Spatio-Temporal Bayesian On-line Changepoint Detection with Model Selection

Inference for non-stationary spatio-temporal processes

Jeremias Knoblauch[1] & Theodoros Damoulas[2]

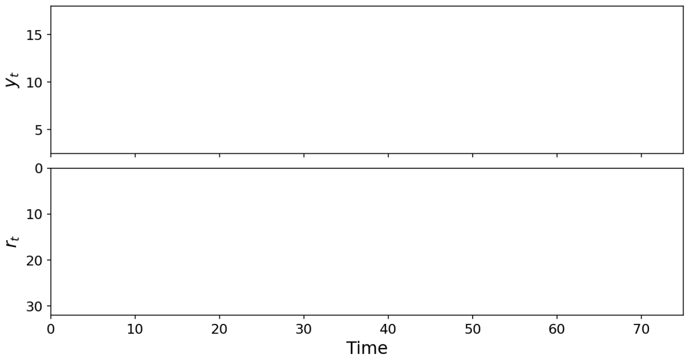
July 12, 2018

[1,2]University of Warwick, Department of Statistics
[2]University of Warwick, Department of Computer Science
[2]The Alan Turing Institute for Data Science and AI

Standard Bayesian On-line Changepoint (CP) Detection

Idea due to Adams and MacKay (2007):

(1) Define **Run-length at** $t = r_t \iff$ there was a CP at time $t - r_t$.
(2) **Inference on last CP** via $p(r_t|y_{1:t})$ rather than on *all* CPs
(3) Resulting complexity: $\mathcal{O}(t)$ **rather than** $\mathcal{O}(\prod_{i=1}^{t} i)$.



Standard BOCPD: Probabilistic model & Inference

$$r_t | r_{t-1} \sim H(r_t, r_{t-1}) \qquad \text{[conditional CP prior]} \tag{1a}$$

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}) \qquad \text{[parameter prior]} \tag{1b}$$

$$\boldsymbol{y}_t | \boldsymbol{\theta} \sim f(\boldsymbol{y}_t | \boldsymbol{\theta}) \qquad \text{[observation density prior]} \tag{1c}$$

For efficient inference, we need (fast) integrability of

$$f(\boldsymbol{y}_t | \boldsymbol{y}_{1:(t-1)}, r_{t-1}) = \int_{\Theta} f(\boldsymbol{y}_t | \boldsymbol{\theta}) \pi(\boldsymbol{\theta} | \boldsymbol{y}_{(t-r_{t-1}):(t-1)}) d\boldsymbol{\theta}. \tag{2}$$

$\Longrightarrow$ On-line inference [Prediction, CP detection] via **recursion**:

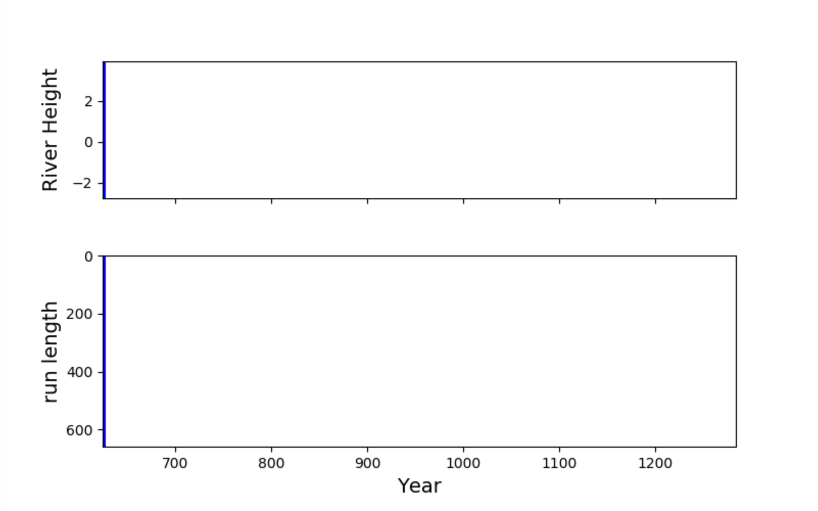
$$p(\boldsymbol{y}_{1:t}, r_t) = \sum_{r_{t-1}} \left\{ f(\boldsymbol{y}_t | \boldsymbol{y}_{1:(t-1)}, r_{t-1}) H(r_t, r_{t-1}) p(\boldsymbol{y}_{1:(t-1)}, r_{t-1}) \right\}. \tag{3}$$

Observations:

(1) Assumes that the same model (f, π) holds in each segment

(2) CPs are shifts in the parameterization $\boldsymbol{\theta}$ of that model

Standard BOCPD: Illustration using AR(1) on Nile data



Standard BOCPD: Limitations

Limitations:

- So far: inherently **univariate** method
- Each segment described by the **same model** $m = (f, \pi)$

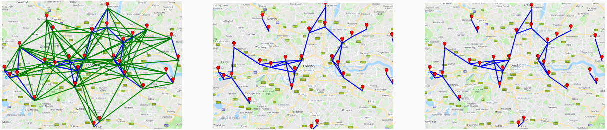
Contributions:

- Construct inherently **multivariate** models m
- Allow **multiple models** $\{m_1, m_2, \ldots, m_K\}$ for the segments

Extension I: Multivariate/spatio-temporal models

Steps: (1) Specify dependency structure, (2) impose on VAR model.

$$\boldsymbol{Y}_t = \alpha + \boldsymbol{B}\boldsymbol{Z}_t \quad + \boldsymbol{A}_1\boldsymbol{Y}_{t-1} \quad + \boldsymbol{A}_2\boldsymbol{Y}_{t-2} \quad + \boldsymbol{A}_3\boldsymbol{Y}_{t-3} \quad + \varepsilon_t$$



Advantages:

(1) **Sparsity** via $\boldsymbol{A}_l$

(2) **Predictive power** of spatial dependence

(3) Allows fast on-line computation & **incremental updates**

Graph uncertain? $\Longrightarrow$ Model selection on graphs

Extension II: Model Selection

Idea: We allow a change of models at CP locations
Model Universe: $\mathcal{M}$, a finite & countable set of admissible models
New Random Variable: m_t, the model at time t

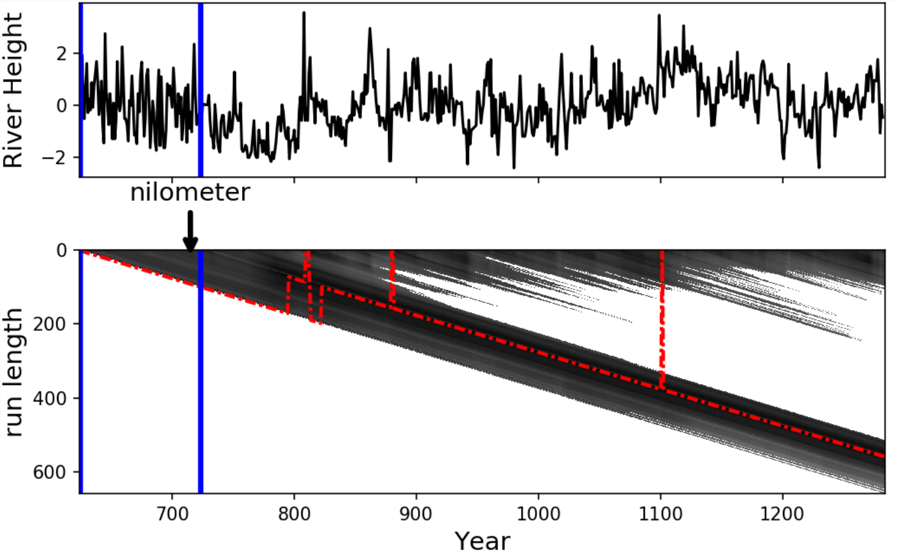
$$
\begin{aligned}
r_t|r_{t-1} &\sim H(r_t, r_{t-1}) && \text{[conditional CP prior]} && (4a)\\
m_t|m_{t-1}, r_t &\sim q(m_t|m_{t-1}, r_t) && \text{[conditional model prior]} && (4b)\\
\boldsymbol{\theta}_m|m_t &\sim \pi_{m_t}(\boldsymbol{\theta}_{m_t}) && \text{[parameter prior]} && (4c)\\
\boldsymbol{y}_t|m_t, \boldsymbol{\theta}_{m_t} &\sim f_{m_t}(\boldsymbol{y}_t|\boldsymbol{\theta}_{m_t}) && \text{[observation density prior]} && (4d)
\end{aligned}
$$

where $q(m_t|m_{t-1}, r_t) = \mathbb{1}_{\{r_t>0\}}\delta(m_{t-1}) + \mathbb{1}_{\{r_t=0\}}q(m_t)$.

$\Longrightarrow$ Prediction, CP detection, **Model Selection** via **recursion:**

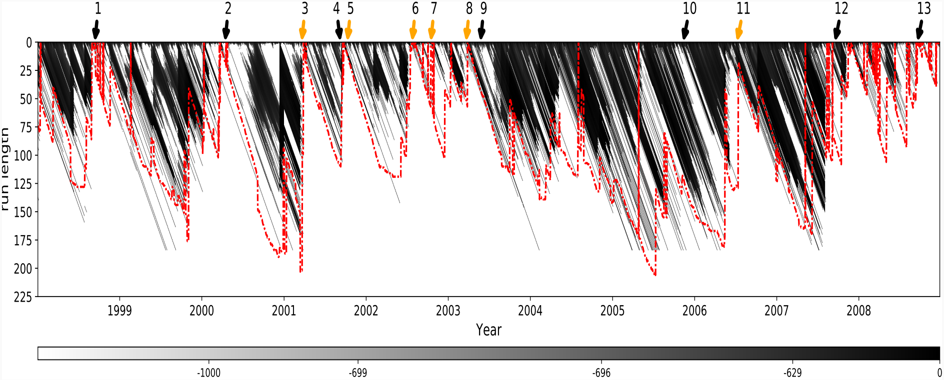
$$
p(\boldsymbol{y}_{1:t}, r_t, m_t) = \sum_{m_{t-1}, r_{t-1}} \Big\{ f_{m_t}(\boldsymbol{y}_t|\boldsymbol{y}_{1:(t-1)}, r_{t-1}) q(m_t|\boldsymbol{y}_{1:(t-1)}, r_{t-1}, m_{t-1}) \\
H(r_t, r_{t-1}) p(\boldsymbol{y}_{1:(t-1)}, r_{t-1}, m_{t-1}) \Big\}
$$

Improved CP detection [on Nile data]



Improved CP detection [compared to GP CP methods]

[on 30 Industry Portfolio return data]



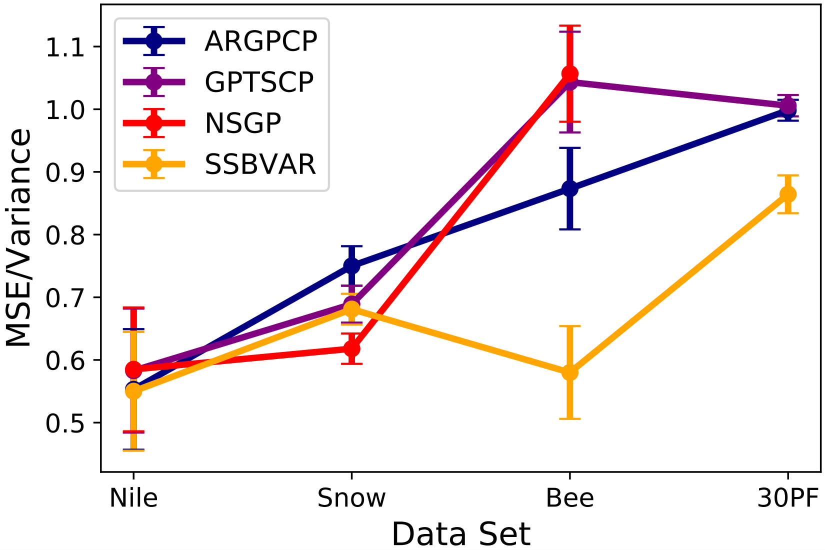
CPs found with GPs by Saatçi et al. (2010) in **black**, some **new CPs** found by BOCPDMS are:

(8) Iraq war,

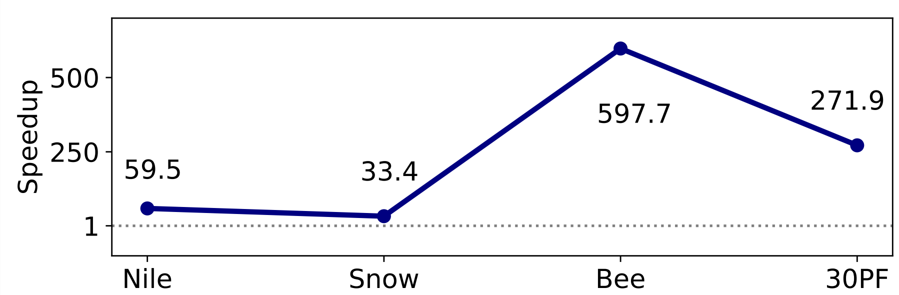
(11) Iran announces successful enrichment of Uranium,

Improved Prediction of multiple VARs vs single GP models

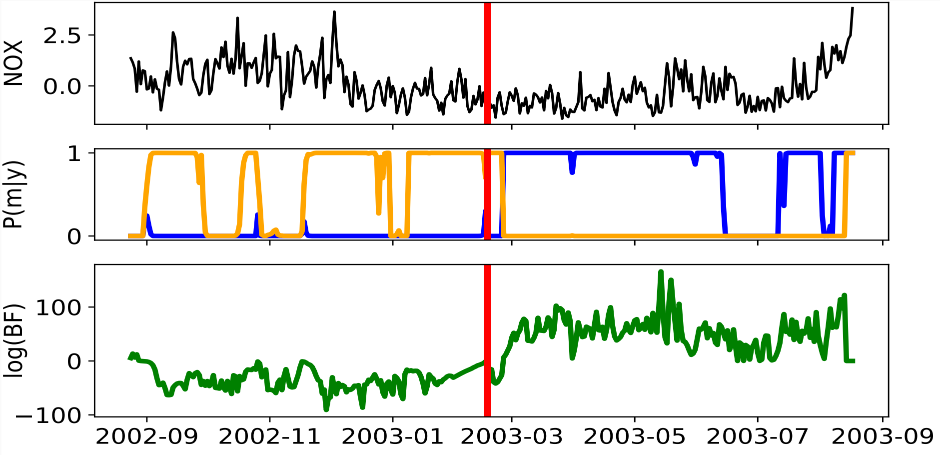
[MSE values for GP changepoint models as in Saatçi et al. (2010)]



✔ Fast computation: E.g., much faster than GP CP-models



New capability: Inference on shifting multivariate dynamics



Panel 1: NOX levels in London with **congestion charge introduction**

Panel 2: Model posteriors for the two VAR models

Panel 3: Corresponding log Bayes Factors

Summary: Novelty & Implications

Novel Features added to BOCPD:

- **Multivariate** modelling of dependencies between data streams
- Generalizing BOCPD to **model selection**

Practical Implications:

- Improved **CP detection**
- Improved **prediction**, especially in multivariate data
- New capability: Inference on **shifting** multivariate **dynamics**

More at Poster #149

Main References

Adams, R. P. and MacKay, D. J. (2007). Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*.

Fearnhead, P. and Liu, Z. (2007). On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):589–605.

Saatçi, Y., Turner, R. D., and Rasmussen, C. E. (2010). Gaussian process change point models. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 927–934.