Zanco Journal of Human Sciences (ZJHS)
ISSN (Print):2218-0222
ISSN (Online): 2412-396X

Vol.28 Issue 6 2024
ID No.2296 (PP 343 - 360)
https://doi.org/10.21271/zjhs.28.6.19

Research Article

Data De-Noise for Multivariate T^2 and S-Charts Using Multivariate wavelets

Dler Hussein Kadir/Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University-Erbil, Erbil, Kurdistan Region, Iraq.
Rebaz Othman Yahya/Department of Business Administration, Cihan University-Erbil, Kurdistan Region, Iraq.
Azhin Muhammed Khudhur/Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University-Erbil, Erbil, Kurdistan Region, Iraq.
Taha Hussein Ali/Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University-Erbil, Erbil, Kurdistan Region, Iraq.

CORRESPONDENCE
Azhin Muhammed Khudhur
Azhin.khudhur@su.edu.krd

Reiceved 22/05/2024
Accepted 10/09/2024
Published 15/12/2024

Keywords:
Multivariate charts,
Multivariate wavelet,
De-noise,
Threshold,
Quality control.

Abstract

In this research, proposed multivariate charts were created corresponding to T^2 and S-charts that are robust to noise data by using multivariate wavelet shrinkage, that dealt with the contamination problem before constructing Shewhart charts, through several different wavelets with (Baye), and (SURE) threshold methods, based on the rule of soft thresholding. It is then compared with the classical method proposed by Shewhart based on total variance (trace of the variance matrix), generalized variance (determinant of the variance matrix), and process capability. A MATLAB program designed to obtain the most efficient charts with the least contamination is used to simulate and use real data to get the most efficient charts with the least contamination. Based on the study's conclusions, the proposed charts are more efficient than the classical method in de-noising the data.

About the Journal
Zanco Journal of Humanity Sciences (ZJHS) is an international, multi-disciplinary, peer-reviewed, double-blind and open-access journal that enhances research in all fields of basic and applied sciences through the publication of high-quality articles that describe significant and novel works; and advance knowledge in a diversity of scientific fields. https://zancojournal.su.edu.krd/index.php/JAHS/about

1. Introduction

Quality control charts are developed from a combination of statistical techniques and principles that enable enterprises to oversee and regulate their process quality. One of the most important statistical theories used in quality control charts is variability. Variability in this context refers to dissimilarities between data points within a process. This implies that variation must be reduced while controlling quality to achieve stable and consistent processes. There are two main types of quality control charts:

There are two types of charts: variable charts and attribute charts. Variable graphs, for example, are used with continuous variables such as weight, length, or time whereas attributes show discrete variables like several errors.

However, Quality Control Charts use statistics over time (such as mean, standard deviation, or range) to analyze process data. By connecting the dots with a line, we can see if there is any trend that could indicate a problem with the system (as a principle of QC charts, if more than two consecutive second points are out of the chart range then the product under consideration is out of control).

Some examples include Control Charts, Run Charts or Pareto Charts etcetera. Among these approaches, the Control chart is widely adopted because it helps track whether processes are within set limits.

Statistical Quality Control also covers another important concept called “Control Limits”. These limits act as upper and lower bounds on a quality control chart signifying whether or not it is always under control. Any point outside (more than two consecutive or sequential beyond points) of those limits therefore demands some corrective measure since it may indicate trouble with whatever process is being monitored. Moreover, Statistical Process Control (SPC) represents an additional significant statistical philosophy embedded into quality control graphs where acceptable levels guide SPC about statistical tools usage when monitoring critical product features e.g., through X-bar & R-charts hence enhancing quick discovery of even slightest alterations indicative wrongnesses in a given process.

In summary, statistics play a key role in improvement activities within QCCs for both processes as well products.

2. Methodology

2.1. The Hotelling T^2 Control Chart

Hotelling’s T^2 chart is a multivariate control chart, which facilitates monitoring of the mean and variability of processes having two or more correlated variables by use of statistical control charts. The Hotelling T^2 statistic is used to measure the distance between the process mean vector and either a target value or a reference set. In this case, a change in correlation structure among variables can be detected using the T^2 statistic that is computed based on the process covariance matrix (Hotelling, 1947).

Control limits for the Hotelling T^2 chart are based on F-distribution, which is used to test whether variances between two groups are equal. Control limits depend on the sample size chosen for the test, the number of variables, and the significance level (Santos-Fernández, 2012).

To provide quality control in statistics process control Hotelling T^2 charts are usually combined with other tools like Shewhart charts. These will therefore detect some changes that can’t be seen using any statistics (Fuchs & Kenett, 1998).

In conclusion, the Hotelling T^2 Control Chart is a powerful tool that helps organizations maintain high-quality standards and identifies potential problems before they become big headaches.

2.1.1. The following steps are often followed when constructing a T^2 Control Chart

1. Get multivariate data about the operation being analyzed.

2. Compute both the mean vector and covariance matrix from the data.

3. Calculate each sample's T^2 statistic.

4. Set upper/lower control limits based upon desired significance level/sample size.

5. Record the upper/lower control limits and T^2 in a chart.

6. Watch overtime and take corrective measures if points exceed these limits (Juran & Gryna 1993).

2.1.2. The model of the T^2 Control Chart consists of:

1. Mean vector: A vector that gives the mean value for each variable in multivariate data. Control charts assume that the vector is a known parameter constant.

2. Covariance matrix: Represents how much and how different various quantities are related to each other and themselves. It is also assumed by control charts that the covariance matrix remains constant over time.

3. T^2 Statistic: This chi-square distributed number measures the distance between the sample mean vector and target mean vector, normalized by the covariance matrix of the population. The significance level and sample size dictate the set up of control limits based on T^2 distribution.

4. Control limits are calculated from the T^2 statistic and the desired significance level or α-value used in Minitab. When these points go consecutively beyond those limits then they are considered out-of-control.

5. Sampling plan: Specifies the number of samples to take and the frequency of sampling. Control charts assume independent and identically distributed samples (each sample contains the same variables) (Montgomery, 2019).

2.2. The Hotelling T^2 Control Chart Distribution

Consider that $\frac{x-\mu}{\frac{\delta}{\sqrt{n}}}$ is the sampling distribution of the student's t distribution. This distribution is used for testing hypotheses (Ho:$\mu = 0$) and constructing confidence intervals as a univariate statistic.

The multivariate match of Student's t-distribution is Hotelling's T^2 distribution. Similar to how Student's t works in one dimension, important for creating confidence regions and testing multiple hypotheses.

$x_1, x_2, \dots, x_n \sim N(\mu, \delta^2)$, $\quad \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$\frac{s^2(n-1)}{\sigma^2} \sim x^2_{(n-1)}$

$$T = \frac{(\underline{\underline{x}} - \mu)}{\frac{s}{\sqrt{n}}} = \frac{\underline{x} - \mu}{\frac{\sigma^2}{n} * \frac{\frac{s^2(n-1)}{\sigma^2}}{n-1}} = \frac{z \sim N(0,1)}{x^2_{(n-1)}} \sim t_{n-1} \quad (1)$$

Definition Suppose,

$x \sim N_p(0, I_p)$, $\mu \sim W_p(I_p, n)$

Where W is Wishart distribution x and μ are independent in equation 1, then the quantity $t^2 = n\, x^T \mu^{-1} x$ Hotelling T^2 distribution with parameters (p) and (n), $t^2 \sim T^2(p, n)$.

Proof:

$x \sim N_p(\mu, \Sigma)$, $\mu \sim W_p(\Sigma, n)$

Where Σ Full Rank, then $n(x-\mu)^T \mu^{-1}(x-\mu)$

Proof:

$$\text{Let } y = \Sigma^{-\frac{1}{2}}(x-\mu), \qquad y \sim N_p(0, I) \quad (2)$$

$$\text{Let } z = \Sigma^{-\frac{1}{2}} \mu \Sigma^{-\frac{1}{2}} \sim W_p(I_p, n) \quad (3)$$

Further, let

$$ny^T zy \sim T^2(n, p)$$

$$z^{-1} = (\Sigma^{-\frac{1}{2}} \mu \Sigma^{-\frac{1}{2}})^{-1} = \Sigma^{\frac{1}{2}} \mu^{1-} \Sigma^{\frac{1}{2}} \quad (4)$$

$$ny^T zy = n(x-\mu)^T \Sigma^{-\frac{1}{2}} \Sigma^{\frac{1}{2}} \mu^{-1} \Sigma^{\frac{1}{2}} \Sigma^{-\frac{1}{2}} (x-\mu) \quad (5)$$

Then,

$$t^2 = n(x-\mu)^T \mu^{-1}(x-\mu) \sim T^2(p, n) \quad (6)$$

The upper control limit for a Phase I control chart changes from UCL $\chi^2_{\alpha,p}$

$$UCL_{T^2} = \frac{p(m-1)(n-1)}{mn - m - p + 1} F_{\alpha,p,mn-m-p+1} \quad (7)$$

$$LCL_{T^2} = 0 \quad (8)$$

Where x is a future subgroup mean to be observed, and the control limit changes to:

$$UCL_{T^2} = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \quad (9)$$

$$LCL_{T^2} = 0 \quad (10)$$

Where m is the number of subgroups of size n used in the Phase I study to estimate the process mean vector and covariance matrix (Montgomery, 2019).

3. Shewhart Control Chart (S-Chart)

A Shewhart control chart is a statistical process control tool that tracks and monitors the result of a process over time. As time passes, data points are plotted on the chart along with UCL and LCL (upper and lower control limits) that are calculated from the data. The basis for this chart lies in viewing variation in processes as being due to two causes, common-cause variation, and special-cause variation. It can be said that the process has some deviations when its data lies outside the control limits established by monitoring it at different times. This means there may be an underlying reason for the problem. In many industries where quality management is important, Shewhart charts are widely used to track trends in processes and reduce variations leading to better product or service value (Best & Neuhauser, 2006).

The mean and standard deviation determine UCL and LCL values. Variability in the process is represented by standard deviation whereas the mean represents the center of the process. UCL & LCL are usually set up at three standard deviation units away from the mean or median each/standard deviation multiplied by three above/below mean or median respectively; this implies that 99.7% of observations will fall within these limits if the process is statistically stable.

This graph was developed by Walter A. Shewhart in the 1920s on his notion that fluctuations can be defined into either:

Examination of common cause variation versus special cause variation; Common cause refers to natural variability while specific cause refers to a single event or factor .From the data being observed, Shewhart charts can be classified into two main groups:

1. Variable control chart: These are used to keep an eye on measurements of process variables such as weight, size, temperature, and time. The X-bar and R-charts are the most commonly used variable control charts which show sample means and ranges over time.
2. Attribute control chart: They are designed to monitor whether some attributes or features are present in a process or not e.g. number of defects or percentage of conforming articles. P-Chart is the most common attribute chart used which plots the proportion of conforming units over time (Montgomery, 2019).

4. Control Chart $\bar{X}$ and S

$\bar{X}$ and S control plots are crucial in statistical process control. They monitor the mean and identify deviations from the standardization of the process over time. These charts aid in determining if a process is stable and regulated by monitoring fluctuations in semantic patterns and standard deviations. The $\bar{X}$ chart records the process's primary tendency, whereas the S chart checks its dispersion or variability. They work together to give a holistic perspective of process stability, which is especially important in the manufacturing and quality control industry sectors (Montgomery, 2019).

4.1. $\overline{X}$ – Chart

The X-bar chart is a control chart used to govern a statistical process and monitor its rate over time. It is beneficial for detecting variations in processes that involve repeated measurements, such as those used in manufacturing. The $\bar{X}$ chart detects changes in the overall process and indicates whether it is under control or requires monitoring. It charts the mean monitoring pattern at any moment in time. The Upper Control Limit (UCL) and Lower Control Limit (LCL) are commonly set three standard deviations higher or lower than the entire process. This ensures that 99.7% of data points fall inside these limits assuming the process is in control (Montgomery, 2019).

CL = Central Line ($\bar{\bar{x}}$) (11)

LCL = Lower Control Limit = $(\bar{\bar{x}} - A_3 * \bar{s})$ (12)

UCL = Upper Control Limit = $(\bar{\bar{x}} + A_3 * \bar{s})$ (13)

4.2. S - Chart

An S-chart, also known as a standard deviation chart, is a control chart utilized in statistical control of processes to track how a process changes or disperses over time. in contrast to an R-chart, which measures the range, an S-chart focuses on the process's standard deviation, offering a more accurate evaluation of change, particularly when working with large sample numbers. The S-chart aids in the identification of changes or trends in process change by charting the standard deviation of samples at any given moment in time. If the process is in control, the dots on the S chart should be within the upper and lower control limits, which are usually three standard deviations from the process average (Montgomery, 2019).

CL= Central Line ($\bar{s}$) (14)

LCL = Lower Control Limit = $(B_3 * \bar{s})$ (15)

UCL= Upper Control Limit = $(B_4 * \bar{s})$ (16)

5. Multivariate Wavelet

A multifarious wavelet is a transform that scrutinizes signals or data concerning more than one dimension or variable. This is important in many areas, such as image processing, data compression, and denoising for instance, because of its ability to capture correlations, and interrelationships between different dimensions of your data (Mallat, 1999).

The Functional Magnetic Resonance Imaging (fMRI) data processing technique is one practical application of multivariate wavelets in neuroimaging. These are 3D times series that track variations in blood supply within the brain resulting from neuronal responses. By using multivariate wavelet analyses, it became possible to isolate typical brain activity patterns associated with particular tasks or behaviors. Another illustration is the multivariate financial time series data analysis employing multivariate wavelets. Multivariate wavelet analysis subdivides data into various frequency bands and identifies patterns of correlation and causation between variables.

Multivariate Wavelets have been employed across multiple fields such as image processing and computer vision, geophysics, and finance among others due to their versatility. Their wide application can be seen in some works like image denoising; image compression; feature extraction; and statistical computing (Nason, 2008).

6. Wavelet De-Noising

Wavelet De-noising is a technique for signal processing that is increasingly being used in many fields such as image, audio, and biomedical signal processing. It decomposes the signal into wavelet coefficients using wavelet transforms; dilates or compresses the coefficients to eliminate noise while keeping critical parts of the signal intact; then it finally de-noises them. However, this works by truncated signals inverse transformed using wavelet coefficients. This involves applying the wavelet transform (Donoho and Johnstone, 1995).

Among the other advantages of wavelet denoising is its capability to effectively handle non-Gaussian or non-stationary noise, save sharp edges and components within the sign, and execute denoising computationally effectively. Nevertheless, it has some benefits in choosing the right wavelets and thresholding techniques for particular signals and noises (Mallat 1999).

Wave-late de-noising is applied in different areas which include: medical imaging where it can be applied to De-noise MRI or CT scans; speech/audio processing where recordings may be de-noised; and financial time series analysis where market data can be de-noised completely (Chang et al., 2000).

7. Threshold Bayes and Sure Threshold

Two of the common methods for thresholding wavelet coefficients in wavelet de-noising are Bayesian thresholding and Stein's Unbiased Risk Estimation (SURE).

It is a statistical method that uses a prior model of wavelet coefficients and estimates the posterior distribution over the coefficients given some observed noise signal – then one chooses an appropriate threshold using a probability criterion like the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC). Researchers have discovered that this method performs well when there is no knowledge and information about the amount of noise present earlier (Johnstone & Silverman, 1997).

On the other hand, SURE thresholding is driven by data hence does not need any prior information on noise level. One thus estimates the level of the threshold by minimizing unbiased risk estimate for mean squared error between the de-noised signal and true signal. SURE has been shown to work well either when the noise level is known or when the signal has a low dimensional structure (Donoho & Johnstone, 1995).

In some situations, both Bayesian as well as SURE exceed traditional thresholds such as hard or soft in terms of performance. The choice usually depends on application specifics together with signal-noise characteristics (Cai & Silverman, 2001).

In their study on parameters of simple linear regression Ordinary Least Square (OLS) and mean squares error (Linear Profile Monitoring) quality control charts; Kareem et al., 2019 employed classical ordinary least squares against robust estimators that use M-Estimator with a Bi-Square function. They used MATLAB language program to estimate weights due to the Bi-Square function before comparing suggested charts and robust estimates for three parameters (intercept, slope, and variance (σ^2)) against classical method in the presence of outlier values under normal distribution randomly generated data (Simulation). According to the findings proposed charts along with robust estimates were more accurate efficient connected than the conventional approach where there are outlier values thereby making them better choices for quality control charts.

An original way of building a bivariate F-control was proposed by Ali et al. (2018) for building a bivariate F-control chart to monitor measurable factors and ensure better quality control. This involved three steps which are: testing bivariate normality, constructing a bivariate S-Chart then using the links between distributions of T^2 and F statistics associated with Shewart T^2 -Chart based on data collected from Erbil laboratory on quality attributes Yield Stress and Elongation (Strain) for steel products manufactured by Erbil Steel. They used computer programs such as SPSS, Excel, and MATLAB to create quality control charts for the T^2-Charts and the F-Chart. The authors concluded that the F-Chart is more accurate than the T^2-Chart, despite the number of samples employed in the suggested chart design being bigger. However, the process was found to be out of control since there is one point outside the boundaries of control, implying that the outcome does not satisfy the needed criteria based on the T^2-Charts and the F-Chart.

8. Proposed Charts

The research proposal aims at developing a multivariate quality control chart based on Shewhart's (Hotelling's) method T^2 and determinant of the S-Chart, after dealing with the data contamination problem using wavelet shrinking when Creation of Control Chart ((Phase I).

Suppose we have variable X representing a contaminated data matrix with dimension$(n \times p)$ and using multivariate wavelet such as Daubechies wavelet of order 2 (Db2), Symlet wavelet of order 3 (Sym3), and Biorthogonal wavelet of order 1.3 (Bior1.3), we get the coefficients of the Discrete Wavelet Transformation (DWT) see equation 17 (Ali et al., 2024), as follows:

$$DWT = WX \tag{17}$$

After estimating the level of threshold using Bayes and Sure methods and using the soft threshold rule, we obtain the Modified Discrete Wavelet Transform Coefficients (MDWT) in equation 18:

$$MDWT = WX^* \tag{18}$$

The inverse of the MDWT matrix is taken, and we get the filtered data matrix:

$$X^* = \text{inv(MDWT)} \tag{19}$$

The filtered data matrix in equation 19 is now used in construction wavelet T^2-Chart and wavelet |S|-Chart:

Wavelet T^2-Chart:

The points drawn (see equation 20) on the chart are:

$$T_n^2 = n(\bar{x}^* - \bar{\bar{x}}^*)' S_w^{-1} (\bar{x}^* - \bar{\bar{x}}^*) \tag{20}$$

Where $\bar{x}^*$ is the means vector$(p \times 1)$ for all m samples, $\bar{\bar{x}}^*(p \times 1)$ is the averaged vector over all m samples, and S_w is the covariance matrix $(p \times p)$ for the filtered data matrix. Likewise, Shewhart's chart depicts the upper and lower control limits and indicates the tabular value of F.

9. Wavelet |S|-Chart

The points drawn on the chart are $|S_{wj}|$, $j = 1, 2, \dots, p$. While the target, upper, and lower control limits are:

$$UCL = (|S_w|/b_1)\,(b_1 + 3b_2^{\frac{1}{2}}) \quad (21)$$

$$CL = |S_w| \quad (22)$$

$$LCL = (|S_w|/b_1)\,(b_1 - 3b_2^{1/2}) \quad (23)$$

Where b_1 and b_2 are parameters of Wavelet |S|-Chart

10. Evaluation Criteria

This paper aims to compare the effectiveness of traditional and modern methods in estimating multi-variate mean vector and covariance matrix, using measures of General Variance (GV) and Process Capability (Cp) as prescribed by Montgomery (2019) see equation 24 and equation 25.

$$GV = |S| \quad (24)$$

$$Cp = \frac{(x_i) - min\,(x_i)}{6 \times GV} \quad (25)$$

The lowest value for the generalized variance and the greatest process capability is the best.

11. Application aspect

Our paper examined, through a simulation study, how well or accurately one can estimate the multivariate mean vector and covariance matrix using current versus suggested control limits.

This comparison includes modeling a multivariate quality control chart and applying it to real-world data via generalized variance and process capacity metrics. The investigation was supported by a MATLAB program designed specifically for this analysis (version 2022a).

11.1. Simulation Study

A correlation coefficient matrix was generated for each of the three samples (p = 3) for the multivariate normal distribution:

p = (1, 0.7, 0.8, 0.7 1, 0.9, 0.8, 0.9, 1)

Our process of generating data was comprehensive where we had 125 observations for m = 25 subsamples each having n=510 observations. We also included random contamination drawn from a standard normal distribution to contaminate the generated data. The first simulation experiment (250 observations) was done for three variables using multivariate wavelet (Db2) and sure thresholding method with the soft rule as depicted in the following figure:

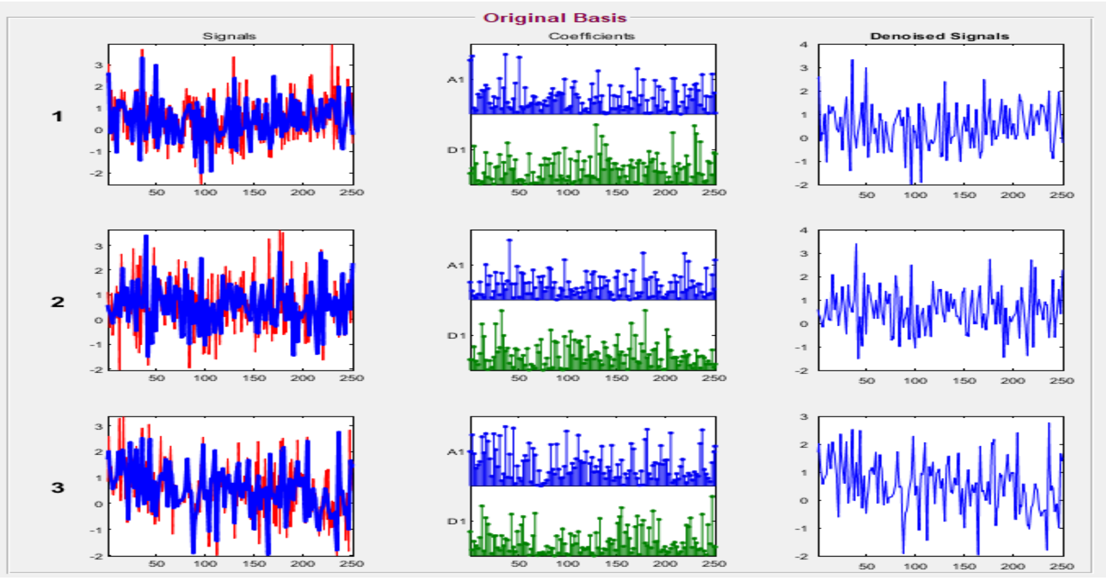

Fig. 1. Evaluation of the data by using multivariate wavelet.

Figure (1) shows the three variables (1, 2, and 3) for generated data (red color) and filtered data (blue color), the second column represents the approximation (A1) and detail (D1) coefficients for the three variables at first level, and the final column represents de-noised data (filtered data).

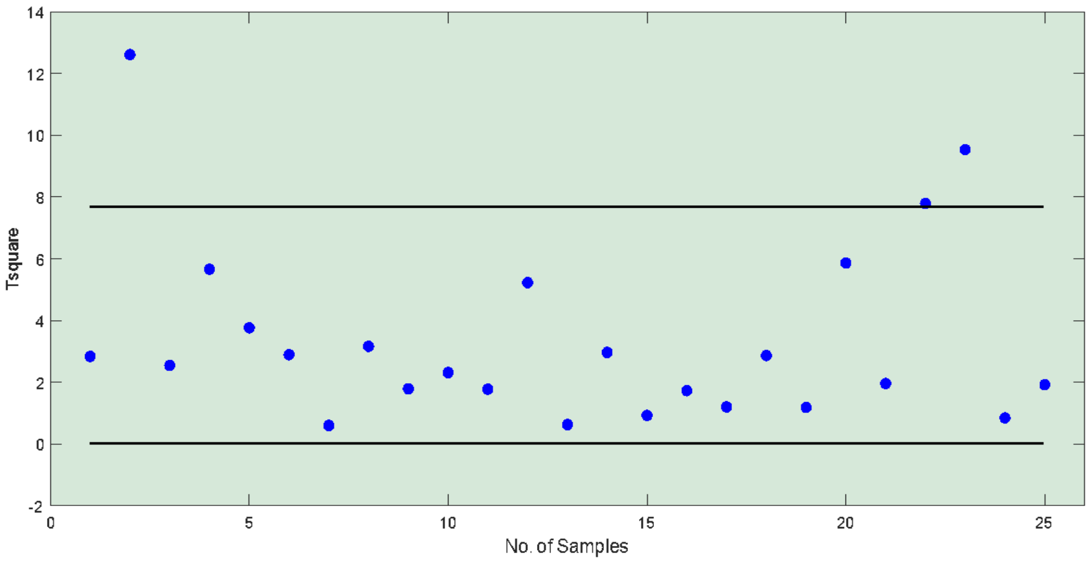

Fig. 2. Classical T^2-Chart.

Control limits are exceeded by three points, as shown in Figure (2).

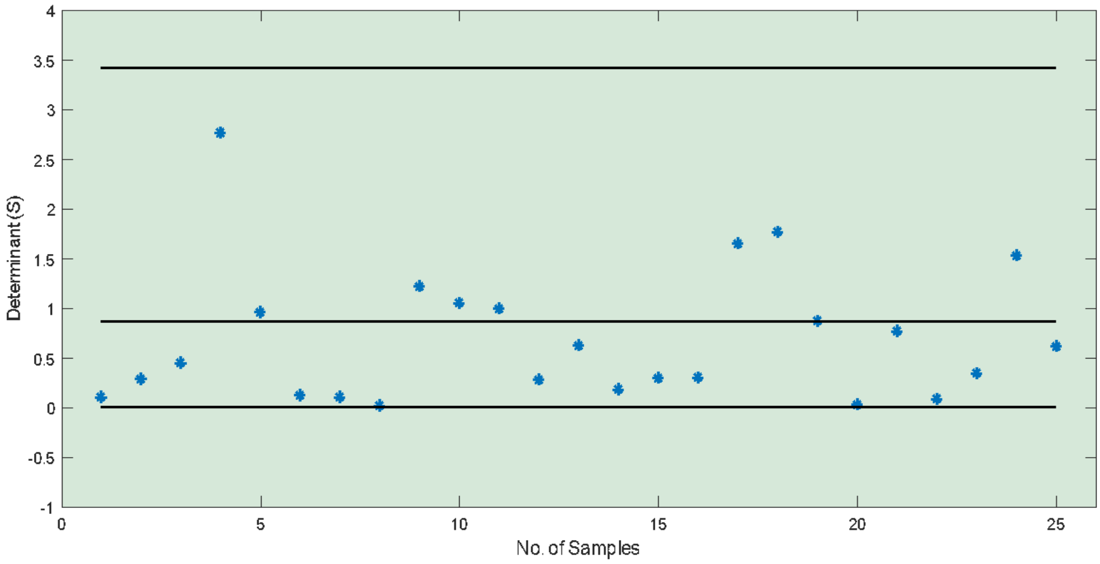

Fig.3. Classical Determinant (S)-Chart.

Figure (3) shows that all points were in control.

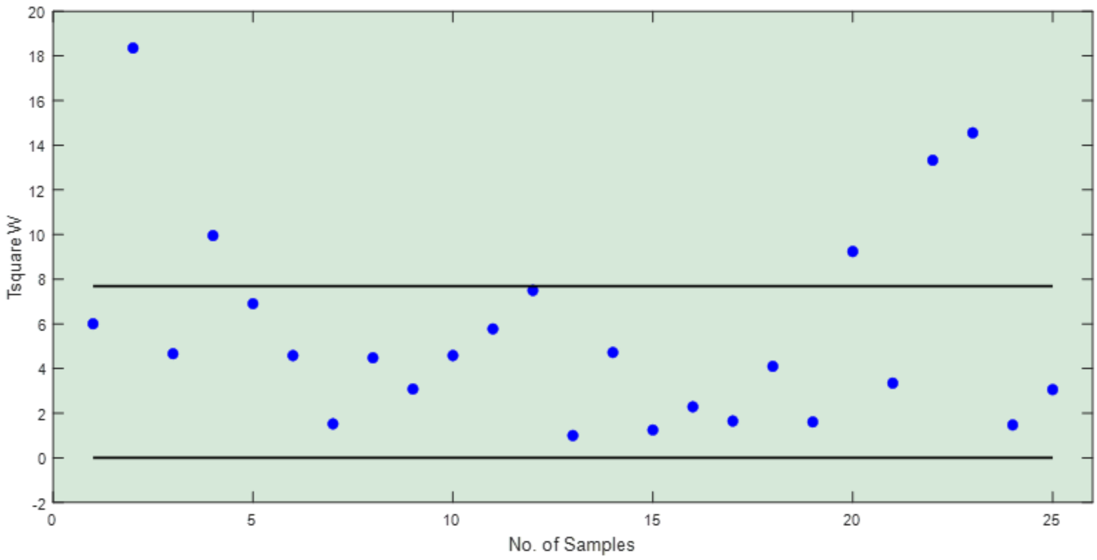

Fig. 4. Wavelet T^2-Chart.

Figure (4) depicts five data points outside of control boundaries. Figure 5 presents the suggested wavelet |S|-Chart (Phase I).

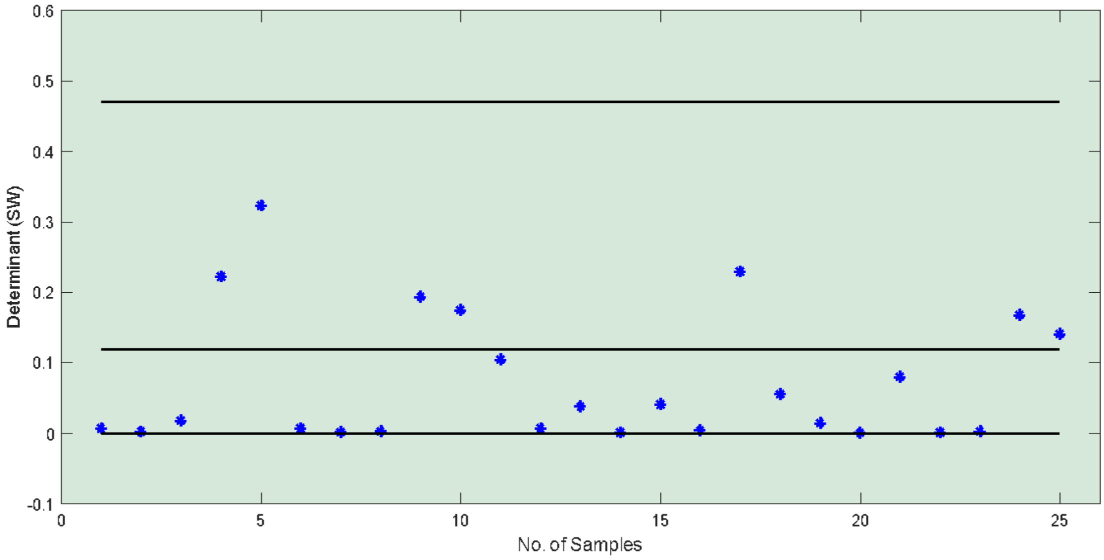

Fig. 5. Wavelet |S|-Chart.

Figure (5) shows that all points are under control hence this is proof enough of the effectiveness of our proposed charts. Thus, the simulation outcomes for classical and new charts are demonstrated in the following table.

Table 1: Effects of the 1st investigation

Chart's Name	Cp	Generalized of Variance	Total of Variance	CL	UCL
Classical T square	0.9286	1.3287	3.3147	------	7.6862
Classical |S|	0.9566	1.2521	3.2523	0.8657	3.4159
Wavelet-T^2	**1.8391**	**0.2375**	**1.8825**	------	7.6862
Wavelet-|S|	**2.1576**	**0.1726**	**1.6912**	0.1193	0.4707

Table (1) indicates that the proposed methods using wavelet T^2 and |S|-charts outperformed the classical method in terms of Cp, general, and overall variance, where the Cp value for proposed charts was greater than the classical charts and (GV and TV) were lower than those on the classical charts because the data has been adjusted to account for contamination factors.

Hence, this trial was meticulously repeated one thousand times when dealing with sample sizes five and ten, which is an expression of our research being comprehensive., The number of total observations (125 and 250), respectively for three variables, and m = 25. Then the average of GV and Cp were calculated for proposed and classical T^2-charts, and Tables 2 and 3 demonstrate the results.

Table 2: For 1000 experiments when n = 5 and m = 25, here is the average result

Chart	Wavelet	Thresholding	GV	Cp
Wavelet1-T^2	Db2	Bayes	**0.1456**	**2.1544**
Wavelet2-T^2	Db2	Sure	0.1781	2.0103
Wavelet3-T^2	Sym3	Bayes	0.1459	2.0854
Wavelet4-T^2	Sym3	Sure	0.1771	1.9598
Wavelet5-T^2	Bior1.3	Bayes	0.1639	1.7558
Wavelet6-T^2	Bior1.3	Sure	0.1997	1.7383
Classical T square	-------	--------	1.2217	0.9384

Table 3: For 1000 experiments when n = 10 and m = 25, here is the average result

Charts	Wavelet	Thresholding	GV	Cp
Wavelet1-T^2	Db2	Bayes	**0.1528**	**2.2429**
Wavelet2-T^2	Db2	Sure	0.1722	2.1485
Wavelet3-T^2	Sym3	Bayes	0.1554	2.1625
Wavelet4-T^2	Sym3	Sure	0.1762	2.0746
Wavelet5-T^2	Bior1.3	Bayes	0.1660	1.8295
Wavelet6-T^2	Bior1.3	Sure	0.1840	1.8126
Classical T square	-------	--------	1.2391	0.9883

All customary charts performed worse than each recommended one. In both cases of n = 5 and n=10. As an illustration, when proposed charts (Wavelet-T^2) were used, their average variances were lower than classical charts' (1.2217 and 1.2391 correspondingly). On the other hand, the average Cp for the proposed charts was larger than that of classical ones (0.9384 and 0.9883). Only one among all these suggested charts – Wavelet1-T^2 – proved to be better than any other chart suggested here besides itself. On top of it all, the Bayes threshold method worked better than the Sure method. GV had the best value at a sample size of 5 with 125 observations. However, at a sample size of 10 with 250 observations, Cp performed best.

The average of GV and Cp were calculated for proposed and classical |S|-charts, and Tables 4 and 5 illustrate the results.

Table 4: The overall outcomes of 1000 investigations with m=25 and n = 5

Charts	Wavelet	Thresholding	Total Variance	Cp
Wavelet1-|S|	Db2	Bayes	**0.0692**	**3.1785**
Wavelet2-|S|	Db2	Sure	0.0940	2.8339
Wavelet3-|S|	Sym3	Bayes	0.0693	3.0853
Wavelet4-|S|	Sym3	Sure	0.0932	2.7703
Wavelet5-|S|	Bior1.3	Bayes	0.0842	2.4923
Wavelet6-|S|	Bior1.3	Sure	0.1125	2.3644
Classical T square	-------	--------	1.2090	0.9482

Table 5: The overall outcomes of 1000 investigations with m=25 and n = 10

Chart	Wavelet	Thresholding	Total Variance	Cp
Wavelet1-|S|	Db2	Bayes	**0.1097**	**2.6564**
Wavelet2-|S|	Db2	Sure	0.1270	2.5130
Wavelet3-|S|	Sym3	Bayes	0.1138	2.5343
Wavelet4-|S|	Sym3	Sure	0.1325	2.4017
Wavelet5-|S|	Bior1.3	Bayes	0.1184	2.1747
Wavelet6-|S|	Bior1.3	Sure	0.1347	2.1643
Classical T square	-------	--------	1.2359	0.9905

When n=5 and n=10, the proposed charts outperformed the classical charts. This was due to the lower average common spread of the suggested charts (wavelet-|S|) compared to the classical charts (1.2090 and 1.2359). The mean Cp for recommended graphs consistently exceeded those of traditional ones (0.9482 and 0.9905). Among all the alternatives considered, only one proved to be better Wavelet1-|S|. Bayes threshold was better than the Sure methods.

At a sample size of (5) with (125) observations, the value of GV was the best, while at a sample size of 10 with (250) observations.

11.2. Real Data

Ali et al. (2018) used to obtain the actual data, which represents quality properties of Yield Stress and Elongation for steel products from the Erbil Steel manufactory. Number of total observations (165) for the number of subsamples (m = 33) and n = 5 observations for all subsamples for two variables. On this basis, the classic charts in Figures (6) and (7) were composed:

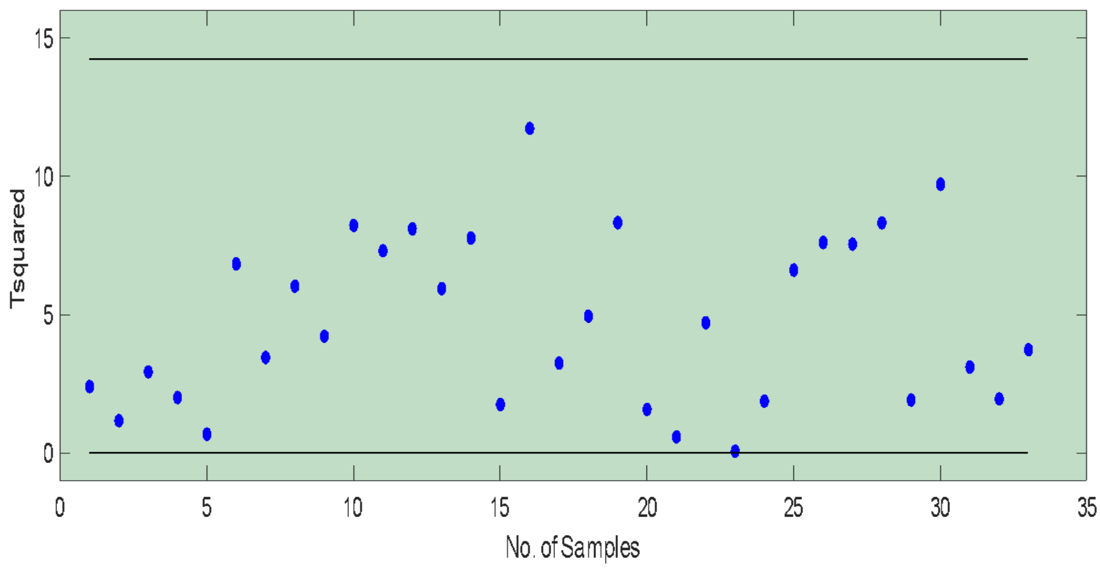

Fig. 6. Classical T2-Chart for real data.

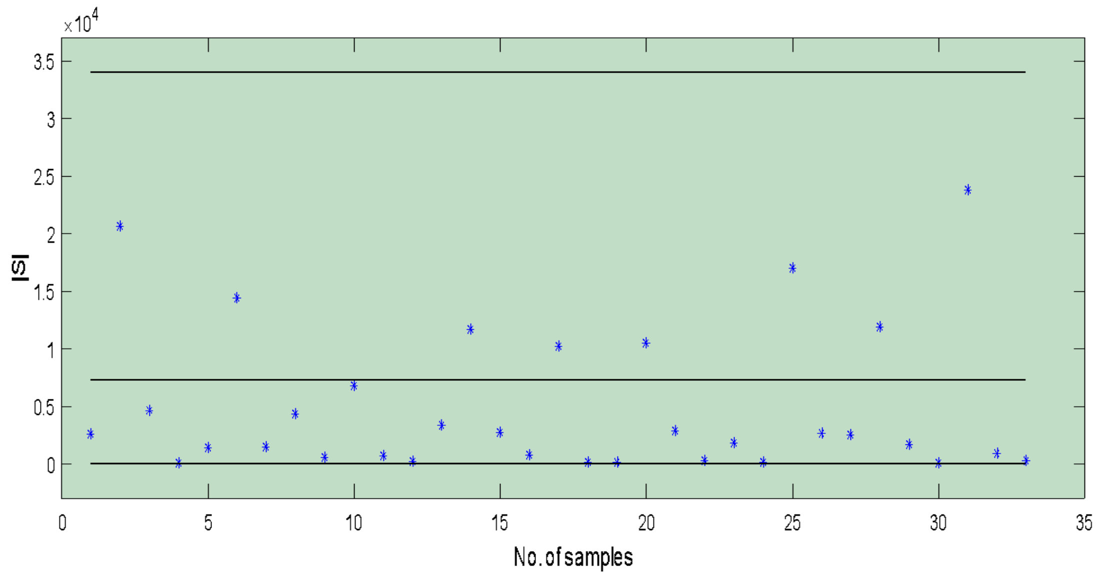

Fig. 7. Classical Determinant (S)-Chart for real data.

Figures (6) and (7) reveal that all points on both graphs are within the authority boundaries so that they can be dealt with in the stage (Phase-I) and used for control and monitoring in the future (Phase-II). By using multivariate wavelet (Db2) and Bayes Thresholding method with soft rule, they are shown in figure (8):

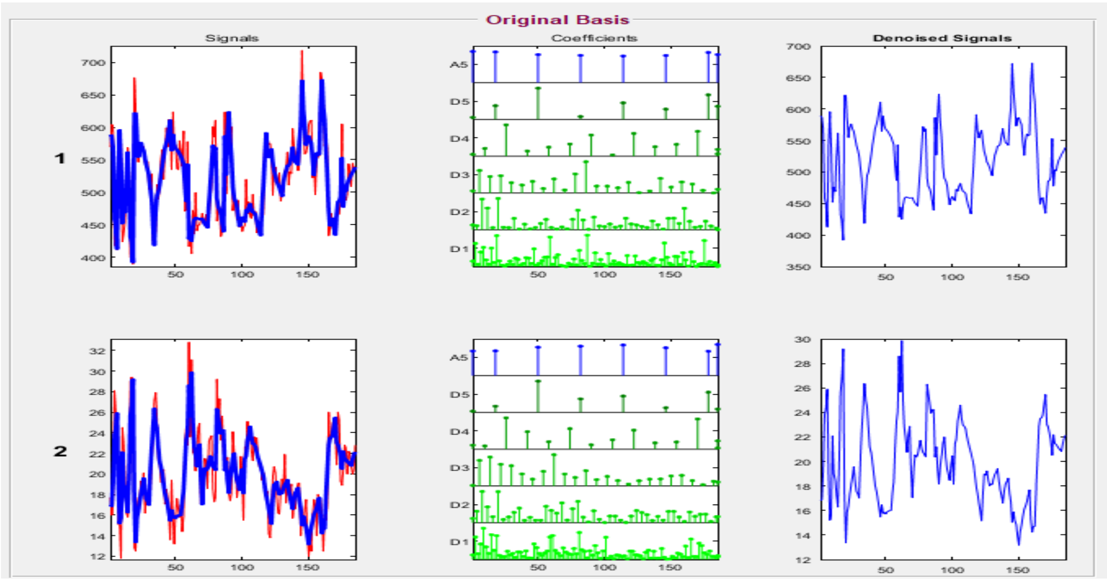

Fig. 8. Analysis of the data by using multivariate wavelet.

Figure (8) shows the two variables (1, and 2) for real data (red color) and filtered data (blue color), the second column represents the approximation (A5) and details (D1-D5) coefficients for the two variables at the fifth level, and the final column represents de-noised data (filtered data). The proposed T^2 and |S|-Charts (Phase-I) are configured as in figures (9) and (10).

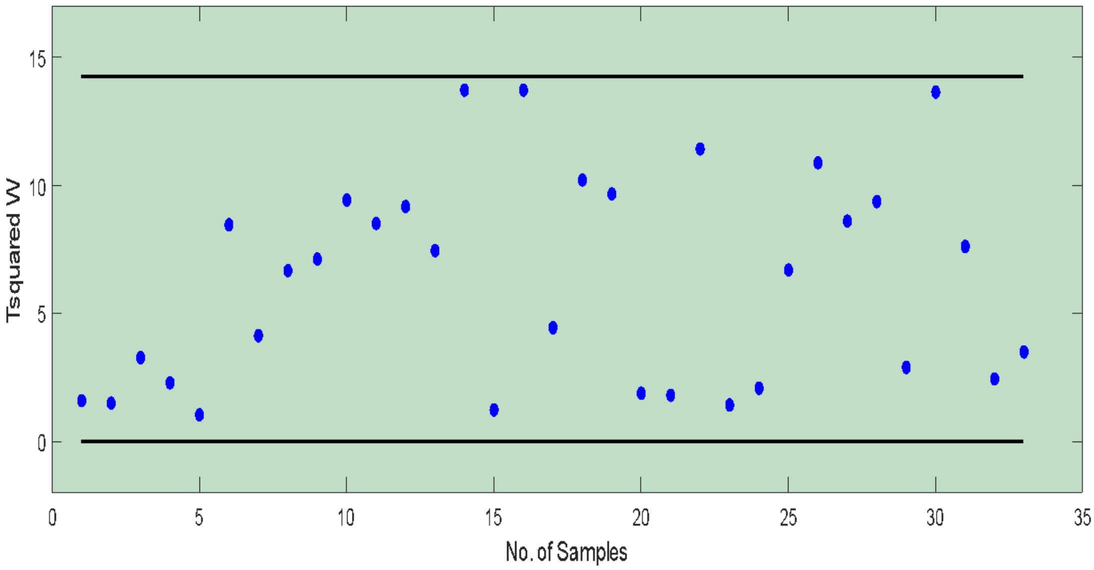

Fig. 9. Wavelet T^2-Chart for real data.

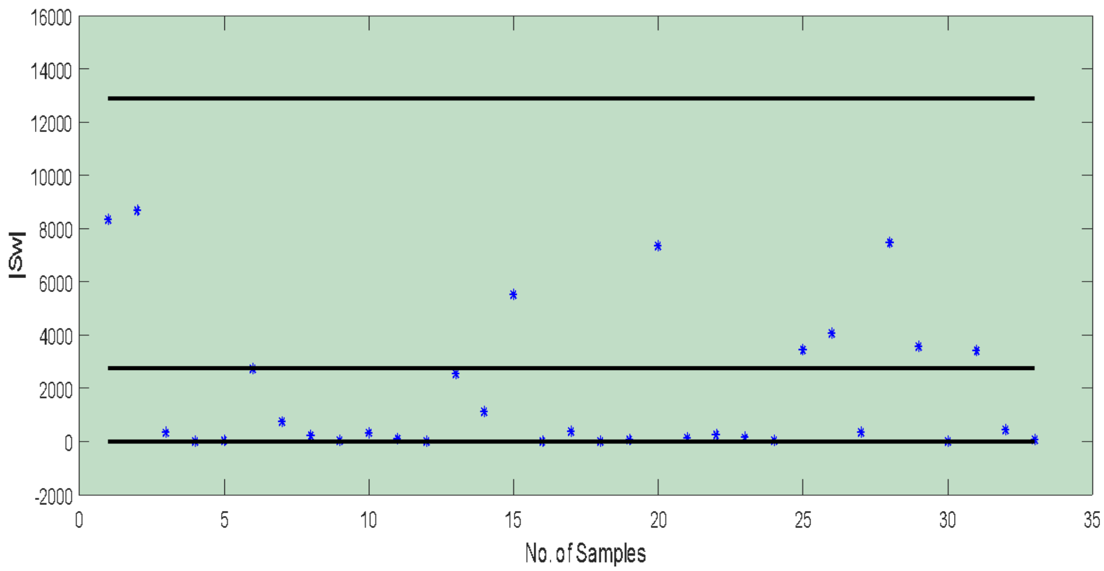

Fig. 10. Wavelet |S|-Chart for real data.

The outcomes of the actual data for both standard and suggested charts are outlined in Table 6:

Table 6: Effects of actual data

Charts	Cp	Generalized of Variance	Total of Variance	CL	UCL
Classical T square	0.7686	23477	3948.7	------	14.2366
Classical |S|	1.1958	9699.4	2217.7	7274.5	34003
Wavelet-T^2	**1.0680**	**11056**	**3223.7**	------	14.2366
Wavelet-|S|	**1.8720**	**3682.0**	**1587.2**	2761.5	12908

According to Table (6), the proposed methods were more successful than the classical approach in terms of Cp, general, and total variance. It was shown that the Cp values for proposed charts were higher than those for classical charts, while GV and TV were lower due to removing contaminants.

12 Conclusion & Recommendations

After a comprehensive investigation of both simulated and real data, we have arrived at the following key discoveries and recommendations:

12.1. Conclusions

Depending on Generalized Variance and Cp criteria from Table 2 to 5 we conclude some highlighted conclusions from 1 to 4.

1. Every proposal made under the wavelet theory has surpassed what was previously considered the best classical T^2 statistic chart.
2. The proposed chart (Db2 wavelet with Bayes thresholding) was better than the other proposed charts, for T^2 and |S|-charts. Depend on Generalized Variance and Cp criteria from Table 2 to 5.
3. Bayes threshold was better than the Sure methods.

4. At a small sample size, the value of GV was the best, while at a big sample size, Cp was the best.

5. For actual data, the new charts are relied upon and are employed to control and observe the iron product in the Erbil factory.

12.2. Recommendations

1. Using the proposed chart (phase I) when contamination is present in the construction of the T^2 and |S|-Charts (based on figure 8.), and especially the proposed chart (Wavelet1)
2. Studying the use of the EUMMA-Chart in construction from a prospective perspective.
3. A prospective study will examine charts by using multivariate wavelet analysis with robust methods to construct T^2-Charts and |S|.

References

-ALI, T. H., RAHIM, A. G. & SALEH, D. M. 2018. Construction of Bivariate F-Control Chart with Application. *Construction of Bivariate F-Control Chart with Application*, 116-133.

-Ali, T.H., Sedeeq, B.S., Saleh, D.M. and Rahim, A.G., 2024. Robust multivariate quality control charts for enhanced variability monitoring. Quality and Reliability Engineering International, 40(3), pp.1369-1381.

-BEST, M. & NEUHAUSER, D. 2006. Walter A Shewhart, 1924, and the Hawthorne factory. *BMJ quality & safety,* 15**,** 142-143.

-CAI, T. T. & SILVERMAN, B. W. 2001. Incorporating information on neighbouring coefficients into wavelet estimation. *Sankhyā: The Indian Journal of Statistics, Series B***,** 127-148.

-CHANG, S. G., YU, B. & VETTERLI, M. 2000. Adaptive wavelet thresholding for image denoising and compression. *IEEE transactions on image processing,* 9**,** 1532-1546.

-DONOHO, D. L. & JOHNSTONE, I. M. 1995. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the american statistical association,* 90**,** 1200-1224.

-FUCHS, C. & KENETT, R. S. 1998. *Multivariate quality control: theory and applications,* New York, Chapman and Hall/CRC.

-HOTELLING, H. 1947. Multivariate quality control-illustrated by the air testing of sample bombsights. *Techniques of statistical analysis*.

-JOHNSTONE, I. M. & SILVERMAN, B. W. 1997. Wavelet threshold estimators for data with correlated noise. *Journal of the royal statistical society: series B (statistical methodology),* 59**,** 319-351.

-JURAN, J. M. & GRYNA, F. M. 1993. Quality Planning and Analysis: From Product Development through Use. . discusses T^2 control charts and provides a description of the steps for constructing them. McGraw-Hill.

-KAREEM, N. S., MOHAMMAD, A. S. & ALZUBAYDI, T. H. A. 2019. Construction robust simple linear regression profile Monitoring. *journal of kirkuk University for Administrative and Economic Sciences***,** 242-257.

-MALLAT, S. 1999. *A wavelet tour of signal processing,* Elsevier.

-MONTGOMERY, D. C. 2019. *Introduction to statistical quality control*, John wiley & sons.

-NASON, G. P. 2008. *Wavelet methods in statistics with R*, Springer.

-SANTOS-FERNÁNDEZ, E. 2012. *Multivariate statistical quality control using R,* Springer Science & Business Media.

کەمکردنەوەی پیسبوونی داتا بۆ هێڵکارییەکانی فرەگۆڕاوەکان T^2 و S بە بەکارهێنانی شەپۆلی فرەگۆڕاوەکان

دلێر حوسێن قادر
بەشی ئامار و زانیارییەکان،کۆلێژی بەڕێوەبردن و ئابووری، زانکۆی سەلاحەددین-هەولێر،هەولێر،هەرێمی کوردستان،عێراق
dler.kadir@su.edu.krd

ڕێباز عوسمان یەحیا
بەشی کارگێڕی کار، زانکۆی جیهان – هەولێر
rebaz.othman@cihanuniversity.edu.iq

ئەژین محمد خضر
بەشی ئامار و زانیارییەکان،کۆلێژی بەڕێوەبردن و ئابووری، زانکۆی سەلاحەددین-هەولێر، هەولێر،هەرێمی کوردستان،عێراق
azhin.khudhur@su.edu.krd

تەها حوسێن عەلی
بەشی ئامار و زانیارییەکان،کۆلێژی بەڕێوەبردن و ئابووری، زانکۆی سەلاحەددین-هەولێر، هەولێر،هەرێمی کوردستان،عێراق
taha.ali@su.edu.krd

پوختە

لەم توێژینەوەیەدا پێشنیاری پێکهێنانی هێڵکاری نوێ کراوە بۆ فرەگۆڕاوەکان کە گونجاوە لەگەڵ هێڵکارییەکانی T^2 و S کە بەرگرییان لە پیسبوونی داتا هەیە بە بەکارهێنانی کەمبوونەوەی شەپۆلی کە چارەسەری کێشەی پیسبوونی داتا دەکات هێڵکاریەکانی شیوارت لە ڕێگەی چەندین شەپۆلی جیاواز لەگەڵ ڕێگاکانی بڕینی سنوور (Bayes) و (SURE) ، و لەسەر بنچینەی ڕێسای بڕینی سنووری نەرم. پاشان بەراوردکردنی ڕێگای پێشنیارکراو و ڕێگای ئاسایی بۆ شیوارت لەسەر بنەمای کۆی جیاوازی (کۆی توخمەکانی تیرەی سەرەکی ماتریکسی جیاواز)، جیاوازی گشتی (دیاریکردنی ماتریکسی جیاواز) و توانای کرداری بۆ بەدەستهێنانی کاراترین هێڵکارییەکان بە پیسبوونی کەمتر لە ڕێگەی داتای خەمڵێنراو و و داتای ڕاستەقینە و بەکارهێنانی بەرنامەی ماتلاب کە بۆ ئەم مەبەستە دیزاین کراوە. لێکۆڵینەوەکە دەریخستووە کە هێڵکاریە پێشنیارکراوەکان کێشەی پیسبوونی داتا چارەسەر دەکەن و کاراترن لە ڕێگای ئاسایی.

ووشە سەرەکییەکان: هێڵکاری فرەگۆڕ، شەپۆلی فرەگۆڕ، کەمکردنەوەی پیسبوون، سنوور.

تقليل تلوث البيانات للوحات T^2 وS متعدد المتغيرات باستخدام المويجات متعدد المتغيرات

دلير حسين قادر
قسم الإحصاء والمعلوماتية، كلية الإدارة و الإقتصاد، جامعة صلاح الدين – أربيل- أربيل،اقليم كوردستان،العراق
dler.kadir@su.edu.krd

ريباز عثمان يحيى
قسم إدارة الأعمال، جامعة جيهان – أربيل
rebaz.othman@cihanuniversity.edu.iq

أزين محمد خضر
قسم الإحصاء والمعلوماتية، كلية الإدارة و الإقتصاد، جامعة صلاح الدين – أربيل، أربيل،اقليم كوردستان،العراق
azhin.khudhur@su.edu.krd

طه حسين علي
قسم الإحصاء والمعلوماتية، كلية الإدارة و الإقتصاد، جامعة صلاح الدين – أربيل، أربيل،اقليم كوردستان،العراق
taha.ali@su.edu.krd

ملخص

تم في هذا البحث اقتراح تكوين لوحات جديدة لمتعدد المتغيرات مقابلة للوحاتT^2 و S حصينة ضد تلوث البيانات باستخدام التقليص المويجي الذي يعالج مشكلة تلوث البيانات قبل بناء لوحات شيوارت وذلك من خلال عدة مويجات مختلفة مع طرائق قطع العتبة (Bayes) و(SURE) ، وعلى أساس قاعدة قطع العتبة الناعمة. ثم المقارنة بين الطريقة المقترحة والتقليدية للباحث شيوارت اعتمادا على التباين الكلي (مجموع عناصر القطر الرئيسي لمصفوفة التباين)، التباين العام (محدد مصفوفة التباين) والقدرة العملية للحصول على أكفأ لوحات ذات تلوث أقل من خلال المحاكاة والبيانات الحقيقية وباستخدام برنامج بلغة ماتلاب مصمم لهذا الغرض. وتوصلت الدراسة إلى أن اللوحات المقترحة عالجت مشكلة تلوث البيانات وذات كفاءة عالية أكثر من الطريقة التقليدية.

الكلمات المفتاحية: الرسوم البيانية متعددة المتغيرات، المويجات متعدد المتغيرات ، تقليل تلوث ، العتبة.