Spatiotemporal Bayesian On-line Changepoint Detection with Model Selection

Jeremias Knoblauch, Department of Statistics
Theodoros Damoulas, Departments of Statistics & Computer Science
j.knoblauch@warwick.ac.uk
t.damoulas@warwick.ac.uk

Motivation

Standard Bayesian On-line Changepoint Detection [1] (BOCPD) segments non-stationary data into stationary segments described by a **single model**. Inspired by [2], we generalize the method to **multiple models**, yielding BOCPD with model selection (BOCPDMS) and adapt it to spatio-temporal data via Vector Autoregressions (VARs).

Probabilistic Model

$r_t|r_{t-1} \sim H(r_t, r_{t-1})$ — **Conditional run-length prior**: Current segment's length, so $r_t = k \iff$ CP at $t-k$.

$m_t|\{r_t = 0\} \sim q(m_t)$ — **Model prior** at time t, given that there is a CP at t. (If $r_t > 0$, $m_t = m_{t-1}$.)

$\boldsymbol{\theta}_{m_t} \sim \pi_{m_t}(\boldsymbol{\theta}_{m_t})$ — **Parameter prior** for the model at time t

$\boldsymbol{y}_t \sim f_{m_t}(\boldsymbol{y}_t|\boldsymbol{\theta}_{m_t})$ — **Observation density** at time t

Recursion

Efficient inference requires that the posterior predictives

$$f_{m_t}(\boldsymbol{y}_t|\boldsymbol{y}_{1:(t-1)}, r_{t-1}) = \int_{\Theta_{m_t}} f_{m_t}(\boldsymbol{y}_t|\boldsymbol{\theta}_{m_t})\pi_{m_t}(\boldsymbol{\theta}_{m_t}|\boldsymbol{y}_{(t-r_{t-1}):(t-1)})d\boldsymbol{\theta}_{m}$$

are available in closed form and proceeds recursively through

$$p(\boldsymbol{y}_1, r_1 = 0, m_1) = q(m_1)\int_{\Theta_{m_1}} f_{m_1}(\boldsymbol{y}_1|\boldsymbol{\theta}_{m_1})\pi_{m_1}(\boldsymbol{\theta}_{m_1})d\boldsymbol{\theta}_{m_1} = q(m_1)f_{m_1}(\boldsymbol{y}_1|\boldsymbol{y}_0)$$

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

with the model posterior at time $t-1$ serving as prior at t via

$$q(m_t|\boldsymbol{y}_{1:(t-1)}, r_{t-1}, m_{t-1}) = \begin{cases} q(m_t) & \text{if } r_{t-1} = 0, \\ q(m_{t-1}|\boldsymbol{y}_{(t-r_{t-1}-1):(t-1)}) & \text{if } r_{t-1} > 0. \end{cases}$$

Inference

Evidence: $p(\boldsymbol{y}_{1:t}) = \sum_{r_t, m_t} p(\boldsymbol{y}_{1:t}, r_t, m_t)$

run-length and model: $p(r_t, m_t|\boldsymbol{y}_{1:t}) = p(\boldsymbol{y}_{1:t}, r_t, m_t)/p(\boldsymbol{y}_{1:t})$

Prediction: $p(\boldsymbol{y}_{t+1}|\boldsymbol{y}_{1:t}) = \sum_{r_t, m_t} f_{m_t}(\boldsymbol{y}_{t+1}|\boldsymbol{y}_{1:t}, r_t)p(r_t, m_t|\boldsymbol{y}_{1:t})$

Run-length marginal posterior: $p(r_t|\boldsymbol{y}_{1:t}) = \sum_{m_t} p(r_t, m_t|\boldsymbol{y}_{1:t})$

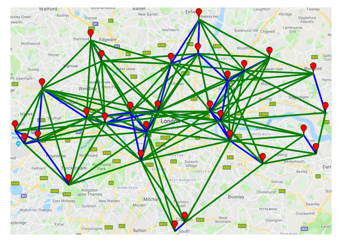
Model marginal posterior: $p(m_t|\boldsymbol{y}_{1:t}) = \sum_{r_t} p(r_t, m_t|\boldsymbol{y}_{1:t})$.

Segmentation: $\text{MAP}_t = \max_{r,t}\{\text{MAP}_{t-r-1} \cdot p(r_t = r, m_t = m|\boldsymbol{y}_{1:t})\}$

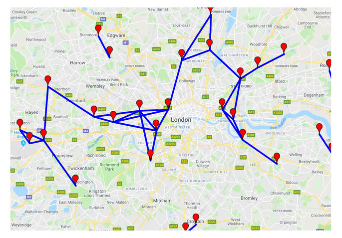
Spatially Structured Bayesian VAR (SSBVAR)

Ingredient 1: Standard Bayesian VAR model
Ingredient 2: Neighbourhoods/sparsity imposed via $\boldsymbol{A}_{1:L}$

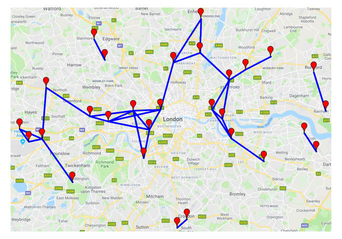
$$\sigma^2 \sim \text{IG}(a,b), \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{\Omega}), \quad \text{vec}([\boldsymbol{\alpha}, \boldsymbol{A}_{1:L}]) \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{V}_c),$$
$$\boldsymbol{Y}_t = \boldsymbol{\alpha} + \boldsymbol{A}_1\boldsymbol{Y}_{t-1} + \boldsymbol{A}_2\boldsymbol{Y}_{t-2} + \boldsymbol{A}_3\boldsymbol{Y}_{t-3} + \boldsymbol{\varepsilon}_t$$



$i \in N_1(j)$ if $d(i,j) < 20km$



$i \in N_2(j)$ if $d(i,j) < 10km$

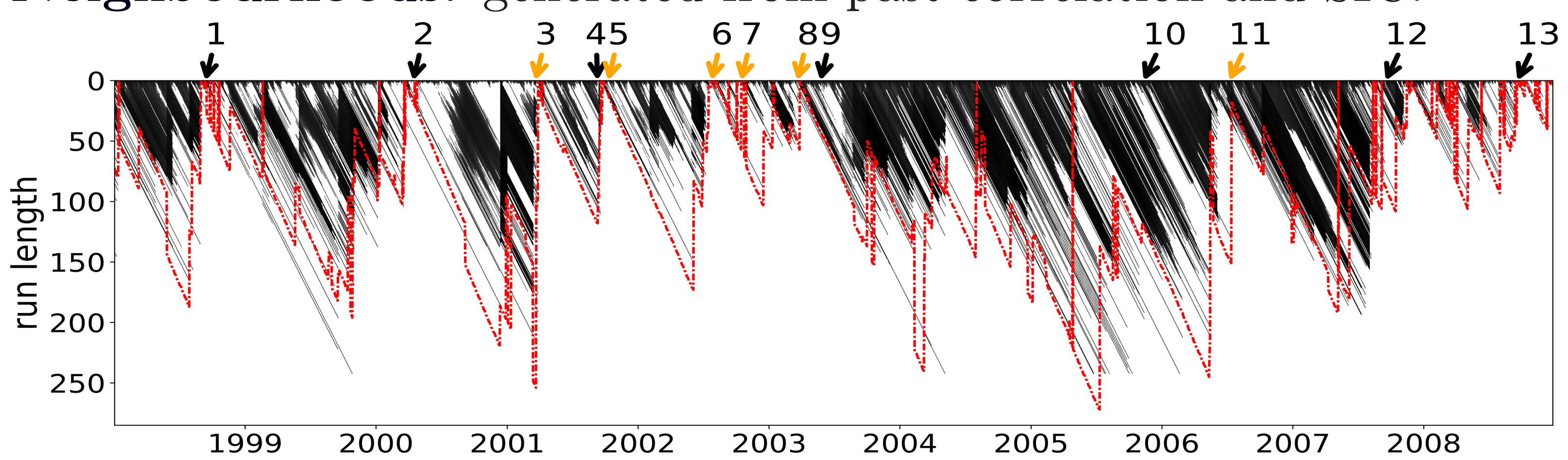


$i \in N_3(j)$ if $d(i,j) < 10km$

Spatial Structure: $[A_l]_{(i,j)} = 0$ if location i is not a neighbour of (=connected to) location j at the l-th lag, i.e. if $i \notin N_l(j)$.

Application I: Subprime Mortgage crisis

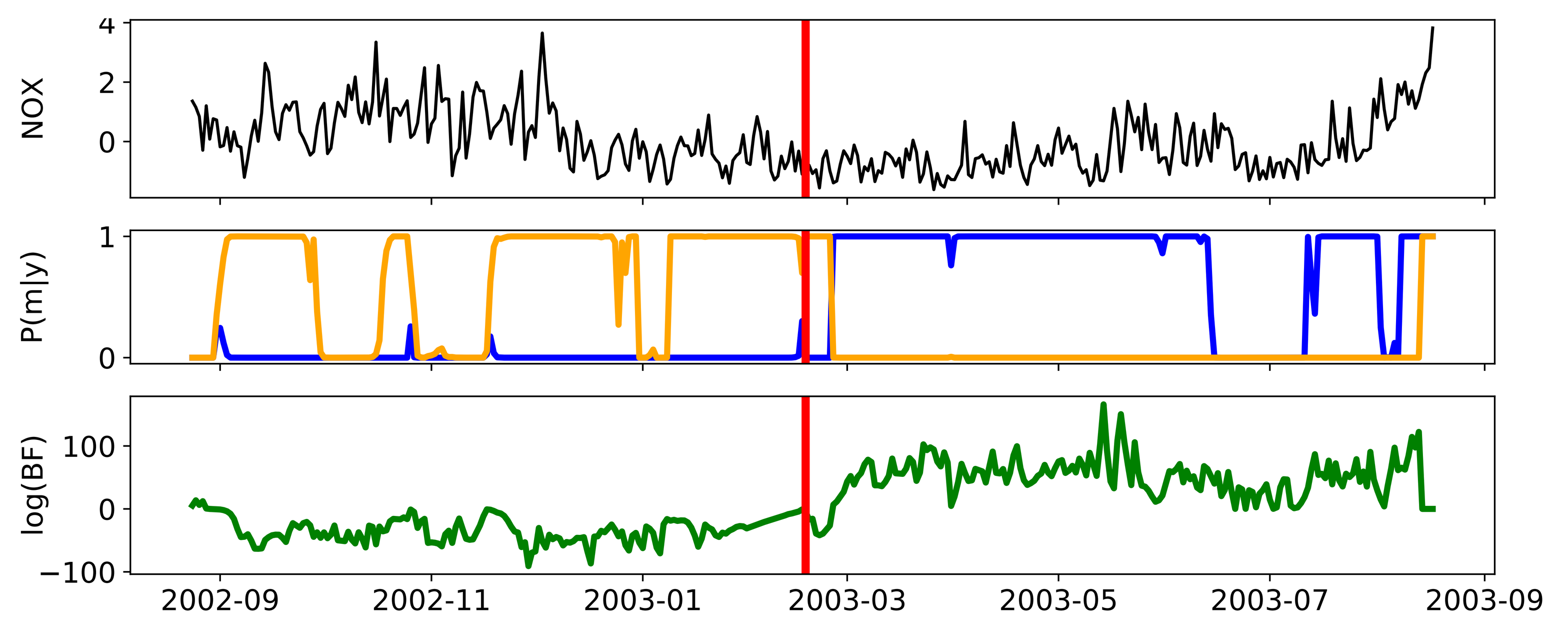
Data: Daily returns of 30 industry portfolios.
Neighbourhoods: generated from past correlation and SIC.



Legend: run-length posterior (grayscale) and its **maximum**. Events found in [4] indicated using **black**, additionally found ones using **orange** arrows.
Selected events: (1) Asia Crisis, (2) DotCom bubble, (3) OPEC cuts output by 4%, (4) 9/11, (5) Afghanistan war, (6) 2002 stock market crash, (7) Bombing in Bali, (8) Iraq war, (9) Major US tax cuts, (10) US election, (11) Iran successfully enriches Uranium, (12) Northern Rock bank run, (13) Lehman Brothers collapse.

Application II: London Congestion Charge

Data: daily mean NOX levels for 29 stations across London.
Neighbourhoods: Generated from Euclidean & road distances.



Legend: **congestion charge introduction**. **Panel 1:** NOX in Brent. **Panel 2:** Model posteriors for best-fitting models. **Panel 3:** Corresponding Log Bayes Factors.

Prediction compared to GP CP models

	Nile		Snowfall		Bee		30 Portfolio	
Method	MSE	NLL	MSE	NLL	MSE	NLL	MSE	NLL
SSBVAR	**.550** (.095)	**1.13** (.068)	.681 (.025)	.923 (.0231)	**1.74** (.222)	**3.57** (.166)	**25.93** (.906)	48.32 (.964)
*Best GP CP	.553 (.096)	1.15 (.056)	**.618** (.024)	**-1.98** (.056)	2.62 (.195)	4.07 (.150)	29.95 (.50)	— —

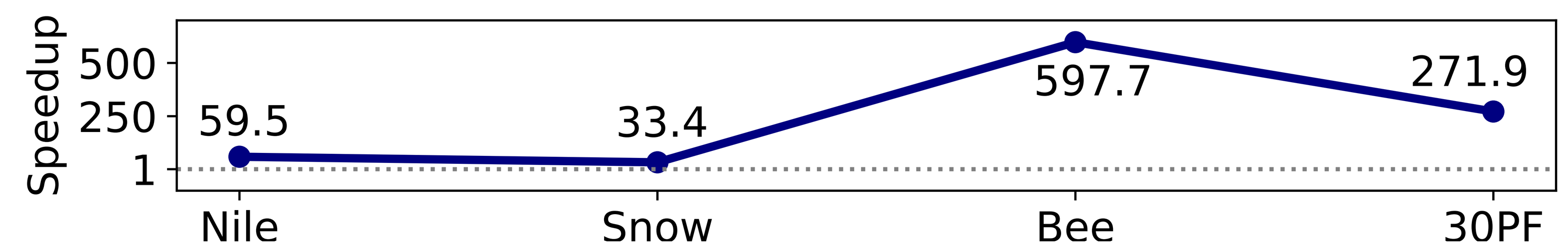
*Numbers taken from [4].

Complexity compared to GP CP models

Processing $y_{1:T} \in \mathcal{R}^{T \times d}$ using a VAR, tracking K most likely (r_t, m_t):

BOCPDMS: $\mathcal{O}(\sum_{t=1}^T K \cdot C_{p,d}) = \mathcal{O}(TKC_{p,d}) = \mathcal{O}(TK)$

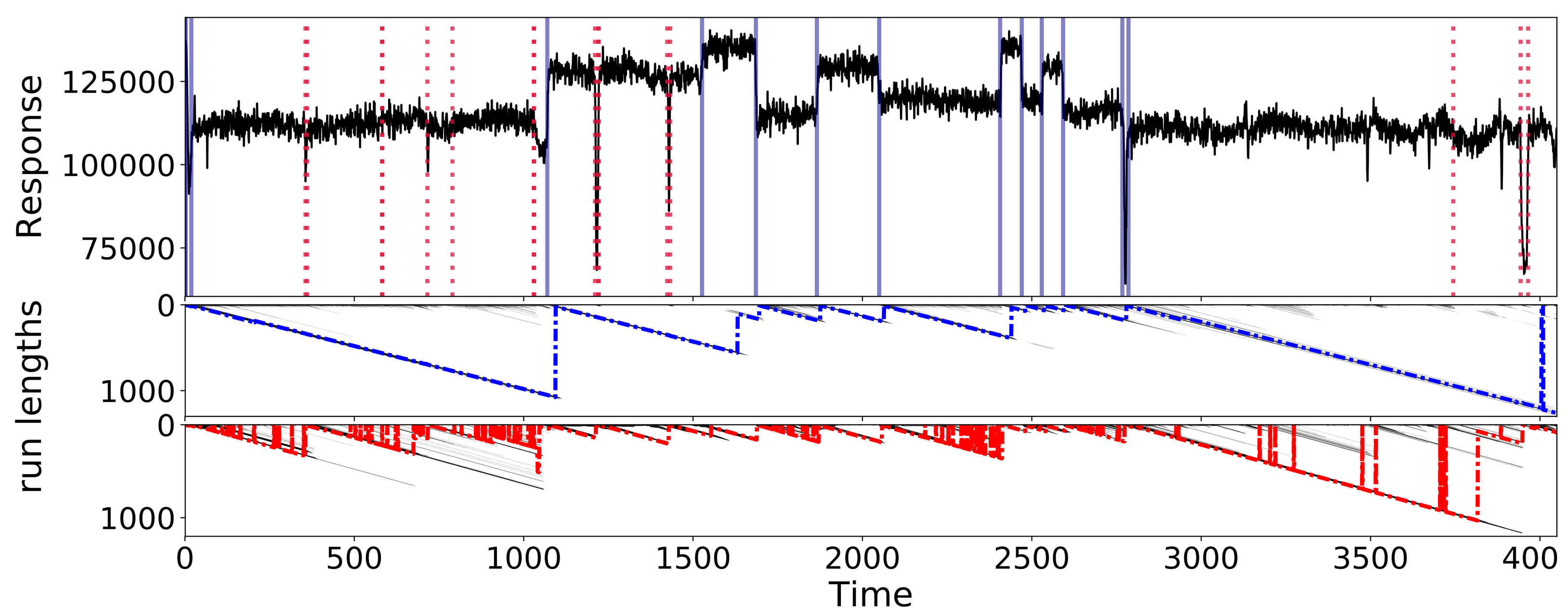
GP CP models [4]: $\mathcal{O}(TK^3)$



Conclusion & Future directions

This paper generalized BOCPD to model selection and inherently multivariate data. The resulting inference outperforms GP CP methods in terms of CP detection, prediction and computation time.

In a **follow-up paper** [3], we have made BOCPDMS robust to outliers and model misspecification. The picture below shows the analysis of the outlier-prone well-log data with **standard** vs. **robust** BOCPDMS.



References

[1] Ryan Prescott Adams and David JC. MacKay. Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*, 2007.
[2] Paul Fearnhead and Zhen Liu. On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society: Series B*, 69 (4), pp. 589–605, 2007.
[3] Jeremias Knoblauch, Jack Jewson and Theodoros Damoulas. Doubly Robust Bayesian Inference for Non-Stationary Streaming Data with β-Divergences. *arXiv:1806.02261*, 2018.
[4] Yunus Saatçi, Ryan D. Turner and Carl E. Rasmussen. Gaussian process change point models. *Proceedings of the 27th International Conference on Machine Learning*, pp. 927–934, 2010.