

Article

On Phase-I Monitoring of Process Location Parameter with Auxiliary Information-Based Median Control Charts

Shahid Hussain ^{1,2,*}, Sun Mei ^{1,*}, Muhammad Riaz ³ and Saddam Akber Abbasi ⁴ 

¹ Faculty of Science, Institute of Applied Systems Analysis, Jiangsu University, Zhenjiang 212013, China

² Department of Mathematics, COMSATS University Islamabad, Attock Campus, Attock 43600, Pakistan

³ Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia; riazm@kfupm.edu.sa

⁴ Department of Mathematics, Statistics, and Physics, Qatar University, Doha 2713, Qatar; sabbasi@qu.edu.qa or saddamabbasi@yahoo.com

* Correspondence: shahidhussain@ujs.edu.cn or shahid_libra82@hotmail.com (S.H.); sunm@ujs.edu.cn (S.M.)

Received: 9 March 2020; Accepted: 21 April 2020; Published: 2 May 2020



Abstract: A control chart is often used to monitor the industrial or services processes to improve the quality of the products. Mostly, the monitoring of location parameters, both in Phase I and Phase II, is done using a mean control chart with the assumption that the process is free from outliers or the estimators are correctly estimated from in-control samples. Generally, there are question marks about such kind of narratives. The performance of the mean chart is highly affected in the presence of outliers. Therefore, the median chart is an attractive alternative to the mean chart in this situation. The control charts are usually implemented in two phases: Phase I (retrospective) and Phase II (prospective/monitoring). The efficiency of any control chart in Phase II depends on the accuracy of control limits obtained from Phase I. The current study focuses on the Phase I analysis of location parameters using median control charts. We examined the performance of different auxiliary information-based median control charts and compared the results with the usual median chart. Standardized variance and relative efficacy are used as performance measures to evaluate the efficiency of median estimators. Moreover, the probability to signal measure is used to evaluate the performance of proposed control charts to detect any potential changes in the process. The results revealed that the proposed auxiliary information based median control charts perform better in Phase I analysis. In addition, a practical illustration of an industrial scenario demonstrated the significance of the proposed control charts, in which the monitoring of concrete compressive strength is emphasized.

Keywords: auxiliary information; median control charts; Phase I monitoring; probability to signal; relative efficiency; standardized variance

1. Introduction

Quality improvement is a never-ending process in manufacturing industries. There is always a need to increase the quality of a product by reducing process variations. Statistical Process Control (SPC) is the core field of statistics that supports essential methods to complete this objective. SPC consists of a toolkit that is employed to monitor industrial processes and to boost the quality of the product(s) by reducing the variability in the product(s) [1]. The SPC toolkit contains a Pareto chart, scatter diagram, cause and effect diagram, check sheet, histogram and control chart, the last one being the most prominent and frequently used tool in the SPC toolkit. A control chart, initially developed in 1924 by Walter A. Shewhart, later published in [2], is a graphical display of the quality characteristic

used for the monitoring of the process. The main innovative idea of Shewhart is to define the variability of a production process using statistical thoughts and further divide the variation of a process into common and special causes. At present, the choice of an appropriate statistical control chart has become a prime concern of professionals for quality control in manufacturing industries [1].

The monitoring of process parameters (location or dispersion) by applying control charts is mainly categorized into two phases: Phase I (retrospective phase) and Phase II (prospective/monitoring phase). In Phase I, we attempt to adjust a process for stability (in-control state) with respect to different process parameters, which usually happens at the first monitoring process (e.g., the output of production from the first installation of a machine). There is a need for the small dataset from the process to apply the Phase I control chart in order to know whether it is in-control or needs an amendment, if the process is not in an in-control state. Usually, this exercise needs to be repeated a number of times until the process is in-control. The next step for online monitoring, using a controlled structure in Phase I, is to construct a control chart. The main goal for Phase II is the quick signaling of an out-of-control situation of the process parameters. A control chart assumes that the process parameters of the in-control samples are either known or correctly estimated from Phase I samples. In reality, the process parameters are unknown or estimated from an inadequate number of samples, which influences the results of the control chart in terms of detection ability (see, [3,4]). The effectiveness of any Phase II control chart is determined by the accurate limits of the Phase I control chart. The main reason for the difference in performance is the variability of the parameter estimates. A number of researchers have studied the performances of control charts in Phase I analysis [5–7].

In SPC, it is a common practice to use mean (\bar{Y}) or the variance (S^2) in control charts for process monitoring. However, other estimators like median (\tilde{Y}) have received less attention for this purpose. The mean chart is a more effective chart to detect out-of-control signals as compared to the median chart under normality. However, in abnormal cases, where we come across outliers, the median chart is a good substitute for the mean chart. The median chart is robust and very simple compared to the mean chart in the presence of outliers (contamination) or small deviations from normality [8]. These characteristics (properties) strengthen the use of the median chart in a long running process or without checking possible outliers. There is a relatively small body of literature concerned with median control charts (see [9–13]). Furthermore, to investigate the effect of estimated parameters on the deduction ability of the median control chart, Castagliola and Figueiredo [14] presented the Shewhart median charting structure with estimated parameters. They also gave recommendations about the selection of the number of initial subgroups by comparison of Shewhart median chart in both known and unknown (estimated) parameter cases. The exponentially weighted moving average (EWMA) median chart with estimated parameter for the monitoring of location parameters was proposed by Castagliola et al. [8]. They showed that the large number of initial samples can reduce the difference between known and estimated parameters. Castagliola et al. [15] proposed the estimated parameters study for the cumulative sum (CUSUM) structure of the median control chart. Tran [16] suggested the investigation of run rules median chart in the existence of measurement errors. Recently, Teng et al. [17] presented the run rules median chart with estimated parameters. Motivated by the above research, we proposed and investigated various median charts for the effective monitoring of process locations under contaminated situations.

The supplementary extra (or auxiliary) information correlated or associated with the variable of interest results into the reduction of the variance. The auxiliary information, i , enhances the precision of estimator(s) or improves the sampling plan. There are several studies in literature of survey sampling available for the use of auxiliary information and it has been conclusively shown that the use of auxiliary information enhances the accuracy of estimators for population mean [18], population median [19] or population variance [20].

In SPC literature, the use of auxiliary information has been proven successful to improve the efficiency of control chart structures. Zhang [21] used the auxiliary variable regressed by the variable of interest in the cause-selecting control chart. The study of [22,23] introduced control charts for location

and dispersion monitoring, respectively, by using a regression estimator, and showed the enhancement in the efficiency of the control chart based on the auxiliary variable. Some other useful literature includes [24,25] for location parameter, and [13,26] for dispersion parameter.

The International Organization for Standardization (ISO) is a global organization of national standard bodies which encourage international trade in goods and services, promote international cooperation in intellectual, scientific, technological and economical activities and support the development of international standardization activities. No universal agreement has been reached in the past on the use of control charts. Various companies had various rules. A number of international regulatory standards, such as ISO 7870-1 [27], ISO 7870-2 [28], ISO 7870-3 [29], ISO 7870-4 [30], ISO 7870-5 [31] and ISO 7870-6 [32] are being developed and are widely accepted with massive business exchanges in countries of different levels of quality. This is in addition to the international regulatory standards [33] and [34]. The international standard provides a guide on the application and knowledge of the methods of statistical control of a process by the Shewhart control chart known as ISO 7870-2:2013 Control charts. There are different parts for ISO 7870-2:2013, in which the second part is related to Shewhart control charts and deals with statistical methods for process control using Shewhart charts. Some additional information related to Shewhart charts, like process capability, analysis of trend pattern and use of warning limits is briefly included.

The control process in Phase I can be iterative and control limits are generally seen as trail limits, as defined in [1]. When statistical checks are developed in order to satisfy the user, any in-control data (also known as reference data) in unknown quantities (parameters) are estimated to lead to the development of control charts in order to allow for the efficient process monitoring in Phase II. It is therefore no wonder that the effectiveness of the Phase II process monitoring critically depends on the performance of the related Phase I study. Motivated by the above-cited literature, auxiliary information-based median control charts for location monitoring, particularly for Phase I analysis, is the main idea of this study, because Phase I analysis for any monitoring plays a crucial role in the monitoring of parameters. The organization of this article is as follows: the details about median estimators with their biases and mean square errors are presented in Section 2. The control charting structure of auxiliary information based median control charts are given in Section 3. Section 4 presents the information about performance measures used to evaluate the charting structures. The results discussion and comparative analysis is provided in Section 5. An illustrative example with a real dataset for the application of control charts is included in Section 6, and finally, the summary and conclusions of the study are given in Section 7.

2. Auxiliary Information Based Location Median Estimators and Their Properties

Let us assume that the characteristic of interest (study variable) is represented by Y and the auxiliary characteristic (variable) is represented by X —follow the bivariate normal distribution with means μ_y and μ_x , respectively, with standard deviations σ_y and σ_x , medians M_y and M_x , coefficient of variations C_y and C_x , coefficient of kurtosis $\beta_2(y)$ and $\beta_2(x)$ and correlation coefficient ρ_{yx} . Similarly, let (y_i, x_i) for $(i = 1, 2, 3, \dots, n)$ be a bivariate random sample of size n . The notations for sample quantities are explained as; the sample means of Y and X are \bar{y} and \bar{x} , respectively, sample standard deviations are s_y and s_x and the sample medians are \widehat{M}_y and \widehat{M}_x . On the basis of this preliminary information, we considered some auxiliary information-based efficient median estimators for the estimation of the median of quality characteristic M_y . Let the correlation coefficient between $(\widehat{M}_x, \widehat{M}_y)$ be $\rho_{(\widehat{M}_x, \widehat{M}_y)} = \rho_c = 4P_{11}(x, y) - 1$, where $P_{11}(x, y) = P(X \leq M_x \cap Y \leq M_y)$ is the probability of the population units with $X \leq M_x$ and $Y \leq M_y$. It is further assumed the limiting distribution of (X, Y) is a continuous distribution with marginal densities $f_x(x)$ and $f_y(y)$ and also the quantities $f_x(M_x)$ and $f_y(M_y)$ are positive.

These median estimators M_a ($a = 1, 2, \dots, 6$) with their respective bias, $B(\cdot)$ And mean square error, $MSE(\cdot)$ are presented below:

(i) The usual median estimator

The most commonly used estimator for the median is sample median, which is an unbiased estimator defined as:

$$M_1 = \widehat{M}_y \tag{1}$$

Up to the first degree of approximation, the MSE of the usual estimator is

$$MSE(M_1) = \lambda M_y^2 C_{M_y}^2 = Var(M_1) \tag{2}$$

where $\lambda = \frac{1}{4}(\frac{1}{n} - \frac{1}{N})$ and $C_{M_y} = \frac{1}{[M_y f_y(M_y)]}$.

(ii) Ratio estimator

The ratio estimator for the population median with known population median of auxiliary variable (M_x) has been introduced by Kuk and Mak [35], defined as:

$$M_2 = \widehat{M}_R = \widehat{M}_y \left(\frac{M_x}{\widehat{M}_x} \right) \tag{3}$$

The bias and MSE of the ratio estimator up to the first-degree approximation is

$$\left. \begin{aligned} B(M_2) &= \lambda M_y (C_{M_x}^2 - C_{M_{yx}}) \\ MSE(M_2) &= \lambda M_y^2 (C_{M_y}^2 + C_{M_x}^2 - 2C_{M_{yx}}) \end{aligned} \right\} \tag{4}$$

where $C_{M_x} = \frac{1}{[M_x f_x(M_x)]}$ and $C_{M_{yx}} = \rho_c C_{M_y} C_{M_x}$. The constants λ and C_{M_y} are already defined above.

(iii) Ratio type estimators

Singh et al. [36] introduced a variety of ratio-type estimators for the population median with a known coefficient of variation (C_x) and coefficient of kurtosis ($\beta_2(x)$) of auxiliary variable, which are defined as:

$$M_3 = \widehat{M}_y \left(\frac{M_x + C_x}{\widehat{M}_x + C_x} \right) \tag{5}$$

$$M_4 = \widehat{M}_y \left(\frac{M_x + \beta_2(x)}{\widehat{M}_x + \beta_2(x)} \right) \tag{6}$$

$$M_5 = \widehat{M}_y \left(\frac{M_x \beta_2(x) + C_x}{\widehat{M}_x \beta_2(x) + C_x} \right) \tag{7}$$

$$M_6 = \widehat{M}_y \left(\frac{M_x C_x + \beta_2(x)}{\widehat{M}_x C_x + \beta_2(x)} \right) \tag{8}$$

The biases of the above mentioned ratio-type estimators for population median up to the first-degree approximation are

$$\left. \begin{aligned} B(M_3) &= \lambda M_y (1 + \phi_1^2 C_{M_x}^2 - \phi_1 C_{M_{yx}}) \\ B(M_4) &= \lambda M_y (1 + \phi_2^2 C_{M_x}^2 - \phi_2 C_{M_{yx}}) \\ B(M_5) &= \lambda M_y (1 + \phi_3^2 C_{M_x}^2 - \phi_3 C_{M_{yx}}) \\ B(M_6) &= \lambda M_y (1 + \phi_4^2 C_{M_x}^2 - \phi_4 C_{M_{yx}}) \end{aligned} \right\} \tag{9}$$

Similarly, the mean square errors (MSEs) for above mentioned ratio-type estimators for population median up to the first-degree approximation are

$$\left. \begin{aligned} \text{MSE}(M_3) &= \lambda M_y^2 (C_{M_y}^2 + \phi_1^2 C_{M_x}^2 - 2\phi_1 C_{M_{yx}}) \\ \text{MSE}(M_4) &= \lambda M_y^2 (C_{M_y}^2 + \phi_2^2 C_{M_x}^2 - 2\phi_2 C_{M_{yx}}) \\ \text{MSE}(M_5) &= \lambda M_y^2 (C_{M_y}^2 + \phi_3^2 C_{M_x}^2 - 2\phi_3 C_{M_{yx}}) \\ \text{MSE}(M_6) &= \lambda M_y^2 (C_{M_y}^2 + \phi_4^2 C_{M_x}^2 - 2\phi_4 C_{M_{yx}}) \end{aligned} \right\} \tag{10}$$

where $\phi_1 = \frac{M_x}{M_x + C_x}$, $\phi_2 = \frac{M_x}{M_x + \beta_2(x)}$, $\phi_3 = \frac{M_x \beta_2(x)}{M_x \beta_2(x) + C_x}$ and $\phi_4 = \frac{M_x C_x}{M_x C_x + \beta_2(x)}$. The other constant terms are already defined above.

3. Proposed Phase I Monitoring of Process Location by Median Charts

The location monitoring of the process is carried out using different median estimators defined in Section 2. The abovementioned median estimators M_a ($a = 1, 2, \dots, 6$) have already been used in literature for the monitoring of location parameters in Phase II analysis. For example, the usual median estimator (M_1) has been extensively examined in different charting structures under known and unknown parameters (cf. [8,14,15,17,37]. The other auxiliary information-based median estimators ($M_2 - M_6$) have only been studied in Phase II studies [12,13,38]. Therefore, none of the above-mentioned median estimators ($M_1 - M_6$) have been studied under the case of Phase I control charting structure.

Assume that there are m (> 1) independent subgroups each with size n (> 1). The data point are represented by Y_{ji} , which symbolizes the i^{th} observation in j^{th} subgroup, $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, m$. In addition, assume that the underlying process distribution is continuous (e.g., normal). If the mean (μ_0) and standard deviation (σ_0) in-control process parameters are not known, then the initial subgroups ($j = 1, 2, \dots, m$) each of size n $\{Y_{j,1}, Y_{j,2}, \dots, Y_{j,n}\}$ (called Phase I dataset) are used to estimate the unknown parameters. Suppose that the subgroups are independent (within and between subgroups) and $Y_{ji} \sim N(\mu_0, \sigma_0^2)$. There are numerous methods to estimate μ_0 and σ_0 , but the most common method for estimating them is defined as:

$$\hat{\mu}_0 = \frac{1}{m} \sum_{j=1}^m \bar{Y}_j \tag{11}$$

$$\hat{\sigma}_0 = \frac{1}{c_{4,n}} \left(\frac{1}{m} \sum_{j=1}^m S_j \right) \tag{12}$$

In Equations (11) and (12), the mean and standard deviation of j^{th} subgroup are denoted by \bar{Y}_j and S_j , respectively, and in normal case, the constant $c_{4,n}$ is defined as $c_{4,n} = E\left(\frac{S_j}{\sigma_0}\right)$. Generally, this combination is used in the case of the mean (\bar{Y}) control chart, but not suitable for the median (\tilde{Y}) control chart. We used the following estimates of μ_0 and σ_0 to increase the efficiency of the median control chart.

$$\hat{\mu}'_0 = \text{median}(\text{medians}) = \begin{cases} \tilde{Y}_{j,(\frac{m+1}{2})} & \text{if } m \text{ is odd} \\ \frac{\tilde{Y}_{j,(\frac{m}{2})} + \tilde{Y}_{j,(\frac{m}{2}+1)}}{2} & \text{if } m \text{ is even} \end{cases} \tag{13}$$

where $\{\tilde{Y}_{j,(1)}, \tilde{Y}_{j,(2)}, \dots, \tilde{Y}_{j,(n)}\}$ are the ordered medians of j^{th} subgroup. Similarly, the standard deviation (σ_0) is estimated with the median absolute deviation (MAD), which is defined as the median of the absolute deviations from the median, mathematically represented as $MAD_i = 1.4826 * \text{median}(|Y_i - \tilde{Y}_j|)$,

and is a robust measure in the case of outliers or non-normality. The details about MAD can be seen in [26,39–41]. An unbiased MAD-based estimator of σ_0 is described as:

$$\hat{\sigma}'_0 = \frac{1}{d_{2,MAD}} \left(\frac{1}{m} \sum_{j=1}^m MAD_j \right) \tag{14}$$

where $d_{2,MAD}$ is a constant formulated for the normal distribution. If the process parameters μ_0 and σ_0 are unknown, then replace them by their estimators $\hat{\mu}'_0$ and $\hat{\sigma}'_0$ to generate the data for the estimation of g_a (mean of estimator M_a) and d_a (standard deviation of estimator M_a). The value of $\hat{\mu}'_0$ is calculated as the median of medians for m subgroups and $\hat{\sigma}'_0$ is the median absolute deviation (MAD) of m subgroups. The Phase I control chart is the plot of all charting statistics of m subgroups on the estimated control limits. The estimated control limits for random variable charting statistic M_a are defined as

$$\left. \begin{aligned} \widehat{LCL}_a &= \hat{\mu}'_0 - K_a \frac{\hat{\sigma}'_0}{\sqrt{n * d_{2,MAD}(a)}} \\ \widehat{CL}_a &= \hat{\mu}'_0 \\ \widehat{UCL}_a &= \hat{\mu}'_0 + K_a \frac{\hat{\sigma}'_0}{\sqrt{n * d_{2,MAD}(a)}} \end{aligned} \right\} \text{for } a = 1, 2, \dots, 6 \tag{15}$$

In the known parameter case, the lower control limit (LCL) and upper control limit (UCL) are constant, but for unknown parameter case, \widehat{LCL}_a and \widehat{UCL}_a are not constant and treated as random variables. It should be noted that, in the case of unknown parameters, the value of K_a depends on the choices of n, m, ρ_{xy} and the prefixed false alarm rate (α) or the in-control average run length (ARL_0). For a given probability of type-I error,

$$P(\widehat{LCL}_a \leq M_a \leq \widehat{UCL}_a) = 1 - \alpha, \text{ for } a = 1, 2, \dots, 6 \tag{16}$$

We will use the respective sample statistic M_a (for $a = 1, 2, \dots, 6$) as the plotting quantity on its corresponding charting structure. Furthermore, after introducing the shift in the mean of the study variable, an out-of-control signal will be observed after setting the limits for a specific control chart satisfying Equations (15) and (16). The process location parameter is said to be in-control if M_a falls inside its respective control limits, otherwise it is out-of-control.

4. Performance Evaluation of Median Control Charts in Phase I

In this section, by using some performance measures, we compared the efficiency of all median control charts mentioned in Section 2. In the Phase I and Phase II analysis, there are several measures available to assess the performance of the control chart, including variance, mean square error (MSE), relative mean square error (RMSE), relative efficiency (RE), average time to signal (ATS), probability to detect out-of-control data in Phase I, average run length (ARL), probability to signal and different run length properties (cf. [1,22,26,42–45] and the references therein). We considered the relative efficiency (RE) measure for the comparison of estimators and the probability to signal measure to compare the performances of the proposed control charts using the Monte Carlo simulations approach. We computed the standardized variance of the median estimators (denoted as SV_{M_a}), proposed by Rousseeuw and Croux [46] and the relative efficiencies of the median estimators as exercised by [47,48] for comparison, and to assess the efficiency of the median estimators considered in this study. The SV_{M_a} of a median estimator M_a (for $a = 1, 2, \dots, 6$) is calculated as

$$SV_{M_a} = \frac{nVAR(M_a)}{[E(M_a)]^2} \tag{17}$$

To get the natural measuring of the accuracy for a parameter, the denominator $[E(M_a)]^2$ is required (cf. [49]). By using SV_{M_a} , for any median estimator, the mathematical formula to compute relative efficiency (RE) is described as

$$RE_{M_a} = \frac{\min(SV_{M_a})}{SV_{M_a}} \tag{18}$$

SV_{M_a} and RE_{M_a} are computed by generating 10^5 samples of sizes $n = 5, 7, 10$ and 12 and considering correlations $\rho = 0.20, 0.50, 0.80$ and 0.95 by using the normal distribution. SV_{M_a} are presented numerically in Table 1 and RE_{M_a} are presented graphically in Figure 1 for comparison purposes. In each plot, the relative efficiencies are plotted on the y-axis and sample sizes are on the x-axis.

Table 1. $d_{2,MAD}$ and standardized variance for different median estimators.

n	ρ	M_1	M_2	M_3	M_4	M_5	M_6	M_1	M_2	M_3	M_4	M_5	M_6
		$d_{2,MAD}$						Standardized Variance (SV_{M_a})					
5	0.20	0.5363	0.7254	0.7094	0.5966	0.7199	0.5349	0.0576	0.1032	0.0989	0.0707	0.1017	0.0572
	0.50	0.5363	0.6210	0.6075	0.5222	0.6163	0.5037	0.0576	0.0760	0.0729	0.0544	0.0749	0.0508
	0.80	0.5362	0.4558	0.4462	0.4084	0.4524	0.4605	0.0575	0.0412	0.0396	0.0334	0.0406	0.0425
	0.95	0.5358	0.2676	0.2626	0.2917	0.2657	0.4238	0.0575	0.0143	0.0138	0.0171	0.0141	0.0361
7	0.20	0.4584	0.6172	0.6041	0.5103	0.6127	0.4577	0.0589	0.1051	0.1008	0.0725	0.1036	0.0586
	0.50	0.4585	0.5321	0.5210	0.4490	0.5283	0.4318	0.0589	0.0784	0.0753	0.0563	0.0773	0.0522
	0.80	0.4581	0.3999	0.3917	0.3566	0.3970	0.3960	0.0588	0.0445	0.0427	0.0356	0.0439	0.0440
	0.95	0.4578	0.2458	0.2412	0.2585	0.2441	0.3645	0.0587	0.0169	0.0163	0.0188	0.0166	0.0373
10	0.20	0.3720	0.4936	0.4835	0.4109	0.4901	0.3704	0.0554	0.0965	0.0927	0.0673	0.0952	0.0549
	0.50	0.3717	0.4204	0.4119	0.3577	0.4174	0.3477	0.0553	0.0702	0.0675	0.0511	0.0692	0.0484
	0.80	0.3715	0.3079	0.3018	0.2790	0.3057	0.3177	0.0552	0.0378	0.0363	0.0311	0.0373	0.0404
	0.95	0.3714	0.1829	0.1797	0.2005	0.1817	0.2931	0.0552	0.0134	0.0129	0.0161	0.0132	0.0344
15	0.20	0.3186	0.4244	0.4158	0.3538	0.4214	0.3181	0.0609	0.1073	0.1031	0.0749	0.1058	0.0607
	0.50	0.3186	0.3687	0.3614	0.3132	0.3662	0.3008	0.0609	0.0811	0.0780	0.0588	0.0800	0.0543
	0.80	0.3189	0.2843	0.2788	0.2534	0.2823	0.2776	0.0610	0.0483	0.0465	0.0385	0.0477	0.0463
	0.95	0.3189	0.1895	0.1861	0.1913	0.1883	0.2571	0.0610	0.0215	0.0208	0.0220	0.0212	0.0397

Usually, samples in rational subgroups are obtained from the process introduced by Shewhart, to apply the control charting structure. The main aim of the rational subgroups is to maximize the chances of differences between subgroups and to minimize differences within subgroups in the presence of assignable causes [1]. For our study, we considered 30 subgroups ($m = 30$) each of size n ($n = 5, 7, 10$ and 15) from the normal probability distribution. We divided subgroups into two portions as m_0 and m_1 (as $m = m_0 + m_1$). The stable subgroups are considered to be m_0 while m_1 (remaining subgroups) have variations in the form of shifts in parameter. The probability of signaling for $m_1 = 3, 6, 9$ and 12 was computed, which means that 3, 6, 9 or 12 samples out of 30 are considered to be contaminated with shift (δ). Phase I analysis is aimed preliminary, as quickly as possible, at the identification of these inconsistent (contaminated) samples. For the in-control (stable) samples, observations with a normal in-control mean and variance as $\mu_0 = 5$ and $\sigma_0 = 1$, respectively, were taken into account without loss of generality, whereas for the out-of-control situation, the sample observations are contaminated with the shift in mean of $\mu_1 = \mu_0 + \delta\sigma_0$, where δ is the shift amount. The mean vector $\underline{\mu}$ with the in-control mean of study variable Y as μ_0 and known mean of auxiliary variable X as μ_X and sigma matrix Σ with population variances of study and auxiliary variables as σ_Y^2 and σ_X^2 , respectively for this bivariate normally distributed process are expressed as:

$$\underline{\mu} = \begin{pmatrix} \mu_0 + \delta\sigma_Y \\ \mu_X \end{pmatrix} = \begin{pmatrix} 5 + \delta \\ 5 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_Y^2 & Cov(Y, X) \\ Cov(X, Y) & \sigma_X^2 \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{19}$$

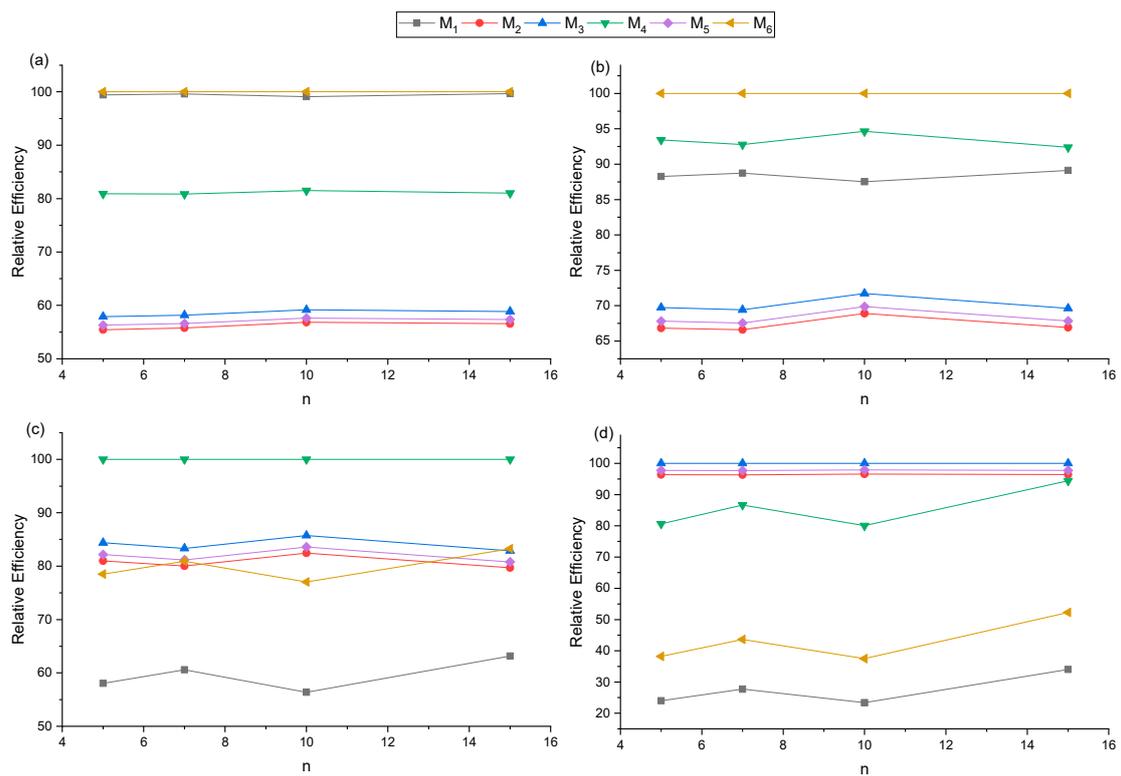


Figure 1. Comparison of relative efficiency of different median estimators at correlations: (a) $\rho = 0.30$; (b) $\rho = 0.50$; (c) $\rho = 0.80$; (d) $\rho = 0.95$.

For the dataset of m samples (subgroups), the control limits specified in Section 3 are determined for the structures described in Section 2. For all control charts, the values of control limit multipliers (K) are correctly chosen to achieve a pre-selected value of α (false alarm probability). We presented values of K for $n = 10$ and correlations as $\rho = 0.20, 0.50, 0.80$ and 0.95 in Table 2.

Table 2. Control chart multipliers (K_a) for obtaining $\alpha = 0.01$ using $n = 10$ at different levels of ρ .

ρ	M_1	M_2	M_3	M_4	M_5	M_6
0.20	5.824	7.712	7.581	6.210	7.630	5.640
0.50	5.800	7.011	6.720	5.474	6.872	5.340
0.80	5.680	4.910	4.803	4.360	4.838	4.900
0.95	5.712	2.905	2.806	3.151	2.859	4.462

First of all, we observe the probability that a single sample will be signaled (denoted by α^*) and then, by using the relationship $\alpha = 1 - (1 - \alpha^*)^m$, we can calculate the overall false alarm probability (α) for m samples. One may see [6,26,50,51] for further details. Once the control charts limits are fixed, median estimators M_a (for $a = 1, 2, \dots, 6$) were used to detect out-of-control signals according to their respective control limits. We repeat this process 10,000 times to fix the $\alpha = 0.01$ and chose values of K using $m = 30$ and $n = 5, 7, 10$ and 15 for all median control charts. We also considered different values of m_1 as 3, 6, 9 and 12 and considered correlations $\rho = 0.20, 0.50, 0.80$ and 0.95 to investigate the signaling probabilities. Such m_1 values indicate the different situations in which the samples out of 30 are known to have a change in mean with shift δ . The probability of signaling various δ values for $m_1 = 3, 6, 9$ and 12 are graphically illustrated (see Figures 2–8).

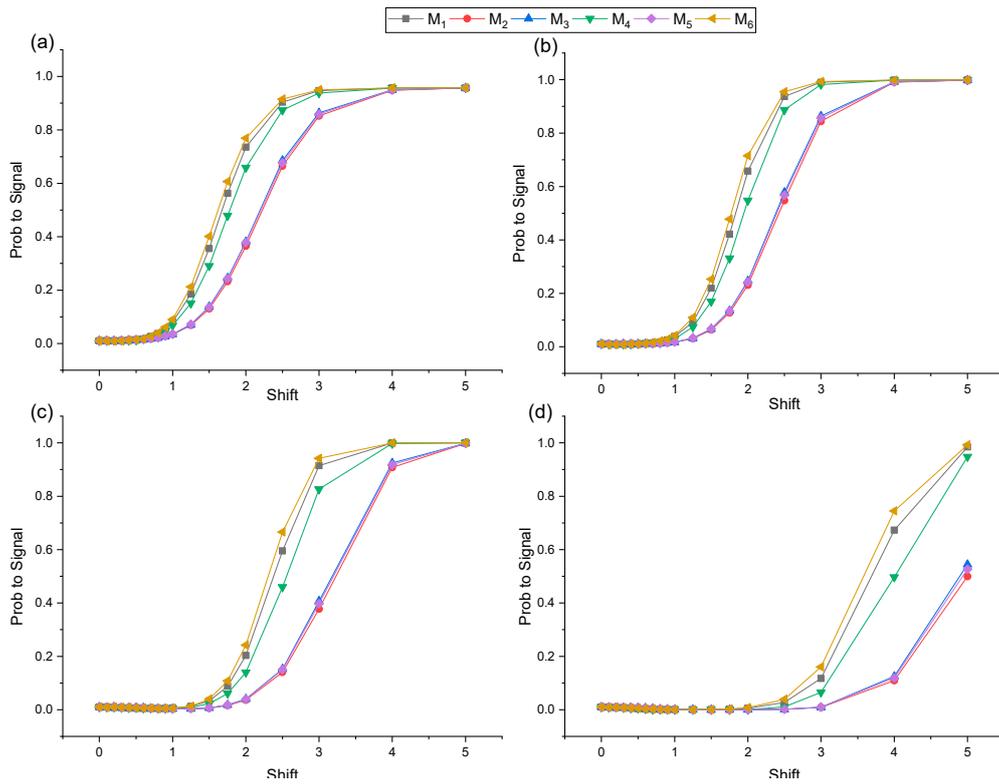


Figure 2. Probability to signal plot for different auxiliary information based median charts using $n = 10$, $m = 30$, $\rho = 0.30$ and $\alpha = 0.01$ at (a) $m_1 = 3$, (b) $m_1 = 6$, (c) $m_1 = 9$, (d) $m_1 = 12$.

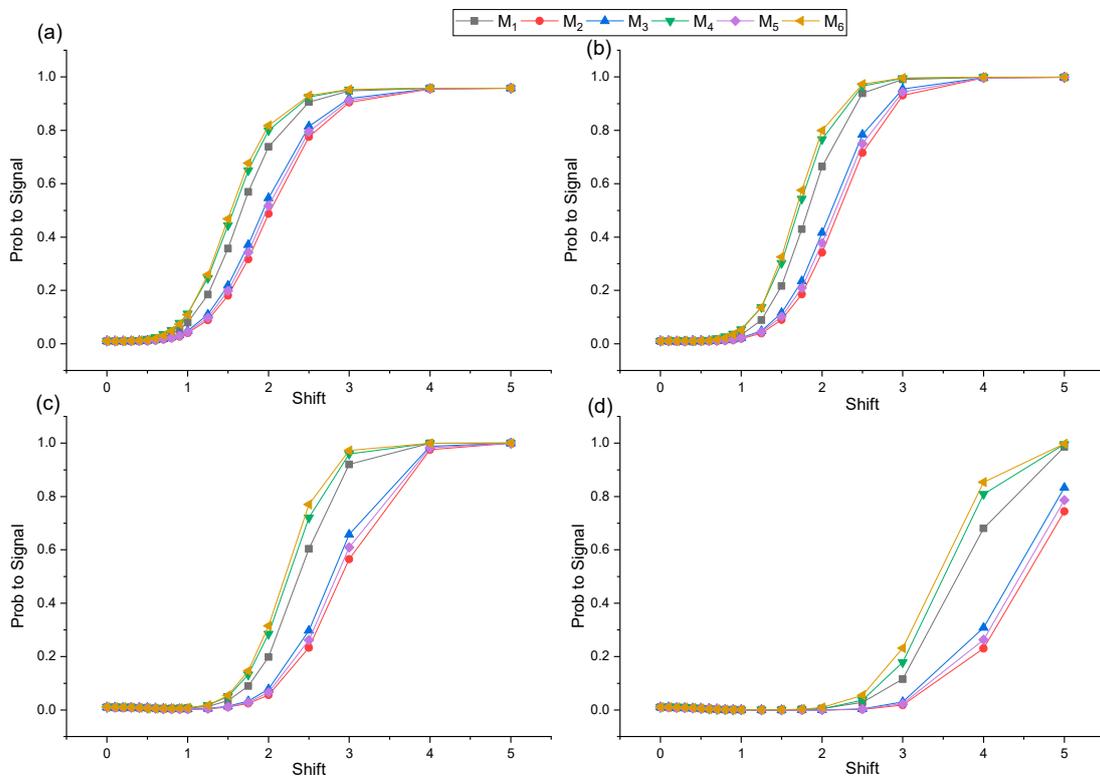


Figure 3. Probability to signal plot for different auxiliary information based median charts using $n = 10$, $m = 30$, $\rho = 0.50$ and $\alpha = 0.01$ at (a) $m_1 = 3$, (b) $m_1 = 6$, (c) $m_1 = 9$, (d) $m_1 = 12$.

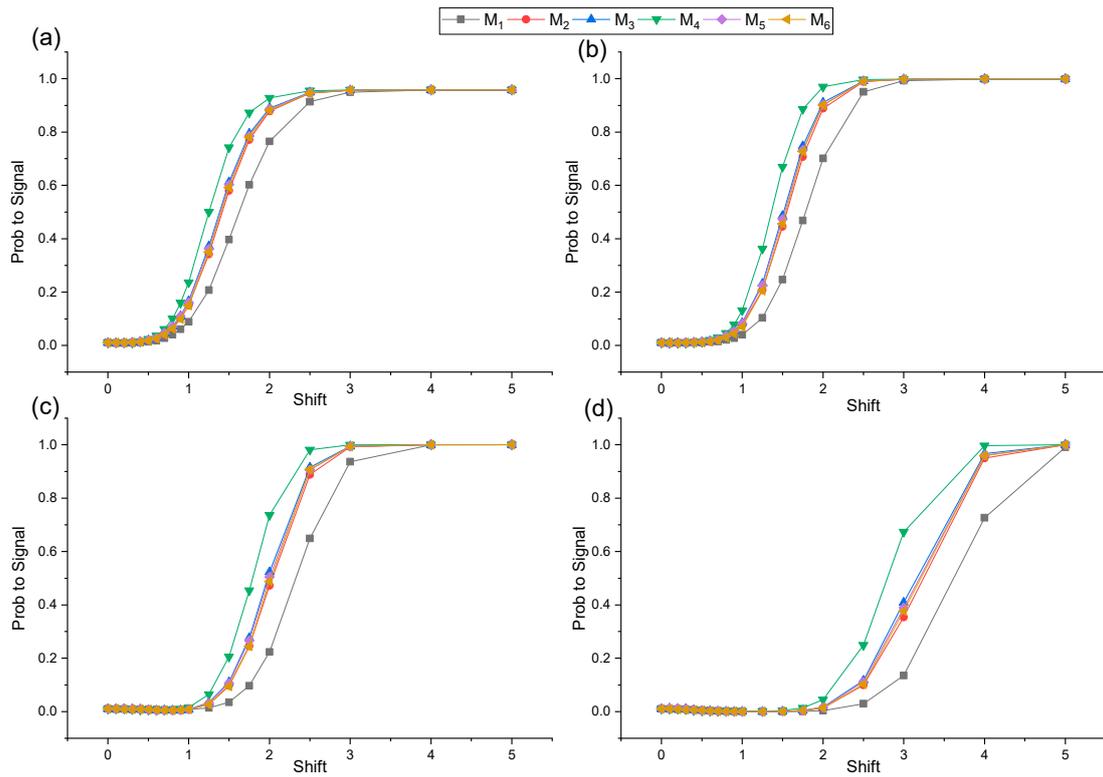


Figure 4. Probability to signal plot for different auxiliary information based median charts using $n = 10$, $m = 30$, $\rho = 0.80$ and $\alpha = 0.01$ at (a) $m_1 = 3$, (b) $m_1 = 6$, (c) $m_1 = 9$, (d) $m_1 = 12$.

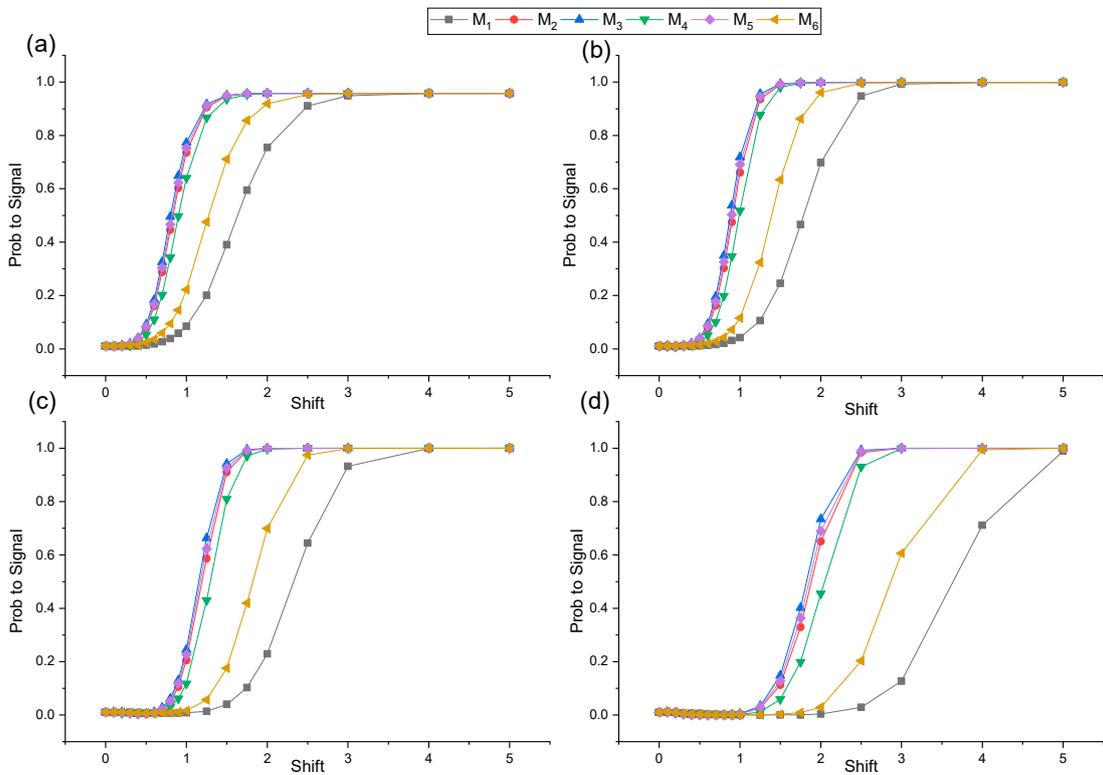


Figure 5. Probability to signal plot for different auxiliary information based median charts using $n = 10$, $m = 30$, $\rho = 0.95$ and $\alpha = 0.01$ at (a) $m_1 = 3$, (b) $m_1 = 6$, (c) $m_1 = 9$, (d) $m_1 = 12$.

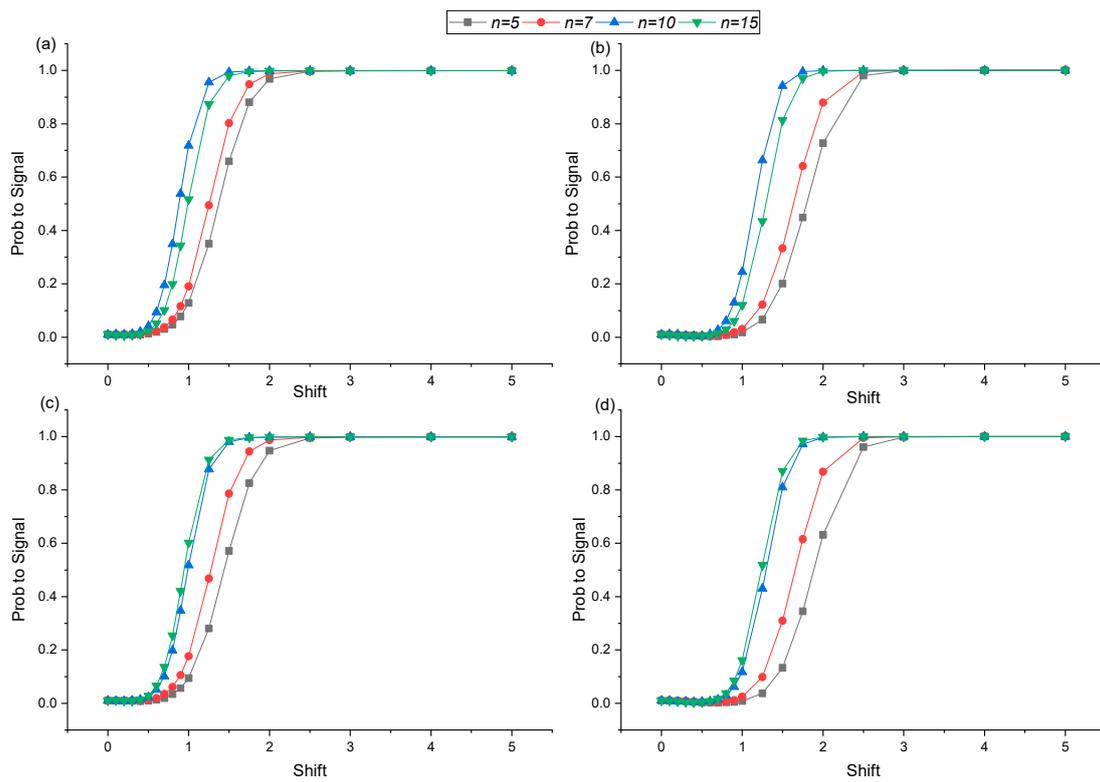


Figure 6. Probability to signal plot for different sample sizes using $m = 30$, $\rho = 0.95$ and $\alpha = 0.01$ of: (a) M_3 and $m_1 = 6$; (b) M_3 and $m_1 = 9$; (c) M_4 and $m_1 = 6$; (d) M_4 and $m_1 = 9$.

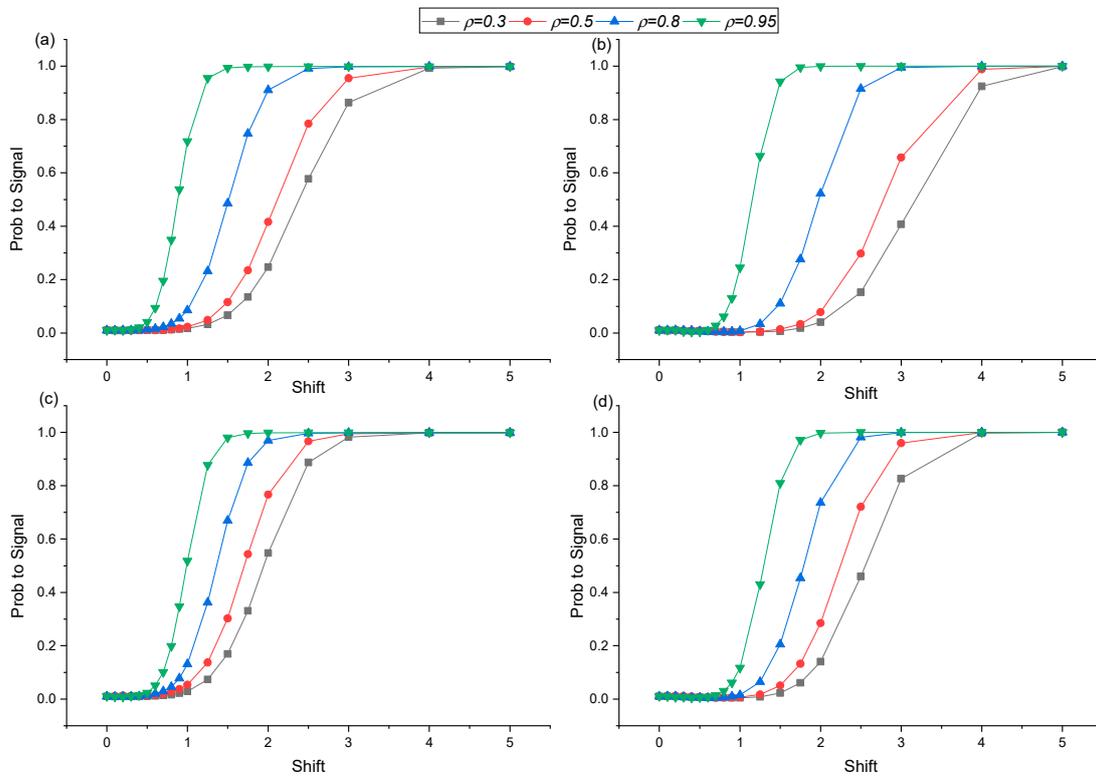


Figure 7. Probability to signal plot for different correlations using $m = 30$, $n = 10$ and $\alpha = 0.01$ of: (a) M_3 and $m_1 = 6$; (b) M_3 and $m_1 = 9$; (c) M_4 and $m_1 = 6$; (d) M_4 and $m_1 = 9$.

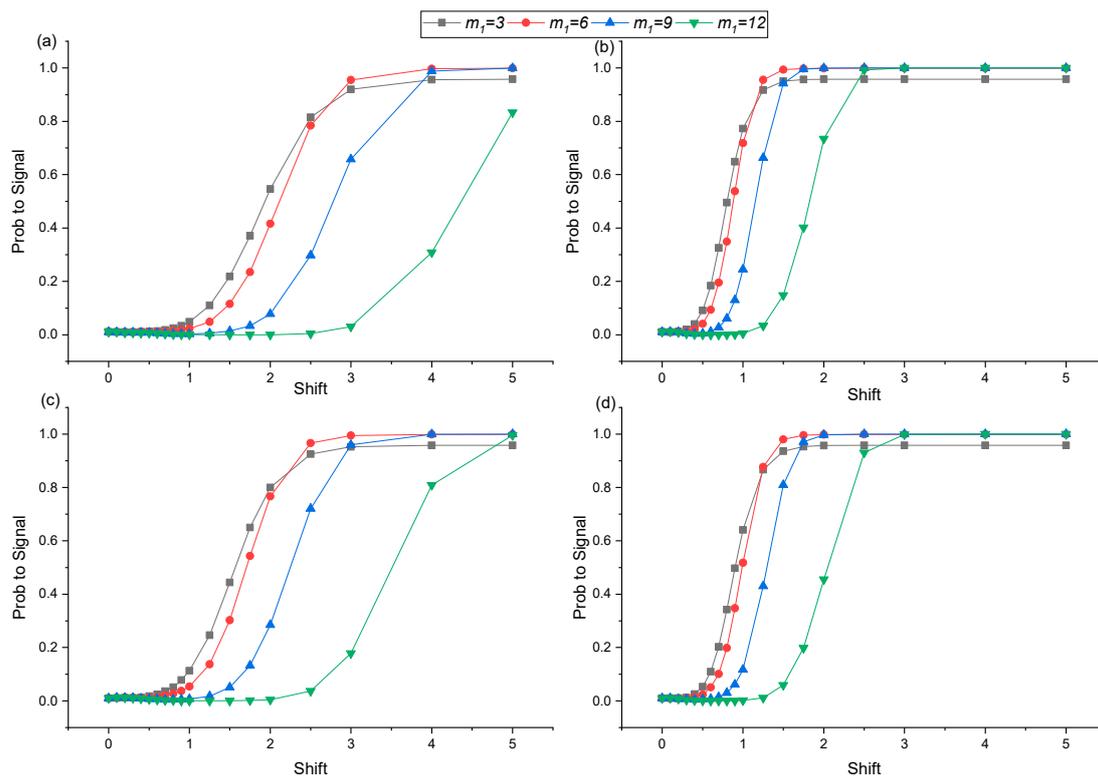


Figure 8. Probability to signal plot for different m_1 values using $m = 30$, $n = 10$ and $\alpha = 0.01$ of: (a) M_3 and $\rho = 0.50$; (b) M_3 and $\rho = 0.95$; (c) M_4 and $\rho = 0.50$; (d) M_4 and $\rho = 0.95$.

5. Discussion and Comparative Analysis

In this section, we present a comparative debate based on calculations conducted in Section 4 for the various median control charts investigated in this study. The charts that have the highest probability of signaling out-of-control samples are considered better than others for a fixed false alarm probability (α). The results include the values of a constant $d_{2,MAD}(n)$ and standardized variance SV_{M_a} defined in Equation (17), and are presented in Table 1. Figure 1 represents the graphical comparison of estimators considered in this study with respect to their relative efficiencies, defined in Equation (18). Table 2 represents the control chart multipliers K_a , defined in Equation (15) for correlations $\rho = 0.20, 0.50, 0.80$ and 0.95 and $n = 10$.

Similarly, the comparison of the median control charts (M_3 and M_3) with respect to correlations is presented in Figure 7 for $m = 30$, $n = 10$, $m_1 = 6$ and 9 . Figure 8 represents the effect of unstable (or stable) subgroups effect on the efficiency of median charts (M_3 and M_3) for $m = 30$, $n = 10$, $\rho = 0.5$ and 95 . The signaling probabilities of the median charts are provided in Table 3 for $m = 30$, $n = 10$, $\rho = 0.80$, $m_1 = 3, 6, 9$ and 12 at various values of shift (δ). Similarly, the signaling probabilities for the median control charts are also presented in graphical forms for $m = 30$, $n = 10$, $m_1 = 3, 6, 9$ and 12 $\rho = 0.20, 0.50, 0.80$ and 0.95 at various values of shift (δ) in Figures 2–5. The effect of sample size on the performance of the median control charts (M_3 and M_3) is presented in Figure 6 for $m = 30$, $m_1 = 6$ and 9 $\rho = 0.95$.

From Table 1, we can observe that the auxiliary information-based median estimators have less standardized variance as compared to the usual median estimator. Furthermore, within the comparison of auxiliary information based median estimators, M_6 has the lowest standardized variance for a small correlation between the study and auxiliary variables (ρ), M_4 has the lowest standardized variance for a moderate correlation (ρ) and M_3 has the lowest standardized variance for a large value of the correlation (ρ). Alternatively, M_2 has the largest standardized variance for a low correlation and the

usual median estimator M_1 has the largest standardized variance for a moderate and high correlation. Similar behavior can be observed from Figure 1 in the form of relative efficiencies.

Table 3. Probability to signal values for the different median charts when $m = 30, n = 10, \rho = 0.80$ at various levels of δ and m_1 when $\alpha = 0.01$.

	M_1	M_2	M_3	M_4	M_5	M_6	M_1	M_2	M_3	M_4	M_5	M_6
δ	$m_1=3$						$m_1=6$					
0.0	0.0104	0.0101	0.0101	0.0102	0.0104	0.0101	0.0104	0.0101	0.0101	0.0102	0.0104	0.0101
0.1	0.0092	0.0094	0.0091	0.0105	0.0101	0.0104	0.0082	0.0096	0.0095	0.0088	0.0097	0.0108
0.2	0.0091	0.0094	0.0088	0.0107	0.0102	0.0101	0.0099	0.0097	0.0100	0.0097	0.0099	0.0088
0.3	0.0094	0.0115	0.0106	0.0101	0.0117	0.0115	0.0093	0.0101	0.0101	0.0098	0.0111	0.0106
0.4	0.0110	0.0136	0.0146	0.0143	0.0142	0.0132	0.0105	0.0111	0.0115	0.0104	0.0121	0.0113
0.5	0.0132	0.0186	0.0200	0.0241	0.0207	0.0173	0.0102	0.0133	0.0131	0.0134	0.0141	0.0116
0.6	0.0172	0.0275	0.0289	0.0365	0.0290	0.0252	0.0134	0.0158	0.0162	0.0206	0.0165	0.0137
0.7	0.0273	0.0429	0.0460	0.0609	0.0463	0.0404	0.0146	0.0207	0.0225	0.0293	0.0222	0.0195
0.8	0.0404	0.0677	0.0712	0.1001	0.0710	0.0625	0.0207	0.0339	0.0347	0.0459	0.0358	0.0294
0.9	0.0608	0.1025	0.1099	0.1598	0.1092	0.0992	0.0271	0.0489	0.0542	0.0779	0.0524	0.0442
1.0	0.0886	0.1504	0.1662	0.2358	0.1610	0.1497	0.0395	0.0768	0.0852	0.1309	0.0825	0.0692
1.3	0.2077	0.3419	0.3720	0.5002	0.3620	0.3497	0.1033	0.2081	0.2316	0.3625	0.2249	0.2056
1.5	0.3971	0.5802	0.6125	0.7416	0.6028	0.5918	0.2467	0.4450	0.4855	0.6692	0.4730	0.4558
1.8	0.6018	0.7708	0.7939	0.8726	0.7851	0.7809	0.4681	0.7075	0.7474	0.8859	0.7322	0.7272
2.0	0.7650	0.8779	0.8896	0.9277	0.8857	0.8819	0.7010	0.8880	0.9104	0.9692	0.9028	0.9009
2.5	0.9137	0.9459	0.9486	0.9546	0.9475	0.9460	0.9505	0.9888	0.9912	0.9965	0.9903	0.9901
3.0	0.9491	0.9564	0.9568	0.9576	0.9567	0.9564	0.9926	0.9976	0.9979	0.9985	0.9977	0.9977
4.0	0.9576	0.9577	0.9577	0.9578	0.9577	0.9578	0.9986	0.9987	0.9988	0.9988	0.9987	0.9987
5.0	0.9577	0.9577	0.9577	0.9578	0.9577	0.9578	0.9988	0.9988	0.9988	0.9988	0.9988	0.9988
	$m_1 = 9$						$m_1 = 12$					
0.0	0.0104	0.0101	0.0101	0.0102	0.0104	0.0101	0.0104	0.0101	0.0101	0.0102	0.0104	0.0101
0.1	0.0089	0.0102	0.0103	0.0104	0.0114	0.0108	0.0094	0.0115	0.0121	0.0099	0.0125	0.0101
0.2	0.0094	0.0108	0.0104	0.0110	0.0105	0.0083	0.0086	0.0114	0.0107	0.0101	0.0126	0.0084
0.3	0.0080	0.0090	0.0094	0.0085	0.0100	0.0082	0.0088	0.0096	0.0103	0.0083	0.0102	0.0066
0.4	0.0092	0.0089	0.0087	0.0067	0.0089	0.0085	0.0073	0.0058	0.0060	0.0047	0.0065	0.0062
0.5	0.0082	0.0066	0.0064	0.0061	0.0072	0.0080	0.0054	0.0041	0.0037	0.0030	0.0042	0.0040
0.6	0.0075	0.0047	0.0049	0.0050	0.0050	0.0060	0.0040	0.0026	0.0024	0.0014	0.0025	0.0018
0.7	0.0068	0.0050	0.0051	0.0054	0.0053	0.0047	0.0026	0.0010	0.0011	0.0016	0.0014	0.0012
0.8	0.0063	0.0042	0.0049	0.0067	0.0046	0.0050	0.0016	0.0006	0.0006	0.0009	0.0006	0.0006
0.9	0.0058	0.0055	0.0052	0.0100	0.0056	0.0062	0.0009	0.0000	0.0000	0.0008	0.0000	0.0003
1.0	0.0081	0.0076	0.0080	0.0155	0.0085	0.0082	0.0005	0.0000	0.0002	0.0010	0.0000	0.0001
1.3	0.0138	0.0280	0.0339	0.0643	0.0310	0.0269	0.0003	0.0000	0.0001	0.0010	0.0001	0.0001
1.5	0.0352	0.0971	0.1109	0.2059	0.1063	0.0947	0.0001	0.0008	0.0009	0.0040	0.0008	0.0007
1.8	0.0969	0.2456	0.2765	0.4538	0.2663	0.2448	0.0005	0.0030	0.0042	0.0137	0.0034	0.0045
2.0	0.2236	0.4724	0.5228	0.7366	0.5046	0.4876	0.0032	0.0140	0.0167	0.0444	0.0156	0.0138
2.5	0.6493	0.8884	0.9157	0.9813	0.9060	0.9076	0.0294	0.0994	0.1168	0.2489	0.1114	0.1020
3.0	0.9369	0.9923	0.9948	0.9993	0.9940	0.9941	0.1354	0.3540	0.4088	0.6732	0.3891	0.3771
4.0	0.9995	0.9999	0.9999	1.0000	0.9999	0.9999	0.7266	0.9497	0.9668	0.9967	0.9610	0.9590
5.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9910	0.9998	0.9999	1.0000	0.9998	0.9998

The charting structure of the M_6 control chart showed the highest signaling probability or best detection ability to detect an out-of-control signal at large values of shift (δ) followed by the usual median control chart M_1 for the small correlation coefficient (ρ), while the control chart structure of M_2 performs worst for the same situation at different values of m_1 (see Figure 2). Similarly, the control charting structure of M_5 performs better for small shifts (*i.e.*, $\delta < 0.8$). Similarly, at $\rho = 0.5$, control charting structures M_4 and M_6 have the largest signaling probabilities at small (*i.e.*, $\delta < 1.0$) and large (*i.e.*, $\delta \geq 1.0$) shift values, respectively. Meanwhile, the structure of M_2 chart has the lowest signaling probabilities (see Figure 3).

From Figure 4 and Table 3, we observed that the design structure of the M_4 median chart exhibited the best performance while the M_1 median chart structure showed the worst performance at both large and small shift values for $\rho = 0.8$ and $m_1 = 3$ and 6. With the increasing value of m_1 , the outer design structure of M_4 performs at a large value of shift (*i.e.*, $\delta \geq 1.0$) while the M_1 structure showed the best performance at small shift values (*i.e.*, $\delta < 1.0$).

The design structure of the M_3 median chart showed the best performance at $\rho = 0.95$ for different values of m_1 (see Figure 5). There are some other charts which showed their performances at different values of m_1 . For example, the $m_1 = 3$, M_3 chart has the largest signaling probabilities at shift values ($\delta \leq 2.0$), and at shift values ($2.0 < \delta \leq 3.0$), the M_4 control chart has the largest signaling probabilities, and for ($\delta \geq 3.0$), the M_6 chart has the highest detection ability. Similarly, for $m_1 = 6$, the M_3 median chart showed the best performance at $\delta \leq 2.5$ and the outer design structure of the M_4 chart performs at $\delta > 2.5$. For $m_1 = 9$ and 12, the usual median control chart M_1 showed the best performance at a small shift (*i.e.*, $\delta \leq 0.8$) and the structure of the M_3 median chart has the largest signaling probabilities at moderate and large shift values (*i.e.*, $\delta > 0.8$).

Performance tests for the median charts ($M_1 - M_6$) indicate that all charts boost the probability of signaling in order to detect an out-of-control signal for an increased sample size (n) and correlation coefficient (ρ). We present this comparison for charts M_3 and M_4 at $\rho = 0.95$ and $m_1 = 6$ and 9 in Figures 6 and 7. From Figure 6, the M_3 median design structure showed the best performance at $n = 10$, while the M_4 median control chart showed the best performance at $n = 15$. Both charting structures M_3 and M_4 exhibited poor performance at a small sample size.

All charting structures improved the performance with increments in correlation. From Figure 7, we observe that the design structures of the M_3 and M_4 median control charts exhibited the best performance at a large correlation coefficient (*i.e.*, $\rho = 0.95$) and poor performance at a small value of correlation coefficient (*i.e.*, $\rho = 0.3$).

The comparison of different values of m_1 for the M_3 and M_4 median control charts in Figure 8 revealed that, for $\rho = 0.5$, both the M_3 and M_4 median control charts have the largest signaling probabilities at $m_1 = 3$ against small and moderate values of shift (*i.e.*, $\delta \leq 2.5$) followed by at $m_1 = 6$. Similarly, for $\rho = 0.5$, both charting structures M_3 and M_4 have the following pattern of largest signaling probabilities at different values of shifts as: largest signaling probabilities at $m_1 = 3$ against $\delta \leq 1.0$, at $m_1 = 6$ against $1.0 < \delta \leq 2.0$, at $m_1 = 9$ against $2.0 < \delta \leq 3.0$ and at $m_1 = 12$ against shift values $\delta > 3.0$.

6. Practical Implementation

In this section, we considered the practical implantation of our proposed Phase I monitoring of location parameters with some real processes. For this purpose, we first considered some real process data, which are available online, and we selected the study variable and correlated the auxiliary variable from the process data. From the dataset, we selected random subgroups m , each of size n . After this, we calculated control limits for in-control initial sample m_1 to obtain prefixed false alarm rate α . After obtaining the control limits, we added shift δ in the mean of study variable and calculated plotting statistics, and then observed the out-of-control samples. All of this algorithm is developed using R software.

6.1. Case Study-I

In this section, a real-life example is shown in the application of the proposed Phase I median control charts. For this purpose, we considered real data from the cement industry where the compression strength of concrete is crucial for the quality of cement production. In a laboratory experiment, this set of data on concrete compressive strength for a particular mixture were calculated and were firstly proposed in [52], before being used in many research articles (see [13]).

In this multivariate process, the study variable concrete compressive strength was monitored with some other auxiliary variables, such as cement, blast, fly ash, water, superplasticizer, coarse aggregate,

fine aggregate, and age. For our study, we considered bivariate data on the measurement for concrete compressive strength (measured in MPa), containing the study variable Y and the quantity of cement (measured in kg in m^3 mixture) as auxiliary variable X .

These variables were not scaled, so first we scaled these variables to make their mean set to five, with a variance of one for the similarity of the simulation study, mentioned in Section 4. The dataset contains 1030 observations, and we randomly selected 30 subgroups (as $m = 30$) each of size 10 (as $n = 10$). For this dataset, we considered the values of m_1 as six and nine, which means that out of 30 subgroups, six and nine samples are contaminated with shift (δ) 2.6 and 3, respectively. For illustration purposes, we considered two median estimators (M_4 and M_6) based on auxiliary variables and compared their performance with the usual median estimator (M_1). First, we constructed their control limits based on in-control samples (m_0) and then analyzed their detection ability by introducing the shift (δ) in the mean of study variable. The values of plotting statistics (M_1, M_4 and M_6) for both cases $m_1 = 6$ with $\delta = 2.5$ and $m_1 = 9$ with $\delta = 3$ are presented in Table 4 and their control limits (LCL and UCL) are presented in Table 5. Similarly, the graphical display of plotting statistics with their respective control limits against $m_1 = 6$ with $\delta = 2.5$ and $m_1 = 9$ with $\delta = 3$ are shown in Figure 9.

Table 4. Plotting statistics of the median charts based on M_1, M_4 and M_6 for concrete data.

	$m_1=6, \delta=2.5$			$m_1=9, \delta=3.0$		
	M_1	M_4	M_6	M_1	M_4	M_6
1	4.972	5.084	5.016	4.972	5.084	5.016
2	4.677	5.096	4.836	4.677	5.096	4.836
3	4.440	4.462	4.448	4.440	4.462	4.448
4	4.777	3.859	4.362	4.777	3.859	4.362
5	5.106	5.292	5.179	5.106	5.292	5.179
6	5.243	4.689	5.006	5.243	4.689	5.006
7	7.375	6.721	7.099	7.375	6.721	7.099
8	4.758	5.061	4.874	4.758	5.061	4.874
9	5.387	4.500	4.993	5.387	4.500	4.993
10	4.375	4.545	4.441	4.375	4.545	4.441
11	5.801	5.165	5.529	5.801	5.165	5.529
12	4.685	5.506	4.982	4.685	5.506	4.982
13	6.705	5.721	6.273	6.705	5.721	6.273
14	7.016	6.393	6.753	7.016	6.393	6.753
15	4.413	5.203	4.698	4.413	5.203	4.698
16	4.761	4.937	4.830	4.761	4.937	4.830
17	4.176	3.901	4.061	4.176	3.901	4.061
18	4.711	5.155	4.879	4.711	5.155	4.879
19	4.723	5.407	4.975	4.723	5.407	4.975
20	5.082	4.649	4.899	5.082	4.649	4.899
21	5.665	5.029	5.392	5.665	5.029	5.392
22	3.917	4.549	4.148	6.917	7.549	7.148
23	4.035	4.328	4.147	7.035	7.328	7.147
24	4.158	4.688	4.355	7.158	7.688	7.355
25	7.840	7.275	7.599	8.340	7.775	8.099
26	6.288	6.797	6.476	6.788	7.297	6.976
27	6.906	7.704	7.194	7.406	8.204	7.694
28	6.593	7.253	6.834	7.093	7.753	7.334
29	8.377	9.223	8.688	8.877	9.723	9.188
30	7.898	7.641	7.793	8.398	8.141	8.293

Table 5. Control limits of the median charts based on M_1, M_4 and M_6 for the concrete data.

	M_1		M_4		M_6	
	LCL	UCL	LCL	UCL	LCL	UCL
$m_1 = 6, \delta = 2.5$	2.836	7.217	3.521	6.730	3.170	6.805
$m_1 = 9, \delta = 3.0$	2.297	8.333	3.236	7.259	2.916	7.278

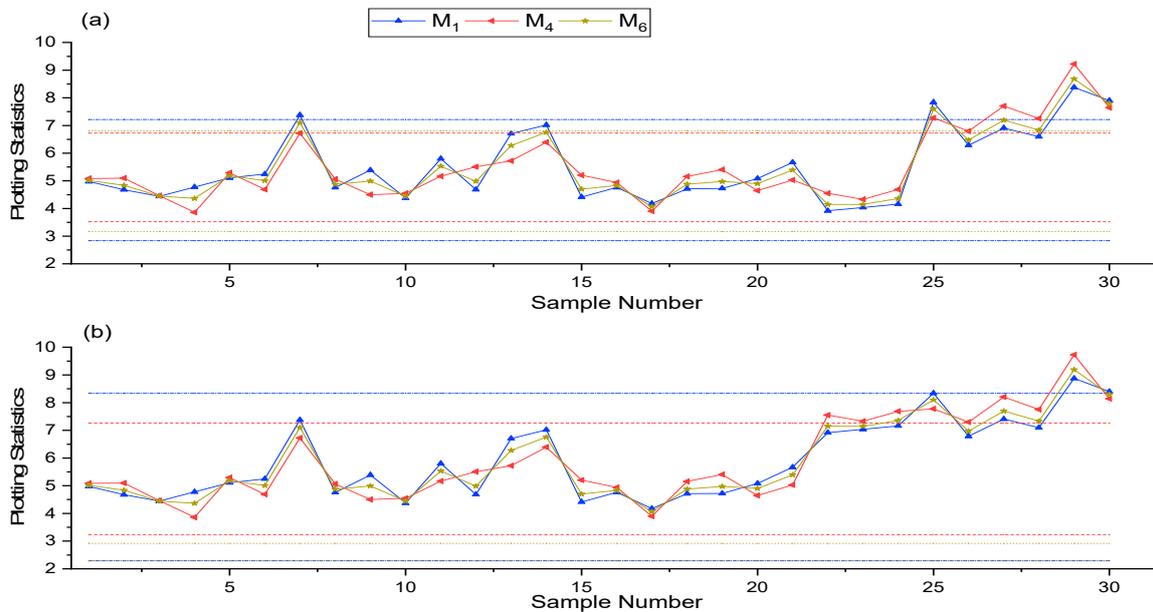


Figure 9. Graphical displays of plotting statistics corresponding to their control limits at (a) $m_1 = 6$ and $\delta = 2.5$; (b) $m_1 = 9$ and $\delta = 3.0$.

From the results, it is observed that the plotting statistic M_1 detects shift at subgroup numbers 25, 29 and 30, whereas M_4 detects shift at subgroups 25, 26, 27, 28, 29 and 30. Similarly, M_6 detects shift at subgroups 25, 27, 28, 29 and 30 in case of $m_1 = 6$ with $\delta = 2.5$. These results showed the superiority of M_4 chart over M_6 and M_1 . Similarly, in the case of $m_1 = 9$ with $\delta = 3$, again, the M_4 chart detects all nine out-of-control shifts, whereas M_6 and M_1 detect six and three, respectively, out-of-control subgroups. Similar findings of a higher detection ability of the control structure M_4 may be seen in Figure 9. It should be noted in both of the m_1 choices in the two cases, the charting structure M_4 displayed the superior detection ability compared to M_6 and M_1 . The superiority of the M_4 control chart may also be expected with the comparison of other charts considered in this study, which is based on the findings of Section 5.

6.2. Case Study-II

In this section, we provide another case study to demonstrate the application of our proposed auxiliary information-based median control charts for the Phase I monitoring of location parameters. For this purpose, we used real-life data of the non-iso-thermal continuous stirred tank chemical reactor (CSTR) process, as first provided by Marlin [53]. At a sampling interval of 30 s, the dataset was collected with 1024 values. The CSTR process mechanism includes nine process variables, and Yoon and MacGregor [54] and Xiangrong et al. [55] provide details of those variables of the process. The CSTR process has been extensively used for defect detection and diagnosis (see [54,55]). In this study, we considered outlet temperature T (measured in K) as the study variable Y and cooling temperature T_C (measured in K) as the auxiliary variable X . We randomly selected 30 subgroups (as $m = 30$) each of size 10 (as $n = 10$) from the dataset consisting of 1024 values. Again, for this dataset, we considered the values of m_1 as three and six, which means that out of 30 subgroups, three and six samples are

contaminated with shift (δ) 2.0 and 3.0, respectively. Again, in this case study, for illustration purposes, we considered two median estimators (M_4 and M_6) based on the auxiliary variable and compared their performance with the usual median estimator (M_1).

Initially, we constructed their control limits based on m_0 (in-control samples), which are presented in Table 6 and the values of plotting statistics (M_1 , M_4 and M_6) for both cases $m_1 = 3$ with $\delta = 2.0$ and $m_1 = 6$ with $\delta = 3$ are presented in Table 7. The graphical presentation of results are presented in Figure 10.

Table 6. Control limits of the median charts based on M_1, M_4 and M_6 for the non-iso-thermal continuous stirred tank chemical reactor (CSTR) data.

	M_1		M_4		M_6	
	LCL	UCL	LCL	UCL	LCL	UCL
$m_1 = 3, \delta = 2.0$	3.267	6.067	3.525	6.582	3.545	6.636
$m_1 = 6, \delta = 3.0$	2.531	8.057	3.386	6.779	2.995	7.280

Table 7. Plotting statistics of the median charts based on M_1, M_4 and M_6 for CSTR Data.

	$m_1=3, \delta=2.0$			$m_1=6, \delta=3.0$		
	M_1	M_4	M_6	M_1	M_4	M_6
1	5.297	4.485	4.939	5.297	4.485	4.939
2	5.291	5.737	5.461	5.291	5.737	5.461
3	3.671	4.632	4.003	3.671	4.632	4.003
4	4.629	5.094	4.804	4.629	5.094	4.804
5	5.988	5.534	5.798	5.988	5.534	5.798
6	5.630	5.527	5.588	5.630	5.527	5.588
7	5.445	4.693	5.117	5.445	4.693	5.117
8	5.243	5.035	5.158	5.243	5.035	5.158
9	5.092	4.940	5.030	5.092	4.940	5.030
10	4.821	4.709	4.776	4.821	4.709	4.776
11	5.010	5.147	5.064	5.010	5.147	5.064
12	4.388	4.940	4.593	4.388	4.940	4.593
13	5.685	5.071	5.423	5.685	5.071	5.423
14	3.023	3.543	3.211	3.023	3.543	3.211
15	3.208	4.571	3.642	3.208	4.571	3.642
16	5.490	5.442	5.471	5.490	5.442	5.471
17	3.291	3.774	3.469	3.291	3.774	3.469
18	6.795	6.190	6.539	6.795	6.190	6.539
19	4.334	4.137	4.253	4.334	4.137	4.253
20	5.525	5.340	5.449	5.525	5.340	5.449
21	4.509	4.748	4.602	4.509	4.748	4.602
22	4.456	4.232	4.364	4.456	4.232	4.364
23	4.131	4.684	4.336	4.131	4.684	4.336
24	5.373	5.360	5.368	5.373	5.360	5.368
25	4.979	4.745	4.882	7.979	7.745	7.882
26	5.456	5.444	5.452	8.456	8.444	8.452
27	5.034	6.101	5.413	8.034	9.101	8.413
28	7.921	7.457	7.726	8.921	8.457	8.726
29	6.153	6.213	6.177	7.153	7.213	7.177
30	7.036	6.865	6.966	8.036	7.865	7.966

From the results of CSTR dataset, it is observed that the plotting statistic M_1 detects only two out-control samples at subgroup numbers 28 and 30, whereas M_4 detects four out-control samples at subgroups 17, 28, 29 and 30. Similarly, M_6 detects five out-control samples at subgroups 14, 17, 28, 29 and 30 in case of $m_1 = 3$ with $\delta = 2.0$.

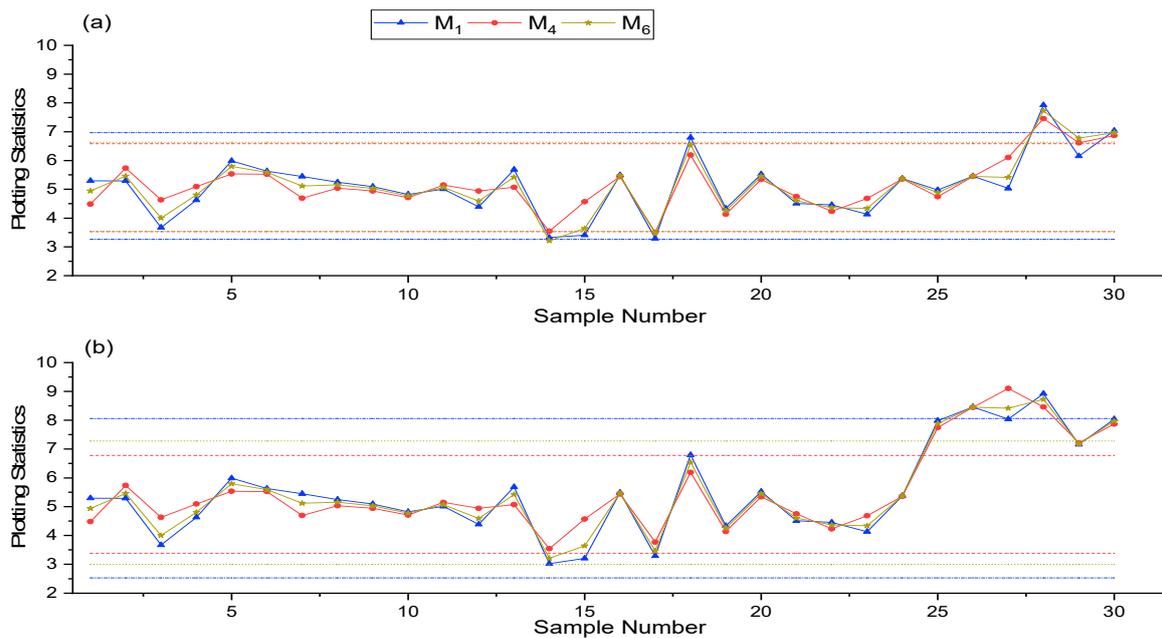


Figure 10. Graphical displays of plotting statistics corresponding to their control limits at (a) $m_1 = 3$ and $\delta = 2.0$; (b) $m_1 = 6$ and $\delta = 3.0$.

These results showed the superiority of the M_6 chart over M_4 and M_1 . Similarly, in the case of $m_1 = 6$ with $\delta = 3$, the M_4 chart detects all six out-of-control shifts at subgroups 25, 26, 27, 28, 29 and 30, whereas the M_6 chart detects five out-of-control shifts at subgroups 25, 26, 27, 28 and 30 and M_1 detects only two out-of-control shifts at subgroups 26 and 28.

Similar findings of a higher detection ability of the auxiliary information-based control charts can be seen in Figure 10. The superiority of auxiliary information-based control charts may also be expected with the other combinations of m_1 at different shifts, which is based on the findings of Section 5.

7. Conclusions

In this paper, we proposed the Phase I analysis of different auxiliary information-based median control charts for location monitoring. The control chart is a very common tool of SPC used to monitor process inconsistencies in quality characteristic(s) of interest. In the context of Phase I analysis, control charts often play a significant role alongside the use of both statistical exploratory (e.g., graphic) and confirmatory processes (e.g., hypothesis testing). They lead to a better understanding of what is really happening over time, to identify and promote the elimination of root(s) of assignable causes. A thorough Phase I analysis is the key part of the overall control and monitoring of the statistical process. Shewhart control charts are especially useful for Phase I analysis as they are simple to create and interpret, and they are efficient at detecting large changes in process parameters. There is no study available in literature which presents the Phase I analysis of location parameters using median charts. Therefore, we considered several median estimators to construct control limits in Phase I. This study examined the selection of an appropriate auxiliary information-based median control chart to monitor the process location parameters effectively in Phase I. We analyzed five auxiliary information-based median charts and compared them with usual median control chart. The efficiency of all six median estimators is evaluated in terms of standardized variance and relative efficiency, while probability to signal measure is used to evaluate the performance of control charts considered in this study. All auxiliary information-based control charts perform well in Phase I analysis. Furthermore, within the comparison of auxiliary information-based control charts, for small values of m_1 , M_6 chart performs well, whereas, M_4 and M_3 control charts showed the best performance at a large value of m_1 . The design structures of M_5 and M_2 also show reasonable performance for some special cases.

The findings of the simulation study were also verified with the real-life industrial example which describes the applications of the proposed charts of this study.

Author Contributions: Study planning, mathematical derivations, calculation of results and draft writing were done by S.H. and S.A.A.; under the supervision of S.M. and M.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research fund was supported by the National Science Foundation of China (No. 71774070).

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

References

1. Montgomery, D.C. *Introduction to Statistical Quality Control*, 7th ed.; John Wiley & Sons: New York, NY, USA, 2012.
2. Shewhart, W.A. *Economic Control of Quality of Manufactured Product*; Van Nostrand, D., Ed.; reprinted by the American Society for Quality Control in 1980; ASQ Quality Press: Milwaukee, WI, USA, 1931.
3. Hu, X.; Castagliola, P.; Zhou, X.; Tang, A. Conditional design of the EWMA median chart with estimated parameters. *Commun. Stat. Theory Methods* **2019**, *48*, 1871–1889. [[CrossRef](#)]
4. Faraz, A.; Woodall, W.H.; Heuchenne, C. Guaranteed conditional performance of the S² control chart with estimated parameters. *Int. J. Prod. Res.* **2015**, *53*, 4405–4413. [[CrossRef](#)]
5. Chakraborti, S.; Human, S.W.; Graham, M. Phase I Statistical Process Control Charts: An Overview and Some Results. *Qual. Eng.* **2008**, *21*, 52–62. [[CrossRef](#)]
6. Jones-Farmer, L.A.; Woodall, W.H.; Steiner, S.H.; Champ, C.W. An Overview of Phase I Analysis for Process Improvement and Monitoring. *J. Qual. Technol.* **2014**, *46*, 265–280. [[CrossRef](#)]
7. Atalay, M.; Testik, M.C.; Duran, S.; Weiß, C.H. Guidelines for automating Phase I of control charts by considering effects on Phase-II performance of individuals control chart. *Qual. Eng.* **2019**, *32*, 223–243. [[CrossRef](#)]
8. Castagliola, P.; Maravelakis, P.; Figueiredo, F. The EWMA median chart with estimated parameters. *IIE Trans.* **2015**, *48*, 66–74. [[CrossRef](#)]
9. Khoo, M.B. A Control Chart Based on Sample Median for the Detection of a Permanent Shift in the Process Mean. *Qual. Eng.* **2005**, *17*, 243–257. [[CrossRef](#)]
10. Sheu, S.-H.; Yang, L. The Generally Weighted Moving Average Median Control Chart. *Qual. Technol. Quant. Manag.* **2006**, *3*, 455–471. [[CrossRef](#)]
11. Tran, K.P.; Castagliola, P.; Balakrishnan, N. On the Performance of Shewhart median Chart in the Presence of Measurement Errors. *Qual. Reliab. Eng. Int.* **2016**, *33*, 1019–1029. [[CrossRef](#)]
12. Hussain, S.; Song, L.; Ahmad, S.; Riaz, M. On Auxiliary Information Based Improved EWMA Median Control Charts. *Sci. Iran.* **2017**, *25*, 954–982. [[CrossRef](#)]
13. Hussain, S.; Song, L.-X.; Ahmad, S.; Riaz, M. A new auxiliary information based cumulative sum median control chart for location monitoring. *Front. Inf. Technol. Electron. Eng.* **2019**, *20*, 554–570. [[CrossRef](#)]
14. Castagliola, P.; Figueiredo, F. The median chart with estimated parameters. *Eur. J. Ind. Eng.* **2013**, *7*, 594. [[CrossRef](#)]
15. Castagliola, P.; Figueiredo, F.O.; Maravelakis, P. The CUSUM median chart for known and estimated parameters. *REVSTAT-Stat. J.* **2019**, *17*, 345–370.
16. Tran, K.P. Run Rules median control charts for monitoring process mean in manufacturing. *Qual. Reliab. Eng. Int.* **2017**, *33*, 2437–2450. [[CrossRef](#)]
17. Teng, H.; Xuelong, H.; Anan, T.; Min, Z. Evaluation of the Run Rules median chart with estimated parameters. In Proceedings of the 2019 Chinese Control and Decision Conference (CCDC), Nanchang, China, 3–5 June 2019; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2019; pp. 2213–2219.
18. Kadilar, G. Özel A New Exponential Type Estimator for the Population Mean in Simple Random Sampling. *J. Mod. Appl. Stat. Methods* **2016**, *15*, 207–214. [[CrossRef](#)]
19. Shabbir, J.; Gupta, S. A generalized class of difference type estimators for population median in survey sampling. *Hacet. J. Math. Stat.* **2017**, *46*, 1015–1028.
20. Asghar, A.; Sanaullah, A.; Hanif, M. Generalized Exponential Type Estimator for Population Variance in Survey Sampling. *Revista Colombiana Estadística* **2014**, *37*, 213–224. [[CrossRef](#)]
21. Zhang, G. Cause-selecting control charts—A new type of quality control charts. *QR J.* **1985**, *12*, 221–225.

22. Riaz, M. Monitoring process mean level using auxiliary information. *Stat. Neerlandica* **2008**, *62*, 458–481. [[CrossRef](#)]
23. Riaz, M. Monitoring process variability using auxiliary information. *Comput. Stat.* **2007**, *23*, 253–276. [[CrossRef](#)]
24. Hussain, S.; Song, L.; Mehmood, R.; Riaz, M. New Dual Auxiliary Information-Based EWMA Control Chart with an Application in Physicochemical Parameters of Ground Water. *Iran. J. Sci. Technol. Trans. A Sci.* **2018**, *43*, 1171–1190. [[CrossRef](#)]
25. Chong, Z.L.; Khoo, M.B.; Lee, M.H.; Haq, A. A variable sample size and sampling interval control chart for monitoring the process mean using auxiliary information. *Qual. Technol. Quant. Manag.* **2018**, *16*, 389–406. [[CrossRef](#)]
26. Abbasi, S.A.; Riaz, M.; Miller, A.; Ahmad, S. On the Performance of Phase I Dispersion Control Charts for Process Monitoring. *Qual. Reliab. Eng. Int.* **2014**, *31*, 1705–1716. [[CrossRef](#)]
27. ISO7870-1. *Control Charts-Part 1: General Guidelines*; ISO: Geneva, Switzerland, 2019.
28. ISO7870-2. *Control Charts-Part 2: Shewhart Control Charts*; ISO: Geneva, Switzerland, 2013.
29. ISO7870-3. *Process Acceptance Control Charts-Part 3*; ISO: Geneva, Switzerland, 2012.
30. ISO7870-4. *Cumulative Sum Charts-Part 4*; ISO: Geneva, Switzerland, 2011.
31. ISO7870-5. *Specialized Control Charts-Part 5*; ISO: Geneva, Switzerland, 2014.
32. ISO7870-6. *EWMA Control Charts-Part 6 (Under Development)*; ISO: Geneva, Switzerland, 2014.
33. ASTM D6299. *Applying Statistical Quality Control Assurance and Control Charting Techniques to Evaluate Analytical Measurement System Performance*; American Society for Testing and Materials: West Conshohocken, PA, USA, 2013.
34. ASTM E2587. *Use of Control Charts in Statistical Process Control*; American Society for Testing and Materials: West Conshohocken, PA, USA, 2012.
35. Kuk, A.Y.C.; Mak, T.K. Median Estimation in the Presence of Auxiliary Information. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1989**, *51*, 261–269. [[CrossRef](#)]
36. Singh, H.P.; Singh, S.; Puertas, S.M. Ratio type estimators for the median of finite populations. *Allg. Stat. Arch.* **2003**, *87*, 369–382.
37. Castagliola, P. An-EWMA control chart for monitoring the process sample median. *Int. J. Reliab. Qual. Saf. Eng.* **2001**, *8*, 123–135. [[CrossRef](#)]
38. Ahmad, S.; Riaz, M.; Abbasi, S.A.; Lin, Z. On efficient median control charting. *J. Chin. Inst. Eng.* **2013**, *37*, 358–375. [[CrossRef](#)]
39. Abbasi, S.A.; Miller, A. D chart: An efficient alternative to monitor process dispersion. In Proceedings of the World Congress on Engineering and Computer Science, San Francisco, CA, USA, 19–21 October 2011.
40. Leys, C.; Ley, C.; Klein, O.; Bernard, P.; Licata, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* **2013**, *49*, 764–766. [[CrossRef](#)]
41. Howell, D.C. *Median Absolute Deviation*; Wiley statsRef, Statistics Reference Online: Hoboken, NJ, USA, 2014.
42. Riaz, M.; Mehmood, R.; Ahmad, S.; Abbasi, S.A. On the Performance of Auxiliary-based Control Charting under Normality and Nonnormality with Estimation Effects. *Qual. Reliab. Eng. Int.* **2012**, *29*, 1165–1179. [[CrossRef](#)]
43. Alwan, L.C. *Statistical Process Analysis*; McGraw-Hill/Irwin: New York, NY, USA, 2000.
44. Riaz, M. An improved control chart structure for process location parameter. *Qual. Reliab. Eng. Int.* **2011**, *27*, 1033–1041. [[CrossRef](#)]
45. Schoonhoven, M.; Nazir, H.Z.; Riaz, M.; Does, R.J. Robust location estimators for the X control chart. *Qual. Control Appl. Stat.* **2013**, *58*, 25–26. [[CrossRef](#)]
46. Rousseeuw, P.J.; Croux, C. Alternatives to the median absolute deviation. *J. Am. Stat. Assoc.* **1993**, *88*, 1273–1283. [[CrossRef](#)]
47. Abbasi, S.A.; Miller, A. On Proper Choice of Variability Control Chart for Normal and Non-normal Processes. *Qual. Reliab. Eng. Int.* **2011**, *28*, 279–296. [[CrossRef](#)]
48. Nazir, H.Z.; Riaz, M.; Does, R.J.M.M.; Abbas, N. Robust CUSUM Control Charting. *Qual. Eng.* **2013**, *25*, 211–224. [[CrossRef](#)]
49. Bickel, P.J.; Lehmann, E.L. Descriptive statistics for nonparametric models. III. Dispersion. In *Selected works of EL Lehmann*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 499–518.

50. Jones, L.A.; Champ, C.W. A Distribution-Free Phase I Control Chart for Subgroup Scale. *J. Qual. Technol.* **2010**, *42*, 373–387. [[CrossRef](#)]
51. Shiau, J.-J.H.; Sun, J.-H. A new strategy for Phase I analysis in SPC. *Qual. Reliab. Eng. Int.* **2009**, *26*, 475–486. [[CrossRef](#)]
52. Yeh, I.-C. Modeling of strength of high-performance concrete using artificial neural networks. *Cem. Concr. Res.* **1998**, *28*, 1797–1808. [[CrossRef](#)]
53. Marlin, T.E. *Process Control: Designing Processes and Control Systems for Dynamic Performance*; McGraw-Hill Higher Education Singapore: Singapore, 2000.
54. Yoon, S.; MacGregor, J.F. Fault diagnosis with multivariate statistical models part I: Using steady state fault signatures. *J. Process. Control.* **2001**, *11*, 387–400. [[CrossRef](#)]
55. Xiangrong, S.; Yan, L.; ZhengShun, F.; Jun, L. A multivariable statistical process monitoring method based on multiscale analysis and principal curves. *Int. J. Innov. Comput. Inf. Control* **2013**, *9*, 1781–1800.



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).