii. $\mathrm{Cov}[X_t, X_{t-s}]$ depends only on the difference $t-s$, that is, the covariance between two elements of the sequence of random variables depends only on the lag between them, and not on time t.

A time series which satisfies conditions (i) and (ii) above is said to be *weakly stationary* or just *stationary*.

Transformations are usually applied to make a sequence of observed time series stationary, with the differencing technique being the most common one. A differenced time series is defined as the sequence $\{\nabla^d X_t\}_{t \geq d+1}$, where each $\nabla^d X_t$ corresponds to the difference $X_t - X_{t-d}$, with $t \geq d+1$. When (3.6) is applied to a differenced time series, we can write

$$\nabla^d X_t = W_t = \sum_{i=1}^{p} \alpha_i W_{t-i} + \sum_{j=0}^{q} \beta_j Z_{t-j} \tag{3.7}$$

and then we have what is called the *autoregressive integrated moving average process* of order (p, d, q) or simply ARIMA(p, d, q), where d is related to the lag applied in differencing the original time series. The term *integrated* refers to the fact that forecasts based on the transformed series must be summed (or "integrated") to

produce forecasts for the original time series (Chatfield, 2003). Within this approach, the main tasks when fitting ARIMA models for forecasting are determining the terms p, d and q and parameter estimation.

A very influential contribution in this field was made by Box and Jenkins (1970), which defined a set of strategies to determine how to produce forecasts using ARIMA models. A rough summary of these strategies can be basically divided into steps which involve the determination of the differencing order d to make a time series stationary, the identification of the order of the autoregressive and moving average terms of the model (based on graphical techniques and statistical tests), parameter estimation through the use of non-linear least squares or maximum likelihood and, finally, model adequacy. The application of ARIMA models together with these strategies are usually called *Box-Jenkins methodology.*

The first models based on Box-Jenkins methodology for traffic forecasting can be found in Ahmed and Cook (1979) and Levin and Tsao (1980). More recently, Hamed *et al.* (1995) developed univariate models to predict five series of 1-min flows collected at the city of Amman (Jordan), and concluded that an ARIMA$(0,1,1)$ was the best model for all series. Given their model structure, they also mentioned that such a model could be easily implemented, as it has a simple formula and, to estimate the traffic flow at time $t+1$, only the forecast residual and flow observed at time t would be required. Additionally, Williams and Hoel (2003) applied a seasonal version of an ARIMA$(1,0,1)$ (Chatfield, 2003) using weekly differencing for 15-min flows, and found that this model provided better forecasts when compared to random walks and historical average models.

3.4 VARMA models in traffic forecasting

Multivariate versions of ARMA models can be defined by extending the equation (3.6) to the case where interest is on a time series vector $\boldsymbol{X}_t$. These multivariate models, namely *vector* autoregressive moving average (VARMA) models (Chatfield 2003 and Prado and West, 2010) have been recently applied to traffic forecasting. Chandra and Al-Deek (2009) used VAR models (which are particular cases of VARMA models without the moving average term) to forecast flows from three data collection sites from a motorway in Florida (US), and they give better results when compared to ARIMA models independently fitted for each of the sites. However, they point out the issue of having a large number of parameters to be estimated in VAR models, and this problem can be worse in VARMA models or when forecasting flows from a high number of data collection sites. Moreover, a constant variance was assumed in their models to forecast flows from the whole day, which can be a very restrictive assumption (as shown in Subsection 2.4.1 of Chapter 2).

Both ARIMA and VARMA models can be estimated through least squares or maximum likelihood, which can be heavily affected by outliers. Furthermore, these estimation procedures cannot be feasible when applying VARMA models to forecast flows in large networks. The fact that there are equations like (3.6) for each of the components of a time series vector when using a VARMA model results in a large number of parameters to be estimated. As an alternative, Min and Wynter (2011) developed a VARMA model which considers the network structure and average speeds between data collection sites to decrease the number of parameters to be estimated.

Prado and West (2010) describe some Bayesian approaches to parameter estimation in VARMA models, pointing out some advantages of considering this alternative, especially because of the large number of parameters involved. Mai *et al.* (2011)

adopted a Bayesian approach via MCMC to develop VARMA models to forecast 15-min flows from three sites of a busy signalized intersection in Dublin, Ireland. Similar to Min and Wynter (2011), they also considered the structure of the intersection to minimize the number of parameters to be estimated. One of their multivariate models gave a slighty better performance than its related univariate version. The authors point out that the reason for such a result is that their VARMA models take into account the spatial correlation among the data collection sites, which may not have a high influence on forecasts when using 15-min data, since vehicles can travel a considerable distance during 15 minutes. They also expect that their multivariate model would give better results if smaller time intervals for data aggregation were considered. However, the fact that the model uses MCMC simulation for parameter estimation can be a drawback when forecasting flows in real-time.

3.5 Neural networks in traffic modelling

Neural networks are mathematical models which try to replicate the human brain process of recognizing patterns and decision making that constantly appear in our daily lives. These models aim to automate these replications to perform specific tasks, which are usually divided in two types: *unsupervised pattern recognition* problems, where the main objective is to identify groups of similar observations in a data set (like cluster analysis, for example), and *supervised pattern recognition* problems, where the goal is to produce an output of interest given a data set. The latter aims to solve the same type of problems in which regression and time series analysis can be applied.

The idea which forms the basis to develop a neural network (NN) for a supervised pattern recognition task consists of creating a system in which a set of unit inputs (explanatory variables), usually defined as the *input layer*, is connected with the

output (dependent variable), defined as the *output layer*. These layers are connected by a set of *nodes* (also called processing units or neurons), where the mechanism for combining the inputs to produce an output is defined. Usually the nodes are organized in one layer, called *the hidden layer*, but more than one layer is possible.

Consider the following example to understand how a neural network can be used in traffic modelling. Suppose that the flow forecast at time t at a road site of interest, represented by $\hat{X}_t$, must be made given flows at times $t-1$, $t-2$ and $t-3$, namely x_{t-1}, x_{t-2} and x_{t-3}. A typical NN structure to tackle this type of problem is shown in Figure 3.2 (adapted from Chatfield, 2001). In this NN, the lagged flow values x_{t-1}, x_{t-2} and x_{t-3} form the input layer, the hidden layer comprises two nodes (although it can have several nodes) and each input node is connected with the hidden layer nodes. Given this structure, the forecast is made by measuring the "strength" of each connection based on data and then calculating combinations of these connections. Within this framework, the forecasting procedure based on a neural network can be

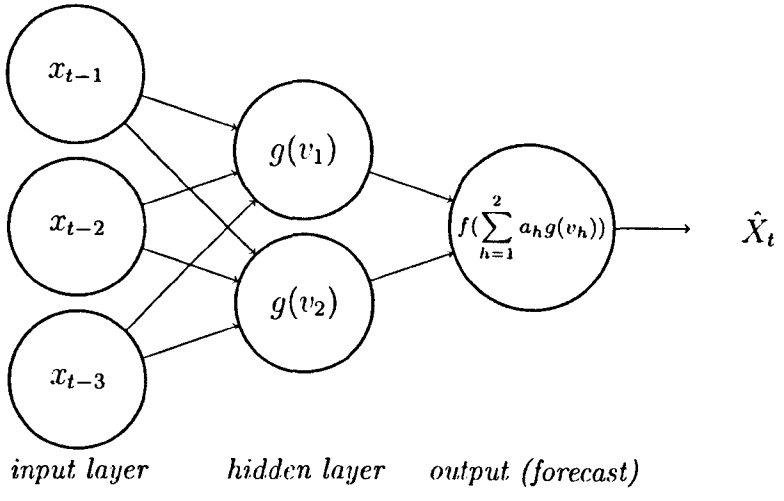

FIGURE 3.2: Neural network example for traffic forecasting

roughly summarized by the following steps. First, an average v_h for each node in the hidden layer is calculated based on a linear combination of the input variables and weights (with these weights being "strength" measures of each connection) which,

again, are defined based on data. That is, $v_h = \sum_{i=1}^{3} w_{ih} x_{t-i}$ is calculated for each node h, $h = 1, 2$, in the hidden layer. Still in the hidden layer, functions are also chosen to be applied to each v_h, usually called *activation functions* and denoted by $g(v_h)$, $h = 1, 2$ in Figure 3.2, with the choice of these functions depending on the output domain. Finally, the flow forecast $\hat{X}_t$ is calculated as a function of a linear combination of the results of these activation functions and weights, a_h, associated with the hidden layer nodes, that is, $\hat{X}_t = f(\sum_{h=1}^{2} a_h g(v_h))$ in our example.

The NN architecture comprises the number of input variables, nodes, hidden layers and also the activation functions. Furthermore, different NN methodologies can vary according to the way the layers are connected. The NN structure described in the example above is usually called a *multilayer perceptron* or *feed-forward network*. In this widely used structure, the units and nodes have only one-way connections in the direction from the input layer to the output layer, and these connections are always from one layer to a later layer (Ripley, 1996).

Training an NN is a crucial step, where the weights of all the connections from the network are estimated based on a dataset. This step involves a trial-and-error procedure in which a non-linear optimization algorithm is applied to minimize some measure of forecasting error, using training and test samples (Chatfield, 2001). Once the network is trained, forecasts are made as new data arrive.

Due to its flexible structure, neural networks can be a suitable choice to develop traffic forecasting problems addressing important requirements from this area, such as the need to consider highly non-linear relationships in a multivariate context (as will be shown in Chapter 6), and also the need to have a tool to provide multiple step ahead forecasts in a simple manner. Indeed, several studies can be found in the literature showing the implementation of NNs in traffic modelling problems. For a list of papers using NN in traffic forecasting, see Vlahogianni *et al.* (2004) and Jiang and Adeli (2005).

Despite these advantages, the downside of the NN flexibility is the difficulty to find an optimal structure given the available data and the traffic forecasting problem under study. Current practise shows that rules of thumb and trial-and-error procedures are adopted based on a set of possible competing structures, although alternatives to efficiently optimize the NN architecture have been proposed (Vlahogianni *et al.*, 2005).

There are also some other potential problems when using NNs for traffic forecasting. Since each input is connected to each node with a weight being assigned to each connection, it can be easily verified that, even for a small number of input variables, a considerable number of parameters need to be estimated from data, which can lead to overfitting problems. Moreover, the large number of parameters involved in the non-linear optimization required in NN training requires a lot of data and, additionally, problems such as local minima and instability of the results given the initial parameter values can arise.

Furthermore, although an NN can be suitable for active traffic management systems operating in on-line environments, there are crucial issues that still need to be addressed in its development. First, while the main objective of traffic modelling is to describe the evolution of the traffic in a way that forecasting algorithms are consistent with the dynamically evolving nature of this system, the usual neural network procedure is static, in the sense that the weights of the network connections are estimated given a dataset and its estimates are constant unless the network is retrained. An alternative method of NN training based on the Kalman filter technique is proposed in Chen and Grant-Muller (2001), which results in better forecasts when compared to standard NN training approaches. Also, while sudden changes in the traffic flow can be acommodated in the NN estimation due to its flexibility, changes in the network structure require new training of the model. Moreover, subjective information concerning things such as road blockages or weather conditions

are very difficult to include in NNs. Indeed, some sources of uncertainty cannot be accommodated into a NN forecasting model because the model would need to learn about it first.

3.6 State space models

State space models have been successfully applied in the field of time series analysis. This is possibly due to their flexibility and generality, which make these models a suitable choice for a variety of problems that have become of great interest to statisticians over the years. Comprehensive studies of time series using state space modelling approaches can be found in Durbin and Koopman (2001), West and Harrison (1997) and Harvey (1989).

Given a time series $\{Y_t\}_{t\geq 1}$, the main principle underlying state space models is that there exists an unobservable *state process* $\{\boldsymbol{\theta}_t\}_{t\geq 0}$ which generates the observed values of the time series. The specification of a state space model (SSM) then consists of defining a relationship between the *observable process* $\{Y_t\}_{t\geq 1}$, and the unobservable process $\{\boldsymbol{\theta}_t\}_{t\geq 0}$, together with how the state process $\{\boldsymbol{\theta}_t\}_{t\geq 0}$ evolves in time.

Figure 3.3 gives a pictorial representation of a SSM. In this graph, the arrows from $\boldsymbol{\theta}_k$ to $\boldsymbol{\theta}_{k+1}$, $k = 0, \ldots, t+1$, represent the evolution of the state process $\{\boldsymbol{\theta}_t\}_{t\geq 0}$ in time. In a state space model, a *Markovian property* is defined for the dynamic stochastic evolution of the state process, that is, $\boldsymbol{\theta}_t$ depends only on $\boldsymbol{\theta}_{t-1}$ and a random error (not considered in the graph for simplicity). Also, given that $\boldsymbol{\theta}_t$ generates Y_t (represented in the graph by the arrows from $\boldsymbol{\theta}_i$ to Y_i, $i = 1, \ldots, t+1$), a description of the observable time series can be formulated, also considering a random error. Some properties of Figure 3.3 are further discussed in Chapter 4.

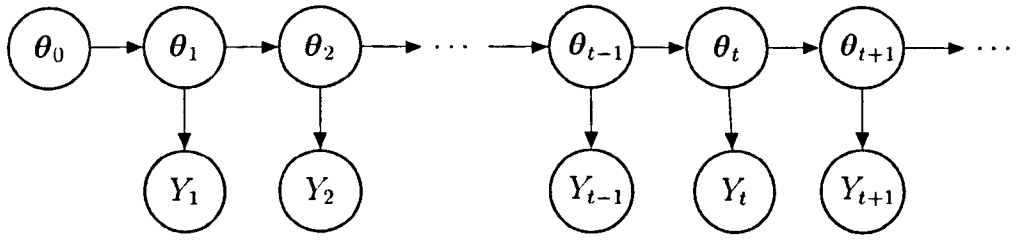

FIGURE 3.3: Graphical representation of a state space model (Petris *et al.*, 2009). θ_0 represents the initial information about the state process $\{\theta_t\}_{t\geq 0}$

3.6.1 Dynamic linear models

Dynamic linear models form an important class of SSMs, and their formulation relies on the assumption of a linear relationship between observable and state processes and also on the linearity assumption between consective values of the state process. In addition, normal distributions for the model error components are assumed. The development of the dynamic linear model (DLM) is extensively described in West and Harrison (1997). Additionally, Prado and West (2010) show recent developments regarding these models.

A univariate DLM for a time series $\{Y_t\}_{t\geq 1}$ is defined by the following observation and system equations, and initial information.

observation equation:

$$Y_t = \boldsymbol{F}_t^\top \boldsymbol{\theta}_t + v_t, \quad v_t \sim N(0, V_t) \tag{3.8}$$

system equation:

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \quad \boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{W}_t) \tag{3.9}$$

initial information:

$$\boldsymbol{\theta}_0 | D_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0) \tag{3.10}$$

In this definition, the p-dimensional regression vector $\boldsymbol{F}_t$ and the $p \times p$ evolution matrix $\boldsymbol{G}_t$ are assumed known, $\boldsymbol{\theta}_t$ is the p-dimensional state vector, the scalar V_t and the $p \times p$ matrix $\boldsymbol{W}_t$ are, respectively, the observation and evolution variances

and the sequences $\{v_t\}_{t\geq 1}$ and $\{\boldsymbol{w}_t\}_{t\geq 1}$ are independent normally distributed errors which are independent of each other. In (3.10), the p-dimensional mean vector $\boldsymbol{m}_0$ and the $p \times p$ covariance matrix $\boldsymbol{C}_0$ are assumed known, and D_0 represents all the available information at time 0.

A univariate time series model can be specified by defining each component of the set $\{\boldsymbol{F}_t, \boldsymbol{G}_t, V_t, \boldsymbol{W}_t\}$ in a DLM. This specification allows the inclusion of separate components representing different characteristics of the time series under study, such as trend and seasonality terms. It makes the DLM a structural model (Harvey, 1989) and a very flexible approach to represent non-stationary time series.

Given the system equation (3.9), the DLM can be viewed as a dynamic extension of the normal linear regression model, whose form can be represented by the observation equation (3.8).

3.6.1.1 DLM estimation and time series forecasting

Estimation of the state process $\{\boldsymbol{\theta}_t\}_{t\geq 1}$ and forecasts for Y_t given $\boldsymbol{\theta}_t$ are the main tasks when developing a DLM $\{\boldsymbol{F}_t, \boldsymbol{G}_t, V_t, \boldsymbol{W}_t\}$. The Bayesian solution for this problem when V_t and $\boldsymbol{W}_t$ are assumed known and constant is equivalent to the Kalman filter recursive equations (Kalman, 1960). However, V_t and $\boldsymbol{W}_t$ are unknown in most applications.

Consider a DLM with a constant but unknown observational variance V. Recall the (static) linear regression model of the form $y_i = \boldsymbol{X}^\top \boldsymbol{\beta} + \epsilon_i$, where $\boldsymbol{X}$ is a vector of covariates and ϵ_i has a normal distribution with null mean and unknown constant variance σ^2. A Bayesian conjugate analysis in this static model can be developed by assuming that $(\boldsymbol{\beta}, 1/\sigma^2)$ has a multivariate normal/gamma distribution (for details, see Gelman *et al.*, 1995). This distribution can also be used in a dynamic modelling context for a conjugate analysis of $(\boldsymbol{\theta}_t, \phi)$ in a DLM $\{\boldsymbol{F}_t, \boldsymbol{G}_t, V, V\boldsymbol{W}_t\}$ where ϕ is

the observational precision $1/V$. Notice that the evolution variance $\boldsymbol{W}_t$ is scaled by the observation variance V in this DLM, that is, $\boldsymbol{W}_t$ in equation (3.9) is now $V\boldsymbol{W}_t$ instead.

For practical purposes, interest relies on the marginal posterior distribution $(\boldsymbol{\theta}_t|D_t)$, where D_t represents all the available information at time t. Similar to the (static) linear regression model, when assuming a multivariate normal/gamma distribution for $(\boldsymbol{\theta}_t, \phi|D_t)$, the marginal $(\boldsymbol{\theta}_t|D_t)$ follows a multivariate T distribution.

The DLM sequential estimation procedure to obtain the posterior distribution $(\boldsymbol{\theta}_t, \phi)$ in time can be found in West and Harrison (1997). Let $(\boldsymbol{\theta}_{t-1}, \phi)|D_{t-1}$ be the posterior distribution at $t-1$ for a DLM $\{\boldsymbol{F}_t, \boldsymbol{G}_t, V, V\boldsymbol{W}_t\}$, where $\boldsymbol{F}_t$, $\boldsymbol{G}_t$ and $\boldsymbol{W}_t$ are known. Given $(\boldsymbol{\theta}_{t-1}, \phi)|D_{t-1}$, the marginal distribution $\boldsymbol{\theta}_{t-1}|D_{t-1}$ and system equation (3.9) can be used to obtain the prior distribution $\boldsymbol{\theta}_t|D_{t-1}$. With this prior, the one-step ahead forecast distribution of Y_t can be calculated via the observation equation (3.8). The posterior of the parameters is then updated when y_t is observed via a conjugate analysis, resulting in the posterior distribution $(\boldsymbol{\theta}_t, \phi)|D_t$.

These steps can be repeated as time evolves and as new data are observed. The key idea is that forecast distributions for Y_t are obtained given the prior for $\boldsymbol{\theta}_t$, and the posterior distribution of $(\boldsymbol{\theta}_t, \phi)$ is updated as new data arrive. This procedure is usually called the *Bayesian forecasting system* (West and Harrison, 1997).

Let $\{\boldsymbol{F}_t, \boldsymbol{G}_t, V, V\boldsymbol{W}_t\}$ be a DLM for a time series $\{Y_t\}_{t\geq 1}$. The distributions and updating equations for this DLM are described as follows (for details and proofs, see West and Harrison, 1997).

At time $t-1$, all beliefs about the parameters $(\boldsymbol{\theta}_{t-1}, \phi)$, where $\phi = V^{-1}$, are represented by, respectively, the posterior distributions

$$(\boldsymbol{\theta}_{t-1}|D_{t-1}) \sim T_{n_{t-1}}(\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}) \quad \text{and} \quad (\phi|D_{t-1}) \sim \text{Gamma}\,(n_{t-1}/2, n_{t-1}S_{t-1}/2)$$

which are respectively a multivariate T distribution with degress of freedom n_{t-1}, mean vector $\boldsymbol{m}_{t-1}$ and covariance matrix $\boldsymbol{C}_{t-1}$, and a gamma distribution with parameters $n_{t-1}/2$ and $n_{t-1}S_{t-1}/2$. The estimate of the observational variance V at time $t-1$ is S_{t-1}.

Given the system equation (3.9), the posterior distribution of $\boldsymbol{\theta}_{t-1}$ at time $t-1$ dynamically evolves into the prior distribution

$$(\boldsymbol{\theta}_t|D_{t-1}) \sim T_{n_{t-1}}(\boldsymbol{a}_t, S_{t-1}\boldsymbol{R}_t), \tag{3.11}$$

with parameters

$$\boldsymbol{a}_t = \boldsymbol{G}_t\boldsymbol{m}_{t-1} \tag{3.12}$$

and

$$\boldsymbol{R}_t = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t^\top + \boldsymbol{W}_t. \tag{3.13}$$

Given the distribution of $\boldsymbol{\theta}_t|D_{t-1}$, which represents the prior beliefs about the state vector $\boldsymbol{\theta}$ at time t, forecasts for Y_t are obtained through the observation equation (3.8). The one-step ahead forecast distribution for Y_t is then

$$(Y_t|D_{t-1}) \sim T_{n_{t-1}}(\boldsymbol{f}_t, Q_t)$$

which is a univariate T distribution with parameters n_{t-1}, $\boldsymbol{f}_t$ and Q_t, where

$$f_t = \boldsymbol{F}_t^\top \boldsymbol{a}_{t-1} \qquad \text{and} \qquad Q_t = \boldsymbol{F}_t^\top \boldsymbol{R}_t \boldsymbol{F}_t + S_{t-1}.$$

Finally, after y_t is observed, the beliefs about $(\boldsymbol{\theta}_{t-1}, \phi)$ are updated, resulting in the posterior distributions

$$(\boldsymbol{\theta}_t|D_t) \sim T_{n_t}(\boldsymbol{m}_t, \boldsymbol{C}_t) \qquad \text{and} \qquad (\phi|D_t) \sim \text{Gamma}\,(n_t/2, n_tS_t/2)$$

where

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t e_t \ , \qquad \text{and} \qquad \boldsymbol{C}_t = \frac{S_t}{S_{t-1}} \left[\boldsymbol{R}_t - \boldsymbol{A}_t \boldsymbol{A}_t^\top Q_t\right]$$

with $e_t = y_t - f_t$ and $\boldsymbol{A}_t = \boldsymbol{R}_t \boldsymbol{F}_t Q_t^{-1}$. Also,

$$n_t = n_{t-1} + 1 \ , \qquad \text{and} \qquad S_t = S_{t-1} + \frac{S_{t-1}}{n_t}\left[\frac{e_t^2}{Q_t} - 1\right].$$

The repetition of these steps over time leads to a *sequential* estimation of $(\{\boldsymbol{\theta}_t\}_{t\geq 1}, \phi)$ and $\{Y_t\}_{t\geq 1}$ forecasting.

When $\boldsymbol{W}_t$ is unknown, the sequential estimation for a DLM can be adapted as follows. Suppose the state process $\{\boldsymbol{\theta}_t\}_{t\geq 1}$ is constant in time. In this case, no error is considered in the dynamic evolution of this process, and the variance $\boldsymbol{W}_t$ of the evolution error $\boldsymbol{w}_t$ in the system equation (3.9) is a matrix containing only null values. Also, from equation (3.13), when $\boldsymbol{W}_t = 0$, the variance of the prior distribution for $(\boldsymbol{\theta}_t | D_{t-1})$ at time t is

$$\boldsymbol{R}_t = [\boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}_t^\top] = \boldsymbol{P}_t \tag{3.14}$$

Now, suppose $\boldsymbol{W}_t$ is positive definite, resulting in a stochastic evolution of the state process $\{\boldsymbol{\theta}_t\}_{t\geq 1}$. The evolution error $\boldsymbol{w}_t$ in system equation (3.9) represents in this case a loss of information when moving from $\boldsymbol{\theta}_{t-1}$ to $\boldsymbol{\theta}_t$ in the state process. The effect of this loss of information is controlled by the magnitude of $\boldsymbol{W}_t$ relative to $\boldsymbol{P}_t$ in equation (3.14). For $\delta \in (0, 1]$, we can then write

$$\boldsymbol{W}_t = \frac{(1-\delta)}{\delta} \boldsymbol{P}_t \tag{3.15}$$

and, with this, we have a relationship between $\boldsymbol{W}_t$ and $\boldsymbol{P}_t$ through a *discount factor* δ. The effect of the $\boldsymbol{w}_t$ on the dynamic evolution of $\{\boldsymbol{\theta}_t\}_{t\geq 1}$ can be controlled by δ: when $\delta = 1$, a constant state process is assumed for a DLM, whereas the smaller the δ, the bigger the variability induced by $\boldsymbol{w}_t$ in the evolution of $\{\boldsymbol{\theta}_t\}_{t\geq 1}$. Writing

$\boldsymbol{W}_t$ as a function of $\boldsymbol{P}_t$ in the variance of the prior distribution $\boldsymbol{\theta}_t|D_{t-1}$ in equation (3.13), we now have

$$\boldsymbol{R}_t = [\boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t^\top]\delta^{-1}. \tag{3.16}$$

In practise, the value of the discount factor is chosen by comparing the forecast accuracy of fitted DLMs based on different δ values, usually between 0.9 and 0.99.

3.6.2 Matrix normal dynamic linear models

Let $\{\boldsymbol{Y}_t\}_{t\geq 1}$ be a multivariate time series vector, where $\boldsymbol{Y}_t = [Y_t(1),\ldots,Y_t(n)]^\top$. Quintana and West (1987) developed an extension of the univariate DLM described in Subsection 3.6.1 by assuming that each component $Y_t(i)$ follows a DLM $\{\boldsymbol{F}_t(i),\boldsymbol{G}_t(i),V(i),V(i)\boldsymbol{W}(i)\}$ such that $\boldsymbol{F}_t(i) = \boldsymbol{F}_t$, $\boldsymbol{G}_t(i) = \boldsymbol{G}_t$ and $\boldsymbol{W}_t(i) = \boldsymbol{W}_t$, for $i = 1,\ldots,n$. From this idea, if each $Y_t(i)$ is modelled with a DLM $\{\boldsymbol{F}_t,\boldsymbol{G}_t,V(i),V(i)\boldsymbol{W}_t\}$, a closed-form Bayesian inference for univariate DLMs can be extended to multivariate time series. The series $\{\boldsymbol{Y}_t\}_{t\geq 1}$ then follows a *matrix normal DLM* (MNDLM) with state matrix $\boldsymbol{\Theta}_t$, where each column of $\boldsymbol{\Theta}_t$ corresponds to the state vector of each component of $\boldsymbol{Y}_t$. This model setup has also been called the *multivariate DLM* (Prado and West, 2010).

The MNDLM for a time series $\{\boldsymbol{Y}_t\}_{t\geq 1} = \{Y_t(i),\ldots,Y_t(n)\}_{t\geq 1}$ is defined by the following observation and system equations.

Observation equation:

$$\boldsymbol{Y}_t^\top = \boldsymbol{F}_t^\top\boldsymbol{\Theta}_t + \boldsymbol{v}_t,\ \boldsymbol{v}_t \sim N(\boldsymbol{0},\boldsymbol{\Sigma}), \tag{3.17}$$

system equation:

$$\boldsymbol{\Theta}_t = \boldsymbol{G}_t\boldsymbol{\Theta}_{t-1} + \boldsymbol{\Omega}_t,\ \ \boldsymbol{\Omega}_t \sim N(\boldsymbol{0},\boldsymbol{W}_t,\boldsymbol{\Sigma}), \tag{3.18}$$

The MNDLM assumes that $\boldsymbol{Y}_t$ is a function of the state evolution matrix Θ_t with the addition of a *vector* of observational errors $\boldsymbol{v}_t = [v_t(1), \ldots, v_t(n)]$ following a *multivariate* normal distribution with a cross-sectional *covariance* matrix $\boldsymbol{\Sigma}$. This matrix represents the covariance structure among the components of $\boldsymbol{Y}_t$ at each time t. In addition, the dynamic stochastic evolution of the state evolution matrix Θ_t, represented by equation (3.18), is a function of the state matrix at time $t-1$ with the addition of a *matrix* of evolution errors $\boldsymbol{\Omega}_t = [\boldsymbol{\omega}_t(1), \ldots, \boldsymbol{\omega}_t(n)]$. This matrix of evolution errors follows a matrix variant of the normal distribution. This variant, namely the *matrix-variate* (or just matrix) normal distribution, has a null vector mean, left variance matrix $\boldsymbol{W}_t$ and right variance matrix $\boldsymbol{\Sigma}$ (West and Harrison, 1997), denoted $N(\boldsymbol{0}, \boldsymbol{W}_t, \boldsymbol{\Sigma})$ in (3.18). Full details of the matrix normal distribution from a Bayesian perspective can be found in Dawid (1981). Also, $\boldsymbol{\omega}_t(i)$ and $v_t(i)$ are independent sequences such that each of them contains independent normally distributed errors, $i = 1, \ldots, n$.

Under the MNDLM, the methodology for a closed-form conjugate analysis of $(\Theta_t, \boldsymbol{\Sigma})$ is a direct multivariate extension of the normal-gamma conjugate inference for $(\boldsymbol{\theta}_t, \phi)$ in the univariate DLM. In the multivariate case, $(\Theta_t, \boldsymbol{\Sigma})$ follows a matrix normal/inverse Wishart distribution (Dawid, 1981). The following theorem summarizes the Bayesian sequential estimation and one-step ahead forecasting for the MNDLM.

Theorem 3.1. *(adapted from West and Harrison, 1997) Let* $\{\boldsymbol{Y}_t\}_{t\geq 1}$ *be governed by a MNDLM with equations (3.17) and (3.18). Suppose the initial prior for* $(\Theta_0, \boldsymbol{\Sigma})$ *follows a matrix normal/inverse Wishart distribution with* n_0 *degrees of freedom, mean vector* $\boldsymbol{m}_0$*, left variance matrix* $\boldsymbol{C}_0$ *and right variance matrix* $\boldsymbol{S}_0$*:*

$$(\Theta_0, \boldsymbol{\Sigma}|D_0) \sim NW_{n_0}^{-1}(\boldsymbol{m}_0, \boldsymbol{C}_0, \boldsymbol{S}_0).$$

The posterior distributions of $(\Theta_t, \boldsymbol{\Sigma})$ *and one-step ahead forecasts, at each time* t*, are as follows.*

(i) Posteriors at $t-1$:

$$(\Theta_{t-1}, \Sigma | D_{t-1}) \sim NW_{n_{t-1}}^{-1}(\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}, \boldsymbol{S}_{t-1}).$$

(ii) Priors at t:

$$(\Theta_t, \Sigma | D_{t-1}) \sim NW_{n_{t-1}}^{-1}(\boldsymbol{a}_t, \boldsymbol{R}_t, \boldsymbol{S}_{t-1}),$$

where

$$\boldsymbol{a}_t = \boldsymbol{G}_t \boldsymbol{m}_{t-1} \quad \textit{and} \quad \boldsymbol{R}_t = \boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}_t^\top + \boldsymbol{W}_t$$

(iii) One-step ahead forecast:

$$(\boldsymbol{Y}_t | \Sigma, D_{t-1}) \sim N(\boldsymbol{f}_t, Q_t \Sigma),$$

with marginal

$$(\boldsymbol{Y}_t | D_{t-1}) \sim T_{n_{t-1}}(\boldsymbol{f}_t, Q_t \boldsymbol{S}_{t-1}),$$

where

$$\boldsymbol{f}_t^\top = \boldsymbol{F}_t^\top \boldsymbol{a}_t \quad \textit{and} \quad Q_t = \boldsymbol{F}_t^\top \boldsymbol{R}_t \boldsymbol{F}_t^\top + 1.$$

(iv) Posteriors at t:

$$(\Theta_t, \Sigma | D_t) \sim NW_{n_t}^{-1}(\boldsymbol{m}_t, \boldsymbol{C}_t, \boldsymbol{S}_t).$$

with

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t \boldsymbol{e}_t^\top \quad \textit{and} \quad \boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t \boldsymbol{A}_t^\top Q_t,$$
$$n_t = n_{t-1} + 1 \quad \textit{and} \quad \boldsymbol{S}_t = n_{t-1}^{-1} \left[n_{t-1} \boldsymbol{S}_{t-1} + \boldsymbol{e}_t \boldsymbol{e}_t^\top Q_t^{-1} \right],$$

where

$$\boldsymbol{A}_t = \boldsymbol{R}_t \boldsymbol{F}_t / Q_t \quad \textit{and} \quad \boldsymbol{e}_t = \boldsymbol{y}_t - \boldsymbol{f}_t$$

Proof. See Quintana (1985, 1987). □

As $\boldsymbol{F}_t$, $\boldsymbol{G}_t$ and $\boldsymbol{W}_t$ are common to each $Y_t(i)$ in the MNDLM, so are the variances $\boldsymbol{C}_t$, $\boldsymbol{R}_t$ and Q_t. Hence, the use of discount factors for the evolution variance matrix $\boldsymbol{W}_t$ is straightforward and the same as for a univariate DLM (West and Harrison, 1997).

The MNDLM can be appropriate when analysing time series that are supposed to follow the same structure, since very similar DLMs must be defined for each time series component under this multivariate model. As an example, Prado and West (2010) applied the MNDLM to a 12-dimensional multivariate time series of exchange rates from different countries over a 10 year period. Their interest in particular was to analyse the covariance structure of the multivariate time series, so they applied principal components analysis in the estimate of the observational covariance matrix $\boldsymbol{\Sigma}$ to understand which components and exchange rates could better explain the variability in the series.

3.6.3 State space models in traffic modelling

Okutani and Stephanedes (1984) developed one of the first Kalman filter based traffic flow models, where their goal was to forecast flows at four sites from a simple Japanese urban network. Although they did not mention in the paper that the proposed model had a state space formulation, the Kalman filter was used to estimate what they called the state vector component. The model also considered lagged information from upstream sites as covariates to forecast downstream sites. They

showed that their model outperformed alternative solutions based on differential equations which took into account only past measurements of the site under study.

A state-space approach was also adopted by Whittaker *et al.* (1997), considering a motorway network with 500 data collecting sites in Holland. The variables collected in each site were flow, occupancy and speed. The approach considered a Bayesian multivariate time series model describing flow for each site as a function of lagged flow in upstream sites and also lagged occupancies, through the fundamental diagram of traffic (illustrated in Figure 3.1 of Section 3.2).

As pointed out by Whittaker *et al.* (1997), due to computational difficulties involving Kalman filter estimates due to model complexity, the resulting approach could not be implemented in real-time, and just some preliminary results considering a small subset of the network were presented in their paper. Also, it was assumed a time invariant network, that is, changes in the network could not be accommodated by the model.

Compared to this approach, Stathopoulos and Karlaftis (2003) developed a simpler state space model for traffic flows taking into account lagged values of upstream sites. The model proposed was applied to a network with five data collection sites, and had a better performance when compared to ARIMA models.

Tebaldi *et al.* (2002) used a dynamic linear model to forecast traffic flows in a motorway located in Seattle (US), with five loop detectors installed. Data available were from nine days of traffic counts during the morning period, collected on a minute-by-minute basis. Their model also considered lagged data from upstream sites as being informative about downstream sites, together with the inclusion of a component representing the smoothed trend of the traffic volume over time in the regression vector $\boldsymbol{F}_t$ via splines. The days with available data were considered as a sample of a population of days in which traffic flows could be generated. In

this sense, a *hierarchical* approach was considered in the development of the model, following the ideas presented in Lindley and Smith (1972). For the same data, a hierarchical static regression model was also developed, excluding the time-varying property of the parameters allowed in a DLM formulation.

Univariate static and dynamic linear models were fitted for each of the five data collection sites separately in Tebaldi *et al.* (2002). The results showed that the dynamic linear model provided better forecasts than the static model. The main reason for this difference was due to the DLM's ability to capture the dynamics of the traffic during the day. While the parameters of a DLM could change at each minute, thus having a bigger impact on traffic forecasts through the application of the Bayesian forecasting system, forecasts using the static model were calculated using the same parameter estimates, regardless of the time of the day.

Still in Tebaldi *et al.* (2002), lagged information about downstream sites was included in the developed models. The objective was to verify their effect on forecasting upstream sites. The inclusion of this type of data improved the forecasts just for one specific day, in which unexpected low traffic flows were observed. This can be an indication that downstream sites can be useful when forecasts are needed during congestion periods.

Considering the developments described in this subsection, state space models have been shown to be useful when applied to traffic forecasting. However, the state space models described here were only fitted to simple networks. Except for Whittaker *et al.* (1997), the multivariate nature of the models was in the number of covariates used to forecast the time series of interest, and not in forecasting the multivariate time series of traffic flows.

When models are required to forecast *simultaneously* a multivariate time series of

flows in a network, the resulting SSM formulation can be quite complex and unfeasible to be implemented in real time, as pointed out by Whittaker *et al.* (1997). From this perspective, the MNDLM could be an appropriate choice due to its closed form Bayesian sequential updating equations. However, in the MNDLM, the components of a multivariate time series of flows must have the same DLM structure. This is not necessarily reasonable when modelling flows from the same network. As a simple example, suppose an MNDLM must be developed to forecast flows for three consecutive sites $S(1)$, $S(2)$ and $S(3)$ in a road section, where traffic flows from $S(1)$ to $S(2)$ to $S(3)$. Assume further that, as in Tebaldi *et al.* (2002), lagged flows from upstream sites are informative about downstream sites. In this case, lagged flows from site $S(2)$ should be included in the regression vector $\boldsymbol{F}_t(3)$ when modelling flows at $S(3)$, whereas the regression vector for a flow model to $S(2)$ would require lagged flows observed at site $S(1)$. Such model structure is not allowed in an MNDLM, which requires the same $\boldsymbol{F}_t(i)$ when modelling a time series $\boldsymbol{Y}_t = [Y_t(1), \ldots, Y_t(n)]^\top$, $i = 1, \ldots, n$ (as shown in Subsection 3.6.2).

3.7 Summary

Despite the variety of approaches to develop short-term forecasting models, there are traffic modelling aspects that still must be addressed. As an example, it can be very difficult to build these models in a way that accurately captures the dynamically evolving nature of traffic data, since mathematical approaches are based on static deterministic relationships and both ARIMA and VARMA models are based on the assumption of stationarity of the time series. In addition, while neural networks can have the flexibility to adapt to sudden changes in traffic patterns, it can be quite difficult to determine the optimal configuration of the model and, as mentioned earlier, the model has to learn about new traffic patterns first. Moreover, it could

not be found in the literature how to easily calculate forecast limits in traffic flow models, as well as evidences of their calibration, as will be discussed in Section 5.3 of Chapter 5.

On the other hand, state space models can be a suitable alternative for traffic modelling. Indeed, these models are flexible to fit non-stationary time series, real-time forecasting can be done via a Bayesian sequential parameter estimation, and there is also the possibility of including external information through the use of intervention techniques. However, this model can become quite complex when applied to large networks, as shown in Whittaker *et al.* (1997).

Chapter 4

The multiregression dynamic model

4.1 Introduction

The multiregression dynamic model (MDM) is introduced in this chapter. Since the MDM uses a graph to represent multivariate time series, some important graph theory definitions and graphical modelling concepts are firstly presented.

This chapter also introduces the linear regression dynamic model (LMDM), which is a particular case of the MDM for which computations are particularly straightforward. An application of the LMDM using the Manchester network data is also presented. This example includes the elicitation of a graph representing the time series of flows observed at all Manchester network data collection sites.

4.2 Graphical models

Understanding the dependence structure among sets of random variables is a crucial problem in statistics. This problem is becoming even more important and challenging with the increasing availability of large-scale data sets in several fields.

Graphical models have been shown to be a powerful approach to represent complex dependence structures across variables, and their use has been widespread in many areas. For an overview of the sucessful use of graphical models in a variety of statistical applications, see Lauritzen (2003) and Jordan (2004).

4.2.1 Some definitions

Some of the definitions that will be used throughout the thesis are presented in this section. A comprehensive discussion of fundamental graph-theoretic definitions for graphical models can be found in Whittaker (1990) and Lauritzen (1996).

A *graph* $\mathcal{G}$ is defined as a pair (V, E), where V is a finite set of *nodes* (or *vertices*) and E is a subset of the Cartesian product $V \times V$ that contains *ordered* pairs of nodes, called *edges* of $\mathcal{G}$. A pictorial representation of a graph uses circles or ovals to describe nodes. Figure 4.1 shows a graph with nodes X_1, X_2, X_3 and X_4. From a statistical perspective, the nodes in a graph represent random variables or parameters. On the other hand, the edges connecting pairs of nodes in the graph define a relationship between nodes.

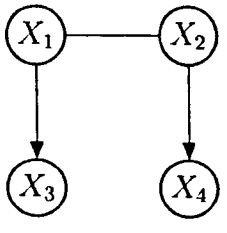

FIGURE 4.1: A graph.

An edge $(X, Y) \in E$ is called *directed* if and only if $(Y, X) \notin E$, otherwise (X, Y) is an *undirected* edge. A directed edge is represented by an arrow (or arc) in a graph. In Figure 4.1, there is an arc from X_1 to X_3 and from X_2 to X_4. Arcs in a graph describes a *direct influence* of a node X on Y. This influence can describe a *causal* relationship between X and Y.

When there is an arc from X to Y in a graph, X is called a *parent* of Y. Following the kinship analogy, Y is called a *child* of X. The sets of parents and children of a node X are denoted as $\mathrm{pa}(X)$ and $\mathrm{ch}(X)$ respectively. In the graph of 4.1, X_1 is a parent of X_3 and X_4 is a child of X_2, for example. A *root node* is any node without parents in the graph, that is, X is a root node if $\mathrm{pa}(X) = \emptyset$. The root nodes in the graph of Figure 4.1 are X_1 and X_2.

Undirected edges are represented by lines in a graph. These edges represent dependence between nodes without specifying the direction of the influence. For example, there is an undirected edge between X_1 and X_2 in the graph of Figure 4.1. In a graph, X and Y are also defined to be *neighbours* if they are connected by an undirected edge. An *undirected graph* is any graph that contains only undirected edges.

4.2.1.1 Directed acyclic graphs

Let $\mathcal{G} = (V, E)$ be a graph. A *path* of length n from α to β is any sequence $(\alpha_0 = \alpha, \ldots, \alpha_n = \beta)$ of distinct nodes such that $(\alpha_{i-1}, \alpha_i) \in E$ for all $i = 1, \ldots, n$. A *directed path* is any path $(\alpha_0, \ldots, \alpha_n)$ that has at least one directed edge. Given two nodes α and β, if there is a directed path from α to β, α is called an *ancestor* of β. Similarly, β is called a *descendant* of α. The set of descendants of α is denoted as

$\mathrm{de}(\alpha)$, whereas the set of *non-descendants* of α is denoted by $\mathrm{nd}(\alpha)$, such that

$$\mathrm{nd}(\alpha) = V \backslash (\mathrm{de}(\alpha) \cup \alpha),$$

where "\" means "excludes".

A *cycle* is any directed path that starts and ends at the same node. A graph which contains only directed edges and does not have any cycles, is called a *directed acyclic graph* (DAG). Figure 4.2 shows a DAG example. DAGs are commonly used when interest relies on the *causal structure* among variables. Section 4.8 shows how to elicit DAGs for time series of flows in road traffic networks.

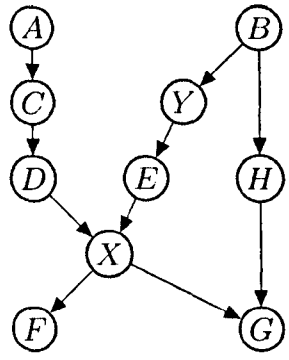

FIGURE 4.2: A directed acyclic graph (DAG).

Let $\boldsymbol{X} = [X_1, \ldots, X_n]^\top$ be a vector of random variables and let $\mathcal{G}$ be a DAG for $\boldsymbol{X}$. A *complete ordering* of the elements of $\boldsymbol{X}$, such that there is an arrow from X_i to X_j in $\mathcal{G}$ only when $i < j$, is the condition that ensures that $\mathcal{G}$ does not have any paths with cycles (Whittaker, 1990). It means that a DAG can only be elicited when there is a complete ordering across the components of $\boldsymbol{X}$.

DAGs form the basis of the *multiregression dynamic model* which will be described in Section 4.4.

4.2.1.2 Chain graphs

In contrast to DAGs, chain graphs can be used when there are only partial orderings among the components of $\boldsymbol{X}$. A chain graph (CG) can be viewed as a mixture of undirected graphs and DAGs. In a CG, the components of $\boldsymbol{X}$ are divided into *blocks*, and each block corresponds to a group of variables where no direction of dependence (such as a causal structure) are specified. Then, an ordering is defined among these blocks of variables. The arcs in the CG follow this block ordering, connecting nodes from a lower-numbered block to a higher-numbered block.

To understand how chain graphs can be elicited, consider the following example, adapted from Wermuth and Lauritzen (1990). Suppose of interest is the relationship between anxiety and smoking habits of individuals, as well as their relationship with other variables which may influence these characteristics. Suppose also that a cross-sectional observational study was conducted, where the following variables were measured on a group of survey respondents:

$$
\begin{aligned}
X_1 &= \text{gender, respondent;} \\
X_2 &= \text{socioeconomic status, respondent;} \\
X_3 &= \text{country of study;} \\
X_4 &= \text{Age when entering college, respondent;} \\
X_5 &= \text{smoking habits, respondent's parents;} \\
X_6 &= \text{anxiety measure, respondent;} \\
X_7 &= \text{smoking habits, respondent.}
\end{aligned}
$$

Figure 4.3 shows a chain graph for $\boldsymbol{X} = [X_1, \ldots, X_7]^\top$, motivated by subject matter knowledge from psychology theories (for details, see Wermuth and Lauritzen, 1990).

In this example, anxiety (X_6) and smoking habits (X_7) are defined as responses. The CG of Figure 4.3 is described below.

Suppose that there is no psychological evidence of a potential influence of anxiety on smoking habits, as well as no evidence that the anxiety of an individual is influenced by its smoking habits. Based on this idea, a *symmetric association* (Wermuth and Lauritzen, 1990) is assumed to occur between these variables, such that they are considered to be on an equal footing in the analysis. In this sense, X_6 and X_7 form a block of variables in the CG in Figure 4.3, and this symmetric association is represented by an undirected edge between these variables.

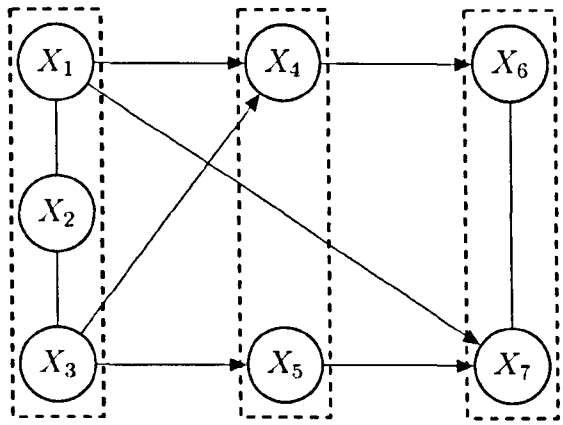

FIGURE 4.3: A chain graph example (adapted from Wermuth and Lauritzen, 1990).

The variables $\{X_1, \ldots, X_5\}$ are considered as potential influences on $\{X_6, X_7\}$, and these influences may be represented by directed edges from $\{X_1, \ldots, X_5\}$ to $\{X_6, X_7\}$. Furthermore, still based on subject matter knowledge, the subset of potential influencing variables $\{X_1, \ldots, X_5\}$ can be divided into a set of background variables formed by gender of respondent (X_1), socioeconomic status of the respondent (X_2) and country of study (X_3). An association structure assumption between these variables is also considered, represented by undirected edges between them. Finally, these background variables $\{X_1, X_2, X_3\}$ are assumed to be possible potential influences on the remaining block $\{X_4, X_5\}$. These influences may be also represented by directed edges from $\{X_1, X_2, X_3\}$ to $\{X_4, X_5\}$.

With all these assumptions, the CG in Figure 4.3 is formed by the *ordered blocks* B_1, B_2, and B_3, where $B_1 = \{X_1, X_2, X_3\}$, $B_2 = \{X_4, X_5\}$ and $B_3 = \{X_6, X_7\}$. The directed influences from a lower ordered block to a higher ordered block are represented by directed edges in the graph.

The *boundary* of a node α in a CG $\mathcal{G} = (V, E)$ is the set formed by parents and neighbours of $\alpha \in E$. This set is denoted as $\text{bd}(\alpha)$. For example, in Figure 4.3, $\text{bd}(X_6) = \{X_4, X_7\}$. Also, a subset of nodes $H \in V$ is called an *ancestral set* if $\text{bd}(\beta) \in H$ for all $\beta \in H$. In the CG of Figure 4.3, $H = \{X_1, X_2, X_3, X_4\}$ is an ancestral set, because

$$\text{bd}(X_1) = \{X_2\}, \quad \text{bd}(X_2) = \{X_1, X_3\}, \quad \text{bd}(X_3) = \{X_2\}, \quad \text{bd}(X_4) = \{X_1, X_3\}$$

and $\text{bd}(X_i) \in H$, $i = 1, 2, 3, 4$.

The *chain components* of a CG $\mathcal{G} = (V, E)$ are the connected sets of nodes when removing all directed edges in $\mathcal{G}$. The chain components of the chain graph in Figure 4.3 are $\{X_1, X_2, X_3\}$, $\{X_4\}$, $\{X_5\}$ and $\{X_6, X_7\}$.

Chain graphs are fundamental to the development of the multivariate time series model to be presented in Chapter 7.

4.2.2 Conditional independence and global Markov properties

A graphical model enables a high-dimensional statistical problem to be split up into small and manageable pieces. Representing a set of random variables by a graph not only provides a pictorial representation of the variables under study, but it also introduces *modularity* into the problem. This modularity means that simple local computations can be achieved, instead of a single (and possibly complex) one.

Conditional independence is the main concept underlying this modularity. Given a vector of random variables, a graph can be used to encode statements associated with *irrelevance* of one variable X when knowing about another variable Y, *given* the value of a variable Z. The association between such conditional independence relations and the graph plays a key role in the development of graphical modelling techniques in statistics. And there are a series of properties, namely *global Markov properties*, that translate the structure of graphs into conditional independence statements.

Let $\boldsymbol{X} = [X_1, \ldots, X_n]^\top$ be an ordered vector of random variables. Suppose that a DAG $\mathcal{G}$ can be elicited for $\boldsymbol{X}$, where X_1 is a root node. Using global Markov properties which will be described in the next subsection, the $n-1$ conditional independence statements

$$X_i \perp\!\!\!\perp [X_1, \ldots, X_{i-1}] \backslash \text{pa}(X_i) | \text{pa}(X_i) \tag{4.1}$$

can be derived from $\mathcal{G}$, where $i = 2, \ldots, n$. The conditional independence statements in (4.1), together with $\mathcal{G}$ are usually defined as a *Bayesian network* (Cowell *et al.*, 1999, Smith, 2010).

As an example, consider the vector of random variables $\boldsymbol{X} = [X_1, \ldots, X_4]^\top$. The joint density distribution of X can be writen as

$$f(\boldsymbol{x}) = f(x_1) f(x_2|x_1) f(x_3|x_1, x_2) f(x_4|x_1, x_2, x_3) \tag{4.2}$$

for any convenient order of $\boldsymbol{X}$ elements. Now, suppose the DAG $\mathcal{G}$ in Figure 4.4 can be used to represent the conditional independence structure among the components of $\boldsymbol{X}$. In this DAG, X_1 and X_2 are root nodes (that is, $\text{pa}(X_i) = \emptyset$, $i = 1, 2$) whereas $\text{pa}(X_3) = \{X_1, X_2\}$ and $\text{pa}(X_4) = X_3$. Hence, using (4.1) to rewrite equation (4.2), we have that

$$f(\boldsymbol{x}|\mathcal{G}) = f(x_1) f(x_2) f(x_3|x_1, x_2) f(x_4|x_3).$$

Rewriting the density function of $\boldsymbol{X}$ conditioning on $\mathcal{G}$ simplifies some factors of the original density function (4.2). For example, while x_4 is conditioned on $\{x_1, x_2, x_3\}$ in (4.2), the conditional independence structure of $\mathcal{G}$ implies that the conditional density of x_4 can be represented in (4.2) based on its parent x_3 only.

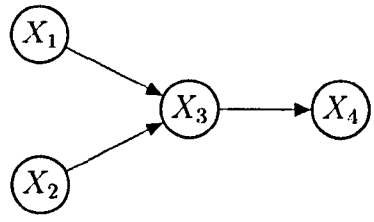

FIGURE 4.4: A DAG for $\boldsymbol{X} = [X_1, \ldots, X_4]^\top$.

A *statistical model* for $\boldsymbol{X}$ is defined as a family of density functions

$$\mathcal{M} = \{f(\boldsymbol{x}|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\},$$

where Θ is the parameter space (Cox, 2006). A graphical model characterizes a statistical model in the sense that $\mathcal{M}$ can be redefined in order to ensure that all the conditional independence relations translated from a graph hold for all densities in the statistical model. For example, given $\mathcal{G}$, a statistical model for $\boldsymbol{X}$ can be

$$\mathcal{M}_{\mathcal{G}} = \{f(\boldsymbol{x}|\boldsymbol{\theta}, \mathcal{G}) : \boldsymbol{\theta} \in \Theta_{\mathcal{G}}\},$$

such that

$$f(\boldsymbol{x}|\boldsymbol{\theta}, \mathcal{G}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}, \mathrm{pa}(x_i)).$$

Given a chain graph $\mathcal{C} = (V, E)$ and its ordered blocks $B_1, \ldots, B_k$, the joint probability distribution of a vector $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ factorizes through the distribution of B_1 together with the distribution of each block B_i conditioned on the union of previous blocks $B_1, \ldots, B_{i-1}$, $i = 2, \ldots, k$ (Edwards, 2000). Hence, we have,

$$f(\boldsymbol{x}) = f(B_1)f(B_2|B_1)f(B_3|B_1 \cup B_2)\ldots f(B_k|B_1 \cup B_2 \ldots \cup B_{k-1}). \tag{4.3}$$

Again, using global Markov properties which will be described in the next subsection, the $n-1$ conditional independence statements

$$X_i \perp\!\!\!\perp \mathrm{nd}(X_i) \backslash \mathrm{bd}(X_i) | \mathrm{bd}(X_i) \tag{4.4}$$

can be deduced from a chain graph (Lauritzen, 1996). Since

$$[X_1, \ldots, X_{i-1}] \in \mathrm{nd}(X_i), \quad i = 1, \ldots, n,$$

it is also true that

$$X_i \perp\!\!\!\perp [X_1, \ldots, X_{i-1}] \backslash \mathrm{bd}(X_i) | \mathrm{bd}(X_i), \tag{4.5}$$

and these statements can be used to decompose each conditional distribution in (4.3). For example, recall the (chain) graph in Figure 4.1 of Subsection 4.2.1 for $\boldsymbol{X} = \{X_1, X_2, X_3, X_4\}$, and suppose that $B_1 = \{X_1, X_2\}$, $B_2 = X_3$ and $B_3 = X_4$. From equation (4.3), the joint distribution of $\boldsymbol{X}$ is

$$\begin{aligned} f(\boldsymbol{x}) &= f(B_1) f(B_2 | B_1) f(B_3 | B_1 \cup B_2) \\ &= f(x_1, x_2) f(x_3 | x_1, x_2) f(x_4 | x_1, x_2, x_3) \end{aligned}$$

and, following (4.5), we have the conditional independence statements

$$X_3 \perp\!\!\!\perp X_2 | X_1, \quad \textit{and} \quad X_4 \perp\!\!\!\perp \{X_1, X_3\} | X_2.$$

Hence,

$$f(\boldsymbol{x}) = f(x_1, x_2) f(x_3 | x_1) f(x_4 | x_2).$$

4.2.2.1 Global Markov properties for DAGs and CGs

The development of global Markov properties to associate the structure of graphs with conditional independence statements has been the focus of extensive research in the graphical modelling field. Lauritzen (1996) details a variety of Markov properties of different graphical structures. This section focuses on global Markov properties of DAGs and CGs.

Pearl (1988) introduced the d-separation criterion, from which conditional independence statements can be translated from a DAG. An equivalent criterion, namely the *moralization criterion*, was defined in Lauritzen *et al.* (1990). The moralization criterion can be better understood by defining an algorithm to verify whether, given a DAG $\mathcal{G} = (V, E)$, an arbitrary conditional independence statement, say $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} | \boldsymbol{Z}$, for sets of variables $\{\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}\} \in V$, is true. This algorithm, described by Dawid (2002) in its present form, is defined by the following steps:

> *Step 1:* (ancestral graph): remove from the DAG any node which is neither in $\boldsymbol{X} \cup \boldsymbol{Y} \cup \boldsymbol{Z}$ nor an ancestor of a node in this set, together with any edges in or out of such nodes.
>
> *Step 2:* (moralization): add an undirected edge between any two remaining nodes which have a common child, but are not already connected by an arrow. Then replace all arcs in the moralized graph by undirected edges.
>
> *Step 3*: (separation): the statement $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} | \boldsymbol{Z}$ is true if, in the resulting undirected graph based on steps 1 and 2 above, all the paths which join a node in $\boldsymbol{X}$ to one in $\boldsymbol{Y}$ are intersected by $\boldsymbol{Z}$. In this case, $\boldsymbol{Z}$ is said to *separate* $\boldsymbol{X}$ and $\boldsymbol{Y}$.

To illustrate how this algorithm can be used, consider the DAG in Figure 4.2, taken from Cowell *et al.* (1999). Suppose that of interest is verifying whether

$A \perp\!\!\!\perp B | \{X, Y\}$. The ancestral graph induced by $A \cup B \cup X \cup Y$ is given in Figure 4.5(a) (step 1 of the the moralization criterion algorithm). The only nodes which have a common child in this graph are D and E. These nodes must be connected by an undirected edge to moralize the graph (step 2). Figure 4.5(b) shows the ancestral moralized graph for $A \cup B \cup X \cup Y$ replacing all arcs by undirected edges (step 2).

The possible paths from A to B in the graph of Figure 4.5(b) are A–C–D–E–Y–B and A–C–D–X–E–Y–B. Since these paths are intersected by $\{X, Y\}$, the statement $A \perp\!\!\!\perp B | \{X, Y\}$ can be deduced from the DAG in Figure 4.2.

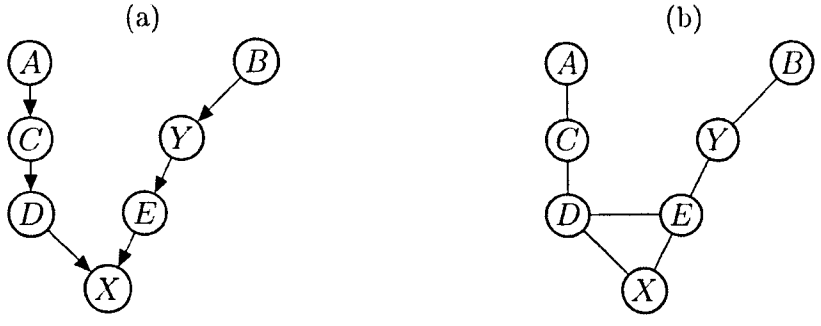

FIGURE 4.5: Ancestral graph (a) and moralized ancestral graph (b) to verify whether $A \perp\!\!\!\perp B | X, Y$ given the DAG in Figure 4.2 (Cowell *et al.*, 1999).

The global Markov property for CGs (Frydenberg, 1990) can be described by an algorithm very similar to the one defined above for DAGs. In the criterion for CGs, the main difference compared to the moralization criterion for DAGs is in the moralization step: a CG is moralized by adding undirected edges between nodes that have children in a *common* chain component, and then replacing the arcs of the resulting graphs by undirected edges (Cowell *et al.*, 1999).

Now, let $\mathcal{G} = (V, E)$ be a CG. Suppose that of interest is whether $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} | \boldsymbol{Z}$, for sets of variables $\{\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}\} \in V$. An algorithm to verify such a statement is given by the following steps:

Step 1 (ancestral set): define a subgraph containing the smallest ancestral set of $\{\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}\}$.

Step 2 (moralization): add an undirected edge between any two remaining nodes which have children in a common chain component of $\mathcal{G}$, but are not already connected by an arrow. Then replace any arcs in the moralized graph by undirected edges.

Step 3 (separation): the statement $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} | \boldsymbol{Z}$ is true if, in the resulting undirected subgraph based on steps 1 and 2 above, $\boldsymbol{Z}$ separates $\boldsymbol{X}$ and $\boldsymbol{Y}$.

As an example, suppose we want to verify whether $F \perp\!\!\!\perp D | E$ in the graph $\mathcal{G}$ of Figure 4.6(a), provided by Edwards (2000). Figure 4.6(b) shows the smallest ancestral graph of $\{F, D, E\}$. The chain components of this graph are $\{A, B, C\}$,$\{D, E\}$ and $\{F, G\}$. Figure 4.6(c) shows the the moralized version of this ancestral graph, where the orange lines are undirected edges due to moralization. Since this moralized graph in Figure 4.6(c) contains a path, D–G–F, which does not pass through E, the nodes F and D are not separated by E. Hence, the statement $F \perp\!\!\!\perp D | E$ is not true given the graph $\mathcal{G}$ of Figure 4.6(a).

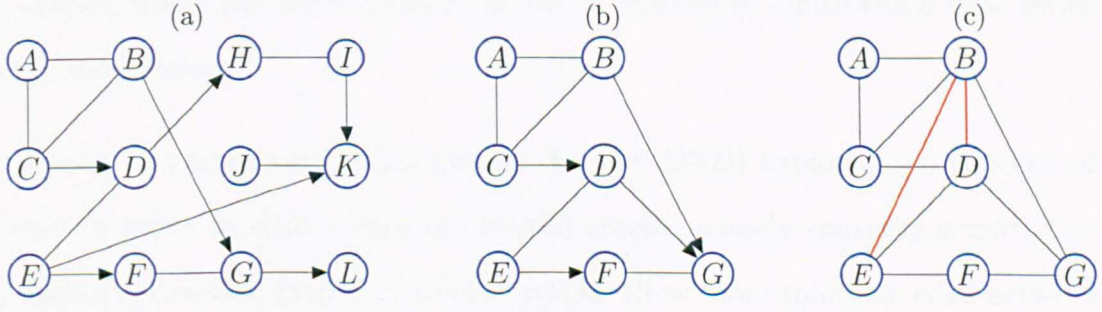

FIGURE 4.6: A chain graph example (a), together with ancestral graph (b) and moralized ancestral graph (c) to verify whether $F \perp\!\!\!\perp D | E$ (Edwards, 2000).

4.3 Graphical models for time series

4.3.1 Introduction

Some of the first ideas to combine graphical models with multivariate time series can be found in Queen and Smith (1993), which developed the multiregression dynamic model to be described in Section 4.4. Brillinger (1996) also proposed the use of graphs to represent time series, providing some examples where structural time series models (which has the dynamic linear model presented in Subsection 3.6.1 of Chapter 3 as an example) were defined from graphical representations of multivariate processes. Brillinger also stressed the importance of applying the work developed on causality by Pearl (1995) in the context of econometrics, since "economy can be viewed as a vast system or network of interconnected processes" (Brillinger, 1996).

Along these lines, Dahlhaus (2000) applied concepts from frequency domain analysis of time series (Chatfield, 2003) to elicit undirected graphs known as *partial correlation graphs*, where the dependence structure of stationary multivariate time series can be characterized.

In contrast to partial correlation graphs, Eichler (2000) explored the concept of Granger causality to elicit a class of directed graphs, namely *causality graphs*. Unlike partial correlation graphs, causality graphs allow more than one edge between two nodes. Dahlhaus and Eichler (2003) give an overview of causality and partial correlation graphs as well as examples of their application.

The crucial issue when applying graphical models for time series is the *temporal* dependence among the components of a time series vector. The DAG of the state space model given in Figure 3.3 of Chapter 3 is an example of this application. At each time t, the DAG represents the direct influence of the state vector $\boldsymbol{\theta}_t$ on the observed value Y_t, represented by an arc from $\boldsymbol{\theta}_t$ to Y_t. This dependence is replicated

at each time frame. The temporal dependence between $\{\boldsymbol{\theta}_{t-1}, Y_{t-1}\}$ and $\{\boldsymbol{\theta}_t, Y_t\}$ is characterized by the arcs between $\boldsymbol{\theta}_{t-1}$ and $\boldsymbol{\theta}_t$ and arc omissions between Y_{t-1} and Y_t. This DAG then encodes the conditional independence statements (Smith, 2010)

$$Y_t \perp\!\!\!\perp \{Y_1, \ldots, Y_{t-1}\}, \boldsymbol{\theta}_{t-1} | \boldsymbol{\theta}_t \quad i = 1, \ldots, n$$

and

$$\boldsymbol{\theta}_t \perp\!\!\!\perp \{Y_1, \ldots, Y_{t-1}\}, \{\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_{t-2}\} | \boldsymbol{\theta}_{t-1} \quad i = 2, \ldots, n.$$

The idea of replicating a graph at each time t taking into account the temporal dependence structure among the components of time series models forms the basis of what is usually called the *dynamic graphical model* (DGM). A particular case of a DGM is when the dynamic graphical representation of a multivariate time series is defined by a DAG, resulting then in a *dynamic Bayesian network* (DBN). See Korb and Nicholson (2010) for an introduction to DBNs and Xiang *et al.* (2011) for an example of this class of model.

4.3.2 Dynamic graphical models

Apart from the multiregression dynamic model which will be described in Section 4.4, two classes of models named *dynamic graphical models* can be found in the literature, both based on state space modelling approaches. The first one, developed by Queen and Smith (1992), will be presented in Chapter 7. In this subsection, the focus is on the class of DGMs developed by Carvalho and West (2007).

The key idea of Carvalho and West was to include an undirected graph in the structure of the matrix normal DLM (MNDLM) described in Chapter 3. To understand the concept of this DGM, recall the definition of the MNDLM for a time series vector $\boldsymbol{Y}_t = [Y_t(1), \ldots, Y_t(n)]^\top$, given as follows.

Observation equation:

$$\boldsymbol{Y}_t^\top = \boldsymbol{F}_t^\top \Theta_t + \boldsymbol{v}_t,\ \boldsymbol{v}_t \sim N(\boldsymbol{0}, \Sigma_t), \tag{4.6}$$

System equation:

$$\Theta_t = \boldsymbol{G}_t \Theta_{t-1} + \Omega_t,\ \ \Omega_t \sim N(\boldsymbol{0}, \boldsymbol{W}_t, \Sigma_t), \tag{4.7}$$

Initial information:

$$\Theta_0 | D_0 \sim N(\boldsymbol{p}_0, \boldsymbol{W}_0, \Sigma_0),$$

As previously described, a closed-form conjugate analysis for (Θ_t, Σ_t) can be obtained via a matrix normal/inverse Wishart distribution.

Consider the covariance matrix Σ_t in observation equation (4.6). Given the precision matrix $\Psi_t = \Sigma_t^{-1}$, possible independencies across the components of $\boldsymbol{Y}_t$ can be accommodated by defining zero off-diagonal elements in Ψ_t, allowing a *sparsity structure* in the model. This sparsity structure can be further represented by an undirected graph $\mathcal{G} = (V, E)$ such that $V = \{Y_1, \ldots, Y_n\}$ and with undirected edges between Y_i and Y_j, $i \neq j$, if and only if $\psi(i, j) = 0$, where $\psi(\cdot, \cdot)$ are elements of Ψ_t.

Using the factorization of $f(\boldsymbol{y}_t | \Sigma_t^{-1}, \mathcal{G})$ (Lauritzen, 1996), Carvalho and West (2007) developed a closed-form conjugate analysis for (Θ_t, Σ_t) based on the matrix normal/hyper-inverse Wishart distribution. The hyper-inverse Wishart distribution is the extension of the Wishart one for graphical models (Dawid and Lauritzen, 1993, Roverato, 2002).

An attractive feature of this class of DGM is the possibility to accommodate uncertainty with respect to the undirected graph for Ψ_t. A stochastic graphical search procedure developed in Jones *et al.* (2005) is then applied to calculate posterior probabilities for possible graph configurations sequentially over time.

Within this framework, the DGM of Carvalho and West (2007) extends the MNDLM to accommodate sparsity structures in the cross-sectional covariance matrix $\boldsymbol{\Sigma}_t$ in equations (4.6) and (4.7). Although this extension simplifies modelling of high-dimensional of time series, this model still requires that the regression vector $\boldsymbol{F}_t$ in equation (4.6) must be common for the components of the time series of interest. This makes this DGM not suitable for multivariate flow forecasting, as discussed in Section 3.6.3 of Chapter 3.

Applications of this DGM can be found in Carvalho and West (2007), Prado and West (2010) and Wang *et al.* (2011). Wang and West (2009) developed an extension of this model to matrix-valued time series.

4.4 The multiregression dynamic model (MDM)

Multiregression dynamic models (Queen and Smith, 1993) are a class of DGMs which also combine graphs with state space models. Whereas Carvalho and West (2007) use an undirected graph to represent conditional independence relationships thus allowing sparsity in the covariance structure of a multivariate time series, a *causal driving mechanism* between the elements of a time series vector $\boldsymbol{Y}_t$ is the main assumption underlying the MDM.

With this assumption, a DAG can be used to represent any conditional independence relationships *related to causality* across time series components at each time frame. Thus, the MDM can introduce modularity in a high-dimensional time series problem by breaking a multivariate model into simpler univariate components.

4.4.1 Model definition

Let $\boldsymbol{Y}_t = [Y_t(1), \ldots, Y_t(n)]^\top$ be an n-dimensional time series vector, where $y_t(i)$ is the observed value of $Y_t(i)$ and $\boldsymbol{y}^t(i) = [y_1(i), \ldots, y_t(i)]^\top$ is the observed process associated with $Y_t(i)$, $i = 1, \ldots, n$. Let $\boldsymbol{\theta}_t^\top = [\boldsymbol{\theta}_t(1)^\top, \ldots, \boldsymbol{\theta}_t(n)^\top]$ be the state vectors of $\boldsymbol{Y}_t$, where $\dim(\boldsymbol{\theta}_t(i)) = s_i$, $i = 1, \ldots, n$. For notational convenience, define

$$\boldsymbol{X}_t(i) = [Y_t(1), \ldots, Y_t(i-1)]^\top, \quad i = 2, \ldots, n$$

and

$$\boldsymbol{Z}_t(i) = [Y_t(i+1), \ldots, Y_t(n)]^\top, \quad i = r+1, \ldots, n-1,$$

with observed processes

$$\boldsymbol{x}^t(i) = \{y_t(1), \ldots, y_t(i-1)\}_{t \geq 1}$$

and

$$\boldsymbol{z}^t(i) = \{y_t(i+1), \ldots, y_t(n)\}_{t \geq 1}$$

respectively. Suppose further a conditional independence structure related to causality such that the variables are ordered and indexed. Then, a DAG can be elicited for the time series so that, at each time $t \in \mathbb{N}$, for $i = 2, \ldots, n$,

$$Y_t(i) \perp\!\!\!\perp \{[Y_t(1), \ldots, Y_t(i-1)] \backslash \mathrm{pa}(Y_t(i))\} | \mathrm{pa}(Y_t(i))$$

and

$$Y_t(i) \perp\!\!\!\perp \{[\boldsymbol{Y}^t(1), \ldots, \boldsymbol{Y}^t(i-1)] \backslash \mathrm{pa}(\boldsymbol{Y}^t(i))\} | \{\mathrm{pa}(\boldsymbol{Y}^t(i)), \boldsymbol{Y}^{t-1}(i)\}$$

The MDM is then defined by the following observation and system equations and initial information.

Observation equations:

$$Y_t(i) = \boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i) + v_t(i), \;\; v_t(i) \sim (0, V_t(i)), \; i = 1, \ldots, n. \tag{4.8}$$

System equation:

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \;\; \boldsymbol{w}_t \sim (\boldsymbol{0}, \boldsymbol{W}_t) \,, \tag{4.9}$$

Initial information:

$$\boldsymbol{\theta}_0 | D_0 \sim (\boldsymbol{m}_0, \boldsymbol{C}_0). \tag{4.10}$$

In the MDM, the s_i-dimensional vector $\boldsymbol{F}_t(i)$ is a function of $\boldsymbol{x}^t(i)$ and $\boldsymbol{y}^{t-1}(i)$, *but not* $\boldsymbol{z}^t(i)$. Also, $V_t(i)$ are observational variances, $i = 1, \ldots, n$, and the $s \times s$ matrices $\boldsymbol{G}_t$, $\boldsymbol{W}_t$ and $\boldsymbol{C}_0$ are defined as,

$$\begin{aligned} \boldsymbol{G}_t &= \text{blockdiag}(\boldsymbol{G}_t(1), \ldots, \boldsymbol{G}_t(n)), \\ \boldsymbol{W}_t &= \text{blockdiag}(\boldsymbol{W}_t(1), \ldots, \boldsymbol{W}_t(n)) \end{aligned}$$

and

$$\boldsymbol{C}_0 = \text{blockdiag}(\boldsymbol{C}_0(1), \ldots, \boldsymbol{C}_0(n)),$$

such that $s = \sum_{i=1}^n s_i$ and the $s_i \times s_i$ matrices $\boldsymbol{G}_t(i)$ and $\boldsymbol{W}_t(i)$ may be functions of $\boldsymbol{x}^{t-1}(i)$ and $\boldsymbol{y}^{t-1}(i)$, *but not* $\boldsymbol{z}^{t-1}(i)$, $i = 1, \ldots, n$. In addition, $\boldsymbol{w}_t^\top = (\boldsymbol{w}_t(1)^\top \; \cdots \; \boldsymbol{w}_t(n)^\top)$, and $v_t(1), \ldots, v_t(n)$ and $\boldsymbol{w}_t(1), \ldots, \boldsymbol{w}_t(n)$, are independent sequences of independent errors.

Similar to the DLM estimation and forecasting approach described in Subsection 3.6.1 of Chapter 3, the prior distribution for $\boldsymbol{\theta}_t | D_{t-1}$ is obtained from the posterior distribution $\boldsymbol{\theta}_{t-1} | D_{t-1}$ and system equation (4.9) for any time t. Forecast distributions for each $Y_t(i)$ conditional on $\text{pa}(Y_t(i))$ are then found separately via (4.8).

Within this framework, the MDM allows the following n conditional distributions,

$$Y_t(i)|\{D_{t-1}, \boldsymbol{F}_t(i), \boldsymbol{\theta}_t(i)\} \sim (\boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i), V_t(i)), \quad i = 1, \ldots, n. \tag{4.11}$$

Note that distributional assumptions for both observation and evolution errors are not required in equations (4.8) and (4.9). This results in very flexible conditional distributions for $Y_t(i)|\{D_{t-1}, \boldsymbol{F}_t(i), \boldsymbol{\theta}_t(i)\}$, $i = 1, \ldots, n$.

Since $\boldsymbol{F}_t(i)$ may be a function of $\boldsymbol{x}^t(i)$, and both $Y_t(i)$ and $\boldsymbol{x}^t(i)$ are simultaneously observed at a given time t, *marginal* distributions for each $Y_t(i)$ are required. The moments of these marginal distributions can be analytically calculated when assuming normal observation and evolution error distributions and also by imposing linearity with respect with the parameters in both observation and system equations (4.8) and (4.9). A detailed example of a particular MDM considering these assumptions is presented in Section 4.7.

The independence across the state vectors $\boldsymbol{\theta}_t(i)$, $i = 1, \ldots, n$, is an important result which allows independent forecasts for each $Y_t(i)$ using the conditional distributions in (4.11), and also allows each $\boldsymbol{\theta}_t(i)$ to be updated separately in $Y_t(i)$'s univariate conditional model. This result is formalized by the following theorem.

Theorem 4.1. *Let* $\{\boldsymbol{Y}_t\}_{t\geq 1}$ *be governed by an MDM. If* $\perp\!\!\!\perp_{i=1}^n \boldsymbol{\theta}_0(i)$, *then*

$$\perp\!\!\!\perp_{i=1}^n \boldsymbol{\theta}_t(i)|\boldsymbol{y}^t$$

and

$$\boldsymbol{\theta}_t(i) \perp\!\!\!\perp \boldsymbol{z}^t(i)|\boldsymbol{x}^t(i), \boldsymbol{y}^t(i).$$

Proof. See Queen and Smith (1993). □

This theorem enables independent sequential Bayesian inference for each $\boldsymbol{\theta}_t(i)$, $i = 1, ..., n$, under the MDM, after $\boldsymbol{y}_t$ is observed at each time frame. That is, given that $\perp\!\!\!\perp_{i=1}^{n} \boldsymbol{\theta}_0(i)$, which is a consequence of setting $\boldsymbol{C}_0$ to be block diagonal in equation (4.10) of the MDM, the state vectors $\boldsymbol{\theta}_t(i)$, $i = 1, ..., n$, *remain independent after sampling* (as termed by Smith, 2010). Therefore, the MDM tackles the problem of modelling multivariate time series by considering separate univariate models for each $Y_t(i)|\{\boldsymbol{\theta}_t(i), \text{pa}(Y_t(i))\}$, $i = 1, ..., n$, allowing local computations in possible high-dimensional settings.

4.5 The linear multiregression dynamic model (LMDM)

The *linear multiregression dynamic model* (LMDM) is a particular case of the MDM when assuming linear relationships with respect with the parameters and normal distributions for the random errors in equations (4.8) and (4.9). In the LMDM, a normal distribution is also considered for the initial information about $\boldsymbol{\theta}_0$ in (4.10). With this approach, the LMDM uses the DAG to model the multivariate time series by n univariate separate regression DLMs: one each for $Y_t(1)$ and $Y_t(i)|\text{pa}(Y_t(i))$, $i = 2, ..., n$. Each time series has its parents as linear regressors, while root nodes are modelled by any suitable DLMs. As such, the LMDM is computationally simple and DLM techniques can be readily applied.

Formally, the LMDM is defined as follows.

Observation equations:

$$Y_t(i) = \boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i) + v_t(i), \; v_t(i) \sim N(0, V_t(i)), \; i = 1, \ldots, n. \qquad (4.12)$$

System equation:

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim N(0, \boldsymbol{W}_t). \tag{4.13}$$

Initial information:

$$\boldsymbol{\theta}_0|D_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0).$$

Similar to the MDM definition, the s_i-dimensional vector $\boldsymbol{F}_t(i)$ contains an arbitrary, but known, function of the parents $\text{pa}(Y_t(i))$ and possibly other known exogenous variables; $\boldsymbol{\theta}_t(i)$ is the s_i-dimensional parameter vector for $Y_t(i)$, and $\boldsymbol{\theta}_t^\top = (\boldsymbol{\theta}_t(1)^\top \cdots \boldsymbol{\theta}_t(n)^\top)$; $V_t(1), \ldots, V_t(n)$ are the scalar observation variances; $\boldsymbol{m}_0$ and $\boldsymbol{C}_0$ are the moments for $\boldsymbol{\theta}_0|D_0$; the $s \times s$ matrices $\boldsymbol{G}_t$, $\boldsymbol{W}_t$, and $\boldsymbol{C}_0$ are block diagonal, where $s = \sum_{i=1}^n s_i$; $\boldsymbol{w}_t^\top = (\boldsymbol{w}_t(1)^\top \cdots \boldsymbol{w}_t(n)^\top)$, and $v_t(1), \ldots, v_t(n)$ and $\boldsymbol{w}_t(1), \ldots, \boldsymbol{w}_t(n)$, are independent sequences of independent errors.

Even though the LMDM conditional distributions are assumed as being normal, the joint forecast distribution of $\mathbf{Y}_t$ can yield highly non-Gaussian patterns, as well as their marginal distributions (Queen and Smith, 1993). Also, analytical expressions for the marginal forecast distributions of $Y_t(i)$ cannot generally be obtained. However, marginal forecast moments for $Y_t(i)$ are readily available using the standard probability identities

$$E[Y_t(i)|D_{t-1}] = E\{E[Y_t(i)|D_{t-1}, \text{pa}(Y_t(i))]\}, \tag{4.14}$$

$$V[Y_t(i)|D_{t-1}] = E\{V[Y_t(i)|D_{t-1}, \text{pa}(Y_t(i))]|D_{t-1}\} + V\{E[Y_t(i)|D_{t-1}, \text{pa}(Y_t(i))]|D_{t-1}\}, \tag{4.15}$$

and, given two variables $Z_t(1)$ and $Z_t(2)$,

$$E[Z_t(1)Z_t(2)] = E\{E[Z_t(2), Z_t(1)|Z_t(2)]\} = E\{Z_t(2)E[Z_t(1)|Z_t(2)]\} \quad (4.16)$$

Section 4.7 shows an example of the LMDM including the calculation of these marginal moments.

4.6 Assessing MDM performance

Different measures are considered to verify the accuracy of the models to be developed in this thesis, due to the nature of traffic data and the characteristics of interest in the models. The forecast accuracy measures to be adopted are the *median square error*, the *log-predictive likelihood* and the *mean interval score*, defined as follows.

Let $\{y_t\}$ be a time series observed at times $t = 1, \ldots, T$ and $\{f_t\}$ be a sequence of forecasts for $\{y_t\}$. The median squared error (MedianSE) is

$$\text{MedianSE} = \text{median}\left\{[y_i - f_i]^2\right\}_{i=1,\ldots,T}.$$

The smaller the MedianSE, the more support there is for the corresponding model. The MedianSE is preferred to the mean squared error in the evaluation of flow forecasting models due to the large number of possible outliers in traffic data (Queen *et al.*, 2007).

When evaluating model forecast performance in this thesis, it is important to consider not only the performance of point forecasts but also the quality of its precision. Thus, a measure which assesses the accuracy of the multivariate forecast distribution as a whole, rather than just the point forecasts, is preferred. Such a measure is the joint log-predictive likelihood (LPL). After observing the time series up to

time T, the LPL evaluates the log of the density of the joint one-step ahead forecast distribution at time t at the observed value $\boldsymbol{y}_t = (y_t(1), \ldots, y_t(n))^T$, and aggregates these over all values $t = 1, \ldots, T$. In the LMDM, because of the conditional independence structure across $Y_t(1), \ldots, Y_t(n)$, the density of the joint one-step ahead forecast distribution at time t evaluated at the observed value $\boldsymbol{y}_t$ is given by

$$f(\boldsymbol{y}_t|D_{t-1}) = \prod_{i=1}^{n} f(y_t(i)|\text{pa}(y_t(i)), D_{t-1}),$$

where $f(y_t(i)|\text{pa}(y_t(i)), D_{t-1})$ is the one-step forecast density for $Y_t(i)$ conditional on its parents evaluated at $\boldsymbol{y}_t$. Thus, the LPL for the LMDM is calculated as

$$\text{LPL} = \sum_{t=1}^{T} \left[\sum_{i=1}^{n} \log\{f(y_t(i)|\text{pa}(y_t(i)), D_{t-1})\} \right].$$

The larger the value of the LPL, the more support there is for the corresponding model.

Because the forecast variance directly affects the forecast limits, an alternative, decision theoretically principled way of comparing forecast performance, is through the mean interval score (Gneiting and Raftery, 2007). Denote the lower and upper $100(1-\alpha)\%$ forecast limits for $Y_t(i)$ by $l_t(i)$ and $u_t(i)$, respectively. Then, for each t, the interval score for $(l_t(i), u_t(i))$ for each observation $y_t(i)$ is

$$\begin{aligned} S_\alpha^{int}(l_t(i), u_t(i), y_t(i)) &= (u_t(i) - l_t(i)) + \frac{2}{\alpha}(l_t(i) - y_t(i))I(y_t(i) < l_t(i)) + \\ &+ \frac{2}{\alpha}(y_t(i) - u_t(i))I(y_t(i) > u_t(i)), \end{aligned}$$

where I is an indicator function. This measure is based on the range of the forecast limits $(l_t(i), u_t(i))$ and also considers a penalty in the situation that these limits do not cover the observation $y_t(i)$. This results in a negatively oriented score, that is, the smaller the measure, the better the model from which the forecast limits were

calculated. The mean interval score is then calculated over all observations in a time series. This idea can be extended to the multivariate MDM setting by simply calculating the mean interval score over all observations for each time series.

4.7 An LMDM example

To illustrate the definition of an LMDM and the derivation of marginal moments of forecast distributions for time series vector components, a DAG from Queen *et. al* (2007) and Queen and Albers (2009) will be used. This DAG represents a traffic network, hereafter called the *London network*, formed by an intersection of the M25 orbital motorway with the motorways A2 and A282 in the east of London, UK. An aerial view of the London network is given in Figure 4.7(a).

Data from this network are available for 21 weeks in the form of hourly vehicle counts in 17 sites of the network. The schematic diagram in Figure 4.7(b) shows the distribution of these data collection sites over the intersection.

The sites 167, 168, 170A and 170B of the London network will be used in this example (see Figure 4.7(b)). Thus, we have the vector,

$$\mathbf{Y}_t = [Y_t(167), Y_t(168), Y_t(170A), Y_t(170B)]^\top,$$

which represents hourly vehicle flows in these four sites.

The DAG for this subnetwork, described in Queen *et al.* (2007) is given in Figure 4.8. From the data collecting sites diagram of this network, it is known that vehicles on site 167 can leave the A282 motorway and go to site 168, or they can continue on the road, and go to sites 170A or 170B. Hence, we have that $Y_t(167)$ is a parent of $Y_t(170) = Y_t(170A) + Y_t(170B)$ and $Y_t(168)$, for example. Also, following Queen

FIGURE 4.7: The London network. (a) Aerial photograph (©2012 DigitalGlobe, GeoEye, Infoterra Ltd & Bluesky, The GeoInformation Group, Map data ©2012 Google) and (b) schematic diagram: the grey diamonds are the data collection sites, each of which is numbered. The arrows indicate the direction of traffic flow on each part of the network.

et al. (2007), both $Y_t(168)$ and $Y_t(170A)$ are logical functions of their parents, and are known since their parents are known. Using the terminology of the WinBUGS software (Lunn *et al.*, 2000), they are called *logical variables* and are represented by a double oval in the DAG.

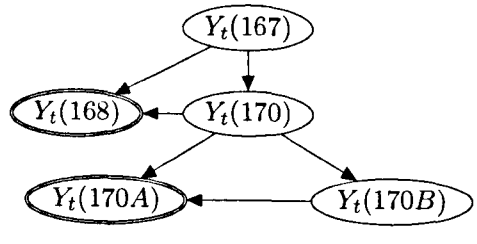

FIGURE 4.8: DAG for a subset of the London network based on sites 167, 168, $170A$ and $170B$.

Let

$$Y_t(167) = Y_t(1), \quad Y_t(170) = Y_t(170A) + Y_t(170B) = Y_t(2), \quad Y_t(170B) = Y_t(3),$$

$$Y_t(168) = Y_t(4) \quad \text{and} \quad Y_t(170A) = Y_t(5).$$

Given the DAG in Figure 4.8, the LMDM for this subnetwork is defined as follows.

Observation equations:

$$\begin{aligned} Y_t(1) &= \boldsymbol{F}_t(1)^\top \boldsymbol{\theta}_t(1) + v_t(1), & v_t(1) &\sim N(0, V_t(1)) \\ Y_t(2) &= \boldsymbol{F}_t(2)^\top \boldsymbol{\theta}_t(2) + v_t(2), & v_t(2) &\sim N(0, V_t(2)) \\ Y_t(3) &= \boldsymbol{F}_t(3)^\top \boldsymbol{\theta}_t(3) + v_t(1), & v_t(3) &\sim N(0, V_t(3)) \end{aligned}$$

where,

$$\boldsymbol{F}_t(1)^\top = [1\ 0 \ldots 0]\ , \qquad \boldsymbol{\theta}_t(1)^\top = [\theta_1(1) \ldots \theta_{24}(1)]$$

$$\boldsymbol{F}_t(2)^\top = [y_t(1)\ 0 \ldots 0]\ , \qquad \boldsymbol{\theta}_t(2)^\top = [\theta_1(2) \ldots \theta_{24}(2)]$$

$$\boldsymbol{F}_t(3)^\top = [y_t(2)\ 0 \ldots 0]\ , \qquad \boldsymbol{\theta}_t(3)^\top = [\theta_1(3) \ldots \theta_{24}(3)]$$

Logical variables:

$$Y_t(4) = Y_t(1) - Y_t(2)$$

$$Y_t(5) = Y_t(2) - Y_t(3)$$

In this LMDM, the regression vectors $\boldsymbol{F}_t(2)$ and $\boldsymbol{F}_t(3)$ contain the observed values of the parents of the time series $Y_t(2)$ and $Y_t(3)$, respectively. Since $Y_t(1)$ does not have parents, $\boldsymbol{F}_t(1)$ contains the scalar 1 followed by a vector of zeros. To accommodate the different flow patterns within the day (as shown in Chapter 2), a seasonal DLM model (West and Harrison, 1997) will be used for each time series (apart from the logical variables). Hence, the vectors $\boldsymbol{\theta}_t(1)$, $\boldsymbol{\theta}_t(2)$ and $\boldsymbol{\theta}_t(3)$ will

contain one parameter for each hour of the day. Thus we have,

$$\boldsymbol{\theta}_t(i)^\top = [\theta_1(i), \theta_2(i), \ldots, \theta_{24}(i)]$$

for $i = 1, 2, 3$.

System equation:

$$\boldsymbol{\theta}_t = \boldsymbol{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim N(\,0, \boldsymbol{W}_t)$$

where,

$$\boldsymbol{\theta}_t{}^\top = [\boldsymbol{\theta}_t(1)^\top \;\; \boldsymbol{\theta}_t(2)^\top \;\; \boldsymbol{\theta}_t(3)^\top] = [\theta_1(1) \ldots \theta_{24}(1) \;\; \theta_1(2) \ldots \theta_{24}(2) \;\; \theta_1(3) \ldots \theta_{24}(3)],$$

$$\boldsymbol{\theta}_{t-1} = [\theta_{24}(1) \ldots \theta_1(1) \;\; \theta_{24}(2) \ldots \theta_1(2) \;\; \theta_{24}(3) \ldots \theta_1(3)]^\top,$$

$$\boldsymbol{G} = \text{blockdiag}\{\boldsymbol{G}(1), \boldsymbol{G}(2), \boldsymbol{G}(3)\},$$

where the 24×24 square matrix $\boldsymbol{G}(i)$, $i = 1, 2, 3$, is

$$\boldsymbol{G}(i) = \begin{pmatrix} a & 1-a & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & & & \ddots & 1 \\ 1 & 0 & \ldots & \ldots & 0 \end{pmatrix},$$

$$\boldsymbol{W}_t = \text{blockdiag}\{\boldsymbol{W}_t(1), \boldsymbol{W}_t(2), \boldsymbol{W}_t(3)\}.$$

The structure of the matrices $\boldsymbol{G}(1)$, $\boldsymbol{G}(2)$ and $\boldsymbol{G}(3)$ is such that the hour parameters are "rotated" at each time as required. Also, the scalar a in $G(i)$, $i = 1, 2, 3$ is included to allow a small effect of the parameter for the corresponding hour from the previous day, which was found to improve forecasts (Queen *et al.*, 2007).

Initial information:

$$(\boldsymbol{\theta}_0 | D_0) \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0)$$

where,

$$\boldsymbol{C}_0 = \text{blockdiag}\{\boldsymbol{C}_0(1), \boldsymbol{C}_0(2), \boldsymbol{C}_0(3)\},$$

with $\boldsymbol{C}_0(i)$ being a 24×24 square matrix and assumed known, $i = 1, 2, 3$.

Hence, from (4.11), we have the following conditional one-step ahead forecast distributions,

$$Y_t(1)|D_{t-1} \sim N(\boldsymbol{F}_t(1)^\top \boldsymbol{\theta}_t(1), V_t(1)),$$
$$Y_t(2)|D_{t-1}, y_t(1) \sim N(\boldsymbol{F}_t(2)^\top \boldsymbol{\theta}_t(2), V_t(2)),$$

and

$$Y_t(3)|D_{t-1}, y_t(2) \sim N(\boldsymbol{F}_t(3)^\top \boldsymbol{\theta}_t(3), V_t(3)).$$

Note that these are *conditional* forecast distributions. After observing $\boldsymbol{y}_t$, the distribution for each $\boldsymbol{\theta}_t(i)$ can be updated separately (in closed form) within the (conditional) DLM for $Y_t(i)|\text{pa}(Y_t(i))$, because of the initial independence of $\boldsymbol{\theta}_0(i)$, $i = 1, 2, 3$, through the block diagonal form for $\boldsymbol{C}_0$ and the block diagonal forms of $\boldsymbol{W}_t$ and $\boldsymbol{G}_t$ in the LMDM.

The derivations of first and second moments of both conditional and marginal one-step ahead forecast distributions for this LMDM are in Appendix A.

4.8 Building an LMDM for the Manchester network

This section describes the elicitation of a DAG and the associated LMDM for the Manchester network presented in Chapter 2. The elicitation process defined here

generalizes the one developed in Queen *et al.* (2007) by assuming that all types of road layouts can be split up into combinations of two types of junction: joins and forks. Moreover, a causal driving mechanism motivates this elicitation, in the sense that traffic flows at upstream sites can be informative about traffic flows at downstream sites.

4.8.1 Forks and joins

Traffic networks are basically a series of junctions of two types: forks and joins. A fork, in which vehicles from a single site $S(1)$ move to two sites $S(2)$ and $S(3)$, is illustrated in Figure 4.9(a). A join, in which traffic from two sites, $S(4)$ and $S(5)$, merge to a single site $S(6)$, is illustrated in Figure 4.9(b). Elicitation of a DAG and LMDM for each fork and join in a network are the building blocks for eliciting a DAG and LMDM for an entire network.

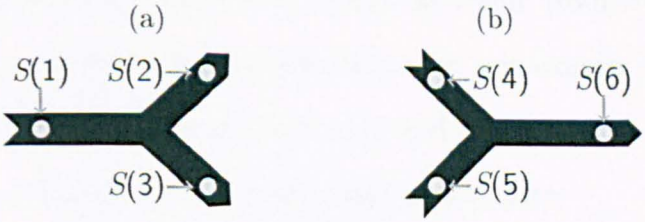

FIGURE 4.9: (a) a fork and (b) a join. In each diagram, the arrows denote the direction of travel and the circles are the sites.

Let $Y_t(i)$ be the number of vehicles passing site $S(i)$ during time t. Following Queen *et al.* (2007) and using upstream flows to model downstream flows, the observation equations of an LMDM for $Y_t(1), Y_t(2), Y_t(3)$ and $Y_t(4), Y_t(5), Y_t(6)$ can be defined as

$$\begin{aligned} &Y_t(1) = \mu_t(1) + v_t(1), \quad Y_t(2) = \alpha_t y_t(1) + v_t(2), \quad Y_t(3) = y_t(1) - y_t(2), \\ &Y_t(4) = \mu_t(4) + v_t(4), \quad Y_t(5) = \mu_t(5) + v_t(5), \quad\quad Y_t(6) = y_t(4) + y_t(5). \end{aligned} \tag{4.17}$$

In (4.17), the $\mu_t(\cdot)$ parameters are level parameters, while parameter α_t represents the proportion of traffic flowing from $S(1)$ to $S(2)$, and $v_t(\cdot)$ are normal error terms.

In Queen *et al.* (2007) the normality of the errors $v_t(\cdot)$ is justified by appealing to the Poisson approximation to normality for large means. The flows from the Manchester network cannot be considered either Poisson or normal, as will be seen in Chapter 5. However, the flow variance does increase as a function of the flow mean. West and Harrison (1997) propose using a variance law within a normal DLM to model such non-normal data. Thus, in order to take advantage of the computational simplicity of the LMDM and the ease with which established DLM techniques can be incorporated into the model, normal errors will be used for $v_t(\cdot)$ and, in Chapter 5, the LMDM will be extended to incorporate a variance law to accommodate the non-normality of the data.

The series $Y_t(3)$ and $Y_t(6)$ are modelled as logical variables. This is because all traffic from $S(1)$ must flow to $S(2)$ and $S(3)$, while all traffic from $S(4)$ and $S(5)$ flows to $S(6)$. Of course, these logical relationships are not exactly true because some vehicles will be between sites at the start and end of the period. This error should, however, be small enough to make this model appropriate.

DAGs representing the fork and join are given in Figure 4.10. Because the model for $Y_t(2)$ depends on $Y_t(1)$, $Y_t(1)$ is a parent of $Y_t(2)$, and hence there is an arc from $Y_t(1)$ to $Y_t(2)$ in the DAG, and so on. Logical variables are denoted by double ovals in the DAG. Joining together the DAGs of individual forks and joins provides a general method for eliciting a DAG and associated LMDM for an entire network. Figure 4.11 shows the full DAG for the Manchester network.

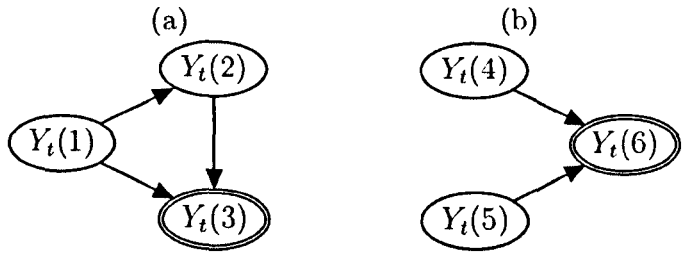

FIGURE 4.10: DAGs representing (a) a fork and (b) a join. The double ovals represent logical variables.

4.8.2 Model parameters

Although each of the observation equations for $Y_t(1)$, $Y_t(4)$ and $Y_t(5)$ in (4.17) are algebraically the same for each time t, the actual parameters, $\mu_t(1)$, $\mu_t(4)$ and $\mu_t(5)$, will exhibit a diurnal cycle, as clearly shown in Figure 2.3(a) of the Chapter 2. This diurnal cycle can be modelled by a seasonal factor DLM as considered in Section 4.7, thus having a mean flow level parameter for each time period in the day. A Fourier form DLM (West and Harrison, 1997, Section 8.6) or using splines to represent the smooth flow trend over the day (as in Tebaldi *et al.*, 2002) could be considered.

The advantage of a seasonal factor model is its interpretability, which, as demonstrated in Queen and Albers (2009), is especially helpful at times of modelling change via intervention (the technique of intervention allows information regarding a change in the time series to be fed into the model to maintain forecast performance — see West and Harrison, 1997, Section 11.2). When flow data are aggregated to small time intervals such as 5 minutes, a seasonal factor model can cause numerical instability problems with the Kalman filter computations because of the large number of parameters on possibly different scales. In this case, either a Fourier or a smooth trend model would be preferable for parsimony. However, for 15-min data, a seasonal factor model does not have such problems and computation is fast and efficient. In Chapter 5, 15-min data are modelled and a seasonal factor model is used to model

seasonality, whereas Chapter 6 models 5-min data and splines will be used instead there.

The parameter α_t in (4.17) represents the proportion of traffic flowing from parent to child which, as illustrated in Figure 4.12, can vary systematically at different times of the day. The diurnal pattern exhibited by the parameter α_t can also be modelled by a seasonal factor model as described in Queen and Albers (2009). Again, there are computational problems in doing this for 5-min data. Thus, for the 15-min data in Chapter 5, a seasonal factor is used for α_t, while for the 5-min data in Chapter 6 splines are used to model the seasonality.

4.8.3 Linear relationship between parent and child

The LMDM equation for $Y_t(2)$ in (4.17) assumes a linear relationship between parent and child. Figure 4.13, showing typical plots of 15-min flows for parent versus child at different times of the day, illustrates why this is a realistic assumption. A linear relationship would explain most of the variation between parent and child in each plot, although the relationship is not the same throughout the day. This is simply a consequence of the diurnal cycle of the proportion parameter α_t, as demonstrated in Figure 4.12. Notice that there seems to be two separate regimes in the plot for 17:15–17:29. This is the result of some unusual flows requiring intervention at the parent site.

4.8.4 Contemporaneous flows as regressors

As mentioned in Chapter 2, the distances between data collection sites in the Manchester network are such that vehicles are usually counted at several data sites in the same 5-min interval. As a result, the flows at sites upstream to a site $S(i)$ at time

t are helpful in forecasting the flows at $S(i)$ *at the same time* t. Since the LMDM can accommodate contemporaneous flows (which are not known at time t) as linear regressors, this model is suitable for forecasting flows in networks where vehicles are simultaneously observed at different data collection sites, as is the case in the Manchester network.

On the other hand, Tebaldi *et al.* (2002) considered lagged upstream flows for forecasting at time t. As these lagged flows are known at time t, this model can be a simpler approach when compared to the LMDM to forecast flows in an on-line environment. The comparison of the LMDM with univariate DLMs with lagged flows as regressors (as, for example, in Tebaldi *et al.*, 2002) is therefore an important issue to support the use of the LMDM for real-time traffic flow forecasting.

To make this comparison, both models were fitted using 15-min and 5-min flows between 07:00–20:59 (ignoring the quiet night-time period) during May 2010 at sites 9206B and 9200B of the Manchester network . To evaluate the quality of one-step ahead forecasts, both MedianSE and LPL were calculated for each model. Table 4.1 shows MedianSE and LPL for both models for both 15-min and 5-min flows. The LMDM did indeed perform better than univariate DLMs with lagged flows as regressors.

TABLE 4.1: Median SE and LPL for LMDMs and univariate DLMs using 15-min and 5-min flows of the Manchester network

Model for $Y_t(9206B), Y_t(9200B)$	15-min data		5-min data	
	MedianSE	LPL	MedianSE	LPL
Univariate DLMs with lagged flows	2876	-7198	914	-3834
LMDM with contemporaneous flows	1154	-1288	232	-3404

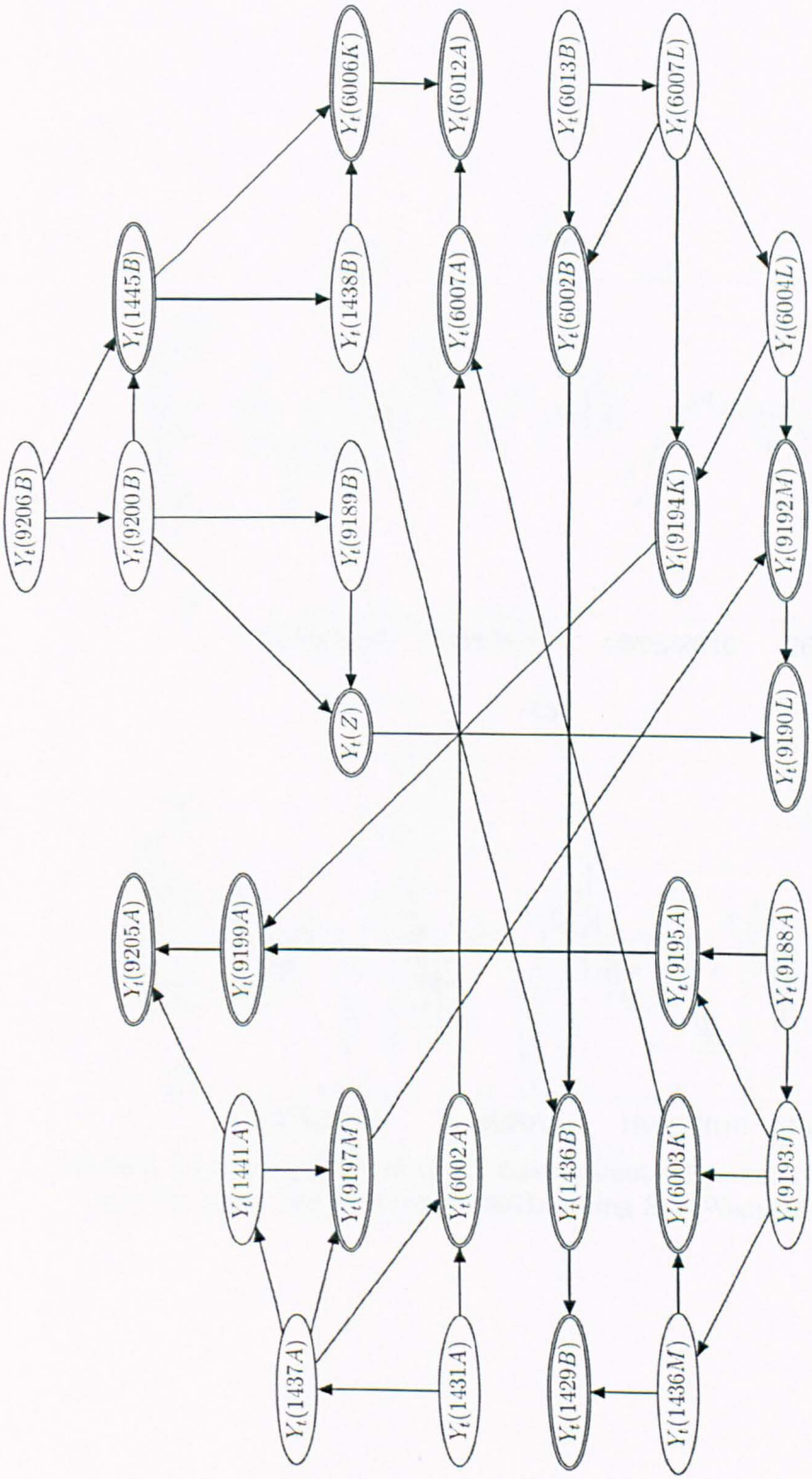

FIGURE 4.11: DAG for traffic data collection sites in the Manchester network.

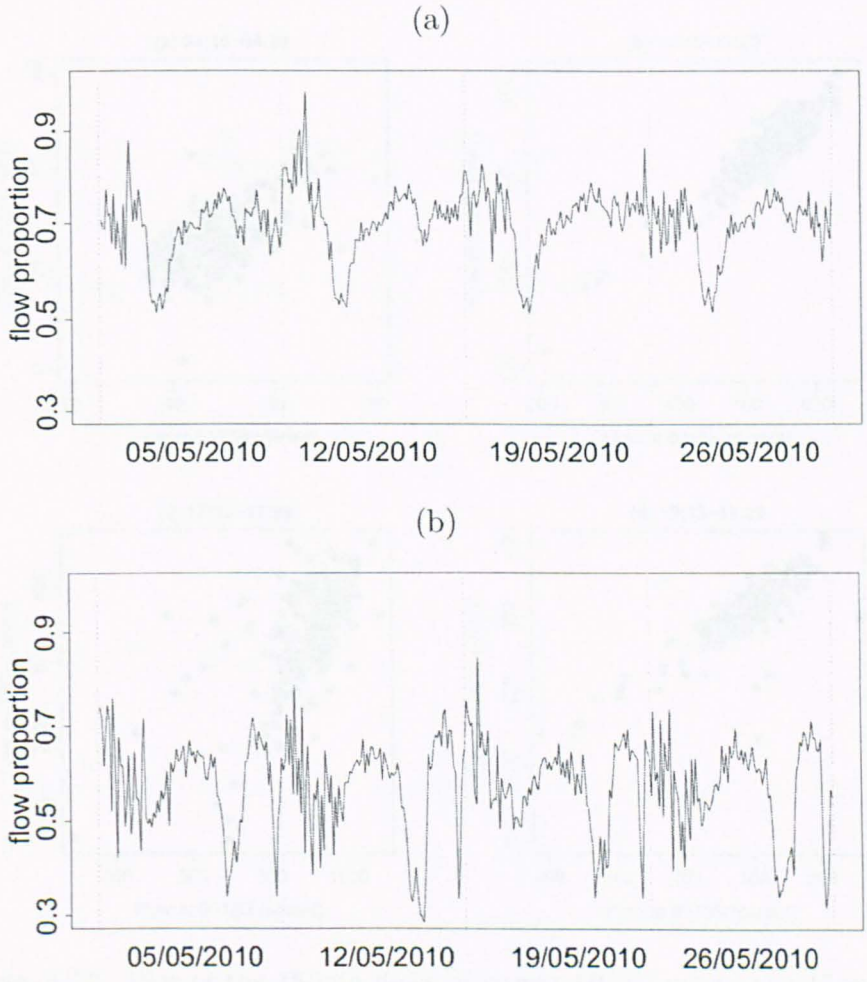

FIGURE 4.12: Proportion of traffic flowing from (a) parent 1431A to child 1437A and (b) parent 6013B to child 6007L during four Wednesdays in May 2010.

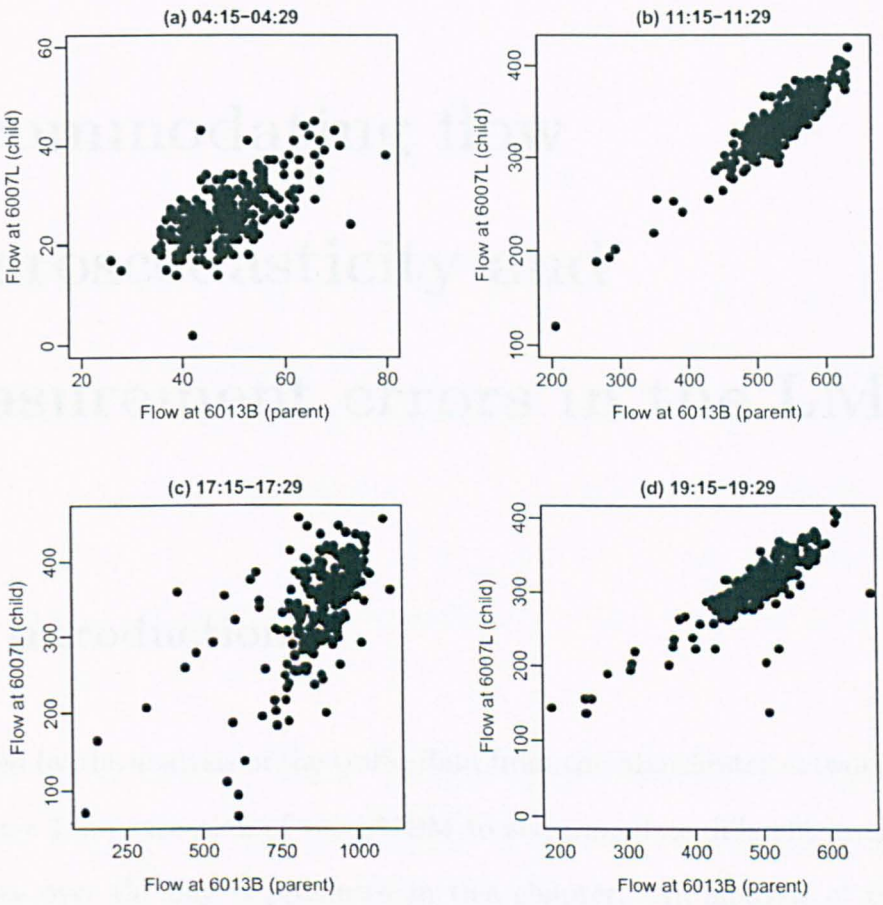

FIGURE 4.13: Plot of the 15-min flows of parent 6013B versus the 15-min flows of its child 6007L for some periods of the day (plots on different scales).

Chapter 5

Accommodating flow heteroscedasticity and measurement errors in the LMDM

5.1 Introduction

Motivated by the analysis of the traffic data from the Manchester network described in Chapter 4, an extension of the LMDM to accommodate different levels of traffic variability over the day is presented in this chapter. An analysis of the forecast limits provided by this extended LMDM is done by comparing its approximate and simulation-based forecast limits, and also by verifying the coverage of the approximate ones. Following these developments, this chapter also shows how measurement errors due to malfunctions in traffic data collection devices can be better accommodated within the LMDM.

5.2 Modelling flow heteroscedasticity

Consider an LMDM for an n-dimensional time series vector $\boldsymbol{Y}_t = [Y_t(1), \ldots, Y_t(n)]^\top$, with observation and system equations

$$Y_t(i) = \boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i) + v_t(i), \quad v_t(i) \sim N(0, V_t(i)), \quad i = 1, \ldots, n, \tag{5.1}$$

and

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim N(0, \boldsymbol{W}_t) \tag{5.2}$$

respectively, as defined in Section 4.5 of Chapter 4. In all previous LMDM applications in traffic forecasting, the observation variance $V_t(i)$ was assumed unknown and, following West and Harrison (1997), a conjugate analysis was done following a multivariate normal/gamma distribution for $(\boldsymbol{\theta}_t(i), \phi_t(i))$, where $\phi_t(i)$ is the observational precision $1/V_t(i)$, $i = 1, \ldots, n$. This conjugate analysis considers $V_t(i) = V(i)$ in (5.1), that is, the observation variance $V_t(i)$ is assumed constant over time for each time series $Y_t(i)$, $i = 1, \ldots, n$.

However, the assumption of a constant observational variance V_t is unrealistic when looking at traffic flow data. An example of flow heterocedasticity was shown in Section 2.4 of Chapter 2, where boxplots of flows from the Manchester network were shown.

Some approaches to deal with time-varying variance flows can be found in Kamarianakis *et al.* (2005) and Tsekeris and Stathopoulos (2006), in which traffic forecasts with a measure of their uncertainty are obtained through the combination of ARIMA models (described in Section 3.3 of Chapter 3) with a variance model called (generalized) autoregressive conditional heteroscedasticity models ((G)ARCH, Chatfield, 2003). However, as shown in Section 3.7 of Chapter 3, ARIMA models might not

be a reasonable choice for real-time multivariate flow forecasting, given the requirements of an on-line traffic management environment.

In order to accommodate the assumption of non-constant observation variances in the LMDM while being able to use the simple structure of the standard DLM for conditional univariate distributions, some model extensions described by West and Harrison (1997) can be applied in the context of the LMDM. This section shows a description of these extensions as well as their applications in traffic modelling.

5.2.1 Variance laws for V_t

It was suggested in Section 4.8.1 of Chapter 4 to extend the LMDM to accommodate non-normal data by using a variance law, such that the mean is related to the variance. This variance law would also accommodate a non-constant $V_t(i)$. Since the LMDM uses simple normal DLMs for each $Y_t(i)|\text{pa}(Y_t(i))$, it can easily incorporate a variance law into each of its conditional separate DLMs using extensions described in West and Harrison (1997), thus producing a novel approach for accommodating non-normal data and non-constant $V_t(i)$ in multivariate state space models.

Suppose an LMDM for a time series $\boldsymbol{Y}_t$ is given by observation and system equations (5.1) and (5.2), respectively. In a variance law model, the observation variance $V_t(i)$ in the observation equation (5.1) is (West and Harrison, 1997)

$$V_t(i) = k(\mu_t(i))V(i),\ i = 1, \ldots, n, \tag{5.3}$$

where $\mu_t(i)$ and $V(i)$ are the underlying level and observation variance, respectively, of the series $Y_t(i)$, $i = 1, \ldots, n$. The function $k(\mu_t(i))$ represents the change in observation variance associated with $\mu_t(i)$, which depends on the context and nature

of the data. Some context-specific variance law examples can be found in Stevens (1974), Boylan and Johnston (1996) and Nobre *et al.* (2001).

Figure 5.1 shows scatterplots of log mean versus log variance of flows at site 9206B from the Manchester network for two periods: 19:00-06:59 and 07:00-18:59. Both plots show (different) roughly linear relationships between log mean and log variance at this site. Similar relationships can also be observed at other sites. These empirical relationships suggest a traffic variance law model of the form

$$\log(\mathrm{Var}(Y_t(i))) = \beta(i)\log(\mu_t(i)), \tag{5.4}$$

where $\beta(i)$ is a parameter representing the slope of the relationship between log mean and log variance of flows at a given site i.

Rather than determining precise values for $k(\mu_t(i))$, more important is that $k(\mu_t(i))$ "changes markedly as the level of the series changes markedly" (West and Harrison, 1997). The empirical flow mean-variance relationship shown in Figure 5.1 therefore suggests modelling the change in observation variance associated with the level $\mu_t(i)$ by

$$k(\mu_t(i)) = \exp(\beta(i)\log(\mu_t(i))),$$

with different $\beta(i)$ values for the two periods 19:00-06:59 and 07:00-18:59. An alternative would be to have an intercept parameter in (5.4). However, based on the application to be described in Section 5.2.3, this was not found to improve flow forecasts.

The parameter $\mu_t(i)$ in equation (5.3) is the unknown mean of $Y_t(i)$, $i = 1, \ldots, n$. When considering a similar variance modelling issue in DLMs in the related application of road safety research, Bijleveld *et al.* (2010) use the observations themselves as proxies for the unknown mean. Since the focus here is on traffic flow forecasting, $\mu_t(i)$ is estimated by its forecast, denoted $f_t(i)$, which can be obtained from the

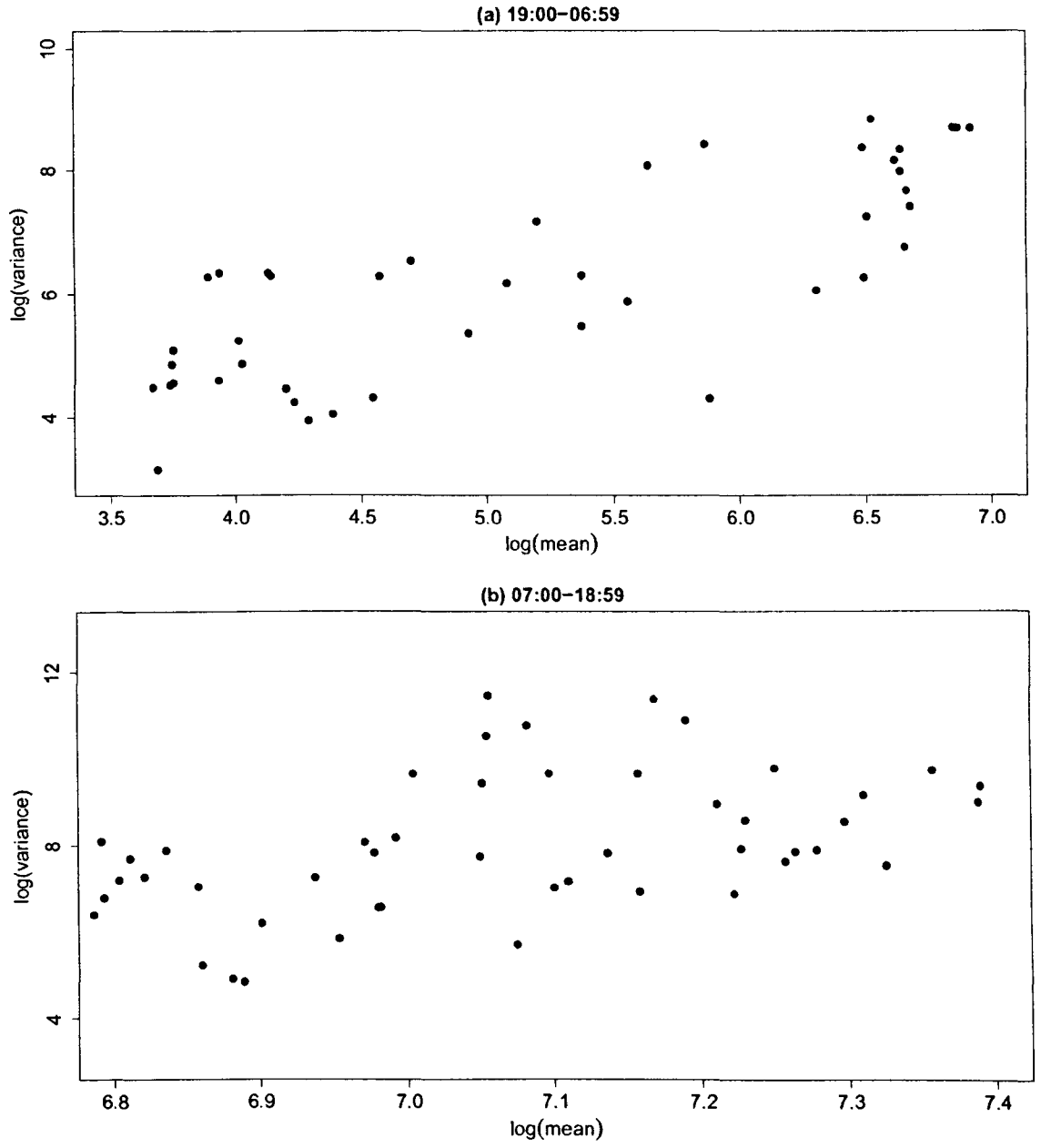

FIGURE 5.1: Flow mean versus flow variance (log scale, calculated using all Wednesdays in 2010) at site 9206B: (a) the 48 15-minute periods during 19:00–06:59 and (b) the 48 15-minute periods during 07:00–18:59 (plots on different scales).

DLM sequential estimation and forecasting procedure, described in Subsection 3.6.1 of Chapter 3. This motivates a variance law in which $V_t(i)$ in (5.1) is replaced by

$$V_t(i) = \exp(\beta(i)\log(f_t(i)))V(i). \tag{5.5}$$

The underlying observation variance $V(i)$ can be estimated on-line dynamically as data are observed using usual variance learning techniques, as described in Subsection 3.6.1 of Chapter 3. In addition, the parameter $\beta(i)$ can be estimated from flow means and variances using historical data, with different $\beta(i)$ values for the two periods 19:00–06:59 and 07:00–18:59.

5.2.2 Discount factors for the observational variance V_t

The LMDM can accommodate non-constant observational variances not only through variance laws, but also by considering discount factors, in an approach similar to one usually adopted to model evolution variances in a DLM, described in Subsection 3.6.1 of Chapter 3. The use of discount factors for V_t was firstly proposed in Ameen and Harrison (1985), and since then it has been extensively applied in both univariate and multivariate DLMs as shown in West and Harrison (1997) and Prado and West (2010).

Consider again an LMDM for a time series $\boldsymbol{Y}_t$, where a DLM $\{\boldsymbol{F}_t(i), \boldsymbol{G}_t(i), V_t(i), \boldsymbol{W}_t(i)\}$ is assumed for each $Y_t(i)|\text{pa}(Y_t(i))$, as defined in Section 3.6.1 of Chapter 3. In this model, inference is done through a conjugate analysis for $(\boldsymbol{\theta}_t(i), \phi(i))$, based on a multivariate normal/gamma distribution with $\phi(i) = V^{-1}(i)$ (as shown in Subsection 3.6.1 of Chapter 3). Now, assume the precision $\phi(i)$ in the DLM for $Y_t(i)|\text{pa}(Y_t(i))$ can change over time so that given the posterior

$$\phi_{t-1}(i)|D_{t-1} \sim \text{Gamma}\,(a_{t-1}, b_{t-1})\,, \tag{5.6}$$

the prior for $\phi_t(i)$ is

$$\phi_t(i)|D_{t-1} \sim \text{Gamma}\,(\delta a_{t-1}, \delta b_{t-1})\,, \tag{5.7}$$

for some $\delta \in (0, 1]$. With this assumption, while the prior mean for $\phi_t(i)$ is the same as the posterior mean for $\phi_{t-1}(i)$, the prior variance for $\phi_t(i)$ is larger than the posterior variance for $\phi_{t-1}(i)$. This represents an increase in uncertainty about $\phi_t(i)$, when compared to $\phi_{t-1}(i)$, after observing $y_{t-1}(i)$. Smaller values of δ increase this uncertainty more than larger values do. Hence, different uncertainty levels about $\phi_t(i)$ can be assumed by considering different δ values, similar to the case when discount factors are used to model unknown evolution variances $\boldsymbol{W}_t(i)$ (again, as described in Subsection 3.6.1 of Chapter 3). That is, smaller values of δ are suitable when the observation variance is unstable over time, whereas larger δ values are suitable when the observation variance is more static.

Following these ideas, the updating equations presented in Section 3.6.1 of Chapter 3 are easily adapted. Hence, given an DLM $\{\boldsymbol{F}_t(i), \boldsymbol{G}_t(i), V_t(i), \boldsymbol{W}_t(i)\}$, after $y_t(i)$ is observed, the beliefs about $(\boldsymbol{\theta}_{t-1}(i), \phi_t(i))$ are updated such that the parameters $n_t(i)$ and $S_t(i)$ are now

$$n_t(i) = \delta n_{t-1}(i) + 1 \qquad \text{and} \qquad S_t(i) = \delta S_{t-1}(i) + \frac{S_{t-1}(i)}{n_t(i)}\left[\frac{e_t(i)^2}{Q_t(i)} - 1\right],$$

where δ is the discount factor for the time-varying observational variance $V_t(i)$. This idea can also be combined with the variance law so that, in (5.5), $V(i)$ can also evolve dynamically.

A discount factor can also be used in matrix normal DLMs to assume a time-varying observation covariance matrix $\boldsymbol{\Sigma}_t$. As described in Section 3.6.2 of Chapter 3, in a matrix normal DLM for a multivariate time series vector $\boldsymbol{Y}_t$, conjugate analysis for $(\boldsymbol{\Theta}_t, \boldsymbol{\Sigma})$ is done through a matrix normal/inverse Wishart distribution. By using a similar argument considered for the univariate case, a time-varying $\boldsymbol{\Sigma}_t$ can be assumed by using a discount factor $\delta \in (0, 1]$. In this multivariate case, the updating equations for estimation and forecasting in a matrix normal DLM (presented in 3.6.2

of Chapter 3) can be also easily adapted. Further details appear in Quintana (1985, 1987) as well as in West and Harrison (1997).

5.2.3 Example

To verify the effect of considering time-varying observational variances when forecasting traffic flows, LMDMs were fitted using 15-min flows from a subset of data collection sites of the Manchester network. Given the DAG for this network, presented in Subsection 4.8.1 of Chapter 4, let $\boldsymbol{Y}_t$ be a multivariate time series formed by all the four root nodes $Y_t(9206B)$, $Y_t(6013B)$, $Y_t(9188A)$ and $Y_t(1431A)$, as well as one of each of their respective children, $Y_t(9200B)$, $Y_t(6007L)$, $Y_t(9193J)$ and $Y_t(1437A)$. For notational convenience, redefine

$$Y_t(9206B) = Y_t(1),\ Y_t(6013B) = Y_t(2),\ Y_t(9188A) = Y_t(3),\ Y_t(1431A) = Y_t(4),$$
$$Y_t(9200B) = Y_t(5),\ Y_t(6007L) = Y_t(6),\ Y_t(9193J) = Y_t(7),\ Y_t(1437A) = Y_t(8),$$

and set $\boldsymbol{Y}_t = [Y_t(1), \ldots, Y_t(8)]^\top$.

Still based on the DAG for the Manchester network, $\boldsymbol{Y}_t$ can be modelled through an LMDM with observation and system equations

$$Y_t(i) = \boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i) + v_t(i),\ v_t(i) \sim N(0, V_t(i)).\ i = 1, \ldots, 8, \qquad (5.8)$$

and

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim N(0, \boldsymbol{W}_t), \qquad (5.9)$$

where $\boldsymbol{\theta}_t = [\boldsymbol{\theta}_t(1)^\top \ldots \boldsymbol{\theta}_t(8)^\top]$. In this LMDM, the daily cycle observed in flows is modelled using a seasonal factor representation as considered in Sections 4.7 and 4.8 of Chapter 4. Hence, for root nodes, each $\boldsymbol{\theta}_t(i)$ in (5.8) and (5.9) is a 96-dimensional

vector of mean flow level parameters when using 15-min traffic data (one mean flow level parameter for each 15-minute period in the day) with corresponding 96-dimensional vector

$$\boldsymbol{F}_t(i)^\top = (1\ 0\ \ldots\ 0), \quad i = 1, 2, 3, 4.$$

The child series $Y_t(i)$, $i = 5, 6, 7, 8$, have their parents as linear regressors, where the regression coefficients represent the proportions of vehicles flowing from parent to child. As shown in Subsection 4.8.2 of Chapter 4, these proportions also exhibit a daily pattern, being then also modelled by a seasonal factor representation as considered in Sections 4.7 and 4.8 of Chapter 4. Therefore, for children of root nodes, each $\boldsymbol{\theta}_t(i)$ in (5.8) and (5.9) is a 96-dimensional vector of proportion parameters with corresponding 96-dimensional vector

$$\boldsymbol{F}_t(i)^\top = (\mathrm{pa}(y_t(i))\ 0\ \ldots\ 0), \quad i = 5, 6, 7, 8.$$

Additionally, in equation (5.9), we have that

$$\boldsymbol{G}_t = \mathrm{blockdiag}(\boldsymbol{G}(1), \ldots, \boldsymbol{G}(8)),$$

with 96×96 evolution matrices

$$\boldsymbol{G}(i) = \begin{pmatrix} a & 1-a & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & & & \ddots & 1 \\ 1 & 0 & \ldots & \ldots & 0 \end{pmatrix},$$

The matrix $\boldsymbol{G}_t$ in (5.9) then 'cycles' through the mean level parameters to ensure that the correct 15-min mean flow level parameter is used at time t. Similar to what was considered in the example described in Section 4.7 of Chapter 4, the scalar a

in $G(i)$, $i = 1,\ldots,8$, is included to allow a small effect of the parameter for the corresponding 15-min interval from the previous day. The optimal value of a in this example was found to be 0.01.

With this LMDM, each root node

$$Y_t(9206B),\ Y_t(6013B),\ Y_t(9188A) \text{ and } Y_t(1431A)$$

follows a univariate DLM, and univariate DLMs are also used to model

$$Y_t(9200B)|Y_t(9206B),\ Y_t(6004L)|Y_t(6013B),$$

$$Y_t(9193J)|Y_t(9188A) \text{ and } Y_t(1437A)|Y_t(1431A).$$

Four LMDMs based on equations (5.8) and (5.9) were used for forecast $\boldsymbol{Y}_t$. These four models are described as follows.

- Model A assumes a constant $V(i)$ and uses variance learning techniques (described in Subsection 3.6.1 of Chapter 3) to estimate $V(i)$ on-line dynamically as data are observed. This was the modelling approach used so far to forecast flows using the LMDM (as in Queen *et al.*, 2007, and Queen and Albers, 2009);
- Model B assumes a time-varying $V_t(i)$ using the variance law (5.5) with a dynamically evolving underlying variance $V(i)$ as in (5.6) and (5.7);
- Model C assumes a time-varying $V_t(i)$ using the variance law (5.5) with a dynamically evolving underlying variance $V(i)$ as in (5.6) and (5.7) for period 19:00-06:59, while using a dynamically evolving underlying variance $V(i)$ as in (5.6) and (5.7) but no variance law for period 07:00–18:59 (because of the weaker mean-variance relationship in this period, as shown in Figure 5.1);
- Model D assumes a time-varying $V_t(i)$ using the variance law (5.5) only.

Historical data from February to April 2010 were used to estimate the two values of β in (5.5) for the two periods 19:00-06:59 and 07:00-18:59, and were also used, in the absence of expert information, to elicit priors. On-line one-step ahead forecasts were then obtained for Wednesday flows in May and June 2010.

For models B and C, a value for δ to assume a time-varying $V_t(i)$ is needed for each series. Following West and Harrison (1997), the LPL (described in Section 4.6 of Chapter 4) could be used as an informal guide to choosing the δ which gives the best forecast performance for these data. However, since it was found that the LPL can be sensitive to outliers, δ was chosen to minimise the mean interval score (MIS, also described in Section 4.6 of Chapter 4). Furthermore, Models A-D all have the same forecast means and so only an assessment of the forecast limits is required.

As illustration of the parent and child forecast performance using Models A–D, Table 5.1 shows the values of LPL and MIS when forecasting the 4 parent root nodes together with their associated (non logical variable) children. The LPL and MIS quoted in Table 5.1 for each series for Models B and C, are those obtained when using the δ which minimised the MIS for that series and model.

Although Model A performs the best in terms of LPL for the first pair of series in Table 5.1, Model B performs the best in terms of MIS for these series, and in all other cases, the best performing model is Model B, which uses the variance law and also allows the underlying variance $V(i)$ to evolve dynamically.

TABLE 5.1: LPL and MIS for forecasting using Models A–D.

Series	LPL				MIS			
	A	B	C	D	A	B	C	D
$(Y_t(9206B), Y_t(9200B))$	$-10,001$	$-10,040$	$-10,230$	$-10,266$	691	498	541	635
$(Y_t(9188A), Y_t(9193J))$	$-8,010$	$-7,710$	$-7,852$	$-8,394$	407	294	336	396
$(Y_t(1431A), Y_t(1437A))$	$-9,615$	$-9,077$	$-9,140$	$-9,158$	595	414	453	487
$(Y_t(6013B), Y_t(6007L))$	$-9,137$	$-8,466$	$-8,724$	$-9,157$	441	272	347	385

As another illustration of the forecast performance, Figure 5.2 shows the observed flows on a specific day for root node 1431A and its child 1437A, together with their one-step ahead forecast means $f_t(i)$ and one-step ahead forecast limits defined as $f_t(i) \pm 2\sqrt{\text{Var}(Y_t(i)|D_{t-1})}$. The forecasts were calculated considering Models A and B, since Model B performs the best amongst the time-varying models. The effect of the variance law and dynamically evolving underlying variance is clearly visible at both sites: the range of the forecast limits given by Model B is more variable than the range given by Model A.

Note that there are some flows observed during the morning and afternoon peak periods that lie outside the forecast limits based on Model A but lie *inside* the forecast limits provided by Model B. As time t increases, in a variance law model, the observation variance estimate, $\widehat{V}_t(i)$, has the form of an exponentially weighted moving average of the forecast error (West and Harrison, 1997, p. 363), so that the most recent forecast error has a larger weight than the forecast errors observed in the past. This results in scaling the variance of the forecast distributions from Model B by $\widehat{V}_t(i)$. Model B therefore adapts more quickly to correct for large forecast errors than Model A does. This means that a variance law model automatically increases uncertainty in the forecasts, which can be useful when intervention may be required but expert information is not available.

5.3 Forecast limits in the LMDM

5.3.1 Approximate and simulation-based LMDM forecast limits

When considering plots of forecasts together with the observed values, it is common to include an indication of the uncertainty associated with the forecasts. This was

done in the example described in Subsection 5.2.3 by considering the forecast limits as the marginal forecast mean $\pm(2 \times$ marginal forecast standard deviation). The uncertainty of the forecasts are often represented by forecast limits calculated in this way.

For normally distributed forecast distributions, roughly 95% of observations should lie within these forecast limits and the forecast limits are approximately 95% (equal-tailed) prediction intervals. However, the marginal forecast distributions in the LMDM are not normal and, furthermore, they cannot usually be calculated analytically.

Even though recent advances in MCMC and sequential Monte Carlo techniques can simulate estimates of the true forecast limits in real-time, the approximation based on marginal forecast moments illustrated in Section 4.7 of Chapter 4 and used in Subsection 5.2.3 is far simpler and faster. However, if the forecast limits are calculated using the marginal forecast moments in the LMDM, one question that remains is how close this approximation is to the true 95% forecast limits.

To answer this question, consider once again the forecast limits of site 1437A obtained by this approximation (as shown in Figure 5.2(b)). The 'true' 95% forecast limits of the marginal forecast distributions for site 1437A would be the 2.5% and 97.5% percentiles of the marginal forecast distributions. These can be estimated at each time t via simulation: simulate samples from the marginal forecast distributions by simulating the joint forecast distribution of parent Y_t(1431A) and child Y_t(1437A) via the normal forecast distribution for Y_t(1431A) and the conditional normal forecast distribution for Y_t(1437A)$|Y_t$(1431A).

Figure 5.3 shows the approximate forecast limits for site 1437A based on marginal moments, together with the estimated 'true' forecast limits based on simulation. The plot shows the same day as was considered in Figure 5.2 in which there were

some unusual traffic flows which created a high level of flow uncertainty. As can be seen, even when there is a lot of forecast uncertainty, the forecast limits based on marginal moments are in fact close to the simulated true limits — certainly a good enough approximation given their ease and speed of calculation. When considering all flow series considered in Table 5.1, forecast intervals provided by both models also have similar MIS.

5.3.2 Coverage of LMDM forecast limits

In Figure 5.2, the forecast limits are quite wide at times and most observations lie within them. However, for a well-calibrated model, approximately only 95% of observations should lie within the forecast limits. Over the whole forecast period, Model B actually is well-calibrated for the root nodes with roughly 95% of observations lying within the forecast limits for each series: the wide forecast limits in Figure 5.2(a) are a result of increased forecast uncertainty due to unexpected observations on that particular day. On the other hand, for each root node, Model A underestimates the forecast uncertainty with a coverage of roughly only 89%.

When forecasting child variables, however, Model B overestimates the forecast uncertainty with roughly 98% of observations falling within the forecast limits for each series, while this time Model A is well-calibrated with a coverage of roughly 95%. This suggests that, for child variables, there are factors affecting the variation that are not accounted for in Model B. One possible element missing from Model B is the use of data for other traffic variables affecting flows. Indeed, it will be shown in Chapter 6 how the LMDM can accommodate these extra variables as predictors thus providing better calibrated forecast limits.

5.4 Accommodating measurement errors in the LMDM

5.4.1 Measurement errors

When building DAGs and LMDMs for forks and joins in Subsection 4.8.1 of Chapter 4, $Y_t(3)$ and $Y_t(6)$ were both modelled as logical variables without errors. However, as is common for data in a variety of applications, loop detector data are prone to measurement errors due to device malfunctions (as discussed in Chen *et al.*, 2003 and Bickel *et al.*, 2007) so that modelling $Y_t(3)$ and $Y_t(6)$ as logical variables may not be a realistic assumption in practice.

To illustrate, consider the fork consisting of sites 1431A, 1437A and 6002A in Figure 2.2 of Section 2.3 in Chapter 2. As noted in Section 4.8 of Chapter 4, it would be unrealistic to expect $Y_t(6002A)$ to be exactly equal to $Y_t(1431A) - Y_t(1437A)$ because of time-lag effects. However, when examining the errors $Y_t(1431A) - (Y_t(1437A) + Y_t(6002A))$, it is apparent that not only time lag effects affect these errors but also other factors which can be associated with loop detector malfunctions.

Figure 5.4 shows a histogram and q-q plot of these errors observed in the period 21:00-22:59 during 2010 with 5% of the extreme errors excluded from the plot. The most extreme errors were removed because these would be dealt with using intervention to maintain forecast performance. The inclusion of such extreme errors in the plot therefore gives an unrealistic picture of the measurement errors. From the histogram in Figure 5.4(a), it is clear that the errors are nearly all positive with some significant variability, while the q-q plot in Figure 5.4(b), suggests that an assumption of normally distributed measurement error seems reasonable for 95% of the data and is worth considering as a simple model. The normal distribution can

therefore form the basis for a simple model to accommodate measurement errors illustrated in Figure 5.4.

5.4.2 Accommodating measurement error

Consider again the fork of Figure 4.9(a) in Subsection 4.8.1 of Chapter 4. In Section 4.8 of Chapter 4, the layout of the sites and direction of traffic flow suggested that the model for $Y_t(3)|Y_t(1), Y_t(2)$ could be simply $Y_t(3) = y_t(1) - y_t(2)$. However, the histogram of Figure 5.4(a) suggests an alternative model which accommodates measurement error of the form

$$Y_t(3) = (y_t(1) - y_t(2))\theta_t(3)^{(1)} + \theta_t(3)^{(2)} + v_t(3), \tag{5.10}$$

where $\theta_t(3)^{(2)}$ is the level of the measurement error and $v_t(3) \sim N(0, V_t(3))$, for some $V_t(3)$. As vehicles from $S(1)$ can only go to $S(2)$ or $S(3)$, it is reasonable to assume a prior for $\theta_t(3)^{(1)}$ with mean 1 and very small variance. Note that the measurement errors at $S(1)$ and $S(2)$ are taken into account automatically through the model parameters and observation variances $V_t(1)$ and $V_t(2)$. The DAG representing this new model is the same as in Figure 4.10(a) in Subsection 4.8.1 of Chapter 4 except that the double oval (representing a logical variable) is now an ordinary single oval.

The distribution of the errors in the Manchester network actually differs with the time of day, as illustrated in Figure 5.5. Moreover, the mean of the error follows the usual pattern of the flow observed during the day, where the plots of flow time series presented in Section 2.4 of Chapter 2 is an example. To account for this, a seasonal factor model can be used for $\theta_t(3)^{(2)}$ in the same way as for modelling the diurnal cycle of $\mu_t(i)$ in Section 4.8.2 of Chapter 4 . Figure 5.5 also shows the error variability changing through the day. In fact for the Manchester network, as with the flows themselves, there is a roughly linear relationship between the logs of the

means and variances of the errors during periods 19:00–06:59 and 07:00–18:59. The variability of $V_t(3)$ can then be accommodated by using a variance law LMDM as in (5.5), combined with a dynamically evolving underlying variance $V(3)$ as in (5.6) and (5.7).

An analogous model can be defined to allow for measurement error in a join.

5.4.3 Forecast performance

Model (5.10) and the logical model without an error term (described in Subsection 4.8.1 of Chapter 4) were used to obtain one-step ahead forecasts for the four children of root nodes considered as logical variables in the DAG for the Manchester network presented in Subsection 4.8.1 of Chapter 4, namely Y_t(6002A), Y_t(1445B), Y_t(6002B) and Y_t(9195A). A variance law and dynamically evolving observation variance as described in Section 5.2 were used with each model. As in the previous section, historical data from February to April 2010 were used to estimate the β parameters for the variance law model and for eliciting priors, while on-line one-step ahead forecasts were obtained for Wednesday flows in May and June 2010.

The MedianSE for each series when using these two models is shown in columns 2 and 3 of Table 5.2. Neither the LPL nor the MIS are appropriate for model comparison here: the LPL cannot be calculated for the model without an error and the MIS is not appropriate because the error model naturally has wider forecast limits. Table 5.2 also shows for each series (in columns 4 and 5) the means and standard deviations of the relative measurement errors (that is, 100 × (observed measurement error at time t)/$Y_t(i)$).

As can be seen in Table 5.2, the error model performs significantly better than the logical model in terms of MedianSE for two of these series and slightly worse for the other two series. Notice that the series which show the greatest improvement

TABLE 5.2: MedianSE for the error model (5.10) and logical model without an error term, together with the means and standard deviations of the relative measurement errors.

	MedianSE		Relative measurement errors	
Series	Error model	Logical model	Mean	Standard devation
$Y_t(6002A)$	142	882	31.2	27.6
$Y_t(1445B)$	969	1211	9.0	59.8
$Y_t(6002B)$	180	159	−1.2	8.1
$Y_t(9195A)$	618	616	0.4	3.3

in using the error model in comparison to the logical model are those for which the relative measurement errors are high. However, although the error model gives greater improvement in forecast performance when the relative measurement errors are high, high relative measurement errors also mean an increase in the uncertainty of the resulting forecasts which, in turn, means that forecast limits are wider for series with high relative measurement error than for series with low relative measurement errors. Although the choice of which of two children at a fork should be considered to be the logical variable is arbitrary, the relative measurement errors for each of the children should be considered when making a decision.

As with the time-varying variance model of Section 5.2, the forecast limits for each (child) series in Table 5.2 overestimate the forecast uncertainty, with a coverage of roughly 97% for each series when using the logical model, and roughly 99% for each series when using the error model. Again, as discussed in Section, 5.3 this is indicative that there are factors (such as extra traffic variables, possibly) affecting the variability which are not captured by the model.

5.5 Discussion

This chapter proposed a methodology to allow for time-varying observation variances in the LMDM when forecasting traffic flows. However, the assumption of constant

variance may also not be reasonable when using the LMDM for other applications. As an example, the flow variability of goods to be distributed over a chain of supermarkets can be affected by seasonal effects due to holidays and seasons of the year. These seasonal effects can also, for example, be responsible for non-constant variability of electricity flow distribution to residential areas. The use of discount factors for observation variances can be also applied when using LMDMs in these applications, and variance laws can also be developed by identifying variance relationships given the context-specific knowledge and analysis of available data.

In addition, the normal model developed in Section 5.4 for incorporating the measurement error at nodes previously considered as logical variables in the LMDM is only a simple model and other distributions may be more appropriate. As an example, a mixture of distributions may work well. However, as discussed in Chapters 2 and 3, traveller information systems and some traffic management systems require real-time forecasts, and so the computational costs of considering alternative approaches for error modelling must be carefully taken into account.

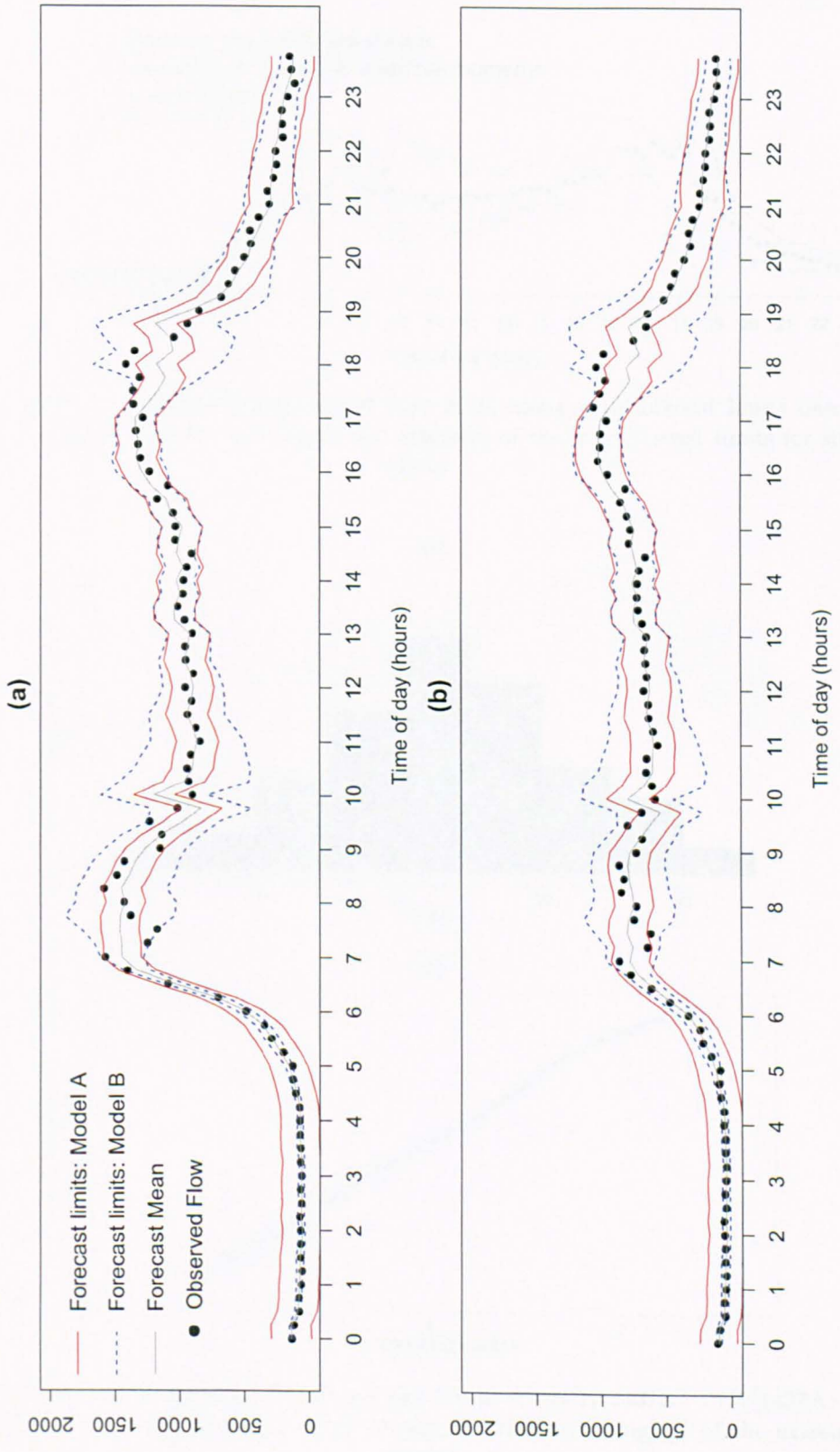

FIGURE 5.2: Observed flows on 19 May 2010, along with forecast means and forecast limits based on Models A (red) and B (blue) sites (a) 1431A and (b) 1437A.

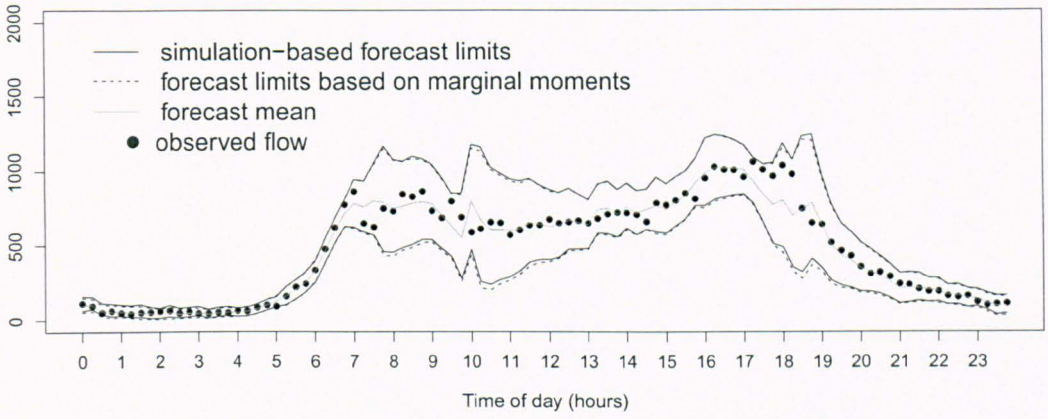

FIGURE 5.3: Observed flows on 19 May 2010, along with forecast limits based on marginal moments and simulated estimates of the true forecast limits for site 1437A.

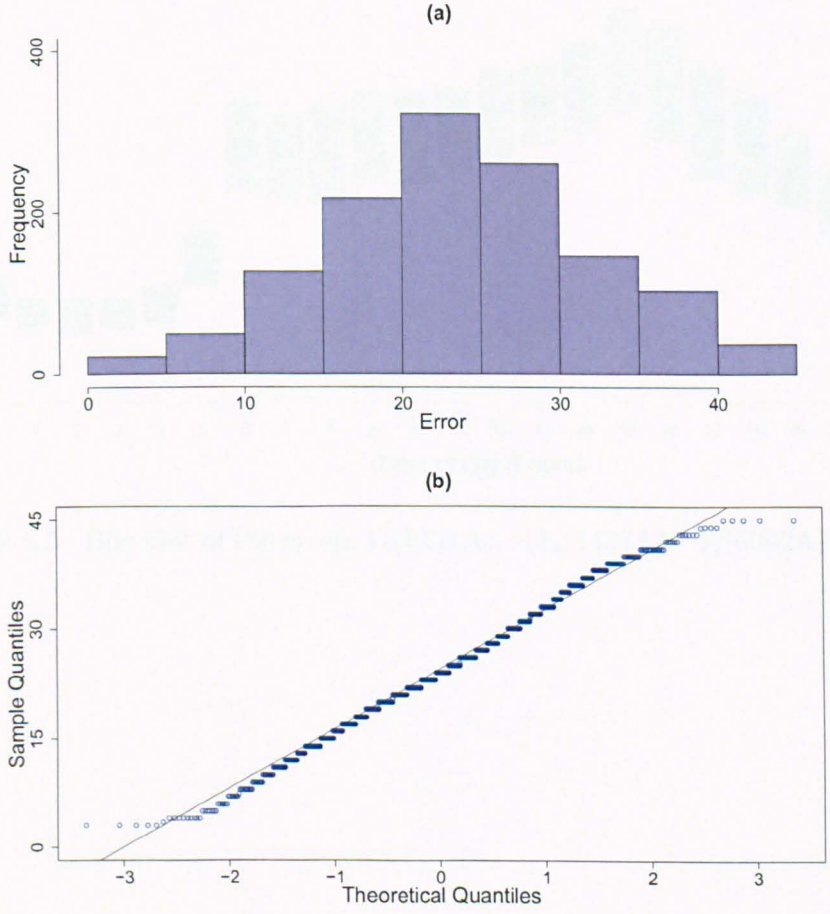

FIGURE 5.4: Histogram (a) and q-q plot (b) of errors $Y_t(1431\text{A}) - (Y_t(1437\text{A}) + Y_t(6002\text{A}))$ in the period 21:00-22:59 during 2010 (excluding 5% of the extreme errors).

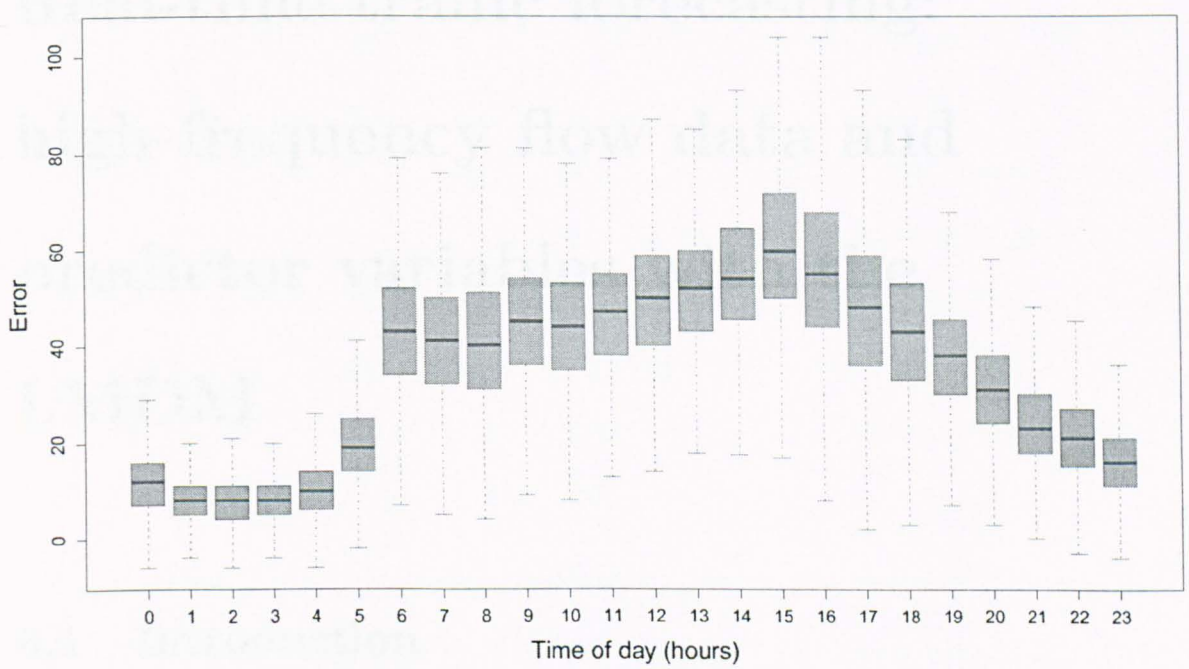

FIGURE 5.5: Box plot of the errors $Y_t(1431\text{A}) - (Y_t(1437\text{A}) + Y_t(6002\text{A}))$ in 2010.

Chapter 6

Real-time traffic forecasting: high-frequency flow data and predictor variables with the LMDM

6.1 Introduction

This chapter shows LMDM extensions based on cubic splines to accommodate two important issues in traffic flow modelling. The problem of modelling flows using high-frequency flow data suitable for on-line traffic management purposes is addressed. In addition, methodology for including extra traffic variables as predictors in the LMDM is presented.

6.2 Traffic data aggregation

Like in any data analysis task, preliminary data processing is an important step for the development of traffic flow models. This involves data aggregation, which relates to the time interval adopted to build the time series of flows and other available variables from the road traffic network under study.

Data aggregation is extensively discussed in the traffic modelling literature. This is an important issue because different traffic features can be observed depending on the aggregation level. If a long time interval (such as, say, 24 hours) is used to aggregate flows, the resulting forecasting model will not capture traffic patterns of interest for real-time applications, where morning and afternoon peak times are the periods in which traffic flow models are of their most use. On the other hand, the smaller the time interval for aggregation, the more variability there is in traffic flows and the more difficult is to develop accurate forecasting models. As an example, Abdulhai *et al.* (1999) studied how the accuracy of flow forecasting models decreases as flow time series are aggregated into smaller time intervals.

Although different time intervals for traffic flow aggregation can be suitable depending on the purpose of application of the forecasting models (Vlahogianni *et al.*, 2004), 15-min intervals seem to be the common choice (Vlahogianni and Karlaftis, 2011 and Chang *et al.*, 2012). This time interval is also recommended in the *Highways Capacity Manual* (2010) to assess the quality of road networks, such as evaluation of their capacity. However, as discussed by Vlahogianni and Karlaftis (2011) and Hurdle *et al.* (1997), flows may significantly change during a 15-min period.

For real-time traffic control systems, where the Managed Motorways project described in Chapter 2 is an example, time intervals smaller than 15-min may be suitable to support immediate actions to alleviate congestion. What must therefore be considered are traffic time series on higher frequencies (as suggested by Vlahogianni

et al., 2004). However, the time interval must be carefully chosen in order to avoid modelling high-frequency variations which may not be of interest for traffic management purposes. Indeed, Huang and Sadek (2009) pointed out that forecasting models using 1-min traffic data are not particularly useful from a practical viewpoint. In this context, data aggregation using 5-min intervals has been commonly used for real-time traffic modelling (see, for example, Castro Neto *et al.*, 2009 and Min and Wynter, 2011).

The LMDM has so far been previously applied to forecast hourly traffic flows (as in Queen *et al.*, 2007 and Queen and Albers, 2009), whereas 15-min flows were considered for the LMDM extensions described in Chapter 5. Given the crucial role of flow aggregation based on higher frequencies for on-line traffic management systems, the next section shows how the LMDM can be adapted to cope with 5-min data. Data available from the Manchester network are suitable for these developments, as loop detectors in this motorway collect traffic variables on a minute-by-minute basis.

6.3 Modelling the daily flow cycle

Chapter 5 showed how the LMDM can be extended in order to accommodate important traffic data features, such as the heteroscedasticity of flows over the day and measurement errors. The example described in Subsection 5.2.3 of Chapter 5 considered flows from the Manchester network aggregated into 15-min intervals. There, a seasonal factor representation was used to model the daily cycle in the LMDM, where a 96-dimensional state vector models the seasonality of each time series. This model setup implies that the state vector of each root node in the LMDM contains one mean flow level parameter for each 15-min period in the day. For the non-root nodes, the parameters in the state vector represents the proportion of traffic flowing from parent to child for each 15-min interval.

When the dimension of the state vector gets large, numerical problems can arise when updating the variances associated with a DLM (Prado and West, 2010). This can be tackled by using alternative equations in the Kalman filter algorithm, as described in Durbin and Koopman (2001) and implemented in the *R DLM* package (Petris *et al.*, 2009). However, when dealing with 5-min data, the dimension of the state vector is so large ($12 \times 24 = 288$) that it becomes important to consider alternatives to the seasonal factor representation to keep model parsimony. To address this problem, the daily cycles in each univariate DLM within the LMDM can be represented using cubic splines. This approach follows from Tebaldi *et al.* (2002), where splines were used to model 1-min traffic flow cycles within DLMs.

6.3.1 Cubic splines

Cubic splines are widely used in regression models in order to relax the linearity assumption for continuous regressors, as illustrated in Harrell (2001), Hastie *et al.* (2001) and Fox and Weisberg (2011). A cubic spline has the basic form:

$$f(x) = \sum_{m=1}^{M} \beta_m h_m(x), \tag{6.1}$$

where

$$h_m(x) = \begin{cases} x^m & m = 1, 2, 3, \\ (x - k_{m-3})^3 & m = 4, \ldots, M, \end{cases}$$

for values $k_1, ..., k_{M-3}$ with

$$a < k_1 < k_2 < ... < k_{M-3} < b,$$

where $[a, b] \in \mathbb{R}$ is the domain of x. When $x - k_{m-3}$ is negative, then $h_m(x) = 0$.

The functions $h_1(x), \ldots, h_M(x)$ are called *spline basis functions*, $k_1, ..., k_{M-3}$ are the *spline knots* and $\beta_1, \ldots, \beta_M$ are parameters. In the context of regression, the idea is to consider the spline basis functions as regressor variables and then estimate the parameters $\beta_1, \ldots, \beta_M$.

6.3.2 Cubic splines in the LMDM

Let $\boldsymbol{Y}_t = [Y_t(1), \ldots, Y_t(n)]^\top$ be an n-dimensional time series vector where the conditional independence structure across its components can be represented by a DAG, such that an LMDM can be defined for $\boldsymbol{Y}_t$. Recall the observation and system equations of an LMDM for $\boldsymbol{Y}_t$, described in Section 4.5 of Chapter 4 and given as

$$Y_t(i) = \boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i) + v_t(i), \; v_t(i) \sim N(0, V_t(i)) \; i = 1, \ldots, n, \tag{6.2}$$

and

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim N(0, \boldsymbol{W}_t), \tag{6.3}$$

respectively. The use of splines as described in Subsection 6.3.1 for (static) regression can be extended to LMDMs as follows. Consider first a root node. The daily cycle can be modelled in a time series using a spline to fit one full cycle. In this case, x would be time t and $k_1, \ldots, k_{M-3}$ would represent times over the cycle. For example, for 5-min data with a daily cycle over 24 hours, k_1 could be 13, for example, representing the time period 01:00–01:04 and t would be the current time (which at 02:00–02:04, say, would be $t = 25$). Prior data can be used to calculate the spline basis functions $h_1(x), \ldots, h_M(x)$ which can then be evaluated at each 5-min time period $x = t$. The regression vector $\boldsymbol{F}_t(i)$ for root node $Y_t(i)$ in (6.2) then has the

form:

$$\boldsymbol{F}_t(i)^\top = (h_1(t) \ \cdots \ h_M(t)). \tag{6.4}$$

Although this form of $\boldsymbol{F}_t(i)$ only models the daily cycle of flows, it is possible that $\boldsymbol{F}_t(i)$ has additional elements like exogenous regressors for each $Y_t(i)$'s DLM, as will be discussed in Section 6.4.

For $\boldsymbol{F}_t(i)$ in (6.4), the associated state vector in (6.2) is:

$$\boldsymbol{\theta}_t(i)^\top = (\beta_{t1} \ \cdots \beta_{tM}) \tag{6.5}$$

where $\beta_{t1}, \ldots, \beta_{tM}$ are dynamic versions of the associated parameters in (6.1), which evolve through the system equation (6.3) with state evolution matrix $\boldsymbol{G}_t(i)$ being the M-dimensional identity matrix.

As an example, suppose that the daily cycle of root node $Y_t(1)$ is to be represented by a spline with (for simplicity) just two knots, k_1 and k_2, and that an exogenous regressor, X_t, is also to be included in $Y_t(1)$'s DLM. Then the observation equation for $Y_t(1)$ has the form

$$Y_t(1) = \sum_{m=1}^{5} \beta_{tm} h_m(t) + \alpha_t x_t + v_t(1), \qquad v_t(1) \sim N(0, V_t(1)),$$

so that in (6.2),

$$\begin{aligned} \boldsymbol{F}_t(1)^\top &= (h_1(t) \ \cdots \ h_5(t) \ \ x_t) \\ \boldsymbol{\theta}_t(1)^\top &= (\beta_{t1} \ \cdots \ \beta_{t5} \ \ \alpha_t). \end{aligned}$$

In this case the evolution matrix $\boldsymbol{G}_t(1) = \text{blockdiag}(\boldsymbol{I}_5, g)$ where $\boldsymbol{I}_5$ is the 5-dimensional identity matrix and g is some scalar in $\mathbb{R}$ for parameter α_t's evolution.

A child in the LMDM is modelled as having its parents as linear regressors. For

example, if $Y_t(3)$ has parents $Y_t(1)$ and $Y_t(2)$, then the simplest observation equation for $Y_t(3)$ would be

$$Y_t(3) = \alpha_t(1)y_t(1) + \alpha_t(2)y_t(2) + v_t(3), \qquad v_t(3) \sim N(0, V_t(3))$$

so that $\boldsymbol{F}_t(3)^\top = (y_t(1) \;\; y_t(2))$ and $\boldsymbol{\theta}_t(3)^\top = (\alpha_t(1) \;\; \alpha_t(2))$. For traffic flow data, the regression parameters ($\alpha_t(1)$ and $\alpha_t(2)$ in the example above) exhibit a daily pattern (as in Subsection 4.8.2 of Chapter 4). A spline can be used to model the daily cycle by setting each regression parameter to the form $\sum_{m=1}^{M} \beta_{tm}h_m(t)$. Thus, in general the regression and state vectors $\boldsymbol{F}_t(i)$ and $\boldsymbol{\theta}_t(i)$ for child variable $Y_t(i)$ with (for simplicity) single parent $\text{pa}(Y_t(i))$ have the forms

$$\boldsymbol{F}_t(i)^\top = (\text{pa}(y_t(i))h_1(t) \cdots \text{pa}(y_t(i))h_M(t)), \tag{6.6}$$

and

$$\boldsymbol{\theta}_t(i)^\top = (\beta_{t1} \cdots \beta_{tM}). \tag{6.7}$$

Again, $\boldsymbol{\theta}_t(i)$ evolves through the system equation (6.3) with state evolution matrix $\boldsymbol{G}_t(i)$ being the M-dimensional identity matrix.

As another example, suppose now that $Y_t(3)$ has parents $Y_t(1)$ and $Y_t(2)$ and the daily cycles exhibited by $Y_t(1)$ and $Y_t(2)$'s regression parameters are to be represented by splines with three and two knots, respectively. Suppose further that an exogenous regressor, Z_t, is also to be included in $Y_t(3)$'s model. Then the observation equation for $Y_t(3)$ has the form

$$Y_t(3) = y_t(1)\sum_{m=1}^{6} \beta_{tm}^{(1)}h_m(t)^{(1)} + y_t(2)\sum_{m=1}^{5} \beta_{tm}^{(2)}h_m(t)^{(2)} + \gamma_t z_t + v_t(3), \qquad v_t(3) \sim N(0, V_t(3)),$$

so that in (6.2),

$$\begin{aligned} \boldsymbol{F}_t(3)^\top &= (h_1(t)^{(1)} \cdots h_6(t)^{(1)} \; h_1(t)^{(2)} \cdots h_5(t)^{(2)} \; z_t) \\ \boldsymbol{\theta}_t(3)^\top &= (\beta_{t1}^{(1)} \cdots \beta_{t6}^{(1)} \; \beta_{t1}^{(2)} \cdots \beta_{t5}^{(2)} \; \gamma_t). \end{aligned}$$

In this case the evolution matrix $\boldsymbol{G}_t(3) = \text{blockdiag}(\boldsymbol{I}_6, \boldsymbol{I}_5, g)$ where $\boldsymbol{I}_k$ is the k-dimensional identity matrix and g is some scalar in $\mathbb{R}$ for parameter γ_t's evolution.

To take advantage of the computational simplicity of a fully conjugate LMDM, normal priors need to be specified for the state vectors. When using the seasonal factor model for modelling the daily cycle exhibited by regression parameters in the child model, as described in Subsection 4.8.2 of Chapter 4, the regression parameters are proportions and so normal priors are not ideal. However, when using splines to model the regression parameters' daily cycles, using normal priors is not a problem as the spline regression parameters do not have any restrictions on their values.

6.3.3 Example

In order to compare the performance of using cubic splines and seasonal factors for modelling daily cycles in the LMDM, both models were used to forecast four separate bivariate series of the Manchester network. From the DAG described in Subsection 4.8 of Chapter 4, these four separate bivariate series was formed by considering the four root nodes of the Manchester network together with one of their children each: specifically, the four bivariate series considered were $(Y_t(9206B), Y_t(9200B))$, $(Y_t(6013B), Y_t(6007L))$, $(Y_t(9188A), Y_t(9193J))$ and $(Y_t(1431A), Y_t(1437A))$. To also assess whether dynamic estimation of the spline parameters $\beta_{t1}, \ldots, \beta_{tM}$ (as in (6.5) and (6.7)) improves forecast performance, forecasts for these four bivariate series were also obtained using a static version of the cubic spline LMDM, by using a system equation (6.3) with no error term $\boldsymbol{w}_t$.

Although Harrell (2001) suggests that the positions of knots $k_1, \ldots, k_{M-3}$ associated with the splines basis functions in (6.4) are not important when fitting splines for static regression purposes, it was found that LMDMs for traffic flows give better results when concentrating the positions of $k_1, \ldots, k_{M-3}$ during morning and afternoon peak periods. Harrell (2001) also recommends using only 3 to 5 knots for static regression. However, when using splines to represent daily cycles of flows, 15 to 20 knots were typically found to perform much better for all time series considered in this example. This is a small number when compared to the 288 parameters required to use a seasonal factor representation to model 5-min flow data. Moreover, overfitting is controlled because fitted splines were found to not vary very much over time and the parameters $\beta_{t1}, \ldots, \beta_{tM}$ evolve dynamically to capture any drift in time.

Observed flows from July to August 2010 were used to estimate the spline basis functions and form priors for all state vectors for these three models (so that each model had the same equivalent priors). Following the developments presented in Section 5.2 of Chapter 5, a time varying observation variance was considered by using variance laws and discount factors for $V_t(i)$ in equation (6.2) for each of the three defined models. Again, historical data from July to August 2010 was used to estimate parameters associated with the variance laws for each $V_t(i)$. One-step ahead forecasts for flows were then obtained for September 2010.

Each model and each series requires two separate discount factors to be specified: one for estimating $\phi_t(i) = V_t^{-1}(i)$ (that is, the precision of the observation variance $V_t(i)$ in equation (6.2)) and one for estimating $\boldsymbol{W}_t$. Usually these discount factors are chosen by comparing the forecast accuracy of different models varying discount factor values (as suggested by West and Harrison, 1997 and applied in Subsection 5.2.3 of Chapter 5). However, the high number of models and the level of complexity makes this optimization a demanding task. For example, the optimization of the combination of both discount factors for $\boldsymbol{W}_t$ and $\phi_t(i)$ for a child time series would

depend on the optimization of the discount factors for $\boldsymbol{W}_t$ and $\phi_t(i)$ in its parent. Due to this, based on preliminary tests, the chosen value for the discount factors for both $\boldsymbol{W}_t$ and $\phi_t(i)$ in all models in this example was 0.99.

Table 6.1 shows the LPLs for the three models for each bivariate series. The number of spline basis functions for static and dynamic spline LMDMs varied between 15 and 20 among the considered time series, and the LPL values shown in Table 6.1 are for the number of basis functions which performed best for each model and each series. In Table 6.1, the dynamic spline versions of the LMDMs produce the largest LPL values for all bivariate series, indicating that the dynamic spline LMDMs provide the most accurate forecasts.

TABLE 6.1: LPLs for LMDMs with different seasonal representations.

	LMDM		
Bivariate series	seasonal factors	static splines	dynamic splines
$Y_t(9206B), Y_t(9200B)$	-8,794	-8,857	-8,570
$Y_t(6013B), Y_t(6007L)$	-8,159	-8,157	-7,886
$Y_t(9188A), Y_t(9193J)$	-8,581	-8,647	-8,336
$Y_t(1431A), Y_t(1437A)$	-9,056	-8,839	-8,597

An alternative parsimonious approach to using cubic splines for modelling the daily cycle would be to use a Fourier representation, as in West and Harrison (1997), Section 8.6. The standard Fourier representation can be used directly for modelling the daily cycle exhibited by the parents, but the model would need to be adapted somewhat for modelling the daily cycle in the proportion regression parameters in the models for child variables. When modelling the four root nodes, however, the Fourier representation was found to perform worse than both the seasonal factor model and the splines, and so Fourier models were not pursued further.

6.4 Non-linear traffic predictor variables in the LMDM

As described in Chapter 2, the data collection process used by the Highways Agency in England includes minute-by-minute measurements of flow, together with occupancy, speed and headway. Section 2.5 of Chapter 2 particularly described the relationship between each of these extra variables and flow using the Manchester network data.

Although there is great interest and an extensive literature concerning traffic flow modelling, few models deal with the analysis of flow in conjunction with the other variables. Based on a survey carried out by Vlahogianni *et al.* (2004), of forty traffic models where flow was considered, just seven used other extra variables, while none considered all three. From a statistical perspective, Ahmed and Cook (1979) and Levin and Tsao (1980) fitted independent ARIMA models for flow and occupancy forecasting, while Whittaker *et al.* (1997) tackled a similar problem using state space models. Neural networks have also been used for modelling flow in conjunction with other variables, for example in Innamaa (2000), Abdulhai *et. al.* (1999) and Gilmore and Abe (1995). Multivariate forecasting of flow, speed and occupancy using k-nearest neighbour classifiers has also been considered by Clark (2003). More recently, Chandra and Al-Deek (2009) considered vector autoregressive models to forecast flows using speed as a predictor variable.

In the LMDM, exogenous variables can be easily introduced into the model as regressors (as X_t and Z_t were in the earlier examples). Figure 6.1 shows the scatterplots of flow at time t versus occupancy, speed and headway at previous time $t-1$ at site 9188A for three separate months (they were firstly presented in Section 2.5 of Chapter 2). These scatterplots suggest non-linear relationships between flow and all the possible predictor variables of interest. As previously mentioned, plots of

flow at t versus the other traffic variables at $t-1$ look broadly comparable at the other sites. Adopting a similar approach used when modelling the daily flow cycle described in Section 6.3, splines can be used to model the non-linear relationships between flow at time t and the exogenous variables at time $t-1$. These splines can then be incorporated into the LMDM as regressors.

For traffic control, it is preferable to include all three of the splines (for occupancy, speed and headway) as regressors for forecasting flows. This is because one predictor variable may be better for predicting possible changes in flow behaviour (such as congestion) than other predictors at one time, and a different variable may be better for predicting possible changes in flow at another time. Thus, although the model with all three predictors may not necessarily be the most parsimonious, it will be more responsive to traffic conditions and so, from a practical point of view, will be the most useful model for traffic control. Additionally, the fact that certain regressors give the best forecast performance on the data considered in this thesis is not a guarantee that the same regressors will give the best performance for *future* data. Therefore, the focus here is to present a model which uses all three predictors rather than searching for a subset of predictors which performs best for this particular dataset.

The developments presented in this section only considers using the values of the traffic variables at time $t-1$ for forecasting flows at time t. Different lags could be used, so that values of the traffic variables at time $t-k$, for $k>1$, could be used instead of, or in addition to, $t-1$. Whichever lags are used, splines can still be used to model the relationships between the traffic variables at $t-k$ and flow at t, and the same methods proposed in this section can then be used to incorporate these splines into the LMDM as regressors.

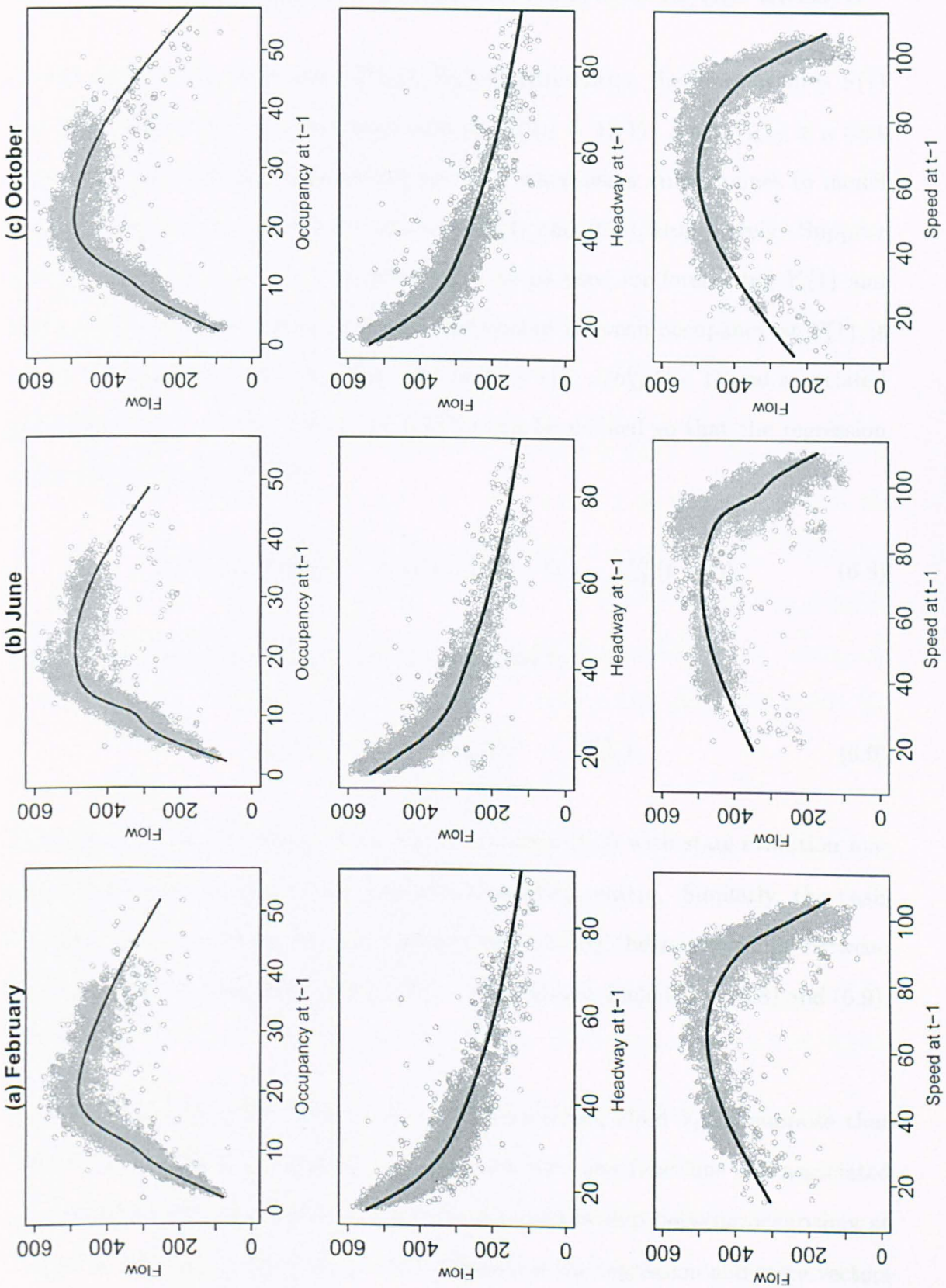

FIGURE 6.1: Scatterplots of flows at site 9188A at time t versus occupancy, headway and speed at $t-1$, in (a) February, (b) June and (c) October 2010.

6.4.1 Incorporating the predictor variables in the LMDM

Consider a bivariate time series $(Y_t(1), Y_t(2))$, representing the flows at sites $S(1)$ and $S(2)$, where $Y_t(1)$ is a root node and $\text{pa}(Y_t(2)) = Y_t(1)$. Since $Y_t(1)$ is a root node, the regression and state vectors for $Y_t(1)$ when using cubic splines to model the daily cycle in the LMDM are given by (6.4) and (6.5), respectively. Suppose that occupancy at time $t-1$ at site $S(1)$ is to be used for forecasting $Y_t(1)$ and that a cubic spline (6.1) represents the relationship between occupancy at $S(1)$ at time $t-1$ and $Y_t(1)$ with basis functions $h_1^{O_1}(t-1), \ldots, h_{M_1}^{O_1}(t-1)$ and associated parameters $\beta_{t1}^{O_1}, \ldots, \beta_{tM_1}^{O_1}$. Then, an LMDM can be defined so that the regression vector (6.4) is augmented to

$$\boldsymbol{F}_t(1)^\top = (h_1(t) \ \cdots \ h_M(t) \ \ h_1^{O_1}(t-1) \ \cdots \ h_{M_1}^{O_1}(t-1)) \tag{6.8}$$

and the associated state vector (6.5) is augmented to

$$\boldsymbol{\theta}_t(1)^\top = (\beta_{t1} \ \cdots \beta_{tM} \ \beta_{t1}^{O_1} \ \cdots \ \beta_{tM_1}^{O_1}). \tag{6.9}$$

As usual, $\boldsymbol{\theta}_t(1)$ evolves through the system equation (6.3) with state evolution matrix $\boldsymbol{G}_t(1)$ being the $(M + M_1)$-dimensional identity matrix. Similarly, the basis functions and parameters for cubic splines representing the relationships between $Y_t(1)$ and headway and speed at $S(1)$ at $t-1$ can also be included in (6.8) and (6.9), respectively.

To use occupancy at site $S(2)$ at $t-1$ for forecasting child $Y_t(2)$, suppose that $h_1^{O_2}(t-1), \ldots, h_{M_2}^{O_2}(t-1)$ and $\beta_{t1}^{O_2}, \ldots, \beta_{tM_2}^{O_2}$ are the basis functions and associated parameters of the cubic spline representing the relationship between occupancy at time $t-1$ and flow at time t at site $S(2)$. Based on the regression and state vectors

for a child in the LMDM given in (6.6) and (6.7),

$$\begin{aligned}\boldsymbol{F}_t(2)^\top &= (y_t(1)h_1(t) \cdots y_t(1)h_M(t) \;\; h_1^{O_2}(t-1) \cdots h_{M_2}^{O_2}(t-1)),\\ \boldsymbol{\theta}_t(2)^\top &= (\beta_{t1} \cdots \beta_{tM} \; \beta_{t1}^{O_2} \cdots \beta_{tM_2}^{O_2}).\end{aligned}$$

State vector $\boldsymbol{\theta}_t(2)$ evolves through the system equation (6.3) with state evolution matrix $\boldsymbol{G}_t(2)$ being the $(M + M_2)$-dimensional identity matrix. The basis functions and parameters for cubic splines representing the relationships between $Y_t(2)$ and headway and speed at $S(2)$ at $t-1$ can similarly be included in $\boldsymbol{F}_t(2)$ and $\boldsymbol{\theta}_t(2)$.

The extension to the case where a child has more than one parent is straightforward. Again, consider the scenario in which $Y_t(3)$ has parents $Y_t(1)$ and $Y_t(2)$, where the daily cycle for the regression parameters is represented by splines with three knots for $Y_t(1)$ and two knots for $Y_t(2)$, and the exogenous variable Z_t needs to be included in $Y_t(3)$'s model. Then a spline representing the relationship between occupancy at time $t-1$ and flow at time t at site $S(3)$ can be additionally incorporated into the model by setting

$$\begin{aligned}\boldsymbol{F}_t(3)^\top &= (h_1(t)^{(1)} \cdots h_6(t)^{(1)} \;\; h_1(t)^{(2)} \cdots h_5(t)^{(2)} \;\; z_t \;\; h_1^{O_3}(t-1) \cdots h_{M_3}^{O_3}(t-1))\\ \boldsymbol{\theta}_t(3)^\top &= (\beta_{t1}^{(1)} \cdots \beta_{t6}^{(1)} \;\; \beta_{t1}^{(2)} \cdots \beta_{t5}^{(2)} \;\; \gamma_t \;\; \beta_{t1}^{O_3} \cdots \beta_{tM_3}^{O_3}).\end{aligned}$$

In this case the evolution matrix $\boldsymbol{G}_t(3) = \text{blockdiag}(\boldsymbol{I}_6, \boldsymbol{I}_5, g, \boldsymbol{I}_{M_3})$ where $\boldsymbol{I}_k$ is the k-dimensional identity matrix and g is some scalar in $\mathbb{R}$ for parameter γ_t's evolution.

Fitted cubic splines are represented by black lines in the scatterplots of flow versus the predictor variables given in Figure 6.1. Unlike the splines fitted for the daily flow cycle in Subsection 6.3.3, in this case the position of the spline knots did not have a considerable effect on the final curve. Also, as suggested by Harrell (2001), four knots were used as a default for fitting all the splines representing extra traffic

variables used in the LMDM.

A comparison of the scatterplots between the columns of Figure 6.1 suggests that the relationships between flow at time t and the predictor variables at time $t-1$ do not vary very much over time. As a consequence, spline fitting would not have to be updated on a frequent basis. However, even if frequent spline fitting were required, fitting is computationally very quick, so it could be used in real-time. Similar conclusions are valid when looking at the same scatterplots for traffic data collected at other months during 2010 and also for different data collection sites. This is also very useful because it means that huge amounts of data are not required before the models can be used.

As mentioned in Chapter 2, only forecasts for Wednesdays are considered throughout this thesis. In a model for all weekdays, it would be parsimonious to have a single spline for all days of the week. In fact, preliminary data analysis indicates that splines fitted using data from Wednesdays and splines fitted using all weekdays give similar results. Furthermore, the forecast performance of the two models using these two sets of splines is very similar. Thus, data for all weekdays is used here for fitting the splines.

6.4.2 Model performance

In order to assess the effect of including occupancy, headway and speed as exogenous regressors on the accuracy of the forecasts provided by the LMDM, various models were compared for several separate subsets of sites in the Manchester network. In particular, forecast models were run for:

- all root nodes;

- four separate bivariate time series formed by the four root nodes together with one of their children;
- four separate trivariate time series formed by the four root nodes together with one of their children and one grandchild.

The reason for this approach was to evaluate the inclusion of predictor variables at a root node on the flow forecasts of its descendants in the DAG.

In the absence of expert information, historical data from July and August 2010 were used to elicit priors. The priors used were comparable across models so that, for example, the spline parameters $\beta_{t1}, \ldots, \beta_{tM}$ representing the daily cycle for a series $Y_t(i)$, used the same priors for all models for that series, and so on. These historical data were also used to estimate the basis functions for all splines used in the models. As considered in the example presented in Subsection 6.3.3, the observation variance $V_t(i)$ was modelled for each $Y_t(i)$ using variance laws and discount factors as described in Section 5.2 of Chapter 5. As such, for each series, the parameter associated with the variance law was the same for all models and was calculated using the prior data. Once again, all discount factors for all models and series were set to be 0.99. On-line one-step ahead forecasts were then obtained for Wednesday flows from September to October 2010.

6.4.3 Root nodes

For each of the root nodes, five (univariate) DLMs were considered:

- Model D (daily cycle model) only uses the daily cycle patterns to forecast flows at time t, modelled via splines as described in Subsection 6.3.2;
- Model O is Model D with the addition of the single predictor variable occupancy at time $t-1$, modelled via splines as described in Subsection 6.4.1;

- Model S is Model D with the addition of the single predictor variable speed at time $t-1$, modelled via splines as described in Subsection 6.4.1;
- Model H is Model D with the addition of the single predictor variable headway at time $t-1$, modelled via splines as described in Subsection 6.4.1;
- Model F (full model) uses cubic splines to model the daily cycle patterns and also uses cubic splines for occupancy, headway and speed measurements at time $t-1$ to forecast flows as described in Subsection 6.4.1.

Table 6.2 gives the LPL for each of these models for all the root nodes. All the models with the daily cycle pattern and one single predictor variable (Models O, S and H) provide better forecasts than Model D, with Model H being the best one for almost all sites. When comparing these models with Model F, Models H and O provide slightly better forecasts for site 9206B and Model H also shows a marginal improvement over Model F for 6013B, whereas Model F is the best among all models for sites 9188A and 1431A. A model using Model D with the inverse of headway was also considered (since this variable can be viewed as the inverse of flow and that is also suggested by the scatterplots in Figure 6.1), but gave worse performance compared to model H.

Although Model F does not necessarily gives the best forecasts for all sites, as mentioned earlier, from a traffic modelling perspective it is sensible to retain all the variables in the model.

TABLE 6.2: LPLs for various models for all root nodes of the Manchester network.

Model	Root node			
	$Y_t(9206B)$	$Y_t(6013B)$	$Y_t(9188A)$	$Y_t(1431A)$
Daily cycle model (D)	-8,259	-7,182	-8,362	-8,128
O	-8,077	-7,138	-8,068	-7,913
S	-8,162	-7,165	-8,141	-7,966
H	-8,075	-7,127	-8,025	-7,931
Full model (F)	-8,084	-7,128	-8,003	-7,887

6.4.4 Children and grandchildren of the root nodes

In order to assess the effects of including predictor variables for root nodes and children, forecasts were obtained using LMDMs for the same four (separate) bivariate series that were considered in Table 6.1. For each bivariate series, three LMDMs were considered:

- Model D/D uses Model D for both root node and child;
- Model F/D uses Model F for root node and Model D for child (with parents as regressors);
- Model F/F uses Model F for both root node and child.

To also evaluate the effect of considering parent flows when forecasting flows of children, independent DLMs using all predictor variables (for both parent and child) were also fitted for each of the bivariate series. The LPLs for all of these models are shown in Table 6.3.

From Table 6.3 it is clear that Model F/F is the best model among all possible alternatives for each bivariate series. What's more, Model F/F provides better forecasts than independent DLMs using all predictor variables for each of the bivariate series. Thus, the inclusion of parent information in addition to the predictor variables when forecasting a child, is better than simply including the predictor variables.

TABLE 6.3: LPLs for different models for bivariate time series from the Manchester network.

Bivariate series	LMDM			Independent
	D/D	F/D	F/F	DLMs (F/F)
$Y_t(9206B), Y_t(9200B)$	-15,156	-15,064	-14,986	-15,467
$Y_t(6013B), Y_t(6007L)$	-13,620	-13,559	-13,520	-13,852
$Y_t(9188A), Y_t(9193J)$	-14,591	-14,219	-14,208	-14,517
$Y_t(1431A), Y_t(1437A)$	-15,163	-14,862	-14,817	-15,402

Notice also that Model F/D provides better forecasts when compared to Model D/D for all bivariate series in Table 6.3. Thus, using the predictor variables in the LMDM seems to improve not only the forecasts at the same site that occupancy, headway and speed were measured, but also affects the quality of the forecasts of its descendants in the DAG.

Similar conclusions can be made when looking at trivariate time series forecasts based on results from Table 6.4, which shows LPLs for LMDMs for series formed by root nodes together with one of their children and one of their associated grandchildren. In this case, model F/D/D, for example, means an LMDM for a trivariate time series using Model F for root node and Model D (with parents as regressors) for its child and grandchild.

TABLE 6.4: LPLs for different models for trivariate time series from the Manchester network

Trivariate series	LMDM			Independent
	D/D/D	F/D/D	F/F/F	DLMs (F/F/F)
$Y_t(9206B), Y_t(9200B), Y_t(9189B)$	-21,935	-21,811	-21,741	-22,740
$Y_t(6013B), Y_t(6007L), Y_t(6004L)$	-19,033	-18,972	-18,923	-19,566
$Y_t(9188A), Y_t(9193J), Y_t(1436M)$	-20,207	-19,791	-19,764	-20,598
$Y_t(1431A), Y_t(1437A), Y_t(1441A)$	-21,879	-21,509	-21,426	-22,418

As another illustration of model improvement when considering occupancy, headway and speed for flow forecasting, Figure 6.2 shows the observed flows on a specific day for (child) site 9200B, together with forecast means and one-step ahead forecast limits (forecast mean $\pm 2\times$ forecast standard deviation). The forecasts were calculated considering LMDMs D/D and F/F for the bivariate time series $(Y_t(9206B), Y_t(9200B))$. The F/F model has narrower forecast limits than the D/D model for the whole day, which is an indication that the inclusion of the predictor variables in the model decreases the uncertainty about flows when compared to an LMDM modelling the daily cycle alone. Notice also that the F/F model captures the deviations from the usual flow patterns that occur during the periods 07:30-08:30 and 15:00-17:00, providing much more accurate forecasts than the D/D model. These

periods correspond to peak times in the network, times when in fact flow forecasting models are most useful.

In Figure 6.2, most observations lie within their respective forecast limits. This should happen for approximately only 95% of observations in a well-calibrated model. Over the whole forecast period, both daily cycle (D) and full (F) models are well-calibrated for the root nodes with roughly 95% of observations lying within the forecast limits for each series. When forecasting child variables as well, however, Model D/D overestimates the forecast uncertainty with roughly 99% to 100% of observations falling within the forecast limits for each series, while this time model F/F is well-calibrated with a coverage of roughly 95%. A similar behaviour was observed for models D/D/D and F/F/F when including grandchild variables. As proposed in Subsection 5.3.2 of Chapter 5, these coverages are a strong indication that there are factors affecting the flow variation which are captured by the inclusion of extra variables as predictors in the model.

6.5 Discussion

The methodology proposed in this chapter tackles both the problem of using extra traffic variables for enhancing flow forecasts as well as using a suitable flow aggregation in order to provide forecasts in an on-line traffic environment. Neither of these issues are often considered in traffic modelling.

Cubic splines were introduced into the LMDM for accommodating the daily cycle exhibited by traffic flows. This approach avoids the use of high-dimensional state parameters and possible computational problems when using seasonal factor representations, while providing better forecasts and allowing the model to be implemented while dealing with data aggregated into small intervals (5-mins here). Cubic splines

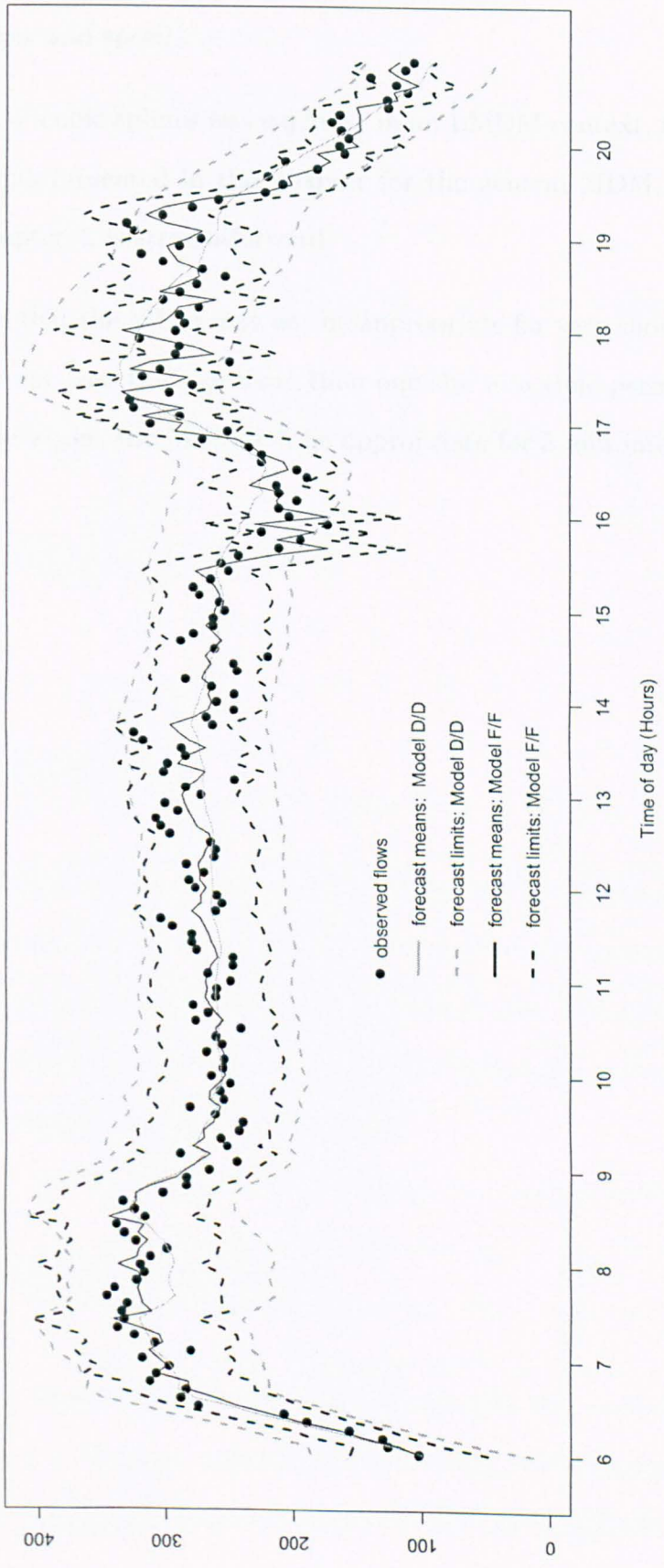

FIGURE 6.2: Observed flows at site 9200B on 27 October 2010, along with forecast means and forecast limits based on LMDMs D/D and F/F for bivariate series $(Y_t(9206B), Y_t(9200B))$.

were also used to model the non-linear relationships which exist between flows and occupancy, headway and speed.

Although the use of cubic splines was explored in an LMDM context, the extension of the developments presented in this chapter for the general MDM, described in Section 4.4 of Chapter 4, is straightforward.

It should be noted that the MDM may not be appropriate for very short intervals in which vehicles do not pass through more than one site in a time period. However, for many traffic networks, the MDM will be appropriate for 5-min intervals.

Chapter 7

Extending the MDM: the dynamic chain graph model

7.1 Introduction

This chapter introduces an extension to the MDM, namely the *dynamic chain graph model* (DCGM). In this new model, a chain graph is used to represent the association structure across time series components at each time frame. This provides a more general graphical representation of multi-dimensional time series and accommodates structures not previously considered in the MDM.

7.2 An MDM restriction

It was shown in previous chapters how the MDM can cope with crucial traffic modelling issues in order to forecast multivariate time series of flows given the requirements of an on-line traffic management environment. Although the MDM seems

to be a suitable model for forecasting traffic flows in real-time, there is still a restriction which prevents this model from accommodating a particular characteristic which may be present in traffic networks.

To illustrate this restriction, consider the root nodes of the DAG presented in Figure 4.11 of Chapter 4, which represent the entrances of the Manchester road network. Queen *et al.* (2008) showed that, in an LMDM for a multivariate time series $\boldsymbol{Y}_t$,

$$cov(Y_t(i), Y_t(j)|D_{t-1}) = 0$$

for any two root nodes $Y_t(i)$ and $Y_t(j)$. This actually holds for the general MDM, since root nodes are marginally independent in a DAG (Edwards, 2000). However, this is not necessarily the case for traffic flow data, where root nodes can be highly correlated. As an example, Figure 7.1 shows scatterplots of flows observed during Wednesdays from June to November 2010 at the root nodes 1431A and 9206B of the Manchester network for four separate 5-min intervals. Figure 7.1(a) indicates a strong relationship between flows at both sites when there is traffic build-up at the beginning of the day (06:00-06:04). Figures 7.1(b) to 7.1(d) show scatterplots of flows at sites 1431A and 9206B for three 5-min intervals observed during afternoon peak times: a relationship between flows of this pair of root nodes can also be identified here, although there is a larger variability in flows.

An approach to tackle this problem can be developed by using a chain graph (CG). The correlation association structure among the root nodes, together with the representation of conditional independence related to causality across time series of flows at the non-root nodes, can be jointly accommodated by using this graphical representation.

Consider the DAG for the Manchester network shown in Figure 4.11 of Chapter 4. For simplicity and ease of presentation, we will consider a subset of the time

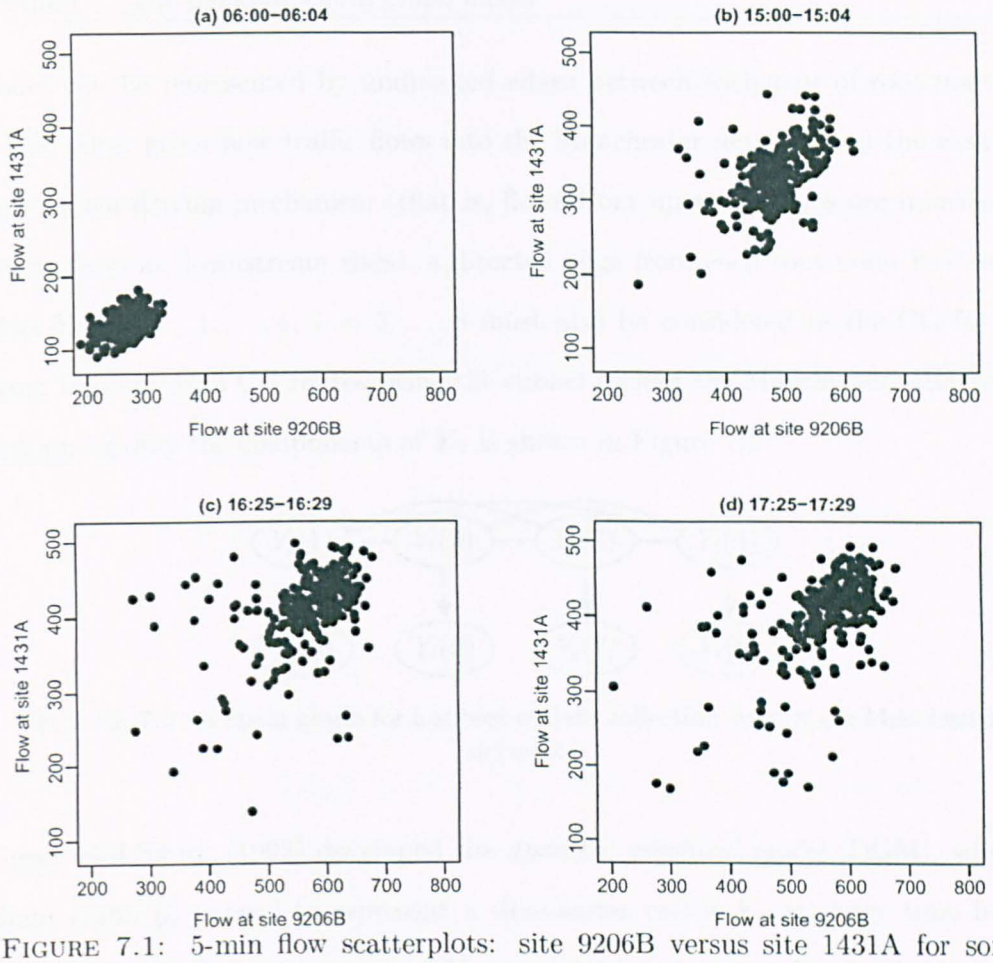

FIGURE 7.1: 5-min flow scatterplots: site 9206B versus site 1431A for some periods of the day: (a) 06:00-06:04; (b) 15:00-15:04; (c) 16:25-16:29; (d) 17:25-17:29 (flows observed during Wednesdays from June to November 2010).

series in this network. Suppose a multivariate time series is formed by all the four root nodes $Y_t(9206B)$, $Y_t(6013B)$, $Y_t(9188A)$ and $Y_t(1431A)$, as well as one of each of their respective children, $Y_t(9200B)$, $Y_t(6007L)$, $Y_t(9193J)$ and $Y_t(1437A)$. For notational convenience, redefine

$$Y_t(9206B) = Y_t(1),\ Y_t(6013B) = Y_t(2),\ Y_t(9188A) = Y_t(3),\ Y_t(1431A) = Y_t(4),$$
$$Y_t(9200B) = Y_t(5),\ Y_t(6007L) = Y_t(6),\ Y_t(9193J) = Y_t(7),\ Y_t(1437A) = Y_t(8),$$

and set $\boldsymbol{Y}_t = [Y_t(1), \ldots, Y_t(8)]^\top$.

Motivated by the scatterplots in Figure 7.1, and similar to the CG elicitation example presented in Chapter 4, an *association structure* across the flows observed at the root

nodes can be represented by undirected edges between each pair of root nodes in a CG. Also, given how traffic flows into the Manchester network and the assumed flow causal driving mechanism (that is, flows from upstream sites are informative about flows at downstream sites), a directed edge from each root node $Y_t(i)$ to its *child* $Y_t(j)$, $i = 1, \ldots, 4$, $j = 5, \ldots, 8$ must also be considered in the CG for $\boldsymbol{Y}_t$. From these ideas, a CG representing the subnetwork of the Manchester intersection containing only the components of $\boldsymbol{Y}_t$ is shown in Figure 7.2.

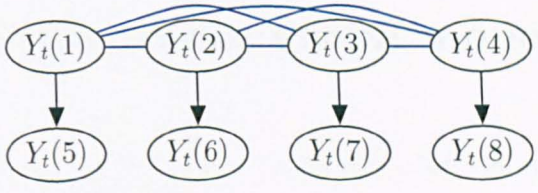

FIGURE 7.2: A chain graph for a subset of data collection sites of the Manchester network

Queen and Smith (1992) developed the *dynamic graphical model* (DGM), where a chain graph $\mathcal{G}_t$ is used to represent a time series vector $\boldsymbol{Y}_t$ at every time frame. In that model, the components of $\boldsymbol{Y}_t$ are partitioned into ordered subsets, called p-sets, and the directed edges in $\mathcal{G}_t$ follow the order of p-sets. This DGM is defined by imposing two conditions:

i. each p-set is a chain component;

ii if there is a directed edge from a variable in a p-set to a variable in another p-set, then there must be a directed edge from *every* variable in the first p-set to *every* variable in the second.

These conditions may not be reasonable for eliciting CGs for road traffic networks. Consider the CG in Figure 7.2. Given item (ii) described above, the DGM from Queen and Smith would require directed edges from, say, $Y_t(1)$ to *all* remaining non root nodes $\{Y_t(5), Y_t(6), Y_t(7), Y_t(8)\}$. Considering the causal driving mechanism

assumed for traffic flows, directed edges from each root node $Y_t(i)$ to *every* non-root $Y_t(j)$, $i = 1,\ldots,4$, $j = 5,\ldots,8$, would not make sense given how traffic flows in the Manchester network. Therefore, the resulting chain graph based on the DGM requirement described in item (ii) may not be suitable to represent network flow structures.

The next section introduces a model where such a requirement is no longer needed.

7.3 The dynamic chain graph model (DCGM)

Let $\boldsymbol{Y}_t = [Y_t(1),\ldots,Y_t(r),Y_t(r+1),\ldots,Y_t(n)]^\top$, be an n-dimensional time series vector, and let $\boldsymbol{Y}_t(R) = [Y_t(1),\ldots,Y_t(r)]^\top$ be the root nodes. Hence, $\boldsymbol{Y}_t = [\boldsymbol{Y}_t(R)^\top, Y_t(r+1),\ldots,Y_t(n)]^\top$ where r can take any integer value between 1 and $n-1$.

For notational convenience, define

$$\begin{aligned}
\boldsymbol{X}_t(i) &= [Y_t(1),\ldots,Y_t(i-1)]^\top, \quad i = r+1,\ldots,n,\\
\boldsymbol{X}_t^*(i) &= [Y_t(r+1),\ldots,Y_t(i-1)]^\top, \quad i = r+2,\ldots,n,\\
\boldsymbol{Z}_t(i) &= [Y_t(i+1),\ldots,Y_t(n)]^\top, \quad i = r+1,\ldots,n-1.
\end{aligned}$$

For $i = r+1$, $\boldsymbol{X}_t^*(i)$ is defined to be $\emptyset$, as is $\boldsymbol{Z}_t(i)$ for $i = n$.

To illustrate this notation, suppose $\boldsymbol{Y}_t = [\boldsymbol{Y}_t(R)^\top, Y_t(3), Y_t(4), Y_t(5)]^\top$ and $r = 2$, so that $\boldsymbol{Y}_t(R) = [Y_t(1), Y_t(2)]^\top$. The subvectors $\boldsymbol{X}_t(i)$, $\boldsymbol{X}_t^*(i)$ and $\boldsymbol{Z}_t(i)$ for each $Y_t(i)$, $i = 3,4,5$, are

$$\begin{aligned}
&\boldsymbol{X}_t(3) = \boldsymbol{Y}_t(R), \qquad \boldsymbol{X}_t^*(3) = \emptyset, \qquad \boldsymbol{Z}_t(3) = [Y_t(4), Y_t(5)]^\top,\\
&\boldsymbol{X}_t(4) = [\boldsymbol{Y}_t(R)^\top, Y_t(3)]^\top, \qquad \boldsymbol{X}_t^*(4) = Y_t(3), \qquad \boldsymbol{Z}_t(4) = Y_t(5)
\end{aligned}$$

and

$$\boldsymbol{X}_t(5) = [\boldsymbol{Y}_t(R)^\top, Y_t(3), Y_t(4)]^\top, \qquad \boldsymbol{X}_t^*(5) = [Y_t(3), Y_t(4)]^\top, \qquad \boldsymbol{Z}_t(5) = \emptyset.$$

Let $\boldsymbol{\theta}_t^\top = [\boldsymbol{\theta}_t(1)^\top, \ldots, \boldsymbol{\theta}_t(n)^\top]$ be the state vectors of $\boldsymbol{Y}_t$, where $\dim(\boldsymbol{\theta}_t(i)) = s_i$, $i = 1, \ldots, n$, and $\boldsymbol{\theta}_t(R)^\top = [\boldsymbol{\theta}_t(1)^\top, \ldots, \boldsymbol{\theta}_t(r)^\top]$ be the state vector of $\boldsymbol{Y}_t(R)$, where $\dim(\boldsymbol{\theta}_t(R)) = \sum_{i=1}^r s_i = q$. In addition, let $\boldsymbol{y}^t(i) = [y_1(i), \ldots, y_t(i)]^\top$ be the observed process associated with $Y_t(i)$, $i = 1, \ldots, n$.

Consider a chain graph $\mathcal{G}$ for $\boldsymbol{Y}_t$ with an association structure across $Y_t(1), \ldots, Y_t(R)$ so that these are joined pairwise by undirected edges. Still in this chain graph, suppose a conditional independence structure where the components of $\boldsymbol{Y}_t$ can be grouped into ordered blocks $B_1, \ldots, B_k$, such that $B_1 = \boldsymbol{Y}_t(R)$ and the blocks $B_2, \ldots, B_k$ each have one time series component only so that $B_2 = Y_t(r+1), \ldots, B_k = Y_t(n)$. Then, the chain graph $\mathcal{G}$ can be elicited for the time series $\boldsymbol{Y}_t$ so that, at each time $t \in \mathbb{N}$, for $i = r+1, \ldots, n$,

$$Y_t(i) \perp\!\!\!\perp [Y_t(1), \ldots, Y_t(i-1)] \backslash \mathrm{bd}(Y_t(i)) | \mathrm{bd}(Y_t(i)) \tag{7.1}$$

and

$$Y_t(i) \perp\!\!\!\perp [\boldsymbol{Y}^t(1), \ldots, \boldsymbol{Y}^t(i-1)] \backslash \mathrm{bd}(\boldsymbol{Y}^t(i)) | \{\mathrm{bd}(\boldsymbol{Y}^t(i)), \boldsymbol{Y}^{t-1}(i)\}, \tag{7.2}$$

where, as defined in Subsection 4.2.1 of Chapter 4, $\mathrm{bd}(Y_t(i))$ is the boundary of $Y_t(i)$. Since

$$\mathrm{bd}(Y_t(i)) = \mathrm{pa}(Y_t(i)) \cup \mathrm{ne}(Y_t(i))$$

where $\mathrm{pa}(Y_t(i))$ and $\mathrm{ne}(Y_t(i))$ are, respectively, the parents and neighbours of $Y_t(i)$, and each $Y_t(i)$, $i = r+1, \ldots, n$, forms a block with one component only, we have that $\mathrm{ne}(Y_t(i)) = \emptyset$, $i = r+1, \ldots, n$. Therefore, at each time $t \in \mathbb{N}$ and for $i = r+1, \ldots, n$,

equations (7.1) and (7.2) can be simplified, respectively, to

$$Y_t(i) \perp\!\!\!\perp [Y_t(1), \ldots, Y_t(i-1)] \backslash \mathrm{pa}(Y_t(i)) | \mathrm{pa}(Y_t(i)) \tag{7.3}$$

and

$$Y_t(i) \perp\!\!\!\perp [\boldsymbol{Y}^t(1), \ldots, \boldsymbol{Y}^t(i-1)] \backslash \mathrm{pa}(\boldsymbol{Y}^t(i)) | \{\mathrm{pa}(\boldsymbol{Y}^t(i)), \boldsymbol{Y}^{t-1}(i)\}, \tag{7.4}$$

which are similar to the conditional independence statements induced by a DAG in the MDM definition, described in Section 4.4 of Chapter 4.

The dynamic chain graph model (DCGM) is defined by the following observation and system equations and initial information.

Observation equations:

$$\boldsymbol{Y}_t(R) = \boldsymbol{F}_t(R)^\top \boldsymbol{\theta}_t(R) + \boldsymbol{v}_t(R), \ \boldsymbol{v}_t(R) \sim (\boldsymbol{0}, \boldsymbol{\Sigma}_t(R)), \tag{7.5}$$

$$Y_t(i) = \boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i) + v_t(i), \ v_t(i) \sim (0, V_t(i)), \ i = r+1, \ldots, n. \tag{7.6}$$

System equations:

$$\boldsymbol{\theta}_t(R) = \boldsymbol{G}_t(R)\boldsymbol{\theta}_{t-1}(R) + \boldsymbol{w}_t(R), \ \ \boldsymbol{w}_t(R) \sim (\boldsymbol{0}, \boldsymbol{W}_t(R)), \tag{7.7}$$

$$\widetilde{\boldsymbol{\theta}}_t = \boldsymbol{G}_t \widetilde{\boldsymbol{\theta}}_{t-1} + \boldsymbol{w}_t, \ \ \boldsymbol{w}_t \sim (\boldsymbol{0}, \boldsymbol{W}_t) \ , \tag{7.8}$$

where $\widetilde{\boldsymbol{\theta}}_t^\top = [\boldsymbol{\theta}_t(r+1)^\top, \ldots, \boldsymbol{\theta}_t(n)^\top]$.

Initial information:

$$\boldsymbol{\theta}_0(R) | D_0 \sim (\boldsymbol{p}_0, \boldsymbol{S}_0), \tag{7.9}$$

$$\widetilde{\boldsymbol{\theta}}_0 | D_0 \sim (\boldsymbol{m}_0, \boldsymbol{C}_0). \tag{7.10}$$

The $q \times r$ dynamic regression matrix $\boldsymbol{F}_t(R)$, the $q \times q$ state evolution matrix $\boldsymbol{G}_t(R)$ and the $q \times q$ evolution variance-covariance matrix $\boldsymbol{W}_t(R)$ are allowed to be a

function of $\boldsymbol{y}^{t-1}(R)$ *but not* $\{\boldsymbol{y}^t(r+1),\ldots,\boldsymbol{y}^t(n)\}$ or $\boldsymbol{y}_t(R)$; $\boldsymbol{\Sigma}_t(R)$ is an $r \times r$ observational variance-covariance matrix and $V_t(i)$ are scalar observational variances, $i = r+1,\ldots,n$. The s_i-dimensional column vector $\boldsymbol{F}_t(i)$ is a known function of $\boldsymbol{x}^t(i)$ and $\boldsymbol{y}^{t-1}(i)$, *but not* $\boldsymbol{z}^t(i)$, $i = r+1,\ldots,n$ or $y_t(i)$. In addition, the r-dimensional vector $\boldsymbol{v}_t(R)$, the q-dimensional vector $\boldsymbol{w}_t(R)$, the s-dimensional vector $\boldsymbol{w}_t$, where $s = \sum_{i=r+1}^{n} s_i$, and $v_t(i)$, $i = r+1,\ldots,n$, are independent sequences of independent errors.

The $s \times s$ matrices $\boldsymbol{G}_t$, $\boldsymbol{W}_t$ and $\boldsymbol{C}_0$ are defined as

$$\boldsymbol{G}_t = \text{blockdiag}(\boldsymbol{G}_t(r+1),\ldots,\boldsymbol{G}_t(n)),$$
$$\boldsymbol{W}_t = \text{blockdiag}(\boldsymbol{W}_t(r+1),\ldots,\boldsymbol{W}_t(n))$$

and

$$\boldsymbol{C}_0 = \text{blockdiag}(\boldsymbol{C}_0(r+1),\ldots,\boldsymbol{C}_0(n))$$

where $\boldsymbol{G}_t(i)$ and $\boldsymbol{W}_t(i)$ may be functions of $\boldsymbol{x}^{t-1}(i)$ and $\boldsymbol{y}^{t-1}(i)$, *but not* $\boldsymbol{z}^{t-1}(i)$, $i = r+1,\ldots,n$. The mean vectors $\boldsymbol{p}_0$ and $\boldsymbol{m}_0$ and covariance matrices $\boldsymbol{S}_0$ and $\boldsymbol{C}_0$ of the initial distributions (7.9) and (7.10) are assumed known.

In the DCGM, a separate multivariate state space model is defined for $\boldsymbol{Y}_t(R)$, such that a non-zero covariance structure between the root nodes of the graph representing $\boldsymbol{Y}_t$ can be considered. This cross covariance structure can be sequentially estimated in time by making inferences on $\boldsymbol{\Sigma}_t(R)$ of equation (7.5). The remaining time series components, $Y_t(r+1),\ldots,Y_t(n)$ are still modelled in an MDM fashion, through the observation equations (7.6) and system equation (7.8). With this approach, all the modelling techniques developed so far for the MDM can still be used within the DCGM, but now allowing a covariance structure among the time series components represented by the root nodes in the chain graph.

7.4 Theoretical results of the DCGM

Let $\boldsymbol{Y}_t = [Y_t(1), \ldots, Y_t(n)]^\top$ be a time series vector modelled by an MDM. The independence after sampling of each state vector $\boldsymbol{\theta}_t(i)$, $i = 1, \ldots, n$, is one of the main properties of the MDM, and it allows local computations when modelling high-dimensional time series. Now, let $\boldsymbol{Y}_t = [Y_t(1), \ldots, Y_t(r), Y_t(r+1), \ldots, Y_t(n)]^\top$, be an n-dimensional time series vector modelled by a DCGM such that $\boldsymbol{Y}_t(R) = [Y_t(1), \ldots, Y_t(r)]^\top$ are the root nodes. In the DCGM, not only each state vector $\boldsymbol{\theta}_t(i)$, $i = r+1, \ldots, n$, remains independent after sampling, but each $\boldsymbol{\theta}_t(i)$ is also independent of the state vector $\boldsymbol{\theta}_t(R)$ after $\boldsymbol{y}_t$ is observed.

To prove this DCGM property, the following Lemma is required.

Lemma 7.1. *(Dawid, 1979). Let $\boldsymbol{X}$,$\boldsymbol{Y}$,$\boldsymbol{Z}$ and $\boldsymbol{W}$ be any random vectors. Then we have that $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y}, \boldsymbol{Z} | \boldsymbol{W} \Leftrightarrow \boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} | (\boldsymbol{Z}, \boldsymbol{W})$ and $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Z} | \boldsymbol{W}$.*

Lemma 7.1 can now be used to prove the following theorem. For ease of presentation, let $\boldsymbol{y}^t$ denote all available information at time t during the proof. This will simplify representations of information subsets at a given time t.

Theorem 7.2. *Let $\{\boldsymbol{Y}_t\}_{t \geq 1}$ be governed by a DCGM with r root nodes, where r can be any integer between 1 and $n-1$, and let $\boldsymbol{\theta}_t^\top = [\boldsymbol{\theta}_t(1)^\top, \ldots, \boldsymbol{\theta}_t(n)^\top]$ be the state vector of $\boldsymbol{Y}_t$. For notational convenience, let*

$$\boldsymbol{\theta}_t(R)^\top = [\boldsymbol{\theta}_t(1)^\top, \ldots, \boldsymbol{\theta}_t(r)^\top],$$

$$\boldsymbol{\phi}_t(i)^\top = [\boldsymbol{\theta}_t(1)^\top, \ldots, \boldsymbol{\theta}_t(i-1)^\top], \quad i = r+1, \ldots, n,$$

$$\boldsymbol{\phi}_t^*(i)^\top = [\boldsymbol{\theta}_t(r+1)^\top, \ldots, \boldsymbol{\theta}_t(i-1)^\top], \quad i = r+2, \ldots, n,$$

and

$$\boldsymbol{\psi}_t(i)^\top = [\boldsymbol{\theta}_t(i+1)^\top, \ldots, \boldsymbol{\theta}_t(n)^\top], \quad i = r+1, \ldots, n-1,$$

be the state vectors of $\boldsymbol{Y}_t(R)$, $\boldsymbol{X}_t(i)$, $\boldsymbol{X}_t^*(i)$ *and* $\boldsymbol{Z}_t(i)$ *(defined at the start of Section 7.3) respectively. If we have,*

$$\boldsymbol{\theta}_{t-1}(R) \perp\!\!\!\perp \{\boldsymbol{\theta}_{t-1}(r+1), \ldots, \boldsymbol{\theta}_{t-1}(n)\}, \{\boldsymbol{y}^{t-1}(r+1), \ldots, \boldsymbol{y}^{t-1}(n)\} | \boldsymbol{y}^{t-1}(R) \quad (7.11)$$

$$\boldsymbol{\phi}_{t-1}^*(i) \perp\!\!\!\perp \boldsymbol{y}^{t-1}(i), \boldsymbol{z}^{t-1}(i), \boldsymbol{\theta}_{t-1}(R) | \boldsymbol{x}^{t-1}(i), \quad i = r+2, \ldots, n, \quad (7.12)$$

$$\boldsymbol{\theta}_{t-1}(i) \perp\!\!\!\perp \boldsymbol{z}^{t-1}(i), \boldsymbol{\theta}_{t-1}(R), \boldsymbol{\phi}_{t-1}^*(i) | \boldsymbol{x}^{t-1}(i), \boldsymbol{y}^{t-1}(i), \quad i = r+1, \ldots, n, \quad (7.13)$$

$$\boldsymbol{\psi}_{t-1}(i) \perp\!\!\!\perp \boldsymbol{\theta}_{t-1}(R), \boldsymbol{\phi}_{t-1}^*(i), \boldsymbol{\theta}_{t-1}(i) | \boldsymbol{y}^{t-1}, \quad i = r+1, \ldots, n-1, \quad (7.14)$$

then, the following conditional independence statements must also be true:

$$\boldsymbol{\theta}_t(R) \perp\!\!\!\perp \{\boldsymbol{\theta}_t(r+1), \ldots, \boldsymbol{\theta}_t(n)\}, \{\boldsymbol{y}^t(r+1), \ldots, \boldsymbol{y}^t(n)\} | \boldsymbol{y}^t(R) \quad (7.15)$$

$$\boldsymbol{\phi}_t^*(i) \perp\!\!\!\perp \boldsymbol{y}^t(i), \boldsymbol{z}^t(i), \boldsymbol{\theta}_t(R) | \boldsymbol{x}^t(i), \quad i = r+2, \ldots, n, \quad (7.16)$$

$$\boldsymbol{\theta}_t(i) \perp\!\!\!\perp \boldsymbol{z}^t(i), \boldsymbol{\theta}_t(R), \boldsymbol{\phi}_t^*(i) | \boldsymbol{x}^t(i), \boldsymbol{y}^t(i), \quad i = r+1, \ldots, n, \quad (7.17)$$

$$\boldsymbol{\psi}_t(i) \perp\!\!\!\perp \boldsymbol{\theta}_t(R), \boldsymbol{\phi}_t^*(i), \boldsymbol{\theta}_t(i) | \boldsymbol{y}^t, \quad i = r+1, \ldots, n-1, \quad (7.18)$$

This theorem can be proved by defining a chain graph representing the hypotheses (7.11) to (7.14), the observation equations (7.5) and (7.6) and system equations (7.7) and (7.8). The conditional independence statements (7.15) to (7.18) are then verified by applying global Markov properties in the resulting chain graph. Preliminary chain graphs will be considered during the following proof in order to elicit the final graph. The proof holds for all $r = 1, \ldots, n-1$ and $i = r+1, \ldots, n$. However, for clarity of presentation, the graphs used in the proof consider the particular case in which $r = 2$.

Proof. Figure 7.3 shows a chain graph representing the components of $\boldsymbol{y}^{t-1}$ and $\boldsymbol{\theta}_{t-1}$ for some $i = 3, \ldots, n-1$ (note that, for $i = 3$, $\boldsymbol{x}^{*t-1}(i)$ and $\boldsymbol{\phi}_{t-1}^*(i)$ are simply null).

The hypothesis (7.11) implies omission of arcs between $\{\boldsymbol{y}^{t-1}(i) \mid i = 3, \ldots, n\}$ and $\boldsymbol{\theta}_{t-1}(R) = \{\boldsymbol{\theta}_{t-1}(1), \boldsymbol{\theta}_{t-1}(2)\}$, and between $\boldsymbol{\theta}_{t-1}(R)$ and $\{\boldsymbol{\theta}_{t-1}(r+3), \ldots, \boldsymbol{\theta}_{t-1}(n)\}$.

Applying Lemma 7.1 in hypotheses (7.12) and (7.13), we have the conditional independence statements

$$\boldsymbol{\phi}^*_{t-1}(i) \perp\!\!\!\perp \boldsymbol{\theta}_{t-1}(R) | \boldsymbol{y}^{t-1}, \tag{7.19}$$

$$\boldsymbol{\phi}^*_{t-1}(i) \perp\!\!\!\perp \boldsymbol{y}^{t-1}(i), \boldsymbol{z}^{t-1}(i) | \boldsymbol{x}^{t-1}(i), \tag{7.20}$$

$$\boldsymbol{\theta}_{t-1}(i) \perp\!\!\!\perp \boldsymbol{\theta}_{t-1}(R), \boldsymbol{\phi}^*_{t-1}(i) | \boldsymbol{y}^{t-1} \tag{7.21}$$

and

$$\boldsymbol{\theta}_{t-1}(i) \perp\!\!\!\perp \boldsymbol{z}^{t-1}(i) | \boldsymbol{x}^{t-1}(i), \boldsymbol{y}^{t-1}(i). \tag{7.22}$$

Hence, hypothesis (7.14) and statements (7.19) and (7.21) imply the omission of arcs between $\{\boldsymbol{\theta}_{t-1}(i) \mid i = 3, \ldots, n\}$ and $\boldsymbol{\theta}_{t-1}(R)$, and also between $\{\boldsymbol{\theta}_{t-1}(k) \mid k \neq i\}$ and $\boldsymbol{\theta}_{t-1}(i)$, $i, k = 3, \ldots, n$. In addition, hypothesis (7.11) and statements (7.20) and (7.22) imply arc omissions between $\{\boldsymbol{y}^{t-1}(k) \mid k > i\}$ and $\boldsymbol{\theta}_{t-1}(i)$, for $i = 1, \ldots, n-1$, $k = 3, \ldots, n$.

Figure 7.4 shows a chain graph representing associations between components of $\boldsymbol{y}^{t-1}$, $\boldsymbol{\theta}_{t-1}$ and $\boldsymbol{\theta}_t$ based on system equations (7.7) and (7.8). In this figure, associations between $\boldsymbol{y}^{t-1}$ and $\boldsymbol{\theta}_{t-1}$ described using Figure 7.3 are omitted for ease of presentation. The restriction that $\boldsymbol{G}_t(R)$ and $\boldsymbol{W}_t(R)$ are functions of $\boldsymbol{y}^{t-1}(R)$ and not $\{\boldsymbol{y}^{t-1}(r+1), \ldots, \boldsymbol{y}^{t-1}(n)\}$ justifies arc omissions between $\{\boldsymbol{y}^{t-1}(i) \mid i = 3, \ldots, n\}$ and $\boldsymbol{\theta}_t(R)$. Also, the MDM restriction that $\boldsymbol{G}_t(i)$ and $\boldsymbol{W}_t(i)$ can be functions of $\boldsymbol{x}^{t-1}(i)$ and $\boldsymbol{y}^{t-1}(i)$, but not $\boldsymbol{z}^{t-1}(i)$, justifies omissions of arcs between $\{\boldsymbol{y}^{t-1}(j) \mid j > i\}$ and $\boldsymbol{\theta}_t(i)$, $i = 3, \ldots, n-1$ and $j = 4, \ldots, n$. The arc omissions between $\boldsymbol{\theta}_t(R)$ and $\boldsymbol{\theta}_t(i)$, $i = 3, \ldots, n$, can be justified by noting that the form of equations (7.7) and (7.8) imply that $\boldsymbol{\theta}_t(R) \perp\!\!\!\perp \{\boldsymbol{\theta}_t(r+1), \ldots, \boldsymbol{\theta}_t(n)\} | \boldsymbol{y}^{t-1}, \boldsymbol{\theta}_{t-1}(R)$.

The justification for the remaining arc omissions in Figure 7.4 are based on the block diagonal forms of $\boldsymbol{G}_t$ and $\boldsymbol{W}_t$, which imply that

$$\boldsymbol{\theta}_t(i) \perp\!\!\!\perp \{\boldsymbol{\theta}_{t-1} \backslash \boldsymbol{\theta}_{t-1}(i)\} | \boldsymbol{y}^{t-1}, \boldsymbol{\theta}_{t-1}(i), \quad i = 3, \ldots, n.$$

This justifies omission of arcs between $\{\boldsymbol{\theta}_{t-1}(i) \mid i = 3, \ldots, n\}$ and $\boldsymbol{\theta}_t(R)$, and also between $\{\boldsymbol{\theta}_{t-1}(k) \mid k \neq i\}$ and $\boldsymbol{\theta}_t(i)$, $i, k = 3, \ldots, n$.

Figure 7.5 shows a chain graph representing associations between components of $\boldsymbol{y}^{t-1}$, $\boldsymbol{\theta}_{t-1}$, $\boldsymbol{\theta}_t$ and $\boldsymbol{y}_t$, based on observation equations (7.5) and (7.6). In this figure, associations between $\boldsymbol{y}^{t-1}$, $\boldsymbol{\theta}_{t-1}$ and $\boldsymbol{\theta}_t$ described using Figures 7.3 and Figures 7.4 are omitted for ease of presentation. These observation equations and the forms of $\boldsymbol{F}_t(R)$ and $\boldsymbol{F}_t(i)$, $i = 3, \ldots, n$, ensure that,

$$\boldsymbol{y}_t(R) \perp\!\!\!\perp \boldsymbol{\theta}_{t-1} | \boldsymbol{y}^{t-1}(R), \boldsymbol{\theta}_t(R)$$

and

$$y_t(i) \perp\!\!\!\perp \boldsymbol{\theta}_{t-1} | \boldsymbol{x}^t(i), \boldsymbol{y}^{t-1}(i), \boldsymbol{\theta}_t(i), \quad i = 3, \ldots, n.$$

These conditional independence statements justify omission of arcs between $\boldsymbol{\theta}_{t-1}(k)$ to $y_t(i)$, $i, k = 1, \ldots, n$. Arcs between $\{\boldsymbol{\theta}_t(i) \mid i = 3, \ldots, n\}$ and $\boldsymbol{y}_t(R)$ and arcs between $\{\boldsymbol{\theta}_t(k) \mid k \neq i\}$ and $y_t(i)$, $i, k = 3, \ldots, n$, can also be omitted given (7.5) and (7.6), as these equations imply that

$$\boldsymbol{y}_t(R) \perp\!\!\!\perp \{\boldsymbol{\theta}_t(r+1), \ldots, \boldsymbol{\theta}_t(n)\} | \boldsymbol{y}^{t-1}(R), \boldsymbol{\theta}_t(R)$$

and

$$y_t(i) \perp\!\!\!\perp \{\boldsymbol{\theta}_t \backslash \boldsymbol{\theta}_t(i)\} | \boldsymbol{x}^t(i), \boldsymbol{y}^{t-1}(i), \boldsymbol{\theta}_t(i). \quad i = 3, \ldots, n.$$

Since $\boldsymbol{F}_t(R)$ may be a function of $\boldsymbol{y}^{t-1}(R)$ but is not a function of $\{\boldsymbol{y}^{t-1}(r+1), \ldots, \boldsymbol{y}^{t-1}(n)\}$ and given that $\boldsymbol{F}_t(i)$ is a known function of $\boldsymbol{x}^t(i)$ and $\boldsymbol{y}^{t-1}(i)$, but

not $\boldsymbol{z}^t(i)$, we also have that,

$$\boldsymbol{y}_t(R) \perp\!\!\!\perp \{\boldsymbol{y}^{t-1}(r+1), \ldots, \boldsymbol{y}^{t-1}(n)\} | \boldsymbol{y}^{t-1}(R)$$

and

$$y_t(i) \perp\!\!\!\perp \boldsymbol{z}^{t-1}(i) | \boldsymbol{x}^t(i), \boldsymbol{y}^{t-1}(i), \ i = 3, \ldots, n.$$

This justifies missing arcs between $\{\boldsymbol{y}^{t-1}(i) \mid i = 3, \ldots, n\}$ and $\boldsymbol{Y}_t(R)$, and between $\{\boldsymbol{y}^{t-1}(k) \mid k > i\}$ and $y_t(i)$, $i = 3, \ldots, n$.

Figure 7.6 shows the moralized resulting chain graph built on preliminary chain graphs in Figure 7.3 (Box A), Figure 7.4 (Box B) and Figure 7.5 (Box C). The chain components of this graph which are not single nodes are $\{\boldsymbol{y}^{t-1}(1), \boldsymbol{y}^{t-1}(2), \}$, $\{\boldsymbol{\theta}_{t-1}(1), \boldsymbol{\theta}_{t-1}(2)\}$,$\{\boldsymbol{\theta}_t(1), \boldsymbol{\theta}_t(2)\}$ and $\{y_t(1), y_t(2)\}$, and the grey undirected edges are the result of moralization, as described in Section 4.2.2.1 of Chapter 4. For the general case where r is any integer between 1 and $n-1$, the chain graph would have the same general form, only there would be chain components $\{\boldsymbol{y}^{t-1}(1), \ldots, \boldsymbol{y}^{t-1}(r)\}$, $\{\boldsymbol{\theta}_{t-1}(1), \ldots, \boldsymbol{\theta}_{t-1}(r)\}$, $\{\boldsymbol{\theta}_t(1), \ldots, \boldsymbol{\theta}_t(r)\}$ and $\{y_t(1), \ldots, y_t(r)\}$, in which all nodes within a chain component are joined pairwise by undirected edges.

The theorem is proved by verifying that, for each statement (7.15) to (7.18), of the form $A \perp\!\!\!\perp B | C$, the (conditioning) component C separates the (conditioned) components A and B.

Figures 7.7 to 7.10 show the chain graph of Figure 7.6 highlighting the conditioned components (orange and brown nodes) and the conditioning components (violet nodes) of each statement (7.15) to (7.18). The global Markov property for chain graphs, described in Chapter 4, can then be used to verify each of these conditional independence statements. □

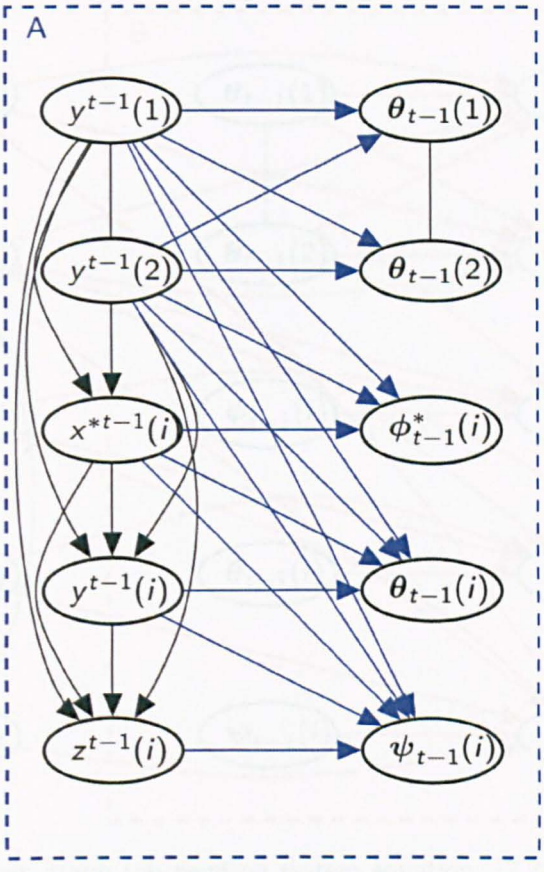

FIGURE 7.3: Chain graph for the inductive hypothesis (7.11) to (7.14) of the Theorem 7.2.

The following corollary ensures the independence after sampling of $\boldsymbol{\theta}_t(R)$ and each $\boldsymbol{\theta}_t(i)$, $i = r+1, \ldots, n$, under the DCGM.

Corollary 7.3. *If* $\boldsymbol{\theta}_0(R) \perp\!\!\!\perp \{\boldsymbol{\theta}_0(r+1), \ldots, \boldsymbol{\theta}_0(n)\}$ *and* $\perp\!\!\!\perp_{i=r+1}^{n} \boldsymbol{\theta}_0(i)$, *then, for all* t,

$$\boldsymbol{\theta}_t(R) \perp\!\!\!\perp \{\boldsymbol{\theta}_t(r+1), \ldots, \boldsymbol{\theta}_t(n)\} | \boldsymbol{y}^t,$$

$$\perp\!\!\!\perp_{i=r+1}^{n} \boldsymbol{\theta}_t(i) | \boldsymbol{y}^t$$

and

$$\boldsymbol{\theta}_t(R) \perp\!\!\!\perp \{\boldsymbol{y}^t(r+1), \ldots, \boldsymbol{y}^t(n)\} | \boldsymbol{y}^t(R),$$

$$\boldsymbol{\theta}_t(i) \perp\!\!\!\perp \{\boldsymbol{y}^t(i+1), \ldots, \boldsymbol{y}^t(n)\} | \boldsymbol{y}^t(1), \ldots, \boldsymbol{y}^t(i), \quad i = r+1, \ldots, n.$$

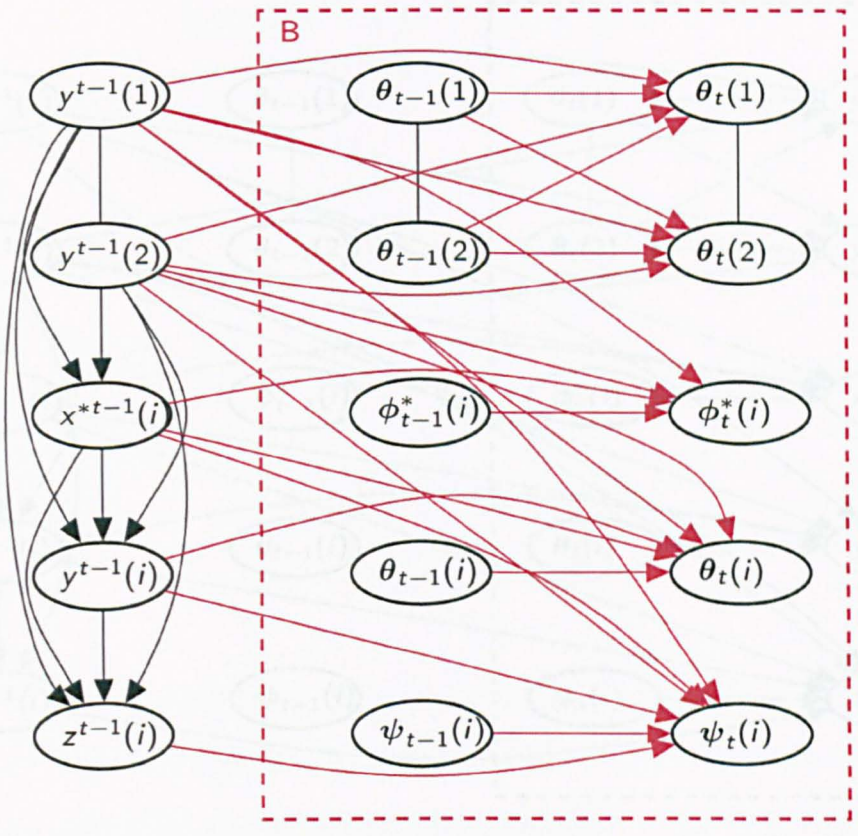

FIGURE 7.4: Chain graph representing system equations (7.7) and (7.8) of the DCGM definition.

Proof. Figure 7.11 shows a chain graph representing the initial independence hypotheses $\boldsymbol{\theta}_0(R) \perp\!\!\!\perp \{\boldsymbol{\theta}_0(r+1), \ldots, \boldsymbol{\theta}_0(n)\}$ and $\perp\!\!\!\perp_{i=r+1}^{n} \boldsymbol{\theta}_0(i)$, (box A), system equations (7.7) and (7.8) at time 1 (box B) and the observation equations (7.5) and (7.6) at time 1 (box C). The grey undirected edges in Figure 7.11 are the result of moralization.

From the moralized graph in Figure 7.11, the conditional independence statements (7.15) to (7.18) can be deduced for time $t = 1$. Therefore, when $\boldsymbol{\theta}_0(R) \perp\!\!\!\perp \{\boldsymbol{\theta}_0(r+1), \ldots, \boldsymbol{\theta}_0(n)\}$ and $\perp\!\!\!\perp_{i=r+1}^{n} \boldsymbol{\theta}_0(i)$, the conditional independence statements (7.15) to (7.18) are true for time $t = 1$ and, by induction, from Theorem 7.2, conditional independence statements (7.15) to (7.18) must be true for all time t.

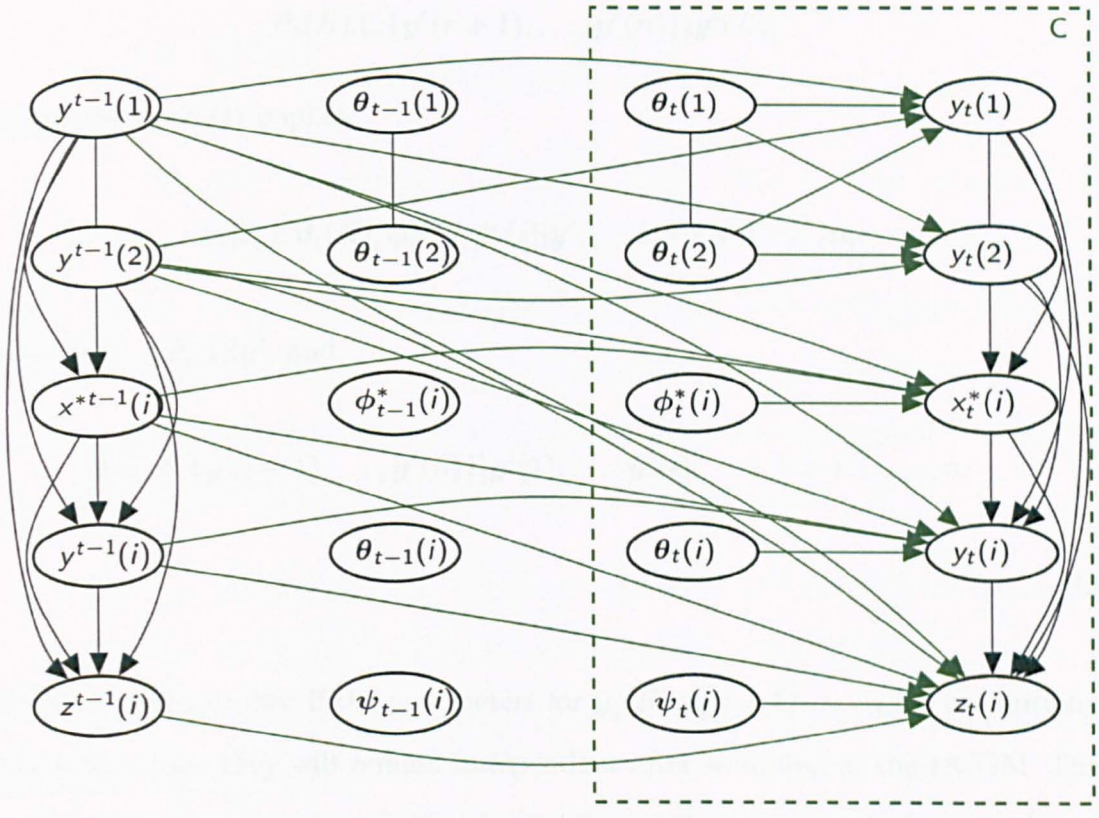

FIGURE 7.5: Chain graph representing the observation equations (7.5) and (7.6) of the DCGM definition.

Conditional independence statements (7.15) to (7.18) can be then combined into two statements:

$$\boldsymbol{\theta}_t(R) \perp\!\!\!\perp \{\boldsymbol{\theta}_t(r+1), \ldots, \boldsymbol{\theta}_t(n)\}, \{\boldsymbol{y}^t(r+1), \ldots, \boldsymbol{y}^t(n)\} | \boldsymbol{y}^t(R) \tag{7.23}$$

and

$$\boldsymbol{\theta}_t(i) \perp\!\!\!\perp \boldsymbol{z}^t(i), \boldsymbol{\theta}_t(R), \boldsymbol{\phi}^*_t(i), \boldsymbol{\psi}_t(i), | \boldsymbol{x}^t(i), \boldsymbol{y}^t(i), \quad i = r+1, \ldots, n. \tag{7.24}$$

Then, by Lemma 7.1, statement (7.23) implies

$$\boldsymbol{\theta}_t(R) \perp\!\!\!\perp \{\boldsymbol{\theta}_t(r+1), \ldots, \boldsymbol{\theta}_t(n)\} | \boldsymbol{y}^t$$

and

$$\boldsymbol{\theta}_t(R) \perp\!\!\!\perp \{\boldsymbol{y}^t(r+1), \ldots, \boldsymbol{y}^t(n)\} | \boldsymbol{y}^t(R),$$

and statement (7.24) implies

$$\boldsymbol{\theta}_t(i) \perp\!\!\!\perp \boldsymbol{\theta}_t(R), \boldsymbol{\phi}_t^*(i), \boldsymbol{\psi}_t(i) | \boldsymbol{y}^t, \quad i = r+1, \ldots, n,$$

so that $\perp\!\!\!\perp_{i=r+1}^{n} \boldsymbol{\theta}_t(i) | \boldsymbol{y}^t$, and

$$\boldsymbol{\theta}_t(i) \perp\!\!\!\perp \{\boldsymbol{y}^t(i+1), \ldots, \boldsymbol{y}^t(n)\} | \boldsymbol{y}^t(1), \ldots, \boldsymbol{y}^t(i), \quad i = r+1, \ldots, n.$$

□

Corollary 7.3 means that if the parameters for $\boldsymbol{y}_t(R), y_t(r+1), \ldots, y_t(n)$ are initially independent, then they will remain independent after sampling in the DCGM. The fact that separate priors are specified for $\boldsymbol{\theta}_0(R)$ and $\widetilde{\boldsymbol{\theta}}_0$ in (7.9) and (7.10), and that $\boldsymbol{C}_0$ in (7.10) is block diagonal, ensures that the condition of initial independence holds. Thus the parameters $\boldsymbol{\theta}_t(R), \boldsymbol{\theta}_t(r+1), \ldots, \boldsymbol{\theta}_t(n)$ can be updated separately within the respective models for $\boldsymbol{Y}_t(R)$, $Y_t(r+1)|\text{pa}(y_t(r+1)), \ldots, Y_t(n)|pa(y_t(n))$. Forecasts can also be obtained separately for $\boldsymbol{Y}_t(R)$ and $Y_t(r+1)|pa(y_t(r+1)), \ldots, Y_t(n)|pa(y_t(n))$. Thus, the n-dimensional model is decomposed into an r-dimensional multivariate model (with $r < n$), plus $n - r$ separate univariate models.

As with the multiregression dynamic model, the marginal forecasts for $\boldsymbol{Y}_t(r+1), \ldots, \boldsymbol{Y}_t(n)$ are required and calculation of the marginal forecast moments is exactly the same as for the multiregression dynamic model (Queen & Smith, 1993; Queen *et al.*, 2008). Because $\boldsymbol{Y}_t(R)$ do not condition on any other variables, no further calculations are required to obtain the marginal forecasts for that vector.

7.5 The linear dynamic chain graph model (LDCGM)

Suppose a time series $\boldsymbol{Y}_t = [Y_t(1), \ldots, Y_t(n)]^\top$ with r root nodes is modelled by a DCGM. The non-root nodes $\{Y_t(r+1), \ldots, Y_t(n)\}$ in this case can be modelled using a regression vector $\boldsymbol{F}_t(i)$ in observation equation (7.6) which must include a function of the parents of $Y_t(i)$, $i = r+1, \ldots, n$. Since all the components of $\boldsymbol{Y}_t$ are simultaneously observed, the marginal distributions of the forecast distribution for each non-root node $\boldsymbol{Y}_t(i)$ is required for forecasting purposes.

When regression is linear and the errors in equations (7.5) to (7.8) are normally distributed, then the marginal moments of the forecast distributions for $Y_t(i)$, $i = r+1, \ldots, n$ can be calculated analytically. In this case, the model is a *linear dynamic chain graph model* (LDCGM).

Let $\boldsymbol{Y}_t = [Y_t(1), \ldots, Y_t(n)]^\top$ be a time series and $\boldsymbol{Y}_t(R) = [Y_t(1), \ldots, Y_t(r)]^\top$ be the r root nodes. Given a chain graph for $\boldsymbol{Y}_t$, the LDCGM is defined by observation equations, system equations and initial information as follows.

Observation equations:

$$\boldsymbol{Y}_t(R)^\top = \boldsymbol{F}_t(R)^\top \Theta_t(R) + \boldsymbol{v}_t(R), \ \boldsymbol{v}_t(R) \sim N(\boldsymbol{0}, \Sigma_t(R)), \tag{7.25}$$

$$Y_t(i) = \boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i) + v_t(i), \ v_t(i) \sim N(0, V_t(i)), \ i = r+1, \ldots, n. \tag{7.26}$$

System equations:

$$\Theta_t(R) = \boldsymbol{G}_t(R)\Theta_{t-1}(R) + \Omega_t(R), \ \ \Omega_t(R) \sim N(\boldsymbol{0}, \boldsymbol{W}_t(R), \Sigma_t(R)), \tag{7.27}$$

$$\widetilde{\boldsymbol{\theta}}_t = \boldsymbol{G}_t \widetilde{\boldsymbol{\theta}}_{t-1} + \boldsymbol{w}_t, \ \ \boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{W}_t) \, . \tag{7.28}$$

Initial information:

$$\Theta_0(R)|D_0 \sim N(\boldsymbol{p}_0, \boldsymbol{W}_0(R), \Sigma_0(R)),$$

$$\widetilde{\boldsymbol{\theta}}_0|D_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0).$$

In this DCGM, the observation equation (7.25) and the system equation (7.27) represent a matrix normal DLM, as described in Subsection 3.6.2 of Chapter 3. Hence, the observation equation (7.25) uses row vector $\boldsymbol{Y}_t(R)^\top$ rather than $\boldsymbol{Y}_t(R)$, and the matrix $\Sigma_t(R)$ represents the cross-sectional time-varying covariance structure among the root nodes $\boldsymbol{Y}_t(R)$ at each time t. Additionally, each $Y_t(1), \ldots, Y_t(r)$ has the same p-dimensional regression vector $\boldsymbol{F}_t(R)$, with $p \times r$ state matrix

$$\Theta_t(R) = [\boldsymbol{\theta}_t(1), \ldots, \boldsymbol{\theta}_t(r)],$$

where $\boldsymbol{\theta}_t(i)$ is the p-dimensional state vector for $Y_t(i)$ and has same evolution matrix $\boldsymbol{G}_t(R)$, $i = 1, \ldots, r$.

The dynamic evolution of the state process of $\boldsymbol{Y}_t(R)$ is a function of the state matrix $\Theta_t(R)$ and a matrix of evolution errors $\boldsymbol{\Omega}_t(R)$ which follows a matrix normal distribution with null vector mean, left variance matrix $\boldsymbol{W}_t(R)$ and right variance matrix $\Sigma_t(R)$ (again, as in Subsection 3.6.2 of Chapter 3).

The remaining time series components, $\{Y_t(r+1), \ldots, Y_t(n)\}$, are modelled through equations (7.26) and (7.28). The form of these equations means that an LMDM is used to model the non-root nodes of $\boldsymbol{Y}_t(i)$ with some extra parents from $\boldsymbol{Y}_t(R)$. Therefore, each $\boldsymbol{Y}_t(i)$, $i = r+1, \ldots, n$ follows a conditional regression DLM and forecasts for $\{Y_t(r+1), \ldots, Y_t(n)\}$ can be obtained by applying the techniques to calculate marginal moments under the LMDM, as illustrated in Section 4.7 of Chapter 4.

7.6 Example

To compare the effect of using different graphical representations of road traffic networks on the quality of flow forecasts, particular cases of both the LMDM (in which root nodes are uncorrelated) and the LDCGM were used for the Manchester network data. The 8-dimensional time series $\boldsymbol{Y}_t = [Y_t(1), \ldots, Y_t(8)]^{\mathrm{T}}$ defined in Section 7.2 was used for this comparison.

5-min data from November 2010 were used to elicit priors for both models, and one-step ahead forecasts were obtained for flows observed from 15:00 to 19:59 during Wednesdays in December 2010 (the reason for not using data from the whole day will be explained later). This particular month was chosen because traffic disruptions caused by heavy snow were observed at the end of 2010. Hence, the models could be compared given that an explicit factor (adverse weather conditions) was affecting the traffic being modelled.

Based on the chain graph in Figure 7.2, the LDCGM defined in Section 7.5 was used to forecast $\boldsymbol{Y}_t$. Therefore, given $r = 4$, the vector of root nodes $\boldsymbol{Y}_t(R)$ is modelled by a matrix normal DLM, with a $p \times 4$ state matrix $\boldsymbol{\Theta}_t(R)$, where each column is a p-dimensional state vector for each time series component in $\boldsymbol{Y}_t(R)$.

In equation (7.25), $\boldsymbol{F}_t(R)$ is a vector common to all the time series of root nodes. It is then difficult to apply the methodology described in Chapter 6 to include extra traffic variables available for each root node under the LDCGM. Due to this restriction, occupancy, speed and headway values were not used in the matrix normal DLM for $\boldsymbol{Y}_t(R)$ and

$$\boldsymbol{F}_t(R) = [h_1(t), \ldots, h_M(t)]^T \quad \text{and} \quad \boldsymbol{G}_t(R) = \boldsymbol{I}_M,$$

where $h_1(t), \ldots, h_M(t)$ are the spline basis functions representing the cycle from the period 15:00 to 19:59, and $\boldsymbol{G}_t(R)$ is an identity matrix of order M. The state

evolution matrix $\Theta_t(R)$ is given by

$$\begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \beta_{1,3} & \beta_{1,4} \\ \beta_{2,1} & \beta_{2,2} & \beta_{2,3} & \beta_{2,4} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{M,1} & \beta_{M,2} & \beta_{M,3} & \beta_{M,4} \end{pmatrix}_t .$$

The form of this evolution matrix results in four separate spline functions modelling the seasonal cycle of each root node in $\boldsymbol{Y}_t(R)$. Hence, $\Theta_t(R)$ can accommodate different seasonal patterns for $Y_t(1), \ldots, Y_t(4)$. The flow seasonality patterns at all sites of the Manchester network are fairly similar, with any difference being accounted for in $\Theta_t(R)$.

In this example, both covariance matrix $\boldsymbol{\Sigma}_t(R)$ and evolution variance matrix $\boldsymbol{W}_t(R)$ in (7.25) and (7.27) were assumed unknown. The evolution variance matrix $\boldsymbol{W}_t(R)$ was modelled via a discount factor, as described in Subsection 3.6.2 of Chapter 3. A discount factor was also used to model the time-varying covariance matrix $\boldsymbol{\Sigma}_t(R)$, as described in Section 5.2.2 of Chapter 5. However, as pointed out by Prado and West (2010), variance modelling via discount factors is suitable when the (co)variances have a behaviour similar to a random walk, that is, when the (co)variances have a smooth and gradual random changing. This is a reasonable assumption between 15:00 to 19:59 for these data, so this is the reason why only data during this period each day are considered here.

In addition, unlike what was observed for root node flow variances (described in Subsection 5.2.1 of Chapter 5), it is not necessarily the case that covariances between root nodes change with the mean. Hence, variance laws were not considered to model the observational variance $\boldsymbol{\Sigma}_t(R)$ in equations (7.25) and (7.27).

Given $\boldsymbol{Y}_t(R)$, each $Y_t(i)|\text{pa}(Y_t(i))$, is modelled by a separate univariate DLM

$\{\boldsymbol{F}_t(i), \boldsymbol{G}_t(i), V_t(i), V_t(i)\boldsymbol{W}_t(i)\}$, $i = 5, \ldots, 8$. Hence, a Bayesian conjugate analysis of $(\boldsymbol{\theta}_t(i), \phi_t(i))$ through a normal-gamma prior together with a discount factor for $V_t(i)$, as used in Section 5.2 of Chapter 5 and Subsection 6.4.2 of Chapter 6, can be applied for the non-root node modelling under the LDCGM. Additionally, discount factor techniques for the evolution variances $\boldsymbol{W}_t(i)$ and the assumption of a time-varying observational variance of the form

$$V_t(i) = \exp\{\alpha \log[E(Y_t(i)|D_{t-1})]\}\phi_t(i)^{-1}$$

where

$$\log(\text{variance of flow}) = \alpha \log(\text{mean flow})$$

such that α is estimated from historical data (again as in Section 5.2 of Chapter 5 and Subsection 6.4.2 of Chapter 6) was also considered for $Y_t(5), \ldots, Y_t(8)$ in this example.

The regression vector $\boldsymbol{F}_t(i)$, $i = 5, \ldots, 8$, is built based on splines to model the daily flow cycle and also to include values of occupancy, speed and headway observed at time $t-1$ for each series $Y_t(i)$, $i = 5, \ldots, 8$. Following the ideas developed in Sections 6.3 and 6.4.1 of Chapter 6, each regression vector $\boldsymbol{F}_t(i)$ is defined as follows.

$$\boldsymbol{F}_t(i)^T = [\text{pa}(y_t(i))\boldsymbol{L}_i^{D\top} \;\; \boldsymbol{L}_i^{O\top} \;\; \boldsymbol{L}_i^{S\top} \;\; \boldsymbol{L}_i^{H\top}], \quad i = 5, \ldots, 8, \tag{7.29}$$

where,

$$\begin{aligned}
\boldsymbol{L}_i^D &= [h_{i,1}^D(t-1), \ldots, \; h_{i,M1}^D(t-1)]^\top, \\
\boldsymbol{L}_i^O &= [h_{i,1}^O(t-1), \ldots, \; h_{i,M2}^O(t-1)]^\top, \\
\boldsymbol{L}_i^S &= [h_{i,1}^S(t-1), \ldots, \; h_{i,M3}^S(t-1)]^\top
\end{aligned}$$

and

$$\boldsymbol{L}_i^H = [h_{i,1}^H(t-1), \ldots, h_{i,M4}^H(t-1)]^\top$$

The regression vectors in (7.29) contains the spline basis vector $\boldsymbol{L}_i^D$ to model the proportion of vehicles going from $\text{pa}(y_t(i))$ to $y_t(i)$ as well as $\boldsymbol{L}_i^O$, $\boldsymbol{L}_i^S$ and $\boldsymbol{L}_i^H$, which are spline basis vectors representing the relationship between $y_t(i)$ and occupancy, speed and headway observed at site $S(i)$ at time $t-1$ respectively. Given this regression vector structure, each parameter vector associated with basis vectors $\text{pa}(y_t(i))\boldsymbol{L}_i^D$, $\boldsymbol{L}_i^O$, $\boldsymbol{L}_i^S$ and $\boldsymbol{L}_i^H$ can then be included in $\boldsymbol{\theta}_t(i)$. It is also assumed that $\boldsymbol{G}_t = \text{blockdiag}\{\boldsymbol{G}_t(5), \ldots, \boldsymbol{G}_t(8)\}$, where $\boldsymbol{G}_t(i) = \boldsymbol{I}_{ni}$, which is an identity matrix of order $ni = \dim(\boldsymbol{F}_t(i))$, $i = 5, \ldots, 8$.

A DAG for the $\boldsymbol{Y}_t$ can be obtained by removing the blue undirected edges from the CG in Figure 7.2. Given this DAG, a LMDM can be defined for $\boldsymbol{Y}_t$ such that each root node $Y_t(i)$, $i = 1, \ldots, 4$, follows a univariate DLM, as assumed in the previous chapters. For fairness of comparison with the LDCGM, occupancy, speed and headway observed at root nodes were not used as predictors in the DLMs for $Y_t(i)$, $i = 1, \ldots, 4$, and each observation variance $V_t(i)$, $i = 1, \ldots, 4$ was modelled using discount factor techniques only. Additionally, the seasonal pattern of each root node in the LMDM was modelled using splines with

$$\boldsymbol{F}_t(i)^\top = (h_1(t) \ \cdots \ h_M(t)) \quad \text{and} \quad \boldsymbol{\theta}_t(i)^\top = (\beta_{t1} \ \cdots \beta_{tM})$$

where each $\boldsymbol{F}_t(i)$ contains the spline basis functions representing the cycle from the period 15:00 to 19:59 for each $Y_t(i)$, $i = 1, \ldots, 4$.

Each non-root node in the LMDM was modelled by separate conditional DLMs for $Y_t(i)|\text{pa}(Y_t(i))$, $i = 5, \ldots, 8$, in the same fashion as these nodes were modelled within the LDCGM.

7.6.1 Results

Table 7.1 shows the log predictive likelihood (LPL) values for the LDCGM and the LMDM used to forecast a subset of the Manchester network as described above. The first row of the table shows LPLs based on forecasts from both models for the root nodes only. The matrix normal DLM for $\boldsymbol{Y}_t(R)$ considered under the LDCGM framework provided better forecasts than the independent DLMs assumed by the LMDM, and this improvement clearly affects the overall LPL when looking at all eight time series considered in the example (second row of Table 7.1).

TABLE 7.1: LPLs for LMDM and LDCGM using different $\boldsymbol{Y}_t$ subsets

node setup	model	
	LMDM	LDCGM
$\boldsymbol{Y}_t(R)$ (root nodes)	-6,728	-6,154
$\boldsymbol{Y}_t$ (all eight nodes)	-11,488	-10,914

Figure 7.12 shows observed flow for $Y_t(1)$ on 22 December 2010 from 15:00 to 19:59, together with one-step ahead forecast means and 90% forecast limits for both the LMDM and the LDCGM. These models provide very similar forecast means, hence the superposition of the lines representing the forecast means from both models in Figure 7.12. However, although their forecast limits follow the same pattern during the period covered by the plot (that is, bigger forecast ranges at periods with bigger flow variability) forecast limits from the LDCGM are slightly smaller than the ones from the LMDM and so slightly more informative.

Although the one-step ahead forecast means and limits are similar for the two models in Figure 7.12, the advantage of using the LDCGM can be clearly seen when looking at multivariate forecasts, which can be highly influenced by the covariance structure among the root nodes.

Figure 7.13 shows twelve plots, each with a single (two-dimensional) observed flow, represented by a black dot. Each black dot represents the flow observed during a

5-min interval during 18:00-18:59 on 29 December 2010 at the *pair* of root nodes $(Y_t(1), Y_t(2))$ of the Manchester network. The red ellipses in Figure 7.13 represent 90% forecast regions for root nodes $(Y_t(1), Y_t(2))$ based on the forecast distributions obtained from the LMDM, whereas the blue ellipses in Figure 7.13 represent the 90% forecast regions for the same nodes based on the forecast distributions obtained from the LDCGM. These forecast distributions would be obtained *before* the flows represented by black dots in the plots were observed.

For most 5-min intervals during 18:00-18:59, the observed flows are far from the centre of forecast regions for both models, which are represented by the grey dots in the plots (the centre of the forecast regions of both LMDM and LDCGM are very similar). This is possibly due to the extra variability in traffic flows during December 2010 which are not being captured by the models, since both LMDM and LDCGM use splines to model the *smoothed* daily flow trend only.

In all the twelve plots, the LDCGM forecast regions are smaller than the forecast regions based on the LMDM. Moreover, the LDCGM forecast regions clearly indicate a positive correlation among the node pairs in each plot: such positive correlations support the root node scatterplots shown in Figure 7.1 and discussed in Section 7.2. On the other hand, the LMDM does not exhibit such a correlation structure between pairs of root nodes.

Figure 7.13 also illustrates how the LDCGM accommodates the uncertainty due to unexpected traffic flow levels better than the LMDM does: the LDCGM managed to provide a forecast region with a smaller area than the LMDM forecast region, while better accommodating the uncertainty due to the high flow variations. Indeed, the LDCGM forecast region covers the flow observed during 18:05-18:09 and 18:20-18:24, while the LMDM forecast region does not.

7.7 Discussion

When using a DAG to represent the conditional independence statements related to causality across a time series vector of flows, the lack of edges among the root nodes leads to the asssumption that flows at the entrances of the network are *marginally* independent. This assumption does not seem to hold for traffic data, as described in Section 7.2.

The representation of the association structure between traffic flow time series by a chain graph, where the root nodes are connected by undirected edges, induces a symmetric association structure among these root nodes. In the example presented in Section 7.6, the assumption of symmetry among the root nodes represents the hypothesis that the adverse weather conditions observed during December 2010 affects all the root nodes of the chain graph in Figure 7.2 *equally*, and that each non-root node is independent of the adverse weather condition given its root node parent.

The elicitation of a chain graph for traffic networks is an extension of the DAG elicitation procedure described in Chapter 4. A chain graph can be obtained from the DAG by just encoding the research hypotheses representing symmetric associations between the root nodes. In the resulting chain graph, the chain components are the blocks formed by the root nodes and each non-root node. Thus, a partial causal ordering can be defined among these chain components.

Although the example described in Section 7.6 focuses on the problem of forecasting traffic flows, the DCGM has much wider applicability to any application involving multivariate time series which exhibit causal and symmetric association structures. For example, suppose that of interest is forecasting an econometric n-dimensional time series $\boldsymbol{Y}_t$, where $r = 2$, $n > 2$, and the root nodes $Y_t(1)$ and $Y_t(2)$ represent time series of gross domestic product and inflation rate for a country of interest. Suppose a chain graph $\mathcal{G}$ can be elicited for $\boldsymbol{Y}_t$, where an association structure represents

the symmetry between gross domestic product and inflation rate, which would be described by an undirected edge between $Y_t(1)$ and $Y_t(2)$ in $\mathcal{G}$. Assume further that the root nodes may be potential causes of the non-root nodes, and these causes would be represented by arcs from $\{Y_t(1), Y_t(2)\}$ to $Y_t(i)$ in $\mathcal{G}$, $i = r+3, \ldots, n$. A DCGM could be formulated for these series such that a multivariate single model is defined for $Y_t(1)$ and $Y_t(2)$ whereas the other series $Y_t(i)$, $i = r+3, \ldots, n$ are modelled by $n-2$ (conditional) separate univariate models.

The DCGM also extends the dynamic graphical model developed by Queen and Smith (1992) as follows. Suppose that a time series $\boldsymbol{Y}_t = [Y_t(1), \ldots, Y_t(8)]^{\mathrm{T}}$ is modelled by this DGM using a chain graph $\mathcal{G}$. Given two ordered p-sets (A, B), the DGM from Queen and Smith requires that, if there is a directed edge from a variable in a p-set A to a variable in B, then there must be a directed edge from *every* variable in A to *every* variable B. Based on the model definition given in Section 7.3, this requirement is no longer needed in the DCGM.

The flexibility of the DCGM allows the use of any multivariate model for the time series of root nodes. The traffic network considered for all the applications in this thesis has only four root nodes. However, when dealing with higher-dimensional time series, sparsity could be included in the root node modelling by considering a graphical model to represent the covariance structure among the root nodes, with the dynamic graphical model of Carvalho and West (2007), described in Section 4.3.2 of Chapter 4, as an example.

Due to the matrix normal DLM structure for root nodes in the LDCGM, variance laws (as in Section 5.2 of Chapter 5) and the use of extra traffic variables as predictors (as in Section 6.4 of Chapter 6) were not considered for the root nodes in the example of Section 7.6. Even with these restrictions, there was evidence that it is worth considering a chain graph for the root nodes when forecasting traffic flows. However, when modelling traffic data for the whole day, a better multivariate model must be

developed which uses the extra variables as predictors and which also accommodates the covariance structure of the root nodes when discount factor modelling is not appropriate.

The econometric example introduced earlier in which $Y_t(1)$ is gross domestic product and $Y_t(2)$ is inflation rate is another example of when an alternative multivariate model to the matrix normal DLM may be more appropriate for the root nodes. Although symmetry is assumed between gross domestic product and inflation rate, it is not necessarily the case that $Y_t(1)$ and $Y_t(2)$ should be modelled using a common regression vector $\boldsymbol{F}_t$ as required by a matrix normal DLM.

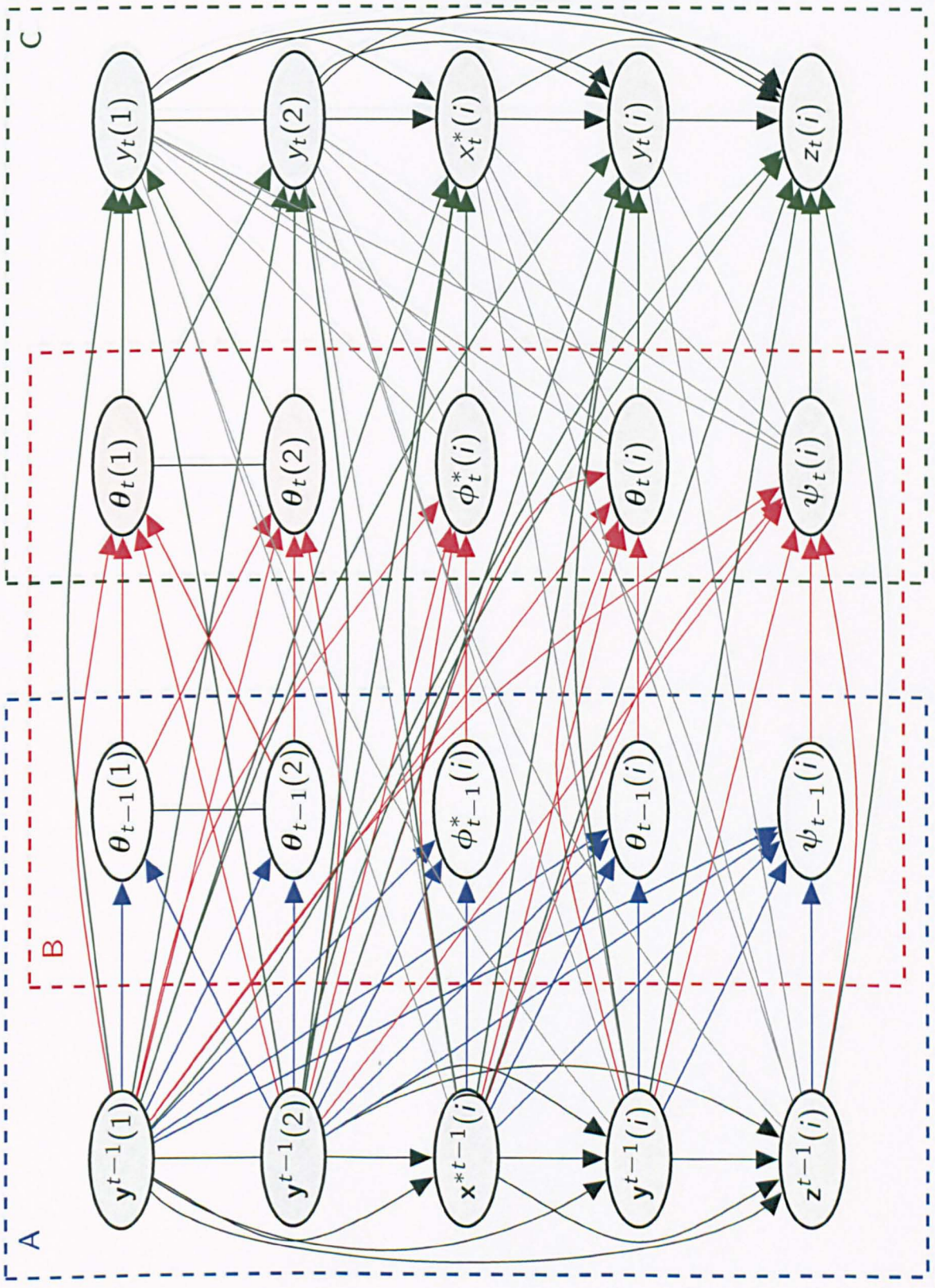

FIGURE 7.6: Moralized chain graph for the inductive hypothesis (7.11) to (7.14) (Box A), system equations (Box B) and observation equations (Box C) of the DCGM.

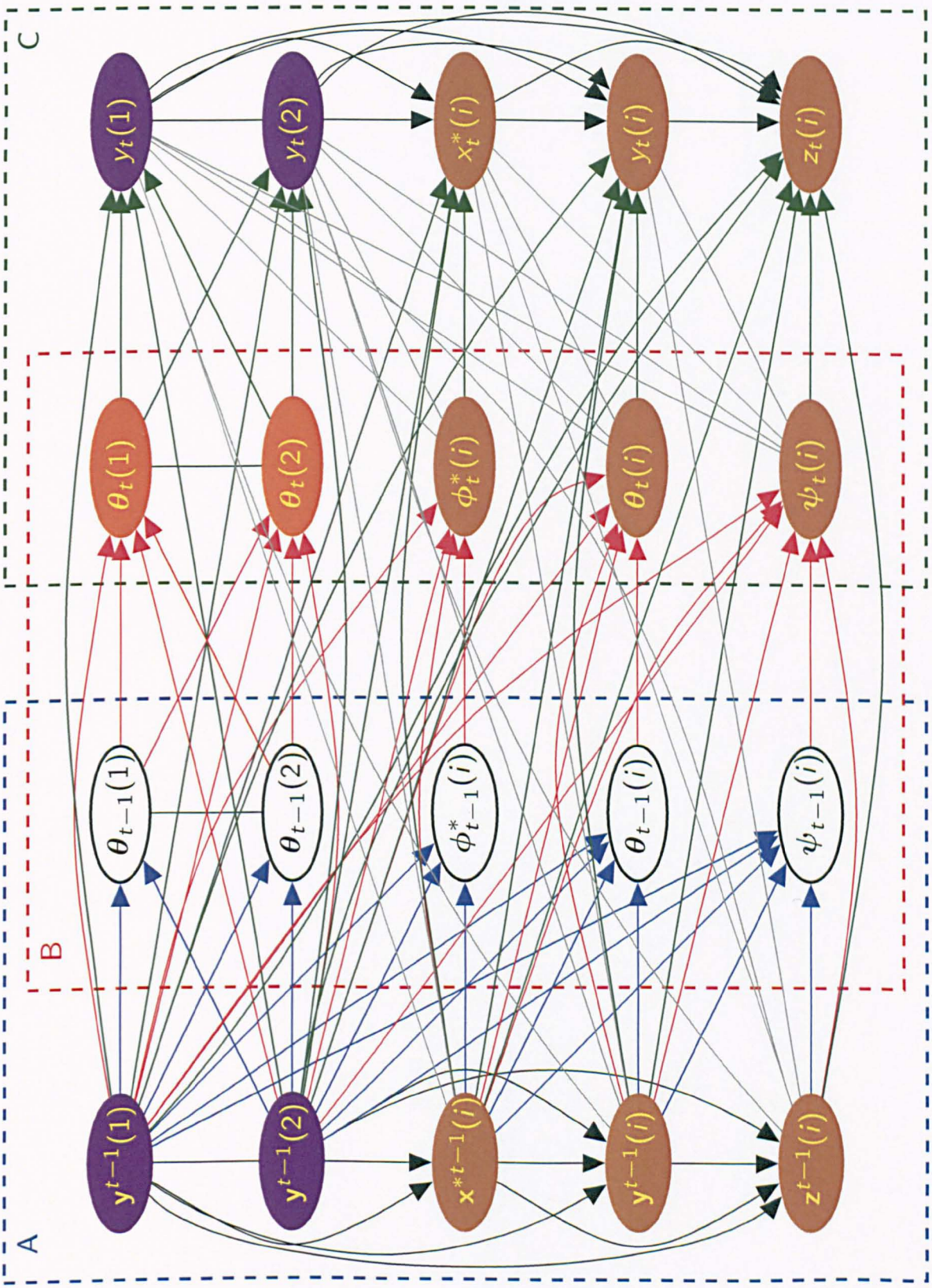

FIGURE 7.7: Moralized chain graph of Figure 7.6, highlighting statement (7.15), of the form $A \perp\!\!\!\perp B|C$, from theorem 7.2. C are the violet nodes and A and B are orange and brown nodes respectively

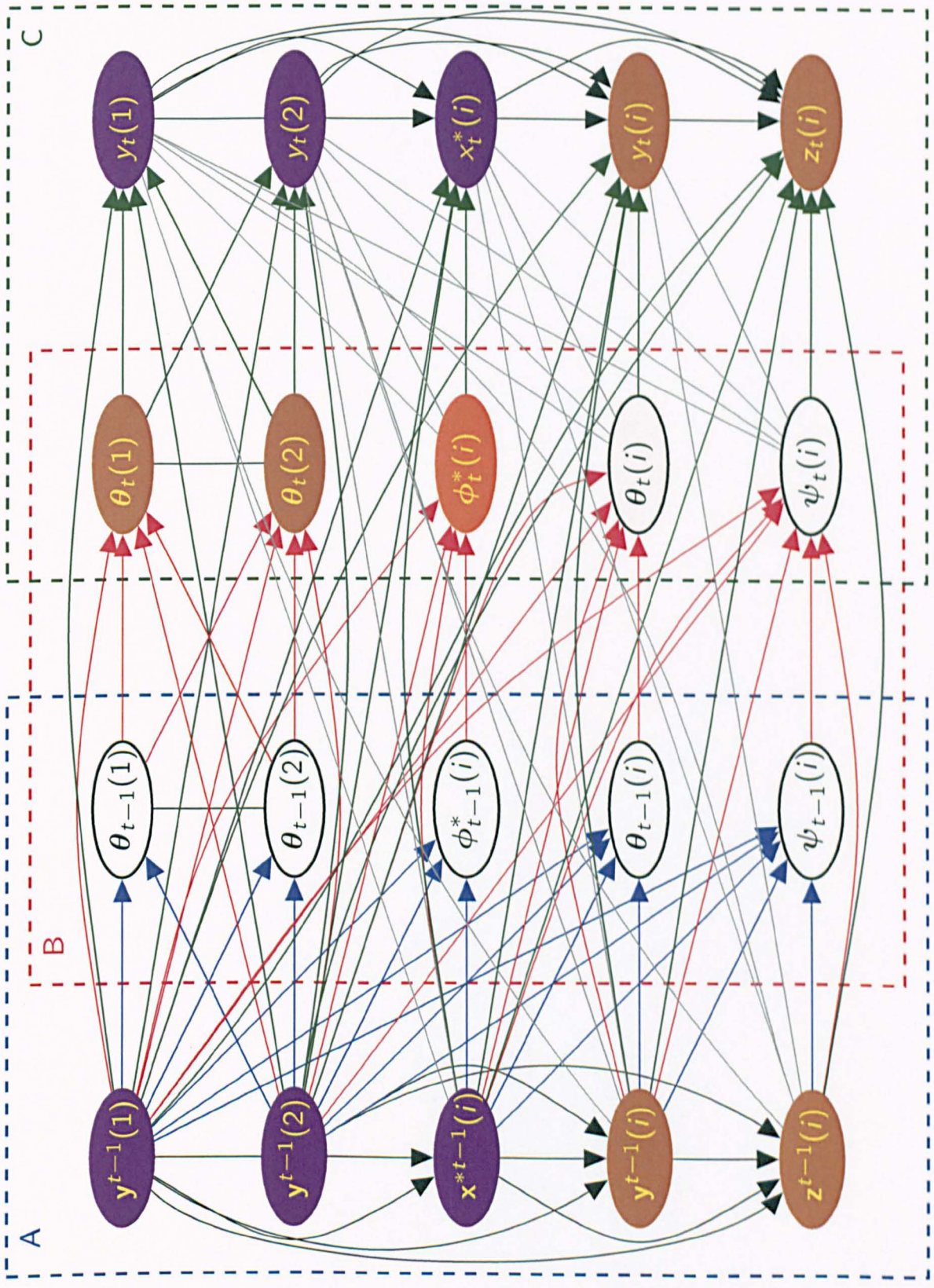

FIGURE 7.8: Moralized chain graph of Figure 7.6, highlighting statement (7.16), of the form $A \perp\!\!\!\perp B|C$, from theorem 7.2. C are the violet nodes and A and B are orange and brown nodes respectively

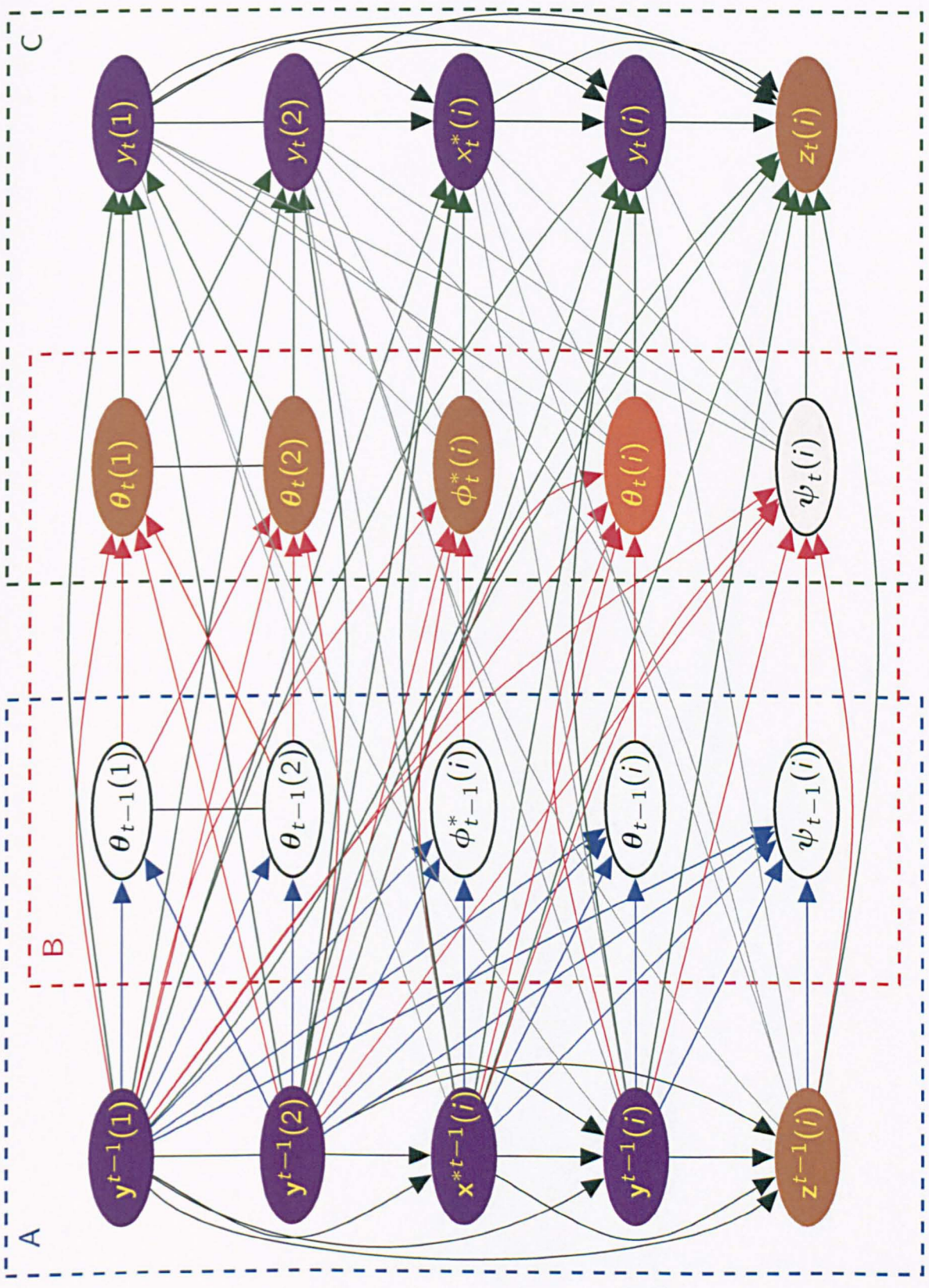

FIGURE 7.9: Moralized chain graph of Figure 7.6, highlighting statement (7.17), of the form $A \perp\!\!\!\perp B|C$, from Theorem 7.2. C are the violet nodes and A and B are orange and brown nodes respectively

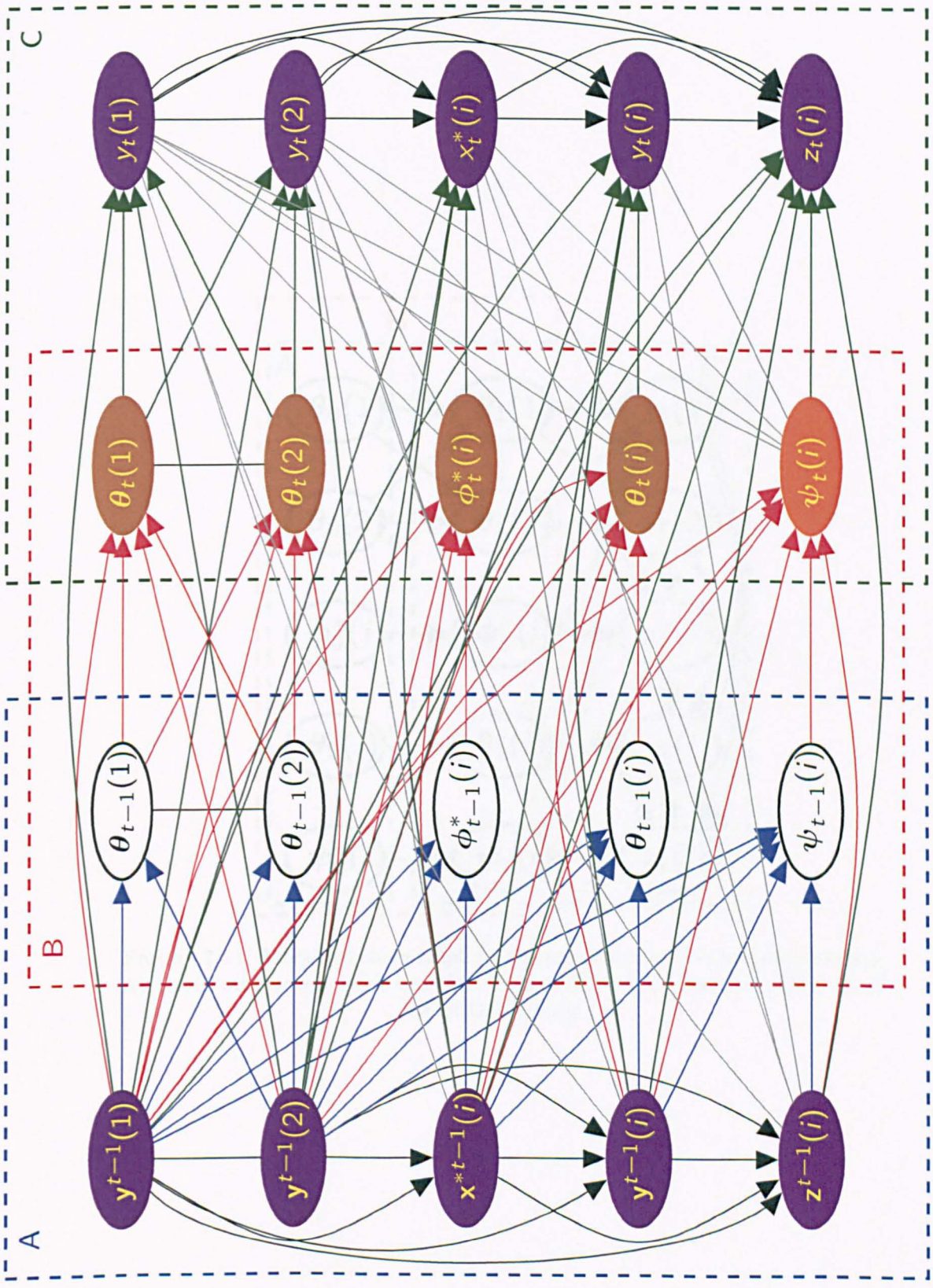

FIGURE 7.10: Moralized chain graph of Figure 7.6, highlighting statement (7.18), of the form $A \perp\!\!\!\perp B|C$, from Theorem 7.2. C are the violet nodes and A and B are orange and brown nodes respectively

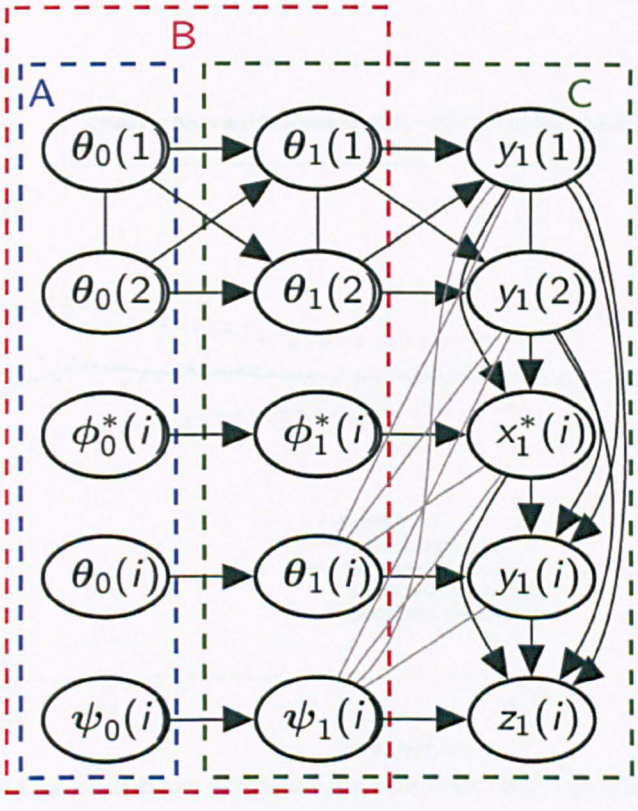

FIGURE 7.11: Moralized chain graph for the hypothesis of the Corollary 7.3 (Box A), along with system equations (Box B) and observation equations (Box C) of the DCGM at time 1.

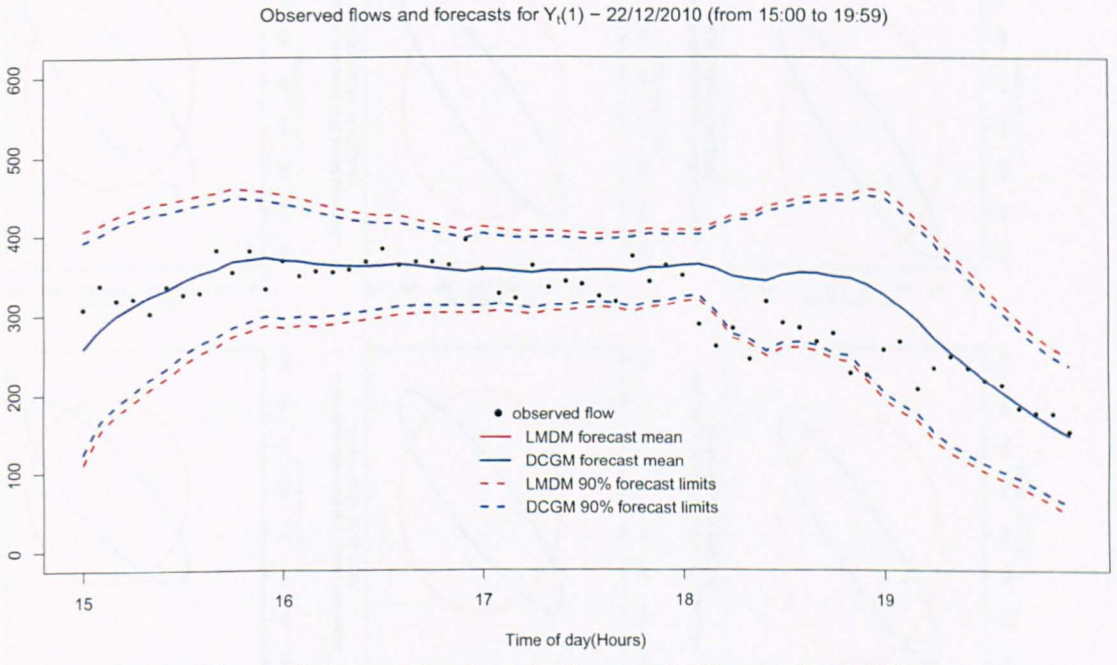

FIGURE 7.12: Observed flows and univariate forecast limits at site 9206B on 22 Dec 2010

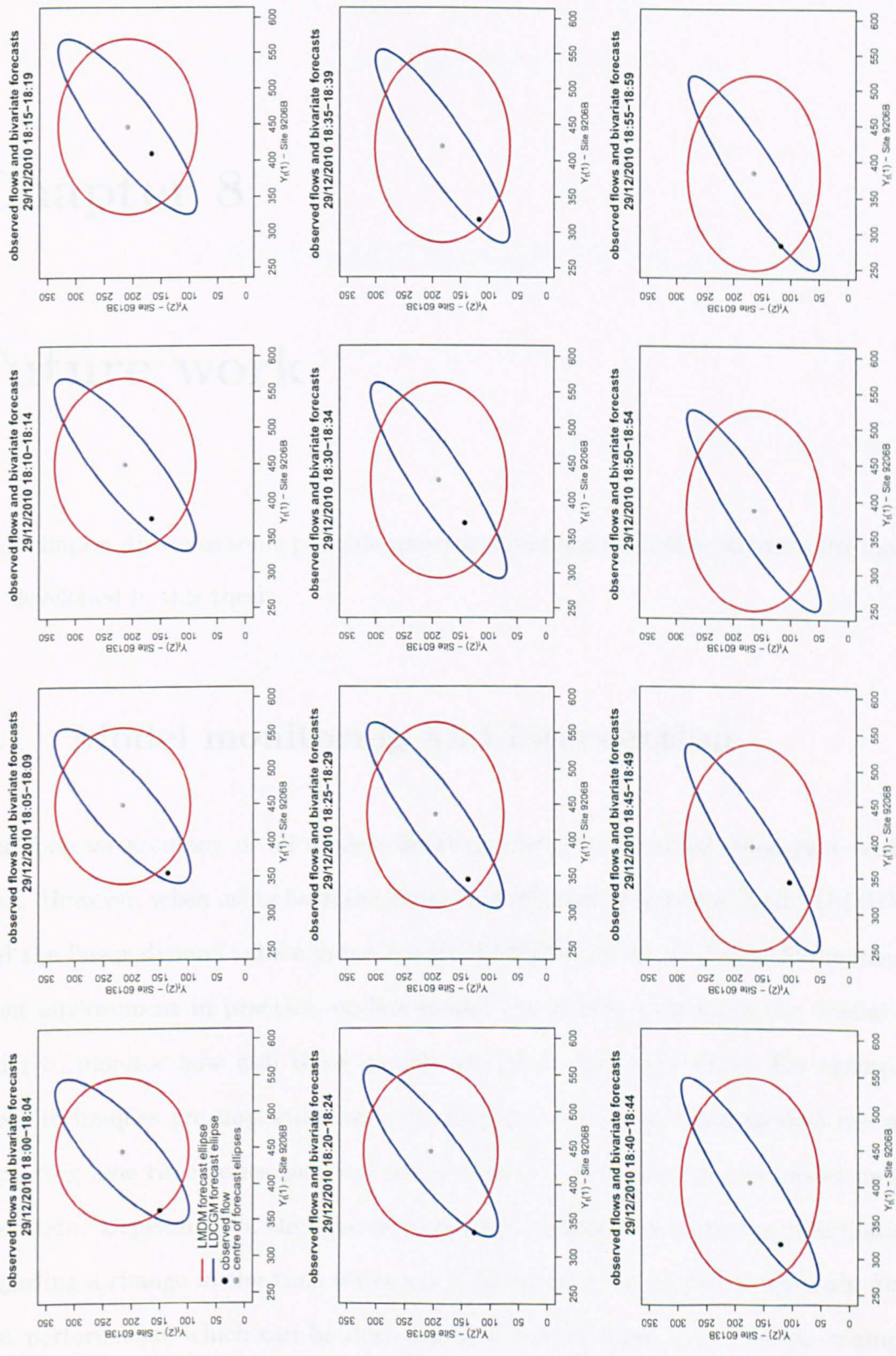

FIGURE 7.13: Observed flows and bivariate forecast limits at root node pair $(Y_t(1), Y_t(2))$ during 18:00-18:59 on 29 December 2010

Chapter 8

Future work

This chapter discusses some possible research directions to further improve the models developed in this thesis.

8.1 Model monitoring and intervention

The forecast accuracy of all models developed here was verified using past traffic data. However, when using both the linear multiregression dynamic model (LMDM) and the linear dynamic chain graph model (LDCGM) in an on-line traffic management environment in practice, on-line model monitoring techniques are crucial in order to monitor how well these models are performing over time. For example, these techniques are important tools to identify when forecasting models are not capturing new time series patterns not previously considered in the model under operation. Depending on the type or frequency of these new patterns, information regarding a change in the time series must be fed into the model to maintain forecast performance, which can be done via intervention techniques. For an example of intervention in LMDMs, see Queen and Albers (2009).

Given that both the LMDM and the LDCGM are sets of (conditional) dynamic linear models, it should be relatively straightforward to adapt established monitoring and intervention techniques for DLMs (as described in West and Harrison, 1997) into their context.

8.2 Using downstream flows to forecast upstream flows

When using dynamic linear models to forecast road traffic flows, Tebaldi *et al.* (2002) used lagged information from downstream sites to forecast flows at upstream sites. This improved forecasts at periods where unexpected low traffic flows were observed, and it can be an indication that flows from downstream sites can be useful when forecasting flows during congestion periods.

The effect of using lagged downstream flows to forecast upstream flows could be tested when using LMDMs and LDCGMs. When verifying this effect in LDCGMs, it should be noted that the root nodes in this model share the same regression vector $\boldsymbol{F}_t(R)$ in the matrix-normal model assumed for them (as explained in Section 7.6 of Chapter 7). In this case, the effect of using lagged downstream flows on forecasting upstream flows could be used for non-root nodes only.

Still from the traffic modelling perspective, the graphical structure of both multi-regression dynamic models and dynamic chain graph models allow the use of *contemporaneous* flows from upstream sites to model downstream sites, and the use of such information improves flow forecasts when compared to the inclusion of lagged flow information, as shown in Section 4.8.4 of Chapter 4. However, the inclusion of contemporaneous flows from downstream sites is still not possible, due to the conditional independence structure implied by these models, defined in Section 4.4

of Chapter 4 for the MDM and in Section 7.3 of Chapter 7 for the DCGM. In this sense, graphical structure adaptations of both MDMs and DCGMs need to be developed in order to allow contemporaneous information from downstream sites to forecast flows at upstream sites.

8.3 Modelling all weekdays

As shown in Section 2.4 of Chapter 2, the variability of flows may have different patterns depending on the day of week. Due to this, flows from Wednesdays only were considered when using the Manchester network data to apply all the methodologies proposed in this thesis. However, both the LMDM and the LDCGM can be further improved for forecasting traffic flows by considering variability between days. Extensions of these models can be possibly developed along the lines of *hierarchical dynamic models* (Gamerman and Migon, 1993), where the theory of hierarchical models introduced by Lindley and Smith (1972) are developed in the dynamic linear model context.

8.4 General chain graph structures for the dynamic chain graph model

In Section 7.3 of Chapter 7, the dynamic chain graph model (DCGM) was defined for a time series $\boldsymbol{Y}_t = [Y_t(1), \ldots, Y_t(n)]^\top$ with r root nodes, $r < n$. This definition assumes the time series being represented by a chain graph with blocks $B_1, \ldots, B_k$ such that the blocks $B_2, \ldots, B_k$ each have one time series component only. The DCGM uses a separate multivariate state space model for the root nodes $\boldsymbol{Y}_t(R) =$

$[Y_t(1), \ldots, Y_t(r)]^{\top}$, whereas conditionally univariate state space models are defined for each remaining time series component $Y_t(r+1), \ldots, Y_t(n)$.

The DCGM can be extended to consider the case where any of blocks $B_2, \ldots, B_k$ have more than one time series component. The main steps for this extension are defining a general graph to represent the conditional independence structures induced by the extended DCGM and proving that, in this general case, the state vectors $\boldsymbol{\theta}_t(i)$, $i = r+1, \ldots, n$, remain independent after sampling, together with the independence between each $\boldsymbol{\theta}_t(i)$ and state vector $\boldsymbol{\theta}_t(R)$ after $\boldsymbol{y}_t$ is observed.

Appendix A

Moments of LMDM one-step ahead forecast distributions

We derive in this appendix the first and second moments of both conditional and marginal one-step ahead forecast distributions of the LMDM example described in Section 4.7 of Chapter 4.

A.1 Mean and variance of conditional one-step ahead forecast distributions

The Bayesian sequential procedure for a DLM described in section 3.6.1.1 of Chapter 3 is used to obtain the moments of the forecast distribution of each $Y_t(i)$ conditional on its parents $\text{pa}(Y_t(i))$ as follows. Suppose that, at a given time t, the prior information for each $\boldsymbol{\theta}_t(i)$, $i = 1, 2, 3$, is

$$\boldsymbol{\theta}_t(i)|D_{t-1} \sim N(\boldsymbol{a}_t(i), \boldsymbol{R}_t(i)).$$

Hence, the means of the conditional forecast distributions of $Y_t(i)$ are, respectively,

$$E[Y_t(1)|D_{t-1}] = a_t^{(1)}(1),$$

$$E[Y_t(2)|D_{t-1}, y_t(1)] = a_t^{(1)}(2)y_t(1) \tag{A.1}$$

$$E[Y_t(3)|D_{t-1}, y_t(2)] = a_t^{(1)}(3)y_t(2),$$

where $a_t^{(1)}(i)$ is the first element of the vector $\boldsymbol{a}_t(i)$, which is the prior mean of $\boldsymbol{\theta}_t(i)$, $i = 1, 2, 3$. Writing these equations in a matrix form, we have,

$$\begin{bmatrix} E[Y_t(1)|D_{t-1}] \\ E[Y_t(2)|D_{t-1}, y_t(1)] \\ E[Y_t(3)|D_{t-1}, y_t(2)] \end{bmatrix} = \begin{bmatrix} a_t^{(1)}(1) \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ a_t^{(1)}(2) & 0 & 0 \\ 0 & a_t^{(1)}(3) & 0 \end{bmatrix} \begin{bmatrix} y_t(1) \\ y_t(2) \\ y_t(3) \end{bmatrix} \tag{A.2}$$

The variance of the conditional distributions for $Y_t(i)$, $i = 1, 2, 3$, are, respectively,

$$[Y_t(1)|D_{t-1}] = \boldsymbol{F}_t(1)^\top \boldsymbol{R}_t(1)\boldsymbol{F}_t(1) + S_{t-1}(1) \tag{A.3}$$

$$V[Y_t(2)|D_{t-1}, y_t(1)] = \boldsymbol{F}_t(2)^\top \boldsymbol{R}_t(2)\boldsymbol{F}_t(2) + S_{t-1}(2) \tag{A.4}$$

$$V[Y_t(3)|D_{t-1}, y_t(2)] = \boldsymbol{F}_t(3)^\top \boldsymbol{R}_t(3)\boldsymbol{F}_t(3) + S_{t-1}(3) \tag{A.5}$$

where $\boldsymbol{R}_t(i)$ is the prior covariance for $\boldsymbol{\theta}_t(i)$ and $S_{t-1}(i)$ is a posterior estimate for $V_t(i)$ at time $t-1$, obtained via conjugate Bayesian inference for the precision $\phi(i) = 1/V_t(i)$, $i = 1, 2, 3$ (again, as described in section 3.6.1.1 of Chapter 3) Now, for any matrices $\boldsymbol{U}, \boldsymbol{Z} \in \mathbb{R}^n$, we have the identity

$$\boldsymbol{U}^\top \boldsymbol{Z} = \text{trace}\{\boldsymbol{U}\boldsymbol{Z}^\top\} \tag{A.6}$$

Setting $\boldsymbol{B}_t(i)\boldsymbol{B}_t(i)^\top = \boldsymbol{R}_t(i)$, $i = 2, 3$, and applying identity (A.6) in equations (A.4) and (A.5), the conditional forecast variances for $Y_t(2)$ and $Y_t(3)$ can be written as,

respectively,

$$V[Y_t(2)|D_{t-1}, y_t(1)] = trace\{\boldsymbol{B}_t(2)^T \boldsymbol{F}_t(2)\boldsymbol{F}_t(2)^\top \boldsymbol{B}_t(2)\} + S_{t-1}(2) \qquad \text{(A.7)}$$

and

$$V[Y_t(3)|D_{t-1}, y_t(2)] = trace\{\boldsymbol{B}_t(3)^T \boldsymbol{F}_t(3)\boldsymbol{F}_t(3)^\top \boldsymbol{B}_t(3)\} + S_{t-1}(3). \qquad \text{(A.8)}$$

A.2 Mean and variance of marginal one-step ahead forecast distributions

Since the time series that compose the vector $\boldsymbol{Y}_t$ are observed simultaneously and so $\text{pa}(Y_t(i))$ is not known when forecasting $\boldsymbol{y}_t$, the LMDM uses the moments of the marginal forecast distribution for each $Y_t(i)$, $i = 1, 2, 3$ (unconditional on $\text{pa}(Y_t(i))$) for one-step ahead forecasting. Using the identity (4.14), we can write

$$E[\boldsymbol{Y}_t|D_{t-1}] = E\left\{ \begin{bmatrix} E[Y_t(1)|D_{t-1}] \\ E[Y_t(2)|D_{t-1}, y_t(1)] \\ E[Y_t(3)|D_{t-1}, y_t(2)] \end{bmatrix} |D_{t-1} \right\},$$

and, from equation (A.2),

$$E[\boldsymbol{Y}_t|D_{t-1}] = E\left\{ \begin{bmatrix} a_t^{(1)}(1) \\ 0 \\ 0 \end{bmatrix} |D_{t-1} \right\} + E\left\{ \begin{bmatrix} 0 & 0 & 0 \\ a_t^{(1)}(2) & 0 & 0 \\ 0 & a_t^{(1)}(3) & 0 \end{bmatrix} \begin{bmatrix} Y_t(1) \\ Y_t(2) \\ Y_t(3) \end{bmatrix} |D_{t-1} \right\}.$$

Hence,

$$\begin{bmatrix} E[Y_t(1)|D_{t-1}] \\ E[Y_t(2)|D_{t-1}] \\ E[Y_t(3)|D_{t-1}] \end{bmatrix} = \begin{bmatrix} a_t^{(1)}(1) \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ a_t^{(1)}(2) & 0 & 0 \\ 0 & a_t^{(1)}(3) & 0 \end{bmatrix} \begin{bmatrix} E[Y_t(1)|D_{t-1}] \\ E[Y_t(2)|D_{t-1}] \\ E[Y_t(3)|D_{t-1}] \end{bmatrix}$$

and the marginal forecast means for $\boldsymbol{Y}_t$ are

$$E[\boldsymbol{Y}_t|D_{t-1}] = \begin{bmatrix} a_t^{(1)}(1) \\ a_t^{(1)}(1)a_t^{(1)}(2) \\ a_t^{(1)}(1)a_t^{(1)}(2)a_t^{(1)}(3) \end{bmatrix}. \tag{A.9}$$

The variance of $Y_t(1)$ is given by (A.3) and, being a root node, its marginal forecast moments are not required. The marginal variances for $Y_t(2)$ and $Y_t(3)$ can be derived by applying the identity (4.15) in equations (A.7) and (A.8), resulting in

$$\begin{aligned} V[Y_t(2)|D_{t-1}] &= E\{V[Y_t(2)|D_{t-1}, y_t(1)]|D_{t-1}\} + V\{E[Y_t(2)|D_{t-1}, y_t(1)]|D_{t-1}\} \\ &= E\{trace\{\boldsymbol{B}_t(2)^T \boldsymbol{F}_t(2)\boldsymbol{F}_t(2)^\top \boldsymbol{B}_t(2)\}|D_{t-1}\} + S_{t-1}(2) \\ &\quad + V\{a_t^{(1)}(2)Y_t(1)|D_{t-1}\} \end{aligned}$$

from equations (A.7) and (A.1). Then, we can write $V[Y_t(2)|D_{t-1}]$ as

$$E\left\{\text{trace}\left\{\begin{bmatrix} b_{(1,1),t}(2) & b_{(2,1),t}(2) & \dots & b_{(24,1),t}(2) \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ b_{(1,24),t}(2) & b_{(2,24),t}(2) & \dots & b_{(24,24),t}(2) \end{bmatrix} \begin{bmatrix} Y_t(1) \\ 0 \\ \vdots \\ 0 \end{bmatrix} [Y_t(1)\ 0 \dots 0] \times \right.\right.$$

$$\left.\left. \times \begin{bmatrix} b_{(1,1),t}(2) & b_{(1,2),t}(2) & \dots & b_{1,24}(2) \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ b_{(24,1),t}(2) & b_{(24,2),t}(2) & \dots & b_{(24,24),t}(2) \end{bmatrix} \right\} |D_{t-1} \right\} +$$

$$+ S_{t-1}(2) + \{a_t^{(1)}(2)\}^2 V\{Y_t(1)|D_{t-1}\}$$

where $b_{(i,j),t}(2)$ is the $(i,j)^{th}$ element of the matrix $\boldsymbol{B}_t(2)$. Therefore, we can write the marginal variance of $Y_t(2)|D_{t-1}$ as

$$\begin{aligned} V[Y_t(2)|D_{t-1}] &= \left\{\sum_{i=1}^{24}(b_{(1,i),t}(2)^2)\right\} E[Y_t(1)^2|D_{t-1}] + S_{t-1}(2) \\ &\quad + \left\{a_t^{(1)}(2)\right\}^2 V[Y_t(1)|D_{t-1}] \\ &= \left\{\sum_{i=1}^{24}(b_{(1,i),t}(2)^2)\right\} \{V[Y_t(1)|D_{t-1}] + E[Y_t(1)|D_{t-1}]^2\} + S_{t-1}(2) \\ &\quad + \left\{a_t^{(1)}(2)\right\}^2 V[Y_t(1)|D_{t-1}]. \end{aligned}$$

Similarly, we have

$$\begin{aligned} V[Y_t(3)|D_{t-1}] &= \left\{\sum_{i=1}^{24}(b_{(1,i),t}(3)^2)\right\} \{V[Y_t(2)|D_{t-1}] + E[Y_t(2)|D_{t-1}]^2\} \\ &\quad + S_{t-1}(3) + \left\{a_t^{(1)}(3)\right\}^2 V[Y_t(2)|D_{t-1}]. \end{aligned}$$

Essentially, in this LMDM, the marginal forecast moments $Y_t(1)$ are used to obtain the marginal forecast moments for $Y_t(2)$, which in turn are used to find the marginal

forecast moments of $Y_t(3)$.

The marginal covariance between any $Y_t(i)$ and $Y_t(j)$ is

$$\text{Cov}(Y_t(i), Y_t(j)|D_{t-1}) = E(Y_t(i), Y_t(j)|D_{t-1}) - E(Y_t(i)|D_{t-1})E(Y_t(j)|D_{t-1}). \quad \text{(A.10)}$$

Using (A.2) and (A.9), and applying identity (4.16) in (A.10), we have

$$\begin{aligned}
\text{Cov}(Y_t(1), Y_t(2)|D_{t-1}) &= E(Y_t(1), Y_t(2)|D_{t-1}) - E(Y_t(1)|D_{t-1})E(Y_t(2)|D_{t-1}) \\
&= E(Y_t(1)E[Y_t(2)|y_t(1), D_{t-1}]|D_{t-1}) - a_t^{(1)}(1)^2 a_t^{(1)}(2) \\
&= E(Y_t(1)a_t^{(1)}(2)Y_t(1)|D_{t-1}) - a_t^{(1)}(1)^2 a_t^{(1)}(2) \\
&= a_t^{(1)}(2)E(Y_t(1)^2|D_{t-1}) - a_t^{(1)}(1)^2 a_t^{(1)}(2) \\
&= a_t^{(1)}(2)[V(Y_t(1)|D_{t-1}) + E(Y_t(1)|D_{t-1})^2] - a_t^{(1)}(1)^2 a_t^{(1)}(2),
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(Y_t(1), Y_t(3)|D_{t-1}) &= E(Y_t(1), Y_t(3)|D_{t-1}) - E(Y_t(1)|D_{t-1})E(Y_t(3)|D_{t-1}) \\
&= E(Y_t(3)E[Y_t(1)|y_t(3), D_{t-1}]|D_{t-1}) - E(Y_t(1)|D_{t-1})E(Y_t(3)|D_{t-1}) \\
&= E(Y_t(1)|D_{t-1})E(Y_t(3)|D_{t-1}) - E(Y_t(1)|D_{t-1})E(Y_t(3)|D_{t-1}) \\
&= 0
\end{aligned}$$

and

$$\begin{aligned}
Cov(Y_t(2), Y_t(3)|D_{t-1}) &= E(Y_t(2), Y_t(3)|D_{t-1}) - E(Y_t(2)|D_{t-1})E(Y_t(3)|D_{t-1}) \\
&= E(Y_t(2)E[Y_t(3)|y_t(2)]|D_{t-1}) - a_t^{(1)}(1)^2 a_t^{(1)}(2)^2 a_t^{(1)}(3) \\
&= E(Y_t(2)a_t^{(1)}(3)Y_t(2)|D_{t-1}) - a_t^{(1)}(1)^2 a_t^{(1)}(2)^2 a_t^{(1)}(3) \\
&= a_t^{(1)}(3)E(Y_t(2)^2|D_{t-1}) - a_t^{(1)}(1)^2 a_t^{(1)}(2)^2 a_t^{(1)}(3) \\
&= a_t^{(1)}(3)\{V(Y_t(2)|D_{t-1}) + E(Y_t(2)|D_{t-1})^2\} \\
&\quad -a_t^{(1)}(1)^2 a_t^{(1)}(2)^2 a_t^{(1)}(3).
\end{aligned}$$

A.3 Moments of logical variables

The marginal moments calculated above can be used to obtain the marginal mean and variance for the logical variables, given as

$$E(Y_t(4)|D_{t-1}) = E(Y_t(1)|D_{t-1}) - E(Y_t(2)|D_{t-1}),$$
$$V(Y_t(4)|D_{t-1}) = V(Y_t(1)|D_{t-1}) + V(Y_t(2)|D_{t-1}) - 2Cov(Y_t(1), Y_t(2)|D_{t-1})$$

and

$$E(Y_t(5)|D_{t-1}) = E(Y_t(2)|D_{t-1}) - E(Y_t(3)|D_{t-1}),$$
$$V(Y_t(5)|D_{t-1}) = V(Y_t(2)|D_{t-1}) + V(Y_t(3)|D_{t-1}) - 2Cov(Y_t(2), Y_t(3)|D_{t-1}).$$

The marginal covariance between the logical variables $Y_t(4)$ and $Y_t(5)$ is as follows (assume that all expectations and covariances in (A.11) are conditioned on D_{t-1}).

$$\begin{aligned}
\text{Cov}\{Y_t(4), Y_t(5)\} &= E\{Y_t(4)Y_t(5)\} - E\{Y_t(4)\}E\{Y_t(5)\} \\
&= E\{[Y_t(1) - Y_t(2)][Y_t(2) - Y_t(3)]\} \\
&\quad -E\{Y_t(1) - Y_t(2)\}E\{Y_t(2) - Y_t(3)\} \\
&= E\{Y_t(1)Y_t(2)\} - E\{Y_t(1)Y_t(3)\} - E\{\{Y_t(2)\}^2\} + E\{Y_t(2)Y_t(3)\} \\
&\quad -E\{Y_t(1)\}E\{Y_t(2)\} + E\{Y_t(1)\}E\{Y_t(3)\} + \{E\{Y_t(2)\}\}^2 \\
&\quad -E\{Y_t(2)\}E\{Y_t(3)\} \\
&= \text{Cov}\{Y_t(1), Y_t(2)\} - \text{Cov}\{Y_t(1), Y_t(3)\} - \text{V}\{Y_t(2)\} \\
&\quad -\text{Cov}\{Y_t(2), Y_t(3)\} \\
&= \text{Cov}\{Y_t(1), Y_t(2)\} - \text{V}\{Y_t(2)\} - \text{Cov}\{Y_t(2), Y_t(3)\}.
\end{aligned}$$

The remaining covariances can be easily obtained by applying identity (4.16) in equation (A.10).

Bibliography

Abdulhai, B., Porwal, H., and Recker, W. (1999). Short-term freeway traffic flow prediction using genetically-optimized time-delay-based neural networks. *Working Papers, California Partners for Advanced Transit and Highways (PATH), Institute of Transportation Studies (UCB)*, UC Berkeley.

Ahmed, M. S., and Cook, A. R. (1979). Analysis of freeway traffic time series data by using Box-Jenkins techniques. *Transportation Research Record*, **773**, 47-49.

Algers, S., Bernauer, E., Boero, M., Breheret, L., Di Taranto, C., Dougherty, M., Fox, K., and Gabard, J.F. (1997). Review of micro-simulation models. Review report of the SMARTEST project. Available at http://www.its.leeds.ac.uk/smartest/deliv3.html [Accessed June 15, 2010].

Ameen, J.R.M., and Harrison, P.J. (1985). Normal discount Bayesian models. In *Bayesian Statistics 2.* J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith (Eds.), Oxford University Press, Oxford, 741-751.

Ashton, W. D. (1966). *Theory of Road Traffic Flow.* Methuen's Monographs on Applied Probability and Statistics. John Wiley, London.

Bellomo, N., Delitala, M., and Coscia, V. (2002). On the mathematical theory of vehicular traffic flow - I. Fluid dynamic and kinetic modelling. *Mathematical Models and Methods in Applied Sciences*, **12**, 1801-1843.

Bickel, P., Chen, C., Kwon, J., Rice, J., Van Zwet, E., and Varaiya, P. (2007). Measuring traffic. *Statistical Science*, **22(4)**, 581-597.

Bijleveld, F., Commandeur, J., Koopman, S.J., and van Montfort K. (2010). Multivariate non-linear time series modeling of exposure and risk in road safety research. *Journal of the Royal Statistical Society (Series C)*, **59**, 145-161.

Box, G., and Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

Boylan, I.E., and Johnston, F.R (1996). Variance laws for inventory management. *International Journal of Production Economics*, **45**, 343-352.

Brillinger, D.R. (1996). Remarks concerning graphical models for time series and point processes. *Revista de Econometria*, **16**, 1-23.

Carvalho C. M., and West M. (2007). Dynamic matrix-variate graphical models. *Bayesian Analysis*, **2**, 69-98.

Castro-Neto M., Jeong Y. S., Jeong M. K., and Han, L. D. (2009). Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert Systems with Applications*, **36(2)**, 6164-6173.

Chandler, R. E., Herman, R., and Montroll, E. W. (1958). Traffic dynamics: studies in car following. *Operations Research*, **6**, 165-184.

Chandra, S. R., and Al-Deek, H. (2009). Predictions of freeway traffic speeds and volumes using vector autoregressive models. *Journal of Intelligent Transportation Systems*, **13**, 53-72.

Chang, H., Lee, Y., Yoon, B., and Baek, S. (2012). Dynamic near-term traffic flow prediction: system oriented approach based on past experiences. *Intelligent Transport Systems, (IET)*, **6**, 292-305.

Chatfield, C. (2001). *Time Series Forecasting.* Chapman & Hall, London.

Chatfield, C. (2003). *The Analysis of Time Series: An Introduction.* Chapman & Hall, London.

Chen, C., Kwon, J., Rice, J, Skabardonis, A., and Varaiya, P. (2003). Detecting errors and imputing missing data for single loop surveillance systems. *Transportation Research Record*, **1855**, 160-167.

Chen, H., and Grant-Muller, S. (2001). Use of sequential learning for short-term traffic flow forecasting. *Transportation Research Part C*, **9**, 319-336.

Clark, S. (2003). Traffic prediction using multivariate non-parametric regression. *Journal of Transportation Engineering*, **129(2)**, 161-168.

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J. (1999). *Probabilistic Networks and Expert Systems.* Springer, New York.

Cox, D. R. (2006). *Principles of Statistical Inference.* Cambridge University Press, Cambridge.

Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika*, **51**, 157-172.

Dahlhaus, R., and Eichler, M. (2003). Causality and graphical models for time series. In: *Highly Structured Stochastic Systems*, eds. P. Green, N. Hjort, and S. Richardson, Oxford University Press, Oxford, 115-137.

Dawid, A. P. (1979). Conditional independence in statistical theory (with Discussion). *Journal of the Royal Statistical Society (Series B)*, **41**, 1-31.

Dawid, A.P. (1981). Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika*, **68**, 265-274.

Dawid, A. P., and Lauritzen, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Annals of Statistics*, **21(3)**, 1272-1317.

Dawid, A. P. (2002). Influence diagrams for causal modelling and inference. *International Statistical Review*, **70(2)**, 161-189.

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

Edie, L. C. (1961). Car-following and steady-state theory for noncongested traffic. *Operations Research*, **9**, 66-76.

Edwards, D. (2000). *Introduction to Graphical Modelling.* Springer Texts in Statistics, Springer-Verlag, New York (second edition).

Eichler, M. (2000). Granger-causality graphs for multivariate time series. Technical report, University of Heidelberg, Germany.

Fox, J., and Weisberg, S. (2011). *An R Companion to Applied Regression.* Sage, Thousand Oaks, CA (second edition).

Frydenberg, M. (1990). The chain graph markov property. *Scandinavian Journal of Statistics*, **17(4)**, 333-353.

Gamerman, D., and Migon, H. Dynamic hierarchical models. *Journal of the Royal Statistical Society (Series B)*, **55**, 629-642.

Gartner, N., Messer, C. J., and Rathi, A. K. (2001). Traffic flow theory, a state-of-the-art report. *TRB Committee on Traffic Flow Theory and Characteristics (AHB45)*, Revised Ed. Available at http://www.tfhrc.gov/its/tft/tft.htm. [Accessed June 15, 2010].

Gazis, D. C., Herman, R., and Rothery, R. W. (1961). Nonlinear follow the leader models of traffic flow. *Operations Research*, **9**, 545-567.

Gelman, A., Carlin, B., Stern, H.S., and Rubin, D.B., (1995). *Bayesian Data Analysis.* Chapman & Hall, London.

Gibbens, R., and Werft, W. (2005). Data gold mining: MIDAS and journey time predictors. *Significance*, **2**, 102-105.

Gilchrist, R. (1988). Three-dimensional relationships in traffic flow theory variables. Unpublished Master's degree report, Department of Civil Engineering, McMaster University, Hamilton, Ontario, Canada.

Gilchrist, R. S., and Hall, F. L. (1989). Three dimensional relationships among traffic flow theory variables. *Transportation Research Record*, **1225**, 99-108.

Gilmore, J. F., and Abe, N. (1995). Neural network models for traffic control and congestion prediction. *Journal of Intelligent Transportation Systems*, **3**, 231-252.

Gneiting, T., and Raftery, A.E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359-378.

Greenberg, H. (1959). An analysis of traffic flow. *Operations Research*, **7**, 78-85.

Greenshields, B. D. (1934). A photographic method of studying traffic behavior. *Highway Research Board Proceedings*, **13**, 382-399.

Greenshields, B. D (1935). A study of traffic capacity. *Highway Research Board Proceedings*, **14**, 448-474.

Hamed, M. M., Al-Masaeid, H. R., and Bani Said, Z. M. (1995). Short-term prediction of traffic volume in urban arterials. *ASCE Journal of Transportation Engineering*, **121(3)**, 249-254.

Harrell, F. E. (2001). *Regression modelling strategies: with applications to linear models, logistic regression and survival analysis.* Springer-Verlag, New York.

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

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning.* Springer-Verlag, New York.

Helbing, D. (1996). Gas-kinetic derivation of Navier-Stokes-like traffic equations. *Physical Review E*, **53(3)**, 2266-2381.

Herman, R., and Potts, R. B. (1961). Single-lane traffic theory and experiment *in: Theory of traffic flow*, R. Herman (ed.), Elsevier, Amsterdam.

Highways Agency (2012a). *Highways Agency's future delivery programmes.* Available at: http://www.highways.gov.uk/our-road-network/managing-our-roads/highways-agencys-future-delivery-programmes/ [Accessed October 02 2012].

Highways Agency (2012b). *Managed Motorways.* Available at: http://www.highways.gov.uk/our-road-network/managing-our-roads/managed-motorways/ [Accessed October 02 2012].

Highway Capacity Manual (2000). (Washington DC: Transportation Research Board, National Research Council)

Hoogendoorn, S. P., and Bovy, P. H. L. (2001). State-of-the-art of vehicular traffic flow modelling. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, **215 (4)**, 2041-3041.

Huang, S., and Sadek, A. W. (2009). A novel forecasting approach inspired by human memory: the example of short-term traffic volume forecasting. *Transportation Research Part C: Emerging Technologies*, **17(5)**, *510-525.*

Hurdle, V. F, Mark I. Merlo, M. I., and Robertson, R. D. (1997). Study of speed-flow relationships on individual freeway lanes. *Transportation Research Record: Journal of the Transportation Research Board,* **1591**, 7-13.

Innamaa, S. (2000). Short-term prediction of traffic situation using MLP-neural networks, *in: Proceedings of the 7th World Congress on Intelligent Transportation Systems*, Turin, Italy.

Jiang, X., and Adeli, H. (2005). Dynamic wavelet neural network model for traffic flow forecasting. *Journal of Transportation Engineering*, **10**, 771-779.

Jones, B., Carvalho, C. M. , Dobra, A., Hans, C., Carter, C., and West, M. (2005). Experiments in stochastic computation for high-dimensional graphical models. *Statistical Science*, **20(4)**, 388-400.

Jordan, M. (2004). Graphical models. *Statistical Science*, **19(1)**, 140-155.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering (Transactions ASME)*, **82**, 35-45.

Kamarianakis, Y., Kanas, A., and Prastacos, P. (2005). Modeling traffic flow volatility dynamics in an urban network. *Transportation Research Record*, **1923**, 18-27.

Karlaftis, M. G., and Vlahogianni, E. I. (2011). Statistical methods versus neural networks in transportation research: differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, **19(3)**, 387-399.

Korb, K. B., and Nicholson, A. E (2010). *Bayesian Artificial Intelligence.* Chapman & Hall, London (second edition).

Kühne, R. (2008). Foundations of traffic flow theory I: Greenshields legacy highway traffic. In: *Proceedings Symposium on the Fundamental Diagram: 75 years.* Woods Hole, Massachusetts.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N., and Leimer, H. G, (1990). Independence properties of directed Markov fields. *Networks*, **20**, 491-505.

Lauritzen, S. L. (1996). *Graphical Models.* Clarendon Press, Oxford.

Lauritzen, S. L. (2003). Some modern applications of graphical models. In: *Highly Structured Stochastic Systems*, eds. P. Green, N. Hjort, and S. Richardson, Oxford University Press, Oxford, 115-137.

Leutzbach, W. (1988). *An Introduction to the Theory of Traffic Flow.* Springer-Verlag, Berlin.

Levin, M., and Tsao, Y. D. (1980). On forecasting freeway occupancies and volumes. *Transportation Research Record*, **773**, 47-49.

Lighthill, M. H., and Whitham, G. B. (1955). On kinematic waves II: a theory of traffic flow on long, crowded roads. *Proc. R. Soc. Lond. Ser. A*, **229**, 317-345.

Lindley D. V., and Smith A. F. M. (1972). Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society (Series B)*, **34**, 1-41.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325-337.

Mai, T., Ghosh, B., and Wilson, S.P. (2011). Multivariate short-term traffic flow forecasting using Bayesian vector autoregressive moving average model. Technical Report, Trinity College Dublin (TCD-Statistics-11-05).

Min, W., and Wynter, L. (2011). Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, **19(4)**, 606-616.

Nihan, N. L., and Holmesland, K. O. (1980). Use of the Box and Jenkins time series techniques in traffic forecasting. *Transportation*, **9**, 25-143.

Nobre F.F., Monteiro A.B., Telles P.R., Williamson G.D (2001). Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology. *Statistics in Medicine*, **20**, 3051-3069.

Okutani, I., and Stephanedes, Y. J. (1984). Dynamic prediction of traffic volume through Kalman filtering theory. *Transportation Research B*, **18(1)**, 1-11.

Papageorgiou, M. (1998). Some remarks on macroscopic traffic flow modelling. *Transportation Research A*, **32**, 323-329.

Payne, H. J. (1971). Models for freeway traffic and control. *In Mathematical Models of Public Systems* (Ed. G. A. Bekey), **1**, 51-61.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan and Kaufman, San Mateo.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, **82(4)**, 669-709.

Petris, G., Petrone, S., and Campagnoli, P. (2009). *Dynamic Linear Models With R*. Springer, New York.

Prado, R., and West, M. (2010). *Time Series: Modelling, Computation and Inference*. Chapman & Hall, New York.

Queen, C. M., and Smith, J. Q. (1992). Dynamic graphical models. In *Bayesian Statistics 4*. J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith (Eds.), Oxford University Press, Oxford, 741-751.

Queen, C. M., and Smith, J. Q. (1993). Multiregression dynamic models. *Journal of the Royal Statistical Society (Series B)*, **55**, 849-870.

Queen, C. M., Wright, B. J., and Albers, C. J. (2007). Eliciting a directed acyclic graph for a multivariate time series of vehicle counts in a traffic network. *Australian and New Zealand Journal of Statistics*, **49**, 221-239.

Queen, C. M., Wright, B. J., and Albers, C. J. (2008). Forecast covariances in the linear multiregression dynamic model. *Journal of Forecasting*, **27**, 175-191.

Queen, C. M., and Albers, C.J. (2009). Intervention and causality: forecasting traffic flows using a dynamic Bayesian network. *Journal of the American Statistical Association*, **104**, 669-681.

Quintana, J.M. (1985). A dynamic linear matrix-variate regression model. *Research Report* 83, Department of Statistics, University of Warwick.

Quintana, J.M. (1987). Multivariate Bayesian forecasting models. *Unpublished Ph.D. thesis*, University of Warwick.

Quintana, J.M., and West, M. (1987). Multivariate time series analysis: new techniques applied to international exchange rate data. *The Statistician*, **36**, 275-281.

Richards, P. I. (1955). Shock waves on the highway, *Operations Research*, **4**, 42-51.

Ripley, B. D. (1996). *Pattern Recognition and Neural Networks.* Cambridge University Press, Cambridge.

Roverato, A. (2002). Hyper-inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scandinavian Journal of Statistics*, **29**, 391-411.

Smith, J. Q. (2010). *Bayesian Decision Analysis: Principles and Practice.* Cambridge University Press, Cambridge.

Stathopoulos, A., and Karlaftis, M. G. (2003). A multivariate state-space approach for urban traffic flow modelling and prediction. *Transportation Research C*, **11(2)**, 121-135.

Stephanopoulos, G., and Michalopoulos, P. G. (1979). Modelling and analysis of traffic queue dynamics at signalized intersections. *Transportation Research Journal*, **13**, 295-307.

Stephanopoulos, G., and Michalopoulos, P. G. (1981). An application of shock wave theory to traffic signal control. *Transportation Research Journal*, **5**, 35-51.

Stevens, C.F. (1974). On the variability of demand for families of items. *Operational Research Quarterly*, **25**, 411-420.

Sultan, B., Meekums, R., and Brown, M. (2008). The impact of active traffic management on motorway operation. *Road Transport Information and Control - RTIC 2008 and ITS United Kingdom Members' Conference, 20-22 May 2008*, 1-8.

Tebaldi, C., West, M., and Karr, A. F. (2002). Statistical analysis of freeway traffic flows. *Journal of Forecasting*, **21**, 39-68.

Tsekeris, T., and Stathopoulos, A. (2006). Real-time traffic volatility forecasting in urban arterial networks. *Transportation Research Record: Journal of the Transportation Research Board*, **1964**, 146-156.

Van Arem, B., Kirby, H. R , Martie J., and Whittaker, J. C. (1997). Recent advances and applications in the field of short-term traffic forecasting. *International Journal of Forecasting*, **13**, 1-12.

Vlahogianni, E. I., and Golias, J. C., and Karlaftis, M. G. (2004). Short-term traffic forecasting: overview of objectives and methods. *Transport Reviews*, **24(5)**, 533-557.

Vlahogianni, E. I., and 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 C*, **13(2)**, 211-234.

Vlahogianni, E. I., and Karlaftis, M. G. (2011). Temporal aggregation in traffic data: implications for statistical characteristics and model choice. *Transportation Letters*, **3**, 37-49.

Wang, H., and West, M. (2009). Bayesian analysis of matrix normal graphical models. *Biometrika*, **96**, 821-834.

Wang, H., Reeson, C., and Carvalho, C. M. (2011). Dynamic financial index models: modeling conditional dependencies via graphs. *Bayesian Analysis*, **6**, 639-664.

Wermuth, N., and Lauritzen, S. L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models (with discussion). *Journal of the Royal Statistical Society, (Series B)*, **52**, 21-72.

West, M., and P.J. Harrison, (1997). *Bayesian Forecasting and Dynamic Linear Models*. Springer, New York (second edition).

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. John Wiley & Sons, New York.

Whittaker, J., Garside, S., and Lindeveld, K. (1997). Tracking and predicting network traffic process. *International Journal of Forecasting*, **13**, 51-61.

Williams, B. M and Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA Process: theoretical basis and empirical results, *ASCE Journal of Transportation Engineering*, **129(6)**, 664-672.

Wilson, R.E (2010). *Trafficdata: Individual Vehicle Data from the M42 Motorway*. Available at http://www.enm.bris.ac.uk/trafficdata. [Accessed June 15, 2010].

Xiang, Y., Smith, J. Q., and Kroes, K. (2011). Multiagent bayesian forecasting of structural Time invariant dynamic systems with graphical models. *International Journal of Approximate Reasoning*, **52(7)**, 960-977.

Yule, G.U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers *Philosophical Transactions of the Royal Society of London*, Ser. A, **226**, 267-298.