# Bayesian regularization of non-homogeneous dynamic Bayesian networks by globally coupling interaction parameters

**Marco Grzegorczyk**
Department of Statistics, TU Dortmund, Germany
grzegorczyk@statistik.tu-dortmund.de

**Dirk Husmeier**
School of Mathematics and Statistics, Glasgow University, UK
dirk.husmeier@glasgow.ac.uk

**Abstract:** To relax the homogeneity assumption of dynamic Bayesian networks (DBNs), various recent studies have combined DBNs with multiple changepoint processes. The underlying assumption is that the parameters associated with time series segments delimited by multiple changepoints are **a priori** independent. Under weak regularity conditions, the parameters can be integrated out in the likelihood.. However, the assumption of prior independence is unrealistic in many real-world applications. We therefore propose a Bayesian coupling scheme to introduce information sharing among the segment-specific interaction parameters. We investigate the effect this model improvement has on the network reconstruction accuracy in a reverse engineering context, where the objective is to learn the structure of a gene regulatory network.

**MOTIVATION:** Non-homogeneous (NH) dynamic Bayesian networks (DBN) combine a node-specific multiple changepoint process with a conventional DBN.
Model parameters, and possibly network structures, are allowed to vary between time series segments.

**Previous work:** Robinson & Hartemink (JMLR 2010), Lebre et al. (BMC Systems Biology 2010), Grzegorczyk & Husmeier (Mach.Learn. 2011), based on Punskaya et al. (IEEE Trans. Sig. Proc. 2002).

**Shortcoming of these methods:** no information sharing among time series segments, leading to large posterior uncertainty (Bayesian) or overfitting (classical) when data are sparse.

**Objective:** Incorporate prior knowledge that changes are gradual, while maintaining the changepoint structure for analytical tractability.

**BACKGROUND:**
We formulate the NH-DBN in terms of a Bayesian piecewise linear regression model.

**Time-shifted target:** $\mathbf{y}_{g,h} = (y_{g,(\tau_{g,h}+1)}, \ldots, y_{g,\tau_{g,(h+1)}})^{\mathsf{T}}$

node (gene) · time segment · changepoints

**Design matrix:** $\mathbf{X}_{\pi_g,h} = (\mathbf{x}_{\pi_g,(\tau_{g,h}+1)}, \ldots, \mathbf{x}_{\pi_g,\tau_{g,(h+1)}})$

parent nodes · time segment · changepoints

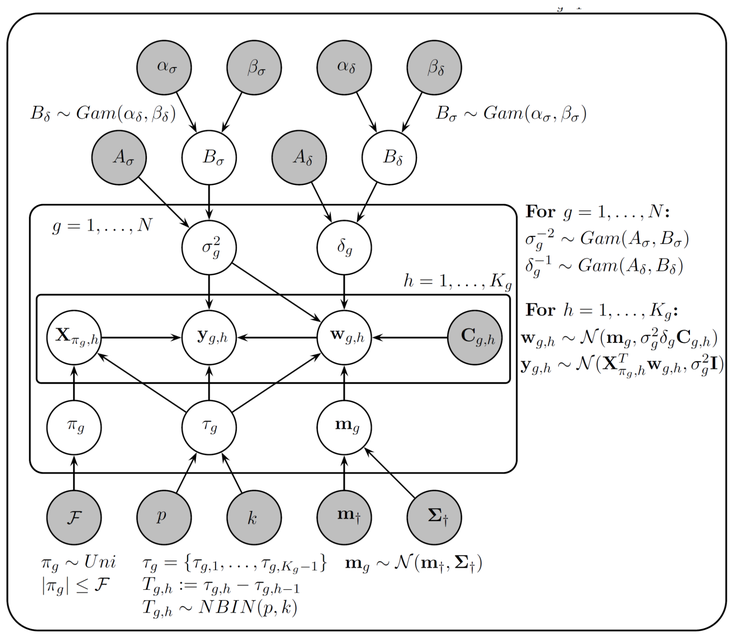
**Noise model:**
$$P(\mathbf{y}_{g,h}|\mathbf{X}_{\pi_g,h}, \mathbf{w}_{g,h}, \sigma^2_{g,h}) = \mathcal{N}(\mathbf{X}^{\mathsf{T}}_{\pi_g,h}\mathbf{w}_{g,h}, \sigma^2_{g,h}\mathbf{I})$$

**Interaction parameter prior:**
$$P(\mathbf{w}_{g,h}|\mathbf{m}_g, \sigma^2_{g,h}, \delta_g) = \mathcal{N}(\mathbf{w}_{g,h}|\mathbf{m}_g, \delta_g\sigma^2_{g,h}\mathbf{C}_{g,h})$$

**The computation of the marginal likelihood is analytically tractable:**
$$P(\mathbf{y}_{g,h}|\mathbf{X}_{\pi_g,h}, \delta_g, \mathbf{m}_g) = \int_0^\infty \left\{ \int P(\mathbf{y}_{g,h}, \mathbf{w}_{g,h}|\mathbf{X}_{\pi_g,h}, \sigma^2_{g,h}, \delta_g, \mathbf{m}_g) d\mathbf{w}_{g,h} \right\} d\sigma^2_{g,h}$$

**METHODOLOGY:**

**The proposed model can be illustrated with a probabilistic graphical model and plate diagram:**




**Various coupling schemes can be distinguished:**
Overview of the coupling schemes **(S1-S9)** for the noise variance hyperparameters.
**No coupling**: The noise variance hyperparameters are d-separated, i.e., they have separate level-2 hyperparameters which are fixed.
**Weak coupling**: The noise Variance hyperparameters are not d-separated, i.e., they share a set of common level-2 hyper-parameters which are flexible.
**Hard coupling**: There are common noise variance hyperparameters (with fixed level-2 hyperparameters).

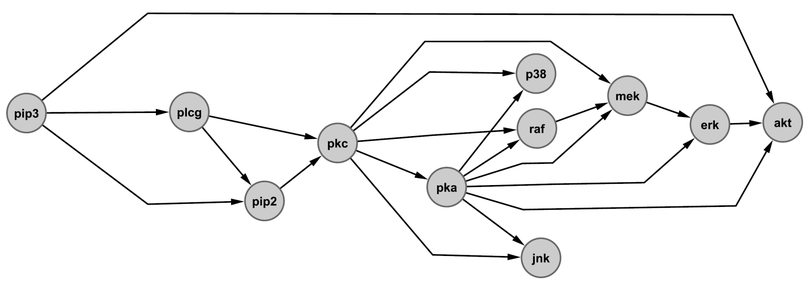
| segments $h = 1, \ldots, K_g$ | nodes ($g = 1, \ldots, N$): no coupling | weak coupling | hard coupling |
|---|---|---|---|
| no coupling | **(S1)** $\sigma^{-2}_{g,h} \sim Gam(A_{\sigma,g,h}, B_{\sigma,g,h})$ $A_{\sigma,g,h}$ and $B_{\sigma,g,h}$ fixed | **(S2)** $\sigma^{-2}_{g,h} \sim Gam(A_{\sigma,h}, B_{\sigma,h})$ $A_{\sigma,h}$ and $B_{\sigma,h}$ flexible i.e. $\{\sigma^2_{g,h}\}_g$ coupled $\forall h$ | **(S3)** $\sigma^2_{g,h} = \sigma^2_h$ $\sigma^{-2}_h \sim Gam(A_{\sigma,h}, B_{\sigma,h})$ $A_{\sigma,h}$ and $B_{\sigma,h}$ fixed |
| weak coupling | **(S4)** $\sigma^{-2}_{g,h} \sim Gam(A_{\sigma,g}, B_{\sigma,g})$ $A_{\sigma,g}$ and $B_{\sigma,g}$ flexible i.e. $\{\sigma^2_{g,h}\}_h$ coupled $\forall g$ | **(S5)** $\sigma^{-2}_{g,h} \sim Gam(A_\sigma, B_\sigma)$ $A_\sigma$ and/or $B_\sigma$ flexible i.e. $\{\sigma^2_{g,h}\}_{g,h}$ coupled | **(S6)** $\sigma^2_{g,h} = \sigma^2_h$ $\sigma^{-2}_h \sim Gam(A_\sigma, B_\sigma)$ $A_\sigma$ and/or $B_\sigma$ flexible i.e. $\{\sigma^2_h\}_h$ coupled |
| hard coupling | **(S7)** $\sigma^2_{g,h} = \sigma^2_g$ $\sigma^{-2}_g \sim Gam(A_{\sigma,g}, B_{\sigma,g})$ $A_{\sigma,g}$ and $B_{\sigma,g}$ fixed | **(S8)** $\sigma^2_{g,h} = \sigma^2_g$ $\sigma^{-2}_g \sim Gam(A_\sigma, B_\sigma)$ $A_\sigma$ and/or $B_\sigma$ flexible i.e. $\{\sigma^2_g\}_g$ coupled | **(S9)** $\sigma^2_{g,h} = \sigma^2$ $\sigma^{-2} \sim Gam(A_\sigma, B_\sigma)$ $A_\sigma$ and $B_\sigma$ fixed |

**SELECTED**

**SIMULATED DATA** were generated from an NH-DBN with segment-specific parameters of the following form: For each node g:
$$\mathbf{w}^\dagger_{g,\star} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad \mathbf{w}_{g,\star} = \frac{\mathbf{w}^\dagger_{g,\star}}{|\mathbf{w}^\dagger_{g,\star}|_2}$$
and then for each node-specific segment h:
$$\mathbf{w}^\dagger_{g,h} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad \tilde{\mathbf{w}}_{g,h} = \frac{\mathbf{w}^\dagger_{g,h}}{|\mathbf{w}^\dagger_{g,h}|_2} \quad \mathbf{w}_{g,h} = \frac{\mathbf{w}_{g,\star} + \varepsilon\tilde{\mathbf{w}}_{g,h}}{|\mathbf{w}_{g,\star} + \varepsilon\tilde{\mathbf{w}}_{g,h}|_2}$$

Time series of length m=41 with 4 equidistant time segments (h=1,2,3,4) were generated.
Observational white Gaussian noise at a target signal-to-noise ratio (**SNR**) was added.
As the underlying network

**Figure:** The RAF protein signalling pathway as reported in Sachs et al. (Science, 2005). The RAF network consists of 11 nodes (proteins) and 20 directed edges.

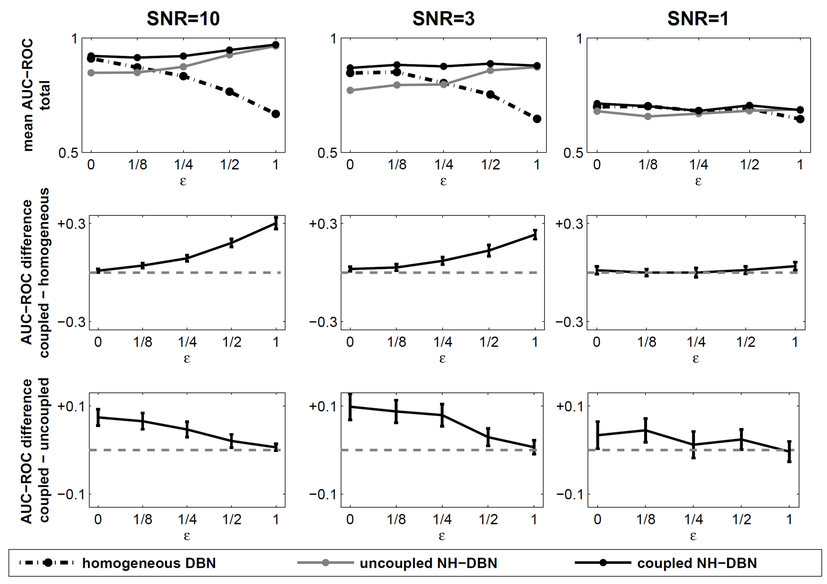



**RESULTS** The results show that for low perturbations (ε→0), the proposed method outperforms uncoupled NH-DBNs, while for large perturbations (ε→ 1) it outperforms conventional homogeneous DBNs

**Figure:** Network reconstruction accuracy (average AUC-ROC scores) for 3 signal-to-noise ratios (SNRs). The global parameter vector of amplitude 1 was perturbed with random perturbations of amplitude ε (abscissa).

The top row shows the total AUC-ROC scores.

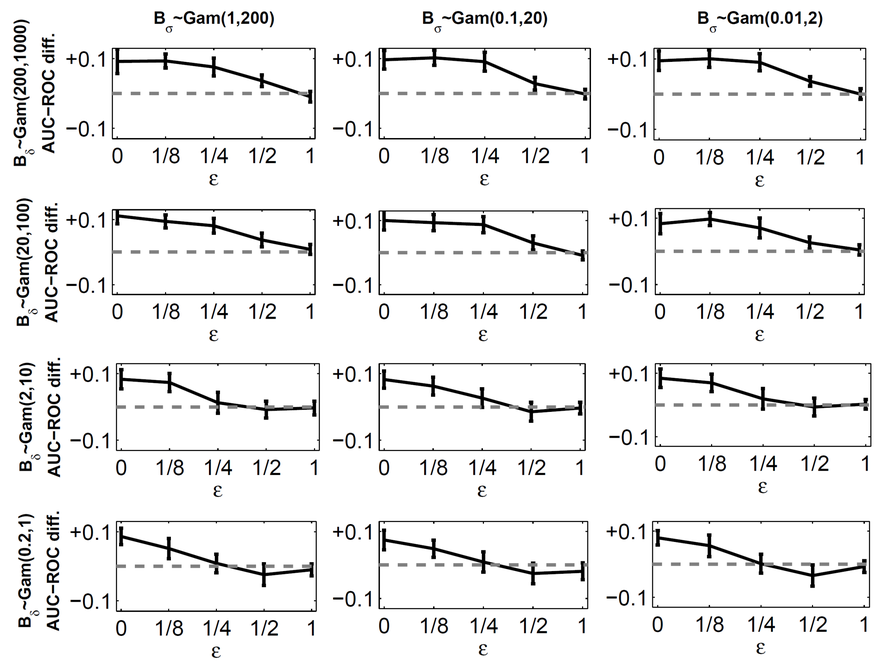
The centre row and the bottom row show the AUC-ROC differences.




**The results above were obtained with a specific choice of the highest-level hyperparameters:** $B_\sigma \sim Gam(1,200)$ *and* $B_\delta \sim Gam(200,1000)$.

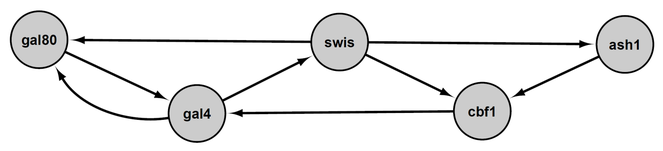
The proposed method is robust with respect to a variation of the level-3 hyperparameters as long as the prior on the scale of the coupling strengths $\delta^{-1}$ is not too vague.

**Figure:** AUC-ROC differences between the proposed coupled NH-DBN and the uncoupled NH-DBN **for various level-3 hyperparameter settings.**




**SYNTHETIC BIOLOGY IN YEAST:**
Cantone et al. (2009) synthetically designed a gene network in Saccharomyces cerevisiae (yeast). The authors measured expression levels of five genes in vivo with quantitative real-time polymerase chain reaction (PCR) at 37 time points over 8 hours. In about the middle of this time period, they changed the environment by switching the carbon source from galactose to glucose.

**Figure:** The S. cerevisiae network, designed in Cantone et al. (2009), consists of five nodes (genes).




For the Saccharomyces cerevisiae time series we assume the **segmentation to be unknown** so that the **changepoints have to be inferred** from the data. To this end, we impose a negative binomial distribution **NBIN(p,k)** with hyperparameters **p** and **k** onto the distance between neighbouring changepoints.

We vary the hyperparameter **p** to obtain different segmentations (i.e. numbers of changepoints per gene).

**As a consequence of the proposed coupling scheme, we obtain increased robustness with respect to a variation of the hyperparameters of the changepoint process.**

**Figure: Network reconstruction accuracy of the coupled and the uncoupled NH-DBN plotted against the mean number of changepoints per gene.**