Bayesian Analysis (2017) **12**, Number 2, pp. 491–509

Dynamic Chain Graph Models for Time Series Network Data

Osvaldo Anacleto* and Catriona Queen†

Abstract. This paper introduces a new class of Bayesian dynamic models for inference and forecasting in high-dimensional time series observed on networks. The new model, called the dynamic chain graph model, is suitable for multivariate time series which exhibit symmetries within subsets of series and a causal drive mechanism between these subsets. The model can accommodate high-dimensional, non-linear and non-normal time series and enables local and parallel computation by decomposing the multivariate problem into separate, simpler sub-problems of lower dimensions. The advantages of the new model are illustrated by forecasting traffic network flows and also modelling gene expression data from transcriptional networks.

Keywords: chain graph, multiregression dynamic model, network traffic flow forecasting, gene expression networks, network data, time series.

1 Introduction

Multivariate time series are often observed on a network or graph. Despite the ever-increasing research on network modelling, statistical dynamic modelling on networks has not been explored much so far (Kolaczyk, 2009, Chapter 9). This paper proposes a new class of multivariate Bayesian dynamic models (West and Harrison, 1997) for time series networks, called dynamic chain graph models.

Although Bayesian dynamic models have long been successfully applied in many different application areas (see, for example Velozo et al., 2014; Xiao et al., 2015; Nascimento et al., 2015; Quirós et al., 2015), few of these models can accommodate complex high-dimensional series, while not being too demanding computationally. Furthermore, established Bayesian multivariate dynamic models such as the matrix normal dynamic linear model (Quintana and West, 1987) and its Gaussian graphical model extension (Carvalho and West, 2007) are only suitable for multivariate series whose component univariate series are similar and share a common structure.

The multiregression dynamic model (MDM) (Queen and Smith, 1993) is an alternative model that does not require all the component univariate time series to have a common structure. This model assumes a conditional independence and causal structure among time series components at each time step, as expressed by a directed acyclic graph (Lauritzen, 1996). This graph is then used to decompose the n-dimensional model into n separate conditional models, each of which is a univariate Bayesian dynamic model. As such, the MDM can accommodate arbitrarily high-dimensional structures, while enabling local and parallel computation.

*The Roslin Institute and R(D)SVS, University of Edinburgh, UK, osvaldo.anacleto@roslin.ed.ac.uk
†Department of Mathematics and Statistics, The Open University, UK, catriona.queen@open.ac.uk

 DOI: 10.1214/16-BA1010

The MDM has been extensively used in a variety of applications: brand sales forecasting (Queen, 1994), traffic flow forecasting (Queen et al., 2007; Queen and Albers, 2009; Anacleto et al., 2013a,b), functional magnetic resonance imaging (Costa et al., 2015; Oates et al., 2015a,b), financial portfolio analysis (Zhao et al., 2015) and electricity demand forecasting (Zhao, 2015). However, the directed acyclic graph used in the MDM may be too restrictive: only directional associations between individual series can be accommodated, and no symmetric associations are allowed.

Queen and Smith (1992) developed a Bayesian dynamic model called the dynamic graphical model, in which a chain graph (Wermuth and Lauritzen, 1990) represents a multivariate time series $\boldsymbol{Y}_t$ at each time t, with both directional and symmetric associations between the individual time series. In that model, $\boldsymbol{Y}_t$ is partitioned into ordered chain components, which are *complete* undirected graphs and are connected by directed edges. However, if two chain components are connected in this model, then there must be a directed edge from *every* variable in the first chain component to *every* variable in the second. Also, the dynamic graphical model cannot accommodate chain components as sparse undirected graphs, which simplify estimation of the observation covariance matrix (Carvalho and West, 2007; Prado and West, 2010).

This paper presents a Bayesian dynamic model based on a chain graph which does not rely on the stringent assumptions required for the dynamic graphical model developed in Queen and Smith (1992). This new model not only enables sparsity on chain components, but it also allows for unrestricted directed edges between them, thus accommodating more complex dependence patterns among multivariate time series components. It is shown that, like the MDM, computation in this new model is simplified and parallelizable since the multivariate problem is decomposed into separate, simpler sub-problems of lower dimensions, although, unlike the MDM, not all of these will be univariate.

The proposed model is illustrated using two application areas: road traffic flow forecasting and gene expression modelling. Both applications are examples where the MDM cannot capture all types of dependencies among the time series components.

2 Basic graph theory concepts and notation

Before defining the new model in the next section, some notation and terminology is introduced. The terms *network* and *graph* are interchangeably used throughout this paper. A chain graph is defined as a pair (V, E), where V is a finite set of *nodes* (or *vertices*) and E is a subset of ordered pairs of nodes, called *edges*.

Figure 1 shows an example of a chain graph for a 7-dimensional vector $\boldsymbol{X}$ with chain components $\{X_1, X_2\}$, $\{X_3, X_4\}$, $\{X_5, X_6\}$ and $\{X_7\}$. If there is a directed edge from X_i to X_j, then X_i is a parent of X_j while X_j is a child of X_i, and if X_k and X_l are connected by an undirected edge, then they are neighbours. For set of variables $\boldsymbol{A}$, the parents, children and neighbours of $\boldsymbol{A}$ are denoted, respectively, $\mathrm{pa}(\boldsymbol{A})$, $\mathrm{ch}(\boldsymbol{A})$ and $\mathrm{ne}(\boldsymbol{A})$. The boundary of set $\boldsymbol{A}$ is $\mathrm{bd}(\boldsymbol{A}) = \mathrm{pa}(\boldsymbol{A}) \cup \mathrm{ne}(\boldsymbol{A})$. A path of length n from α to β is any sequence $(\alpha_0 = \alpha, \ldots, \alpha_n = \beta)$ of distinct nodes such that $(\alpha_{i-1}, \alpha_i) \in E$ for all $i = 1, \ldots, n$. If there is a path from X_i to X_j, but no path from X_j to X_i, then

X_j is a descendant of X_i: the descendants of the set of variables $\boldsymbol{A}$ is denoted $\mathrm{de}(\boldsymbol{A})$. The non-descendants of $\boldsymbol{A}$ are $\mathrm{nd}(\boldsymbol{A}) = \boldsymbol{X} \backslash (\mathrm{de}(\boldsymbol{A}) \cup \boldsymbol{A})$. Then, for each X_i, $i = 1, \ldots, n$ (Lauritzen, 1996),

$$X_i \perp\!\!\!\perp \{\mathrm{nd}(X_i) \backslash \mathrm{bd}(X_i)\} \mid \mathrm{bd}(X_i). \tag{1}$$

For example, in Figure 1, $\mathrm{pa}(X_6) = \{X_3, X_4\}$, $\mathrm{ch}(X_6) = \{X_7\}$, $\mathrm{ne}(X_6) = \{X_5\}$, $\mathrm{de}(X_6) = \{X_7\}$, $\mathrm{nd}(X_6) = \{X_1, X_2, X_3, X_4, X_5\}$ and so $X_6 \perp\!\!\!\perp \{X_1, X_2)\} \mid \{X_3, X_4, X_5\}$.

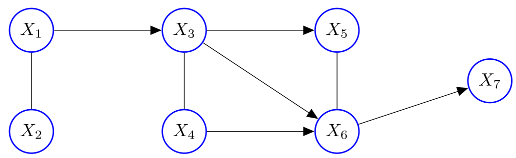

Figure 1: Example of a chain graph.

3 The dynamic chain graph model

The *dynamic chain graph model* (DCGM) is defined for time series which can be represented over time by a dynamic chain graph. The DCGM makes explicit use of the dynamic chain graph structure of the series to simplify computation. Before defining the DCGM in Section 3.2, a dynamic chain graph representation for time series, together with the associated conditional independence relationships between the component time series, is presented in Section 3.1.

Firstly, some notation is required. Let $\{\boldsymbol{Y}_t\}_{t \in \mathbb{N}}$ be an n-dimensional time series and suppose that $\boldsymbol{Y}_t$ is partitioned into N vector time series of dimensions $r_1, \ldots, r_N$ with $\sum_{i=1}^{N} r_i = n$, so that $\boldsymbol{Y}_t^{\mathrm{T}} = (\boldsymbol{Y}_t(1)^{\mathrm{T}}, \ldots, \boldsymbol{Y}_t(N)^{\mathrm{T}})$ where, for $i = 1, \ldots, N$, $\boldsymbol{Y}_t(i)^{\mathrm{T}} = (Y_{t1}(i), \ldots, Y_{tr_i}(i))$. Let $\boldsymbol{Y}^t = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_t)^{\mathrm{T}}$, $\boldsymbol{Y}^t(i) = (\boldsymbol{Y}_1(i), \ldots, \boldsymbol{Y}_t(i))^{\mathrm{T}}$ and $\boldsymbol{Y}_j^t(i) = (\boldsymbol{Y}_{1j}(i), \ldots, \boldsymbol{Y}_{tj}(i))^{\mathrm{T}}$, and let $\boldsymbol{y}_t$, $\boldsymbol{y}_t(i)$ and $y_{tj}(i)$ be the realizations of $\boldsymbol{Y}_t$, $\boldsymbol{Y}_t(i)$ and $Y_{tj}(i)$, respectively.

3.1 Representing multivariate time series with dynamic chain graphs

The dependence structure of a dynamic chain graph can be divided into a set of *intra time-slice dependencies*, which represent associations among time series components in a fixed time $t \in \mathbb{N}$, and a set of *inter time-slice dependencies*, which represent associations among time series components across time.

Intra time-slice dependencies

Suppose that, at each time $t \in \mathbb{N}$, there is an association structure between all individual time series within each vector series $\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(N)$ so that, for each $i = 1, \ldots, N$,

$Y_{t1}(i), \ldots, Y_{tr_i}(i)$ form an undirected chain component in a chain graph. Suppose also that for each time $t \in \mathbb{N}$ there is a causal drive through the system and a conditional independence structure (which is the same structure for each time $t \in \mathbb{N}$) so that $\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(N)$ are ordered chain components in the chain graph with $\mathrm{pa}(Y_{tj}(i)) \subseteq \{\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(i-1)\}$. For example, consider the time series $\boldsymbol{Y}_t$ represented at time t by the chain graph given in Figure 2. Here there are three chain components (so that $N = 3$): $\{Y_{t1}(1), Y_{t2}(1)\}$, $\{Y_{t1}(2), Y_{t2}(2), Y_{t3}(2)\}$ and $\{Y_{t1}(3), Y_{t2}(3)\}$, with $r_1 = 2$, $r_2 = 3$ and $r_3 = 2$. Thus $\boldsymbol{Y}_t^{\mathrm{T}} = (\boldsymbol{Y}_t(1)^{\mathrm{T}}, \boldsymbol{Y}_t(2)^{\mathrm{T}}, \boldsymbol{Y}_t(3)^{\mathrm{T}})$ where $\boldsymbol{Y}_t(1) = (Y_{t1}(1), Y_{t2}(1))^{\mathrm{T}}$, $\boldsymbol{Y}_t(2) = (Y_{t1}(2), Y_{t2}(2), Y_{t3}(2))^{\mathrm{T}}$ and $\boldsymbol{Y}_t(3) = (Y_{t1}(3), Y_{t2}(3))^{\mathrm{T}}$.

Inferring intra-time slice dependencies from data is an important research area. This problem is not the focus here, but it is briefly mentioned in Section 6.

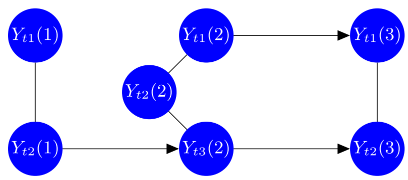

Figure 2: Example of intra-time slice dependencies in a chain graph representation of a 7-dimensional time series at time t.

Inter time-slice dependencies

The chain graphs defined above for the time series vectors $\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(N)$ at each $t \in \mathbb{N}$ can be linked together by assuming inter time-slice dependencies, so that a *dynamic chain graph* for the time series $\{\boldsymbol{Y}_t\}_{t \in \mathbb{N}}$ is obtained. For a time series up to time t, represented by $\boldsymbol{Y}^t(1), \ldots, \boldsymbol{Y}^t(N)$, the inter time-slice dependencies of a time series component $Y_{tj}(i)$, $i = 1, \ldots, N$, $j = 1, \ldots, r_i$, are represented by its parents at previous time steps, which are allowed to be from $\{\boldsymbol{Y}^k(1), \ldots, \boldsymbol{Y}^k(i)\}$, $k = 1, \ldots, t-1$. Together with the contemporaneous parents defined in the previous section, in a dynamic chain graph we then have

$$\mathrm{pa}(Y_{tj}(i)) \subseteq \{\boldsymbol{Y}^t(1), \ldots, \boldsymbol{Y}^t(i-1), \boldsymbol{Y}^{t-1}(i)\}. \tag{2}$$

The parent set $\mathrm{pa}(Y_{tj}(i))$ is assumed to be the same at each time $t \in \mathbb{N}$.

Figure 3 shows a dynamic chain graph example based on the chain graph of Figure 2 and considering all time points up to time t. Notice that if $Y_{tj}(l)$ is a parent of $Y_{tk}(m)$, it is not necessary that $\boldsymbol{Y}_j^{t-1}(l)$ is also a parent of $Y_{tk}(m)$, nor that $\boldsymbol{Y}_j^{t-1}(l)$ is a parent of $Y_{tj}(l)$.

Dynamic chain graph conditional independence structure

The dynamic chain graph defines a conditional independence structure among contemporaneous variables as stated in the following theorem.

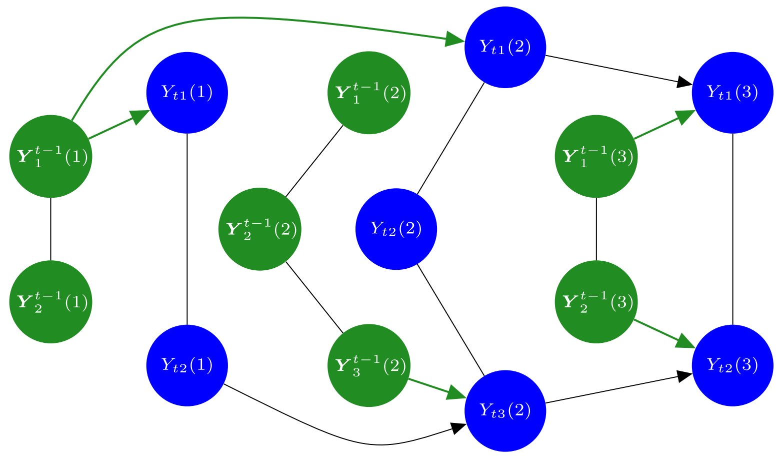

Figure 3: Example of a dynamic chain graph for the 7-dimensional time series with intra-time slice dependencies (black edges) defined in Figure 2. Inter time-slice dependencies are represented by green edges. The figure shows inter time-slice dependences for the series at a arbitrary time such that these dependencies are the same for any time step $t > 1$. Undirected edges between processes represent intra-time slice relationships. For example, the undirected edge connecting $\boldsymbol{Y}_1^{t-1}(1)$ and $\boldsymbol{Y}_2^{t-1}(1)$ means that there is an undirected edge between $Y_{i1}(1)$ and $Y_{i2}(1)$ for $i = 1, \ldots, t-1$.

Theorem 1. *Let $\{\boldsymbol{Y}_t\}_{t\in\mathbb{N}}$ be represented by a dynamic chain graph where $\boldsymbol{Y}_t$ can be decomposed, for each time $t \in \mathbb{N}$, into a set of ordered chain components $\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(N)$, with such ordering remaining constant over time. Then, the following conditional independence statements hold for each time $t \in \mathbb{N}$:*

$$\boldsymbol{Y}_t(i) \perp\!\!\!\perp [\{\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(i-1)\} \setminus pa(\boldsymbol{Y}_t(i))] \mid pa(\boldsymbol{Y}_t(i)). \tag{3}$$

Proof. For notational convenience, define

$$\begin{array}{rcll} \boldsymbol{X}_t(i)^{\mathrm{T}} & = & (\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(i-1)), & i = 2, \ldots, N, \\ \boldsymbol{Z}_t(i)^{\mathrm{T}} & = & (\boldsymbol{Y}_t(i+1), \ldots, \boldsymbol{Y}_t(N)), & i = 1, \ldots, N-1, \end{array}$$

where for $i = 1$, $\boldsymbol{X}_t(i)$ is $\emptyset$, as is $\boldsymbol{Z}_t(i)$ for $i = N$.

For the dynamic chain graph representing $\boldsymbol{Y}^t(1), \ldots, \boldsymbol{Y}^t(N)$ as described above, conditional independence statements (1) imply that for each $t \in \mathbb{N}$, $i = 1, \ldots, N$ and $j = 1, \ldots, r_i$,

$$Y_{tj}(i) \perp\!\!\!\perp \{\mathrm{nd}(Y_{tj}(i)) \setminus \mathrm{bd}(Y_{tj}(i))\} \mid \mathrm{bd}(Y_{tj}(i)). \tag{4}$$

Now, the set $\mathrm{nd}(Y_{tj}(i))$ consists of $\{\boldsymbol{Y}_t(i) \setminus Y_{tj}(i)\}$ together with $\boldsymbol{X}^t(i)$, $\boldsymbol{Y}^{t-1}(i)$ and $\boldsymbol{Z}^{t-1}(i)$, while $\mathrm{bd}(Y_{tj}(i)) = \mathrm{pa}(Y_{tj}(i)) \cup \mathrm{ne}(Y_{tj}(i))$, where $\mathrm{pa}(Y_{tj}(i)) \subseteq \left\{\boldsymbol{X}^t(i), \boldsymbol{Y}^{t-1}(i)\right\}$ from (2), and $\mathrm{ne}(Y_{tj}(i)) \subseteq \{\boldsymbol{Y}_t(i) \setminus Y_{tj}(i)\}$. Therefore, statement (4) becomes

$$Y_{tj}(i) \perp\!\!\!\perp \left[\left\{\boldsymbol{X}^t(i), \{\boldsymbol{Y}^t(i) \setminus Y_{tj}(i)\}, \boldsymbol{Z}^{t-1}(i)\right\} \setminus \mathrm{bd}(Y_{tj}(i))\right] \mid \mathrm{bd}(Y_{tj}(i))$$

In particular, this means that

$$Y_{tj}(i) \perp\!\!\!\perp [\{\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(i-1)\} \setminus \mathrm{bd}(Y_{tj}(i))] \mid \mathrm{bd}(Y_{tj}(i)).$$

But since $\mathrm{ne}(Y_{tj}(i)) \subseteq \{\boldsymbol{Y}_t(i) \setminus Y_{tj}(i)\}$ for each $j = 1, \ldots, r_i$, then collectively $\mathrm{bd}(\boldsymbol{Y}_t(i)) \equiv \mathrm{pa}(\boldsymbol{Y}_t(i))$, so that the conditional independence statements in (3) hold. □

As will be seen later, Theorem 1 is important for the proposed model as it plays a key role in breaking the multivariate problem into N smaller subproblems. Next the new model will be defined.

3.2 Model definition

For time series $\{\boldsymbol{Y}_t\}_{t\in\mathbb{N}}$ represented by a dynamic chain graph as described above, the DCGM is defined for all $t \in \mathbb{N}$ as follows. The initial information available is denoted by D_0.

$$\textit{Observation equations:} \quad \boldsymbol{Y}_t(i) = \boldsymbol{F}_t(i)^{\mathrm{T}}\boldsymbol{\theta}_t(i) + \boldsymbol{v}_t(i), \quad \boldsymbol{v}_t(i) \sim (\boldsymbol{0}, \boldsymbol{\Sigma}_t(i)), \quad i = 1, \ldots, N, \tag{5}$$

$$\textit{System equation:} \quad \boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim (\boldsymbol{0}, \boldsymbol{W}_t), \tag{6}$$

$$\textit{Initial information:} \quad \boldsymbol{\theta}_0 \mid D_0 \sim (\boldsymbol{m}_0, \boldsymbol{C}_0). \tag{7}$$

Here $\boldsymbol{F}_t(i)^{\mathrm{T}} = (\boldsymbol{F}_{t1}(i)^{\mathrm{T}}, \ldots, \boldsymbol{F}_{tr_i}(i)^{\mathrm{T}})$, where s_j-dimensional vector $\boldsymbol{F}_{tj}(i)$, $j = 1, \ldots, r_i$, is allowed to be an arbitrary, but known, function of the contemporaneous values of $\mathrm{pa}(y_{tj}(i))$ and $\boldsymbol{y}^{t-1}(1), \ldots, \boldsymbol{y}^{t-1}(i)$, but not $\boldsymbol{y}^t(i+1), \ldots, \boldsymbol{y}^t(N)$ or $\boldsymbol{y}_t(i)$. The vector $\boldsymbol{\theta}_t^{\mathrm{T}} = (\boldsymbol{\theta}_t(1)^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_t(N)^{\mathrm{T}})$ is a s-dimensional state vector, where $\boldsymbol{\theta}_t(i)$ is the s_i-dimensional state vector for $\boldsymbol{Y}_t(i)$ with $s_i = \sum_{j=1}^{r_i} s_j$ and $s = \sum_{i=1}^{N} s_i$. The $r_i \times r_i$ matrix $\boldsymbol{\Sigma}_t(i)$ is the observation covariance matrix for $\boldsymbol{Y}_t(i)$. The $s \times s$ matrices $\boldsymbol{G}_t = \mathrm{blockdiag}(\boldsymbol{G}_t(1), \ldots, \boldsymbol{G}_t(N))$ and $\boldsymbol{W}_t = \mathrm{blockdiag}(\boldsymbol{W}_t(1), \ldots, \boldsymbol{W}_t(N))$, where $\boldsymbol{G}_t(i)$ and $\boldsymbol{W}_t(i)$ are, respectively, the $s_i \times s_i$ state evolution matrix and state evolution covariance matrix for $\boldsymbol{\theta}_t(i)$, $i = 1, \ldots, N$, are allowed to be functions of $\boldsymbol{y}^{t-1}(1), \ldots, \boldsymbol{y}^{t-1}(i)$, but not $\boldsymbol{y}^{t-1}(i+1), \ldots, \boldsymbol{y}^{t-1}(N)$. In the s-dimensional vector $\boldsymbol{w}_t^{\mathrm{T}} = (\boldsymbol{w}_t(1)^{\mathrm{T}}, \ldots, \boldsymbol{w}_t(N)^{\mathrm{T}})$, $\boldsymbol{w}_t(i)$ is the s_i-dimensional system error vector for $\boldsymbol{\theta}_t(i)$, $i = 1, \ldots, N$. The s-dimensional vector $\boldsymbol{m}_0$ and $s \times s$ matrix $\boldsymbol{C}_0 = \mathrm{blockdiag}(\boldsymbol{C}_0(1), \ldots, \boldsymbol{C}_0(N))$ are moments of $\boldsymbol{\theta}_0 \mid D_0$. Errors $\boldsymbol{v}_t(1), \ldots, \boldsymbol{v}_t(N)$ and $\boldsymbol{w}_t(1), \ldots, \boldsymbol{w}_t(N)$ are mutually independent of each other and through time.

To illustrate the DCGM, consider once again the time series represented by the dynamic chain graph in Figure 3. Separate observation equations (5) are specified for $\boldsymbol{Y}_t(1)$, $\boldsymbol{Y}_t(2)$ and $\boldsymbol{Y}_t(3)$: $\boldsymbol{F}_t(1)^{\mathrm{T}} = (\boldsymbol{F}_{t1}(1)^{\mathrm{T}}, \boldsymbol{F}_{t2}(1)^{T})$, such that $\boldsymbol{F}_{t1}(1)^{\mathrm{T}}$ is a function of $\boldsymbol{y}_1^{t-1}(1)$; $\boldsymbol{F}_t(2)^{\mathrm{T}} = (\boldsymbol{F}_{t1}(2)^{\mathrm{T}}, \boldsymbol{F}_{t2}(2)^{T}, \boldsymbol{F}_{t3}(2)^{\mathrm{T}})$, such that $\boldsymbol{F}_{t1}(2)^{\mathrm{T}}$ is a function of $\boldsymbol{y}_1^{t-1}(1)$ and $\boldsymbol{F}_{t3}(2)^{T}$ is a function of $\boldsymbol{y}_3^{t-1}(2)$ and $\boldsymbol{y}_{t2}(1)$; $\boldsymbol{F}_t(3)^{\mathrm{T}} = (\boldsymbol{F}_{t1}(3)^{\mathrm{T}}, \boldsymbol{F}_{t2}(3)^{T})$, such that $\boldsymbol{F}_{t1}(3)^{\mathrm{T}}$ is a function of $\boldsymbol{y}_1^{t-1}(3)$ and $\boldsymbol{y}_{t1}(2)$ and $\boldsymbol{F}_{t2}(3)^{T}$ is a function of $\boldsymbol{y}_2^{t-1}(3)$ and $\boldsymbol{y}_{t3}(2)$. Additionally, $\boldsymbol{F}_t(1)^{\mathrm{T}}$, $\boldsymbol{F}_t(2)^{\mathrm{T}}$ and $\boldsymbol{F}_t(3)^{\mathrm{T}}$ can also be functions of exogenous variables.

Corollary 1 to follow presents a key result for the DCGM. Corollary 1 follows on from a theorem which is provided in the Supplementary Material for the paper (Anacleto and Queen, 2016). The proofs of this theorem and of the corollary are also provided in the Supplementary Material.

Corollary 1. *If* $\perp\!\!\!\perp_{i=1}^{N} \boldsymbol{\theta}_0(i)$*, then under the DCGM, for all* $t \in \mathbb{N}$*,*

1. $\perp\!\!\!\perp_{i=1}^{N} \boldsymbol{\theta}_t(i) \mid \boldsymbol{y}^t$*, and*
2. $\boldsymbol{\theta}_t(i) \perp\!\!\!\perp \boldsymbol{y}^t(i+1), \ldots, \boldsymbol{y}^t(N) \mid \boldsymbol{y}^t(1), \ldots, \boldsymbol{y}^t(i)$*, for* $i = 1, \ldots, N-1$*.*

Corollary 1 means that if $\boldsymbol{\theta}_t(1), \ldots, \boldsymbol{\theta}_t(N)$ start independent, then they remain so after sampling: initial independence is ensured since $\boldsymbol{C}_0$ is block diagonal in (7). Corollary 1 and Theorem 1 together mean that each parameter vector $\boldsymbol{\theta}_t(i)$ can be updated separately within the conditional multivariate model for $\boldsymbol{Y}_t(i) \mid \mathrm{pa}(\boldsymbol{y}_t(i))$, and conditional forecasts for $\boldsymbol{Y}_t(i) \mid \mathrm{pa}(\boldsymbol{y}_t(i))$ can be found separately, since the joint forecast distribution can be expressed as

$$f(\boldsymbol{y}_t \mid \boldsymbol{y}^{t-1}) \quad = \quad \prod_{i=1}^{N} \int_{\boldsymbol{\theta}_t(i)} f(\boldsymbol{y}_t(i) \mid \mathrm{pa}(\boldsymbol{y}_t(i)), \boldsymbol{\theta}_t(i)) f(\boldsymbol{\theta}_t(i) \mid \boldsymbol{x}^{t-1}(i), \boldsymbol{y}^{t-1}(i)) d\boldsymbol{\theta}_t(i).$$

Another consequence of $\boldsymbol{\theta}_t(1), \ldots, \boldsymbol{\theta}_t(N)$ remaining independent after sampling, is that the smoothing distributions $\boldsymbol{\theta}_{t-k}(i) \mid \boldsymbol{y}^t$, $k = 1, \ldots, t-1$, can be calculated separately for each $i = 1, \ldots N$. (For example, when normal errors are assumed, each smoothing density has the form as given in West and Harrison, 1997, page 113). The DCGM therefore decomposes the n-dimensional model into N separate conditional multivariate models of smaller dimensions. This decomposition greatly simplifies model computations, breaking what can be a highly complex multivariate problem into more manageable parts.

It is worth emphasizing that the regression vectors, $\boldsymbol{F}_{tj}(i)$, $i = 1, \ldots, N$, $j = 1, \ldots, r_i$, are functions of the *contemporaneous* values of the parents of $Y_{tj}(i)$, which are unknown at time $t-1$ when forecasts for $Y_{tj}(i)$ are required. Although the regression vector at time t is usually known before time t in the Bayesian dynamic linear model (DLM) framework, the idea of having unknown random variables in the regression vector is not new: both the MDM and Queen and Smith (1992) dynamic graphical model also allow the regression vectors to be functions of the (unknown) contemporaneous values of parents, while Wang et al. (2011) also consider DLMs with random vectors. To obtain forecasts for $\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(N)$ in the DCGM, marginal forecasts (marginalizing over $\mathrm{pa}(\boldsymbol{y}_t(i))$) are required. Although the marginal distributions are not generally simple distributional forms, it is usually straightforward to calculate the marginal forecast moments from the conditional ones using the identities $E(X) = E[E(X \mid Z)]$ and $V(X) = E[V(X \mid Z)] + V[E(X \mid Z)]$.

The MDM is a special case of the DCGM in which all the chain components are single values so that $N = n$. In both models, the set of contemporaneous variables

$\mathrm{pa}(y_{tj}(i))$ are used as regressors when modelling $Y_{tj}(i)$ and both models break the multivariate problem into simpler sub-problems. However, whereas the MDM breaks the n-dimensional problem into n univariate ones, the dynamic chain graph model breaks the problem into N separate multivariate models for the chain components.

In this respect the proposed model is like the dynamic graphical model of Queen and Smith (1992). The DCGM is, however, far more general: if $Y_{tk}(l)$ is a parent of $Y_{tj}(i)$ in a chain graph representing $\boldsymbol{Y}_t$, then the dynamic graphical model would require *all* components of the vector $\boldsymbol{Y}_t(l)$ to be parents to *all* components of the vector $\boldsymbol{Y}_t(i)$, whereas in the DCGM $Y_{tj}(i)$ can have any number and combination of component series of $\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(i-1)$ as parents and the other components of $\boldsymbol{Y}_t(i)$ need not have the same parents. Also, in the dynamic graphical model, all component series within a chain component must be pairwise connected, whereas this need not be the case in the DCGM. Further, unlike the dynamic graphical model, no distributional assumptions are made for the priors or error distributions in the DCGM, $\boldsymbol{F}_t(i)$ in (5) need not be a linear function of its parents, and no assumptions are made regarding the multivariate dynamic models for $\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(N)$.

The dynamic graphical model uses matrix normal DLMs (Quintana and West, 1987) to model each of $\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(N)$. This model is conjugate and thus computationally simple and quick to use. However, although the matrix normal DLM can be used to model $\boldsymbol{Y}_t(1)$ in the DCGM, it is not appropriate for $\boldsymbol{Y}_t(i) \mid \mathrm{pa}(\boldsymbol{y}_t(i))$, $i = 2, \ldots, N$: the matrix normal DLM would require each of the individual series $Y_{t1}(i), \ldots, Y_{tr_i}(i)$, $i = 2, \ldots, N$, to have the *same* regression vector so that $\boldsymbol{F}_{t1}(i) = \cdots = \boldsymbol{F}_{tr_i}(i)$, whereas in the DCGM each $\boldsymbol{F}_{tj}(i)$, $j = 1, \ldots, r_i$, is potentially different because it is a function of $\mathrm{pa}(y_{tj}(i))$, and $\mathrm{pa}(y_{tj}(i))$ is not necessarily equal to $\mathrm{pa}(y_{tk}(i))$, for $j \neq k$.

The simplest models for $\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(N)$ are DLMs where $\boldsymbol{F}_{tj}(i)$, $j = 1, \ldots, r_i$, $i = 1, \ldots, N$, is a linear function of regressor(s) $\mathrm{pa}(y_{tj}(i))$ and all distributions in (5)–(7) are normal. This is the *linear* dynamic chain graph model (LDCGM). In the next section, the LDCGM is applied to forecast road traffic network flows.

4 Application: forecasting traffic network flows

Anacleto et al. (2013a,b) used a linear version of the MDM, the LMDM, to forecast flows in a road traffic network at the intersection of three busy motorways near Manchester, UK. Figure 4(a) shows a schematic diagram of the network: arrows represented by the roadways indicate the direction of travel and circles denote the flow data collection sites which are labelled by identification numbers. In this paper, an LDCGM is used to forecast flows in part of this network and the performance of the LDCGM and the LMDM is compared.

Time series data of 5-minute counts of vehicles passing over induction loops (see Li, 2009) in the Manchester network for November and December 2010 are available from the Highways Agency in England (`http://www.highways.gov.uk/`). Let $Y_t(k)$ denote the traffic flow (5-minute vehicle counts) at site k at time t. Anacleto et al. (2013b) elicited a directed acyclic graph (DAG) to represent these traffic flow series.

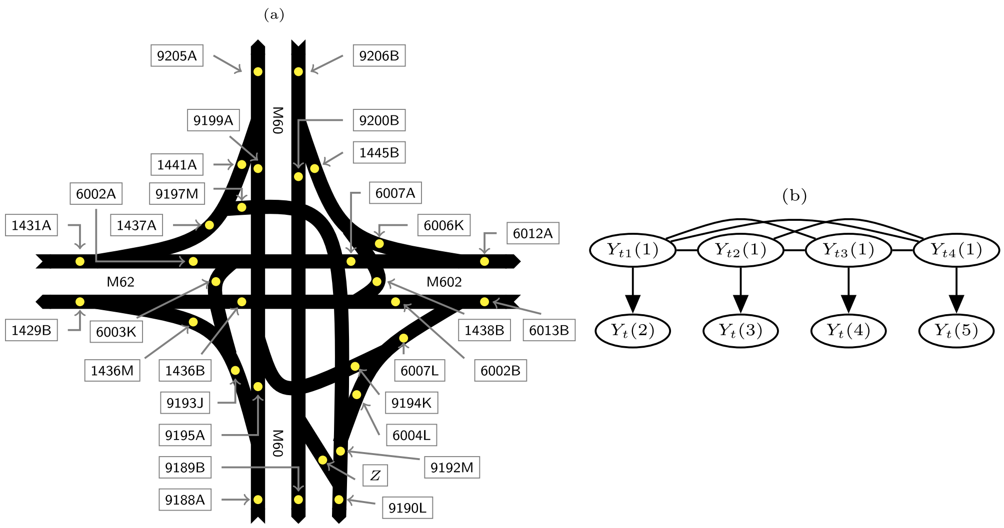

Figure 4: (a) Schematic diagram of a traffic network near Manchester. (b) A chain graph for a subset of data collection sites of the Manchester network.

In that DAG, all variables have one or two parents except for the time series at the four entrances to the network, namely $Y_t(9206B)$, $Y_t(6013B)$, $Y_t(9188A)$ and $Y_t(1431A)$, which do not have parents: these variables without parents are referred to as root nodes.

Queen et al. (2008) showed that, for any two root nodes $Y_t(k)$ and $Y_t(l)$ being modelled by an LMDM, the forecast covariance between $Y_t(k)$ and $Y_t(l)$ is 0. This result also holds for the general MDM. However, this is an unrealistic assumption for the Manchester network traffic flow series, where the root nodes (the series at the entrances to the network) can be highly correlated (Anacleto, 2012). In this case, a chain graph representing the root nodes in a chain component may be a more suitable representation of the flow series.

For clarity of presentation, consider a small subset of the Manchester network comprising the flow series at the entrances to the network and four of the adjacent downstream flows (the four root nodes with one of each of their respective children in the DAG representation). For notational convenience, let

$$\begin{array}{llll} Y_t(9206B) = Y_{t1}(1), & Y_t(6013B) = Y_{t2}(1), & Y_t(9188A) = Y_{t3}(1), & Y_t(1431A) = Y_{t4}(1), \\ Y_t(9200B) = Y_t(2), & Y_t(6007L) = Y_t(3), & Y_t(9193J) = Y_t(4), & Y_t(1437A) = Y_t(5), \end{array}$$

and set $\boldsymbol{Y}_t = (\boldsymbol{Y}_t(1)^{\mathrm{T}}, Y_t(2), \ldots, Y_t(5))^{\mathrm{T}}$ where $\boldsymbol{Y}_t(1)^{\mathrm{T}} = (Y_{t1}(1), \ldots, Y_{t4}(1))$. A chain graph representation of $\boldsymbol{Y}_t$ is given in Figure 4(b): directed edges from $Y_{tj}(1)$ to $Y_t(j+1)$, $j = 1, \ldots, 4$, represent parent child relationships from the original DAG representation in Anacleto et al. (2013b), and undirected edges represent associations between pairs of root nodes.

4.1 An LDCGM for the traffic network

The chain graph in Figure 4(b) has only one multivariate chain component, $\boldsymbol{Y}_t(1)$, while the other chain components are single series. In this case an LDCGM can be defined in which a matrix normal DLM is used to model $\boldsymbol{Y}_t(1)$, while conditional univariate DLMs are used to model $Y_t(2), \ldots, Y_t(5)$.

A matrix normal DLM for $\boldsymbol{Y}_t(1)$ is specified in terms of *row* vector $\boldsymbol{Y}_t(1)^{\mathrm{T}}$ and $s_1 \times 4$ *matrix* parameter $\boldsymbol{\Theta}_t(1) = (\boldsymbol{\theta}_{t1}(1), \ldots, \boldsymbol{\theta}_{t4}(1))$, where $\boldsymbol{\theta}_{tj}(1)$ is the s_1-dimensional state vector for $Y_{tj}(1)$, $j = 1, \ldots, 4$. $Y_{t1}(1), \ldots, Y_{t4}(1)$ each has the same regression vector $\boldsymbol{F}_t(1)$, and each state vector $\boldsymbol{\theta}_{t1}(1), \ldots, \boldsymbol{\theta}_{t4}(1)$ has the same dimension (s_1) and the same state evolution matrix $\boldsymbol{G}_t(1)$.

Denoting $\widetilde{\boldsymbol{\theta}}_t^{\mathrm{T}} = (\boldsymbol{\theta}_t(2)^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_t(5)^{\mathrm{T}})$, an LDCGM for $\boldsymbol{Y}_t$ in the Manchester network is defined as follows for times $t \in \mathbb{N}$.

Observation equations:

$$\boldsymbol{Y}_t(1)^{\mathrm{T}} = \boldsymbol{F}_t(1)^{\mathrm{T}}\boldsymbol{\Theta}_t(1) + \boldsymbol{v}_t(1)^{\mathrm{T}}, \qquad \boldsymbol{v}_t(1) \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_t(1)), \tag{8}$$

$$Y_t(i) = \boldsymbol{F}_t(i)^{\mathrm{T}}\boldsymbol{\theta}_t(i) + v_t(i), \;\; v_t(i) \sim N(0, V_t(i)), \quad i = 2, \ldots, 5. \tag{9}$$

System equations:

$$\boldsymbol{\Theta}_t(1) = \boldsymbol{G}_t(1)\boldsymbol{\Theta}_{t-1}(1) + \boldsymbol{\Omega}_t(1), \boldsymbol{\Omega}_t(1) \sim N(\boldsymbol{0}, \boldsymbol{W}_t(1), \boldsymbol{\Sigma}_t(1)), \tag{10}$$

$$\widetilde{\boldsymbol{\theta}}_t = \widetilde{\boldsymbol{G}}_t\widetilde{\boldsymbol{\theta}}_{t-1} + \widetilde{\boldsymbol{w}}_t, \widetilde{\boldsymbol{w}}_t \sim N(\boldsymbol{0}, \widetilde{\boldsymbol{W}}_t). \tag{11}$$

Initial information:

$$\boldsymbol{\Theta}_0(1) \mid D_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0(1), \boldsymbol{\Sigma}_0(1)), \tag{12}$$

$$\widetilde{\boldsymbol{\theta}}_0 \mid D_0 \sim N(\widetilde{\boldsymbol{m}}_0, \widetilde{\boldsymbol{C}}_0). \tag{13}$$

The s_1-dimensional vector $\boldsymbol{F}_t(1)$ may be a function of $\boldsymbol{y}^{t-1}(1)$ but not $\boldsymbol{y}^t(2), \ldots, \boldsymbol{y}^t(5)$; s_i-dimensional vector $\boldsymbol{F}_t(i)$ is a linear function of $\mathrm{pa}(y_t(i))$, $i = 2, \ldots, 5$; 4×4 matrix $\boldsymbol{\Sigma}_t(1)$ defines a cross-sectional covariance structure across $\boldsymbol{Y}_t(1)$; $V_t(i)$ is the scalar observation variance for $Y_t(i)$, $i = 2, \ldots, 5$; $\boldsymbol{G}_t(1)$ is the $s_1 \times s_1$ state evolution matrix for $\boldsymbol{\Theta}_t(1)$; $\widetilde{\boldsymbol{G}}_t = \mathrm{blockdiag}(\boldsymbol{G}_t(2), \ldots, \boldsymbol{G}_t(5))$ is the state evolution matrix for $\widetilde{\boldsymbol{\theta}}_t$; $\boldsymbol{\Omega}_t(1)$ is the $s_1 \times 4$ matrix of system errors for $\boldsymbol{\Theta}_t(1)$ with matrix normal distribution (Dawid, 1981), with $s_1 \times 4$ mean matrix of zeros, $s_1 \times s_1$ left covariance matrix $\boldsymbol{W}_t(1)$ and 4×4 right covariance matrix $\boldsymbol{\Sigma}_t(1)$; $\widetilde{\boldsymbol{w}}_t$ is the system error vector for $\widetilde{\boldsymbol{\theta}}_t$; $\widetilde{\boldsymbol{W}}_t = \mathrm{blockdiag}(\boldsymbol{W}_t(2), \ldots, \boldsymbol{W}_t(5))$ is the state evolution covariance matrix for $\widetilde{\boldsymbol{\theta}}_t$; $\boldsymbol{\Theta}_0(1) \mid D_0$ has a matrix normal distribution with $s_1 \times 4$ mean matrix $\boldsymbol{m}_0$, $s_1 \times s_1$ left covariance matrix $\boldsymbol{C}_0(1)$ and 4×4 right covariance matrix $\boldsymbol{\Sigma}_0(1)$; and $\widetilde{\boldsymbol{m}}_0$ and $\widetilde{\boldsymbol{C}}_0$ are the moments of $\widetilde{\boldsymbol{\theta}}_0 \mid D_0$. All model errors are mutually independent of each other and independent through time.

Matrix $\boldsymbol{\Sigma}_t(1)$ and variances $V_t(i)$, $i = 2, \ldots, 5$, are estimated sequentially on-line using conjugate inverse Wishart and gamma priors, respectively: see West and Harrison (1997, pages 108–112, 603–604). Conjugacy allows quick and easy computation.

To evaluate the effect of the joint modelling of $Y_{t1}(1), \ldots, Y_{t4}(1)$ in a chain component with the LDCGM, forecasts were also obtained using an LMDM with no such association structure. The graph for this LMDM is the DAG obtained by removing the undirected edges from the chain graph in Figure 4(b), so that $Y_{t1}(1), \ldots, Y_{t4}(1)$ are unconnected root nodes and $Y_{tj}(1)$ is a parent of $Y_t(j+1)$, for $j = 1, \ldots, 4$.

Series $Y_t(2), \ldots, Y_t(5)$ are modelled in exactly the same way via (9), (11) and (13) in both the LMDM and the LDCGM, whereas in the LMDM each $Y_{tj}(1)$, $j = 1, \ldots, 4$, is modelled by a separate DLM of the form:

$$\begin{aligned}
&\textit{Obs. equation:} && Y_{tj}(1) = \boldsymbol{F}_{tj}(1)^{\mathrm{T}}\boldsymbol{\theta}_{tj}(1) + v_{tj}(1), \quad v_{tj}(1) \sim N(0, V_{tj}(1)) && (14)\\
&\textit{Sys. equation:} && \boldsymbol{\theta}_{tj}(1) = \boldsymbol{G}_{tj}(1)\boldsymbol{\theta}_{t-1,j}(1) + \boldsymbol{w}_{tj}(1), \ \boldsymbol{w}_{tj}(1) \sim N(\boldsymbol{0}, \boldsymbol{W}_{tj}(1)) && (15)\\
&\textit{Initial info.:} && \boldsymbol{\theta}_{0j}(1) \mid D_0 \sim N(\boldsymbol{m}_{0j}(1), \boldsymbol{C}_{0j}(1)). && (16)
\end{aligned}$$

Following Anacleto et al. (2013b), because of differences in flow patterns for different weekdays, for clarity of presentation only flows from Wednesdays are used here. In the absence of expert information, data from November were used to elicit all priors, while one-step ahead forecasts are obtained for December. Heavy snow caused several periods of disruption to the network traffic during December 2010. The models are thus compared when an explicit factor was affecting the traffic flows: it is at such times when forecasting is of most use for traffic control.

Traffic flow series exhibit daily patterns which both models need to accommodate. Following Anacleto et al. (2013a), cubic splines can model these daily patterns so that the regression vectors $\boldsymbol{F}_t(1)$ in (8) and $\boldsymbol{F}_{t1}(1), \ldots, \boldsymbol{F}_{t4}(1)$ in (14) contain fixed basis functions, while $\boldsymbol{\Theta}_t(1)$ in (8) and $\boldsymbol{\theta}_{t1}(1), \ldots, \boldsymbol{\theta}_{t4}(1)$ in (14) contain dynamically evolving spline parameters for individual series which are estimated sequentially online. In the matrix normal DLM, the same $\boldsymbol{F}_t(1)$, and hence basis functions, are used for each series $Y_{t1}(1), \ldots, Y_{t4}(1)$. The daily patterns exhibited by $Y_{t1}(1), \ldots, Y_{t4}(1)$ are similar, and variation in patterns is accommodated through each series having different parameters. Evolution matrices, $\boldsymbol{G}_t(1)$ in (10) and $\boldsymbol{G}_{t1}(1), \ldots, \boldsymbol{G}_{t4}(1)$, in (15) are identity matrices.

For both the LMDM and the LDCGM, $Y_t(2), \ldots, Y_t(5)$ are modelled in the same way: separate regression DLMs are defined for $Y_t(2), \ldots, Y_t(5)$ where each $Y_t(i)$, $i = 2, \ldots, 5$, has $\mathrm{pa}(y_t(i)) = y_{t,i-1}(1)$ as a linear regressor. The parameters for these regressors exhibit daily patterns, and, following Anacleto et al. (2013a), these can also be modelled by cubic splines so that $\boldsymbol{F}_t(i)$ and $\boldsymbol{\theta}_t(i)$ in (9) contain fixed basis functions and dynamically evolving spline parameters, respectively. Matrix $\tilde{\boldsymbol{G}}_t(i)$ in (11) is an identity matrix. Exogenous variables — namely, speed, occupancy and headway — are available at each traffic site, and are also considered in the DLMs for $Y_t(2), \ldots, Y_t(5)$ using splines: see Anacleto et al. (2013a) for details.

In the LDCGM, however, the matrix normal DLM for $\boldsymbol{Y}_t(1)$ requires $Y_{t1}(1), \ldots,$ $Y_{t4}(1)$ to have the *same* regression vector $\boldsymbol{F}_t(1)$. Thus, it is not possible for $Y_{tj}(1)$'s model to include exogenous variables (i.e. speed, occupancy and headway) at that site as predictors, without also including exogenous variables at all the other sites in $\boldsymbol{Y}_t(1)$ as predictors as well. Thus these predictors are not included for $\boldsymbol{Y}_t(1)$, and for fairness, are also not included when modelling $Y_{t1}(1), \ldots, Y_{t4}(1)$ as root nodes in the LMDM. As an alternative, predictors in $\boldsymbol{Y}_t(1)$ could be included by using the seemingly unrelated regression model proposed in Wang (2010), which requires MCMC for parameter estimation. However, since the emphasis here is the evaluation of the effect of capturing both directed and undirected relationships with the LDCGM in comparison to just capturing directed relationships with the LMDM, a simpler model, such as the matrix normal DLM, can be used.

For both models, the observation variances $V_t(2), \ldots, V_t(5)$ in (9) are estimated online following variance laws (West and Harrison, 1997, Chapter 10.7) which relate the observation variance with mean flow, introduced in Anacleto et al. (2013b) to account for heterogeneity in traffic flow series. The cross-sectional covariance matrix, $\boldsymbol{\Sigma}_t(1)$, is also estimated online. However, covariances between $Y_{t1}(1), \ldots, Y_{t4}(1)$ don't necessarily change with the mean, so this matrix is estimated using the discounting variance learning techniques of West and Harrison (1997, page 608) alone. For fairness of comparison variance laws are not used to model each $V_{tj}(1)$ in equation (14), so that the scalar observational variances of $Y_{t1}(1), \ldots, Y_{t4}(1)$ in the LMDM are also modelled using discounting techniques only. Discount factors for all observation variances and $\boldsymbol{\Sigma}_t(1)$ vary around 0.90. Prado and West (2010) point out that variance learning via discount factors is only suitable when the (co)variances have a smooth and gradual random change. This is a reasonable assumption between 15:00 to 19:59 for these data, but not at other times. Thus only data between 15:00 and 19:59 are considered here.

Discount factors for all evolution covariance matrices in the LMDM and the LDCGM vary around 0.97, estimated on-line using standard discounting techniques (see West and Harrison, 1997, page 193). All discount factors are chosen by comparing the forecast accuracy of different models for each chain component, obtained through combination of observation and evolution discount factor values ranging from 0.80 to 1.

4.2 Forecast performance

Because of the heteroscedasticity of traffic flow series, the joint log-predictive likelihood (LPL), which assesses the precision of forecasts as well as point forecasts, is used when evaluating model forecast performance. The LPL calculates the log of the joint one-step ahead forecast distribution for $\boldsymbol{Y}_t$ *before* $\boldsymbol{y}_t$ is observed, and then evaluates this at the observed value $\boldsymbol{y}_t$. The LPL is the aggregate of all these values over all time points. Anacleto et al. (2013a) provides details of the LPL for the LMDM and this is easily adapted for the LDCGM.

Table 1 shows the LPL values when forecasting Wednesday traffic flows in December 2010 using the LMDM and the LDCGM. The first row of Table 1 shows the forecast performance when only the four series in $\boldsymbol{Y}_t(1)$ are modelled: clearly the matrix normal DLM for $\boldsymbol{Y}_t(1)$ used in the LDCGM provides better forecasts than the independent DLMs assumed for $\boldsymbol{Y}_t(1)$ under the LMDM. From the second row of Table 1, the LDCGM also performs better than the LMDM when all eight series are considered. The one-step ahead forecast means are very similar for both models, while the forecast variances for the LDCGM are slightly smaller, and so slightly more informative, than those for the LMDM. However, the real advantage of using the LDCGM is seen when considering multivariate forecasts. Figure 5 shows $(y_{t1}(1), y_{t2}(1))$, represented by a dot, at three consecutive time intervals. The 90% forecast regions for $(Y_{t1}(1), Y_{t2}(1))$

	LPL	
Series considered	LMDM	LDCGM
$\boldsymbol{Y}_t(1)$ only	$-6,728$	$-6,154$
$\boldsymbol{Y}_t(1),\ldots,Y_t(5)$	$-11,488$	$-10,914$

Table 1: LPL values for the LMDM and the LDCGM.

for the LMDM and the LDCGM are represented by black and grey ellipses, respectively. In each plot, the forecast regions are smaller, thus more informative, for the LDCGM. The LDCGM forecast regions also clearly indicate a positive correlation between $(Y_{t1}(1), Y_{t2}(1))$, which the other model does not. Forecast regions show positive correlations amongst root nodes at other times too (Anacleto, 2012).

In Figure 5 the observed flows are not close to the centre of the forecast regions for either model. This could be due to variability in traffic flows which is not captured by the models. Neither model uses the exogenous variables speed, occupancy and headway, which may have captured some of this variability. The LDCGM does, however, perform better in Figure 5 than the LMDM: for example, in Figure 5(b), the observed flow lies within the forecast region for the former, but does not for the latter.

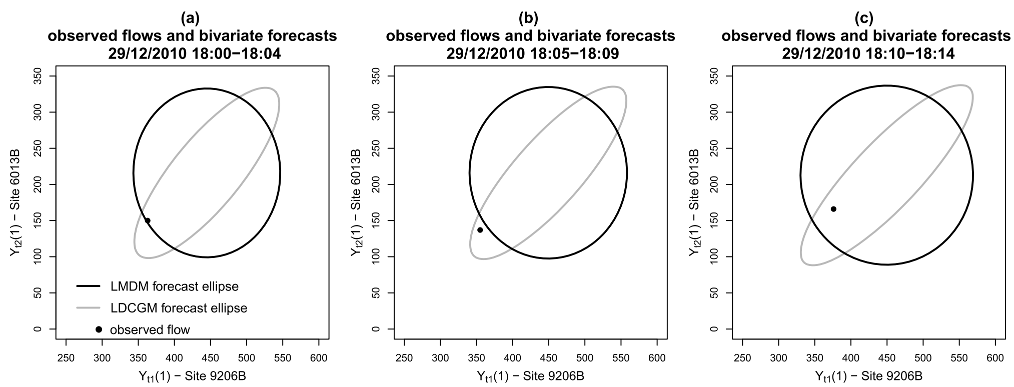

Figure 5: Observed flows $(\cdot)$ and bivariate forecast limits at a pair of root nodes.

The matrix normal DLM is not ideal for modelling $\boldsymbol{Y}_t(1)$: the covariances vary too much between times 20:00 and 14:59 to estimate $\boldsymbol{\Sigma}_t(1)$ through variance learning discounting, and the exogenous variables speed, occupancy and headway, cannot be used. However, even with these restrictions, it has been shown that an LDCGM is worth consideration as an alternative to the LMDM for modelling traffic flows.

5 Application: modelling time series gene expression data

The graph in Figure 4(b) was elicited by exploiting the direction of traffic in the network as the causal driving mechanism across the time series. Alternatively, graphs can be also inferred from data. As an example from biology, gene expression information, obtained through measurements of DNA transcription, can be used to infer networks describing regulatory mechanisms among genes (Kolaczyk, 2009). These networks can validate known gene associations and also allow the discovery of new gene relationships.

To capture the dynamics of biological processes, time-varying expression data may be preferable to transcriptional measurements obtained at steady-state. Dynamic Bayesian networks (DBNs) have been extensively applied to analyse time series gene expression data, as these models can capture feedback mechanisms which are ubiquitous in gene regulation (Lèbre, 2009; Husmeier et al., 2011). Feedback mechanisms are captured in DBNs by replicating a DAG at each time-slice, with arrows connecting these DAGs based on the causal flow of time.

DCGMs can directly address two current limitations with DBNs for time series gene expression data. Firstly, DBNs usually assume that parents of a expression time series of a gene at a given time only take values at the previous time points. However, it is experimentally challenging to define a suitable sampling rate to collect time-varying transcriptional data so that this assumption is valid (Bar-Joseph et al., 2012). Secondly, even though symmetric associations between expression of different genes are common in transcriptional networks (Cantone et al., 2009), current DBNs for gene expression data cannot capture both directed and undirected edges in a graph.

The DCGM was used to model two gene expression datasets. The first dataset was obtained from an experiment involving the plant *Arabidopsis Thaliana*, as described in Smith et al. (2004). The expression of 800 genes at 11 different time points were measured and, following the correlation analysis in Opgen-Rhein and Strimmer (2007), 92 genes with the most significant connections were considered. The second dataset consists of gene expression measurements at 18 time points from an experiment aimed at understanding the developmental process of the mammary gland in mice (Stein et al., 2004). For the second dataset, the focus is on 30 genes identified using cluster analysis by Abegaz and Wit (2013) as providing the best separation between developmental stages. Section S2 of the Supplementary Material provides a full description of the DCGM application to both datasets. The results of the application strongly suggest that the DCGM can improve modelling of gene expression data.

6 Final remarks

This paper presents a novel Bayesian dynamic model — the DCGM — for multivariate time series which assume a chain graph representation of the conditional independence structure among time series components. The new model deals with high-dimensional time series by *decoupling* multivariate time series of lower dimensions for sequential inference, which can then be *recoupled* for forecasting and decision analysis. This decoupling of high dimensional time series for sequential inference and recoupling for forecasting and decision making is also at the heart of the MDM and a model recently developed by Gruber and West (2015), although for these models the decoupling involves the univariate time series components, rather than subsets of multivariate series. Under the DCGM, state parameters of the time series subsets remain independent after sampling, which allows sequential and parallel inference of these subsets. The paper demonstrates how the DCGM improves time series modelling when there is evidence that conditional independence among time series components are better represented by both directed and undirected relationships in a graph, rather than directed relationships alone.

Application of the DCGM was illustrated here using traffic flow and gene expression networks. The model does, however, have much wider applicability to any multivariate time series which exhibits symmetric associations between groups of series together with a conditional independence and causal structure. The DCGM can also be used for any chain graph application (such as can be found, for example, in Cox and Wermuth, 1996) which may be part of a longitudinal study over time. What's more, the fact that the model is very general and does not specify a particular multivariate model to use for each chain component, nor impose linearity and normality, increases its potential application areas.

The traffic flow time series analysed in this paper were obtained on a set of sites distributed over space, and so can be viewed as being generated from a spatio-temporal process. Multivariate time series from such processes are available in a variety of areas, and Cressie and Wikle (2011, Chapter 2.4) suggest that chain graphs are a natural template for representing such data. The DCGM could therefore be a potential candidate when modelling time series originating from spatio-temporal processes.

Whereas inference of DAGs and undirected graphs from data is a lively research area (see, for example, Scutari, 2013; Mohammadi and Wit, 2015; Wang, 2015), models for inferring chain graphs from data has received little attention (McCarter and Kim, 2014). A structural learning method for chain graphs using time series data has been recently proposed by Abegaz and Wit (2013). However, their model is based on vector autoregressive (VAR) processes, therefore relying on stringent assumptions of those models. In this context, following recent successful developments of DAG inference methods using the MDM (see Costa et al., 2015; Oates et al., 2015a,b), the DCGM is an important building block for inferring chain graphs from time series data.

Supplementary Material

Supplementary material for paper: Dynamic chain graph models for time series network data (DOI: 10.1214/16-BA1010SUPP; .pdf). Supplementary material available online includes the theorem for which Corollary 1 is a consequence, together with the proofs of that theorem and Corollary 1. It also includes the description and results of the application of the DCGM to two gene expression datasets, as mentioned in Section 5.

References

Abegaz, F. and Wit, E. (2013). "Sparse time series chain graphical models for reconstructing genetic networks." *Biostatistics*, 14(3): 586–599. 504, 505

Anacleto, O. (2012). "Bayesian dynamic graphical models for high-dimensional flow forecasting in road traffic networks." Ph.D. thesis, The Open University. 499, 503

Anacleto, O. and Queen, C. (2016). "Supplementary material for paper: Dynamic chain graph models for time series network data." *Bayesian Analysis*. doi: `http://dx.doi.org/10.1214/16-BA1010SUPP`. 497

Anacleto, O., Queen, C., and Albers, C. J. (2013a). "Forecasting multivariate road traffic flows using Bayesian dynamic graphical models, splines and other traffic variables." *Australian & New Zealand Journal of Statistics*, 55(2): 69–86. MR3079021. doi: `http://dx.doi.org/10.1111/anzs.12026`. 492, 498, 501, 502

Anacleto, O., Queen, C., and Albers, C. J. (2013b). "Multivariate forecasting of road traffic flows in the presence of heteroscedasticity and measurement errors." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(2): 251–270. MR3045876. doi: `http://dx.doi.org/10.1111/j.1467-9876.2012.01059.x`. 492, 498, 499, 501, 502

Bar-Joseph, Z., Gitter, A., and Simon, I. (2012). "Studying and modelling dynamic biological processes using time-series gene expression data." *Nature Reviews Genetics*, 13(8): 552–564. 504

Cantone, I., Marucci, L., Iorio, F., Ricci, M. A., Belcastro, V., Bansal, M., Santini, S., Di Bernardo, M., Di Bernardo, D., and Cosma, M. P. (2009). "A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches." *Cell*, 137(1): 172–181. 504

Carvalho, C. M. and West, M. (2007). "Dynamic matrix-variate graphical models." *Bayesian Analysis*, 2(1): 69–97. MR2289924. doi: `http://dx.doi.org/10.1214/07-BA204`. 491, 492

Costa, L., Smith, J., Nichols, T., Cussens, J., Duff, E. P., and Makin, T. R. (2015). "Searching multiregression dynamic models of resting-state fMRI networks using integer programming." *Bayesian Analysis*, 10(2): 441–478. MR3420889. doi: `http://dx.doi.org/10.1214/14-BA913`. 492, 505

Cox, D. R. and Wermuth, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation*, volume 67. CRC Press. MR1456990. 505

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. John Wiley & Sons. MR2848400. 505

Dawid, A. P. (1981). "Some matrix-variate distribution theory: notational considerations and a Bayesian application." *Biometrika*, 68(1): 265–274. MR0614963. doi: `http://dx.doi.org/10.1093/biomet/68.1.265`. 500

Gruber, L. F. and West, M. (2015). "GPU-accelerated Bayesian learning in simultaneous graphical dynamic linear models." *Bayesian Analysis*, 11(1): 125–149. MR3447094. doi: `http://dx.doi.org/10.1214/15-BA946`. 504

Husmeier, D., Werhli, A. V., and Grzegorczyk, M. (2011). "Advanced Applications of Bayesian Networks in Systems Biology." In Stumpf, M. P. H., Balding, D. J., and Girolami, M. (eds.), *Handbook of Statistical Systems Biology*, 270–289. Wiley Online Library. MR2920207. doi: `http://dx.doi.org/10.1002/9781119970606`. 504

Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models*. Springer, New York. MR2724362. doi: http://dx.doi.org/10.1007/978-0-387-88146-1. 491, 504

Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press. MR1419991. 491, 493

Lèbre, S. (2009). "Inferring dynamic genetic networks with low order independencies." *Statistical Applications in Genetics and Molecular Biology*, 8(1): 1–38. MR2476387. doi: http://dx.doi.org/10.2202/1544-6115.1294. 504

Li, B. (2009). "A non-Gaussian Kalman filter with application to the estimation of vehicular speed." *Technometrics*, 51(2): 162–172. MR2668171. doi: http://dx.doi.org/10.1198/TECH.2009.0017. 498

McCarter, C. and Kim, S. (2014). "On Sparse Gaussian Chain Graph Models." In *Advances in Neural Information Processing Systems*, 3212–3220. 505

Mohammadi, A. and Wit, E. C. (2015). "Bayesian structure learning in sparse Gaussian graphical models." *Bayesian Analysis*, 10(1): 109–138. MR3420899. doi: http://dx.doi.org/10.1214/14-BA889. 505

Nascimento, F. F., Gamerman, D., and Lopes, H. F. (2015). "Time-varying extreme pattern with dynamic models." *TEST*, 1–19. 491

Oates, C., Smith, J., Mukherjee, S., and Cussens, J. (2015a). "Exact estimation of multiple directed acyclic graphs." *Statistics and Computing*, 1–15. http://dx.doi.org/10.1007/s11222-015-9570-9 492, 505

Oates, C. J., Costa, L., and Nichols, T. E. (2015b). "Toward a multisubject analysis of neural connectivity." *Neural Computation*, 27: 151–170. 492, 505

Opgen-Rhein, R. and Strimmer, K. (2007). "From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data." *BMC Systems Biology*, 1(1): 37. 504

Prado, R. and West, M. (2010). *Time Series: Modeling, Computation, and Inference*. CRC Press. MR2655202. 492, 502

Queen, C. and Smith, J. (1992). "Dynamic graphical models." *Bayesian Statistics*, 4: 741–751. 492, 497, 498

Queen, C. M. (1994). "Using the multiregression dynamic model to forecast brand sales in a competitive product market." *The Statistician*, 87–98. 492

Queen, C. M. and Albers, C. J. (2009). "Intervention and causality: forecasting traffic flows using a dynamic Bayesian network." *Journal of the American Statistical Association*, 104(486): 669–681. MR2751447. doi: http://dx.doi.org/10.1198/jasa.2009.0042. 492

Queen, C. M. and Smith, J. Q. (1993). "Multiregression dynamic models." *Journal of the Royal Statistical Society. Series B (Methodological)*, 849–870. MR1229884. 491

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 & New Zealand Journal of Statistics*, 49(3): 221–239. MR2405393. doi: http://dx.doi.org/10.1111/j.1467-842X.2007.00477.x. 492

Queen, C. M., Wright, B. J., and Albers, C. J. (2008). "Forecast covariances in the linear multiregression dynamic model." *Journal of Forecasting*, 27(2): 175–191. MR2420180. doi: http://dx.doi.org/10.1002/for.1050. 499

Quintana, J. M. and West, M. (1987). "An analysis of international exchange rates using multivariate DLM's." *The Statistician*, 275–281. 491, 498

Quirós, A., Wilson, S. P., Diez, R. M., Solana, A. B., and Tamames, J. A. H. (2015). "Brain activity detection by estimating the signal-to-noise ratio of fMRI time series using dynamic linear models." *Digital Signal Processing*, 47: 205–211. MR3425327. doi: http://dx.doi.org/10.1016/j.dsp.2015.06.008. 491

Scutari, M. (2013). "On the prior and posterior distributions used in graphical modelling (with discussion)." *Bayesian Analysis*, 8(3): 505–532. MR3102220. doi: http://dx.doi.org/10.1214/13-BA819. 505

Smith, S. M., Fulton, D. C., Chia, T., Thorneycroft, D., Chapple, A., Dunstan, H., Hylton, C., Zeeman, S. C., and Smith, A. M. (2004). "Diurnal changes in the transcriptome encoding enzymes of starch metabolism provide evidence for both transcriptional and posttranscriptional regulation of starch metabolism in Arabidopsis leaves." *Plant Physiology*, 136(1): 2687–2699. 504

Stein, T., Morris, J. S., Davies, C. R., Weber-Hall, S. J., Duffy, M.-A., Heath, V. J., Bell, A. K., Ferrier, R. K., Sandilands, G. P., and Gusterson, B. A. (2004). "Involution of the mouse mammary gland is associated with an immune cascade and an acute-phase response, involving LBP, CD14 and STAT3." *Breast Cancer Research*, 6(2): R75–91. 504

Velozo, P. L., Alves, M. B., and Schmidt, A. M. (2014). "Modelling categorized levels of precipitation." *Brazilian Journal of Probability and Statistics*, 28(2): 190–208. MR3189493. doi: http://dx.doi.org/10.1214/12-BJPS201. 491

Wang, H. (2010). "Sparse seemingly unrelated regression modelling: Applications in finance and econometrics." *Computational Statistics & Data Analysis*, 54(11): 2866–2877. MR2720481. doi: http://dx.doi.org/10.1016/j.csda.2010.03.028. 501

Wang, H. (2015). "Scaling it up: Stochastic search structure learning in graphical models." *Bayesian Analysis*, 10(2): 351–377. MR3420886. doi: http://dx.doi.org/10.1214/14-BA916. 505

Wang, H., Reeson, C., and Carvalho, C. M. (2011). "Dynamic financial index models: Modeling conditional dependencies via graphs." *Bayesian Analysis*, 6(4): 639–664. MR2869960. 497

Wermuth, N. and Lauritzen, S. L. (1990). "On substantive research hypotheses, conditional independence graphs and graphical chain models." *Journal of the Royal Statistical Society. Series B (Methodological)*, 21–50. MR1049302. 492

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models (2nd ed.)*. New York, NY, USA: Springer-Verlag New York, Inc. MR1482232. 491, 497, 500, 502

Xiao, S., Kottas, A., and Sansó, B. (2015). "Modeling for seasonal marked point processes: An analysis of evolving hurricane occurrences." *The Annals of Applied Statistics*, 9(1): 353–382. MR3341119. doi: `http://dx.doi.org/10.1214/14-AOAS796`. 491

Zhao, Z. Y. (2015). "Bayesian Multiregression Dynamic Models with Applications in Finance and Business." Ph.D. thesis, Duke University. MR3337678. 492

Zhao, Z. Y., Xie, M., and West, M. (2015). "Dynamic dependence networks: Financial time series forecasting & portfolio decisions." *Applied Stochastic Models in Business and Industry* (to appear). doi: `http://dx.doi.org/10.1002/asmb.2161`. 492

Acknowledgments

The authors thank the Highways Agency in England for providing the traffic data used in this paper, and Professor Tom Freeman, Dr Tom Michoel and Dr Chris Oates for valuable discussions. The authors also thank an Editor, an Associative Editor and a referee whose constructive feedback helped improve the paper. Osvaldo Anacleto was a research student at the Open University while doing part of this work, which was completed with funding from the BBSRC Institute Strategic Programme Grant (ISPG1; OA).