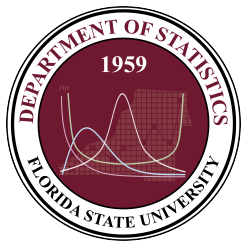


A Bayesian Approach for Nonparametric Multivariate Process Monitoring using Universal Residuals

Daniel A. Timme

Florida State University

26 April 2023

Introduction

Profile Definition

$$y_i^t = f(\boldsymbol{x}_i^t) + \varepsilon_i^t, \quad i \in \{1, \dots, n\} = [n], \quad t = 1, 2, \dots$$

where $y_i^t \in \mathbb{R}$, $\boldsymbol{x}_i^t \in \mathbb{R}^d$, and $\varepsilon_i^t \in \mathbb{R}$

Change-Point Detection

$$H_0 : f^0 = f^1 = \cdots = f^T$$
$$H_A : f^0 = f^1 = \cdots = f^\tau \neq f^{\tau+1} = \cdots = f^T$$

For fixed $\tau < T$, for all $i \in [n], t \in \{1 - m, \dots T\}$,

$$y_i^t = \begin{cases} f(\boldsymbol{x}_i^t) + \varepsilon_i^t, & t \leq \tau \quad \text{“in-control”} \\ h(\boldsymbol{x}_i^t) + \varepsilon_i^t, & t > \tau \quad \text{“out-of-control”} \end{cases}$$

where $t \leq 0$ denotes historical, known IC profiles.

Goals

Model Assumptions

- ✓ Linear & Nonlinear f
- ✓ Localized Change h
- ✓ Nonparametric
- ✓ Multivariate predictor

Performance

- ✓ Computationally fast
- ✓ Low FAR even at large τ
- ✓ Fast detection (low ARL_1)

Performance Metrics

In-control Average Run Length

Trial	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
1	✓	✓	✓	✓	×	
2	✓	✓	✓	×		
3	✓	×				
4	✓	✓	×			
5	✓	✓	✓	✓	✓	×

$$\text{ARL}_0 \simeq \frac{1}{N}\sum_{j=1}^{N} T_j = 4$$

Out-of-control Average Run Length & False Alarm Rate

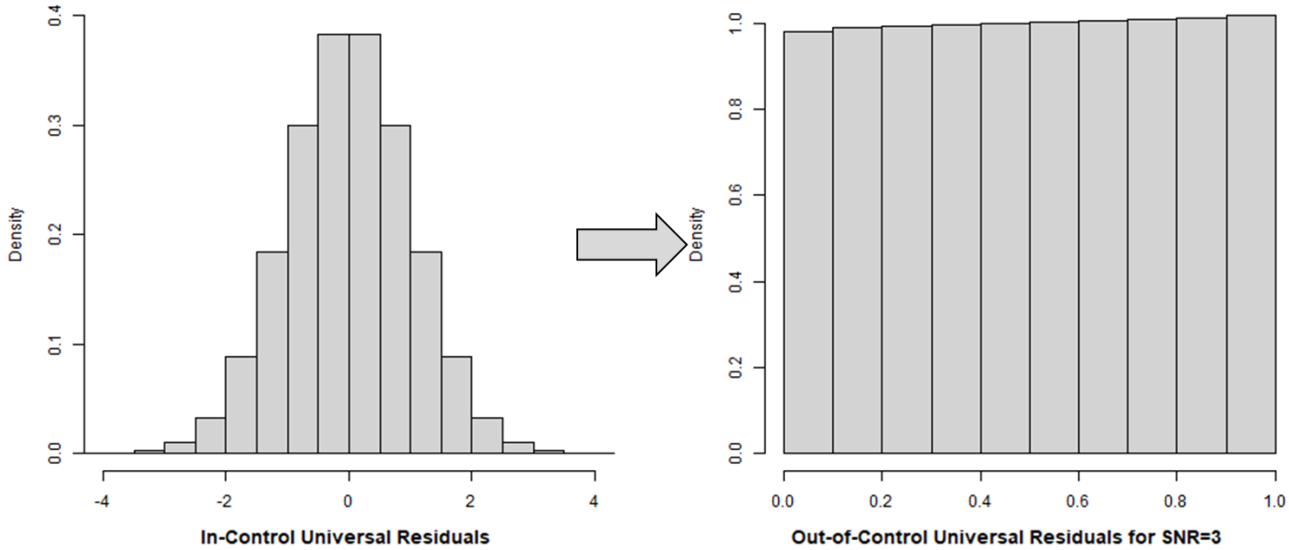
Trial	$t=1$	$t=2$	...	$t=\tau-1$	$t=\tau$	$t=\tau+1$	$t=\tau+2$
1	✓	✓		✓	✓	×	
2	✓	✓		✓	×	×	
3	✓	✓		✓	✓	×	
4	✓	✓		×	✓	×	
5	✓	✓		✓	✓	✓	×

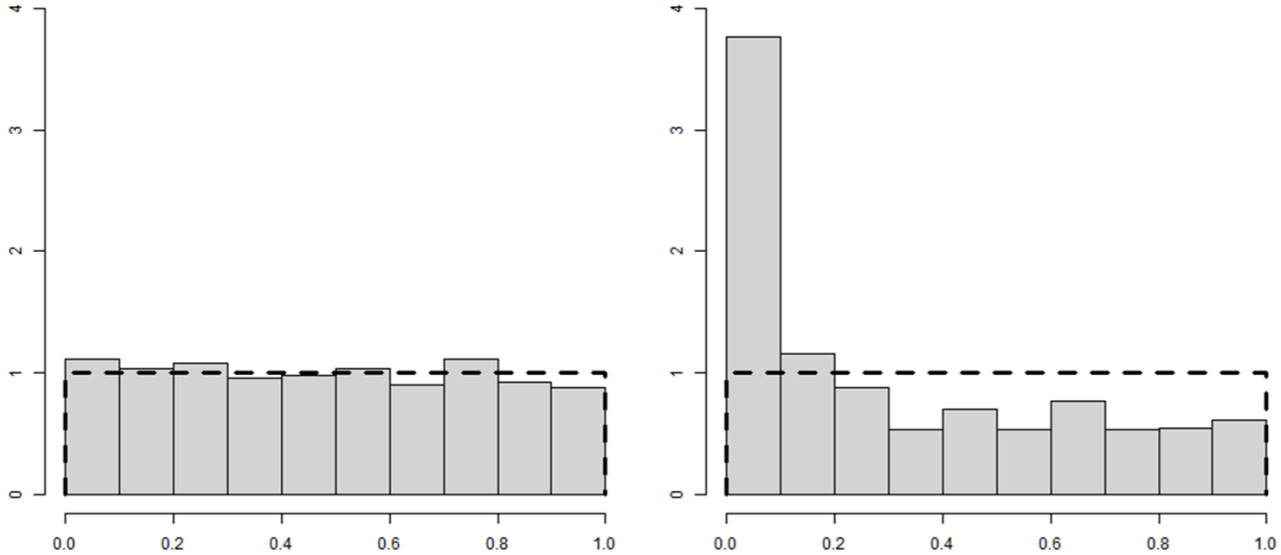
✓: No Change Point Detected
×: Change-Point Detected

$$\text{ARL}_1 \simeq \frac{1}{N}\sum_{j=1}^{N} (T_j - \tau) = 1.2$$

$$\text{FAR} \simeq \frac{N_{FA}}{N + N_{FA}} \approx 0.2857$$

Universal Residuals





Bayesian Inference on the Hypotheses

p_1: Multinomial **p_2: Dirichlet-Multinomial**

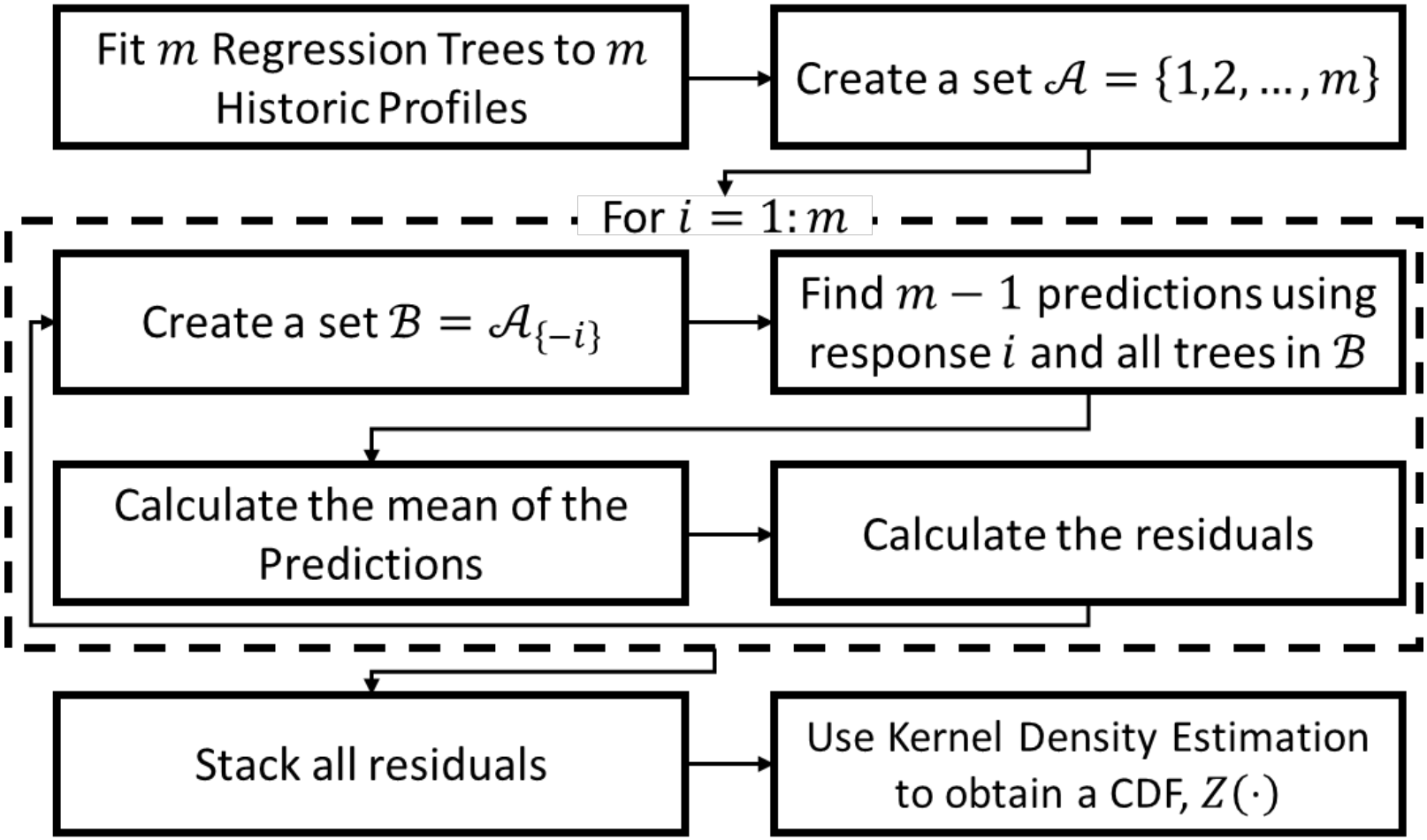
$$H_{\tau \geq t} : p\left(\vec{n}^{\,j} | H_{t,j}\right) = \prod_{j=1}^{t} p_1\left(\vec{n}^{\,j} | N, \vec{p}_{\,0}\right)$$

$$H_{\tau = q} : p\left(\vec{n}^{\,j} | H_{t,j}\right) = \prod_{j=1}^{q} p_1\left(\vec{n}^{\,j} | N, \vec{p}_{\,0}\right) \prod_{j=q+1}^{t} p_1\left(\vec{n}^{\,j} | N, (1, \cdots, 1)\right) \; \forall q \in [1, t-1]$$

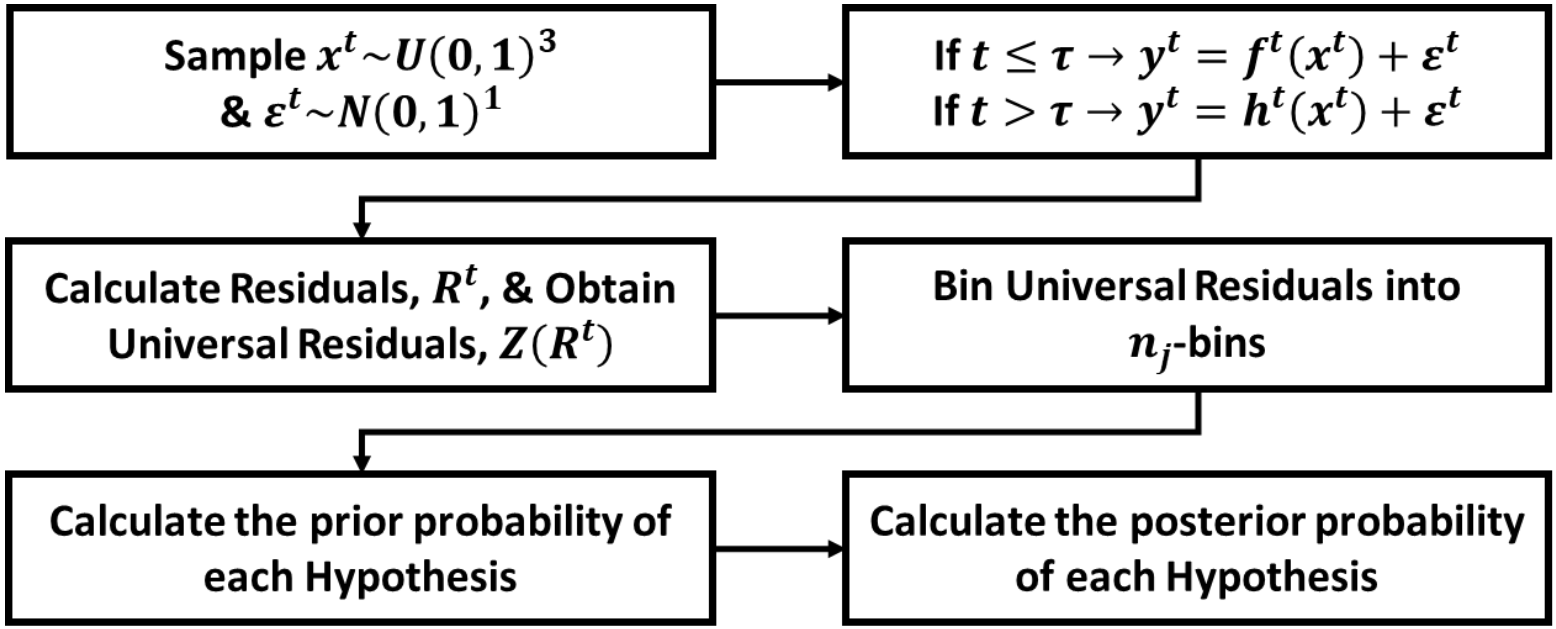
$$H_{\tau = 0} : p\left(\vec{n}^{\,j} | H_{t,j}\right) = \prod_{j=1}^{t} p_2\left(\vec{n}^{\,j} | N, (1, \cdots, 1)\right)$$

$$p\left(H_t | \text{data}\right) = \frac{p\left(\text{data} | H_t\right) \pi\left(H_t\right)}{\sum_{j=0}^{t} p\left(\text{data} | H_j\right) \pi\left(H_j\right)}$$

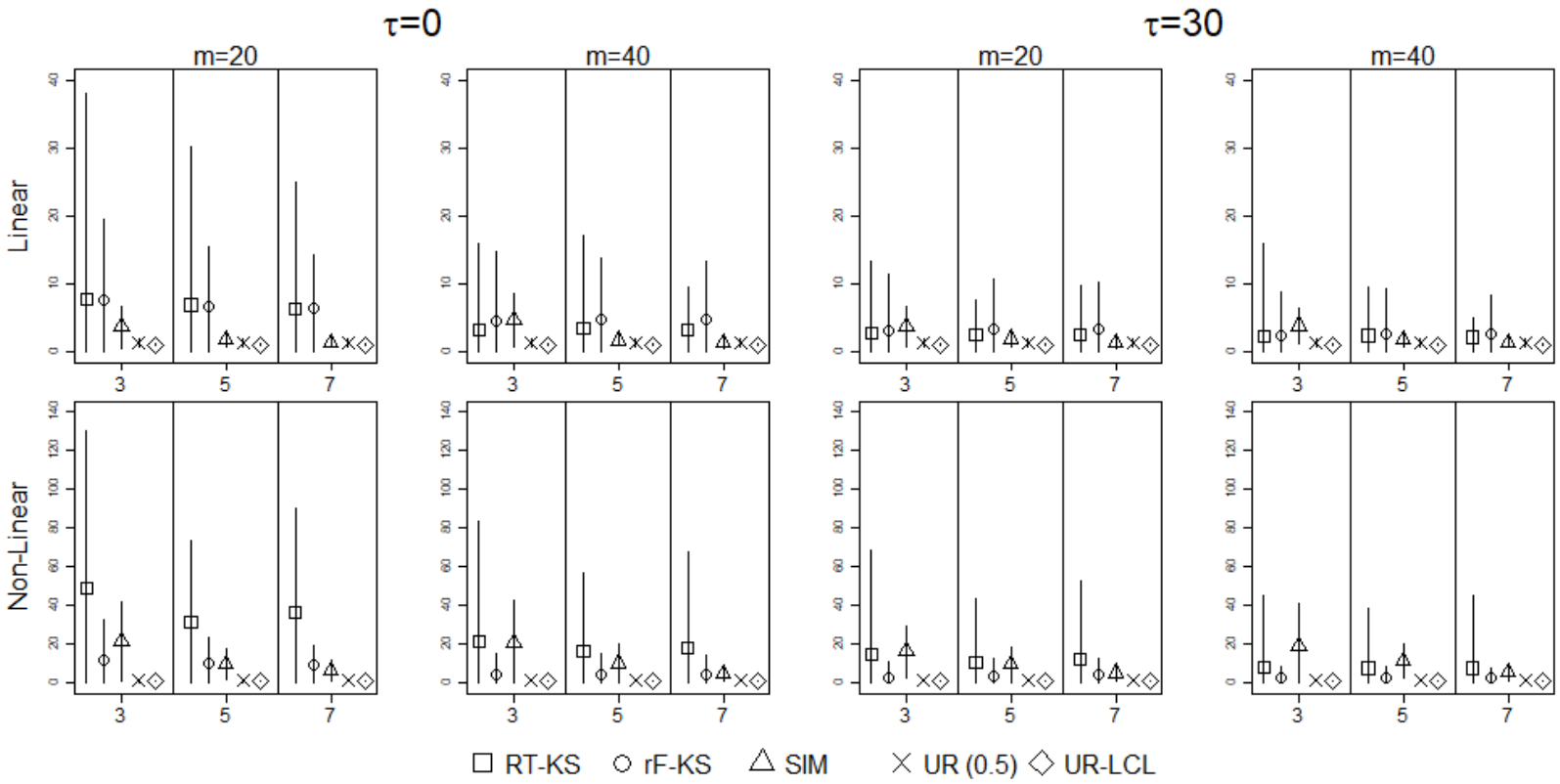
Obtaining eCDF from Historic Profiles



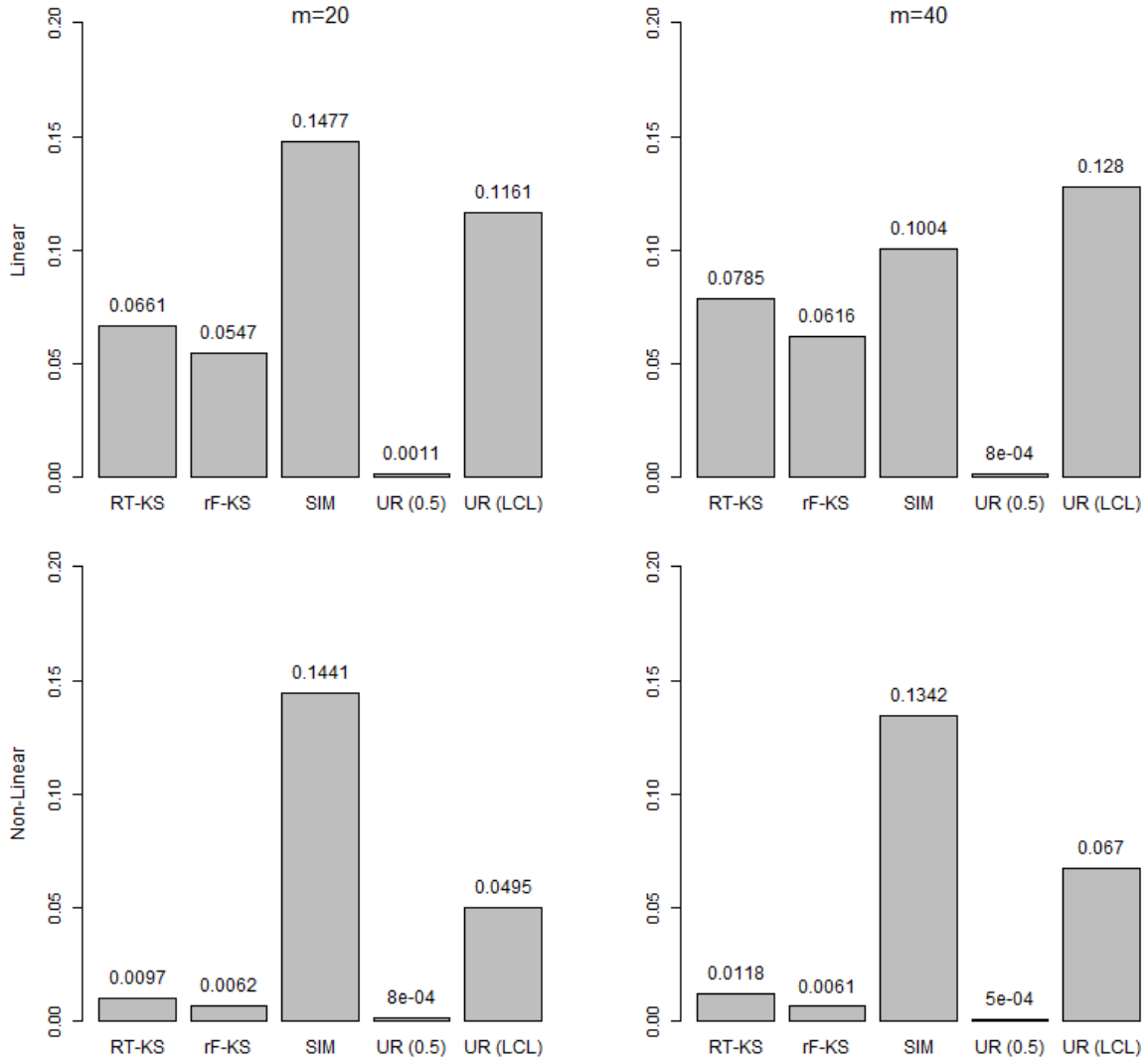
Simulation Flow Chart

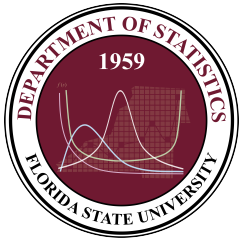


Simulation Results: ARL_1



Simulation Results: False Alarm Rate





Conclusion

Other Observations

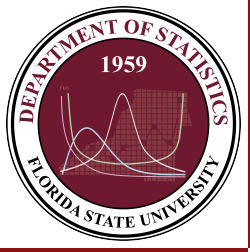
- Maintains small ARL_1 and FAR for $\tau = \{100, 500, 1000\}$
- Faster computation than previous methods
- Versatile method

ARL_1

- Out-performs other methods for ARL_1 with both control limits

False Alarm Rate

- Out-performs other methods with the 0.5 control limit



Questions?

Simulation Setup: Signal to Noise Ratio (SNR)

Signal-to-Noise (SNR) Ratio

- Localized Change

$$SNR = va^2$$