Forecasting traffic flows in road networks: A graphical dynamic model approach.

Catriona M Queen[1] and Casper J Albers[2]

Department of Mathematics and Statistics, The Open University,

Milton Keynes, MK7 6AA, UK.

July 29, 2008

Abstract

Congestion on roads is a major problem worldwide. Many roads now have induction loops implanted into the road surface providing real-time traffic flow data. These data can be used in a traffic management system to monitor current traffic flows in a network so that traffic can be directed and managed efficiently. Reliable short-term forecasting and monitoring models of traffic flows are crucial for the success of any traffic management system.

Traffic flow data are invariably multivariate so that the flows of traffic upstream and downstream of a particular data collection site S in the network are very informative about the flows at site S. Despite this, most of the short-term forecasting models of traffic flows are univariate and consider the flow at site S in isolation. In this paper we use a Bayesian graphical dynamic model called the Linear Multiregression Dynamic Model (LMDM) for forecasting traffic flow. An LMDM is a multivariate model which uses a graph in which the nodes represent time series of flows at the various data collection sites, and the links between nodes represent the conditional independence and causal structure between flows at different sites. All computation in LMDMs is performed locally, so that model computation is always simple, even for arbitrarily complex road networks. This allows the model to work in real-time, as required by any traffic management system. LMDMs are also non-stationary and can readily accommodate changes in traffic flows. This is an essential property for any model for use with traffic management systems where series often exhibit temporary changes due to congestion or accidents, for example. Finally, LMDMs are often easily interpretable by non-statisticians, making them easy-to-use and understand.

The paper will focus on the problem of forecasting traffic flows in two separate motorway networks in the UK.

Key words: traffic modelling; multivariate forecasting; conditional independence; dynamic linear model; linear multiregression dynamic model

[1]Email: C.Queen@open.ac.uk

[2]Email: C.J.Albers@open.ac.uk

1 Introduction

Congestion is a major problem on many roads worldwide. Congestion has serious consequences, including environmental and health problems caused by vehicle emissions and economic problems due to the costs to road users (both in time and money) caused by longer and unreliable journey times. Many roads now have induction loops implanted into the road surface providing real-time traffic flow data. These data can be used as part of a traffic management system to monitor traffic flows and reduce congestion by taking actions, such as imposing variable speed limits or diverting traffic onto alternative routes. Reliable short-term forecasting and monitoring models of traffic flows are crucial for the success of any such traffic management system.

Several different techniques have been used to produce short-term forecasts of traffic flows, including ARIMA-type models (Williams and Hoel, 2003; Lee and Fambro, 1999); generalised linear models (Chang and Miaou, 1999); nonparametric statistical methods (Davis and Nihan, 1991; Smith *et al.*, 2002); neural networks (Dougherty and Cobbett, 1997; Zhang, 2000; Yasdi, 1999); dynamic neural networks (Chen and Grant-Muller, 2001); state space models (Whittaker *et al.*, 1997; Stathopoulos and Karlaftis, 2003); dynamic hierarchical regression models (Tebaldi *et al.*, 2002); and STARIMA models (Kamarianakis and Prastacos, 2005).

The flows of traffic upstream and downstream of a particular data collection site S in the network are very informative about the flows at site S. Despite this, most of the short-term forecasting models of traffic flow consider the flow at site S in isolation. Some models do make use of traffic flow at other data collection sites in the network by using their lagged values when modelling the flow at site S (Tebaldi *et al.*, 2002; Kamarianakis and Prastacos, 2005; Stathopoulos and Karlaftis, 2003). Whittaker *et al.*, (1997) and Sun *et al.* (2006) additionally use conditional independence so that only lagged flows of adjacent data collection sites are required in their models. With all these models, the lags used depend on the expected time it takes vehicles to travel between sites. For example, if it is expected to take one minute to travel from a particular upstream site S1 to S, then the flow at S1 at lag 1 would be used in the model to forecast 1-minute flows at S. However,

the distance between sites may be such that vehicles are counted at a number of different sites during the same time period, as is the case in this paper, Whitlock and Queen (2000) and Queen *et al.* (2007). In this case the flow at site S1 at lag 0 is useful for forecasting the flow at site S. The models used in this paper, Whitlock and Queen (2000) and Queen *et al.* (2007), not only use traffic flows at upstream sites for modelling site S and conditional independence to reduce the number of upstream sites included in the model, but, unlike other models, also allow the inclusion of flows from other data collection sites at lag 0.

This paper follows Whitlock and Queen (2000) and Queen *et al.* (2007) in using graphical dynamic models for forecasting traffic flows. In particular, the Linear Multiregression Dynamic Model (LMDM) is used. Using information regarding how traffic flows through a network, an LMDM uses a graph (see, for example, Cowell *et al.*, 1999) to represent the conditional independence and causal structure across the traffic flow time series. This graph provides a simple and useful pictorial representation of the relationships between flows at different sites. It is also used to break the multivariate forecasting problem into separate (conditional) univariate dynamic models (West and Harrison, 1997). Consequently, LMDMs are computationally simple, no matter how complex the network is that is being modelled. Thus the model can work in real-time, as required by any traffic management system.

LMDMs are non-stationary and can readily accommodate changes in traffic flows. This is an essential property for any model for use with traffic management systems where series often exhibit sudden changes occurring in response to congestion, road traffic accidents or adverse weather conditions. It is during these periods of change when a traffic management system is, in fact, of most use. As a consequence, any good model for short-term forecasting of traffic flows needs to be dynamic, and models which assume stationarity, such as ARIMA, will not be generally flexible enough for most networks. Although neural networks can accommodate sudden changes in traffic flow, the neural network needs to be trained to model each specific change, making the models impractical for everyday real-time use for many networks. It is simple to model any changes in traffic flows using LMDMs by using established intervention techniques for univariate dynamic models. Further, the graphical structure of the model means that fewer interventions

are required, since the graph automatically propagates a model change at a single site to other affected sites downstream in the network. What's more, assessing the intervention required is often also simpler because intervention parameters are often only required for individual series at the source of the problem. The use of intervention for this class of model is investigated in detail in Queen and Albers (2008). The graphical structure of the model also means that changes in the network as a whole, both temporary (due to road works, for example) and permanent (due to changes to road layout, for example), can be accommodated whilst retaining much of the original model information.

An added bonus when using LMDMs to model traffic flows, is that model parameters are often interpretable and easily understandable. This is because the possible traffic routes in the network are used to derive the graph and, in turn, the model. As a consequence, specialist statistical knowledge is not usually required when interpreting the model, which is particularly helpful when intervention is required. Thus the models are easy-to-use and understandable for non-statisticians.

The paper is organised as follows. Section 2 describes the two particular road networks which are the focus of the paper. Section 3 gives a brief overview of LMDMs. Section 4 demonstrates how a traffic network can be represented by a directed acyclic graph and Section 5 outlines the general form of the LMDM used for forecasting traffic networks. In Section 6 the powerful forecasting technique of intervention is addressed. Finally, Section 7 provides a discussion of the use of the LMDM for forecasting traffic networks and ideas for future research in the area.

2 The traffic networks

This paper focuses on forecasting vehicle counts at two separate motorway networks in the UK: the M25/A2/A282 intersection east of London, and the M60/M62/M602 intersection west of Manchester. Figure 1 shows aerial photographs of the two junctions.

The traffic data are in the form of counts of vehicles passing over induction loops in the road surface at a number of data collection sites in each network. Diagrams of the two intersections showing the layout of the data collection sites are given in Figure 2. Each network is such that traffic flows into the network, through a number of data collection

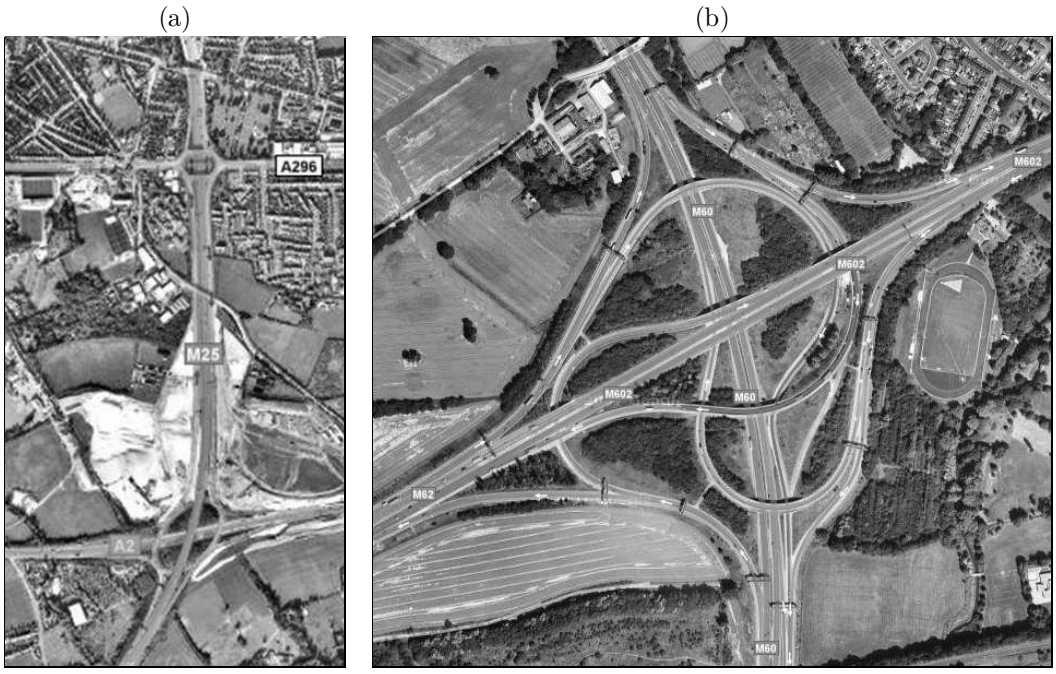


Figure 1: Aerial photographs of the junctions of interest (taken from Google Maps): (a) M25/A2/A282 junction and (b) M60/M62/M602 junction.

sites, and then out of the network. During normal conditions it will only take a few minutes for a vehicle to traverse each network.

The data collected for the M25/A2/A282 junction is in the form of hourly counts of the number of vehicles passing each data collection point, whereas for the M60/M62/M602 junction the data are minute counts. In this paper, the M60/M62/M602 junction minute data are aggregated into hourly data. Using hourly data for both networks will enable the proposed methodology to be tested on two separate networks. Using hourly data also allows a slightly simpler situation to be modelled in that the effects of everyday congestion do not need to be accommodated into the model as well (Kirby *et al.*, 1997). It is hoped that the model developed for hourly data will form the basis of a suitable model for shorter time periods.

Figure 3 shows time series plots for a typical week for the two junctions. Both plots clearly show the daily pattern of vehicle counts, with peaks in the morning and evening rush hours. What's more, it can be clearly seen that the daily pattern is different on

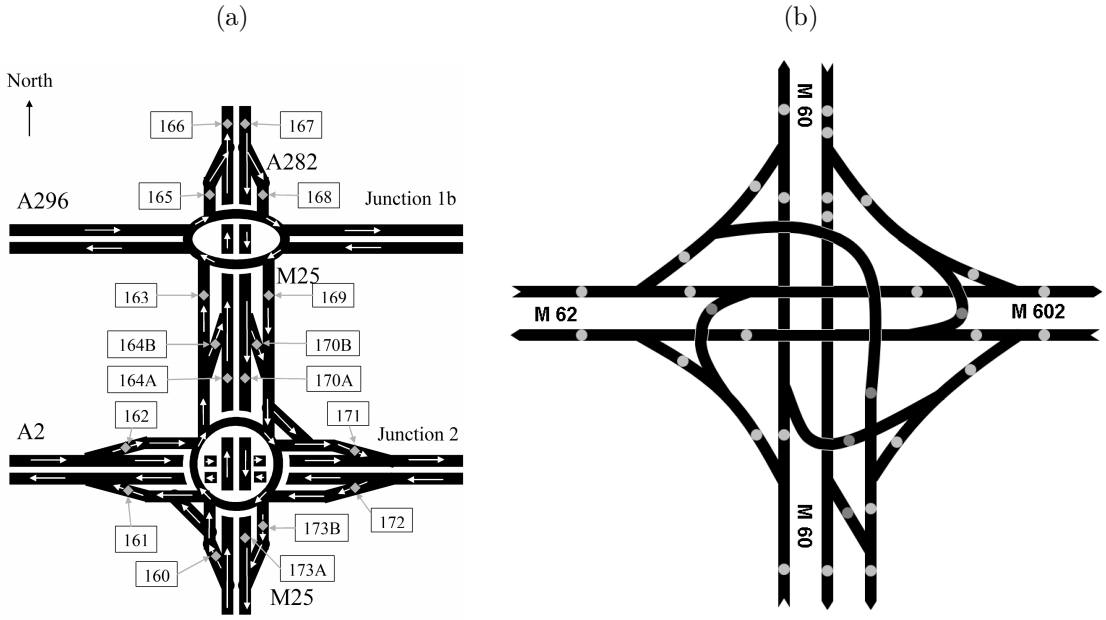


Figure 2: Diagrams showing the layout of data collection sites used in (a) the M25/A2/A282 junction and (b) the M60/M62/M602 junction. The grey diamonds in (a) are the data collection sites, each of which is numbered. The white arrows in (a) indicate the direction of traffic flow on each part of the network. In (b) the data collection sites are indicated by circles on the diagram, but for clarity the individual data collection sites are not specifically numbered and the direction of flow is indicated only at entrance/exit points. In (b) missing data are indicated with a darker grey circle.

a weekday than it is at the weekend. What is not so obvious from the plots but is apparent after further investigation (using component analysis), is that the daily pattern for Mondays and Fridays is also different to the pattern for Tuesdays to Thursdays. For clarity when presenting the model and results, only traffic flows for Tuesday–Thursday each week are considered here, so that all data follow the same daily pattern.

Vehicles pass through several data collection sites on their route through the network. Thus traffic flows at sites upstream and downstream to a particular site S are highly informative about the flows at site S. What's more, because hourly data are considered and it only takes a matter of minutes to travel through either network in normal traffic conditions, most vehicles are counted at a number of different sites during the same time period. Thus vehicle counts at sites upstream and downstream to S are informative about

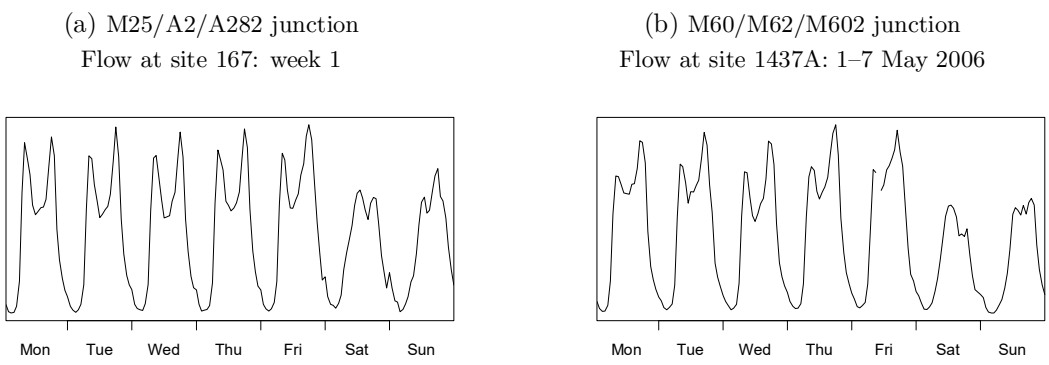


Figure 3: Plot of time series for a typical week for (a) the M25/A2/A282 junction and (b) the M60/M62/M602 junction.

the vehicle count at S in the same time period. The model developed in this paper for forecasting the vehicle counts across the network makes use of this. The model is a graphical dynamic model called the Linear Multiregression Dynamic Model (LMDM) and, through the use of conditional independence, uses information about the forecasts of vehicle counts at adjacent upstream sites to site S when forecasting vehicle counts at S. The LMDM is introduced in a little more detail in the next section.

3 Linear multiregression dynamic models

This section provides a brief overview of the LMDM. For a full account of the model, see Queen & Smith (1993).

Consider a multivariate time series $\boldsymbol{Y}_t = (Y_t(1), \ldots, Y_t(n))^T$. Suppose that the series is ordered and that the same conditional independence and causal structure is defined across the series through time, so that at each time $t = 1, 2, \ldots$, we have

$$Y_t(i) \perp\!\!\!\perp \{\{Y_t(1), \ldots, Y_t(i-1)\} \setminus \mathrm{pa}(Y_t(i))\} \,|\mathrm{pa}(Y_t(i)) \qquad \text{for } i = 2, \ldots n$$

which reads "$Y_t(i)$ is independent of $\{Y_t(1), \ldots, Y_t(i-1)\} \setminus \mathrm{pa}(Y_t(i))$ given $\mathrm{pa}(Y_t(i))$" (Dawid, 1979), where the notation "\" reads "excluding" and $\mathrm{pa}(Y_t(i)) \subseteq \{Y_t(1), \ldots, Y_t(i-1)\}$. Each variable in the set $\mathrm{pa}(Y_t(i))$ is called a *parent* of $Y_t(i)$ and $Y_t(i)$ is known as a *child* of each variable in the set $\mathrm{pa}(Y_t(i))$. Any series for which $\mathrm{pa}(Y_t(\cdot)) = \varnothing$ is known as a *root node* and is listed before any children in the ordered series $\boldsymbol{Y}_t$.

The conditional independence relationships at each time point t can be represented by a DAG, where there is a directed arc to $Y_t(i)$ from each of its parents in $\text{pa}(Y_t(i))$. To illustrate, Figure 4 shows a DAG for five time series at time t, where $\text{pa}(Y_t(2)) = \varnothing$, $\text{pa}(Y_t(3)) = \{Y_t(1)\}$, $\text{pa}(Y_t(4)) = \{Y_t(1), Y_t(3)\}$ and $\text{pa}(Y_t(5)) = \{Y_t(2), Y_t(4)\}$. Note that both $Y_t(1)$ and $Y_t(2)$ are root nodes.

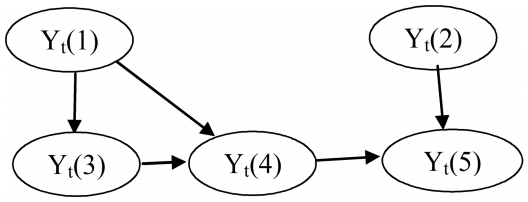


Figure 4: DAG representing five time series at time t.

As $Y_t(i)$ is independent of $\{Y_t(1), \ldots, Y_t(i-1)\} \backslash \text{pa}(Y_t(i))$ given $\text{pa}(Y_t(i))$, a forecasting model for $Y_t(i)$ need only depend on $\text{pa}(Y_t(i))$, rather than *all* the series at time t. An LMDM uses this idea and models the multivariate time series by n separate univariate models – for $Y_t(1)$ and $Y_t(i)|\text{pa}(Y_t(i))$, $i = 2, \ldots, n$. Explicitly, denoting the information available at time t by D_t, an LMDM has the following system equation and n observation equations for all times $t = 1, 2, \ldots$.

$$\begin{array}{lrcll}
\text{Observation equations:} & Y_t(i) & = & \boldsymbol{F}_t(i)^\top \boldsymbol{\theta}_t(i) + v_t(i), & v_t(i) \sim N(0, V_t(i)), 1 \leq i \leq n \\
\text{System equation:} & \boldsymbol{\theta}_t & = & G_t \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, & \boldsymbol{w}_t \sim N(\boldsymbol{0}, W_t) \\
\text{Initial Information:} & (\boldsymbol{\theta}_0 \,|\, D_0) & \sim & N(\boldsymbol{m}_0, C_0). &
\end{array}$$

The vector $\boldsymbol{F}_t(i)$ contains the parents $\text{pa}(Y_t(i))$ and possibly other known variables (which may include $Y_1(i), \ldots, Y_{t-1}(i)$ and $\text{pa}(Y_1(i)), \ldots, \text{pa}(Y_{t-1}(i))$); $\boldsymbol{\theta}_t(i)$ is the parameter vector for $Y_t(i)$; $V_t(1), \ldots V_t(n)$ are the scalar observation variances; $\boldsymbol{\theta}_t^\top = (\boldsymbol{\theta}_t(1)^\top, \ldots, \boldsymbol{\theta}_t(n)^\top)$; and the matrices G_t, W_t and C_0 are all block diagonal. The error vectors, $\boldsymbol{v}_t^\top = (v_t(1), \ldots, v_t(n))$ and $\boldsymbol{w}_t^\top = (\boldsymbol{w}_t(1)^\top, \ldots, \boldsymbol{w}_t(n)^\top)$, are such that $v_t(1), \ldots, v_t(n)$ and $\boldsymbol{w}_t(1), \ldots, \boldsymbol{w}_t(n)$ are mutually independent and $\{\boldsymbol{v}_t, \boldsymbol{w}_t\}_{t \geq 1}$ are mutually independent with time.

For those series with parents, each univariate model in the LMDM is simply a regression DLM with the parents as linear regressors. For root nodes, any suitable univariate DLM

may be used. For example, consider the DAG in Figure 4. As $Y_t(1)$ and $Y_t(2)$ are both root nodes, each of these series can be modelled separately in an LMDM using any suitable univariate DLMs. $Y_t(3)$, $Y_t(4)$ and $Y_t(5)$ all have parents and so these would be modelled by (separate) univariate regression DLMs with the single regressor $Y_t(1)$ for $Y_t(3)$'s model, the two regressors $Y_t(1)$ and $Y_t(3)$ for $Y_t(4)$'s model and the two regressors $Y_t(2)$ and $Y_t(4)$ for $Y_t(5)$'s model.

As long as $\boldsymbol{\theta}_t(1), \boldsymbol{\theta}_t(2), \ldots, \boldsymbol{\theta}_t(n)$ are mutually independent a priori, each $\boldsymbol{\theta}_t(i)$ can be updated separately in closed form from $Y_t(i)$'s (conditional) univariate model. A forecast for $Y_t(1)$ and the conditional forecasts for $Y_t(i)|\text{pa}(Y_t(i))$, $i = 2, \ldots, n$, can also be found separately using established DLM results (see West and Harrison (1997) for details). For example, in the context of the DAG in Figure 4, forecasts can be found separately (using established DLM results) for

$$Y_t(1), \quad Y_t(2), \quad Y_t(3)|Y_t(1), \quad Y_t(4)|Y_t(1), Y_t(3) \quad \text{and} \quad Y_t(5)|Y_t(2), Y_t(4).$$

The joint one-step ahead forecast distribution of $\boldsymbol{Y}_t$ can then be expressed as the product of $Y_t(1)$'s forecast distribution and the individual univariate conditional forecast distributions, $Y_t(i) \,|\, \text{pa}(Y_t(i))$, $i = 2, \ldots, n$. Even though regression is linear so that each of the univariate forecast distributions for $Y_t(1)$ and $Y_t(i) \,|\, \text{pa}(Y_t(i))$, $i = 2, \ldots, n$, is Gaussian, these models can yield highly non-Gaussian joint forecast distributions. As such, they are analogous to non time series graphical models in that although the sub-problems can be fairly simple to work with (in this case univariate DLMs), the joint distribution can be highly complex (Cowell *et al.*, 1999).

The values for $Y_t(i)$ and $\text{pa}(Y_t(i))$ are, of course, observed simultaneously. So, although individual (conditional) forecasts of $Y_t(i)|\text{pa}(Y_t(i)), i = 2, \ldots, n$, are easily obtained from established DLM results, the marginal forecast for each $Y_t(i)$, without conditioning on the values of its parents, is in fact required. Unfortunately, the marginal forecast distributions for $Y_t(i)$, $i = 2, \ldots, n$, will not generally be of a simple form. However, (under quadratic loss), the mean and variance of the marginal forecast distributions for $Y_t(i)$, $i = 2, \ldots, n$, are adequate for forecasting purposes, and these can be calculated. So, returning to the context of the DAG in Figure 4, this means that the marginal forecast means and variances

for $Y_t(3)$, $Y_t(4)$ and $Y_t(5)$ need to be calculated. Note that as $Y_t(1)$ and $Y_t(2)$ do not have any parents, their marginal forecasts have already been calculated from DLM theory.

For calculating the marginal forecast variance for $Y_t(i)$, the marginal forecast covariance matrix for $\text{pa}(Y_t(i))$ is required. Returning to the example DAG in Figure 4 to illustrate, this means that the marginal forecast variance of $Y_t(5)$, for example, requires the forecast covariance for $Y_t(2)$ and $Y_t(4)$, without conditioning on either of their parents $Y_t(1)$ and $Y_t(3)$ — that is, the marginal covariance of $Y_t(2)$ and $Y_t(4)$. Queen *et al.* (2008) present a simple algebraic form for calculating these marginal forecast covariances.

It is important to note that whereas two DAGs may exhibit the same conditional independence statements, they can yield quite different LMDMs. For example, consider the two DAGs in Figure 5. In both DAGs $A_t \perp\!\!\!\perp C_t \mid B_t$. However, they would yield quite different LMDMs. For LMDMs, the DAG needs to represent the conditional independence structure *related to causality*, so that (following Wermuth & Lauritzen, 1990) variables which are hypothesized to be causally linked should be connected by a directed arc following the direction of causation. Queen and Albers (2008) show how the use of intervention in the LMDM (see Section 6) is a powerful tool for identifying the causal structure of a multivariate time series.

Figure 5: Two directed acyclic graphs both representing $A_t \perp\!\!\!\perp C_t \mid B_t$, but yielding different Linear Multiregression Dynamic Models.

4 Representing a traffic network by a DAG

Queen *et al.* (2007) describes in detail how a DAG can be elicited to represent the M25/A2/A282 traffic time series. The basic ideas, which apply to any traffic network, will be described here. For full details, see Queen *et al.* (2007).

Road traffic networks are basically a series of junctions of two types: forks and joins. A fork is illustrated in Figure 6(a), in which vehicles from data collection site S1 move to sites S2 and S3. A join is illustrated in Figure 6(b), in which traffic from two sites, S4 and

S5, merge to site S6.

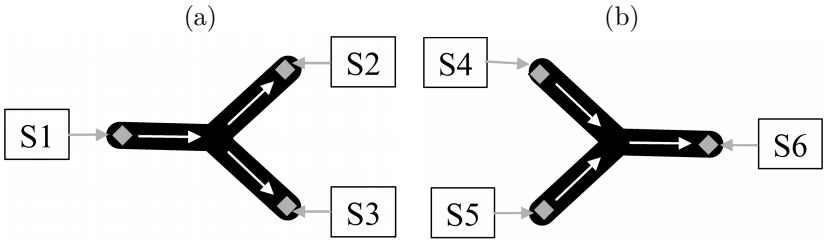


Figure 6: The two basic junctions found in road traffic networks: (a) a fork and (b) a join. In each diagram, the arrows denote the direction of travel and the grey diamonds are the data collection sites.

Let $Y_t(\mathrm{S}i)$ be the count at site Si at hour t so that $Y_t(\mathrm{S}i) \sim Po(\mu_t(\mathrm{S}i))$. In Figure 6(a), traffic from S1 only flows to S2 and S3. So $Y_t(\mathrm{S1})$ should be (approximately) equal to the sum of $Y_t(\mathrm{S2})$ and $Y_t(\mathrm{S3})$. It is therefore possible to define the conditional distribution

$$Y_t(\mathrm{S2})|Y_t(\mathrm{S1}) \sim Bi(Y_t(\mathrm{S1}), \alpha_t)$$

where $\alpha_t = \mu_t(\mathrm{S2})/\mu_t(\mathrm{S1})$, the proportion of traffic flowing from S1 to S2. The traffic counts $Y_t(\cdots)$ are, in general, large so that this binomial distribution can be approximated to normality so that the conditional distribution for $Y_t(\mathrm{S2})|Y_t(\mathrm{S1})$ is normal with mean $\alpha_t Y_t(\mathrm{S1})$. Thus a DAG would have $Y_t(\mathrm{S1})$ as a parent of $Y_t(\mathrm{S2})$ and the observation equation for $Y_t(\mathrm{S2})$ in the LMDM would be of the form

$$Y_t(\mathrm{S2}) = \alpha_t Y_t(\mathrm{S1}) + v_t(\mathrm{S1}), \qquad v_t(\mathrm{S1}) \sim N(0, V_t(\mathrm{S1})).$$

Conditional on $Y_t(\mathrm{S1})$ and $Y_t(\mathrm{S2})$, $Y_t(\mathrm{S3})$ should be (approximately) equal to $Y_t(\mathrm{S1})$ minus $Y_t(\mathrm{S2})$. In this case, $Y_t(\mathrm{S3})$ has parents $Y_t(\mathrm{S1})$ and $Y_t(\mathrm{S2})$ in a DAG, and, in the terminology of the WinBUGS software (http://www.mrc-bsu.cam.ac.uk/bugs/), $Y_t(\mathrm{S3})$ is a logical variable denoted by an double oval in the DAG. The resulting DAG is given in Figure 7.

Approximating $Y_t(\mathrm{S1})$'s Poisson distribution to normality gives the LMDM observation

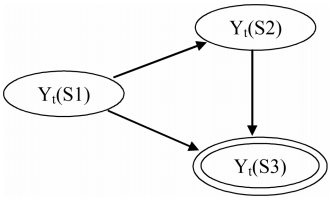


Figure 7: DAG representing the fork junction given in Figure 6(a).

equations

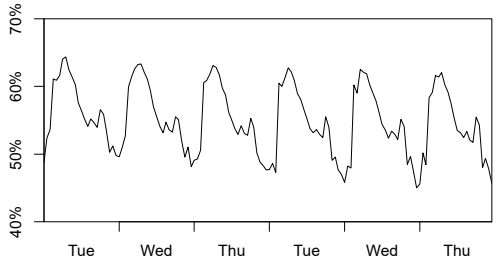
$$\begin{aligned} Y_t(\text{S1}) &= \mu_t(\text{S1}) + v_t(\text{S1}), & v_t(\text{S1}) &\sim N(0, V_t(\text{S1})), \\ Y_t(\text{S2}) &= \alpha_t Y_t(\text{S1}) + v_t(\text{S2}), & v_t(\text{S2}) &\sim N(0, V_t(\text{S2})), \\ Y_t(\text{S3}) &= Y_t(\text{S1}) - Y_t(\text{S2}). \end{aligned} \tag{4.1}$$

Note that although the observation equations are algebraically the same for each time t, the actual parameters, $\mu_t(\text{S1})$ and α_t, will vary depending on which hour of the day t is. For example, if time t is the hour 1:00-2:00 a.m. when roads are (usually) quiet, $\mu_t(\text{S1})$ will be quite small, but if time t is the hour 8:00-9:00 a.m. during the morning rush hour, $\mu_t(\text{S1})$ will be much larger. The time plots shown in Figure 3 are for two root nodes in the networks. It is clear how the μ_t parameters for these series depend on the hour t. On the other hand, the parameter vector $\boldsymbol{\alpha}_t$ represents the proportion of traffic flowing from the associated parent to the child and is affected by changing routes during the day. For example, the proportion of traffic leaving at a particular motorway exit may be high during the morning rush hour, but low during the evening rush hour when drivers are returning home. An illustration of how α_t varies during the day is given in Figure 8, which shows plots of the hourly proportion over several days of traffic flowing from (a) site 167 to site 170a in the M25/A2/A282 network and (b) site 1437A to site 1441A in the M60/M62/M602 network.

For the join junction in Figure 6(b), all traffic from S4 and S5 flow to S6. So, $Y_t(\text{S6})$ should be (approximately) equal to the sum of $Y_t(\text{S4})$ and $Y_t(\text{S5})$, so that $Y_t(S6)$ is a logical variable. This is represented by the DAG given in Figure 9.

Approximating $Y_t(\text{S4})$ and $Y_t(\text{S5})$'s Poisson distributions to normality gives the LMDM

(a) M25/A2/A282 junction
Proportion from site 167 to site 170a



(b) M60/M62/M602 junction
Proportion from site 1437A to site 1441A

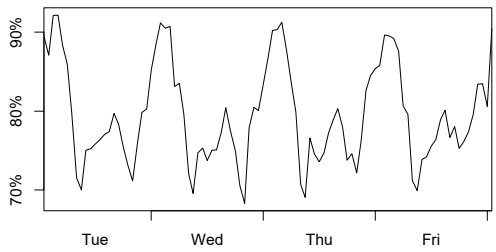


Figure 8: Plots of the hourly proportion of traffic flowing from (a) site 167 to site 170a in the M25/A2/A282 network and (b) site 1437A to site 1441A in the M60/M62/M602 network.

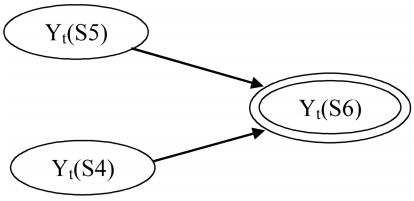


Figure 9: DAG representing the join junction given in Figure 6(b).

observation equations

$$\begin{aligned} Y_t(\mathrm{S4}) &= \mu_t(\mathrm{S4}) + v_t(\mathrm{S4}), \qquad v_t(\mathrm{S4}) \sim N(0, V_t(\mathrm{S4})), \\ Y_t(\mathrm{S5}) &= \mu_t(\mathrm{S5}) + v_t(\mathrm{S5}), \qquad v_t(\mathrm{S5}) \sim N(0, V_t(\mathrm{S5})), \\ Y_t(\mathrm{S6}) &= Y_t(\mathrm{S4}) + Y_t(\mathrm{S5}). \end{aligned} \tag{4.2}$$

Once again, although the observation equations are algebraically the same for each time t, the actual parameters will vary depending on which hour of the day t is.

By using these ideas for each fork and join, A DAG and LMDM can be elicited for an entire network.

Although traffic networks are simply a series of forks and joins, unfortunately not every fork and join always has data collection sites for each entry and each exit. This is the case for the M25/A2/A282 junction. For example, in Figure 2(a), consider site 160 which counts vehicles exiting the northbound carriageway of the M25. Directly after site 160

there is a fork with no data collection sites. There are also no sites on the roundabout at Junction 2. Eliciting a suitable DAG for cases such as this is a little more complicated and is fully described in Queen *et al.* (2007).

4.1 Missing data

Missing data are a common problem for traffic networks. Data collection sites can be broken or simply not commissioned leading to short- or long-term missing measurements for a site. Although it is important that any model for a traffic management system can continue to produce accurate forecasts when faced with missing data, few models developed so far can do so. In contrast, missing data are not a huge problem for LMDMs.

If a data collection site is missing some measurements, but otherwise working, following West and Harrison (1997), the missing data can be accommodated in the LMDM simply by not updating posteriors for any time periods for which data are missing. The M60/M62/M602 network has problems with missing data of this sort whereas the M25/A2/A282 network does not.

For data collection sites for which all data are missing, since the models are Bayesian, any missing data can be treated as an unknown parameter. Whitlock and Queen (2000) showed how MCMC techniques could be used to forecast in the M25/A2/A282 junction where there are no data for several data collection sites. However, because the missing sites are at the edges of the network, there is no way of assessing how well these techniques worked at the missing sites. The M60/M62/M602 junction is different, in that the missing sites are in the centre of the network represented by logical variables. It is thus simple to produce forecasts for the missing sites, and sites further downstream, without resorting to numerical techniques. The quality of the forecasts at the missing data sites can be assessed to a certain extent by assessing the quality of the forecasts of downstream sites.

An alternative approach for the M25/A2/A282 junction was used in Queen *et al.* (2007). In that paper, any site for which there are no data is treated as if they are simply not there. Using this technique means that some forks and joins do not have data collection sites for each entry and each exit. As mentioned earlier, this does make elicitation of the DAG a little less straightforward. However, it does have the advantage

that model evaluation is easier as forecasts are made only for series which are observed. The DAG shown in this paper representing the M25/A2/A282 junction was elicited using this method.

The DAGs elicited for the two networks are given in Figures 10 and 11. Even with data collection sites with no data, the M60/M62/M602 junction is simply a series of forks and joins, and elicitation of its DAG is straightforward following the methods outlined earlier. The missing data in the M25/A2/A282 junction cause more of a problem and in fact cause the DAG to separate into two parts and even cause some nodes to be totally disconnected from the rest of the DAG. For full details regarding the elicitation of the DAG representing the M25/A2/A282 junction, see Queen *et al.* (2007). In the DAG of Figure 10, the two variables Z_1 and Z_2 are simply logical functions of their parents created to overcome problems of collinearity (caused by the missing data), and are not important in this paper.

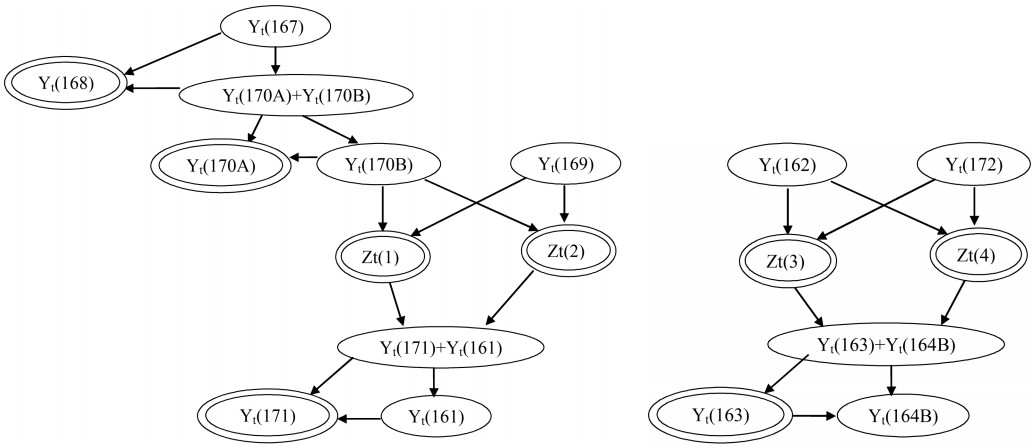


Figure 10: DAG representing the M25/A2/A282 junction.

5 Using the LMDM for forecasting in the traffic networks

The form of the LMDM observation equations for observed forks and joins (as shown in Figure 6) are given in equations 4.1 and 4.2. In general, the observation equations have the following forms:

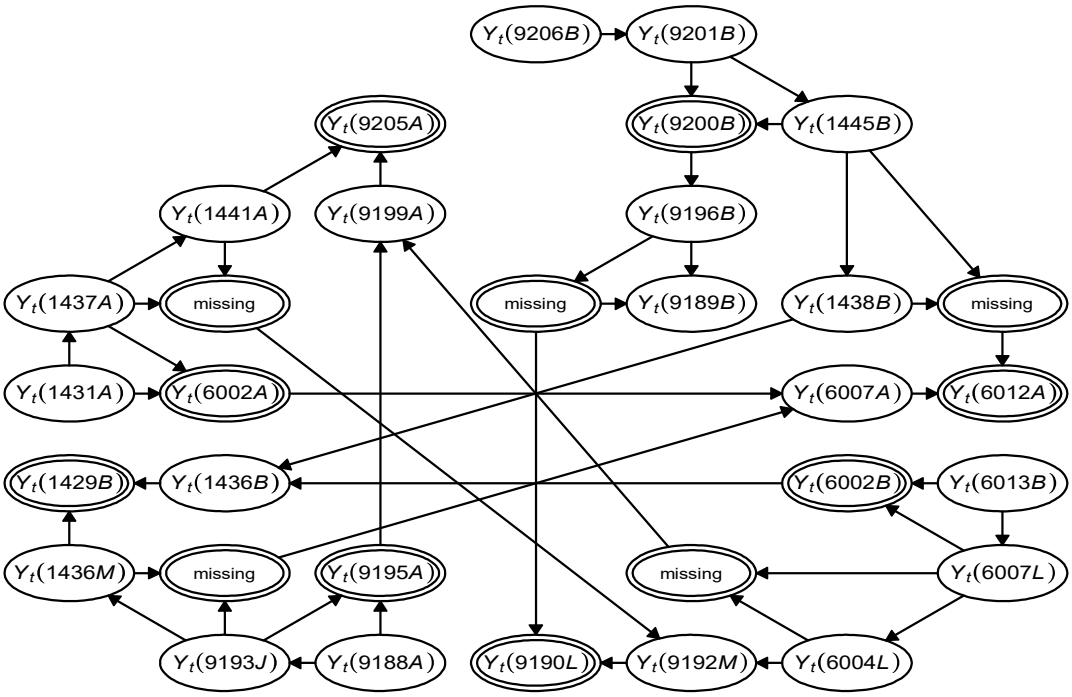


Figure 11: DAG representing the M60/M62/M602 junction.

For $Y_t(\mathrm{S}i)$ with no parents:

$$Y_t(\mathrm{S}i) = \mu_t(\mathrm{S}i) + v_t(\mathrm{S}i), \qquad v_t(\mathrm{S}i) \sim N(0, V_t(\mathrm{S}i)). \tag{5.1}$$

For $Y_t(\mathrm{S}i)$ with parents:

$$Y_t(\mathrm{S}i) = \boldsymbol{\alpha}_t \boldsymbol{pa}(Y_t(\mathrm{S}i)) + v_t(\mathrm{S}i), \qquad v_t(\mathrm{S}i) \sim N(0, V_t(\mathrm{S}i)), \tag{5.2}$$

where $\boldsymbol{pa}(Y_t(\mathrm{S}i))$ is the vector of $Y_t(\mathrm{S}i)$'s parents and $\boldsymbol{\alpha}_t$ is a vector of parameters, each of which represents the proportion (which may equal 1) of traffic flowing from the associated parent to $Y_t(\mathrm{S}i)$.

To account for the seasonality exhibited by each $Y_t(\mathrm{S}i)$ for those series without parents, a seasonal factor model will be used. This is preferred to a Fourier model here as it is easier to interpret and thus simplifies intervention. For the M25/A2/A28 junction, full-form Fourier models were found to give identical mean squared errors, and reduced form Fourier models performed less well. The seasonal effects approach is not used as there is no obvious typical value to estimate or obvious advantage to using this approach.

To use a seasonal factor model for equation 5.1, define a separate level parameter for each hour of the day, so that $\mu_t^{(1)}(\mathrm{S}i)$ represents the level parameter for hour 1 at time t (ie 00:00–01:00), $\mu_t^{(2)}(\mathrm{S}i)$ represents the level parameter for hour 2 at time t (ie 01:00–02:00), and so on. Suppose for simplicity that t is the hour 00:00–01:00. Define $\boldsymbol{\mu}_t(\mathrm{S}i)^\top = (\mu_t^{(1)}(\mathrm{S}i), \ldots, \mu_t^{(24)}(\mathrm{S}i))$, then $Y_t(\mathrm{S}i)$'s observation equation can be written as

$$Y_t(\mathrm{S}i) = \boldsymbol{F}_t^\top \boldsymbol{\mu}_t(\mathrm{S}i) + v_t(\mathrm{S}i), \qquad v_t(\mathrm{S}i) \sim N(0, V_t(\mathrm{S}i)),$$

where $\boldsymbol{F}_t^\top = (1, 0, \ldots, 0)$.

Time $t+1$ is hour 2 (ie 01:00–02:00) requiring

$$\boldsymbol{\mu}_{t+1}(\mathrm{S}i)^\top = (\mu_{t+1}^{(2)}(\mathrm{S}i), \ldots, \mu_{t+1}^{(24)}(\mathrm{S}i), \mu_{t+1}^{(1)}(\mathrm{S}i)).$$

In order to 'cycle' through the hour parameters at each time as required, the system equation has the form

$$\boldsymbol{\mu}_{t+1}(\mathrm{S}i) = G\boldsymbol{\mu}_t(\mathrm{S}i) + \boldsymbol{w}_{t+1}(\mathrm{S}i), \qquad \boldsymbol{w}_{t+1}(\mathrm{S}i) \sim N(0, W_{t+1}(\mathrm{S}i))$$

where

$$G = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & 1 \\ 1 & 0 & \cdots & \cdots & 0 \end{pmatrix}. \tag{5.3}$$

Using this model essentially links the current hour parameter with the parameter for the corresponding hour the previous day.

An alternative model which has been found to perform slightly better, is to link the current hour parameter (hour $t+1$) with the parameter for the previous hour (hour t), as well as the parameter for the corresponding hour the previous day (hour $t+1-24$),

giving, for $a \in [0,1]$,

$$G = \begin{pmatrix} a & 1-a & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & 1 \\ 1 & 0 & \cdots & \cdots & 0 \end{pmatrix}.$$

The optimal value of a for the M25/A2/A282 network was found to be 0.01.

To account for the seasonality displayed by the α parameters in equation 5.2, a separate proportion parameter is defined for each hour of the day. Assume for clarity of presentation that $Y_t(\mathrm{S}i)$ only has one single parent. As before, suppose that time t is hour 1. Then define $\boldsymbol{\alpha}_t^\top = (\alpha_t^{(1)}, \ldots, \alpha_t^{(24)})$, so that equation 5.2 can be written in the form

$$Y_t(\mathrm{S}i) = \boldsymbol{F}_t^\top \boldsymbol{\alpha}_t + v_t(\mathrm{S}i), \qquad v_t(\mathrm{S}i) \sim N(0, V_t(\mathrm{S}i))$$

where $\boldsymbol{F}_t^\top = (\mathrm{pa}(Y_t(\mathrm{S}i)), 0, \ldots, 0)$. In order to 'cycle' through the hour parameters at each time t as required, the system equation has the form

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

where, once again, G is given by (5.3). Thus the current hour parameter is again linked with the parameter for the corresponding hour of the previous day.

The value of the observation variances $V_t(\mathrm{S}i)$ can be estimated on-line using standard DLM variance learning techniques (see for example West and Harrison (1997), Section 2.5) and the evolution variances W_{t+1} can be estimated using established DLM discounting techniques (see for example West and Harrison (1997), Section 2.4)

Obviously the LMDM assumes a linear relationship between parent and child. Two typical plots of parent versus child are shown in Figure 12, demonstrating that this is not an unreasonable assumption for these two networks.

6 Intervention

Traffic time series often exhibit sudden changes due to such events as road works or adverse weather conditions, and it is essential that any forecasting model can accommodate such changes. In the LMDM this is done using established intervention techniques developed for univariate DLMs (Harrison and Stevens, 1976; West and Harrison, 1986, 1989, 1997).

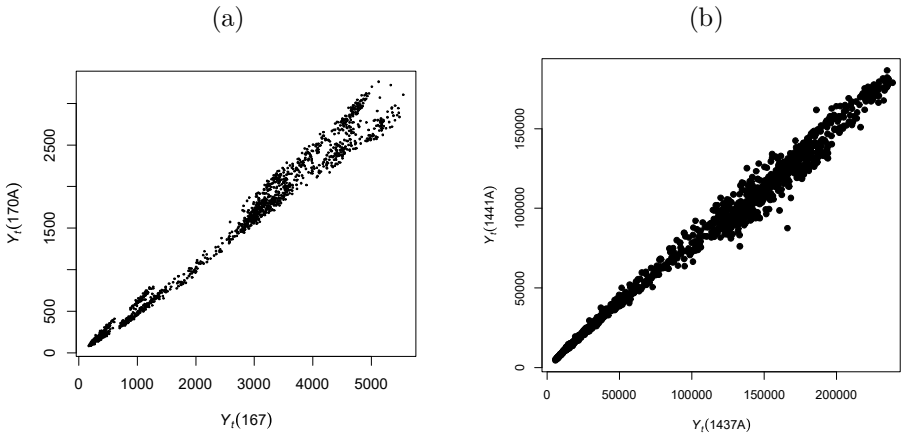


Figure 12: Plots of (a) parent $Y_t(167)$ against its child $Y_t(170\text{A})$ for the M25/A2/A282 junction and (b) parent $Y_t(1437\text{A})$ against its child $Y_t(1441\text{A})$ for the M60/M62/M602 junction.

Intervention is a very powerful forecasting tool for the LMDM. The graphical structure of the LMDM means that any intervention at a particular site automatically propagates to other sites downstream in the network. Because of this, fewer interventions are generally required and interventions are also simpler to assess since intervention is only usually required for individual series at the source of the problem.

Intervention in the LMDM usually works as follows. Consider the observation equations for series with and without parents given in equations 5.1 and 5.2. The observation equation for an individual $Y_t(\text{S}i)$ can then be written generically as

$$Y_t(\text{S}i) = \boldsymbol{F}_t^\top \boldsymbol{\theta}_t + v_t, \qquad v_t \sim N(0, V_t),$$

where $\boldsymbol{F}_t$ is a function of $\boldsymbol{pa}(Y_t(\text{S}i))$ if $Y_t(\text{S}i)$ has parents, otherwise $\boldsymbol{F}_t^\top = (1, 0, \ldots, 0)$, and $\boldsymbol{\theta}_t$ is $\boldsymbol{\alpha}_t(\text{S}i)$ if $Y_t(\text{S}i)$ has parents, otherwise $\boldsymbol{\theta}_t$ is $\boldsymbol{\mu}_t(\text{S}i)$. Intervention for $Y_t(\text{S}i)$ in the LMDM then involves manipulating $Y_t(\text{S}i)$'s observation equation so that

$$Y_t(\text{S}i) = \boldsymbol{F}_t^\top \boldsymbol{\theta}_t + \boldsymbol{h}_t + \boldsymbol{v}_t, \qquad \boldsymbol{v}_t \sim (0, V_t + H_t)$$

for some vector $\boldsymbol{h}_t$ and matrix H_t. The system equation for the above $\boldsymbol{\theta}$ parameter for $Y_t(\text{S}i)$ can be written generically as

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

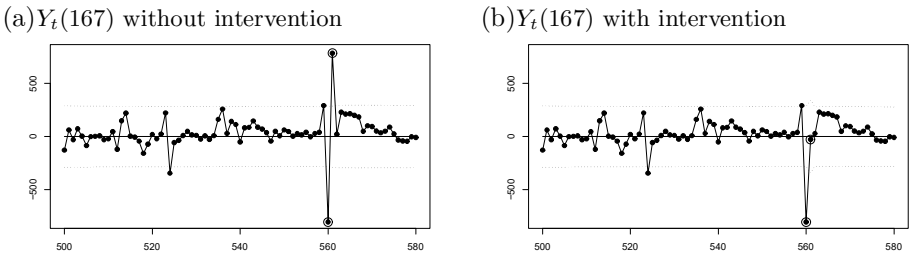


Figure 13: Plots of the one-step forecast errors (solid line) and ± 1.96 forecast standard deviations (dotted lines) obtained using the Linear Multiregression Dynamic Model between times 500 and 580 for $Y_t(167)$. The plot on the left shows the forecast errors obtained without intervention and that on the right shows the forecast errors following intervention for $Y_{561}(167)$. The observations at times 560 and 561 are circled for each series in each plot.

When intervening for $\boldsymbol{\theta}_t$, the system equation is manipulated so that

$$\boldsymbol{\theta}_t = G_t^* \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim (\mathbf{0}, W_t^*)$$

for suitable matrices G_t^* and W_t^*.

Queen *et al.* (2007) and Queen and Albers (2008) demonstrate how effective intervention is for improving forecast performance when forecasting vehicle counts in the M25/A2/A282 network. For illustration, one of the interventions used for the M25/A2/A282 network presented in Queen and Albers (2008) will be discussed briefly here. Figure 13(a) shows a plot of the one-step forecast errors together with the ± 1.96 forecast standard deviation error bars for $Y_t(167)$ between times 500 and 580. At time 560 (07.00-08.00 on Thursday week 8) there is an unusually large negative forecast error, followed by an unusually large positive forecast error. The forecast errors for times 560 and 561 are circled on the plot. The same pattern of forecast errors is also evident at sites downstream from site 167, such as 168, 170A or 170B, but not at other sites, such as site 169.

Following the unexpectedly large negative forecast error at time 560, intervention was used for $Y_t(167)$ at time 561 as follows. The observation $y_{560}(167)$ was unexpected and so was treated simply as an outlier. During the following time period ($t = 561$), as the road blockage clears and vehicles start moving, the delayed vehicles (from the previous hour) are expected to pass site 167, in addition to the vehicles that arrive during hour $t = 561$.

The expected number of cars delayed from hour 560 is $e_{560}(167) = f_{560}(167) - y_{560}(167)$, where $f_{560}(167)$ is the one-step forecast (at time 559) for $Y_{560}(167)$. The observation error variance is also increased by 10,000 at time 561 to allow for increased uncertainty and to allow the model to adapt to the change rapidly. The value of 10,000 was chosen so that the observation variance at the time of intervention is approximately 50% larger than before the intervention. (Incidentally, the MSE is fairly robust to the choice of how much the observation variance should increase to at intervention.) Thus intervention adjusts the observation equation at hour 561 so that for $t = 561$,

$$Y_t(167) = \boldsymbol{F}_t(167)^\top \boldsymbol{\theta}_t(167) + e_{t-1}(167) + v_t(167), \qquad v_t(167) \sim N(0, V_t(167) + 10000).$$

The effects of intervening for $Y_{561}(167)$ on the one-step ahead forecasts can be clearly seen in Figure 13(b). The intervention not only improves the one-step forecast error for $Y_{561}(167)$, but it also improves the one-step forecast error for downstream sites. On the other hand, the intervention for $Y_{561}(167)$ has no affect on the one-step forecast error of other sites, such as 169, which are not downstream from 167.

7 Discussion

Queen *et al.* (2007) illustrated the model's performance for forecasting hourly traffic flows for the M25/A2/A282 network over a time period of 14 weeks of consecutive Tuesdays, Wednesdays and Thursdays. The first two weeks of data were used to create initial priors for the model's parameters. One-step forecast errors, both with and without the intervention at $t = 561$ (see Section 6 and Figure 13), were used to compute mean squared errors. Since these mean squared errors are not robust with respect to outliers, median squared errors were computed as well. These were compared to those when running univariate DLM's for all the nodes, and thus ignoring the multivariate nature of the network. Using the LMDM provides, in general, better results than using univariate DLM's, and after the intervention at $t = 561$ the median squared error of $Y_t(561)$ (over the 12-week forecasting period) and its direct descendants reduces by a further 1% to 5%. Queen and Albers (2008) demonstrate that the model's performance may also be improved by using intervention in the parameters of the system equation.

For the M60/M62/M602 network, we do not have complete results yet (work in progress), but initial work suggests that similar performance improvements can be attained by implementing the LMDM. Some stretches in this network contain more than one traffic counting station (see Figure 2). We can use this information to improve our model by estimating how much of the forecasting error is due to measurement errors in the counting devices.

As part of future research, we aim to compare our model's performance with other methods, such as neural networks and ARIMA methods. At the moment, comparison has only been done with univariate DLMs. (The natural model to compare the LMDM's performance with is a standard multivariate DLM. However, this model gives very unnatural results due to the high correlation between the series, which affects the underlying algorithm's precision.) An important advantage of the LMDM is that intervention will automatically carry through to descendant nodes, whereas alternative methods usually require more complicated work to carry out an intervention.

A second topic for future research is to study the distributional assumptions in the LMDM. At the moment, we have implemented the 'standard DLM assumptions' such as using normal priors for parameters. Since many parameters are proportion parameters, it seems natural to model a beta prior for them. Also, since the data are essentially counts, there might be a gain in performance if variance laws are used to allow the errors to be modelled by a Poisson distribution.

Acknowledgements

We would like to thank Kent County Council and the Highways Agency for providing us with the data used in this paper.

References

Chang, J.L. and Miaou, S.P. (1999) Real-time prediction of traffic flows using dynamic generalized linear models. *Transportation Research Record* **1678**, 168-178.

Chen, H. and Grant-Muller, S. (2001) Use of sequential learning for short-term traffic flow forecasting. *Transportation Research Part C*, **9**, 319-336.

Cowell, R.G., Dawid, A.P., Lauritzen, S.L. and Spiegelhalter, D.J. (1999) *Probabilistic Networks and Expert Systems.* Springer-Verlag New York.

Davis, G.A. and Nihan, N.L. (1991) Nonparametric regression and short-term freeway traffic forecasting. *ASCE Journal of Transportation Engineering*, **117** (2), 178-188.

Dawid, A.P. (1979). Conditional independence in statistical theory (with Discussion). *Journal of the Royal Statistical Society, Series B*, **41**, 1–31.

Dougherty, M. and Cobbett, M. (1997) Short-term inter-urban forecasts using neural networks. *International Journal of Forecasting*, **13**, 21-31.

Harrison, P.J. and Stevens, C. (1976). Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society, Series B*, **38**, 205–247.

Kamarianakis, Y. and Prastacos, P. (2005) Space-time modeling of traffic flow. *Computers and Geosciences*, **31**, 119-133.

Kirby, H.R., Watson, S.M. and Dougherty, M.S. (1997) Should we use neural networks or statistical models for short-term motorway traffic forecasting? *International Journal of Forecasting*, **13**, No. 1, 43–50.

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

Queen, C.M. and Albers, C.J. (2008) Intervention and causality in a dynamic Bayesian network. *Open University Statistics Group Technical Report* 08/01 (statistics.open.ac.uk).

Queen, C.M. and Smith, J.Q. (1993) Multiregression dynamic models. *Journal of the Royal Statistical Society, B*, **55**, No 4, 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** (3), 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.

Smith, B.L., Williams, B.M. and Oswald, R.K. (2002) Parametric and nonparametric traffic volume forecasting. *Transportation Research Part C - Emerging Technologies*, **10** (4), 303-321.

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

Sun, S.L., Zhang, C.S. and Yu, G.Q. (2006) A Bayesian network approach to traffic flows forecasting. *IEEE Transactions on Intelligent Transportation Systems*, **7** (1), 124–132.

Tebaldi, C., West, M. and Karr, A.K. (2002) Statistical analyses of freeway traffic flows. *Journal of Forecasting*, **21**, 39–68.

Wermuth, N., and Lauritzen, S.L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models. *J. Royal Statist. Soc. Ser. B* **52**, 21–50.

West, M. and Harrison, P.J. (1986). Monitoring and adaptation in Bayesian forecasting models. *Journal of the American Statistical Association*, **81**, 741–750.

West, M. and Harrison, P.J. (1989). Subjective intervention in formal models. *Journal of Forecasting*, **8**, 33–53.

West, M. and Harrison, P.J. (1997) *Bayesian Forecasting and Dynamic Models* (2nd edition) Springer-Verlag. New York.

Whitlock, M.E. and Queen, C.M. (2000) Modelling a traffic network with missing data. *Journal of Forecasting*, **19**, 7, 561-574.

Whittaker, J., Garside, S. and Lindveld, K. (1997) Tracking and predicting a 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.

Yasdi, R. (1999) Prediction of road traffic using a neural network approach. *Neural Computing and Applications*, **8**, 135-142.

Zhang, H.M. (2000) Recursive prediction of traffic conditions with neural network models. *ASCE Journal of Transportation Engineering*, **126** (6), 472-481.

View publication stats