A New Partially Segment-Wise Coupled Piece-Wise Linear Regression Model for Statistical Network Structure Inference

Shafiee Kamalabad, Mahdi; Grzegorczyk, Marco

Published in:
Computational Intelligence Methods for Bioinformatics and Biostatistics

DOI:
10.1007/978-3-030-34585-3_13

A New Partially Segment-Wise Coupled Piece-Wise Linear Regression Model for Statistical Network Structure Inference

Mahdi Shafiee Kamalabad and Marco Grzegorczyk(✉)

Bernoulli Institute, Groningen University,
Nijenborgh 9, 9747 Groningen, The Netherlands
m.a.grzegorczyk@rug.nl
http://www.math.rug.nl/stat/People/Marco

Abstract. We propose a new non-homogeneous dynamic Bayesian network with partially segment-wise sequentially coupled network parameters. The idea is to infer the segmentation of a time series of network data using multiple changepoint processes, and to model the data in each segment by linear regression models. The conventional uncoupled models infer the network interaction parameters for each segment separately, without any systematic information-sharing among segments. More recently, it was proposed to couple the network interaction parameters sequentially among segments. The idea is to enforce the parameters of any segment to stay similar to those of the previous segment. This coupling mechanism can be disadvantageous, as it enforces coupling and does not feature any options to uncouple. We propose a new consensus model that infers for each individual segment whether it should be coupled to (or better should stay uncoupled from) the preceding one.

Keywords: Network structure learning · Dynamic Bayesian networks · Bayesian piece-wise linear regression · Partial segment-wise coupling

1 Introduction

In systems biology, dynamic Bayesian network models (DBNs) have become popular tools for learning the structures of regulatory networks from data. For temporal data, the usual assumption is that all regulatory interactions are subject to a time lag. There is then no acyclicity constraint on the network structure, and the parent nodes of each node can be learned separately. A common approach is then to make use of independent regression models to infer the parents (=covariates) of each individual network node, and to merge the individual parent sets in

This work was supported by the European Cooperation in Science and Technology (COST) [COST Action CA15109 European Cooperation for Statistics of Network Data Science (COSTNET)].

M. Raposo et al. (Eds.): CIBB 2018, LNBI 11925, pp. 139–152, 2020.
https://doi.org/10.1007/978-3-030-34585-3_13

form of a network; see Sect. 2.3 for a more detailed description. A shortcoming of dynamic Bayesian networks is that they are homogeneous models, so that the network parameters are not allowed to change over time. The same set of network parameters applies to all time points. For many applications, this assumption of homogeneity is unrealistic. For example, in cellular networks the strengths of the regulatory interactions often depend on unobserved external factors, such as cellular or experimental conditions, which do not necessarily stay constant over time. This renders the traditional dynamic Bayesian network models inappropriate for many biological applications. Therefore various non-homogeneous DBN models (NH-DBNs) have been proposed in the computational biology literature. The proposed NH-DBNs can be grouped into two classes: (i) NH-DBNs that allow only the network parameters to vary in time and (ii) NH-DBNs that allow the network structure and the network parameters to be time-dependent. We here focus on NH-DBNs that assume the network structure to be time-invariant (i), as this assumption is more realistic for our applications in Sect. 4.

When the network structure (i.e. the collection of covariate sets) does not change over time, while the network parameters (i.e. the regression coefficients) are time-dependent, piece-wise linear regression models can be used to model the data. A multiple changepoint process is applied to infer the segmentation of the data into disjoint segments, and the data within each segment are modelled by linear regression. The joint network structure, the number of changepoints and their locations, and the segment-specific network parameters (regression coefficients) are unknown and have to be inferred from the data. With the conventional uncoupled models (see, e.g., the work by Lèbre et al. [1]), the segment-specific network parameters have to be learned for each segment separately, without information-sharing among segments. For short time series, the segmentation leads to even shorter segments, sometimes containing a few data points only. This can lead to inflated inference uncertainties, so that the segment-specific network parameters cannot be properly learned from the data.

To address this issue, a model with fully sequentially coupled network parameters was proposed by Grzegorczyk and Husmeier [2]. The key idea is to allow for information-exchange among segments by a sequential coupling scheme. The posterior expectation of the network parameters of each segment h are used as prior expectation for the next segment $h+1$. In the fully sequentially coupled model from Grzegorczyk and Husmeier [2], node-specific coupling strength parameters regulate the variance of the network parameter priors and so the effective coupling strength between all pairs $(h, h+1)$ of neighboring segments. A disadvantage of this fully coupled model is that each segment $h \geq 2$ is enforced to be coupled to the preceding segment $h-1$. That is, the regression coefficient prior distributions for each segment $h \geq 2$ are automatically centered around their posterior expectations from the preceding segment $h-1$, and the coupling strength parameter only regulates the variance of these prior distributions. Low coupling strength parameters yield peaked prior distributions while high coupling strength parameters yield vague prior distributions around the posterior expectations from the preceding segment. In particular, in the fully sequentially

coupled model from Grzegorczyk and Husmeier [2] there is no option to uncouple a segment $h \geq 2$ from the previous segment $h-1$. Therefore, for networks with dissimilar segment-specific network parameters (regression coefficients), this information coupling scheme can become very counter-productive: Uncoupling can effectively only be achieved by making the network parameter prior distributions vague, so as to allow the regression coefficients to get dissimilar from their prior expectations (= the posterior expectations from the preceding segment).

In this paper, we address this disadvantage of the fully sequentially coupled model from Grzegorczyk and Husmeier [2]. We extend the fully coupled model by introducing a new option to uncouple segments. This yields a new model, which we refer to as the new partially segment-wise coupled model. The new models infers for each individual segment $h \geq 2$ whether it is coupled to (or uncoupled from) the preceding segment $h-1$. Hence, our new partially segment-wise coupled model can infer the best trade-off between the uncoupled model from Lèbre et al. [1] and the fully sequentially coupled model from Grzegorczyk and Husmeier [2]. The uncoupled model and the fully coupled model are the limiting cases, where either *all* segments are uncoupled or *all* segments are coupled.

In a complementary work [3], we have generalized the fully sequentially coupled model from Grzegorczyk and Husmeier [2] by introducing segment-specific coupling strength parameters. Although segment-specific coupling strengths increase the model flexibility, in this generalized fully coupled model [3] each segment stays coupled to the previous one and the uncoupled model from Lèbre et al. [1] cannot be reached as limiting case.

2 Methods

2.1 The New Partially Segment-Wise Coupled Model

Consider a piece-wise Bayesian linear regression model with response variable Y and covariate set $\pi = \{X_1, \ldots, X_k\}$. We assume that there are T temporal data points and that they can be divided into H disjoint segments with segment-specific regression coefficients. Let $\mathbf{y}_h$ be the response vector and $\mathbf{X}_h$ be the design matrix for segment h, where $\mathbf{X}_h$ includes a column of 1's for the intercept. For each segment $h = 1, \ldots, H$ we then have:

$$\mathbf{y}_h \sim \mathcal{N}(\mathbf{X}_h \mathbf{w}_h, \sigma^2 \mathbf{I}) \tag{1}$$

where $\mathbf{w}_h = (\mathbf{w}_{h,0}, \ldots, \mathbf{w}_{h,k})^{\mathsf{T}}$ is the regression coefficient vector for segment h, and σ^2 is the noise variance parameter, which we assume to have an inverse Gamma distribution:

$$\sigma^{-2} \sim \mathrm{GAM}(\alpha_\sigma, \beta_\sigma)$$

Onto the regression coefficient vectors $\mathbf{w}_h$ $(h = 1, \ldots, H)$ we impose the following novel prior distributions:

$$\mathbf{w}_h \sim \mathcal{N}(\boldsymbol{\mu}_h, \boldsymbol{\Sigma}_h) \text{ with } \boldsymbol{\mu}_h := \delta_h \tilde{\mathbf{w}}_{h-1} \text{ and } \boldsymbol{\Sigma}_h := \lambda_{c^{\delta_h}} \lambda_{u^{1-\delta_h}} \sigma^2 \mathbf{I} \tag{2}$$

where $\delta_1 := 0$, $\delta_h \in \{0, 1\}$ for $h > 1$, $\tilde{\mathbf{w}}_0 := \mathbf{0}$. This yields:

$$\mathbf{w}_h \sim \begin{cases} \mathcal{N}(\mathbf{0}, & \lambda_u \sigma^2 \mathbf{I}) \quad \text{if } h = 1 \text{ or } \quad \delta_h = 0 \\ \mathcal{N}(\tilde{\mathbf{w}}_{h-1}, \lambda_c \sigma^2 \mathbf{I}) & \text{if } h > 1 \text{ and } \delta_h = 1 \end{cases} \tag{3}$$

λ_c and λ_u are free hyperparameters, onto which we impose inverse Gamma prior distributions:

$$\lambda_c^{-1} \sim \text{GAM}(\alpha_c, \beta_c)$$
$$\lambda_u^{-1} \sim \text{GAM}(\alpha_u, \beta_u)$$

The newly introduced indicator variables $\delta_h \in \{0, 1\}$ indicate whether segment $h \geq 2$ is coupled to segment $h - 1$ ($\delta_h = 1$) or not ($\delta_h = 0$), and $\tilde{\mathbf{w}}_{h-1}$ ($h \geq 2$) is the posterior expectation of $\mathbf{w}_{h-1}$; see Eq. (5) below.
The subscripts 'u' and 'c' indicate whether the hyperparameters apply to uncoupled ($\delta_h = 0$) or coupled ($\delta_h = 1$) segments. The posterior distribution (full conditional distribution) of $\mathbf{w}_h$ is:

$$\mathbf{w}_h | (\sigma^2, \lambda_c, \lambda_u, \delta_h, \mathbf{y}_h) \sim \mathcal{N}(\tilde{\mathbf{w}}_h, \sigma^2 \tilde{\boldsymbol{\Sigma}}_h) \tag{4}$$

where

$$\tilde{\boldsymbol{\Sigma}}_h = \lambda_c^{-\delta_h} \lambda_u^{-(1-\delta_h)} \mathbf{I} + \mathbf{X}_h^\mathsf{T} \mathbf{X}_h$$
$$\tilde{\mathbf{w}}_h = \left(\tilde{\boldsymbol{\Sigma}}_h \right)^{-1} \left(\lambda_c^{-\delta_h} \lambda_u^{-(1-\delta_h)} \boldsymbol{\mu}_{h-1} + \mathbf{X}_h^\mathsf{T} \mathbf{y}_h \right) \tag{5}$$

We assume that the new indicator variables $\delta_2, \ldots, \delta_H$ follow a Bernoulli distribution:

$$\delta_h | p \sim \text{Ber}(p)$$

where the probability hyperparameter $p \in [0, 1]$ is Beta distributed:

$$p \sim \text{Beta}(a, b)$$

Our new model is then a consensus model between an uncoupled model and a fully coupled model:

- If $\delta_h = 0$ for all h, then $P(\mathbf{w}_h) = \mathcal{N}(\mathbf{0}, \lambda_u \sigma^2 \mathbf{I})$ for all h. This is an **uncoupled model** without information sharing among segments. The prior expectations of all regression coefficients $\mathbf{w}_{h,j}$ ($j = 0, \ldots, k$) are 0, and λ_u can be interpreted as a signal-to-noise ratio parameter.
- If $\delta_h = 1$ for $h \geq 2$, then $P(\mathbf{w}_h) = \mathcal{N}(\tilde{\mathbf{w}}_{h-1}, \lambda_c \sigma^2 \mathbf{I})$ for $h \geq 2$. This refers to the **fully sequentially coupled model** from [2]. The prior expectation of each individual regression coefficient $\mathbf{w}_{h,j}$ is its posterior expectation $\tilde{\mathbf{w}}_{h-1,j}$ from the previous segment ($j = 0, \ldots, k$), and λ_c can be interpreted as coupling (strength) parameter, where λ_c^{-1} refers to the coupling strength.
- Our new **partially segment-wise coupled model** infers the values of the binary variables δ_h ($h \geq 2$) from the data, so as to find the best trade-off between the uncoupled and the fully coupled model.

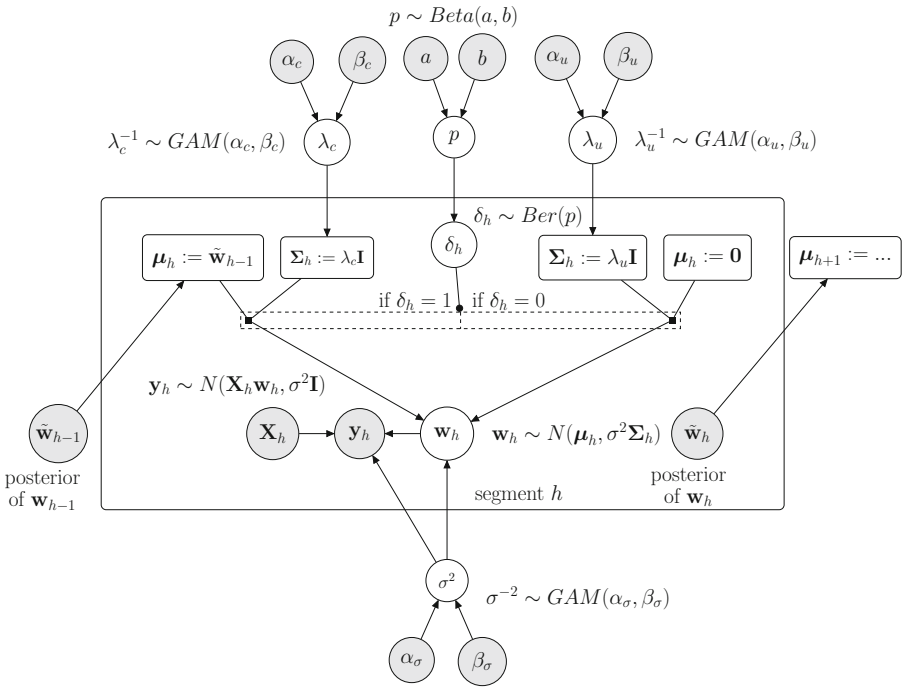

Fig. 1. Graphical representation of the new partially segment-wise coupled model. Parameters that have to be inferred are in white circles. The data and the fixed hyperparameters are in grey circles. The rectangles contain definitions that deterministically depend on the parent nodes. Everything within the plate is segment-specific.

Figure 1 shows a graphical model representation of our new partially coupled model and lists all prior distributions along with their hyperparameters.

For $\delta_h \sim \text{Ber}(p)$ with $p \sim \text{Beta}(a, b)$ the joint marginal density of $\{\delta_h\} := (\delta_2, \ldots, \delta_H)$ is:

$$p(\{\delta_h\}) = \int_0^1 p(p) \prod_{h=2}^{H} p(\delta_h|p)\, dp \tag{6}$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a + \sum_{h=2}^{H} \delta_h)\Gamma(b + \sum_{h=2}^{H}(1-\delta_h))}{\Gamma(a+b+(H-1))}$$

The marginal likelihood, $p(\mathbf{y}|\boldsymbol{\pi}, \boldsymbol{\tau}, \lambda_u, \lambda_c, \{\delta_h\})$, where $\boldsymbol{\pi} = \{X_1, \ldots, X_k\}$ is the covariate set and $\boldsymbol{\tau}$ denotes the data segmentation into the H segments $(h = 1, \ldots, H)$, can be computed analytically by applying the rule from Sect. 2.3

in Bishop [4]:

$$p(\mathbf{y}|\lambda_u, \lambda_c, \{\delta_h\}) = \frac{\Gamma(\frac{T}{2} + a_\sigma)}{\Gamma(a_\sigma)} \cdot \frac{\pi^{-T/2} \cdot (2b_\sigma)^{a_\sigma}}{\left(\prod_{h=1}^{H} \det(\mathbf{C}_h)\right)^{1/2}} \cdot \left(2b_\sigma + \Delta^2\right)^{-(\frac{T}{2}+a_\sigma)} \quad (7)$$

where σ^2 and $\mathbf{w}_1, \ldots, \mathbf{w}_H$ have been integrated out, and

$$\mathbf{C}_h := \mathbf{I} + \mathbf{X}_h \boldsymbol{\Sigma}_h \mathbf{X}_h^\mathsf{T}$$
$$\Delta^2 := \sum_{h=1}^{H} (\mathbf{y}_h - \delta_h \mathbf{X}_h \tilde{\mathbf{w}}_{h-1})^\mathsf{T} \mathbf{C}_h^{-1} (\mathbf{y}_h - \delta_h \mathbf{X}_h \tilde{\mathbf{w}}_{h-1})$$

2.2 Covariate Set and Data Segmentation Learning

The data segmentation $\boldsymbol{\tau}$ and the covariate sets $\boldsymbol{\pi}$ are usually unknown and the major objective is to infer them from the data. Given N covariates, we assume that all subsets $\boldsymbol{\pi} \subset \{X_1, \ldots, X_N\}$ are equally likely a priori. For the data segmentation we use changepoint sets; i.e. we use a set of $H-1$ changepoints $\boldsymbol{\tau} := \{\tau_1, \ldots, \tau_{H-1}\}$ to divide the temporal data points $\mathcal{D} = \{\mathcal{D}_1, \ldots, \mathcal{D}_T\}$ into H segments. Data point $\mathcal{D}_t$ is in segment h if $\tau_{h-1} < t \leq \tau_h$, where $\tau_0 := 0$ and $\tau_H := T$ are two pseudo-changepoints . We assume that the distances between neighboring changepoints are geometrically distributed with a fixed hyperparameter $\mathrm{p} \in (0, 1)$:

$$\begin{aligned} p(\boldsymbol{\tau}|\mathrm{p}) &= \left(\prod_{h=1}^{H-1} (1-\mathrm{p})^{\tau_h - \tau_{h-1} - 1} \mathrm{p}\right) (1-\mathrm{p})^{\tau_H - \tau_{H-1} - 1} \\ &= (1-\mathrm{p})^{(T-1)-(H-1)} \mathrm{p}^{H-1} \end{aligned} \quad (8)$$

The posterior distribution then takes the form:

$$\begin{aligned} p(\lambda_u, \lambda_c, \{\delta_h\}, \boldsymbol{\pi}, \boldsymbol{\tau}|\mathcal{D}) \propto p(\boldsymbol{\pi}) \cdot p(\boldsymbol{\tau}|\mathrm{p}) \cdot p(\lambda_u) \cdot p(\lambda_c) \cdot p(\{\delta_h\}) \\ \cdot p(\mathbf{y}|\boldsymbol{\pi}, \boldsymbol{\tau}, \lambda_u, \lambda_c, \{\delta_h\}) \end{aligned} \quad (9)$$

where the marginal likelihood $p(\mathbf{y}|\boldsymbol{\pi}, \boldsymbol{\tau}, \lambda_u, \lambda_c, \{\delta_h\})$ was defined in Eq. 7. Reversible Jump Markov Chain Monte Carlo (RJMCMC) simulations can be used to generate posterior samples from the posterior distribution in Eq. 9. In each RJMCMC iteration we re-sample the parameters in λ_u, λ_c and $\{\delta_h\}$ from their full conditional distributions (Gibbs sampling), and we perform Metropolis-Hastings moves to sample covariate sets $\boldsymbol{\pi}$ and changepoint sets $\boldsymbol{\tau}$.

Covariate Set Inference. For sampling covariate sets $\boldsymbol{\pi}$ we implement 3 move types: 'covariate additions (A)', covariate removals (R)', and 'covariate exchanges (E)'. Each move proposes to replace $\boldsymbol{\pi}$ by a new covariate set $\boldsymbol{\pi}^*$ having one covariate more (A) or less (R) or exchanged (E). When randomly

selecting the move type and the involved covariate(s), we get the acceptance probability:

$$A(\boldsymbol{\pi} \to \boldsymbol{\pi}^*) = \min\left\{1, \frac{p(\mathbf{y}|\boldsymbol{\pi}^*, \boldsymbol{\tau}, \lambda_u, \lambda_c, \{\delta_h\})}{p(\mathbf{y}|\boldsymbol{\pi}, \boldsymbol{\tau}, \lambda_u, \lambda_c, \{\delta_h\})} \cdot \frac{p(\boldsymbol{\pi}^*)}{p(\boldsymbol{\pi})} \cdot HR_{\boldsymbol{\pi}}\right\}$$

where the move type (A, R, or E) specific Hastings ratios are:

$$HR_{\boldsymbol{\pi},A} = \frac{n - |\boldsymbol{\pi}|}{|\boldsymbol{\pi}^*|}, \quad HR_{\boldsymbol{\pi},R} = \frac{|\boldsymbol{\pi}|}{n - |\boldsymbol{\pi}^*|}, \quad HR_{\boldsymbol{\pi},E} = 1$$

Changepoint Set Inference. For sampling changepoint sets $\boldsymbol{\tau}$ we also implement 3 move types: 'changepoint birth (B)', 'changepoint death (D)', and 'changepoint re-allocation (R)' moves. Each move proposes to replace $\boldsymbol{\tau}$ by a new changepoint set $\boldsymbol{\tau}^*$ having one changepoint added (B) or deleted (D) or re-allocated (R). We randomly select the move type, the involved changepoint and the new changepoint location. Changepoint moves also affect the numbers of parameters in the collection $\{\delta_h\}$. For each segment that stays unchanged we keep the old parameter. For altering segments we re-sample the corresponding parameters. To this end, we flip coins to get candidates for the involved δ_h's. This yields a new collection of binary variables $\{\delta_h\}^*$ and the Metropolis-Hastings acceptance probability is:

$$A([\boldsymbol{\tau}, \{\delta_h\}] \to [\boldsymbol{\tau}^*, \{\delta_h\}^*])$$
$$= \min\left\{1, \frac{p(\mathbf{y}|\boldsymbol{\pi}, \boldsymbol{\tau}^*, \lambda_u, \lambda_c, \{\delta_h\}^*)}{p(\mathbf{y}|\boldsymbol{\pi}, \boldsymbol{\tau}, \lambda_u, \lambda_c, \{\delta_h\})} \cdot \frac{p(\boldsymbol{\tau}^*)}{p(\boldsymbol{\tau})} \cdot \frac{p(\{\delta_h\}^*)}{p(\{\delta_h\})} \cdot HR_{\boldsymbol{\tau}}\right\}$$

where the move type specific (B, D, or R) Hastings ratios are

$$HR_{\boldsymbol{\tau},B} = \frac{T - 1 - |\boldsymbol{\tau}^*|}{|\boldsymbol{\tau}|} \cdot \frac{1}{2}, \quad HR_{\boldsymbol{\tau},D} = \frac{|\boldsymbol{\tau}^*|}{T - 1 - |\boldsymbol{\tau}|} \cdot 2, \quad HR_{\boldsymbol{\tau},R} = 1 \quad (10)$$

2.3 Network Structure Learning

If n variables $Z_1, \ldots, Z_n$ have been observed over time, the conventional dynamic Bayesian network assumption is that the regulatory interactions are subject to a time lag. A network edge $Z_i \to Z_j$ then indicates that Z_j at time point $t+1$ depends on the value of Z_i at time point t. The task of learning a network among the n variables can then be separated into n independent regression tasks. In the j-th regression model $Y := Z_j$ is the response and the other $N := n-1$ variables $\{Z_1, \ldots, Z_{j-1}, Z_{j+1}, \ldots, Z_n\}$ are the potential covariates. Having inferred a covariate set $\boldsymbol{\pi}^j$ for each Z_j, a network can be built by merging the covariate sets in form of a network $\mathcal{N} := \{\boldsymbol{\pi}^1, \ldots, \boldsymbol{\pi}^n\}$. There is the edge $Z_i \to Z_j$ in the network $\mathcal{N}$ if and only if $Z_i \in \boldsymbol{\pi}^j$.

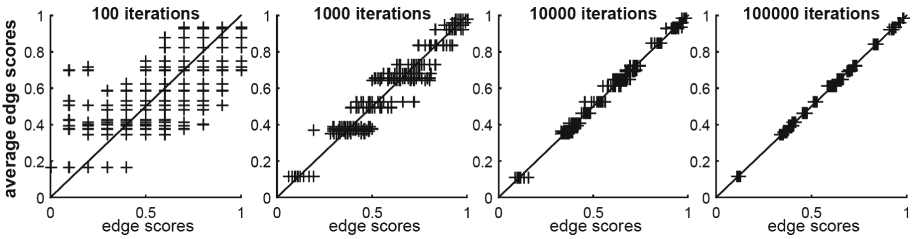

Fig. 2. Convergence diagnostic for the yeast data. For different MCMC simulation run length, $V \in \{100, 1000, 10000, 100000\}$, we performed 10 RJMCMC simulations with the new partially segment-wise coupled model. We used the hyperparameter p = 0.05 for the changepoint prior in Eq. 8. For each V there is a scatter plot where the simulation-specific edge scores (vertical axis) are plotted against the average scores for that V (horizontal axis).

For each network variable $Y := Z_j$ we generate a posterior sample from Eq. 9, $\{\lambda_u^{(j,w)}, \lambda_c^{(j,w)}, \{\delta_h\}^{(j,w)}, \boldsymbol{\pi}^{(j,w)}, \boldsymbol{\tau}^{(j,w)}\}_{w=1,\ldots,W}$, and from the covariate sets we form a network sample: $\mathcal{N}^{(w)} = \{\boldsymbol{\pi}^{(1,w)}, \ldots, \boldsymbol{\pi}^{(n,w)}\}_{w=1,\ldots,W}$.

The score $\hat{e}_{i,j} \in [0,1]$ of the edge $Z_i \rightarrow Z_j$ is the fraction of sampled networks that contain this edge; i.e. $\hat{e}_{i,j}$ is the estimated marginal posterior probability of $Z_i \rightarrow Z_j$.

If the true network is known and has M edges, we evaluate the network reconstruction accuracy as follows: for each threshold $\xi \in [0,1]$ we extract the n_ξ edges whose scores $\hat{e}_{i,j}$ exceed ξ, and we count the number of true positives T_ξ among them. Plotting the precisions $P_\xi := T_\xi / n_\xi$ against the recalls $R_\xi := T_\xi / M$, gives the precision-recall curve. We refer to the area under the precision-recall curve as AUC ('area under curve') value. The higher the AUC value, the higher the network reconstruction accuracy.

3 Hyperparameter Settings and RJMCMC Simulation Run Lengths

For the empirical cross-model comparison, we re-use the hyperparameters from the earlier works by Lèbre et al. [1] and Grzegorczyk and Husmeier [2]:

$$\sigma^{-2} \sim GAM(\alpha_\sigma = \nu, \beta_\sigma = \nu)$$

with $\nu = 0.005$, and

$$\lambda_u^{-1} \sim GAM(\alpha_u = 2, \beta_u = 0.2)$$
$$\lambda_c^{-1} \sim GAM(\alpha_c = 3, \beta_c = 3)$$

For our new partially segment-wise coupled model we use the same hyperparameters with the extension: $\delta_h \sim BER(\mathrm{p})$ with $\mathrm{p} \sim BETA(a = 1, b = 1)$.

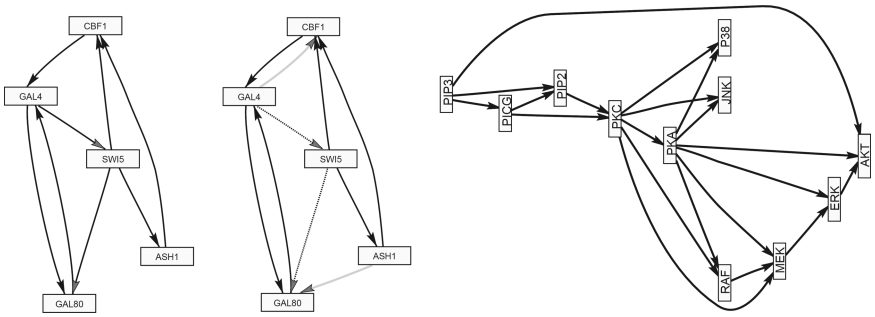

Fig. 3. Network topologies. *Left*: The true yeast network with $n = 5$ nodes and $M = 8$ edges. *Centre*: Network prediction inferred with the new partially segment-wise coupled model. We extracted the $M = 8$ edges with the highest scores. The two grey solid edges correspond to false positives. The two grey dotted edges refer to false negatives. *Right*: RAF pathway with $n = 11$ nodes and $M = 20$ edges.

For each of the three models we run RJMCMC simulation for $V = 100,000$ iterations. Setting the burn-in phase to $0.5V$ (50%) and thinning out by the factor 10 during the sampling phase, yields $W = 0.5V/10 = 5000$ samples from each posterior distribution. To check for convergence, we compared the samples of independent simulations, using standard trace plot diagnostics as well as scatter plots of the estimated edge scores. For the data sets, analyzed here, the diagnostics indicated almost perfect convergence already after $V = 10,000$ iterations; see Fig. 2 for some example scatter plot diagnostics.

4 Results

4.1 Synthetic RAF Protein Pathway Data

We generate synthetic RAF pathway data and assume the data segmentation to be known, i.e. we keep the changepoints in $\boldsymbol{\tau}$ fixed. The RAF pathway [5] has $n = 11$ nodes and $M = 20$ edges, as shown in the right panel of Fig. 3. We generate data with $H = 4$ segments having 10 data points each. For each node Z_j and its parent nodes in $\boldsymbol{\pi}^j$ we sample the regression coefficients for $h = 1$ from standard Gaussian distributions and collect them in a vector $\mathbf{w}_1^j$, which we normalize to Euclidean norm 1, $\mathbf{w}_1^j \leftarrow \mathbf{w}_1^j/|\mathbf{w}_1^j|$. For the segments $h = 2, 3, 4$ we use: $\mathbf{w}_h^j = \mathbf{w}_{h-1}^j$ ($\delta_h = 1$, coupled) or $\mathbf{w}_h^j = -\mathbf{w}_{h-1}^j$ ($\delta_h = 0$, uncoupled). The design matrices $\mathbf{X}_h^j$ contain a first column of 1's for the intercept and the segment-specific values of the parent nodes, shifted by one time point. To the values of Z_j: $\mathbf{z}_h^j = \mathbf{X}_h^j \mathbf{w}_h^j$ we add Gaussian noise with standard deviation $\sigma = 0.05$. For all eight coupling scenarios $(\delta_2, \delta_3, \delta_4) \in \{0, 1\}^3$, we generate 25 data sets with different regression coefficients.

Figure 4 compares the network reconstruction accuracies in terms of average AUC value differences. For 6 out of 8 scenarios the AUC differences are in

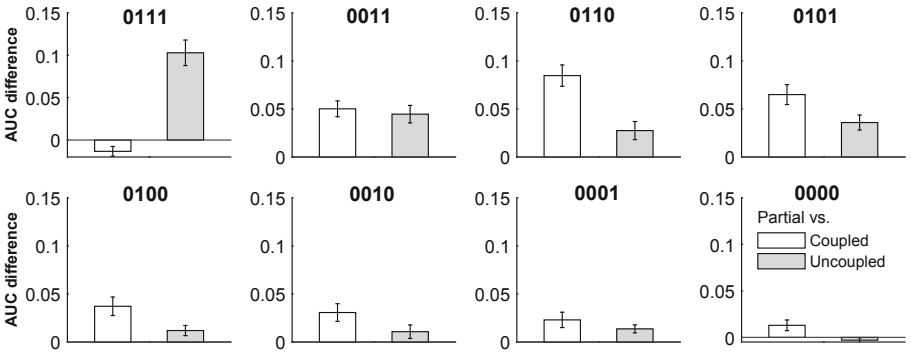

Fig. 4. Network reconstruction improvements for RAF data. For each scenario $(\delta_2, \delta_3, \delta_4)$ there are two bars of AUC differences between the new and the uncoupled model (white), and the new and the coupled model (grey). Positive values are in favor of the new model. The error bars give 95% t-test confidence intervals.

favor of the new partially coupled model. Only when all segments $h \geq 2$ are coupled ('111') or all segments are uncoupled ('000'), the new model performs slightly worse than the fully coupled or the uncoupled model, respectively. For the new partially coupled model, Fig. 5 shows the posterior probabilities that the segments $h = 2, 3, 4$ are coupled. The trends are in good agreement with the true coupling scenarios, i.e. the new model correctly infers if the regression coefficients are similar (identical) or different (opposite signs).

4.2 Saccharomyces Cerevisiae Gene Expression Data

Cantone et al. [6] synthetically designed a network in *S. cerevisiae* (yeast) with $n = 5$ genes and $M = 8$ edges, and then measured gene expression data under galactose- and glucose-metabolism: 16 measurements were taken in galactose and 21 measurements were taken in glucose. This is an ideal benchmark data set, as the network structure is known, so the network reconstruction accuracies can be cross-compared on real wet-lab data. The true network topology is shown in the left panel of Fig. 3. We pre-process the data as described in [2]. Here we assume the changepoint(s) to be unknown, so that we infer them from the data. As the number of changepoints grows with the hyperparameter p of the changepoint prior in Eq. 8, we implement the models with different values p. The average AUC scores of the models are shown in Fig. 6. The uncoupled model is consistently inferior to the new partially coupled model. The new partially coupled model also performs better than the coupled model. One exemption occurs for $\mathrm{p} = 0.1$, where the coupled model is slightly superior.

The center panel of Fig. 3 shows a network prediction that was obtained with the new partially segment-wise coupled NH-DBN model.

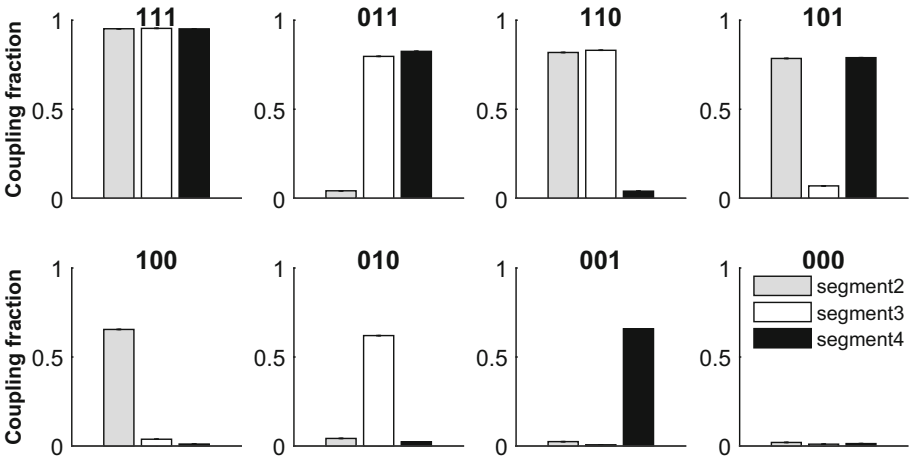

Fig. 5. Diagnostic plot for the new partially segment-wise coupled model on RAF pathway data. For each scenario $(\delta_2, \delta_3, \delta_4) \in \{0,1\}^3$ there is a bar chart of the posterior probabilities that segment h is coupled to segment $h-1$ $(h = 2,3,4)$. It can be clearly seen that the trends are in good agreement with the true underlying coupling scenarios. There are high coupling fractions for the coupled segments (with $\delta_h = 1$) and low coupling fractions for the uncoupled segments (with $\delta_h = 0$).

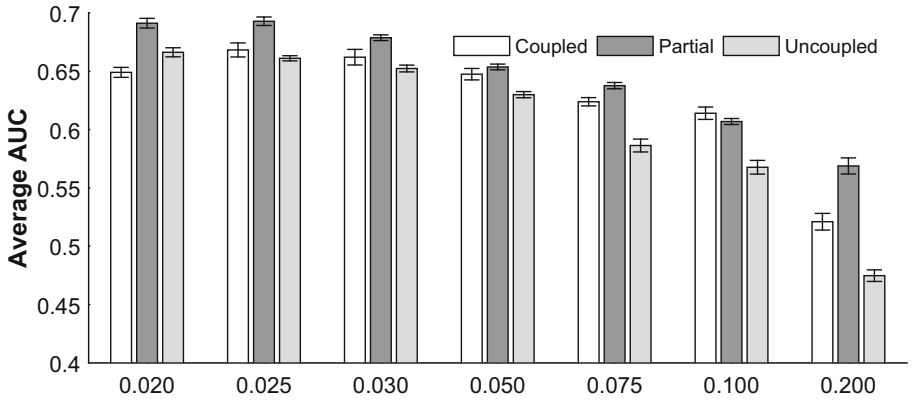

Fig. 6. Network reconstruction accuracy for *S. cerevisiae* network. We implemented the models with seven different hyperparameters p of the changepoint prior in Eq. 8. There is a bar chart for each p $\in \{0.02, , 0.025.0.03, 0.05, 0.075, 0.1, 0.2\}$, and the bars show the model-specific average AUC scores; the error bars indicate standard deviations.

4.3 Arabidopsis Thaliana Gene Expression Data

The circadian clock network in *A. thaliana* optimizes the gene regulatory processes w.r.t. the daily dark:light cycles (photo periods). In four experiments Ara-

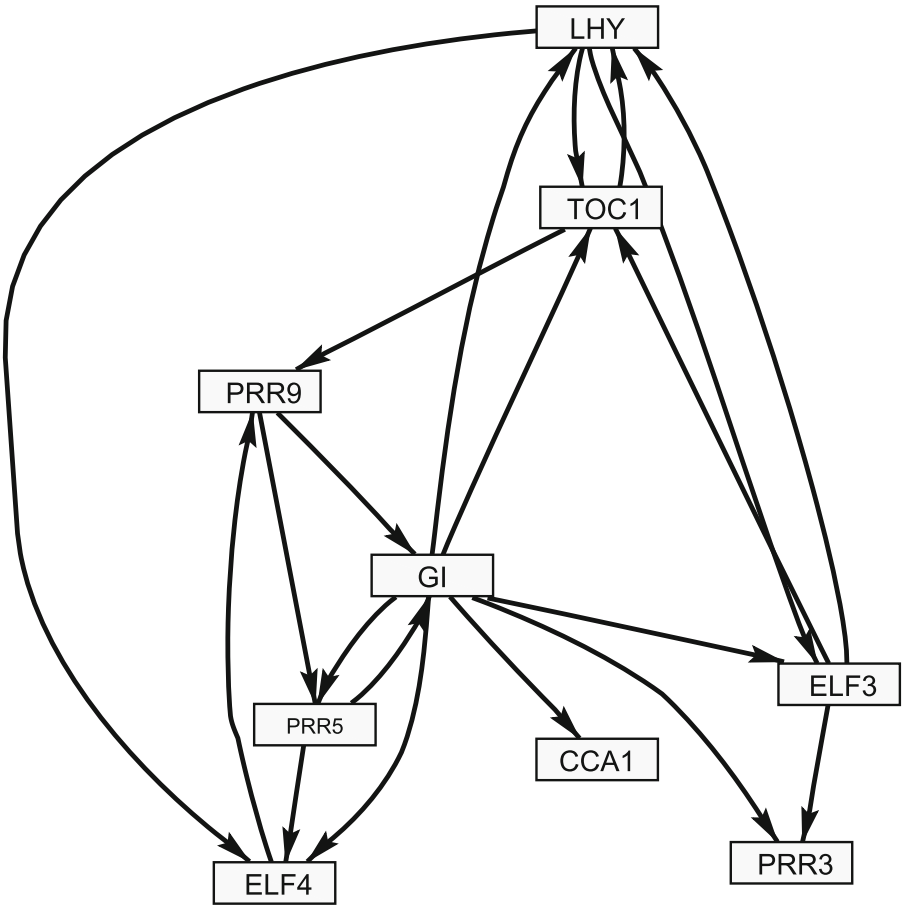

Fig. 7. Circadian clock network in *Arabidopsis thaliana*, inferred with the newly proposed partially segment-wise coupled NH-DBN model. We have implemented the model with the hyperparameter p = 0.1 for the changepoint process prior in Eq. 8. The figure shows the 20 edges with the highest marginal edge posterior probabilities (edge scores). Many of the inferred edges are consistent with the plant biology literature; see main text for further details.

bidopsis plants were entrained in different dark:light cycles, before gene expressions were measured under constant light condition over 24- and 48-hours. We follow earlier studies [2], and merge the four time series to one single data set with $T = 47$ data points and focus our attention on the $n = 9$ core genes: LHY, TOC1, CCA1, ELF4, ELF3, GI, PRR9, PRR5, and PRR3.

Here we cannot objectively cross-compare the network reconstruction accuracies, as the true underlying network topology is not known. Figure 7 shows the structure of the network that was inferred with the new partially segment-wise coupled model, using p = 0.1 for the changepoint process prior in Eq. 8. For

obtaining the prediction in Fig. 7, we extracted the 20 edges with the highest marginal edge posterior probabilities (edge scores). A proper biological evaluation of the predicted network topology is beyond the scope of this paper, but we note that many of the inferred edges are consistent with the plant biology literature. For example, the feedback loop between the two genes LHY and $TOC1$ is one of the most important key features of the circadian clock network; see, e.g., the early work by Locke et al. [8]. Many of the other predicted network edges have been reported in more recent works. For example, the five edges $LHY \rightarrow ELF3$, $LHY \rightarrow ELF4$, $GI \rightarrow TOC1$, $ELF3 \rightarrow PRR3$ and $ELF4 \rightarrow PRR9$ can all be found in the Arabidopsis circadian clock network of Herrero et al. [7].

5 Conclusion

We have proposed a new partially segment-wise coupled non-homogeneous dynamic Bayesian network model (NH-DBN). Our new model is a consensus model between the standard uncoupled NH-DBN model and the fully coupled non-homogeneous dynamic Bayesian network model from [2]; see Fig. 1 for a graphical model representation.

Our empirical results on synthetic RAF pathway data (see Fig. 4) and on *S. cerevisiae* gene expression data (see Fig. 6) show that the new partially segment-wise coupled model reaches higher network reconstruction accuracies than its competitors, namely the uncoupled NH-DBN and the fully sequentially coupled NH-DBN. We have also seen that the new partially coupled model can correctly infer from the data whether the parameters of neighboring segments are similar or dissimilar; see Fig. 5 for a diagnostic plot.

In the third and last empirical case study, we have applied the new partially segment-wise coupled model to gene expression data from the circadian clock network in *Arabidopsis thaliana*. The inferred network topology (shown in Fig. 7) is consistent with the plant biology literature.

References

1. Lèbre, S., Becq, J., Devaux, F., Lelandais, G., Stumpf, M.P.H.: Statistical inference of the time-varying structure of gene-regulation networks. BMC Syst. Biol. 4, Article 130 (2010)
2. Grzegorczyk, M., Husmeier, D.: A non-homogeneous dynamic Bayesian network with sequentially coupled interaction parameters for applications in systems and synthetic biology. Stat. Appl. Genet. Mol. Biol. (SAGMB) **11**(4), Article 7 (2012)
3. Shafiee Kamalabad, M., Grzegorczyk, M.: Improving nonhomogeneous dynamic Bayesian networks with sequentially coupled parameters. Stat. Neerl. **72**(3), 281–305 (2018)
4. Bishop, C.M.: Pattern Recognition and Machine Learning, 1st edn. Springer, Singapore (2006)
5. Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D.A., Nolan, G.P.: Causal protein-signaling networks derived from multiparameter single-cell data. Science **308**, 523–529 (2005)

6. Cantone, I., et al.: A yeast synthetic network for in Vivo assessment of reverse-engineering and modeling approaches. Cell **137**, 172–181 (2009)
7. Herrero, E., et al.: EARLY FLOWERING4 recruitment of EARLY FLOWERING3 in the nucleus sustains the Arabidopsis circadian clock. Plant Cell Online **24**(2), 428–443 (2012)
8. Locke, J.C.W., et al.: Experimental validation of a predicted feedback loop in the multi-oscillator clock of Arabidopsis thaliana. Mol. Syst. Biol. **2**(1), online article (2006)