Nonparametric Threshold Estimation of Autocorrelated Statistics in Multivariate Statistical Process Monitoring

Taylor Grimm[1], Kathryn Newhart[2], and Amanda Hering[1]

September 28, 2023

[1] Department of Statistical Science, Baylor University

[2] Department of Geography and Environmental Engineering, United States Military Academy

Presentation Outline

1. Statistical Process Monitoring (SPM)
2. Parametric and Nonparametric Thresholds
3. Simulation Study
4. Case Study - Wastewater Treatment
5. Other Work

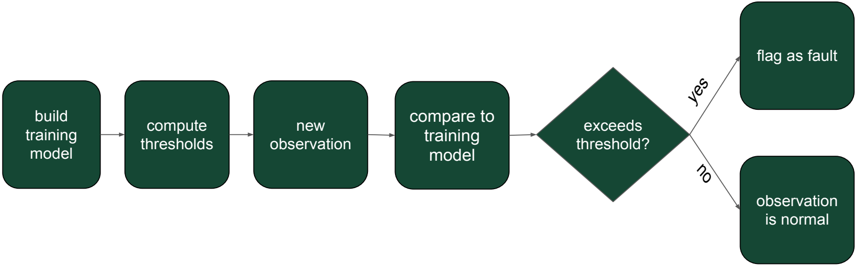
Statistical Process Monitoring (SPM)

Method for **real-time** detection of abnormal behavior in a set of variables.

Common applications:

- Water/wastewater treatment
- Manufacturing processes (e.g., military munitions)
- Healthcare (e.g., biomedical or clinical management variables)

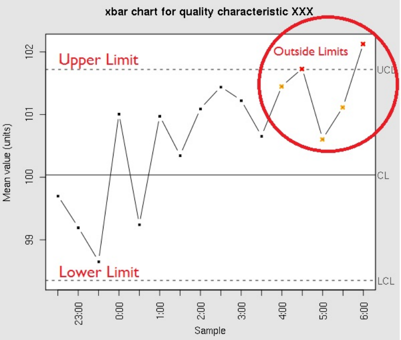
General SPM process:



Traditional SPM

In traditional SPM, the following apply:

- Upper and lower limits are set for each process variable.
- An observation is flagged as a fault if it goes outside the control limits.
- Control limits are often set at 3 standard deviations above/below the in-control mean.



(https://www.statisticshowto.com/statistical-process-control/)

Phases of SPM

Phase I

Determine in-control (IC) process parameters.

- What does normal operating behavior look like?

Phase II

Determine if a new observation is IC or out-of-control (OC).

- Is a new observation different from what normal operating behavior looks like?

How do we define "different"?

We set thresholds (control limits) to classify new observations as IC or OC.

- New observations that exceed a threshold are deemed "sufficiently different" and are flagged as OC.

Multivariate Statistical Process Monitoring (MSPM)

MSPM allows us to

- monitor many variables simultaneously
- detect changes in means and/or covariances
- condense information from many variables into one or two *monitoring statistics*

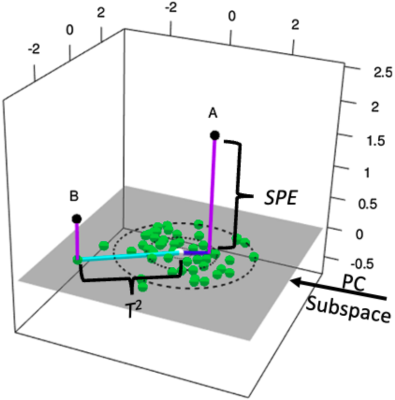
Some common methods include

- multivariate exponentially weighted moving average (MEWMA)
- multivariate cumulative sum (MCUSUM)
- principal component analysis (PCA) methods

In high-dimensional processes, PCA-based methods are the most common.

PCA Monitoring Statistics

MSPM with PCA is done with two monitoring statistics: T^2 and squared prediction error (SPE).



A set of projected observations and two original observations ($\mathbf{A}, \mathbf{B}$) (Kazor et al., 2016).

MSPM with PCA

Hotelling's T^2 at time t is computed as

$$\begin{aligned} T_t^2 &= (\mathbf{x}_t^T \mathbf{P}_k)(\mathbf{\Lambda}_k)^{-1}(\mathbf{x}_t^T \mathbf{P}_k)^T \\ &= \mathbf{y}_t^T (\mathbf{\Lambda}_k)^{-1} \mathbf{y}_t, \end{aligned}$$

and SPE can computed as

$$SPE_t = \left\| \left[\mathbf{I} - \mathbf{P}_k (\mathbf{P}_k)^T \right] \mathbf{x}_t \right\|^2 .$$

- $\mathbf{P}$ = projection matrix
- $\mathbf{\Lambda}$ = diagonal matrix of eigenvalues
- k = number of principal components retained in the model

Parametric Thresholds of T^2 and SPE

T^2 (Hotelling, 1947)

$$T^2_\alpha = \frac{k(n^2-1)}{n(n-k)} F_\alpha(k, n-k),$$

- n = number of training observations
- k = reduced dimension (number of retained PC's)
- α = IC false alarm rate
- $F_\alpha(k, n-k)$ = the upper α quantile of the F distribution with k and $n-k$ degrees of freedom

SPE (Jackson, 1979)

$$SPE_\alpha = a\chi^2_{\nu,\alpha}.$$

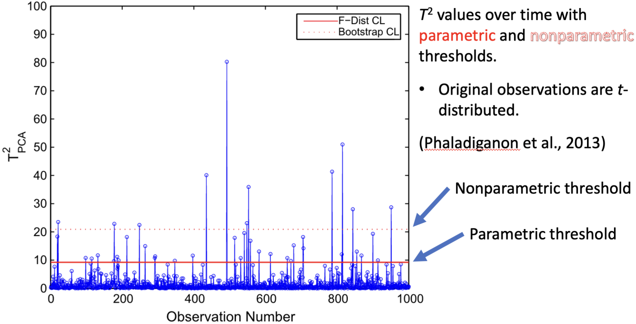
Issues with Parametric Thresholds

Assumptions are often violated.

- Data are often non-normal and heavily skewed.
- Observations are usually autocorrelated.

Why is this bad?

- Using thresholds based on the aasumptions of normality or independence results in unreliable and inaccurate alarms. (Montgomery and Mastrangelo, 1991; Ma et al., 2018).



T^2 values over time with parametric and nonparametric thresholds.

- Original observations are *t*-distributed.

(Phaladiganon et al., 2013)

Adjustments for Assumption Violations

Adjusting for autocorrelation

- Fit a statistical or machine learning model to each variable, then monitor the residuals.
- (PCA) Incorporate lags of the variables into the model (called dynamic PCA, or DPCA).

Adjusting for non-normality

- Use a different control chart altogether, such as a nonparametric control chart built on order statistics.
- Estimate the $(1 - \alpha)$ upper quantile of the IC monitoring statistic.

Nonparametric Thresholds

Kernel Density Estimation (KDE)

For monitoring statistics $z_1, \ldots, z_n$ during the training period, we estimate $f(\cdot)$ by

$$\hat{f}_h(z) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{z - z_i}{h}\right), \tag{1}$$

where K is a kernel, and h is the bandwidth.

- Find q_α such that $\int_0^{q_\alpha} \hat{f}_h(t)dt = 1 - \alpha$.

Bootstrapping

Obtain random samples, with replacement, from $\mathbf{z} = (z_1, \ldots, z_n)'$.

- Calculate the average $1 - \alpha$ sample quantile across all bootstrap samples.

Nonparametric Thresholds

KDE and Bootstrapping

- Do not assume multivariate normality of the original data.
- Both assume independence (in the monitoring statistics, not in the original data).
  - often violated (Mason and Young, 2002; Vanhatalo and Kulahci, 2015; Rato et al., 2016; Vanhatalo et al., 2017)

Dependence in Original Data vs. Monitoring Statistics

Dependence in Data

- Violates independence assumption of parametric thresholds.
- Naturally occurs with high sampling frequencies.

Dependence in Monitoring Statistics

- Violates independence assumption of nonparametric estimators.
- Monitoring statistics are often autocorrelated, even when the original observations are independent.
- Adjusting for dependence in the original data (e.g. using dynamic PCA) still results in autocorrelated monitoring statistics.

How does dependence in the monitoring statistics affect nonparametric estimation performance?

Time Series Models

Autoregressive order 1 (AR(1)) model:

$$z_t = \phi z_{t-1} + \epsilon_t \tag{2}$$

- z_t is the monitoring statistic at time index t.
- ϕ is the AR coefficient.
- ϵ_t's are "white noise" error terms that are independent and identically distributed with mean zero and finite variance.

The autoregressive moving average (ARMA) model combines the AR and moving average (MA) models into a single model. The ARMA(1, 1) is:

$$z_t = \phi z_{t-1} + \epsilon_t + \theta \epsilon_{t-1} \tag{3}$$

- θ is the MA coefficient.

Nonparametric Estimation Methods: KDE

Baseline

- Sample quantile

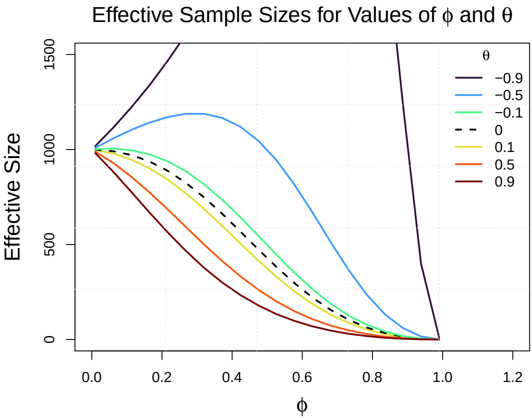
Kernel Density Estimation (plug-in bandwidth estimators)

- Silverman (SLVM)

$$h = 0.9n^{-1/5} \min\left\{\mathrm{sd}(\mathbf{z}), \frac{\mathrm{IQR}(\mathbf{z})}{1.34}\right\} \tag{4}$$

- Scott (SCOTT)
  - Replace 0.9 with 1.06 in (4).
- Adjusted Silverman (ADJ-SLVM)
  - Replace n in (4) with the estimated effective sample size $\hat{n}_{\text{eff}}$.
- Adjusted Scott (ADJ-SCOTT)
  - Replace 0.9 with 1.06 and n with $\hat{n}_{\text{eff}}$ in (4).

Plots: Effective Sample Size



Effective sample sizes for different parameters when $n = 1000$.

$$\hat{n}_{\text{eff}}^{\text{AR}(1)} = \frac{n}{1 + \frac{1}{n}\sum_{i \neq j}\sum \frac{\hat{\phi}^{|i-j|}}{1-\phi^2}} \qquad \hat{n}_{\text{eff}}^{\text{ARMA}(1,\,1)} = \frac{n}{1 + \frac{1}{n}\sum_{i \neq j}\sum \frac{(1+\hat{\theta}\hat{\phi})(\hat{\phi}+\hat{\theta})}{1-\hat{\phi}^2}\hat{\phi}^{|i-j|-1}}$$

Nonparametric Estimation Methods: Bootstrap

Bootstrap

- Standard bootstrap (BOOT)
- Moving block bootstrap (MB-BOOT)
  - Randomly sample blocks of size ℓ instead of individual observations.
- Random block bootstrap (RB-BOOT)
  - Same as MB-BOOT, but randomly generate block sizes according to a Geometric distribution with mean ℓ.

For MB-BOOT and RB-BOOT, we select ℓ by finding ℓ such that

$$|\hat{\rho}(\ell)| \leq 0.05,$$

where $\hat{\rho}$ is the estimated theoretical autocorrelation function (ACF) evaluated at lag ℓ.

We restrict ℓ to be no larger than $n/2$.

Dependent Bootstrap Block Sizes

Select the value of ℓ such that:

AR(1)

$$|\hat{\rho}(\ell)| = |\hat{\phi}^{\ell}| \leq 0.05$$

ARMA(1, 1)

$$|\hat{\rho}(\ell)| = \left| \frac{(1+\hat{\theta}\hat{\phi})(\hat{\phi}+\hat{\theta})}{1+2\hat{\theta}\hat{\phi}+\hat{\theta}^2} \hat{\phi}^{\ell-1} \right| \leq 0.05$$

Restrict ℓ to be no larger than $n/2$.

Simulation Study: Outline

Goal: Evaluate how well different nonparametric methods estimate the 0.995 quantile of an autocorrelated monitoring statistic.

Factor	Levels
AR(1)	$\phi = 0, 0.1, 0.5, 0.9$
ARMA(1, 1)	$\phi = 0.1, \theta = -0.9, -0.5, 0.1$ $\phi = 0.5, \theta = -0.9, -0.1, 0.5$ $\phi = 0.9, \theta = -0.5, -0.1, 0.9$
Sample Size (n)	100, 500, 1000, 5000
Error Distribution	$F_{5,20}$, $F_{5,n-5}$, Gamma$(0.05, 0.05)$

1. Find true quantile.
2. Generate data.
3. Estimate quantile using each method.
4. Repeat Steps 2-3 1,000 times.
5. Calculate the bias of each method.

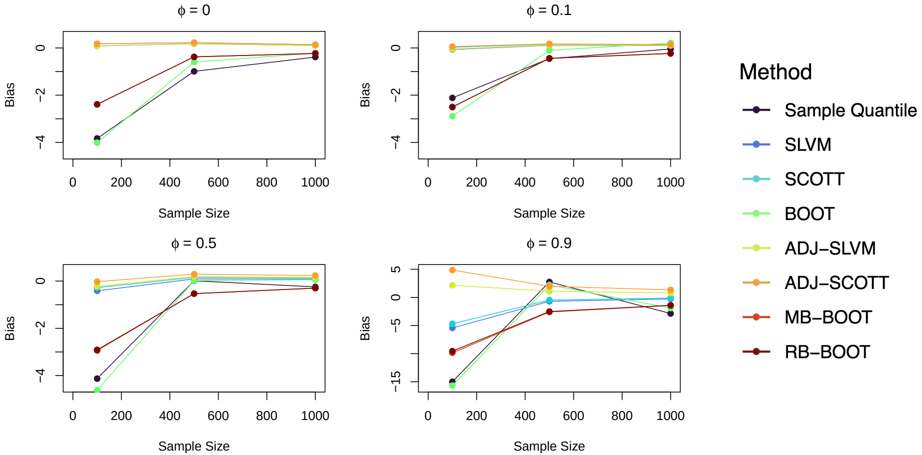
Simulation Study: Step 1

1. Find true quantile.
   (i) Simulate a series of length 1,000,000 with the desired autocorrelation structure. Use error terms that come from one of the distributions below.
      - $F_{5,20}$ distribution
      - $F_{5,n-5}$ distribution
      - Gamma(0.05, 0.05) distribution (shape and rate)

   (ii) Use the observed 0.995 sample quantile of the series generated in (i) as the true quantile of the underlying process.
2. Generate data.
3. Estimate quantile using each method.
4. Repeat Steps 2-3 1,000 times. Calculate the average quantile across the 1,000 simulated datasets.
5. Calculate the bias of each method by subtracting the true quantile computed from Step 1 from the mean estimated quantile from Step 4.

Simulation Study: Results

Bias of Estimators: $F_{5,n-5}$ Errors



- When $n = 5000$, there is little difference in the bias between methods.

Simulation Study: Results

Bias of each estimator for the 99.5th percentile of the AR(1) sampling distribution using $F_{5,n-5}$ errors. The bias closest to zero in each row is shown in bold.

		Sample	Assuming Independence			Accounting for Dependence			
ϕ	n	Quantile	SLVM	SCOTT	BOOT	ADJ-SLVM	ADJ-SCOTT	MB-BOOT	RB-BOOT
0	100	−3.83	**0.08**	0.18	−4.01	**0.08**	0.18	−2.39	−2.39
	500	−0.99	**0.18**	0.22	−0.60	**0.18**	0.22	−0.38	−0.38
	1000	−0.39	**0.10**	0.14	−0.22	**0.10**	0.14	−0.23	−0.23
	5000	0.04	**0.02**	0.04	−0.06	**0.02**	0.04	−0.08	−0.08
0.1	100	−2.11	−0.07	**0.03**	−2.89	−0.04	0.06	−2.51	−2.50
	500	−0.45	0.11	0.16	**−0.10**	0.12	0.17	−0.44	−0.44
	1000	**−0.04**	0.10	0.13	0.21	0.11	0.14	−0.24	−0.23
	5000	−0.07	0.03	0.05	**0.01**	0.03	0.05	−0.07	−0.07
0.5	100	−4.13	−0.41	−0.28	−4.61	−0.18	**0.04**	−2.94	−2.91
	500	**0.01**	0.09	0.16	0.02	0.21	0.32	−0.53	−0.53
	1000	−0.25	0.09	0.14	**0.06**	0.18	0.26	−0.30	−0.30
	5000	0.20	**0.03**	0.05	0.17	0.07	0.11	−0.09	−0.09
0.9	100	−14.99	−5.44	−4.69	−15.67	**2.41**	5.18	−9.86	−9.55
	500	2.75	−0.69	**−0.47**	2.21	4.52	6.46	−2.59	−2.52
	1000	−2.86	−0.27	**−0.12**	2.52	3.66	1.34	−1.42	−1.41
	5000	−1.55	**0.12**	0.20	−1.61	1.18	1.66	−0.22	−0.22

Simulation Study: Summary

Overall Results

- The best estimator to choose depends on the strength of the autocorrelation and sample size.
- KDE methods tend to perform better than the other methods.
- The dependent bootstraps tend to outperform the standard bootstrap when dependence is strong.
- There is a substantial difference in estimator performance when n is small and dependence is strong.

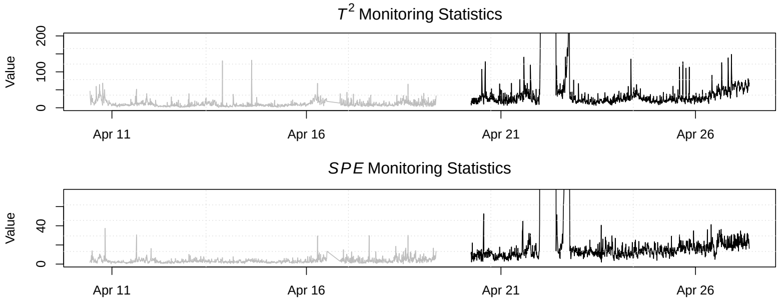
Case Study: Wastewater Treatment

Data from a demonstration-scale wastewater treatment facility in Golden, Colorado, in April 2010.

A change went unnoticed for several days, resulting in a 2 month period before the system was able to return to its normal fully operational state.

- Observations are 10 minute averages of 32 process variables from April 10 to May 10, 2010.
- The training period contains 1235 observations, and the testing period contains 1031 observations.
- Monitoring statistics are computed from the residuals of an adaptive lasso fit to 20 process variables (Klanderman et al., 2020).

Case Study: Assessing Dependence



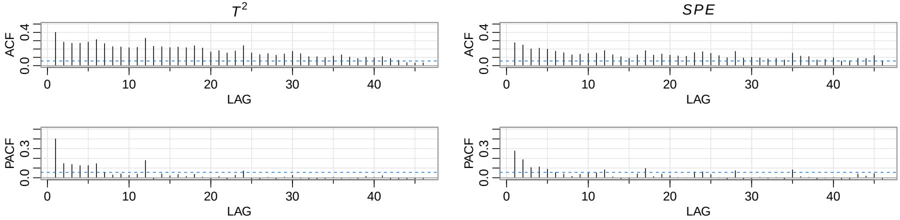
Time series plots of T^2 and SPE for the wastewater treatment data.

Estimated time series model parameters, effective sample sizes, and block sizes for T^2 and SPE using an AR(1) or ARMA(1, 1) model.

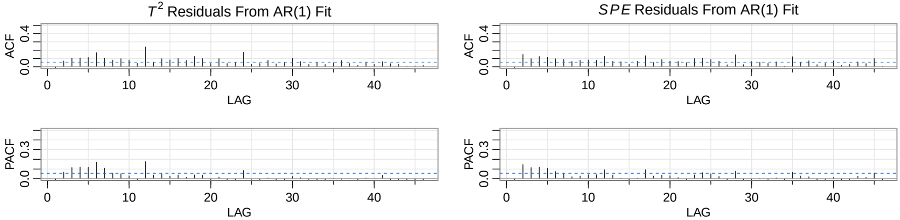
Model	Monitoring Statistic	$\hat{\phi}$	$\hat{\theta}$	$\hat{n}_{\text{eff}}$	ℓ
AR(1)	T^2	0.40		529	4
	SPE	0.28		701	3
ARMA(1, 1)	T^2	0.97	-0.83	41	66
	SPE	0.96	-0.84	79	42

Case Study: Assessing Dependence

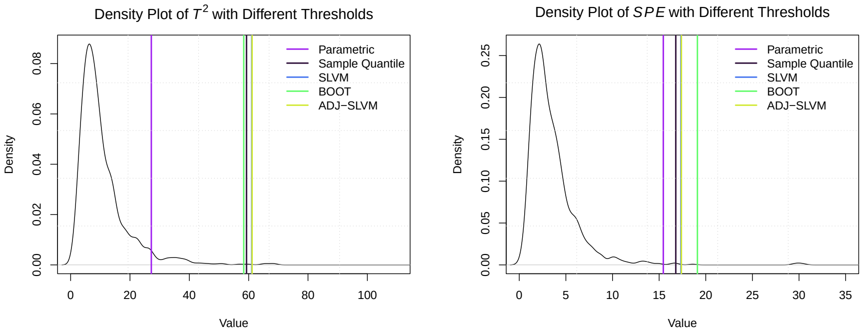
Original Monitoring Statistics



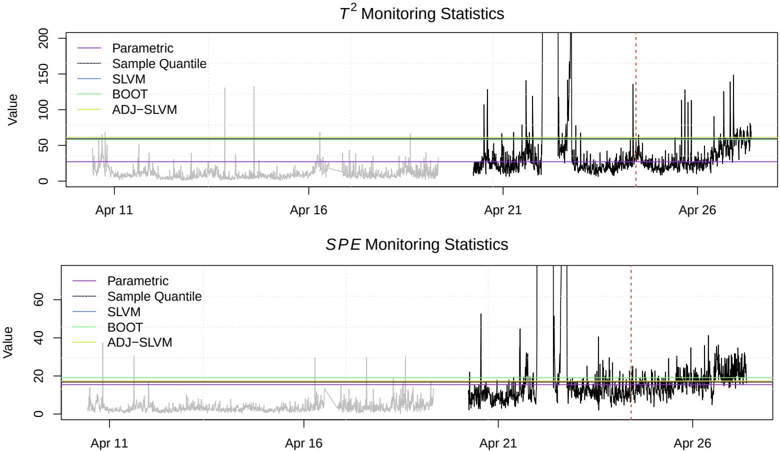
Residuals of AR(1) Fit



Case Study: Assessing Dependence



Density plots of T^2 and SPE from the training period with vertical lines for different thresholds.



T^2 (top) and SPE (bottom) monitoring statistics during the training and **testing** periods.

Case Study: Proportion of OC Observations

Proportion and number (in parentheses) of monitoring statistics flagged as OC during training and testing for selected methods when $\alpha = 0.005$.

Monitoring Statistic	Method	Training (# of Days / 1235)	Testing (# of Days / 1031)
T^2	Parametric	0.0478 (59)	0.481 (496)
	Sample Quantile	0.0057 (7)	0.153 (158)
	SLVM	0.0049 (6)	0.143 (147)
	BOOT	0.0057 (7)	0.157 (162)
	ADJ-SLVM	0.0049 (6)	0.142 (146)
SPE	Parametric	0.0073 (9)	0.434 (446)
	Sample Quantile	0.0057 (7)	0.375 (387)
	SLVM	0.0049 (6)	0.344 (355)
	BOOT	0.0041 (5)	0.277 (286)
	ADJ-SLVM	0.0049 (6)	0.343 (354)

Case Study: AR(1) and ARMA(1, 1) Thresholds

Nonparametric thresholds for T^2 and *SPE* with effective sample sizes and block sizes computed based on an AR(1) or ARMA(1, 1) structure.

Monitoring Statistic	Correlation Structure	ADJ-SLVM	ADJ-SCOTT	MB-BOOT	RB-BOOT
T^2	AR(1)	61.2	61.3	57.8	58.2
	ARMA(1, 1)	61.4	61.3	57.4	57.4
SPE	AR(1)	17.4	17.5	19.4	19.2
	ARMA(1, 1)	17.6	17.7	18.9	18.8

- For this case study, the adjustment used for effective sample size and block sizes is not practically meaningful.

Case Study: Training Period Sizes

Thresholds were also computed for different training period sample sizes.

- Parametric thresholds are mostly unchanged.
- Nonparametric thresholds vary greatly.
  - Smaller training periods may not capture the full range of normal operating behavior, resulting in major changes in threshold values.

Parametric and nonparametric thresholds for T^2 and SPE using different sizes of training periods.

Monitoring Statistic	Training Size	Parametric	Sample Quantile	SLVM	BOOT	ADJ-SLVM
T^2	100	28.4	23.4	24.5	22.7	24.6
	200	27.7	47.3	47.9	45.5	48.1
	300	26.8	67.5	68.9	68.8	69.1
	400	26.4	93.0	95.9	86.9	95.7
	500	26.2	105.7	107.2	96.9	107.3
	1235	27.2	59.3	61.0	58.3	61.2
SPE	100	11.8	11.2	12.0	10.6	12.0
	200	10.5	10.5	11.4	10.3	11.4
	300	12.9	13.3	13.8	13.9	13.8
	400	12.4	14.8	15.0	14.5	15.0
	500	14.2	18.9	19.1	18.0	19.1
	1235	15.4	16.8	17.3	19.1	17.4

Other Work

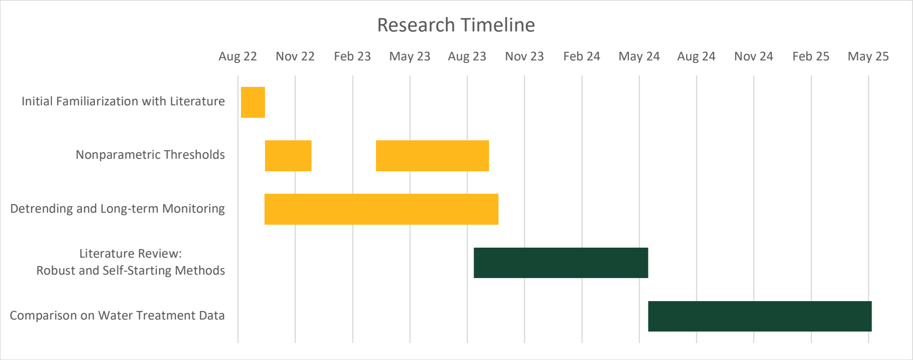
Questions of Interest

1. Are there better ways to use all process variables for fault detection? (detrending)
2. What adjustments need to be made to allow for MSPM to be applied long term?
3. How do we build MSPM models if there is no fault-free training period? What if there is very little historical data?

Research Papers

- A paper on nonparametric thresholds (this presentation) will be submitted to *Quality and Reliability Engineering International*.
- A paper addressing 1 & 2 will be submitted to a journal like *Water Research* or *Water Science & Technology*.
- The next project is to write a literature review on robust and self-starting SPM methods. (question 3 above)

Research Timeline



Acknowledgement

Special Thanks:

Alyssa & Emryn
Dr. Mandy Hering and Dr. Kate Newhart
Other students who have listened to me give at least 5 presentations on MSPM

Funding Source:

This work is supported by the National Alliance for Water Innovation (NAWI), funded by the US Department of Energy (DOE), Energy Efficiency and Renewable Energy Office, Advanced Manufacturing Office under Funding Opportunity Announcement DE-FOA-0001905, Project 5.17 Data-Driven Fault Detection and Process Control for Potable Reuse with Reverse Osmosis.

References

Hotelling, H. (1947) "Multivariate quality control" In Eisenhart, C., Hastay, M. W., and Wallis, W., editors, Techniques of Statistical Analysis, pages 111–184. McGraw-Hill, New York.

Jackson, J. E. and Mudholkar, G. S. (1979) "Control procedures for residuals associated with principal component analysis," Technometrics, 21(3):341–349.

Kazor, K., Holloway, R., Cath, T., and Hering, A. S. (2016) "Comparison of linear and nonlinear dimension reduction techniques for automated process monitoring of a decentralized wastewater treatment facility," Stochastic Environmental Research and Risk Assessment, 30: 1527-1544.

Klanderman, M., Newhart, K.B., Cath. T.Y., Hering, A.S. (2020) "Fault isolation for a complex decentralized wastewater treatment facility," Journal of the Royal Statistical Society, Series C., 69, 931-951.

Klanderman, M. C., Newhart, K. B., Cath, T. Y., and Hering, A. S. (2020) "Case studies in real-time fault isolation in a decentralized wastewater treatment facility," Journal of Water Process Engineering, 38: 101556.

Ku, W., Storer, R. H., and Georgakis, C. (1995). "Disturbance detection and isolation by dynamic principal component analysis," Chemometrics and Intelligent Laboratory Systems, 30(1):179–196.

References

Ma, X., Zhang, L., Hu, J., and Palazoglu, A. (2018). “A model-free approach to reduce the effect of autocorrelation on statistical process control charts,” Journal of Chemometrics, 32(12).

Mason, R. L. and Young, J. C. (2002). “Multivariate statistical process control with industrial applications,” Society for Industrial and Applied Mathematics.

Montgomery, D. C. and Mastrangelo, C. M. (1991). “Some statistical process control methods for autocorrelated data,” Journal of Quality Technology, 23(3):179–193.

Newhart, K.B., Klanderman, M.C., Hering, A.S., Cath, T.Y. (2023) “A holistic evaluation of multivariate statistical process monitoring in a biological and membrane treatment system,” in review.

Odom, G.J., Newhart, K.B., Cath, T.Y., Hering, A.S. (2018) “Multi-state multivariate statistical process control,” Applied Stochastic Models in Business and Industry, 34(6), 880-892.

Phaladiganon, P., Kim, S. B., Chen, V. C. P., and Jiang, W. (2013). “Principal component analysis-based control charts for multivariate nonnormal distributions,” Expert Systems with Applications, 40(8):3044–3054.

References

Rato, T., Reis, M., Schmitt, E., Hubert, M., and De Ketelaere, B. (2016). "A systematic comparison of PCA-based statistical process monitoring methods for high-dimensional, time-dependent Processes," AIChE Journal, 62(5):1478–1493.

Tibshirani, R. (1996) "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society, Series B (Methodological), 58(1), 267–288.

Vanhatalo, E. and Kulahci, M. (2015). "The effect of autocorrelation on the Hotelling T^2 control chart," Quality and Reliability Engineering International, 31(8):1779–1796.

Vanhatalo, E., Kulahci, M., and Bergquist, B. (2017). "On the structure of dynamic principal component analysis used in statistical process monitoring," Chemometrics and Intelligent Laboratory Systems, 167:1–11.

Zou, H. (2006) "The adaptive lasso and its oracle properties," Journal of the American Statistical Association, 101(476), 1418–1429.

Principal Component Analysis

1. Apply eigenvalue decomposition to the correlation matrix $\mathbf{R}$ of the variables we want to monitor:

$$\mathbf{R} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$$

   - $\mathbf{P}$ is the projection matrix.
   - $\boldsymbol{\Lambda}$ is the diagonal matrix of the eigenvalues of $\mathbf{R}$.

2. Keep $\mathbf{P}_k$ and $\boldsymbol{\Lambda}_k$, the first k columns of $\mathbf{P}$ and $\boldsymbol{\Lambda}$, where k is the # of principal components required to capture 80% of the variability.
3. Project the original data $\mathbf{X}$ into the PCA subspace by

$$\mathbf{Y} = \mathbf{X}\mathbf{P}_k.$$

Parametric Threshold of SPE

SPE (Jackson, 1979)

$$SPE_\alpha = a\chi^2_{\nu,\alpha},$$

- n = number of observations in the training period
- k = number of principal components retained
- α = upper tail area

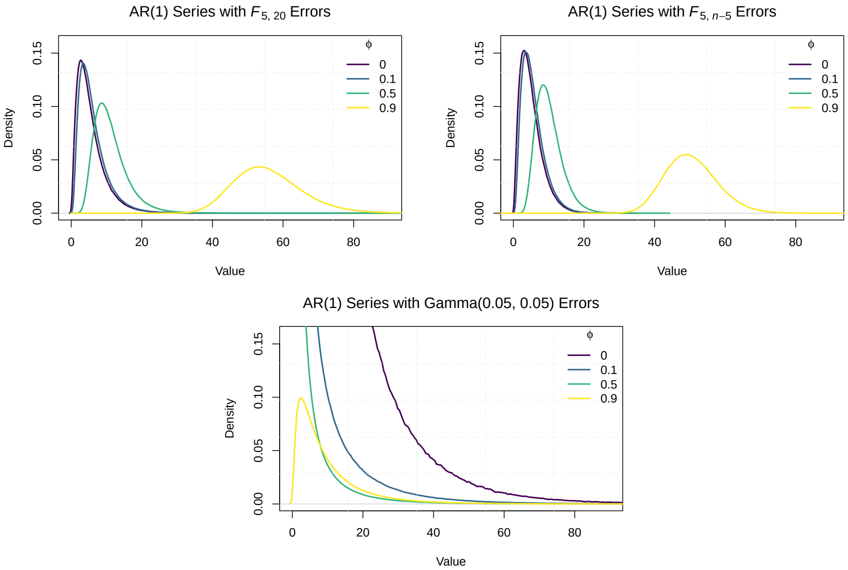
Determining a and ν:

Let

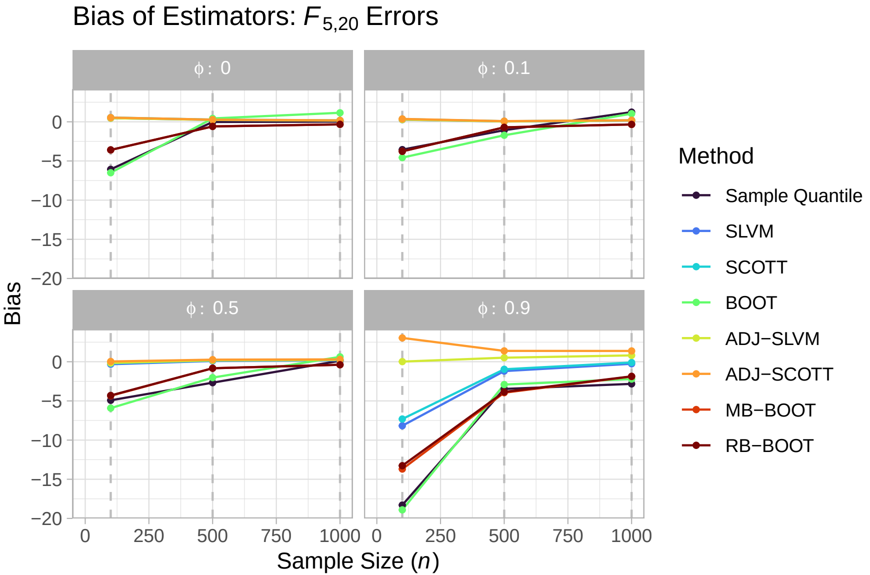
$$\overline{SPE} = \frac{1}{n}\sum_{i=1}^{n} SPE_i = a\nu,$$
$$\frac{1}{n-1}\sum_{i=1}^{n}(SPE_i - \overline{SPE})^2 = 2a^2\nu.$$

Solve for a and ν.

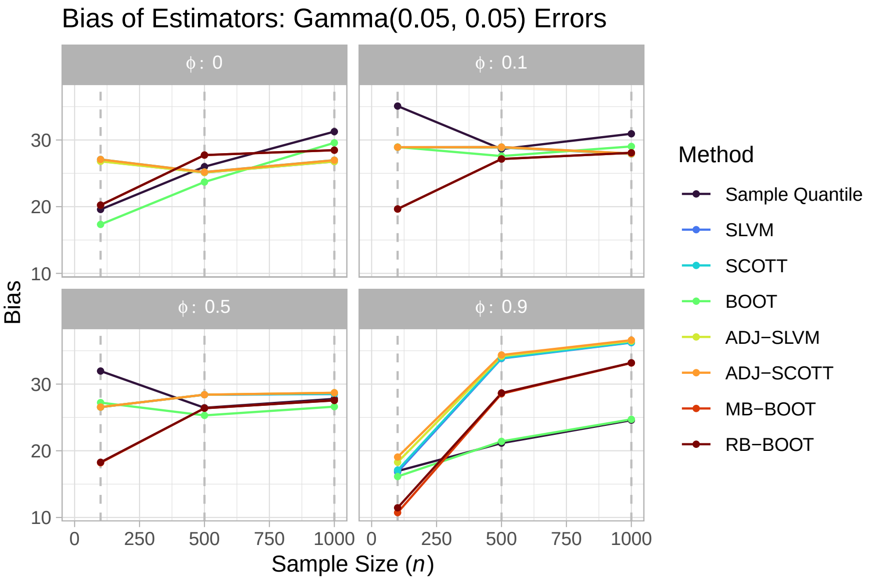
Simulation Study: True AR(1) Distributions



Simulation Study: Results



Simulation Study: Results



Bias: AR(1) Series

Bias of each estimator for the 99.5th percentile of the AR(1) sampling distribution using F errors with 5 and 20 degrees of freedom.

		Sample	Assuming Independence			Accounting for Dependence			
ϕ	n	Quantile	SLVM	SCOTT	BOOT	ADJ-SLVM	ADJ-SCOTT	MB-BOOT	RB-BOOT
0	100	−6.07	**0.47**	0.54	−6.50	**0.47**	0.54	−3.60	−3.59
	500	**−0.02**	0.24	0.28	0.43	0.24	0.28	−0.59	−0.59
	1000	**0.02**	0.15	0.18	1.15	0.15	0.18	−0.33	−0.33
	5000	**−0.01**	0.04	0.05	0.09	0.04	0.05	−0.08	−0.09
0.1	100	−3.55	**0.27**	0.35	−4.57	0.28	0.37	−3.78	−3.77
	500	−1.07	**0.04**	0.07	−1.72	**0.04**	0.09	−0.71	−0.71
	1000	1.23	**0.16**	0.19	1.03	0.17	0.20	−0.34	−0.34
	5000	0.14	**0.04**	0.05	0.08	**0.04**	0.05	−0.09	−0.09
0.5	100	−4.92	−0.32	−0.20	−5.92	−0.15	**0.03**	−4.35	−4.30
	500	−2.67	**0.10**	0.16	−2.02	0.18	0.27	−0.83	−0.83
	1000	**0.13**	0.18	0.22	0.62	0.24	0.30	−0.38	−0.38
	5000	0.11	0.11	0.13	−0.09	0.14	0.17	**−0.04**	**−0.04**
0.9	100	−18.29	−8.17	−7.31	−18.91	**0.02**	3.04	−13.70	−13.27
	500	−3.47	−1.19	−0.96	−2.91	**0.52**	1.38	−3.95	−3.86
	1000	−2.83	−0.26	**−0.10**	−2.21	0.81	1.38	−1.88	−1.87
	5000	−3.25	0.39	0.47	−3.42	0.86	1.12	**−0.02**	**−0.02**

Bias: AR(1) Series

Bias of each estimator for the 99.5th percentile of the AR(1) sampling distribution with Gamma(0.05, 0.05) errors.

		Sample	Assuming Independence			Accounting for Dependence			
ϕ	n	Quantile	SLVM	SCOTT	BOOT	ADJ-SLVM	ADJ-SCOTT	MB-BOOT	RB-BOOT
0	100	−9.91	−2.66	−**2.41**	−12.15	−2.69	−**2.41**	−9.26	−9.25
	500	−3.50	−4.37	−4.29	−5.79	−4.37	−4.29	−1.76	−**1.75**
	1000	1.76	−2.72	−2.52	**0.09**	−2.72	−2.52	−1.02	−1.02
	5000	0.43	−3.15	−2.90	−**0.06**	−3.15	−2.90	−0.43	−0.43
0.1	100	5.98	−0.20	−0.20	−**0.17**	−0.20	−0.20	−9.47	−9.45
	500	−0.48	−0.26	−0.18	−1.52	−0.23	−**0.17**	−1.96	−1.97
	1000	1.82	−1.21	−1.15	−**0.08**	−1.18	−1.14	−1.04	−1.04
	5000	0.33	−1.79	−1.52	**0.23**	−1.74	−1.45	−0.45	−0.45
0.5	100	3.19	−2.22	−2.22	−**1.54**	−2.22	−2.21	−10.56	−10.49
	500	−2.32	−0.36	−0.36	−3.46	−0.36	−**0.35**	−2.38	−2.39
	1000	−0.98	−0.26	−0.24	−2.15	−**0.07**	−**0.07**	−1.21	−1.22
	5000	−1.09	−0.28	−0.26	−1.24	−**0.25**	−**0.25**	−0.47	−0.46
0.9	100	−20.70	−20.85	−20.56	−21.51	−19.37	−**18.59**	−26.93	−26.20
	500	−16.51	−3.82	−3.76	−16.25	−3.48	−**3.29**	−9.11	−8.99
	1000	−13.07	−1.44	−1.40	−12.94	−1.21	−**1.07**	−4.49	−4.47
	5000	−8.43	−0.76	−0.74	−8.11	−0.65	−**0.58**	−1.26	−1.27

Standard Errors: AR(1) Series

Standard error of each estimator for the 99.5th percentile of the sampling distribution using F errors with 5 and $n-5$ degrees of freedom.

		Sample	Assuming Independence			Accounting for Dependence			
ϕ	n	Quantile	SLVM	SCOTT	BOOT	ADJ-SLVM	ADJ-SCOTT	MB-BOOT	RB-BOOT
0	100	2.77	3.64	3.59	2.45	3.64	3.59	2.45	2.45
	500	1.39	1.44	1.41	1.27	1.44	1.41	1.28	1.27
	1000	1.07	1.02	1.00	0.94	1.02	1.00	0.94	0.94
	5000	0.48	0.46	0.45	0.46	0.46	0.45	0.46	0.46
0.1	100	2.83	3.71	3.67	2.49	3.71	3.66	2.49	2.49
	500	1.36	1.41	1.38	1.24	1.40	1.37	1.24	1.24
	1000	1.06	1.03	1.01	0.96	1.02	1.01	0.96	0.96
	5000	0.46	0.44	0.44	0.44	0.44	0.43	0.44	0.44
0.5	100	3.16	3.85	3.79	2.86	3.79	3.72	2.79	2.81
	500	1.51	1.54	1.52	1.42	1.51	1.48	1.41	1.40
	1000	1.15	1.12	1.11	1.08	1.10	1.09	1.07	1.07
	5000	0.52	0.49	0.49	0.50	0.49	0.48	0.49	0.49
0.9	100	6.60	7.01	7.11	6.40	7.89	8.33	5.94	6.09
	500	3.97	3.96	3.92	3.87	3.87	3.85	3.48	3.50
	1000	3.15	3.11	3.08	3.07	2.99	2.95	2.83	2.84
	5000	1.43	1.39	1.39	1.41	1.36	1.34	1.38	1.38

Standard Errors: AR(1) Series

Standard error of each estimator for the 99.5th percentile of the sampling distribution using F errors with 5 and 20 degrees of freedom.

		Sample	Assuming Independence			Accounting for Dependence			
ϕ	n	Quantile	SLVM	SCOTT	BOOT	ADJ-SLVM	ADJ-SCOTT	MB-BOOT	RB-BOOT
0	100	5.11	7.23	7.19	4.47	7.23	7.19	4.46	4.47
	500	2.79	3.15	3.12	2.65	3.15	3.12	2.64	2.65
	1000	2.03	2.04	2.02	1.85	2.04	2.02	1.86	1.87
	5000	0.97	0.95	0.94	0.92	0.95	0.94	0.92	0.92
0.1	100	5.00	7.05	7.01	4.38	7.05	7.00	4.38	4.38
	500	2.67	2.91	2.88	2.53	2.90	2.87	2.52	2.53
	1000	2.06	2.08	2.06	1.89	2.08	2.06	1.90	1.89
	5000	0.96	0.94	0.93	0.91	0.94	0.93	0.91	0.91
0.5	100	5.51	7.12	7.06	4.89	7.04	6.96	4.77	4.79
	500	2.92	3.13	3.10	2.82	3.09	3.05	2.77	2.78
	1000	2.29	2.29	2.28	2.16	2.27	2.25	2.14	2.14
	5000	1.05	1.03	1.02	1.01	1.02	1.01	1.01	1.01
0.9	100	9.55	10.08	10.18	9.21	11.03	11.52	8.47	8.73
	500	6.86	6.98	6.92	6.63	6.77	6.68	5.91	5.95
	1000	5.37	5.45	5.41	5.30	5.28	5.20	4.92	4.94
	5000	2.62	2.59	2.58	2.59	2.54	2.51	2.56	2.55

Standard error of each estimator for the 99.5th percentile of the sampling distribution with Gamma(0.05, 0.05) errors.

		Sample	Assuming Independence			Accounting for Dependence			
ϕ	n	Quantile	SLVM	SCOTT	BOOT	ADJ-SLVM	ADJ-SCOTT	MB-BOOT	RB-BOOT
0	100	12.77	16.78	16.86	10.99	16.79	16.85	11.03	11.02
	500	7.45	6.95	6.98	6.98	6.95	6.98	6.99	6.97
	1000	5.87	5.39	5.52	5.37	5.39	5.52	5.38	5.38
	5000	2.80	2.45	2.48	2.65	2.45	2.48	2.65	2.65
0.1	100	12.40	17.78	17.78	10.76	17.78	17.78	10.70	10.71
	500	7.58	8.80	8.82	7.27	8.80	8.82	7.24	7.25
	1000	5.90	5.87	5.89	5.23	5.88	5.89	5.24	5.23
	5000	2.78	2.68	2.73	2.67	2.70	2.73	2.65	2.66
0.5	100	13.83	18.00	18.00	12.26	18.00	18.00	11.75	11.83
	500	8.56	9.55	9.55	8.51	9.55	9.55	8.29	8.27
	1000	6.15	6.34	6.37	5.81	6.43	6.39	5.74	5.72
	5000	3.01	3.01	3.01	2.90	3.01	3.01	2.88	2.88
0.9	100	18.99	20.11	20.17	18.18	20.51	20.75	16.18	16.74
	500	17.60	18.38	18.36	17.14	18.28	18.23	15.09	15.22
	1000	13.53	13.97	13.97	13.29	13.91	13.87	12.17	12.21
	5000	6.35	6.38	6.38	6.33	6.36	6.35	6.25	6.24

Bias: ARMA(1, 1) Series

Bias of each estimator for the 99.5th percentile of the sampling distribution using F errors with 5 and $n - 5$ degrees of freedom.

			Sample	Assuming Independence			Accounting for Dependence			
ϕ	θ	n	Quantile	SLVM	SCOTT	BOOT	ADJ-SLVM	ADJ-SCOTT	MB-BOOT	RB-BOOT
0.1	−0.9	100	−1.25	0.17	0.32	−1.99	**0.15**	0.28	−2.38	−2.37
		500	−0.24	0.20	0.28	−0.24	**0.19**	0.26	−0.39	−0.39
		1000	−0.04	0.18	0.23	**0.01**	0.17	0.22	−0.19	−0.19
		5000	0.14	0.10	0.12	−0.13	0.09	0.12	**−0.02**	**−0.02**
	−0.5	100	−1.74	**0.01**	0.11	−2.35	−0.02	0.08	−2.44	−2.43
		500	−0.37	**0.15**	0.21	−0.44	**0.15**	0.20	−0.40	−0.40
		1000	−0.57	**0.12**	0.16	−0.27	**0.12**	0.16	−0.21	−0.22
		5000	−0.19	**0.05**	0.08	−0.07	**0.05**	0.07	**−0.05**	**−0.05**
	0.1	100	−2.27	−0.10	**0.00**	−3.13	−0.09	0.02	−2.55	−2.53
		500	**0.08**	0.12	0.17	0.15	0.13	0.18	−0.43	−0.43
		1000	0.77	**0.09**	0.13	0.57	0.10	0.13	−0.24	−0.24
		5000	**−0.01**	0.02	0.04	0.09	0.03	0.04	−0.07	−0.07
0.5	−0.9	100	−1.15	0.16	0.28	−2.09	**0.06**	0.14	−2.29	−2.28
		500	−0.59	0.18	0.24	−0.24	**0.14**	0.18	−0.38	−0.37
		1000	−0.61	0.16	0.20	**−0.09**	0.12	0.15	−0.19	−0.18
		5000	−0.11	0.05	0.07	0.05	**0.04**	0.05	−0.05	−0.05
	−0.1	100	−3.54	−0.30	−0.19	−4.17	−0.21	**−0.05**	−2.80	−2.78
		500	−0.09	0.08	0.14	**0.04**	0.13	0.21	−0.51	−0.51
		1000	0.40	0.09	0.14	0.34	**0.13**	0.19	−0.27	−0.28
		5000	−0.12	0.03	0.05	**0.02**	0.05	0.08	−0.08	−0.08
	0.5	100	−7.60	−1.37	−1.12	−7.98	−0.83	**−0.41**	−4.32	−4.28
		500	0.46	**0.13**	0.23	0.84	0.33	0.51	−0.80	−0.80
		1000	−0.43	**0.15**	0.22	−0.54	0.30	0.42	−0.38	−0.38
		5000	−0.07	**0.04**	0.08	−0.07	0.12	0.19	−0.13	−0.13
0.9	−0.5	100	−6.16	−1.91	−1.67	−6.73	**1.04**	2.27	−4.51	−4.38
		500	0.83	−0.16	**−0.06**	0.48	0.82	1.28	−1.00	−0.99
		1000	**0.01**	−0.06	**0.01**	−0.05	0.51	0.79	−0.57	−0.57
		5000	−1.01	**0.02**	0.05	−0.92	0.25	0.38	−0.15	−0.15
	−0.1	100	−12.94	−4.75	−4.11	−13.69	**2.17**	4.65	−8.70	−8.44
		500	1.95	−0.58	**−0.38**	2.03	3.60	5.18	−2.25	−2.19
		1000	−1.80	−0.21	**−0.07**	−1.45	1.95	2.85	−1.23	−1.22
		5000	−1.67	**0.09**	0.16	−1.53	0.92	1.30	−0.21	−0.22
	0.9	100	−30.80	−10.73	−9.22	−31.33	**4.55**	9.81	−19.40	−18.76
		500	7.12	−1.44	**−1.01**	4.01	9.24	13.09	−5.13	−4.94
		1000	−7.14	−0.54	**−0.26**	−3.80	9.20	12.77	−2.79	−2.75
		5000	−2.56	**0.14**	0.28	−2.50	3.58	4.99	−0.51	−0.51

Bias - ARMA(1, 1) Series

Bias of each estimator for the 99.5th percentile of the sampling distribution using F errors with 5 and 20 degrees of freedom.

			Sample	Assuming Independence			Accounting for Dependence			
ϕ	θ	n	Quantile	SLVM	SCOTT	BOOT	ADJ-SLVM	ADJ-SCOTT	MB-BOOT	RB-BOOT
0.1	−0.9	100	−3.62	0.55	0.67	−5.17	**0.53**	0.64	−3.50	−3.49
		500	−0.97	0.24	0.29	−1.26	**0.23**	0.28	−0.62	−0.62
		1000	1.69	**0.21**	0.26	1.97	**0.21**	0.25	−0.29	−0.29
		5000	0.35	0.08	0.10	0.20	0.07	0.09	**−0.06**	**−0.06**
	−0.5	100	−3.32	0.44	0.53	−4.81	**0.43**	0.51	−3.57	−3.55
		500	−1.54	**0.16**	0.20	−1.61	**0.16**	0.19	−0.65	−0.64
		1000	0.96	**0.19**	0.22	1.65	**0.19**	0.22	−0.29	−0.29
		5000	−0.21	0.07	0.09	**0.02**	0.07	0.09	−0.05	−0.05
	0.1	100	−3.74	**0.20**	0.28	−4.62	0.21	0.30	−3.85	−3.82
		500	−1.35	**0.02**	0.07	−1.76	0.03	0.07	−0.73	−0.73
		1000	1.26	**0.15**	0.18	1.04	**0.15**	0.18	−0.35	−0.36
		5000	0.22	**0.02**	0.03	0.07	**0.02**	0.04	−0.11	−0.11
0.5	−0.9	100	−2.52	0.62	0.71	−4.23	**0.55**	0.61	−3.41	−3.37
		500	−1.03	0.18	0.22	−1.29	**0.14**	0.17	−0.62	−0.62
		1000	0.98	0.20	0.23	1.81	**0.17**	0.20	−0.28	−0.29
		5000	**0.01**	0.07	0.09	0.02	0.06	0.07	−0.06	−0.06
	−0.1	100	−4.64	−0.10	**0.01**	−5.39	**−0.01**	0.13	−4.14	−4.10
		500	−2.10	**0.07**	0.12	−1.96	0.12	0.18	−0.79	−0.79
		1000	−0.30	**0.17**	0.21	0.62	0.20	0.25	−0.37	−0.37
		5000	−0.09	**0.07**	0.09	−0.05	0.08	0.11	**−0.07**	**−0.07**
	0.5	100	−8.73	−2.31	−2.05	−9.40	−1.76	**−1.31**	−6.47	−6.41
		500	−1.31	**0.03**	0.13	−1.57	0.24	0.42	−1.23	−1.23
		1000	3.58	**0.14**	0.21	2.83	0.28	0.40	−0.51	−0.51
		5000	0.39	0.15	0.18	0.58	0.21	0.27	**−0.06**	**−0.06**
0.9	−0.5	100	−7.77	−2.41	−2.19	−8.62	**0.33**	1.55	−6.26	−6.08
		500	−2.54	−0.36	**−0.26**	−2.43	0.45	0.85	−1.51	−1.50
		1000	0.35	−0.04	**0.03**	0.17	0.45	0.71	−0.78	−0.77
		5000	−1.13	0.13	0.16	−1.07	0.35	0.47	**−0.08**	**−0.08**
	−0.1	100	−15.68	−7.01	−6.30	−16.45	**0.93**	3.84	−12.01	−11.63
		500	−3.14	−1.03	**−0.83**	−2.77	2.76	4.32	−3.44	−3.37
		1000	−1.93	−0.23	**−0.08**	−1.74	1.88	2.80	−1.67	−1.65
		5000	−2.55	0.37	0.44	−2.66	1.22	1.62	**−0.01**	**−0.01**
	0.9	100	−35.97	−16.03	−14.27	−37.11	**2.00**	8.28	−26.80	−25.93
		500	−2.72	−2.52	**−2.07**	−4.96	9.35	13.84	−7.84	−7.63
		1000	−4.84	−0.58	**−0.26**	−3.97	9.48	13.35	−3.74	−3.70
		5000	−6.46	0.76	0.91	−6.69	4.40	5.92	**−0.04**	**−0.04**

Bias of each estimator for the 99.5th percentile of the sampling distribution with Gamma(0.05, 0.05) errors.

			Sample	Assuming Independence			Accounting for Dependence			
ϕ	θ	n	Quantile	SLVM	SCOTT	BOOT	ADJ-SLVM	ADJ-SCOTT	MB-BOOT	RB-BOOT
0.1	−0.9	100	6.74	−0.02	**−0.01**	0.50	−0.02	−0.01	−9.24	−9.18
		500	0.23	−0.02	**−0.01**	−0.83	−0.03	**−0.01**	−1.81	−1.82
		1000	2.61	−1.13	−1.12	**0.58**	−1.14	−1.12	−0.97	−0.98
		5000	1.09	−0.38	−0.38	0.93	−0.38	−0.38	**−0.37**	**−0.37**
	−0.5	100	6.52	**−0.05**	**−0.05**	0.20	**−0.05**	**−0.05**	−9.26	−9.21
		500	0.04	−0.04	**−0.03**	−0.66	−0.04	**−0.03**	−1.82	−1.84
		1000	2.38	−1.15	−1.14	**0.41**	−1.15	−1.14	−0.97	−0.96
		5000	0.87	−0.37	**−0.36**	0.61	−0.37	**−0.36**	**−0.36**	**−0.36**
	0.1	100	34.51	28.43	28.43	28.92	28.43	28.43	**19.27**	**19.30**
		500	28.07	28.55	28.56	27.28	28.55	28.56	**26.72**	**26.72**
		1000	30.35	**27.64**	27.68	28.54	27.65	27.68	27.66	27.66
		5000	28.86	**27.70**	27.86	28.73	27.71	27.87	28.25	28.25
0.5	−0.9	100	6.88	**0.13**	**0.13**	0.55	**0.13**	**0.13**	−9.07	−9.00
		500	−0.17	0.14	0.14	−0.72	**0.13**	0.14	−1.74	−1.74
		1000	2.75	−0.88	−0.84	**0.32**	−0.92	−0.88	−0.90	−0.87
		5000	0.94	−0.23	**−0.22**	0.79	−0.26	−0.24	−0.32	−0.31
	−0.1	100	4.40	−1.22	−1.22	**−0.53**	−1.22	−1.21	−9.93	−9.83
		500	−1.30	**−0.24**	**−0.24**	−2.78	**−0.24**	**−0.24**	−2.16	−2.17
		1000	**0.23**	−0.46	−0.30	−1.28	−0.36	−0.30	−1.12	−1.14
		5000	−0.79	−0.31	−0.29	−0.85	−0.29	**−0.27**	−0.45	−0.45
	0.5	100	30.65	25.41	25.48	29.51	25.52	25.61	**19.25**	19.38
		500	28.42	34.04	34.06	**27.44**	34.08	34.12	32.69	32.72
		1000	**28.43**	35.58	35.56	28.45	35.67	35.56	34.67	34.66
		5000	**29.46**	35.66	35.66	30.09	35.66	35.67	35.35	35.37
0.9	−0.5	100	**0.27**	−3.86	−3.83	−4.68	−3.71	−3.62	−12.05	−11.65
		500	−1.73	−1.32	−1.30	−3.76	−1.22	**−1.17**	−3.41	−3.41
		1000	−1.91	−0.68	−0.65	−2.16	−0.52	**−0.47**	−1.94	−1.96
		5000	−2.72	−0.32	−0.32	−2.73	−0.28	**−0.26**	−0.64	−0.66
	−0.1	100	−16.23	−17.18	−17.03	−17.57	−15.43	**−14.56**	−23.54	−22.91
		500	−14.62	−3.73	−3.67	−14.33	−3.06	**−2.76**	−8.16	−8.08
		1000	−11.94	−1.34	−1.30	−11.77	−0.93	**−0.73**	−4.19	−4.15
		5000	−7.93	−0.64	−0.62	−7.72	−0.43	**−0.33**	−1.12	−1.12
	0.9	100	30.10	28.62	29.21	28.48	35.66	38.39	**17.66**	19.02
		500	39.44	60.10	60.22	**38.88**	63.17	64.50	50.47	50.61
		1000	**42.34**	65.06	65.15	43.89	67.18	68.09	59.26	59.31
		5000	**52.90**	66.76	66.80	53.31	67.73	68.15	65.80	65.79

Standard Error: ARMA(1, 1) Series

Standard error of each estimator for the 99.5th percentile of the sampling distribution using F errors with 5 and $n-5$ degrees of freedom.

			Sample	Assuming Independence			Accounting for Dependence			
ϕ	θ	n	Quantile	SLVM	SCOTT	BOOT	ADJ-SLVM	ADJ-SCOTT	MB-BOOT	RB-BOOT
0.1	−0.9	100	2.76	3.57	3.51	2.41	3.59	3.53	2.42	2.43
		500	1.35	1.37	1.34	1.24	1.37	1.34	1.24	1.24
		1000	1.03	0.98	0.96	0.94	0.98	0.97	0.94	0.94
		5000	0.48	0.45	0.44	0.45	0.45	0.44	0.45	0.45
	−0.5	100	2.77	3.65	3.60	2.43	3.67	3.62	2.44	2.44
		500	1.33	1.37	1.35	1.22	1.38	1.35	1.23	1.23
		1000	1.04	1.00	0.98	0.93	1.00	0.98	0.93	0.93
		5000	0.46	0.44	0.44	0.44	0.44	0.44	0.44	0.44
	0.1	100	2.86	3.73	3.68	2.52	3.72	3.67	2.52	2.53
		500	1.36	1.41	1.39	1.26	1.41	1.38	1.26	1.26
		1000	1.06	1.04	1.02	0.97	1.04	1.02	0.97	0.97
		5000	0.47	0.44	0.44	0.44	0.44	0.44	0.44	0.44
0.5	−0.9	100	2.77	3.64	3.58	2.41	3.69	3.65	2.42	2.44
		500	1.33	1.37	1.34	1.22	1.39	1.37	1.21	1.22
		1000	1.03	0.99	0.97	0.93	1.01	0.99	0.93	0.94
		5000	0.47	0.45	0.44	0.45	0.45	0.45	0.45	0.45
	−0.1	100	3.03	3.80	3.75	2.72	3.76	3.70	2.69	2.70
		500	1.44	1.47	1.45	1.35	1.45	1.43	1.34	1.34
		1000	1.10	1.08	1.07	1.03	1.07	1.05	1.02	1.03
		5000	0.49	0.47	0.46	0.47	0.46	0.46	0.47	0.47
	0.5	100	4.35	4.61	4.56	3.89	4.58	4.55	3.67	3.70
		500	2.16	2.22	2.19	2.09	2.17	2.14	2.03	2.03
		1000	1.60	1.58	1.56	1.54	1.55	1.53	1.51	1.51
		5000	0.74	0.70	0.69	0.71	0.69	0.68	0.70	0.70
0.9	−0.5	100	3.92	4.47	4.43	3.67	4.52	4.71	3.65	3.71
		500	1.92	1.94	1.92	1.86	1.94	1.97	1.76	1.77
		1000	1.47	1.45	1.43	1.41	1.41	1.40	1.36	1.36
		5000	0.65	0.63	0.63	0.64	0.61	0.61	0.63	0.63
	−0.1	100	6.01	6.39	6.46	5.81	7.66	8.24	5.50	5.63
		500	3.54	3.52	3.49	3.44	3.73	3.96	3.09	3.13
		1000	2.79	2.75	2.72	2.72	2.65	2.68	2.51	2.51
		5000	1.27	1.23	1.23	1.25	1.19	1.19	1.22	1.22
	0.9	100	12.37	13.18	13.42	12.03	16.40	17.76	11.26	11.54
		500	7.74	7.66	7.58	7.50	7.24	7.49	6.68	6.79
		1000	6.18	6.08	6.01	6.00	5.56	5.76	5.49	5.52
		5000	2.81	2.74	2.72	2.77	2.57	2.56	2.70	2.69

Standard Error: ARMA(1, 1) Series

Standard error of each estimator for the 99.5th percentile of the sampling distribution using F errors with 5 and 20 degrees of freedom.

			Sample	Assuming Independence			Accounting for Dependence			
ϕ	θ	n	Quantile	SLVM	SCOTT	BOOT	ADJ-SLVM	ADJ-SCOTT	MB-BOOT	RB-BOOT
0.1	−0.9	100	4.96	6.94	6.88	4.33	6.95	6.90	4.33	4.34
		500	2.65	2.89	2.85	2.47	2.90	2.86	2.48	2.48
		1000	1.98	2.02	2.00	1.86	2.03	2.01	1.87	1.86
		5000	0.93	0.91	0.90	0.90	0.91	0.90	0.90	0.90
	−0.5	100	4.97	7.01	6.96	4.31	7.02	6.97	4.34	4.36
		500	2.65	2.91	2.88	2.50	2.91	2.88	2.49	2.49
		1000	1.98	2.03	2.01	1.87	2.03	2.02	1.87	1.87
		5000	0.95	0.92	0.92	0.90	0.92	0.92	0.90	0.91
	0.1	100	5.04	7.06	7.02	4.42	7.06	7.01	4.42	4.44
		500	2.69	2.92	2.89	2.55	2.91	2.88	2.55	2.54
		1000	2.08	2.09	2.08	1.92	2.09	2.08	1.91	1.91
		5000	0.96	0.94	0.93	0.91	0.94	0.93	0.91	0.91
0.5	−0.9	100	4.92	6.95	6.90	4.27	6.99	6.95	4.29	4.32
		500	2.63	2.91	2.88	2.47	2.94	2.92	2.49	2.48
		1000	1.99	2.02	2.00	1.84	2.03	2.02	1.85	1.85
		5000	0.94	0.91	0.91	0.89	0.92	0.91	0.89	0.89
	−0.1	100	5.28	7.10	7.05	4.69	7.06	7.00	4.62	4.65
		500	2.83	3.03	3.00	2.69	3.01	2.97	2.66	2.67
		1000	2.20	2.19	2.18	2.05	2.18	2.16	2.03	2.03
		5000	1.01	0.98	0.98	0.97	0.98	0.97	0.96	0.96
	0.5	100	7.39	7.74	7.69	6.42	7.68	7.62	5.97	6.00
		500	4.19	4.42	4.38	4.15	4.35	4.30	4.03	4.03
		1000	3.23	3.27	3.25	3.12	3.23	3.20	3.11	3.12
		5000	1.48	1.45	1.45	1.44	1.44	1.43	1.43	1.43
0.9	−0.5	100	6.15	7.59	7.53	5.70	7.31	7.36	5.72	5.79
		500	3.35	3.52	3.49	3.30	3.40	3.38	3.16	3.16
		1000	2.65	2.65	2.63	2.53	2.56	2.54	2.44	2.46
		5000	1.26	1.23	1.23	1.23	1.21	1.20	1.21	1.21
	−0.1	100	8.78	9.34	9.40	8.43	10.64	11.31	7.94	8.17
		500	6.04	6.17	6.13	5.88	6.03	6.13	5.27	5.31
		1000	4.74	4.82	4.79	4.69	4.59	4.53	4.38	4.38
		5000	2.32	2.29	2.29	2.30	2.23	2.20	2.27	2.27
	0.9	100	17.66	18.66	18.90	17.13	22.24	23.85	15.97	16.40
		500	13.09	13.20	13.08	12.55	11.82	11.84	11.12	11.25
		1000	10.24	10.36	10.29	10.12	9.53	9.64	9.31	9.34
		5000	5.07	4.99	4.96	5.01	4.65	4.57	4.91	4.90

Standard error of each estimator for the 99.5th percentile of the sampling distribution with Gamma(0.05, 0.05) errors.

			Sample	Assuming Independence			Accounting for Dependence			
ϕ	θ	n	Quantile	SLVM	SCOTT	BOOT	ADJ-SLVM	ADJ-SCOTT	MB-BOOT	RB-BOOT
0.1	−0.9	100	12.16	17.71	17.72	10.51	17.71	17.71	10.54	10.60
		500	7.45	8.67	8.67	7.14	8.66	8.67	7.14	7.12
		1000	5.81	5.81	5.81	5.18	5.81	5.81	5.16	5.17
		5000	2.76	2.75	2.75	2.62	2.75	2.75	2.62	2.62
	−0.5	100	12.20	17.72	17.72	10.58	17.71	17.72	10.61	10.65
		500	7.48	8.67	8.67	7.14	8.67	8.67	7.17	7.13
		1000	5.85	5.85	5.85	5.20	5.85	5.85	5.20	5.19
		5000	2.78	2.78	2.78	2.64	2.78	2.78	2.65	2.64
	0.1	100	12.51	17.80	17.80	10.90	17.80	17.80	10.86	10.86
		500	7.62	8.87	8.87	7.30	8.87	8.87	7.31	7.30
		1000	5.89	5.89	5.90	5.23	5.89	5.90	5.22	5.24
		5000	2.79	2.76	2.77	2.67	2.76	2.77	2.66	2.67
0.5	−0.9	100	12.16	17.71	17.71	10.50	17.71	17.71	10.56	10.61
		500	7.42	8.68	8.68	7.11	8.68	8.68	7.12	7.12
		1000	5.80	5.82	5.83	5.16	5.82	5.83	5.16	5.18
		5000	2.77	2.77	2.77	2.62	2.77	2.77	2.63	2.63
	−0.1	100	13.23	17.93	17.93	11.72	17.93	17.93	11.40	11.49
		500	8.06	9.23	9.23	7.97	9.23	9.23	7.82	7.83
		1000	5.95	6.05	6.13	5.49	6.06	6.08	5.43	5.43
		5000	2.86	2.87	2.87	2.78	2.87	2.87	2.77	2.76
	0.5	100	18.22	18.93	18.94	15.62	18.94	18.95	14.21	14.29
		500	12.88	13.51	13.50	12.42	13.50	13.50	11.93	11.95
		1000	9.08	9.14	9.21	8.33	9.26	9.18	8.27	8.28
		5000	4.31	4.34	4.35	4.18	4.34	4.34	4.13	4.13
0.9	−0.5	100	13.96	18.40	18.40	12.84	18.38	18.36	12.43	12.69
		500	9.01	10.05	10.04	9.14	9.99	9.96	8.66	8.66
		1000	6.63	6.78	6.77	6.37	6.76	6.75	6.09	6.09
		5000	3.33	3.34	3.34	3.26	3.32	3.32	3.21	3.21
	−0.1	100	17.82	19.57	19.57	16.98	19.97	20.24	15.32	15.79
		500	15.61	16.30	16.28	15.36	16.18	16.13	13.63	13.68
		1000	11.93	12.36	12.36	11.76	12.28	12.23	10.82	10.81
		5000	5.67	5.71	5.71	5.65	5.67	5.65	5.59	5.59
	0.9	100	34.27	36.10	36.25	32.92	38.28	39.26	29.34	30.39
		500	32.20	33.47	33.43	31.43	32.75	32.52	27.55	27.73
		1000	25.16	25.93	25.93	24.77	25.44	25.25	22.45	22.47
		5000	11.90	12.01	12.00	11.90	11.86	11.80	11.74	11.75

Simulation Study: Summary

Estimator performance for each dependence structure and error distribution.

Estimator	AR(1)		ARMA(1, 1)	
	F	Gamma	F	Gamma
Sample Quantile				
SLVM	✓		✓	
SCOTT	✓		✓	✓
BOOT	✓			
ADJ-SLVM			✓	
ADJ-SCOTT		✓		✓
MB-BOOT		✓	✓	
RB-BOOT		✓	✓	

✓: Best Overall

✓: Good for weak dependence (but not strong)

✓: Good for strong dependence (but not weak)

Simulation Study: Summary

AR(1) Results

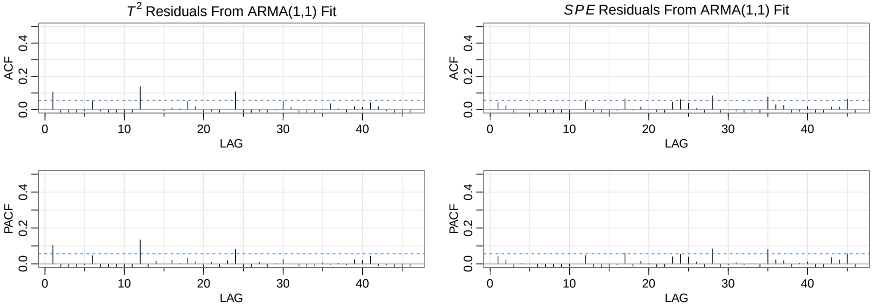
- F errors: SLVM and SCOTT are the best overall, and BOOT performs poorly when n is small or ϕ is large.
- Gamma errors: ADJ-SCOTT is the best, and dependent bootstraps perform better than BOOT as ϕ increases.

ARMA(1, 1) Results

- F errors: ADJ-SLVM is the best for weak dependence, but MB-BOOT and RB-BOOT are the best for strong dependence.
- Gamma errors: SCOTT and ADJ-SCOTT are the best, with SLVM and ADJ-SLVM performing similarly.

Case Study

Assessing Dependence in the Monitoring Statistics



ACF and PACF plots for the residuals of T^2 (left) and *SPE* (right) from an ARMA(1, 1) fit during the training period.

Case Study

AR(1) and ARMA(1,1) Thresholds

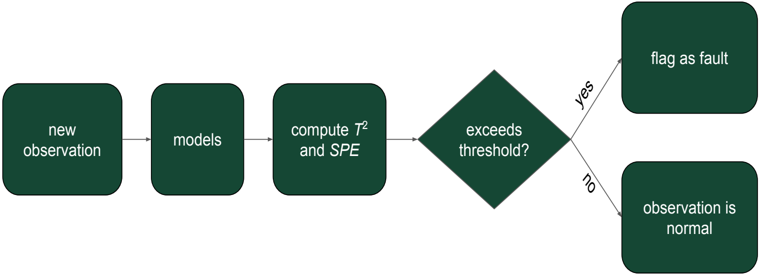
All nonparametric thresholds for T^2 and SPE for the wastewater treatment case study.

Monitoring Statistic	Correlation Structure	SLVM	SCOTT	BOOT	ADJ-SLVM	ADJ-SCOTT	MB-BOOT	RB-BOOT
T^2	AR(1)	61.0	61.3	58.3	61.2	61.3	57.8	58.2
	ARMA(1,1)	61.0	61.3	58.3	61.4	61.3	57.4	57.4
SPE	AR(1)	17.3	17.4	19.1	17.4	17.5	19.4	19.2
	ARMA(1,1)	17.3	17.4	19.1	17.6	17.7	18.9	18.8

MSPM Steps - Outline

To use MSPM:

1. Determine in-control training period
2. Detrend monitoring variables
3. Apply PCA
4. Compute T^2 and SPE (training period)
5. Determine threshold (upper control limit)
6. Monitor new observations



Multivariate Statistical Process Monitoring (MSPM)

1. Determine in-control training period
2. Detrend the monitoring variables
   - Detrending methods:
     - no model fit (raw data)
     - adaptive lasso
     - k-nearest neighbors
     - random forest
     - extreme gradient boosting
   - Obtain residuals from model fits

Lasso

Recall the standard lasso technique (Tibshirani, 1996), which produces the estimates

$$\hat{\boldsymbol{\beta}}_{\text{lasso}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\| \mathbf{y} - \sum_{j=1}^{p} \mathbf{x}_j \beta_j \right\|^2 + \lambda \sum_{j=1}^{p} |\beta_j|,$$

where

$\boldsymbol{\beta}$: coefficient vector
$\mathbf{y}$: response vector
$\mathbf{x}_j$: jth explanatory variable
λ: regularization parameter

Adaptive Lasso

Pick $\gamma > 0$, and define $\hat{\mathbf{w}} = 1/|\hat{\boldsymbol{\beta}}|^\gamma$. Then we obtain the adaptive lasso (Zou, 2006) estimates

$$\hat{\boldsymbol{\beta}}_{\text{adaptive}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \left\| \mathbf{y} - \sum_{j=1}^{p} \mathbf{x}_j \beta_j \right\|^2 + \lambda \sum_{j=1}^{p} \hat{w}_j |\beta_j|.$$

For $\hat{\mathbf{w}}$, we use $1/|\hat{\boldsymbol{\beta}}_{\text{ridge}}|$, where

$$\hat{\boldsymbol{\beta}}_{\text{ridge}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \left\| \mathbf{y} - \sum_{j=1}^{p} \mathbf{x}_j \beta_j \right\|^2 + \lambda \sum_{j=1}^{p} \beta_j^2.$$

MSPM (Cont.)

3. Apply Principal Component Analysis (PCA) to residuals from detrending

- Apply eigenvalue decomposition to the residual correlation matrix $\mathbf{R}$ to obtain

$$\mathbf{R} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$$

where $\mathbf{P}$ is the projection matrix, and $\boldsymbol{\Lambda}$ is the diagonal matrix of the eigenvalues of $\mathbf{R}$.

Keep $\mathbf{P}_k$ and $\boldsymbol{\Lambda}_k$, the first k columns of $\mathbf{P}$ and $\boldsymbol{\Lambda}$, where k is the # of principal components required to capture 80% of the variability.

MSPM (Cont.)

4. Compute monitoring statistics

Hotelling's T^2 is computed as

$$\begin{aligned} T^2 &= (\mathbf{x}_t^T \mathbf{P}_k)(\boldsymbol{\Lambda}_k)^{-1}(\mathbf{x}_t^T \mathbf{P}_k)^T \\ &= \mathbf{y}_t^T (\boldsymbol{\Lambda}_k)^{-1} \mathbf{y}_t, \end{aligned}$$

and SPE can computed as

$$SPE = \left\| \left[\mathbf{I} - \mathbf{P}_k (\mathbf{P}_k)^T \right] \mathbf{x}_t \right\|^2 .$$

MSPM (Cont.)

5. Determine thresholds for T^2 and SPE from training data
   - Use an upper quantile of computed T^2 and SPE values as the threshold to classify future observations as normal or abnormal (fault)
6. Monitor new observations