A Bayesian Framework for Change-Point Detection with Uncertainty Quantification

UCLA

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

[1]Department of Statistics & Data Science, UCLA, [2]Department of Economics and Business, Universitat Pompeu Fabra

ISBA BNP 14

Introduction

Existing change-point detection (CPD) methods generally do not provide uncertainty quantification for the point estimates they return. Bayesian CPD methods can provide uncertainty quantification, but generally lag behind the state-of-the-art in performance and computational cost. We address this gap in the literature by introducing a novel Bayesian method that can efficiently detect multiple structural breaks in the mean and variance of a length T time-series and return α-level credible sets around the estimated locations of the changes. In simulations, our method is competitive with other state-of-the-art CPD methods and returns credible sets that are an order of magnitude smaller than the confidence intervals returned by competitors.

Single Change-Point Models

- **Generic Single Change-Point (SCP) Model**: We introduce a latent change-point variable $\tau \in [T]$ to index a change in the mean $\boldsymbol{\mu}_{1:T}$ and/or precision $\boldsymbol{\Lambda}_{1:T}$ signals:

$$\mathbf{y}_t \mid \boldsymbol{\mu}_t, \boldsymbol{\Lambda}_t \overset{\text{ind.}}{\sim} \mathcal{N}_d\left(\boldsymbol{\mu}_t, \boldsymbol{\Lambda}_t^{-1}\right), \quad \forall\, t \in [T],$$
$$\tau \sim \text{Categorical}(\boldsymbol{\pi}_{1:T}).$$

- **SCP Models**:

	Mean-SCP	Var-SCP	MeanVar-SCP
Model	$\boldsymbol{\mu}_t = \mathbf{b}\mathbb{1}_{\{t\geq\tau\}}$, $\mathbf{b} \sim \mathcal{N}_d(\mathbf{0}, \omega_0^{-1}\mathbf{I}_d)$.	$\lambda_t = s^{\mathbb{1}\{t\geq\tau\}}$, $s \sim \text{Gamma}(u_0, v_0)$.	$\mu_t = b\mathbb{1}_{\{t\geq\tau\}}$, $\lambda_t = s^{\mathbb{1}\{t\geq\tau\}}$, $b \mid s \sim \mathcal{N}(0, (\omega_0 s)^{-1})$, $s \sim \text{Gamma}(u_0, v_0)$.
Multivariate	✓	✗	✗

- **Posterior Quantities**: The conjugate priors in each SCP model ensure closed forms for the marginal likelihood $p(\mathbf{y}_{1:T} \mid \tau)$ and lead to efficient computation of posterior quantities:
  - **Posterior distribution of τ**:
  $$\tau \mid \mathbf{y}_{1:T} \sim \text{Categorical}(\overline{\boldsymbol{\pi}}_{1:T}),$$
  $$\overline{\pi}_t \propto \pi_t p(\mathbf{y}_{1:T} \mid \tau = t).$$
  - **Point-Estimate**: $\hat{\tau}_{\text{MAP}} := \arg\max_{1\leq t\leq T} \overline{\pi}_t$.
  - **α-Level Credible Set**: $\mathcal{CS}(\alpha, \overline{\boldsymbol{\pi}}_{1:T}) := \arg\min_{S\subseteq[T]} |S|$ s.t. $\sum_{t\in S}\overline{\pi}_t \geq 1-\alpha$.

Localization Theory for Single Change-Point

Given true location of change $t_0 \in [T]$, we aim to characterize the smallest *localization error* $\{\epsilon_T\}_{T\geq 1}$ and a minimal set of conditions on $\Delta_T = \min\{t_0, T - t_0 + 1\}$ and the size of $\mathbb{E}[\mathbf{y}_\tau - \mathbf{y}_{\tau-1}]$ and $\text{Var}(\mathbf{y}_\tau)/\text{Var}(\mathbf{y}_{\tau-1})$ so that:

$$\lim_{T\to\infty} \mathbb{P}\left(|t_0 - \hat{\tau}_{\text{MAP}}| \leq \epsilon_T\right) = 1 \text{ and } \lim_{T\to\infty} \frac{\epsilon_T}{\Delta_T} = 0. \tag{1}$$

- **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 $\text{Var}(\mathbf{y}_t) = \boldsymbol{\Lambda}_t^{-1}$. Assume that $\Delta_T \gtrsim \log T$ and $\Delta_T \min_{1\leq t\leq T}\|\boldsymbol{\Lambda}_t^{1/2}\mathbf{b}_0\|_2^2 \gg d\log T$.
- **Assumption 2 (Detectable Variance Change)**: Suppose $\text{Var}(y_t) = (s_0^2)^{\mathbb{1}\{t\geq t_0\}}$ for some $t_0 \in [T]$ and $0 < \underline{s} < s_0 < \overline{s} < \infty$. Assume that $\Delta_T \gtrsim \log T$ and $\Delta_T(s_0^2 - 1)^2 \gg \log T$.

Theorem 1 (SCP Localization Rates)

Let $\mathbf{y}_{1:T}$ be a sequence of independent, sub-Gaussian observations with $\|\mathbf{y}_t\|_{\psi_2} = O(1)$. For each SCP model, the following table summarizes conditions on Δ_T and the signal strength $\kappa(b_0, s_0^2)$ for which $\hat{\tau}_{MAP}$ satisfies (1) with $\epsilon_T = O\left(\frac{\log T}{\kappa(b_0, s_0^2)}\right)$:

Model	*Assumptions*	$\kappa(b_0, s_0^2)$
Mean-SCP	*Assumption 1*, $\text{Var}(\mathbf{y}_t) = \boldsymbol{\Lambda}^{-1}$	$\|\boldsymbol{\Lambda}^{1/2}\mathbf{b}_0\|_2^2$
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, the localization errors above will still hold if we replace $\log T$ with an approximately $\log^2 T$ factor.

Multiple Independent CHange-Point (MICH) Model

We can modularly combine SCP models to allow for 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 Bayesian Inference

- Following the example set in Wang et al. (2020), we use Mean-Field Variational Bayes to find a $q \in Q_{\text{MF}}$ that approximates true posterior of MICH:

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

- 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 $O(T(J+L+K))$.
- Can use value of ELBO to automatically select J, L, and K (MICH-Auto).

Oil Well Lithology

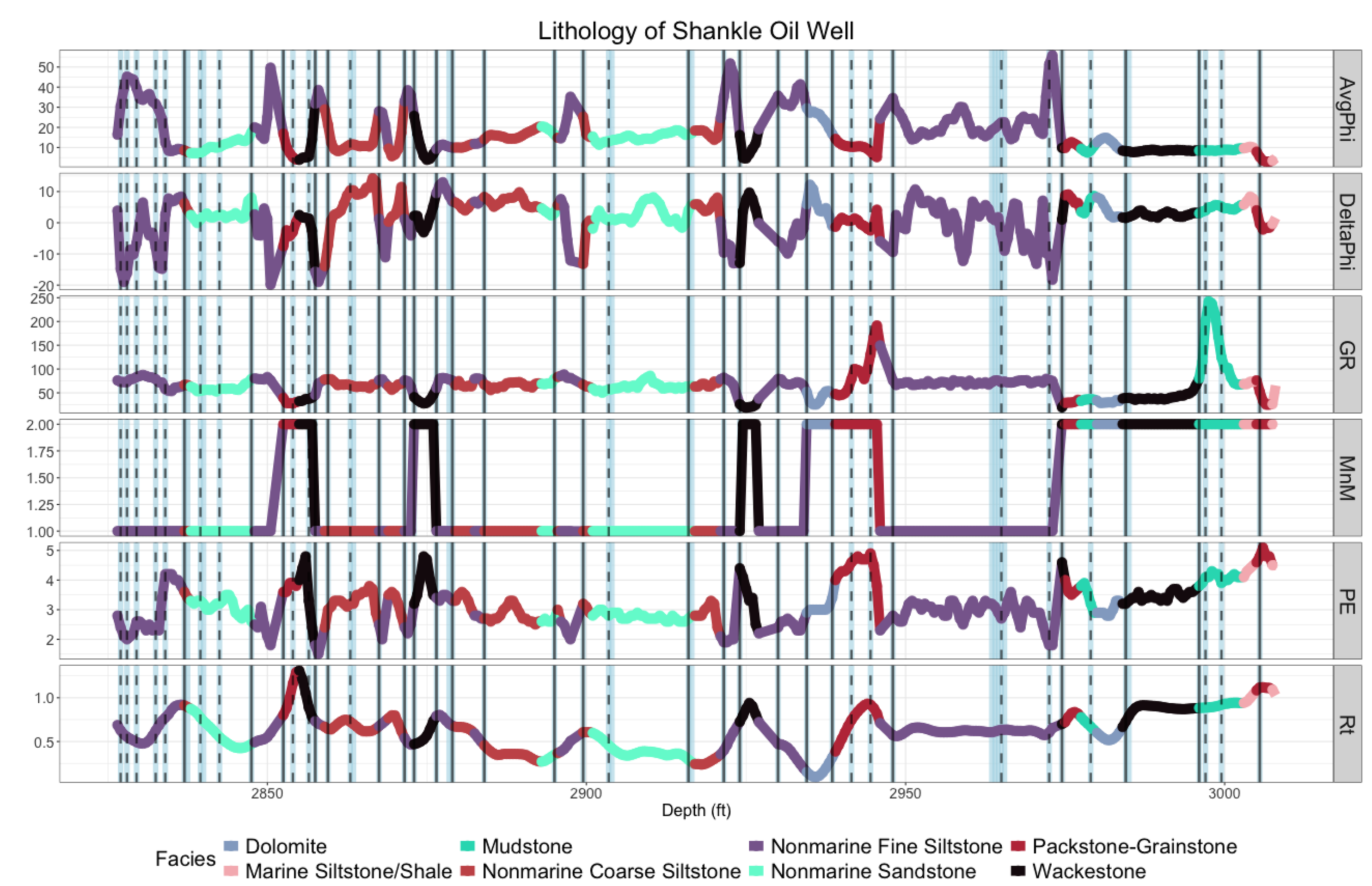

Figure 1: Recordings from six petrophysical measures from the Shankle oil well in Southwest Kansas. $\hat{L} = 42$ estimated changes with 99% credible sets shaded. 26 of the estimated changes or their 99% credible sets are within one index of a true change and 24 of the true changes are covered by a credible set.

Simulation Results

We compare the performance of MICH to state-of-the-art methods E-Divisive (James and Matteson, 2015) ℓ^2-HD (Li et al., 2023), and Inspect (Wang and Samworth, 2017) in a simulation study where a multivariate sequence $\mathbf{y}_{1:T}$ exhibits L mean changes in a fraction p of its d coordinates. MICH clearly beats the other methods in terms of bias and localization error, even in the sparse setting $p = 0.1$.

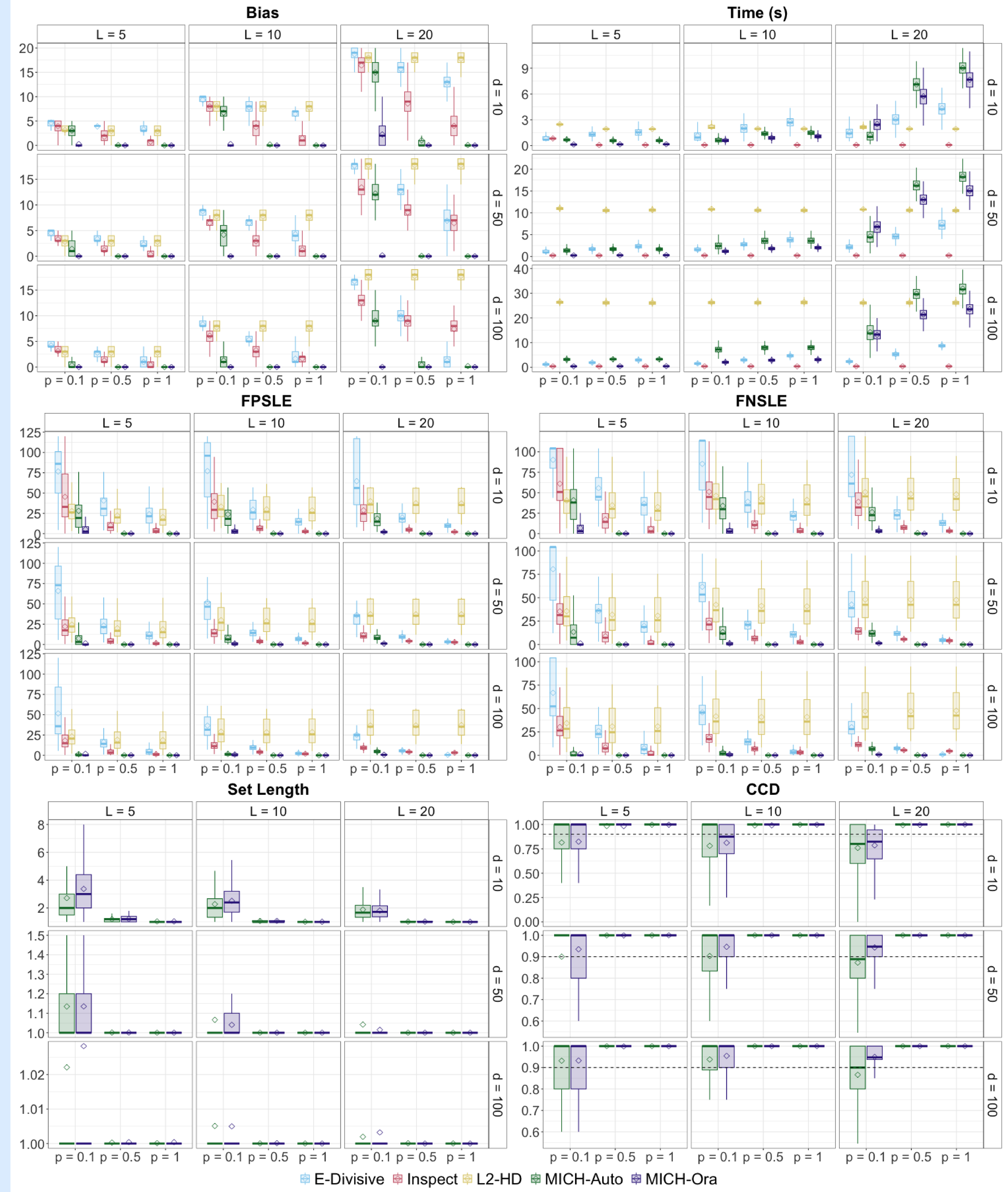

Figure 2: Box-plots of evaluation metrics across 5,000 replicates with $T = 250$ and $\Delta_T = 10$. Diamonds (◇) display mean of each statistic. Bias $|J^* - J|$ assesses each model's ability to estimate correct number of changes (lower is better). FPSLE and FNSLE assess each model's ability to accurately estimate the locations of the changes (lower is better). Set Length and CCD report the average size and coverage of credible/confidence sets for methods that provide uncertainty quantification (dashed line indicates nominal coverage level for $\alpha = 0.1$).

References

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.

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.

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.