Application of Statistical Control Charts to Monitor Turbidity of Drinking Water

Araujo, Nillo Gabriel Alves
University of Santa Catarina State, Department of Production Engineering
Rua Paulo Malschitzki, s/numero
Joinville (89219-710), Brazil
E-mail: nillo.araujo@yahoo.com.br

Henning, Elisa
University of Santa Catarina State, Department of Mathematics
Rua Paulo Malschitzki, s/numero
Joinville (89219-710), Brazil
E-mail: dma2eh@joinville.udesc.br

Alves, Custodio da Cunha
University of Joinville Region, Department of Production Engineering
Rua Paulo Malschitzki, s/numero
Joinville (89219-710), Brazil
E-mail: custodio.alves@univille.net

Zvirtes, Leandro
University of Santa Catarina State, Department of Production Engineering
Rua Paulo Malschitzki, s/numero
Joinville (89219-710), Brazil
E-mail: zvirtes@joinville.udesc.br

Statistical Process Control (SPC) is an efficient and powerful formal technique used for monitoring processes. It consists of a collection of tools for resolving problems useful in obtaining a stable process. The control chart, which originated in the industrial environment, is one such tool, and probably the most technically sophisticated [6].

This paper provides an analysis of the application of statistical control charts for the monitoring of turbidity, an indicator of the quality of drinking water. Because it is the quality of the water consumed by the population, there is a need for a rapid response when there is some deviation, no matter how small. This work arose from studies of the need to continuously monitor water quality indicators. Even if the indicators show values within the limits of the specification, to improve quality, there is a need to assess the variability. As the variability only can be expressed in terms of statistics, control charts have a special role for this purpose [12]. When a process has only random variation, it is said to be under statistical control. However, apart from random causes, processes can undergo the action of special and structural causes, resulting in the occurrence of an extra variation that must be detected.

The application of classical control charts assumes that the quality characteristic data of the process are independent and identically distributed [14]. If these assumptions are violated, the performance of the control chart may be insufficient, losing efficiency in detecting special causes [1]. Autocorrelation is not uncommon in this type of measurement, which limits the direct application of conventional statistical control charts, requiring application of other techniques. One possible solution is to adjust time series models to the data, such as Autoregressive Moving Average (ARIMA), and then apply the control charts to the residuals of the model, which is done in this paper.

In this paper, we cover only one indicator - turbidity, the state of water that is caused by the presence of finely divided particulate or colloidal matter, and microscopic organisms. According to existing legislation [3], the maximum value for turbidity in any sample point should be five turbidity units (NTU). Natural turbidity does not cause direct health problems. However, it is aesthetically unpleasant in drinking water, and suspended solids can provide shelter for pathogenic microorganisms. However, turbidity of anthropogenic origin may be associated with toxic compounds and pathogenic organisms, causing more serious problems.

Statistical Control Charts

A control chart is a visual statistical tool that highlights the presence of special causes. It basically consists of plotting lines representing the Upper Control Limit (UCL) and Lower Control Limit (LCL), the mean or target of the process (Central Line - CL) and the observed points, which represent, in general, a statistic related to the variable of interest. If one or more points are beyond the control limits, it indicates that the process can be out of statistical control, i.e. there may be a problem in the process [10].

The performance of a control chart is typically evaluated through parameters related to the distribution of time required for the chart to emit a signal. The Average Run Length (ARL) is one of these parameters. In a control chart, ARL_0 indicates the average number of samples collected until the emission of a signal during the period under control, while ARL_1 represents the average number of samples collected until the emission of a signal that indicates an out-of-control situation [2].

The most popular control chart for variables is the Shewhart control chart, which achieved success due to its simplicity. The decision rule is based only on examination of the last point: if it is beyond the control limits, it must intervene in the process. This is also a disadvantage, because it ignores any information given by the preceding sequence of points. This makes the Shewhart chart relatively insensitive to small changes in the process, 1,5σ or less [6]. The Cumulative Sum (CUSUM) control chart and the Exponentially Weighted Moving Average (EWMA) chart were developed to correct this deficiency, and are indicated for the monitoring of processes subject to minor deviations. These two charts help in decision making, because they are based on the results presented by a number of samples rather than isolated observations, i.e., the analysis of this chart is a function of current and previous results.

Typical control charts can only be effective if the observations at different times are independent and normally distributed. Violation of these assumptions implies a reduction in the performance of the chart, generating more false positives [4][6]. When the observations are autocorrelated, two approaches, in general, have been applied in several studies. One approach is to apply a control chart with extended limits. The other involves adjusting the data to a time series model and monitoring their residuals [13]. This model is used in this paper. It is also possible to solve the problem of autocorrelation with a less frequent sampling, but this can cause an increase in time to detect a real change in the process [6].

The application of control charts consists of two phases. Phase I is a retrospective study of samples from the process being monitored, where the necessary checks are carried out and parameters estimated. In Phase II, information obtained in Phase I is used in the construction of control charts that are used to test whether the process remains under control when the future observations are monitored [12] [10]. In this study, we only executed Phase I, and the results can be used for future monitoring in order to keep this variable in a state of statistical control (Phase II).

CUSUM and EWMA Control Charts

The CUSUM chart was first proposed by [7], and the decision about the state of the process is based on the information accumulated in the previous samples. If the process is under statistical control, the cumulative sum describes a random curve around the mean. However, if the mean changes for some value above (or below) the nominal value, then an increasing (or decreasing) trend appears in the cumulative sum. The CUSUM works accumulating deviations through the statistics $C_i^+ = \max[0, x_i - (\mu_0 + k) + C_{i-1}^+]$

(upper CUSUM) and $C_i^- = \max[0, (\mu_0 - K) - x_i + C_{i-1}^-]$ (lower CUSUM), with $C_0^+ = C_0^- = 0$. The reference value K is chosen between the nominal value μ_0 and the μ_1, mean out of statistical control value, that we are interested in detecting. So, if the changes are expressed in terms of standard deviations $\mu_1 = \mu_0 + \delta\sigma$, then the value of K is the half of the magnitude of the change $K = (\delta/2)\sigma = |\mu_1 - \mu_0| / 2$.

If C_i^+ or C_i^- exceeds the decision limit H, we say that the process is out of statistical control. The parameters K and H are estimated in order to establish good ARL performance. A reasonable value for H is five times the standard deviation of the process. An ARL_0 = 370 results in K = 0,5 and H = 4,77 [6].

The EWMA chart [9] is similar to the CUSUM. The main difference is that the EWMA provides greater weight to more current information and less weight to older information. The exponentially weighted moving average chart is defined by $Z_i = \lambda X_i + (1-\lambda)Z_{i-1}$ where X_i is the most recently observed value, λ is the parameter that corresponds to the weight given the observations, usually 0,1 or 0,2, and the initial value is the nominal value. Assuming that the observations x_i are independent random variables with variance σ^2, then the variance of z_i is $\sigma_{z_i}^2 = \sigma^2(\lambda/(2-\lambda))(1-(1-\lambda)^{2i}$. The central line is CL = 0 and for calculating the control limits, the equations $UCL = \mu_0 + L\sigma\sqrt{(\lambda/(2-\lambda))(1-(1-\lambda)^{2i}}$ and $LCL = \mu_0 - L\sigma\sqrt{(\lambda/(2-\lambda))(1-(1-\lambda)^{2i}}$ are applied, where L is the width of the control limits. A frequent value for L is 3(three) (conventional three-sigma limits). Being a chart that uses a weighted average of samples, this chart is more robust for the normality assumption [6].

Methodology

This analysis was done with data from the year 2009, from a water treatment plant in the city of Joinville, southern Brazil. Daily, every half hour a sample is taken in order to verify the level of residual chlorine, fluorine, and turbidity, among others. The values taken from the samples are verified according to the parameters of the law. Periodic reports are made available to the local people. For Phase I, 37 samples taken from the plant exit were analyzed, corresponding to a complete day in January 2009. The Jarque Bera test was applied to verify the assumption of normality, and autocorrelation and partial autocorrelation function charts were applied to check the autocorrelation. We used the GNU R(version 2.10.1)[8] for all the analyses, with the contributed packages *forecast*[5] for the ARIMA model, and the package *qcc*[11] for the control charts. The nominal value, in this case, must meet the standards, as the maximum allowed is 1(one) NTU.

Results and Discussion

Fig.1a shows the time-series plot for turbidity. The examination of the plot in Fig. 1a reveals that the behavior of turbidity is not random, indicating that the values of turbidity seem to depend on the previous values. As we can see in Fig. 1b, autocorrelation is present for the turbidity data. To check the normality, the Jarque Bera test was applied, and the data can be considered normally distributed (p =0,3053). Time dependence is not rare in such measurements, and it has a result, an apparent trend that depends on the time parameter. Therefore, typical SPC control charts are not appropriate, due to presence of autocorrelation [12].

An ARIMA (1,1,0) model was fitted and a diagnosis check was used to validate the model. The residuals of the model were independent and normally distributed (p-value = 0,8126). The mean absolute percentage error (MAPE) was 7,15%.

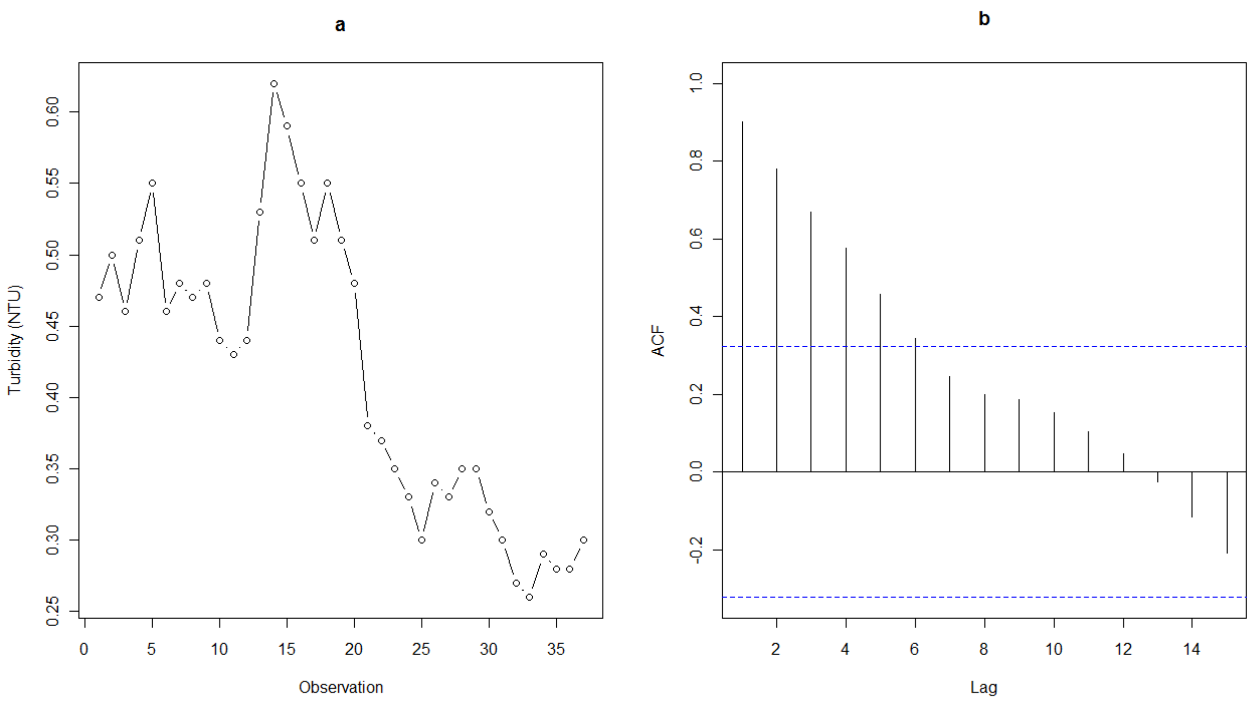

Figure 1: Time series plot (a) and autocorrelation function (b) for turbidity.

We applied a Shewhart control chart on the residuals from the model (Fig 2b). One can see that there are no points beyond the control limits, which occurs when applying the conventional chart to the observed data (Fig 2a). The process is under statistical control and the presence of autocorrelation leads to an indication of more false positives. If autocorrelation were not observed, it could result in unnecessary stops in the process.

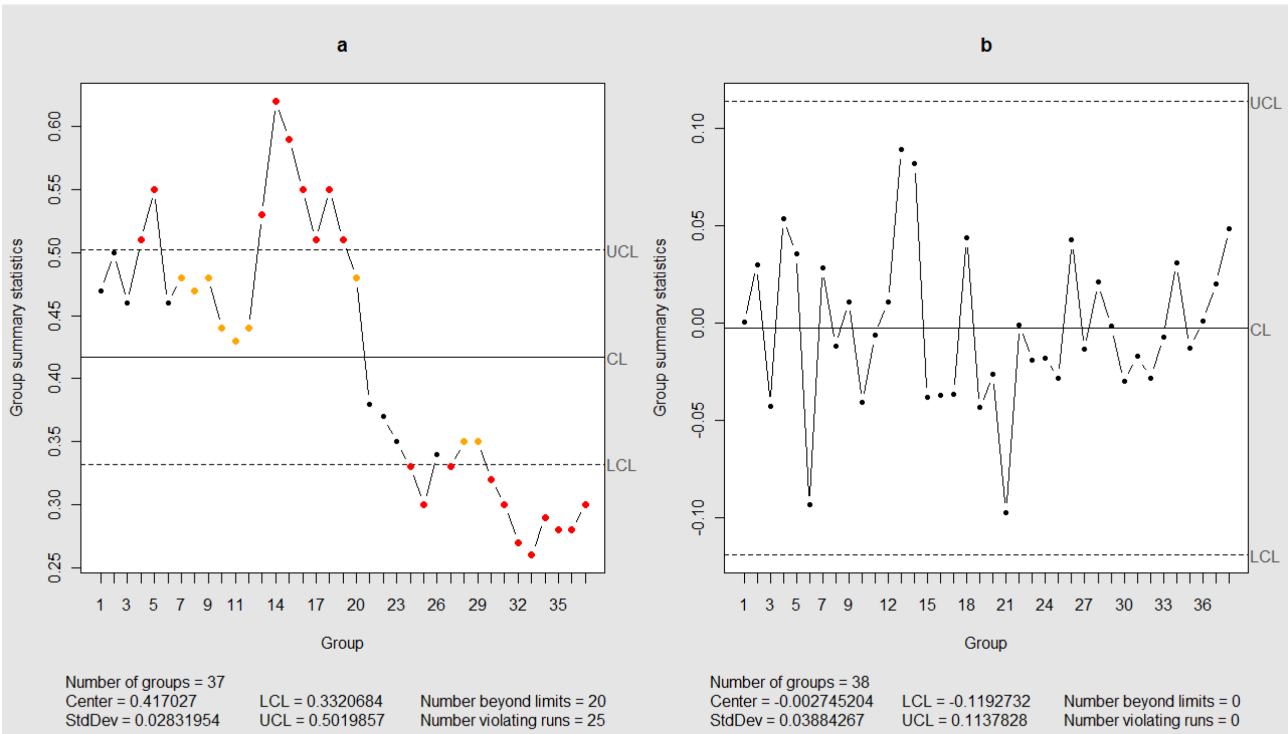

Figure 2: Shewhart Control Charts, applied to turbidity data (a) and to the residuals from the ARIMA model (b).

We applied a CUSUM chart to the residuals of the ARIMA model and this also indicates a control condition (Figure 3b). When the CUSUM chart was applied on the observed data, many false alarms were also indicated (Figure 3a). Although not shown here, the same occurs with the EWMA chart.

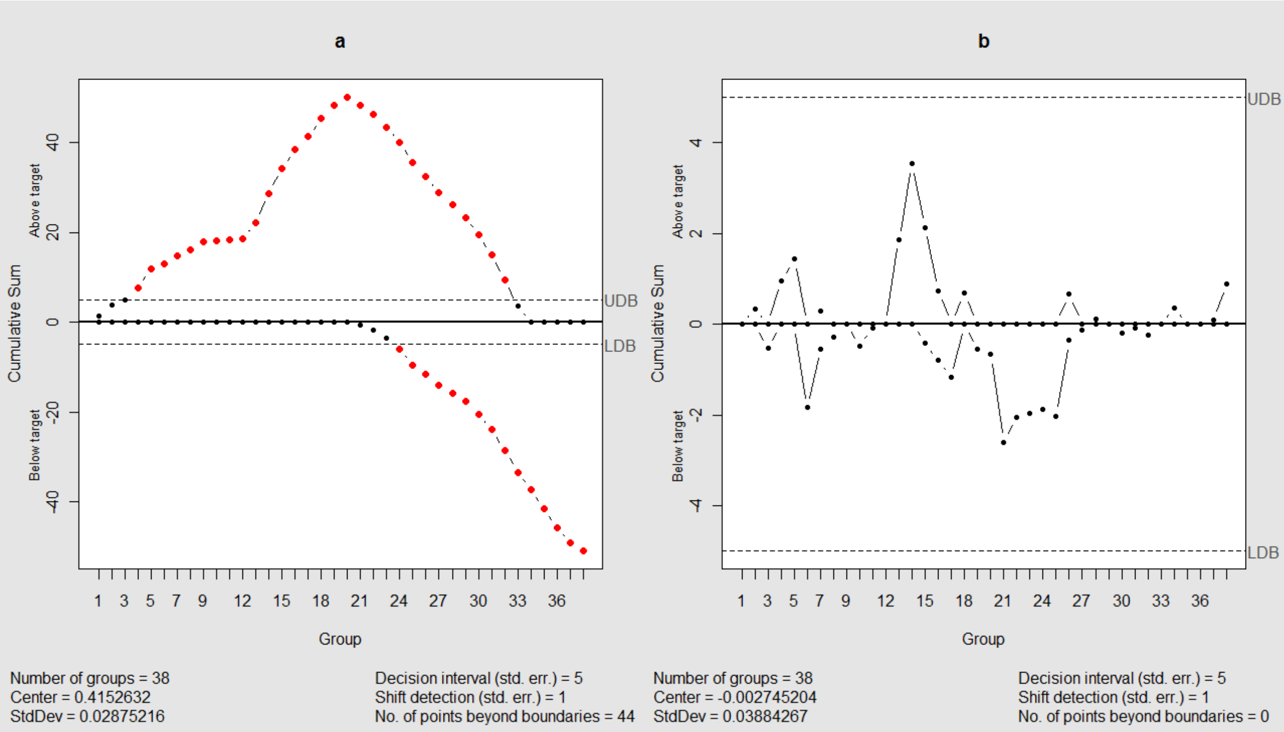

Figure 3: CUSUM (a) applied to turbidity data (a) and to the residuals from the ARIMA model (b).

To complement, we applied a Shewhart chart and a CUSUM chart to monitor some further measures (Figures 4a and 4b). It can be observed that the CUSUM chart signals an increase in turbidity level, starting from the first observation after the change, which does not occur in the Shewhart chart.

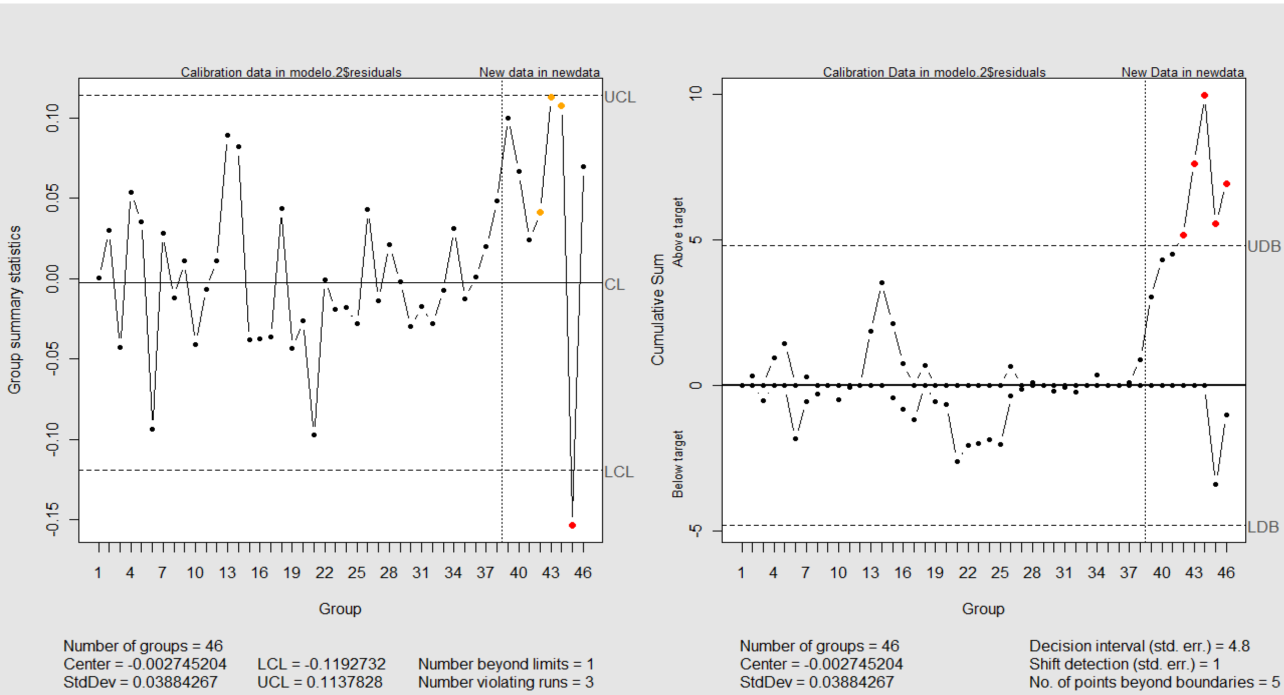

Figure 4: Shewhart (a) and CUSUM (b) charts applied to monitoring new measures.

The supplementary rules (yellow points) indicate that something may be occurring in the process, but also increase the number of false alarms. For this reason are not considered in this paper.

Conclusions and Final Considerations

In this paper, control charts for turbidity were presented in order to propose alternatives for future monitoring of the process. This analysis emphasized the importance of verifying the assumptions of normality and autocorrelation before using control charts to monitor water quality parameters.

In all analyses, the values beyond the control limits were false alarms, caused by the presence of autocorrelation. The methodology allowed us to define the dimensions of quality in the process of treated water turbidity going beyond just meeting the specification limits. The final results obtained with the application of control charts for this process were similar. In choosing a procedure for monitoring, the positives and negatives of each technique must be balanced. While the Shewhart chart is easy to implement, CUSUM and EWMA give more information about the trend of the process, but the latter is more robust on the issue of normality. A suggestion for future studies is to investigate the application of a combined Shewhart CUSUM chart, which allows signal changes of several magnitudes.

REFERENCES

[1] AIZENCHTADT, E.; INGMAN, D.; FRIEDLER, E. (2008), Quality control of wastewater treatment: a new approach, *European Journal of Operations Research*, v. 189, p. 445-458.

[2] ALVES, C.C.; SAMOHYL, R.W. (2004), A utilização dos gráficos de controle CUSUM para o monitoramento de Processos industriais. In: *Anais - XXIV Encontro Nacional de Engenharia de Produção Enegep 2004.* Florianópolis.

[3] BRASIL, Ministério da Saúde, Secretaria de Vigilância em Saúde, Coordenação-Geral de Vigilância em Saúde Ambiental, Portaria no. 518/2004 do Ministério da Saúde. URL: http://portal.saude.gov.br/portal/arquivos/pdf/portaria518 2004.pdf.

[4] COSTA, A. F.B. EPPRECHT, E. K. CARPINETTI, L.C.R. (2005), *Controle Estatístico de Qualidade*. São Paulo: Editora Átlas S.A, 334 p.

[5] HYNDMAN, R. J. (2010), *forecast: Forecasting functions for time series*. R package version 2.08. URL: http://CRAN.R-project.org/package=forecast .

[6] MONTGOMERY, D. C. (2004), *Introdução ao Controle Estatístico de Qualidade*, 4. ed. Rio de Janeiro: LTC.

[7] PAGE, E. S. (1954), Continuous Inspection Schemes. *Biometrika*, v. 41, n. 1/2, p. 100-114.

[8] R DEVELOPMENT CORE TEAM. (2009), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org .

[9] ROBERTS, S. W. (1959), Control Chart Tests Based on Geometric Moving Averages, *Technometrics*, v.1, p. 239-250.

[10] SAMOHYL, R.W. (2009), *Controle Estatístico de Qualidade*, Rio de Janeiro: Elsevier.

[11] SCRUCCA, L. (2004), qcc: an R package for quality control charting and statistical process control. *R News*, v.4, n.1, p.11-17.

[12] SMETI, E.M.; THANASOULIAS, N.C.; KOUSOURIS, L.P.; TZOUMERKARS, P.C. (2007), An approach for the application of statistical process control techniques for quality improvement of treated water, *Desalination*, v. 213, p. 273-281.

[13] STOUMBOS, Z. G.; REYNOLDS, M.R.; RYAN, T.P. WOODALL, W. (2000), The State of Statistical Process Control as We Proceed into the 21st Century. *Journal of the American Statistical Association*, v.95, n. 451, p. 992-998.

[14] WOODALL, W. H.; MONTGOMERY, D.C. (1999), Research issues and ideas in statistical process control. *Journal of Quality Technology*, v. 31, n. 4, p. 376-386.