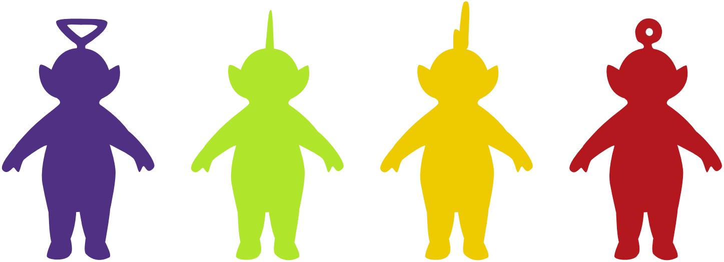
Exact Bayesian inference for off-line change-point detection in tree-structured graphical models

Loïc Schwaller
Statistics for Structures Seminar

AgroParisTech
INRA SCIENCE & IMPACT
Universiteit Leiden

March 17, 2017

Let's play a game...



Let's play a game...

Rules of the game

1. Choose an order.
2. First player

   -  → `stand`
   -  → `sit`

   Other players

   -  → `as previous player`
   -  →
     -  → `stand`
     -  → `sit`

Let's play a game...

Rules of the game

1. Choose an order.
2. First player

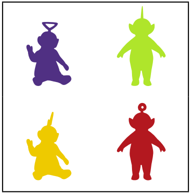
    → `stand`

    → `sit`

   Other players

    → `as previous player`

    →
   -  → `stand`
   -  → `sit`



Let's play a game...

Rules of the game

1. Choose an order.
2. First player

 → stand

 → sit

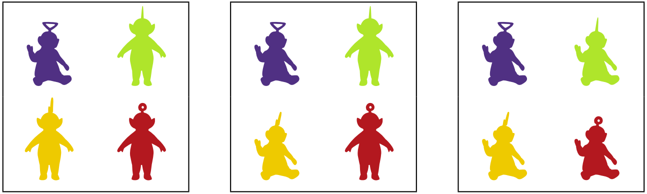
Other players

 → as previous player

 →

 → stand

 → sit



Let's play a game...

Rules of the game

1. Choose an order.
2. First player

    → `stand`

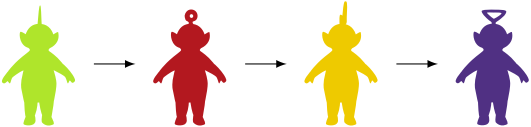
    → `sit`

   Other players

    → `as previous player`

    →

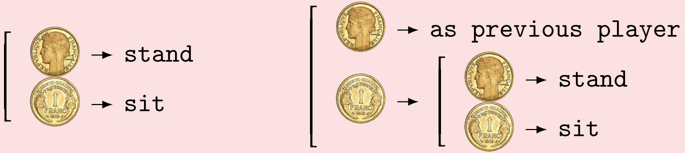
   -  → `stand`
   -  → `sit`

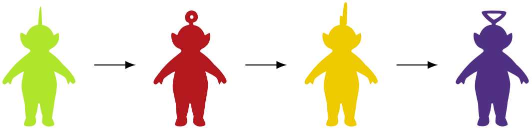


Let's play a game...

Rules of the game

1. Choose an order.
2. First player — Other players





Network Inference

Let's play a game...

New rules

3. Repeat 2 for some time.
4. Change order.
5. Repeat 2 for some time.

Let's play a game...

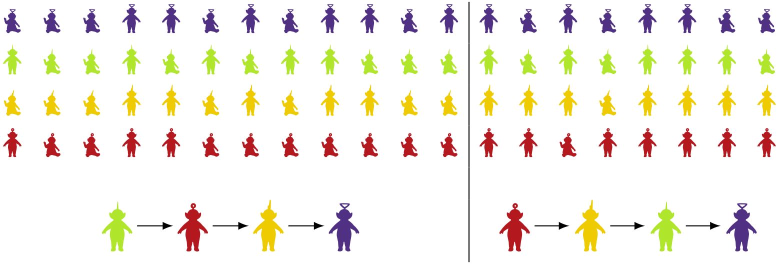
New rules

3. Repeat 2 for some time.
4. Change order.
5. Repeat 2 for some time.

Let's play a game...

New rules

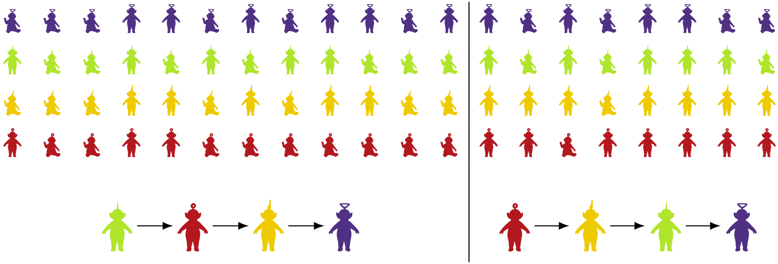
3. Repeat 2 for some time.
4. Change order.
5. Repeat 2 for some time.



Let's play a game...

New rules

3. Repeat 2 for some time.
4. Change order.
5. Repeat 2 for some time.

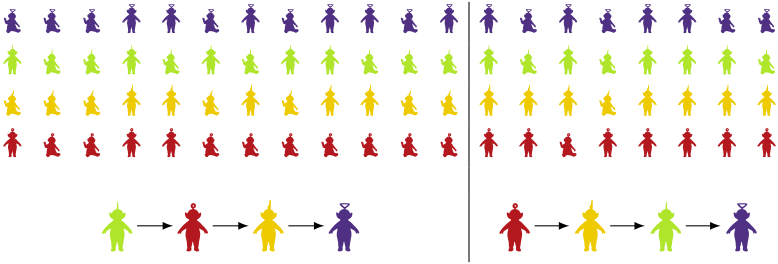


Segmentation

Let's play a game...

New rules

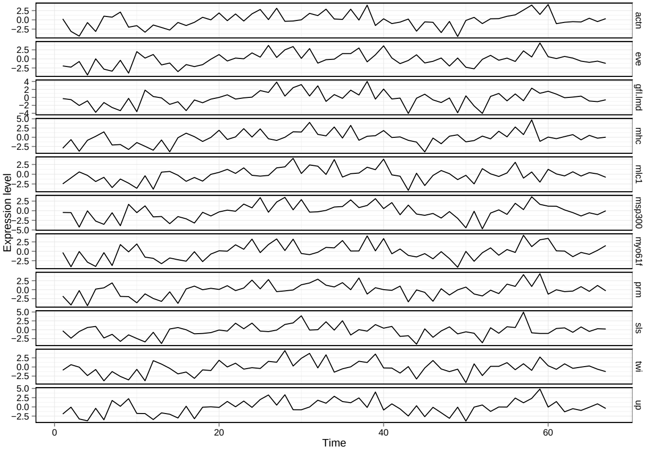
3. Repeat 2 for some time.
4. Change order.
5. Repeat 2 for some time.



Change-point detection

Drosophila life cycle microarray data[1]

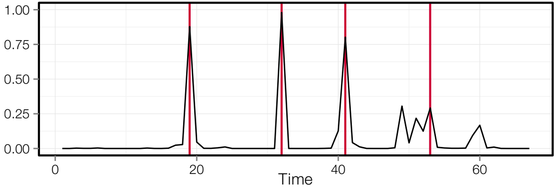
- Expression levels of 11 genes involved in wing muscle development
- 67 time-points

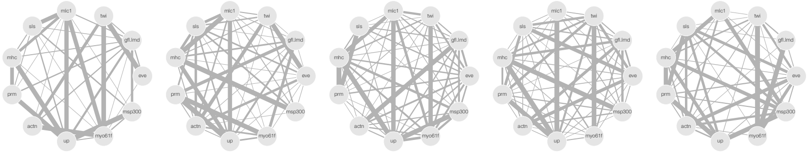


[1] Arbeitman et al. 2002.

Drosophila life cycle microarray data[1]

- Expression levels of 11 genes involved in wing muscle development
- 67 time-points





[1]Arbeitman et al. 2002.

Exact Bayesian inference

- **Bayesian** inference

- **Exact** inference

Exact Bayesian inference

- **Bayesian** inference

Requirement

- ▹ Providing a full probabilistic construction
  - Prior distributions

- **Exact** inference

Exact Bayesian inference

- **Bayesian** inference

Requirement	Tool
▷ Providing a full probabilistic construction ‣ Prior distributions	▷ Graphical models ‣ Markov property ‣ hyper Markov property

- **Exact** inference

Exact Bayesian inference

- **Bayesian** inference

Requirement

- Providing a full probabilistic construction
  - Prior distributions

Tool

- Graphical models
  - Markov property
  - hyper Markov property

- **Exact** inference

Requirement

- Dealing with the combinatorial issue
  - $\binom{N-1}{K-1}$ segmentations
  - $2^{p(p-1)/2}$ undirected graphs

Exact Bayesian inference

- **Bayesian** inference

Requirement	Tool

- ▷ Providing a full probabilistic construction
  - Prior distributions
- ▷ Graphical models
  - Markov property
  - hyper Markov property

- **Exact** inference

Requirement	Tool

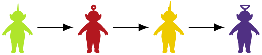
- ▷ Dealing with the combinatorial issue
  - $\binom{N-1}{K-1}$ segmentations
  - $2^{p(p-1)/2}$ undirected graphs
- ▷ Algebraic results
  - Segmentations
  - Spanning trees (Meilă and Jaakkola 2007)

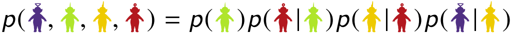
Outline

- **Introduction**
  - Graphical models
  - Algebraic results
- **Model & Inference**
- **Simulations**
- **Applications**
  - *Drosophila* life cycle microarray data
  - fMRI data
- **Extensions & Conclusion**

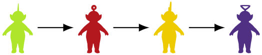
Introduction

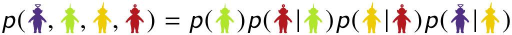
Graphical models

-  is a Markov chain.



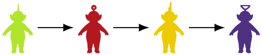
Graphical models

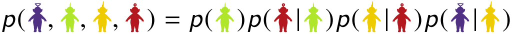
-  is a Markov chain.



- Graphical models extend this reasoning to **arbitrary dependence structures**.

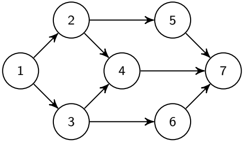
Graphical models

-  is a Markov chain.

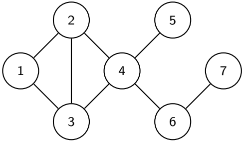


- Graphical models extend this reasoning to **arbitrary dependence structures**.

Directed Acyclic Graphs

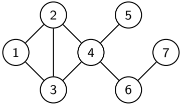


Undirected Graphs



Undirected graphs

- $V = \{1, \ldots, p\}$
- $\mathcal{P}_2(V)$ = subsets of V of size 2

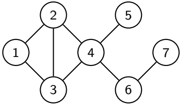


Definition

For $E \subseteq \mathcal{P}_2(V)$, $G = (V, E_G)$ is the undirected graph with **vertices** V and **edges** E_G.

Undirected graphs

- $V = \{1, \ldots, p\}$
- $\mathcal{P}_2(V)$ = subsets of V of size 2



Definition

For $E \subseteq \mathcal{P}_2(V)$, $G = (V, E_G)$ is the undirected graph with **vertices** V and **edges** E_G.

Definition

A **clique** is a fully connected subsets of vertices.

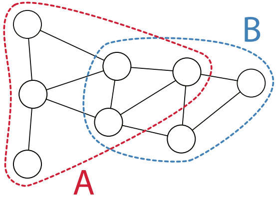
Graph decomposition

- $G = (V, E_G)$
- $A, B \subseteq V$

Definition

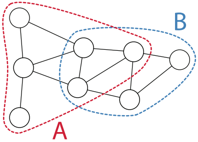
(A, B) is a **decomposition** of G if

- $A \cup B = V$
- $A \cap B$ is a clique
- $A \cap B$ separates A from B



Markov property

- $Y = (Y_1, \dots, Y_p)$ **a random vector** taking values in $\mathcal{Y} = \bigotimes_{i=1}^{p} \mathcal{Y}_i$
- $Y_A := (Y_\alpha)_{\alpha \in A}$, $A \subseteq V$



Definition

A distribution π for Y is **Markov** w.r.t. G, if for all decompositions (A, B) of G,

$$Y_A \perp\!\!\!\perp Y_B \,|\, Y_{A \cap B}$$

$$\mathfrak{M}_G = \{\text{distributions } \pi \text{ Markov w.r.t. } G\}$$

Factorisation

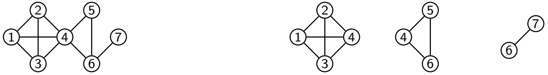
- $Y = (Y_1, \ldots, Y_p)$ **a random vector** taking values in $\mathcal{Y} = \bigotimes_{i=1}^{p} \mathcal{Y}_i$
- $Y_A := (Y_\alpha)_{\alpha \in A}$, $A \subseteq V$
- $Y \sim \pi$ with positive density

Proposition

$$\left(\pi \text{ Markov w.r.t. } G \right) \quad \Leftrightarrow \quad \left(\pi(Y) = \prod_{C \in \mathcal{C}_G} \psi_C(Y_C) \right)$$

- $\mathcal{C}_G$ maximal cliques of G

Example



Graphical models (formally)

Definition (Graphical Model)

An **undirected graphical model** is a couple $(G, \mathcal{F}_G)$ where

- G is an undirected graph,
- $\mathcal{F}_G \subseteq \mathfrak{M}_G$ is a family of distributions Markov w.r.t. G.

Graphical models (formally)

Definition (Graphical Model)

An **undirected graphical model** is a couple $(G, \mathcal{F}_G)$ where

- G is an undirected graph,
- $\mathcal{F}_G \subseteq \mathfrak{M}_G$ is a family of distributions Markov w.r.t. G.

Bayesian inference

ρ distribution on $\mathcal{F}_G$ called **hyperdistribution**

$$\pi \sim \rho$$

- Some Markov property for ρ

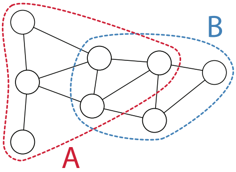
Strong hyper Markov property

- $(G, \mathcal{F}_G)$ graphical model
- ρ hyperdistribution on $\mathcal{F}_G$
- $\pi \sim \rho$

Definition

ρ is said to be **strong hyper Markov** w.r.t. G if, for any decomposition (A, B) of G,

$$\pi_A \perp\!\!\!\perp \pi_{B|A}.$$



Strong hyper Markov property

- $(G, \mathcal{F}_G)$ graphical model
- ρ hyperdistribution on $\mathcal{F}_G$
- $\pi \sim \rho$

Proposition (Dawid and Lauritzen 1993)

If ρ is strong hyper Markov w.r.t. G, then the marginal likelihood

$$p(Y) = \int \pi(Y)\rho(\pi)d\pi$$

is Markov w.r.t. to G.

Strong hyper Markov property

- $(G, \mathcal{F}_G)$ graphical model
- ρ hyperdistribution on $\mathcal{F}_G$
- $\pi \sim \rho$

Proposition (Dawid and Lauritzen 1993)

If ρ is strong hyper Markov w.r.t. G, then the marginal likelihood

$$p(Y) = \int \pi(Y)\rho(\pi)d\pi$$

is Markov w.r.t. to G.

$$p(Y) = \prod_{C \in \mathcal{C}_G} \underbrace{\psi_C(Y_C)}_{\text{locally integrated}}$$

Compatible family

- ρ strong hyper Markov w.r.t. to the **complete graph**
- G graph

Proposition (Dawid and Lauritzen 1993)

There exists a **unique** *hyperdistribution* ρ^G *on* $\mathfrak{M}_G$ *that is* **strong hyper Markov w.r.t.** G *and s.t.*

$$\forall C \in \mathcal{C}_G, \ \rho^G_C = \rho_C.$$

- ρ_C hyperdistribution induced by ρ on π_C
- $\mathcal{C}_G$ maximal cliques of G

Compatible family

- $\mathcal{G} = \{G_1, \ldots, G_u\}$ graph family
- ρ strong hyper Markov w.r.t. to the complete graph

Definition

The **compatible family** built on $\mathcal{G}$ from ρ is given by $\{\rho^G\}_{G \in \mathcal{G}}$

$$C \in \mathcal{C}_{G_i} \cap \mathcal{C}_{G_j}$$

$$\underbrace{\int \pi_C(Y_C)\rho_C^{G_i}(\pi_C)d\pi_C}_{p(Y_C|G_i)} = \underbrace{\int \pi_C(Y_C)\rho_C^{G_j}(\pi_C)d\pi_C}_{p(Y_C|G_j)}$$

In practice

- ρ given by a **parametric distribution**

Y	π	ρ
$\mathbf{R}^p$	Normal	(Normal-)Wishart Geiger and Heckerman 2002
$\{1, \ldots, r\}^p$	Multinomial	Dirichlet Dawid and Lauritzen 1993
$[0; 1]^p$	Copula	Depends on the copula Schwaller et al. 2015

Algebraic magic bag

Algebraic tricks

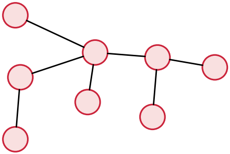
$$\Sigma\ \Pi$$

- Computing **sum-products**
  - For **spanning trees**
  - For **segmentations**

Spanning Trees

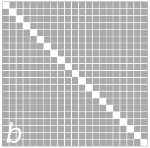
Definition

A **spanning tree** is a connected graph with no cycles.



- $\mathcal{T} := \left\{ T = (V, E_T) \text{ spanning tree on } V \right\}$
- $|\mathcal{T}| = p^{p-2}$
- Maximal cliques = edges

Summing over $\mathcal{T}$



$$Z(b) = \sum_{T \in \mathcal{T}} \prod_{\{i,j\} \in E_T} b_{ij}$$

Theorem (Matrix-Tree, Kirchhoff 1847, Cayley 1889)

$$\Delta_{ij} = \begin{cases} -b_{ij} & \text{if } i \neq j \\ \sum_k b_{kj} & \text{if } i = j \end{cases}$$

All cofactors of Δ are equal to $Z(b)$.

Complexity $= O(p^3)$

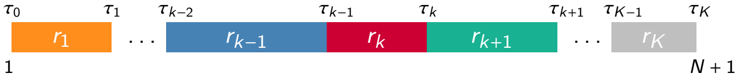
Segmentations

Definition

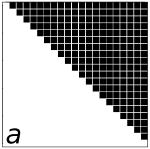
A **segmentation** of $[\![1;N]\!]$ is a partition of $[\![1;N]\!]$ into sets of consecutive elements called segments.

- $m = ([\![\tau_{k-1};\tau_k[\![)_{k=1}^K = (r_k)_{k=1}^K$
  - $1 = \tau_0 < \tau_1 < \ \dots\ < \tau_{K-1} < \tau_K = N+1$ **change-points** of m
  - $r_1, \dots, r_K$ **segments** of m

- $\mathcal{M}_K = \{$ segmentations of $[\![1;N]\!]$ into K segments $\}$
- $|\mathcal{M}_K| = \binom{N-1}{K-1}$



Summing over $\mathcal{M}_K$



$$C_K(a) = \sum_{m \in \mathcal{M}_K} \prod_{[\![s;t [\![\in m} a_{st}$$

Proposition (Rigaill et al. 2012)

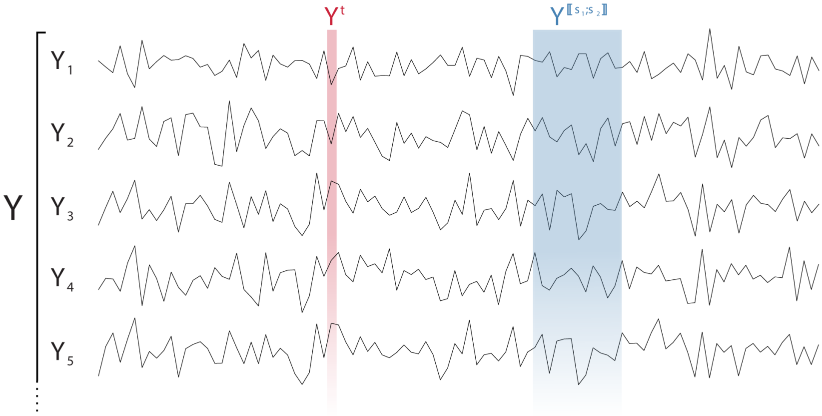
$$C_K(a) = [a^K]_{1,N+1}$$

Complexity $= O(KN^2)$

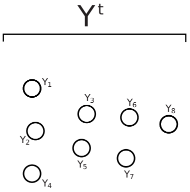
Model & Inference

A segmentation problem

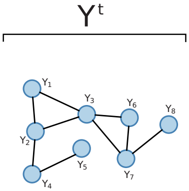
- $Y = \{Y^t\}_{t=1}^N$ multivariate random process of dimension p



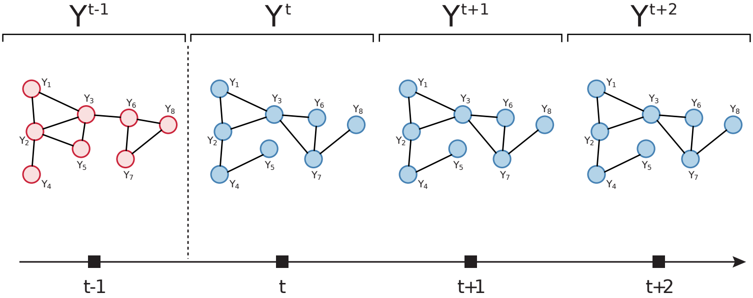
A segmentation problem



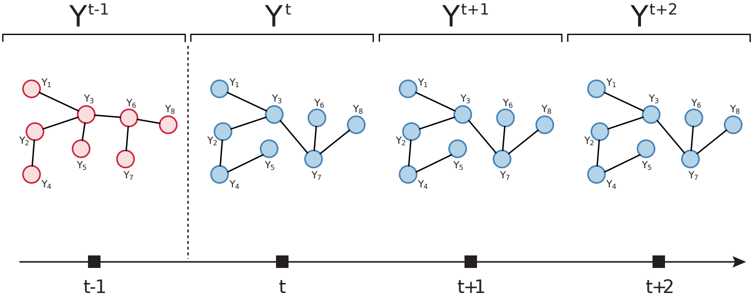
A segmentation problem



A segmentation problem



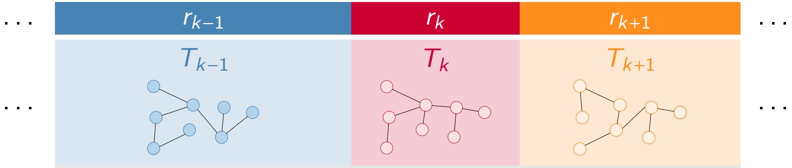
A segmentation problem



Model

$\ldots$ | r_{k-1} | r_k | r_{k+1} | $\ldots$

Model

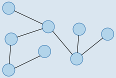


$$p(\mathbf{T}|K) = \prod_k p(T_k) \qquad \mathbf{T} = (T_1, \ldots, T_K)$$

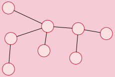
Model

… r_{k-1} | r_k | r_{k+1} …

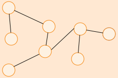
… T_{k-1}



T_k



T_{k+1} …

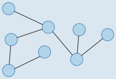
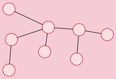
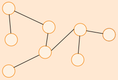


$$p(\mathbf{T}|K) = \prod_k p(T_k) \qquad \mathbf{T} = (T_1, \ldots, T_K)$$

… π_{k-1} Markov w.r.t. T_{k-1} | π_k Markov w.r.t. T_k | π_{k+1} Markov w.r.t. T_{k+1} …

$$p(\Pi|\mathbf{T}) = \prod_k p(\pi_k|T_k) \qquad \Pi = (\pi_1, \ldots, \pi_K)$$

Model

...	r_{k-1}	r_k	r_{k+1}	...
...	T_{k-1} 	T_k 	T_{k+1} 	...

$$p(\mathbf{T}|K) = \prod_k p(T_k) \qquad \mathbf{T} = (T_1, \dots, T_K)$$

...	π_{k-1} Markov w.r.t. T_{k-1}	π_k Markov w.r.t. T_k	π_{k+1} Markov w.r.t. T_{k+1}	...

$$p(\Pi|\mathbf{T}) = \prod_k p(\pi_k|T_k) \qquad \Pi = (\pi_1, \dots, \pi_K)$$

...	$Y^{r_{k-1}} \overset{\text{i.i.d.}}{\sim} \pi_{k-1}$	$Y^{r_k} \overset{\text{i.i.d.}}{\sim} \pi_k$	$Y^{r_{k+1}} \overset{\text{i.i.d.}}{\sim} \pi_{k+1}$	...

$$p(Y|m,\pi) = \prod_k \prod_{t \in r_k} p(Y^t|T_k, \pi_k)$$

Inference

- **Marginal likelihood**

$$p(Y|K) = \sum_{m \in \mathcal{M}_K} \sum_{\mathbf{T} \in \mathcal{T}^K} \int p(Y, \Pi, \mathbf{T}, m|K) d\Pi$$

$$|\mathcal{M}_K| \cdot |\mathcal{T}^K| = \binom{N-1}{K-1} \cdot p^{K(p-2)} \approx \left(\frac{Np^{p-2}}{K}\right)^K$$

Example

$$N = 200 \qquad |\mathcal{M}_4| \approx 1.3 \cdot 10^6$$

$$p = 10 \qquad |\mathcal{T}| = 10^8$$

A complexity result

Proposition (Schwaller and Robin 2016)

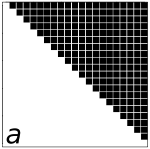
Under some assumptions on prior distributions, the marginal likelihood $p(Y|K)$ can be computed in $O(\max(K, p^3)N^2)$ time from locally integrated quantities on Π.

Prior distributions

"Under some assumptions on prior distributions"

- On **segmentations** .... m
- On **trees** .... $\mathbf{T} = (T_1, \ldots, T_K)$
- On **distributions** .... $\Pi = (\pi_1, \ldots, \pi_K)$

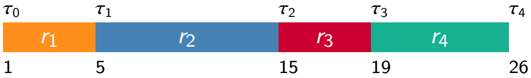
Prior distribution on m



$$p(m|K) = \frac{1}{C_K(a)} \prod_{[\![s;t[\![\in m} a_{st}$$

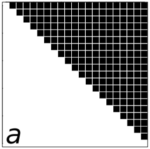
$$C_K(a) = \sum_{m \in \mathcal{M}_K} \prod_{[\![s;t[\![\in m} a_{st}$$

Example



$$p(m|K) = \frac{1}{C_K(a)} a_{1,5}\, a_{5,15}\, a_{15,19}\, a_{19,26}$$

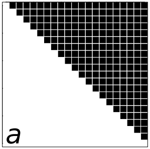
Prior distribution on m



$$p(m|K) = \frac{1}{C_K(a)} \prod_{[\![s;t[\![\in m} a_{st}$$

$$p(Y|K) = \sum_m p(m|K)p(Y|m) = \frac{1}{C_K(a)} \sum_m p(Y|m) \prod_{[\![s;t[\![\in m} a_{st}$$

Prior distribution on m



$$p(m|K) = \frac{1}{C_K(a)} \prod_{\llbracket s;t \llbracket \in m} a_{st}$$

$$p(Y|K) = \sum_m p(m|K)p(Y|m) = \frac{1}{C_K(a)} \underbrace{\sum_m p(Y|m) \prod_{\llbracket s;t \llbracket \in m} a_{st}}_{C_K(A)}$$

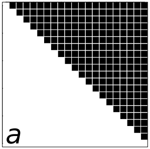
Factorising $p(Y|m)$

$$p(Y|m) = \sum_{\mathbf{T}\in\mathcal{T}^K} \int \underbrace{p(Y,\Pi,\mathbf{T}|m)}_{p(\mathbf{T})p(\Pi|\mathbf{T})p(Y|\Pi,m)} d\Pi$$

$$= \sum_{\mathbf{T}\in\mathcal{T}^K} \prod_k p(T_k) \int \prod_k p(Y^{r_k}|\pi_k) \prod_k p(\pi_k|T_k) d\pi_k$$

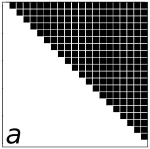
$$= \prod_k \underbrace{\sum_{T\in\mathcal{T}} p(T) \int p(Y^{r_k}|\pi_k) p(\pi_k|T) d\pi_k}_{p(Y^{r_k}) = \text{ marginal likelihood on } r_k}$$

Integrating on m



$$p(m) = \frac{1}{C_K(a)} \prod_{\llbracket s;t \llbracket \in m} a_{st}$$

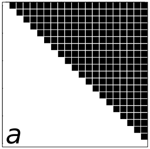
Integrating on m



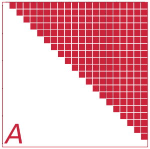
$$p(m) = \frac{1}{C_K(a)} \prod_{[\![s;t[\![\in m} a_{st}$$

$$p(Y|K) = \frac{1}{C_K(a)} \sum_{m \in \mathcal{M}_K} \prod_{[\![s;t[\![\in m} a_{st} \cdot p(Y^{[\![s;t[\![})$$

Integrating on m



$$p(m) = \frac{1}{C_K(a)} \prod_{\llbracket s;t \llbracket \in m} a_{st}$$



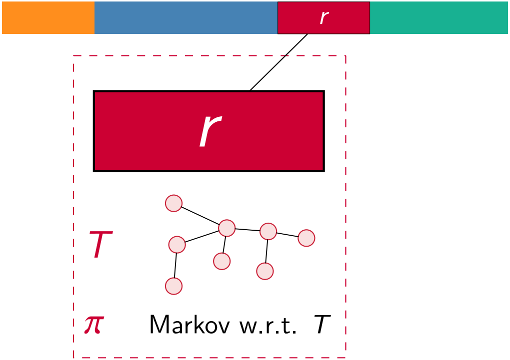
$$p(Y|K) = \frac{C_K(A)}{C_K(a)}$$

$$A_{st} = a_{st} \cdot p(Y^{\llbracket s;t \llbracket})$$

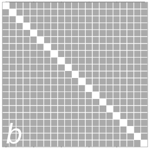
Complexity $= O(KN^2)$

Marginal likelihood on a segment

$$p(Y^r) = \sum_{T \in \mathcal{T}} p(T) \int p(Y^r|\pi) p(\pi|T) d\pi$$



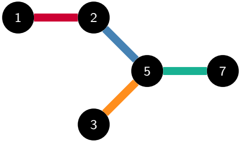
Prior distribution on T



$$p(T) = \frac{1}{Z(b)} \prod_{\{i,j\} \in E_T} b_{ij}$$

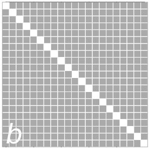
$$Z(b) = \sum_{T \in \mathcal{T}} \prod_{\{i,j\} \in E_T} b_{ij}$$

Example



$$p(T) = \frac{1}{Z(b)} b_{12} b_{25} b_{35} b_{57}$$

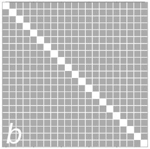
Prior distribution on T



$$p(T) = \frac{1}{Z(b)} \prod_{\{i,j\} \in E_T} b_{ij}$$

$$p(Y^r) = \frac{1}{Z(b)} \sum_{T \in \mathcal{T}} \int p(Y^r|\pi) p(\pi|T) d\pi \prod_{\{i,j\} \in E_T} b_{ij}$$

Prior distribution on T



$$p(T) = \frac{1}{Z(b)} \prod_{\{i,j\} \in E_T} b_{ij}$$

$$p(Y^r) = \frac{1}{Z(b)} \underbrace{\sum_{T \in \mathcal{T}} \int p(Y^r|\pi) p(\pi|T) d\pi \prod_{\{i,j\} \in E_T} b_{ij}}_{\propto Z(B)}$$

Prior distribution on π

- ρ strong hyper Markov w.r.t. to the complete graph
- $\{\rho^T\}_{T\in\mathcal{T}}$ compatible family built from ρ on $\mathcal{T}$

$$p(\pi|T) = \rho^T(\pi) \qquad \forall T \in \mathcal{T}$$

Prior distribution on π

- ρ strong hyper Markov w.r.t. to the complete graph
- $\{\rho^T\}_{T \in \mathcal{T}}$ compatible family built from ρ on $\mathcal{T}$

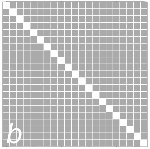
$$p(\pi | T) = \rho^T(\pi) \qquad \forall T \in \mathcal{T}$$

$$p(Y^r | T) = \underbrace{\prod_i p(Y_i^r)}_{U^{(r)}} \prod_{\{i,j\} \in E_T} \frac{p(Y_i^r, Y_j)}{p(Y_i^r) p(Y_j^r)}$$

$$p(Y_i^r, Y_j^r) = \int \pi_{ij}(Y_i^r, Y_j^r) \rho_{ij}(\pi_{ij}) d\pi_{ij}$$

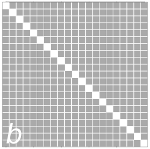
$$p(Y_i^r) = \int \pi_i(Y_i^r) \rho_i(\pi_i) d\pi_i$$

Integrating on T



$$p(T) = \frac{1}{Z(b)} \prod_{\{i,j\} \in E_T} b_{ij}$$

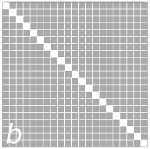
Integrating on T



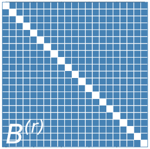
$$p(T) = \frac{1}{Z(b)} \prod_{\{i,j\} \in E_T} b_{ij}$$

$$p(Y^r) = \frac{1}{Z(b)} U^{(r)} \sum_{T \in \mathcal{T}} \prod_{\{i,j\} \in E_T} b_{ij} \frac{p(Y_i^r, Y_j^r)}{p(Y_i^r) p(Y_j^r)}$$

Integrating on T



$$p(T) = \frac{1}{Z(b)} \prod_{\{i,j\} \in E_T} b_{ij}$$

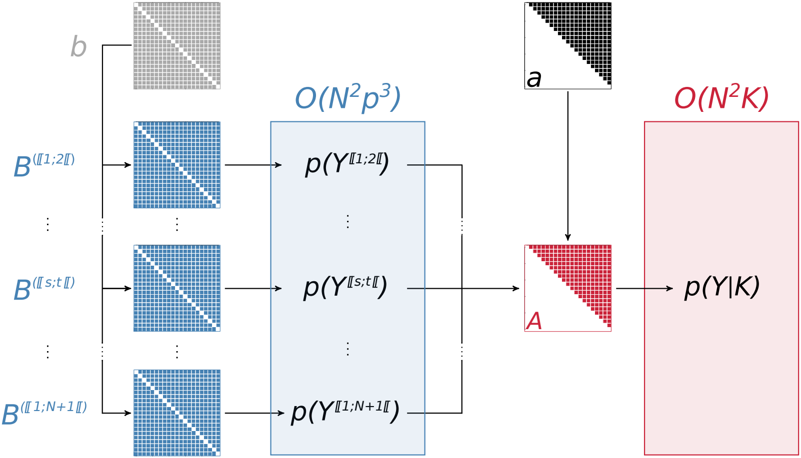


$$p(Y^r) = U^{(r)} \cdot \frac{Z(B^{(r)})}{Z(b)}$$

$$B_{ij}^{(r)} = b_{ij} \cdot \frac{p(Y_i^r, y_j^r)}{p(Y_i^r) p(Y_j^r)}$$

Complexity $= O(p^3)$

In a nutshell



In a nutshell

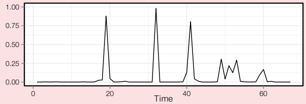
Proposition (Schwaller and Robin 2016)

Under some assumptions on prior distributions, the marginal likelihood $p(Y|K)$ *can be computed in* $O(\max(K, p^3)N^2)$ *time from locally integrated quantities on* Π.

Other quantities

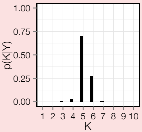
- Posterior probability of events

{there is a change-point at time t}
$\{m$ contains segment $[\![s;t[\![\}$



- Posterior distribution of K

$$p(K|Y) \propto \frac{p(K)[A^K]_{1,N+1}}{[a^K]_{1,N+1}}$$

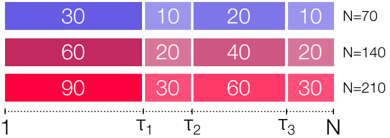


Complexity $= O(\max(p^3, K)N^2)$

Simulations

Data generation

- $p = 10$ variables
- Time-series of size $N = 70$, 140 and 210
- Four segments of lengths $\frac{3}{7}N$, $\frac{1}{7}N$, $\frac{2}{7}N$



- Three structure scenarios
  - spanning trees
  - Erdös-Rényi, $p_C = 2/p$
  - Erdös-Rényi, $p_C = 4/p$
- For each sample size and structure series, 100 datasets with centered Gaussian observations were generated.

Inference

Tree model

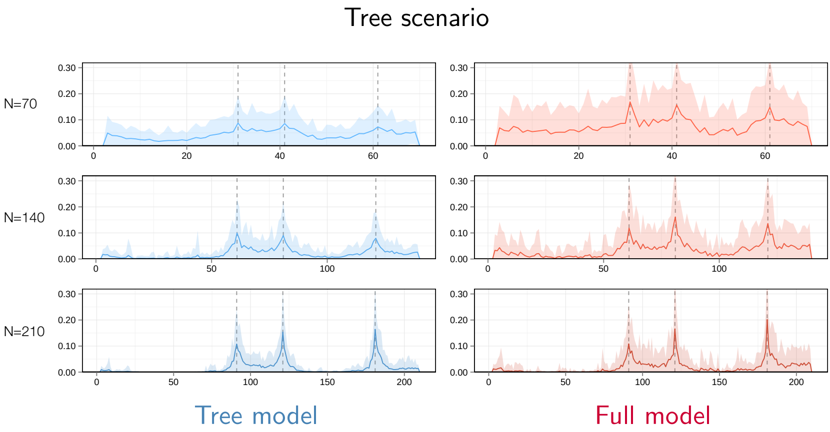
Number of segments	K	$K \sim \mathcal{P}(4)$
Segmentation	m	$m \sim \mathcal{U}(\mathcal{M}_K)$
Trees	$\{T_k\}_{k=1}^{K}$ i.i.d.	$T_k \sim \mathcal{U}(\mathcal{T})$
Precision matrices	$\{\Lambda_r\}_{r\in m}$ ind.	$\Lambda_r \sim h\mathcal{W}(\alpha, \phi, T_{\kappa(r\|m)})$
Observations	$\{Y_t\}_{t=1}^{N}$ ind.	$Y^t \sim \mathcal{N}(\mathbf{0}_p, \Lambda_r) \quad \forall t \in r$

Inference

Full model

Number of segments	K	$K \sim \mathcal{P}(4)$
Segmentation	m	$m \sim \mathcal{U}(\mathcal{M}_K)$
Precision matrices	$\{\Lambda_r\}_{r\in m}$ i.i.d	$\Lambda_r \sim \mathcal{W}(\alpha, \phi)$
Observations	$\{Y_t\}_{t=1}^N$ ind.	$Y^t \sim \mathcal{N}(\mathbf{0}_p, \Lambda_r) \quad \forall t \in r$

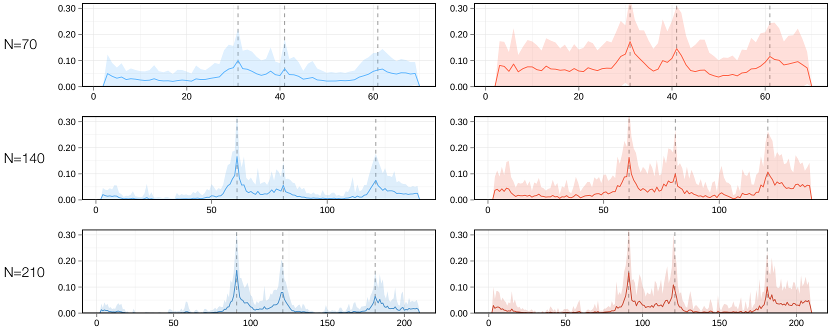
Results



Posterior probability of a change-point occurring at time t, integrated on K

Results

ER, low density scenario

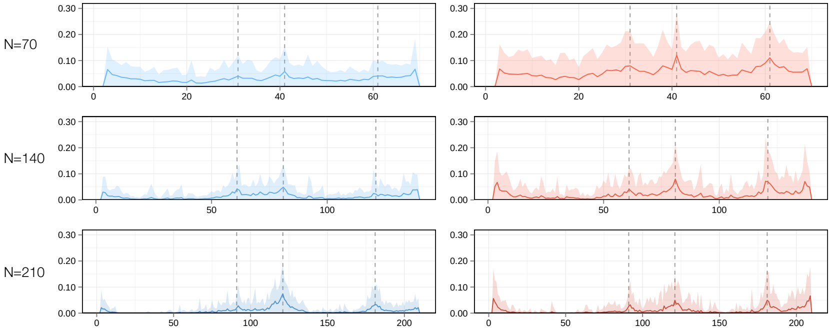


Tree model | Full model

Posterior probability of a change-point occurring at time t, integrated on K

Results

ER, high density scenario



Tree model | Full model

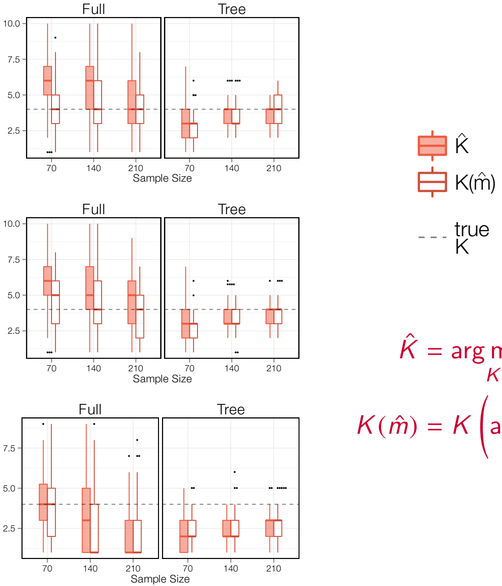
Posterior probability of a change-point occurring at time t, integrated on K

Results

Tree

ER
low density

ER
high density



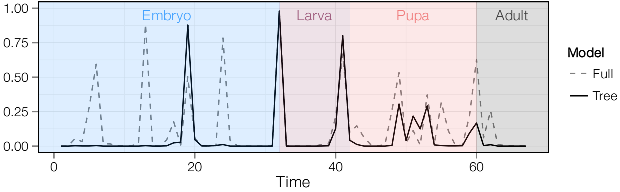
$$\hat{K} = \arg\max_K p(K|y)$$

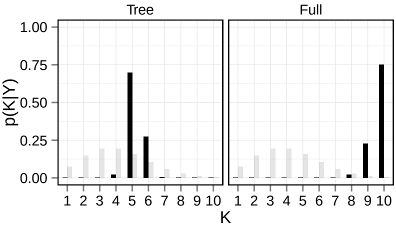
$$K(\hat{m}) = K\left(\arg\max_m p(m|y)\right)$$

Applications

Drosophila life cycle microarray data

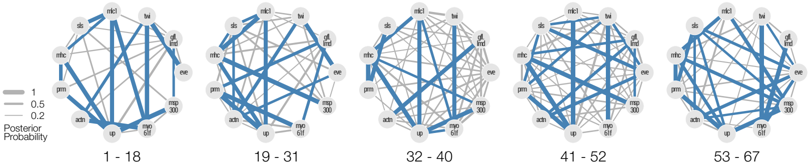
- Expression levels of 11 genes involved in wing muscle development
- 67 time-points





Drosophila life cycle microarray data

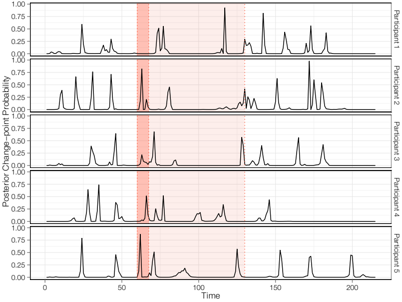
- 5 segments selected by $p(K|Y)$
- Best segmentation found by Neighbourhood Search algorithm (Auger and Lawrence 1989)

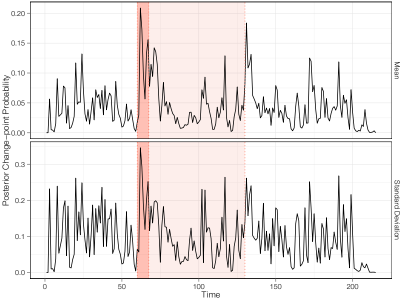


Schwaller L. (2015) **saturnin**: *SpAnning Trees Used foR Network InfereNce*
https://cran.r-project.org/web/packages/saturnin/

fMRI Data[2]

- Averaged blood flow of 5 regions
- 20 participants
- 215 images at 2s intervals

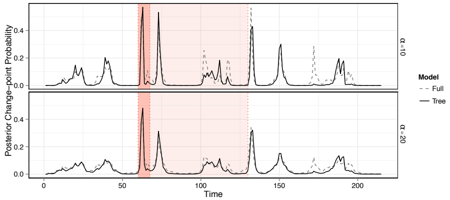


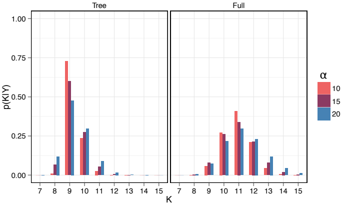


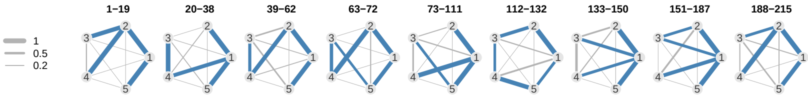
[2] Cribben et al. 2012.

fMRI data[3]

$$p^*_\alpha(y^r) = \sum_{T \in \mathcal{T}} \prod_{u=1}^{U} \left[\int \prod_{t \in r} p(y^{t,u}|\theta_u) p(\theta_u|T) d\theta_u \right]^{1/\alpha}$$







[3] Cribben et al. 2012.

Extensions

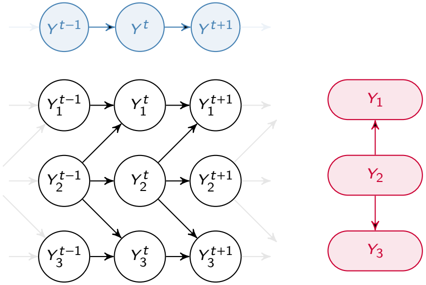
- Covariates
  - Easy to take into account in a Gaussian setting
  - Not so much generally

- Application to **microbial ecology**
  using a copula-based pragmatical approach
  - Learning direct interactions between species
  - Taking environmental covariates into account

B. Jakuschkin et al. (2016) "Deciphering the pathobiome: intra- and interkingdom interactions involving the pathogen Erysiphe alphitoides" *Microbial Ecology*

Extensions

- Temporal dependence
  - Within segments
  - Based on Temporal Interaction Models (Siracusa 2009)



Conclusion & Perspectives

Conclusion & Perspectives

Problem at hand

- **Segmentation** of the dependence structure in a multivariate time-series

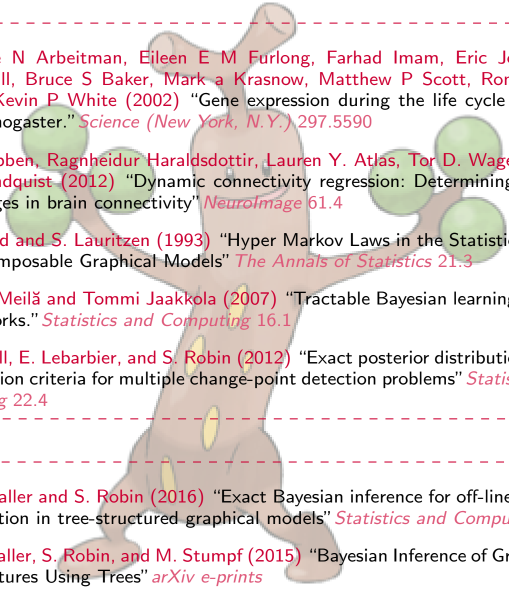
So far

- **Exact & Bayesian** inference in $\mathcal{O}(\max(p^3, K)N^2)$ time
- Using **algebraic results** on
  - Spanning Trees
  - Segmentations
- Using **(hyper) Markov properties**

Perspectives

- Numerical issues
- R package

References



- Michelle N Arbeitman, Eileen E M Furlong, Farhad Imam, Eric Johnson, Brian H Null, Bruce S Baker, Mark a Krasnow, Matthew P Scott, Ronald W Davis, and Kevin P White (2002) “Gene expression during the life cycle of Drosophila melanogaster.” *Science (New York, N.Y.)* 297.5590
- Ivor Cribben, Ragnheidur Haraldsdottir, Lauren Y. Atlas, Tor D. Wager, and Martin a. Lindquist (2012) “Dynamic connectivity regression: Determining state-related changes in brain connectivity” *NeuroImage* 61.4
- A. Dawid and S. Lauritzen (1993) “Hyper Markov Laws in the Statistical Analysis of Decomposable Graphical Models” *The Annals of Statistics* 21.3
- Marina Meilă and Tommi Jaakkola (2007) “Tractable Bayesian learning of tree belief networks.” *Statistics and Computing* 16.1
- G. Rigaill, E. Lebarbier, and S. Robin (2012) “Exact posterior distributions and model selection criteria for multiple change-point detection problems” *Statistics and Computing* 22.4

- L. Schwaller and S. Robin (2016) “Exact Bayesian inference for off-line change-point detection in tree-structured graphical models” *Statistics and Computing*
- L. Schwaller, S. Robin, and M. Stumpf (2015) “Bayesian Inference of Graphical Model Structures Using Trees” *arXiv e-prints*