A Bayesian framework for change-point detection with uncertainty quantification

Davis Berlind[1] Lorenzo Cappello[2] Oscar Madrid Padilla[3]

August 15, 2025

[1] Department of Statistics & Data Science, University of California, Los Angeles
[2] Department of Economics and Business, Universitat Pompeu Fabra; Data Science Center, Barcelona School of Economics
[3] Department of Statistics & Data Science, University of California, Los Angeles

Table of Contents

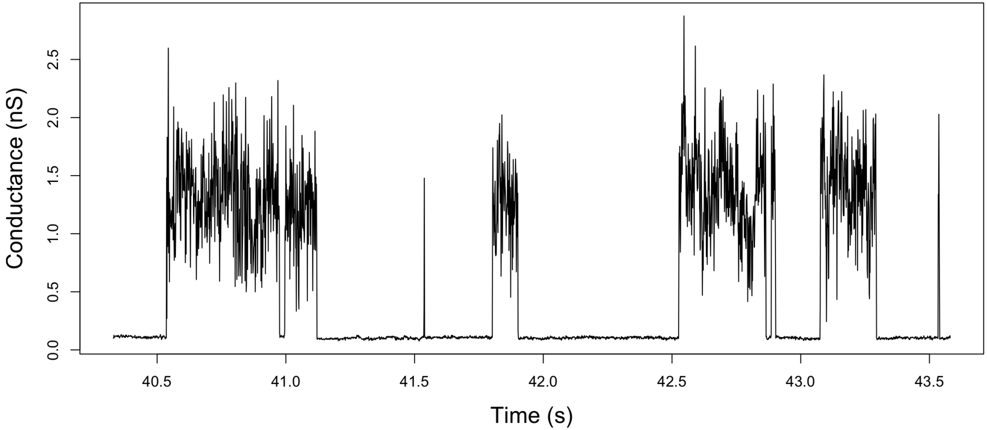
Change-Point Detection

- Change-Point Detection (CPD) is a classical problem in statistical inference (Page, 1954).
- Problem set-up:
  - ▷ T Observations: $\mathbf{y}_{1:T} := \{\mathbf{y}_t\}_{t=1}^{T}$ where $\mathbf{y}_t \in \mathbb{R}^d$
  - ▷ L Change-Points: $\boldsymbol{\tau}_{1:L} \subset \{1, \ldots, T\}$, with $\tau_0 := 1 < \tau_1 < \ldots < \tau_L < \tau_{L+1} := T + 1$, and collection of $L + 1$ distributions $\{F_\ell\}_{\ell=0}^{L}$ with $F_\ell \neq F_{\ell+1}$ such that:

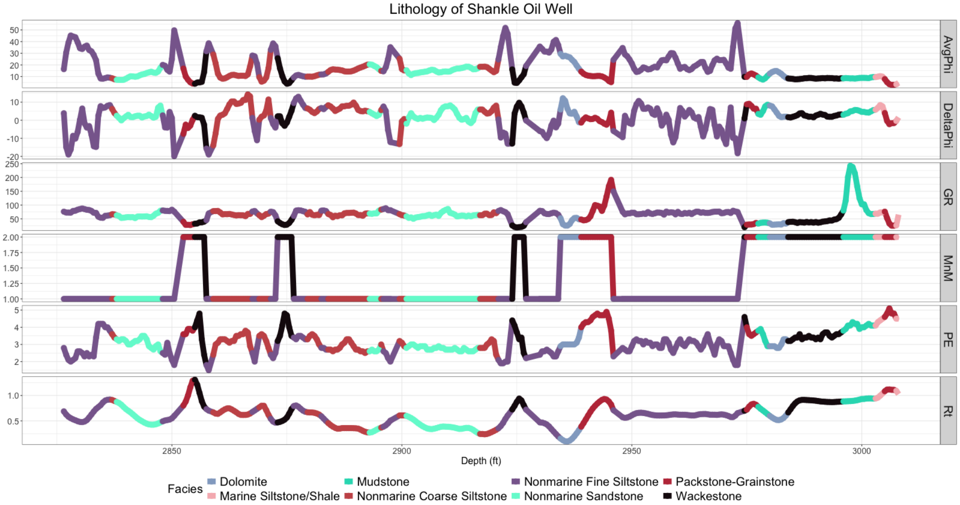
$$\mathbf{y}_t \sim F_\ell, \;\; \forall\, t \in [\tau_\ell, \tau_{\ell+1}).$$

  - ▷ Goal: consistently estimate and perform inference on $\{L, \boldsymbol{\tau}_{1:L}\}$.
- Mean and variance change-points:
  - ▷ Univariate: changes in piece-wise constant mean $\boldsymbol{\mu}_{1:T} := \{\mathbb{E}[y_t]\}_{t=1}^{T}$ and precision $\boldsymbol{\lambda}_{1:T} := \{\mathrm{Var}(y_t)^{-1}\}_{t=1}^{T}$ signals.
  - ▷ Multivariate: changes in piece-wise constant mean signal $\boldsymbol{\mu}_{1:T}$.

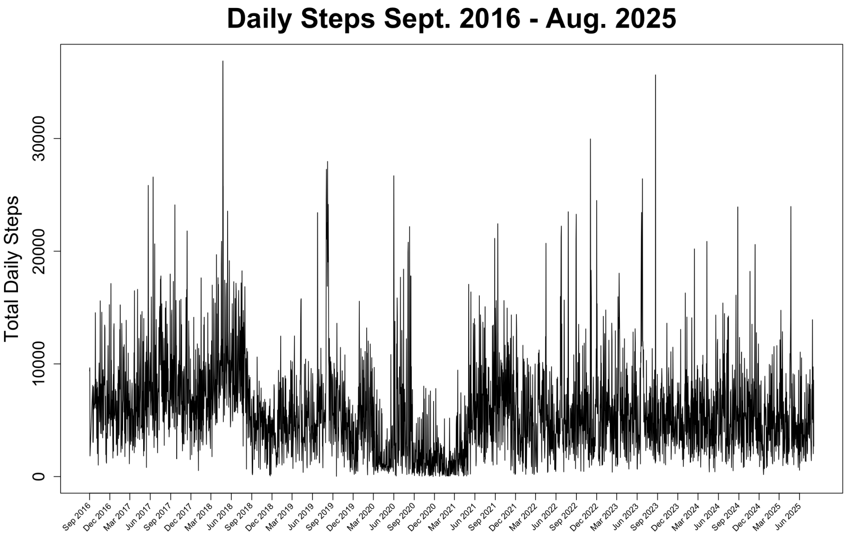
Ion Channel (Hotz et al., 2013)



Oil Well Lithology (Bohling and Dubois, 2003)



Daily Step Count



Uncertainty Quantification

- We would like to quantify the uncertainty around estimates $\hat{\tau}_{1:\hat{L}}$.
- Early attempts limited to a single mean change (Siegmund, 1986; Worsley, 1986; Jirak, 2015; Horváth et al., 2017), required knowledge of L (Bai and Perron, 2003), or only produced approximate sets from some limiting distribution (Bai, 2010).
- SMUCE (Frick et al., 2014) advanced the state-of-the-art, but returns CIs that can be overly conservative with undesirable coverage properties as α decreases (Fryzlewicz, 2024).
- Methods for multivariate data and variance changes remain underdeveloped.

Bayesian CPD

- Issues with existing Bayesian CPD methods:
  - Do not scale beyond small T.
  - Generally lack theoretical guarantees for $\hat{\tau}_{1:\hat{L}}$.
  - Posterior distributions can be difficult to interpret.
- Proposal:
  - Introduce Bayesian single change-point (SCP) models with optimal localization properties.
  - Modularly combine SCP models and approximate posterior distribution using variational Bayes.

Table of Contents

Single Change-Point Model

- Change-point $\tau \in \{1, \ldots, T\}$ with $\mathbb{P}(\tau = t) = \pi_t$
- Posterior: $\mathbb{P}(\tau = t \mid \mathbf{y}_{1:T}) := \overline{\pi}_t \propto \pi_t p(\mathbf{y}_{1:T} \mid \tau = t)$.
- MAP Estimator: $\hat{\tau}_{\text{MAP}} := \underset{1 \leq t \leq T}{\arg\max}\, \overline{\pi}_t$.
- α-Level Credible Sets:

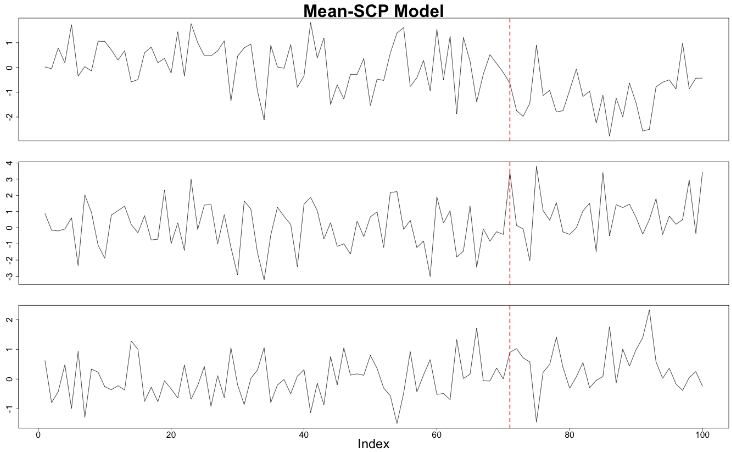
$$\mathcal{CS}(\alpha, \overline{\boldsymbol{\pi}}_{1:T}) := \underset{S \subseteq [T]}{\arg\min} |S| \quad \text{s.t.} \sum_{t \in S} \overline{\pi}_t \geq 1 - \alpha.$$

Single Change-Point Models

- Three Bayesian models for a single change-point in $\mathbf{y}_{1:T}$:
  - ▷ Change in mean ($d \geq 1$).
  - ▷ Change in variance ($d = 1$).
  - ▷ Change in mean and variance ($d = 1$).

Multivariate Mean Single Change-Point (Mean-SCP) Model

$$
\begin{aligned}
\mathbf{y}_t \mid \boldsymbol{\mu}_t, \boldsymbol{\Lambda}_t &\overset{\text{ind.}}{\sim} \mathcal{N}_d(\boldsymbol{\mu}_t, \boldsymbol{\Lambda}_t^{-1}) \\
\boldsymbol{\mu}_t &= \mathbf{b}\mathbb{1}\{t \geq \tau\} \\
\mathbf{b} &\sim \mathcal{N}_d(\mathbf{0}, \omega_0^{-1}\mathbf{I}_d) \\
\tau &\sim \text{Categorical}(\boldsymbol{\pi}_{1:T}) \\
\mathbf{b} &\perp\!\!\!\perp \tau
\end{aligned}
$$



Mean-SCP Posterior

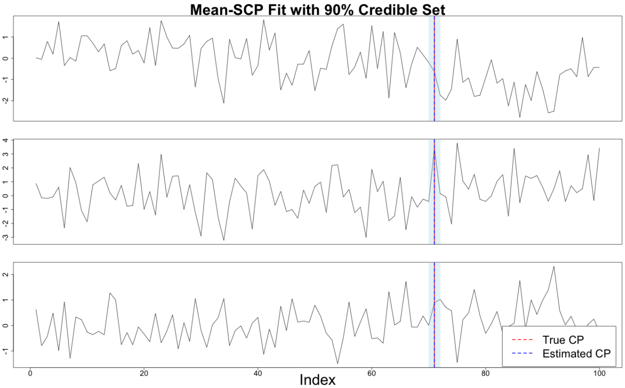
$$\mathbf{b} \mid \tau = t,\ \mathbf{y}_{1:T} \sim \mathcal{N}_d\left(\overline{\mathbf{b}}_t, \overline{\boldsymbol{\Omega}}_t^{-1}\right)$$

$$\tau \mid \mathbf{y}_{1:T} \sim \text{Categorical}(\overline{\boldsymbol{\pi}}_{1:T})$$

$$\overline{\boldsymbol{\Omega}}_t = \omega_0 \mathbf{I}_d + \sum_{t'=t}^{T} \boldsymbol{\Lambda}_{t'}$$

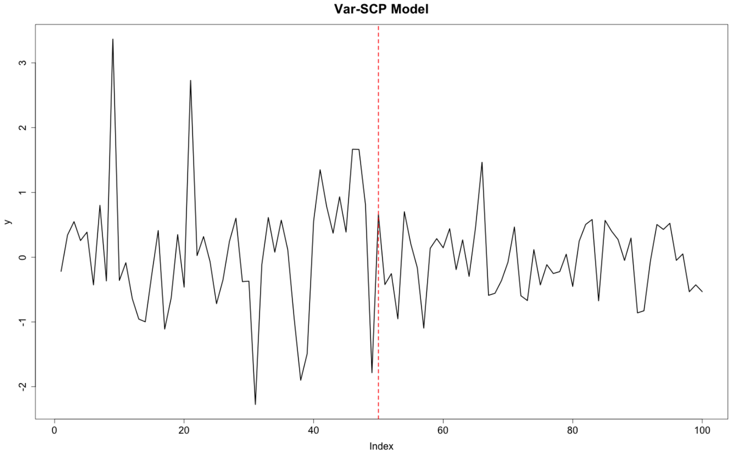
$$\overline{\mathbf{b}}_t = \overline{\boldsymbol{\Omega}}_t^{-1} \sum_{t'=t}^{T} \boldsymbol{\Lambda}_{t'} \mathbf{y}_{t'}$$

$$\overline{\pi}_t \propto \pi_t |\overline{\boldsymbol{\Omega}}_t|^{-\frac{1}{2}} \exp\left[\frac{\|\overline{\boldsymbol{\Omega}}_t^{\frac{1}{2}} \overline{\mathbf{b}}_t\|_2^2}{2}\right]$$



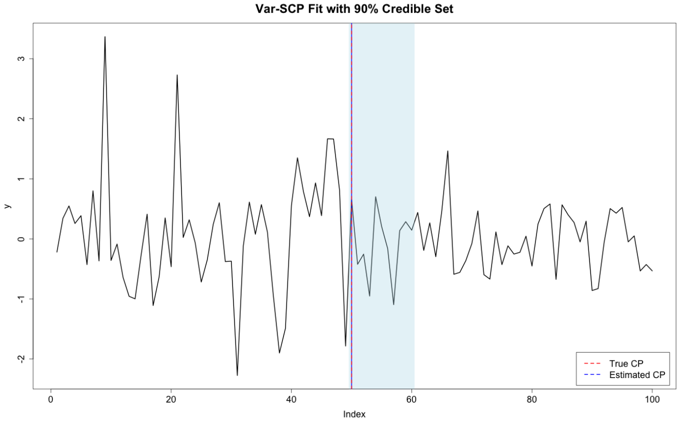
Variance Single Change-Point (Var-SCP) Model

$$y_t \mid \lambda_t \overset{\text{ind.}}{\sim} \mathcal{N}(0, \lambda_t^{-1})$$
$$\lambda_t = \omega_t s^{\mathbb{1}\{t \geq \tau\}}$$
$$s \sim \text{Gamma}(u_0, v_0)$$
$$\tau \sim \text{Categorical}(\boldsymbol{\pi}_{1:T})$$
$$s \perp\!\!\!\perp \tau$$



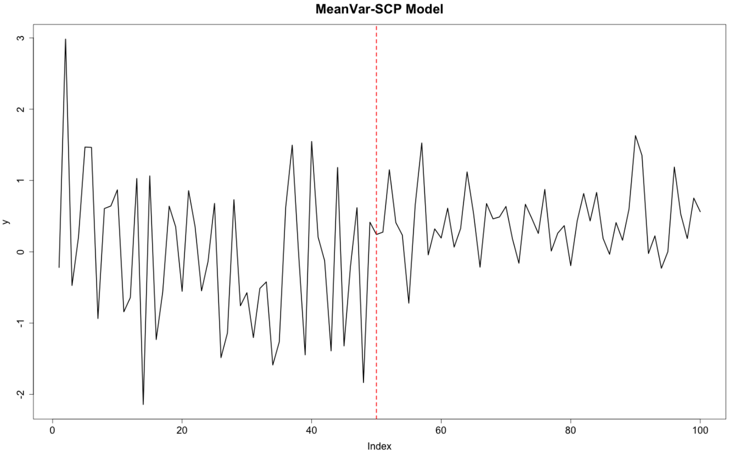
Var-SCP Posterior

$$
\begin{aligned}
s \mid \tau = t,\ \mathbf{y}_{1:T} &\sim \text{Gamma}\left(\overline{u}_t, \overline{v}_t\right) \\
\tau \mid \mathbf{y}_{1:T} &\sim \text{Categorical}(\overline{\boldsymbol{\pi}}_{1:T}) \\
\overline{u}_t &= u_0 + \frac{T - t + 1}{2} \\
\overline{v}_t &= v_0 + \frac{1}{2} \sum_{t'=t}^{T} \omega_{t'} y_{t'}^2 \\
\overline{\pi}_t &\propto \frac{\pi_t \Gamma(\overline{u}_t)}{\overline{v}_t^{\overline{u}_t}} \exp\left(-\frac{1}{2} \sum_{t'=1}^{t-1} \omega_{t'} y_{t'}^2 \right)
\end{aligned}
$$



Mean-Variance Single Change-Point (MeanVar-SCP) Model

$$
\begin{aligned}
y_t \mid \mu_t, \lambda_t &\overset{\text{ind.}}{\sim} \mathcal{N}(\mu_t, \lambda_t^{-1}) \\
\mu_t &= b\mathbb{1}\{t \geq \tau\} \\
\lambda_t &= \omega_t s^{\mathbb{1}\{t \geq \tau\}} \\
b \mid s &\sim \text{Normal}(0, (\omega_0 s)^{-1}) \\
s &\sim \text{Gamma}(u_0, v_0) \\
\tau &\sim \text{Categorical}(\boldsymbol{\pi}_{1:T}) \\
\{b, s\} &\perp\!\!\!\perp \tau
\end{aligned}
$$



MeanVar-SCP Posterior

$$b \mid s, \tau = t, \mathbf{y}_{1:T} \sim \mathcal{N}(\overline{b}_t, (\overline{\omega}_t s)^{-1})$$

$$s \mid \tau = t, \mathbf{y}_{1:T} \sim \text{Gamma}(\overline{u}_t, \overline{v}_t)$$

$$\tau \mid \mathbf{y}_{1:T} \sim \text{Categorical}(\overline{\boldsymbol{\pi}}_{1:T})$$

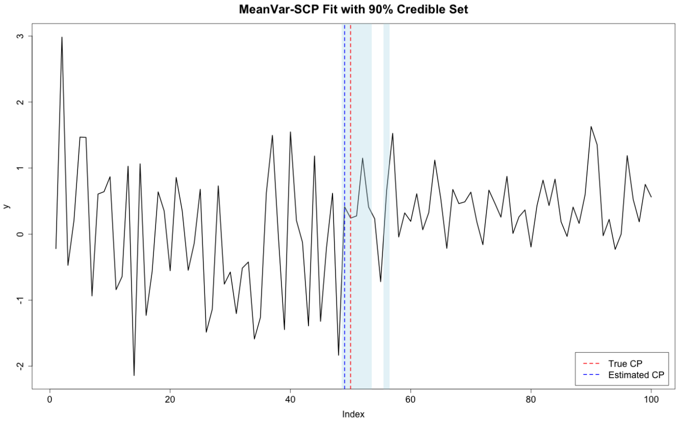
$$\overline{\omega}_t = \omega_0 + \sum_{t'=t}^{T} \omega_{t'}$$

$$\overline{b}_t = \sum_{t'=t}^{T} \frac{\omega_{t'} y_{t'}}{\overline{\omega}_t}$$

$$\overline{u}_t = u_0 + \frac{T - t + 1}{2}$$

$$\overline{v}_t = v_0 - \frac{\overline{\omega}_t \overline{b}_t^2}{2} + \frac{1}{2} \sum_{t'=t}^{T} \omega_{t'} y_{t'}^2$$

$$\overline{\pi}_t \propto \frac{\pi_t \Gamma(\overline{u}_t)}{\overline{v}_t^{\overline{u}_t} \overline{\omega}_t^{1/2}} \exp\left(-\frac{1}{2} \sum_{t'=1}^{t-1} \omega_{t'} y_{t'}^2\right)$$



Localization Theory

- True change-point: $t_0 \in \{1, \dots, T\}$.
- Minimum spacing condition: $\Delta_T := \min\{t_0, T - t_0 + 1\} \gtrsim \log T$.
- Consistency: $\lim_{T\to\infty} \mathbb{P}\left(|\hat{\tau}_{\text{MAP}} - t_0| \le \epsilon_T\right) = 1$ and $\lim_{T\to\infty} \frac{\epsilon_T}{\Delta_T} = 0$. (Yu, 2020)

Detectable Mean and Scale Change

Assumption 1 (Detectable Mean Change)

Suppose $\mathbb{E}[\mathbf{y}_t] = \mathbf{b}_0 \mathbb{1}_{\{t \geq t_0\}}$ *for some* $t_0 \in [T]$ *and* $\mathbf{b}_0 \in \mathbb{R}^d$ *and* $\mathrm{Var}(\mathbf{y}_t) = \mathbf{\Lambda}_t^{-1}$. *Assume that* $\Delta_T \gtrsim \log T$ *and* $\Delta_T \min_{1 \leq t \leq T} \|\mathbf{\Lambda}_t^{1/2} \mathbf{b}_0\|_2^2 \gg d \log T$.

Assumption 2 (Detectable Scale Change)

Suppose $\mathrm{Var}(y_t) = (s_0^2)^{\mathbb{1}\{t \geq t_0\}}$ *for some* $t_0 \in [T]$ *and* $0 < \underline{s} < s_0 < \overline{s} < \infty$. *Assume that* $\Delta_T \gtrsim \log T$ *and* $\Delta_T (s_0^2 - 1)^2 \gg \log T$.

- Necessary: consistent localization not possible when $\Delta_T \|\mathbf{b}_0\|_2^2 \lesssim \log T$ (Wang et al., 2020)
- Non-Sparse: suppose $\|\mathbf{b}_0\|_\infty = \mathcal{O}(1)$ and $\Delta_T \geq \log^{1+\varepsilon} T$. Assumption 1 not met if $\|\mathbf{b}_0\|_0 \leq d_0 \lesssim d \log^{-\varepsilon} T$:

$$\Delta_T \|\mathbf{b}_0\|_2^2 \lesssim d_0 \log^{1+\varepsilon} T \lesssim d \log T.$$

Similar assumptions appear in Bai (2010); **?**); Li et al. (2023).

SCP Localization Rates

Theorem 1

Let $\mathbf{y}_{1:T}$ be a sequence of independent, sub-Gaussian observations with $\|\mathbf{y}_t\|_{\psi_2} = \mathcal{O}(1)$ and assume that $\max_{t\in[T]} |\log \pi_t| \leq C_\pi \log T$ for some C_π. For each SCP model, the following table summarizes the minimum spacing Δ_T and signal strength $\kappa(b_0, s_0^2)$ conditions under which $\lim_{T\to\infty} \mathbb{P}(|\hat{\tau}_{MAP} - t_0| \leq \epsilon_T) = 1$, where $\epsilon_T = \mathcal{O}\left(\frac{\log T}{\kappa(b_0, s_0^2)}\right)$:

Model	*Assumptions*	$\kappa(b_0, s_0^2)$
Mean-SCP	*Assumption 1, Var*$(\mathbf{y}_t) = \mathbf{\Lambda}^{-1}$	$\|\mathbf{\Lambda}^{1/2}\mathbf{b}_0\|_2^2$
Var-SCP	*Assumption 2,* $\mathbb{E}[y_t] = 0$	$(s_0^2 - 1)^2$
MeanVar-SCP	*Assumption 1 or 2*	$\max\{\min\{b_0^2, b_0^2/s_0^2\}, (s_0^2 - 1)^2\}$

We also show that when $\mathbf{y}_{1:T}$ is an α-mixing process, then under mild regularity conditions $\mathbb{P}(|\hat{\tau}_{MAP} - t_0| \leq \tilde{\epsilon}_T) = 1$ where $\tilde{\epsilon}_T \propto \epsilon_T \log T$.

Results of Wang and Samworth (2017), Wang et al. (2020), and Wang et al. (2021) show that the minimax optimal localization rate is proportional to $[\Delta_T \kappa(b_0, s_0^2)]^{-1}$.

Detection Rule

Corollary 2

Let ϵ_T be the localization error corresponding to one of SCP models, then for any $\alpha > 0$, $\lim_{T\to\infty} \mathbb{P}\left(|\mathcal{CS}(\alpha, \overline{\boldsymbol{\pi}}_{1:T})| \leq 2\epsilon_T\right) = 1$.

- Detect change-point if $|\mathcal{CS}(\alpha, \overline{\boldsymbol{\pi}}_{1:T})| \leq \log^{1+\delta} T$ for some small $\delta > 0$.

Table of Contents

Multiple Independent CHange-point (MICH) Model

We can modularly combine SCP models to incorporate multiple change-points in $\boldsymbol{\mu}_{1:T}$ and/or $\boldsymbol{\lambda}_{1:T}$:

$$
\begin{aligned}
y_t \mid \mu_t, \lambda_t &\overset{\text{ind.}}{\sim} \mathcal{N}(\mu_t, \lambda_t^{-1}), & 1 \leq t \leq T, \\
\mu_t &:= \mu_0 + \sum_{i=1}^{J+L} \mu_{it} := \sum_{j=1}^{J} b_j \mathbb{1}_{\{t \geq \tau_j\}} + \sum_{\ell=J+1}^{J+L} b_\ell \mathbb{1}_{\{t \geq \tau_\ell\}}, & \\
\lambda_t &:= \lambda_0 \prod_{i=1}^{J+K} \lambda_{it} := \prod_{j=1}^{J} s_j^{\mathbb{1}\{t \geq \tau_j\}} \prod_{k=J+L+1}^{J+L+K} s_k^{\mathbb{1}\{t \geq \tau_k\}}, & \\
\tau_i &\overset{\text{ind.}}{\sim} \text{Categorical}(\boldsymbol{\pi}_{i,1:T}), & 1 \leq i \leq J + L + K, \\
\{b_j, s_j\} &\overset{\text{ind.}}{\sim} \text{Normal-Gamma}(0, \omega_0, u_0, v_0), & 1 \leq j \leq J, \\
b_\ell &\overset{\text{ind.}}{\sim} \mathcal{N}(0, \omega_0^{-1}), & J < \ell \leq J + L, \\
s_k &\overset{\text{ind.}}{\sim} \text{Gamma}(u_0, v_0), & J + L < k \leq J + L + K.
\end{aligned}
$$

Variational Bayes Approximation to MICH

- Could fit MICH with Gibbs sampler, but the discrete, highly correlated, high-dimensional parameters lead to poor mixing.
- Following the example set in Wang et al. (2020), we use Mean-Field Variational Bayes to find a $q \in \mathcal{Q}_{\text{MF}}$ that approximates true posterior of MICH:

$$\mathcal{Q}_{\text{MF}} := \left\{ q : q = \prod_{j=1}^{J} q_j(b_j, s_j, \tau_j) \prod_{\ell=J+1}^{J+L} q_\ell(b_\ell, \tau_\ell) \prod_{k=J+L+1}^{J+L+K} q_k(s_k, \tau_k) \right\}.$$

- Finding $q \in \mathcal{Q}_{\text{MF}}$ that minimizes the KL divergence with the true posterior equivalent to maximizing ELBO:

$$\begin{aligned} \boldsymbol{\Theta} &:= \{\{b_j, s_j, \tau_j\}_{j=1}^{J}, \{b_\ell, \tau_\ell\}_{\ell=J+1}^{J+L}, \{s_k, \tau_k\}_{k=J+L+1}^{J+L+K}\} \\ \text{ELBO}(q) &:= \int q(\boldsymbol{\Theta}) \log \frac{p(\mathbf{y}_{1:T}, \boldsymbol{\Theta})}{q(\boldsymbol{\Theta})}\, d\boldsymbol{\Theta} \\ &= \log p(\mathbf{y}_{1:T}) - \text{KL}(q \,\|\, p). \end{aligned}$$

Fitting MICH with VB

- Computationally efficient backfitting procedure to find q:

Algorithm 1 MICH Variational Approximation

Initialize Posterior Parameters.
repeat
 For $\ell \in \{1,\ldots,L\}$: Subtract out ℓ^{th} mean component from $\boldsymbol{\mu}_{1:T}$ and update q_ℓ by fitting Mean-SCP model to partial residual.
 For $k \in \{1,\ldots,K\}$: Divide out k^{th} scale component from $\boldsymbol{\lambda}_{1:T}$ and update q_k by fitting fit Var-SCP model to partial residual.
 For $j \in \{1,\ldots,J\}$ Partial out j^{th} mean and scale component from $\boldsymbol{\mu}_{1:T}$ and $\boldsymbol{\lambda}_{1:T}$ and update q_j by fitting MeanVar-SCP model to partial residuals.
until Convergence

- Algorithm 1 is equivalent to maximizing the ELBO via coordinate ascent, guaranteeing convergence. Each outer loop of Algorithm 1 is $\mathcal{O}(T(J+L+K))$.
- Can use value of ELBO to automatically select J, L, and K (MICH-Auto).

Multivariate Simulation Study

Generate 5,000 replicates of following simulation with $T = 250$, $\Delta_T = 10$, and $C = \sqrt{10}$, $d \in \{10, 50, 100\}$, $L^* \in \{5, 10, 20\}$, and $p \in \{0.1, 0.5, 1\}$:

i. Draw $\boldsymbol{\tau}_{1:L^*}$ uniformly from $[T]$ subject to the minimum spacing condition $|\tau_{\ell+1} - \tau_\ell| \geq \Delta_T$ with $\tau_0 = 1$ and $\tau_{L^*+1} = T + 1$.

ii. Draw $\{U_i\}_{i=1}^d \sim \text{Uniform}(-2, 2)$ and set $s_i := 2^{U_i}$.

iii. Let A be a set of $d_0 := \lfloor pd \rfloor$ active coordinates drawn uniformly at random from $[d]$.

iv. Set $\boldsymbol{\mu}_0 := 0$, and for each $i \in [d]$ draw $\xi_{\ell,i} \sim \text{Bernoulli}(0.5)$ and set:

$$\mu_{\ell,i} := \mu_{\ell-1,i} + \frac{C(1 - 2\xi_{\ell,i})s_i \mathbb{1}_{\{i \in A\}}}{\sqrt{\min\{\tau_{\ell+1} - \tau_\ell, \tau_\ell - \tau_{\ell-1}\}}}.$$

v. Draw $\mathbf{y}_t \overset{\text{ind.}}{\sim} \mathcal{N}_d\left(\sum_{j=0}^{L^*} \boldsymbol{\mu}_\ell \mathbb{1}_{\{\tau_\ell \leq t < \tau_{\ell+1}\}}, \text{diag}(\mathbf{s}_{1:d})\right)$.

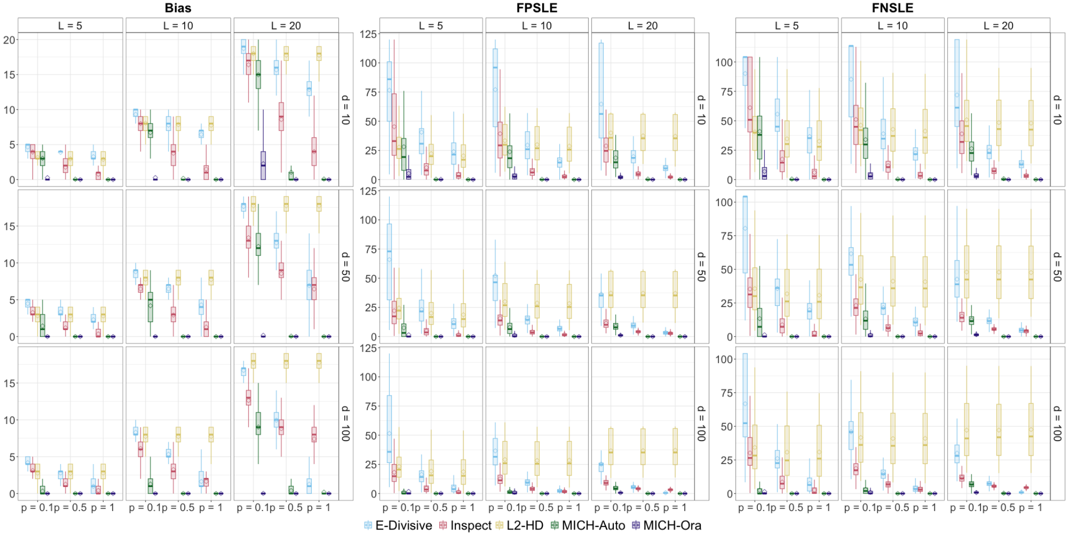
Multivariate Simulation Study

- Calculate bias $|L^* - L|$ and measure accuracy of $\hat{\boldsymbol{\tau}}_{1:\hat{L}}$ with FPSLE and FNSLE statistics:

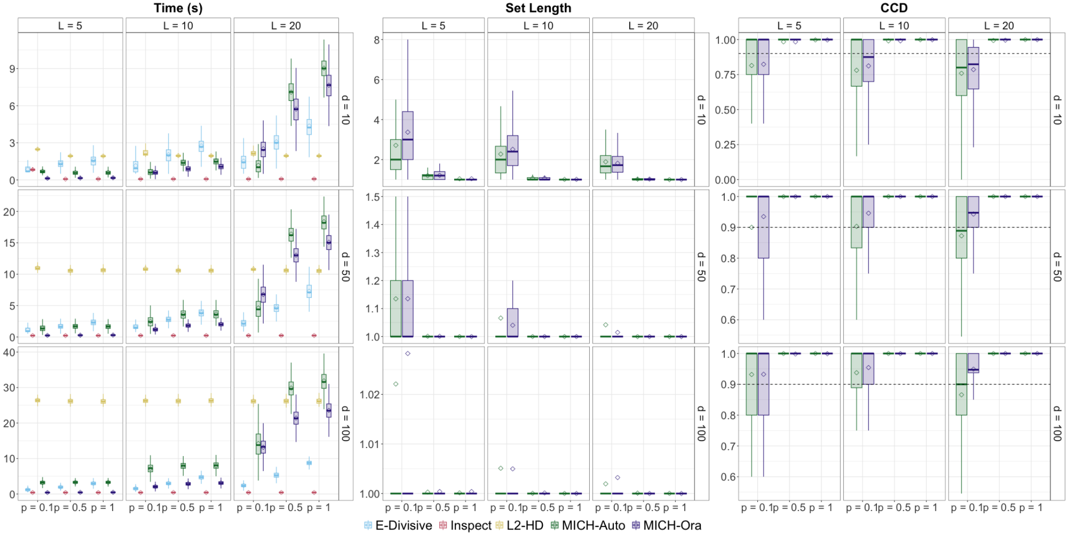
$$
\begin{aligned}
d_{\text{FPSLE}}(\hat{\boldsymbol{\tau}}_{1:\hat{L}} \| \boldsymbol{\tau}_{1:L}) &:= \frac{1}{2(\hat{L}+1)} \sum_{\ell=1}^{\hat{L}+1} |\hat{\tau}_{\ell-1} - \tau_{i_\ell - 1}| + |\hat{\tau}_\ell - \tau_{i_\ell}|, \\
\{i_\ell\}_{\ell=1}^{L+1} &:= \{i \in [L+1] : \tau_{i_\ell - 1} < (\hat{\tau}_{\ell-1} + \hat{\tau}_\ell)/2 \leq \tau_{i_\ell} \; \forall \, \ell \in [L+1]\} \\
d_{\text{FNSLE}}(\hat{\boldsymbol{\tau}}_{1:\hat{L}} \| \boldsymbol{\tau}_{1:L}) &:= d_{\text{FPSLE}}(\boldsymbol{\tau}_{1:L} \| \hat{\boldsymbol{\tau}}_{1:L})
\end{aligned}
$$

- Fit MICH with L set to true value (Ora-MICH) and selected from the ELBO (Auto-MICH) and return 90% credible sets.
- Compare to the E-Divisive method of James and Matteson (2015), the Two-Way MOSUM (ℓ^2-HD) method of Li et al. (2023), and the informative sparse projection (Inspect) method of Wang and Samworth (2017).

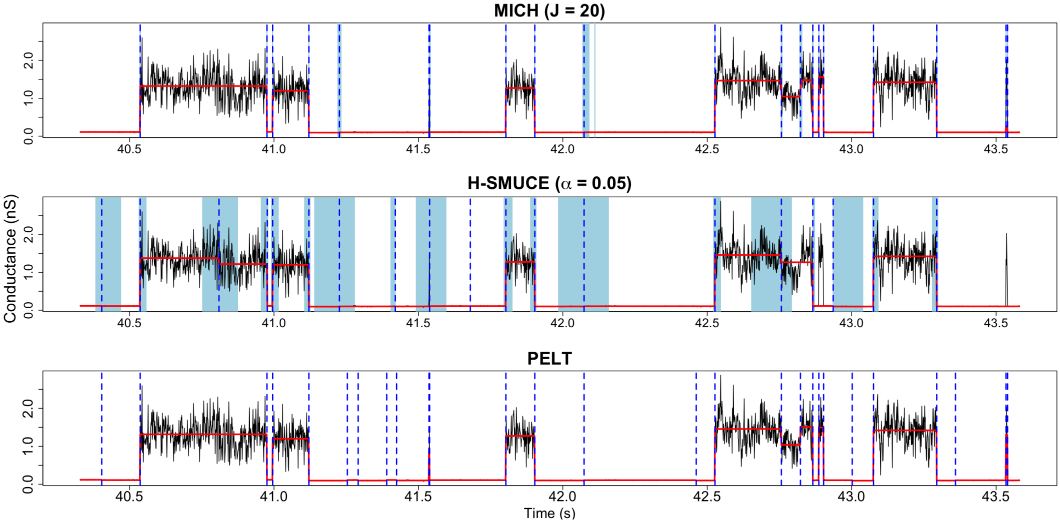
Multivariate Simulation Results



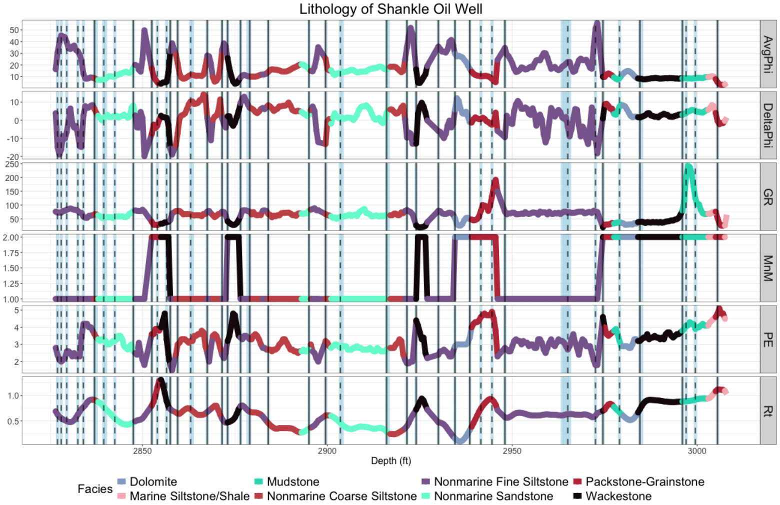
Multivariate Simulation Results



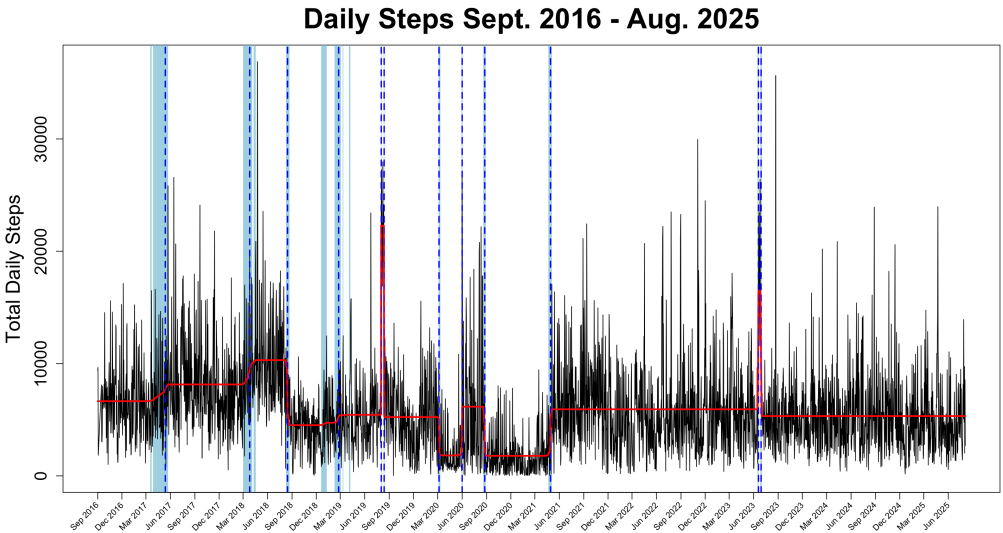
MICH fit of Ion Channel (Hotz et al., 2013)



MICH Fit of Oil Well (Bohling and Dubois, 2003)



MICH Fit of Daily Steps



Paper & Code

arXiv:

R Package:

References I

Bai, J. (2010). Common breaks in means and variances for panel data. *Journal of Econometrics 157*(1), 78–92. Nonlinear and Nonparametric Methods in Econometrics.

Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*(1), 1–22.

Bohling, G. and M. Dubois (2003). An integrated application of neural network and markov chain techniques to the prediction of lithofacies from well logs: Kansas geological survey open-file report 2003-50, 6 p. *Group 6*.

Frick, K., A. Munk, and H. Sieling (2014). Multiscale change point inference. *Journal of the Royal Statistical Society Series B: Statistical Methodology 76*(3), 495–580.

Fryzlewicz, P. (2024). Narrowest significance pursuit: Inference for multiple change-points in linear models. *Journal of the American Statistical Association 119*(546), 1633–1646.

Horváth, L., M. Hušková, G. Rice, and J. Wang (2017, April). Asymptotic Properties Of The Cusum Estimator For The Time Of Change In Linear Panel Data Models. *Econometric Theory 33*(2), 366–412.

Hotz, T., O. M. Schütte, H. Sieling, T. Polupanow, U. Diederichsen, C. Steinem, and A. Munk (2013). Idealizing ion channel recordings by a jump segmentation multiresolution filter. *IEEE Transactions on NanoBioscience 12*(4), 376–386.

James, N. A. and D. S. Matteson (2015). ecp: An r package for nonparametric multiple change point analysis of multivariate data. *Journal of Statistical Software 62*(7), 1–25.

Jirak, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics 43*(6), 2451 – 2483.

Killick, R., P. Fearnhead, and I. A. Eckley (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association 107*(500), 1590–1598.

Li, J., L. Chen, W. Wang, and W. B. Wu (2023). ℓ^2 inference for change points in high-dimensional time series via a two-way mosum.

References II

Page, E. S. (1954). Continuous inspection schemes. *Biometrika 41*(1/2), 100–115.

Pein, F., H. Sieling, and A. Munk (2017). Heterogeneous change point inference. *Journal of the Royal Statistical Society Series B: Statistical Methodology 79*(4), 1207–1227.

Siegmund, D. (1986). Boundary crossing probabilities and statistical applications. *The Annals of Statistics*, 361–404.

Wang, D., Y. Yu, and A. Rinaldo (2020). Univariate mean change point detection: Penalization, CUSUM and optimality. *Electronic Journal of Statistics 14*(1), 1917 – 1961.

Wang, D., Y. Yu, and A. Rinaldo (2021). Optimal covariance change point localization in high dimensions. *Bernoulli 27*(1), 554 – 575.

Wang, G., A. Sarkar, P. Carbonetto, and M. Stephens (2020, 07). A Simple New Approach to Variable Selection in Regression, with Application to Genetic Fine Mapping. *Journal of the Royal Statistical Society Series B: Statistical Methodology 82*(5), 1273–1300.

Wang, T. and R. J. Samworth (2017, 08). High Dimensional Change Point Estimation via Sparse Projection. *Journal of the Royal Statistical Society Series B: Statistical Methodology 80*(1), 57–83.

Worsley, K. J. (1986). Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika 73*(1), 91–104.

Yu, Y. (2020). A review on minimax rates in change point detection and localisation.

Mean-Variance Simulation Study

- Recreate simulation study for mean and variance jumps introduced in Pein et al. (2017). 5,000 replicates for $T \in \{100, 500, 1000\}$ and $J^* \in \{2, 5, 10\}$.
- Calculate bias $|J^* - J|$ and measure accuracy of $\hat{\boldsymbol{\tau}}_{1:\hat{J}}$ with FPSLE and FNSLE statistics:

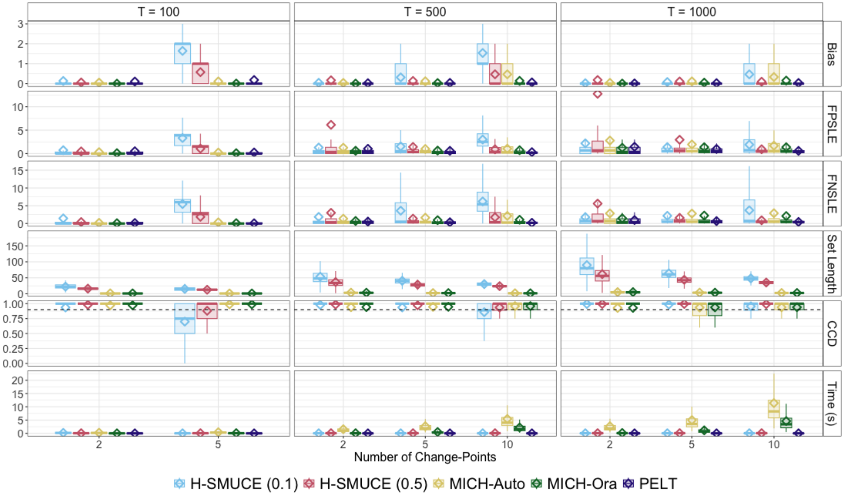
$$d_{\text{FPSLE}}(\hat{\boldsymbol{\tau}}_{1:\hat{J}} \| \boldsymbol{\tau}_{1:J}) := \frac{1}{2(\hat{J}+1)} \sum_{j=1}^{\hat{J}+1} |\hat{\tau}_{j-1} - \tau_{i_j - 1}| + |\hat{\tau}_j - \tau_{i_j}|,$$

$$\{i_j\}_{j=1}^{J+1} := \{i \in [J+1] : \tau_{i_j - 1} < (\hat{\tau}_{j-1} + \hat{\tau}_j)/2 \leq \tau_{i_j} \ \forall\, j \in [J+1]\}$$

$$d_{\text{FNSLE}}(\hat{\boldsymbol{\tau}}_{1:\hat{J}} \| \boldsymbol{\tau}_{1:J}) := d_{\text{FPSLE}}(\boldsymbol{\tau}_{1:J} \| \hat{\boldsymbol{\tau}}_{1:J})$$

- Fit MICH with J set to true value (Ora-MICH) and selected from the ELBO (Auto-MICH) and return 90% credible sets.
- Compare to H-SMUCE (Pein et al., 2017) with $\alpha \in \{0.1, 0.5\}$ and PELT (Killick et al., 2012).

Mean-Variance Simulation Results



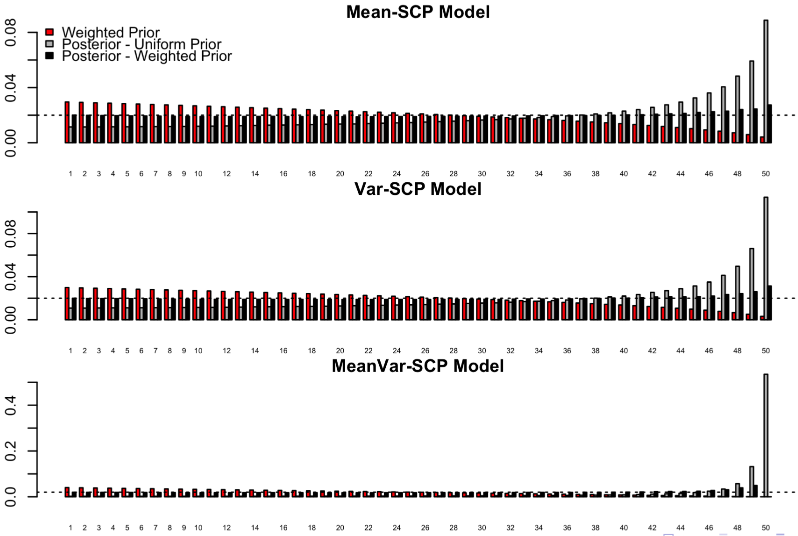
Choice of π_t

- Localization results valid when $\pi_t = T^{-1}$ for each t.
- Uniform prior may reduce power and result in false negatives in small samples.
- Choosing $\boldsymbol{\pi}_{1:T}$ so that:

$$\mathbb{E}\left[\log \overline{\pi}_t - \log \overline{\pi}_{t+1}\right] = 0$$

leads to closed form recursions.

$\mathbb{E}[\boldsymbol{\pi}_{1:T} \mid \mathbf{y}_{1:T}]$ under Null Model



VB Details

- Finding best $q \in \mathcal{Q}_{\text{MF}}$ is equivalent to simple back-fitting procedure.
- Given initial guess of q, define residual mean, precision, and variance correction terms:

$$\tilde{r}_t := y_t - \sum_{j=1}^{J} \frac{\mathbb{E}_{q_j}[\lambda_{jt}\mu_{jt}]}{\mathbb{E}_{q_j}[\lambda_{jt}]} - \sum_{\ell=J+1}^{J+L} \mathbb{E}_{q_\ell}[\mu_{\ell t}]$$

$$\overline{\lambda}_t := \prod_{j=1}^{J} \mathbb{E}_{q_j}[\lambda_{jt}] \prod_{k=J+L+1}^{N} \mathbb{E}_{q_k}[\lambda_{kt}]$$

$$\delta_t := \sum_{j=1}^{J} \left(\frac{\mathbb{E}_{q_j}[\lambda_{jt}\mu_{jt}^2]}{\mathbb{E}_{q_j}[\lambda_{jt}]} - \frac{\mathbb{E}_{q_j}[\lambda_{jt}\mu_{jt}]^2}{\mathbb{E}_{q_j}[\lambda_{jt}]^2} \right) + \sum_{\ell=J+1}^{J+L} \text{Var}_{q_\ell}(\mu_{\ell t})$$

VB Residuals

- Iteratively partial out components and fit single change-point model (modulo correction term δ_t):
  - Mean-SCP to $\tilde{r}_{-\ell t}$ with precision parameters $\overline{\lambda}_t$

$$\tilde{r}_{-\ell t} := \tilde{r}_t + \mathbb{E}_{q_\ell}[\mu_{\ell t}]$$

  - Var-SCP to $\tilde{r}_t$ with precision parameters $\overline{\lambda}_{-kt}$

$$\overline{\lambda}_{-kt} := \mathbb{E}_{q_k}[\lambda_{kt}]^{-1}\overline{\lambda}_t$$

  - MeanVar-SCP to $\tilde{r}_{-jt}$ with scale parameters $\overline{\lambda}_{-jt}$

$$\tilde{r}_{-jt} := \tilde{r}_t + \frac{\mathbb{E}_{q_j}[\lambda_{jt}\mu_{jt}]}{\mathbb{E}_{q_j}[\lambda_{jt}]}$$

$$\overline{\lambda}_{-jt} := \mathbb{E}_{q_j}[\lambda_{jt}]^{-1}\overline{\lambda}_t$$

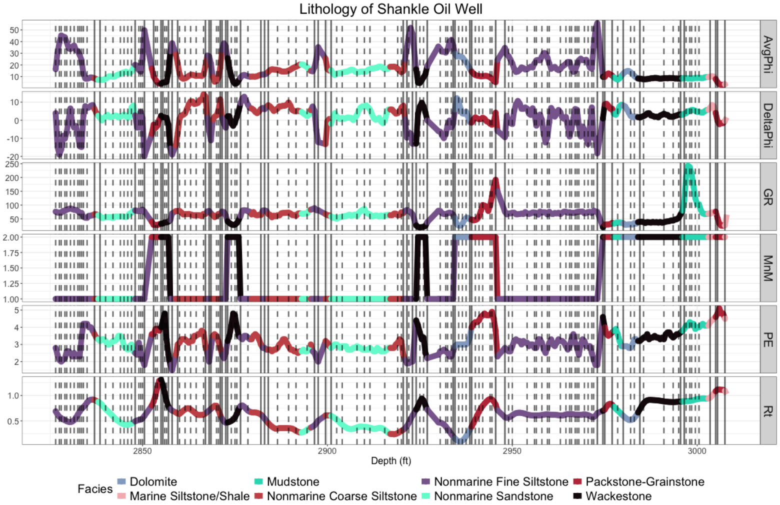
Simulation Details

- Fixing the number of observations T, the number of change-points J, the minimum spacing condition Δ_T, and a constant $C > 0$.
- Drawing $\boldsymbol{\tau}_{1:J^*}$ uniformly from $[T]$ subject to the minimum spacing condition $|\tau_{j+1} - \tau_j| \geq \Delta_T$ with $\tau_0 = 1$ and $\tau_{J+1} = T + 1$.
- Picking standard deviations such that $s_0 := 1$ and $s_j := 2^{U_j}$ where $\{U_j\}_{j=1}^J \overset{\text{i.i.d.}}{\sim} \text{Uniform}(-2, 2)$.
- Letting $\mu_0 := 0$, drawing J^* independent Rademacher variables ξ_j, and setting:

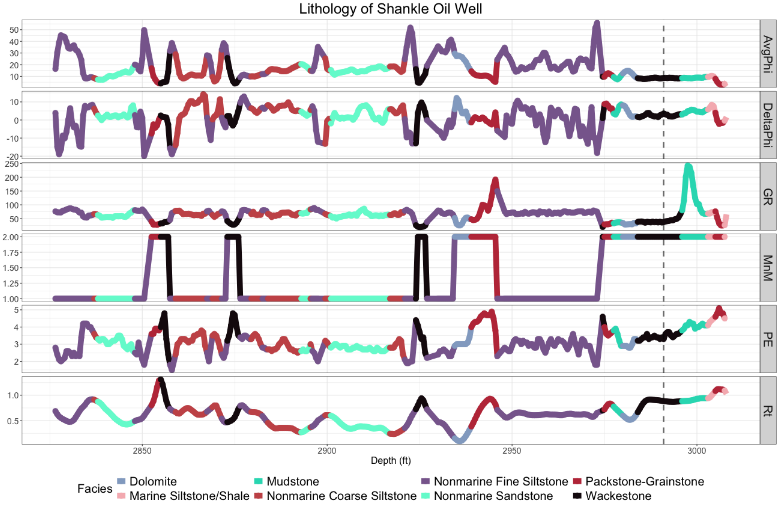
$$\mu_j := \mu_{j-1} + \xi_j C \left(\min\{s_j^{-1}\sqrt{\tau_{j+1} - \tau_j}, s_{j-1}^{-1}\sqrt{\tau_j - \tau_{j-1}}\} \right)^{-1}.$$

- Drawing $y_t \overset{\text{ind.}}{\sim} \mathcal{N}\left(\sum_{j=0}^J \mu_j \mathbb{1}_{\{\tau_j \leq t < \tau_{j+1}\}}, \sum_{j=0}^J \sigma_j \mathbb{1}_{\{\tau_j \leq t < \tau_{j+1}\}}\right)$.

InspectChangepoint Fit of Oil Well (Bohling and Dubois, 2003)



L2hdchange Fit of Oil Well (Bohling and Dubois, 2003)



α-Mixing

Assumption 3

Given the stochastic process $\{y_t\}_{t\geq 1}$, *assume that for any* $t_0 \in \mathbb{N}$, *and some distributions* F_0 *and* F_1, *there are stochastic processes* $\{y_{0,t}\}_{t\geq 1}$ *and* $\{y_{1,t}\}_{t\geq 1}$ *such that* $y_{0,t} \sim F_0$, *and* $y_{1,t} \sim F_1$, *and* $y_t := y_{0,t}\mathbb{1}_{\{t<t_0\}} + y_{1,t}\mathbb{1}_{\{t\geq t_0\}}$. *Additionally, assume that:*

(i) $\{y_{0,t}\}_{t\geq 1}$ *and* $\{y_{1,t}\}_{t\geq 1}$ *are* α*-mixing processes with respective coefficients* $\{\alpha_{0,k}\}_{k\geq 1}$ *and* $\{\alpha_{1,k}\}_{k\geq 1}$ *that satisfy* $\max\{\alpha_{0,k}, \alpha_{1,k}\} \leq e^{-Ck}$ *for some* $C > 0$.

(ii) *There exist constants* $\delta_1, D_1 > 0$ *such that* $\sup_{t\geq 1} \max\{\mathbb{E}\left[|y_{0,t}|^{4+\delta_1}\right], \mathbb{E}\left[|y_{1,t}|^{4+\delta_1}\right]\} \leq D_1$.