



University
of Glasgow

Grzegorzczuk, M., and Husmeier, D. (2009) *Avoiding spurious feedback loops in the reconstruction of gene regulatory networks with dynamic bayesian networks*. *Lecture Notes in Computer Science*, 5780 . pp. 113-124. ISSN 0302-9743

<http://eprints.gla.ac.uk/69429/>

Deposited on: 4 October 2012

Avoiding Spurious Feedback Loops in the Reconstruction of Gene Regulatory Networks with Dynamic Bayesian Networks

Marco Grzegorzcyk¹ and Dirk Husmeier²

¹ Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany
`Grzegorzcyk@statistik.tu-dortmund.de`

² Biomathematics and Statistics Scotland, JCMB, KB, Edinburgh EH9 3JZ, UK
`dirk@bioass.ac.uk`

Abstract. Feedback loops and recurrent structures are essential to the regulation and stable control of complex biological systems. The application of dynamic as opposed to static Bayesian networks is promising in that, in principle, these feedback loops can be learned. However, we show that the widely applied BGe score is susceptible to learning spurious feedback loops, which are a consequence of non-linear regulation and autocorrelation in the data. We propose a non-linear generalisation of the BGe model, based on a mixture model, and demonstrate that this approach successfully represses spurious feedback loops.

1 Introduction

In systems biology, there has been increased interest in learning regulatory networks and signalling pathways from postgenomic data. Following up on the seminal paper by Friedman et al. [1], Bayesian networks have been widely applied to this end. Their popularity partially stems from the tractability of the *marginal* likelihood of the network structure, which is the consistent scoring scheme for model selection in the Bayesian context. The practical computation requires the integration of the likelihood over the entire parameter space, though. To obtain a closed-form expression, two probabilistic models with their respective conjugate prior distributions have been employed in the past: the multinomial distribution with the Dirichlet prior, leading to the so-called BDe score [2], and the linear Gaussian distribution with the normal-Wishart prior, leading to the BGe score [3]. These approaches are restricted in that they either require the data to be discretised (BDe) or can only capture linear regulatory relationships (BGe). A non-linear non-discretised model based on heteroscedastic regression has been proposed by Imoto et al. [4]. However, this approach no longer allows the marginal likelihood to be obtained in closed-form and requires a restrictive approximation (the Laplace approximation) to be adopted. Another non-linear model based on node-specific Gaussian mixture models has been proposed [5]. Here, Ko et al. resort to the Bayesian information criterion BIC of Schwarz [6] for model selection, which is only a good approximation to the marginal likelihood in the limit of very large data sets. Recently, we proposed a generalisation

of the BGe score [8] based on a combination of a mixture model with the allocation sampler proposed in [7]. In this approach the latent variable allocation is common to the whole network, though, which results in a heterogeneous linear rather than a proper non-linear model. Our work aims to generalise our earlier work [8] by the introduction of node-specific change-points. This model is similar to the model by Ko et al. [5], with the difference that the marginal likelihood is computed properly, rather than approximated by BIC.

2 Problems of the BGe Score

When the objective is to infer regulatory networks from time series, as is typically the case in systems biology, the restriction of the model to linear processes can result in the prediction of spurious feedback loops. Consider the simple example shown in Figure 1. The graph shows two interacting nodes. Node X is a regulator of node Y , and it also has a regulatory feedback loop acting back on itself. Node Y is regulated by node X , but does not contain a feedback loop. The figure shows both the state space representation, i.e. the recurrent graph, and the corresponding dynamic Bayesian network. Note that the latter is a valid DAG obtained by the standard procedure of unfolding the state space graph in time. First assume that the data generation processes are consistent with the BGe model assumption, e.g.: $X(t+1) = X(t) + c + \sigma_x \cdot \phi_X(t)$ and $Y(t+1) = w \cdot X(t) + m + \sigma_y \cdot \phi_Y(t)$ where $w, m, c, \sigma_x, \sigma_y$ are constants, and $\phi(\cdot)$ are iid normally distributed random variables. Under fairly general regularity conditions, the marginal likelihood and, hence, the BGe score is a consistent estimator. This implies that the correct model structure will be learned as $m \rightarrow \infty$, where m is the data set size. Next, consider the scenario of a non-linear regulatory influence that X exerts on Y :

$$X(t+1) = X(t) + c + \sigma_x \cdot \phi_X(t), \quad Y(t+1) = f(X(t)) + \sigma_y \cdot \phi_Y(t) \quad (1)$$

for some non-linear function $f(\cdot)$. This non-linear function cannot be modelled with a linear Bayesian network based on the BGe model. Consequently, the prediction of $Y(t+1)$ from $X(t)$ will tend to be poor. Note that for sufficiently small noise levels, the $Y(t)$'s will exhibit a strong autocorrelation, by virtue of the autocorrelation of the $X(t)$'s, and the regulatory influence of $X(t)$ on $Y(t+1)$. If

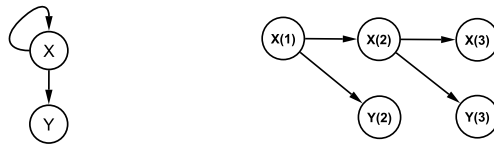


Fig. 1. State space graph and corresponding dynamic Bayesian network. Left: Recurrent state space graph containing two nodes. Node X has a recurrent feedback loop and acts as a regulator of node Y . Right: The same graph unfolded in time.

the latter regulatory influence cannot be learned owing to the linear restriction of our model, the next best explanation is a direct modelling of the autocorrelation between the $Y(t)$'s themselves. This autocorrelation corresponds to a feedback loop of Y acting back on itself in the state-space graph, or, equivalently, an edge from $Y(t)$ to $Y(t + 1)$ in the dynamic Bayesian network. We would therefore conjecture that the linear restriction of the Bayesian network model may result in the prediction of spurious feedback loops and, hence, to the reconstruction of wrong network structures. Ruling out feedback loops altogether, as we did in [8], will not provide a sufficient remedy for this problem, as some nodes – X in the example above – will exhibit regulatory feedback loops (e.g. in molecular biology: transcription factors regulating their own transcription), and it is generally not known in advance where these nodes are.

3 Methodology

3.1 The Dynamical BGe Network

Dynamical Bayesian networks (DBNs) are flexible models for representing probabilistic relationships between interacting variables X_1, \dots, X_N . The graph \mathcal{G} of a DBN describes the relationships between the variables, which have been measured at equidistant time points $t = 1, \dots, m$, in the form of conditional probability distributions. An edge pointing from X_i to X_j means that the realisation of X_j at time point t , symbolically: $X_j(t)$, is influenced by the realisation of X_i at time point $t - 1$, symbolically: $X_i(t - 1)$. $\pi_n = \pi_n(\mathcal{G})$ denotes the parent node set of node X_n in \mathcal{G} , i.e. the set of all nodes from which an edge points to node X_n in \mathcal{G} . Given a data set \mathcal{D} , where $\mathcal{D}_{n,t}$ and $\mathcal{D}_{(\pi_n,t)}$ are the t th realisations $X_n(t)$ and $\pi_n(t)$ of X_n and π_n , respectively, DBNs are based on the following homogeneous Markov chain expansion:

$$P(\mathcal{D}|\mathcal{G}, \boldsymbol{\theta}) = \prod_{n=1}^N \prod_{t=2}^m P(X_n(t) = \mathcal{D}_{n,t} | \pi_n(t-1) = \mathcal{D}_{(\pi_n,t-1)}, \boldsymbol{\theta}_n) \quad (2)$$

where $\boldsymbol{\theta}$ is the total parameter vector, composed of subvectors $\boldsymbol{\theta}_n$, which specify the local conditional distributions in the factorisation. The BGe model [3] specifies the distributional form $P(\mathcal{D}|\mathcal{G}, \boldsymbol{\theta} = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}\})$ as multivariate Gaussian distribution with expectation vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and assumes a normal-Wishart distribution as prior distribution $P(\{\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}\})$. The local probability distributions $P(X_n | \pi_n, \boldsymbol{\theta}_n)$ are then given by conditional Gaussian distributions so that only linear relationships between X_n and its parent nodes π_n can be modelled. Under fairly weak conditions imposed on the parameters $\boldsymbol{\theta}$ (prior independence and modularity) and prior distribution $P(\boldsymbol{\theta})$ (conjugacy), the parameters can be integrated out analytically, as shown by Geiger and Heckerman [3], and the marginal likelihood satisfies the same expansion rule as the Bayesian network with fixed parameters:

$$P(\mathcal{D}|\mathcal{G}) = \prod_{n=1}^N \prod_{t=2}^m P(X_n(t) = \mathcal{D}_{n,t} | \pi_n(t-1) = \mathcal{D}_{(\pi_n,t-1)}) = \prod_{n=1}^N \Psi(\mathcal{D}_n^{\pi_n}) \quad (3)$$

where $\mathcal{D}_n^{\pi_n} := \{(\mathcal{D}_{n,t}, \mathcal{D}_{\pi_n, t-1}) : 2 \leq t \leq m\}$ is the subset of the data pertaining to node X_n and parent set π_n . For the BGe model the factors $\Psi(\mathcal{D}_n^{\pi_n})$ can be computed according to Eqn. (15) and (24) in [3].

3.2 The New Mixture/Change-Point BGe Model

To obtain a more flexible model for DBNs we generalise Eq. (2) with a node-specific mixture model:

$$P(\mathcal{D}|\mathcal{G}, \mathbf{V}, \mathbf{K}, \boldsymbol{\theta}) = \prod_{n=1}^N \prod_{t=2}^m \prod_{k=1}^{\mathcal{K}_n} P(X_n(t) = \mathcal{D}_{n,t} | \pi_n(t-1) = \mathcal{D}_{(\pi_n, t-1)}, \boldsymbol{\theta}_n^k)^{\delta_{V_n(t), k}} \quad (4)$$

where $\delta_{V_n(t), k}$ is the Kronecker delta, \mathbf{V} is a matrix of latent variables $V_n(t)$, $V_n(t) = k$ indicates that the t th realisation of node X_n , symbolically $X_n(t)$, has been generated by the k th component of a mixture with \mathcal{K}_n components, and $\mathbf{K} = (\mathcal{K}_1, \dots, \mathcal{K}_n)$. Note that the matrix \mathbf{V} divides the data into several disjointed subsets, each of which can be regarded as pertaining to a separate BGe model with parameters $\boldsymbol{\theta}_n^k$. The probability model defined in Eq.(4) is effectively a mixture model with local probability distributions $P(X_n | \pi_n, \boldsymbol{\theta}_n)$ and it can hence, in principle, approximate any probability distribution arbitrarily closely. The vectors \mathbf{V}_n are node-specific, i.e. different nodes can have different break-points, so that the proposed model has a higher flexibility in modelling non-linear relationships than the BGM model proposed in [8]. Different from the free allocation of latent variables in [8], we change the assignment of data points to mixture components via a change-point process. This effectively reduces the complexity of the latent variable space and incorporates our prior belief that, in a time series, adjacent time points are likely to be assigned to the same component. Conditional on the latent variables \mathbf{V} and under the assumption that the regularity conditions defined in Geiger and Heckerman [3] are satisfied, the marginal likelihood can be computed in closed form:

$$P(\mathcal{D}|\mathcal{G}, \mathbf{V}, \mathbf{K}) = \int P(\mathcal{D}|\mathcal{G}, \mathbf{V}, \mathbf{K}, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta} = \prod_{n=1}^N \prod_{k=1}^{\mathcal{K}_n} \Psi(\mathcal{D}_n^{\pi_n}[k, \mathbf{V}_n]) \quad (5)$$

where $\mathcal{D}_n^{\pi_n}[k, \mathbf{V}_n] := \{(\mathcal{D}_{n,t}, \mathcal{D}_{\pi_n, t-1}) : V_n(t) = k, 2 \leq t \leq m\}$ is the subset of the data pertaining to node X_n and its parents π_n that has been assigned to the k th mixture component, symbolically: $V_n(t) = k$. In the absence of genuine prior knowledge about the regulatory network structure, we assume for $P(\mathcal{G})$ a uniform distribution on graphs, subject to a fan-in restriction of $|\pi_n| \leq 3$.

As prior probability distribution on the number of mixture components \mathcal{K}_n , $P(\mathcal{K}_n)$, we take an iid truncated Poisson distribution with shape parameter $\lambda = 1$ restricted to $1 \leq \mathcal{K}_n \leq 10$. We assume that the prior distributions $P(\mathbf{V}_n | \mathcal{K}_n)$ of the latent variable vectors \mathbf{V}_n conditional on \mathcal{K}_n are iid, and we identify \mathcal{K}_n with $\mathcal{K}_n - 1$ break-points: $\mathbf{b}_n = \{b_{n,1}, \dots, b_{n, \mathcal{K}_n - 1}\}$ on the continuous interval $[2, m]$. For notational convenience we introduce the pseudo-break-points $b_{n,0} = 2$

and $b_{n,\mathcal{K}_n} = m$. For node X_n the observation at time point t is assigned to the k th component, symbolically $V_n(t) = k$, if $b_{n,k-1} \leq t < b_{n,k}$. Following [10] we assume that the break-points are distributed as the even-numbered order statistics of $\mathcal{L} := 2(\mathcal{K}_n - 1) + 1$ points $u_1, \dots, u_{\mathcal{L}}$ uniformly and independently distributed on the interval $[2, m]$. The motivation for this prior, instead of taking \mathcal{K}_n uniformly distributed points, is to encourage *a priori* an equal spacing between the break-points, i.e. to discourage mixture components that contain only a few observations. The joint probability distribution of the proposed mixture BGe model is given by:

$$\begin{aligned} P(\mathcal{G}, \mathbf{V}, \mathbf{K}, \mathcal{D}) &= P(\mathcal{D}|\mathcal{G}, \mathbf{V}, \mathbf{K}) \cdot P(\mathcal{G}) \cdot P(\mathbf{V}|\mathbf{K}) \cdot P(\mathbf{K}) \\ &= P(\mathcal{G}) \cdot \prod_{n=1}^N \left\{ P(\mathbf{V}_n|\mathcal{K}_n) \cdot P(\mathcal{K}_n) \cdot \prod_{k=1}^{\mathcal{K}_n} \Psi(\mathcal{D}_n^{\pi_n}[k, \mathbf{V}_n]) \right\} \end{aligned} \quad (6)$$

Here $P(\mathcal{G})$ is the graph prior, and $P(\mathcal{K}_n)$ the Poisson prior on the number of mixture components for the n th node. The local marginal likelihood terms $\Psi(\mathcal{D}_n^{\pi_n}[k, \mathbf{V}_n])$, which result from Eq.(5), can be computed independently for each k using Eqn. (15) and (24) in [3]. Note that each vector \mathbf{V}_n acts as a filter which divides the data of X_n into \mathcal{K}_n different compartments, for which separate independent BGe scores can be computed in closed-form. When a mixture component is empty, then $\Psi(\mathcal{D}_n^{\pi_n}[k, \mathbf{V}_n]) = 1$. The term $P(\mathbf{V}|\mathbf{K}) = \prod_{n=1}^N \{P(\mathbf{V}_n|\mathcal{K}_n)$ is the prior distribution on the node-specific allocation vectors \mathbf{V}_n , which is induced by the even-numbered order statistics prior on the break-point locations \mathbf{b}_n . Deriving a closed-form expression is involved. However, the MCMC scheme we discuss in the next section does not sample \mathbf{V}_n directly, but is based on local modifications of \mathbf{V}_n based on birth, death and reallocation moves. All that is required for the acceptance probabilities of these moves are $P(\mathbf{V}_n|\mathcal{K}_n)$ ratios, which are straightforward to compute.

3.3 MCMC Inference

We now describe an MCMC algorithm to obtain a sample $\{\mathcal{G}^i, \mathbf{V}^i, \mathbf{K}^i\}_{i=1, \dots, I}$ from the posterior distribution $P(\mathcal{G}, \mathbf{V}, \mathbf{K}|\mathcal{D}) \propto P(\mathcal{G}, \mathbf{V}, \mathbf{K}, \mathcal{D})$ of Eq. (6). We combine the structure MCMC algorithm for Bayesian networks of Madigan and York [9] with the change-point model of Green [10], and draw on the fact that conditional on the allocation vectors \mathbf{V} , the model parameters can be integrated out to obtain the marginal likelihood terms $\Psi(\mathcal{D}_n^{\pi_n}[k, \mathbf{V}_n])$. Note that this approach is equivalent to the idea underlying the allocation sampler proposed in [7]. The resulting algorithm is effectively an RJMCMC scheme [10] in the discrete space of network structures and latent allocation vectors, where the Jacobian in the acceptance criterion is always 1 and can be omitted. With probability $p_G = 0.5$ we perform a structure MCMC move on the current graph \mathcal{G}^i and leave the latent variable matrix and the numbers of mixture components unchanged, symbolically: $\mathbf{V}^{i+1} = \mathbf{V}^i$ and $\mathbf{K}^{i+1} = \mathbf{K}^i$. A new candidate graph \mathcal{G}^{i+1} is randomly drawn out of the set of graphs $\mathcal{N}(\mathcal{G}^i)$ that can be reached from the

current graph \mathcal{G}^i by deletion or addition of a single edge. The proposed graph \mathcal{G}^{i+1} is accepted with probability:

$$A(\mathcal{G}^{i+1}|\mathcal{G}^i) = \min \left\{ 1, \frac{P(\mathcal{D}|\mathcal{G}^{i+1}, \mathbf{V}^i, \mathbf{K}^i)}{P(\mathcal{D}|\mathcal{G}^i, \mathbf{V}^i, \mathbf{K}^i)} \cdot \frac{P(\mathcal{G}^{i+1})}{P(\mathcal{G}^i)} \cdot \frac{|\mathcal{N}(\mathcal{G}^i)|}{|\mathcal{N}(\mathcal{G}^{i+1})|} \right\} \quad (7)$$

where $|\cdot|$ is the cardinality and the marginal likelihood terms have been specified in Eq. (5). The graph is left unchanged, symbolically $\mathcal{G}^{i+1} := \mathcal{G}^i$, if the move is not accepted. We note that the subsequent analysis will be based on the marginal posterior probabilities of individual edges, which can be estimated for each edge from the MCMC sample $\mathcal{G}^1, \dots, \mathcal{G}^I$ by the fraction of graphs in the sample that contain the edge of interest. With the complementary probability $1 - p_G$ we leave the graph \mathcal{G}^i unchanged, and perform a move on $(\mathbf{V}^i, \mathbf{K}^i)$, where \mathbf{V}_n^i is the latent variables vector of X_n in \mathbf{V}^i , and $\mathbf{K}^i = (\mathcal{K}_1^i, \dots, \mathcal{K}_N^i)$. We randomly select a node X_n and change its current number of components \mathcal{K}_n^i via a break-point birth or death move, or its latent variable vector \mathbf{V}_n^i by a break-point re-allocation move. The break-point birth (death) move increases (decreases) \mathcal{K}_n^i by 1 and may also have an effect on \mathbf{V}_n^i . The break-point reallocation move leaves \mathcal{K}_n^i unchanged and may have an effect on \mathbf{V}_n^i . If with probability $(1 - p_G)/N$ a break-point move on $(\mathcal{K}_n^i, \mathbf{V}_n^i)$ is performed, we randomly draw the move type. Under fairly mild regularity conditions (ergodicity, detailed balance), discussed in [10], the MCMC sampling scheme converges to the desired posterior distribution. To ensure detailed balance, the acceptance probabilities for the three break-point moves $(\mathcal{K}_n^i, \mathbf{V}_n^i) \rightarrow (\mathcal{K}_n^{i+1}, \mathbf{V}_n^{i+1})$ are chosen of the form $\min(1, R)$, see [10], with

$$R = \frac{\prod_{k=1}^{\mathcal{K}_n^{i+1}} \Psi(\mathcal{D}_n^{\pi_n}[k, \mathbf{V}_n^{i+1}])}{\prod_{k=1}^{\mathcal{K}_n^i} \Psi(\mathcal{D}_n^{\pi_n}[k, \mathbf{V}_n^i])} \times A \times B \quad (8)$$

where $A = P(\mathbf{V}_n^{i+1}|\mathcal{K}_n^{i+1})P(\mathcal{K}_n^{i+1})/P(\mathbf{V}_n^i|\mathcal{K}_n^i)P(\mathcal{K}_n^i)$ is the prior probability ratio, and B is the inverse proposal probability ratio. The exact form of these factors depends on the move type. (i) For a break-point reallocation (r) we randomly select one of the existing break-points $b_{n,j} \in \{b_{n,1}, \dots, b_{n,\mathcal{K}_n-1}\}$, and the replacement value $b_{n,j}^\dagger$ is drawn from a uniform distribution on $[b_{n,j-1}, b_{n,j+1}]$ where $b_{n,0} = 2$ and $b_{n,\mathcal{K}_n} = m$. Hence, the proposal probability ratio is one, the prior probabilities $P(\mathcal{K}_n^{i+1}) = P(\mathcal{K}_n^i)$ cancel out, and the remaining prior probability ratio $P(\mathbf{V}_n^{i+1}|\mathcal{K}_n^{i+1})/P(\mathbf{V}_n^i|\mathcal{K}_n^i)$ can be obtained from p.720 in [10]:

$$A_r = \frac{(b_{n,j+1} - b_{n,j}^\dagger)(b_{n,j}^\dagger - b_{n,j-1})}{(b_{n,j+1} - b_{n,j})(b_{n,j} - b_{n,j-1})}, \quad B_r = 1 \quad (9)$$

If there is no break-point ($\mathcal{K}_n^i = 1$) the move is rejected and the Markov chain is left unchanged. (ii) If a break-point birth move (b) on \mathcal{K}_n^i is proposed, the location of the new break-point b^\dagger is randomly drawn from a uniform distribution on the interval $[2, m]$; the proposal probability for this move is $b_{\mathcal{K}_n^i}/(m - 2)$, where $b_{\mathcal{K}_n^i}$ is the (\mathcal{K}_n^i -dependent) probability of selecting a birth move. The reverse death move, which is selected with probability $d_{(\mathcal{K}_n^i+1)}$, consists in discarding

randomly one of the $\mathcal{K}_n^i - 1 + 1 = \mathcal{K}_n^i$ change-points. The inverse proposal probability ratio is thus given by $B = d_{(\mathcal{K}_n^i+1)}(m-2)/b_{\mathcal{K}_n^i} \mathcal{K}_n^i$. The prior probability ratio is given by the expression at the bottom of p.720 in [10] (slightly modified to allow for the fact that \mathcal{K}_n components correspond to $\mathcal{K}_n - 1$ break-points), and we get:

$$A_b = \frac{P(\mathcal{K}_n^i+1)}{P(\mathcal{K}_n^i)} \frac{2\mathcal{K}_n^i(2\mathcal{K}_n^i+1)}{(m-2)^2} \frac{(b_{n,j+1}-b^\dagger)(b^\dagger-b_{n,j})}{(b_{n,j+1}-b_{n,j})}, \quad B_b = \frac{d_{(\mathcal{K}_n^i+1)}(m-2)}{b_{\mathcal{K}_n^i} \mathcal{K}_n^i} \quad (10)$$

For $\mathcal{K}_n^i = \mathcal{K}_{MAX}$ the birth of a new break-point is invalid and the Markov chain is left unchanged. Note that the ratio of the proposal probabilities for birth versus death moves $d_{(\mathcal{K}_n^i+1)}/b_{\mathcal{K}_n^i}$ can be chosen such that it cancels out against the prior ratio $P(\mathcal{K}_n^i+1)/P(\mathcal{K}_n^i)$, and the expression simplifies:

$$A_b B_b = \frac{2(2\mathcal{K}_n^i+1)}{(m-2)} \frac{(b_{n,j+1}-b^\dagger)(b^\dagger-b_{n,j})}{(b_{n,j+1}-b_{n,j})} \quad (11)$$

(iii) A break-point death move (d) is the reverse of the birth move, and we get:

$$A_d B_d = \frac{(m-2)}{2(2\mathcal{K}_n^i-1)} \frac{(b_{n,j+1}-b_{n,j})}{(b_{n,j+1}-b^\dagger)(b^\dagger-b_{n,j})} \quad (12)$$

4 Data

We have evaluated our method, which we henceforth refer to as the Mix-BGe model, on various synthetic data sets. For illustration purposes we present the results obtained for two studies with small networks. The first network consists of two domain nodes X and Y . The true network structure is shown in Figure 1. The dynamics of the system are given by the non-linear state-space equations (see Eq. (1)), with the non-linear function $f(\cdot) = \sin(\cdot)$. We generated 40 observations by applying Eq. (1), setting the drift term $c = 2\pi/41$ to ensure that the complete period $[0, 2\pi]$ of the sinusoid is involved. The second network is a generalisation of this two node domain where three nodes Y_1 , Y_2 , and Y_3 are regulated by X . The regulatory relationships are again realised by sinusoids, whereby we shift the periods. More precisely, we set: $Y_i(t+1) = \sin(X(t) + \tau_i \cdot \pi) + \sigma_y \cdot \phi_{y,i}(t)$ where $\tau_1 = 0$, $\tau_2 = 2/3$, and $\tau_3 = 4/3$. Again we set $c = 2\pi/41$, initialised the variables at $t = 1$ randomly, and generated 40 further observations. We have also applied our method to three gene expression time series of the Interferon regulatory factors Irf1, Irf2 and Irf3 from bone marrow derived macrophages, which we analysed in our earlier work [8]. Data of Irf1, Irf2 and Irf3 were collected at 25×30 minute time intervals under three external conditions: (1) infection with Cytomegalovirus (CMV), (2) treatment with Interferon Gamma (IFN γ), and (3) infection with Cytomegalovirus after pre-treatment with IFN γ (CMV+IFN γ). Finally we have applied our method to two gene expression time series from *Arabidopsis thaliana* cells, which were sampled at 13×2 hour time intervals. As in [8] we focus on 9 genes: LHY, CCA1, TOC1, ELF4, ELF3, GI, PRR9, PRR5, and

PRR3, which are known to be involved in circadian regulation [8]. The expressions were measured independently in 2 plants under experimentally generated constant light condition but with different pre-histories. In the experiment T_{20} (T_{28}), the plant was entrained in a 10h:10h (14h:14h) light/dark-cycle.

5 Simulations

In all our simulations, data were standardised to zero mean and marginal variance of 1 for all dimensions. The hyperparameters of the normal-Wishart prior were chosen as uninformative as possible subject to certain regulatory conditions discussed in [3]. For the synthetic (real) data we set both the burn-in and the sampling-phase lengths of our MCMC simulations to 50,000 (500,000) each and sampled every 1,000 iterations during the sampling-phase¹. For the real data we started 5 independent MCMC simulations from different initialisations on each data set, and we computed the potential scale reduction factor (PSRF) based on the marginal edge posterior probabilities to monitor convergence. As we observed a sufficient degree of convergence for all these data sets ($PSRF < 1.2$), we report only the results of the empty-seeded MCMC runs. For the evaluation of the results, we proceeded in different ways. For the synthetic study based on the two-node network of Figure 1, we computed the marginal posterior probabilities for the individual edges. Our main interest was to test our conjecture from Sect. 2 that the BGe score is susceptible to inferring a spurious self loop for node Y . We wanted to test whether this susceptibility could be reduced by the proposed Mix-BGe model. For the second synthetic simulation study, we assessed the network reconstruction accuracy via the area under the ROC (receiver operator characteristic) curve: AUC; this is a standard criterion that has been applied in numerous related articles. In this study we also compare the Mix-BGe model with our BGM model [8], whereby we exchanged the random allocation of the latent variables of our original BGM model by a change-point process. This slight modification ensures a fair comparison, and actually improved the performance of the BGM model on the synthetic data sets. Finally, for the real data, we focused on the self-loops again. Since we do not know the true network for the *Arabidopsis thaliana* data, computing AUC scores is impossible. However, as discussed in Sect. 2, we conjecture that in the presence of temporal autocorrelation in the signals of the (unknown) regulators, many downstream nodes will show spurious self-loops when the network is reconstructed with the BGe model, whereas this susceptibility should be reduced with the proposed Mix-BGe model. We therefore take as an alternative figure of merit the average marginal posterior probability of a self-loop compared to the average marginal posterior probability of a non-self-loop:

$$\xi = \frac{1}{N_{sl}} \sum_{sl=1}^{N_{sl}} P(e_{sl}|\mathcal{D}) - \frac{1}{N_{nl}} \sum_{nl=1}^{N_{nl}} P(e_{nl}|\mathcal{D}) \quad (13)$$

¹ We note that even for the largest network with $N = 9$ nodes each single simulation was accomplished within a few hours using non-optimised Matlab[©] code on a SunFire X4100M2 machine with AMD Opteron 2224 SE dual-core processor.

where e_{sl} is an edge corresponding to a self-loop, e_{nl} is an edge corresponding to a non-self-loop, N_{sl} is the total number of self loops, and N_{nl} is the total number of non-self-loops. Lower values of ξ are taken as an indication of a better performance.

6 Results

Panels (a) and (b) of Figure 2 show the AUC scores for the synthetic data. Both figures are laid out as matrices, in which the rows and columns correspond to different standard deviations of the noise in X (rows) and in Y (columns). Note that an increase of the noise in X reduces the autocorrelation of X , while increasing the noise in Y blurs the functional dependence of $Y(t+1)$ on $X(t)$. The autocorrelation of Y is jointly influenced by both noise levels. The proposed Mix-BGe model consistently outperforms both the linear BGe model, as well as the common change-point BGM model of Grzegorzczak et al. [8]. Only when the noise in the signal of the hub node X is large (right-most columns) does the proposed Mix-BGe model fail to achieve an improvement. As discussed above, this is a consequence of an increased mis-classification of latent variables, owing to the nature of the change-point process. This could in principle be addressed by combining the node-specific change-points of the present paper with the allocation sampler used in [8] - albeit at the cost of a greatly inflated configuration space in latent space. Panels (c) and (d) in Figure 2 show the marginal posterior probabilities of the four possible edges in the two-node network of Figure 1 and the non-linear state space process of Eq. (1). The results in panel (c) of Figure 2 were obtained with the linear BGe model and show a clear propensity for inferring the spurious self-loop $Y \rightarrow Y$, in confirmation of our earlier conjecture (see Sect. 2). Compare this with the results for the proposed Mix-BGe model, shown in panel (d) of Figure 2. Here, the spurious self-loop $Y \rightarrow Y$ is suppressed in favour of the correct edge $X \rightarrow Y$. There are two noise regimes in which the spurious self-loop $Y \rightarrow Y$ has a marginal posterior probability that is higher than or equal to that of the correct edge $X \rightarrow Y$. One noise regime is where both noise levels in X and in Y are low (top left corner in panels (c) and (d) of Figure 2). Here, the autocorrelation of Y is so high that the spurious self-loop $Y \rightarrow Y$ is still favoured over the true edge $X \rightarrow Y$; this is a consequence of the fact that the functional dependence of $Y(t+1)$ on $X(t)$ is only learned approximately (namely approximated by a mixture model). The second regime is where both noise levels are high (top right corners in the panels (c) and (d) of Figure 2). High noise in Y blurs the functional dependence of $Y(t+1)$ on $X(t)$, while high noise in X leads to a high mis-classification of latent variables and, consequently, a deterioration of the model accuracy; this is a consequence of the fact that latent variables are not allocated individually, as in [8], but according to a change-point process. However, in the majority of noise scenarios, the marginal posterior probability of the correct edge $X \rightarrow Y$ is significantly higher than that of the self-loop $Y \rightarrow Y$. This suggests that the proposed Mix-BGe model is successful at suppressing spurious feedback loops. Finally, Figure 3 shows the results for the real data. It is seen that for four out of five data sets,

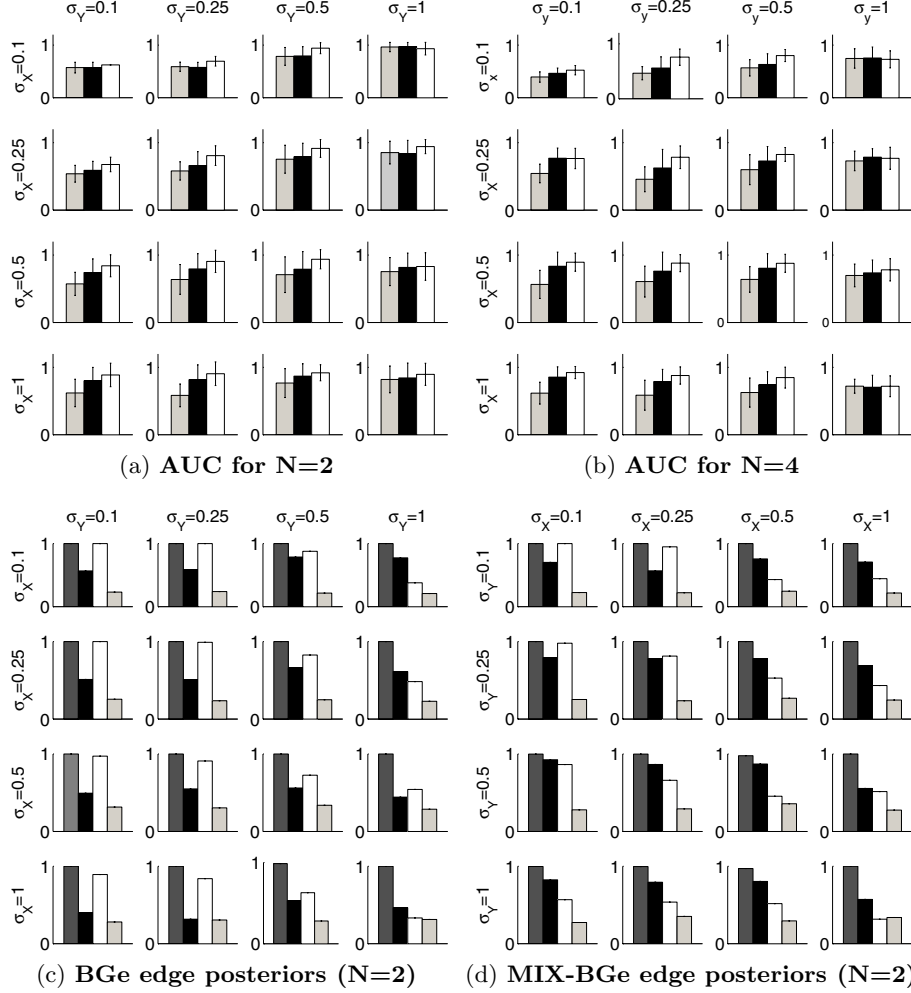


Fig. 2. AUC scores and marginal edge posterior probabilities for the synthetic data. All four panels are laid out as matrices, whose cells correspond to standard deviations σ_X and σ_Y of the noise in X and Y (or Y_i). All histograms show averages (means/std. deviations) from 20 independent data instantiations. Panels (a) and (b) were obtained for data generated from the 4-node-network described in Section 4 and show average AUC score histograms for BGe (left grey bar) and MIX-BGe (right white bar). The centre black bar shows the AUC score for a mixture model in which each change-point applies to *all* the variables, as proposed in [8]. In (c) and (d) the histograms show the posterior probabilities of the edges in the 2-node network, as obtained with BGe (c) and the MIX-BGe model (d). Each histogram contains 4 bars, which represent the average posterior probabilities of the 4 possible edges: Left: self-loop $X \rightarrow X$ (true); centre left: $X \rightarrow Y$ (true); centre right: self-loop $Y \rightarrow Y$ (false); right: $Y \rightarrow X$ (false). It is seen that BGe has a high propensity for learning the spurious feedback loop $Y \rightarrow Y$, while MIX-BGe tends to learn an increased probability of the correct edge $X \rightarrow Y$ (centre left bars).

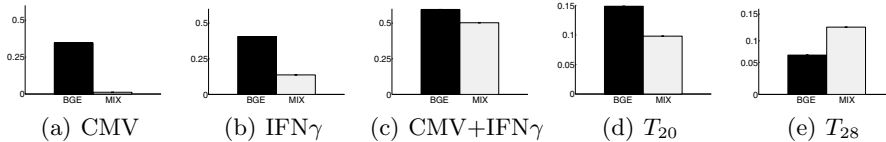


Fig. 3. Results on the Interferon and *Arabidopsis thaliana* gene expression time series. The histograms shows the self-loop score ξ of Eq.(13) for the BGe model (dark bar) and the proposed MIX-BGe model (light bar). Lower values are taken as an indication of a better performance. The histograms in (a)-(c) were obtained from the Interferon regulatory factor gene expression time series: (a) infection with CMV, (b) pre-treatment with IFN γ , and (c) infection and pre-treatment. The histograms in (d)-(e) were obtained from the *Arabidopsis thaliana* gene expression time series: (d) 10h:10h light/dark entrainment, (e) 14h:14h light/dark entrainment.

employing the proposed Mix-BGe model leads to a significant suppression of the marginal posterior probabilities of potentially spurious self loops. We note that the chosen criterion is not based on a proper gold-standard, as we do not know the true number of genuine feedback loops. The difference between the results shown in panels (d) and (e) of Figure 3 may appear as surprising. It cannot be ruled out, though, that the entrainment with different light-dark cycles may indeed lead to the activation of different recurrent pathways, especially given that the interactions between day and evening genes in *Arabidopsis thaliana* is intrinsically of a recurrent nature [12]. Overall, our findings summarised in Figure 3 point to the general reduction of the posterior probability of feedback loops inferred with Mix-BGe as compared with BGe. Given that BGe is intrinsically susceptible to inferring spurious feedback loops, as discussed in Section 2, this points to a potential improvement in the network reconstruction accuracy.

7 Discussion

We have demonstrated that when learning dynamic Bayesian networks from time series data, the presence of temporal autocorrelations in the signals of the regulating nodes renders an approach based on the linear BGe score susceptible to spurious feedback loops. We have proposed a non-linear generalisation of the BGe score based on a mixture model and node-specific change-point processes; this is also a generalisation of the BGM model, where the allocation of data points to mixture components was not node-specific, but affected all nodes simultaneously [8]. Our simulations have shown that the network reconstruction accuracy is improved, and that spurious feedback loops are avoided.

We note that there is a close similarity between our model and the one proposed in [13]. The essential difference is that the model in [13] learns separate network structures for different time series segments. This assumption is reasonable for some scenarios, like morphogenesis. However, for most cellular processes on a shorter time scale, it is questionable whether it is the structure rather than

just the strength of the regulatory interactions that changes with time. The practical problem is potential model over-flexibility. Owing to the high costs of postgenomic high-throughput experiments, time series in systems biology are typically rather short. Modelling short time series segments with separate network structures will almost inevitably lead to inflated inference uncertainty. For that reason we have constrained the network structure to remain invariant and only allow the interaction parameters to change. As a direction for future research one might consider the implementation of a hybrid scheme with a soft rather than hard constraint on the network structures, based on the hierarchical Bayesian model proposed in [14].

References

1. Friedman, N., Linial, M., Nachman, I., Pe'er, D.: Using Bayesian networks to analyze expression data. *Journal of Computational Biology* 7, 601–620 (2000)
2. Cooper, G.F., Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data. *Machine Learning* 9, 309–347 (1992)
3. Geiger, D., Heckerman, D.: Learning Gaussian networks. In: *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, pp. 235–243 (1995)
4. Imoto, S., Kim, S., Goto, T., Aburatani, S., Tashiro, K., Kuhara, S., Miyano, S.: Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network. *Journal of Bioinformatics and Computational Biology* 1(2), 231–251 (2003)
5. Ko, Y., Zhai, C., Rodriguez-Zas, S.L.: Inference of gene pathways using Gaussian mixture models. In: *BIBM International Conference on Bioinformatics and Biomedicine*, pp. 362–367. Fremont, CA (2007)
6. Schwarz, G.: Estimating the dimension of a model. *Annals of Statistics* 6(2), 461–464 (1978)
7. Nobile, A., Fearnside, A.T.: Bayesian finite mixtures with an unknown number of components: The allocation sampler. *Statistics and Computing* 17(2), 147–162 (2007)
8. Grzegorzczak, M., Husmeier, D., Edwards, K.D., Ghazal, P., Millar, A.J.: Modelling non-stationary gene regulatory processes with a non-homogeneous Bayesian network and the allocation sampler. *Bioinformatics* 24, 2071–2078 (2008)
9. Madigan, D., York, J.: Bayesian graphical models for discrete data. *International Statistical Review* 63, 215–232 (1995)
10. Green, P.J.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82(4), 711–732 (1995)
11. Gelman, A., Rubin, D.B.: Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4), 457–472 (1992)
12. Salome, P., McClung, C.: The Arabidopsis thaliana clock. *Journal of Biological Rhythms* 19, 425–435 (2004)
13. Lèbre, S.: Analyse de processus stochastiques pour la génomique: étude du modèle MTD et inférence de réseaux bayésiens dynamiques. PhD thesis. Évry-Val-d'Essone (2008)
14. Werhli, A.V., Husmeier, D.: Gene regulatory network reconstruction by Bayesian integration of prior knowledge and/or different experimental conditions. *Journal of Bioinformatics and Computational Biology* 6(3), 543–572 (2008)