Journal of Basics and Applied Sciences Research (JOBASR)
ISSN (print): 3026-9091, ISSN (online): 1597-9962
Volume 3(4) July 2025
DOI: https://dx.doi.org/10.4314/jobasr.v3i4.13

Enhancing Multivariate Process Monitoring in Soybean Meal Production Using Robust Statistical Quality Control Techniques

Kor M[1*] and Nworah C[2]

[1]Department of Statistics, Federal University Dutsin-Ma, Katsina State, Nigeria.
[2]Department of Mathematics, Federal University Dutsin-Ma, Katsina State, Nigeria

[*]Corresponding Author Email: kmoses@fudutsinma.edu.ng

ABSTRACT

Statistical process monitoring of many quality variables independently can lead to confounding results. A production process with two or more correlated quality features, require the use of Multivariate Statistical Process Control and Multivariate Capability Analysis. This study employed Multivariate Statistical Quality Control techniques to analyze the soybean meal production at Hule and Sons Nigeria Limited. Quality characteristic of interest were percentage of crude residual oil in the meal after extraction using solvent extraction method; percentage moisture content; free fatty acid (FFA); amount of phosphorus and the flash point. The study investigates the stability and capability of the multivariate process using advanced statistical process control techniques. Hoteling's T^2 square control chart, applied on the transformed data, indicated statistical control with no out-of-control signals detected as all the points fall below the upper control limit of 13.19. In contrast, Robust principal component analysis (ROBPCA) identified one observation exceeding the 95% control limit which suggests the presence of an assignable variation which the conventional method may overlook. Furthermore, the orthogonal distance (OD) chart revealed two samples outside the 95% control threshold, although all the OD values were almost all zeros, indicating potential collinearity or numerical issues in orthogonal projections. The process capability indices were $MC_p = 0.3864$; $MC_{pk} = 0.3451$; MC_{pm}=0.0086 and $MC_{pmk} = 0.0076$, signifying a substantial deviation from the desired performance standard. These results highlight the limitations of relying solely on traditional control charts and emphasize the importance of incorporating robust multivariate techniques for more sensitive and reliable process monitoring in high-dimensional data.

Keywords:
Residual oil, moisture content, free fatty acid, phosphorous, flash point, robust principal components analysis, process capability indices.

INTRODUCTION

In the industrial setting, turning eagle eye on the quality of products and services is a key intelligence for organizations, manufacturing industries, distributors, transportation companies, banks and fintech and many evolving tech industries. Building quality into the production process is one of the ways to ensure a quality product (Akinnola, 2009). Montgomery (2009) noted that "quality has become one of the most important consumer decision parameters in consumer choice for competing products and services".

He pointed seven dimensions of quality to include Performance, Reliability, Durability, Serviceability, Aesthetics, Features, Perceived quality, and Conformance to Standards (Montgomery, 2000). Focusing search light of statistical quality control to food processing/manufacturing industry is sacrosanct as it is an industry that produces edible consumer products. (Grigg,1997).

Statistical analysis is key to all forms of quantitative measures. To achieve reliable quality in any production environment, the use of statistical quality control of the process is a way to go. Statistical process control provides a systematic methodology with many techniques for maintaining quality in a production process.

How to cite this article: Kor M and Nworah C. (2025). Enhancing Multivariate Process Monitoring in Soybean Meal Production Using Robust Statistical Quality Control Techniques. *Journal of Basics and Applied Sciences Research*, *3*(4), 115-124. https://dx.doi.org/10.4314/jobasr.v3i4.13

Statistical quality Control as the use of valid analytical statistical methods to identify the presence of special causes of variation in a process (Raza & Payam, 2009). Statistical process control (SPC) involves using statistical methods (descriptive statistics and inferential statistics)in capturing and analyzing variability in a setup, be it production or service provision.

The foundation for Statistical Process Control technique as quality control tool was laid by Walter Shewart in 1920s (Marilyn & Robert, 2007

Control charts are handy tools for monitoring quality in production environments. They are used torun analysis of the process parameters to determine if a controlled process is within or out of control, helping in distinguishing between assignable cause of variation and common cause of variation (Smith, 2004). Samples collected from the process are prospectively used to monitor departures from the in-control process, hence deeming the system is in control or not (Jensen, Jones-Farmer, Champ & Woodall, 2006).

Obadara and Alaka (2013) examined how accreditation influences quality assurance in Nigerian universities by applying the Statistical Quality Assurance (SQA) method. Their study found a significant link between accreditation and factors such as resource input, process quality, and output quality, but no notable connection with the quality of academic content. They further emphasized that educational quality can be evaluated through input, process, content, and output dimensions. Consequently, these indicators were employed in their study to assess quality assurance.Awariefe and Ogbereyivwe (2024) applied both linear and nonlinear statistical models such as ARIMA and neural networks to forecast foreign direct investment in Nigeria.

Chang S.I &Ghafarias P (2025) in a comprehensive review explored the application of artificial intelligence (AI) and machine learning (ML) algorithms in statistical process monitoring, covering univariate, multivariate, profile, and image data. The study categorized AI methods into classification, pattern recognition, time series application, and generative AI. It is also shown that deep learning for multivariate statistical in-process control can be applied in discrete manufacturing, like the metal forming process. Biegel, T. et al (2022) investigated the use of deep autoencoder-based monitoring approaches, experimenting with reconstruction error and latent representation to enhance fault detection in multivariate processes.

There are many situations in which the simultaneous monitoring or control of two or more related quality characteristics is necessary. Monitoring these quality characteristics in a univariate sense can lead to confounding issues. The more the number of quality variable, the more the univariate processmonitoring procedure becomes inefficient. In this work, we apply the robust multivariate statistical approach to monitor multiple quality variable in soybean meal production.

MATERIALS AND METHODS

This study employed production data sourced from the Department of Quality Assurance at Hule and Sons Oil Company, Wanune, Benue State, Nigeria. The dataset comprises 40 samples drawn from different batches of soybean meal production. Five key quality characteristics were measured: residual crude oil percentage, moisture content, free fatty acid (FFA) level, phosphorus content, and flash point. These variables are considered critical to ensuring product quality and safety in edible oil production.

Prior to analysis, the dataset underwent preprocessing to ensure reliability and consistency. Missing values were addressed using appropriate imputation techniques, while extreme outliers were examined and treated based on domain knowledge and statistical thresholds. Each variable was standardized to zero mean and unit variance to enable valid multivariate analysis, particularly for principal component analysis and control chart development.

All statistical analyses were conducted using **R version 4.3.0**. The following packages were employed: mvnormtest for assessing multivariate normality, Hotelling for constructing T^2 control charts, rrcov for implementing robust principal component analysis (ROBPCA), and qcc for process capability assessment.

The assumption of multivariate normality was verified using Mardia's test of multivariate skewness and kurtosis. Box's M test was applied to check homogeneity of covariance matrices across subgroups. Linearity among variables was assessed through scatterplot matrices, while multicollinearity was evaluated using correlation coefficients and variance inflation factors (VIF).

The Hotelling's T^2 Control Chart was deployed in analyzing the data for the soybean oil meal production.The Hoteling's $T^2$is takes it root from multivariate normal distribution (MVN), a keymultivariate statistical analysis where sampling distributions of multivariate distributions tend to be approximately normal owing to the central limit theorem. We have that if a univariate random variable is normally distributed with mean μ and variance σ^2 it has a density function:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[(x-\mu)/\sigma]^2}{2}} \quad (1)$$

where $-\infty < x < \infty$.

The numerator of exponent in (1)can be written as

$$[(x-\mu)]^2/\sigma^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu) \quad (2)$$

In the case of multivariate normal distribution with the number of random variables is p≥ 2 , (2) can be generalized as

$$(x-\mu)'\textstyle\sum^{-1}(x-\mu) \quad (3)$$

Equation (3) is the Mahalanobis distance, where μ is a p ×1 vector of the expected values. That is

$$\mu' = [\mu_1\mu_2 \;,,, \mu_p]$$

and $\sum$ is a $p \times p$ variance-covariance matrix, given as

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \quad (4)$$

Now, substituting in (1) by (3) and the constant $\frac{1}{\sqrt{2\pi\sigma^2}}$ by $\frac{1}{(2\pi)^{p/2}|\sum|^{1/2}}$, we have

$$f(x) = \frac{1}{(2\pi)^{p/2}|\sum|^{1/2}} e^{-\frac{(x-\mu)'\sum^{-1}(x-\mu)}{2}} \quad (5)$$

where $-\infty < x_i < \infty$ and $|\sum|$is the determinant of the covariance matrix.

Fromm the t test statistics, we have

$$t = \frac{\bar{X}-\mu}{S/\sqrt{n}} \quad (6)$$

Taking the square of both sides, we have

$$t^2 = \frac{(\bar{X}-\mu)^2}{S^2/n} = n(\bar{X}-\mu)(S^2)^{-1}(\bar{X}-\mu) \quad (7)$$

Now,if p random variables, X_1 , X_2 , … Xp are monitored jointly according to the p-multivariate normal distribution, the set of quality characteristic means is represented by p × 1 vector as:

$$\bar{X} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}$$

If there are m-subgroups, then the sample means and variances fare calculated from each subgroup as

$$\bar{X}_{jk} = \frac{1}{n}\textstyle\sum_{i=1}^{n} X_{ijk} \quad (8)$$

and

$$S_{jk}^2 = \frac{1}{n-1}\textstyle\sum_{i=1}^{n}(X_{ijk} - \bar{X}_{jk})^2 \quad (9)$$

j= 1, 2, ..., p; k=1, 2, ..., m.

where X_{ijk} is the i-th observation on the j-th quality characteristic in the k-th subgroup. The covariance between quality characteristic j and quality characteristic h in the k-th subgroup is

$$S_{ijk} = \frac{1}{n-1}\textstyle\sum_{i=1}^{n}(X_{ijk} - \bar{X}_{jk})(X_{ijk} - \bar{X}_{hk}) \quad (10)$$

k=1, 2, ..., m, $j \neq h$.

Taking the averages of the statistics$\bar{X}_{jk}$, S_{jk}^2 and S_{ijk} over all m –subgroups we obtain

$$\bar{\bar{X}}_j = \frac{1}{m}\sum_{i=h}^{m} \bar{X}_{jk} \quad j = 1,2,\dots,p$$

$$\bar{S}_j^{\,2} = \frac{1}{m}\sum_{i=k}^{m} S_{jk}^2 \quad j = 1,2,\dots,p$$

$$\bar{S}_{hj} = \frac{1}{m}\sum_{i=k}^{m} s_{jhk} \quad j = 1,2,\dots,p$$

Where j≠ h and $\bar{\bar{X}}_j$ is the i-th element of the p× 1sample mean vector $\bar{\bar{X}}$and p × p average of sample covariance matrices S is formed as:

$$S = \begin{bmatrix} \bar{S}_1^{\,2} & \cdots & \bar{S}_{1p} \\ \vdots & \ddots & \vdots \\ \bar{S}_{pm}^{\,2} & \cdots & \bar{S}_p^{\,2} \end{bmatrix} \quad (11)$$

Replacing μ with $\bar{\bar{X}}$ and $\sum$ with S in (7), the statistics becomes

$$T^2 = n(\bar{X} - \bar{\bar{X}})'S^{-1}(\bar{X} - \bar{\bar{X}}) \quad (12)$$

Equation (12) above is called the Hoteling T^2 Control chart. The control limits for the T^2 control chart for phase I and phase II are respectively given by:

$$Upper\ Contro\ Limit(UCL) = \frac{p(m-1)(n-1)}{mn-m-p+1)} F_{\alpha,\ p,\ m-p+mn-1} \quad (13)$$

Lower Control Limit (LCL) = 0

and $UCL = \frac{p(m+1)(n-1)}{mn-m-p+1)} F_{\alpha,\ p,\ m-p+mn-1}$

LCL = 0 (14)

where F is the Fisher Distribution with p and m-p+mn-1 degrees of freedom at α level of significance.

The control limits here are different from specification limits. The control limits describe what the process is capable of producing while specification limits describe how the product should be produced to meet customer's expectations.

Let $N_P(\mu, \sum)$ denote a p-variate normal distribution with location μ and known covariance $\sum$. Let X_1 , X_2 , . . . $X_n \sim N_P(\mu, \sum)$ be n independent identically distributed random variables. If $\bar{X}$ is the sample mean with variance $S_{\bar{X}} = \sum/_n$. It is shown that

$$T^2 = n(\bar{X}-\mu)'(S_{\bar{X}})^{-1}(\bar{X}-\mu) \sim \chi_p^2 \quad (15)$$

2.1 Hoteling's T^2 control chart for Individual observations

If we have m sample with only 1 sample size n with p number of quality features , then the Hoteling's T^2 Statistic becomes:

$$T^2 = (X - \bar{X})'S^{-1}(X - \bar{X}) \quad (16)$$

Tracy, Young and Mason (1992) proposed a Phase I control limit as:

$$UCL = \frac{(m-1)^2}{m}\beta_{\alpha,p/2,(m-p-1)/2}$$

where m is the sample size, β is the Beta distribution with (p/2, m-p-1)/2 degrees of freedom.
Ryan (1989) defined the phase II control limits for the statistics as

$$UCL = \frac{p(m+1)(m-1)}{m(m-p)}F_{\alpha}, p, m-p$$

$$LCL = 0$$

The computation of the sample covariance matrix S, is according to Sullivan and woodall (1996) which is given as

$$S = \sum_{i=1}^{m}\frac{(x_i-\bar{x})(x_i-\bar{x})^T}{m-1} \quad (17)$$

2.2 Principal components procedure for T^2 Chart

Jackson (1991) recommended the use the principal components procedure to aid in the interpretation of an out-of-control signal. He gives an alternative form of the T^2 statistic as:

$$T^2 = (X-\bar{X})`S^{-1}(X-\bar{X}) = \sum_{i=1}^{p}\frac{z_i^2}{\lambda_i} \quad (18)$$

where $\lambda_1 > \lambda_2 > \cdots > \lambda_p$ are the eigenvalues of the estimated covariance matrix S and the z_i i=1,2,...,p are the corresponding principal components.
That is

$$z_i = U_i`(X-\bar{X})$$

The estimated covariance matrix S is a positive definite symmetric matrix. Thus, its singular value decomposition is given as $S = UAU`$ where U is a p × p orthogonal matrix whose columns are the normalized eigenvectors $U_i$of S, and A is a diagonal matrix whose elements are the corresponding eigenvalues i.e.

$$U = U_1, U_2, \ldots, U_p$$

and

$$A = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p \end{bmatrix}$$

Now if we substitute for$S^{-1} = UA^{-1}U`$ into 3.8, we have

$$T^2 = (X-\bar{X})`S^{-1}(X-\bar{X}) = (X-\bar{X})`UA^{-1}U`(X-\bar{X}) \quad (19)$$

A Hotelling's T^2 statistic for a single observation also can be written as

$$T^2 = (X-\bar{X})`S^{-1}(X-\bar{X}) = Y`R^{-1}Y \quad (20)$$

where R is the estimated correlation matrix and Y is the standardized observation vector of x, i.e.

$$R = \begin{bmatrix} 1 & \cdots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & 1 \end{bmatrix}$$

$$[y_1, y_2, \ldots, y_p] = \left[\frac{X_1-\bar{X}}{S_1}, \frac{X_2-\bar{X}}{S_1}, \ldots, \frac{X_p-\bar{X}}{S_p}\right]$$

Using a transformation similar to (10), the above T^2 can be written as

$$T^2 = \sum_{i=1}^{p}\frac{w_i^2}{\tau_i} \quad (21)$$

where $\tau_1 > \tau_2 > \cdots > \tau_p$ are the eigenvalues of the correlation matrix R, and $w_1, w_2, \ldots, w_p$ are the corresponding principal components of the matrix.

2.3 Robust PCA for Hoteling's T^2 Control Chart

Traditional multivariate process control techniques such as Hoteling's T^2 control charts operates under the assumption of multivariate normality of the data. However, in practical terms, most of time, we have data with outliers, missing values and non-Gaussian noise. These factors affect the authenticity of the results obtained from standard PCA-based T^2 charts. To take care of this, we adopt robust principal components analysis (ROBP in this work to compare with the traditional Hoteling's control chart and ensure accurate results.

If $D \in R^{n\times p}$ is the matrix of the observed data, representing n observations of p variables, Robust PCA is decomposed into:

$$D = L + S \quad (22)$$

Where L is the low-rank matrix for the underlying clean signal and S is the sparse matrix for anomaly detection. The solution is obtained by solving:

$$\min_{L,S} ||S||_* + \lambda||S||_1 \quad (23)$$

The standard PCA is performed on L by computing P and T. For each observation i, T^2 is computed as

$$T_i^2 = t_i^T\Lambda^{-1}t_i \quad (24)$$

Here:

i. T_i^2 is the statistic for the i-th observation, which measures how far each observation lies from the center of the data in the principal component space.
ii. t_i is the score vector for the i-th observation. The vector $t_i \epsilon R^a$ contains the coordinates of the i-th observation in a reduced a-dimensional space.
iii. $\Lambda$is a diagonal matrix of the eigenvalues (variances) corresponding to the retained principal components.
iv. Λ^{-1} is the inverse of the diagonal matrix Λ.

The control limit for T^2 , having *p* principal components and *n* observation at significant level α is given by:

$$CL_{T^2} = \frac{p(n^2-1)}{n(n-p)} F_{p,n-p,1-\alpha}$$

$F_{p,n-p,1-\alpha}$ is the upper 100(1-α)% percentile of the F-distribution p and n-p degrees of freedom.

2.4 Multivariate Process Capability Indices by Use of Principal Component Analysis

Over the years, several indices grounded in principal component analysis (PCA) have been introduced. Among the most widely recognized are those put forward by Wang and Chen (1998), Wang (2005), and Xekalaki and Perakis (2002). This method typically starts with PCA, which generates uncorrelated variables and facilitates the reduction of data dimensionality. These indices are based on the spectral decomposition of the covariance matrix

$$S = UDU' \quad (25)$$

where U is the eigenvectors matrix and D the diagonal matrix of the eigenvalues

D=diag($\lambda_1, , \lambda_2, \ldots , \lambda_p$) (26)

The engineering specifications (Upper, Lower Specification and Target) aretransformed as

$LSL_{PC} = u_i' LSL$;

$USL_{PC} = u_i' USL$;

$T_{PC} = u_i'$T

where LSL is the lower specification limit ; USL is the upper specification limit and T is the target value; i = 1, 2,. . . ,p. The ith principal component results in $PC_i = u_i'$x

Normally the first components are responsible for most of the variability, therefore the dimensionality can be reduced without significant loss of information. The problem consists of how many components should be retained.

The proposal by Wang and Chen (1998) is the multivariate extension of the univariate C_P, C_{pk}, C_{pm}, and C_{pmk}indices.

$$MC_p = \left(\prod_{i=1}^{v} C_P; PC_i\right)^{1/v} \quad (27)$$

where

$$C_P; PC_i = \frac{USL_{Pc_i} - LSL_{Pc_i}}{6\sigma p C_i} \quad (28)$$

v is the number of principal component and and

$$\sigma p C_i = \sqrt{\lambda_i}$$

Similarly, MC_{pk}, MC_{pm}, and MC_{pmk} are obtained by replacing $C_P; pC_i$ by $C_{Pk;}pC_i C_{Pm}; pC_i$ and $C_{Pmk;}pC_i$ respectively, where

$$C_{Pk;}pC_i = min\left\{\frac{USL_{Pc_i} - \mu}{3\sigma p C_i}, \frac{\mu - LSL_{Pc_i}}{3\sigma p C_i}\right\} \quad (29)$$

$$C_{Pm}; pC_i = \frac{USL_{Pc_i} - LSL_{Pc_i}}{6\sqrt{\sigma^2 Pc_i + (\mu - T)^2}} \quad (30)$$

$$C_{Pmk;}pC_i = \frac{USL_{Pc_i} - LSL_{Pc_i}/2 - \left|\mu - [(USL_{Pc_i} - LSL_{Pc_i})/2]\right|}{3\sqrt{\sigma^2 Pc_i + (\mu - T)^2}} \quad (31)$$

RESULTS AND DISCUSSION

Henze and Zickler (1990) normality test was carried on the data which returned a p-value less than 0.05 indicating that the data is not from a normal population.

The data was transformed using Johnson Transformation. The transformed data (Table 4.2) returns a p value of 0.63 which shows that there is no evidence to reject the assumption of multivariate normality.

The sample mean for X_1 (% oil content) is

$$\bar{X}_1 = \frac{\sum_{i=1}^{40} x_{i1}}{40}$$

= (-0.4825+ 1.52 + (1.0063) + … + 0.06)/50 = -0.19

Similarly, the sample mean for X_2 (% moisture)

$$\bar{X}_2 = \frac{\sum_{i=1}^{40} x_{i2}}{40}$$

= (-1.871 + 0.5801 + 1.3813 + …+ (-0.6174)/40 = -0.05.

In a same vain, $\bar{X}_3 = \frac{\sum_{i=1}^{40} x_{i3}}{40} = -0.02$

$$\bar{X}_4 = \frac{\sum_{i=1}^{40} x_{i4}}{40} = -0.00$$

$$\bar{X}_5 = \frac{\sum_{i=1}^{40} x_{i5}}{40} = 0.03$$

The sample mean vector is thus

$$\bar{X} = \begin{bmatrix} -0.19 \\ -0.05 \\ -0.12 \\ 0.00 \\ 0.03 \end{bmatrix}$$

The covariance matrix S, is computed as

$$S = \sum_{i=1}^{40} \frac{(x_i-\bar{x})(x_i-\bar{x})^T}{40-1} = \frac{1}{40}\left\{\begin{bmatrix}-0.4825\\-1.8771\\0.7877\\-0.2046\\-0.7700\end{bmatrix} - \begin{bmatrix}-0.1900\\-0.0500\\-0.1200\\0.0000\\0.0300\end{bmatrix}\right\} \times$$

$$\left\{\begin{bmatrix}-0.4825\\-1.8771\\0.7877\\-0.2046\\-0.7700\end{bmatrix}^T - \begin{bmatrix}-0.1900\\-0.0500\\-0.1200\\0.0000\\-0.0300\end{bmatrix}^T\right\} + \cdots$$

$$+\left\{\begin{bmatrix}0.2115\\-0.6173\\-0.2622\\-0.2046\\-04769\end{bmatrix} - \begin{bmatrix}-0.1900\\-0.0500\\-0.1200\\0.0000\\0.0300\end{bmatrix}\right\} \times \left\{\begin{bmatrix}-0.4825\\-1.8771\\0.7877\\-0.2046\\-0.7700\end{bmatrix}^T - \begin{bmatrix}-0.1900\\-0.0500\\-0.1200\\0.0000\\-0.0300\end{bmatrix}^T\right\}$$

$$S = \begin{bmatrix}+1.1000 & +0.0650 & +0.0500 & -0.0940 & +0.2600\\ +0.0650 & +0.7000 & -0.2600 & -0.1300 & -0.0690\\ +0.0500 & -0.2600 & +1.3000 & -0.0930 & +0.4600\\ -0.0940 & -0.1300 & -0.0930 & +0.8400 & -0.0300\\ +0.2600 & -0.0690 & +0.4600 & -0.3800 & +1.1000\end{bmatrix}$$

The sample correlation matrix is obtained as

$$r = \begin{bmatrix}1.00 & 0.65 & 0.50 & 0.15 & 0.26\\ 0.65 & 1.00 & 0.26 & 0.13 & 0.69\\ 0.50 & 0.26 & 1.00 & 0.93 & 0.46\\ 0.15 & 0.13 & 0.93 & 1.00 & 0.38\\ 0.26 & 0.69 & 0.46 & 0.38 & 1.00\end{bmatrix}$$

TheT_i^2 $(i = 1,2,\dots,40)$ were computed according to equation (3.7):

$$T^2 = (X-\bar{X})'S^{-1}(X-\bar{X})$$

T_1^2

$$= \left\{\begin{bmatrix}-0.4825\\-1.8771\\0.7877\\-0.2046\\-0.7700\end{bmatrix} - \begin{bmatrix}-0.1900\\-0.0500\\-0.1200\\0.0000\\0.0300\end{bmatrix}\right\}$$

$$\times \begin{bmatrix}+1.1000 & +0.0650 & +0.0500 & -0.0940 & +0.2600\\ +0.0650 & +0.7000 & -0.2600 & -0.1300 & -0.0690\\ +0.0500 & -0.2600 & +1.3000 & -0.0930 & +0.4600\\ -0.0940 & -0.1300 & -0.0930 & +0.8400 & -0.0300\\ +0.2600 & -0.0690 & +0.4600 & -0.3800 & +1.1000\end{bmatrix}^{-1}$$

$$\times \left\{\begin{bmatrix}-0.4825\\-1.8771\\0.7877\\-0.2046\\-0.7700\end{bmatrix}^T - \begin{bmatrix}-0.1900\\-0.0500\\-0.1200\\0.0000\\-0.0300\end{bmatrix}^T\right\}$$

T_1^2 = 6.34.

T_2^2 up to T_{40}^2 are computed similarly and the results are contained in table 4.

The phase I control limits of the Hotelling's control chart is established according to equation (8):

$$UCL = \frac{(m-1)^2}{m}\beta_{\alpha,p/2,(m-p-1)/2}$$

$$= \frac{(40-1)^2}{40}\beta_{0.05,5/2,(40-5-1)/2}$$

$$= \frac{(39)^2}{40}\beta_{0.05,5/2,(34)/2}$$

UCL=13.19

Figure 1 gives a graphical display of the Hotelling's chart. There is no point falling beyond the control limit which shows that the production process is under statistical control. Now this control parameters can be extended for performing a control in the future production (Phase II) using the in-control mean and covariance obtained

The orthogonal transformation of the correlated dataset was carried out to obtain a linear combination of variables called principal component. This reduced the dimensionality of the dataset. The resulting Eigen values are:

1.8729233 1.2215127 0.8607345 0.5919661 0.468

The corresponding Eigen vectors are:

$$\begin{bmatrix}-0.3237755 & 0.7004251 & -0.33949011 & 0.3369650 & 0.41925542\\ 0.1576874 & 0.3970698 & 0.40674297 & -0.7196757 & 0.36619297\\ -0.6826633 & -0.5105007 & 0.08815733 & -0.1273150 & 0.49937859\\ 0.1636387 & -0.1559277 & -0.84261959 & -0.4803055 & 0.09059646\\ -0.6144081 & 0.2584874 & -0.03907856 & -0.3487390 & -0.65767842\end{bmatrix}$$

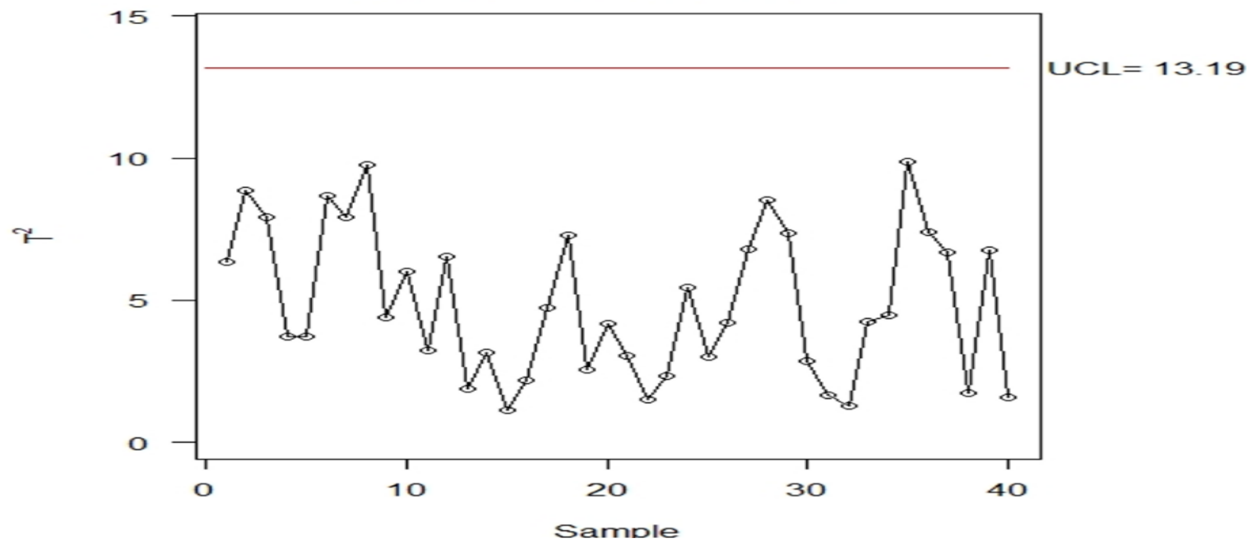

Figure 1. Classical Hoteling T2 Control chart

The Roust PCA gives the linear combination of the original variables which represents pattern in most of the data (inliers), indicating an acceptable trend. Observation 4 (fig 2 & Table 1) are flagged as outliers as they are significantly distant from the PCA center. All the orthogonal distance values are 0.00. This means that the data lies completely within PCA subspace. This indicates that there is low noise or perfect projection. The inference is that T^2 is performing all the heavy lifting by identifying points far from the multivariate center. We infer further that OD contributes almost nothing due to perfect PCA projection. Hence focus should be on T^2 for interpretation unless OD assumes non-zero values in a more complex dataset.

Table 1 T^2 Values for the PCA

61.95362587	54.2946792	145.68585269	73.29713566	191.3770554
43.67885046	87.71779681	43.45568025	58.35724141	49.37359666
63.70591007	96.90630613	89.71681798	67.28853908	55.3656548
71.27914223	58.64037584	81.99974636	41.66991526	54.55759202
38.09542325	52.58403791	91.80106364	55.14493309	50.87094673
59.18618879	41.75638343	40.40345916	76.20778094	60.62638916
59.29599745	79.85149729	76.46696943	52.63666993	47.16055385
41.47253579	38.01591923	54.69944963	69.01903703	33.4628298

Explained Variance: The explained variance tells how much of the variance of the original data is captured by each principal component. The first principal component alone explains 98.56% of the total variance while second principal component accounts for about 1% (Table 2**).**This effectively reduced the dimension of the data to 2 principal components without loss of important information.Using the number of components as 2 with alpha level of 0.05, we obtained the control limit of the Robust PCA of the Hoteling's chart as 6.0 (Fig 2).

*Table 2*Explained Variance

PC	% Variance Explained	Cumulative	Recommendation
PC1	98.58%	98.58%	Extremely dominant
PC2	1.00%	99.58%	Worth keeping
PC3	0.42%	99.99%	No much effect
PC4 to 5	0%	100%	Likely noise

Loadings: The loadings in the PCA show how much each original variable contributes to each principal component. They are the direction vectors which show the weight of the principal components as a linear combination of the original variables. Tab gives the loadings of the PCA of our data. In PC1, there is strong positive loading from variable5 (0.5266) variable3 (0.5161) while there is a strong negative loading from variable2 (-0.5051. this suggest that PC1 contrast Variable2 with variable3 and variable5. PC3 recorded remarkably high loading from variable1 (0.9281) indicating that variable1 dominates PC3. Similarly, PC4 shows strong positive from variable3 (.6768) and strong negative influence from variable5 (-0.5296)

Table 3 Loadings of the PCA

0.20500782 -0.19442909 0.92808741 -0.11721133 -0.21233081
-0.50514593 -0.46612201 0.19652117 0.36182527 0.59834753
0.51611548 -0.42757371 -0.17509979 0.67681603 -0.24912982
0.3989554 0.61974086 0.20936514 0.34175788 0.54417244
0.52655369 -0.42193552 -0.15981077 -0.52959002 0.48857605

Table 4 OD Scores

7.77156117e	16 1.91337404e	15 1.30654508e	15 1.43735632e	15
4.80002970e	15 4.48405047e	16 2.01987376e	15 9.09180473e	16
4.75554754e	16 1.62719651e	15 1.97669590e	15 2.82538263e	15
2.12208802e	15 1.95888305e	15 9.52481673e	16 7.13053648e	16
1.02507921e	15 1.83942212e	15 6.77599955e	16 8.23363432e	16
7.77651595e	16 8.46232179e	16 2.43237678e	15 1.34155937e	15
1.23629204e	15 1.19316697e	15 7.31189546e	16 8.69586270e	16
1.55926071e	15 7.23277762e	16 7.67179373e	16 1.24763165e	15
1.80902966e	15 1.12675331e	15 7.17898916e	16 2.89776717e	16
5.38200579e	16 1.02507921e	15 1.55133557e	15 2.07819583e	16

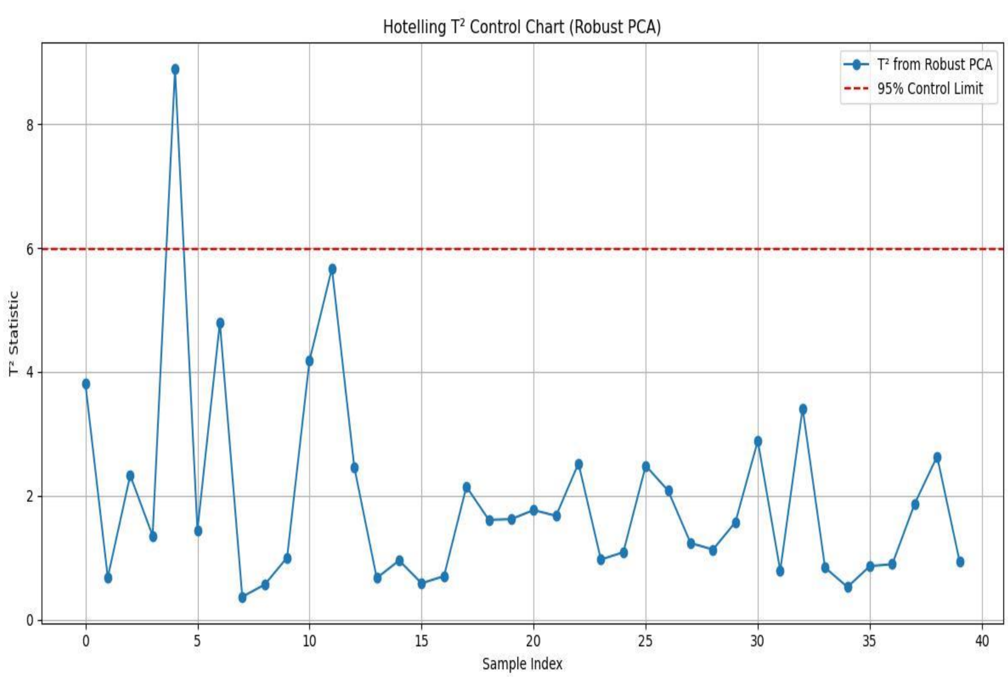

Figure 3Hoteling’s Control Chart of Robust PCA

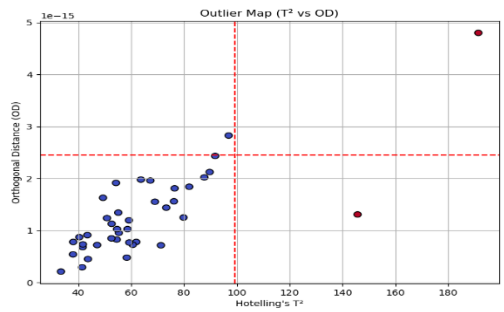

Figure 4 Outlier Plot

CONCLUSION

The Hotelling's T² chart indicated that the process was statistically in control—with all values falling below the control limit of 13.19—ROBPCA revealed an outlier exceeding the 95% threshold. This divergence highlights the limitation of conventional methods in detecting subtle process variations. Similar insights were noted by Tracy, Young, and Mason (1992) and Jackson (1991), who emphasized the enhanced sensitivity of robust multivariate approaches in detecting non-normality or masked signals in high-dimensional datasets.

Moreover, the process capability indices (Cp = 0.3864 and Cpk = 0.0086) reflect poor process performance, suggesting that despite appearing stable, the production process consistently fails to meet customer-defined specification limits. This aligns with Montgomery (2009), who stated that stability without capability renders a quality control process ineffective in practice.

These results imply that relying solely on traditional univariate or non-robust methods can result in overlooked anomalies, leading to suboptimal product quality. The study emphasizes the need for industries, particularly food processors, to integrate advanced MSQC techniques in routine monitoring to improve both detection accuracy and decision-making.

Further work should be focus on identifying the root causes of these anomalies and improving process design to enhance overall capability.

REFERENCE

Akinola, A. A. (2009). *Total quality management and business success*. Ibadan: Greenline Publishers.

Awariefe, C., &Ogbereyivwe, O. (2024).Time series modelling and forecasting foreign direct investment using linear and nonlinear models: The case of Nigeria. *Journal of Basics and Applied Sciences Research*, *2*(1), 46–53. https://doi.org/10.33003/jobasr-2024-v2i1-33

Biegel, T., Bauer, M., & Riedl, M. (2022). Fault detection in multivariate processes using deep autoencoder-based monitoring. *IEEE Transactions on Industrial Informatics*, 18(4), 2765–2774. https://doi.org/10.1109/TII.2021.3075642

Chang, S. I., &Ghafarias, P. (2025). Artificial intelligence and machine learning in statistical process monitoring: A comprehensive review. *Computers & Industrial Engineering*, [Volume pending].

Grigg, N. P. (1997). Monitoring and controlling quality in food processing operations. *Total Quality Management*, 8(6), 365–376. https://doi.org/10.1080/0954412979560

Henze, N., &Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods*, **19**(10), 3595–3617. https://doi.org/10.1080/03610929008830400

Jackson, J. E. (1991). *A user's guide to principal components*. New York: Wiley.

Jensen, W. A., Jones-Farmer, L. A., Champ, C. W., & Woodall, W. H. (2006). Effects of parameter estimation on control chart performance: A review. *Journal of Quality Technology*, 38(4), 349–364.

Marilyn, M. H., & Robert, D. H. (2007). *Statistical thinking: Improving business performance* (2nd ed.). Duxbury Press.

Montgomery, D. C. (2009). *Introduction to Statistical Quality Control* (6th ed.). John Wiley & Sons.

Obadara, E. O., & Alaka, A. A. (2013). Accreditation and quality assurance in Nigerian universities. *Journal of Education and Practice*, 4(8), 124–130.

Raza, S. A., & Payam, S. (2009). Statistical process control as a quality control tool. *Journal of Quality and Technology Management*, 5(1), 21–34.

Ryan, T. P. (1989). *Statistical methods for quality improvement*. New York: Wiley.

Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. Van Nostrand.

Smith, G. M. (2004). *Statistical process control and quality improvement* (5th ed.). Pearson Prentice Hall.

Smith, S. (2004). *Control charts and process monitoring*. Journal of Quality Technology, 36(1), 2–12.

Sullivan, J. H., & Woodall, W. H. (1996). A comparison of multivariate control charts for individual observations. *Journal of Quality Technology*, 28(4), 398–408.

Tracy, N. D., Young, J. C., & Mason, R. L. (1992). Multivariate control charts for individual observations. *Journal of Quality Technology*, 24(2), 88–95.

Wang, F. K., & Chen, K. S. (1998). Process capability indices for multivariate processes. *Computers & Industrial Engineering*, 35(1–2), 45–48.

Wang, F. K. (2005). An extension of process capability index to multivariate processes. *International Journal of Production Research*, 43(12), 2551–2560.

Xekalaki, E., & Perakis, K. (2002). Multivariate capability indices: A review. *Quality and Reliability Engineering International*, 18(2), 111–118.