processes

MDPI

Article

Robust Max-Half-Mchart Based on the Cellwise Minimum Covariance Determinant

Syafi' Bariq' Syihabuddin Hidayatullah, Muhammad Ahsan * and Wibawati

Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia; syafibariq@gmail.com (S.B.S.H.); wibawati@its.ac.id (W.)
* Correspondence: muh.ahsan@its.ac.id

Abstract

One of the main tools in Statistical Process Control (SPC) for monitoring quality is the control chart. The Max-Half-Mchart is a Shewhart-type simultaneous multivariate control chart designed to detect shifts in both process mean and variability. However, outliers can distort the estimation of process parameters used to set control limits, leading to masking and swamping effects. Recent studies have highlighted the importance of cellwise contamination, which can reduce the effectiveness of casewise robust estimators. To overcome this limitation, this study develops a robust Max-Half-Mchart using the cellwise Minimum Covariance Determinant (cellMCD) estimator for location and covariance estimation. The proposed chart was evaluated through simulation studies, average run length analysis, and applications to synthetic and real OPC cement quality data. Simulation results under different correlation levels and contamination proportions show that the proposed chart provides more stable outlier detection performance than the conventional Max-Half-Mchart and the Fast-MCD-based Max-Half-Mchart, with better discrimination between normal and contaminated observations. The ARL analysis also indicates faster detection of small to moderate shifts. In the synthetic-data application, it achieved an Accuracy of 0.9899 and an AUC of 0.9939 under 20% contamination, and in the real-data application it detected seven out-of-control signals. Overall, the findings demonstrate that incorporating cellMCD into the Max-Half-Mchart provides a more robust and effective approach for multivariate process monitoring under cellwise contamination.

Keywords: simultaneous multivariate control charts; max-half-mchart; cellwise contamination; cellMCD

Check for updates

Academic Editor: Jie Zhang

Received: 22 February 2026
Revised: 22 March 2026
Accepted: 27 March 2026
Published: 31 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Statistical Process Control (SPC) is an important approach to monitor product quality and keep a production process consistent. SPC monitors the process mean and variability across several quality characteristics, so process problems can be detected early before products go beyond specification limits [1]. One of the main SPC tools is the control chart, introduced by Walter A. Shewhart in 1924 to distinguish common cause variation from special cause variation in a process [2]. Based on the number of quality characteristics, control charts are classified into univariate and multivariate charts. Univariate charts monitor one characteristic at a time, while multivariate control charts monitor two or more correlated characteristics at the same time.

A wide range of multivariate control charts has been developed in the literature. Hotelling's T^2 chart extends the t-test and the $\overline{X}$ chart to detect shifts in the process mean

vector [3]. To better detect small mean shifts, MCUSUM and MEWMA charts have been proposed [4,5]. For monitoring changes in multivariate variability, the generalized variance chart uses the determinant of the sample covariance matrix [1]. For individual observations, a successive difference approach has also been introduced [6]. Other variability measures include IGV, vector variance, and methods that combine generalized variance with the trace of the covariance matrix [7–9]. In addition, some charts focus on changes in the correlation structure [10], and the $\mathrm{Tr}(R^2)$ chart has been proposed for subgroup data [11].

Simultaneous control charts have also been developed to monitor the process mean and variability at the same time. A Shewhart-type simultaneous chart was first introduced using a modified boxplot for subgroup data [12]. Later, the Max Chart was proposed by combining the $\overline{X}$ chart and the S chart [13]. For the bivariate case, a Max Chart based on Hotelling's T^2 statistic and the generalized variance was developed for subgroup data [14]. This idea was then extended to the multivariate setting and became known as the Max-Mchart [15]. Further work includes the Max-Mchart Chart with a gamma distribution transformation [16]. In addition, the Max-Mchart and the Max-Half-Mchart have been developed for individual observations using standard normal and half normal transformations, and they have also been studied for subgroup data [17,18].

However, conventional control charts are known to be sensitive to outliers. Outliers can affect the estimation of process parameters used to set control limits, which may result in wider or stretched control limits [19]. A key weakness of classical estimators is their vulnerability to masking and swamping effects. Masking refers to a loss of detection power when the true number of outliers is larger than assumed, so extreme outliers may not be detected, leading to false negatives [20]. In contrast, swamping occurs when normal observations are incorrectly flagged as outliers, leading to false positives [21].

Several studies have been developed to address the problem of casewise outliers. Robust estimators such as the Minimum Volume Ellipsoid (MVE) and the Minimum Covariance Determinant (MCD) were introduced as outlier resistant estimators of location and covariance. MVE searches for the minimum volume ellipsoid covering at least h observations, while MCD selects an h-subset with the smallest covariance determinant [22]. To reduce computational cost, the Fast MCD algorithm was proposed [23]. The efficiency of MCD was later improved by reweighting, leading to RMCD [24], and a deterministic version was introduced as Det MCD [25]. For fat or high-dimensional data, the Minimum Regularized Covariance Determinant (MRCD) was proposed to stabilize robust covariance estimation [26]. MRCD was further extended through a kernel approach, known as Kernel MRCD, to provide more flexibility for non-elliptical data [27]. However, these robust estimators were primarily developed for casewise contamination, where entire observations are treated as outlying units, and therefore may be less effective when contamination occurs only in a subset of cells.

More recently, attention has shifted to cellwise outliers [28]. Cellwise contamination may cause outlier propagation; even if only a small fraction of cells is contaminated, many rows can become contaminated because a single outlying cell is enough to affect an entire observation. The expected proportion of contaminated rows is $1-(1-\varepsilon)^p$, where ε is the fraction of contaminated cells, and p is the number of variables [28]. In high-dimensional settings, a small ε can therefore lead to a large fraction of contaminated rows, which may reduce the accuracy of casewise robust methods. A widely used method for cellwise detection is Detecting Deviating Data Cells (DDC), which predicts each cell and flags cells that deviate strongly from their predictions [29]. Several approaches have also been proposed to obtain covariance estimators that are robust under cellwise contamination, for example, using rank-based correlations, robust pairwise correlations, or wrapping transformations to produce a positive semidefinite covariance matrix [30–33]. However, these methods follow a two-step strategy, where outlier detection is separated from parameter estimation. To

overcome this limitation, the cellMCD estimator was introduced to jointly estimate location and covariance while detecting cellwise outliers, using an observed likelihood framework for incomplete data and a penalty term that limits the number of flagged cells [34]. This estimator is robust under cellwise contamination, efficient for clean data, and implemented with C-steps to ensure improvement of the objective function at each iteration.

The development of robust estimators has motivated various modifications of conventional control charts that are sensitive to outliers. Several studies have incorporated robust estimators into multivariate control charts, including MVE-, RMCD-, MVV-, Fast-MCD-, Det-MCD-, and MRCD-based approaches for Hotelling's T^2 and Max-Half-Mchart frameworks [35–40]. In general, these studies reported improved monitoring performance compared with their conventional counterparts under contaminated data. However, most of these robust approaches were developed under the casewise contamination framework, where entire observations are treated as outlying units. As a result, their effectiveness may decrease when contamination occurs only in a subset of cells within an observation.

Based on these considerations, this study addresses the limitations of casewise robust estimation under cellwise contamination by proposing a robust Max-Half-Mchart based on the cellwise Minimum Covariance Determinant (cellMCD) estimator. Its performance is then compared with that of the conventional Max-Half-Mchart and the Fast-MCD-based Max-Half-Mchart. The remainder of this paper is organized as follows: Section 2 introduces the proposed cellMCD-based Max-Half-Mchart. Section 3 describes the methodology and implementation procedures. Section 4 evaluates the chart's performance in detecting outliers and process shifts, and also presents its applications to synthetic and real data. Finally, Section 5 provides conclusions and directions for future research.

2. Materials and Methods

2.1. Max-Half-Mchart

The Max-Mchart is known to have limitations in providing accurate results due to its use of transformations based on the standard normal distribution. This limitation motivates an alternative approach to avoid negative-domain issues when computing quantiles. One proposed solution is to use the half-normal distribution, which has a positive support from 0 to ∞. This idea forms the basis for developing the Max-Half-Mchart.

The Max-Half-Mchart is constructed using the same combined framework as the Max-Mchart, namely Hotelling's T^2 and the successive difference approach. The simultaneous statistics for monitoring the process mean and variability in the Max-Half-Mchart for individual observations are defined as follows [17]:

$$Z_i^{IH} = \mathrm{Q}^{-1}\left[H_p\left\{(\mathbf{x}_i - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_0)\right\}\right], \tag{1}$$

and

$$V_i^{IH} = \mathrm{Q}^{-1}\left[H_p\left\{\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_{i-1})'\boldsymbol{\Sigma}_0^{-1}(\mathbf{x}_i - \mathbf{x}_{i-1})\right\}\right], \tag{2}$$

where $Q(\cdot)$ denotes the cumulative distribution function of the standard half-normal distribution, and $H_p(\cdot)$ denotes the cumulative distribution function of the chi square distribution with p degrees of freedom. Because the half-normal transformation produces statistics on a positive support, the absolute value is no longer required. The Max-Half-Mchart statistic is then defined by

$$M_i^{IH} = \max\left\{Z_i^{IH}, V_i^{IH}\right\},\ i = 2, 3 \ldots, n. \tag{3}$$

The Max-Half-Mchart uses only an upper control limit, since $M_i^{IH} > 0$. Because the distribution of M_i^{IH} is not available in closed form and is typically skewed, the

upper control limit is estimated using a nonparametric procedure such as a bootstrap approach [17].

2.2. Cellwise Minimum Covariance Determinant (CellMCD)

The Minimum Covariance Determinant (MCD) estimator is theoretically defined by selecting an h-subset of observations that minimizes the determinant of the covariance matrix. MCD is a robust estimator designed mainly for casewise outliers, where entire rows are atypical. However, in several practical applications, contamination is cellwise. Only some entries in a row are outlying, while the remaining cells still contain valid information. Under dominant cellwise contamination, discarding whole rows may remove useful information and reduce efficiency. To address this issue, cellMCD was developed as a robust estimator of location and covariance for cellwise contaminated data, as a cellwise extension of MCD [34].

CellMCD introduces a binary weight matrix $\mathbf{W} = (w_{ij}) \in \{0,1\}^{n \times p}$, where n denotes the number of observations, p the number of variables, and $w_{ij} = 1$ indicates that cell (i, j) is retained as clean, whereas $w_{ij} = 0$ indicates that the cell is flagged as outlying or treated as missing. The notation $\mathbf{x}_i^{(w_i)}$, $d(w_i)$, $\boldsymbol{\mu}^{(w_i)}$, and $\boldsymbol{\Sigma}^{(w_i)}$ follows the observed likelihood framework for incomplete data [41]. The squared distance term is defined as the partial Mahalanobis distance $\mathrm{MD}(\mathbf{x}_i, \boldsymbol{w}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sqrt{\left(\mathbf{x}_i^{(w_i)} - \boldsymbol{\mu}^{(w_i)}\right)' \left(\boldsymbol{\Sigma}^{(w_i)}\right)^{-1} \left(x_i^{(w_i)} - \boldsymbol{\mu}^{(w_i)}\right)}$ computed on the observed cells [42]. The matrix $\mathbf{W}$ is not specified in advance but is estimated jointly with the model parameters. The trimming level h is applied to the number of unflagged cells per column rather than to the number of unflagged cases, so that h represents the minimum number of retained cells required in each column. Following Raymaekers and Rousseeuw, h should not be chosen too low, since this may lead to poorly defined covariance estimates and numerical instability. Therefore, $h = 0.75n$ is recommended and used as the default setting in the cellMCD implementation. The cellMCD objective function is defined as follows [34]:

$$G(\boldsymbol{x}_i, \boldsymbol{w}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{q}) = \sum_{i=1}^{n} \left(\ln \left|\boldsymbol{\Sigma}^{(w_i)}\right| + d(\boldsymbol{w}_i)\ln(2\pi) + \mathrm{MD}^2(\boldsymbol{x}_i, \boldsymbol{w}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma})\right) + \sum_{j=1}^{p} q_j \left\| \mathbf{1}_d - \mathbf{W}_{\cdot j} \right\|_0, \tag{4}$$

under the constraints $\lambda_d(\boldsymbol{\Sigma}) \geq a$ and $\| \mathbf{W}_{\cdot j} \|_0 \geq h$, for $j = 1, \ldots, p$. Then, q_j denotes the penalty constant used in the objective function to control the number of flagged cells in column j. The penalty parameter q_j is computed as

$$q_j = \chi^2_{1,p} + \ln(2\pi) + \ln(C_j), \tag{5}$$

where $\chi^2_{1,p}$ denotes the p-quantile of the chi-square distribution with one degree of freedom. The term C_j is the conditional variance of variable j based on the initial covariance estimate $\hat{\boldsymbol{\Sigma}}_0$, defined by $1/(\hat{\boldsymbol{\Sigma}}_0^{-1})_{jj}$. The value $\theta = 0.99$ is recommended because it leads to more efficient estimates [34].

In the first stage, the data are robustly standardized columnwise using [29]. Next, initial estimates are obtained using the DDCW estimator [43]. The algorithm then applies the C-step (Part A), which updates the weight matrix $\mathbf{W}$ at iteration k while keeping $\hat{\boldsymbol{\mu}}^{(k)}$ and $\hat{\boldsymbol{\Sigma}}^{(k)}$ fixed. For each cell, the update is based on the change in the objective function when comparing $w_{ij} = 1$ with $w_{ij} = 0$. This change is denoted by Δ_{ij} and defined as follows:

$$\begin{aligned}\Delta_{ij} &= G(\boldsymbol{x}_i, \boldsymbol{w}_i = 1, \hat{\boldsymbol{\mu}}^{(k)}, \hat{\boldsymbol{\Sigma}}^{(k)}, \boldsymbol{q}) - G(\boldsymbol{x}_i, \boldsymbol{w}_i = 0, \hat{\boldsymbol{\mu}}^{(k)}, \hat{\boldsymbol{\Sigma}}^{(k)}, \boldsymbol{q}) \\ &= \ln(C_{ij}) + \ln(2\pi) + \frac{(x_{ij} - \hat{x}_{ij}^{(k)})^2}{C_{ij}^{(k)}} - q_j\end{aligned} \tag{6}$$

where $\hat{x}_{ij} = \mu_j + \boldsymbol{\Sigma}_{j,o}\boldsymbol{\Sigma}_{o,o}^{-1}(\mathbf{x}_{i,o} - \boldsymbol{\mu}_o)$ and $C_{ij} = \Sigma_{j,j} - \boldsymbol{\Sigma}_{j,o}\boldsymbol{\Sigma}_{o,o}^{-1}\boldsymbol{\Sigma}_{o,j}$ are the conditional mean and conditional variance given the currently unflagged (observed) cells in row i (indexed by o). If $\Delta_{ij} \leq 0$ holds for at least h indices i in column j, the objective is minimized by setting $w_{ij} = 1$ for those cells and $w_{ij} = 0$ for the remaining cells. Otherwise, the minimum is attained by setting $w_{ij} = 1$ for the h observations with the smallest Δ_{ij} values and $w_{ij} = 0$ for the others. After cycling through all columns, the updated weight matrix $\mathbf{W}^{(k+1)}$ is obtained.

After obtaining $\mathbf{W}^{(k+1)}$ from the C-step (Part A), the C-step (Part B) updates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ using an EM-type framework for incomplete data, where flagged cells $(w_{ij} = 0)$ are treated as missing. For each observation i, the index set of retained (observed) cells is defined as $o_i^{(k)} = \left\{ j : w_{ij}^{(k+1)} = 1 \right\}$ and $m_i^{(k)} = \left\{ j : w_{ij}^{(k+1)} = 0 \right\}$, with $o_i^{(k)} \cup m_i^{(k)} = \{1, \ldots, p\}$ and $o_i^{(k)} \cap m_i^{(k)} = \emptyset$. The data vector $\mathbf{x}_i$, mean $\boldsymbol{\mu}$, and covariance matrix $\boldsymbol{\Sigma}$ are partitioned into observed and missing blocks, which allows structured imputation and parameter updating.

In the E-step, for rows with missing cells $\left(m_i^{(k)} \neq \emptyset\right)$, the missing block is imputed by the conditional mean given the observed block,

$$\hat{x}_{i,m_i^{(k)}}^{(k)} = \hat{\boldsymbol{\mu}}_{m_i^{(k)}}^{(k)} + \hat{\boldsymbol{\Sigma}}_{m_i^{(k)},o_i^{(k)}}^{(k)} (\hat{\boldsymbol{\Sigma}}_{o_i^{(k)},o_i^{(k)}}^{(k)})^{-1} (x_{i,o_i^{(k)}} - \hat{\mu}_{o_i^{(k)}}^{(k)}). \tag{7}$$

and the corresponding conditional covariance is computed to correct for the variance contribution of imputation,

$$\mathbf{C}_i^{(k)} = \hat{\boldsymbol{\Sigma}}_{m_i^{(k)},m_i^{(k)}}^{(k)} - \hat{\boldsymbol{\Sigma}}_{m_i^{(k)},o_i^{(k)}}^{(k)} (\hat{\boldsymbol{\Sigma}}_{o_i^{(k)},o_i^{(k)}}^{(k)})^{-1} \hat{\boldsymbol{\Sigma}}_{o_i^{(k)},m_i^{(k)}}^{(k)}. \tag{8}$$

These quantities define the imputed matrix $\mathbf{X}_{\text{imp}}^{(k)}$, where $x_{ij}^{\text{imp},(k)} = x_{ij}$ for observed cells, and $x_{ij}^{\text{imp},(k)} = \hat{x}_{ij}^{(k)}$ for flagged cells.

In the M-step, the mean estimator is updated as the average of the imputed observations,

$$\hat{\boldsymbol{\mu}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} x_i^{imp,(k)}, \tag{9}$$

and the covariance estimator is obtained from the sample covariance of $\mathbf{X}_{\text{imp}}^{(k)}$ plus a bias-correction matrix $\mathbf{B}^{(k)}$. The matrix $\mathbf{B}^{(k)}$ aggregates the contributions of $\mathbf{C}_i^{(k)}$ on the missing blocks for each row and is averaged across all observations. The updated covariance is therefore

$$\hat{\boldsymbol{\Sigma}}^{(k+1)} = \frac{1}{n}(x_{ij}{}^{imp,(k)} - 1_n\hat{\boldsymbol{\mu}}^{(k)\prime})'(x_{ij}{}^{imp,(k)} - 1_n\hat{\boldsymbol{\mu}}^{(k)\prime}) + \frac{1}{n}\mathbf{B}^{(k)}, \tag{10}$$

To keep the covariance matrix well-conditioned, the constraint $\lambda_{\min}\left(\hat{\boldsymbol{\Sigma}}^{(k+1)}\right) \geq a$ is enforced with $a = 10^{-4}$ by truncating small eigenvalues, and the resulting matrix is denoted by $\overset{\sim}{\boldsymbol{\Sigma}}^{(k+1)}$. The C-step (Parts A and B) is repeated until convergence. After convergence at iteration k, the standardized-scale estimates are $\hat{\boldsymbol{\mu}}_s = \hat{\boldsymbol{\mu}}^{(k)}$ and $\hat{\boldsymbol{\Sigma}}_s = \overset{\sim}{\boldsymbol{\Sigma}}^{(k)}$. The estimates on the original scale are obtained by back-transformation as

$$\overline{\mathbf{x}}_{cellMCD} = \mathbf{m} + \hat{\boldsymbol{\mu}}_s \cdot \mathbf{s} \tag{11}$$

$$\mathbf{S}_{cellMCD} = \mathbf{D}\,\hat{\mathbf{\Sigma}}_s\mathbf{D} \tag{12}$$

where $\mathbf{D} = \mathrm{diag}(\mathbf{s})$.

2.3. CellMCD Based Max-Half-Mchart

This study adapts the Max-Half-Mchart by substituting the classical estimators of the mean vector $\bar{\mathbf{x}}$ and covariance matrix $\mathbf{S}$ with robust estimates of location and scatter in Equations (11) and (12). The replacements are obtained from the cellMCD estimator, which is designed to reduce the influence of cellwise contamination. As a result, the control statistics become more reliable under outliers. The statistics for monitoring the process mean vector and variability for individual observations based on cellMCD are defined as follows:

$$Z^{IH}_{cellMCD_i} = \mathrm{Q}^{-1}\left[H_p\left\{(\mathbf{x}_i - \bar{\mathbf{x}}_{cellMCD})'\mathbf{S}^{-1}_{cellMCD}(\boldsymbol{x}_i - \bar{\mathbf{x}}_{cellMCD})\right\}\right], \tag{13}$$

and

$$V^{IH}_{cellMCD_i} = \mathrm{Q}^{-1}\left[H_p\left\{\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_{i-1})'\mathbf{S}^{-1}_{cellMCD}(\mathbf{x}_i - \mathbf{x}_{i-1})\right\}\right], \tag{14}$$

Then, the cellMCD-based Max-Half-Mchart statistic is constructed by combining the mean and variability components into a single monitoring statistic. The cellMCD-based Max-Half-Mchart statistic is given by:

$$M^{IH}_{cellMCD_i} = \max\left\{Z^{IH}_{cellMCD_i}, V^{IH}_{cellMCD_i}\right\},\ i = 2, 3\ldots, n \tag{15}$$

Estimating the upper control limit (UCL) for the Max-Half-Mchart is carried out using a resampling strategy based on bootstrap methods [17]. In practice, calibrating the chart to achieve a target in-control average run length (ARL) is computationally demanding. For this reason, Monte Carlo simulation is used to approximate the in-control behavior of the chart and to link candidate UCL values to the resulting ARL [38]. In this study, the UCL is therefore determined through a combined bootstrap–Monte Carlo procedure, which is described in Algorithm 1.

Algorithm 1 Procedure to Compute and Plot the Proposed cellMCD-Based Max-Half-Mchart

Step 1. Set $\alpha = 0.00273$.

Step 2. Compute the robust mean vector $\bar{\mathbf{x}}_{cellMCD}$ and robust covariance matrix $\mathbf{S}_{cellMCD}$ using Equations (11) and (12).

Step 3. Calculate the mean statistic $Z^{IH}_{cellMCD_i}$ and the variability statistic $V^{IH}_{cellMCD_i}$ using Equations (13) and (14). Then form the simultaneous statistic $M^{IH}_{cellMCD_i}$ using Equation (15).

Step 4. Estimate the upper control limit $UCL_{cellMCD}$ for the proposed chart using the bootstrap–Monte Carlo procedure in Algorithm 2.

Step 5. Plot $M^{IH}_{cellMCD_i}$ against $UCL_{cellMCD}$.

1. If $Z^{IH}_{cellMCD_i} > UCL_{cellMCD}$, classify the signal as M^{++} and conclude a shift in the process mean.
2. If $V^{IH}_{cellMCD_i} > UCL_{cellMCD}$, classify the signal as V^{++} and conclude a shift in process variability.
3. If $Z^{IH}_{cellMCD_i} > UCL_{cellMCD}$ and $V^{IH}_{cellMCD_i} > UCL_{cellMCD}$, classify the signal as B^{++} and conclude shifts in both mean and variability.

Algorithm 2 Bootstrap and Monte Carlo Control Limit

Step 1. Set $\alpha = 0.00273$.

Step 2. Obtain the robust estimates $\bar{\mathbf{x}}_{cellMCD}$ and $\mathbf{S}_{cellMCD}$ using Equations (11) and (12).

Step 3. For $l = 1, \ldots, 1000$, repeat:

a. Generate n observations from the multivariate normal distribution $N_p(\bar{\mathbf{x}}_{cellMCD}, \mathbf{S}_{cellMCD})$

b. Compute the cellMCD-based Max-Half-Mchart statistic $M^{IH}_{cellMCD_i}$ for $i = 2, 3, \ldots, n$ using the simulated sample.

c. Apply bootstrap resampling (with replacement) to obtain 1000 bootstrap values of $M^{IH}_{cellMCD_i}$.

d. For replication l, compute the empirical $100(1-\alpha)$ th percentile of the 1000 bootstrap values, denoted by $M^{IH}_{cellMCD_{(100(1-\alpha)),l}}$

Step 4. Estimate the UCL by averaging these percentile values across the 1000 Monte Carlo replications:

$$UCL_{cellMCD} = \frac{1}{1000}\sum_{i=1}^{1000} M^{IH}_{cellMCD_{(100(1-\alpha)),l}}.$$

Step 5. Using the estimated UCL, compute the in-control average run length (ARL_0) as the mean run length over 1000 replications.

$$\mathrm{ARL}_0 = \frac{\sum_{i=1}^{1000} RL_i}{1000}.$$

Step 6. Repeat Step 5 until $\mathrm{ARL}_0 \approx 370$. At each iteration, update the control limit as $UCL_{cellMCD,\, i+1} = UCL_{cellMCD,\, i} \pm 0.005$, decreasing the UCL when $\mathrm{ARL}_0 > 370$ and increasing the UCL when $\mathrm{ARL}_0 < 370$.

3. Methodology

This section presents the methodological framework used in this study. The proposed cellMCD-based Max-Half-Mchart is implemented in two main stages. First, the cellMCD-based Max-Half-Mchart statistics are computed. Second, the upper control limit is calibrated using a combined bootstrap–Monte Carlo procedure, since the distribution of the charting statistic is not available in closed form and the target in-control ARL must therefore be approximated numerically. Algorithms 1 and 2 summarize these two stages of the proposed procedure.

The performance evaluation of the control chart includes ARL under process shifts and a confusion matrix assessment for outlier detection. First, ARL is used to assess how quickly the chart detects process shifts. Second, a confusion-matrix-based evaluation is used to assess the chart's detection capability in distinguishing contaminated (out-of-control) observations from uncontaminated (in-control) observations under cellwise contamination. In the confusion-matrix-based evaluation, the classification results are divided into correct decisions and misclassification errors, as presented in Table 1.

Table 1. Confusion matrix table.

Actual	Detection	
	Outlier	**Normal**
Outlier	True Positives (TPs)	False Negatives (FNs)
Normal	False Positives (FPs)	True Negatives (TNs)

True positives (TPs) refer to outlier observations that are correctly identified as outliers. True negatives (TNs) refer to normal observations that are correctly identified as normal. Misclassification errors are also grouped into two types. False positives (FPs) occur when a normal observation is incorrectly flagged as an outlier, which leads to a false alarm. False negatives (FNs) occur when an outlier observation is missed and incorrectly classified as normal, which results in an undetected outlier.

These errors are summarized using the false positive rate (FP Rate) and false negative rate (FN Rate) [44].

$$FP\ Rate = \frac{FP}{TN + FP} \tag{16}$$

$$FN\ Rate = \frac{FN}{TP + FN} \tag{17}$$

In addition, the overall detection ability is measured using the area under the curve (AUC), defined as [45]:

$$AUC = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{18}$$

The specific simulation settings, contamination scenarios, and parameter configurations used for outlier-detection and shift-detection evaluation are presented in Section 4.

4. Results

4.1. Detecting Outlier

In this section, the performance of the robust Max-Half-Mchart is evaluated through a simulation study to assess its ability to detect outliers. The simulation begins by generating multivariate normal data $\mathbf{X} \sim N_p(\mu_0, \Sigma_0)$, where the mean vector is $\mu_0 = (0, \ldots, 0)'$ of dimension $p \times 1$. The sample size in each simulated dataset is set to $n = 1000$, and the number of quality characteristics is fixed at $p = 5$. The variance of each quality characteristic is set to 1, and an exchangeable correlation structure ρ is assumed across characteristics. Therefore, the covariance matrix is defined as $\mathbf{\Sigma}_0 = \begin{pmatrix} 1 & \rho & \ldots & \rho \\ \rho & 1 & \ldots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \ldots & 1 \end{pmatrix}$, with $\rho \in \{0.3,\ 0.5,\ 0.7\}$.

Next, cellwise contamination is introduced by replacing a fraction ε of cells with contaminated values. For each column of the data matrix, $n\varepsilon$ cell indices are randomly selected for contamination, and the selected cells are then replaced by contaminated values controlled by the parameter γ. A larger γ indicates more severe deviations in the contaminated cells. In this study, $\gamma = 5$ is used. This contamination mechanism is implemented using the *generateData* function from the R package cellWise 2.5.7 [46]. The outlier proportion is varied across four scenarios: 5%, 10%, 15%, and 20%. The proportion of cellwise outlier contamination is limited to 20% because it remains below the theoretical breakdown point of the cellMCD estimator, which is approximately 25% under the default choice $h = 0.75n$, given by $(n - h + 1)/n \approx 0.25$ [34].

For performance evaluation, each observation that contains at least one contaminated cell is treated as a contaminated observation and assigned label 1; otherwise, it is assigned label 0. Furthermore, because the Max-Half statistic is computed based on successive difference $(t-1, t)$, a *pair label* is defined for each pair. The pair is assigned label 1 if at least one of the two observations in $(t-1, t)$ is contaminated, and label 0 if both are normal. The proposed robust Max-Half-Mchart based on cellMCD is compared with the conventional chart and the robust Max-Half-Mchart based on Fast-MCD. The chart performance is considered satisfactory when it achieves high Accuracy and AUC, along with a low false positive rate (FP rate) and false negative rate (FN rate).

Based on the simulation results, detailed results are presented in Tables 2–4. Table 2 shows that the conventional Max-Half-Mchart performs poorly under all combinations of correlation levels and outlier proportions. In general, its Accuracy decreases as the contamination proportion increases, while the FN rate becomes extremely high, and the AUC remains close to 0.5. For instance, when $\rho = 0.3$, the Accuracy drops from about 0.63 at 5% outliers to around 0.11 at 20% outliers, while the FN rate increases from about 0.93 to nearly 1.00. A similar pattern is observed at $\rho = 0.5$ and $\rho = 0.7$. These results indicate that the conventional chart is highly sensitive to cellwise contamination and has limited ability to distinguish contaminated observations from normal ones.

Table 3 presents the performance of the Fast-MCD-based Max-Half-Mchart. Compared with the conventional chart, this robust version shows a clear improvement, particularly when the contamination proportion is still low, and the correlation among variables is moderate to high. Under favorable conditions, the chart achieves very high Accuracy and AUC, together with a low FN rate. However, its performance still declines as the contamination level increases, especially when the correlation is weak. For example, at $\rho = 0.3$, the Accuracy decreases substantially, and the FN rate rises sharply when the outlier proportion increases from 10% to 20%. This finding suggests that Fast-MCD improves detection performance relative to the conventional approach, but its robustness is still limited when cellwise contamination becomes more severe.

Table 2. Simulation result for conventional Max-Half-Mchart.

ρ	%Out	Conventional Max-Half-Mchart			
		Accuracy	**FP Rate**	**FN Rate**	**AUC**
0.3	5	0.6278	0.0000	0.9273	0.5363
	10	0.3532	0.0000	0.9926	0.5037
	15	0.1977	0.0000	0.9992	0.5004
	20	0.1069	0.0000	0.9999	0.5001
0.5	5	0.6366	0.0000	0.9044	0.5478
	10	0.3559	0.0000	0.9892	0.5054
	15	0.1975	0.0000	0.9985	0.5007
	20	0.1076	0.0000	0.9997	0.5002
0.7	5	0.6557	0.0000	0.8579	0.5710
	10	0.3602	0.0000	0.9823	0.5089
	15	0.1984	0.0000	0.9975	0.5013
	20	0.1081	0.0000	0.9994	0.5003

Table 3. Simulation result for Max-Half-Mchart based on Fast-MCD.

ρ	%Out	Fast-MCD-Based Max-Half-Mchart			
		Accuracy	**FP Rate**	**FN Rate**	**AUC**
0.3	5	0.9150	0.0025	0.2080	0.8948
	10	0.8262	0.0008	0.2660	0.8666
	15	0.4353	0.0000	0.7018	0.6491
	20	0.2131	0.0000	0.8816	0.5592
0.5	5	0.9559	0.0025	0.1059	0.9458
	10	0.9390	0.0019	0.0925	0.9528
	15	0.7933	0.0001	0.2571	0.8714
	20	0.6166	0.0000	0.4293	0.7853
0.7	5	0.9896	0.0025	0.0222	0.9877
	10	0.9882	0.0023	0.0169	0.9904
	15	0.9067	0.0009	0.1160	0.9416
	20	0.8119	0.0002	0.2107	0.8945

Table 4. Simulation result for Max-Half-Mchart based on cellMCD.

ρ	%Out	cellMCD-Based Max-Half-Mchart			
		Accuracy	**FP Rate**	**FN Rate**	**AUC**
0.3	5	0.9073	0.0019	0.2279	0.8851
	10	0.8741	0.0015	0.1924	0.9031
	15	0.8764	0.0010	0.1535	0.9227
	20	0.8956	0.0008	0.1167	0.9413
0.5	5	0.9495	0.0018	0.1231	0.9375
	10	0.9301	0.0014	0.1064	0.9461
	15	0.9297	0.0011	0.0872	0.9558
	20	0.9400	0.0009	0.0671	0.9660
0.7	5	0.9885	0.0020	0.0258	0.9861
	10	0.9838	0.0015	0.0241	0.9872
	15	0.9836	0.0011	0.0202	0.9894
	20	0.9862	0.0011	0.0153	0.9918

Table 4 shows that the cellMCD-based Max-Half-Mchart provides the most stable performance across all simulation scenarios. In contrast to the conventional and Fast-MCD-based charts, the proposed chart maintains high Accuracy and AUC, very small FP rates, and relatively low FN rates, even as the outlier proportion increases. A notable pattern is that the FN rate tends to decrease rather than increase when contamination becomes larger, while the AUC remains high and even improves slightly in several cases. This pattern is observed consistently across low, moderate, and high correlation settings. Overall, Table 4 indicates that the cellMCD-based Max-Half-Mchart is the most robust and reliable among the competing methods for detecting cellwise contamination.

To make the performance comparison among charts clearer, radar charts are provided in Figures 1–3. These figures present the comparison of Accuracy, FN rate, and AUC under three correlation settings, (a) $\rho = 0.3$, (b) $\rho = 0.5$, and (c) $\rho = 0.7$, with outlier proportions

of 5%, 10%, 15%, and 20%. The FP rate is not visualized because its differences across charts are very small and thus do not provide additional insight.

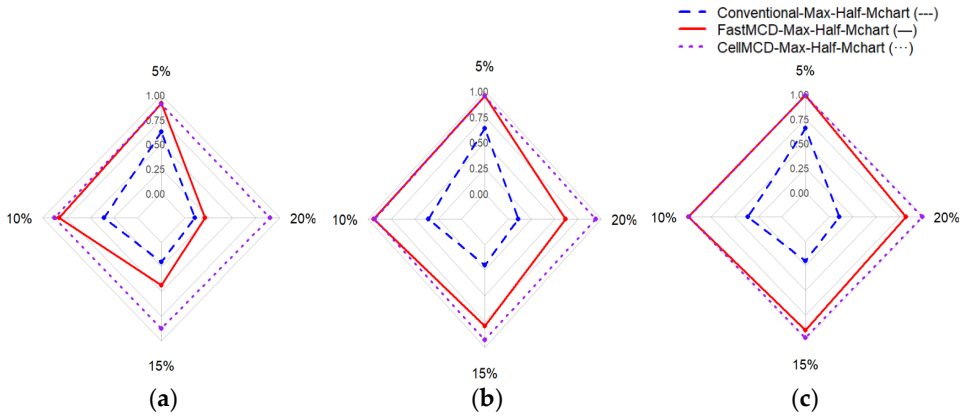


Figure 1. Accuracy performance for all charts based on various correlations: (**a**) $\rho = 0.3$, (**b**) $\rho = 0.5$, and (**c**) $\rho = 0.7$.

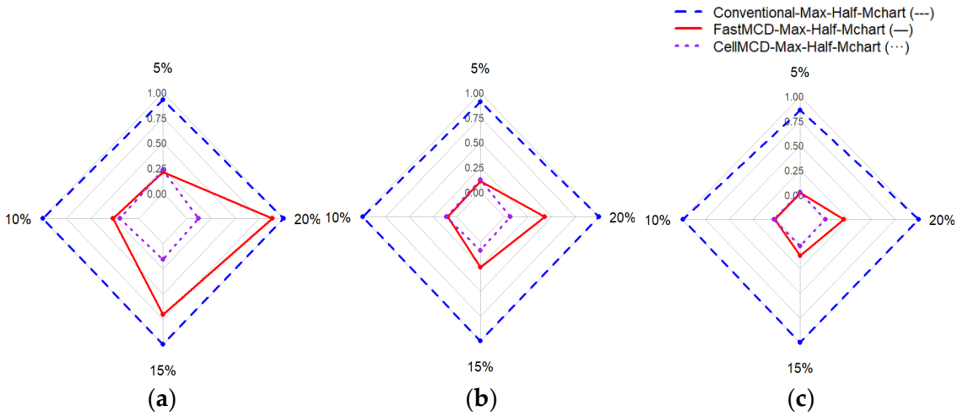


Figure 2. FN rate comparison for all charts based on various correlations: (**a**) $\rho = 0.3$, (**b**) $\rho = 0.5$, and (**c**) $\rho = 0.7$.

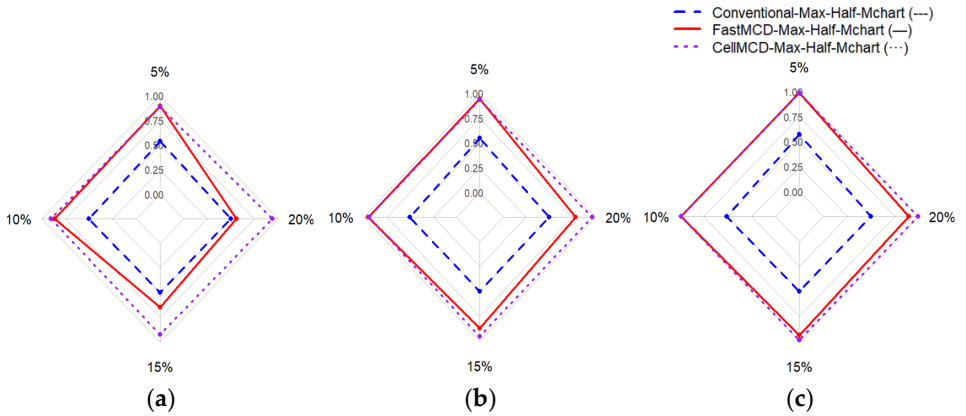


Figure 3. AUC performance for all charts based on various correlations: (**a**) $\rho = 0.3$, (**b**) $\rho = 0.5$, and (**c**) $\rho = 0.7$.

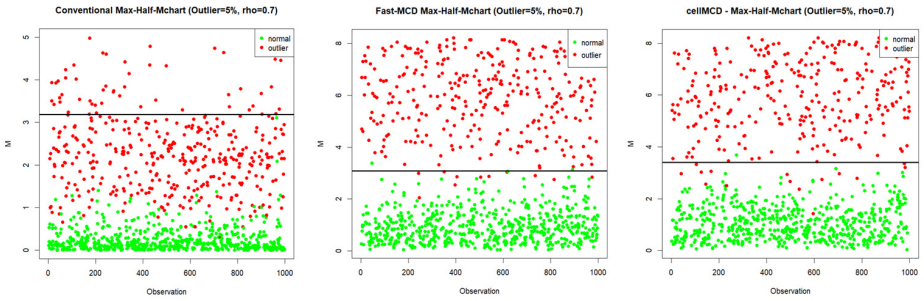
Figure 1 shows that the cellMCD-based Max-Half-Mchart generally forms the largest area across most scenarios, indicating the highest and most stable Accuracy as the outlier proportion increases. Figure 2 provides a more comprehensive view of outlier detection ability. The cellMCD-based Max-Half-Mchart produces the smallest area in all panels, reflecting the lowest FN rates. In contrast, the Fast-MCD–based Max-Half-Mchart lies in

between: it performs better than the conventional chart, but under higher contamination levels it still exhibits relatively larger FN rates in several cases.

Figure 3 shows a pattern consistent with the previous metrics. The cellMCD-based Max-Half-Mchart achieves the highest and most stable AUC values across all combinations of ρ and outlier proportions, indicating strong discrimination between normal and contaminated observations. The Fast-MCD-based Max-Half-Mchart performs well under low to moderate contamination, but its performance tends to decrease when the outlier proportion increases, particularly under low correlation. Conversely, the conventional chart consistently shows the lowest AUC values, suggesting weak separation capability under cellwise contamination.

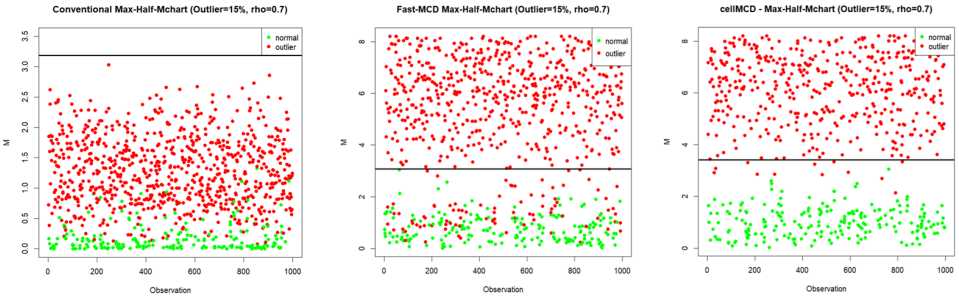
To clarify how each chart produces signals on the simulated data, Figures 4–6 present the Max-Half-Mchart statistics plots under three contamination levels, namely 5%, 15%, and 20%. Visually, the cellMCD-based Max-Half-Mchart provides the clearest separation between normal observations and outliers. Contaminated observations more frequently exceed the UCL, whereas normal observations tend to remain below the UCL. This pattern is consistent across all contamination levels, even when the outlier proportion increases to 20%.

In contrast, the conventional chart and the Fast-MCD-based Max-Half-Mchart still show many contaminated observations falling below the UCL, especially under higher contamination, indicating a higher false negative rate. Therefore, Figures 4–6 support the results reported in the previous tables and confirm that the cellMCD-based Max-Half-chart is the most reliable approach for detecting cellwise contamination, as it maintains high detection sensitivity without excessively increasing false alarms.



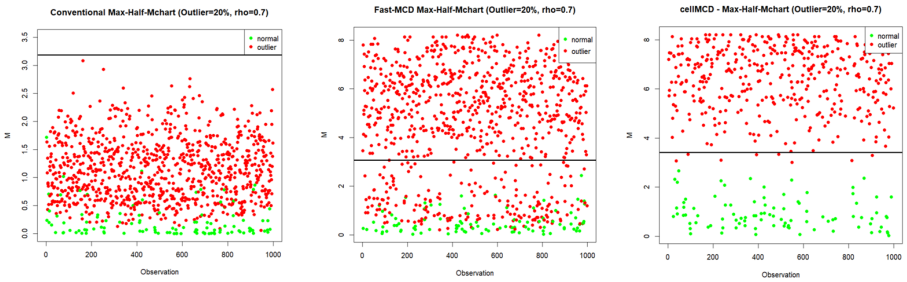
(**a**) (**b**) (**c**)

Figure 4. Statistics plot of simulation on 5% outliers using (**a**) Max-Half-Mchart, (**b**) Max-Half-Mchart-FMCD, and (**c**) Max-Half-Mchart-cellMCD.



(**a**) (**b**) (**c**)

Figure 5. Statistics plot of simulation on 15% outliers using (**a**) Max-Half-Mchart, (**b**) Max-Half-Mchart-FMCD, and (**c**) Max-Half-Mchart-cellMCD.



(**a**) (**b**) (**c**)

Figure 6. Statistics plot of simulation on 20% outliers using (**a**) Max-Half-Mchart, (**b**) Max-Half-Mchart-FMCD, and (**c**) Max-Half-Mchart-cellMCD.

4.2. Detecting Shifts in the Process

In statistical process control performance evaluation, one of the fundamental metrics is the Average Run Length (ARL). ARL is widely used to assess the efficiency of a control chart because it reflects how quickly the chart can detect a process change (shift). A control chart is considered desirable when it has a large ARL_0 under in-control conditions, indicating a low false alarm rate. Conversely, a smaller ARL_1 indicates faster detection of real process changes. Therefore, a smaller ARL_1, especially one closer to 1, indicates faster and more effective detection of process shifts [1].

In this study, ARL values were estimated using a simulation procedure. The simulations considered several shift scenarios, including shifts in the process mean, shifts in the covariance matrix, and simultaneous shifts in both. ARL_0 corresponds to the case when the process is in control, that is, when a $= 0$ and $b = 1$. Meanwhile, ARL_1 is evaluated under shifted conditions by changing the mean vector through **a**, the covariance structure through b, or both simultaneously. Experiments were conducted for a quality-characteristic dimension of $p = 5$ and three correlation settings $\rho \in \{0.3, 0.5, 0.7\}$. Under these settings, the performance of each chart can be compared objectively and comprehensively based on its ability to detect process shifts. For comparison purposes, the ARL values of the conventional and Fast-MCD-based charts were adopted from the preprint study by [38] and treated as benchmark results.

Table 5 shows the ARL values of the conventional Max-Half-Mchart for $\rho = 0.3$. When the mean vector shifts from $\mathbf{a} = [0, 0]'$ to $\mathbf{a} = [0.25, 0.25]'$ while the covariance matrix remains in control ($b = 1$), the ARL of the conventional Max-Half-Mchart decreases from 373 to 310. The ARL further decreases from 372 to 80 when the covariance matrix is inflated by a factor of 1.5 ($b = 1.5$) while the mean vector remains in control ($a = 0$).

Based on Table 6, the ARL of the robust Fast-MCD-based Max-Half-Mchart decreases from 371 to 326 under the same mean shift $\mathbf{a} = [0, 0]'$ to $\mathbf{a} = [0.25, 0.25]'$ with $b = 1$. It also decreases from 371 to 80 when the covariance matrix is inflated by a factor of 1.5 ($b = 1.5$) while the mean remains in control. Table 7 shows that the robust cellMCD-based Max-Half-Mchart achieves a faster response, with the ARL decreasing from 371 to 263 under the mean shift and from 371 to 62 when the covariance matrix is inflated by a factor of 1.5 ($b = 1.5$) with the mean still in control. These results indicate that the cellMCD-based robust Max-Half-Mchart detects both mean shifts and covariance changes more quickly than the other charts. A similar conclusion is observed for larger shifts, where the cellMCD-based chart also provides the fastest detection for changes in both the mean and the covariance structure.

Tables 8–10 report the ARL values of the three charts for $\rho = 0.7$. Under small shifts, the robust cellMCD-based Max-Half-Mchart detects both mean shifts and covariance changes faster than the other charts. However, under large shifts, the conventional chart provides faster detection for both mean shifts and covariance shifts. To make the comparison clearer, Figure 7 presents a graphical visualization of the detection performance of the three charts at $\rho = 0.7$ for mean shifts, covariance shifts, and simultaneous shifts in both.

Table 5. ARL of the conventional Max-Half-MChart for $p = 5$ and $\rho = 0.3$.

a	b												
	1	**1.25**	**1.5**	**1.75**	**2**	**2.25**	**2.5**	**2.75**	**3**	**3.25**	**3.5**	**3.75**	**4**
0	372.94	80.06	28.39	14.18	8.91	6.00	4.67	3.37	2.95	2.54	2.22	2.00	1.85
0.25	310.41	72.76	26.31	13.54	8.58	5.78	4.60	3.66	2.85	2.53	2.20	2.00	1.85
0.5	215.25	52.65	22.32	12.56	7.71	5.41	4.12	3.30	2.77	2.50	2.13	1.98	1.75
0.75	122.80	39.15	17.88	10.65	6.81	4.88	3.61	3.06	2.57	2.23	2.13	1.89	1.71
1	63.71	25.38	12.36	8.14	5.21	4.37	3.46	2.95	2.47	2.18	1.99	1.77	1.64
1.25	31.20	14.70	8.63	6.25	4.37	3.48	3.04	2.63	2.29	1.97	1.85	1.68	1.60
1.5	16.26	8.80	6.37	4.23	3.53	3.02	2.51	2.20	1.94	1.81	1.71	1.59	1.48
1.75	8.44	5.37	4.35	3.26	2.93	2.50	2.21	1.97	1.79	1.69	1.56	1.47	1.46
2	5.26	3.91	3.18	2.66	2.30	2.12	1.95	1.85	1.63	1.51	1.43	1.43	1.38
2.25	3.37	2.73	2.35	2.11	1.92	1.75	1.67	1.51	1.50	1.43	1.37	1.33	1.25
2.5	2.41	2.07	1.92	1.73	1.62	1.63	1.52	1.45	1.36	1.39	1.33	1.24	1.24
2.75	1.72	1.72	1.54	1.48	1.47	1.34	1.35	1.28	1.32	1.25	1.21	1.20	1.19
3	1.42	1.34	1.33	1.28	1.30	1.30	1.24	1.24	1.19	1.17	1.16	1.15	1.13

Table 6. ARL of the Fast-MCD-based Max-Half-MChart for $p = 5$ and $\rho = 0.3$.

a	b												
	1	**1.25**	**1.5**	**1.75**	**2**	**2.25**	**2.5**	**2.75**	**3**	**3.25**	**3.5**	**3.75**	**4**
0	371.13	80.56	29.22	14.51	9.41	6.22	4.63	3.84	3.02	2.57	2.25	2.17	1.88
0.25	326.69	70.36	28.05	13.81	8.47	5.83	4.53	3.53	2.92	2.50	2.24	1.97	1.88
0.5	222.40	59.14	23.22	12.57	7.86	5.30	4.01	3.28	2.82	2.43	2.23	1.96	1.85
0.75	129.17	39.18	18.81	10.13	6.78	4.84	4.00	3.17	2.61	2.32	2.16	1.88	1.78
1	65.58	23.70	12.51	8.22	5.65	4.39	3.38	2.81	2.54	2.25	2.00	1.80	1.71
1.25	31.82	15.46	9.08	6.34	4.69	3.55	2.99	2.47	2.23	2.05	1.81	1.77	1.64
1.5	16.50	9.28	6.15	4.76	3.64	3.03	2.49	2.30	2.10	1.87	1.68	1.61	1.56
1.75	8.97	5.83	4.30	3.39	3.01	2.34	2.24	1.88	1.80	1.66	1.57	1.48	1.46
2	5.08	3.78	3.19	2.66	2.36	2.08	1.96	1.79	1.63	1.57	1.51	1.43	1.37
2.25	3.33	2.66	2.53	2.19	1.95	1.81	1.69	1.56	1.47	1.41	1.38	1.34	1.31
2.5	2.37	2.04	1.90	1.78	1.70	1.53	1.50	1.44	1.36	1.32	1.27	1.31	1.23
2.75	1.72	1.71	1.59	1.50	1.49	1.44	1.36	1.36	1.27	1.24	1.22	1.21	1.19
3	1.44	1.43	1.37	1.31	1.35	1.30	1.30	1.22	1.19	1.18	1.16	1.14	1.16

Table 7. ARL of the cellMCD-based Max-Half-MChart for $p = 5$ and $\rho = 0.3$.

a	b												
	1	**1.25**	**1.5**	**1.75**	**2**	**2.25**	**2.5**	**2.75**	**3**	**3.25**	**3.5**	**3.75**	**4**
0	371.18	62.39	23.23	12.83	7.76	5.59	4.28	3.40	2.85	2.44	2.07	2.01	1.82
0.25	263.72	59.82	23.67	12.23	7.66	5.40	4.14	3.20	2.72	2.45	2.11	1.94	1.69
0.5	183.27	49.78	19.93	10.88	6.99	5.14	3.86	3.05	2.71	2.26	2.08	1.86	1.75
0.75	118.93	35.40	16.55	9.75	6.47	4.51	3.53	2.91	2.71	2.35	2.01	1.81	1.74
1	59.12	23.33	11.85	7.44	5.13	4.04	3.25	2.74	2.39	2.02	1.90	1.74	1.61
1.25	32.83	14.73	8.24	5.73	4.33	3.46	2.89	2.50	2.27	1.96	1.78	1.69	1.54
1.5	18.20	9.49	5.98	4.63	3.41	2.86	2.45	2.20	1.91	1.76	1.73	1.57	1.47
1.75	9.47	6.08	4.09	3.47	2.75	2.42	2.11	1.94	1.76	1.61	1.55	1.45	1.41
2	5.47	3.99	3.19	2.83	2.36	2.14	1.89	1.71	1.66	1.56	1.47	1.41	1.35
2.25	3.45	2.85	2.51	2.16	1.83	1.83	1.68	1.55	1.50	1.43	1.35	1.30	1.25
2.5	2.53	2.03	1.90	1.74	1.66	1.61	1.45	1.42	1.38	1.32	1.33	1.26	1.20
2.75	1.80	1.68	1.58	1.53	1.43	1.40	1.34	1.29	1.26	1.24	1.21	1.21	1.16
3	1.46	1.40	1.40	1.38	1.30	1.28	1.22	1.22	1.19	1.17	1.14	1.15	1.13

Table 8. ARL of the conventional Max-Half-MChart for $p = 5$ and $\rho = 0.7$.

a	b												
	1	**1.25**	**1.5**	**1.75**	**2**	**2.25**	**2.5**	**2.75**	**3**	**3.25**	**3.5**	**3.75**	**4**
0	372.44	77.84	28.33	14.09	8.70	5.85	4.56	3.48	2.93	2.59	2.30	2.00	1.87
0.25	308.73	71.61	26.26	14.57	9.26	6.15	4.32	3.52	2.92	2.48	2.20	1.98	1.87
0.5	251.44	63.97	24.55	13.34	7.82	5.84	4.27	3.53	2.92	2.47	2.18	1.95	1.87
0.75	194.19	51.68	21.28	11.60	7.64	5.18	4.22	3.40	2.83	2.37	2.10	1.94	1.80
1	126.91	37.13	17.74	9.83	6.47	4.70	3.81	3.09	2.75	2.23	2.10	1.92	1.79
1.25	72.12	27.71	13.50	7.89	5.46	4.52	3.64	2.88	2.53	2.16	1.94	1.75	1.63
1.5	40.55	18.17	10.45	6.68	5.18	3.79	3.16	2.61	2.32	2.04	1.87	1.79	1.64
1.75	25.97	11.60	7.67	5.51	4.27	3.32	2.83	2.57	2.12	1.92	1.76	1.59	1.51
2	14.65	8.97	5.80	4.17	3.36	2.91	2.56	2.30	1.99	1.83	1.70	1.55	1.50
2.25	9.48	5.83	4.48	3.57	2.86	2.46	2.28	2.00	1.86	1.72	1.61	1.45	1.39
2.5	6.20	4.30	3.46	2.95	2.52	2.22	2.07	1.84	1.66	1.57	1.53	1.42	1.38
2.75	4.23	3.37	2.78	2.42	2.10	1.90	1.79	1.69	1.58	1.49	1.46	1.41	1.35
3	3.27	2.48	2.31	2.04	1.89	1.75	1.67	1.53	1.49	1.40	1.32	1.31	1.27

Table 9. ARL of the Fast-MCD-based Max-Half-MChart for $p = 5$ and $\rho = 0.7$.

a	b												
	1	**1.25**	**1.5**	**1.75**	**2**	**2.25**	**2.5**	**2.75**	**3**	**3.25**	**3.5**	**3.75**	**4**
0	370.54	77.70	29.56	14.84	9.36	6.13	4.53	3.72	3.09	2.51	2.31	1.98	1.87
0.25	337.26	75.36	29.64	14.69	9.26	6.27	4.42	3.64	2.96	2.50	2.16	1.97	1.87
0.5	273.39	69.17	26.64	12.82	8.02	5.88	4.42	3.46	3.00	2.44	2.16	1.94	1.86
0.75	202.89	54.29	22.03	11.90	7.58	5.57	4.23	3.37	2.87	2.36	2.16	2.02	1.84
1	124.60	39.58	17.35	9.97	6.77	5.32	3.92	3.20	2.55	2.33	2.16	1.93	1.83
1.25	78.04	26.61	14.87	8.16	6.27	4.35	3.58	2.94	2.55	2.23	1.92	1.86	1.71
1.5	45.70	19.12	10.44	6.92	5.06	4.01	3.25	2.67	2.28	1.99	1.92	1.76	1.67
1.75	25.59	12.95	7.85	5.57	4.46	3.48	2.92	2.53	2.18	1.92	1.77	1.65	1.53
2	15.94	8.75	5.59	4.45	3.58	2.96	2.57	2.29	2.10	1.82	1.67	1.62	1.48
2.25	9.96	6.14	4.67	3.71	3.02	2.61	2.21	2.07	1.83	1.81	1.62	1.54	1.46
2.5	6.44	4.74	3.55	2.87	2.47	2.23	2.06	1.93	1.76	1.63	1.56	1.45	1.39
2.75	4.59	3.54	2.85	2.47	2.18	1.91	1.81	1.68	1.64	1.48	1.44	1.41	1.31
3	3.16	2.58	2.36	2.09	1.94	1.77	1.65	1.55	1.47	1.37	1.34	1.33	1.30

Table 10. ARL of the cellMCD-based Max-Half-MChart for $p = 5$ and $\rho = 0.7$.

a	b												
	1	**1.25**	**1.5**	**1.75**	**2**	**2.25**	**2.5**	**2.75**	**3**	**3.25**	**3.5**	**3.75**	**4**
0	370.86	57.12	23.12	12.50	8.10	5.47	3.95	3.41	2.83	2.55	2.12	1.98	1.87
0.25	233.34	55.11	22.16	12.10	7.86	5.22	4.02	3.28	2.66	2.40	2.18	1.99	1.76
0.5	193.92	52.54	20.88	12.01	7.50	4.96	3.92	3.25	2.57	2.31	2.14	1.83	1.77
0.75	155.75	40.85	17.86	10.21	6.81	4.80	3.60	3.03	2.57	2.42	1.99	1.82	1.74
1	118.61	35.23	15.91	9.00	6.29	4.49	3.67	3.02	2.56	2.12	2.00	1.80	1.65
1.25	74.11	23.64	13.00	8.24	5.14	4.05	3.23	2.98	2.29	2.10	1.85	1.71	1.62
1.5	41.96	18.23	10.36	6.47	4.71	3.71	2.90	2.61	2.36	1.94	1.77	1.64	1.59
1.75	26.91	12.99	7.85	5.45	3.97	3.17	2.64	2.44	2.10	1.86	1.72	1.62	1.52
2	17.68	8.84	6.13	4.56	3.36	2.87	2.57	2.19	1.95	1.83	1.63	1.52	1.44
2.25	10.49	6.56	4.77	3.70	3.05	2.53	2.25	1.94	1.85	1.63	1.60	1.48	1.45
2.5	6.88	4.78	3.64	2.99	2.52	2.18	2.05	1.78	1.72	1.58	1.47	1.42	1.35
2.75	4.72	3.67	3.00	2.51	2.23	2.02	1.86	1.71	1.60	1.50	1.42	1.38	1.30
3	3.52	2.86	2.47	2.12	1.90	1.81	1.60	1.56	1.46	1.43	1.33	1.31	1.28

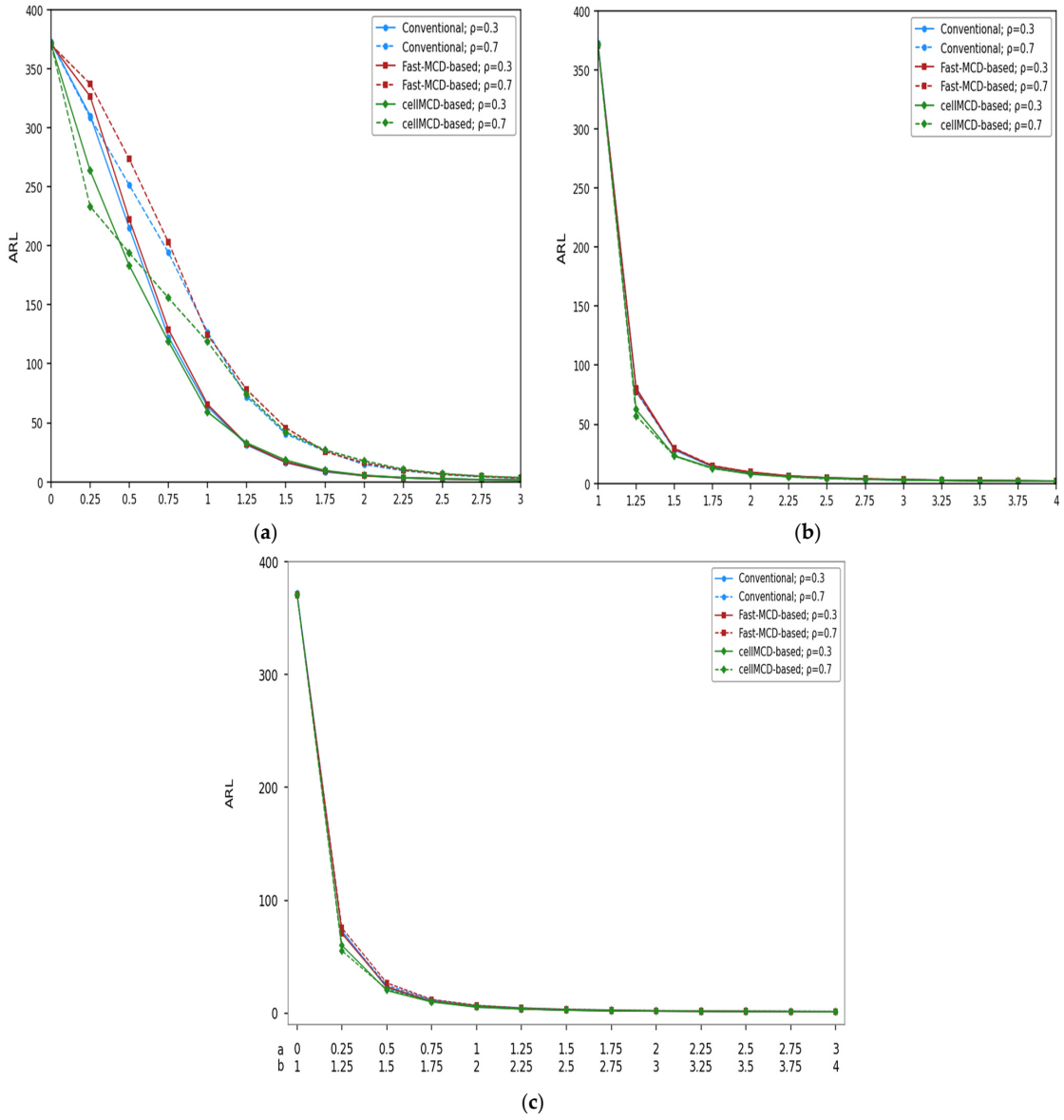


Figure 7. Comparison of ARL values for all charts: (**a**) mean shift, (**b**) covariance matrix shift, and (**c**) both.

Figure 7a compares the ARL values when a shift occurs in the mean vector (parameter a) while the covariance matrix remains in control ($b = 1$). Figure 7b shows the ARL comparison when the mean vector stays in control ($a = 0$) and a shift occurs in the covariance matrix, represented by the inflation factor b. Figure 7c presents the ARL comparison when shifts occur simultaneously in both the mean vector and the covariance matrix. In general, the robust cellMCD-based Max-Half-Mchart gives the fastest detection for small to moderate shifts, as indicated by the consistently smaller ARL values in the early part of the curves. This pattern is observed not only when the shift occurs in the mean vector but also when the covariance matrix changes and when both components shift simultaneously. Therefore, the advantage of the proposed chart is fairly consistent across the three shift scenarios, especially in the region where early detection is the most important.

However, Figure 7 also shows a trade-off. As the shift magnitude becomes larger, the differences among the charts become smaller, and in several cases the conventional Max-Half-Mchart eventually gives the smallest ARL values. This indicates that the main strength of the cellMCD-based chart lies in detecting subtle to moderate process changes, whereas its relative advantage becomes less pronounced for large shifts that are already

easy to detect by all charts. In addition, the effect of correlation is most visible in the mean-shift case, where higher correlation tends to increase ARL for small to moderate shifts. By contrast, for covariance shifts and simultaneous shifts, the ARL curves under different correlation levels are much closer, indicating that correlation has only a minor effect on detection speed in these two scenarios.

4.3. Application Scheme

4.3.1. Application to Synthetic Data

In this section, the proposed cellMCD-based Max-Half-Mchart is further evaluated using synthetic data. Synthetic data are employed because the data-generating mechanism can be controlled explicitly, allowing a clearer comparison of the competing charts under cellwise contamination. The synthetic data generation follows the same procedure as described in Section 4.1, but with a simplified design for illustration. In this subsection, two contamination levels are considered, namely 5% for Scenario 1 and 20% for Scenario 2, and the sample size is set to $n = 100$. Under these two scenarios, the performance of the conventional Max-Half-Mchart, the Fast-MCD-based Max-Half-Mchart, and the cellMCD-based Max-Half-Mchart is evaluated using Accuracy, AUC, false positive rate (FP rate), and false negative rate (FN rate).

Table 11 summarizes the performance comparison of the three charts under the two synthetic-data scenarios. In general, the conventional Max-Half-Mchart shows the weakest performance, particularly under the higher contamination level, where its ability to distinguish contaminated observations from normal ones becomes very limited. The Fast-MCD-based Max-Half-Mchart performs better than the conventional chart, especially under low contamination, but its performance still declines when the contamination level becomes more severe. In contrast, the cellMCD-based Max-Half-Mchart provides the most stable classification performance across both scenarios, as indicated by high Accuracy and AUC, very low FP rates, and the smallest FN rate among the competing charts.

Table 11. Performance Comparison of All Charts on Synthetic Data.

Max-Half-Mchart	Scenario	Evaluation Metrics			
		Accuracy	FP Rate	FN Rate	AUC
Conventional	1	0.6263	0.0000	0.9024	0.5488
	2	0.1515	0.0000	1.0000	0.5000
Fast-MCD	1	0.9798	0.0000	0.0541	0.9730
	2	0.8889	0.0000	0.1642	0.9562
cellMCD	1	0.9899	0.0182	0.0000	0.9909
	2	0.9899	0.0000	0.0122	0.9939

Figures 8 and 9 provide a graphical illustration of these findings. In each plot, the dashed vertical line separates Phase I (in-control) from Phase II (contaminated), while the red horizontal line indicates the upper control limit (UCL). In Scenario 1, the conventional chart detects only a few out-of-control observations in Phase II, whereas the Fast-MCD-based chart is more responsive but still misses several contaminated points. By contrast, the cellMCD-based chart shows a clearer separation, with most contaminated observations exceeding the UCL while the in-control observations remain below it. This pattern becomes even more apparent in Scenario 2, where the contamination level increases to 20%. Under this setting, the conventional chart fails to provide adequate out-of-control signals, and the Fast-MCD-based chart still misses part of the contaminated observations, while the cellMCD-based chart remains the most consistent. Overall, the synthetic-data results are

consistent with the simulation study and confirm that incorporating cellMCD into the Max-Half-Mchart framework yields the most stable and reliable performance for detecting cellwise contamination.

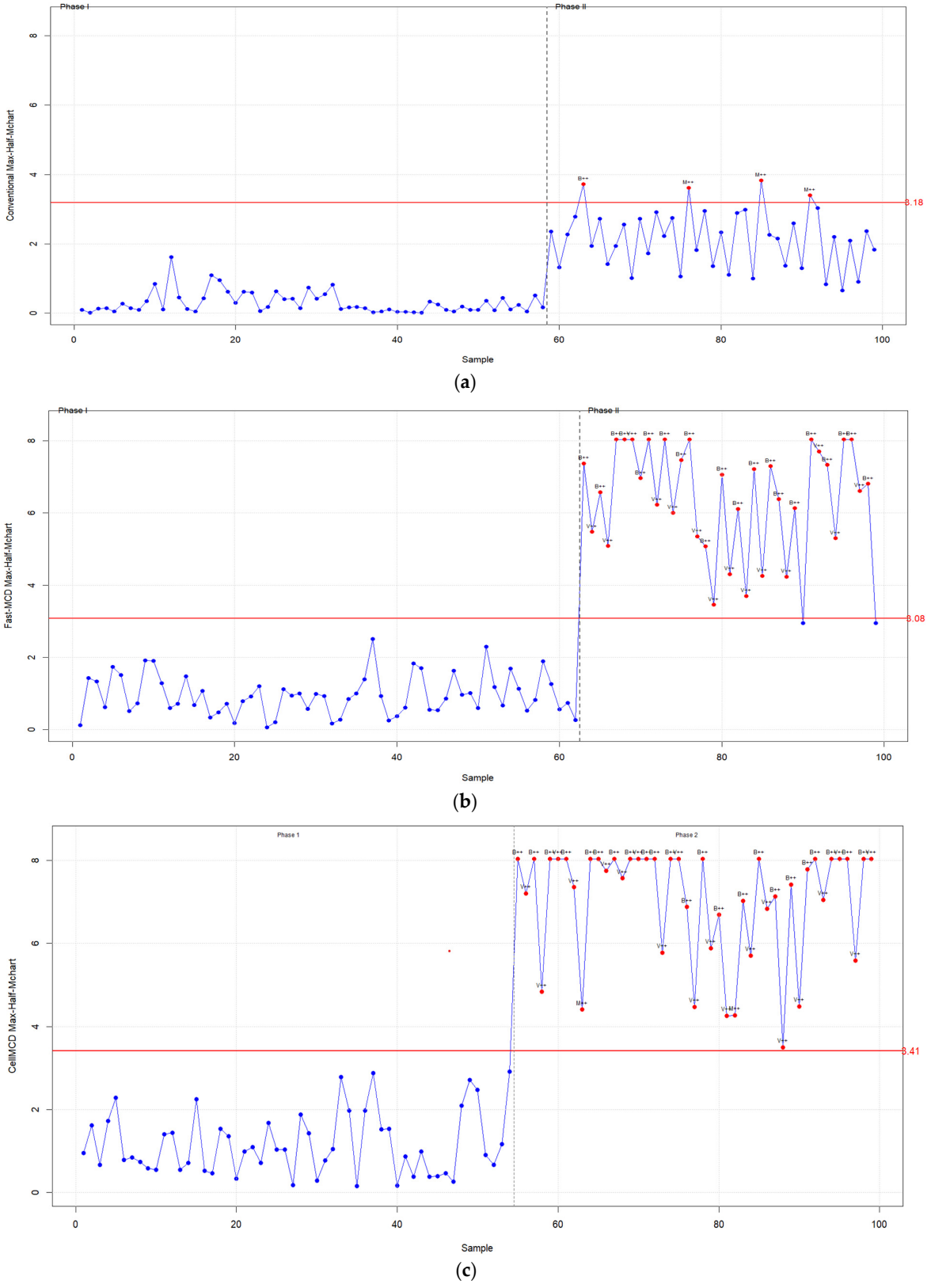
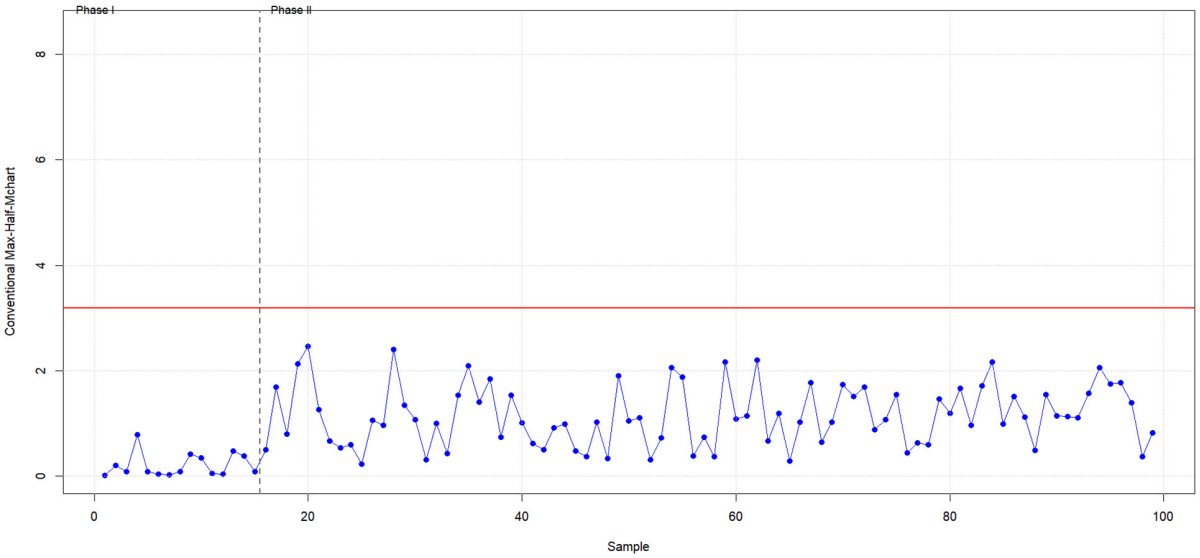
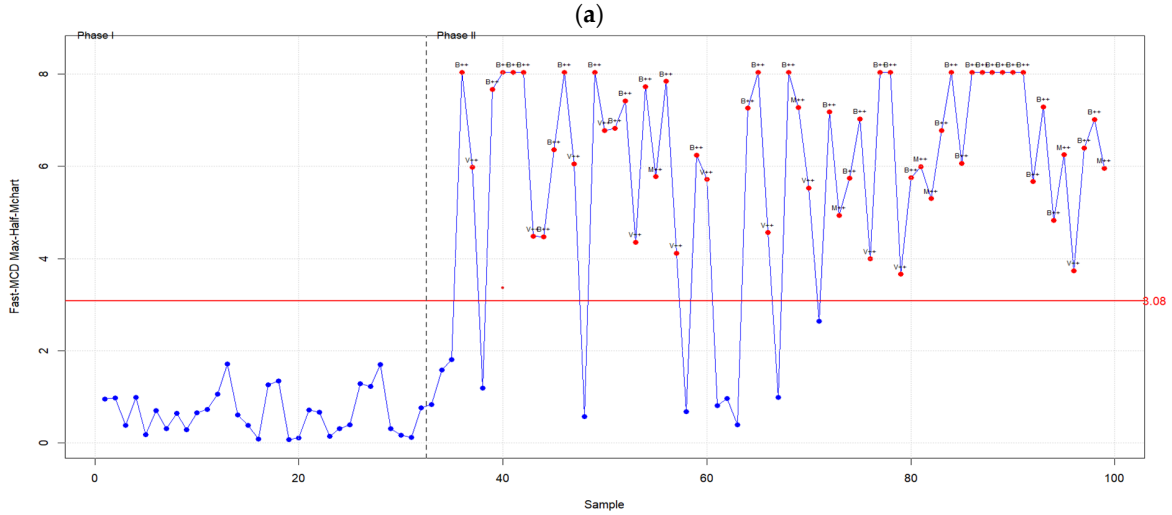


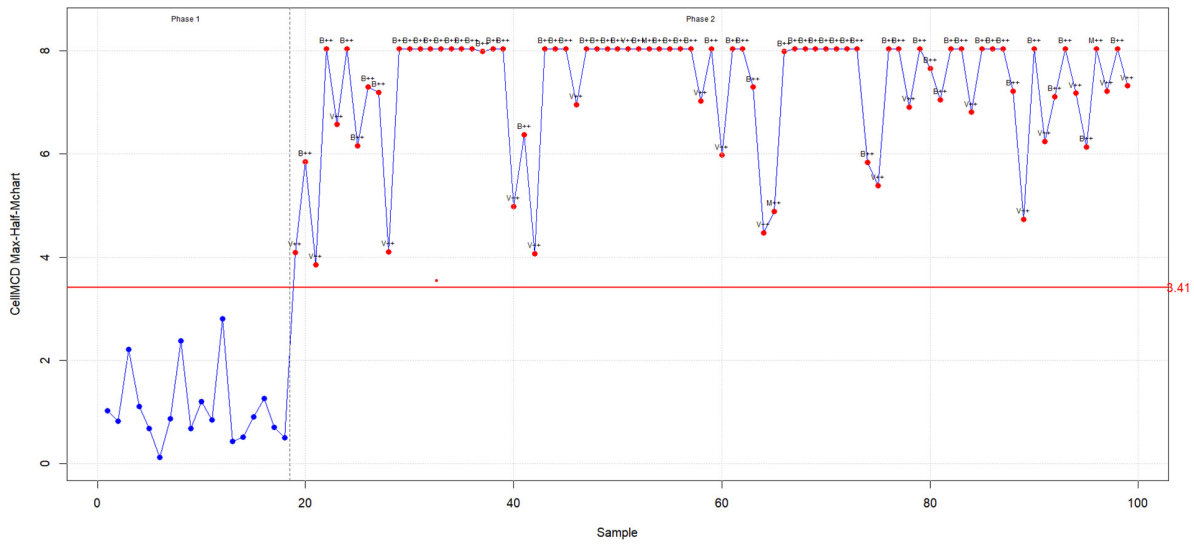
Figure 8. Statistics plots of all charts under Scenario 1: (**a**) Max-Half-Mchart, (**b**) Max-Half-Mchart-FMCD, and (**c**) Max-Half-Mchart-cellMCD.



(**a**)



(**b**)



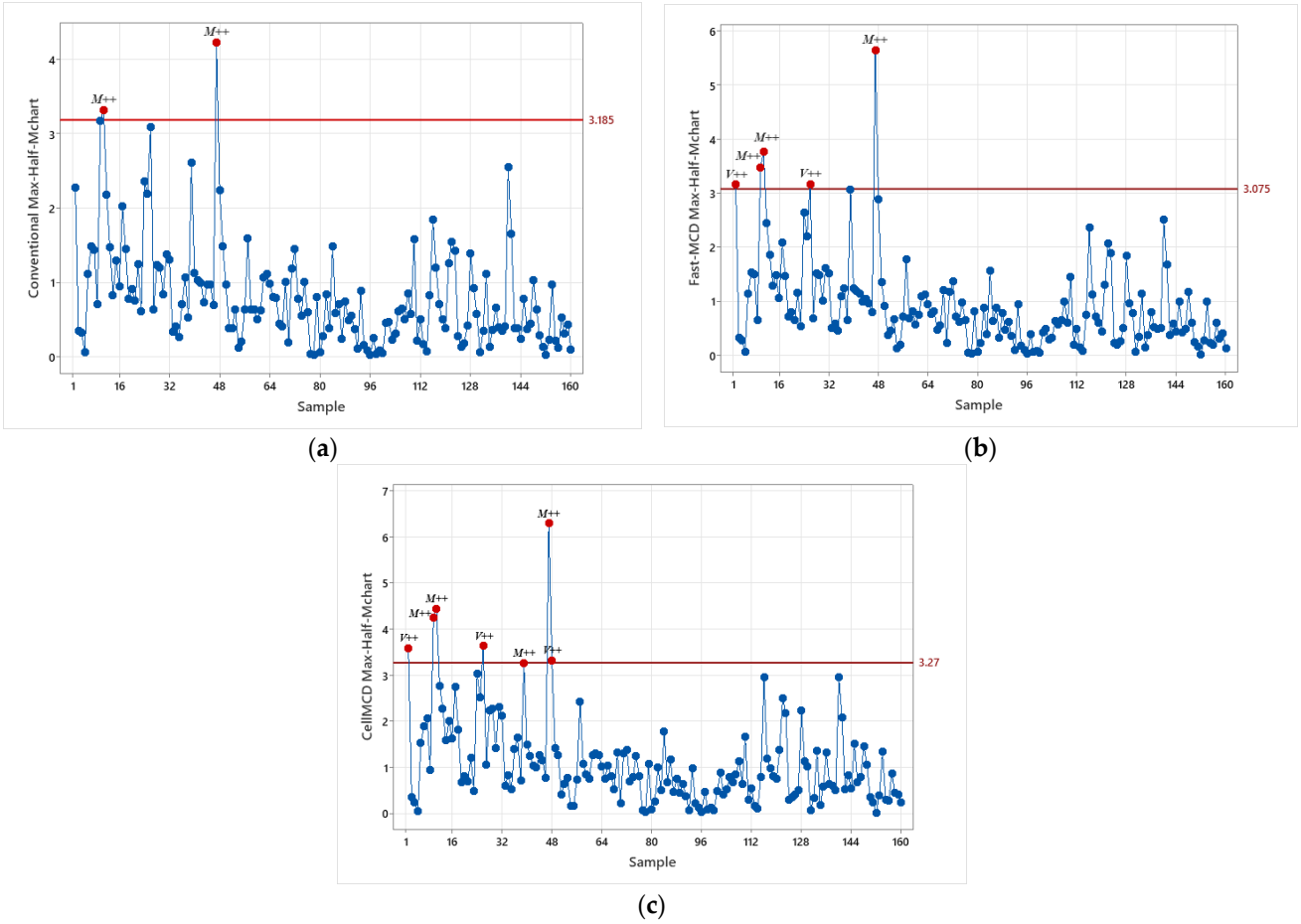
(**c**)

Figure 9. Statistics plots of all charts under Scenario 2: (**a**) Max-Half-Mchart, (**b**) Max-Half-Mchart-FMCD, and (**c**) Max-Half-Mchart-cellMCD.

4.3.2. Application to Real Data

The proposed robust cellMCD-based Max-Half-Mchart is applied to real OPC cement quality data [38]. The monitoring results of the cellMCD-based chart are compared with two other approaches, namely the conventional chart using the classical mean and covariance estimators and the robust Fast-MCD-based Max-Half-Mchart. This comparison aims to evaluate how clearly each chart separates in-control and out-of-control conditions.

The application results are illustrated in Figure 10, which shows the monitoring outputs of the conventional Max-Half-Mchart, the Fast-MCD-based Max-Half-Mchart, and the cellMCD-based Max-Half-Mchart. The red horizontal line represents the upper control limit (UCL), and points above the UCL are interpreted as out-of-control signals. The annotated points indicate the type of detected shift. The label M^{++} denotes that the out-of-control signal is mainly driven by a mean shift, whereas the label V^{++} indicates a signal driven by a variability shift. In the conventional chart, the detected signals are dominated by M^{++}, suggesting that the early deviations are primarily associated with changes in the process mean. In the Fast-MCD-based and cellMCD-based charts, V^{++} signals also appear at several early observations, indicating that robust approaches can detect changes related to process variability that may not be clearly captured by the conventional approach.



(**a**) (**b**) (**c**)

Figure 10. Monitoring result of all charts under real datasets: (**a**) conventional Max-Half-Mchart, (**b**) Fast-MCD-based Max-Half-Mchart, (**c**) cellMCD-based Max-Half-Mchart.

Based on Figure 10, the conventional chart detects out-of-control signals only at observations 10 and 46, and both are classified as M^{++}, meaning that the detected deviations mainly reflect mean shifts. The Fast-MCD-based Max-Half-Mchart detects more out-of-control signals, including V^{++} at observations 1 and 25 and M^{++} at observations 9, 10,

and 46. Furthermore, the cellMCD-based Max-Half-Mchart provides the most sensitive detection with seven out-of-control signals, namely V^{++} at observations 1, 25, and 47 and M^{++} at observations 9, 10, 38, and 46. These results are summarized in Table 12.

Table 12. Performance comparison of all charts on monitoring cement OPC datasets.

Control Chart	Out-of-Control	Mean Shift	Variability Shift
Max-Half-Mchart	2	2	0
Robust Max-Half-Mchart Based on Fast-MCD	5	3	2
Robust Max-Half-Mchart Based on cellMCD	7	4	3

5. Conclusions and Future Research

This study proposed a robust Max-Half-Mchart based on the cellwise Minimum Covariance Determinant (cellMCD) estimator for simultaneous monitoring of process mean and variability under cellwise contamination. The main methodological contribution of this work lies in integrating a cellwise robust estimator of location and covariance into the Max-Half-Mchart framework, making the chart more robust to cellwise outliers that may trigger outlier propagation. This is particularly important because, under cellwise contamination, even a small proportion of contaminated cells may affect many observations, thereby reducing the effectiveness of conventional and casewise robust approaches. Overall, the results indicate that the proposed chart provides a more reliable monitoring scheme than the conventional Max-Half-Mchart and the Fast-MCD-based Max-Half-Mchart in the presence of cellwise outliers. In the outlier-detection study, the proposed method showed more stable performance across different correlation structures and contamination levels, with better ability to distinguish between normal and contaminated observations. In the process-shift study, the chart also demonstrated improved sensitivity, particularly for small to moderate shifts. The applications to synthetic data and real OPC cement quality data further support the practical usefulness of the proposed approach for identifying out-of-control observations and distinguishing between changes in process mean and variability.

For future work, evaluations can be extended to settings with different numbers of quality characteristics, including high-dimensional cases where cellwise contamination may have a stronger impact on conventional and casewise robust methods. Additional comparisons with other robust techniques may also be conducted to further benchmark the cellMCD-based Max-Half-Mchart against alternative approaches [47,48]. Moreover, the procedure can be adapted for subgrouped observations and tested on other industrial datasets to examine its consistency and reliability across different contexts. Integrating machine learning methods to enhance detection and decision support is another promising direction [49,50].

Author Contributions: S.B.S.H.: methodology, software, simulation study, data curation, formal analysis, visualization, writing—original draft. M.A.: conceptualization, methodology, supervision, validation, investigation, resources, writing—review and editing. W.: supervision, validation, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study is available on request from the corresponding author. The data is not publicly available due to privacy.

Acknowledgments: The authors gratefully acknowledge financial support from the Institut Teknologi Sepuluh Nopember.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Montgomery, D.C. *Introduction to Statistical Quality Control*; John Wiley & Sons: New York, NY, USA, 2020.
2. Shewhart, W.A. *Economic Control of Quality of Manufactured Product*; Books On Demand; American Society for Quality Control: Milwaukee, WI, USA, 1980.
3. Hotelling, H. Multivariate Quality Control Illustrated by Air Testing of Sample Bombsights. *Tech. Stat. Anal.* **1947**, 111–184.
4. Woodall, W.H.; Ncube, M.M. Multivariate CUSUM Quality-Control Procedures. *Technometrics* **1985**, *27*, 285. [CrossRef]
5. Lowry, C.A.; Woodall, W.H.; Champ, C.W.; Rigdon, S.E. A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics* **1992**, *34*, 46. [CrossRef] [PubMed]
6. Khoo, M.B.C.; Quah, S.H. Multivariate control chart for process dispersion based on individual observations. *Qual. Eng.* **2003**, *15*, 639–642. [CrossRef]
7. Djauhari, M.A. Improved Monitoring of Multivariate Process Variability. *J. Qual. Technol.* **2005**, *37*, 32–39. [CrossRef]
8. Djauhari, M.A.; Mashuri, M.; Herwindiati, D.E. Multivariate Process Variability Monitoring. *Commun. Stat. Theory Methods* **2008**, *37*, 1742–1754. [CrossRef]
9. Djauhari, M.A. A Multivariate Process Variability Monitoring Based on Individual Observations. *Mod. Appl. Sci.* **2010**, *4*, 91–96. [CrossRef]
10. Sindelar, M.F. Multivariate Statistical Process Control for Correlation Matrices. Doctoral Dissertation, University of Pittsburgh, Pittsburgh, PA, USA, 2007.
11. Mashuri, M.; Haryono, H.; Aksioma, D.F.; Wibawati, W.; Ahsan, M.; Khusna, H. Tr(R2) Control Charts Based on Kernel Density Estimation for Monitoring Multivariate Variability Process. *Cogent Eng.* **2019**, *6*, 1665949. [CrossRef]
12. White, E.M.; Schroeder, R. A Simultaneous Control Chart. *J. Qual. Technol.* **1987**, *19*, 1–10. [CrossRef]
13. Chen, G.; Cheng, S.W. Max Chart: Combining X-Bar Chart And S Chart. 1998. Available online: http://about.jstor.org/terms (accessed on 20 December 2025).
14. Khoo, M.B.C. A new bivariate control chart to monitor the multivariate process mean and variance simultaneously. *Qual. Eng.* **2005**, *17*, 109–118. [CrossRef]
15. Thaga, K.; Gabaitiri, L. Multivariate Max-Chart. *Econ. Qual. Control.* **2006**, *21*, 113–125. [CrossRef]
16. Sabahno, H.; Amiri, A.; Castagliola, P. A new adaptive control chart for the simultaneous monitoring of the mean and variability of multivariate normal processes. *Comput. Ind. Eng.* **2021**, *151*, 106524. [CrossRef]
17. Kruba, R.; Mashuri, M.; Prastyo, D.D. The effectiveness of Max-half-Mchart over Max-Mchart in simultaneously monitoring process mean and variability of individual observations. *Qual. Reliab. Eng. Int.* **2021**, *37*, 2334–2347. [CrossRef]
18. Kruba, R.; Mashuri, M.; Prastyo, D.D. Max-Half-Mchart: A Simultaneous Control Chart Using a Half-Normal Distribution for Subgroup Observations. *IEEE Access* **2021**, *9*, 105369–105381. [CrossRef]
19. Rocke, D.M. Robust control charts. *Technometrics* **1989**, *31*, 173–184. [CrossRef]
20. Bendre, S.M.; Kale, B.K. Masking Effect on Tests for Outliers in Normal Samples. 1987. Available online: http://biomet.oxfordjournals.org/ (accessed on 18 January 2026).
21. Hadi, A.S.; Imon, A.H.M.R.; Werner, M. Detection of outliers. *WIREs Comp. Stat.* **2009**, *1*, 57–70. [CrossRef]
22. Rousseeuw, P.J. Multivariate Estimation with High Breakdown Point. *Math. Stat. Appl.* **1985**, *37*, 283–297.
23. Rousseeuw, P.J.; Van Driessen, K. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **1999**, *41*, 212–223. [CrossRef]
24. Willems, G.; Pison, G.; Rousseeuw, J.; Van Aelst, S. A robust Hotelling test. *Metrika* **2012**, *55*, 125–138. [CrossRef]
25. Hubert, M.; Rousseeuw, P.J.; Verdonck, T. A deterministic algorithm for robust location and scatter. *J. Comput. Graph. Stat.* **2012**, *21*, 618–637. [CrossRef]
26. Boudt, K.; Rousseeuw, P.J.; Vanduffel, S.; Verdonck, T. The minimum regularized covariance determinant estimator. *Stat. Comput.* **2020**, *30*, 113–128. [CrossRef]
27. Schreurs, J.; Vranckx, I.; Hubert, M.; Suykens, J.A.K.; Rousseeuw, P.J. Outlier detection in non-elliptical data by kernel MRCD. *Stat. Comput.* **2021**, *31*, 1–18. [CrossRef]
28. Alqallaf, F.; Van Aelst, S.; Yohai, V.J.; Zamar, R.H. Propagation of outliers in multivariate data. *Ann. Stat.* **2009**, *37*, 311–331. [CrossRef]
29. Rousseeuw, P.J.; Van Den Bossche, W. Detecting Deviating Data Cells. *Technometrics* **2018**, *60*, 135–145. [CrossRef]
30. Öllerer, V.; Croux, C. Robust High-Dimensional Precision Matrix Estimation. In *Modern Nonparametric, Robust and Multivariate Methods*; Springer International Publishing: Cham, Switzerland, 2015; pp. 325–350. [CrossRef]

31. Öllerer, V.; Alfons, A.; Croux, C. The shooting S-estimator for robust regression. *Comput. Stat.* **2016**, *31*, 829–844. [CrossRef]
32. Tarr, G.; Müller, S.; Weber, N.C. Robust estimation of precision matrices under cellwise contamination. *Comput. Stat. Data Anal.* **2016**, *93*, 404–420. [CrossRef]
33. Raymaekers, J.; Rousseeuw, P.J. Fast Robust Correlation for High-Dimensional Data. *Technometrics* **2021**, *63*, 184–198. [CrossRef]
34. Raymaekers, J.; Rousseeuw, P.J. The Cellwise Minimum Covariance Determinant Estimator. *J. Am. Stat. Assoc.* **2024**, *119*, 2610–2621. [CrossRef]
35. Vargas, N.J.A. Robust Estimation in Multivariate Control Charts for Individual Observations. *J. Qual. Technol.* **2003**, *35*, 367–376. [CrossRef]
36. Chenouri, S.; Steiner, S.H.; Variyath, A.M. A Multivariate Robust Control Chart for Individual Observations. *J. Qual. Technol.* **2009**, *41*, 259–271. [CrossRef]
37. Ali, H.; Yahaya, S.S.S.; Omar, Z. Robust Hotelling Control Chart with Consistent Minimum Vector Variance. *Math. Probl. Eng.* **2013**, *2013*, 1–7. [CrossRef]
38. Ahsan, M.; R, A.P.S.; Mashuri, M.; Wibawati, W.; Safira, D.A.; Lee, M.H. Robust Maximum Half-Normal Multivariate Control Chart Based on Det-MCD. *Appl. Sci.* 2026; *under processing*.
39. Prasetya, I.K.; Ahsan, M.; Mashuri, M.; Lee, M.H. Bootstrap and MRCD Estimators in Hotelling's T2 Control Charts for Precise Intrusion Detection. *Appl. Sci.* **2024**, *14*, 7948. [CrossRef]
40. Ahsan, M.; Mashuri, M.; Amalia, R.N.; Fahri, F.; Safira, D.A.; Lee, M.H. Robust Multivariate Simultaneous Control Chart Based on Minimum Regularized Covariance Determinant (MRCD). *Appl. Sci.* **2026**, *16*, 2924. [CrossRef]
41. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum Likelihood from Incomplete Data Via the *EM* Algorithm. *J. R. Stat. Soc. Series B Stat. Methodol.* **1977**, *39*, 1–22. [CrossRef]
42. Little, R.; Rubin, D. *Statistical Analysis with Missing Data*, 3rd ed.; Wiley: Sydney, Australia, 2019. [CrossRef]
43. Raymaekers, J.; Rousseeuw, P. Handling Cellwise Outliers by Sparse Regression and Robust Covariance. *J. Data Sci. Statistics Vis.* **2021**, *1*. [CrossRef]
44. Ahsan, M.; Mashuri, M.; Kuswanto, H.; Prastyo, D.D. Intrusion Detection System Using Multivariate Control Chart Hotelling's T2 Based on PCA. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2018**, *8*, 1905–1911. [CrossRef]
45. Hidayatulloh, S.; Mustajab, M.A.; Ramdhani, Y. Penggunaan Otimasi Atribut Dalam Peningkatan Akurasi Prediksi Deep Learning Pada Bike Sharing Demand. *INFOTECH J.* **2023**, *9*, 54–61. [CrossRef]
46. Raymaekers, J.; Rousseeuw, P. Cellwise: Analyzing Data with Cellwise Outliers. *R Package Version 2.5.7*, CRAN: Vienna, Austria, 2016. [CrossRef]
47. Kalina, J.; Tichavský, J. The minimum weighted covariance determinant estimator for high-dimensional data. *Adv. Data Anal. Classif.* **2022**, *16*, 977–999. [CrossRef]
48. Centofanti, F.; Hubert, M.; Rousseeuw, P.J. Cellwise and Casewise Robust Covariance in High Dimensions. 2025. Available online: http://arxiv.org/abs/2505.19925 (accessed on 15 February 2026).
49. Ahsan, M.; Khusna, H.; Wibawati; Lee, M.H. Support vector data description with kernel density estimation (SVDD-KDE) control chart for network intrusion monitoring. *Sci. Revol.* **2023**, *13*, 19149. [CrossRef] [PubMed]
50. Nur, S.; Ahsan, M.; Khusna, H. Multivariate Exponentially Weighted Moving Average (MEWMA) and Multivariate Exponentially Weighted Moving Variance (MEWMV) Chart Based on Residual XGBoost Regression for Monitoring Water Quality. *Eng. Lett.* **2023**, *31*, 1001–1008.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.