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

Loïc Schwaller, Stéphane Robin

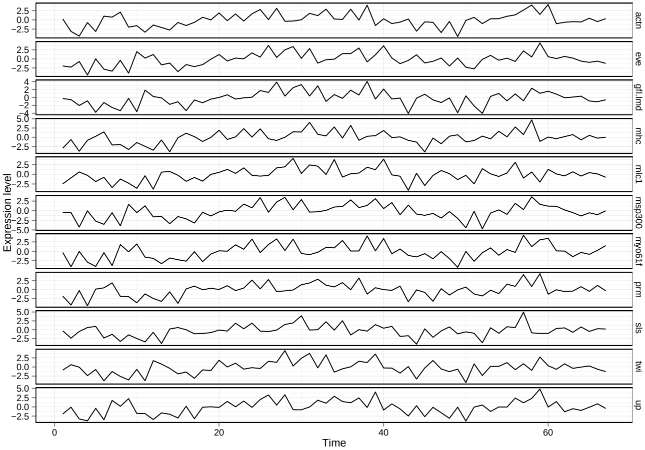
AgroParisTech INRA SCIENCE & IMPACT

Colloque Apprentissage de Réseaux :
de la Théorie aux Applications en BioLogie et Ecologie

October 14, 2016

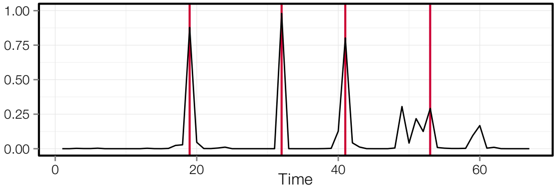
A segmentation problem

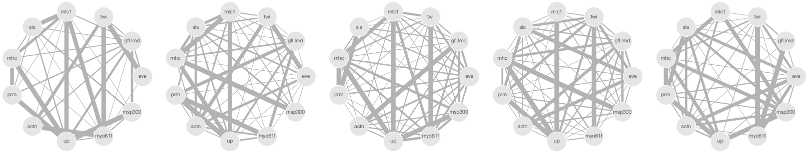
- Expression levels of 11 genes involved in wing muscle development[1]
- 67 time-points



[1] Arbeitman et al. 2002.

- Expression levels of 11 genes involved in wing muscle development[1]
- 67 time-points

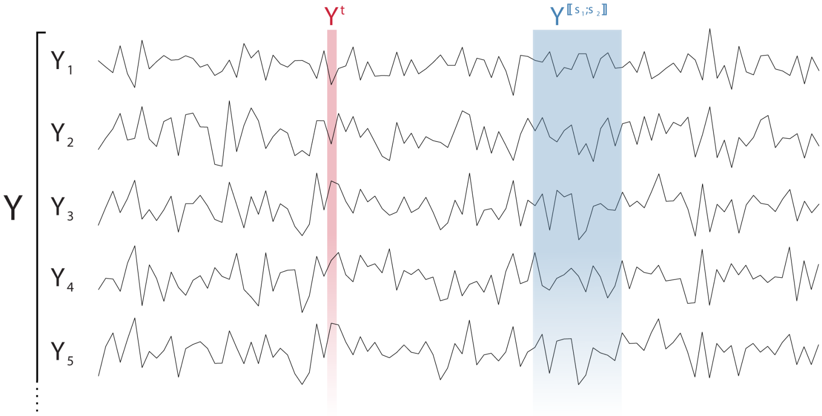




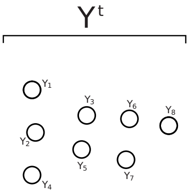
[1] Arbeitman et al. 2002.

A segmentation problem

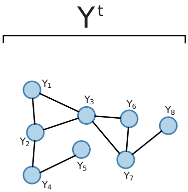
- $Y = \{Y^t\}_{t=1}^N$ multivariate random process of dimension p



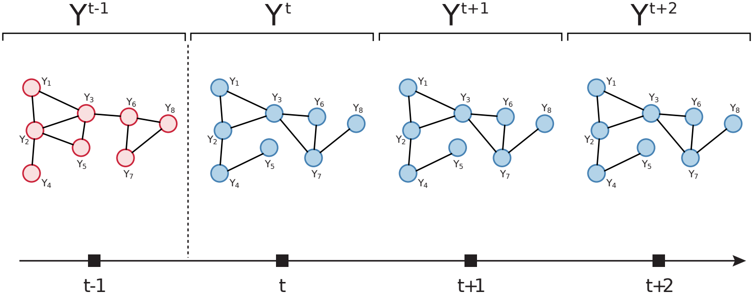
A segmentation problem



A segmentation problem

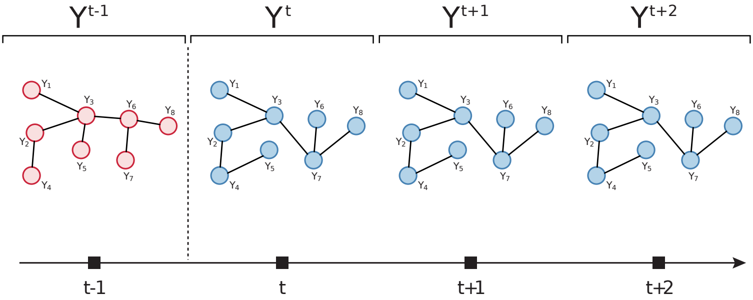


A segmentation problem



A segmentation problem

$$p(Y^t|T) = \prod_i p_i(Y_i^t) \prod_{\{i,j\}\in E_T} \frac{p_{ij}(Y_i^t, Y_j^t)}{p_i(Y_i^t)p_j(Y_j^t)}$$



Spanning tree assumption

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
  - p^{p-2} spanning trees

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 - p^{p-2} spanning trees	▷ Algebraic results - Segmentations - Spanning trees (Meilă and Jaakkola 2007)

Algebraic magic bag

Algebraic tricks

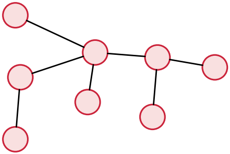
$$\sum \prod$$

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

Spanning Trees

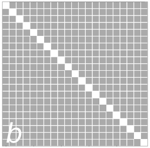
Definition

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 $= \mathcal{O}(p^3)$

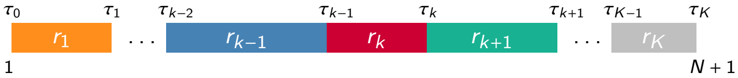
Segmentations

Definition

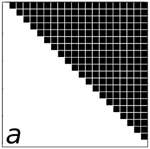
A **segmentation** of $[\![1; N]\!]$ is a partition of $[\![1; N]\!]$ into sets of consecutive elements called segments.

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

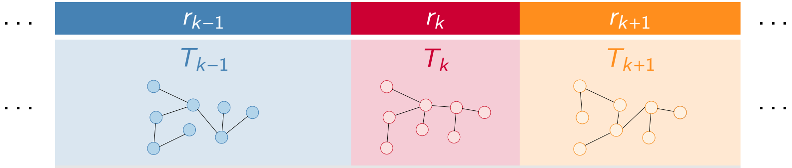
$$C_K(a) = [a^K]_{1,N+1}$$

Complexity $= \mathcal{O}(KN^2)$

Model & Inference

Model

Model

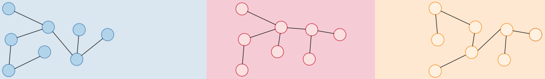


$$p(\mathbf{T}|K) = \prod_k p(T_k) \qquad \mathbf{T} = (T_1, \ldots, T_K)$$

Model

… r_{k-1} | r_k | r_{k+1} …

… T_{k-1} | T_k | T_{k+1} …

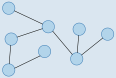
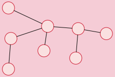
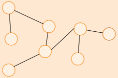


$$p(\mathbf{T}|K) = \prod_k p(T_k) \qquad \mathbf{T} = (T_1, \ldots, T_K)$$

… 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} 	…

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

…	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}	…

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

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

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

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

$N = 200$	$\lvert\mathcal{M}_4\rvert \approx 1.3 \cdot 10^6$
$p = 10$	$\lvert\mathcal{T}\rvert = 10^8$

A complexity result

Proposition (Schwaller and Robin 2016)

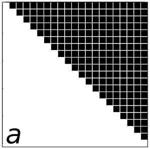
Under some assumptions on prior distributions, the marginal likelihood $p(Y|K)$ can be computed in $\mathcal{O}(\max(K, p^3)N^2)$ time from locally integrated quantities on Π.

Prior distributions

"Under some assumptions on prior distributions"

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

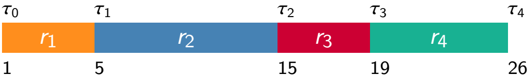
Prior distribution on m



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

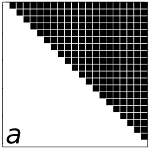
$$C_K(a) = \sum_{m \in \mathcal{M}_K} \prod_{[\![s;t[\![\in m} a_{st}$$

Example



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

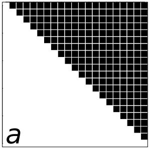
Prior distribution on m



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

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

Prior distribution on m



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

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

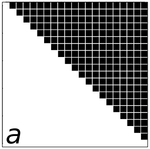
Factorising $p(Y|m)$

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

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

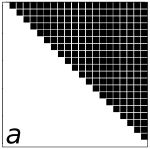
$$= \prod_k \underbrace{\sum_{T \in \mathcal{T}} p(T) \int p(Y^{r_k}|\pi_k) p(\pi_k|T) d\pi_k}_{p(Y^{r_k}) = \text{ marginal likelihood on } r_k}$$

Integrating on m



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

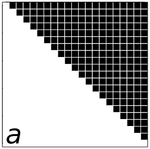
Integrating on m



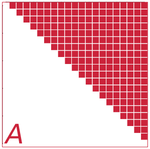
$$p(m) = \frac{1}{C_K(a)} \prod_{[\![s;t[\![\in m} a_{st}$$

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

Integrating on m



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



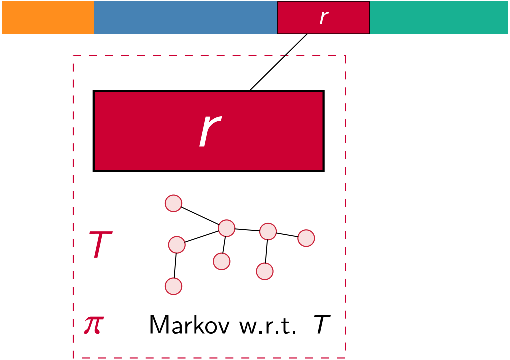
$$p(Y|K) = \frac{C_K(A)}{C_K(a)}$$

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

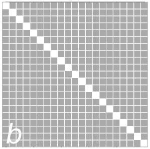
Complexity $= \mathcal{O}(KN^2)$

Marginal likelihood on a segment

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



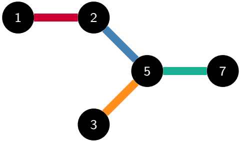
Prior distribution on T



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

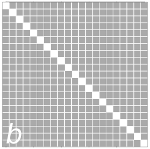
$$Z(b) = \sum_{T \in \mathcal{T}} \prod_{\{i,j\} \in E_T} b_{ij}$$

Example



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

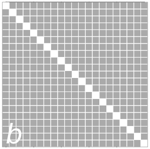
Prior distribution on T



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

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

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

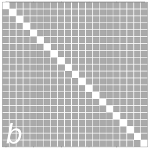
- Under some assumption on $\{p(\pi|T)\}_{T\in\mathcal{T}}$

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

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

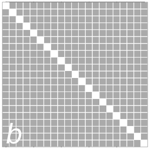
$$p(Y_i^r) = \int \pi_i(Y_i^r)\rho_i(\pi_i)d\pi_i$$

Integrating on T



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

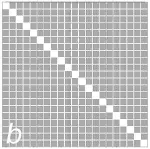
Integrating on T



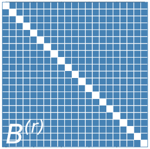
$$p(T) = \frac{1}{Z(b)} \prod_{\{i,j\} \in E_T} b_{ij}$$

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

Integrating on T



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

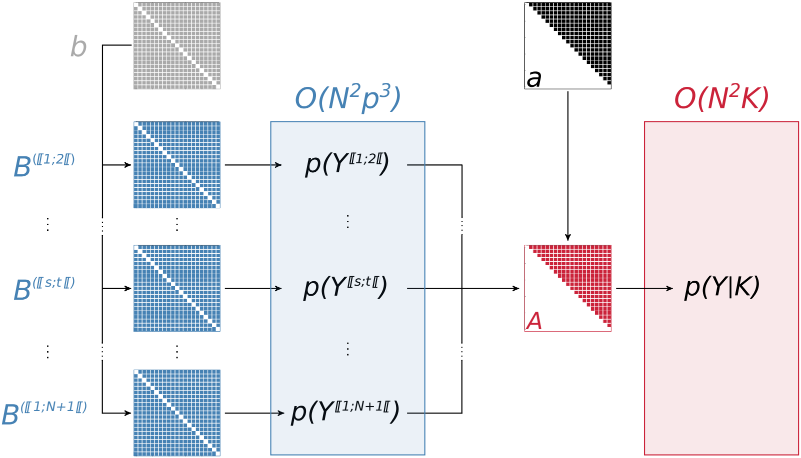


$$p(Y^r) = U^{(r)} \cdot \frac{Z(B^{(r)})}{Z(b)}$$

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

Complexity $= \mathcal{O}(p^3)$

In a nutshell



In a nutshell

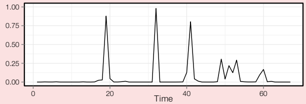
Proposition (Schwaller and Robin 2016)

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

Other quantities

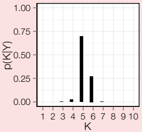
- Posterior probability of events

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



- Posterior distribution of K

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

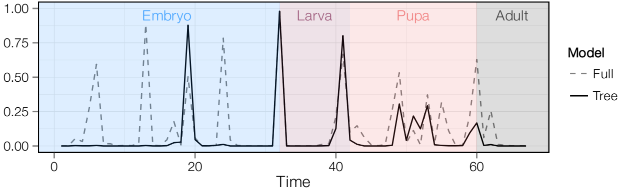


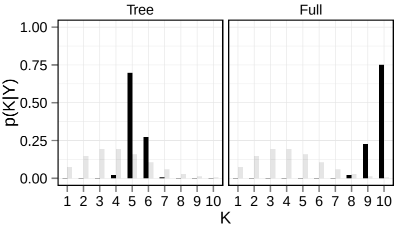
Complexity $= \mathcal{O}(\max(p^3, K)N^2)$

Application

Drosophila life cycle microarray data

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





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

- Temporal dependency within segments
- 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
- Arthur Cayley (1889) "A theorem on trees" *Quaterly Journal of Mathematics* 23
- G. Kirchhoff (1847) "Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird" *Annalen der Physik* 148.12
- 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*