

Intervention and causality: forecasting traffic flows using a dynamic Bayesian network

Catriona M. Queen Casper J. Albers

The Open University, Milton Keynes, UK

September 29, 2008

Abstract

Real-time traffic flow data across entire networks can be used in a traffic management system to monitor current traffic flows so that traffic can be directed and managed efficiently. Reliable short-term forecasting models of traffic flows are crucial for the success of any traffic management system.

The model proposed in this paper for forecasting traffic flows is a multivariate Bayesian dynamic model called the multiregression dynamic model (MDM). This model is an example of a dynamic Bayesian network and is designed to preserve the conditional independences and causal drive exhibited by the traffic flow series.

Sudden changes can occur in traffic flow series in response to such events as traffic accidents or roadworks. A traffic management system is particularly useful at such times of change. To ensure that the associated forecasting model continues to produce reliable forecasts, despite the change, the MDM uses the technique of external intervention. This paper will demonstrate how intervention works in the MDM and how it can improve forecast performance at times of change.

External intervention has also been used in the context of Bayesian networks to identify causal relationships between variables, and in dynamic Bayesian networks to identify lagged causal relationships between time series. This paper goes beyond the identification of lagged causal relationships previously addressed using intervention in dynamic Bayesian networks, to show how intervention in the MDM can be used to identify contemporaneous causal relationships between time series.

Keywords: multivariate time series, Bayesian forecasting, dynamic linear model, multiregression dynamic model

⁰Catriona M. Queen and Casper J. Albers are lecturer and research fellow, respectively, at the Department of Mathematics and Statistics, Open University, Walton Hall, Milton Keynes, MK7 6AA, U.K. (emails: *c.queen@open.ac.uk* and *c.j.albers@open.ac.uk*). The authors would like to thank Kent County Council for providing the data used in this paper, and also the referees and Associate Editor for their useful comments.

1. INTRODUCTION

As the number of vehicles on roads worldwide continues to rise, the problem of keeping traffic flowing is becoming increasingly important. Many roads now have induction loops implanted into the road surface at various sites, providing real-time traffic flow data across entire traffic networks. These data can be used as part of a traffic management system to monitor traffic flows and reduce congestion by, for example, imposing variable speed limits or diverting traffic onto alternative routes. Reliable short-term forecasting models of traffic flows are crucial for the success of any such traffic management system.

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, only a few short-term forecasting models make use of this fact by using lagged values at other data collection sites when modelling the flow at site S (Tebaldi, West and Karr 2002; Kamarianakis and Prastacos 2005; Stathopoulos and Karlaftis 2003). Whittaker, Garside and Lindveld (1997) and Sun, Zhang and Yu (2006) additionally use conditional independence so that only lagged flows of adjacent data collection sites are required in their models. When the distance between sites is such that vehicles are counted at a number of different sites during the same time period, the flow at other sites at lag 0 is useful for forecasting the flow at site S . The models used in this paper, Whitlock and Queen (2000) and Queen, Wright and Albers (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.

Following Whitlock and Queen (2000) and Queen *et al.* (2007), this paper uses a multivariate Bayesian dynamic model, called the multiregression dynamic model

(MDM) (Queen and Smith 1993), for forecasting traffic flows. The MDM uses any conditional independence and causal structure across the time series and is an example of a dynamic Bayesian network. A Bayesian network (BN) is a directed acyclic graph in which variables are represented by nodes and arcs between nodes represent conditional dependencies between the variables. A dynamic BN is a Bayesian network for a sequence of variables such as a time series or stochastic process. Dynamic BNs of various forms have received a lot of interest in recent years (see for example Kjærulff 1995; Brillinger 1996; Farrow, Goldstein and Spiropoulos 1997; Dahlhaus 2000; Dahlhaus and Eichler 2003; Sun *et al.* 2006; Smith and Figueroa 2007).

In the MDM, at each time t , the observable component series $Y_t(1), \dots, Y_t(n)$ of the n -dimensional time series, and their associated state vectors $\theta_t(1), \dots, \theta_t(n)$, are represented by a BN. These individual BNs are linked together over time to form a dynamic BN. As in Sun *et al.* (2006), the direction of traffic flow produces the causal drive in the system and the possible routes through the network are used to define a conditional independence structure across the time series. The MDM then uses the conditional independences and causal driving mechanism through the system, as represented by the dynamic BN, to break down the multivariate model into simpler univariate regression dynamic linear models (DLMs) (West and Harrison 1997) where contemporaneous traffic flows at upstream links in the network are used as regressors. Tebaldi *et al.* (2002) also use regression DLMs when modelling traffic flows. As in this paper, they use traffic flows at upstream links in the network as linear regressors. However, whereas the vehicle counts in this paper are for one-hour intervals, those in Tebaldi *et al.* (2002) are for one-minute intervals, and so regression on lagged flows (rather than contemporaneous flows) is required.

Traffic flows can exhibit sudden changes in response to events such as congestion, road traffic accidents, roadworks or adverse weather conditions. It is during these pe-

riods of change when a traffic management system is often of most use and it is crucial to the success of a traffic management system that its associated forecasting model continues to perform well despite the change. To maintain forecast performance, the MDM uses the technique of external intervention in which forecast distributions are manipulated externally to the system. This is a long established technique in the context of DLMS (West and Harrison 1986, 1989, 1997).

External intervention has also been used in the context of BNs with the different aim of identifying causal relationships between variables (Pearl 1995, 2000; Spirtes, Glymour and Scheines 2000; Lauritzen 2000; Lauritzen and Richardson 2002; Dawid 2002). Here, intervention for a random variable X_i means that the value of X_i is manipulated externally and is assigned the value x_i . Then X_i is causal for X_j if intervention with respect to X_i affects X_j 's distribution. Didelez (2003) and Eichler and Didelez (2007) consider the effects of external intervention in a dynamic setting for dynamic BNs. Here, a time series $\{X_t(i)\}$ is said to be causal for the time series $\{X_t(j)\}$ if an intervention with respect to $X_t(i)$ affects the predictions of $X_{t+k}(j)$, for some future time $t + k$. Thus intervention in dynamic BNs has been used to investigate lagged causal relationships between time series.

In this paper it will be shown how, in addition to maintaining good forecast performance, intervention in the MDM can aid the identification of *contemporaneous* causal relationships between time series. This goes beyond the identification of lagged causal relationships previously addressed using intervention in dynamic BNs. The MDM uses contemporaneous causal relationships between component series explicitly: different contemporaneous causal relationships produce different MDMs. It is therefore important that the correct causal relationships between traffic flow series are used when specifying the MDM. In normal traffic conditions, these are dictated by the direction of traffic flow and possible routes through the network: generally

flow series at adjacent upstream sites to a particular site S are causal for the flow series at site S . However, queuing traffic can change causal relationships, because a queue at site S can feed upstream so that the flow series at site S can become (temporarily) causal for the flows at upstream sites, as well as for the flows at downstream sites. The MDM offers a method of identifying causal relationships on-line, ensuring that the ‘correct’ causal relationships, and consequently the ‘correct’ MDM, is always used.

Although the focus of this paper is the use of intervention in the MDM with respect to forecasting traffic flows, there are many potential application areas for the MDM, including modelling economic indicators such as energy consumption and GDP, environmental resource management problems, industrial product distribution flow problems and medical patient physiological monitoring. As a specific example, monthly brand sales in a competitive market are modelled using an MDM in Queen (1994) and Queen, Wright and Albers (2008). Here, the competition in the market is the causal drive in the system and is used to elicit a conditional independence structure across the time series: this is demonstrated in Queen (1997) and Queen, Smith and James (1994). Goldstein, Farrow and Spiropoulos (1993), Farrow *et al.* (1997) and Farrow (2003) also use dynamic BNs to represent brand relationships when forecasting time series of brand sales. Guo and Brown (2001) use a generalisation of the MDM to model bivariate hormone time series. Like the MDM, their model uses univariate structural models where structural parameters are functions of the other series. Their model reduces to the MDM in the special case in which the relationship between the two series is in one direction only. Fosen *et al.* (2006) also use similar ideas to the MDM, using parents from a dynamic BN as regressors, when analysing a trial of cancer patients with liver cirrhosis.

The intervention techniques and methods for identifying causal relationships de-

tailed in this paper are potentially applicable to any series suitable for modelling using an MDM. Many series can exhibit sudden changes for which intervention is appropriate to maintain forecast performance. For example, sales series often exhibit sudden changes in response to marketing activity. Although the conditional independence structure related to causality is fairly straightforward to elicit for traffic networks under normal traffic conditions (see Queen *et al.* 2007), this is not necessarily the case for other potential application areas. Furthermore, like the traffic flow application, causal relationships across a multivariate time series may change over time. For example, when forecasting sales data in competitive markets, marketing activity can result in changes in the causal relationships between brand sales. Thus the methods presented in this paper represent a significant advance for using the MDM in practice, as well as for the notoriously difficult problem of identifying contemporaneous causal relationships in general multivariate time series problems.

The paper will focus on a single traffic network near London, UK. The network will be described in Section 2, while an MDM is defined for it in Section 3. Details of how intervention in the MDM works are presented in Section 4 before looking at intervention in practice in the London network in Section 5. Section 6 shows how the MDM can be used to identify contemporaneous causal relationships between series. Finally, Section 7 offers some concluding remarks.

2. THE LONDON NETWORK

This paper focuses on using intervention when forecasting vehicle counts at the junction of three major roads — the M25, A2 and A296 — east of London, UK. Figure 1(a) shows an aerial photograph of the network. The traffic data are hourly counts of vehicles passing over induction loops in the road surface at a number of data collection sites in the network. A diagram of the network showing the layout of

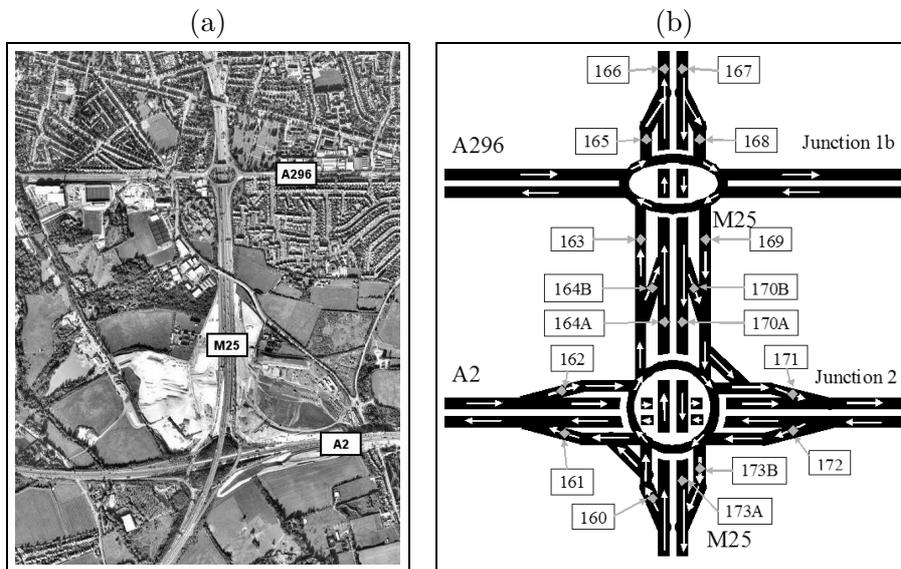


Figure 1: The London network (a) aerial photograph (taken from Google Maps) (b) schematic diagram: the grey diamonds are the (numbered) data collection sites and the white arrows indicate the direction of traffic flow.

the data collection sites is given in Figure 1(b). The data used here were collected between 30th January 1995 and 26th June 1995. Data for this network are available at <http://trads.hatris.co.uk>.

The network is such that traffic flows into the network, through a number of data collection sites, and then out of the network. During normal conditions it will only take a few minutes for a vehicle to traverse the network, so most vehicles are counted at a number of different sites during the same time period.

The vehicle counts have a strong seasonal pattern with peaks in the morning and evening rush hours. The daily pattern is different on a weekday than it is at the weekend, and also slightly different for Monday and Friday. It is possible to incorporate these differences into the model, but for clarity of presentation, only traffic flows for Tuesday–Thursday each week are considered here. The hourly counts

in this network will be modelled by an MDM, as described in the next section.

3. A MULTIREGRESSION DYNAMIC MODEL FOR THE LONDON NETWORK

In this section, an MDM (Queen and Smith, 1993) will be defined for the London network. Let $Y_t(i)$ be the vehicle count for site i at hour t . Then let $\mathbf{Y}_t = (Y_t(1), \dots, Y_t(n))^\top$ denote the n -dimensional multivariate time series and let $\mathbf{Y}^t = (\mathbf{Y}_1, \dots, \mathbf{Y}_t)^\top$ and $\mathbf{Y}^t(i) = (Y_1(i), \dots, Y_t(i))^\top$.

Suppose that the variables are ordered and indexed so that there is a conditional independence structure related to causality so that at each time $t \in \mathbb{N}$, for $i = 2, \dots, n$,

$$\begin{aligned} Y_t(i) &\perp\!\!\!\perp (\{Y_t(1), \dots, Y_t(i-1)\} \setminus pa(Y_t(i))) | pa(Y_t(i)), \\ Y_t(i) &\perp\!\!\!\perp \{\{\mathbf{Y}^t(1), \dots, \mathbf{Y}^t(i-1)\} \setminus pa(\mathbf{Y}^t(i))\} | (pa(\mathbf{Y}^t(i)), \mathbf{Y}^{t-1}(i)). \end{aligned}$$

The notation $A \perp\!\!\!\perp B | C$ reads “ A is independent of B given C ” (Dawid 1979), “ \setminus ” reads “excluding” and $pa(Y_t(i)) \subseteq \{Y_t(1), \dots, Y_t(n)\}$. Each variable in the set $pa(Y_t(i))$ is a parent of $Y_t(i)$ so that in a BN representing the conditional independence relationships, there is a directed arc to $Y_t(i)$ from each of its parents. Thus over time the conditional independence relationships are represented by a dynamic BN.

In Queen *et al.* (2007), the possible routes through the London network were used to elicit the conditional independences related to causality across \mathbf{Y}_t at each time t . A BN representing these relationships, suitable for use with an MDM, was elicited. Because some of the data collection sites were not operational, the BN separates out into two parts. The larger of these parts, representing vehicle counts at sites labelled 167, 168, 170A, 170B, 169, 161 and 171, is presented here in Figure 2.

In the BN of Figure 2, $Y_t(170AB)$ denotes $Y_t(170A) + Y_t(170B)$ and $Y_t(161.171)$ denotes $Y_t(161) + Y_t(171)$. Each node $\theta_t(i)$ is the parameter vector associated with

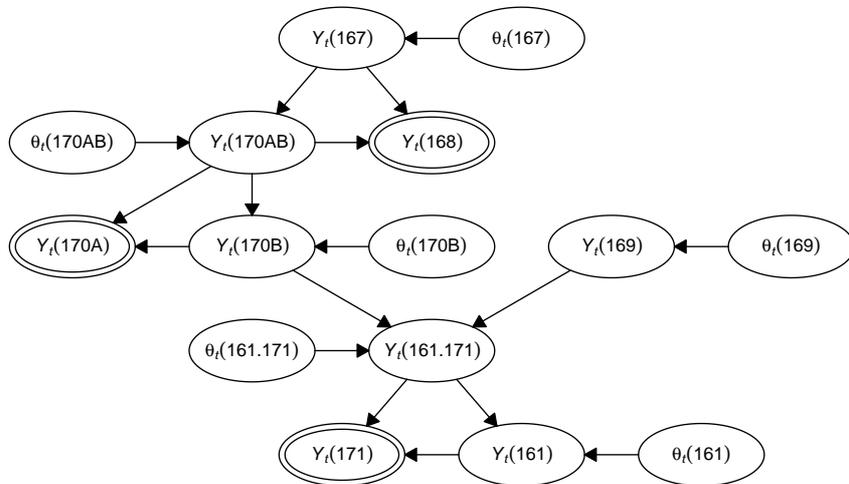


Figure 2: Part of the BN for each time t representing traffic flows in the London network of Figure 1.

$Y_t(i)$ in the MDM. These are mutually independent at each time t within the MDM framework. Three variables are logical functions of their parents:

$$\begin{aligned}
 Y_t(168) &= Y_t(167) - Y_t(170AB), \\
 Y_t(170A) &= Y_t(170AB) - Y_t(170B), \\
 Y_t(171) &= Y_t(161.171) - Y_t(161).
 \end{aligned} \tag{1}$$

Following the terminology of WinBUGS software, these are called logical variables and denoted by double ovals. Note that all these logical variables are also general time series. However, it is not possible to model them directly, because then all the parameter vectors would no longer be mutually independent. (For full details regarding the BN of Figure 2 and how it was elicited, see Queen *et al.*, 2007.)

Denote the information available at time t by D_t . The MDM for the n -dimensional vector time series \mathbf{Y}_t over times $t = 1, 2, \dots$, is defined by the n observation equations,

the system equation, and information at time $t - 1$ as follows.

$$\begin{aligned} \text{Observation equations: } & Y_t(i) = \mathbf{F}_t(i)^\top \boldsymbol{\theta}_t(i) + v_t(i), & v_t(i) &\sim (0, V_t(i)), \quad 1 \leq i \leq n \\ \text{System equation: } & \boldsymbol{\theta}_t = G_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, & \mathbf{w}_t &\sim (\mathbf{0}, W_t) \\ \text{Information: } & (\boldsymbol{\theta}_{t-1} | D_{t-1}) &\sim (\mathbf{m}_{t-1}, C_{t-1}). \end{aligned}$$

The m_i -dimensional vector $\mathbf{F}_t(i)$ contains an arbitrary, but known, function of the parents $pa(Y_t(i))$ and possibly other known variables (which may include \mathbf{Y}^{t-1}); $\boldsymbol{\theta}_t(i)$ is the m_i -dimensional parameter vector for $Y_t(i)$; $V_t(1), \dots, V_t(n)$ are the scalar observation variances; $\boldsymbol{\theta}_t^\top = (\boldsymbol{\theta}_t(1)^\top, \dots, \boldsymbol{\theta}_t(n)^\top)$ is the m -dimensional parameter vector; \mathbf{m}_{t-1} and C_{t-1} are the (posterior) moments for $\boldsymbol{\theta}_{t-1}$ at time $t - 1$; and the block diagonal $m \times m$ matrices G_t , W_t , and C_{t-1} are assumed known (and are not functions of $pa(Y_t(i))$). The error vectors, $\mathbf{v}_t^\top = (v_t(1), \dots, v_t(n))$ and $\mathbf{w}_t^\top = (\mathbf{w}_t(1)^\top, \dots, \mathbf{w}_t(n)^\top)$, are such that $v_t(1), \dots, v_t(n)$ and $\mathbf{w}_t(1), \dots, \mathbf{w}_t(n)$ are mutually independent and $\{\mathbf{v}_t, \mathbf{w}_t\}_{t \in \mathbb{N}}$ are mutually independent with time.

The MDM uses the conditional independence structure to model the multivariate time series by n separate univariate models — for $Y_t(1)$ and $Y_t(i)|pa(Y_t(i))$, $i = 2, \dots, n$. Note that no distributional assumptions have been placed on the error terms or the distribution for $\boldsymbol{\theta}_{t-1}$. Also, there is no specific requirement that $\mathbf{F}_t(i)$ be a linear function of $pa(Y_t(i))$, just that the function is known. Thus the MDM is a very general model. When $\mathbf{F}_t(i)$ is a linear function of $pa(Y_t(i))$ and the error distributions are normal, then this is the special case of the *linear multiregression dynamic model*. In this case, each $Y_t(i)$ with parents is modelled by a regression DLM with its parents as (linear) regressors, and each $Y_t(i)$ without parents is modelled by any appropriate DLM. Linear MDMs are particularly simple to use analytically and will be used in this paper to forecast traffic flows in the London network.

From the BN in Figure 2, neither $Y_t(167)$ nor $Y_t(169)$ have parents so each of these is modelled separately by univariate DLMS. For $i = 167, 169$, to account for the

seasonality exhibited by each $Y_t(i)$, a seasonal factor model is used with a separate level parameter for each hour of the day. (This was shown to perform as well as the Fourier model, and has the added advantage of interpretability, which is helpful when using intervention.) Thus in this case, $\boldsymbol{\theta}_t(i)$ is a 24-dimensional vector with a seasonal factor for each of the 24 hours, $\mathbf{F}_t(i)^\top = (1, 0, \dots, 0)$ and, for $a \in [0, 1]$,

$$G_t(i) = \begin{pmatrix} a & 1-a & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & 1 \\ 1 & 0 & \cdots & \cdots & 0 \end{pmatrix}. \quad (2)$$

So the system equation links the current hour parameter (hour t) with the parameter for the previous hour (hour $t-1$), as well as the parameter for the corresponding hour the previous day (hour $t-24$). The optimal value of a for the London network was found to be 0.01.

From Figure 2, series $Y_t(170AB)$, $Y_t(170B)$, $Y_t(161.171)$ and $Y_t(161)$ all have parents. Thus each is modelled by a regression DLM with its parents as regressors. In this case the regression parameters represent the proportion of traffic flowing from parents to $Y_t(i)$. These proportions exhibit a seasonal daily pattern which remains stable over time. Thus a separate regression parameter is defined for each hour for each parent, and, in the case where $Y_t(i)$ has a single parent, $G_t(i)$ is given by (2) with $a = 0$. Thus the current hour parameter is linked with the parameter for the corresponding hour of the previous day.

The remaining series, $Y_t(168)$, $Y_t(170A)$ and $Y_t(171)$, are logical variables and their forecasts are simply derived from those of their parents in accordance with (1).

The value of the observation variances $V_t(i)$ are estimated on-line using standard DLM variance learning techniques (see West and Harrison (1997), Section 2.5) and the evolution variance W_t is estimated using established DLM discounting techniques

(see West and Harrison (1997), Section 2.4).

As long as $\boldsymbol{\theta}_t(1), \boldsymbol{\theta}_t(2), \dots, \boldsymbol{\theta}_t(n)$ are mutually independent initially (i.e. C_0 is block diagonal), then the block diagonal form of G_t and W_t ensures that the $\boldsymbol{\theta}_t(i)$ remain mutually independent and that each $\boldsymbol{\theta}_t(i)$ can be updated separately in closed form from $Y_t(i)$'s (conditional) univariate model. Forecasts for $Y_t(1)$ and $Y_t(i)|pa(Y_t(i)), i = 2, \dots, n$, and the k -step forecasts for $Y_{t+k}(1)$ and $Y_{t+k}(i)|pa(Y_{t+k}(i)), k \in \mathbb{N}, i = 2, \dots, n$, could also then be found separately, often using established dynamic model results. However, the values of $pa(Y_t(i))$ are not available when forecasting $Y_t(i)$, since $Y_t(i)$ and $pa(Y_t(i))$ are observed simultaneously. Similarly, the values of $pa(Y_{t+k}(i))$ are not available when forecasting $Y_{t+k}(i)$. So the marginal forecasts for each $Y_t(i)$ and $Y_{t+k}(i)$, $k \in \mathbb{N}$, without conditioning on the values of parents, are required. Unfortunately, the marginal forecast distributions for $Y_t(i)$ and $Y_{t+k}(i)$, $k \in \mathbb{N}, i = 2, \dots, n$, will not generally be of a simple form. However, (under quadratic loss) the marginal moments of the forecast distributions are adequate for forecasting purposes, and these can be easily found for many MDMs. (See Queen and Smith (1993) and Queen *et al.* (2008) for further details.)

Because the MDM breaks the multivariate model into univariate components, computations in the MDM are fast and efficient, no matter how large and complex the network is. The software used in this paper was written in R and had a speed of over 1200 forecasts per second.

It is important to realise that although two BNs can represent the same conditional independences, they can represent quite different conditional independences related to causality and consequently, quite different MDMs. For example, consider the BN in Figure 2 and a second BN which is exactly the same except that the arc between $Y_t(167)$ and $Y_t(170AB)$ is reversed. These two BNs are probabilistically equivalent. However, the conditional independences related to causality re-

late to two quite different MDMs. For the BN in Figure 2 an MDM would model $Y_t(167)$ and $Y_t(170AB)|Y_t(167)$, whereas for the second BN, an MDM would model $Y_t(167)|Y_t(170AB)$ and $Y_t(170AB)$.

4. INTERVENTION IN THE MULTIREGRESSION DYNAMIC MODEL

Intervention in the BN context is usually ‘atomic’ where X_i is assigned a specific (single) value x_i . Intervention in the context of the DLM is instead ‘random’ where the *distribution* of the random variable is manipulated.

The need for intervention in the time series context can be triggered by poor model forecast performance or by expert information regarding external events (such as planned roadworks). As pointed out by West and Harrison (1997), detecting when intervention is required, and monitoring the intervention afterwards, can be informal involving the forecaster’s subjective judgement (by examining the pattern and/or magnitude of the one-step forecast errors, for example), or can be based on more formal monitoring techniques. Formal monitoring techniques are long-established for DLMs (see West and Harrison, 1997, Chp. 11). Because the MDM decomposes into (conditional) DLMs, it should be relatively straightforward to develop such monitors for the MDM. In this paper, however, the focus lies in the use of intervention and informal monitoring only is used. The development of a suitable formal monitor for the MDM will be addressed in future work.

As is assumed in Lauritzen (2000) and Dawid (2002) with respect to BNs, intervention in the DLM always precedes observation, so that intervention at time t is done before forecasts are made and the series is observed at time t . In the MDM it is possible to intervene (separately) for any number and combination of the individual series $Y_t(1), \dots, Y_t(n)$ and/or their associated parameter vectors $\theta_t(1), \dots, \theta_t(n)$ at

any particular time t . Thus, sequential decisions are made at each time point regarding intervention on the time series, as in Dawid and Didelez (2005). For simplicity here, only intervention at the single time point t for a single component $Y_t(i)$, and its associated parameter vector $\boldsymbol{\theta}_t(i)$, will be considered.

The observation equation for $Y_t(i)$ specifies the distribution $Y_t(i)|(\boldsymbol{\theta}_t(i), pa(Y_t(i))) \sim (\mathbf{F}_t(i)^\top \boldsymbol{\theta}_t(i), V_t(i))$. Intervention for $Y_t(i)$ in the MDM involves manipulating this distribution, to give the intervention distribution

$$Y_t(i)|(\boldsymbol{\theta}_t(i), pa(Y_t(i)), \text{intervention}) \sim (\mathbf{F}_t(i)^\top \boldsymbol{\theta}_t(i) + h_t(i), V_t(i) + H_t(i)), \quad (3)$$

for suitable scalars $h_t(i)$ and $H_t(i)$. The system equation specifies the distribution $\boldsymbol{\theta}_t| \boldsymbol{\theta}_{t-1} \sim (G_t \boldsymbol{\theta}_{t-1}, W_t)$. So when intervening for $\boldsymbol{\theta}_t(i)$, the part of the system equation associated with $\boldsymbol{\theta}_t(i)$ is manipulated to produce the intervention distribution

$$\boldsymbol{\theta}_t(i)|(\boldsymbol{\theta}_{t-1}(i), \text{intervention}) \sim (G_t(i)^* \boldsymbol{\theta}_{t-1}(i), W_t(i)^*),$$

for suitable matrices $G_t(i)^*$ and $W_t(i)^*$. The values of $h_t(i)$ and $G_t(i)^*$ reflect the expected change in $Y_t(i)$ and $\boldsymbol{\theta}_t(i)$, respectively, at time t , and $H_t(i)$ and $W_t(i)^*$ reflect the uncertainty regarding the changes.

Because the MDM is a dynamic BN, intervention for $Y_t(i)$ or $\boldsymbol{\theta}_t(i)$ not only affects the manipulated distributions, but also affects other variables in the dynamic BN. This makes intervention in the MDM a very powerful forecasting technique. To investigate the effects of interventions, this paper will use influence diagrams.

An influence diagram is a generalisation of a BN with random nodes (drawn as ovals) and decision nodes (drawn as rectangles), which can be used to represent and solve Bayesian decision problems (Howard and Matheson 1984; Shachter 1986, 1988; Oliver and Smith 1990). The value of a decision node arises through external intervention by a decision maker. Arcs leading into random nodes represent conditional

dependencies, whereas arcs leading into decision nodes imply that information regarding its parents is assumed available before a decision is made.

Following the ideas of Dawid (2002), introduce indicator variables $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$, where $\sigma(Y_t(i))$ only takes the value 1 when intervention for $Y_t(i)$ occurs, and $\sigma(\boldsymbol{\theta}_t(i))$ only takes the values 1 when intervention for $\boldsymbol{\theta}_t(i)$ occurs. Note that, although similar in concept, the intervention variables used here are different to those used in Dawid (2002), where atomic interventions were used with a finite set of possible interventions. With random intervention the distribution for $Y_t(i)$ or $\boldsymbol{\theta}_t(i)$ can be manipulated arbitrarily at intervention. Thus, $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$ are simply indicators as to whether intervention takes place or not. The conditional distributions can then be defined: $Y_t(i)|(\boldsymbol{\theta}_t(i), pa(Y_t(i)), \sigma(Y_t(i)))$ and $\boldsymbol{\theta}_t(i)|(\boldsymbol{\theta}_{t-1}(i), \sigma(\boldsymbol{\theta}_t(i)))$.

The intervention variables, $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$, are decision variables whose values are controlled by the forecaster using the forecasting model. Following the ideas of Dawid (2002), $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$, will be added as decision nodes to a dynamic BN of the MDM to produce an influence diagram of the MDM. The MDM uses the conditional independence structure related to causality represented in the influence diagram. As such, the descendants of $\sigma(Y_t(i))$ will be affected by intervention for $Y_t(i)$, and the descendants of $\sigma(\boldsymbol{\theta}_t(i))$ will be affected by interventions for $\boldsymbol{\theta}_t(i)$.

In order to draw an influence diagram which represents the general structure of all MDMs, define the following notation. Let

$$\begin{aligned}\mathbf{X}_t(i)^\top &= (Y_t(1), \dots, Y_t(i-1)) \\ \mathbf{Z}_t(i)^\top &= (Y_t(i+1), \dots, Y_t(n)) \\ \boldsymbol{\alpha}_t(i)^\top &= (\boldsymbol{\theta}_t(1)^\top, \dots, \boldsymbol{\theta}_t(i-1)^\top) \\ \boldsymbol{\beta}_t(i)^\top &= (\boldsymbol{\theta}_t(i+1)^\top, \dots, \boldsymbol{\theta}_t(n)^\top).\end{aligned}$$

Thus, for $i = 2, \dots, n-1$, the time series can be written as $\mathbf{Y}_t^\top = (\mathbf{X}_t(i)^\top, Y_t(i), \mathbf{Z}_t(i)^\top)$,

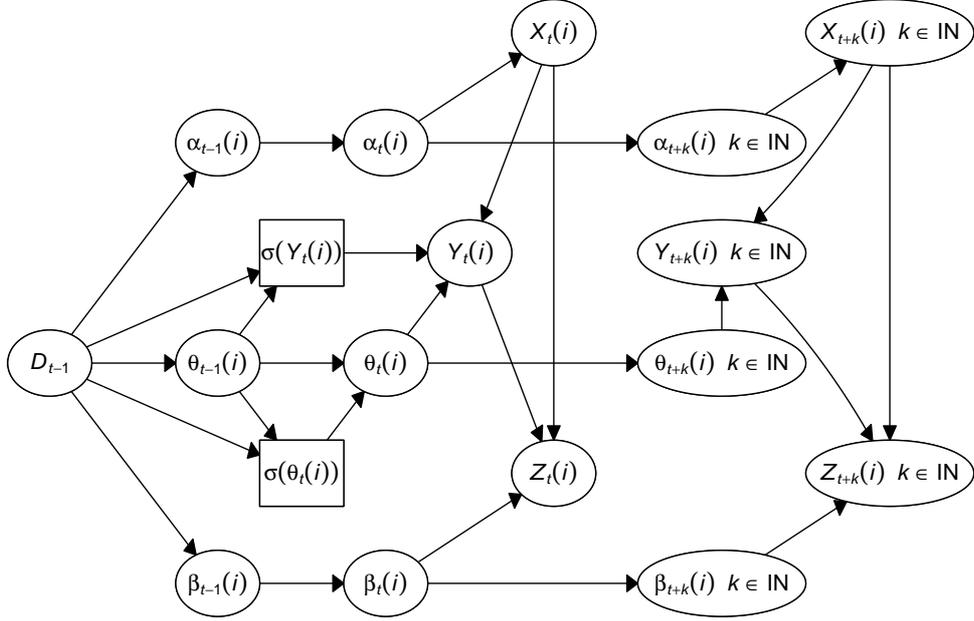


Figure 3: Influence diagram for the multiregression dynamic model before \mathbf{Y}_t is observed, together with the intervention decision variables, $\sigma(Y_t(i))$ and $\sigma(\theta_t(i))$.

with parameter vector $\theta_t^\top = (\alpha_t(i)^\top, \theta_t(i)^\top, \beta_t(i)^\top)$. Then since $pa(Y_t(i)) \subseteq \mathbf{X}_t(i)$, the vector $\mathbf{F}_t(i)$ is a known function of $\mathbf{X}_t(i)$, and $Y_t(i) | (\mathbf{X}_t(i), \theta_t(i))$ has some distribution with mean $\mathbf{F}_t(i)^\top \theta_t(i)$ and variance $V_t(i)$.

Suppose that $\mathbf{Y}_1, \dots, \mathbf{Y}_{t-1}$ have been observed. An influence diagram representing the MDM before \mathbf{Y}_t is observed, is given in Figure 3. Also included in the influence diagram are the intervention decision variables, $\sigma(Y_t(i))$ and $\sigma(\theta_t(i))$.

In the influence diagram there are arcs leading from D_{t-1} and $\theta_{t-1}(i)$ to both $\sigma(Y_t(i))$ and $\sigma(\theta_t(i))$ because the decision to intervene at time t is often a reflection of past behaviour of the series and/or its model. (It is also possible that the decision to intervene is entirely dictated by information external to the model.)

The other arcs in the influence diagram are direct consequences of the MDM. The parameter vectors $\boldsymbol{\alpha}_{t-1}(i)$, $\boldsymbol{\theta}_{t-1}(i)$ and $\boldsymbol{\beta}_{t-1}(i)$ each have the single parent D_{t-1} , since the posterior distributions $\boldsymbol{\alpha}_{t-1}(i)|D_{t-1}$, $\boldsymbol{\theta}_{t-1}(i)|D_{t-1}$ and $\boldsymbol{\beta}_{t-1}(i)|D_{t-1}$ represent all knowledge of $\boldsymbol{\alpha}_{t-1}(i)$, $\boldsymbol{\theta}_{t-1}(i)$ and $\boldsymbol{\beta}_{t-1}(i)$, respectively, after the first $t - 1$ observations are made. The block diagonal form of C_{t-1} ensures that $\boldsymbol{\theta}_{t-1}(1), \dots, \boldsymbol{\theta}_{t-1}(n)$ are all mutually independent given D_{t-1} and so there are no arcs between $\boldsymbol{\alpha}_{t-1}(i)$, $\boldsymbol{\theta}_{t-1}(i)$ and $\boldsymbol{\beta}_{t-1}(i)$. The block diagonal forms of $G_t, G_{t+1}, \dots, G_{t+k}$ and $W_t, W_{t+1}, \dots, W_{t+k}$ ensure that repeated use of the system equation specifies separate distributions for $\boldsymbol{\theta}_t(i)|\boldsymbol{\theta}_{t-1}(i)$ and each $\boldsymbol{\theta}_{t+k}(i)|\boldsymbol{\theta}_t(i)$, $i = 1, \dots, n$, for $k \in \mathbb{N}$. Therefore $\{\boldsymbol{\alpha}_{t+k}(i), k \in \mathbb{N}\}$, $\{\boldsymbol{\theta}_{t+k}(i), k \in \mathbb{N}\}$ and $\{\boldsymbol{\beta}_{t+k}(i), k \in \mathbb{N}\}$ have the single parents $\boldsymbol{\alpha}_t(i)$, $\boldsymbol{\theta}_t(i)$ and $\boldsymbol{\beta}_t(i)$, respectively, which in turn have single parents $\boldsymbol{\alpha}_{t-1}(i)$, $\boldsymbol{\theta}_{t-1}(i)$ and $\boldsymbol{\beta}_{t-1}(i)$, respectively. The n observation equations at times t and $t + k$, $k \in \mathbb{N}$ define the distributions for $Y_t(i)|(\boldsymbol{\theta}_t(i), pa(Y_t(i)))$ and $Y_{t+k}(i)|(\boldsymbol{\theta}_{t+k}(i), pa(Y_{t+k}(i)))$, $i = 1, \dots, n$. So $\mathbf{X}_t(i)$ has parent $\boldsymbol{\alpha}_t(i)$, $\{\mathbf{X}_{t+k}(i), k \in \mathbb{N}\}$ has parent $\{\boldsymbol{\alpha}_{t+k}(i), k \in \mathbb{N}\}$, $Y_t(i)$ has parents $\boldsymbol{\theta}_t(i)$ and $\mathbf{X}_t(i)$, $\{Y_{t+k}(i), k \in \mathbb{N}\}$ has parents $\{\boldsymbol{\theta}_{t+k}(i), k \in \mathbb{N}\}$ and $\{\mathbf{X}_{t+k}(i), k \in \mathbb{N}\}$, $\mathbf{Z}_t(i)$ has parents $\boldsymbol{\beta}_t(i)$, $\mathbf{X}_t(i)$ and $Y_t(i)$, and $\{\mathbf{Z}_{t+k}(i), k \in \mathbb{N}\}$ has parents $\{\boldsymbol{\beta}_{t+k}(i), k \in \mathbb{N}\}$, $\{\mathbf{X}_{t+k}(i), k \in \mathbb{N}\}$ and $\{Y_{t+k}(i), k \in \mathbb{N}\}$. Additionally $Y_t(i)$ and $\boldsymbol{\theta}_t(i)$ have parents $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$, respectively.

4.1 EFFECTS OF INTERVENTION IN THE MDM BEFORE OBSERVING \mathbf{Y}_t

The effects of intervention for $Y_t(i)$ and $\boldsymbol{\theta}_t(i)$ in the MDM can be easily seen by looking at the descendants of $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$, respectively, in the influence diagram of Figure 3.

As $\sigma(Y_t(i))$ has descendants $Y_t(i)$ and $\mathbf{Z}_t(i)$, intervention for $Y_t(i)$ affects $Y_t(i)$'s forecast and $\mathbf{Z}_t(i)$'s forecast only. Note that intervention for $Y_t(i)$ does not affect the forecasts for $\mathbf{X}_t(i)$, nor the priors for any parameters at time $t + k$, nor the k -step

ahead forecasts for \mathbf{Y}_{t+k} .

On the other hand, $\sigma(\boldsymbol{\theta}_t(i))$ has descendants $\boldsymbol{\theta}_t(i)$, $Y_t(i)$, $\mathbf{Z}_t(i)$, $\{\boldsymbol{\theta}_{t+k}(i), k \in \mathbb{N}\}$, $\{Y_{t+k}(i), k \in \mathbb{N}\}$ and $\{\mathbf{Z}_{t+k}(i), k \in \mathbb{N}\}$. Therefore intervention for $\boldsymbol{\theta}_t(i)$ affects: the prior for $\boldsymbol{\theta}_t(i)$, the one-step ahead forecasts for $Y_t(i)$ and $\mathbf{Z}_t(i)$, the priors for $\{\boldsymbol{\theta}_{t+k}(i), k \in \mathbb{N}\}$, and the k -step ahead forecasts for $\{Y_{t+k}(i), k \in \mathbb{N}\}$ and $\{\mathbf{Z}_{t+k}(i), k \in \mathbb{N}\}$. Intervening for $\boldsymbol{\theta}_t(i)$ does not affect the one-step, or k -step, ahead forecasts for $\mathbf{X}_t(i)$, nor the priors for any parameters for components of \mathbf{Y}_t other than $Y_t(i)$, either at time t (that is, $\boldsymbol{\alpha}_t(i)$, $\boldsymbol{\beta}_t(i)$) or at future time periods (that is, $\{\boldsymbol{\alpha}_{t+k}(i), k \in \mathbb{N}\}$, $\{\boldsymbol{\beta}_{t+k}(i), k \in \mathbb{N}\}$).

The influence diagram given in Figure 3 is a generic representation of the structure of *all* MDMs. For a specific MDM, not every variable in $\mathbf{Z}_t(i)$ is necessarily a descendant of $Y_t(i)$. For example, in the dynamic BN of the London network given in Figure 2, when considering intervention for $Y_t(170B)$, say, then $Y_t(169)$ could be included in $\mathbf{Z}_t(170B)$ although $Y_t(169)$ is not a descendant of $Y_t(170B)$ and so will not be affected by intervention for $Y_t(170B)$. As a result, intervention for $Y_t(i)$ will not necessarily affect the forecasts for *all* variables in $\mathbf{Z}_t(i)$, but only the forecasts of those $Y_t(j) \in \mathbf{Z}_t(i)$ which are descendants of $Y_t(i)$. Similarly, intervention for $\boldsymbol{\theta}_t(i)$ will not affect the forecasts of *all* variables in $\mathbf{Z}_t(i)$ or $\{\mathbf{Z}_{t+k}(i), k \in \mathbb{N}\}$, but only those $Y_t(j)$ and $Y_{t+k}(j)$ for which $Y_t(j) \in \mathbf{Z}_t(i)$ are descendants of $Y_t(i)$.

4.2 EFFECTS OF INTERVENTION IN THE MDM AFTER OBSERVING \mathbf{Y}_t

After observing \mathbf{Y}_t , the influence diagram in Figure 3 is no longer appropriate to represent the MDM. Each of the arcs $(\boldsymbol{\alpha}_t(i), \mathbf{X}_t(i))$, $(\boldsymbol{\theta}_t(i), Y_t(i))$ and $(\boldsymbol{\beta}_t(i), \mathbf{Z}_t(i))$ needs to be reversed to reflect the fact that the posterior distributions $\boldsymbol{\alpha}_t(i)|\mathbf{Y}_t$, $\boldsymbol{\theta}_t(i)|\mathbf{Y}_t$ and $\boldsymbol{\beta}_t(i)|\mathbf{Y}_t$ are now of interest (rather than the distributions $Y_t(i)|\boldsymbol{\theta}_t(i)$ specified by the observation equations used for forecasting each $Y_t(i)$). Following

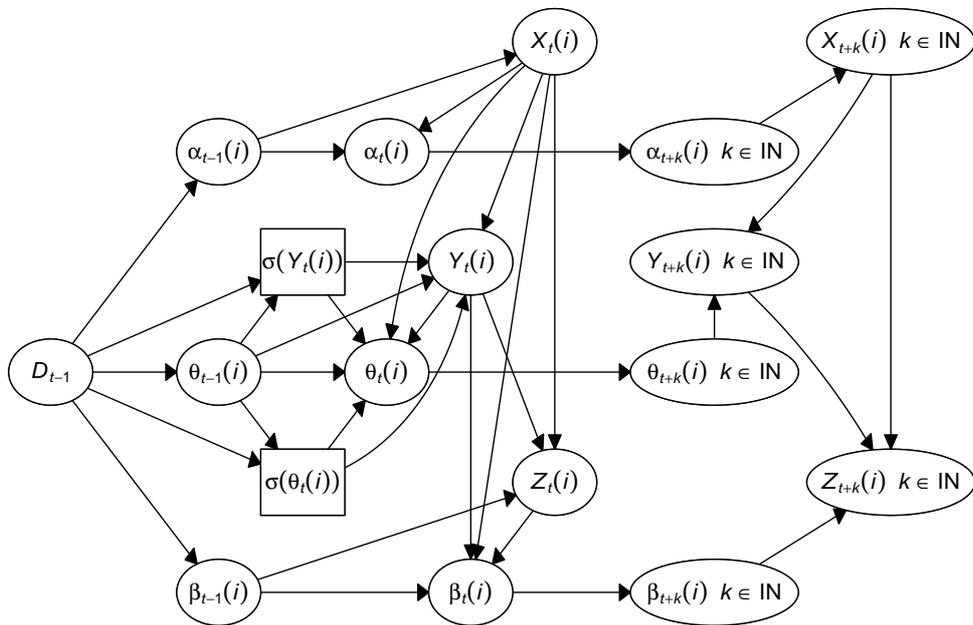


Figure 4: Influence diagram for the multiregression dynamic model, together with intervention indicator decision variables $\sigma(Y_t(i))$ and $\sigma(\theta_t(i))$, *after* \mathbf{Y}_t is observed.

Howard and Matheson's (1984) Arc Reversal Theorem, extra arcs need to be introduced into the influence diagram. Explicitly, reversing the arc between any two nodes A and B means that A must inherit B's parents and B must inherit A's parents. The new influence diagram, after observing \mathbf{Y}_t , is given in Figure 4. Notice how the arc reversals have introduced several new arcs into the influence diagram.

It is important to note that the interventions for $Y_t(i)$ and/or $\theta_t(i)$ still precede the observation \mathbf{Y}_t . However, as the influence diagram representing the MDM changes after observing \mathbf{Y}_t , so the effects of the interventions will change after \mathbf{Y}_t is observed. The effects of intervention after \mathbf{Y}_t is observed are easily seen from the influence diagram in Figure 4 by again looking at the descendants of $\sigma(Y_t(i))$ and

$\sigma(\boldsymbol{\theta}_t(i))$, respectively. This time both $\sigma(Y_t(i))$ and $\sigma(\boldsymbol{\theta}_t(i))$ have the *same* descendants: $Y_t(i)$, $\boldsymbol{\theta}_t(i)$, $\mathbf{Z}_t(i)$, $\boldsymbol{\beta}_t(i)$, $\{\boldsymbol{\theta}_{t+k}(i), k \in \mathbb{N}\}$, $\{\boldsymbol{\beta}_{t+k}(i), k \in \mathbb{N}\}$, $\{Y_{t+k}(i), k \in \mathbb{N}\}$ and $\{\mathbf{Z}_{t+k}(i), k \in \mathbb{N}\}$. Thus after observing \mathbf{Y}_t , intervention for $Y_t(i)$ and $\boldsymbol{\theta}_t(i)$ both affect the posterior for $\boldsymbol{\theta}_t(i)$ and also the posterior for $\boldsymbol{\beta}_t(i)$, the priors for future $\boldsymbol{\theta}_{t+k}(i)$ and also the priors for future $\boldsymbol{\beta}_{t+k}(i)$, the k -step forecasts for $Y_{t+k}(i)$ and the k -step forecasts for $\mathbf{Z}_{t+k}(i)$. It is, however, important to note that although intervention for $Y_t(i)$ and $\boldsymbol{\theta}_t(i)$ both affect the same variables after $\mathbf{Y}_t(i)$ is observed, they affect the variables in *different* ways. Neither intervention for $Y_t(i)$ nor $\boldsymbol{\theta}_t(i)$ affect the posterior for $\mathbf{X}_t(i)$'s parameter vector $\boldsymbol{\alpha}_t(i)$, the prior for future parameters $\boldsymbol{\alpha}_{t+k}$, nor the k -step forecasts for $\mathbf{X}_{t+k}(i)$.

Notice that before \mathbf{Y}_t is observed, intervening for $Y_t(i)$ affects a different set of variables than intervening for $\boldsymbol{\theta}_t(i)$ does. On the other hand, after \mathbf{Y}_t is observed, intervening for $Y_t(i)$ affects exactly the same set of variables (although in different ways) as intervening for $\boldsymbol{\theta}_t(i)$. It is interesting to note that neither intervention affects the forecasts of $\{\mathbf{X}_{t+k}(i), k \in \mathbb{N}\}$, nor the distributions of its parameters.

5. INTERVENTION IN THE LONDON NETWORK

The specific method of intervention to be used when forecasting traffic flows naturally depends on the particular change to be accommodated for a given series. In this section, two specific interventions are considered — one for an observed series $Y_t(i)$ and the other for two state vectors. Both of these are typical of the type of interventions commonly required when forecasting traffic flows.

5.1 INTERVENTION FOR A $Y_t(i)$

Poor forecast performance, as identified by large forecast errors, can be an indication that intervention might be useful. Such an instance of poor forecast performance

occurs in series $Y_t(167)$, where an unusually large negative forecast error occurs at time 560 followed by an unusually large positive forecast error. This pattern of forecast errors is consistent with a slowdown in traffic flow, for example due to a temporary block in the road following a crash, followed by an increase in traffic flow as the problem is resolved and delayed cars move through the network. Such patterns are not uncommon in traffic networks. A plot of the one-step ahead forecast errors with ± 1.96 forecast standard deviation error bars for $Y_t(167)$ between times 500 and 580 is shown in Figure 5(a). The forecast errors for times 560 and 561 are circled on the plot. The same pattern of forecast errors is also evident in $Y_t(167)$'s children (see Figure 5(c)), and in other descendants (see Figure 5(e)). This is due to the fact that traffic from site 167 flows to sites 168, 170A or 170B, so any changes in traffic flow at site 167 will have a knock-on effect to the traffic flows at these sites, and, in turn, to the flows at sites further downstream. It is precisely these kinds of relationships between flows at different sites in the network which the BN was designed to represent. Figure 5(g) shows a plot of the one-step forecast errors for $Y_t(169)$ over the same time period. It is interesting to note that $Y_t(169)$ is not a descendant of $Y_t(167)$ and does not show the same large forecast errors at times 560 and 561.

In order to improve forecast performance of $Y_t(167)$ and its descendants, intervention was used for $Y_t(167)$ as follows. The observation $y_{560}(167)$ was unexpected and so was treated simply as an outlier (since it would not usually be known in advance that there would be a hold-up). Following the large forecast error for hour 560, a decision was made to intervene for $Y_{561}(167)$. As the road blockage clears and vehicles start moving, the delayed vehicles (from hour 560) 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 $Y_{560}(167)$. Thus $h_t(i)$ in the intervention distribution (3)

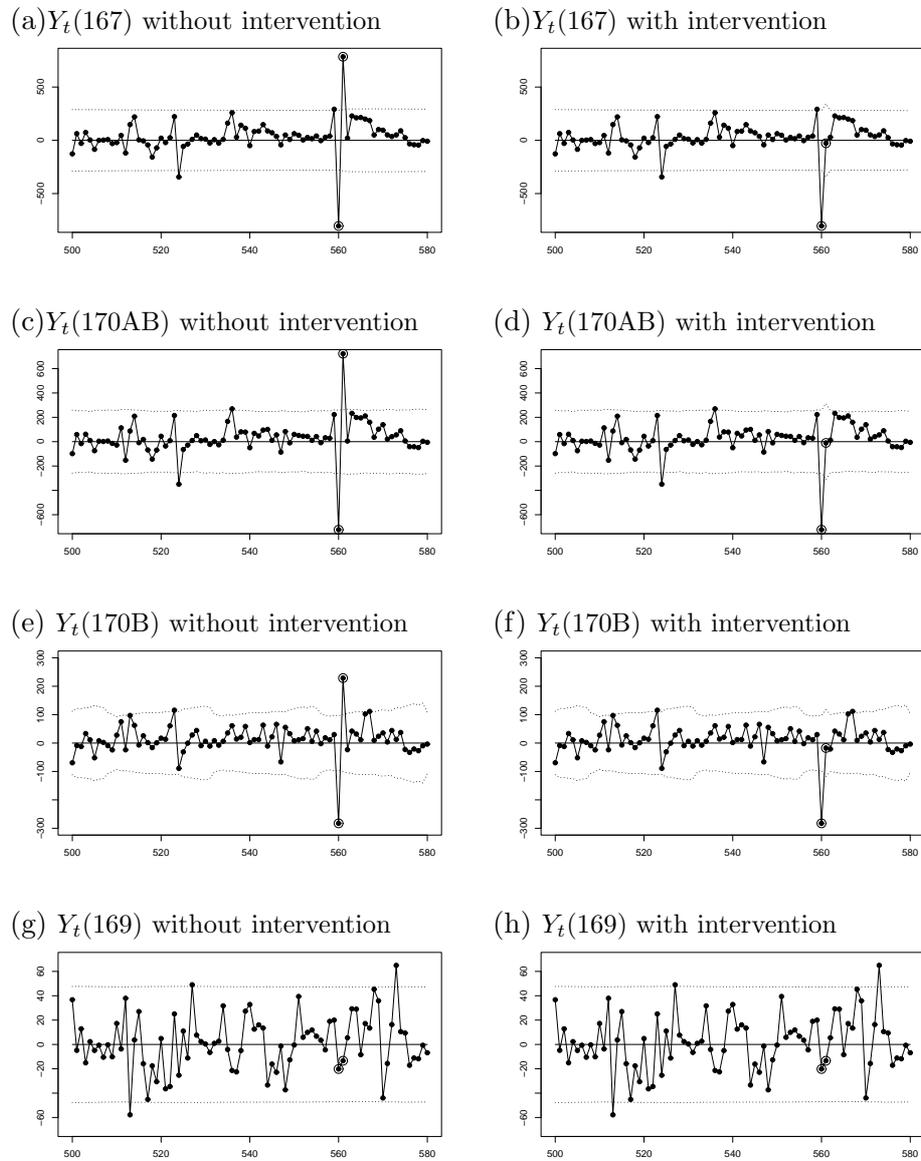


Figure 5: Plots of the one-step forecast errors (solid line) and ± 1.96 forecast standard deviations (dotted lines) obtained using the linear multiregression dynamic model between times 500 and 580 for $Y_t(167)$, one of its children $Y_t(170AB)$, one of its ‘grandchildren’ $Y_t(170B)$ and a non-descendant $Y_t(169)$. Plots on the left are the forecast errors obtained without intervention and those on the right are the forecast errors following intervention for $Y_{561}(167)$. The observations at times 560 and 561 are circled for each series in each plot.

was set to be $e_{560}(167)$. The value of $H_t(i)$ in the intervention distribution (3) was set to be 10,000. This value was fairly arbitrary: it was chosen to be large enough to reflect increased uncertainty and to let the model adapt quickly after intervention. (The MSE was in fact found to be fairly robust with respect to the choice of $H_t(i)$, provided that it is large enough, around 50% of $V_t(167)$ in this case.)

From Section 4.1, before \mathbf{Y}_t is observed, intervention for $Y_t(i)$ should affect the one-step forecast for $Y_t(i)$ and its descendants, but should not affect the one-step forecasts for any non-descendants. This is indeed the case. Intervention for $Y_{561}(167)$ not only improves the one-step forecast error for $Y_{561}(167)$ (Figure 5(a)(b)), but it also improves the one-step forecast error for its children (Figure 5(c)(d)), its ‘grandchildren’ (Figure 5(e)(f)) and indeed its ‘great grandchildren’ (not shown). On the other hand, it is clearly seen that the intervention for $Y_{561}(167)$ has no effect on the one-step forecast error of $Y_{561}(169)$ (Figure 5(g)(h)), a non-descendant of $Y_t(167)$.

In Section 4.1 it was shown that an intervention for $Y_t(i)$ will not affect any k -step ahead forecasts before \mathbf{Y}_t is observed. In contrast, in Section 4.2 it was shown that, after \mathbf{Y}_t is observed, the intervention *does* affect the k -step forecasts for $Y_t(i)$ and its descendants, but not the k -step forecasts of any non-descendants. To demonstrate the effect the intervention for $Y_{561}(167)$ has on the k -step ahead forecasts, Table 1 shows the k -step forecast means and standard deviations for some of the series at time $t = 585$ (24 hours later).

Columns 2–5 of Table 1 show 25-step forecast means and standard deviations made at time $t = 560$, after \mathbf{Y}_{560} has been observed and after any intervention for $Y_{561}(167)$ has been done, but before observing \mathbf{Y}_{561} . Columns 2–3 show the 25-step forecast means and standard deviations when there is no intervention and columns 4–5 show the results when there is intervention. As expected from Section 4.1, the k -step forecast distributions are exactly the same for all the series regardless of intervention.

Table 1: The 24- and 25-step forecast means and standard deviations for time 585, with and without intervention for $Y_{561}(167)$. The 25-step forecasts were made at time $t = 560$ after \mathbf{Y}_{560} has been observed and after any intervention has been done, but before observing \mathbf{Y}_{561} . The 24-step forecasts were made at time $t = 561$ after additionally observing \mathbf{Y}_{561} .

Series	Before \mathbf{Y}_{561} observed (25-step forecasts)				After \mathbf{Y}_{561} observed (24-step forecasts)			
	No intervention		Intervention		No intervention		Intervention	
	mean	sd	mean	sd	mean	sd	mean	sd
$Y_t(167)$	4,653	152	4,653	152	4,357	146	5,457	195
$Y_t(168)$	326	87	326	87	399	96	470	111
$Y_t(170AB)$	4,212	61	4,212	61	4,112	59	4,846	69
$Y_t(170B)$	1,431	76	1,431	76	1,372	73	1,617	85
$Y_t(161)$	1,618	302	1,618	302	1,925	337	2,060	359
$Y_t(169)$	514	26	514	26	507	24	507	24

Columns 6–9 of Table 1 show 24-step forecast means and standard deviations made at time $t = 561$, after any intervention for $Y_{561}(167)$ has been done, and after observing \mathbf{Y}_{561} . Columns 6–7 show the 24-step forecast means and standard deviations when there is no intervention and columns 8–9 show the results after intervening for $Y_{561}(167)$. As expected from Section 4.2, this time it is clearly evident that the intervention does affect k -step forecasts as the distributions are quite different for $Y_t(167)$ and all its descendants (even as far away as $Y_t(161)$). Notice, that also as expected, the k -step forecast moments for non-descendant $Y_t(169)$ are the same regardless of whether intervention took place or not.

From Figure 2, it can be seen that $Y(167)$ does not have any parents. As a consequence, the MDM models $Y_t(167)$ by any suitable univariate DLM. Thus the effects of intervention for $Y_{561}(167)$ within the DLM are also illustrated. In particular, it can be clearly seen that before \mathbf{Y}_{561} is observed, intervention for $Y_{561}(167)$ does not

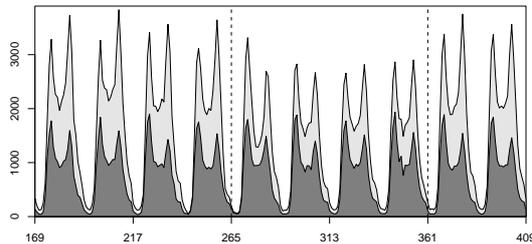


Figure 6: Plot of $Y_t(161)$ (in dark grey) and $Y_t(171)$ (on top in light grey) between times 169 and 409 (exactly 4 days before and 2 days after the period of reduced flows). Dotted vertical lines mark the start and end of the reduced flows.

affect the forecasts for $Y_{585}(167)$. However, after \mathbf{Y}_{561} is observed, intervention does affect the forecast distributions for $Y_{585}(167)$.

5.2 INTERVENTION FOR TWO STATE VECTORS

The second intervention considered involves the state vectors for $Y_t(161.171)$ and for $Y_t(161)$, denoted as $\boldsymbol{\theta}_t(161.171)$ and $\boldsymbol{\theta}_t(161)$, respectively. At time t , $Y_t(161.171)$ is the number of vehicles leaving the M25 to join the A2, $Y_t(171)$ is the number of these vehicles who travel eastbound on the A2, and $Y_t(161)$ is the number who travel westbound.

From time 265 until time 360 (a period of four days) there was a reduction in the number of vehicles leaving the M25 to join the A2 ($Y_t(161.171)$) and a reduction in the number of vehicles who travelled eastbound ($Y_t(171)$). Traffic flow westbound ($Y_t(161)$) remained, however, at the same level. This can be seen on the plot of the two series $Y_t(161)$ (in dark grey) and $Y_t(171)$ (on top in light grey) given in Figure 6. This sort of pattern in traffic flow is consistent with the expected consequences of roadworks eastbound on the A2, where drivers are forewarned of this on the M25 so that fewer vehicles leave the M25 to join the A2 eastbound.

Intervention is required in order to accommodate these reductions in flows. As

the reductions in flows are persistent over four days and not just one or two time points, intervention for the state vectors associated with $Y_{265}(161.171)$ and $Y_{265}(171)$ would be appropriate. However, since $Y_t(171)$ is a logical variable, the equivalent intervention for the parameters associated with $Y_{265}(161.171)$ and $Y_{265}(161)$ will be used instead. Thus, intervention for $\theta_{265}(161.171)$ and $\theta_{265}(161)$ is required.

The idea behind the intervention at time $t = 265$, is to scale the mean for $\theta_{265}(161.171)$ by some value α , for $0 < \alpha < 1$, and, because $Y_t(161)$ does not change at time 265, to scale the mean for $\theta_{265}(161)$ by $1/\alpha$. At time $t = 361$, traffic flows return to their pre-intervention levels, so a further intervention is required scaling the mean for $\theta_{361}(161.171)$ by $1/\alpha$ and the mean for $\theta_{361}(161)$ by α .

When using intervention for planned roadworks at time 265, ideally expert information should be used to estimate α . However, unfortunately no expert information was available for these data. It is possible that a prior could be placed on α , and α could then be estimated on-line from the data. For simplicity though, in order to illustrate the affect of the intervention *as if* good expert information were available, here the data between times 265 and 360 are used to estimate α . Using these data, the flows for $Y_t(171)$ roughly decrease by 1/3 from time 265. So, since the traffic flows for $Y_t(161)$ and $Y_t(171)$ are similar, this suggests an estimate for α of 5/6. Uncertainty concerning the intervention is incorporated into the model by increasing the variances for $\theta_{265}(161.171)$ and $\theta_{265}(161)$. These are regression parameters and, as such, are not expected to vary greatly over time and so, because of the scale of the system error variance, are only increased by 0.01 and 0.005, respectively. The resulting intervention distributions for time $t = 265$ are as follows.

$$\begin{aligned} & \theta_{265}(161.171) | (\theta_{264}(161.171), \text{intervention}) \\ & \sim N(5/6 \times G_{265}(161.171) \theta_{264}(161.171), W_{265}(161.171) + 0.01I), \end{aligned}$$

$$\begin{aligned} \boldsymbol{\theta}_{265}(161) | (\boldsymbol{\theta}_{264}(161), \text{intervention}) \\ \sim N(6/5 \times G_{265}(161)\boldsymbol{\theta}_{264}(161), W_{265}(161) + 0.005I), \end{aligned}$$

where $G_{265}(161.171)$, $G_{265}(161)$, $W_{265}(161.171)$ and $W_{265}(161)$ are the pre-intervention matrices as defined by the system equation, and I is the identity matrix. At time $t = 361$, the mean for $\boldsymbol{\theta}_{361}(161.171)$ is instead scaled by $6/5$, and that for $\boldsymbol{\theta}_{361}(161)$ by $5/6$. The system error variances are again increased by 0.01 and 0.005 , respectively.

In Section 4.1, it was shown that before $Y_t(i)$ is observed, an intervention for $\boldsymbol{\theta}_t(i)$ will affect the one-step forecasts for $Y_t(i)$ and its descendants, but not affect any one-step forecasts of non-descendants. From Figure 7 it is clearly seen how the interventions for $\boldsymbol{\theta}_t(161.171)$ and $\boldsymbol{\theta}_t(161)$ have affected the one step forecasts for $Y_t(161.171)$ (plot (a)) and descendant $Y_t(171)$ (plot (c)). The one-step forecasts for $Y_t(161)$ are also affected, but to a much lesser extent (see plot (b)). This is because $Y_t(161)$ did not exhibit any change at the intervention period and so a change in its forecasts and forecast errors were not actually required. As expected, these interventions have no effect on the one-step forecasts for non-descendants.

Figure 7(d) shows a plot of the one-step prior mean for the first element of $\boldsymbol{\theta}_t(161)$. This element of $\boldsymbol{\theta}_t(161)$ is the parameter for 00:00–01:00 each day and is only updated every 24 hours. The effects on $\boldsymbol{\theta}_t(161)$ of the interventions at times 265 and 361 can be clearly seen: at time 265 the prior mean steps up to a new level which is sustained until the intervention at time 361 returns the mean to its former level.

In contrast to intervention for $Y_t(i)$ which does not affect any k -step forecasts before \mathbf{Y}_t is observed, but does after \mathbf{Y}_t is observed, intervention for $\boldsymbol{\theta}_t(i)$ affects the k -step forecasts of $Y_t(i)$ and its descendant both before *and* after \mathbf{Y}_t is observed. This is demonstrated in Table 2 which shows the k -step forecast means and standard deviations for $Y_t(161.171)$ and $Y_t(161)$ at time $t = 289$ (24 hours after the intervention

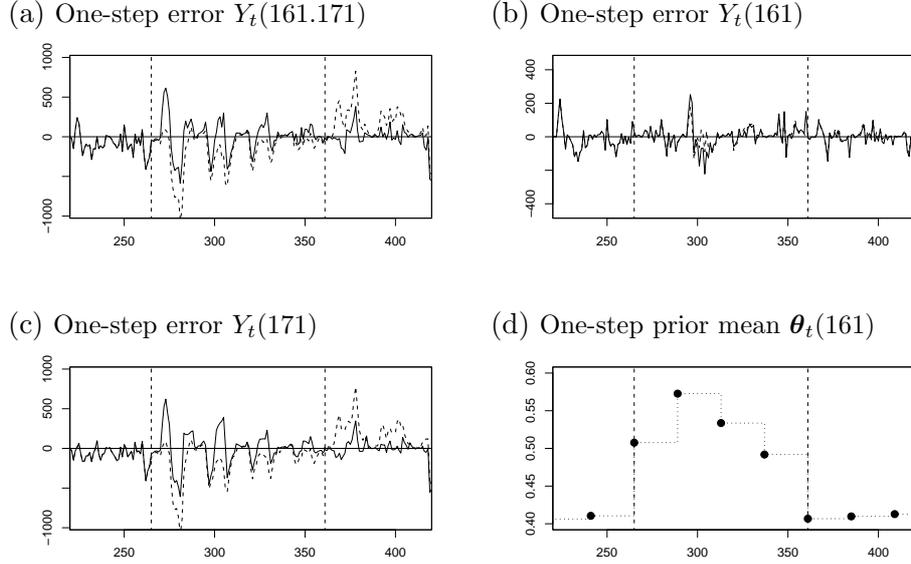


Figure 7: (a)–(c) One-step forecast errors both with intervention for $\theta_t(161.171)$ and $\theta_t(161)$ (solid line) and without intervention (dotted line) for $Y_t(161.171)$, $Y_t(161)$ and their descendant $Y_t(171)$. (d) One-step prior mean for the first element of $\theta_t(161)$. On all plots, dotted vertical lines mark times 265 and 361, when intervention took place.

at time 265).

Columns 2–5 of Table 2 show the 25-step forecast means and standard deviations for time 289 made at time 264 after Y_{264} has been observed and after any interventions for $\theta_{265}(161.171)$ and $\theta_{265}(161)$ have been done, but before \mathbf{Y}_{265} is observed. Columns 6–9 of Table 2 show the 24-step forecast means and standard deviations for time 289 made at time 265 after any interventions for $\theta_{265}(161.171)$ and $\theta_{265}(161)$ have been done, and after \mathbf{Y}_{265} is observed. Columns 2–3 and 6–7 show the k -step forecast means and standard deviations when there is no intervention, and columns 4–5 and 8–9 show the results when there is intervention. From this table, the effects of the interventions on k -step forecasts both before \mathbf{Y}_{265} is observed and after \mathbf{Y}_{265} is observed are clearly

Table 2: The 24- and 25-step forecast means and standard deviations for time 289, with and without intervention for $\theta_{265}(161.171)$ and $\theta_{265}(161)$. The 25-step forecasts were made at time $t = 264$ after \mathbf{Y}_{264} has been observed and after any intervention has been done, but before observing \mathbf{Y}_{265} . The 24-step forecasts were made at time $t = 265$ after additionally observing \mathbf{Y}_{265} .

Series	Before \mathbf{Y}_{265} observed (25-step forecasts)				After \mathbf{Y}_{265} observed (24-step forecasts)			
	No intervention		Intervention		No intervention		Intervention	
	mean	sd	mean	sd	mean	sd	mean	sd
$Y_t(161.171)$	527	593	415	740	601	454	499	737
$Y_t(161)$	318	84	447	147	294	74	372	133

seen. Notice that the interventions will also affect the k -step forecasts both before and after \mathbf{Y}_{265} is observed for the descendant $Y_t(171)$ as this is simply a logical function of its parents.

6. IDENTIFICATION OF CAUSAL RELATIONSHIPS BETWEEN TIME SERIES

Because the MDM is defined to preserve the conditional independence structure related to causality across a time series over time, the BN for the time series at time t represents contemporaneous causal relationships between the component series. The forecast performance of an MDM is therefore informative about these assumed contemporaneous causal relationships.

As mentioned in Section 3, two BNs can represent the same conditional independence statements, but have quite different conditional independence structures related to causality, and hence quite different MDMs. Suppose that there are several possible MDMs for \mathbf{Y}_t , each of which has the same conditional independence structure, but each with a different conditional independence structure related to causality. Following the ideas of multiprocess DLMS (Harrison and Stevens, 1976),

each of these competing MDMs can be modelled simultaneously using a *multiprocess MDM*. In a multiprocess MDM, each competing MDM has an associated probability of being the ‘correct’ model, that is, the corresponding BN represents the ‘correct’ contemporaneous causal relationships. As data are observed, these probabilities are updated, providing on-line assessments of causal relationships between time series.

As is known from BN theory, without intervention it can be extremely difficult to identify which causal relationships are ‘correct’ from a set of competing models with the same conditional independence structure. As a consequence, at times when there is no intervention, there should be little difference between the various model probabilities in a multiprocess MDM. However, this is not the case when intervention is used. As it was shown in Section 4, intervention for $Y_t(i)$ (or $\theta_t(i)$) will affect the distributions of its descendants. Therefore any MDM which continues to provide good forecasts for the descendants of $Y_t(i)$ is likely to be representing the ‘correct’ causal structure, and the associated probability for that model will be larger than the others.

Explicitly, suppose there are m competing BNs for \mathbf{Y}_t with associated MDMs M_1, \dots, M_m . Let $p_{t-1}(j)$ denote the prior probability that model M_j is ‘correct’ given D_{t-1} , for $j = 1, \dots, m$. The likelihood for observing $\mathbf{Y}_t = \mathbf{y}_t$ under model M_j is the observed value of the one-step forecast distribution $f(\mathbf{y}_t | D_{t-1}, M_j)$, which in the MDM is the product of the observed values of individual univariate conditionals. The posterior probability that M_j is ‘correct’ is thus given by

$$p_t(j) \propto p_{t-1}(j) \prod_{i=1}^n f(y_t(i) | pa(y_t(i)), D_{t-1}, M_j).$$

The model probabilities are therefore updated over time and any decision at time t as to which causal structure is ‘correct’ is based on the posterior probabilities $p_t(1), \dots, p_t(m)$.

The multiprocess MDM will be illustrated for the London network. Suppose for simplicity that there are only two competing models:

- Model M_1 : The MDM already used in Section 3 with BN given in Figure 2.
- Model M_2 : An MDM using a BN which is the same as in Figure 2 except that the arc between $Y_t(167)$ and $Y_t(170AB)$ is reversed.

Notice that the influence diagrams for both models represent the same conditional independence structure, but represent different conditional independences related to causality. As there are only two possible models here, the ratio of the posterior model probabilities, $p_t(1)/p_t(2)$, is of particular interest.

In order to illustrate how the multiprocess MDM can identify causal relationships, the ratio of the posterior model probabilities is calculated at two separate times — at time 540 when no intervention is required, and at time 561 following the intervention for $Y_{561}(167)$ as detailed in Section 5.1. Time $t = 540$ was chosen fairly randomly to illustrate the value of the ratio of posterior model probabilities when the model is performing well and no intervention is required. When using model M_1 , it was shown in Section 5.1 how intervention for $Y_{561}(167)$ has a dramatic effect on the one-step forecasts and resulting forecast errors of $Y_{561}(167)$ and its descendants (see Figure 5). Because $Y_t(167)$ has so many descendants in the influence diagram for M_1 , the effects of the intervention for $Y_{561}(167)$ are seen across almost the entire network. In model M_2 , however, $Y_t(167)$ only has the single descendant $Y_t(168)$ and so intervention for $Y_{561}(167)$ will have only a limited affect on the one-step forecasts over the network as a whole. In particular, intervention for $Y_{561}(167)$ will have no affect in reducing the large one-step forecast errors for $Y_{561}(170AB)$ and its descendants. So, although the same intervention for $Y_{561}(167)$ should be used for M_1 and M_2 (after all, the same change is observed for $Y_{561}(167)$), the *effects* of the intervention are quite different for

the two models and the observed likelihoods and hence the posterior probabilities, $p_t(1)$ and $p_t(2)$, will reflect this.

The initial two weeks of data were used to estimate priors for the parameters for both models. The models were then run (separately) and updated sequentially as usual up to time $t = 539$. For fairness, at time 539 both models are assigned the same prior probability so that $p_{539}(1) = p_{539}(2) = 0.5$. After observing $\mathbf{Y}_{540} = \mathbf{y}_{540}$, the ratio of posterior probabilities, $p_{540}(1)/p_{540}(2)$, is calculated to be 1.06. This is approximately equal to 1 as expected, illustrating how it is difficult to identify causality when no intervention is used. The models were then again run (separately) and updated sequentially as usual up to time $t = 560$. At time $t = 561$, intervention was performed for $Y_{561}(167)$ using the same method as presented in Section 5.1 so that the distribution for $Y_{561}(167)|(\boldsymbol{\theta}_{561}(167), M_j)$ was adjusted at intervention by adding $e_{560}(167)$ to the mean and adding 10,000 to the variance. For fairness, at time 560 again both models are assigned the same prior probability, so that $p_{560}(1) = p_{560}(2) = 0.5$. After observing $\mathbf{Y}_{561} = \mathbf{y}_{561}$, the ratio of posterior probabilities, $p_{561}(1)/p_{561}(2)$, is this time calculated to be 1.7×10^{30} providing overwhelming support for model M_1 as opposed to model M_2 . Notice how the ratio of posterior probabilities is much larger at the time of intervention ($t = 561$), than at the time of no intervention ($t = 540$), thus illustrating how much easier it is to identify causal relationships at the time of an intervention than when there is no intervention. For this particular application, the context of the problem heuristically suggests that, under normal traffic conditions, $Y_t(167)$ is causal for $Y_t(170AB)$, since traffic flows from site 167 to sites 170A and 170B. It is therefore reassuring that the multiprocess MDM so clearly confirms this when intervention is used.

As mentioned in Section 1, the causal relationships between traffic flow series can change temporarily in response to queueing traffic. Queueing traffic in itself can

cause a change in traffic flows which can require intervention to maintain forecast performance at the source of the problem. The multiprocess MDM is then able to identify the ‘correct’ contemporaneous causal structure so that the ‘correct’ MDM can be used to maintain forecast performance across the entire multivariate series. This is, in fact, an important modelling issue when forecasting traffic flows in practice. To accommodate changes in causal relationships, multiprocess MDMs can be used at each time point to provide an on-line assessment of the most likely causal relationships at each time t . Further investigation of this will be the focus of future research.

7. CONCLUDING REMARKS

This paper has shown how intervention in the MDM can be an extremely useful tool for forecasting traffic flows, enabling forecast performance to be maintained despite changes in the system. It is simple to implement since intervention is only required for the variable(s) at the root of the problem: the effects of any intervention is automatically passed on to other series also affected by the problem. Intervention can also be used to identify contemporaneous causal relationships between the series, going beyond the previous research using intervention for identifying lagged causal relationships between series.

REFERENCES

- Brillinger, D.R. (1996), “Remarks concerning graphical models for time series and point processes,” *Revista de Econometria*, **16**, 1–23.
- 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. University Press Oxford, pp. 115–137.
- Dawid, A.P. (1979), “Conditional independence in statistical theory” (with discussion) *Journal of the Royal Statistical Society, Series B*, **41**, 1–31.

- (2002), “Influence diagrams for causal modelling and inference,” *International Statistical Review*, **70**, 161–189.
- Dawid, A.P., and Didelez, V. (2005), “Graphical models and sequential decisions,” Technical Report 262, University College London, UK
- Didelez, V. (2003), “Comments on graphical models for stochastic processes,” in *Highly Structured Stochastic Systems*, (eds.) P. Green, N. Hjort and S. Richardson. University Press Oxford.
- Eichler, M., and Didelez, V. (2007), “Causal reasoning in graphical time series models,” in *Proceedings of the 23rd Annual Conference on Uncertainty in Artificial Intelligence*, (eds.) Ronald Parr and Linda van der Gaag, AUAI Press, Arlington, Virginia, pp 109–116.
- Farrow, M. (2003), “Practical building of subjective covariance structures for large complicated systems,” *The Statistician*, **52**, 553–573.
- Farrow M., Goldstein M., and Spiropoulos T. (1997), “Developing a Bayes linear decision support system for a brewery,” in *The Practice of Bayesian Analysis*, (eds.) S. French and J. Q. Smith, Arnold: London, pp. 71-106.
- Fosen, J., Ferkingstad, E., Borgan, Ø., and AalenFosen, OO. (2006), “Dynamic path analysis a new approach to analyzing time-dependent covariates,” *Lifetime Data Analysis*, **12**, 143–167.
- Goldstein, M., Farrow, M. and Spiropoulos, T. (1993), “Prediction under the influence: Bayes linear influence diagrams for prediction in a large brewery,” *The Statistician*, **42**, 445-459.
- Guo, W. and Brown, M.B. (2001), “Cross-related structural time series models,” *Statistica Sinica*, **11**, 961–979.
- Harrison, P.J., and Stevens, C. (1976), “Bayesian forecasting” (with discussion), *Journal of the Royal Statistical Society, Series B*, **38**, 205–247.
- Howard, R.A., and Matheson, J.E. (1984), “Influence diagrams,” in *Readings in the Principles and Applications of Decision Analysis*, (eds.) R.A. Howard and J.E. Matheson. Menlo Park, CA: Strategic Decisions Group.
- Kamarianakis, Y., and Prastacos, P. (2005), “Space-time modeling of traffic flow,” *Computers and Geosciences*, **31**, 119-133.
- Kjærulff, U. (1995), “dHugin: a computational system for dynamic time-sliced Bayesian networks,” *International Journal of Forecasting*, **11**, 89–111.
- Lauritzen, S.L. (2000), “Causal inference from graphical models,” in *Complex Stochastic Systems*, (eds.) O.E. Barndorff-Nielsen, D.R. Cox and C. Klüppelberg, Chapter 2, London: CRC Press, pp 63–107.
- Lauritzen, S.L., and Richardson, T.S. (2002), “Chain graph models and their causal interpretation” (with discussion), *Journal of the Royal Statistical Society, Series B*,

64, 321–361.

Oliver, R.M., and Smith, J.Q. (1990), *Influence Diagrams, Belief Nets and Decision Analysis*. Chichester, UK: John Wiley and Sons.

Pearl, J. (1995), “Causal diagrams for empirical research (with Discussion),” *Biometrika*, **82**, 669–710.

——— (2000), *Causality*. Cambridge University Press.

Queen, C.M. (1994), “Using the multiregression dynamic model to forecast brand sales in a competitive product market,” *The Statistician*, **43**, 87–98.

——— (1997), “Model elicitation in competitive markets,” in *The Practice of Bayesian Analysis*, (eds.) S. French and J. Q. Smith, London: Arnold, pp. 229–243.

Queen, C.M., Smith, J.Q., and James, D.M. (1994), “Bayesian forecasts in markets with overlapping structures,” *International Journal of Forecasting*, **10**, 209–233.

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.

——— (2008), “Forecast covariances in the linear multiregression dynamic model,” *Journal of Forecasting*, **27**, 175–191.

Shachter, R.D. (1986), “Evaluating influence diagrams,” *Operations Research*, **34**, 871–882.

——— (1988), “Probabilistic Inference and Influence Diagrams,” *Operations Research*, **36**, 589–604.

Smith, J.Q., and Figueroa, L.J. (2007), “A Causal Algebra for Dynamic Flow Networks,” in *Advances in Probabilistic Graphical Models*, Springer, pp 39–54.

Spirtes, P., Glymour, C., and Scheines, R. (2000), *Causation, Prediction and Search*. (2nd ed.) New York, N.Y.: MIT Press.

Stathopoulos, A., and Karlaftis, G.M. (2003), “A multivariate state space approach for urban traffic flow modeling and prediction,” *Transportation Research Part C*, **11**, 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**, 124–132.

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

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

——— (1989), “Subjective intervention in formal models,” *Journal of Forecasting*, **8**, 33–53.

——— (1997), *Bayesian Forecasting and Dynamic Models* (2nd ed.) Springer-Verlag, New York.

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

Whittaker, J., Garside, S., and Lindveld, K. (1997), “Tracking and predicting a network traffic process,” *International Journal of Forecasting*, **13**, 51-61.