

# Application of Statistical Monitoring Using Autocorrelated Data and With the Influence of Multicollinearity in a Steel Process

Cristie Diego Pimenta<sup>1</sup>, Messias Borges Silva<sup>2</sup>, Fernando Augusto Silva Marins<sup>3</sup>, Aneirson Francisco da Silva<sup>4</sup>

<sup>1</sup>Department of Production, University of Guaratinguetá (Feg-Unesp), São Paulo, Brazil

Correspondence: Cristie Diego Pimenta, Department of Production, University of Guaratinguetá (Feg-Unesp), São Paulo, Brazil

Received: April 16, 2021 Accepted: June 17, 2021 Online Published: June 22, 2021

doi:10.5539/ijsp.v10n4p96

URL: <https://doi.org/10.5539/ijsp.v10n4p96>

## Abstract

The purpose of this article is to demonstrate a practical application of control charts in an industrial process that has data auto-correlated with each other. Although the control charts created by Walter A. Shewhart are very effective in monitoring processes, there are very important statistical assumptions for Shewhart's control charts to be applied correctly. Choosing the correct Control Chart is essential for managers to be able to make coherent decisions within companies. With this study, it was possible to demonstrate that the original data collected in the process, which at first appeared to have many special causes of variation, was actually a stable process (no anomalies present). However, this finding required the use of autoregressive models, multivariate statistics, autocorrelation and normality tests, multicollinearity analysis and the use of the EWMA control chart. It was concluded that it is of fundamental importance to choose the appropriate control chart for monitoring industrial processes, to ensure that changes in processes do not incorporate non-existent variations and do not cause an increase in the number of defective products.

**Keywords:** control chart, autocorrelated data, principal component analysis, multicollinearity

## 1. Introduction

The objective of this work is to suggest a procedure for choosing the appropriate control chart for a specific situation. In addition, demonstrate the practical application of this methodology in an industrial process that has the data auto-correlated with each other.

Statistical Process Control (CEP) or also called control charts was created by Walter A. Shewhart in 1924. At first, few believed in the effectiveness of this new technique, but over time, Shewhart's control charts became essential for monitoring processes (MACHADO, 2006).

CEP is one of the most fundamental methodologies developed to assist in process control and was used in the rise of Japan, as a leading nation in quality and from that the world has awakened to the importance of obtaining products through statistically stable and capable processes to meet customer requirements (RAMOS, 2005).

The variability in a process reflects the differences between the units produced. Part of this variability is inherent to the process and comes from small disturbances or random causes (CLARO, 2008). Most of the time, the presence of random causes does not demand any reaction and, when the process has only this natural variability, it is said to be under statistical control. Processes under statistical control have predictable results, within statistical limits established through a preliminary set of data. However, a process can suffer major disturbances, the effect of which changes the parameters of the distribution of the random variable  $X$ , either by moving its mean away from the target value or by increasing its variability, which is generated by a special cause (CLARO, 2008).

The special causes of variation are when the data are not random and do not come from the Normal distribution of probability (Gaussian curve). In addition, Control charts are intended to distinguish special causes from common causes of variation (COSTA; EPPRECHT; CARPINETTI, 2005).

Generally, products that meet customer requirements come from stable processes, with natural and random variations. Statistical process control (CEP) is a problem-solving tool useful in achieving process stability and improving the ability to meet specifications by reducing variability (MONTGOMERY, 2004).

## 2. Statistical Methods

### 2.1 Conventional Control Chart ( $\bar{X}$ and MR)

The control chart  $\bar{X}$  and MR (individual value and moving range) of Shewhart it is not good for detecting changes in the process that are smaller than one sigma. Thus, statistical methods such as the EWMA graph (Exponentially Weighted Moving Average) have been widely used for the multivariate case in which the vector of means is subject to small changes (MACHADO, 2006).

The classic Shewhart scheme for control chart of measurements of process input variables ( $X$ 's), consists of taking samples at regular time intervals. For each sample, the value of an appropriate statistic is determined. Then these values are plotted on charts where the control limits were calculated (LCL is the lower control limit and UCL is the upper control limit), whose spacing in relation to the central line (CL) is established based on the natural variability of the measures of  $X$ 's (values) of a product quality characteristic resulting from a process under control (MACHADO, 2006).

The method for calculating the control limits for the  $\bar{X}$  and MR control chart are basic and can be found in Ramos (2005) or in Costa; Epprecht; Carpinetti (2005).

It is very important to note that conventional control charts require the data to be independent of each other and for these control charts are not suitable when there is permanent autocorrelation. This happens even if this autocorrelation is weak (MACHADO, 2006).

If the data has a strong autocorrelation over time, then the control chart will produce many false alarms that can lead to incorrect statistical analysis and this can influence incorrect decision making (MONTGOMERY, 2004). However, the assumption of uncorrelated observations is not always satisfied in some manufacturing processes. Examples include chemical processes in which consecutive measurements of a process variable or a product characteristic tend to have a high degree of correlation (MONTGOMERY, 2004).

Nowadays, data autocorrelation is very frequent in many batch and continuous processes, mainly in the chemical industry and in the metallurgical industry. In general, these processes rarely produce independent observations, so they cannot be monitored by conventional Shewhart charts (COSTA; EPPRECHT; CARPINETTI, 2005). Therefore, it is very important to check before starting the monitoring of a process, if the data from measurements are independent or if they are autocorrelated, because an inadequate control chart, which produces excessive false alarms, will end up being interpreted incorrectly by the personnel involved in monitoring the process.

### 2.2 EWMA Control Chart (Exponentially Weighted Moving Average)

A major disadvantage of any Shewhart control chart is that it uses only the information about the process contained in the last plot point, and ignores any information given by the entire sequence of points. This characteristic makes the Shewhart control chart relatively insensitive to small changes in the process in the magnitude of  $1.5 \sigma$  or less (MONTGOMERY, 2004).

EWMA is also widely used in estimation methods and for forecasting time series and is used in statistical process control charts. The EWMA consists of the exponentially weighted average of the observations: the weights decrease according to the age of the observations, in geometric progression, whose factor is determined by the damping constant  $\lambda$ . In the application of EWMA in statistical process control, the most usual is to apply it to individual observations of measurable characteristics.

The exponentially weighted moving average (EWMA) control chart (shown in Figure 1) is indicated when it is necessary to detect small changes. The red lines represent the control limits (UCL and LCL) and the green line represents the process average (CL).

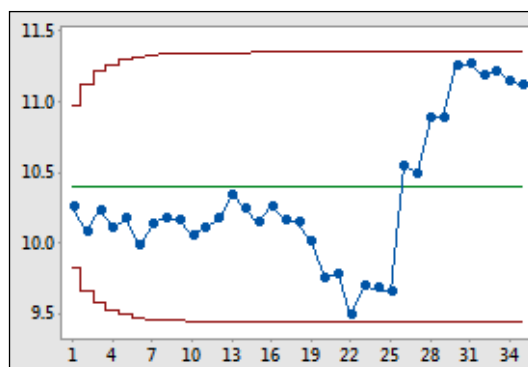


Figure 1. EWMA chart example

This chart was introduced by Roberts (1959) and this chart is defined as:

$$Z_t = \lambda x_t + (1-\lambda) Z_{t-1} \quad (1)$$

Where  $\lambda$  is the damping constant such that  $0 < \lambda \leq 1$  and the initial value (required with the first sample at  $t = 1$ ). The index  $t = 1, 2, \dots$  represents the order number of the observation and for the initial value,  $Z_0 = \mu_0$  (which is the target value or the mean value in control of the variable  $X$ ) is used.

Assuming that the  $X_i$ 's are independent and identically distributed, with variance  $\sigma_x^2$ , the variance of  $Z_t$  is given by:

$$\sigma^2(Z_t) = \left[ 1 - (1-\lambda)^{2t} \right] \frac{\lambda}{2-\lambda} \sigma_x^2. \quad (2)$$

The upper control limit (UCL), the middle line (ML) and the lower control limit (LCL) of the EWMA chart are given by:

$$\begin{aligned} \text{LSC} &= \mu_0 + K \sigma_x \sqrt{\left[ 1 - (1-\lambda)^{2t} \right] \frac{\lambda}{2-\lambda}}, \\ \text{LM} &= \mu_0, \end{aligned} \quad (3)$$

$$\text{LIC} = \mu_0 - K \sigma_x \sqrt{\left[ 1 - (1-\lambda)^{2t} \right] \frac{\lambda}{2-\lambda}}, \quad (4)$$

Since  $K$  is the opening coefficient of the control limits ("K-sigma limits") and  $\sigma_x$  is the standard deviation of the process when in control.

The parameters of an EWMA control chart are, therefore, the damping constant  $\lambda$  and the opening coefficient of the control limits  $K$ . The process is considered out of control and actions must be taken whenever  $Z_t$  is outside the range of limits of control. While the process is in control,  $Z_t$  oscillates around the midline of the graph; if there is a change in the mean of variable  $X$ ,  $Z_t$  will grow (or decrease) until it reaches the new value of the mean and then it will oscillate around itself. At each restart of the process after the elimination of the special cause and consequent readjustments,  $Z_0 = \mu_0$ .

For small values of the damping constant  $\lambda$ , the EWMA scheme detects maladjustments of small magnitude more quickly. Small values of  $\lambda$  make the historical data (observations prior to the last available) have a large weight in the calculation of  $Z_t$  and, conversely, large values of  $\lambda$  make the last observation have a large weight in the calculation of  $Z_t$ . For  $\lambda = 1$ , the EWMA graph will be equal to the Shewhart graph, and the historical data no longer have any influence on the calculation of  $Z_t$ .

The choice of parameters to be used in the EWMA control chart should be based on the analysis in Table 1.

Table 1. Average sequence lengths for various EWMA control schemes (MONTGOMERY, 2004)

Average change (multiple of $\sigma$ )	K= 3.054 $\lambda= 0.40$	K= 2.998 $\lambda= 0.25$	K= 2.962 $\lambda= 0.20$	K= 2.814 $\lambda= 0.10$	K= 2.615 $\lambda= 0.05$
0	500	500	500	500	500
0.25	224	170	150	106	84.1
0.50	71.2	48.2	41.8	31.3	28.8
0.75	28.4	20.1	18.2	15.9	16.4
1.00	14.3	11.1	10.5	10.3	11.4
1.50	5.9	5.5	5.5	6.1	7.1
2.00	3.5	3.6	3.7	4.4	5.2
2.50	2.5	2.7	2.9	3.4	4.2
3.00	2.0	2.3	2.4	2.9	3.5
4.00	1.4	1.7	1.9	2.2	2.7

Table 1 provides important information on how to choose the parameters to adjust the control limits in order to ensure the correct monitoring of the EWMA control chart. The first column "Change in average" refers to how sensitive the process will be to detect a change in average (in relation to the process  $\sigma$ ). The other columns show examples of the  $K$  and  $\lambda$  values to be used and what would be the performance and characteristics of the EWMA control chart based on the sample size that will be needed to ensure the correct monitoring of the process. For example, if the adjustments are  $K = 3.054$  and  $\lambda = 0.40$  and if you want to verify a change in mean variation of  $0.50 \sigma$ , then a sample of size 71.2 will be

required, that is, you will need to obtain 72 measurements to place on the EWMA control. Using Table 1, it will be possible to determine exactly what the sensitivity level of the control chart will be in order to detect changes in the variation of the mean and determine what should be the minimum sample size necessary for the EWMA control chart to be effective in this detection.

### 2.3 The Sample Autocorrelation Function and Autoregressive Models (ARIMA)

According to Schneider and Pruett (1994) the sample autocorrelation function is defined by:

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^N (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^N (X_t - \bar{X})^2} \quad \text{para } k = 0, 1, 2, \dots, N$$

$$\bar{X} = \frac{1}{N} \sum_{t=1}^N X_t \quad (5)$$

In most manufacturing processes, autocorrelation is positive, that is, a value below the large average at time (t-1) tends to be followed also by a value below at time t, and vice versa, which ends up causing observations to be ordered chronologically in a line graph (CLARO, 2008). In processes with negative autocorrelation, consecutive observations alternate above and below the global average, also creating a typical pattern in the graph, which reveals the absence of randomness (CLARO, 2008).

Models called ARIMA were created by Box and Jenkins in 1970 and these models are considered stochastic and describe a large number of industrial processes (CASTILLO, 2002). ARIMA is the acronym used by the combination of autoregressive (AR), integrated (I) and moving average (MR), however this model is routinely also called by the acronym (AR). According to Castillo (2002), the observations of an ARIMA model are calculated by:

$$X_t = \mu + \sum_{k=1}^p \phi_k X_{t-k} + e_t - \sum_{k=1}^q \theta_k e_{t-k} \quad \text{onde } t=1, 2, 3, \dots, T \quad (6)$$

Autoregressive models are represented by:

$$E(X_t) = \mu = \frac{\zeta}{1-\phi}$$

$$\gamma_0 = \sigma_X^2 = \frac{\sigma_e^2}{1-\phi^2}$$

$$\gamma_k = \frac{\phi^k \sigma_e^2}{1-\phi^2}$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k \quad \text{para } k = 0, 1, 2, \dots \quad (7)$$

Using  $\theta_k=0, \forall k$ . The value of  $X_t$  is a linear combination of the most recent past “p” values, plus the mean ( $\mu$ ) and the residual (or random error), being  $e_t \sim \text{IIDN}(0, \sigma_e^2)$ , which incorporates what is new in the series at time t and cannot be explained by past values. The values of  $X_{t-1}, X_{t-2}, \dots$  are independent of  $e_t$ . The autoregressive coefficient of the lag “k” is represented by  $\phi_k$ . For this model the values are regressed on themselves, using the same procedure for linear regression (CASTILLO, 2002).

The first-order autoregressive model  $\text{AR}_{(1)}$ , it is often adopted to represent manufacturing processes and this model is often applied in CEP (statistical process control). Observations of this process can be written using  $\theta_k=0, \forall k$  and  $\theta_k=0$  for  $k > 1$ . In this model the properties will be calculated by:

$$\begin{aligned}
 E(X_t) &= \mu = \frac{\zeta}{1-\phi} \\
 \gamma_0 &= \sigma_X^2 = \frac{\sigma_e^2}{1-\phi^2} \\
 \gamma_k &= \frac{\phi^k \sigma_e^2}{1-\phi^2} \\
 \rho_k &= \frac{\gamma_k}{\gamma_0} = \phi^k \quad \text{para } k = 0, 1, 2, \dots
 \end{aligned}
 \tag{8}$$

The model parameter is  $\phi$  (the subscript is usually suppressed on first order models) and  $\zeta$  is a constant called drifting parameter (CASTILLO, 2002). For the process to be stationary the condition is  $|\phi| < 1$ . When  $|\phi| \geq 1$  the process is called non-stationary. In this case, the mean and the variance are not constant over time and there is no tendency for the process to return to the target value without any external interference and this makes it impossible to use the conventional control charts created by Walter Shewart.

Berthouex, Hunter and Pallesen (1978) is the first scientific publication on the monitoring of processes using residues from models of the ARIMA family as control statistics, for cases in which there is autocorrelation of the original values. The autocorrelated processes are non-random in nature, so it was proposed to adjust their observations (measurements) for a model of the ARIMA family and to monitor the values using the resulting residues (CLARO, 2008). As these residues have the properties of independence and normality, they can be used in a graph for individual observations, with control limits calculated in the conventional way, in which it is possible to detect the special causes.

Montgomery (2004) proposed the use of the EWMA graph as an approximation to the ARIMA models to adjust the observations of the process. This monitoring method, called EWMA, contains a moving central line and is based on the fact that the conditional average of a stationary time series is constantly changing and the sequence of EWMA values allows to estimate these changes. The information on the state of statistical control and the dynamics of the process, represented respectively by the EWMA and the original observations, are gathered in the same graph.

Claro (2008) mentioned that when considering the monitoring of the mean of an autocorrelated process, it is necessary to create some way to represent the variable of interest, to be monitored in the control graph. Among the various possibilities for this, it is appropriate to use the residues after adjusting some autoregressive model. Monitoring the process through the model residues implies the following drawbacks: (a) requires intermediate calculations from the user before the value is plotted on the chart; (b) when the autocorrelation is positive, or when the displacements of the mean are small, the efficiency of the statistical device is low; (c) the interpretation of waste letters may not be so simple for a good part of the users; (d) adjusting and maintaining an appropriate model for each variable can be extremely laborious.

#### 2.4 Multiple Linear Regression, Residual Analysis and Multicollinearity

The multiple regression technique is very efficient to develop statistical models that quantify the influence of the input variables of the process to predict the output variables (BENYOUNIS AND OLABI, 2008).

Montgomery and Runger (2009), explain that multiple regression is used for situations involving more than one regressor, as:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \tag{9}$$

In this expression  $Y$  represents the dependent variable, the independent variables are represented by  $x_1, x_2, \dots, x_n$  and  $\varepsilon$  is the random error term. The unknown parameters are  $\beta_0, \beta_1, \beta_2$  and  $\beta_n$ . In this model, the  $\beta_0$  parameter is the intersection of the plane,  $\beta_1, \beta_2$  and  $\beta_n$  are the partial regression coefficients.

The models that include interaction effects, according to Montgomery and Runger (2003), can be analyzed by the multiple regression method. An interaction between two variables can be represented by a cross term, because if we admit that  $x_3 = x_1 x_2$  and  $\beta_3 = \beta_{12}$ , then the model, including terms of interaction, uses:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \varepsilon \tag{10}$$

The coefficient of multiple determination ( $R^2$ ) is a measure of the amount of reduction in the variability of Y, obtained by the use of the Regressors  $X_1, X_2, \dots, X_k$ . The range of variation of this coefficient is zero to one ( $0 \leq R^2 \leq 1$ ). If the value of  $R^2$  is close to one, it means that the various variables X measures, are responsible almost entirely for the variability of Y. Otherwise,  $R^2$  will display a value close to zero (MONTGOMERY; RUNGER, 2003).

The residuals of a regression model are  $e_i = y_i - \hat{y}_i$ ,  $i = 1, 2, \dots, n$ , where  $y_i$  is a real observation and  $\hat{y}_i$  is the corresponding adjusted value from the regression model.

The estimation of the model parameters requires the assumption that the errors (residuals) are random variables, not correlated, with zero mean and constant variance. For the application of residues in statistical process control (CEP), errors must be normally distributed (MONTGOMERY; RUNGER, 2003).

The premises of normality and stability of the residues must be verified, which can be done by examining the residual graph (MONTGOMERY; RUNGER, 2003).

Selection criteria can be classified into qualitative and quantitative. The qualitative criteria consider the link between the model and the studied process, its interpretability and comprehensibility. The quantitative criteria take into account the quality of the adjustment, the complexity and the generability of the model and must also be interpretable, make sense and be understandable (MYUNG; PITT; KIM, 2003).

In multiple regression problems, we expect to find dependencies between the response variable Y and the regressors  $X_j$ . In most regression problems, however, we also find dependencies between regressors  $X_j$ . In situations where these dependencies are strong, we say that there is multicollinearity. Multicollinearity can have serious effects on the estimates of the regression coefficients and on the general applicability of the estimated model (MONTGOMERY; RUNGER, 2003).

According to Montgomery and Runger (2003), the variance inflation factor (FIV) for  $\beta_j$  is used to measure multicollinearity in the model as:

$$FIV(\beta_j) = 1/(1 - R^2) \quad j = 1, 2, \dots, k \quad (11)$$

An alternative to eliminate multicollinearity is the application of Principal Component Analysis (MONTGOMERY, 2013).

### 2.5 Principal Component Analysis (Multivariate Statistics)

Principal Components Analysis (PCA) is a mathematical procedure that uses an orthogonal transformation (vector orthogonalization) to convert a set of observations of possibly correlated variables into a set of variable values linearly not correlated with major component calls. The number of principal components is always less than or equal to the number of original variables depending on the application area, the PCA is also known as a discrete Karhunen-Loève (KLT) transform, Hotelling transform or orthogonal decomposition itself (POD).

The PCA was invented in 1901 by Karl Pearson. It is now most commonly used as an exploratory data analysis tool and to make predictive models. PCA can be done by decomposition into autovalues (own values) of a covariance matrix, usually after centralizing (and normalizing or using Z-scores) the data matrix for each attribute. The PCA results are generally discussed in terms of component scores (scores), also called factor scores (the transformed variable values correspond to a particular data point) and the weight by which each normalized variable The original must be multiplied to obtain the component score.

Its main objective is to explain the structure of variance and covariance of a random vector, composed of random p-variables, by constructing linear combinations of the original variables. These linear combinations are called main components and are not correlated with each other (MINGOTI, 2007).

Obtaining the main components involves the decomposition of the covariances matrix of the random vector of interest. If any transformation of this random vector is made, the components must be determined using the covariances matrix relative to the transformed vector. A very usual transformation is the standardization of the variables of the vector by the respective means and standard deviations, generating new variables centered in zero and with variances equal to one. In this case, the main components are determined from the covariances matrix of the original standardized variables, which is equivalent to extracting the main components using the correlation matrix of the original variables (MINGOTI, 2007).

Montgomery (2004) states that the main objective of the main components is to find the new set of orthogonal directions that define the maximum variability in the original data, which, hopefully, will lead to a description of the process requiring considerably Less variable than the original P's. The information contained in the complete set of P main components is exactly equivalent to the information contained in the complete set of all the original process

variables.

In general, we will want to retain sufficient components to explain a reasonable proportion of the total variability of the process, but there are no clear guidelines on how much variability needs to be explained in order to produce an effective procedure of Process monitoring (MONTGOMERY, 2004).

Johnson and Wichern (2007) demonstrate that only the components corresponding to the autovalues (latent roots) of magnitudes greater than one should be considered. Another way to define the number of components is through the percentage of variance explained. In this case, the researcher should judge if the components sufficiently explain the relationship between the original  $P$  variables. Generally, a good degree of explanation is higher than 75% and the property of the main components is the independence between them. In this way, they can replace the original variables and eliminate the multicollinearity problem (JOHNSON; WICHERN, 2007).

According to Mingoti (2007), being the auto vector  $e_i$  denoted by  $e_i = (e_{i1} \ e_{i2} \dots \ e_{ip})'$ . Consider the random vector  $Y = O'X$ , with  $O_{p \times p}$  being the orthogonal matrix of the  $p \times p$  dimension, consisting of the normalized vectors of Matrix  $\sum_{p \times p}$ , such as:

$$O_{p \times p} = \begin{bmatrix} e_{11} & e_{21} & \dots & e_{p1} \\ e_{12} & e_{22} & \dots & e_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1p} & e_{2p} & \dots & e_{pp} \end{bmatrix} = [e_1 \ e_2 \dots e_p] \quad (12)$$

The vector's  $Y$  component of  $p$  linear combinations of random variables  $X$ 's, vector has vector of averages equal to the  $\mu$  and the matrix of covariances  $p \times p$ , that is a diagonal matrix, whose elements are equal to there  $\lambda_i$ ,  $i = 1, 2, \dots, p$  this is :

$$\Lambda_{p \times p} = \begin{bmatrix} \lambda_1 & & \mathbf{0} \\ & \lambda_2 & \\ \mathbf{0} & & \lambda_p \end{bmatrix} \quad (13)$$

So, as demonstrated by Mingoti (2007), the random variables that constitute the vector  $Y$  are not correlated with each other.

### 3. Results and Discussion

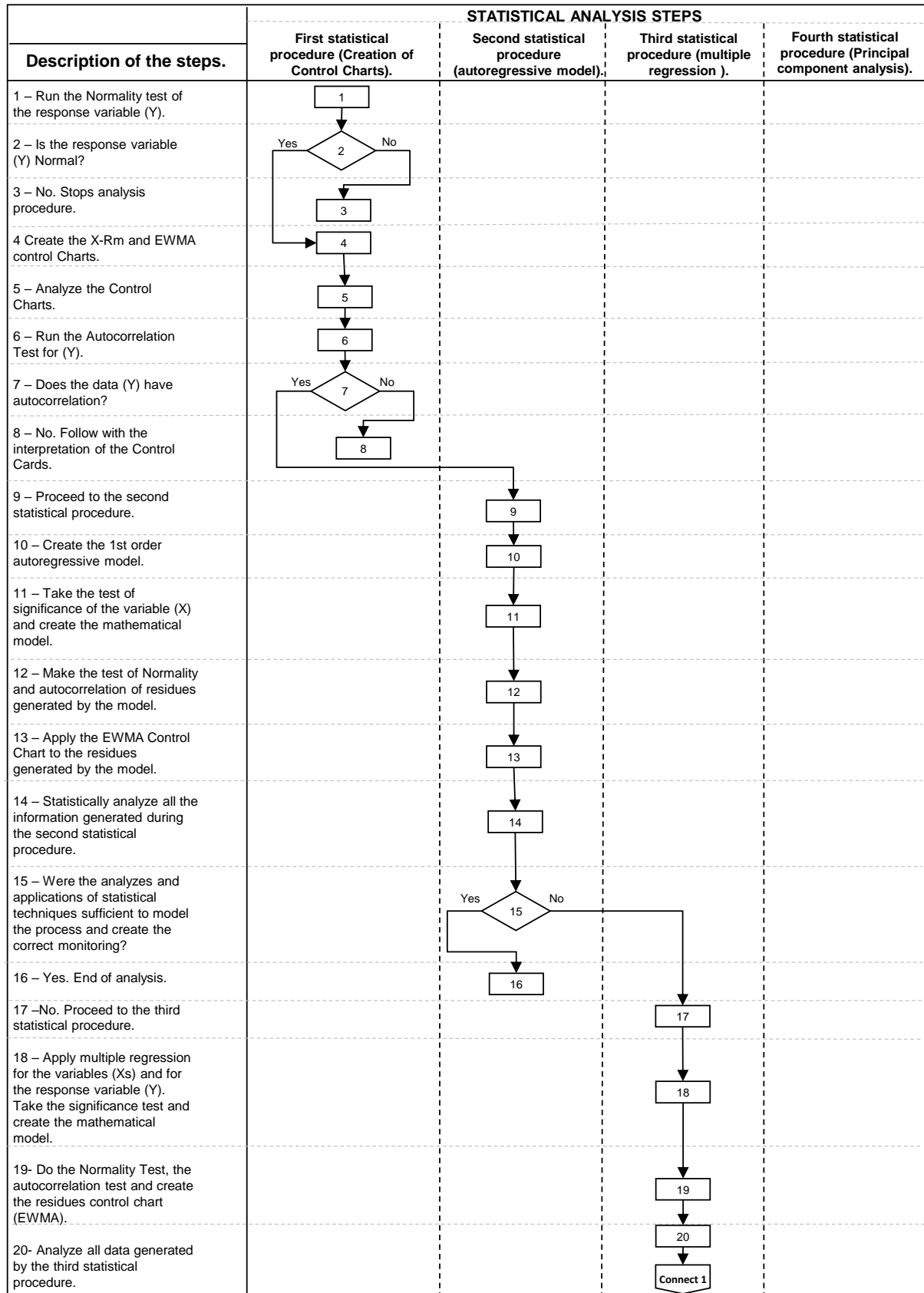
Table 2 shows the process input variables ( $X$ s) and the output variable ( $Y$ ), monitored in the studied process. Statistical procedures will be applied to analyze the data presented in Table 2. These input variables ( $X$ s) are related to a specific steel heat treatment process and the output variable ( $Y$ ) refers to a mechanical property of steel, which is resulting from the adjustments of these factors ( $X$ s). This article will be based on this example of data to demonstrate all the statistical analyzes that will come later, with the aim of portraying the common situations resulting from monitoring processes carried out in industries.

Table 2. Input variables (Xs) and output variable (Y)

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	Y
7,09	50,51	31,08	0,85	0,97	0,62	528,08
6,31	49,95	27,20	0,85	0,95	0,60	526,47
6,77	49,95	22,76	0,82	0,91	0,58	524,37
8,30	49,40	23,87	0,85	1,07	0,61	530,24
8,23	46,07	23,31	0,93	1,03	0,61	529,37
7,19	46,62	29,97	0,83	0,93	0,61	528,47
7,14	49,40	33,86	0,91	1,01	0,61	527,36
10,72	54,95	33,30	0,81	0,94	0,61	527,53
10,41	54,95	28,86	1,05	1,12	0,60	527,58
9,43	54,39	31,64	0,92	1,01	0,60	527,14
10,10	54,39	27,20	0,92	1,07	0,60	527,92
8,99	53,84	28,86	1,20	1,30	0,60	527,58
8,17	45,51	33,86	0,83	0,95	0,61	526,64
8,00	44,96	34,97	0,64	0,83	0,61	528,19
6,12	46,07	31,08	0,87	0,96	0,61	528,08
5,45	47,73	29,42	0,70	0,85	0,62	527,97
6,33	48,29	27,20	0,72	0,87	0,62	528,86
8,18	44,96	23,31	0,86	0,98	0,62	529,41
8,05	46,62	29,42	0,87	0,94	0,62	529,08
8,16	49,40	25,53	0,80	0,93	0,61	528,69
8,47	50,51	26,09	0,97	1,10	0,61	528,53
9,60	52,73	26,09	0,87	1,03	0,61	528,69
9,78	52,73	23,31	1,14	1,19	0,61	528,86
10,11	51,62	22,76	0,81	0,89	0,61	529,41
7,98	49,95	25,53	0,79	0,96	0,61	529,58
6,73	48,29	27,75	0,98	1,05	0,62	528,30
7,06	49,95	26,09	0,84	0,99	0,62	528,53
9,67	49,40	28,31	0,74	0,84	0,62	529,30
9,78	48,29	24,98	0,84	0,93	0,61	529,86
8,97	46,07	31,64	0,78	0,90	0,61	529,80
9,37	47,73	32,19	1,17	1,27	0,61	529,91
7,70	47,18	25,53	1,17	1,20	0,61	529,69
8,08	46,62	22,76	0,74	0,84	0,62	530,03
8,52	46,07	22,20	0,84	1,00	0,62	530,86
8,83	44,96	24,98	0,98	1,05	0,62	530,19
7,95	47,18	26,09	0,88	0,95	0,62	529,58
8,56	47,73	23,87	0,81	0,96	0,62	530,25
8,03	47,18	29,97	0,90	0,99	0,62	529,36
8,18	46,62	28,86	0,82	0,92	0,62	529,03
9,04	47,73	27,20	0,95	1,05	0,62	529,58

A Flowchart was created to describe in detail all the phases of statistical analysis that will be necessary for the correct application of the Control Chart. In this article, the control chart will have the function of providing statistical monitoring of Variable Y (response variable). The detailed phases are described in Figure 2.





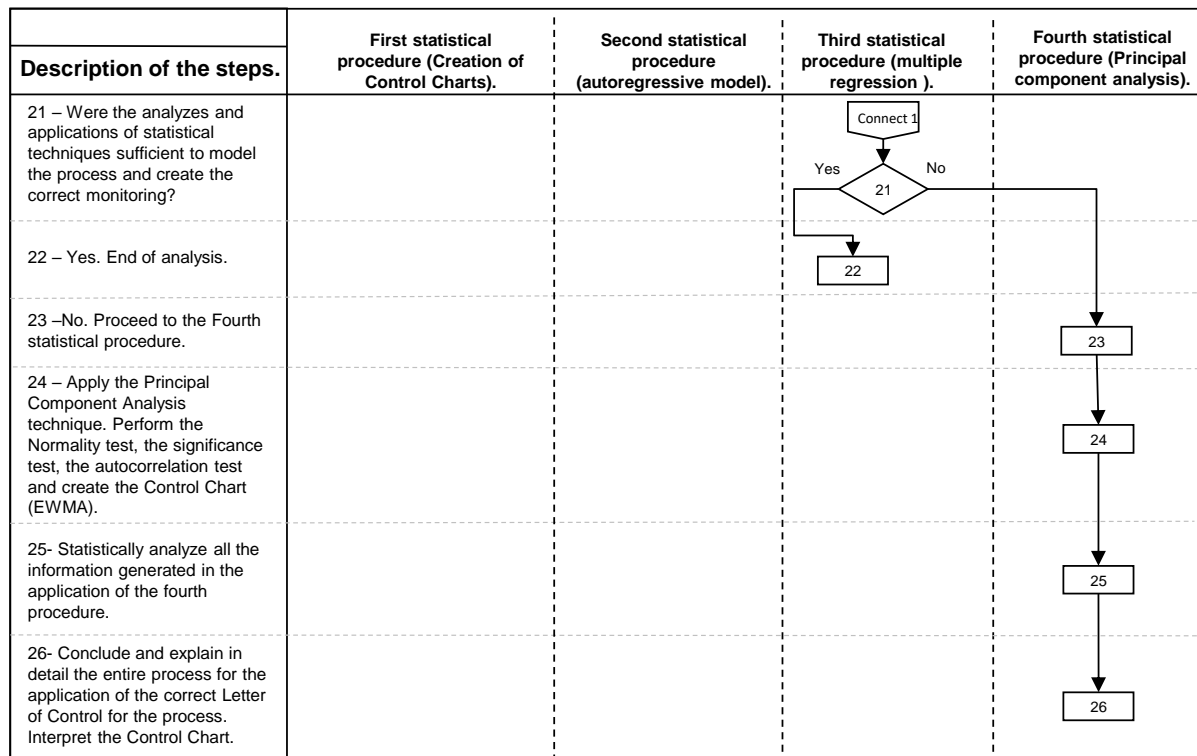


Figure 2. Flowchart of the phases of the statistical analysis used for the application of the control chart for the process

### 3.1 Normality Test of the Variable (Y) and Application of the Control Chart

First, the data Normality test was performed (on variable Y) and it was concluded that the data can be considered Normal, since the P-value is greater than 0.05, using 95% confidence (Figure 3).

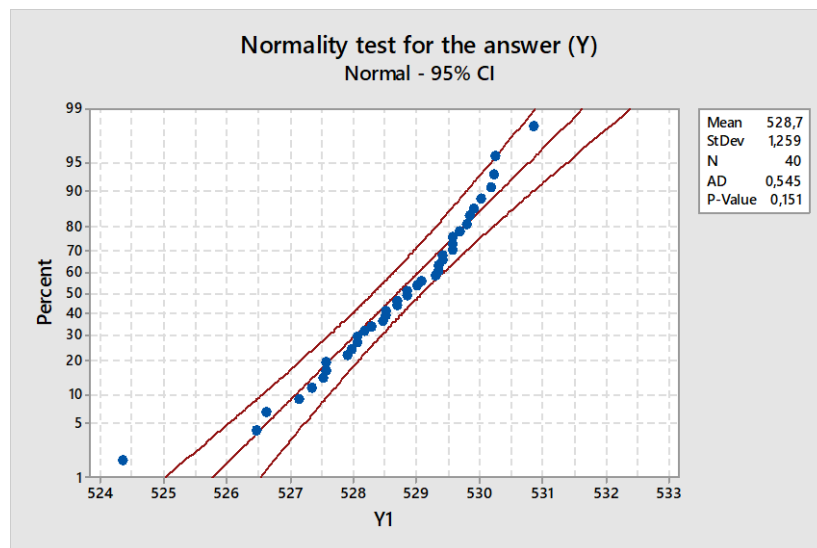


Figure 3. Data Normality Test (Variable Y)

Then, after verifying the Normality of the data (Y), which is a requirement for the application of the control charts, the control chart for individual values and moving range was applied (Figure 4). In this case, there is no subgroup within the sampling of this process, as the measurements are individual and a pre-established frequency was used to collect this data.

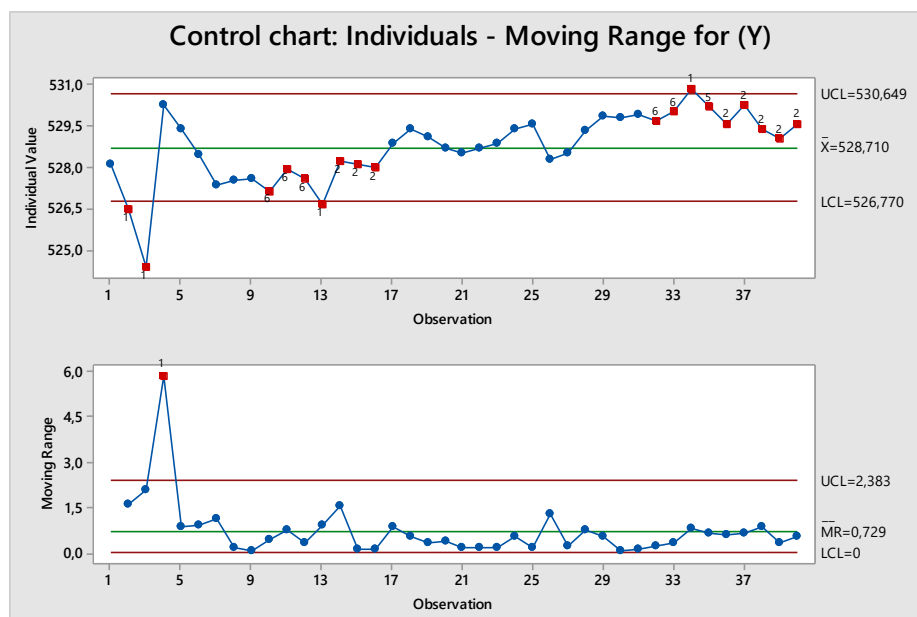


Figure 4. Control Chart for Individual Values and Moving Range (Variable Y)

Analyzing the control chart (Figure 4), it is possible to observe that the process is unstable, as there are many special causes of variation (anomalies), being possible to observe some points outside the control limits both in the individual value graph (X) and on the moving range (MR) graph. In addition, there are sequences of several points below the midline and also above the midline, which is characterized as the absence of randomness, according to the criteria adopted to analyze the Shewhart control charts. This is the conventional analysis normally performed by technicians within the industries. However, it is known that changing the configurations of industrial processes in large companies can cause the so-called “Super control”, which means the frequent alteration of the process adjustments seeking to improve their performance, which can cause even greater instability throughout the process, as changing the settings in processes requires a very detailed analysis, as this change requires the creation of new standard operating procedures and also validation by engineering. For this reason, the authors of this article decided to apply other statistical methods to verify whether this process is really unstable, as the preliminary analysis had revealed.

It is known that the Shewhart control charts, although very efficient for detecting special causes and for monitoring variations in the process, have some deficiencies. Two of these deficiencies will be cited below:

- Shewhart's control charts are efficient for detecting variations (in changes in the mean) from  $1.5 \sigma$  to more. However, studies prove that these control charts are not efficient for detecting changes in means in processes if the changes are less than  $1.5 \sigma$  (MONTGOMERY, 2004);
- Shewhart's control charts cannot be applied in self-correlated processes, because when there is autocorrelation in the data that will be submitted for monitoring, the value of the previous measurement can influence the result of the next measured value (in the later time) or in the next values. So, when there is evidence of the occurrence of data autocorrelation (over time), current scientific studies prove the inefficiency of Shewhart's control charts for monitoring the process. Scientists since 1960 have studied and proved the inefficiency of Shewhart's control charts for some situations and these scientists have created new control charts that can be more effective in these situations (MONTGOMERY, 2004). Shewart could not have predicted these flaws, as his control charts were created in 1920 and scientists discovered these flaws only about 40 years later.

Returning to the flow of studies demonstrated in this article. In the Figure, data (Y) were used for the application of the EWMA control chart. This procedure was performed only to confirm that the instability of the process really exists and that this instability can be perceived at any level of variation sigma ( $\sigma$ ) in relation to the displacement of the mean, since the EWMA control chart has a much higher sensitivity to detect variations in processes (MONTGOMERY, 2004). For this study the configurations of the parameters of the EWMA chart were  $K=3.0$  e  $\lambda=0.20$ . In these conditions, using the minimum sample size of 40 values, it will be possible to perceive a change in the process average between  $0.50 \sigma$  to  $0.75 \sigma$ , which indicates a very sensitive graph to detect the change in the process average.

The EWMA control chart (Figure 5) also demonstrated the existence of special causes of variation (points outside the control limits), which confirmed the instability of the process.

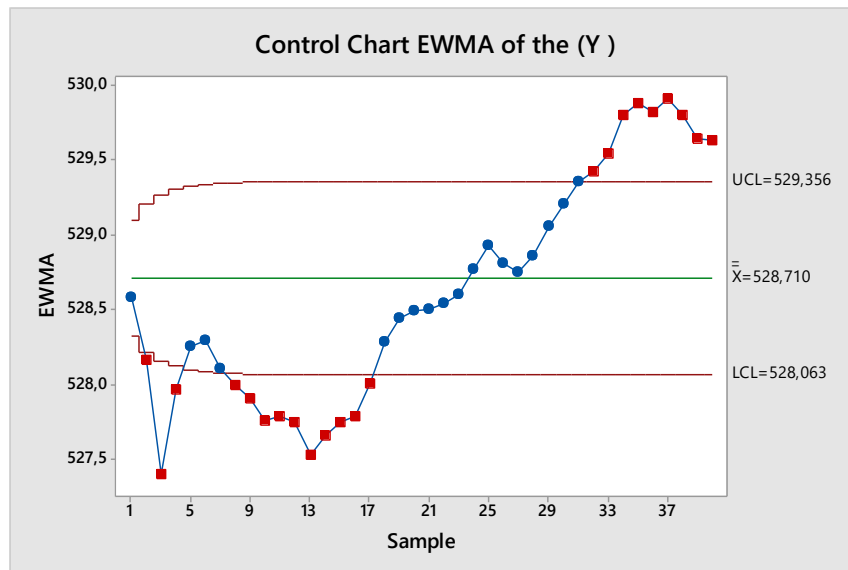


Figure 5. Application of the EWMA control card to the output variable (Y)

To test the second possibility (b, mentioned above), the data autocorrelation calculation (Y) was performed and the result showed that the original data have first-order autocorrelation, that is, the previous measurement, for a reason still unknown, influences and changes the result of the next measurement, rendering statistical analyzes incorrect using Shewhart's conventional control charts. It is possible to verify in the autocorrelation graph (Figure 6) that the first column indicates a 1st order autocorrelation and this implies that it will be necessary to apply forecasting models (AR1) to statistically model this process.

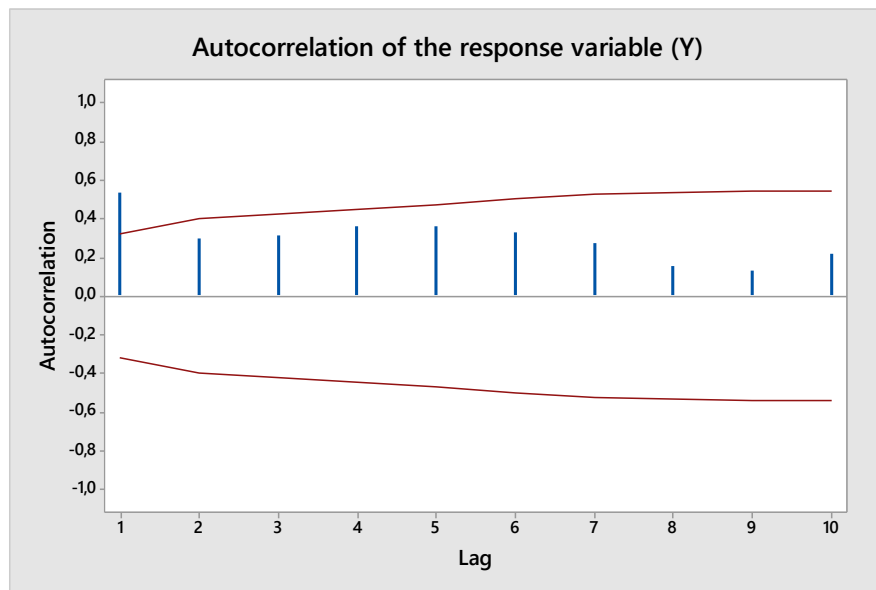


Figure 6. Autocorrelation of the output variable (Y)

#### 4. Creation of the First-Order Autoregressive Forecasting Model ( $AR_1$ )

Using the data contained in Table 3, an  $AR_1$  Model was created, according to the procedure described by Montgomery (2004).

Table 3. Output variable (Y) organized for the creation of the AR<sub>1</sub> model

(Y)	(Y - 1) = X
526,47	528,08
524,37	526,47
530,24	524,37
529,37	530,24
528,47	529,37
527,36	528,47
527,53	527,36
527,58	527,53
527,14	527,58
527,92	527,14
527,58	527,92
526,64	527,58
528,19	526,64
528,08	528,19
527,97	528,08
528,86	527,97
529,41	528,86
529,08	529,41
528,69	529,08
528,53	528,69
528,69	528,53
528,86	528,69
529,41	528,86
529,58	529,41
528,30	529,58
528,53	528,30
529,30	528,53
529,86	529,30
529,80	529,86
529,91	529,80
529,69	529,91
530,03	529,69
530,86	530,03
530,19	530,86
529,58	530,19
530,25	529,58
529,36	530,25
529,03	529,36
529,58	529,03

The AR<sub>1</sub> model was created based only on the variable (Y) in relation to a previous value of that same variable (Y), that is, the model describes what would be the influence that a previously measured value (in Y) would have on the next measurement (in Y). The autoregressive model was generated and, consequently, the residues of this AR<sub>1</sub> model were generated and the statistical test is shown in Table 4.

Table 4. Statistical test for the creation of the AR<sub>1</sub> model

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	241.80	73.30	3.300	0.002	
X	0.543	0.139	3.910	0.000	1.00
S	R-sq	R-sq(adj)			
1.08356	29.28%	27.37%			

Using the residues generated by the AR<sub>1</sub> model, the autocorrelation test was performed and as shown in Figure 7, it is

possible to verify that this procedure using the  $AR_1$  model neutralized the autocorrelation of the residues, as the values are within the specification limits established for the verification of the existence of autocorrelation.

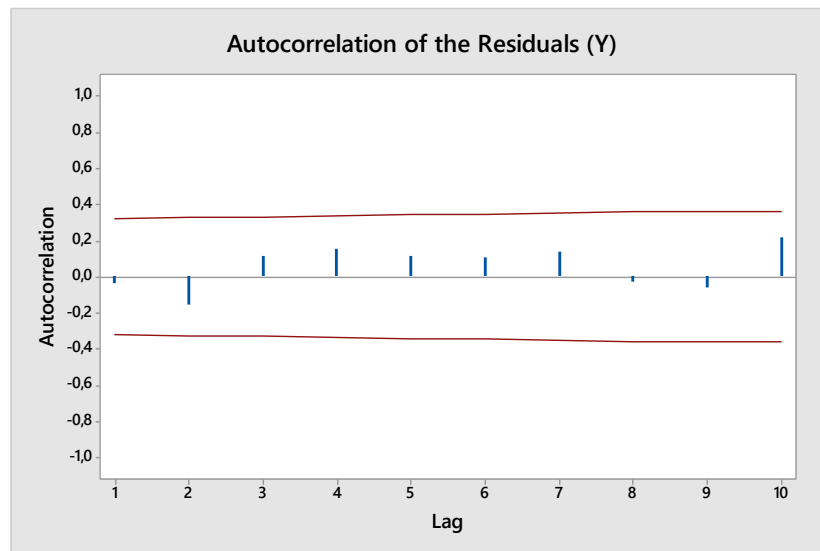


Figure 7. Autocorrelation of residues generated by the  $AR_1$  model

However, the  $AR_1$  model, despite eliminating autocorrelation, demonstrated that the residues are not Normal (Figure 8), with P-value  $<0.05$  (for 95% confidence), which makes it impossible to apply the Control Chart, as normality is a prerequisite for that application.

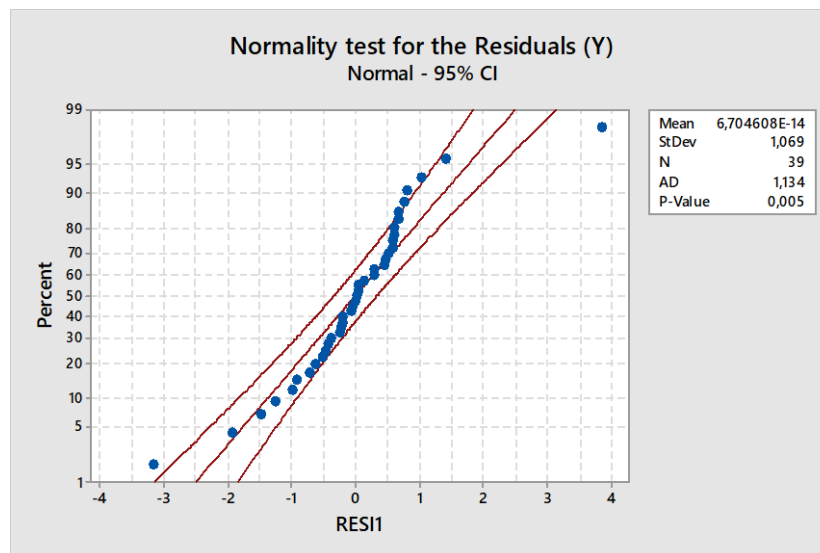


Figure 8. Normality test of residues generated by the  $AR_1$  model

#### 4.1 Multiple Regression Modeling

After realizing that the  $AR_1$  model used for the output variable (Y) did not allow the correct creation of the control chart, multiple regression was used (procedure recommended by Montgomery (2004)) to create a regressive model that will consider the relationship between variables (Xs) and variable (Y). From this modeling it was possible to obtain the residues and from these residues, it was possible to test normality as shown in Figure 9.

Then it was possible to create the model:

$$Y = 463.65 + 0.4261 X_1 - 0.1347 X_2 - 0.0957 X_3 - 3.99 X_4 + 5.35 X_5 + 112.7 X_6 \quad (14)$$

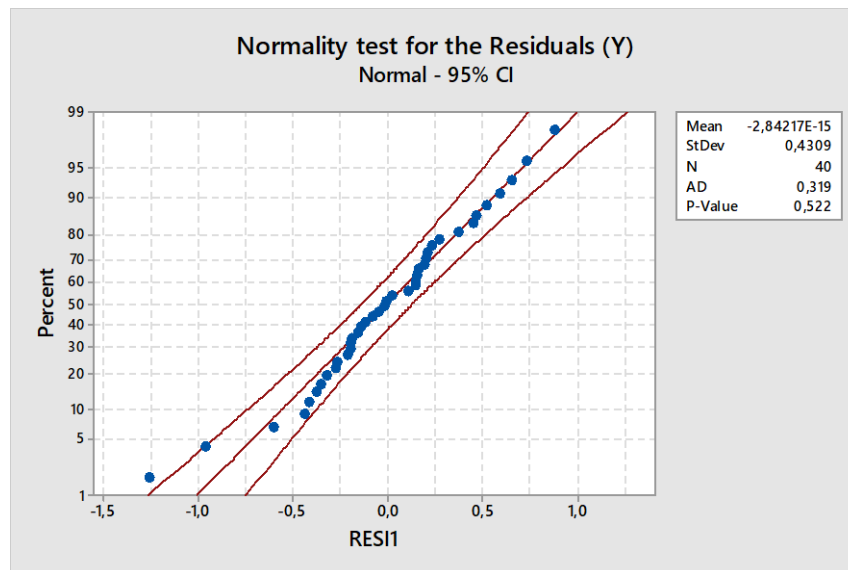


Figure 9. Residual normality test (by multiple regression model)

The residues were considered normal with P-Value  $> 0.05$  (for 95% confidence), meeting the normality criterion that is necessary for the creation of the control chart. In addition, the waste generated by this model is not autocorrelated, which meets yet another important criterion for the application of the correct control chart (Figure 10).

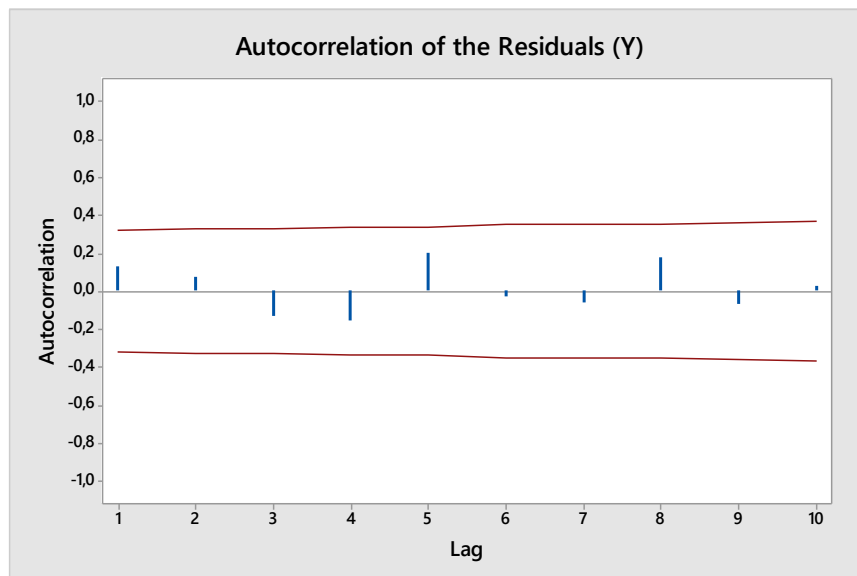


Figure 10. Autocorrelation test of residues (by multiple regression model)

After verifying the normality of the residues and the absence of autocorrelation, the EWMA control chart was used to monitor the process (Figure 11). In the EWMA chart, it was possible to observe that the process is stable, as there are no special causes of variation.

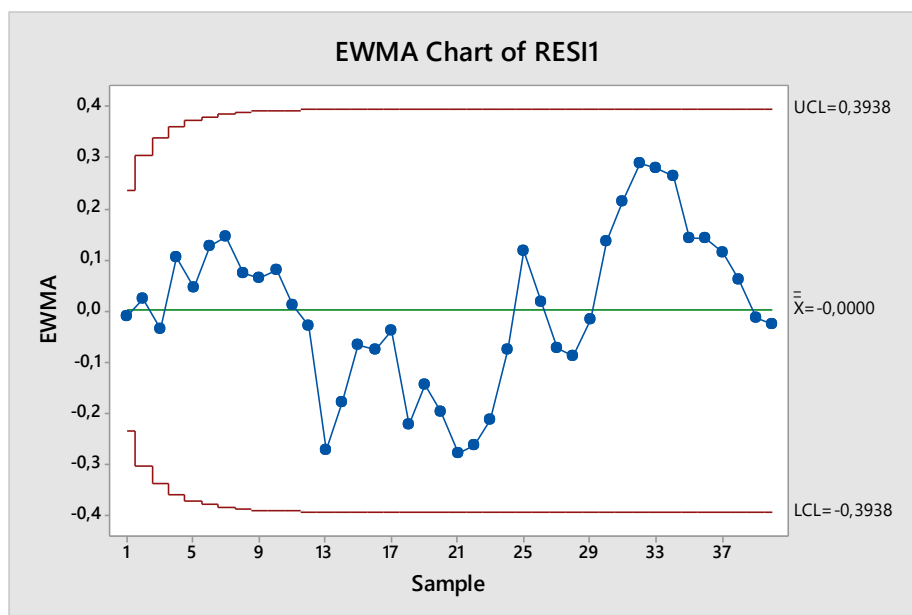


Figure 11. EWMA control chart (for waste)

When analyzing the statistical test generated by the model by multiple regression (Table 5), it was possible to verify the existence of multicollinearity between two factors considered influential (in this case, 90% confidence in the statistical test) and this consequently made this model unstable, as this model may generate unreliable residues, which will make it impossible to use these residues to create the appropriate control chart.

Analyzing the results of the statistical tests (Table 5) it is possible to verify that the factors  $X_4$  and  $X_5$  considered influential in this process (using 90% confidence), have a very high FIV (inflation factor) value ( $> 12$ ), this indicates the existence of high multicollinearity and models that have high multicollinearity cause errors in predictions and consequently the residues become unreliable to be monitored by a control chart. According to Montgomery (2004) when the factor is inflated above 5, the model is strongly influenced by multicollinearity. In this case, it will first be necessary to apply the Principal Component Analysis (multivariate statistics) to remove the multicollinearity, then perform another multiple regression with the components, to verify that the multicollinearity has been eliminated and only after that process, to be able to create a free model of that influence. In addition, it will be necessary to carry out the Normality test and the autocorrelation of the residues to verify that the data meet these criteria, and only then use the control chart.

Table 5. Statistical test for creating the multiple regression model

Term	Coef	SE Coef	T-Value	P-Value	FIV
Constant	463.650	8.100	57.240	0.000	
$X_1$	0.4261	0.0704	6.050	0.000	1.38
$X_2$	-0.1347	0.0351	-3.840	0.001	1.86
$X_3$	-0.0957	0.0214	-4.460	0.000	1.02
$X_4$	-3.990	2.090	-1.910	0.065	12.19
$X_5$	5.35	2.420	2.210	0.034	12.46
$X_6$	112.7	11.9	9.51	0.000	1.38
S	R-sq	R-sq(adj)			
0.4684	88.29%	86.16%			

## 5. Transformation to Principal Component Analysis

It is possible to standardize all the investigated variables by transforming them into a single unit of measurement (Z). Thus, all factors independently of their unit, will have a single unit of measurement to be transformed into main components by multiplying these variables standardized by the constants (PCs) calculated by the Software Minitab.

First were placed in a table the values of the adjustments of the investigated factors and the response variable (Y) and at the end of the Appendix A (in the last lines) were calculated the average and standard deviation of each column of factors, according to the procedure for transformation into main components described by Mingoti (2007).



Then all factor adjustment values were transformed into standardized variables (Appendix A). This transformation was performed by subtracting the factor adjustment value minus the mean value of this factor, then the resulting value was divided by the standard deviation of the respective factor (it is important to inform that only the factors will undergo this transformation, the response variable will not be transformed (Appendix B). For example, (1<sup>st</sup> line of Appendix B):

$$X_1 = (7.09 - 8.28) / 1.25 = -0.955$$

$$X_2 = (50.51 - 48.90) / 2.91 = 0.547$$

$$X_3 = (31.08 - 27.56) / 3.52 = 0.995 \quad (15)$$

All values transformed into a standardized variable regarding the factors are presented in Appendix B.

Then the eigenvalues and eigenvectors were created, which will be used to create the main components later (in Appendix B).

First the factors were standardized and then constants that will be used to transform the standardized values into main components were calculated.

The constants (PCs) correspond to the auto vectors, used for the transformation of the data into principal components, and are presented in Table 6.

Table 6. Auto vectors (PCs, Calculated using Minitab Software)

Variable	PC <sub>1</sub>	PC <sub>2</sub>	PC <sub>3</sub>	PC <sub>4</sub>	PC <sub>5</sub>	PC <sub>6</sub>
X <sub>1</sub>	0,381	0,147	-0,381	0,697	0,449	0,019
X <sub>2</sub>	0,440	0,440	-0,164	0,051	-0,744	-0,055
X <sub>3</sub>	-0,042	0,320	0,853	0,410	0,017	0,000
X <sub>4</sub>	0,545	-0,368	0,227	-0,131	0,083	-0,701
X <sub>5</sub>	0,551	-0,362	0,219	-0,117	-0,012	0,71
X <sub>6</sub>	-0,242	-0,623	-0,037	0,559	-0,488	-0,035

The conversion of the data to principal components must be done by the sum of the multiplication of each adjustment value of each factor (X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, X<sub>4</sub>, X<sub>5</sub>, X<sub>6</sub>), by the constants (PCs) referring to each factor (MINGOTI, 2007).

To apply multiple regression first it was necessary to multiply each of the calculated PCs constants (in Table 6) by the respective factors, that is, by the standardized values of the factors investigated in all experimental conditions.

Then, finally, the component values were calculated to later be created the mathematical model using the main components. Table 6 shows the values of the constants that were calculated. For example, in the first row in Appendix C, each PC<sub>1</sub> for each of the factors was multiplied (X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, X<sub>4</sub>, X<sub>5</sub>, X<sub>6</sub>) for each standard value of each factor in the first row (Table 6).

Example:

$$PC_1 = (0.381 \cdot -0.955) + (0.440 \cdot 0.547) + (-0.042 \cdot 0.995) + (0.545 \cdot -0.209) + (0.551 \cdot -0.206) + (-0.242 \cdot 0.594) = -0.536 \quad (16)$$

Appendix C shows the PC values and contains the response variable (Y) and from these values it was possible to use the multiple regression method and create the mathematical model by principal components.

Table 7 shows the statistics related to the principal components, for the purpose of choosing the components to be used in the mathematical model. The explanation of the criteria for choosing the variables will be detailed later.

Table 7. Main components for each experimental condition

Variable	PC <sub>1</sub>	PC <sub>2</sub>	PC <sub>3</sub>	PC <sub>4</sub>	PC <sub>5</sub>	PC <sub>6</sub>
Eigenvalue	2,4643	1,321 0	0,9948	0,813 1	0,3654	0,0414
Proportion	0,411	0,220	0,166	0,136	0,0610	0,0070
Cumulative	0,411	0,631	0,797	0,932	0,9930	1,0000

The Contents "Autovalues" contained in Table 7 is used to choose the main components, for the modeling (JOHNSON; WICHERN, 2007).

The Contents "Cumulative" defines which percentage of influence has each main component in the response (showing

in cumulative form). In this case,  $PC_1$  explains 41.1% of this phenomenon, the components  $PC_1 + PC_2$  explains 63.1%,  $PC_1 + PC_2 + PC_3$  explain 79.7%,  $PC_1 + PC_2 + PC_3 + PC_4$  explain 93.2%,  $PC_1 + PC_2 + PC_3 + PC_4 + PC_5$  explain 99.3%, while  $PC_1 + PC_2 + PC_3 + PC_4 + PC_5 + PC_6$  explain 100%. The "accumulated" index shown in Table 7 refers to the accumulated sum of the percentages of "explanation" of the phenomenon.

When there is intention to reduce the quantity of components, it is admitted to use the components that accumulate from 70% to 80% of the explanation of the phenomenon. The literature states that with a accumulated within this range it is possible to represent the process well and reduce the amount of variables. However, for this study all components (PCs) will be used, since the application of principal component analysis was only with the objective of create a mathematical model that could make high hit predictions in this process.

A significance test was performed to verify that the principal components and the response variable have a significant relationship and the results revealed that all components are significant (Table 8). Then the mathematical model was created using the main components.

Table 8. Significance test for PCs

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	528.710	0.0740	7139.04	0.000	
$PC_1$	-0.1187	0.0478	-2.4800	0.018	1.00
$PC_2$	-0.7761	0.0662	-11.7200	0.000	1.00
$PC_3$	-0.4383	0.0752	-5.8300	0.000	1.00
$PC_4$	0.6786	0.0832	8.1600	0.000	1.00
$PC_5$	0.0850	0.1240	0.6800	0.499	1.00
$PC_6$	0.7690	0.3690	2.0900	0.045	1.00
S	R-sq	R-sq(adj)			
0.4684	88.29%	86.16%			

Then it was possible to create the model:

$$Y = 528.710 - 0.1187 PC_1 - 0.7761 PC_2 - 0.4383 PC_3 + 0.6786 PC_4 + 0.085 PC_5 + 0.769 PC_6 \quad (17)$$

Analyzing the results of Table 8, it is possible to observe that the application of the principal component analysis methodology was able to remove the multicollinearity of the statistical model, lowering the FIV value from 12.0 to 1.0 and with a value of FIV 1.0, the multicollinearity will no longer influence the statistical model (as described by Montgomery (2004)). Furthermore, the normality analysis of the residuals found that the residuals follow the normal probability distribution, which is fundamental for the use of a mathematical model to make predictions. Figure 12 shows the residual normality graph (with P-value > 0.05 for 95% confidence).

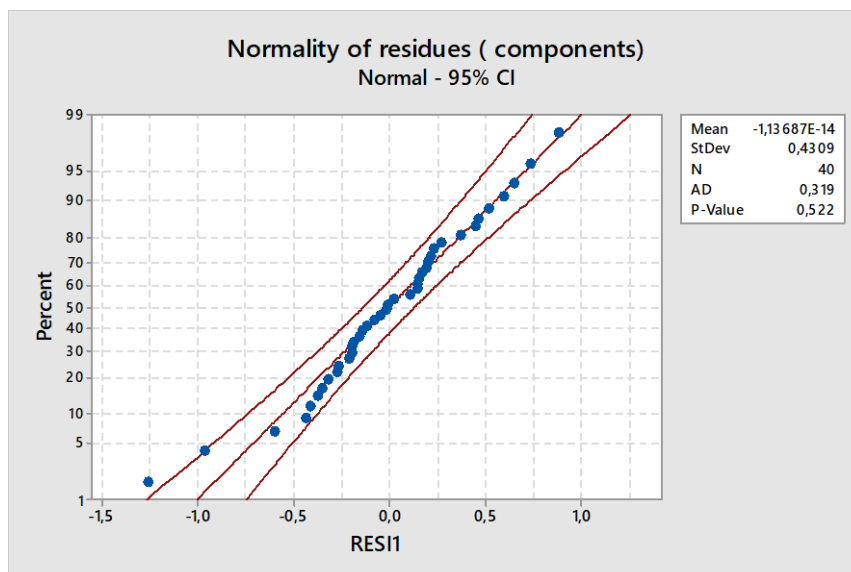


Figure 12. Normality test of residues of the mathematical model (Model by Principal Component Analysis)

As shown in Figure 13, the residues generated by the model created from the main components, in addition to eliminating multicollinearity and being normally distributed, the model also eliminated the autocorrelation of data over time, thus allowing the correct application of the control chart for statistical monitoring.

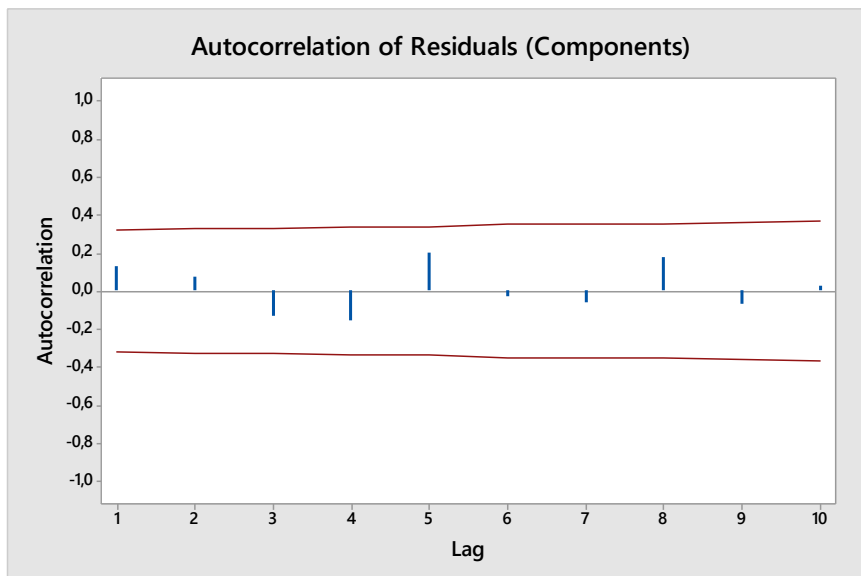


Figure 13. Autocorrelation test of residues (Model by Principal Component Analysis)

Then, the EWMA control card was used (Figure 14). For this study, the parameter settings of the EWMA chart were  $K = 3.0$  and  $\lambda = 0.20$ . In these conditions, with the minimum sample size of 40 values, it will be enough to perceive a change in the process average between  $0.50 \sigma$  to  $0.75 \sigma$ , which indicates a very sensitive graph for the change. In this case, the process was considered stable, without the presence of special causes of variation (Figure 14).

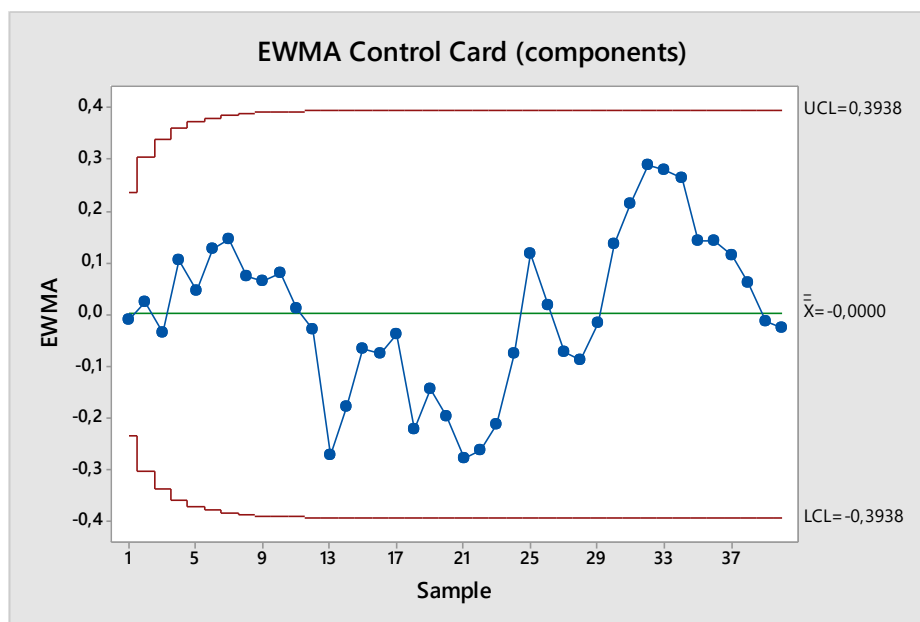


Figure 14. EWMA waste control chart (Model by Principal Component Analysis)

The result of this work was to demonstrate, through statistical procedures, that the incorrect choice of the control chart for process monitoring can lead to dangerous assumptions that, consequently, may influence managers' decision-making. It is concluded that in industrial processes, wrong decisions can bring disastrous consequences, as changes in processes can incorporate non-existent variations, unnecessary adjustments and, with that, cause an increase in the number of defective products.

## 6. Conclusion

With this study it was possible to demonstrate that the data of the original process (Y) that seemed to have many special causes, when using the control chart of individual value (X) and mobile amplitude (MR), is actually a stable process

(without the presence of special causes). However, it was necessary to carry out a careful statistical analysis process to reach this conclusion. Only after the use of multivariate statistics (Principal Component Analysis) combined with the use of other statistical methods, was it possible to eliminate the autocorrelation, multicollinearity and obtain the normality of the model residues, thus allowing the correct application of the process monitoring in the output variable (Y). Then, the appropriate control card was chosen, which in this case was the EWMA (Exponentially Weighted Moving Average).

This article was based on a real situation in a steel heat treatment process to demonstrate that the simple application of conventional control charts in production processes, may in some cases bring misinterpretations about the true situation of the process. It was concluded in this study, that these supposedly incorrect decisions, based on the monitoring data using control charts, may cause successive changes in the process settings in order to stabilize the process. However, these changes in the process, in this case, will not bring stability to the process (since the process is already under statistical control). In fact, this may have the opposite effect, making the process unstable and causing several problems, such as, waste of materials, wasted labor time, costs with analysis in laboratories and carrying out experiments. In addition, changes in processes may incorporate variations that did not previously exist and this would lead to an increase in the number of defective products, which could lead to increased costs for rework, scrap and this could also increase the number of customer complaints.

## References

- Benyounis, K. Y., & Olabi, A. G. (2008). Optimization of different welding processes using statistical and numerical approaches – A reference guide. *Science Direct*, 39, 483-496. <https://doi.org/10.1016/j.advengsoft.2007.03.012>
- Berthouex, P. M., & Hunter, W. G., & Pallesen, L. (1978). Monitoring sewage treatment plants: some quality control aspects. *Journal of Quality Technology*, 10(4), 139-149. <https://doi.org/10.1080/00224065.1978.11980842>
- Castillo, E. D. (2002). *Statistical Process Adjustment for Quality Control* (1st ed.). John Wiley & Sons (New York).
- Claro, F. A. E. (2008). Gráficos de controle de X para monitoramento de processos autocorrelacionados, Universidade do Estado de São Paulo, Guaratinguetá - Brazil, 31-95.
- Costa, A. F. B., & Epprecht, E. K., Carpinetti, L. C. R. (2005). *Controle Estatístico de Qualidade* (2nd ed.). Atlas (São Paulo).
- Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6st ed.). New Jersey: Prentice Hall (New York).
- Machado, M. A. G. (2006). Estudo das propriedades dos gráficos de controle bivariados com amostragem dupla, Universidade do Estado de São Paulo, Guaratinguetá - Brazil, 20-60.
- Mingoti, S. A. (2007). Análise de dados através de métodos de estatística multivariada, uma abordagem aplicada (1st ed.). UFMG (São Paulo).
- Myung, J., & Pitt, M. A., & Kim, W. (2003). Model evaluation, testing and selection, Ohio State University, Department of Psychology, Columbus. *Handbook of Cognition*, 1, 1-45.
- Montgomery, C. D. (2013). *Design and analysis of experiments* (8st ed.). John Wiley & Sons (New York).
- Montgomery, D. C., & Runger, G. C. (2003). *Estatística aplicada e probabilidade para engenheiros* (2nd ed.). LTC (New York).
- Montgomery, D. C., & Runger, G. C. (2009). *Estatística aplicada e probabilidade para engenheiros* (4st ed.). LTC (New York).
- Montgomery, C. D. (2004). *Introdução ao controle estatístico da qualidade* (4st ed.). LTC (New York).
- Ramos, A. W. (2005). CEP para processos contínuos em bateladas (2st ed.). Edgard Blucher (São Paulo).
- Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics*, 1(3), 239-250. <https://doi.org/10.1080/00401706.1959.10489860>
- Schneider, H., & Pruett, J. M. (1994). Control charting issues in the process industries. *Quality Engineering*, 6(3), 347-373. <https://doi.org/10.1080/08982119408918733>

Appendix A. Original table with variables (Xs) and variable (Y)

X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	Y
7,09	50,51	31,08	0,85	0,97	0,62	528,08
6,31	49,95	27,20	0,85	0,95	0,60	526,47
6,77	49,95	22,76	0,82	0,91	0,58	524,37
8,30	49,40	23,87	0,85	1,07	0,61	530,24
8,23	46,07	23,31	0,93	1,03	0,61	529,37
7,19	46,62	29,97	0,83	0,93	0,61	528,47
7,14	49,40	33,86	0,91	1,01	0,61	527,36
10,72	54,95	33,30	0,81	0,94	0,61	527,53
10,41	54,95	28,86	1,05	1,12	0,60	527,58
9,43	54,39	31,64	0,92	1,01	0,60	527,14
10,10	54,39	27,20	0,92	1,07	0,60	527,92
8,99	53,84	28,86	1,20	1,30	0,60	527,58
8,17	45,51	33,86	0,83	0,95	0,61	526,64
8,00	44,96	34,97	0,64	0,83	0,61	528,19
6,12	46,07	31,08	0,87	0,96	0,61	528,08
5,45	47,73	29,42	0,70	0,85	0,62	527,97
6,33	48,29	27,20	0,72	0,87	0,62	528,86
8,18	44,96	23,31	0,86	0,98	0,62	529,41
8,05	46,62	29,42	0,87	0,94	0,62	529,08
8,16	49,40	25,53	0,80	0,93	0,61	528,69
8,47	50,51	26,09	0,97	1,10	0,61	528,53
9,60	52,73	26,09	0,87	1,03	0,61	528,69
9,78	52,73	23,31	1,14	1,19	0,61	528,86
10,11	51,62	22,76	0,81	0,89	0,61	529,41
7,98	49,95	25,53	0,79	0,96	0,61	529,58
6,73	48,29	27,75	0,98	1,05	0,62	528,30
7,06	49,95	26,09	0,84	0,99	0,62	528,53
9,67	49,40	28,31	0,74	0,84	0,62	529,30
9,78	48,29	24,98	0,84	0,93	0,61	529,86
8,97	46,07	31,64	0,78	0,90	0,61	529,80
9,37	47,73	32,19	1,17	1,27	0,61	529,91
7,70	47,18	25,53	1,17	1,20	0,61	529,69
8,08	46,62	22,76	0,74	0,84	0,62	530,03
8,52	46,07	22,20	0,84	1,00	0,62	530,86
8,83	44,96	24,98	0,98	1,05	0,62	530,19
7,95	47,18	26,09	0,88	0,95	0,62	529,58
8,56	47,73	23,87	0,81	0,96	0,62	530,25
8,03	47,18	29,97	0,90	0,99	0,62	529,36
8,18	46,62	28,86	0,82	0,92	0,62	529,03
9,04	47,73	27,20	0,95	1,05	0,62	529,58
8,2887	48,9094	27,5696	0,8809	0,9939	0,6116	Average
1,2515	2,9157	3,5267	0,1452	0,1093	0,0074	Standard deviation

## Appendix B. Transformation of factors into standardized variables

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$
-0,955	0,547	0,995	-0,209	-0,206	0,594
-1,581	0,356	-0,106	-0,209	-0,409	-1,561
-1,213	0,356	-1,365	-0,519	-0,765	-4,247
0,006	0,166	-1,050	-0,209	0,706	0,072
-0,046	-0,975	-1,207	0,366	0,351	0,072
-0,878	-0,785	0,680	-0,431	-0,562	-0,225
-0,917	0,166	1,782	0,233	0,148	-0,449
1,940	2,070	1,624	-0,564	-0,511	-0,747
1,692	2,070	0,365	1,342	1,163	-1,343
0,911	1,879	1,152	0,322	0,148	-1,194
1,448	1,879	-0,106	0,322	0,656	-0,971
0,561	1,689	0,365	2,538	2,787	-0,971
-0,095	-1,165	1,782	-0,431	-0,359	-0,747
-0,228	-1,356	2,096	-1,893	-1,475	-0,598
-1,736	-0,975	0,995	-0,120	-0,308	-0,002
-2,268	-0,404	0,523	-1,450	-1,272	0,743
-1,563	-0,214	-0,106	-1,317	-1,120	0,892
-0,086	-1,356	-1,207	-0,165	-0,105	0,892
-0,192	-0,785	0,523	-0,076	-0,511	0,519
-0,099	0,166	-0,578	-0,608	-0,612	0,445
0,144	0,547	-0,420	0,677	0,960	0,221
1,048	1,308	-0,420	-0,076	0,351	-0,300
1,190	1,308	-1,207	2,051	1,772	-0,449
1,452	0,927	-1,365	-0,564	-0,917	-0,449
-0,245	0,356	-0,578	-0,741	-0,308	-0,151
-1,243	-0,214	0,051	0,765	0,554	0,743
-0,982	0,356	-0,420	-0,298	-0,003	0,743
1,102	0,166	0,208	-1,140	-1,424	0,668
1,195	-0,214	-0,735	-0,342	-0,562	0,072
0,547	-0,975	1,152	-0,785	-0,866	0,221
0,862	-0,404	1,310	2,273	2,483	0,221
-0,472	-0,594	-0,578	2,273	1,874	0,296
-0,166	-0,785	-1,365	-1,096	-1,424	0,817
0,184	-0,975	-1,522	-0,298	0,097	0,892
0,432	-1,356	-0,735	0,765	0,554	1,041
-0,272	-0,594	-0,420	-0,032	-0,359	0,817
0,219	-0,404	-1,050	-0,564	-0,308	0,966
-0,206	-0,594	0,680	0,145	-0,054	1,190
-0,086	-0,785	0,365	-0,519	-0,663	0,743
0,596	-0,404	-0,106	0,588	0,503	0,519

## Appendix C. Main components for each experimental condition

PC <sub>1</sub>	PC <sub>2</sub>	PC <sub>3</sub>	PC <sub>4</sub>	PC <sub>5</sub>	PC <sub>6</sub>	Y
-0,536	0,200	1,008	0,153	-1,124	-0,069	528,080
-0,403	1,088	0,373	-1,924	-0,227	-0,139	526,470
0,074	2,656	-0,889	-3,603	1,205	-0,072	524,370
0,377	-0,485	-0,821	-0,432	-0,199	0,637	530,240
-0,020	-1,129	-0,694	-0,626	0,675	0,042	529,370
-1,198	0,245	0,831	-0,377	0,282	-0,062	528,470
-0,033	0,648	1,944	-0,199	-0,268	-0,069	527,360
1,173	2,574	0,094	1,840	-0,317	-0,018	527,530
3,237	1,198	-0,062	0,371	-0,021	-0,149	527,580
1,672	1,901	0,477	0,476	-0,361	-0,164	527,140
2,155	1,254	-0,697	0,399	-0,257	0,197	527,920
4,096	-0,395	1,044	-0,574	-0,347	0,151	527,580
-0,876	0,797	1,598	0,285	1,188	0,135	526,640
-2,472	1,644	1,367	0,717	1,094	0,371	528,190
-1,367	-0,208	1,575	-0,800	-0,042	-0,113	528,080
-2,735	0,187	0,741	-0,632	-1,176	0,066	527,970
-2,236	-0,023	-0,037	-0,341	-1,075	0,079	528,860
-0,942	-1,452	-0,868	-0,091	0,501	0,082	529,410
-0,890	-0,316	0,499	0,400	0,252	-0,288	529,080
-0,717	0,042	-0,771	0,102	-0,438	-0,035	528,690
1,158	-0,607	-0,147	-0,121	-0,412	0,172	528,530
1,217	0,683	-0,902	0,426	-0,373	0,261	528,690
3,283	-0,752	-0,828	-0,325	-0,091	-0,213	528,860
0,314	1,004	-2,182	0,430	0,122	-0,263	529,410
-0,449	0,4144	-0,688	-0,341	-0,369	0,281	529,580
-0,027	-1,206	0,820	-0,606	-0,704	-0,180	528,300
-0,543	-0,473	-0,139	-0,383	-1,101	0,141	528,530
-1,084	0,820	-0,865	1,552	-0,029	-0,223	529,300
-0,121	0,130	-1,251	0,671	0,626	-0,127	529,860
-1,228	0,485	0,558	1,132	0,828	-0,008	529,800
2,649	-1,505	1,906	0,653	0,761	0,200	529,910
1,782	-2,215	0,699	-0,948	0,242	-0,249	529,690
-1,931	-0,397	-1,563	0,051	0,013	-0,231	530,030
-0,619	-1,370	-1,288	-0,019	0,321	0,304	530,860
0,069	-1,899	-0,313	0,347	0,739	-0,096	530,190
-0,761	-0,804	-0,273	0,110	-0,084	-0,233	529,580
-0,761	-0,764	-1,144	0,352	-0,133	0,169	530,250
-0,607	-0,849	0,733	0,757	-0,206	-0,153	529,360
-1,222	-0,272	0,182	0,611	0,153	-0,091	529,030
0,526	-0,846	-0,026	0,506	0,356	-0,039	529,580

**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).