Modelling dynamic networks
Regularization of non-homogeneous dynamic Bayesian network models by coupling interaction parameters

Marco Grzegorczyk
Johann Bernoulli Institute (JBI)
Rijksuniversiteit Groningen

Presentation at the Van Dantzig Seminar
VU University Amsterdam
9-Oct-2014
Very brief introduction:
Each gene is the code for the synthesis of a specific protein.
**Transcription**: gene $\rightarrow$ mRNA.
**Translation**: mRNA $\rightarrow$ protein.
Proteins are the „functional units“ of the cell.
Proteins are enzymes, transcription factors, etc.
protein 1 is a **transcription factor** for gene 2

protein 2 is an **enzym** and catalyses a metabolic reaction

protein 3 is a **transcription factor** for gene 2
Microarray Chips

Expressions (activities) of thousands of genes in an experimental cell can be measured with Microarray Chips.
(Gen-)Regulatory Network

Gene level

Protein level

Metabolite level

G1 → P1

G2 → P2

G3 → P3

metabolite A

metabolite B
Gen-Regulatory Network

**Goal**: Learn from gene expression data that gene 1 and gene 3 co-regulate gene 2

**Remark**: In gene regulatory networks the protein level is ignored. That is, proteins may build complexes with each other or may have to be activated (e.g. phosphorylated) before they can bind to binding sites of genes.
Protein activation

Cell membrane

P₁ → P₁ phosphorylated

P₃ → P₃ phosphorylated

→ nucleus

→ G₂

→ cell response
Protein activation

Cell membran

P₁ phosphorylated

P₃ phosphorylated

→ cell response

P₁

P₃
**Gen-Regulatory Network**

**Goal:** Learn from gene expression data that gene 1 and gene 3 co-regulate gene 2

**Remark:** In gene regulatory networks **the protein level is ignored**. That is, proteins may build complexes with each other or may have to be activated (e.g. phosphorylated) before they can bind to binding sites of genes.
Medical relevance
e.g. for tumour development
-- simplified example --

gene1 may be a tumour suppressor gene

gene 3 may be an oncogene

gene 2 may cause cell growth and cell division

Healthy condition

 cell division is under control
Medical relevance e.g. for tumour development -- simplified example --

- gene1 may be a tumour suppressor gene
- gene 3 may be an oncogene
- gene 2 may cause cell growth and cell division

Tumour cell - Altered pathway leads to uncontrolled cell division
possibly completely unknown
possibly completely unknown

E.g.: Gene-Microarray experiments

data
(expressions of genes)
possibly completely unknown

E.g.: Gene-Microarray experiments

Machine Learning statistical methods
Extract a network from an n-by-m data matrix

Either m independent (steady-state) observations of the system $X^{(1)},...,X^{(n)}$

Or time series of the system of length m: $(X^{(1)},...,X^{(n)})_{t=1,...,m}$
Dynamic Bayesian networks

Illustration: Simple dynamic Bayesian network (DBN) with three nodes. All interactions are subject to a time delay.
Static/dynamic Bayesian networks

**Static Bayesian networks**

Important feature: Network has to be acyclic

Implied factorisation:

\[ P(A, B) = P(B|B) \cdot P(A|A, B) \]

cycles cannot make sense

**Dynamic Bayesian networks**

Network does **not** have to be acyclic

Implied factorisation:

\[
P(A(t), B(t)|A(t-1), B(t-1)) = P(B(t)|B(t-1)) \cdot P(A(t)|A(t-1), B(t-1))
\]

\((t=2, \ldots, m)\)
Model assumption: **Homogeneous** Markov chain

Example: **4 genes, 10 time points**

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
<th>$t_5$</th>
<th>$t_6$</th>
<th>$t_7$</th>
<th>$t_8$</th>
<th>$t_9$</th>
<th>$t_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^{(1)}$</td>
<td>$X_{1,1}$</td>
<td>$X_{1,2}$</td>
<td>$X_{1,3}$</td>
<td>$X_{1,4}$</td>
<td>$X_{1,5}$</td>
<td>$X_{1,6}$</td>
<td>$X_{1,7}$</td>
<td>$X_{1,8}$</td>
<td>$X_{1,9}$</td>
<td>$X_{1,10}$</td>
</tr>
<tr>
<td>$X^{(2)}$</td>
<td>$X_{2,1}$</td>
<td>$X_{2,2}$</td>
<td>$X_{2,3}$</td>
<td>$X_{2,4}$</td>
<td>$X_{2,5}$</td>
<td>$X_{2,6}$</td>
<td>$X_{2,7}$</td>
<td>$X_{2,8}$</td>
<td>$X_{2,9}$</td>
<td>$X_{2,10}$</td>
</tr>
<tr>
<td>$X^{(3)}$</td>
<td>$X_{3,1}$</td>
<td>$X_{3,2}$</td>
<td>$X_{3,3}$</td>
<td>$X_{3,4}$</td>
<td>$X_{3,5}$</td>
<td>$X_{3,6}$</td>
<td>$X_{3,7}$</td>
<td>$X_{3,8}$</td>
<td>$X_{3,9}$</td>
<td>$X_{3,10}$</td>
</tr>
<tr>
<td>$X^{(4)}$</td>
<td>$X_{4,1}$</td>
<td>$X_{4,2}$</td>
<td>$X_{4,3}$</td>
<td>$X_{4,4}$</td>
<td>$X_{4,5}$</td>
<td>$X_{4,6}$</td>
<td>$X_{4,7}$</td>
<td>$X_{4,8}$</td>
<td>$X_{4,9}$</td>
<td>$X_{4,10}$</td>
</tr>
</tbody>
</table>
Impose changepoints to model non-homogeneous processes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>FIRST SEGMENT</th>
<th></th>
<th>SECOND SEGMENT</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$X^{(1)}$</td>
<td></td>
<td>$X_{1,1}$</td>
<td></td>
<td>$X_{1,2}$</td>
<td>$X_{1,3}$</td>
<td>$X_{1,4}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$X^{(2)}$</td>
<td></td>
<td>$X_{2,1}$</td>
<td></td>
<td>$X_{2,2}$</td>
<td>$X_{2,3}$</td>
<td>$X_{2,4}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$X^{(3)}$</td>
<td></td>
<td>$X_{3,1}$</td>
<td></td>
<td>$X_{3,2}$</td>
<td>$X_{3,3}$</td>
<td>$X_{3,4}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$X^{(4)}$</td>
<td></td>
<td>$X_{4,1}$</td>
<td></td>
<td>$X_{4,2}$</td>
<td>$X_{4,3}$</td>
<td>$X_{4,4}$</td>
</tr>
</tbody>
</table>
Changepoint model

**Our paradigm:** Keep the network topology fixed but the interaction parameters can change with time.

Interaction parameters in the **first** segment
Our paradigm: Keep the network topology fixed but the interaction parameters can change with time.

interaction parameters in the second segment
Introduce gene-specific changepoints to increase flexibility of the models

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
<th>$t_5$</th>
<th>$t_6$</th>
<th>$t_7$</th>
<th>$t_8$</th>
<th>$t_9$</th>
<th>$t_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^{(1)}$</td>
<td>$X_{1,1}$</td>
<td>$X_{1,2}$</td>
<td>$X_{1,3}$</td>
<td>$X_{1,4}$</td>
<td>$X_{1,5}$</td>
<td>$X_{1,6}$</td>
<td>$X_{1,7}$</td>
<td>$X_{1,8}$</td>
<td>$X_{1,9}$</td>
<td>$X_{1,10}$</td>
</tr>
<tr>
<td>$X^{(2)}$</td>
<td>$X_{2,1}$</td>
<td>$X_{2,2}$</td>
<td>$X_{2,3}$</td>
<td>$X_{2,4}$</td>
<td>$X_{2,5}$</td>
<td>$X_{2,6}$</td>
<td>$X_{2,7}$</td>
<td>$X_{2,8}$</td>
<td>$X_{2,9}$</td>
<td>$X_{2,10}$</td>
</tr>
<tr>
<td>$X^{(3)}$</td>
<td>$X_{3,1}$</td>
<td>$X_{3,2}$</td>
<td>$X_{3,3}$</td>
<td>$X_{3,4}$</td>
<td>$X_{3,5}$</td>
<td>$X_{3,6}$</td>
<td>$X_{3,7}$</td>
<td>$X_{3,8}$</td>
<td>$X_{3,9}$</td>
<td>$X_{3,10}$</td>
</tr>
<tr>
<td>$X^{(4)}$</td>
<td>$X_{4,1}$</td>
<td>$X_{4,2}$</td>
<td>$X_{4,3}$</td>
<td>$X_{4,4}$</td>
<td>$X_{4,5}$</td>
<td>$X_{4,6}$</td>
<td>$X_{4,7}$</td>
<td>$X_{4,8}$</td>
<td>$X_{4,9}$</td>
<td>$X_{4,10}$</td>
</tr>
</tbody>
</table>
Non-Homogeneous Dynamic Bayesian Networks (NH-DBN)

**Idea:** Combine a standard DBN with a node-specific multiple changepoint process.

Lèbre, Becq, Devaux, Lelandais, Stumpf (2010)
Statistical inference of the time-varying structure of gene regulation networks
*BMC Systems Biology*

Robinson & Hartemink (2010)
Learning non-stationary dynamic Bayesian networks
*Journal of Machine Learning Research*
What is the problem with these approaches?
**Practical problem:** inference uncertainty in short time series segments

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
<th>$t_5$</th>
<th>$t_6$</th>
<th>$t_7$</th>
<th>$t_8$</th>
<th>$t_9$</th>
<th>$t_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^{(1)}$</td>
<td>$X_{1,1}$</td>
<td>$X_{1,2}$</td>
<td>$X_{1,3}$</td>
<td>$X_{1,4}$</td>
<td>$X_{1,5}$</td>
<td>$X_{1,6}$</td>
<td>$X_{1,7}$</td>
<td>$X_{1,8}$</td>
<td>$X_{1,9}$</td>
<td>$X_{1,10}$</td>
</tr>
<tr>
<td>$X^{(2)}$</td>
<td>$X_{2,1}$</td>
<td>$X_{2,2}$</td>
<td>$X_{2,3}$</td>
<td>$X_{2,4}$</td>
<td>$X_{2,5}$</td>
<td>$X_{2,6}$</td>
<td>$X_{2,7}$</td>
<td>$X_{2,8}$</td>
<td>$X_{2,9}$</td>
<td>$X_{2,10}$</td>
</tr>
<tr>
<td>$X^{(3)}$</td>
<td>$X_{3,1}$</td>
<td>$X_{3,2}$</td>
<td>$X_{3,3}$</td>
<td>$X_{3,4}$</td>
<td>$X_{3,5}$</td>
<td>$X_{3,6}$</td>
<td>$X_{3,7}$</td>
<td>$X_{3,8}$</td>
<td>$X_{3,9}$</td>
<td>$X_{3,10}$</td>
</tr>
<tr>
<td>$X^{(4)}$</td>
<td>$X_{4,1}$</td>
<td>$X_{4,2}$</td>
<td>$X_{4,3}$</td>
<td>$X_{4,4}$</td>
<td>$X_{4,5}$</td>
<td>$X_{4,6}$</td>
<td>$X_{4,7}$</td>
<td>$X_{4,8}$</td>
<td>$X_{4,9}$</td>
<td>$X_{4,10}$</td>
</tr>
</tbody>
</table>
Shortcomings

1. Practical problem
   Short time series
   inference uncertainty

2. Methodological problem
   Prior independence is biologically implausible

Is it plausible to assume a priori that the segment-specific interaction parameters are independent?

**Idea**: Information coupling among segments
Non-homogeneous DBN (uncoupled NH-DBN)

Information coupling with respect to the interaction parameters (coupled NH-DBN)

Grzegorczyk and Husmeier (2012a)
A non-homogeneous dynamic Bayesian network model with sequentially coupled interaction parameters for applications in systems and synthetic biology.
*SAGMB*

Grzegorczyk and Husmeier (2012b)
Bayesian regularization of non-homogeneous dynamic Bayesian networks by globally coupling interaction parameters.
*AISTATS*

Grzegorczyk and Husmeier (2013)
Regularization of Non-Homogeneous Dynamic Bayesian Networks with Global Information-Coupling based on Hierarchical Bayesian models.
*Machine Learning*
Bayesian regression models

### Complete Segmentation Matrix

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
<th>$t_5$</th>
<th>$t_6$</th>
<th>$t_7$</th>
<th>$t_8$</th>
<th>$t_9$</th>
<th>$t_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^{(1)}$</td>
<td>$X_{1,1}$</td>
<td>$X_{1,2}$</td>
<td>$X_{1,3}$</td>
<td>$X_{1,4}$</td>
<td>$X_{1,5}$</td>
<td>$X_{1,6}$</td>
<td>$X_{1,7}$</td>
<td>$X_{1,8}$</td>
<td>$X_{1,9}$</td>
<td>$X_{1,10}$</td>
</tr>
<tr>
<td>$X^{(2)}$</td>
<td>$X_{2,1}$</td>
<td>$X_{2,2}$</td>
<td>$X_{2,3}$</td>
<td>$X_{2,4}$</td>
<td>$X_{2,5}$</td>
<td>$X_{2,6}$</td>
<td>$X_{2,7}$</td>
<td>$X_{2,8}$</td>
<td>$X_{2,9}$</td>
<td>$X_{2,10}$</td>
</tr>
<tr>
<td>$X^{(3)}$</td>
<td>$X_{3,1}$</td>
<td>$X_{3,2}$</td>
<td>$X_{3,3}$</td>
<td>$X_{3,4}$</td>
<td>$X_{3,5}$</td>
<td>$X_{3,6}$</td>
<td>$X_{3,7}$</td>
<td>$X_{3,8}$</td>
<td>$X_{3,9}$</td>
<td>$X_{3,10}$</td>
</tr>
<tr>
<td>$X^{(4)}$</td>
<td>$X_{4,1}$</td>
<td>$X_{4,2}$</td>
<td>$X_{4,3}$</td>
<td>$X_{4,4}$</td>
<td>$X_{4,5}$</td>
<td>$X_{4,6}$</td>
<td>$X_{4,7}$</td>
<td>$X_{4,8}$</td>
<td>$X_{4,9}$</td>
<td>$X_{4,10}$</td>
</tr>
</tbody>
</table>

### Complete Network

- $X^{(1)}$ connected to $X^{(2)}$
- $X^{(2)}$ connected to $X^{(3)}$
- $X^{(3)}$ connected to $X^{(4)}$

The network reveals a hierarchical structure with complete segmentation matrices for each level.
Bayesian regression models

first gene $g=1$

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
<th>$t_5$</th>
<th>$t_6$</th>
<th>$t_7$</th>
<th>$t_8$</th>
<th>$t_9$</th>
<th>$t_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^{(1)}$</td>
<td>$X_{1,1}$</td>
<td>$X_{1,2}$</td>
<td>$X_{1,3}$</td>
<td>$X_{1,4}$</td>
<td>$X_{1,5}$</td>
<td>$X_{1,6}$</td>
<td>$X_{1,7}$</td>
<td>$X_{1,8}$</td>
<td>$X_{1,9}$</td>
<td>$X_{1,10}$</td>
</tr>
<tr>
<td>$X^{(2)}$</td>
<td>$X_{2,1}$</td>
<td>$X_{2,2}$</td>
<td>$X_{2,3}$</td>
<td>$X_{2,4}$</td>
<td>$X_{2,5}$</td>
<td>$X_{2,6}$</td>
<td>$X_{2,7}$</td>
<td>$X_{2,8}$</td>
<td>$X_{2,9}$</td>
<td>$X_{2,10}$</td>
</tr>
<tr>
<td>$X^{(3)}$</td>
<td>$X_{3,1}$</td>
<td>$X_{3,2}$</td>
<td>$X_{3,3}$</td>
<td>$X_{3,4}$</td>
<td>$X_{3,5}$</td>
<td>$X_{3,6}$</td>
<td>$X_{3,7}$</td>
<td>$X_{3,8}$</td>
<td>$X_{3,9}$</td>
<td>$X_{3,10}$</td>
</tr>
<tr>
<td>$X^{(4)}$</td>
<td>$X_{4,1}$</td>
<td>$X_{4,2}$</td>
<td>$X_{4,3}$</td>
<td>$X_{4,4}$</td>
<td>$X_{4,5}$</td>
<td>$X_{4,6}$</td>
<td>$X_{4,7}$</td>
<td>$X_{4,8}$</td>
<td>$X_{4,9}$</td>
<td>$X_{4,10}$</td>
</tr>
</tbody>
</table>
Bayesian regression models

This changepoint divides the observations of node $X^{(1)}$ into $K_{g=1}=2$ disjunct segments.

<table>
<thead>
<tr>
<th>$X^{(1)}$</th>
<th>$X^{(2)}$</th>
<th>$X^{(3)}$</th>
<th>$X^{(4)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{1,1}$</td>
<td>$X_{2,1}$</td>
<td>$X_{3,1}$</td>
<td>$X_{4,1}$</td>
</tr>
<tr>
<td>$h=1$</td>
<td>$h=1$</td>
<td>$h=1$</td>
<td>$h=1$</td>
</tr>
<tr>
<td>$X_{2,2}$</td>
<td>$X_{2,3}$</td>
<td>$X_{2,4}$</td>
<td>$X_{2,5}$</td>
</tr>
<tr>
<td>$X_{2,6}$</td>
<td>$X_{2,7}$</td>
<td>$X_{2,8}$</td>
<td>$X_{2,9}$</td>
</tr>
<tr>
<td>$X_{2,10}$</td>
<td>$X_{3,2}$</td>
<td>$X_{3,3}$</td>
<td>$X_{3,4}$</td>
</tr>
<tr>
<td>$X_{3,5}$</td>
<td>$X_{3,6}$</td>
<td>$X_{3,7}$</td>
<td>$X_{3,8}$</td>
</tr>
<tr>
<td>$X_{3,9}$</td>
<td>$X_{3,10}$</td>
<td>$X_{4,2}$</td>
<td>$X_{4,3}$</td>
</tr>
<tr>
<td>$X_{4,4}$</td>
<td>$X_{4,5}$</td>
<td>$X_{4,6}$</td>
<td>$X_{4,7}$</td>
</tr>
<tr>
<td>$X_{4,8}$</td>
<td>$X_{4,9}$</td>
<td>$X_{4,10}$</td>
<td></td>
</tr>
</tbody>
</table>

$y_{g=1,h=1} = (X_{1,2}, \ldots, X_{1,6})$

$y_{g=1,h=2} = (X_{1,7}, \ldots, X_{1,10})$

changepoint $\tau_{g=1,1} = 6$
Bayesian regression models

For both segments $h=1$ and $h=2$ determine the observations which belong to the parent nodes of $X^{(1)}$.

Note that all interactions are subject to a **time lag of size 1**.
Bayesian regression models

**first gene**
g=1

![Diagram of gene relationships]

\[ y_{g=1,h=1} = (X_{1,2}, \ldots, X_{1,6})^T \quad y_{g=1,h=2} = (X_{1,7}, \ldots, X_{1,10})^T \]

<table>
<thead>
<tr>
<th>(X^{(1)})</th>
<th>(t_1)</th>
<th>(t_2)</th>
<th>(t_3)</th>
<th>(t_4)</th>
<th>(t_5)</th>
<th>(t_6)</th>
<th>(t_7)</th>
<th>(t_8)</th>
<th>(t_9)</th>
<th>(t_{10})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_{1,1})</td>
<td>(X_{1,2})</td>
<td>(X_{1,3})</td>
<td>(X_{1,4})</td>
<td>(X_{1,5})</td>
<td>(X_{1,6})</td>
<td>(X_{1,7})</td>
<td>(X_{1,8})</td>
<td>(X_{1,9})</td>
<td>(X_{1,10})</td>
<td></td>
</tr>
<tr>
<td>(X^{(2)})</td>
<td>(X_{2,1})</td>
<td>(X_{2,2})</td>
<td>(X_{2,3})</td>
<td>(X_{2,4})</td>
<td>(X_{2,5})</td>
<td>(X_{2,6})</td>
<td>(X_{2,7})</td>
<td>(X_{2,8})</td>
<td>(X_{2,9})</td>
<td>(X_{2,10})</td>
</tr>
<tr>
<td>(X^{(3)})</td>
<td>(X_{3,1})</td>
<td>(X_{3,2})</td>
<td>(X_{3,3})</td>
<td>(X_{3,4})</td>
<td>(X_{3,5})</td>
<td>(X_{3,6})</td>
<td>(X_{3,7})</td>
<td>(X_{3,8})</td>
<td>(X_{3,9})</td>
<td>(X_{3,10})</td>
</tr>
<tr>
<td>(X^{(4)})</td>
<td>(X_{4,1})</td>
<td>(X_{4,2})</td>
<td>(X_{4,3})</td>
<td>(X_{4,4})</td>
<td>(X_{4,5})</td>
<td>(X_{4,6})</td>
<td>(X_{4,7})</td>
<td>(X_{4,8})</td>
<td>(X_{4,9})</td>
<td>(X_{4,10})</td>
</tr>
</tbody>
</table>

For both segments h=1 and h=2 determine the observations which belong to the parent nodes of \(X^{(1)}\).
Note that all interactions are subject to a **time lag of size 1**.
Bayesian regression models

\[ y_{g=1,h=1} = (X_{1,2}, \ldots, X_{1,6})^T \quad y_{g=1,h=2} = (X_{1,7}, \ldots, X_{1,10})^T \]

\[ X_{\pi_1 = \{2,3\}, h=1} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_{2,1} & X_{2,2} & \cdots & X_{2,5} \\ X_{3,1} & X_{3,2} & \cdots & X_{3,5} \end{pmatrix} \]

\[ X_{\pi_1 = \{2,3\}, h=1} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_{2,6} & X_{2,7} & \cdots & X_{2,9} \\ X_{3,6} & X_{3,7} & \cdots & X_{3,9} \end{pmatrix} \]
For each gene $g=1,\ldots,G$ and each gene-specific segment $h=1,\ldots,K_g$:

**Likelihood model:**

$$y_{g,h} \sim \mathcal{N}(X_{\pi_{g,h}}^T w_{g,h}, \sigma^2_g I)$$

- target observations
- regressor matrix
- regression coefficients
- noise variance

**Prior on the regression coefficients $w_{g,h}$:**

$$w_{g,h} \sim \mathcal{N}(m_g, \sigma^2_g \delta_g C_{g,h})$$

- noise variance
- SNR hyperparameter

Note that the explicit dependence on the noise variance leads to a fully conjugate prior.
Graphical representation of the regression models

For $g = 1, \ldots, N$:

For $h = 1, \ldots, K_g$:

$w_{g,h} \sim \mathcal{N}(m_g, \sigma^2_g \delta_g C_{g,h})$

$y_{g,h} \sim \mathcal{N}(X^T_{\pi_{g,h}} w_{g,h}, \sigma^2_g I)$

$\tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\}$

$g = 1, \ldots, N$

$h = 1, \ldots, K_g$
Graphical representation of the regression models

For $g = 1, \ldots, N$:

For $h = 1, \ldots, K_g$:

$w_{g,h} \sim \mathcal{N}(m_g, \sigma_g^2 \delta_g C_{g,h})$

$y_{g,h} \sim \mathcal{N}(X_{\pi_g,h}^T w_{g,h}, \sigma_g^2 I)$

In the absence of any genuine prior knowledge: $m_g = 0$ and $C_{g,h} = I$

$\pi_g$

$\tau_g$

$m_g$

$\delta_g$

$\sigma_g^2$

$w_{g,h}$

$C_{g,h}$

$X_{\pi_g,h}$

$y_{g,h}$

$\tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\}$

Segmented data (observed)

Fixed hyperparameters

Changepoint set (segmentation)

$g = 1, \ldots, N$

Parent sets implied by the network

$\mathcal{N}$
In the absence of any genuine prior knowledge: $m_g = 0$ and $C_{g,h} = I$.

For $g = 1, \ldots, N$:

For $h = 1, \ldots, K_g$:

\[ w_{g,h} \sim \mathcal{N}(m_g, \sigma_g^2 \delta_g C_{g,h}) \]

\[ y_{g,h} \sim \mathcal{N}(X_{\pi_g,h}^T w_{g,h}, \sigma_g^2 I) \]
Are these hyperparameters actually known?

For $g = 1, \ldots, N$:

For $h = 1, \ldots, K_g$:

$w_{g,h} \sim \mathcal{N}(m_g, \sigma^2_g \delta_g C_{g,h})$

$y_{g,h} \sim \mathcal{N}(X_{\pi_{g,h}}^T w_{g,h}, \sigma^2_g I)$

$t_g = \{t_{g,1}, \ldots, t_{g,K_g-1}\}$
Graphical representation of the regression models

For $g = 1, \ldots, N$:
- $\sigma_g^{-2} \sim \text{Gam}(A_\sigma, B_\sigma)$
- $\delta_g^{-1} \sim \text{Gam}(A_\delta, B_\delta)$

For $h = 1, \ldots, K_g$:
- $w_{g,h} \sim \mathcal{N}(m_g, \sigma_g^2 \delta_g C_{g,h})$
- $y_{g,h} \sim \mathcal{N}(X_{\pi_g,h}^T w_{g,h}, \sigma_g^2 I)$

$\pi_g, \tau_g, m_g, C_{g,h}$

$g = 1, \ldots, N$,
$h = 1, \ldots, K_g$

$\tau_g = \{\bar{\tau}_{g,1}, \ldots, \bar{\tau}_{g,K_g-1}\}$

Parent sets (networks) must be inferred
Graphical model representation

For $g = 1, \ldots, N$:
\[ \sigma_g^{-2} \sim \text{Gam}(A_\sigma, B_\sigma) \]
\[ \delta_g^{-1} \sim \text{Gam}(A_\delta, B_\delta) \]

For $h = 1, \ldots, K_g$:
\[ \mathbf{w}_{g,h} \sim \mathcal{N}(\mathbf{m}_g, \sigma_g^2 \delta_g \mathbf{C}_{g,h}) \]
\[ \mathbf{y}_{g,h} \sim \mathcal{N}(\mathbf{X}_{\pi_g,h}^T \mathbf{w}_{g,h}, \sigma_g^2 \mathbf{I}) \]

\[ \pi_g \sim \text{Uni} \]
\[ |\pi_g| \leq \mathcal{F} \]
\[ \tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\} \]
With a fixed hyperparameter \( m_g \), there is no information coupling between the segment-specific regression coefficients.

For \( g = 1, \ldots, N \):
\[
\sigma_g^{-2} \sim \text{Gam}(A_{\sigma}, B_{\sigma})
\]
\[
\delta_g^{-1} \sim \text{Gam}(A_\delta, B_\delta)
\]

For \( h = 1, \ldots, K_g \):
\[
w_{g,h} \sim \mathcal{N}(m_g, \sigma_g^2 \delta_g C_{g,h})
\]
\[
y_{g,h} \sim \mathcal{N}(X_{\pi_g,h}^T w_{g,h}, \sigma_g^2 I)
\]
With a fixed hyperparameter $m_g$, there is no information coupling between the segment-specific regression coefficients.
Graphical model representation

For \( g = 1, \ldots, N \): 
\[ \sigma_g^{-2} \sim \text{Gam}(A_\sigma, B_\sigma) \]
\[ \delta_g^{-1} \sim \text{Gam}(A_\delta, B_\delta) \]

For \( h = 1, \ldots, K_g \): 
\[ w_{g,h} \sim \mathcal{N}(m_g, \sigma_g^2 \delta_g C_{g,h}) \]
\[ y_{g,h} \sim \mathcal{N}(X_{\pi_g,h}^T w_{g,h}, \sigma_g^2 I) \]

Main idea from:
Grzegorczyk and Husmeier (2012b)
Bayesian regularization of non-homogeneous dynamic Bayesian networks by globally coupling interaction parameters.

\( \pi_g \sim \text{Uni} \)
\[ |\pi_g| \leq \mathcal{F} \]
\[ \tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\} \]
\[ m_g \sim \mathcal{N}(m_\dagger, \Sigma_\dagger) \]
\[ T_{g,h} = \tau_{g,h} - \tau_{g,h-1} \]
\[ T_{g,h} \sim \text{NBIN}(p, k) \]
Graphical model representation

Main idea from: Grzegorczyk and Husmeier (2012b)
Bayesian regularization of non-homogeneous dynamic Bayesian networks by **globally** coupling interaction parameters.

\[ m_g \text{ variable} \]

so that the segment-specific regression coefficients are coupled

→ information exchange among segments
RJMCMC inference Part 1 of 3

1. Noise variances:
\[
\sigma_g^{-2} | (y_g, \ldots, X_{\pi_g}, \ldots, \delta_g)
\]

2. Regression coefficients:
\[
P(w_{g,h} | y_{g,h}, X_{\pi_g,h}, \sigma_g)
\]

3. Coupling hyperparameters:
\[
P(\delta_g^{-1} | y_g, \ldots, w_{g,\ldots}, \sigma_g^2, \ldots, X_{\pi_g,\ldots})
\]

That is, sample each variable from the conditional distribution, conditional on its Markov blanket.

Conjugate prior distributions: sampling from standard distributions

Collapsing: integrate some variables in the Markov blanket out analytically

can be sampled with standard collapsed and uncollapsed Gibbs sampling steps
4. **Network inference** by a Metropolis Hastings sampling scheme, which changes the network by **adding** and **removing** individual edges:

\[ P(\mathcal{M}|\mathcal{D}, \{\tau_g\}, \delta) \propto P(\mathcal{M}) \prod_g \prod_h P(y_{gh}|X_{\pi_g, h}, \delta_g) \]

- **Network prior**
- **Marginal likelihoods** can be computed in closed form:

5. **Changepoint inference** by a Metropolis Hastings sampling scheme, which changes the segmentation by **adding** and **removing** gene-specific **changepoints**:

\[ P(\{\tau_g\}|\mathcal{D}, \delta, \mathcal{M}) \propto \prod_g P(\tau_g) \prod_h P(y_{gh}|X_{\pi_g, h}, \delta_g) \]

- **Changepoint prior**
- **Marginal likelihoods** can be computed in closed form:
6. The global mean vector $\mathbf{m}_g$ can be sampled with a collapsed Gibbs sampling steps:

$$
\mathbf{m}_g | (\mathbf{w}_{g,1}, \ldots, \mathbf{w}_{g,K_g}) \sim \mathcal{N}(\mathbf{m}_{*,g}, \Sigma_{*,g})
$$

with the sufficient statistics:

$$
\Sigma_{*,g} := \left( \Sigma_\dagger^{-1} + K_g \Sigma_0^{-1} \right)^{-1}
$$

$$
\mathbf{m}_{*,g} := \Sigma_{*,g} (\Sigma_\dagger^{-1} \mathbf{m}_\dagger + \Sigma_0^{-1} \left[ \sum_{h=1}^{K_g} \mathbf{w}_{g,h} \right])
$$

**Overall sampling scheme:**

“Metropolis-Hastings-RJMCMMC scheme within a partially collapsed Gibbs sampler”
Empirical comparison: (1) globally coupled NH-DBN

For $g = 1, \ldots, N$:  
$\sigma_g^{-2} \sim \text{Gam}(A_\sigma, B_\sigma)$
$\delta_g^{-1} \sim \text{Gam}(A_\delta, B_\delta)$

For $h = 1, \ldots, K_g$:
$w_{g,h} \sim \mathcal{N}(m_g, \sigma_g^2 \delta_g C_{g,h})$
$y_{g,h} \sim \mathcal{N}(X_{\pi_g,h}^T w_{g,h}, \sigma_g^2 I)$

$\pi_g \sim \text{Unif}$
$|\pi_g| \leq F$
$\pi_g = \{\pi_{g,1}, \ldots, \pi_{g,K_g-1}\}$
$m_g \sim \mathcal{N}(m_\dagger, \Sigma_\dagger)$
$T_{g,h} := \pi_{g,h} - \pi_{g,h-1}$
$T_{g,h} \sim \text{NBIN}(p, k)$
Empirical comparison: (2) uncoupled NH-DBN

For $g = 1, \ldots, N$:
\[ \sigma_g^{-2} \sim \text{Gam}(A_{\sigma}, B_{\sigma}) \]
\[ \delta_g^{-1} \sim \text{Gam}(A_{\delta}, B_{\delta}) \]

For $h = 1, \ldots, K_g$:
\[ w_{g,h} \sim \mathcal{N}(m_g, \sigma_g^2 \delta_g C_{g,h}) \]
\[ y_{g,h} \sim \mathcal{N}(X_{\pi_g,h}^T w_{g,h}, \sigma_g^2 I) \]

We set: $m_g = 0$. 

\[ \pi_g \sim \text{Uni} \]
\[ \left| \pi_g \right| \leq \mathcal{F} \]
\[ \tau_g = \{\tau_{g,1}, \ldots, \tau_{g,K_g-1}\} \]
\[ T_{g,h} := \tau_{g,h} - \tau_{g,h-1} \]
\[ T_{g,h} \sim \text{NBIN}(p, k) \]
Empirical comparison: (3) Homogeneous DBN

Standard homogeneous dynamic Bayesian network (DBN). There are no changepoints; i.e. there is only one segment for each gene ($K_g = 1$).
Empirical comparison: (4) Sequentially coupled NH-DBN

\[
P(w_{g,h} \mid m_{g,h-1}, \sigma^2_{g,h}, \delta_g, \lambda_g) = \begin{cases} 
\mathcal{N}(w_{g,1} \mid m_{g,0} = 0, \delta_g \sigma^2_{g,h} C_{g,h}), & h = 1 \\
\mathcal{N}(w_{g,h} \mid m_{g,h-1}, \lambda_g \sigma^2_{g,h} C_{g,h}), & h \geq 2
\end{cases}
\]  
(1)

where \( m_{g,h-1} (h \geq 2) \) depends on the preceding segment:

\[
m_{g,h} = \Sigma_{g,h} ([\lambda_g C_{g,h}]^{-1} m_{g,(h-1)} + X_{\pi_{g,h}, y_{g,h}})
\]  
(2)

For \( h \geq 2 \):

The prior expectation of the regression coefficients for segment \( h+1 \), \( m_{g,h} \), depends on the posterior distribution of the regression coefficients \( w_{g,h} \) for segment \( h \).

The coupling strength depends on the hyperparameter \( \lambda_g \).

Main idea from: Grzegorczyk and Husmeier (2012a)

A non-homogeneous dynamic Bayesian network model with sequentially coupled interaction parameters for applications in systems and synthetic biology.

SAGMB
Information coupling

Sequential coupling
- Information is shared between neighbouring segments
- For example: morphogenesis

Global coupling
- Segments are treated as interchangeable and information is shared globally
- For example: different experimental scenarios or environmental conditions
Empirical evaluation

1. Simulated data

2. Data from synthetic biology

3. Data from a real application
Empirical evaluation

1. Simulated data
   
   Known gold standard 🌟
   
   Simulation process does not reflect real biology 😞

2. Data from synthetic biology
   
   Known gold standard 🌟
   
   Real wet lab data 🌟
   
   Regulatory network small 😞

3. Data from a real application
   
   Real wet lab data 🌟
   
   No gold standard 😞
Reconstruction Accuracy

true network

Evaluation of learning performance

extracted network

biological knowledge
(gold standard network)
Example: 2 genes $\rightarrow$ 16 different (dynamic) network structures

Best network: maximum score $P(D|M)$
Ideal scenario: Large data sets, low noise

Identify the best network structure

\[ P(\text{graph}|\text{data}) \]
Realistic: Limited number of experimental replications, high noise

Uncertainty about the best network

$P(\text{graph}|\text{data})$
Sample of high-scoring networks

$P(\text{graph}|\text{data})$
Idea: Model Averaging

Compute marginal posterior probabilities of the edges
Probabilistic inference

data

true regulatory network

marginal edge posterior probabilities

Thresholding

MCMC

concrete network predictions

TP:1/2
FP:0/4

TP:2/2
FP:1/4
From Perry Sprawls

ROC Curve

Area Under the ROC Curve
1. Simulated data

**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.
\[ y_{g,h} \sim \mathcal{N}(X_{\pi_{g,h}}^T \mathbf{w}_{g,h}, \sigma_g^2 \mathbf{I}) \]

**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.
$y_{g,h} \sim \mathcal{N}(X_{\pi_g,h}^T w_{g,h}, \sigma_g^2 I)$

**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.

$w_{g,*} \sim \mathcal{N}(0, 1), \quad \tilde{w}_{g,h} \sim \mathcal{N}(0, 1)$,

$$w_{g,h} = \frac{w_{g,*}}{|w_{g,*}|^2} + \varepsilon \frac{\tilde{w}_{g,h}}{|\tilde{w}_{g,h}|^2}$$
Generate data sets with 4 segments $h=1,...,4$ and 10 observations per segment.

Use three noise levels (SNR=10, 3, and 1)

Use the parameter $\varepsilon$ to vary the similarity of the segment-specific interaction parameters.

$\varepsilon=0$ -> homogeneous data

... $\varepsilon=1$ -> non-homogeneous data

**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.
AUC for SNR=3

- homogeneous DBN
- uncoupled NH-DBN
- coupled NH-DBN

Graph showing AUC values for different values of $\varepsilon$.
AUC difference: coupled NH-DBN – homogeneous DBN

Less homogeneous
AUC difference: coupled NH-DBN – uncoupled NH-DBN

More homogeneous
2. Data from synthetic biology

Synthetic network in yeast, as designed in Cantone et al. (2009)

Carbon-source switch from galactose to glucose

in vivo gene expression levels measured with RT-PCR at 37 time points (in two mediums)
AUC score comparison
sequentially coupled NH-DBN **versus**
uncoupled NH-DBN
for different changepoint prior hyperparameters
(different numbers of changepoints per gene)
AUC score comparison
globally coupled NH-DBN versus uncoupled NH-DBN
for different changepoint prior hyperparameters
(different numbers of changepoints per gene)
AUC score comparison of all three NH-DBNs
3. Data from a real application

Circadian regulation in Arabidopsis

Diagram showing the regulation mechanisms involved in circadian rhythms, with key components such as PHYs, CRYs, LHY, CCA1, ZTL, and TOC1.
Circadian rhythms in *Arabidopsis thaliana*

Collaboration with the Institute of Molecular Plant Sciences at Edinburgh University

4 time series of microarray gene expression data from *Arabidopsis thaliana*.

- **Focus on**: 9 circadian genes: LHY, CCA1, TOC1, ELF4, ELF3, GI, PRR9, PRR5, and PRR3

- The four time series were measured under constant light condition at **13 time points**: 0h, 2h,..., 24h, 26h

- Seedlings entrained with light:dark cycles of different periods
Thin black edges indicate interactions that are inferred with both NH-DBNs. Three edges (dotted) are inferred with the uncoupled NH-DBN only while four edges (bold) are inferred with the coupled NH-DBN only.
Thin black edges indicate interactions that are inferred with both NH-DBNs. Three edges (dotted) are inferred with the uncoupled NH-DBN only while four edges (bold) are inferred with the coupled NH-DBN only.
Thank you for your attention!

Any questions?