Dynamic chain graph models for high-dimensional time series

Osvaldo Anacleto[1] **and Catriona Queen**

University of Edinburgh and The Open University

Abstract

This paper introduces a new class of Bayesian dynamic models for inference and forecasting in high-dimensional time series problems. The new model, called the dynamic chain graph model, is a type of chain graph model suitable for multivariate time series which exhibit symmetries between 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 simplifies computation by decomposing the multivariate problem into separate, simpler sub-problems of lower dimensions. The model is applied to the problem of forecasting road traffic flows at the intersection of three busy motorways near Manchester, UK. The paper also discusses potential application of the model to analysing gene expression data.

Keywords: Bayesian forecasting; chain graph model; dynamic linear model; gene expression modelling; multiregression dynamic model; traffic flow forecasting.

1 Introduction

As data for large scale multivariate problems become more readily available, there is an increasing need for models for high-dimensional complex time series problems. This paper proposes a new class of multivariate Bayesian dynamic models (West & Harrison, 1997), called dynamic chain graph models.

The sequential and dynamic nature of Bayesian dynamic models, together with their integration of data and expert information in the Bayesian framework, make them ideal for modelling time series in many different application areas. Recent examples include environmental modelling (Ippoliti *et al.*, 2012), epidemiology (Schmidt & Pereira, 2011), finance (Quintana *et al.*, 2010) and neuroscience (Lavine *et al.*, 2011). However, few Bayesian dynamic models can accommodate complex high-dimensional multivariate series, while not being too demanding computationally.

Although the standard multivariate dynamic linear model (West & Harrison, 1997, pp 582–584) is computationally simple when the observation covariance matrix is known, the assumption of known covariance matrix is unrealistic and estimation of the matrix in a high-dimensional problem is a non-trivial task. The matrix normal dynamic linear

[1] *Corresponding Author (osvaldo.anacleto@roslin.ed.ac.uk)*

model (Quintana & West, 1987) tackles this problem by estimating the cross-sectional covariance structure sequentially online using a closed form conjugate analysis. Carvalho & West (2007) enhance this model by introducing sparsity through a graphical model representation of the cross-sectional covariance structure. The matrix normal dynamic linear model is, however, only suitable for multivariate series when the component univariate series are similar and have a common structure. So, although this model has been used successfully in practice (see for example Carvalho & West, 2007; Prado & West, 2010), it is not suitable for all high-dimensional problems.

The multiregression dynamic model (MDM) (Queen & Smith, 1993), which is an example of a dynamic Bayesian network (see Queen & Albers, 2009), does not require all the component univariate time series to have a common structure. Instead, this Bayesian dynamic model is suitable for multivariate time series whose components are believed to exhibit a conditional independence and causal structure at each time t, as expressed by a directed acyclic graph (DAG) (Lauritzen, 1996). The model uses the conditional independence structure to decompose the n-dimensional multivariate 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 computation is simplified by using local computation. The MDM is potentially suitable for any application involving flows and has been shown to be promising for modelling multivariate traffic flows in road networks (Queen *et al.* 2007; Queen & Albers, 2009; Anacleto *et al.*, 2013a, 2013b). It has also been used for modelling brand sales in a competitive market (Queen, 1994) and, recently, to model time series of functional magnetic resonance imaging (Costa *et al.*, 2013). The DAG used in the MDM may, however, be too restrictive for some applications: only directional associations between individual series can be accommodated, and no symmetric associations are allowed.

Queen & Smith (1992) developed a multivariate 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 chain components (referred to as p-sets) which are ordered subsets in which all variables within a chain component are pairwise joined by undirected edges representing symmetric associations, and directed edges in the chain graph, representing a causal

drive between processes, follow the order of chain components. The model imposes the condition that if there is a directed edge from a variable in one chain component to a variable in another chain component, then there must be a directed edge from *every* variable in the first chain component to *every* variable in the second. This is, however, a restrictive assumption, which is not always realistic.

In this paper, a new class of models is introduced which is also suitable for multivariate time series whose structure is represented by a chain graph. As with Queen & Smith's (1992) dynamic graphical model, all variables within chain components are pairwise joined by undirected edges and directed edges represent a causal drive between processes. However, unlike Queen & Smith's (1992) model, it is possible for any number and combination of variables in one chain component to be connected by directed edges to any number and combination of variables in another chain component, thus accommodating more complex multivariate time series structures. Like the MDM, computation is simplified 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 paper considers two important application areas of the proposed model: road traffic flow forecasting and gene expression modelling. Real-time traffic flow time series data are available on many roads and can be used as part of a traffic management system for traffic control and traveller information. Although traffic flow data are invariably multivariate, few traffic flow models address this issue. Not only does the proposed model accommodate the multivariate nature of the problem, but it also keeps computation manageable through local computation.

Modelling gene expression data can enhance the understanding of biological systems. Although evaluating the contemporaneous relationships between expression levels for different genes is of great interest to geneticists, this is not possible with currently available gene expression time series models. Moreover, most of these models cannot accommodate both undirected and directed relationships among genes. The paper discusses how the proposed model can tackle these problems.

Before defining the new model in the next section, some notation and terminology is introduced. Figure 1 shows an example of a chain graph with 7 variables and 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 $\mathrm{bd}(\boldsymbol{A}) = \mathrm{pa}(\boldsymbol{A}) \cup \mathrm{ne}(\boldsymbol{A})$. 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),

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

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 $X_6 \perp\!\!\!\perp \{X_1, X_2)\} \mid \{X_3, X_4, X_5\}$.

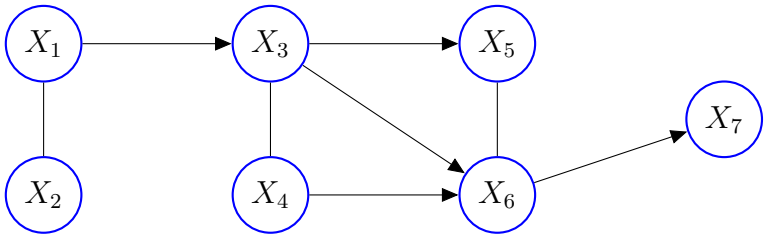

Figure 1: Example of a chain graph

2 The dynamic chain graph model

Let $\{\boldsymbol{Y}_t\}_{t>1}$ 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.

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 $i = 1, \ldots, N$ and $j, k = 1, \ldots, r_i$, $j \neq k$, individual series $Y_{tj}(i)$ and $Y_{tk}(i)$ are pairwise connected by undirected edges in a chain graph. Suppose further that there is a causal drive through the system and a conditional independence structure so that $\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(N)$ are ordered chain components in the chain graph and at each time $t \in \mathbb{N}$, for $i = 2, \ldots, N$ and $j = 1, \ldots, r_i$,

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

Additionally, suppose that the processes over all time points up to and including time t can be represented by a chain graph so that

$$Y_{tj}(i) \perp\!\!\!\perp \{\mathrm{nd}(\boldsymbol{Y}_j^t(i)) \backslash \mathrm{bd}(\boldsymbol{Y}_j^t(i))\} \mid \mathrm{bd}(\boldsymbol{Y}_j^t(i)), \boldsymbol{Y}_j^{t-1}(i).$$

The fact that for $j = 1, \ldots, r_i$, $i = 2, \ldots, N$,

$$[\{\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(i-1)\} \backslash \mathrm{bd}(Y_{tj}(i))] \subseteq \{\mathrm{nd}(Y_{tj}(i)) \backslash \mathrm{bd}(Y_{tj}(i))\},$$

while $\mathrm{bd}(Y_{tj}(i)) = \mathrm{pa}(Y_{tj}(i)) \cup \mathrm{ne}(Y_{tj}(i))$ with $\mathrm{pa}(Y_{tj}(i)) \subseteq \{\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(i-1)\}$ and $\mathrm{ne}(Y_{tj}(i)) = \{\boldsymbol{Y}_t(i) \backslash Y_{tj}(i)\}$, mean that collectively the following conditional independence statements also hold for $i = 2, \ldots, N$:

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

$$\boldsymbol{Y}_t(i) \perp\!\!\!\perp [\{\boldsymbol{Y}^t(1), \ldots, \boldsymbol{Y}^t(i-1)\} \backslash \mathrm{pa}(\boldsymbol{Y}^t(i))] \mid \mathrm{pa}(\boldsymbol{Y}^t(i)), \boldsymbol{Y}^{t-1}(i), \tag{2}$$

where $\mathrm{pa}(\boldsymbol{Y}_t(i))$ and $\mathrm{pa}(\boldsymbol{Y}^t(i))$ are the sets of all parents for $Y_{t1}(i), \ldots, Y_{tr_i}(i)$ and $\boldsymbol{Y}_1(i), \ldots, \boldsymbol{Y}_t(i)$, respectively.

The *dynamic chain graph model* (DCGM) is defined for all $t \in \mathbb{N}$ as follows. The initial information available is denoted by D_0.

$$\textit{Observation equations:} \quad \boldsymbol{Y}_t(i) = \boldsymbol{F}_t(i)^{\mathrm{T}} \boldsymbol{\theta}_t(i) + \boldsymbol{v}_t(i), \quad \boldsymbol{v}_t(i) \sim (\boldsymbol{0}, \boldsymbol{\Sigma}_t(i)), \quad i = 1, \ldots, N, \tag{3}$$

$$\textit{System equation:} \quad \boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim (\boldsymbol{0}, \boldsymbol{W}_t), \tag{4}$$

$$\textit{Initial information:} \quad \boldsymbol{\theta}_0 \mid D_0 \sim (\boldsymbol{m}_0, \boldsymbol{C}_0). \tag{5}$$

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 $\boldsymbol{F}_{tj}(i)$, $j = 1, \ldots, r_i$, contains an arbitrary, but known, function of $pa(\boldsymbol{y}_j^t(i))$ and $\boldsymbol{y}^{t-1}(i)$, but not $\boldsymbol{y}^t(i+1), \ldots, \boldsymbol{y}^t(N)$ or $\boldsymbol{y}_t(i)$. The s-dimensional state vector $\boldsymbol{\theta}_t^{\mathrm{T}} = (\boldsymbol{\theta}_t(1)^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_t(N)^{\mathrm{T}})$, 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 $\boldsymbol{\theta}_t(i)$, $i = 1, \ldots, N$, which may be functions of $\boldsymbol{y}^{t-1}(1), \ldots, \boldsymbol{y}^{t-1}(i)$, but not $\boldsymbol{y}^{t-1}(i+1), \ldots, \boldsymbol{y}^{t-1}(N)$. The s-dimensional vector $\boldsymbol{w}_t^{\mathrm{T}} = (\boldsymbol{w}_t(1)^{\mathrm{T}}, \ldots, \boldsymbol{w}_t(N)^{\mathrm{T}})$

where $\boldsymbol{w}_t(i)$ 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 $\boldsymbol{C}_0 = \text{blockdiag}(\boldsymbol{C}_0(1), \ldots, \boldsymbol{C}_0(N))$ 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 a DCGM, 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: $\{Y_{t1}(1), Y_{t2}(1)\}$, $\{Y_{t1}(2), Y_{t2}(2), Y_{t3}(2)\}$ and $\{Y_{t1}(3), Y_{t2}(3)\}$. 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 $\boldsymbol{Y}_t(3) = (Y_{t1}(3), Y_{t2}(3))^{\mathrm{T}}$. Separate observation equations (3) are specified for $\boldsymbol{Y}_t(1)$, $\boldsymbol{Y}_t(2)$ and $\boldsymbol{Y}_t(3)$: $\boldsymbol{F}_t(1)^{\mathrm{T}} = (\boldsymbol{F}_{t1}(1)^{\mathrm{T}}, \boldsymbol{F}_{t2}(1)^T)$ is a function of $\boldsymbol{y}^{t-1}(1)$; $\boldsymbol{F}_t(2)^{\mathrm{T}} = (\boldsymbol{F}_{t1}(2)^{\mathrm{T}}, \boldsymbol{F}_{t2}(2)^T, \boldsymbol{F}_{t3}(2)^{\mathrm{T}})$ is a function of $\boldsymbol{y}^{t-1}(2)$ and $\boldsymbol{F}_{t2}(2)$ is also a function of $\boldsymbol{y}_2^t(1)$; $\boldsymbol{F}_t(3)^{\mathrm{T}} = (\boldsymbol{F}_{t1}(3)^{\mathrm{T}}, \boldsymbol{F}_{t2}(3)^T)$ is a function of $\boldsymbol{y}^{t-1}(3)$ and $\boldsymbol{F}_{t1}(3)$ is also a function of $\boldsymbol{y}_1^t(2)$, while $\boldsymbol{F}_{t2}(3)$ is also a function of $\boldsymbol{y}_3^t(2)$.

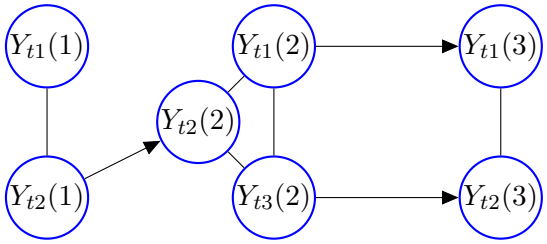

Figure 2: Chain graph to illustrate the DCGM

The following corollary follows on from a theorem presented in the Appendix, which also contains proofs of both theorem and corollary.

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 is a key result which 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 (5). Corollary 1, together with statements (1) and (2), mean that each parameter vector $\boldsymbol{\theta}_t(i)$ can be updated separately within the conditional multivariate model for $\boldsymbol{Y}_t(i) \mid \text{pa}(\boldsymbol{y}_t(i))$, and conditional forecasts for $\boldsymbol{Y}_t(i) \mid \text{pa}(\boldsymbol{y}_t(i))$ can be found

separately, since the joint forecast distribution can be expressed as

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

where $\mathrm{pa}(\boldsymbol{y}^{t-1}(i)) = (\mathrm{pa}(\boldsymbol{y}_1(i)), \ldots, \mathrm{pa}(\boldsymbol{y}_{t-1}(i)))$. 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.

The marginal forecasts for $\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(N)$ are required rather than forecasts conditional on parents. 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 & 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. 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 (3) 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 dynamic linear models (Quintana & West, 1987) to model each of $\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(N)$. As already mentioned, this model is conjugate and thus computationally simple and quick to use. However, although the matrix normal dynamic linear model 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 dynamic linear

model 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 $\text{pa}(y_{tj}(i))$, and $\text{pa}(y_{tj}(i))$ is not necessarily equal to $\text{pa}(y_{tk}(i))$, for $j \neq k$.

The simplest models for $\boldsymbol{Y}_t(1), \ldots, \boldsymbol{Y}_t(N)$ are dynamic linear models where $\boldsymbol{F}_{tj}(i)$, $j = 1, \ldots, r_i$, $i = 1, \ldots, N$, is a linear function of regressor(s) $\text{pa}(y_{tj}(i))$ and all distributions in (3)–(5) are normal. This is the *linear* dynamic chain graph model (LDCGM). In the next section, an LDCGM is used for forecasting road traffic flows.

3 Application: traffic flow forecasting

Anacleto *et al.* (2013a, 2013b) 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 3 shows a schematic diagram of the network: arrows 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 this network and the performance of the LDCGM and the LMDM is compared.

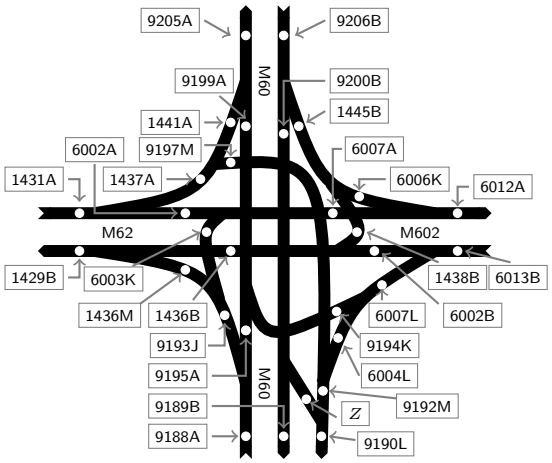

Figure 3: Schematic diagram of a traffic network near Manchester, UK

Let $Y_t(k)$ denote the traffic flow at site k at time t. Anacleto *et al.* (2013a) elicited a DAG to represent these traffic flow series. 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 mod-

elled 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, which joins root nodes pairwise by undirected edges representing an association structure, may be a more suitable representation of the flow series.

For clarity of presentation, consider a small subset of the Manchester network traffic flow series comprising the 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: 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.* (2013a), and undirected edges represent associations between pairs of root nodes.

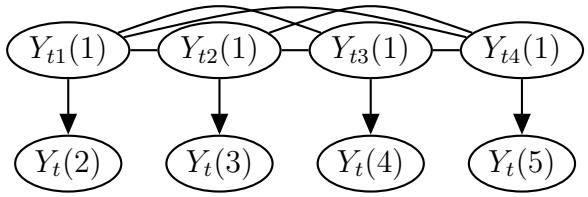

Figure 4: A chain graph for a subset of data collection sites of the Manchester network

3.1 An LDCGM for the traffic network

The chain graph in Figure 4 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 dynamic linear model is used to model $\boldsymbol{Y}_t(1)$, while conditional univariate dynamic linear models are used to model $Y_t(2), \ldots, Y_t(5)$.

A matrix normal dynamic linear model 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:

$$\begin{aligned} \boldsymbol{Y}_t(1)^{\mathrm{T}} &= \boldsymbol{F}_t(1)^{\mathrm{T}}\boldsymbol{\Theta}_t(1) + \boldsymbol{v}_t(1)^{\mathrm{T}}, \qquad \boldsymbol{v}_t(1) \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_t(1)), & (6) \\ Y_t(i) &= \boldsymbol{F}_t(i)^{\mathrm{T}}\boldsymbol{\theta}_t(i) + v_t(i), \qquad v_t(i) \sim N(0, V_t(i)), \quad i = 2, \ldots, 5. & (7) \end{aligned}$$

System equations:

$$\begin{aligned} \boldsymbol{\Theta}_t(1) &= \boldsymbol{G}_t(1)\boldsymbol{\Theta}_{t-1}(1) + \boldsymbol{\Omega}_t(1), \qquad \boldsymbol{\Omega}_t(1) \sim N(\boldsymbol{0}, \boldsymbol{W}_t(1), \boldsymbol{\Sigma}_t(1)), & (8) \\ \widetilde{\boldsymbol{\theta}}_t &= \widetilde{\boldsymbol{G}}_t\widetilde{\boldsymbol{\theta}}_{t-1} + \widetilde{\boldsymbol{w}}_t, \qquad \widetilde{\boldsymbol{w}}_t \sim N(\boldsymbol{0}, \widetilde{\boldsymbol{W}}_t), & (9) \end{aligned}$$

Initial information:

$$\begin{aligned} \boldsymbol{\Theta}_0(1) \mid D_0 &\sim N(\boldsymbol{m}_0, \boldsymbol{C}_0(1), \boldsymbol{\Sigma}_0(1)), & (10) \\ \widetilde{\boldsymbol{\theta}}_0 \mid D_0 &\sim N(\widetilde{\boldsymbol{m}}_0, \widetilde{\boldsymbol{C}}_0). & (11) \end{aligned}$$

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 $pa(y_t(i))$, $i = 2, \ldots, 5$; 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 $\boldsymbol{\Theta}_t(1)$; $\widetilde{\boldsymbol{G}}_t = \text{blockdiag}(\boldsymbol{G}_t(2), \ldots, \boldsymbol{G}_t(5))$ 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 $\widetilde{\boldsymbol{\theta}}_t$; $\widetilde{\boldsymbol{W}}_t = \text{blockdiag}(\boldsymbol{W}_t(2), \ldots, \boldsymbol{W}_t(5))$ 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), pp 108–112, 603–604. Conjugacy allows quick and easy computation.

To evaluate the effect of including an association structure across $Y_{t1}(1), \ldots, Y_{t4}(1)$ in the LDCGM, forecasts were also obtained using an LMDM with no such association structure. The associated graphical structure for this LMDM is the DAG obtained by removing the undirected edges from the chain graph in Figure 4, 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 (7), (9) and (11) in both the LMDM and the LDCGM, whereas in the LMDM each $Y_{tj}(1)$, $j = 1, \ldots, 4$, is modelled by a separate dynamic linear model of the form:

$$\textit{Obs. equation:} \quad Y_{tj}(1) = \boldsymbol{F}_{tj}(1)^{\mathrm{T}}\boldsymbol{\theta}_{tj}(1) + v_{tj}(1), \quad v_{tj}(1) \sim N(0, V_{tj}(1)) \tag{12}$$

$$\textit{Sys. equation:} \quad \boldsymbol{\theta}_{tj}(1) = \boldsymbol{G}_{tj}(1)\boldsymbol{\theta}_{t-1,j}(1) + \boldsymbol{w}_{tj}(1), \ \boldsymbol{w}_{tj}(1) \sim N(\boldsymbol{0}, \boldsymbol{W}_{tj}(1)) \tag{13}$$

$$\textit{Initial info.:} \quad \boldsymbol{\theta}_{0j}(1) \mid D_0 \sim N(\boldsymbol{m}_{0j}(1), \boldsymbol{C}_{0j}(1)). \tag{14}$$

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/). Following Anacleto *et al.* (2013a), 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 traffic during December. 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.* (2013b), cubic splines can model these daily patterns so that the regression vectors $\boldsymbol{F}_t(1)$ in (6) and $\boldsymbol{F}_{t1}(1), \ldots, \boldsymbol{F}_{t4}(1)$ in (12) contain fixed basis functions (estimated from historic data), while $\boldsymbol{\Theta}_t(1)$ in (6) and $\boldsymbol{\theta}_{t1}(1), \ldots, \boldsymbol{\theta}_{t4}(1)$ in (12) contain dynamically evolving spline parameters for individual series which are estimated sequentially online. In the matrix normal dynamic linear model, 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 (8) and $\boldsymbol{G}_{t1}(1), \ldots, \boldsymbol{G}_{t4}(1)$, in (13) 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 dynamic linear models are defined for $Y_t(2), \ldots, Y_t(5)$ where each $Y_t(i)$, $i = 2, \ldots, 5$, has $pa(y_t(i)) = y_{t,i-1}(1)$ as a linear regressor. The parameters for these regressors exhibit daily patterns, and, following Anacleto *et al.* (2013b), these can also be modelled by cubic splines so that $\boldsymbol{F}_t(i)$ and $\boldsymbol{\theta}_t(i)$ in (7) contain fixed basis

functions and dynamically evolving spline parameters, respectively. Matrix $\boldsymbol{G}_t(i)$ in (9) is an identity matrix.

In addition to traffic flow data, data for three other variables are available. These are (a) speed (in kph), (b) occupancy, the percentage of time that vehicles are 'occupying' the inductive loop, and (c) headway, the average time between vehicles passing over the inductive loop (in sec). In Anacleto *et al.* (2013b), flow forecasts using an LMDM were improved by using these variables at time $t-1$ as predictors in the flow model at time t. The relationships between each predictor variable at time $t-1$ and flow at t are non-linear and are modelled by cubic splines in Anacleto *et al.* (2013b) so that, as with splines modelling daily patterns, the basis functions and associated parameters are incorporated into vectors $\boldsymbol{F}_t(i)$ and $\boldsymbol{\theta}_t(i)$ in the LMDM.

In the LDCGM, however, the matrix normal dynamic linear model 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 speed, occupancy and headway at that site as predictors, without also including speed, occupancy and headway 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. There is no such restriction for either model for $Y_t(2), \ldots, Y_t(5)$, and so splines representing the relationships between speed, occupancy and headway at $t-1$ with flow at t at the same site, are incorporated into $\boldsymbol{F}_t(i)$ and $\boldsymbol{\theta}_t(i)$ in (7).

For both models, the observation variances $V_t(2), \ldots, V_t(5)$ in (7) are estimated online following techniques introduced in Anacleto *et al.* (2013a) to account for heterogeneity in traffic flow series. Briefly, each $V_t(i)$ is replaced by

$$V_t(i) = \exp\{\alpha \log[f_t(i)]\}\phi_t(i)^{-1} \tag{15}$$

where α is such that $\log(\text{variance of flow}) = \alpha \log(\text{mean flow})$ and can be estimated using historical data; $f_t(i)$ is the one-step ahead forecast mean for $Y_t(i)$; and $\phi_t(i)$ is the underlying observation precision which is estimated on-line using the discounting variance learning techniques described in West & Harrison (1997), page 359.

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 & Harrison, 1997, page 608) alone. For fairness of comparison, (15) is not used to model each $V_{tj}(1)$

in equation (12), so that the scalar observational variances of $Y_{t1}(1), \ldots, Y_{t4}(1)$ in the LMDM are also modelled using discounting techniques only. Prado & 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.

Matrices $\boldsymbol{W}_t(1), \ldots, \boldsymbol{W}_t(5)$ are estimated on-line using standard discounting techniques for evolution covariance matrices (see West & Harrison, 1997, page 193).

3.2 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.* (2013b) provides details of the LPL for the LMDM and this is easily adapted for the DCGM.

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 dynamic linear model for $\boldsymbol{Y}_t(1)$ used in the LDCGM provides better forecasts than the independent dynamic linear models 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.

Table 1: LPL values for the LMDM and the LDCGM

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

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 $(Y_{t1}(1), Y_{t2}(1))$ 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 predictors 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.

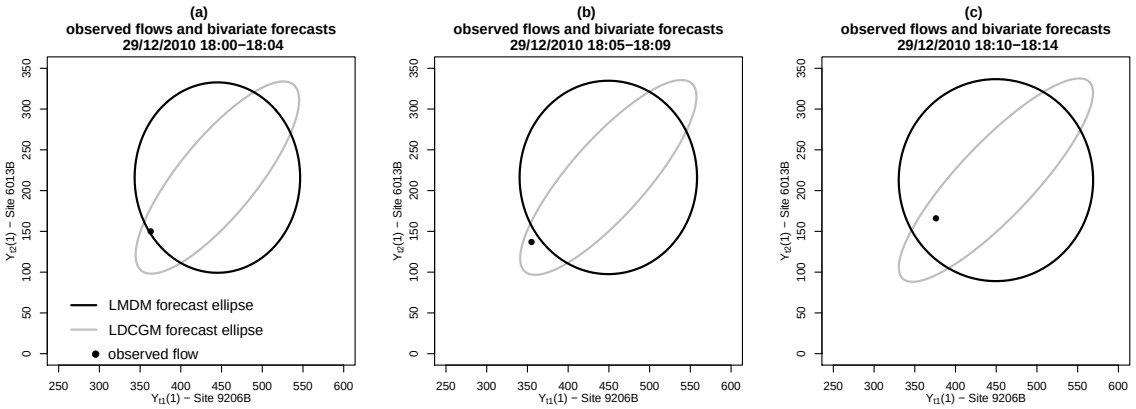

Figure 5: Observed flows $(\cdot)$ and bivariate forecast limits at a pair of root nodes.

The matrix normal dynamic linear model 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 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.

4 Potential application: gene expression modelling

The graph used to represent traffic flows in Section 3 was obtained using the direction of flows in the network as the causal driving mechanism across the time series. However, learning graphs from data is a key task when elicitation techniques are not available. This is one of the main challenges when analysing data on gene expressions.

Gene expression data represent how much of each gene has been translated into proteins, and expressions of different genes make cells perform different functions in biological systems. Understanding of biological systems can be enhanced by analysing high-dimensional time series of gene expression data. In particular, estimating the graphical representation of gene expressions not only can validate known associations between genes, but can also allow the discovery of new gene relationships.

Dynamic Bayesian networks (DBNs) have been extensively applied to model gene expression data, while methods such as LASSO (Meinshausen and Bühlmann, 2006) and score-based MCMC (Husmeier *et al.*, 2011) are commonly used to estimate the graphical structure. When using DBNs to estimate DAGs representing gene relationships, it is usually assumed that the parents of a given time series X_t take values at time period $t-1$ (see, for example Lèbre *et al.*, 2009). However, gene expression relationships can occur on different time scales: for example, there can be relationships occurring on a minute-by-minute level which current models may not capture when using hourly gene expression data. In such a scenario, the MDM is an ideal candidate model as it allows contemporaneous relationships in its DAG.

While current gene expression models only allow estimation of DAG structures, the relationships between genes can sometimes be better represented by a chain graph structure. An example of this occurs when the expression of two genes B and C are affected either by an external process, or the expression of a third gene A which is unavailable. This results in a symmetric association between the expression of genes B and C similar to that observed at the entrances of the Manchester network in Section 3. In this case, the DCGM is an ideal candidate model to estimate the chain graph structure. MCMC methods based on the DCGM are currently being implemented and their application on estimating gene graphs will be reported in a future paper.

5 Discussion

Application of the DCGM was illustrated here in the contexts of traffic flow and gene expression modelling. 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 the potential application areas.

The traffic flow time series analysed in this paper were obtained on a set of sites distributed over space, so that these time series 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 Winkle (2011) suggest that chain graphs are a natural template for representing such data. This suggests that the DCGM can be a potential candidate when modelling time series originating from spatio-temporal processes.

Acknowledgements

The authors thank Tom Michoel and Tom Freeman, both at the Roslin Institute, University of Edinburgh, for valuable discussions regarding gene expression modelling. The authors also thank the Highways Agency in England for providing the data used in this paper. Osvaldo Anacleto was a research student at the Open University while completing most of this work.

References

Anacleto, O. (2012). *Bayesian dynamic graphical models for high-dimensional flow forecasting in road traffic networks.* PhD thesis, The Open University, UK.

Anacleto, O., Queen, C.M. and Albers, C.J. (2013a). Multivariate forecasting of road traffic flows in the presence of heteroscedasticity and measurement errors. *Applied Statistics*, **62(2)**, 251–270.

Anacleto, O., Queen, C.M. and Albers, C.J. (2013b). Forecasting multivariate road traffic flows using Bayesian dynamic graphical models, splines and other traffic variables. *Australian and New Zealand Journal of Statistics*, **55(2)**, 69–86.

Carvalho, C.M. and West, M. (2007). Dynamic matrix-variate graphical models. *Bayesian analysis*, **2(1)**, 69–98.

Costa, L, Smith, J.Q. and Nicholls, T, (2013). On the selection of Multiregression Dynamic Models of fMRI networked time series. CRiSM Research Report, Department of Statistics, University of Warwick.

Cox, D.R. and Wermuth, N. (1996). *Multivariate Dependencies Models, Analysis and Interpretation.* London: Chapman and Hall.

Cressie N. and Wikle C. (2011). *Statistics for Spatio-Temporal Data.* Hoboken: John Wiley & Sons; 2011.

Dawid, A.P. (1981). Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika*, **68**, 265-74.

Husmeier, D., Werhli, A. V., Grzegorczyk, M. (2011). Advanced Applications of Bayesian Networks in Systems Biology In: *Handbook of Statistical Systems Biology* , eds. M. P. H. Stumpf, D. J. Balding and M.Girolami, John Wiley & Sons, London,270–289.

Ippoliti, L., Valentini, P. and Gamerman, D. (2012) Spacetime modelling of coupled spatiotemporal environmental variables. *Journal of the Royal Statistical Society, C*, **61**, 175-200.

Lauritzen, S.L. (1996). *Graphical Models*. Clarendon Press, Oxford.

Lavine, M., Hagland, M.M. and Hochman, D.W. (2011). Dynamic linear model analysis of optical imaging data acquired from the human neocortex. *Journal of Neuroscience Methods*, **199**, 346–362.

Li, B. (2009). A non-Gaussian kalman filter with application to the estimation of vehicular speed. *Technometrics*, **51**, 162-172.

Meinshausen, M. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, **34(3)**, 1436–1462.

Prado, R., and West, M. (2010). *Time Series: Modelling, Computation and Inference*. Chapman & Hall, New York.

Queen, C.M. (1994). Using the multiregression dynamic model to forecast the brand sales in a competitive market. *The Statistician*, **43(1)**, 87-98.

Queen, C.M. and Albers, C.J. (2009). Intervention and causality: forecasting traffic flows using a dynamic Bayesian network. *Journal of the American Statistical Association*, **104**, 669-681.

Queen, C.M. and Smith, J.Q. (1992). Dynamic graphical models. In *Bayesian Statistics 4*. Eds J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith. Oxford University Press.

Queen, C.M. and Smith, J.Q. (1993). Multiregression dynamic models. *Journal of the Royal Statistical Society, B*, **55(4)**, 849-870.

Queen, C.M., Wright, B.J. and Albers, C.J. (2007). Eliciting a directed acyclic graph for a multivariate time series of vehicle counts in a traffic network. *Australian and New Zealand Journal of Statistics*, **49(3)** , 221-239.

Queen, C.M., Wright, B.J. and Albers, C.J. (2008). Forecast covariances in the linear multiregression dynamic model. *Journal of Forecasting*, **27**, 175-191.

Quintana, J.M. and West, M. (1987). Multivariate time series analysis: new techniques applied to international exchange rate data. *The Statistician*, **36**, 275–281.

Quintana, J. M., Carvalho, C. M., Scott, J., and Costigliola, T. (2010). Futures markets, Bayesian forecasting and risk modeling. In *The Handbook of Applied Bayesian Analysis* (ed. A. OHagan and M. West), pp. 343-365. Oxford University Press, Oxford.

Schmidt, A.M. and Pereira, J.B.M. (2011). Modelling time series of counts in epidemiology. *International Statistical Review*, **79(1)**,48-69.

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*, **52**, No. 1, 21–50.

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

Appendix

For notational convenience, define

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

with associated state vectors for $\boldsymbol{X}_t(i)$ and $\boldsymbol{Z}_t(i)$, respectively,

$$\begin{aligned} \boldsymbol{\phi}_t(i)^{\mathrm{T}} &= (\boldsymbol{\theta}_t(1)^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_t(i-1)^{\mathrm{T}}) & i = 2, \ldots, N, \\ \boldsymbol{\psi}_t(i)^{\mathrm{T}} &= (\boldsymbol{\theta}_t(i+1)^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_t(N)^{\mathrm{T}}) & i = 1, \ldots, N-1. \end{aligned}$$

For $i = 1$, $\boldsymbol{X}_t(i)$ and $\boldsymbol{\phi}_t(i)$ are defined to be $\emptyset$, as are $\boldsymbol{Z}_t(i)$ and $\boldsymbol{\psi}_t(i)$ for $i = N$.

Theorem 1 *Let $\{\boldsymbol{Y}_t\}_{t\geq 1}$ be governed by a dynamic chain graph model and suppose that the following conditional independence statements hold at time $t-1$:*

$$\boldsymbol{\phi}_{t-1}(i) \perp\!\!\!\perp \boldsymbol{y}^{t-1}(i), \boldsymbol{z}^{t-1}(i) \mid \boldsymbol{x}^{t-1}(i), \qquad i = 2, \ldots, N, \tag{16}$$

$$\boldsymbol{\theta}_{t-1}(i) \perp\!\!\!\perp \boldsymbol{z}^{t-1}(i), \boldsymbol{\phi}_{t-1}(i) \mid \boldsymbol{x}^{t-1}(i), \boldsymbol{y}^{t-1}(i), \qquad i = 1, \ldots, N, \tag{17}$$

$$\boldsymbol{\psi}_{t-1}(i) \perp\!\!\!\perp \boldsymbol{\phi}_{t-1}(i), \boldsymbol{\theta}_{t-1}(i) \mid \boldsymbol{y}^{t-1}, \qquad i = 1, \ldots, N-1. \tag{18}$$

Then the following conditional independence statements must also be true:

$$\boldsymbol{\phi}_t(i) \perp\!\!\!\perp \boldsymbol{y}^t(i), \boldsymbol{z}^t(i) \mid \boldsymbol{x}^t(i), \qquad i = 2, \ldots, N, \tag{19}$$

$$\boldsymbol{\theta}_t(i) \perp\!\!\!\perp \boldsymbol{z}^t(i), \boldsymbol{\phi}_t(i) \mid \boldsymbol{x}^t(i), \boldsymbol{y}^t(i), \qquad i = 1, \ldots, N, \tag{20}$$

$$\boldsymbol{\psi}_t(i) \perp\!\!\!\perp \boldsymbol{\phi}_t(i), \boldsymbol{\theta}_t(i) \mid \boldsymbol{y}^t, \qquad i = 1, \ldots, N-1. \tag{21}$$

Proof of Theorem 1

To prove that if statements (16) to (18) are true, then so are statements (19) to (21), a chain graph will be elicited representing statements (16) to (18), together with the observation equation (3) and system equation (4). This chain graph will then be used to show that statements (19) to (21) also hold. In the chain graph, $\boldsymbol{Y}_t(i)$ is partitioned into three parts so that $\boldsymbol{Y}_{tj}(i)^* = (Y_{t1}(i), \ldots, Y_{t,j-1}(i))^{\mathrm{T}}$ and $\boldsymbol{Y}_{tj}(i)^+ = (Y_{t,j+1}(i), \ldots, Y_{tr_i}(i))^{\mathrm{T}}$: for $j = 1$, $\boldsymbol{Y}_{tj}(i)^* = \emptyset$, and for $j = r_i$, $\boldsymbol{Y}_{tj}(i)^+ = \emptyset$.

Figure 6 shows a chain graph representing the components of $\boldsymbol{y}^{t-1}$ and $\boldsymbol{\theta}_{t-1}$. For random vectors $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$, Dawid (1979) showed that

$$\boldsymbol{A} \perp\!\!\!\perp (\boldsymbol{B}, \boldsymbol{C}) \mid \boldsymbol{D} \Leftrightarrow \boldsymbol{A} \perp\!\!\!\perp \boldsymbol{B} \mid (\boldsymbol{C}, \boldsymbol{D}) \quad \text{and} \quad \boldsymbol{A} \perp\!\!\!\perp \boldsymbol{C} \mid \boldsymbol{D}.$$

Applying this result to hypothesis (17) implies that:

$$\boldsymbol{\theta}_{t-1}(i) \perp\!\!\!\perp \boldsymbol{\phi}_{t-1}(i) \mid \boldsymbol{y}^{t-1}, \tag{22}$$

$$\boldsymbol{\theta}_{t-1}(i) \perp\!\!\!\perp \boldsymbol{z}^{t-1}(i) \mid \boldsymbol{x}^{t-1}(i), \boldsymbol{y}^{t-1}(i). \tag{23}$$

Hence, hypothesis (18) and statement (22) imply omitting edges between $\{\boldsymbol{\theta}_{t-1}(k) \mid k \neq i\}$ and $\boldsymbol{\theta}_{t-1}(i)$, $i, k = 1, \ldots, N$. In addition, hypothesis (16) and statement (23) imply omitting edges between $\{\boldsymbol{y}^{t-1}(k) \mid k > i\}$ and $\boldsymbol{\theta}_{t-1}(i)$, for $i = 1, \ldots, N-1$, $k = 2, \ldots, N$.

Figure 7 shows a chain graph representing associations between components of $\boldsymbol{y}^{t-1}$ and $\boldsymbol{\theta}_t$, and between $\boldsymbol{\theta}_{t-1}$ and $\boldsymbol{\theta}_t$ based on system equation (4). The edges between $\boldsymbol{y}^{t-1}$ and $\boldsymbol{\theta}_{t-1}$ shown in Figure 6 have been omitted in Figure 7 for clarity. The restriction that $\boldsymbol{G}_t(i)$ and $\boldsymbol{W}_t(i)$ can be functions of $\boldsymbol{y}^{t-1}(1), \ldots, \boldsymbol{y}^{t-1}(i)$, but not $\boldsymbol{y}^{t-1}(i+1), \ldots, \boldsymbol{y}^{t-1}(N)$, justifies omitting edges between $\{\boldsymbol{y}^{t-1}(k) \mid k > i\}$ and $\{\boldsymbol{\theta}_t(i)\}$, $i = 1, \ldots, N-1$, $k = 2, \ldots, N$. The block

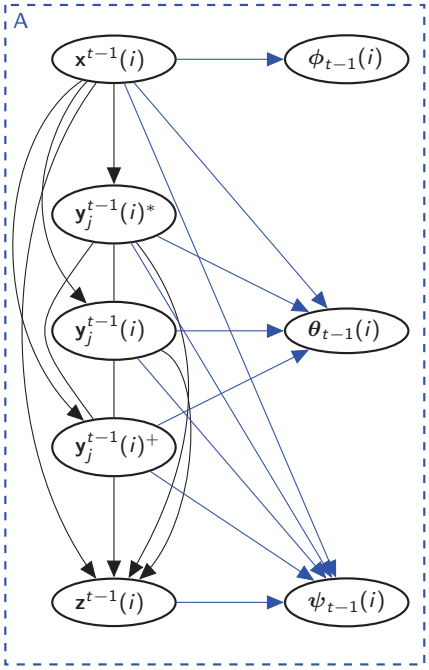

Figure 6: Chain graph for the inductive hypothesis (16) to (18) of Theorem 1.

diagonal forms of $\boldsymbol{G}_t$ and $\boldsymbol{W}_t$ imply that $\boldsymbol{\theta}_t(i) \perp\!\!\!\perp \{\boldsymbol{\theta}_{t-1} \backslash \boldsymbol{\theta}_{t-1}(i)\} \mid \boldsymbol{y}^{t-1}, \boldsymbol{\theta}_{t-1}(i)$, $i = 1, \ldots, N$, justifying omitting edges between $\{\boldsymbol{\theta}_{t-1}(k) \mid k \neq i\}$ and $\boldsymbol{\theta}_t(i)$, $i, k = 1, \ldots, N$.

Figure 8 shows a chain graph representing associations between components of $\boldsymbol{y}^{t-1}$ and $\boldsymbol{y}_t$, between $\boldsymbol{\theta}_{t-1}$ and $\boldsymbol{y}_t$, and between $\boldsymbol{\theta}_t$ and $\boldsymbol{y}_t$, based on observation equation (3). The edges between $\boldsymbol{y}^{t-1}$, $\boldsymbol{\theta}_{t-1}$ and $\boldsymbol{\theta}_t$ shown in Figs. 6 and 7 have been omitted for clarity. The observation equation ensures that,

$$\boldsymbol{y}_t(i) \perp\!\!\!\perp \boldsymbol{\theta}_{t-1} \mid \boldsymbol{x}^t(i), \boldsymbol{y}^{t-1}(i), \boldsymbol{\theta}_t(i), \qquad i = 1, \ldots, N,$$

justifying omitting edges between $\boldsymbol{\theta}_{t-1}(k)$ and $\boldsymbol{y}_t(i)$, $i, k = 1, \ldots, N$. Equation (3) also implies that

$$\boldsymbol{y}_t(i) \perp\!\!\!\perp \{\boldsymbol{\theta}_t \backslash \boldsymbol{\theta}_t(i)\} \mid \boldsymbol{x}^t(i), \boldsymbol{y}^{t-1}(i), \boldsymbol{\theta}_t(i), \qquad i = 1, \ldots, N,$$

justifying omitting edges between $\{\boldsymbol{\theta}_t(k) \mid k \neq i\}$ and $\boldsymbol{y}_t(i)$, $i, k = 1, \ldots, N$. Since $\boldsymbol{F}_t(i)$ is a known function of $\boldsymbol{x}^t(i)$ and $\boldsymbol{y}^{t-1}(i)$, but not $\boldsymbol{z}^{t-1}(i)$, then

$$\boldsymbol{y}_t(i) \perp\!\!\!\perp \boldsymbol{z}^{t-1}(i) \mid \boldsymbol{x}^t(i), \boldsymbol{y}^{t-1}(i), \qquad i = 1, \ldots, N.$$

This justifies omitting edges between $\{\boldsymbol{y}^{t-1}(k) \mid k > i\}$ and $\boldsymbol{y}_t(i)$, $i = 1, \ldots, N-1$, $k = 2, \ldots, N$.

The proof is completed by combining the preliminary chain graphs of Figs. 6 to 8 into a single chain graph, and then using the global Markov property for chain graphs (Frydenberg, 1990) to verify conditional independence statements (19) to (21). The resulting moralized chain graph is shown in Figure 9, in which (grey) undirected edges have been added between any two parents of a common child which are not already connected by an edge. Each conditional independence statement (19) to (21), of the form $\boldsymbol{A} \perp\!\!\!\perp \boldsymbol{B} \mid \boldsymbol{C}$, can now be verified by observing that the (conditioning) set $\boldsymbol{C}$ separates the (conditioned) sets $\boldsymbol{A}$ and $\boldsymbol{B}$. Figures 10 to 12 show the moralized chain graph of Figure 9 highlighting the conditioned components (orange and brown nodes) and the conditioning components (violet nodes) of each statement (19) to (21).

Proof of Corollary 1

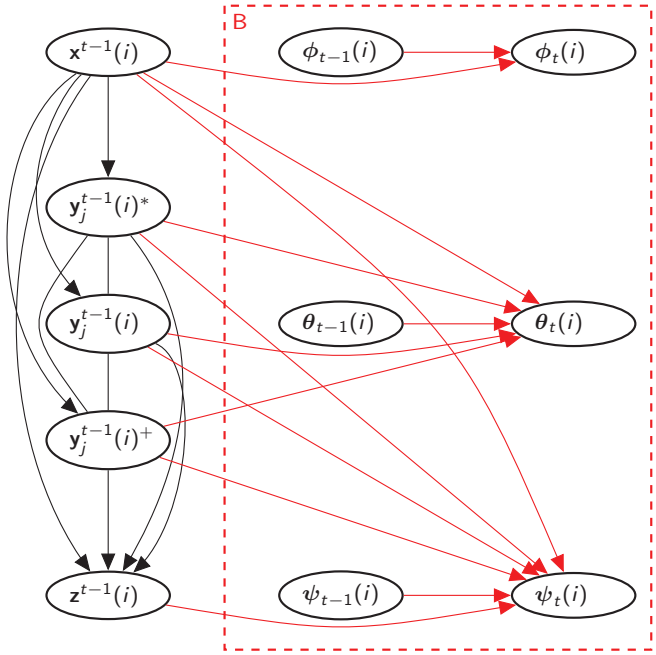

Figure 7: Chain graph representing system equation (4) of the dynamic chain graph model.

Figure 13 shows a chain graph representing the initial independence hypothesis $\perp\!\!\!\perp_{i=1}^{N} \boldsymbol{\theta}_0(i)$, (box A), system equation (4) at time 1 (box B) and observation equation (3) at time 1 (box C). The grey undirected edges in Figure 13 are the result of moralization.

From the moralized graph in Figure 13, the conditional independence statements (19) to (21) can be deduced for time $t = 1$. Therefore, when $\perp\!\!\!\perp_{i=1}^{N} \boldsymbol{\theta}_0(i)$, the conditional independence statements (19) to (21) are true for time $t = 1$ and, by induction from Theorem 1, conditional independence statements (19) to (21) must be true for all times $t \in \mathbb{N}$.

Conditional independence statements (19) to (21) can then be combined into the statement:

$$\boldsymbol{\theta}_t(i) \quad \perp\!\!\!\perp \quad \boldsymbol{z}^t(i), \boldsymbol{\phi}_t(i), \boldsymbol{\psi}_t(i) \mid \boldsymbol{x}^t(i), \boldsymbol{y}^t(i), \qquad i = 1, \ldots, N. \tag{24}$$

Then, again using Dawid's (1979) result that for random vectors $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$, $\boldsymbol{A} \perp\!\!\!\perp (\boldsymbol{B}, \boldsymbol{C}) \mid \boldsymbol{D} \Leftrightarrow \boldsymbol{A} \perp\!\!\!\perp \boldsymbol{B} \mid (\boldsymbol{C}, \boldsymbol{D})$ and $\boldsymbol{A} \perp\!\!\!\perp \boldsymbol{C} \mid \boldsymbol{D}$, statement (24) implies

$$\begin{aligned} &\boldsymbol{\theta}_t(i) \quad \perp\!\!\!\perp \quad \boldsymbol{\phi}_t(i), \boldsymbol{\psi}_t(i) \mid \boldsymbol{y}^t, && i = 1, \ldots, N, \\ &\boldsymbol{\theta}_t(i) \quad \perp\!\!\!\perp \quad \boldsymbol{y}^t(i+1), \ldots, \boldsymbol{y}^t(N) \mid \boldsymbol{y}^t(1), \ldots, \boldsymbol{y}^t(i), && i = 1, \ldots, N-1. \end{aligned} \tag{25}$$

Conditional independence statement (25) can be re-expressed as $\perp\!\!\!\perp_{i=1}^{N} \boldsymbol{\theta}_t(i) \mid \boldsymbol{y}^t$, which concludes the proof.

References

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

Frydenberg, M. (1990). The chain graph Markov property. *Scandinavian Journal of Statistics*, **17**, 333–353.

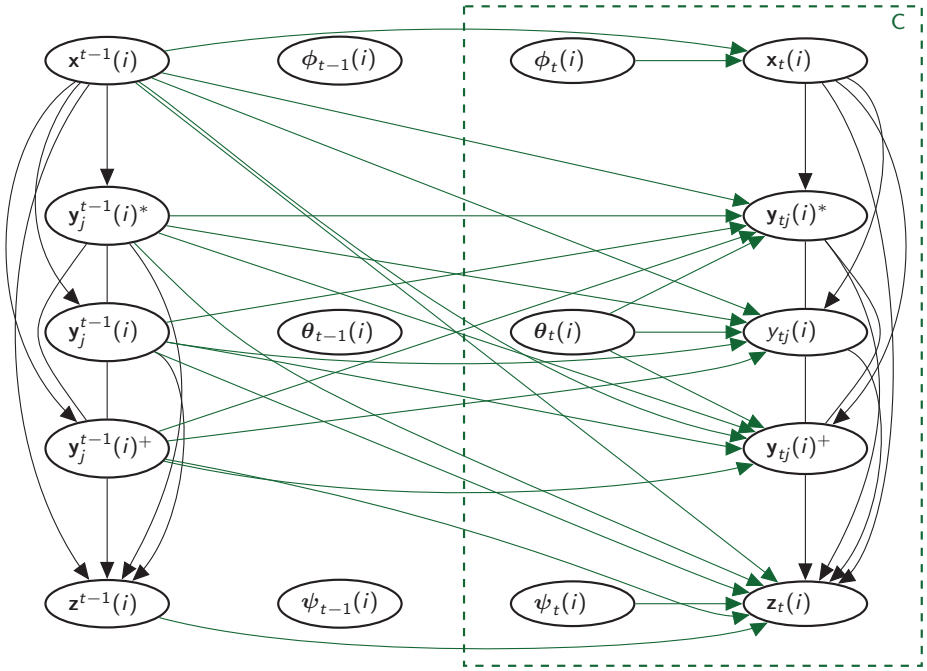

Figure 8: Chain graph representing the observation equation (3) of the dynamic chain graph model.

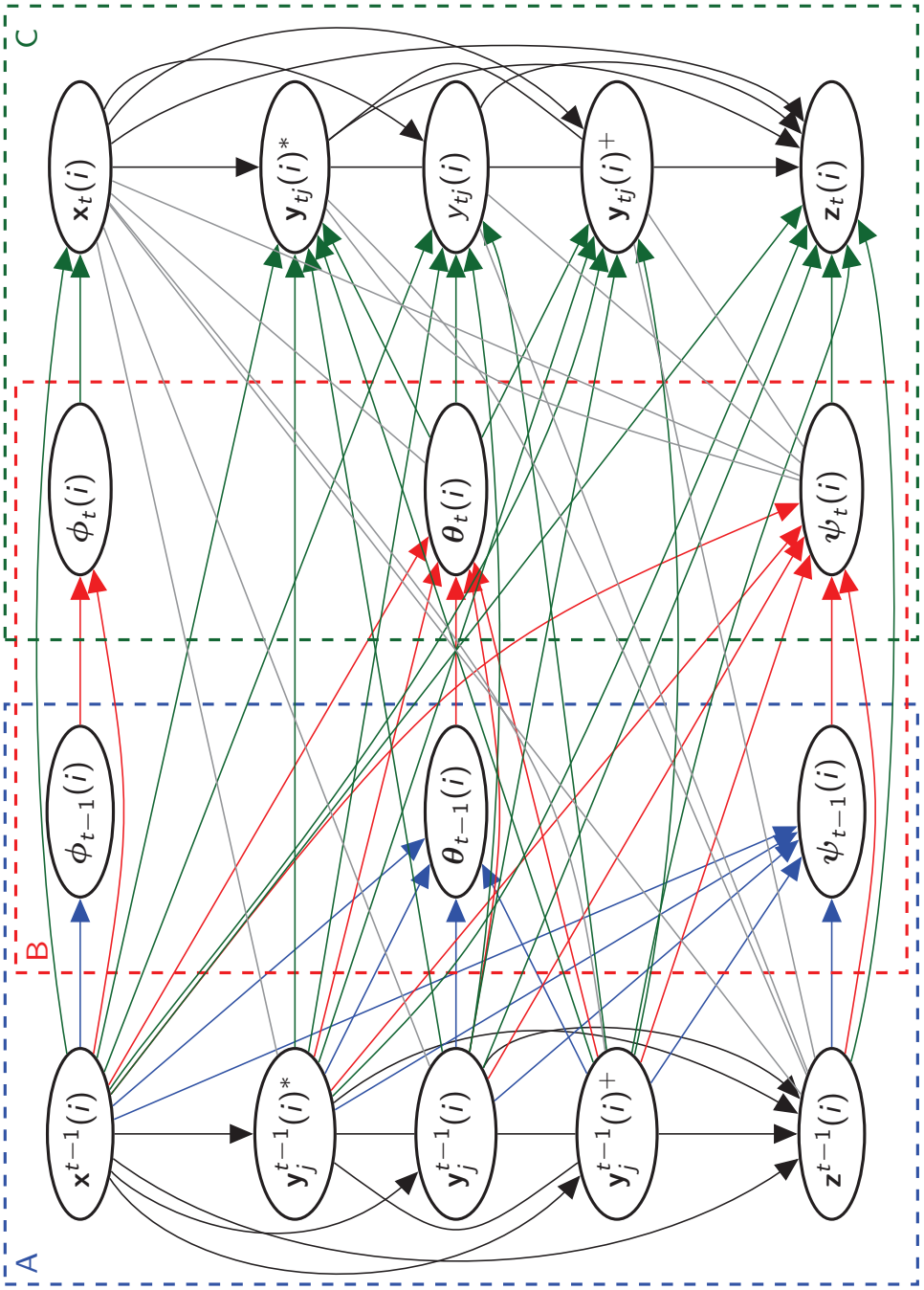

Figure 9: Moralized chain graph for the inductive hypotheses (16) to (18) (Box A), system equations (Box B) and observation equations (Box C) of the dynamic chain graph model.

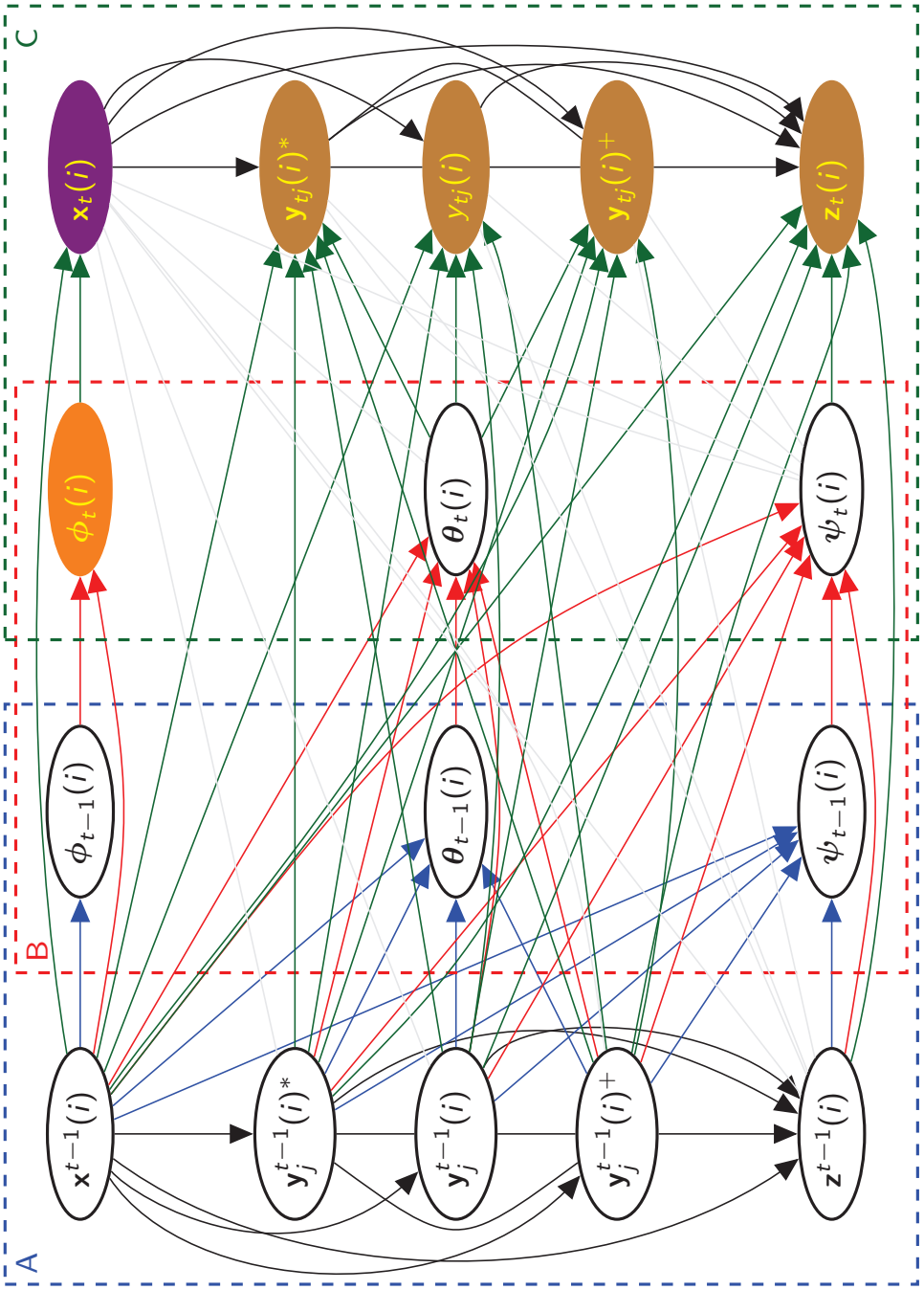

Figure 10: Moralized chain graph of Figure 9, highlighting statement (19), of the form $\boldsymbol{A} \perp\!\!\!\perp \boldsymbol{B} \mid \boldsymbol{C}$, from Theorem 1. $\boldsymbol{C}$ are the violet nodes and $\boldsymbol{A}$ and $\boldsymbol{B}$ are orange and brown nodes, respectively

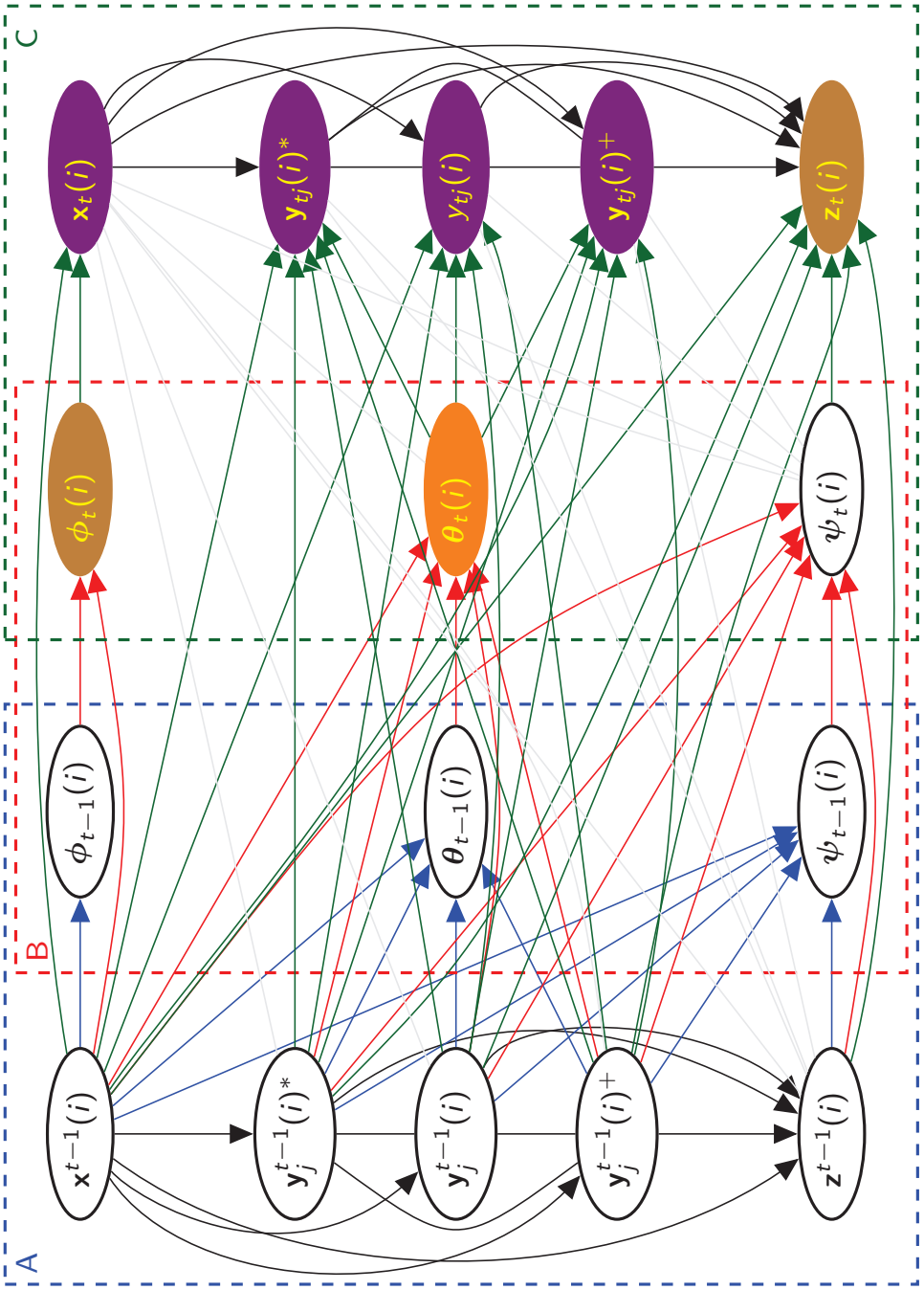

Figure 11: Moralized chain graph of Figure 9, highlighting statement (20), of the form $\boldsymbol{A} \perp\!\!\!\perp \boldsymbol{B} \mid \boldsymbol{C}$, from Theorem 1. $\boldsymbol{C}$ are the violet nodes and $\boldsymbol{A}$ and $\boldsymbol{B}$ are orange and brown nodes, respectively

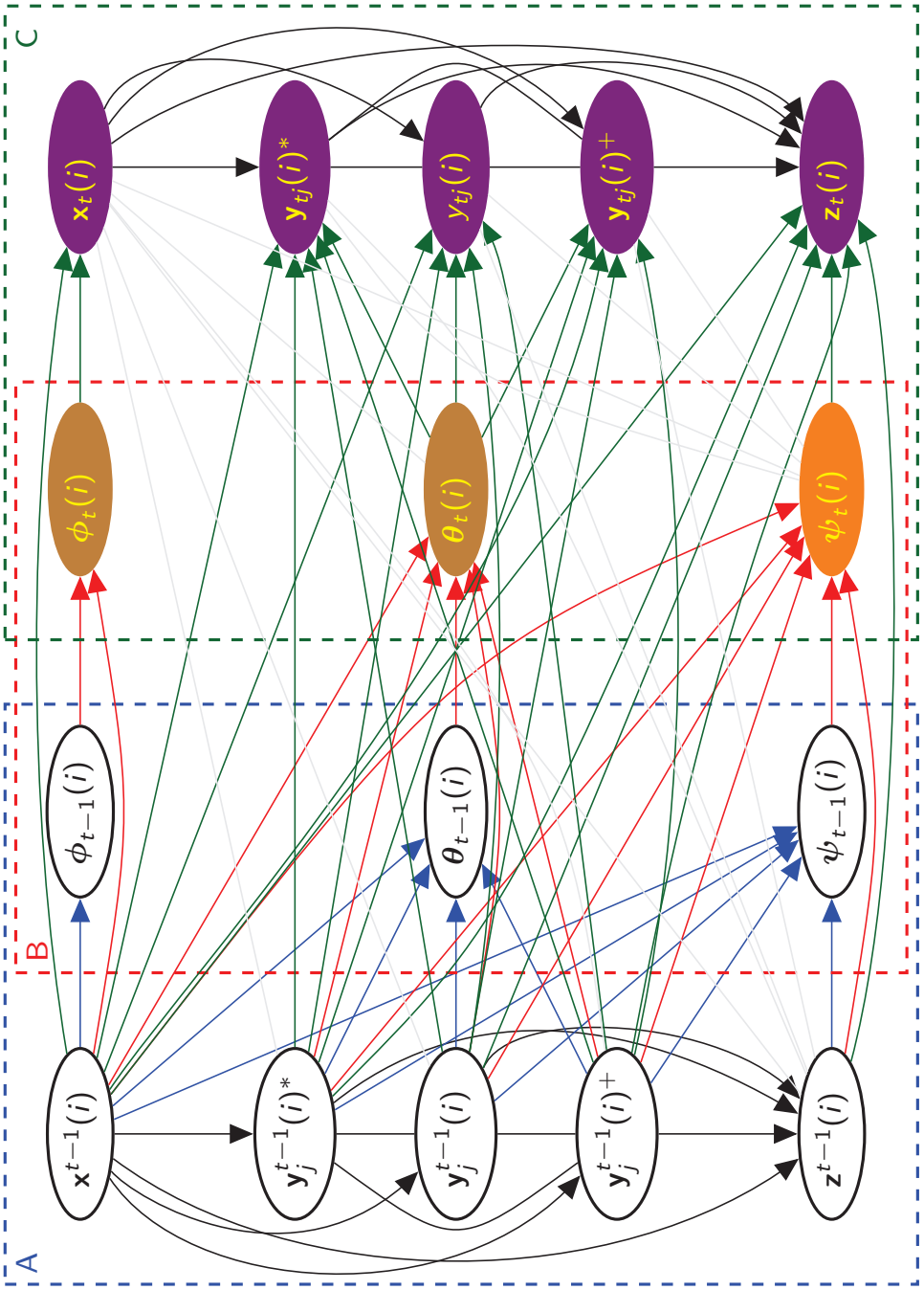

Figure 12: Moralized chain graph of Figure 9, highlighting statement (21), of the form $\boldsymbol{A} \perp\!\!\!\perp \boldsymbol{B} \mid \boldsymbol{C}$, from Theorem 1. $\boldsymbol{C}$ are the violet nodes and $\boldsymbol{A}$ and $\boldsymbol{B}$ are orange and brown nodes, respectively

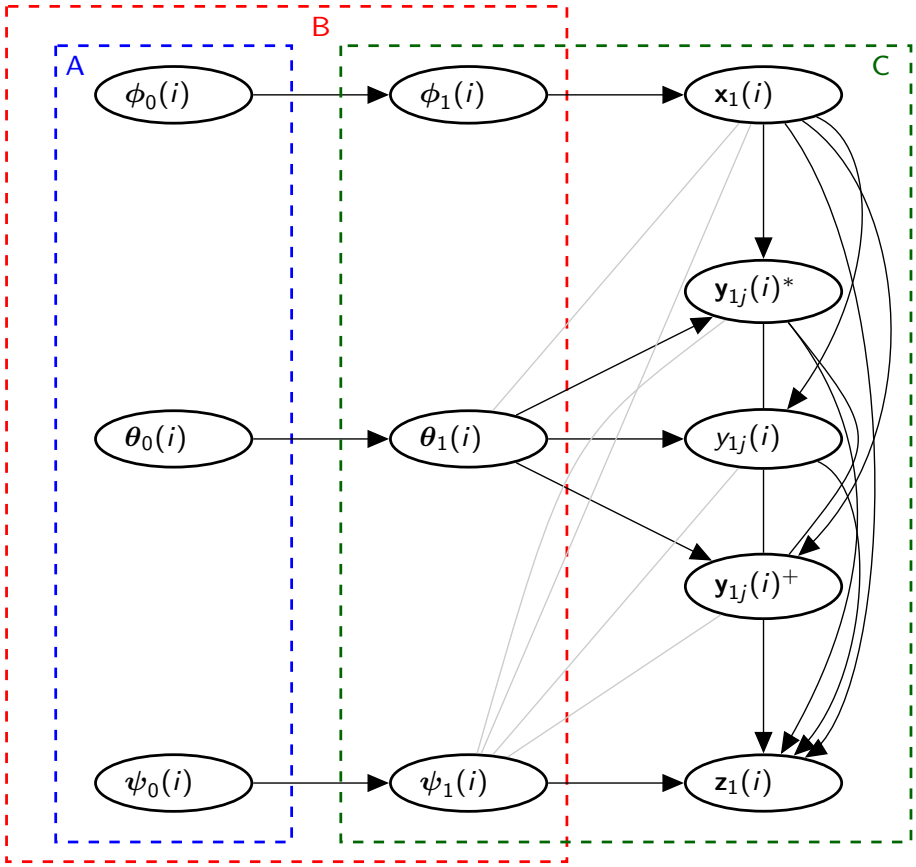

Figure 13: Moralized chain graph for the hypothesis of Corollary 1 (Box A), along with the system equation (Box B) and observation equation (Box C) of the dynamic chain graph model at time 1.