

## Research Article

# Identification of the Out-of-Control Variable on a Multivariate Control Chart Signals in Iron Sheet Production Process

Martin Kyalo Mung'au <sup>\*</sup> , Hellen Wanjiru Waititu ,  
Nyakundi Omwando Cornelious 

Department of Mathematics and Actuarial Science, Catholic University of East Africa, Nairobi, Kenya

## Abstract

Monitoring and diagnosis of an out of control signals in manufacturing processes has become more challenging when it involves two or more correlated variables. In any production process and in the manufacturing industry, regardless of how well designed or carefully maintained it is, a certain amount of inherent or natural variability will always exist. The natural variability is the cumulative effect of many small, essentially unavoidable causes. A process that is operating with only chance causes of variation present is said to be in statistical control. In this paper we propose a method for approaching this problem based on principal components analysis. The principal component analysis method based on the ratios form was used to investigate which of the variable(s) were responsible for the out-of-control signal. A display of matrix of scatter plots was applied to investigate the correlation between the process variables and ellipses were used an approximate control limits. The principal component scores were used as an empirical reference distribution to establish a control region for the process to detect the variables causing the out-of-control signal. These insights derived from the principal component analysis offered valuable cues for identifying potential signals of an out-of-control process. The study findings indicated distinctive patterns of variance within the data, shedding light on potential signals for an out-of-control process.

## Keywords

Process Control, Identification, Out-of-Control Signals, Principal Components Analysis

## 1. Introduction

Multivariate statistical process control (MSPC) techniques have long been integral to industry for evaluating process stability and final product quality. These techniques are crucial in monitoring and maintaining the desired quality levels in manufacturing processes, which are inherently subject to some degree of variability. This variability is acceptable when it is small and caused by random, unidentified factors,

a condition defined by [1] as a statistically in-control process. However, excessive variability introduced by systematic factors renders a process out-of-control, necessitating rigorous statistical process monitoring (SPM) to detect and address these anomalies.

The research by [10] found that Control charts, the cornerstone of SPM provide a graphical representation of sample

\*Corresponding author: martinmungau@gmail.com (Martin Kyalo Mung'au)

Received: 24 July 2024; Accepted: 15 August 2024; Published: 11 September 2024



Copyright: © The Author(s), 2024. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

values related to product quality over time or across sample numbers. These charts are essential for visualizing and analyzing the process behavior, identifying variations that exceed normal levels. According to [7] variability in key quality characteristics typically originates from three primary sources: improperly adjusted or controlled machinery, operator errors, or defective raw materials. Such variability, larger than the background noise, signifies an unacceptable level of process performance and is classified as assignable causes of variation. [9] when a process operates in the presence of assignable causes, it is deemed out-of-control, necessitating immediate corrective actions to restore stability.

In recent years, the study by [11] recognized that the focus of research has shifted towards identifying the specific out-of-control variables after a multivariate control chart signals an anomaly. Traditional univariate methods fall short in complex industrial settings where multiple interrelated quality characteristics need monitoring simultaneously. The study by [12] proposed leveraging Principal Component Analysis (PCA) to identify the specific variables responsible for out-of-control signals in multivariate settings. CA is a powerful statistical tool that reduces data dimensionality while retaining most of the variation present in the dataset, making it easier to detect patterns and pinpoint sources of variation as outlined by [14].

Pioneering work by [3] introduced a PCA-based method for identifying out-of-control variables in multivariate control charts. This method includes deriving theoretical control limits, step-by-step procedural instructions, and a comprehensive evaluation of its strengths and weaknesses. It also offers a graphical technique to address limiting situations and demonstrates superior performance compared to existing PCA methods, largely due to its reduced computational burden.

Research by [2] further emphasized the importance of checking for autocorrelation before applying control charts, as correlated data points can lead to misleading results, such as increased false alarm rates and masked actual correlations between variables. Addressing autocorrelation is crucial for accurate multivariate control chart analysis, ensuring that the identified signals accurately reflect process deviations.

In the context of the Tumaini Mabati Company's iron sheet production process, a high defect rate has been observed, including issues like uneven thickness, surface imperfections, and inadequate tensile strength. These defects lead to material waste, higher production costs, customer complaints, and potential financial losses. This study aims to employ PCA to identify the variable(s) contributing to the out-of-control signals in the company's production process. By pinpointing the exact sources of variability, the company can implement targeted interventions to enhance process stability and product quality, ultimately reducing defect rates and improving overall operational efficiency.

## 2. Methods

### 2.1. Data and Simulation Procedure

A secondary data set from Tumaini Mabati manufacturing company, focusing on box profile iron sheets was used for this analysis. The goal was to determine if the sheets met quality standards.

The principal components analysis (PCA) procedure was demonstrated using data on four process variables  $x_1, x_2, x_3$  and  $x_4$  from a manufacturing process. The first 20 observations from the manufacturing process were plotted against each other in a pairwise manner to obtain a display known as a matrix of scatter plots. The PCA on 20 observations were predicted showing the eigenvalues and Eigen vectors. A scatter plot of the first 20 principal components scores  $Z_{i1}$  and  $Z_{i2}$  were also plotted. Future 10 new observations on the process variables  $x_1, x_2, \dots, x_p$  were obtained and their principal component scores computed and plotted on a control chart. The principal component scores for the new observations were plotted in a different control chart which was used to assist in identifying the scores of the new points.

### 2.2. PCA for Dimensionality Reduction

According to [8], a  $p \times p$  symmetric, nonsingular matrix, such as the variance-covariance matrix  $\Sigma$ , is reduced to a diagonal matrix  $L$  by premultiplying and postmultiplying it by a particular orthonormal matrix  $U$  such that  $U^T \Sigma U = L$ , the diagonal elements of  $L$ ,  $l_1, l_2, \dots, l_p$  are the characteristic roots, or eigenvalues of  $\Sigma$ . The columns of  $U$  are the characteristic vectors, or eigenvectors of  $\Sigma$ . The study by [13] on PCA transforms  $p$  correlated variables  $x_1, x_2, \dots, x_p$  into  $p$  new uncorrelated ones. The main advantage of this method is the reduction of dimensionality. The first few principal components (PCs) typically capture most of the process variability as stated by [6] allowing us to focus on these instead of all original variables.

### 2.3. Proposed Method

Assuming that  $x_i = (x_{1i}, x_{2i}, \dots, x_{pi})^T$  denote the observation (vector)  $i$  for the  $p$  variables of a process. Let  $x_u$  be a  $p$ -dimensional normal distribution  $N_p(\mu_0, \Sigma_0)$ , where  $\mu_0$  is the vector ( $p \times 1$ ) of known means and  $\Sigma_0$  is the known ( $p \times p$ ) variance-covariance matrix.

The PCA model applied is defined as

$$Z_k = \mu_{1k}X_1 + \mu_{2k}X_2 + \mu_{3k}X_3 + \dots + \mu_{pk}X_p$$

Where  $Z_k$  is the  $k$  PC,  $(\mu_{1k}, \mu_{2k}, \dots, \mu_{pk})^T$  is the corresponding  $k$  eigenvector and  $X_1, X_2, \dots, X_p$  are the process variables.

The score for vector  $x_i$  in PC  $k$  is

$$Y_{ki} = \mu_{1k}x_{1i} + \mu_{2k}x_{2i} + \dots + \mu_{pk}x_{pi}$$

When an out-of-control signal in the control chart is detected, the study used principal component analysis to identify the variable or variables that were responsible. For this objective one methodology was developed for the case that the covariance matrix has only positive correlations.

## 2.4. Covariance Matrix with Positive Correlations

In the case where all the variables are positively correlated the first two PCs are weighted average of all the variables. According to [4, 5], one of the existing methods for choosing principal components was to choose  $d \leq p$  significant PCs. The method for the study is based on the ratios of the form:

$$r_{ki} = \frac{(\mu_{k1} + \mu_{k2} + \dots + \mu_{kd})x_{ki}}{Y_{1i} + Y_{2i} + \dots + Y_{di}}$$

Where  $x_{ki}$  is the  $i^{th}$  value of variable  $X_k$ ,  $Y_{ji}$ ,  $j = 1, 2, \dots, d$  is the score of the  $i^{th}$  vector of observations in the  $j^{th}$  PC.

In the ratio, the numerator corresponds to the sum of the contributions of variable  $X_k$  in the first  $d$  PCs in observation (vector)  $d$  whereas the denominator is the sum of scores of observation (vector)  $i$  in the first  $d$  PCs. The rationale of this method is to compute the impact of each of the  $p$  variables on the out-of-control signal by using its contribution to the total score. A multivariate chart is used when there is at least moderate and usually large correlation between the variables

as outlined by [15]. Under such circumstances the first  $d$  PCs account for the largest part of the process variability.

## 3. Result and Discussion

### 3.1. Residuals Versus the Process Variables

The findings reveal a normal quantile-quantile (QQ) plot of the residuals versus the process variable. The data points roughly follow a straight line, indicating that the residuals are normally distributed, which is a good assumption for normality statistical tests as shown in figure 1.

### 3.2. Residuals Versus Order of Data

This scatter plot showed the residuals plotted against the order in which the data were collected. Ideally, there should be no pattern in this plot, indicating that the order in which the data were collected did not affect the results. The plot seems to show some randomness, which a good sign of the data analysis is as indicated in figure 1.

### 3.3. Residuals Versus Fitted Values

This scatter plot displays the residuals versus the fitted values of the response variable. In an ideal scenario, the residuals should be randomly scattered around the horizontal line at zero, with no discernible pattern. This plot seems to show a nearly straight line, suggesting that the model perfectly captures the relationship between the factors and the response as shown in figure 1.

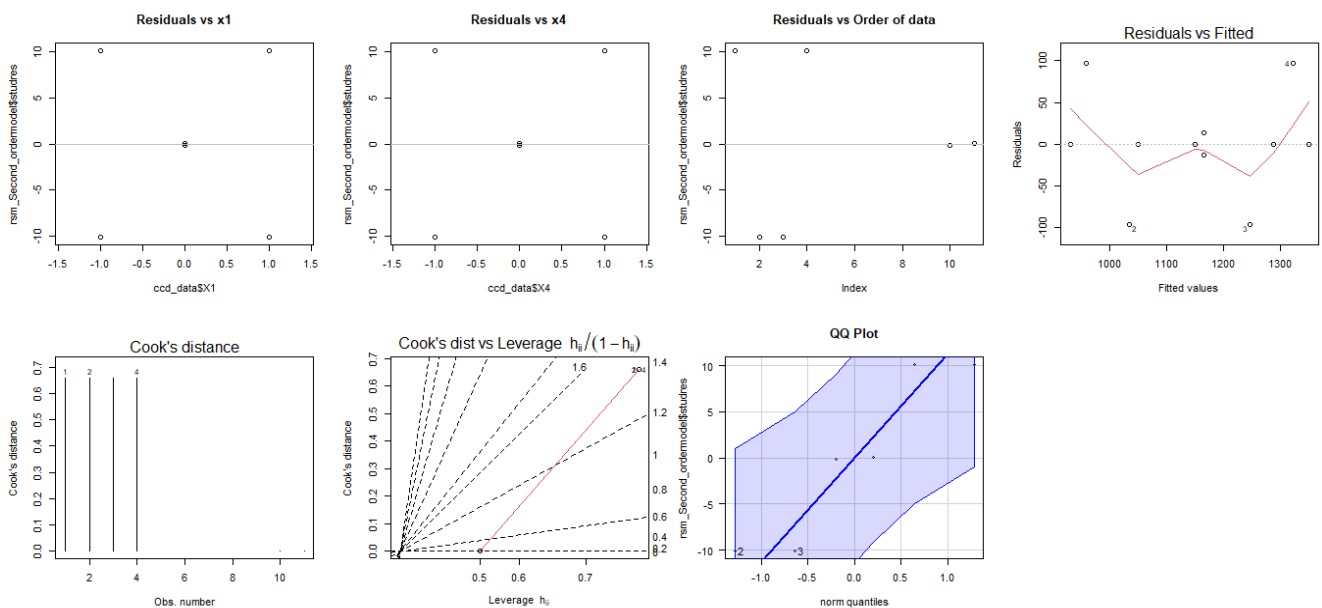


Figure 1. Residuals versus Order of data; fitted values and cooks distance.

### 3.4. Cook's Distance

This plot shows Cook's distance for each observation. Cook's distance is a measure of how influential each data point is on the fitted model. High Cook's distance values indicate observations that may have a large impact on the model and potentially warrant further investigation. The plot showed a few data points with higher Cook's distance, indicating the observations that may have a large impact on the model as shown in [figure 1](#).

### 3.5. Summary Statistics for the Variables of the Study

A standard normal distribution of data assists in creating

reliable and accurate conclusions. Goodness of fit test for Skewness and Kurtosis test were used to find the normality of the data. This part shows important descriptive statistics of the data comprising of the mean, standard deviation, coefficient of variation, skewness, and kurtosis values of Pressure of the press of the rollers, flow of the hydraulic fluid, power applied to digital setting and spacing between the anode and the cathode wheel encoder (gap).

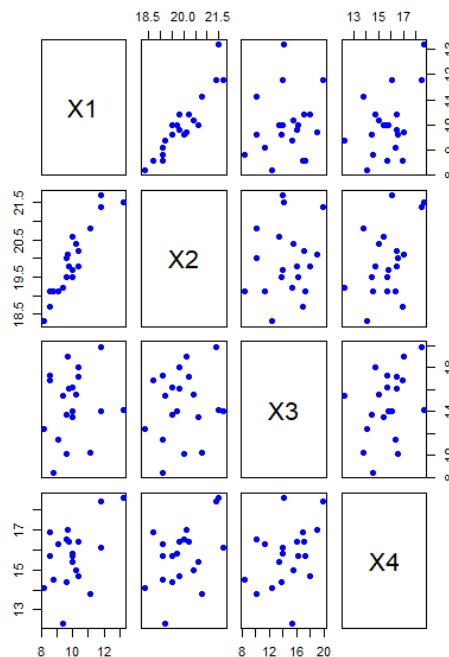
The data of Pressure of the press of the rollers, flow of the hydraulic fluid, power applied to digital setting and spacing between the anode and the cathode wheel encoder (gap) was normally distributed since their skewness and kurtosis test values were falling within the range of  $\pm 3$  and  $\pm 1$  respectively.

**Table 1.** Summary statistics for the variables of the study.

Response	Mean	Standard deviation	Median	Skewness	Kurtosis
Pressure of the rollers	13.508	9.8752	10.3	1.3061	1.1803
Flow of the hydraulic fluid	4.2037	1.5807	4	0.4529	-0.3893
Power applied to digital setting	4.5802	3.1657	4	1.3371	1.1981
Spacing between the anode and the cathode wheel encoder (Gap)	14.6550	3.0755	4	0.3263	-0.4292

### 3.6. Matrix of Scatter Plots for Process Variables

The plots indicate that the first two variables, pressure ( $x_1$ ) and flow ( $x_2$ ) are highly correlated and the other two exhibit moderate correlation as shown in [figure 2](#).



**Figure 2.** Matrix of Scatter plots for the first 20 observations.

### 3.7. Control Chart for Phase I

Figure 3 illustrates a scatter plot for the first 20 Observation principal component score along with 95% confidence contour. The scores were used as an empirical reference distribu-

tion to establish a control region for the process. It is evident that all the scores of  $Z_{i1}$  and  $Z_{i2}$  are inside the ellipses indicating that the process is in control phase.

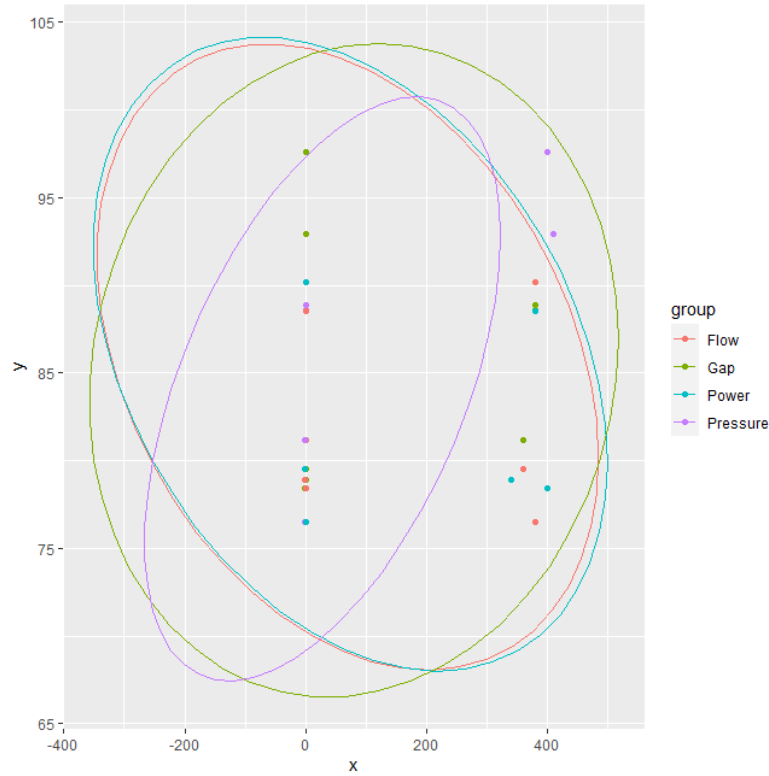


Figure 3. Scatter plot of the first 20 principal component scores  $Z_{i1}$  and  $Z_{i2}$ .

### 3.8. Control Chart for Phase II

The first few new scores are inside the ellipse while some other principal component scores are outside the ellipse. This shows that there has been a shift in the process mean hence evidence for the signal that the process is on an out of control state.

Table 2. Eigen Values.

	Eigenvalues	Percentage	Cumulative Percentage
PC1	2.36083004	59.020751	59.02075
PC2	1.02865177	25.716294	84.73705
PC3	0.53629105	13.407276	98.14432
PC4	0.07422714	1.855679	100.00000

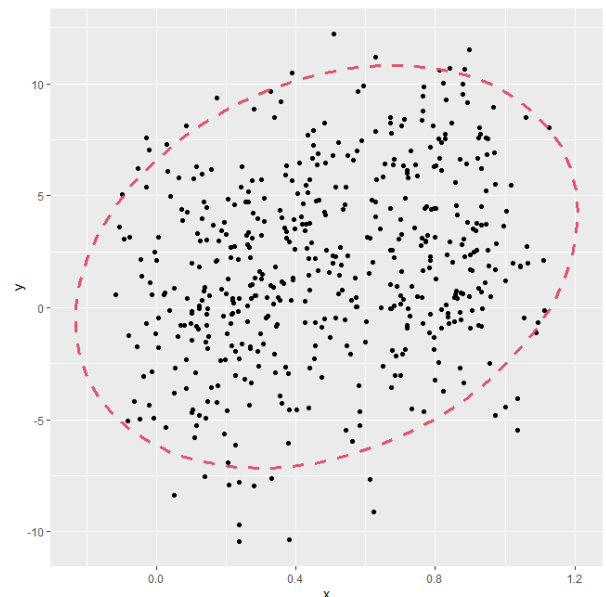


Figure 4. Principal components scores for the new observations.

The analysis unveiled that the first principal component (PC1) accounted for the highest eigenvalue of 2.36083004, representing approximately 59.02% of the total variance. The second PC exhibited an eigenvalue of 1.02865177, contributing to approximately 25.72% of the total variance. The first two components cumulatively explained about 84.74% of the variance, signifying a substantial portion of the dataset's variability. The third principal component (PC3) demonstrated an eigenvalue of 0.53629105, representing approximately 13.41% of the total variance. When combined with PC1 and PC2, these three components captured approximately 98.14% of the total variance, indicating a comprehensive representation of the data set's variability.

**Table 3.** Eigen vectors.

	X1	X2	X3	X4
PC1	0.5965972	0.3071659	-0.2174849	-0.708816752
PC2	0.5923728	0.3288784	-0.2088775	-0.705197616
PC3	0.2642171	0.8266595	0.4965390	-0.016505466
PC4	0.4726051	0.3378198	-0.8139530	-0.001644521

The Eigen vector table, it is evident that each principal component captures distinct patterns of variance within the data. For instance, in PC1, positive coefficients for  $X_1$  and  $X_2$  suggest a strong positive association, implying that variations in these variables contribute significantly to the variance captured by PC1. Similarly, in PC2, while  $X_1$  and  $X_2$  still exhibit positive coefficients, the negative coefficients for  $X_3$  and  $X_4$  indicate an inverse relationship with PC2, suggesting that changes in these variables contribute negatively to the variance explained by PC2. The Eigen vectors reveal the varying degrees of influence each original variable has on different principal components. For instance,  $X_2$  demonstrates a strong positive association with PC3, as indicated by its high positive coefficient, while  $X_3$  exhibits a weaker but still positive association. Conversely,  $X_4$  displays an inverse relationship with PC3, as evidenced by its negative coefficient.

## 4. Conclusions and Recommendations

### 4.1. Conclusion

The methodology for determining an in-control state in the manufacturing process was straightforward: it involved verifying whether the data point representing the current process state fell within the designated target range. This initial check served as a simple yet effective measure to ascertain process stability. However, when out-of-control signals were

detected, the study delved deeper into additional diagnostic displays to pinpoint the sources of variation causing these signals. This thorough examination was crucial for accurately identifying the root causes of deviations from the desired process state.

A pivotal part of this diagnostic approach was the use of principal component analysis (PCA). PCA provided critical insights by highlighting patterns and correlations within the data that might not be immediately apparent. These insights were instrumental in recognizing potential signals indicating an out-of-control process. By leveraging PCA, the study was able to develop more robust detection procedures, ensuring a higher level of accuracy in identifying and addressing issues within the manufacturing process of iron sheets.

Ultimately, the combination of simple in-control state checks and advanced analytical techniques like PCA contributed to a more effective and reliable process control system. This integrated approach not only enhanced the detection of out-of-control signals but also facilitated timely and precise interventions, leading to improvements in the overall quality and consistency of the manufactured iron sheets.

### 4.2. Recommendations

An interesting areas for further research in the domain of multivariate statistical process control are robust design of control charts and nonparametric control charts. The research for multivariate attributes control charts is also a promising task. Further research need to be done, if more than two principal components need to be retained, because the interpretation and use of the charts becomes cumbersome. Furthermore, interpretation of the principal components can be difficult, because they are not the original set of process variables but instead linear combinations of them. The control charts and trajectory plots based on principal component analysis was most effective in detecting shifts in the directions defined by the principal components. Shifts in other directions, particularly directions orthogonal to the retained principal component directions, may be very hard to detect hence the need for further research.

## Abbreviations

PCA	Principal Components Analysis
MSPC	Multivariate Statistical Process Control
SPM	Statistical Process Monitoring
PCs	Principal Components

## Author Contributions

**Martin Kyalo Mung'au:** Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Visualiza-



tion, Writing – original draft, Writing – review & editing

**Hellen Wanjiru Waititu:** Conceptualization, Data curation, Formal Analysis, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing

**Nyakundi Omwando Cornelious:** Conceptualization, Data curation, Formal Analysis, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing

## Conflicts of Interest

The authors declare no conflicts of interest.

## References

- [1] Montgomery, C. D (2001). Designs and analysis of the experiments, 5th edition, Arizon State University.
- [2] Mohamed H Abo-Hawa., MS El-Din., OA Nada., & S El-Kom, (2016). Monitoring production processes using multivariate control charts, International Journal of Advanced Engineering and Global Technology.
- [3] Maravelakis, PE., Bersimis, S., Panaretos J., & Psarakis, S (2002). On identifying the out of control variable in a multivariate control chart. Communications in Statistics - Theory and Methods, 31: 2391-2408.
- [4] Jackson JE (1991). A user guide to principal components; John Wiley: New York.
- [5] Runger, G. C.; Alt, F. B. (1996). Choosing Principal Components for Multivariate SPC. Communications in Statistics A- (Theory and Methods), 25, 909–922.
- [6] Runger GC, Alt FB, Montgomery DC. (1996) Contributors to a multivariate SPC chart signal. Communications in Statistics - Theory and Methods, 25: 2203-2213.
- [7] Bodnar, O., & Schmid, W. (2011). For monitoring the mean of a multivariate Gaussian process. *Journal of Statistical Planning and Inference*, 141(6), 2055–2070.
- [8] B. M. Wise., N. L. Ricker., D. J. Veltkamp & B. R. Kowalski, (1990). “A Theoretical Basis for the Use of Principal Components Models for Monitoring Multivariate Processes,” *Process Control and Quality*, vol. 1, no. 1, pp. 41-51.
- [9] Doganaksoy N., Faltin FW., & Tucker WT. (1991) Identification of out-of-control multivariate characteristic in a multivariable manufacturing environment. Communications in Statistics- Theory and Methods, 20: 2775-2790.
- [10] J. Yu., L. Xi., & X. Zhou, (2009). “Identifying source(s) of out-of-control signals in multivariate manufacturing processes using selective neural network ensemble,” *Engineering Applications of Artificial Intelligence*, vol. 22, no. 1, pp. 141–152.
- [11] Kourtis T. (2005) Application of latent variable methods to process control and multivariate statistical process control in industry. International Journal of Adaptive Control and Signal Processing; 19: 213–246.
- [12] Tai-Fu Li, Sheng Hu, Zheng-yuan Wei, and Zhi-qiang Liao. A, (2013) Framework for Diagnosing the Out-of-Control Signals in Multivariate Process Using Optimized Support Vector Machines.
- [13] Psarakis, S. and Papaleonida, G. E. A. (2007). SPC Procedures for Monitoring Autocorrelated Processes, *Quality Technology and Quantitative Management*, Vol. 4(4), pp. 501-540.
- [14] Ku W, Storer RH, Georgakis C. (1995) Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 30: 179–196.
- [15] Hayter AJ, Tsui K-L. (1994), Identification and quantification in multivariate quality control problems. *Journal of Quality Technology*, 26: 197-208.