Systems biology

Improvements in the reconstruction of time-varying gene regulatory networks: dynamic programming and regularization by information sharing among genes

Marco Grzegorczyk1,∗ and Dirk Husmeier2,∗

1Department of Statistics, TU Dortmund University, Dortmund, Germany and 2Biomathematics and Statistics Scotland (BioSS), Edinburgh, UK

Associate Editor: Joaquin Dopazo

1 INTRODUCTION

Two paradigm shifts have revolutionized molecular biology in the second half of this decade: systems biology, where the objective is to model the whole complexity of cellular processes in a holistic sense, and synthetic biology, which enables biologists to build new molecular pathways in vivo, i.e. in living cells. The combination of both concepts allows the viability of machine learning approaches for network reconstruction to be tested in a rigorous way. As an alternative to mechanistic models (Cao and Ren, 2008; Wang et al., 2010; Wilkinson, 2006; Xiao and Cao, 2008), which provide a powerful approach to the modelling of small systems composed of a few components, and correlation/mutual information based approaches, which do not distinguish between direct and indirect interactions (Butte and Kohane, 2000), dynamic Bayesian networks (DBNs) have emerged as a promising trade-off between over-simplicity and loss of computational tractability (Cantone et al., 2009). The standard assumption underlying DBNs is that of homogeneity: temporal processes and the time-series they generate are assumed to be governed by a homogeneous Markov relation. However, regulatory interactions and signal transduction processes in the cell are usually adaptive and change in response to external stimuli. Following earlier approaches aiming to relax the homogeneity assumption for undirected graphical models (Talih and Hengartner, 2005; Xuan and Murphy, 2007), various recent research efforts have therefore addressed the homogeneity assumption for DBNs. An approach that has become popular recently is based on a combination of a DBN with a multiple changepoint process, and the application of a Bayesian inference scheme via reversible jump Markov chain Monte Carlo (RJMCMC). Robinson and Hartemink (2009) proposed a discrete inhomogeneous DBN, which allows for different structures in different segments of the time series, with a regularization term penalizing differences among the structures. Grzegorczyk and Husmeier (2009) proposed a continuous inhomogeneous DBN, in which the parameters are allowed to vary, while a common network structure provides information sharing among the time series segments. Lèbre (2007) and Lèbre et al. (2010) proposed an alternative continuous inhomogeneous DBN, which is more flexible in that it allows the network structure to vary among the segments. The model proposed in Kolar et al. (2009) is a close cousin of an inhomogeneous DBN. As opposed to the first three approaches, (hyper-)parameters are not consistently inferred within...
We demonstrate that our scheme subsumes the aforementioned Bayesian clustering scheme akin to the weight sharing principle in Grzegorczyk and Husmeier (2009), with corresponding (node-specific) transitions during embryogenesis, one would expect the majority of affected by changing processes in different ways. The approach pursued in Grzegorczyk and Husmeier (2009), Lèbre (2007), and Lèbre et al. (2010) is similar, except that the allocations of time points to components are not node-specific (i.e. \( K_n \),...). Let \( g(t) \) denote the (conjugate) prior distribution of the parameters. Under fairly weak conditions satisfied (and discussed) in Lèbre (2007), Robinson and Hartemink (2009), Grzegorczyk and Husmeier (2009) and Lèbre et al. (2010), the following integral can be solved analytically:

\[
\int g(t) \prod_{n=1}^{N} P(D_k^n | K_n, V_n) \, \, dt
\]

which yields a closed-form expression for the marginal likelihood:

\[
P(D|G, V, K) = \int P(D|G, V, K, \theta) P(\theta) \, \, d\theta
\]

The objective of Bayesian inference is to sample the network structure \( G \), the latent variables \( V = (V_1, \ldots, V_K) \), and the node-specific numbers of segments \( K = (K_1, \ldots, K_n) \) from the posterior distribution \( P(G, V, K | D) \propto P(G, V, K | D) \), where

\[
P(G, V, K | D) = P(G) P(V|K) P(K | D)
\]

In Grzegorczyk and Husmeier (2009) and Lèbre (2007), a truncated Poisson prior is chosen for \( P(K_n) \), and a multiple changepoint process prior for \( P(V_n|K_n) \). The approach in Grzegorczyk et al. (2010) is similar, except that the allocations of time points to components are not node-specific (i.e. \( K_n \) and \( V_n \) do not depend on \( n \)); see above (class 1 vs. 2).

### 2.2 Improved Gibbs sampling based on dynamic programming

To sample from the posterior distribution, \( P(G, V, K | D) \), all previous studies (Grzegorczyk and Husmeier, 2009; Grzegorczyk et al., 2010; Lèbre et al., 2010; Robinson and Hartemink, 2009) follow the same procedure: to sample the network structure \( G \), they follow Madigan and York (1995) and apply Metropolis-Hasting (MH) structure Markov chain Monte Carlo (MCMC), based on single-edge operations; to sample the latent variables \( V, K \), they follow Green (1995) and apply RMCMC, based on changepoint birth, death and reallocation moves. In the present study, we propose an improved scheme based on dynamic programming. The idea is to adapt the method proposed by Fearnhead (2006) in the context of Bayesian mixture models to inhomogeneous DBNs of the form defined in Equation (1). Fearnhead (2006) assumes that the changepoints occur at discrete time points, and he considers weak conditions satisfied (and discussed) in Lèbre (2007), Robinson and Hartemink (2009), Grzegorczyk and Husmeier (2009) and Lèbre et al. (2010), the following integral can be solved analytically:

\[
\int g(t) \prod_{n=1}^{N} P(D_k^n | K_n, V_n) \, \, dt
\]

which yields a closed-form expression for the marginal likelihood:

\[
P(D|G, V, K) = \int P(D|G, V, K, \theta) P(\theta) \, \, d\theta
\]

The objective of Bayesian inference is to sample the network structure \( G \), the latent variables \( V = (V_1, \ldots, V_K) \), and the node-specific numbers of segments \( K = (K_1, \ldots, K_n) \) from the posterior distribution \( P(G, V, K | D) \propto P(G, V, K | D) \), where

\[
P(G, V, K | D) = P(G) P(V|K) P(K | D)
\]

In Grzegorczyk and Husmeier (2009) and Lèbre (2007), a truncated Poisson prior is chosen for \( P(K_n) \), and a multiple changepoint process prior for \( P(V_n|K_n) \). The approach in Grzegorczyk et al. (2010) is similar, except that the allocations of time points to components are not node-specific (i.e. \( K_n \) and \( V_n \) do not depend on \( n \)); see above (class 1 vs. 2).

### 2.2 Improved Gibbs sampling based on dynamic programming

To sample from the posterior distribution, \( P(G, V, K | D) \), all previous studies (Grzegorczyk and Husmeier, 2009; Grzegorczyk et al., 2010; Lèbre et al., 2010; Robinson and Hartemink, 2009) follow the same procedure: to sample the network structure \( G \), they follow Madigan and York (1995) and apply Metropolis-Hasting (MH) structure Markov chain Monte Carlo (MCMC), based on single-edge operations; to sample the latent variables \( V, K \), they follow Green (1995) and apply RMCMC, based on changepoint birth, death and reallocation moves. In the present study, we propose an improved scheme based on dynamic programming. The idea is to adapt the method proposed by Fearnhead (2006) in the context of Bayesian mixture models to inhomogeneous DBNs of the form defined in Equation (1). Fearnhead (2006) assumes that the changepoints occur at discrete time points, and he considers weak conditions satisfied (and discussed) in Lèbre (2007), Robinson and Hartemink (2009), Grzegorczyk and Husmeier (2009) and Lèbre et al. (2010), the following integral can be solved analytically:

\[
\int g(t) \prod_{n=1}^{N} P(D_k^n | K_n, V_n) \, \, dt
\]

which yields a closed-form expression for the marginal likelihood:

\[
P(D|G, V, K) = \int P(D|G, V, K, \theta) P(\theta) \, \, d\theta
\]

The objective of Bayesian inference is to sample the network structure \( G \), the latent variables \( V = (V_1, \ldots, V_K) \), and the node-specific numbers of segments \( K = (K_1, \ldots, K_n) \) from the posterior distribution \( P(G, V, K | D) \propto P(G, V, K | D) \), where

\[
P(G, V, K | D) = P(G) P(V|K) P(K | D)
\]

In Grzegorczyk and Husmeier (2009) and Lèbre (2007), a truncated Poisson prior is chosen for \( P(K_n) \), and a multiple changepoint process prior for \( P(V_n|K_n) \). The approach in Grzegorczyk et al. (2010) is similar, except that the allocations of time points to components are not node-specific (i.e. \( K_n \) and \( V_n \) do not depend on \( n \)); see above (class 1 vs. 2).
We instantiate the model from Equation (4) by following Fearnhead (2006) to adapt this scheme to the inference of inhomogeneous DBNs, note from the sampling scheme proposed in Friedman and Koller (2003), Equation (10).

\[ P(\pi_1, \ldots, \pi_n) \approx \frac{1}{|\pi_1 \times \cdots \times \pi_n|} \sum_{\pi_1, \ldots, \pi_n} \Psi(\pi_1, \ldots, \pi_n) \Psi(\pi_1, \ldots, \pi_n) \]

where \( \Psi(\pi_1, \ldots, \pi_n) \) has been defined in Equation (3). Equation (6) entails a complete enumeration over all parent configurations, which is computationally expensive. In Grzegorzewski and Husmeier (2009) it was found that this sampling scheme is computationally inefficient when applied to inhomogeneous DBNs. We now demonstrate that this scheme is only inefficient when combined with the RJMCMC scheme for sampling \( V \), but that in combination with the dynamic programming scheme for exact sampling of \( V \) from \( P(V | D) \), an overall gain in computational efficiency can be achieved. We empirically corroborate this conjecture in Section 5.1.2.

For the specific class 2 model employed in this study (Grzegorzewski and Husmeier, 2009) we provide the technical details of the traditional RJMCMC and the novel Gibbs sampling procedures in the Supplementary Material.

2.3 Information coupling between nodes based on Bayesian clustering

We instantiate the model from Equation (4) by following Fearnhead (2006) and employing the joint process prior for the changepoint locations defined in Equation (5), i.e. the terms \( \mathbf{K} \) and \( \mathbf{K}_0 \) in Equations (1–4) become obsolete. We extend the model by introducing a cluster function \( C_i \), that allocates the nodes \( X_1, \ldots, X_c \) to \( c \) \((1 \leq c \leq N)\) non-empty clusters, each characterized by its own changepoint vector \( V_{c} \), \( 1 \leq c \leq c^* \):

\[ P(V_{c}, D, C) = P(V_{c} | P(V_{c}), P(D) | D, C) \]

with \( V_{c} = (V_{c1}, \ldots, V_{cn}) \), where \( c \) is the number of non-empty node clusters induced by \( C \). We assume for \( P(C) \) a uniform distribution on all functions \( C \).

For the study in Section 5.1, we used the commonly applied fan-in restriction of 3. When relaxing the fan-in restriction, the computational costs related to Equation (6) increase. However, a set of effective heuristic techniques for approximate computation at controlled computational complexity are available, as discussed in Friedman and Koller (2003).

That gives \( c \) \((1 \leq c \leq N)\) clusters. The key idea behind the model of Equation (7) is to encourage information sharing among nodes with respect to changepoint locations. Moreover, nodes that are in the same cluster \((1 \leq c \leq c^*) \) share the same allocation vector \( V_{c} \) and will be 'penalized' only once. Note that the novel model is a generalization that subsumes both cluster 1 and cluster 2 models as limiting cases. It corresponds to class 1 for \( c = 1 \) and to class 2 for \( c = N \). Inference can follow a slightly extended Gibbs sampling procedure, where we iteratively sample the latent variables from \( P(V_{c} | D, C) \), a new network structure from \( P(G | V_{c}, D, C) \), and a new cluster formation from \( P(C | V_{c}, D, G) \). The first two steps follow the procedure discussed in Section 2.2. For the third step, sampling from \( P(C | V_{c}, D, G) \), we adopt an RJMCMC scheme (Green, 1995) based on cluster birth (b), death (d) and re-clustering (r) moves. In a cluster birth move we randomly select a node cluster \( c \) that contains at least 2 nodes, and we randomly choose a node contained in it. The move tries to re-cluster this node from the \( i \)-th cluster to a new cluster \( c + 1 \). Denote by \( C^* \) the new cluster formation thus obtained. For the \( (c + 1)\)-th cluster we propose a new changepoint allocation vector \( V_{c+1} \) and \( V_{c+1} \) by sampling from the distributions \( P(V_{c+1} | D, C, \pi_{c+1}) \) and \( P(V_{c+1} | D, C, \pi_{c+1}) \), defined in Equation (9), with the dynamic programming (DP) scheme proposed in Fearnhead (2006), as discussed in Section 2.2. In a cluster death move we randomly select one of the clusters that contain only a single node, and we re-allocate this node to one of the other existing clusters, chosen randomly. The first cluster disappears and for cluster \( j \), which absorbs the node, we propose a new changepoint allocation vector \( V_{j} \) from \( P(V_{j} | D, C, \pi_{j}) \) with DP, where \( C^* \) denotes the proposed cluster formation. In a re-clustering move we randomly choose two clusters \( i \) and \( j \) \((i \neq j)\) as follows. First, cluster \( i \) is randomly selected among those that contain at least 2 nodes. Next, cluster \( j \) is randomly selected among the remaining clusters. We then randomly choose one of the nodes from cluster \( i \) and re-allocate the selected node to cluster \( j \). Denote by \( C^* \) the new cluster formation obtained. (Since cluster \( i \) contains at least 2 nodes, this does not affect \( c \)). For both clusters \( i \) and \( j \) we propose new changepoint allocation vectors \( V_{i} \) and \( V_{j} \) from \( P(V_{i} | D, C, \pi_{i}) \), \( P(V_{j} | D, C, \pi_{j}) \) with DP.

The acceptance probabilities of these three RJMCMC moves are given by the product of the likelihood ratio \( LR \), the prior ratio \( PR \), the inverse proposal probability ratio or Hastings factor \( HR \) and the Jacobian \( J \) in the standard way (Green, 1995): \( A_{KL} = \min(1, R_{KL}) \), where \( R_{KL} = LR \times PR \times HR \times J \). Since this is a discrete problem, the Jacobian is \( J = 1 \), and for the chosen uniform prior on \( C \), the prior ratio \( PR = 1 \). For a cluster birth move \((b)\), symbolically \((c, C\rightarrow (c^*, C))\), we thus get \( R_{KL} = LR \times HR \times J \).

\[ R_{KL} = \frac{P(V_{c+1} | D, C, \pi_{c+1}) \times P(V_{c+1} | D, C, \pi_{c+1})}{P(V_{c} | D, C, \pi_{c}) \times P(V_{c} | D, C, \pi_{c})} \times P(V_{c} | D, C, \pi_{c}) \times P(V_{c} | D, C, \pi_{c})} \]

with \( c^* \) the number of clusters induced by \( C \) with at least two nodes, \( c^* \) the number of nodes in the \( i \)-th cluster that was selected, and \( c^* \) the number of clusters induced by \( C^* \) that contain only a single node. In our extended model the DP scheme described in Section 2.2 can be employed to sample the \( j \)-th \((1 \leq j \leq c)\) allocation vector \( V_{j} \), and we have:

\[ P(V_{j} | D, C, \pi_{j}) = \frac{1}{\sum_{j=1}^{c^{*}} q(D, C, \pi_{j})} q(D, C, \pi_{j}) \]

with the sum in Equation (9) over all valid allocation vectors \( V_{j} \) for the variables in the \( j \)-th cluster of \( C^* \).

It follows from Equations (7–8) that all factors except for the \((c + 1)\)-th in the nominator and the \( c \)-th ones cancel out in the likelihood:

\[ LR = \frac{q(D, C, G, V_{c+1}) \times q(D, C, G, V_{c+1})}{q(D, C, G, V_{c}) \times q(D, C, G, V_{c})} \]
3 DATA

3.1 Synthetic RAF-pathway data

The RAF protein signalling transduction pathway, shown in Figure 2, plays a pivotal role in the mammalian immune response and has hence been widely studied in the literature [e.g. Sachs et al. (2005)]. For our simulation study we followed Grzegorczyk and Husmeier (2009) and generated synthetic network data from a slightly modified version of the pathway, in which an extra injection. Example for ‘PIP2’: $PPIP_2(t+1) = \sqrt{1 - \epsilon^2} PPIP_2(t) + \epsilon \times \text{injection}(t+1)$. The realizations of the other nodes are linear combinations of the realizations of their parents at the preceding time points plus iid standard Normally distributed noise injections. For the standard BGe score (Geiger and Heckerman, 1994). As in earlier studies (Grzegorczyk and Husmeier, 2009; Grzegorczyk et al., 2010) we employ a uniform graph prior subject to a maximum fan-in of 3, and we chose the prior parameter distributions in Equations (1) and (2) to vary the signal-to-noise ratio (SNR). The regression coefficients are re-sampled after $t=8$ and after $t=11$, and the coefficients of the other 5 nodes are re-sampled twice independently, after $t=8$ and after $t=13$. We also consider scenario (v): inhomogeneous regularized class 2 data without any autocorrelation: $\epsilon = 1$. For SNR=3 and SNR=10 we generated 10 independent data instantiations for each scenario (i)-(v).

3.2 Gene expression time series from Arabidopsis thaliana

The Arabidopsis data stem from a study related to circadian regulation in plants. To this end, Arabidopsis seedlings grown under four different artificially controlled light/dark cycles were transferred to constant light and harvested at 12–13 time points in 24-hour intervals. From these seedlings, RNA was extracted and assayed on Affymetrix GeneChip oligonucleotide arrays. As Grzegorczyk and Husmeier (2009) we focus on $N=9$ genes, LHY, TOC1, CCA1, ELF4, ELF3, GI, PRR9, PRR5 and PRR3, which from previous studies are known to be involved in circadian regulation (Locke et al., 2005). Details about the data and their preprocessing are available from Grzegorczyk and Husmeier (2009).

3.3 Synthetically generated network in Saccharomyces cerevisiae (yeast)

While systems biology aims to develop a formal understanding of biological processes via the development of quantitative mathematical models, synthetic biology aims to use such models to design unique biological circuits (synthetic networks) in the cell able to perform specific tasks. Conversely, data from synthetic biology can be utilized to assess the performance of models from systems biology. We used a synthetically generated network of five genes in S.cerevisiae (yeast), devised in Cantone et al. (2009) and depicted in Figure 3, which was obtained from synthetically designed yeast cells grown with different carbon sources: galactose (‘switch on’) or glucose (‘switch off’). We took the data from Cantone et al. (2009), which were obtained with quantitative RT-PCR in intervals of 20 min up to 5 h for the first, and in intervals of 10 min up to 3 h for the second condition. In our study, we standardized the data via a log and a z-score transformation.

4 SIMULATION DETAILS

The two improvements proposed in Section 2 can be applied to any of the inhomogeneous DBNs recently proposed in the literature (Grzegorczyk and Husmeier, 2009; Lèbre, 2007; Lèbre et al., 2010; Robinson and Hartemink, 2009). In our empirical simulation study we use the model presented in Grzegorczyk et al. (2010) as class 1 representant. The class 2 model representant is taken from Grzegorczyk and Husmeier (2009). The novel model can be thought of as a regularized consensus of both models: It is effectively a class / model if it infers only one cluster, and it becomes a class 2 model if it infers $N$ clusters such that each node has its own node-specific changepoints. In our simulation study we also include a standard homogeneous dynamic Bayesian network model based on the standard BGe score (Geiger and Heckerman, 1994). As in earlier studies (Grzegorczyk and Husmeier, 2009; Grzegorczyk et al., 2010) we employ a uniform graph prior subject to a maximum fan-in of 3, and we chose the prior parameter distributions in Equations (1) and (2) to vary the signal-to-noise ratio (SNR). The regression coefficients are re-sampled after $t=8$ and after $t=13$. We also consider scenario (v): inhomogeneous regularized class 2 data without any autocorrelation: $\epsilon = 1$. For SNR=3 and SNR=10 we generated 10 independent data instantiations for each scenario (i)-(v).

5 RESULTS

5.1 Convergence diagnostics on gene expression time series from A.thaliana

The objective of the first study was to assess the improvement in convergence and mixing achieved with the dynamically programmed scheme of Section 2.2. To this end, we applied the inhomogeneous DBN of Equation (1) to gene expression time series from the model plant A.thaliana, described in Subsection 3.2. We aimed to reconstruct a regulatory network among 9 genes, which

Hence, $R_{BB} = LR \times HR$ in Equation (8) reduces to:

$$R_{BB} = \frac{\epsilon \Phi^1 Q_{\text{DBN}}(D,C,G) \Phi^2}{\epsilon^2 Q_{\text{DBN}}(D,C,G)}$$

where the terms $Q_{\text{DBN}}(D,C,G) = \sum_q q(D,C,G, V_q^*)$ can be computed effectively with DP. The acceptance probabilities for death and re-clustering moves can be derived analogously as shown in the Supplementary Material.
from previous studies are known to be involved in circadian regulation (Locke et al., 2005). Our model and simulation setup matched the one described in Grzegorczyk and Husmeier (2009). We compared the standard MCMC scheme applied in previous work, MH/RJMCMC (Grzegorczyk and Husmeier, 2009; Lêbre et al., 2010; Robinson and Hartemink, 2009), which is based on RJMCMC (Green, 1995) and structure MCMC (Madigan and York, 1995), with the Gibbs sampling/dynamic programming scheme discussed in Section 2.2. For the latter, we compared three different subschemes, which differ with respect to the prior distribution on the changepoints. The first subscheme imposition of a Poisson prior with truncation threshold \( K_{\text{max}} \leq 10 \) on the number of components, \( P(K_n) \), and the same even-numbered order statistics prior as applied in Grzegorczyk and Husmeier (2009) and Green (1995) on the segmentations, \( P(V_n|K_n) \). The second subscheme is identical, except that the truncation threshold has been lowered to \( K_{\text{max}} \leq 5 \). The third subscheme follows Fearnhead (2006) and uses the prior imposed by the point process prior of Equation (5) with hyperparameters \( p = 0.05 \) and \( a = 2 \). We refer to these four schemes as MH/RJMCMC, Gibbs(\( K_{\text{max}} = 10 \)), Gibbs(\( K_{\text{max}} = 5 \)) and Gibbs-NBIN, respectively. To assess the degree of convergence, we repeated the MCMC simulations from five different initializations with three other models: a standard homogeneous Bayesian network model, a class 1 model with changepoints that are common to all nodes (Grzegorczyk et al., 2010), and a class 2 model with node-specific changepoints (Grzegorczyk and Husmeier, 2009). In our study we evaluated the network reconstruction accuracy with the area under the precision-recall curve (AUC) (Davis and Goadrich, 2006); see Section 5.3 for more details. This is standard in systems biology, with larger scores indicating a better performance.

Figure 2 summarizes the empirical results of our simulation study. (1) **Homogeneous data**: Except for the highest setting of the hyperparameter \( p \), the three inhomogeneous DBNs never perform worse than the homogeneous model, while on the other hand, the homogeneous model is inappropriate and performs substantially worse. (2) **class 1 data**: The class 1 model and the proposed regularized class 2 model perform equally well. Both outperform the class 2 model, except for high values of \( p \). (3) **class 2 data**: The class 1 model cannot accommodate the node-specific changepoints and is outperformed by the proposed regularized class 2 model (the ‘NEW’ model). Interestingly, the latter also shows more stability than the class 2 model with respect to a variation of the hyperparameter \( p \), indicating increased robustness as a consequence of the node clustering. (4) **Regularized class 2 data**: The results are comparable to those for the class 2 data. The class 1 model is consistently inferior to the class 2 model, and the class 2 model is, once again, substantially more susceptible to a variation of \( p \). The mean AUC values are—overall—lower than for the previous case, the class 2 data. This seems to be a consequence of spurious interactions resulting from chance correlations. Setting

---

\[ \text{MH/RJMCMC} (5500 \text{ Gibbs-NBIN}) \text{ steps take 45 min with our MATLAB code on a SunFire X4100M2 machine.} \]

---

\[ \text{Recall that a high value of the hyperparameter } p \text{ implies a low prior penalty for changepoints.} \]
Fig. 2. Network reconstruction accuracy on synthetic data. The figure shows the mean area under the precision-recall curves (AUC) in dependence on the hyperparameter \( p \) of the negative binomial point process prior of Equation (5). For the RAF pathway (bottom right panel) we implemented 5 scenarios of inhomogeneity as explained in Section 5.2. For each scenario there is a panel for SNR=3 and SNR=10; 'noAC' stands for 'no autocorrelation'. The following models were applied to the data, each representing a particular class: (i) homogeneous model: the standard DBN model based on the BGe score, (ii) the class 1 model was taken from Grzegorczyk et al. (2010), (iii) the class 2 model was taken from Grzegorczyk and Husmeier (2009) and (iv) the regularized class 2 model was generated from the class 2 model in Grzegorczyk and Husmeier (2009) as explained in Section 3.1. The mean AUC scores were computed from 10 independent data instantiations.

5.3 Synthetic biology in S.cerevisiae

In the final application we compare the proposed model with other state-of-the art techniques on a topical dataset from synthetic biology. We used a synthetically generated network of five genes in S.cerevisiae (yeast), depicted in Figure 3, which was used in Cantone et al. (2009) to evaluate two state-of-the-art network reconstruction methods: BANJO, a conventional DBN, trained with simulated annealing; and TSNI, an approach based on ordinary differential equations. Both methods, which are described in more detail in Cantone et al. (2009), were applied to gene expression time series obtained from synthetically designed yeast cells grown with different carbon sources: galactose ('switch on') or glucose ('switch off'). BANJO and TSNI were then applied to infer a network [see Cantone et al. (2009) for details], from which, by comparison with the known gold standard, the precision (proportion of correctly predicted interactions out of the total number of predicted interactions) and recall (percentage of true interactions that have been correctly identified) scores were determined. In our study, we used the data described in Section 3.3, applied the proposed regularized class 2 model as described in Sections 2.3, and sampled networks from the posterior distribution with the Gibbs sampling scheme described in Section 2.2. This gives us an ordering of interactions, ranked by their marginal posterior probability, and by plotting precision against recall scores for different thresholds, we obtain the precision-recall (PR) curves (Davis and Goadrich, 2006) shown in Figure 3. Larger areas under the PR curve are indicative of a better reconstruction accuracy; hence in agreement with Cantone et al. (2009) we find that the 'switch on' data are more informative than their 'switch off' counterpart. The scores for BANJO and TSNI, which we took from Cantone et al. (2009), lie clearly and consistently below the 'switch on' PR curve, for different choices of the changepoint process prior—defined by \( p \) in Equation (5). This suggests that the proposed method achieves a genuine and significant improvement over state-of-the-art schemes reported in the recent systems biology literature.
We have proposed two improvements for time-varying DBNs: a Gibbs sampling (GS) scheme based on dynamic programming (DP) as an alternative to RJMCMC, and information coupling between nodes based on Bayesian clustering. The evaluation on real gene expression dataset from *A. thaliana* suggests that GS-DP shows faster mixing and convergence than MH/RJMCMC. A comparative evaluation on synthetic data demonstrates that the new model based on information coupling between nodes compares favourably with earlier models that either employ network-wide (class 1) or node-specific (class 2) change points. On gene expression time series from a recent study of synthetic biology in *S. cerevisiae* the proposed model has outperformed two state-of-the-art network reconstruction methods. These findings suggest that the proposed method makes important contributions both to inference and performance of network reconstruction methods, and hence adds a valuable new tool to the kit of computational systems biology. In our future work we will investigate different choices for the prior on node cluster formations, introduced in Section 2.3, exploring methods from Bayesian non-parametrics based on Dirichlet process priors.

**ACKNOWLEDGEMENTS**

Marco Grzegorczyk is supported by the Graduate School ‘Statistische Modellbildung’ of the Department of Statistics, TU Dortmund University. Dirk Husmeier is supported by the Scottish Government Rural and Environment Research and Analysis Directorate (RERAD) and under the EU FP7 project ‘Timet’.  

**Conflict of Interest:** none declared.

**REFERENCES**

Butte,A. and Kohane,I. (2000) Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Proc. Symp. Biocomput.*, 2000, 418–429.

Cantone,I. et al. (2009) A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. *Cell.*, 137, 172–181.

Cas,J. and Ren,F. (2008) Exponential stability of discrete-time genetic regulatory networks with delays. *IEEE Trans. Neural Netw.*, 19, 520–523.

Davis,J. and Goadrich,M. (2006) The relationship between precision-recall and ROC curves. *JMLR.*, 23, 233–240.

Fearnhead,P. (2006) Exact and efficient Bayesian inference for multiple changepoint problems. *Stat. Comput.*, 16, 203–213.

Freudman,N. and Koller,D. (2000) Being Bayesian about network structure. *Mach. Learn.*, 89, 95–126.

Geweke,D. and Heckerman,D. (1994) Learning Gaussian networks. *UAI*, 10, 235–245.

Gelman,A. and Rubin,D.B. (1992) Inference from iterative simulation using multiple sequences. *Stat. Science*, 7, 457–472.

Green,P. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82, 711–732.

Grzegorczyk,M. and Husmeier,D. (2009) Non-stationary continuous dynamic Bayesian networks. *NIPS*, 22, 662–690.

Grzegorczyk,M. et al. (2008) Modelling non-stationary gene regulatory processes with a non-homogeneous Bayesian network and the allocation sampler. *Bioinformatics*, 24, 2071–2078.

Grzegorczyk,M. et al. (2010) Modelling non-stationary dynamic gene regulatory processes with the BGM model. *Comput. Stat. (Epub ahead of print, doi:10.1007/s00180-010-0201-9).*

Kolar,M. et al. (2009) Sparse learning of varying-coefficient models with structural changes. *NIPS*, 22, 1006–1014.

Lêbre,S. (2007) Stochastic process analysis for Genomics and Dynamic Bayesian Networks inference. Ph.D. Thesis, Université d’Evry-Val-d’Essonne, France.

Lêbre,S. et al. (2010) Statistical inference of the time-varying structure of gene-regulation networks. *BMC Syst. Biol.*, 4, Article 74 or number. 130.

Locke,J. et al. (2005) Extension of a genetic network model by iterative experimentation and mathematical analysis. *Mol. Syst. Biol.*, 1, Article ID or number: 2005.0013.

Madigan,D. and York,J. (1995) Bayesian graphical models for discrete data. *Int. Stat. Rev.*, 63, 215–232.

Nowlan,S.J. and Hinton,G.E. (1992) Simplifying neural networks by soft weight-sharing. *Neural Comput.*, 4, 473–493.

Robinson,J.W. and Hartemink,A.J. (2009) Non-stationary dynamic Bayesian networks. *NIPS*, 21, 1169–1176.

Sachs,K. et al. (2005) Protein-signaling networks derived from multiparameter single-cell data. *Science*, 308, 523–529.

Talbi,M. and Hengartner,N. (2005) Structural learning with time-varying components: Tracking the cross-section of financial time series. *J. Stat. Plann. Est.*, 87, 321–341.

Wang,Y. et al. (2010) Global robust power-rate stability of delayed genetic regulatory networks with noise perturbation. *Cogn. Neurodyn.*, 4, 81–90.

Wilkinson,D. (2006) Stochastic modelling for systems biology. Chapman and Hall/CRC Press, Boca Raton, Florida.

Xuan,X. (2007) Green network model and its application in genetic regulatory networks with time delay. *J. Comput. Biol.*, 14, 815–833.

Xuan,X. and Murphy,K. (2007) Modelling changing dependency structure in multivariate time series. *JMLR.*, 24, 1055–1082.

**6 CONCLUSION**

We have proposed two improvements for time-varying DBNs: a Gibbs sampling (GS) scheme based on dynamic programming (DP) as an alternative to RJMCMC, and information coupling between nodes based on Bayesian clustering. The evaluation on a real gene expression dataset from *A. thaliana* suggests that GS-DP shows faster mixing and convergence than MH/RJMCMC. A comparative evaluation on synthetic data demonstrates that the new model based on information coupling between nodes compares favourably with earlier models that either employ network-wide (class 1) or node-specific (class 2) change points. On gene expression time series from a recent study of synthetic biology in *S. cerevisiae* the proposed model has outperformed two state-of-the-art network reconstruction methods. These findings suggest that the proposed method makes important contributions both to inference and performance of network reconstruction methods, and hence adds a valuable new tool to the kit of computational systems biology. In our future work we will investigate different choices for the prior on node cluster formations, introduced in Section 2.3, exploring methods from Bayesian non-parametrics based on Dirichlet process priors.