

Robust Self-starting Bayesian Control Charts for Multivariate Phase II Observations - Monitoring

Taylor R. Grimm¹, Kathryn B. Newhart², and Amanda S. Hering^{1,*}

September 3, 2024

Abstract

Traditional multivariate control charts require in-control (IC) parameter estimates to be known or estimated from a large set of uncontaminated, historical Phase I observations. Contamination in Phase I data, such as outliers, can result in poor chart performance. In such cases, control charts based on robust estimators of IC parameters are often used to obtain satisfactory fault detection performance. However, some processes need to be monitored when little Phase I data are available prior to monitoring, in which case self-starting approaches, including Bayesian methods, have proven useful. Even so, self-starting methods are sensitive to early outliers or contamination and can quickly be degraded. Few, if any, methods exist to handle situations wherein Phase I data are both limited and contaminated by outliers. Therefore, we propose a robust self-starting Bayesian approach to detect faults in multivariate Phase II data when Phase I data are both limited and contaminated by outliers. The proposed approach is found to substantially outperform existing robust and self-starting methods in several Phase II scenarios, and its application is demonstrated on data from a water treatment process.

Is it just the robust part that is now?

mention simulation studies

repetitive repetitive

put in alphabetical order

update

Keywords: statistical process monitoring · fault detection · robust estimation · self-starting charts · Bayesian methods

Short title: Robust Self-Starting Bayesian Control Chart

¹Department of Statistical Science, Baylor University, One Bear Place 97140, Waco, TX 76798, USA.

²Department of Geography and Environmental Engineering, United States Military Academy, 646 Swift Road, West Point, NY 10996, USA.

*Corresponding Author: mandy_hering@baylor.edu

1 Introduction

A popular approach to detect faults in complex processes in real-time is to use multivariate statistical process monitoring (MSPM) methods, including popular control charts such as Hotelling's T^2 (Hotelling, 1947), multivariate cumulative sum (MCUSUM) (Woodall and Ncube, 1985; Healy, 1987; Crosier, 1988; Pignatiello and Runger, 1990), and multivariate exponentially weighted moving average (MEWMA) (Lowry et al., 1992; Testik et al., 2003) charts. These control charts require the true in-control (IC) process parameters, such as means and variances, to be known. If the IC parameters are not known, they must be estimated using historical process data ~~that have been determined to be stable (IC)~~. This phase of MSPM is referred to as Phase I analysis and is followed by Phase II monitoring, where the process is monitored as new observations are collected, and a signal is given and the process is deemed out-of-control (OC) if any faults (i.e., deviations from IC behavior) are detected.

Although historical data are often available to estimate IC parameters, it is common for these data to be contaminated with outliers. Especially in processes such as water and wastewater treatment, data are often susceptible to contamination due to biased sensors, sensor failure, measurement error, electrical surges, or sudden changes in operation or influent quality. When contaminated data are used to estimate IC parameters, control chart performance can be negatively impacted, causing faults to go undetected during Phase II (Woodall and Montgomery, 1999; Jensen et al., 2006; Zwetsloot et al., 2014). Furthermore, when parameters are estimated using data that contain outliers, the presence of outliers can go undetected. This phenomenon is known as the "masking effect". To address contamination in the historical, ~~(Phase I)~~ dataset, classical estimators of the IC parameters are often replaced with robust estimators, such as the minimum covariance determinant (MCD) or minimum volume ellipsoid (MVE). Using these robust estimators in place of classical estimators has led to improved control chart performance when Phase I data are contaminated with outliers. For instance, robust T^2 charts based on these estimators have been thoroughly studied and have been shown to outperform the classical T^2 chart in the presence of contaminated historical data (e.g., Vargas (2003); Jensen et al. (2007);

remove parentheses here & elsewhere
in this situation

Chenouri et al. (2009); Alfaro and Ortega (2009)). Furthermore, the MCUSUM and MEWMA charts based on the MVE and MCD have also been shown to improve fault detection performance in the presence of contamination (Midi and Shabbak, 2011). While both the MVE and MCD estimators have been studied in-depth, the MCD is often preferred over the MVE because it is easier to compute and has greater asymptotic efficiency (Woodruff and Rocke, 1994; Rocke and Woodruff, 1996; Van Aelst and Rousseeuw, 2009).

In some cases, historical data are not available for parameter estimation in Phase I. This may occur when: 1) a process is new and needs to be monitored immediately; 2) the process has substantially changed since previous data were collected, making historical data unrepresentative of the current process; or 3) obtaining a large sample prior to monitoring is too expensive. In such situations, the IC parameters are unknown and cannot be estimated due to the lack of clean, IC historical data. To address this problem, many self-starting methods have been proposed. These methods allow monitoring to begin when very limited historical data are available by effectively combining Phase I analysis and Phase II monitoring. Instead of first performing a Phase I analysis, estimating IC parameters, and then monitoring observations in a separate Phase II analysis, self-starting methods simultaneously estimate IC parameters while checking for process stability (i.e., monitoring each observation for IC/OC behavior).

Self-starting control charts often work by taking observations that are drawn from an unknown distribution and transforming them to follow a known distribution, such as the standard normal distribution (e.g., Hawkins (1987); Quesenberry (1991); Hawkins and Maboudou-Tchao (2007); Li et al. (2010); Xie and Qiu (2022)). Then, control limits are set based on the known distribution, and IC parameter estimates are updated as new observations are collected. The majority of self-starting charts assume that observations are independent and identically distributed from an unknown normal distribution. For example, the self-starting MEWMA (SS-MEWMA) of Hawkins and Maboudou-Tchao (2007) recursively transforms observations from an unknown multivariate normal distribution to a multivariate standard normal distribution through a two-stage transformation procedure, and a classical MEWMA chart is then applied. Some nonparametric self-starting methods have also been proposed to handle nonnormal data (e.g., Zou

et al. (2012); Zhou et al. (2015); Li et al. (2017)). For example, Zou et al. (2012) proposes a self-starting MEWMA based on spatial ranks for detecting shifts in location. Their method generally outperforms the SS-MEWMA of Hawkins and Maboudou-Tchao (2007) when observations are nonnormal, but SS-MEWMA performs better under normality. Other nonparametric self-starting charts are also designed to accommodate serial dependence in the observations, such as the methods of Xue and Qiu (2021) and Xie and Qiu (2024). While the methods of Xue and Qiu (2021) and Xie and Qiu (2024) are self-starting in the sense that they update IC parameter estimates over time, they also require large IC Phase I samples to achieve stable performance.

*Btw you said
that SS, method
do not require a
separate Phase I
sample. This
statement seems
to contradict
that one.*

A natural approach to the self-starting idea of updating IC parameter estimates is to use Bayesian methods, obtaining updated IC parameter estimates using posterior distribution updates over time. One additional benefit of Bayesian methods is that they allow for seamless integration of any available historical data or expert knowledge through prior distributions. However, few Bayesian methods exist for monitoring multivariate processes. *(Kim et al., 2022). with a few exceptions described here!*

Feltz and Shiau (2001) propose a multivariate empirical Bayesian method for monitoring the mean vector. Parameter estimates are updated as posterior means are computed, with newer observations being given more weight than older observations, and observations are labeled as IC or OC using a quadratic form of the posterior mean at each point in time. More recent multivariate Bayesian approaches have been proposed by Kim and Turkoz (2022) and Kim et al. (2022), both of which can be used to monitor high-dimensional data with minimal computational burden. However, the methods of Kim and Turkoz (2022) and Kim et al. (2022) assume that the IC mean vector is known, which is rarely true. *in practice.*

Although self-starting methods have been shown to effectively detect faults, their performance depends heavily on early IC parameter estimates (Keefe et al., 2015). As a result, any early outliers or faults can negatively affect parameter estimates and result in poor chart performance (Jones-Farmer et al., 2014). Additionally, existing robust methods are unable to update and improve parameter estimates over time, which can be especially useful if the initial Phase I period is small and/or substantially contaminated. Thus, while many methods exist for separately handling the problems of contamination *(robust methods)* and limited historical data *with estimation* *not designed*

(self-starting methods), there is a lack of robust self-starting Phase II control charts that are both robust to Phase I contamination and allow for monitoring to begin using minimal Phase I data with automatic updates to improve IC parameter estimates. Such a method is necessary for processes that are susceptible to outliers and must be monitored shortly after start-up or after substantial changes have occurred in process operation or its environment.

The purpose of this paper is to propose and evaluate a multivariate Bayesian Phase II control chart that: 1) is robust to contamination during Phase I; 2) allows for IC parameter estimates to be updated automatically; and 3) can be used shortly after process start-up. We also compare the performance of the proposed chart to various existing robust and self-starting MSPM methods. The layout of this paper is as follows: We first describe popular robust estimators, introduce the proposed method, and describe various existing MSPM methods in Section 2. Then, we evaluate the performance of each method under multiple various Phase II scenarios via a simulation study in Section 3, apply each method to data from an ultrafiltration water treatment process in Section 4, and conclude in Section 5.

2 Methods

→ needs a bit describing layout of this section.

2.1 Reweighted MCD Estimator

The robust estimator used in this paper is the reweighted MCD (RMCD) estimator. The RMCD estimator is a popular robust estimator of multivariate location and scale. RMCD is a more efficient extension of the MCD estimator that retains the properties of a high breakdown point and affine equivariance (Rousseeuw and Leroy, 1987; Lopuhaä, 1999; Hubert et al., 2018). The MCD estimator finds the subset of ℓ observations that minimizes the determinant of the sample covariance matrix. The MCD estimate of the mean, $\hat{\mu}_{\text{MCD}}$, is the sample mean vector of the resulting ℓ observations, and the MCD estimate of the covariance matrix, $\hat{\Sigma}_{\text{MCD}}$, is the sample covariance matrix of the ℓ observations, multiplied by a constant for consistency (Hubert and Debruyne, 2010). Because there are many possible subsets of ℓ observations, especially for large datasets, the FAST-MCD algorithm of Rousseeuw and Driessen (1999) is typically used to obtain

the MCD estimates. It is important to note that the MCD is only recommended if there are at least at least $5p$ observations in the Phase I dataset, where p is the dimension of the data (Hubert et al., 2018).

The RMCD estimators of the mean vector and covariance matrix can be viewed as one-step M-estimators that use the MCD estimates as initial values (Rousseeuw and Hubert, 2013). Similar to the MCD estimator, the RMCD estimator achieves its highest breakdown point when $\ell = (m_0 + p + 1)/2$, where m_0 is the number of Phase I observations. Let \mathbf{x}_t represent the observed values of p process variables at time t . The RMCD estimates, $\hat{\mu}_{\text{RMCD}}$ and $\hat{\Sigma}_{\text{RMCD}}$, are obtained by first computing the Mahalanobis distances

$$D_t = D(\mathbf{x}_t, \hat{\mu}_{\text{MCD}}, \hat{\Sigma}_{\text{MCD}}) = \sqrt{(\mathbf{x}_t - \hat{\mu}_{\text{MCD}})' \hat{\Sigma}_{\text{MCD}}^{-1} (\mathbf{x}_t - \hat{\mu}_{\text{MCD}})} \quad (1)$$

for all $t \in \{1, 2, \dots, m_0\}$, and then computing the RMCD estimators are

$$\hat{\mu}_{\text{RMCD}} = \frac{\sum_{t=1}^{m_0} W(D_t^2) \mathbf{x}_t}{\sum_{t=1}^{m_0} W(D_t^2)} \quad (2)$$

$$\text{and } \hat{\Sigma}_{\text{RMCD}} = c_{\gamma, p} d_{m_0, p} \frac{\sum_{t=1}^{m_0} W(D_t^2) (\mathbf{x}_t - \hat{\mu}_{\text{RMCD}}) (\mathbf{x}_t - \hat{\mu}_{\text{RMCD}})'}{\sum_{t=1}^{m_0} W(D_t^2)}, \quad (3)$$

where $W(D_t^2)$ is a weight function typically chosen to be $W(D_t^2) = I(D_t^2 < \chi_{p, 0.975}^2)$, where $I(\cdot)$

- > is the indicator function, and $\chi_{p, 0.975}^2$ is the 0.975 quantile of the χ^2 -distribution with p degrees of freedom (Rousseeuw and Driessen, 1999). In Equation 3, c_p and $d_{m_0, p}$ are consistency and finite sample correction factors, respectively, given by Pison et al. (2002) and described by Willems et al. (2002) and Chenouri et al. (2009). RMCD estimates are easy to obtain in most modern statistical software packages and are the default option of some functions in R (R Core Team, 2022), such as the `covMcd()` function in the `robustbase` package (Todorov and Filzmoser, 2009).

↓ here p is
still the
of
variables?

2.2 Robust Self-starting Bayesian Charts

Need to revisit the notation I use in this section for the hierarchical model. Currently, the model technically shows that each observation \mathbf{x}_t can have its own mean and covariance matrix,

dependent on t . However, we usually technically assume that all Phase II observations have the same mean and covariance.

We propose a robust self-starting Bayesian (RSSB) approach for monitoring Phase II observations. This method assumes that the Phase II observations \mathbf{x}_t , $t \in \{1, \dots, n\}$, independently follow a p -dimensional multivariate normal (MVN) distribution, denoted N_p , with conjugate priors. In particular, the MVN mean μ_t is assigned a p -dimensional MVN conditional prior distribution, and the covariance matrix Σ_t is assigned an inverse Wishart (IW) prior distribution. The model is given by

$$\mathbf{x}_t \sim N_p(\mu_t, \Sigma_t)$$

$$\mu_t | \Sigma_t \sim N_p(\hat{\mu}_{t-1}, \Sigma_t / k_{t-1})$$

$$\Sigma_t \sim IW_{\nu_{t-1}}(\Lambda_{t-1}),$$

so what are
 k_{t-1} , ν_{t-1} , Λ_{t-1}
defined to be?
These appear to
come later, so
you need
to state
that here.

where $\hat{\mu}_0 = \hat{\mu}_{\text{RMCD}}$; $k_0 = m_0$; $\nu_0 = m_0 - 1$; and $\Lambda_0 = (m_0 - 1)\hat{\Sigma}_{\text{RMCD}}$. In this way, the prior hyperparameters for the first Phase II observation, \mathbf{x}_1 , are selected so that the prior means for μ_1 and Σ_1 are equal to the RMCD estimates of the Phase I mean vector and covariance matrix, where the Phase I sample is of size m_0 . If the true IC mean vector and/or covariance matrix are known, then the prior hyperparameters can be manually set to reflect those values. As new observations \mathbf{x}_t are collected, the values of $\hat{\mu}_t$, Λ_t , k_t , and ν_t are determined by the following process: for each observation at time t in Phase II, a monitoring statistic, such as a T^2 or MEWMA statistic, is calculated. The computation of these monitoring statistics is shown in Equations (4)-(7) in Section 2.3. If the monitoring statistic does not exceed the control limit at time t , then the estimates of the mean vector and covariance matrix at time t are updated by computing the posterior means of μ_t and Σ_t . If the monitoring statistic exceeds the control limit at time t , then the estimates of the mean vector and covariance matrix are not updated.

Note that because the priors for μ_t and Σ_t are conjugate, the marginal posterior distributions are easily obtained in closed form (Gelman et al., 2013). Specifically, the marginal posterior distribution of $\mu_t | \mathbf{x}_t$ is a multivariate t -distribution with mean μ_t , covariance matrix $\Lambda_t / (k_t(\nu_t -$

$p+1)$, and $\nu_t - p + 1$ degrees of freedom. The posterior distribution of $\Sigma_t | \mathbf{x}_t$ is an inverse Wishart distribution with scale matrix Λ_t and ν_t degrees of freedom. The means of these distributions are given by $E(\mu_t) = \hat{\mu}_t$ and $E(\Sigma_t) = \hat{\Sigma}_t = \Lambda_t / \nu_t$, respectively. Hyperparameters for the posterior distributions are given by

$$\begin{aligned}\hat{\mu}_t &= \frac{k_{t-1}}{k_{t-1} + 1} \hat{\mu}_{t-1} + \frac{1}{k_{t-1} + 1} \mathbf{x}_t \\ \Lambda_t &= \Lambda_{t-1} + \frac{k_{t-1}}{k_{t-1} + 1} (\mathbf{x}_t - \hat{\mu}_{t-1})(\mathbf{x}_t - \hat{\mu}_{t-1})' \\ k_t &= k_{t-1} + 1 \\ \nu_t &= \nu_{t-1} + 1.\end{aligned}$$

Incorporate the step-by-step into the RSSB method.

For convenience, the RSSB method is described step-by-step in Algorithm 1. In this work, we apply the RSSB based on Hotelling's T^2 (RSSB- T^2) and MEWMA (RSSB-MEWMA) monitoring statistics.

Algorithm 1 Robust Self-starting Bayesian Control Chart

- 1: Obtain initial estimates $\hat{\mu}_0 = \hat{\mu}_{\text{RMCD}}$ and $\hat{\Sigma}_0 = \hat{\Sigma}_{\text{RMCD}}$ from the Phase I data.
 - 2: Set $k_0 = m_0$ and $\nu_0 = m_0 - 1$.
 - 3: **for** $t \in \{1, \dots, n\}$ **do**
 - 4: Compute monitoring statistic using $\hat{\mu}_{t-1}$ and $\hat{\Sigma}_{t-1}$ as IC parameter estimates.
 - 5: **if** monitoring statistic < control limit **then**
 - 6: *Process is IC, update parameter estimates:*
 - 7: Set $k_t = k_{t-1} + 1$ and $\nu_t = \nu_{t-1} + 1$.
 - 8: Compute $\hat{\mu}_t$ and $\hat{\Sigma}_t$ using the posterior means of μ_t and Σ_t :
 - 9:
$$\hat{\mu}_t = \frac{k_{t-1}}{k_{t-1} + 1} \hat{\mu}_{t-1} + \frac{1}{k_{t-1} + 1} \mathbf{x}_t$$
 - 10:
$$\hat{\Sigma}_t = \frac{1}{\nu_t} \left(\nu_{t-1} \hat{\Sigma}_{t-1} + \frac{k_{t-1}}{k_{t-1} + 1} (\mathbf{x}_t - \hat{\mu}_{t-1})(\mathbf{x}_t - \hat{\mu}_{t-1})' \right)$$
 - 11: **else** *do not*
 - 12: *Process is OC, don't update parameter estimates:*
 - 13: Set $k_t = k_{t-1}$ and $\nu_t = \nu_{t-1}$.
 - 14: Set $\hat{\mu}_t = \hat{\mu}_{t-1}$ and $\hat{\Sigma}_t = \hat{\Sigma}_{t-1}$.
 - 15: **end if**
 - 16: **end for**
-

Explain 2
classical charts,
2 robust charts,
2 SS charts,
are presented
for comparison.

2.3 Existing Control Charts

We compare our robust self-starting control charts with several existing

The existing multivariate control charts evaluated in this paper are the classical MEWMA chart

(Lowry et al., 1992; Testik et al., 2003); a robust MEWMA chart based on RMCD estimators *that is*

~~similar to~~ Midi and Shabbak (2011), who consider the MCD instead of the RMCD; the classical

T^2 chart (Hotelling, 1947); a robust T^2 chart based on RMCD estimators (e.g., Alfaro and Ortega

(2009); Chenouri et al. (2009); Chenouri and Variyath (2011); Cabana and Lillo (2022)); the self-

starting MEWMA (SS-MEWMA) of Hawkins and Maboudou-Tchao (2007); and a self-starting

T^2 (SS- T^2) that is similar to the SS-MEWMA of Hawkins and Maboudou-Tchao (2007). *We describe each of these in the following sections.*

Classical T^2 [Hotelling, 1947]. This method, denoted hereafter as T^2 , *weird to cite here* assumes independence and is sensitive to the assumption of multivariate normality of the data (Alfaro and Ortega,

2012). Classical estimates of the IC mean ($\hat{\mu}$) and covariance matrix ($\hat{\Sigma}$) are computed by

These need to be defined obtaining the sample mean vector and sample covariance matrix of all observations from the

Phase I sample. These estimates are then used to compute the monitoring statistics at each time

point t . The T^2 statistic at each time t is computed by

$$T_t^2 = (\mathbf{x}_t - \hat{\mu})' \hat{\Sigma}^{-1} (\mathbf{x}_t - \hat{\mu}), \quad (4)$$

which only uses the observed values of the process variables at time t . The chart classifies any observations as OC when

$$T_t^2 > \frac{p(m_0^2 - 1)}{m_0(m_0 - p)} F_{\alpha, p, m_0 - p}, \quad (5)$$

where $F_{\alpha, p, m_0 - p}$ is the upper α quantile of the F -distribution with p numerator and $m_0 - p$ denominator degrees of freedom. Because the T^2 values are computed with no accumulation of information over time, the T^2 chart does not excel at detecting accumulations of small changes over time. However, this method is simple, straightforward to implement, and has been shown to perform well *in* detecting large shifts. *Now state assumptions.*

Classical MEWMA [Lowry et al., 1992; Testik et al., 2003]. Similar to the classical T^2 chart, this method, denoted as MEWMA, assumes independence and multivariate normality of the data. This chart uses classical estimates of the Phase I mean and covariance matrix to compute

make these
subsubsections

more to
end

the monitoring statistics at each time point t . The MEWMA monitoring statistics are found by first computing

$$\mathbf{q}_t = \lambda(\mathbf{x}_t - \hat{\boldsymbol{\mu}}) + (1 - \lambda)\mathbf{q}_{t-1}, \quad (6)$$

which is a weighted sum of the observation vector at time t and the previous MEWMA statistic, where the initial MEWMA statistic, \mathbf{q}_0 , is typically set equal to $\mathbf{0}$. The parameter $\lambda \in [0, 1]$ is selected to weight the contribution of past observations to current values of \mathbf{q}_t . Common choices of λ are 0.05, 0.1, and 0.2. Small values of λ result in better detection of small shifts, and when $\lambda = 1$, the MEWMA chart is equivalent to the T^2 chart. The MEWMA chart gives an OC signal when the MEWMA monitoring statistics

$$\mathbf{q}_t' \widehat{\Sigma}_{\mathbf{q}_t}^{-1} \mathbf{q}_t > h_{\text{MEWMA}}, \quad (7)$$

where $\widehat{\Sigma}_{\mathbf{q}_t} = [\lambda(1 - (1 - \lambda)^{2t})/(2 - \lambda)]\widehat{\Sigma}$, which converges to $(\lambda/(2 - \lambda))\widehat{\Sigma}$ asymptotically, and h_{MEWMA} is a control limit selected to ensure a desired false alarm rate. The MEWMA chart has been shown to be robust to the assumption of independence and performs well even for highly skewed and heavy-tailed distributions when λ is small ($\lambda < 0.02$) (Stoumbos and Sullivan, 2002). Furthermore, the MEWMA chart outperforms the T^2 chart at detecting small shifts, but is worse at detecting large shifts.

SS-MEWMA (Hawkins and Maboudou-Tchao, 2007). This method is one of the most commonly used self-starting methods for monitoring multivariate data. This method assumes independence and multivariate normality of the data. The original data matrix, \mathbf{X} , is first transformed into a matrix of recursive residuals, \mathbf{R} , based on regression fits at time t to all previous time points. Then, \mathbf{R} is transformed columnwise by a probability integral transform so that each column in the resulting matrix \mathbf{U} is independent standard normal. These transformations are fully defined once $p + 1$ observations have been collected, allowing for monitoring to begin at observation $p + 2$, where the matrix \mathbf{U} is updated recursively for each new observation. The classical MEWMA charting scheme is then applied to \mathbf{U} . The primary benefit of SS-MEWMA is that it can be implemented almost immediately upon process start-up and allows for quick detection of faults,

How? Needs to be more specific. \mathbf{U} is a matrix. Eq. 6 does not have a matrix in it.

even when minimal historical data exist. However, SS-MEWMA is not designed to accomodate early shifts or outliers, which can quickly contaminate the parameter estimates and degrade the chart's performance.

SS- T^2 . While Hawkins and Maboudou-Tchao (2007) only implement a MEWMA chart based on the transformed data \mathbf{U} , they ^{also} mention that \mathbf{U} can be combined with any other multivariate control chart of choice. To provide a self-starting method for direct comparison to the T^2 -based methods, we ^{construct such} implement a SS- T^2 chart. This method is identical to the SS-MEWMA except that the matrix \mathbf{U} is monitored using the T^2 chart instead of the MEWMA chart. While this method has not yet been studied in ^{the} literature, we suspect based on the well-studied comparative performance of MEWMA and T^2 , we hypothesize that SS- T^2 will be less sensitive to small shifts than SS-MEWMA. explain how specifically Eq 4 does not have a matrix in it.

Robust T^2 (e.g., Alfaro and Ortega (2009); Chenouri et al. (2009); Chenouri and Variyath (2011); Cabana and Lillo (2022)) This method is equivalent to the classical T^2 chart except that the classical estimates of the IC mean and covariance, $\hat{\mu}$ and $\hat{\Sigma}$, are replaced with robust estimates, such as the RMCD estimates. After computing the RMCD estimates from the Phase I data, plugging these estimates into Equation 4 yields the monitoring statistics

$$T_{t,\text{RMCD}}^2 = (\mathbf{x}_t - \hat{\mu}_{\text{RMCD}})' \hat{\Sigma}_{\text{RMCD}}^{-1} (\mathbf{x}_t - \hat{\mu}_{\text{RMCD}}), \quad (8)$$

and observations are classified as OC when $T_{t,\text{RMCD}}^2$ exceeds the control limit. Because the sampling distribution of the $T_{t,\text{RMCD}}^2$ values is not known in closed-form, control limits are found through simulation to ensure a desired IC false alarm rate or average run length. This chart, which is denoted here as $T_{t,\text{RMCD}}^2$, is able to handle contamination in the training set by allowing the IC mean and covariance to be estimated from a subset of the Phase I observations that does not contain outliers. One drawback of this approach is that using robust estimators will lead to excessive false alarms if the historical data are clean. Additionally, because only a subset of historical observations is used in the final parameter estimates, a lot of potentially useful information is effectively discarded. Also, in this setting, robust estimation tends to need larger sample sizes to get good estimates.

Robust MEWMA (similar to Midi and Shabbak (2011)). This method, like the robust T^2

chart, is equivalent its classical counterpart, the classical MEWMA chart, except that the classical estimates of the IC mean and covariance are replaced with robust estimates, such as the RMCD estimates. The robust MEWMA monitoring statistics are found by first computing

$$\mathbf{q}_t = \lambda(\mathbf{x}_t - \hat{\boldsymbol{\mu}}_{\text{RMCD}}) + (1 - \lambda)\mathbf{q}_{t-1}, \quad (9)$$

and the chart classifies an observation as OC when the monitoring statistic

$$\mathbf{q}_t' \widehat{\boldsymbol{\Sigma}}_{\mathbf{q}_t, \text{RMCD}}^{-1} \mathbf{q}_t > h_{\text{MEWMA}_{\text{RMCD}}},$$

where $\widehat{\boldsymbol{\Sigma}}_{\mathbf{q}_t, \text{RMCD}} = [\lambda(1 - (1 - \lambda)^{2t})/(2 - \lambda)]\widehat{\boldsymbol{\Sigma}}_{\text{RMCD}}$, and $h_{\text{MEWMA}_{\text{RMCD}}}$ is determined through simulation. Similar to the robust T^2 chart, this method, which is denoted as $\text{MEWMA}_{\text{RMCD}}$, accounts for potential contamination in the training data at the cost of a potentially increased false alarm rate. Similar to the MEWMA chart, this method is better suited for detecting small shifts than the robust T^2 chart. ⁻¹³⁻

3 Simulation Study

We evaluate the Phase II true detection rate (TDR) and false alarm rate (FAR) of each method in Section 2 on simulated data. The TDR is the proportion of OC observations that are correctly identified as OC, and the FAR is the proportion of IC observations incorrectly classified as OC. To obtain these values, we first generate a simulated IC Phase I dataset of size m_0 using independent draws from a p -dimensional multivariate standard normal distribution. Then, we randomly introduce contamination by replacing a random $\pi(100)\%$ of observations with draws from a shifted multivariate normal distribution with mean $\boldsymbol{\delta} = (\delta, \dots, \delta)'$. Values of m_0 , p , π , and $\boldsymbol{\delta}$ considered in this simulation study are given in Table 1.

Next, we evaluate the TDR and FAR of each chart on 1000 observations in Phase II under four scenarios that explore three types of faults. The faults we explore are sustained, transient, and isolated faults. Sustained faults occur when one or more variables experience a sudden

Table 1: All simulation study factors for generating data and evaluating the performance of each method.

Factor	Levels
Initial Sample Size (m_0)	30, 50, 100, 250, 500, 1000
Dimension (p)	3, 6
Contamination (π)	0, 0.1, 0.2, 0.3, 0.4
Shift Size (δ)	0, 0.75, 1.5, 2.25, 3

shift that is sustained for a long period of time. Transient faults are when at least one variable is shifted temporarily but returns to normal levels after a period of time. Isolated faults are random outliers, often represented by sudden unexpected spikes in a variable that quickly return to normal levels. Another common type of fault is drift faults, which occur when a variable slowly drifts outside of its normal limits over a long period of time. Drift faults can be difficult to detect because they involve gradual changes over a long period of time. The four scenarios we use in the simulation study to assess the TDR and FAR of each chart in Phase II are:

Example of a realization from each scenario would be helpful.

Scenario 1: Immediate sustained fault after Phase II begins. All 1000 Phase II observations are drawn from a $N_p(\delta, I)$ distribution (OC).

Scenario 2: Initial period of IC observations after Phase I followed by a sustained fault. The first 200 Phase II observations are drawn from a $N_p(0, I)$ distribution (IC), and the next 800 are drawn from a $N_p(\delta, I)$ distribution (OC).

Scenario 3: Multiple transient faults during Phase II. Observations 1-200, 301-500, 601-800, and 901-1000 in Phase II are drawn from a $N_p(0, I)$ distribution (IC). All other observations are drawn from a $N_p(\delta, I)$ distribution (OC).

Scenario 4: Random isolated faults during Phase II. 500 Phase II observations are drawn from a $N_p(0, I)$ distribution (IC), and 500 of the Phase II observations are drawn from a $N_p(\delta, I)$ distribution (OC), representing 50% random isolated faults in Phase II.

Simulation results for Scenarios 1-4 are discussed in Sections 3.1-3.5. In general, the results are similar for both $p = 3$ and $p = 6$, so we only discuss on the $p = 3$ case.

The control limits for each chart are obtained through simulation to ensure a FAR of $\alpha = 0.005$ when no contamination is present. First, we generate a p -dimensional dataset of

size $m_0 + 1000$ from a multivariate standard normal distribution. IC parameters are estimated using the initial m_0 observations, and the number of control limit exceedances for the 1000 subsequent IC observations is computed as the FAR for a given control limit. This procedure is repeated 5000 times, and the mean of the 5000 FAR values is taken as the overall FAR for that control limit. Finally, a bisection search algorithm is used to select a control limit for each chart that results in the desired FAR of 0.005.

*reference
needed*

[include and discuss plots of classical + robust T^2 or MEWMA control limits here to demonstrate effects of p and m_0 on control limit values?]

It could be useful. It's a little bit off-topic, but is there anything useful here that hasn't been reported elsewhere?

3.1 Scenario 1: Immediate Sustained Fault

In Scenario 1, the FAR is not computed because there are no IC observations in Phase II that can incorrectly be classified as OC. In this scenario, all methods obtain parameter estimates using the same set of m_0 initial observations. Because the sustained shift occurs immediately at the beginning of Phase II, RSSB-MEWMA and RSSB- T^2 are *do not* update their parameter estimates. However, SS-MEWMA and SS- T^2 continue to update their parameter estimates with each observation because they are not designed to omit observations classified as OC from parameter estimate updates. Due to the inability of the RSSB methods to update, there is no advantage to their self-starting design, and the RSSB-MEWMA and MEWMA_{RMCD} methods are *equivalent*, as are their T^2 versions. TDR results for Scenario 1 across all values of m_0 and δ when $\pi = 0, 0.2, 0.4$ are given in Figure 1. In this scenario, the best methods are the MEWMA, MEWMA_{RMCD}, and RSSB-MEWMA, all of which *generally* perform similarly. However, the MEWMA is the best overall method in Scenario 1, achieving the highest TDR across all shift sizes regardless of the values of m_0 or π .

The T^2 chart performs better than T^2_{RMCD} when $m_0 < 250$ and $\pi = 0$. The T^2_{RMCD} chart substantially outperforms T^2 whenever there is contamination, but its performance is poor relative to the MEWMA-based methods and generally only outperforms SS-MEWMA in the presence of contamination when δ is large. However, SS-MEWMA outperforms all the T^2 -based methods when π is small and m_0 is large, especially for small shift sizes. In every case, SS-MEWMA

Performance Under Scenario 1; $p = 3$

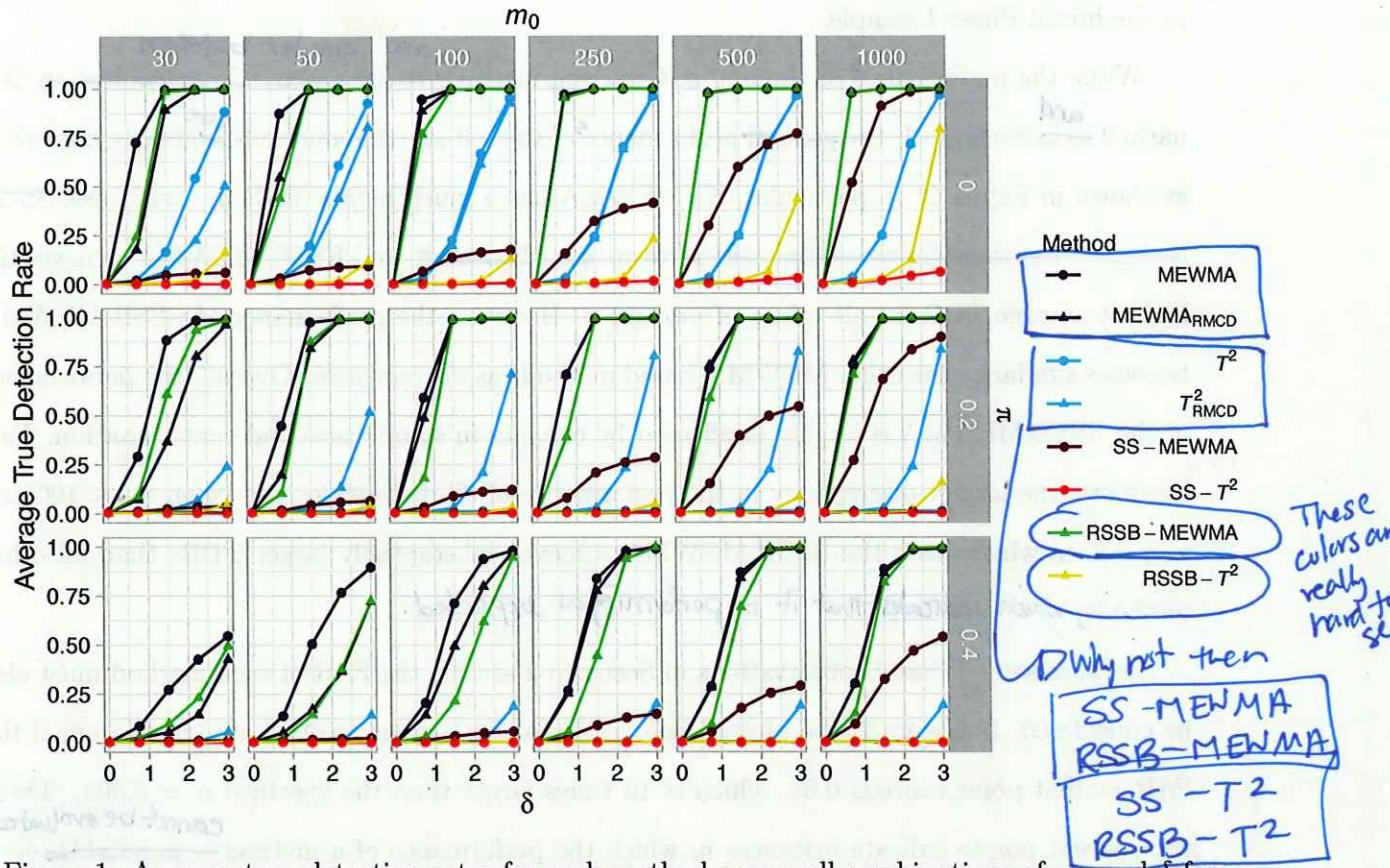


Figure 1: Average true detection rates for each method across all combinations of m_0 and δ for $\pi = 0, 0.2$, and 0.4 . Each robust method is plotted with triangles, and the non-robust methods are plotted with circles.

performs poorly for small values of m_0 , and SS- T^2 performs poorly in all situations. Overall, the performance of the T^2 -based methods is poor when $\pi = 0.4$, and only T^2_{RMCD} performs well when $\pi > 0$. Finally, RSSB- T^2 performs poorly relative to T^2_{RMCD} for all values of m_0 and π , but RSSB- T^2 always outperforms SS- T^2 .

3.2 Scenario 2: Delayed Sustained Fault

Unlike Scenario 1, Scenario 2 gives the RSSB methods an opportunity to update their parameter estimates. This scenario is also more realistic, because Phase II monitoring almost never begins at the exact time ^{that} a fault begins. More often, monitoring begins while the process is IC and continues for some time until a fault occurs. In these cases, it may be beneficial to update

parameter estimates over time to more wholly capture IC behavior that may have been missed in the initial Phase I sample.

While the performance^s of the non self-starting methods is approximately equivalent in Scenario 2 ~~as in~~ and Scenario 1, the general performance^s of the self-starting methods are greatly improved, as shown in Figure 2. In particular, SS-MEWMA has a much higher TDR in every case, SS- T^2 is slightly improved for $\pi = 0$ but still performs poorly overall, and RSSB-MEWMA achieves the highest average TDR for all values of m_0 and π . However, the performance of RSSB-MEWMA becomes similar to the other MEWMA-based methods as m_0 increases. Overall, the performance of the RSSB-MEWMA is largely unaffected by changes in sample size and contamination. Furthermore, the largest discrepancy in the performance of all methods occurs when $m_0 \leq 100$ and $\pi \geq 0.2$, in which cases the RSSB-MEWMA achieves substantially larger TDRs than all other methods, which indicates that it is performing as expected.

Because some Phase II observations in Scenario 2 are IC, the FAR of each method must also be considered. In Figure 2, the plotted line and point for a given method are transparent if the FAR at that point exceeds 0.05, which is 10 times larger than the specified $\alpha = 0.005$. These transparent points indicate instances in which the performance of a method is unreliable, even if the method produces a high TDR, this can likely be attributed to the corresponding large FAR. In general, the FAR is stable for all methods in this scenario. However, when $\pi = 0.3$ or $\pi = 0.4$, the FARs of MEWMA and SS-MEWMA begin to increase as m_0 increases, reaching over 40% for the MEWMA chart when $m_0 = 1000$, $\pi = 0.4$, and $\delta = 3$. The FARs of RSSB-MEWMA also exceed 0.05 on a couple occasions when $m_0 > 100$ and $\pi > 0.2$. Because of largely stable FARs and high TDRs, the RSSB-MEWMA is the best overall method in Scenario 2, and RSSB-MEWMA is especially superior to the other methods when $m_0 \leq 100$ and $\pi \geq 0.2$.

3.3 Scenario 3: Transient Faults

Overall results for Scenario 3 are largely similar to Scenario 2 and are given in Figure 3. In general, the MEWMA-based methods outperform the T^2 -based methods, with RSSB-MEWMA achieving the highest TDRs, especially when $m_0 \leq 100$. As m_0 increases, the performance of

Performance Under Scenario 2; $p = 3$

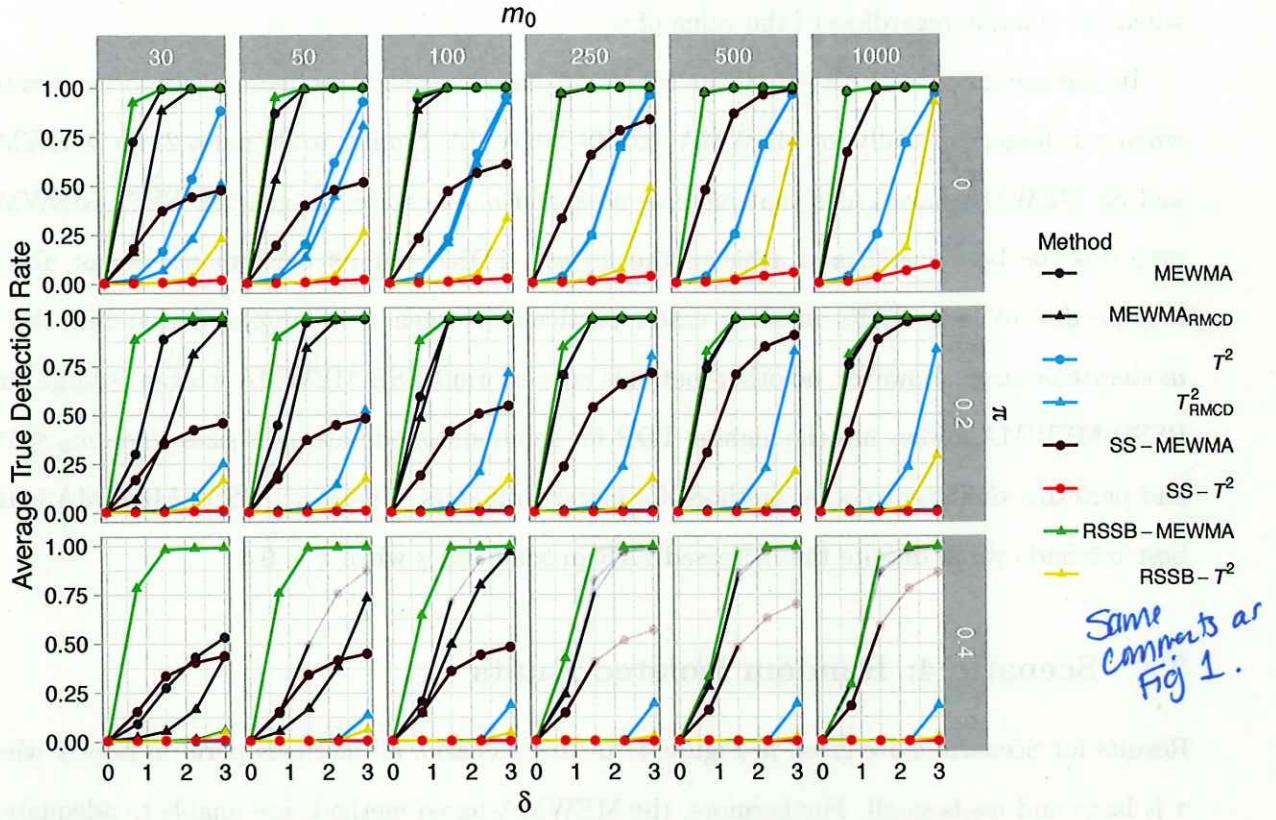


Figure 2: Average true detection rates for each method across all combinations of m_0 and δ for $\pi = 0, 0.2$, and 0.4 . Each robust method is plotted with triangles, and the non-robust methods are plotted with circles. When $\text{FAR} \geq 0.05$ for a given method, the line and points for that method are transparent.

translucent.

all MEWMA-based methods becomes similar. The primary difference in the results between Scenarios 2 and 3 is greatly improved performance for SS-MEWMA and SS- T^2 in Scenario 3 compared to Scenario 2. Here, SS-MEWMA outperforms MEWMA and MEWMA_{RMCD} when $\pi \geq 0.2$, especially when $m_0 \leq 100$. SS- T^2 achieves higher TDRs when $\pi = 0$, but still performs poorly relative to all other methods. The greatly improved performance of SS-MEWMA, along with the performance of RSSB-MEWMA, demonstrates the importance and impact of using self-starting methods to update parameter estimates over time when the initial Phase I sample is small. In this scenario, the performance of the RSSB- T^2 is also improved, particularly for large values of δ . However, RSSB- T^2 only outperforms T^2_{RMCD} when $m_0 \leq 50$ and δ is large. The best method for all values of m_0 and π is RSSB-MEWMA, especially when m_0 is small and π is large.

for very small m_0 (≤ 50) and large shift sizes.

Similar to the other scenarios, the performance of all MEWMA-based methods becomes similar when m_0 is large, regardless of the value of π .

In this scenario, the FARs of all the MEWMA-based methods are more likely to be elevated when π is large, especially for MEWMA and SS-MEWMA. Similar to Scenario 2, the MEWMA and SS-MEWMA have FARs that increase as π and m_0 increase. In general, RSSB-MEWMA performs the best overall and achieves the highest TDRs, but its performance is not always reliable due to large FARs in some cases, particularly when π is large, and the shift size is moderate or large. However, no other methods outperform RSSB-MEWMA when π is large, and RSSB-MEWMA always has the highest TDR for small and moderate shift sizes when $m_0 \leq 100$ and performs similarly to other methods for larger shift sizes. Therefore, RSSB-MEWMA is the best method overall despite the increased FAR in some cases when $\pi = 0.4$.

3.4 Scenario 4: Random Isolated Faults

Results for Scenario 4 are given in Figure 4. In this scenario, all methods perform poorly when π is large and m_0 is small. Furthermore, the MEWMA-based methods are unable to adequately distinguish between IC and OC observations; these methods have FARs that approach 100% in several situations. Even though the corresponding TDRs for these methods are also high, the overall performance of the MEWMA-based methods is poor because almost all observations are classified as OC. Thus, none of the MEWMA-based methods are reliable in Scenario 4.

In contrast to the MEWMA-based methods, the T^2 methods have FARs that are stable and small for all values of m_0 , π , and δ . When $\pi = 0$, T^2 performs the best, with T_{RMCD}^2 performing similarly when $m_0 \geq 100$. Furthermore, T_{RMCD}^2 performs the best when $\pi > 0$. While $\text{RSSB-}T^2$ outperforms T^2 when $\pi > 0$, its performance is always worse than T_{RMCD}^2 . Therefore, T_{RMCD}^2 is the best overall method in Scenario 4. *Why doesn't RSSB do well in this scenario?*

3.5 Summary

The performance of each method on independent multivariate normal observations depends heavily on the types of faults that occur during Phase II and the ability of the self-starting methods to

Performance Under Scenario 3; $p = 3$

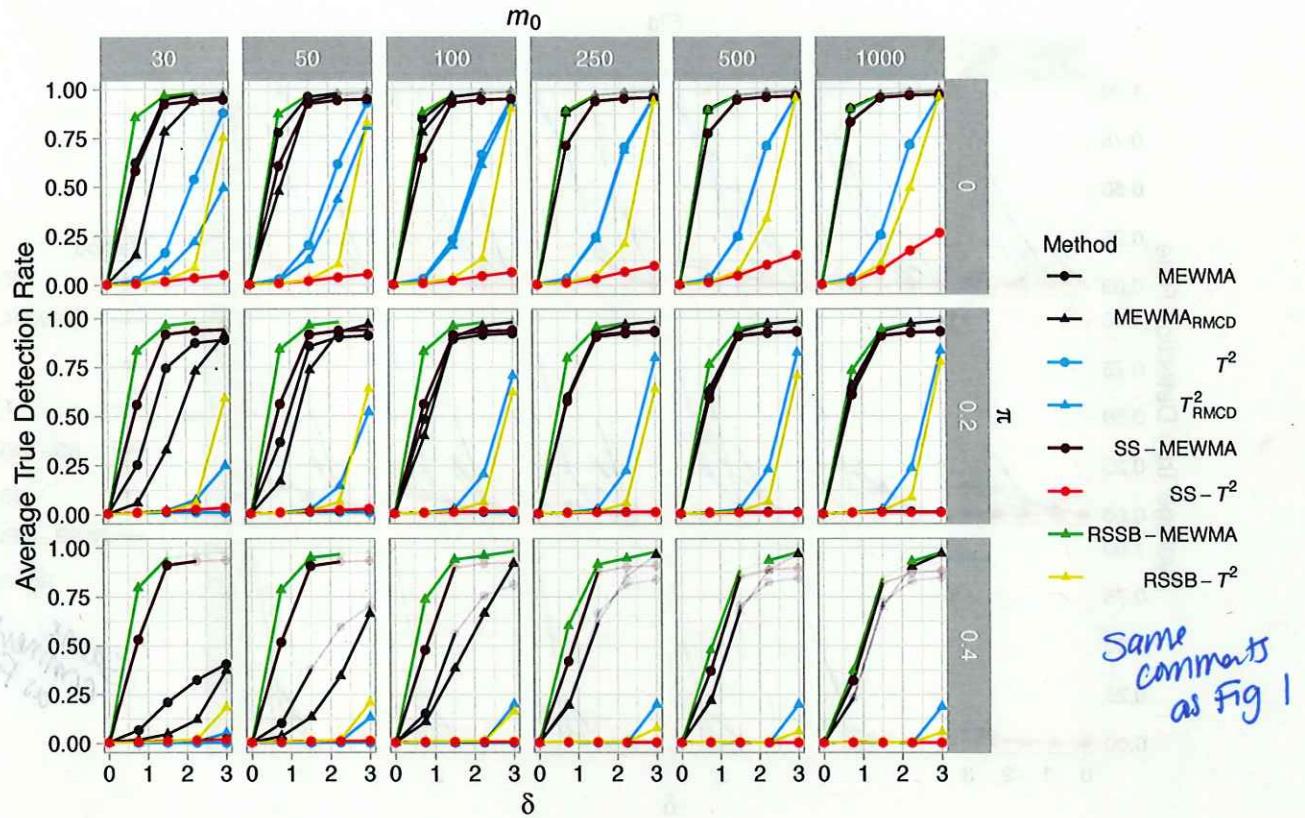


Figure 3: Average true detection rates for each method across all combinations of m_0 and δ for $\pi = 0, 0.2$, and 0.4 . Each robust method is plotted with triangles, and the non-robust methods are plotted with circles. When FAR > 0.05 for a given method, the line and points for that method are transparent. same comments as Fig 1

update. In Scenario 1, neither of the RSSB methods ~~is able to~~ update, and the MEWMA chart performs the best in all scenarios. However, in real applications, sustained faults are generally not expected to begin immediately after Phase I ends, making Scenario 1 unrealistic. Therefore, Scenarios 2-4 are more realistic, with faults being interspersed with IC observations.

In Scenario 2, the process is IC for a period of time after Phase I before the ~~occurrence~~^{initiation} of a sustained fault. In Scenario 3, the process experiences multiple periods of IC behavior interrupted by several transient faults. In Scenario 4, random observations in Phase II are drawn from a shifted distribution to represent isolated faults, such as random sensor or measurement errors. In Scenarios 2 and 3, RSSB-MEWMA generally outperforms all other methods, and its performance is almost completely unaffected by contamination in Phase I. In Scenario 4, the

Performance Under Scenario 4; $p = 3$

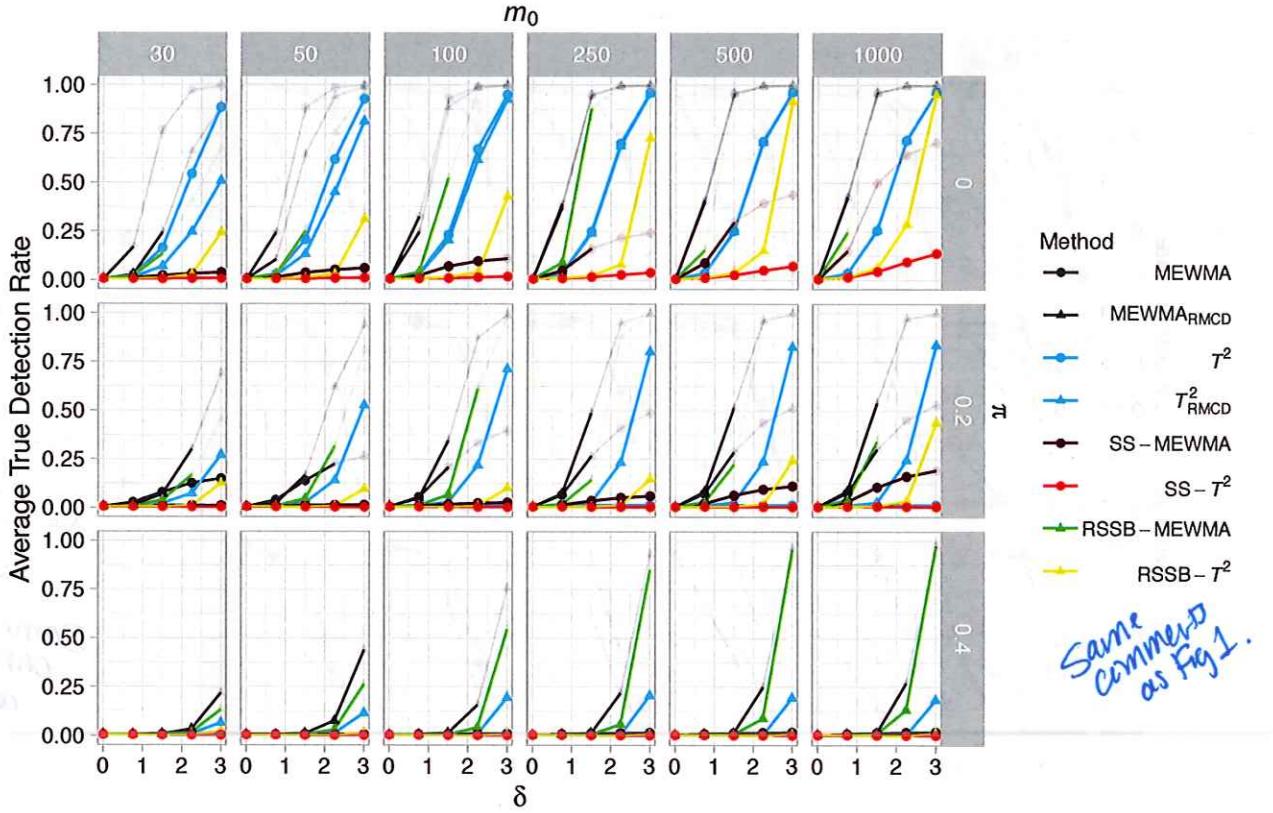


Figure 4: Average true detection rates for each method across all combinations of m_0 and δ for $\pi = 0, 0.2$, and 0.4 . Each robust method is plotted with triangles, and the non-robust methods are plotted with circles. When FAR > 0.05 for a given method, the line and points for that method are transparent. Same comment as Fig 2.

MEWMA-based methods are all unreliable and produce high FARs, and the T^2_{RMCD} performs the best overall.

Comments on computational time of each method would be appropriate to make here.

4 Case Study

In this section, we compare the fault detection performance of each method from Sections 2 and 3 on data from the Pure Water Demonstration facility (PWDF) in Calabasas, California operated by the Las Virgenes-Triunfo Joint Powers Authority. PWDF is a demonstration-scale water treatment facility that receives treated wastewater and purifies it to exceed safe drinking water standards. Water is purified through an advanced treatment process consisting of ultrafiltration

(UF), reverse osmosis, and an ultraviolet-advanced oxidation process. Throughout each stage of the treatment process, several variables, such as feed temperature and filtrate ammonia, are measured by various sensors and analyzers. Additional details regarding the design and operation of PWDF are given by Grimm et al. (2024).

For this case study, we use daily data from the UF portion of the treatment process with a goal of monitoring membrane fouling, filtrate quality, and upstream water quality. To monitor each of these attributes, we apply each control chart from Section 2 to a dataset consisting of the

daily averaged observed values of temperature-corrected permeability, filtrate turbidity, and filtrate ammonia, which are shown in Figure 5. The time period identified for monitoring the treatment process is

1 November 2021 through 30 September 2022. The first 50 observations, *spanning*

December, are used as the initial Phase I data and are shown in gray. During Phase II (black), there are a few notable events that we expect to be flagged as OC. For example, there are a few events in which the filtrate turbidity has a sudden large increase over a short period of time before *being returned* to normal levels, such as the events shortly before 19 Feb and 11 Jun.

Other notable events are the transient fault in ammonia shortly after 19 Feb and the upward shift and large spikes in ammonia before 11 Jun. The behavior of these variables during Phase II most resembles scenario 3 from the simulation study with occasional isolated faults; Phase II is initially IC, allowing for self-starting methods to update their parameter estimates, and then several isolated and transient faults occur sporadically.

Some contamination is present during the Phase I period shown in Figure 5, but the majority of observations behave as expected under normal operating conditions. A clear example of contamination is the beginning of the Phase I period for filtrate turbidity where a large spike occurs before returning to IC levels. To better demonstrate the performance of various methods on real data that may have a high level of contamination, we introduce additional contamination by randomly adding draws from a normal distribution to each variable so that the level of contamination in the resulting Phase I period exceeds 20%. Contamination is introduced by randomly replacing 10 of the observed daily values of temperature-corrected permeability, filtrate turbidity, and filtrate ammonia with draws from a multivariate normal distribution with mean

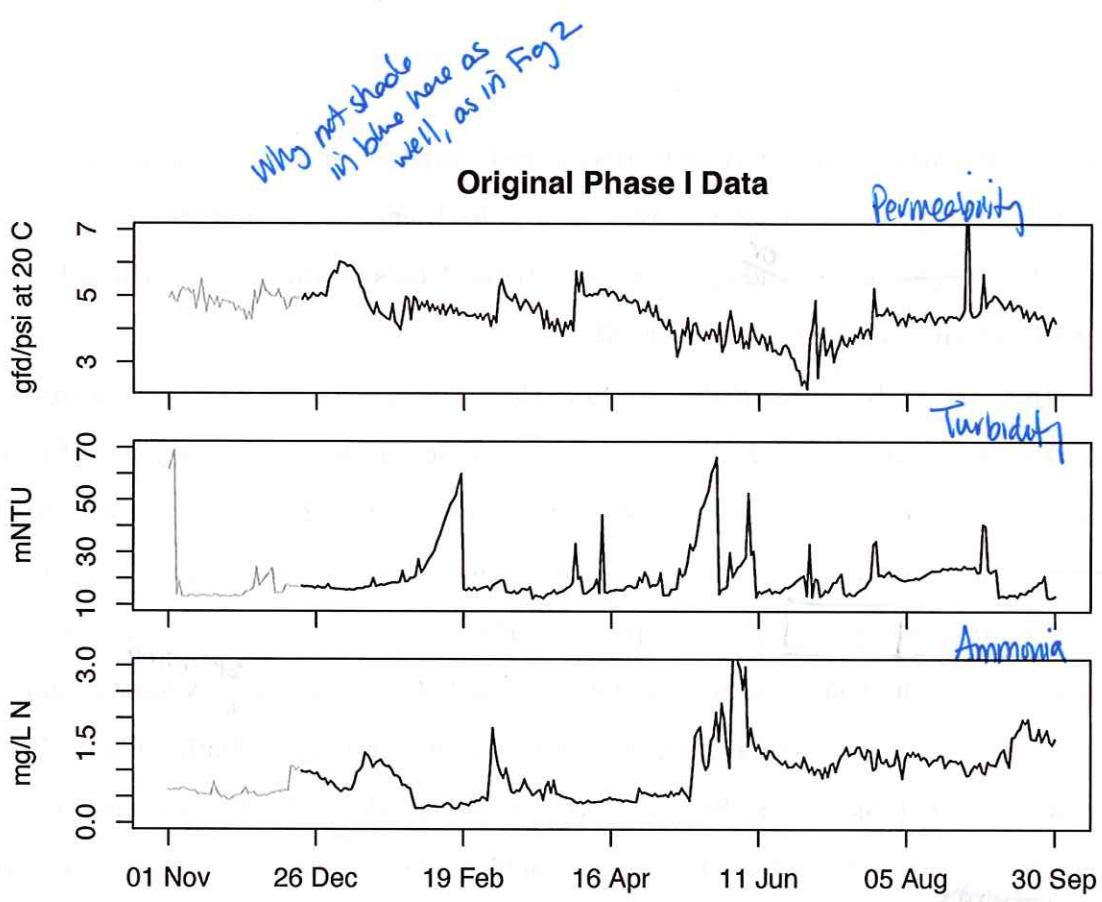


Figure 5: Time series plots of the observed values of temperature-corrected permeability (top), filtrate turbidity (middle), and filtrate ammonia (bottom) from 1 November 2021 through 30 September 2022. The 50 Phase I observations are shown in gray, and Phase II observations are black.

put a legend

vector $(1, 35, 1)'$ and a diagonal covariance matrix with diagonal elements $(0.25^2, 20^2, 0.25^2)$.

This produces contamination for each variable that resembles potential faults during Phase II monitoring. The resulting contaminated Phase I dataset, followed by the original Phase II observations, is given in Figure 6. [Faults for each variable during Phase II are also shown with a shaded background.] → make this for Fig 5 and move this sentence to prior ff.

We apply each method from Sections 2 and 3 to the data shown in Figure 6 with the first 50 days shown in gray used as the Phase I set to obtain initial parameter estimates. Due to the presence of autocorrelation in the data, the typical control limits found in Section 3 must be adjusted to ensure a desired FAR. Therefore, we obtain new control limits for each chart by first fitting a vector autoregressive model of order 1 to the contaminated Phase I data. Then, the estimated time series parameters are used to generate 5000 datasets of size $50 + 1000$, and a bisection search algorithm is used to find the control limit for each chart that ensures a

Do you also apply the methods to data in Fig 5?
Should also show or describe those results.

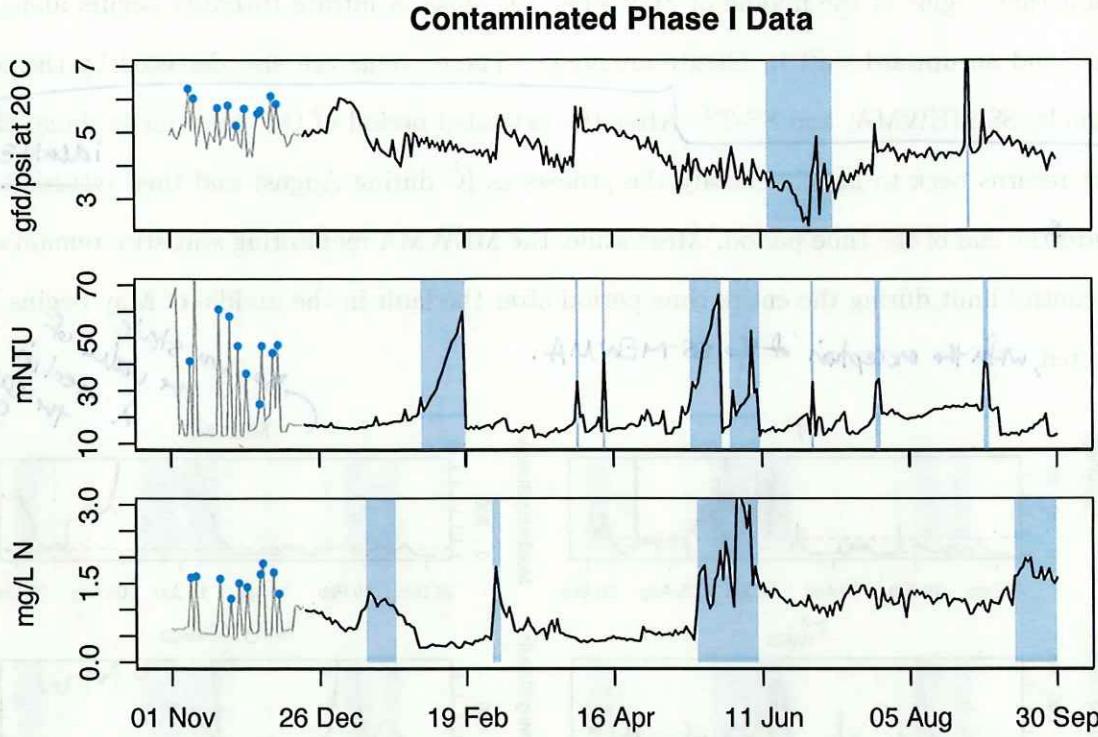


Figure 6: Time series plots of the observed values of temperature-corrected permeability (top), filtrate turbidity (middle), and filtrate ammonia (bottom) from 1 November 2021 through 30 September 2022. The Phase I data have been contaminated (blue points). Background shading indicates a fault in a given variable.

FAR of $\alpha = 0.005$ on the 1000 Phase II observations using the first 50 simulated observations as initial Phase I data, similar to Section 3. Based on the behavior seen in the simulation study for scenario 3, we expect the MEWMA-based methods to quickly detect small shifts but potentially produce excessive false alarms after large shifts occur. We also expect the RSSB- T^2 and T_{RMCD}^2 to perform similarly, quickly detecting large shifts with RSSB- T^2 potentially performing slightly better because m_0 is small, and RSSB- T^2 can update parameter estimates between fault occurrences.

Results for each method are shown in Figure 7. Periods of time when a true fault occurs in any of the three variables are shown with a shaded background. The classical T^2 and SS- T^2 charts do not detect the large spike in filtrate turbidity right before 19 Feb, but the spike is briefly detected by the classical MEWMA chart. However, the same spike is clearly detected by the robust methods and SS-MEWMA. The classical T^2 and MEWMA charts consistently detect

Can't you just say
 not all of
 methods
 detect
 this fault?
 []

a fault that begins in the middle of May when the spike in filtrate turbidity occurs along with spikes and an upward shift in filtrate ammonia. These events are also detected by the robust methods, SS-MEWMA, and SS- T^2 . After the extended period of OC behavior in June, the T^2 chart returns back to largely classify the process as IC during August and then returns to OC behavior toward the end of the time period. Meanwhile, the MEWMA monitoring statistics remain above the control limit during the entire time period after the fault in the middle of May begins to be detected, *with the exception of the SS-MEWMA.*

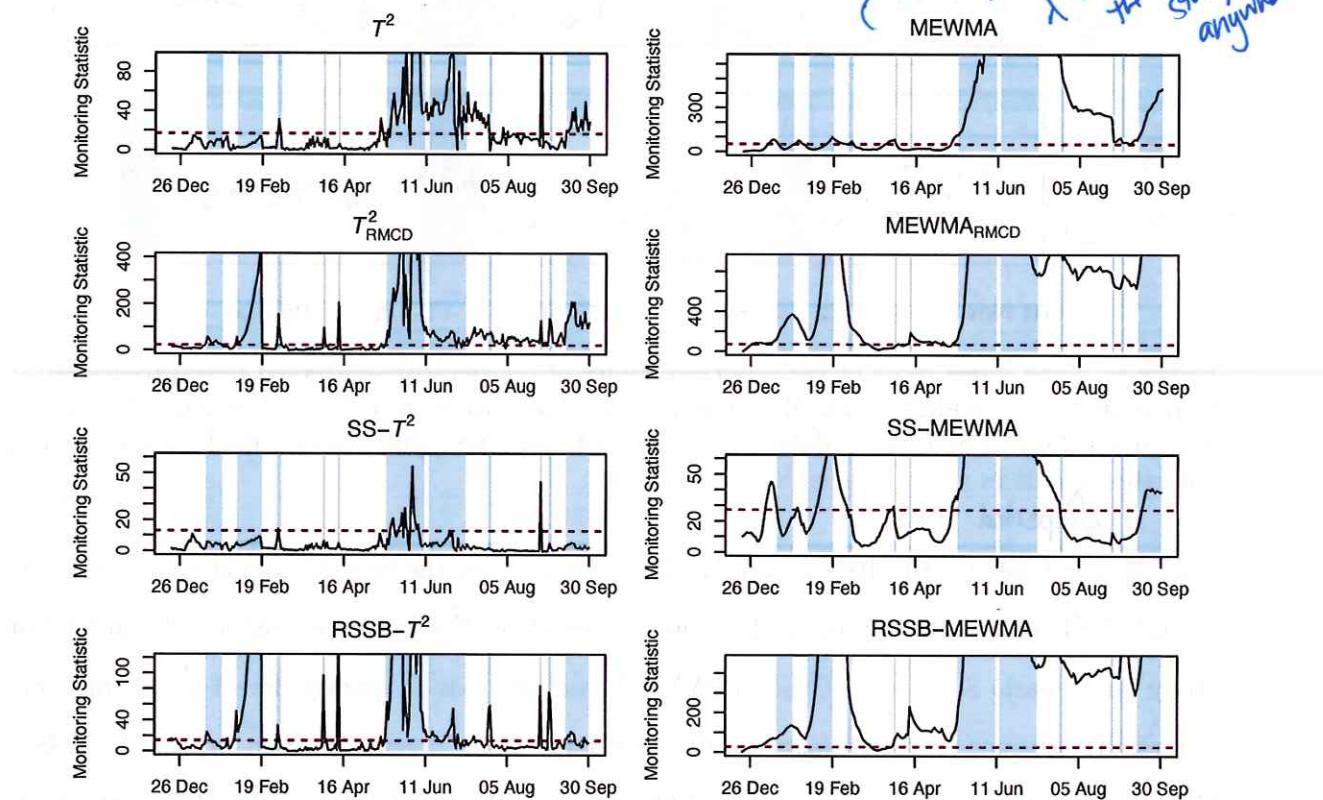


Figure 7: Results from applying each method to case study data using contaminated data from 1 November through 20 December as the Phase I data. Explain what the shaded areas represent.

Say the T^2 methods are in left column, MEWMA are in the right column.

During the first half of the Phase II period being monitored, the performance is similar between $RSSB-T^2$ and T^2_{RMCD} . However, during the second half of Phase II, the performance of $RSSB-T^2$ more closely resembles the performance of T^2 , but $RSSB-T^2$ detects more isolated faults than T^2 . Both $RSSB-T^2$ and T^2_{RMCD} flag several isolated faults during the first half of monitoring that go undetected by T^2 and $SS-T^2$, such as a spike around 16 Apr, which is related

to a spike in filtrate turbidity, and transient faults, such as the temporary increase in filtrate turbidity before 19 Feb. RSSB- T^2 and T_{RMCD}^2 both also detect the spikes in filtrate turbidity and the spikes and upward shift in filtrate ammonia in the middle of May, similar to T^2 . Overall, SS- T^2 performs poorly and only consistently detects the large sustained fault during the middle of May and beginning of June.

One primary difference between the performance of RSSB- T^2 and T_{RMCD}^2 occurs from Jun 2022 - Sep 2022. During this time, almost all observations are flagged as OC by T_{RMCD}^2 , with monitoring statistic values only briefly falling beneath the control limit ~~for~~ ^{for observations} a few occasions. However, the monitoring statistics for RSSB- T^2 ~~is~~ largely beneath the control limit during this same time period, and only isolated faults are flagged until a period of time shortly before 30 Sep when a simultaneous decrease in filtrate turbidity and increase in filtrate ammonia occurs. This difference in performance illustrates the effect of updating the parameter estimates on RSSB- T^2 ; the process is deemed to be primarily IC after the major faults when parameter estimates are occasionally updated ~~for~~ (RSSB- T^2), but not when they stay constant (T_{RMCD}^2).

The classical MEWMA chart ~~generally~~ detects the large spikes that are detected by the T^2 -based methods, such as events shortly before and after 19 Feb and shortly before 16 Apr. ~~However, the isolated faults are not all distinctly detected by MEWMA, and the detection of the identified faults is delayed.~~ Even though the MEWMA chart detects ~~the fault~~ ^{these} that occurs shortly before 19 Feb, the monitoring statistics do not exceed the control limit by a large margin relative to the other control charts, indicating that the fault is not as definitively detected by the MEWMA ~~as the other charts~~. The classical MEWMA chart also detects the event that begins in May, similar to all the other methods.

but the monitoring statistic only just exceeds the threshold, so they are not definitively and strongly detected.

In general, the MEWMA-based methods tend to flag the majority of observations as OC and do not detect isolated faults as well as the T^2 -based methods. The MEWMA-based methods are also overly sensitive to small changes in the variables, resulting in many false alarms; all the MEWMA methods flag observations near the beginning of Phase II before any real faults occur.

After the fault begins in the middle of May, the monitoring statistics remain above the control limit for all MEWMA-based methods except SS-MEWMA, which functions similarly to T^2 and

RSSB- T^2 by classifying some observations as IC prior to classifying the remainder of observations as OC in Sep. However, T^2 and RSSB- T^2 are able to detect clear isolated faults throughout the entire Phase II period, which is not the case for SS-MEWMA.

To summarize the information in Figure 7, Table 2 shows the TDR and FAR of each method during the Phase II period. All of the MEWMA-based methods except SS-MEWMA produce high TDRs. However, all the MEWMA-based methods also produce FARs that exceed 50%, and the robust MEWMA methods produce FARs ^{over} in excess of 80%. SS- T^2 has a low FAR but detects very few of the faults, producing a TDR of only 4%. While T^2_{RMCD} produces the highest TDR of the T^2 -based methods, it also has a FAR of 39%, while T^2 and RSSB- T^2 have much lower FARs of 13% and 9%, respectively, while achieving TDRs of 60% and 71%.

Table 2: Phase II true detection and false alarm rates for each method applied to the contaminated data.

	T^2	T^2_{RMCD}	RSSB- T^2	SS- T^2	MEWMA	MEWMA _{RMCD}	RSSB-MEWMA	SS-MEWMA
TDR	0.60	0.96	0.71	0.04	0.79	0.98	1.00	0.21
FAR	0.13	0.39	0.09	0.07	0.56	0.81	0.89	0.51

The primary difference between T^2 and MEWMA charts is clearly demonstrated in this case study: MEWMA-based methods are sensitive to small changes in process variables, while T^2 -based methods are better suited for detecting large isolated faults. In this case study, most of the faults have large shift sizes, which resulted in the MEWMA-based methods mostly achieving high TDRs while producing a large number of false alarms. The MEWMA_{RMCD}, RSSB-MEWMA, and SS-MEWMA charts also all classify the majority of observations as OC as soon as Phase II begins. Overall, RSSB- T^2 performs the best among all the methods, achieving the best balance of detecting important faults without producing unnecessary false alarms.

5 Conclusion

Fault detection in multivariate processes is commonly done using traditional multivariate control charts, such as T^2 and MEWMA, which are based on classical estimators. However, these methods perform poorly when Phase I data are minimal or are contaminated with faults or OC observations. Robust and self-starting control charts have been proposed to address these issues

separately; however, robust methods are unable to update IC parameter estimates over time and typically require large initial Phase I samples. Furthermore, self-starting methods are sensitive to early parameter estimates and can quickly become degraded if early outliers or process shifts occur.

To simultaneously address the issues of contamination and minimal availability of Phase I data, we proposed a robust, self-starting Bayesian approach with closed-form IC parameter estimate updates. Two charts, RSSB- T^2 and RSSB-MEWMA, were compared to existing methods. While RSSB- T^2 was shown to outperform T^2 in the presence of contamination, its performance is typically worse than T_{RMCD}^2 except when detecting large transient faults over time with minimal Phase I data. RSSB-MEWMA was shown to substantially outperform existing methods when detecting sustained and transient faults in Phase II, especially when contamination in Phase I is strong and the IC parameter estimates are able to be updated, which occurs when the faults do not begin immediately at the start of Phase II, which is usually the case in practice. In a case study on water treatment data, RSSB- T^2 and RSSB-MEWMA were able to detect faults and performed similarly to existing methods, with RSSB- T^2 providing the best balance of detecting important faults while minimizing false alarms.

At first in present tense.

While the performance of the proposed RSSB methods is promising, additional research on these methods is needed. In our simulation study, data are generated exclusively from a multivariate normal distribution. In the future, it would be interesting to evaluate the performance of RSSB-MEWMA and RSSB- T^2 on data from skewed and heavy-tailed distributions to assess their robustness to model misspecification. Additionally, the methods compared in this study assume that observations are independent. This assumption is often violated in practice, especially when data are collected with a high sampling frequency. Therefore, best practices for applying these methods to strongly dependent data should also be studied. Finally, the RSSB methods evaluated here are designed to update their parameter estimates whenever a given observation is classified as IC. However, this design could cause the presence of slow drift faults to go undetected. One solution could be to increase the number of consecutive observations required to be classified as IC before updating parameter estimates.

The number of consecutive observations required before an update may depend on the speed at which the drift changes...

Acknowledgements

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 through a Project Enhancement Grant between Project 5.08 Autonomous Control and Optimization and Project 5.17 Data-Driven Fault Detection and Process Control for Potable Reuse with Reverse Osmosis.

References

- Alfaro, J. L. and Ortega, J. F. (2009). A comparison of robust alternatives to Hotelling's T^2 control chart. *Journal of Applied Statistics*, 36(12):1385–1396.
- Alfaro, J.-L. and Ortega, J.-F. (2012). Robust Hotelling's T^2 control charts under non-normality: The case of t -Student distribution. *Journal of Statistical Computation and Simulation*, 82(10):1437–1447.
- Cabana, E. and Lillo, R. E. (2022). Robust multivariate control chart based on shrinkage for individual observations. *Journal of Quality Technology*, 54(4):415–440.
- Chenouri, S., Steiner, S. H., and Variyath, A. M. (2009). A multivariate robust control chart for individual observations. *Journal of Quality Technology*, 41(3):259–271.
- Chenouri, S. and Variyath, A. M. (2011). A comparative study of phase II robust multivariate control charts for individual observations. *Quality and Reliability Engineering International*, 27(7):857–865.
- Crosier, R. B. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*, 30(3):291–303.
- Feltz, C. J. and Shiao, J. H. (2001). Statistical process monitoring using an empirical Bayes multivariate process control chart. *Quality and Reliability Engineering International*, 17(2):119–124.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.
- Grimm, T. R., Branch, A., Thompson, K. A., Salveson, A., Zhao, J., Johnson, D., Hering, A. S., and Newhart, K. B. (2024). Long-term statistical process monitoring of an ultrafiltration water treatment process. *ACS ES&T Engineering*, 4(6):1492–1506.
- Hawkins, D. M. (1987). Self-starting CUSUM charts for location and scale. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 36(4):299–316.
- Hawkins, D. M. and Maboudou-Tchao, E. M. (2007). Self-starting multivariate exponentially weighted moving average control charting. *Technometrics*, 49(2):199–209.
- Healy, J. D. (1987). A note on multivariate CUSUM procedures. *Technometrics*, 29(4):409–412.
- 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.
- Hubert, M. and Debruyne, M. (2010). Minimum covariance determinant. *WIREs Computational Statistics*, 2(1):36–43.

- Hubert, M., Debruyne, M., and Rousseeuw, P. J. (2018). Minimum covariance determinant and extensions. *WIREs Computational Statistics*, 10(3):e1421.
- Jensen, W. A., Birch, J. B., and Woodall, W. H. (2007). High breakdown estimation methods for phase I multivariate control charts. *Quality and Reliability Engineering International*, 23(5):615–629.
- Jensen, W. A., Jones-Farmer, L. A., Champ, C. W., and Woodall, W. H. (2006). Effects of parameter estimation on control chart properties: A literature review. *Journal of Quality Technology*, 38(4):349–364.
- Jones-Farmer, L. A., Woodall, W. H., Steiner, S. H., and Champ, C. W. (2014). An overview of phase I analysis for process improvement and monitoring. *Journal of Quality Technology*, 46(3):265–280.
- Keefe, M. J., Woodall, W. H., and Jones-Farmer, L. A. (2015). The conditional in-control performance of self-starting control charts. *Quality Engineering*, 27(4):488–499.
- Kim, S. and Turkoz, M. (2022). Bayesian sequential update for monitoring and control of high-dimensional processes. *Annals of Operations Research*, 317(2):693–715.
- Kim, S., Turkoz, M., and Baek, J. W. (2022). High-speed monitoring of multidimensional processes using Bayesian updates. *IEEE Access*, 10:97450–97464.
- Li, W., Pu, X., Tsung, F., and Xiang, D. (2017). A robust self-starting spatial rank multivariate EWMA chart based on forward variable selection. *Computers & Industrial Engineering*, 103:116–130.
- Li, Z., Zhang, J., and Wang, Z. (2010). Self-starting control chart for simultaneously monitoring process mean and variance. *International Journal of Production Research*, 48(15):4537–4553.
- Lopuhaä, H. P. (1999). Asymptotics of reweighted estimators of multivariate location and scatter. *The Annals of Statistics*, 27(5).
- Lowry, C. A., Woodall, W. H., Champ, C. W., and Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics*, 34(1):46.
- Midi, H. and Shabbak, A. (2011). Robust multivariate control charts to detect small shifts in mean. *Mathematical Problems in Engineering*, 2011:1–19.
- Pignatiello, J. J. and Runger, G. C. (1990). Comparisons of multivariate CUSUM charts. *Journal of Quality Technology*, 22(3):173–186.
- Pison, G., Aelst, S. V., and Willems, G. (2002). Small sample corrections for LTS and MCD. *Metrika*, 55(1-2):111–123.

- Quesenberry, C. P. (1991). SPC Q charts for start-up processes and short or long runs. *Journal of Quality Technology*, 23(3):213–224.
- R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rocke, D. M. and Woodruff, D. L. (1996). Identification of outliers in multivariate data. *Journal of the American Statistical Association*, 91(435):1047–1061.
- Rousseeuw, P. and Hubert, M. (2013). High-breakdown estimators of multivariate location and scatter. In Becker, C., Fried, R., and Kuhnt, S., editors, *Robustness and complex data structures: Festschrift in honour of Ursula Gather*, pages 49–66. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Rousseeuw, P. J. and Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223.
- Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust regression and outlier detection*. Wiley series in probability and mathematical statistics. Wiley, New York.
- Stoumbos, Z. G. and Sullivan, J. H. (2002). Robustness to non-normality of the multivariate EWMA control chart. *Journal of Quality Technology*, 34(3):260–276.
- Testik, M. C., Runger, G. C., and Borror, C. M. (2003). Robustness properties of multivariate EWMA control charts. *Quality and Reliability Engineering International*, 19(1):31–38.
- Todorov, V. and Filzmoser, P. (2009). An object-oriented framework for robust multivariate analysis. *Journal of Statistical Software*, 32(3):1–47.
- Van Aelst, S. and Rousseeuw, P. (2009). Minimum volume ellipsoid. *WIREs Computational Statistics*, 1(1):71–82.
- Vargas, N. J. A. (2003). Robust estimation in multivariate control charts for individual observations. *Journal of Quality Technology*, 35(4):367–376.
- Willems, G., Pison, G., Rousseeuw, P. J., and Van Aelst, S. (2002). A robust Hotelling test. *Metrika*, 55(1-2):125–138.
- Woodall, W. H. and Montgomery, D. C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology*, 31(4):376–386.
- Woodall, W. H. and Ncube, M. M. (1985). Multivariate CUSUM quality-control procedures. *Technometrics*, 27(3):285–292.
- Woodruff, D. L. and Rocke, D. M. (1994). Computable robust estimation of multivariate location and shape in high dimension using compound estimators. *Journal of the American Statistical Association*, 89:888–896.

- Xie, X. and Qiu, P. (2022). Robust monitoring of multivariate processes with short-ranged serial data correlation. *Quality and Reliability Engineering International*, 38(8):4196–4209.
- Xie, X. and Qiu, P. (2024). A general framework for robust monitoring of multivariate correlated processes. *Technometrics*, 66(1):40–54.
- Xue, L. and Qiu, P. (2021). A nonparametric CUSUM chart for monitoring multivariate serially correlated processes. *Journal of Quality Technology*, 53(4):396–409.
- Zhou, M., Liu, L., Geng, W., and Zhou, J. (2015). Multivariate control chart based on multivariate Smirnov test. *Communications in Statistics - Simulation and Computation*, 44(6):1600–1611.
- Zou, C., Wang, Z., and Tsung, F. (2012). A spatial rank-based multivariate EWMA control chart. *Naval Research Logistics (NRL)*, 59(2):91–110.
- Zwetsloot, I. M., Schoonhoven, M., and Does, R. J. M. M. (2014). A robust estimator for location in phase I based on an EWMA chart. *Journal of Quality Technology*, 46(4):302–316.

Supplemental Materials: [paper title here]

