

University of Groningen

Advanced non-homogeneous dynamic Bayesian network models for statistical analyses of time series data

Shafiee Kamalabad, Mahdi

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Shafiee Kamalabad, M. (2019). *Advanced non-homogeneous dynamic Bayesian network models for statistical analyses of time series data*. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 2

Partially sequentially segmentwise coupled NH-DBNs

A common Bayesian approach is to employ a multiple changepoint process to divide the time series into disjunct segments with segment-specific network interaction parameters and to model the data in each segment by linear Bayesian regression models. The conventional uncoupled models infer the network interaction parameters for each segment separately. There is no information-sharing among segments. The uncoupled models are very flexible, but for short time series they can be subject to inflated inference uncertainties. It was therefore proposed to couple the network interaction parameters sequentially among segments. The key idea is to enforce the parameters of any segment to stay similar to those of the previous by using the posterior expectation of the network parameters as prior expectation for the consecutive segment. Node-specific coupling parameters regulate the variance of the parameter priors and so the strength of coupling. However, the proposed models are based on coupling mechanisms which can become disadvantageous: First, the models enforce coupling for all segments without any option to uncouple, and second, they couple all pairs of neighbouring segments with the same coupling strength. In this chapter we propose two improved hierarchical Bayesian models to fix those deficits. The partially segment-wise coupled model infers for each segment whether it is coupled to (or uncoupled from) the previous segment. The generalized coupled model introduces segment-specific coupling strengths, so as to allow for a greater model flexibility.

The partially segment-wise coupled model (M3) is a consensus model between the uncoupled model and the fully coupled model. There is a discrete binary variable δ_h for each segment h indicating whether segment h is coupled to the previous segment ($\delta_h = 1$) or uncoupled from the previous segment ($\delta_h = 0$).

Along with the network structure, the changepoints, the segment-specific network interaction parameters and the values of those indicator variables are inferred from the data. The new partially coupled M3 model has the original models as limiting cases: If it couples all segments, it effectively becomes the fully coupled model. If it uncouples all segments, it effectively becomes the conventional uncoupled model. The second model, which we propose here, is the generalised (fully) coupled model, which we refer to as the M4 model. The M4 model generalises the fully coupled model by introducing segment-specific coupling strength hyperparameters. Alternatively, the M4 model can also be thought of as a continuous version of the partially segment-wise coupled model (M3). The M4 model replaces the segment-specific binary indicator variables $\delta_h \in \{0, 1\}$ (uncoupled vs. coupled) by continuous coupling hyperparameters $\lambda_h \in \mathbb{R}^+$. For each pair of neighbouring segments $(h - 1, h)$ there is then a segment-specific continuous coupling strength hyperparameter λ_h .

The work presented in this chapter, is still work in progress. Some parts of this chapter have also been appeared in proceedings of the International conference on Computational Intelligence methods for Bioinformatics and Biostatistics (2018)(see [56]) and proceedings of the International Workshop on Statistical Modelling (2016)(see [54]).

2.1 Methods

2.1.1 Learning dynamic networks with time-varying parameters

Consider a network domain with N random variables Z_1, \dots, Z_N being the nodes. Let \mathbf{D} denote a data matrix whose N rows correspond to the variables and whose $T + 1$ columns correspond to equidistant time points $t = 1, \dots, T + 1$. The element in the i -th row and t -th column, $\mathbf{D}_{i,t}$, is the value of Z_i at time point t . Temporal data are usually modelled with dynamic Bayesian networks, where all interactions are subject to a time lag: An edge $Z_i \rightarrow Z_j$ indicates that $\mathbf{D}_{j,t+1}$ (Z_j at $t + 1$) depends on $\mathbf{D}_{i,t}$ (Z_i at t). Z_i is then called a parent (node) of Z_j .

Because of the time lag, there is no acyclicity constraint in dynamic Bayesian networks, and the parent nodes of each node Z_j can be learned separately. A common approach is to use a regression model where $Y := Z_j$ is the response and the other variables $\{Z_1, \dots, Z_{j-1}, Z_{j+1}, \dots, Z_N\} =: \{X_1, \dots, X_n\}$ are the $n := N - 1$ covariates. Each data point \mathcal{D}_t ($t \in \{1, \dots, T\}$) contains a response value $Y = \mathbf{D}_{j,t+1}$ and the shifted covariate values: $X_1 = \mathbf{D}_{1,t}, \dots, X_{j-1} = \mathbf{D}_{j-1,t}, X_j = \mathbf{D}_{j+1,t}, \dots, X_n = \mathbf{D}_{N,t}$, where $n = N - 1$. Having a covariate set π_j for each Z_j , a network can be built by merging the covariate sets: $\mathcal{G} := \{\pi_1, \dots, \pi_N\}$. There is the edge $Z_i \rightarrow Z_j$ if and only if $X_i \in \pi_j$. As the same regression model is used for each Z_j separately, we describe the models (M1-M4) using the general terminology: Y is the response and X_1, \dots, X_n are the covariates.

To allow for time-dependent regression coefficients, a piece-wise linear regression model can be used. A set of changepoints $\tau := \{\tau_1, \dots, \tau_{H-1}\}$ with $1 \leq \tau_h < T$ divides the data points $\mathcal{D}_1, \dots, \mathcal{D}_T$ into disjunct segments $h = 1, \dots, H$ covering T_1, \dots, T_H consecutive data points, where $\sum T_h = T$. Data point \mathcal{D}_t ($1 \leq t \leq T$) belongs to segment h if $\tau_{h-1} < t \leq \tau_h$, where $\tau_0 := 1$ and $\tau_H := T$.

We assume all covariate sets $\pi \subset \{X_1, \dots, X_n\}$ with up to $\mathcal{F} = 3$ covariates to be equally likely a priori, $p(\pi) = c$, while parent sets with more than \mathcal{F} covariates get a zero prior probability ('fan-in restriction').¹ Further we assume that the distance between changepoints is geometrically distributed with parameter $p \in (0, 1)$, so that

$$p(\tau) = \left(\prod_{h=1}^{H-1} (1-p)^{\tau_h - \tau_{h-1} - 1} \cdot p \right) \cdot (1-p)^{\tau_H - \tau_{H-1} - 1} = (1-p)^{(T-1) - (H-1)} \cdot p^{H-1}$$

With $\mathbf{y} = \mathbf{y}_\tau := \{\mathbf{y}_1, \dots, \mathbf{y}_H\}$ being the set of segment-specific response vectors, implied by τ , the posterior distribution takes the form:

$$p(\pi, \tau, \boldsymbol{\theta} | \mathbf{y}) \propto p(\pi) \cdot p(\tau) \cdot p(\boldsymbol{\theta} | \pi, \tau) \cdot p(\mathbf{y} | \pi, \tau, \boldsymbol{\theta}) \quad (2.1)$$

where $\boldsymbol{\theta} = \boldsymbol{\theta}(\pi, \tau)$ denotes the set of all model parameters, including the segment-specific parameters and those parameters which are shared among segments.

In subsections 2.1.3 to 2.1.6 we assume $\pi \subset \{X_1, \dots, X_n\}$ and the segmentation $\mathbf{y} = \{\mathbf{y}_1, \dots, \mathbf{y}_H\}$ induced by τ to be fixed, and we do not make π and τ explicit anymore. Without loss of generality, we assume further that π contains the first k covariates: $\pi := \{X_1, \dots, X_k\}$. Focusing only on the model-parameters $\boldsymbol{\theta}$, Equation (2.1) reduces to:

$$p(\boldsymbol{\theta} | \mathbf{y}) \propto p(\boldsymbol{\theta}) \cdot p(\mathbf{y} | \boldsymbol{\theta})$$

How to infer the covariate set π and changepoint set τ from the data is subsequently described in Subsection 2.1.7.

2.1.2 A generic Bayesian piece-wise linear regression model

Consider a Bayesian regression model where Y is the response and X_1, \dots, X_k are the covariates. We assume that T data points $\mathcal{D}_1, \dots, \mathcal{D}_T$ have been measured at equidistant time points and that the data can be subdivided into disjunct segments $h \in \{1, \dots, H\}$, where segment h contains T_h data points and has the segment-specific regression coefficient vector β_h . Let \mathbf{y}_h be the response vector and \mathbf{X}_h be the design matrix for segment h , where each \mathbf{X}_h includes a first column of 1's for the intercept. For each segment $h = 1, \dots, H$ we assume a Gaussian likelihood:

$$\mathbf{y}_h | (\beta_h, \sigma^2) \sim \mathcal{N}(\mathbf{X}_h \beta_h, \sigma^2 \mathbf{I}) \quad (2.2)$$

¹The fan-in restriction is biologically motivated, as it is known that genes are rarely regulated by more than 2-3 other regulator genes [31].

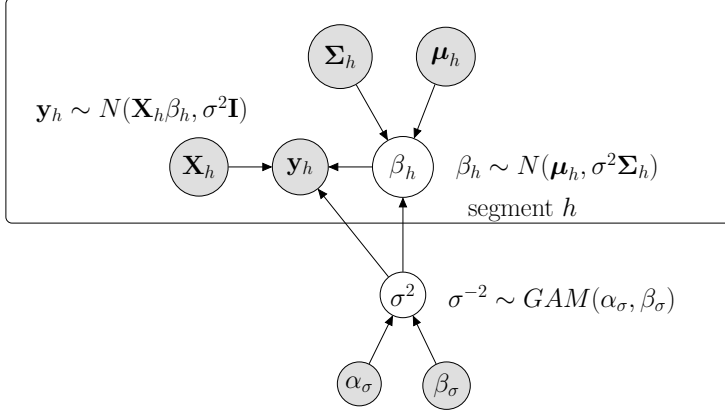


Figure 2.1: Graphical representation of the generic model. Parameters that have to be inferred are represented by white circles. The data and the fixed hyperparameters are represented by grey circles. Circles within the plate are specific for segment h .

where \mathbf{I} is the identity matrix, and σ^2 is the noise variance parameter, which is shared among segments. We impose an inverse Gamma prior on σ^2 , $\sigma^{-2} \sim \text{GAM}(\alpha_\sigma, \beta_\sigma)$, and we assume that the β_h 's have Gaussian prior distributions:

$$\beta_h | (\mu_h, \Sigma_h, \sigma^2) \sim \mathcal{N}(\mu_h, \sigma^2 \Sigma_h) \quad (2.3)$$

Re-using the parameter σ^2 in Equation (2.3), yields a fully-conjugate prior in both β_h and σ^2 (see, e.g., Sections 3.3 and 3.4 in [19]). Figure 2.1 shows a graphical model representation of this generic model. For notational convenience we define:

$$\boldsymbol{\theta} := \{\mu_1, \dots, \mu_H; \Sigma_1, \dots, \Sigma_H\}$$

The full conditional distribution of β_h is (cf. Section 3.3 in [5]):

$$\beta_h | (\mathbf{y}_h, \sigma^2, \boldsymbol{\theta}) \sim \mathcal{N}([\Sigma_h^{-1} + \mathbf{X}_h^T \mathbf{X}_h]^{-1} (\Sigma_h^{-1} \mu_h + \mathbf{X}_h^T \mathbf{y}_h), \sigma^2 (\Sigma_h^{-1} + \mathbf{X}_h^T \mathbf{X}_h)^{-1}) \quad (2.4)$$

and the segment-specific marginal likelihoods with β_h integrated out are:

$$\mathbf{y}_h | (\sigma^2, \boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{X}_h \mu_h, \sigma^2 \mathbf{C}_h(\boldsymbol{\theta})) \quad (2.5)$$

where $\mathbf{C}_h(\boldsymbol{\theta}) := \mathbf{I} + \mathbf{X}_h \Sigma_h \mathbf{X}_h^T$ (cf. Section 3.3 in [5]). From Equation (2.5) we get:

$$p(\sigma^2 | \mathbf{y}, \boldsymbol{\theta}) \propto p(\sigma^2) \cdot \prod_{h=1}^H p(\mathbf{y}_h | \sigma^2, \boldsymbol{\theta}) = (\sigma^{-2})^{\alpha_\sigma + \frac{1}{2} \cdot T - 1} e^{-\sigma^{-2} (b_\sigma + \frac{1}{2} \cdot \Delta^2(\boldsymbol{\theta}))}$$

where $\mathbf{y} := \{\mathbf{y}_1, \dots, \mathbf{y}_H\}$ and $\Delta^2(\boldsymbol{\theta}) := \sum_{h=1}^H (\mathbf{y}_h - \mathbf{X}_h \mu_h)^T \mathbf{C}_h(\boldsymbol{\theta})^{-1} (\mathbf{y}_h - \mathbf{X}_h \mu_h)$. The shape of $p(\sigma^2 | \mathbf{y}, \boldsymbol{\theta})$ implies:

$$\sigma^{-2} | (\mathbf{y}, \boldsymbol{\theta}) \sim \text{GAM} \left(\alpha_\sigma + \frac{1}{2} \cdot T, \beta_\sigma + \frac{1}{2} \cdot \Delta^2(\boldsymbol{\theta}) \right) \quad (2.6)$$

For the marginal likelihood, with β_h ($h = 1, \dots, H$) and σ^2 integrated out, we apply the rule from Section 2.3.7 in [5]:

$$p(\mathbf{y}|\boldsymbol{\theta}) = \frac{\Gamma(\frac{T}{2} + a_\sigma)}{\Gamma(a_\sigma)} \cdot \frac{\pi^{-T/2} \cdot (2b_\sigma)^{a_\sigma}}{\left(\prod_{h=1}^H \det(\mathbf{C}_h(\boldsymbol{\theta}))\right)^{1/2}} \cdot (2b_\sigma + \Delta^2(\boldsymbol{\theta}))^{-\left(\frac{T}{2} + a_\sigma\right)} \quad (2.7)$$

When all parameters in $\boldsymbol{\theta}$ are fixed, the marginal likelihood of the piecewise linear regression model can be computed in closed form. In real-world applications there is normally no prior knowledge about the hyperparameters in $\boldsymbol{\theta}$.

In typical models the (hyper-)hyperparameters in $\boldsymbol{\theta}$ do not have all degrees of freedom, but depend on some free hyperparameters with their own hyperprior distributions. From now on, we will only include the free hyperparameters in $\boldsymbol{\theta}$. In the following subsections we describe four concrete model instantiations: the uncoupled model (M1), the fully coupled model (M2), the partially segment-wise coupled model (M3) and the generalised coupled model (M4), where the latter two are proposed in this chapter.

2.1.3 Model M1: The uncoupled model

A standard approach, akin to the models of [38] and [14], is to set $\boldsymbol{\mu}_h = \mathbf{0}$ and to assume that the matrices $\boldsymbol{\Sigma}_h$ are diagonal matrices $\boldsymbol{\Sigma}_h = \lambda_u \mathbf{I}$, where the parameter $\lambda_u \in \mathbf{R}^+$ is shared among segments and assumed to be inverse Gamma distributed, $\lambda_u^{-1} \sim \text{GAM}(\alpha_u, \beta_u)$. Figure 2.2 shows the uncoupled model (M1) graphically. Using the notation of the generic model, we have:

$$\boldsymbol{\theta} = \{\lambda_u\}, \quad \mathbf{C}_h(\lambda_u) = \mathbf{I} + \lambda_u \mathbf{X}_h \mathbf{X}_h^\top, \quad \Delta^2(\lambda_u) := \sum_{h=1}^H \mathbf{y}_h^\top \mathbf{C}_h(\lambda_u)^{-1} \mathbf{y}_h \quad (2.8)$$

For the posterior distribution of the uncoupled model we have:

$$p(\boldsymbol{\beta}, \sigma^2, \lambda_u | \mathbf{y}) \propto p(\sigma^2) \cdot p(\lambda_u) \cdot \prod_{h=1}^H p(\boldsymbol{\beta}_h | \sigma^2, \lambda_u) \cdot \prod_{h=1}^H p(\mathbf{y}_h | \sigma^2, \boldsymbol{\beta}_h) \quad (2.9)$$

where $\boldsymbol{\beta} := \{\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_H\}$. From Equation (2.9) it follows for the full conditional distribution of λ_u :

$$\begin{aligned} p(\lambda_u | \mathbf{y}, \boldsymbol{\beta}, \sigma^2) &\propto p(\lambda_u) \cdot \prod_{h=1}^H p(\boldsymbol{\beta}_h | \sigma^2, \lambda_u) \\ &\propto (\lambda_u^{-1})^{\alpha_u + \frac{H \cdot (k+1)}{2}} \cdot \exp\left\{-\lambda_u^{-1} \left(b_u + \frac{1}{2} \sigma^{-2} \sum_{h=1}^H \boldsymbol{\beta}_h^\top \boldsymbol{\beta}_h\right)\right\} \end{aligned}$$

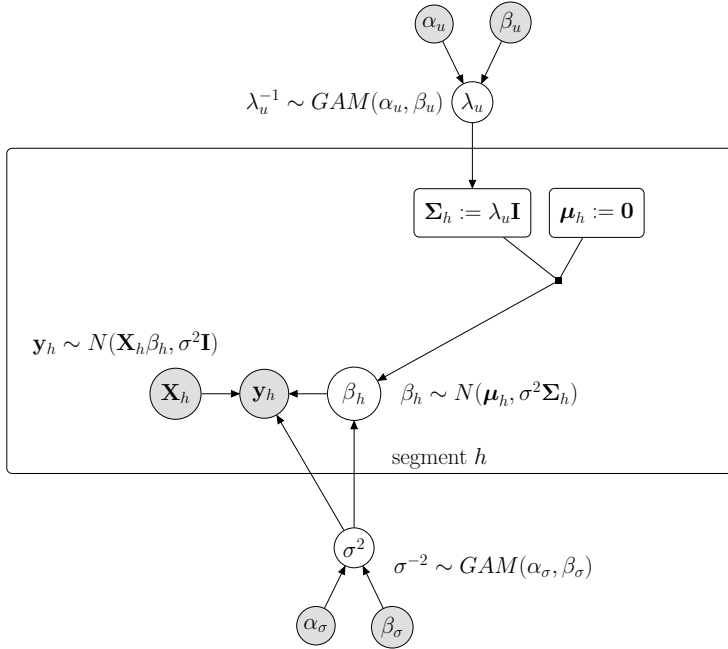


Figure 2.2: Graphical representation of the uncoupled model (M1). Parameters that have to be inferred are represented by white circles. The data and the fixed hyperparameters are represented by grey circles. The two rectangles indicate definitions, which deterministically depend on the parent nodes. Circles and definitions within the plate are segment-specific.

and the shape of the latter density implies:

$$\lambda_u^{-1} | (\mathbf{y}, \boldsymbol{\beta}, \sigma^2) \sim \text{GAM} \left(\alpha_u + \frac{H \cdot (k+1)}{2}, \beta_u + \frac{1}{2} \sigma^{-2} \sum_{h=1}^H \boldsymbol{\beta}_h^\top \boldsymbol{\beta}_h \right) \quad (2.10)$$

Since the full conditional distribution of λ_u depends on σ^2 and $\boldsymbol{\beta}$, those parameters have to be sampled first. From Equation (2.6) an instantiation of σ^2 can be sampled via a collapsed Gibbs-sampling step, with the $\boldsymbol{\beta}_h$'s being integrated out. Subsequently, given σ^2 , Equation (2.4) can be used to sample the $\boldsymbol{\beta}_h$'s. Finally, for each λ_u sampled from Equation (2.10) the marginal likelihood, $p(\mathbf{y} | \lambda_u)$, can be computed by plugging in the expressions from Equation (2.8) into Equation (2.7).

2.1.4 Model M2: The fully coupled model

The (fully) coupled model, proposed by [24], uses the posterior expectation of $\boldsymbol{\beta}_{h-1}$ as prior expectation for $\boldsymbol{\beta}_h$. Only the first segment $h = 1$ has an uninform-

ative prior:

$$\beta_h \sim \begin{cases} \mathcal{N}(\mathbf{0}, \sigma^2 \lambda_u \mathbf{I}) & \text{if } h = 1 \\ \mathcal{N}(\tilde{\beta}_{h-1}, \sigma^2 \lambda_c \mathbf{I}) & \text{if } h > 1 \end{cases} \quad (2.11)$$

where $\tilde{\beta}_{h-1}$ is the posterior expectation of β_{h-1} (cf. Equation (2.4))²

$$\tilde{\beta}_{h-1} := \begin{cases} [\Sigma_1^{-1} + \mathbf{X}_1^\top \mathbf{X}_1]^{-1} (\mathbf{X}_1^\top \mathbf{y}_1) & \text{if } h = 2 \\ [\Sigma_{h-1}^{-1} + \mathbf{X}_{h-1}^\top \mathbf{X}_{h-1}]^{-1} (\lambda_c^{-1} \tilde{\beta}_{h-2} + \mathbf{X}_{h-1}^\top \mathbf{y}_{h-1}) & \text{if } h > 2 \end{cases}$$

The parameter λ_c has been called the ‘*coupling parameter*’ onto which also an inverse Gamma distribution can be imposed, $\lambda_c^{-1} \sim GAM(\alpha_c, \beta_c)$. Using the notation from the generic model (see Figure 2.1), we note that Equation (2.11) corresponds to:

$$\mu_h = \begin{cases} \mathbf{0} & \text{if } h = 1 \\ \tilde{\beta}_{h-1} & \text{if } h > 1 \end{cases}, \quad \Sigma_h = \begin{cases} \lambda_u \mathbf{I} & \text{if } h = 1 \\ \lambda_c \mathbf{I} & \text{if } h > 1 \end{cases}, \quad \mathbf{C}_h(\theta) = \begin{cases} \mathbf{I} + \lambda_u \mathbf{X}_h \mathbf{X}_h^\top & \text{if } h = 1 \\ \mathbf{I} + \lambda_c \mathbf{X}_h \mathbf{X}_h^\top & \text{if } h > 1 \end{cases}$$

$$\theta = \{\lambda_u, \lambda_c\}, \quad \text{and} \quad \Delta^2(\theta) = \sum_{h=1}^H (\mathbf{y}_h - \mathbf{X}_h \tilde{\beta}_{h-1})^\top \mathbf{C}_h(\theta)^{-1} (\mathbf{y}_h - \mathbf{X}_h \tilde{\beta}_{h-1})$$

where $\tilde{\beta}_0 := \mathbf{0}$, $\lambda_u^{-1} \sim GAM(\alpha_u, \beta_u)$ and $\lambda_c^{-1} \sim GAM(\alpha_c, \beta_c)$. As $\tilde{\beta}_{h-1}$ is treated like a fixed hyperparameter when used as input for segment h , we exclude the parameters $\tilde{\beta}_1, \dots, \tilde{\beta}_{H-1}$ from θ . Figure 2.3 shows a graphical representation of the coupled model.

For the posterior we have:

$$p(\beta, \sigma^2, \lambda_u, \lambda_c | \mathbf{y}) \propto p(\sigma^2) \cdot p(\lambda_u) \cdot p(\lambda_c) \cdot p(\beta_1 | \sigma^2, \lambda_u) \quad (2.12)$$

$$\cdot \prod_{h=2}^H p(\beta_h | \sigma^2, \lambda_c) \cdot \prod_{h=1}^H p(\mathbf{y}_h | \sigma^2, \beta_h)$$

In analogy to the derivations in the previous subsection one can derive (cf. [24]):

$$\lambda_u^{-1} | (\mathbf{y}, \beta, \sigma^2, \lambda_c) \sim GAM\left(\alpha_u + \frac{1 \cdot (k+1)}{2}, \beta_u + \frac{1}{2} \sigma^{-2} D_u^2\right) \quad (2.13)$$

$$\lambda_c^{-1} | (\mathbf{y}, \beta, \sigma^2, \lambda_u) \sim GAM\left(\alpha_c + \frac{(H-1) \cdot (k+1)}{2}, \beta_c + \frac{1}{2} \sigma^{-2} D_c^2\right) \quad (2.14)$$

where $D_u^2 := \beta_1^\top \beta_1$ and $D_c^2 := \sum_{h=2}^H (\beta_h - \tilde{\beta}_{h-1})^\top (\beta_h - \tilde{\beta}_{h-1})$.

For each $\theta = \{\lambda_u, \lambda_c\}$ the marginal likelihood, $p(\mathbf{y} | \lambda_u, \lambda_c)$, can be computed by plugging the expressions $\mathbf{C}_h(\theta)$ and $\Delta^2(\theta)$ into Equation (2.7).

2.1.5 Model M3: The partially segment-wise coupled model, proposed here

The first new model, which we propose here, allows each segment $h > 1$ to uncouple from the previous one $h - 1$. We use an uninformative prior for segment $h = 1$, and for all

²Note: $\tilde{\beta}_{h-1}$ in Equation (2.4) is the posterior expectation of β_{h-1} given σ^2 and $\theta = \{\lambda_u, \lambda_c\}$.

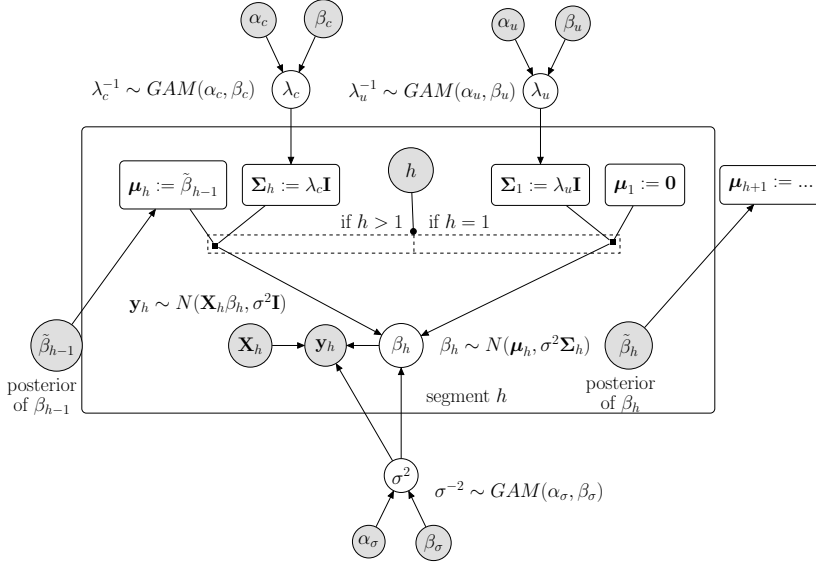


Figure 2.3: Graphical representation of the fully coupled model (M2). See caption of Figure 2.2 for the terminology. Each posterior expectation $\tilde{\beta}_h$ is treated like a fixed parameter vector when used as input for segment $h + 1$. The prior for β_h depends on h .

segments $h > 1$ we introduce a binary variable δ_h which indicates whether segment h is coupled to ($\delta_h = 1$) or uncoupled from ($\delta_h = 0$) the preceding segment $h - 1$:

$$\beta_h \sim \begin{cases} \mathcal{N}(\mathbf{0}, \sigma^2 \lambda_u \mathbf{I}) & \text{if } h = 1 \\ \mathcal{N}(\delta_h \cdot \tilde{\beta}_{h-1}, \sigma^2 \lambda_c^{\delta_h} \lambda_u^{1-\delta_h} \mathbf{I}) & \text{if } h > 1 \end{cases} \quad (2.15)$$

where $\tilde{\beta}_{h-1}$ is again the posterior expectation of β_{h-1} . The new priors from Equation (2.15) yield for $h \geq 2$ the following posterior expectations (cf. Equation (2.4)):

$$\tilde{\beta}_{h-1} = \left(\lambda_c^{-\delta_{h-1}} \lambda_u^{-(1-\delta_{h-1})} \mathbf{I} + \mathbf{X}_{h-1}^\top \mathbf{X}_{h-1} \right)^{-1} \left(\delta_{h-1} \lambda_c^{-1} \tilde{\beta}_{h-2} + \mathbf{X}_{h-1}^\top \mathbf{y}_{h-1} \right)$$

With $\tilde{\beta}_0 := \mathbf{0}$, $\delta_1 := 0$, we have in the generic model notation:

$$\boldsymbol{\mu}_h = \delta_h \tilde{\beta}_{h-1}, \quad \boldsymbol{\Sigma}_h = \lambda_c^{\delta_h} \lambda_u^{1-\delta_h} \mathbf{I}, \quad \boldsymbol{\theta} = \{\lambda_u, \lambda_c, \{\delta_h\}_{h \geq 2}\}, \quad \mathbf{C}_h(\boldsymbol{\theta}) = \mathbf{I} + \lambda_c^{\delta_h} \lambda_u^{1-\delta_h} \mathbf{X}_h \mathbf{X}_h^\top$$

We assume the binary variables $\delta_2, \dots, \delta_H$ to be Bernoulli distributed, $\delta_h \sim BER(p)$, with $p \in [0, 1]$ having a Beta distribution, $p \sim BETA(a, b)$.

- $\delta_h = 0$ ($h \geq 2$) gives model M1 with $P(\beta_h) = \mathcal{N}(\mathbf{0}, \lambda_u \sigma^2 \mathbf{I})$ for all h
- $\delta_h = 1$ ($h \geq 2$) gives model M2 with $P(\beta_h) = \mathcal{N}(\tilde{\beta}_{h-1}, \lambda_c \sigma^2 \mathbf{I})$ for $h \geq 2$.
- The new partially segment-wise coupled model infers the variables δ_h ($h \geq 2$) from the data, so as to find the best trade-off between model M1 and model M2.

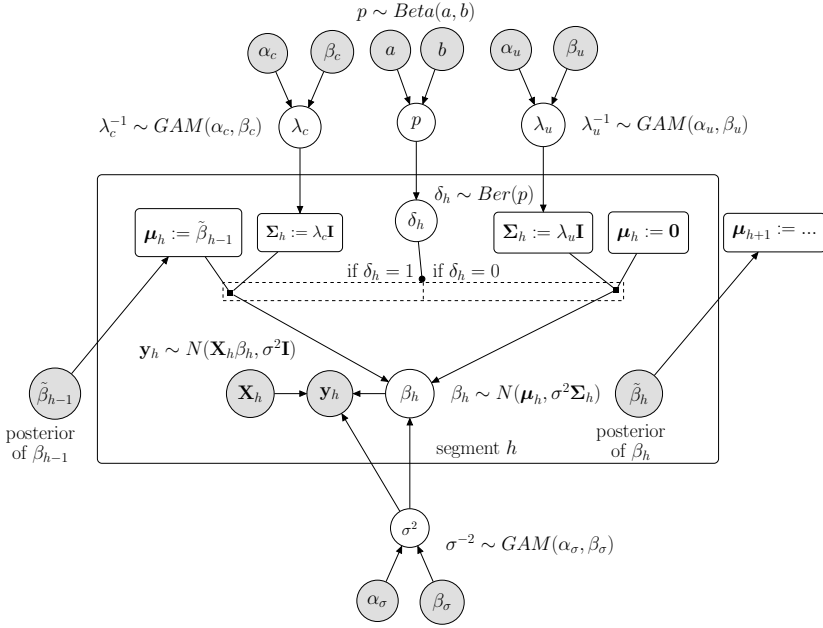


Figure 2.4: Graphical representation of the partially coupled model (M3), proposed here. See caption of Figure 2.3 for the terminology. For each segment h it is inferred from the data whether the prior for β_h should be coupled to ($\delta_h = 1$) or uncoupled from ($\delta_h = 0$) the preceding segment $h - 1$.

A graphical model presentation of the partially coupled model is shown in Figure 2.4. For $\delta_h \sim \text{BER}(p)$ with $p \sim \text{BETA}(a, b)$ the joint marginal density of $\{\delta_h\}_{h \geq 2}$ is:

$$p(\{\delta_h\}_{h \geq 2}) = \int p(p) \prod_{h=2}^H p(\delta_h | p) dp = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a + \sum_{h=2}^H \delta_h) \Gamma(b + \sum_{h=2}^H (1 - \delta_h))}{\Gamma(a+b + (H-1))}$$

For the posterior distribution of the partially segment-wise coupled model we get:

$$p(\beta, \sigma^2, \lambda_u, \lambda_c, \{\delta_h\}_{h \geq 2} | \mathbf{y}) \propto p(\sigma^2) \cdot p(\lambda_u) \cdot p(\lambda_c) \cdot p(\{\delta_h\}_{h \geq 2}) \cdot p(\beta_1 | \sigma^2, \lambda_u) \cdot \prod_{h=2}^H p(\beta_h | \sigma^2, \lambda_u, \lambda_c, \delta_h) \cdot \prod_{h=1}^H p(\mathbf{y}_h | \sigma^2, \beta_h)$$

For the full conditional distributions of λ_u and λ_c we have:

$$p(\lambda_u | \mathbf{y}, \beta, \sigma^2, \lambda_c, \{\delta_h\}_{h \geq 2}) \propto p(\lambda_u) \cdot \prod_{h: \delta_h=0} p(\beta_h | \sigma^2, \lambda_u)$$

$$p(\lambda_c | \mathbf{y}, \beta, \sigma^2, \lambda_u, \{\delta_h\}_{h \geq 2}) \propto p(\lambda_c) \cdot \prod_{h: \delta_h=1} p(\beta_h | \sigma^2, \lambda_c)$$

where $\delta_1 := 0$ fixed. And it follows from the shapes of the densities:

$$\begin{aligned}\lambda_u^{-1}(\mathbf{y}, \boldsymbol{\beta}, \sigma^2, \lambda_c, \{\delta_h\}_{h \geq 2}) &\sim \text{GAM}\left(\alpha_u + \frac{H_u \cdot (k+1)}{2}, \beta_u + \frac{1}{2}\sigma^{-2}D_u^2\right) \\ \lambda_c^{-1}(\mathbf{y}, \boldsymbol{\beta}, \sigma^2, \lambda_u, \{\delta_h\}_{h \geq 2}) &\sim \text{GAM}\left(\alpha_c + \frac{H_c \cdot (k+1)}{2}, \beta_c + \frac{1}{2}\sigma^{-2}D_c^2\right)\end{aligned}$$

where $H_c = \sum_h \delta_h$ is the number of coupled segments, $H_u = \sum_h (1 - \delta_h)$ is the number of uncoupled segments, so that $H_c + H_u = H$, and

$$D_u^2 := \sum_{h:\delta_h=0} \boldsymbol{\beta}_h^\top \boldsymbol{\beta}_h, \quad D_c^2 := \sum_{h:\delta_h=1} (\boldsymbol{\beta}_h - \tilde{\boldsymbol{\beta}}_{h-1})^\top (\boldsymbol{\beta}_h - \tilde{\boldsymbol{\beta}}_{h-1}) \quad (2.16)$$

For each parameter instantiation $\boldsymbol{\theta} = \{\lambda_u, \lambda_c, \{\delta_h\}_{h \geq 2}\}$ the marginal likelihood, $p(\mathbf{y}|\boldsymbol{\theta})$, can be computed with Equation (2.7), where $\mathbf{C}_h(\boldsymbol{\theta})$ was defined above, and

$$\Delta^2(\boldsymbol{\theta}) = \sum_{h=1}^H (\mathbf{y}_h - \delta_h \mathbf{X}_h \tilde{\boldsymbol{\beta}}_{h-1})^\top [\mathbf{I} + \lambda_c^{\delta_h} \lambda_u^{1-\delta_h} \mathbf{X}_h \mathbf{X}_h^\top]^{-1} (\mathbf{y}_h - \delta_h \mathbf{X}_h \tilde{\boldsymbol{\beta}}_{h-1})$$

Moreover, we have for each binary variable δ_k ($k = 2, \dots, H$):

$$p(\delta_k = 1 | \lambda_u, \lambda_c, \{\delta_h\}_{h \neq k}, \mathbf{y}) \propto p(\mathbf{y} | \lambda_u, \lambda_c, \{\delta_h\}_{h \neq k}, \delta_k = 1) \cdot p(\{\delta_h\}_{h \neq k}, \delta_k = 1)$$

so that its full conditional distribution is:

$$\delta_k | (\lambda_u, \lambda_c, \{\delta_h\}_{h \neq k}, \mathbf{y}) \sim \text{BER}\left(\frac{p(\mathbf{y} | \lambda_u, \lambda_c, \{\delta_h\}_{h \neq k}, \delta_k = 1) \cdot p(\{\delta_h\}_{h \neq k}, \delta_k = 1)}{\sum_{j=0}^1 p(\mathbf{y} | \lambda_u, \lambda_c, \{\delta_h\}_{h \neq k}, \delta_k = j) \cdot p(\{\delta_h\}_{h \neq k}, \delta_k = j)}\right)$$

Thus, each δ_k ($k > 1$) can be sampled with a collapsed Gibbs sampling step where $\{\boldsymbol{\beta}_h\}$, σ^2 and p have been integrated out.

2.1.6 Model M4: The generalised coupled model, proposed here

We generalise the fully coupled model M2 by introducing segment-specific coupling parameters λ_h for the segments $h = 2, \dots, H$. This yields:

$$\boldsymbol{\beta}_h \sim \begin{cases} \mathcal{N}(\mathbf{0}, \sigma^2 \lambda_u \mathbf{I}) & \text{if } h = 1 \\ \mathcal{N}(\tilde{\boldsymbol{\beta}}_{h-1}, \sigma^2 \lambda_h \mathbf{I}) & \text{if } h > 1 \end{cases} \quad (2.17)$$

where $\tilde{\boldsymbol{\beta}}_{h-1}$ is again the posterior expectation of $\boldsymbol{\beta}_{h-1}$. For the parameters λ_h we assume that they are inverse Gamma distributed, $\lambda_h^{-1} \sim \text{GAM}(\alpha_c, \beta_c)$, with hyperparameters α_c and β_c . Figure 2.5 gives a graphical model representation. Recalling the generic notation and setting $\tilde{\boldsymbol{\beta}}_0 := \mathbf{0}$ and $\lambda_1 := \lambda_u$, Equation (2.17) gives:

$$\begin{aligned}\boldsymbol{\mu}_h &= \tilde{\boldsymbol{\beta}}_{h-1}, \quad \boldsymbol{\Sigma}_h = \lambda_h \mathbf{I}, \quad \mathbf{C}_h(\boldsymbol{\theta}) = \mathbf{I} + \lambda_h \mathbf{X}_h \mathbf{X}_h^\top, \quad \boldsymbol{\theta} = \{\lambda_u, \{\lambda_h\}_{h \geq 2}\}, \\ \text{and } \Delta^2(\boldsymbol{\theta}) &= \sum_{h=1}^H (\mathbf{y}_h - \mathbf{X}_h \tilde{\boldsymbol{\beta}}_{h-1})^\top \mathbf{C}_h(\boldsymbol{\theta})^{-1} (\mathbf{y}_h - \mathbf{X}_h \tilde{\boldsymbol{\beta}}_{h-1})\end{aligned}$$

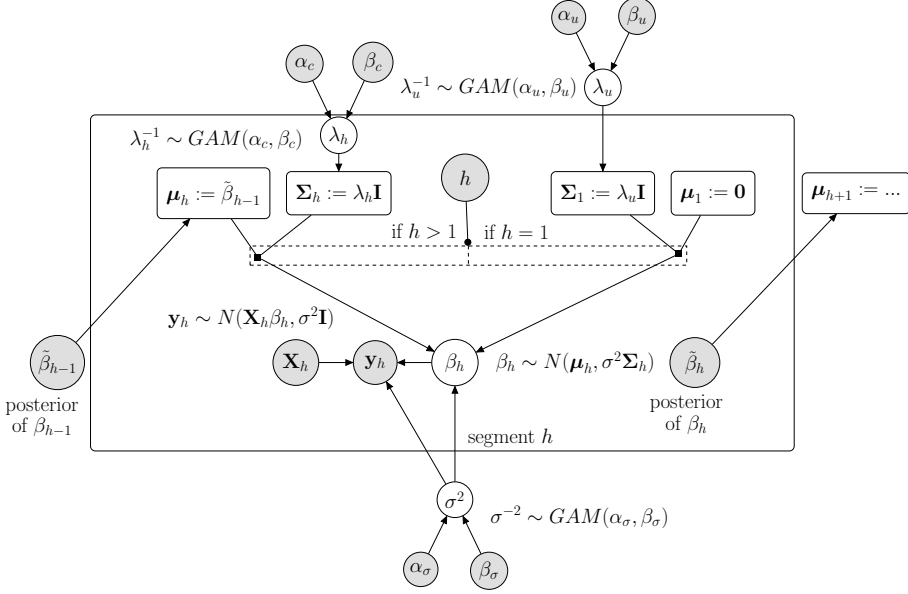


Figure 2.5: Graphical representation of the generalised coupled model (M4), proposed here. See caption of Figures 2.2-2.3 for the terminology. Unlike the M2 model, this new model has segment-specific coupling parameters λ_h ($h > 1$).

For the posterior we have:

$$\begin{aligned}
 p(\beta, \sigma^2, \lambda_u, \{\lambda_h\}_{h \geq 2} | \mathbf{y}) &\propto p(\sigma^2) \cdot p(\lambda_u) \cdot \left(\prod_{h=2}^H p(\lambda_h) \right) \\
 &\quad \cdot p(\beta_1 | \sigma^2, \lambda_u) \cdot \prod_{h=2}^H p(\beta_h | \sigma^2, \lambda_h) \cdot \prod_{h=1}^H p(\mathbf{y}_h | \sigma^2, \beta_h)
 \end{aligned} \tag{2.18}$$

Like for the coupled model M2, it can be derived for $k = 2, \dots, H$:

$$\begin{aligned}
 \lambda_k^{-1} | (\mathbf{y}, \beta, \sigma^2, \lambda_u, \{\lambda_h\}_{h \neq k}) &\sim \text{GAM} \left(\alpha_c + \frac{(k+1)}{2}, \beta_c + \frac{1}{2} \sigma^{-2} D_k^2 \right) \\
 \text{and } \lambda_u^{-1} | (\mathbf{y}, \beta, \sigma^2, \{\lambda_h\}_{h \geq 2}) &\sim \text{GAM} \left(\alpha_u + \frac{(k+1)}{2}, \beta_u + \frac{1}{2} \sigma^{-2} D_u^2 \right)
 \end{aligned}$$

where $D_u^2 := \beta_1^\top \beta_1$ and $D_k^2 := (\beta_k - \tilde{\beta}_{k-1})^\top (\beta_k - \tilde{\beta}_{k-1})$.

For each $\theta = \{\lambda_u, \{\lambda_h\}_{h \geq 2}\}$ the marginal likelihood, $p(\mathbf{y} | \{\lambda_u, \{\lambda_h\}_{h \geq 2}\})$, can be computed with Equation (2.7); using the expressions $C_h(\theta)$ and $\Delta^2(\theta)$ defined above.

2.1.7 Reversible Jump Markov Chain Monte Carlo inference

We use Reversible Jump Markov Chain Monte Carlo simulations [21] to generate posterior samples $\{\pi^{(w)}, \tau^{(w)}, \theta^{(w)}\}_{w=1, \dots, W}$. In each iteration we re-sample the parameters in θ

from their full conditional distributions (Gibbs sampling), and we perform two Metropolis-Hastings moves; one on the covariate set π and one on the changepoint set τ . For the four models Equation (2.1) takes the form:³

$$p(\pi, \tau, \theta | \mathbf{y}) \propto \begin{cases} p(\pi) \cdot p(\tau) \cdot p(\lambda_u) \cdot p(\mathbf{y} | \pi, \tau, \lambda_u) & \text{(M1)} \\ p(\pi) \cdot p(\tau) \cdot p(\lambda_u) \cdot p(\lambda_c) \cdot p(\mathbf{y} | \pi, \tau, \lambda_u, \lambda_c) & \text{(M2)} \\ p(\pi) \cdot p(\tau) \cdot p(\lambda_u) \cdot p(\lambda_c) \cdot p(\{\delta_h\}_{h \geq 2}) \cdot p(\mathbf{y} | \pi, \tau, \lambda_u, \lambda_c, \{\delta_h\}_{h \geq 2}) & \text{(M3)} \\ p(\pi) \cdot p(\tau) \cdot p(\lambda_u) \cdot \left(\prod_{h=2}^H p(\lambda_h)\right) \cdot p(\mathbf{y} | \pi, \tau, \lambda_u, \{\lambda_h\}_{h \geq 2}) & \text{(M4)} \end{cases}$$

For the models M1-M2 the dimension of θ does not depend on τ . For the new models M3-M4 the dimension of θ depends on τ . Model M3 has a discrete parameter $\delta_h \in \{0, 1\}$ for each segment $h > 1$. Model M4 has a continuous parameter $\lambda_h \in \mathbb{R}^+$ for each segment $h > 1$. That is, we have three standard cases of RJMCMC:

M1: $\theta = \{\lambda_u\}$. All segment-specific parameters can be integrated out.

M2: $\theta = \{\lambda_u, \lambda_c\}$. All segment-specific parameters can be integrated out.

M3: $\theta = \{\lambda_u, \lambda_c, \{\delta_h\}_{h \geq 2}\}$. All segment-specific parameters can be integrated out, except for a set of discrete parameters $\{\delta_h\}_{h \geq 2}$ whose cardinality depends on H .

M4: $\theta = \{\lambda_u, \{\lambda_h\}_{h \geq 2}\}$. All segment-specific parameters can be integrated out, except for a set of continuous parameters $\{\lambda_h\}_{h \geq 2}$ whose cardinality depends on H .

The model-specific full conditional distributions for the Gibbs sampling steps have already been provided in subsections 2.1.3 to 2.1.6. For sampling covariate sets π we implement 3 moves: the covariate ‘removal (R)’, ‘addition (A)’, and ‘exchange (E)’ move. Each move proposes to replace π by a new covariate set π^* having one covariate more (A) or less (R) or exchanged (E). When randomly selecting the move type and the involved covariate(s), we get for all models the acceptance probability:

$$A(\pi \rightarrow \pi^*) = \min \left\{ 1, \frac{p(\mathbf{y} | \pi^*, \dots)}{p(\mathbf{y} | \pi, \dots)} \cdot \frac{p(\pi^*)}{p(\pi)} \cdot HR_{\pi} \right\}$$

$$\text{with the Hastings Ratios: } HR_{\pi,R} = \frac{|\pi|}{n - |\pi^*|}, \quad HR_{\pi,A} = \frac{n - |\pi|}{|\pi^*|}, \quad HR_{\pi,E} = 1$$

For sampling changepoint sets τ we also implement 3 move types: the changepoint ‘birth (B)’, ‘death (D)’, and ‘re-allocation (R)’ move. Each move proposes to replace τ by a new changepoint set τ^* having one changepoint added (B) or deleted (D) or re-allocated (R). When randomly selecting the move type, the involved changepoint and the new changepoint location, we get for the models M1 and M2:

$$A(\tau \rightarrow \tau^*) = \min \left\{ 1, \frac{p(\mathbf{y} | \tau^*, \dots)}{p(\mathbf{y} | \tau, \dots)} \cdot \frac{p(\tau^*)}{p(\tau)} \cdot HR_{\tau} \right\}$$

$$\text{with the Hasting Ratios: } HR_{\tau,B} = \frac{T - 1 - |\tau^*|}{|\tau|}, \quad HR_{\tau,D} = \frac{|\tau^*|}{T - 1 - |\tau|}, \quad HR_{\tau,R} = 1$$

For the new models M3 and M4 the changepoint moves also affect the numbers of parameters in $\{\delta_h\}_{h \geq 2}$ and $\{\lambda_h\}_{h \geq 2}$, respectively. For segments that stay identical we keep the

³The likelihoods in the posterior distributions are marginalized over the regression coefficient vectors $\{\beta_h\}$ and the variance σ^2 . In M3, where $\delta_h \sim BER(p)$, the parameter p is also integrated out.

parameters unchanged. For altering segments we re-sample the corresponding parameters. For M3 we flip coins to get candidates for the involved δ_h 's. This yields:

$$A([\boldsymbol{\tau}, \{\delta_h\}] \rightarrow [\boldsymbol{\tau}^*, \{\delta_h\}^*]) = \min \left\{ 1, \frac{p(\mathbf{y}|\boldsymbol{\tau}^*, \{\delta_h\}^*, \dots) p(\boldsymbol{\tau}^*) p(\{\delta_h\}^*)}{p(\mathbf{y}|\boldsymbol{\tau}, \{\delta_h\}, \dots) p(\boldsymbol{\tau}) p(\{\delta_h\})} \cdot HR_{\boldsymbol{\tau}} \cdot c_{\boldsymbol{\tau}} \right\}$$

where $c_{\boldsymbol{\tau},B} = 1/2$ for birth, $c_{\boldsymbol{\tau},D} = 2$ for death, and $c_{\boldsymbol{\tau},R} = 1$ for re-allocation moves. For M4 we re-sample the involved λ_h 's from their priors $p(\lambda_h)$. We obtain:

$$A([\boldsymbol{\tau}, \{\lambda_h\}] \rightarrow [\boldsymbol{\tau}^*, \{\lambda_h\}^*]) = \min \left\{ 1, \frac{p(\mathbf{y}|\boldsymbol{\tau}^*, \{\lambda_h\}^*, \dots) \cdot p(\boldsymbol{\tau}^*)}{p(\mathbf{y}|\boldsymbol{\tau}, \{\lambda_h\}, \dots) \cdot p(\boldsymbol{\tau})} \cdot HR_{\boldsymbol{\tau}} \right\}$$

since the new ratio in $HR_{\boldsymbol{\tau}}$ cancels with the prior ratio $\frac{p(\{\lambda_h\}^*)}{p(\{\lambda_h\})}$.

2.1.8 Edge scores and areas under precision-recall curves (AUC)

For a network with N variables Z_1, \dots, Z_N we infer N separate regression models. For each Z_i we get a sample $\{\boldsymbol{\pi}^{(w)}, \boldsymbol{\tau}^{(w)}, \boldsymbol{\theta}^{(w)}\}_{w=1, \dots, W}$ from the i -th posterior distribution. From the covariate sets we form a sample of graphs $G^{(w)} = \{\boldsymbol{\pi}_1^{(w)}, \dots, \boldsymbol{\pi}_N^{(w)}\}_{w=1, \dots, W}$. The marginal posterior probability that the network has the edge $Z_i \rightarrow Z_j$ is:

$$\hat{e}_{i,j} = \frac{1}{W} \sum_{w=1}^W I_{i \rightarrow j}(G^{(w)}) \quad \text{where} \quad I_{i \rightarrow j}(G^{(w)}) = \begin{cases} 1 & \text{if } X_i \in \boldsymbol{\pi}_j^{(w)} \\ 0 & \text{if } X_i \notin \boldsymbol{\pi}_j^{(w)} \end{cases}$$

We refer to $\hat{e}_{i,j}$ as the 'score' of the edge $Z_i \rightarrow Z_j$.

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. Precision-recall curves have advantages over the traditional Receiver-Operator-Characteristic (ROC) curves (for ROC curves see, e.g., [32]). The advantages were for example shown in [11]. In this thesis we refer to the area under the precision-recall curve as AUC ('area under curve') value.

2.1.9 Hyperparameter settings and simulation details

For the models M1, M2 and M4 we re-use the hyperparameters from the earlier works by [38] and [24]: $\sigma^{-2} \sim GAM(\alpha_\sigma = \nu, \beta_\sigma = \nu)$ with $\nu = 0.005$, $\lambda_u^{-1} \sim GAM(\alpha_u = 2, \beta_u = 0.2)$, and $\lambda_c^{-1} \sim GAM(\alpha_c = 3, \beta_c = 3)$. For M3 we use the same setting with the extension: $\delta_h \sim BER(p)$ with $p \sim BETA(a = 1, b = 1)$. For the new models M3-M4 we also experimented with alternative hyperparameter settings, which we took from [24]. In agreement with the results reported in [24], we found that the models are rather robust w.r.t. the hyperparameter values. The results were only slightly affected by the hyperparameter values.

For all models we run each Reversible Jump Markov Chain Monte Carlo (RJCMCMC) 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. To check for convergence, we compared the samples of independent simulations, using the conventional diagnostics, based on potential scale

reduction factors [20] as well as scatter plots of the estimated edge scores. The latter type of diagnostic has become very common for Bayesian networks (see, e.g., the work by [17]). For most of the data sets, analysed here, the scatter plot diagnostics indicated almost perfect convergence already after $V = 10,000$ iterations; see Figure 2.9(a) for an example. For $V = 100,000$ iterations we consistently observed perfect convergence for all data sets.

2.2 Data

2.2.1 Synthetic network data

For model comparisons we generated various synthetic network data sets. We report here on two studies with realistic network topologies, shown in Figure 2.6. In both studies we assumed the data segmentation to be known. Hence, we kept the changepoints in τ fixed at their right locations and did not perform reversible jump Markov chain Monte Carlo moves on τ .

Study 1: For the RAF pathway [50] with $N = 11$ nodes and $M = 20$ edges, shown in Figure 2.6, we generated data with $H = 4$ segments having $m = 10$ data points each. For each node Z_i and its parent nodes in π_i we sampled the regression coefficients for $h = 1$ from standard Gaussian distributions and collected them in a vector \mathbf{w}_1^i which we normalised to Euclidean norm 1, $\mathbf{w}_1^i \leftarrow \mathbf{w}_1^i / |\mathbf{w}_1^i|$. For the segments $h = 2, 3, 4$ we use: $\mathbf{w}_h^i = \mathbf{w}_{h-1}^i$ ($\delta_h = 1$, coupled) or $\mathbf{w}_h^i = -\mathbf{w}_{h-1}^i$ ($\delta_h = 0$, uncoupled). The design matrices \mathbf{X}_h^i 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 segment-specific values of Z_i : $\mathbf{z}_h^i = \mathbf{X}_h^i \mathbf{w}_h^i$ we element-wise added Gaussian noise with standard deviation $\sigma = 0.05$. For all coupling scenarios $(\delta_2, \delta_3, \delta_4) \in \{0, 1\}^3$, we generated 25 data sets having different regression coefficients.

Study 2: This study is similar to the first one with three changes: (i) We used the yeast network [8] with $N = 5$ nodes and $M = 8$ edges, shown in Figure 2.6, (ii) again we generated data with $H = 4$ segments, but we varied the number of time points per segment $m \in \{2, 3, \dots, 12\}$. (iii) We focused on one scenario: For each node Z_i and its parent nodes in π_i we generated two vectors \mathbf{w}_\diamond^i and \mathbf{w}_\star^i with standard Gaussian distributed entries. We re-normalised the first vector to Euclidean norm 1, $\mathbf{w}_\diamond^i \leftarrow \mathbf{w}_\diamond^i / |\mathbf{w}_\diamond^i|$, and the 2nd vector to norm 0.5, $\mathbf{w}_\star^i \leftarrow 0.5 \cdot \mathbf{w}_\star^i / |\mathbf{w}_\star^i|$. We set $\mathbf{w}_1^i = \mathbf{w}_2^i = \mathbf{w}_\diamond^i$ so that the segments $h = 1$ and $h = 2$ are coupled, and $\mathbf{w}_3^i = \mathbf{w}_4^i = (\mathbf{w}_\diamond^i + \mathbf{w}_\star^i) / (|\mathbf{w}_\diamond^i + \mathbf{w}_\star^i|)$, so that the segments $h = 3$ and $h = 4$ are coupled, while the coupling between $h = 3$ and $h = 2$ is 'moderate'. For each m we generated 25 data matrices with different regression coefficients.

2.2.2 Yeast gene expression data

[8] synthetically designed a network in *S. cerevisiae* (yeast) with $N = 5$ genes, and measured gene expression data under galactose- and glucose-metabolism:

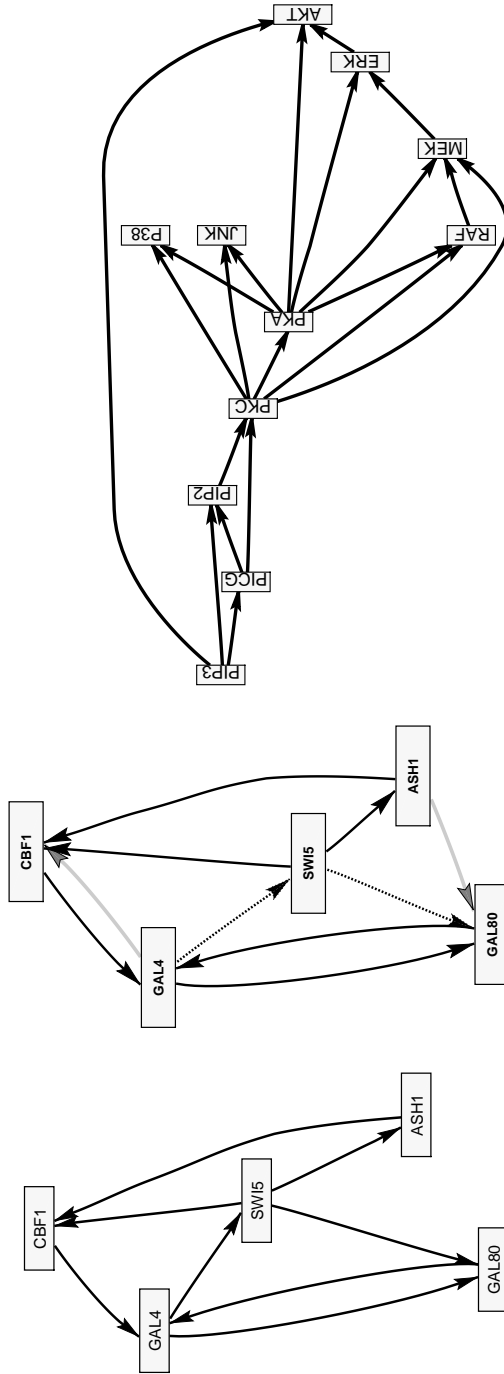


Figure 2.6: Network structures. Left: The true yeast network with $N = 5$ nodes and $M = 20$ edges. Middle: Network prediction obtained with model M3. The grey (dotted) edges correspond to false positives (negatives). Right: RAF pathway with $N = 11$ nodes and $M = 20$ edges.

16 measurements were taken in galactose and 21 measurements were taken in glucose, with 20 minutes intervals in between measurements. Although the network is small, it is an ideal benchmark data set: The network structure is known, so that network reconstruction methods can be cross-compared on real wet-lab data. We pre-process the data as described in [24]. The true network structure is shown in Figure 2.6 (left panel). As an example, a network prediction obtained with the partially coupled model (M3) is shown in the centre panel. For the prediction we extracted the 8 edges with the highest scores.

2.2.3 Arabidopsis gene expression data

The circadian clock in *Arabidopsis thaliana* optimizes the gene regulatory processes with respect to the daily dark:light cycles (photo periods). In four experiments *A. thaliana* plants were entrained in different dark:light cycles, before gene expression data were measured under constant light condition over 24- and 48-hour time intervals. We follow [24] 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.

2.3 Empirical results

We found that the proposed partially coupled model (M3), performed, overall, better than the other models. Thus, we decided to use M3 as reference model.

2.3.1 Results for synthetic network data

We start with the RAF-pathway for which we generated network data for 8 different coupling scenarios. Figure 2.7(a) compares the network reconstruction accuracies in terms of average AUC value differences. For 6 out of 8 scenarios the three AUC differences are clearly in favour of M3. Not surprisingly, for the two extreme scenarios, where all segments $h \geq 2$ are either coupled ('0111') or uncoupled ('0000'), M3 performs slightly worse than the fully coupled models (M2 and M4) or the uncoupled model (M1), respectively. But unlike the uncoupled model (M1) for coupled data ('0111'), and unlike the coupled models (M2 and M4) for uncoupled data ('0000'), the partially coupled model (M3) never performs significantly worse than the respective 'gold-standard' model. For the partially coupled model, Figure 2.7(b) shows the posterior probabilities that the segments $h = 2, 3, 4$ are coupled. The trends are in good agreement with the true coupling mechanism. Model M3 correctly infers whether the regression coefficients stay similar (identical) or change (substantially). The generalised coupled model (M4) can only adjust the segment-specific coupling strengths, but has no option to uncouple. Like the coupled model (M2), it fails when the parameters are subject to drastic changes. When comparing the coupled model (M2) with the generalised coupled model (M4), we see that M2 performs better

when only one segment is coupled, while the new M4 model is superior to M2 if two segments are coupled, see the scenarios ‘0011’, ‘0110’, and ‘0101’.

For the yeast network we generated data corresponding to a ‘0101’ coupling scheme and the change of the parameters (from the 2nd to the 3rd segment) is less drastic than for the RAF pathway data. Figure 2.8 shows how the AUC differences vary with the number of time points T , where $T = 4m$ and m is the number of data points per segment. For sufficiently many data points the effect of the prior diminishes and all models yield high AUC values (see bottom right panel). There are then no substantial differences between the AUC values anymore. However, for the lower sample sizes again the partially coupled model (M3) performs clearly best. For $12 \leq m \leq 28$ model M3 is clearly superior to all other models and for $30 \leq T \leq 40$ it still outperforms the uncoupled (M1) as well as the coupled (M2) model. The performance of the generalised model (M4) is comparable to the performance of the uncoupled model. For moderate sample sizes ($12 \leq T \leq 44$) model M4 is superior to the fully coupled model (M2).

2.3.2 Results for yeast gene expression data

For the yeast gene expression data we assume the changepoint(s) to be unknown and we infer the segmentation from the data. Figure 2.9(a) shows convergence diagnostics for the partially coupled model (M3). It can be seen from the scatter plots that $V = 10,000$ RJMCMC iterations yield already almost perfect convergence. The edge scores of 15 independent MCMC runs are almost identical to each other.

The average AUC scores of the models M1-M4 are shown in Figure 2.9(b). Since the number of inferred changepoints grows with the hyperparameter p of the geometric distribution on the distance between changepoints, we implemented the models with different p 's. The uncoupled model is superior to the coupled model for the lowest p ($p = 0.02$) only, but becomes more and more inferior to the coupled model, as p increases. This result is consistent with the finding in [24], where it was argued that parameter coupling gets more important when the number of segments H increases so that the individual segments become shorter. The new partially coupled model (M3) performs consistently better than the uncoupled and the coupled model (M1-M2). The only exemption occurs for $p = 0.1$ where the coupled model (M2) appears to perform slightly (but not statistically significantly) better than M3. For p 's up to $p = 0.05$ the fully coupled (M2) and the generalised fully coupled model (M4) perform approximately equally well. However, for the three highest p 's the new model M4 performs better than the coupled model (M2) and even outperforms the partially coupled model (M3). While the performances of the models M1-M3 decrease with the number of changepoints, the performance of model M4 stays robust. Subsequently, we re-analysed the yeast data with $K = 1, \dots, 5$ fixed changepoints. For each K we used the first changepoint to separate the two parts of the time series (galactose vs. glucose metabolism). Successively we located the next changepoint in the middle of the longest segment to divide it into 2 segments, until K changepoints

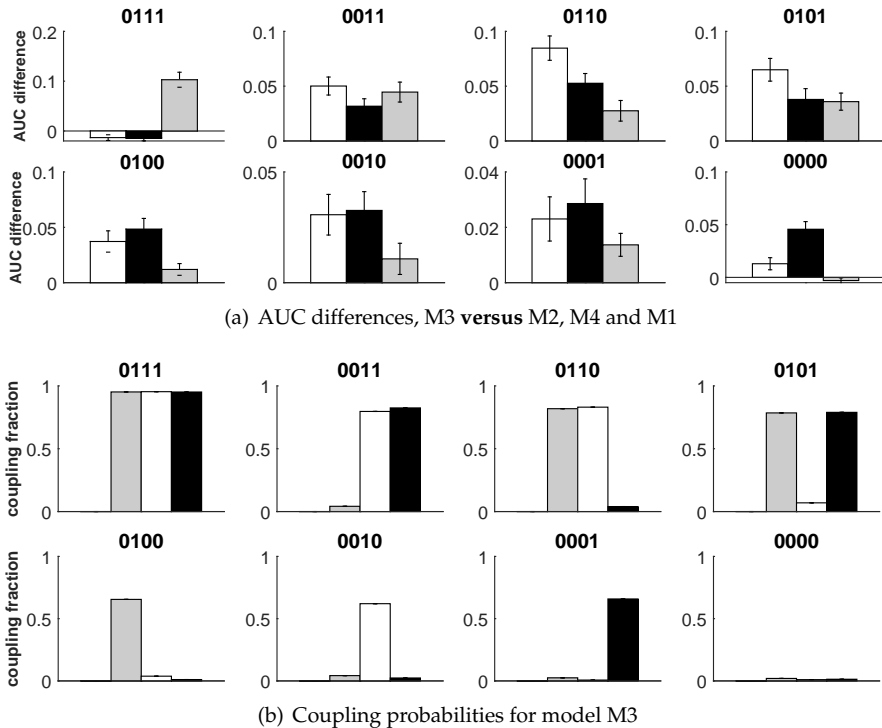


Figure 2.7: Results for synthetic RAF pathway data. We distinguish 8 coupling scenarios ($\delta_1 = 0, \delta_2, \delta_3, \delta_4$). **(a):** Each histogram has three bars for the average AUC differences between the partially coupled model (M3) and the other models: ‘M3 vs. M1 [=Uncoupled]’ (gray), ‘M3 vs. M4 [=Generalised]’ (black), and ‘M3 vs. M2 [=Coupled]’ (white). The error bars indicate 95% t-test confidence intervals. **(b):** Diagnostic for the partially coupled model (M3): The bars give the posterior probabilities $p(\delta_h = 1 | \mathcal{D})$ that segment h is coupled to $h - 1$ ($h = 2, 3, 4$).

were set. Figure 2.10(a-b) show the average AUC scores and the AUC score differences in favour of the partially coupled model (M3). Panel (a) reveals that the partially coupled model (M3) reaches again the highest network reconstruction accuracies. Panel (b) shows that the superiority of M3 is statistically significant, with only one exemption: For $K = 1$ the uncoupled model M1 performs as good as the partially coupled model (M3). Subsequently, we investigated the segment-specific coupling posterior probabilities $p(\delta_h = 1 | \mathcal{D})$ ($h = 2, \dots, H = K + 1$) for the partially coupled model (M3) and the posterior distributions of the coupling parameters $\lambda_u, \lambda_2, \dots, \lambda_{K+1}$ for the generalized model (M4), but we could not find clear trends for any gene. As an example, we provide the results for gene ASH1 in Figure 2.10(c-d). Panel (c) shows that the coupling posterior probabilities of model M3 do not have a clear pattern. However, it becomes obvious that the partially coupled model makes use of segment-wise switches between the

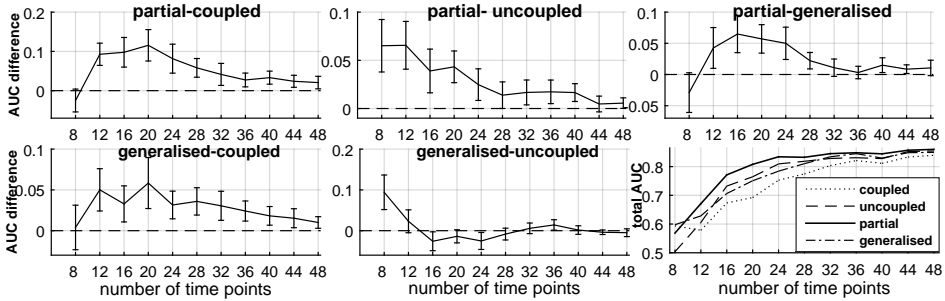
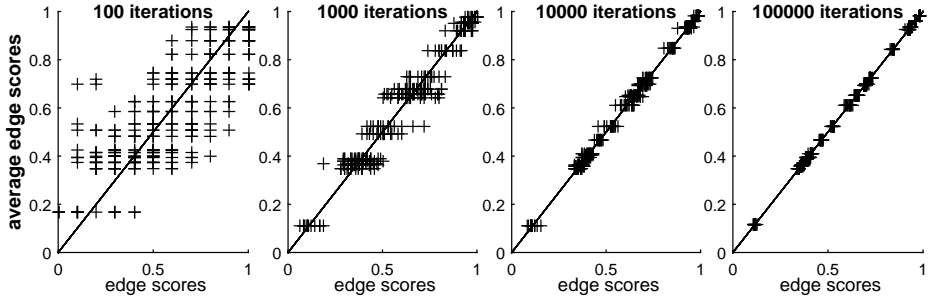
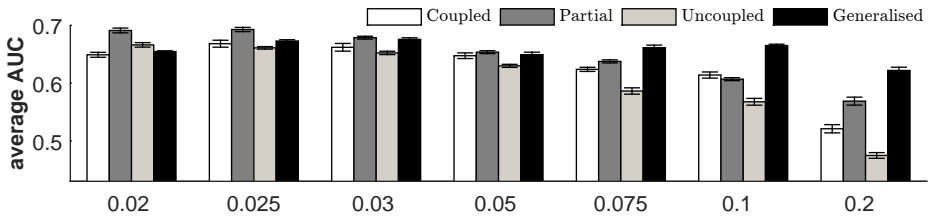


Figure 2.8: Results for synthetic yeast data. Five panels show the average AUC differences plotted against the numbers of data points T . The error bars indicate 95% t-test confidence intervals. The bottom right panel shows the model-specific average AUC values.



(a) Edge score scatter plots for the proposed partially coupled model (M3).



(b) Average AUC results for real yeast data with changepoint inference.

Figure 2.9: Analysis of the real yeast data. Panel (a): For each run length, $V \in \{100, 1000, 10000, 100000\}$ we performed 15 RJMCMC simulations with the partially coupled model (M3). We used the hyperparameter $p = 0.05$ for the changepoint prior. 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). Panel (b): We implemented the models M1-M4 with different hyperparameters p of the geometric distribution for the distance between changepoints. For each p the bars show the model-specific average AUC scores. The error bars indicate standard deviations.

uncoupled and the coupled approach. Panel (d) shows that the distributions of the coupling parameters, $\lambda_2, \dots, \lambda_{K+1}$, of model M4 do not substantially differ among segments. This might be the reason why the generalised coupled model (M4) is not superior to the fully coupled model (M2).

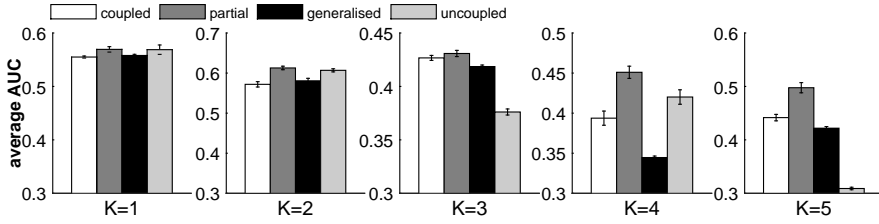
2.3.3 Application to Arabidopsis gene expression data

For the *Arabidopsis thaliana* gene expression data we cannot objectively compare the network reconstruction accuracies of the four models, since the true circadian clock network is not known. We therefore only applied the partially coupled model (M3), which we had found to be the best model in our earlier studies. Figure 2.11 shows the *Arabidopsis thaliana* network, which was reconstructed using the hyperparameter $p = 0.1$ for the geometric distribution on the distance between changepoints. To obtain a network prediction, we extracted the 20 edges with the highest edge scores. Although a proper evaluation of the network prediction is beyond the scope of this paper, we note that several features of the network are consistent with the plant biology literature. E.g. the feedback loop between *LHY* and *TOC1* is the most important key feature of the circadian clock network (see, e.g., the work by [41]). Many of the other predicted edges have been reported in more recent works. E.g. the edges $LHY \rightarrow ELF3$, $LHY \rightarrow ELF4$, $GI \rightarrow TOC1$, $ELF3 \rightarrow PRR3$ and $ELF4 \rightarrow PRR9$ can all be found in the circadian clock network (hypothesis) of [29].

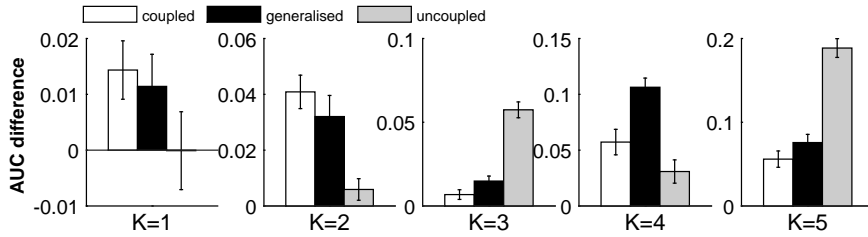
2.4 Discussion and conclusions

We have proposed two new Bayesian piece-wise linear regression models (M3 and M4) for reconstructing regulatory networks from gene expression time series. The partially coupled model (M3), shown in Figure 2.4, is a consensus model between the conventional uncoupled model (M1) and the fully sequentially coupled model (M2), proposed by [24]. In the uncoupled model (M1) the segment-specific regression coefficients have to be learned for each segment separately. In the fully coupled model (M2) each segment is compelled to be coupled to the previous one. The new partially coupled model (M3) combines features of the uncoupled and the fully coupled model. It infers for each segment whether it is coupled to (or uncoupled from) the preceding segment. The generalised coupled model (M4) is a generalisation of the coupled model (M2). Like the fully coupled model (M2), it does not have any option to uncouple, but it possesses segment-specific coupling parameters and allows for different coupling strengths between segments. Figure 2.5 shows a graphical model representation of the generalised model.

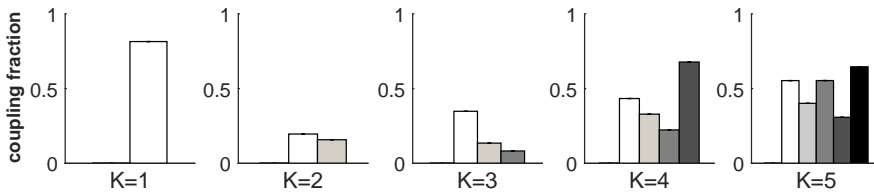
We have cross-compared the two newly proposed models (M3-M4) with the two established models (M1-M2); see Figures 2.2-2.3 for their graphical model representations. Our empirical results show that the generalised coupled model (M4) does not yield consistent improvements over the coupled model (M2). In



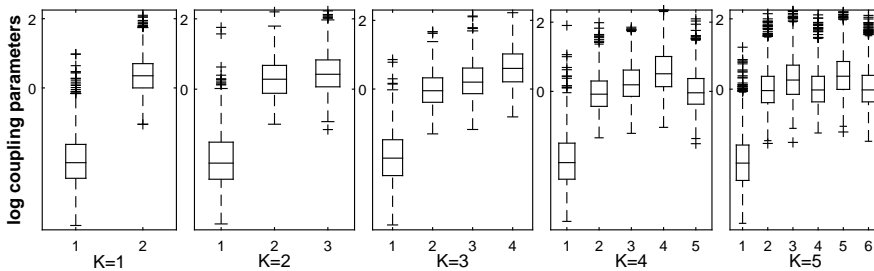
(a) AUCs of M2 (coupled), M3 (partially), M4 (generalised) and M1 (uncoupled).



(b) Partially (M3) vs. M2 (coupled), M4 (generalised) and M1 (uncoupled).



(c) M3: Average segment-specific coupling probabilities for target gene ASH1.



(d) M4: Segment-specific distributions of the coupling parameters for ASH1.

Figure 2.10: Results for real yeast data with fixed changepoints. We imposed $K \in \{1, \dots, 5\}$ fixed changepoints. See Subsection 2.3.2 for more detail. **Panel (a)** show the model-specific average total AUC scores with error bars indicating standard deviations. **Panel (b)** shows the AUC score differences with error bars indicating 95% two-sided t-test confidence intervals. **Panel (c)**: Diagnostic for the partially coupled model (M3): The bars give the posterior probabilities $p(\delta_h = 1 | \mathcal{D})$ that segment h is coupled to $h - 1$ ($h = 2, \dots, K + 1$) for gene ASH1. **Panel (d)**: Diagnostic for the generalized coupled model (M4): In each panel there is a boxplot for each segment showing the distributions of the logarithmic coupling parameters λ_h for gene ASH1.

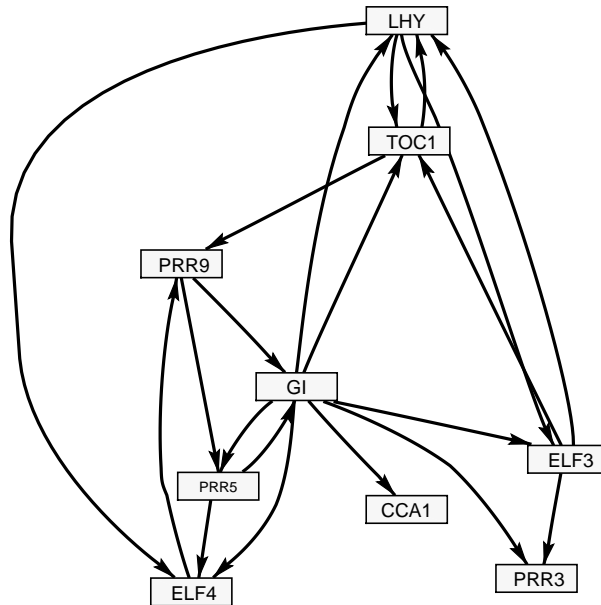


Figure 2.11: Prediction of the circadian clock network in *Arabidopsis thaliana*. The prediction was obtained with the proposed partially coupled model (M3), using the hyperparameter $p = 0.1$ for the geometric distribution on the distance between change-points. The network shows the 20 edges with the highest edge scores.

next chapter, we will try to figure out why the generalised fully coupled model did not perform better than the fully coupled model and we will try to refine this model.

The partially coupled model (M3), on the other hand, led consistently to improved network reconstruction accuracies. It has the desirable feature that it comprises the uncoupled (M1) and the fully coupled (M2) model as limiting cases. As the new partially segment-wise coupled model (M3) infers the best trade-off between M1 and M2 from the data, we would argue that this model should have precedence over the models M1 and M2 with regard to future applications.