Monitoring variables at the outlet of a clarified water loop through multivariate extension of real process capability index

Vásquez, Maura
Statistics School, Central University of Venezuela, Distrito Capital, Caracas, 1050, Venezuela.
E-mail: mauralvasquez@gmail.com
García, Teodoro
Industrial Engineering School, University of Carabobo, Valencia, Carabobo, 2005, Venezuela.
E-mail: tgarcia@uc.edu.ve
Ramirez, Guillermo
Statistics School, Central University of Venezuela, Distrito Capital, Caracas, 1050, Venezuela.
E-mail: guillermo.ramirez.ucv@gmail.com
García, José
Industrial Engineering Department, Papeles Venezolanos C.A., Guacara, Carabobo, 2003, Venezuela.
E-mail: Jose.r.garcia@paveca.com.ve
Martínez, Carlos
Industrial Engineering School, University of Carabobo, Valencia, Carabobo, 2005, Venezuela.
E-mail: cmartin@uc.edu.ve
Camardiel, Alberto
Statistics School, Central University of Venezuela, Distrito Capital, Caracas 1050, Venezuela.
E-mail: acamar09@gmail.com

ABSTRACT

The North-Central area in Venezuela has two restrictions for installing toilet paper factories: First, the only fresh water available in the area should be used for human consumption. On the other hand, all effluents from the area are poured into Valencia Lake. The lake ecosystem presents significant damage due to the effect of biological and chemical wastes. It is recommended to build two closed circuit water treatment plants to minimize the environmental impact and simultaneously supply the huge amounts of water required for the toilet paper factory under study. These treatment plants allowed water to be reused through applying the clarification procedure, which consists of adding flocculants, coagulants and sulfuric acid. The water coming from this type of treatment plant should meet certain quality standards if it is going to be used to manufacture toilet paper. The quality of clarified water depends on the behavior of the following variables: pH, ionic demand, turbidity and total suspended solids. Statistical behaviour of these four variables was monitored applying multivariate control charts. These closed circuits have an efficiency of 95%, a 4.77% of paper sludge which is deposited in a landfill and 0.23% of water that must be restored. The combined effects of both plants produce a daily volume of 5,000,000 cubic meters of clarified water. Comparing this figure with the value of 50 liters proposed by the United Nations as the estimate daily amount required by a person for consumption, cooking and cleaning, this complete process creates a saving equivalent to the daily water requirement for 100 million people. On the other hand, this paper proposes a type C_{pk} index to assess the quality of the clarified water, setting bounds to the probability of the process being within specifications.

Keywords: Clarified water, environmental impact, multivariate process control, multivariate extension of process capability index.

1.- INTRODUCTION

Valencia Lake is the largest natural freshwater lake of Venezuela. It has a surface area of 350 Km^2 and is located in the central-north part of the country, which is the most densely populated. Valencia Lake lies on an east-west tectonic depression between two ranges of mountains: Cordillera de la Costa on the north and the Serrania del Interior in the south. The lake became endorheic about 250 years ago, when the discharge level (427 m) was exceeded due to desiccation.

Intensive human intervention of the watershed and a reduction of groundwater flow have been mentioned for causing that rapid decline of the lake level. A minimum was achieved in 1976 (400.8 m above sea level). However, the lake level has recovered to 405 m with water provided from another watershed for urban consumption. Incoming untreated wastewater from domestic, agricultural and industrial activities of about 2 million people contribute to eutrophication, contamination and salinization of the lake.

The area around the lake has two restrictions for installing toilet paper factories: First, the only fresh water available should be used for human consumption. On the other hand, all effluents from the area are poured into Valencia Lake. It is recommended to build two closed circuit water treatment plants to minimize the environmental impact and simultaneously supply the huge amounts of water required for the toilet paper factory under study. These closed circuits have an efficiency of 95%, a 4.77% of paper sludge which is deposited in a landfill and 0.23% of water that must be restored. The combined effects of both plants produce a daily volume of 5,000,000 cubic meters of clarified water.

The clarification process has three steps: the first one is to add sulfuric acid since to obtain between 6 and 6.9 the water pH. In this range, the coagulant has its best performance. The second step is to add the coagulant. The goals of coagulation are to decrease particle stability and to reduce its mutual repulsion resulting in a stable suspension with no net charge. The third step is to add the flocculant. The purpose of flocculation is to bring particles together to form well settling particles. The quality variables of the clarified water are: pH, total suspended solids, turbidity and ionic charge. The clarified water process is evaluated in an integral way that means the statistical relations between variables will have effect in the final quality of the water.

One of the most important performance indicators is the process capability index. Using this indicator it is possible visualizing how the process is fulfilling all the quality specifications. In addition, the relationships between the quality variables define the appropriate focus. These variables may be independent or correlated. For the independent case, Bothe (1991) proposed a technique to calculate a real capability index C_{pk} to an entire product. If the quality variables are correlated the multivariate techniques must be used. Several authors have worked in multivariate capability indices. Wierda (1993) proposed a multivariate index MC_{pk} calculating the integral of the original variables, Chen (1994) proposed an index over a rectangular solid tolerance zone, Boyles (1996) used the lattices concept to propose an exploratory capability analysis, Wang and Chen (1998) proposed capability indices for multivariate normal distributions using principal component analysis and the geometric mean, and Wang and Du (2000) extent the paper of Wang and Chen (1998) at the multivariate non normal cases, Foster et al. (2005), proposed using the coefficients generated through a new representation named POBREB and Jeh-Nan and Chun-Li (2010), proposed a new multivariate indices named NMC_p and NMC_{pm}. These are some techniques proposed to calculate the multivariate capability indices. This paper proposes a type C_{pk} index to assess the quality of the clarified water, setting bounds to the probability of the process being within specifications.

2.- MATERIALS AND METHODS

2.1 Technical procedures

The samples were taken each four hours. This frequency is greater than the water residence time in the clarifier tank and this helps to solve the autocorrelation problems.

The quality variables considered in this study are the following:

Ionic charge. This characteristic was assessed according to the instructions for use the equipment Mütek PCD 04 and it is measured in microequivalents (μ_{eq}).

Turbidity. It was obtained following the "Standard Methods for the examination of water and wastewater". Method 2130. Turbidity is measured in Nephelometric Turbidity Units (NTU).

Total suspended solids. This variable is obtained following the "Standard Methods for the examination of water and wastewater". Method 2540 D. The TSS mass is measured in mg solids per liter of water (ppm).

pH . The pH was obtained following the test method ASTM E 70-07 "Standard Test Method for pH of Aqueous Solutions With the Glass Electrode".

Table 1 shows the lower specification level, target and upper specification level for the quality variables.

Table 1.- Quality variable specifications

	charge (μ_{eq}/g)	turbidity (ntu)	solids (ppm)	pH
Lower level	-0.12	20	5	6
Target	-0.10	45	12	6.3
Upper level	-0.05	70	18	6.9

2.2 Statistical procedures

Phase I

This phase explores an historical sample of 100 observations, in order to obtain references to build the control limits. The following steps should be followed:

- Assess univariate and multivariate normality (Mardia Tests).
- Draw univariate and multivariate control graphs, to determine if the process is under control.
- Determine the main variability directions, through which the second phase is going to be monitored using a dynamic PCA.
- Compute a multivariate capability index ($C_{pk.mult}$) to evaluate the global behaviour of the process.

Phase II

This phase performs a programmed observation of production in order to monitor the process in real time:

- Monitoring through univariate and multivariate control graphs.
  Control graphs are drawn, based on the references obtained in phase I, to detect special signals.
- Monitoring through dynamic PCA.
  Real time monitoring [7] refers to the principal directions of variability obtained in phase I. The dataset observed in phase II is used as testing sample. To capture the eventual changes in the covariance structure when a new observation is available, a measure of the lack of fit is calculated as the standardized difference between a p-variant observation, say x_i, and its orthogonal projection on the subspace defined by the *a* variability principal directions:

$$e_i^2 = \sum_{\alpha=1}^{a} \frac{x_i^t (I - V_\alpha V_\alpha^t) x_i}{\lambda_\alpha} \qquad (1)$$

This statistic can be approximated by a chi-square distribution $\chi^2_{(p-a)}$. In case of significance, the corresponding observation is considered as a likely "special cause", otherwise it is added to the dataset.

2.3 Process capability indices C_p and C_{pk}

The C_p index compares the difference between the specification limits of the process: upper specification limit – lower specification limit = USL – LSL with the central 99,73% of data (no more than 27 in 10.000 units will fall outside these limits): $(\mu + 3\sigma) - (\mu - 3\sigma) = 6\sigma$, and it is defined as:

$$C_p = \frac{USL-LSL}{6\sigma}$$

It is assumed that the characteristic to be considered is normally distributed. Several authors (Ryan, 2000) state that C_p is a poor measure of process capability since it ignores the process mean. This index measures just the *potential* process capability only if the process mean is centered between the specification limits.

On the other hand, the C_{pk} index is a measure of the capability of one side of the distribution, the side for which the larger of the two proportions nonconforming will result. It is defined as:

$$C_{pk} = (1/3)\, Z_{min}$$

where $Z_{min} = \min(Z_1, Z_2)$, number of standard deviations that the closest specification limit is to μ,
with $Z_1 = (USL - \mu)/\sigma$ and $Z_2 = (\mu - LSL)/\sigma$.
Clearly, C_{pk} is an improvement over C_p since it is a function of the process mean. However, it has the limitation that there is not a one-to-one correspondence between C_{pk} and the percentage of nonconforming units.

2.4 Multivariate capability index - a proposal

In this section we propose a procedure to obtain a type C_{pk} index to evaluate the performance of a multivariate process, setting bounds to the probability of the process being within specifications. We denote the quality characteristics as $X_1, X_2, \ldots X_p$ and the event A_j = {the process meets the requirement in X_j}, j=1,2,...p. The proposed index, denoted by $C_{pk.mult}$, is based on the following expressions:

i.- Prob (process is within specifications) = Prob (A_j , $\forall$ j=1,2,…p) = Prob ($\cap A_j$)

ii.- Prob ($\cap A_j$) $\leq$ min {Prob (A_j), j=1,2,…p}

iii.- Prob ($\cap A_j$) $\geq 1 - \sum$ Prob (A_j^c)

iv.- From ii and iii, bounds for the probability of the process being out of specifications are obtained:
$pl = 1 - \sum \text{Prob}(A_j^c) \leq p = \text{Prob}(\cap Aj) \leq pu = \min\{\text{Prob}(A_j)\ j=1,2,\ldots p\} \Rightarrow$
$ql = 1 - pu \leq q = 1-p \leq qu = 1 - pl$

v.- The index is obtained as:

$$C_{pk.mult} = (⅓)\ \min\{\Phi^{-1}((1-pu)/2), \Phi^{-1}((1+pl)/2)\} \qquad (2)$$

where Φ denotes the cumulative normal distribution function.
It is important to note that $C_{pk.mult}$ is a function of the lower and upper limits of the proportion of non conforming units.

This proposed index is calculated from the original variables as well as from the principal components, following [10] and [5].

3.- RESULTS

3.1 Phase I

In order to reach stability in the process, a monitoring of the historical observation was done. The assumption of multivariate normal distribution was tested by Mardia procedure (p-value = 0.388).

As the quality variables are significantly correlated, there is a stronger need for using multivariate control charts instead of charts for individual variables. Figure 1 shows that the process has reached stability.

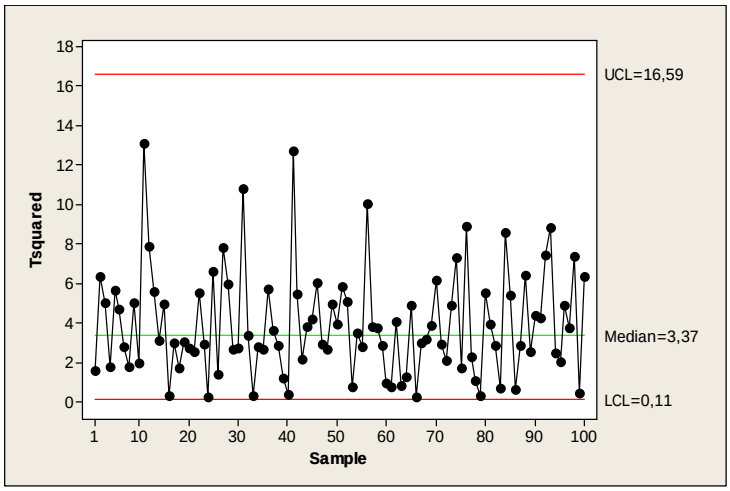

Figure 1.- Tsquared chart

A PCA was done on the covariance matrix for the first 100 observations to understand the structure of the relationships between the quality variables. The first factor accounts for 97.37% of the total process variability (Table 2). The correlations between the original variables and the factors are examined to understand the nature of the component. The first pc is a contrast between the *solid* and *turbidity* measurements and the *ionic charge*. A visual inspection (Figure 2) reveals that the clarified water samples in the beginning of the process are mostly located at the upper right of the diagram, while those at the lower left side correspond to the rest. In addition, we can observe that the entire sample lie inside the tolerance region defined by pc1 and pc2, and besides the target is placed next to the origin.

Table 2.- Results of the PCA

		Correlation		
		pc1	pc2	pc3
Variables	charg	0.67	-0.35	0.17
	turbi	-1.00	0.00	0.00
	solid	-0.67	0.01	0.00
	ph	-0.12	0.00	0.99
Eigenvalue		24.25	0.65	0.00
% explained of total variability		97.37%	2.61%	0.02%

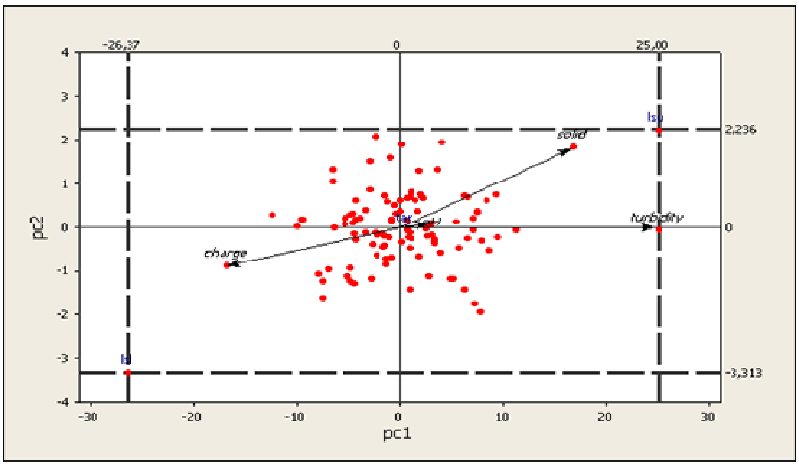

Figure 2.- Factorial Plot

Figure 3 shows the univariate capability index for each variable in the process. Obviously, we need the values of these indexes to obtain $C_{pk.mult}$. It can be seen that Pp and Ppk for *turbidity*, *solids* and *pH* indicate an excellent performance. The same is not true for *charge* whose distribution is not centered between the specification limits.

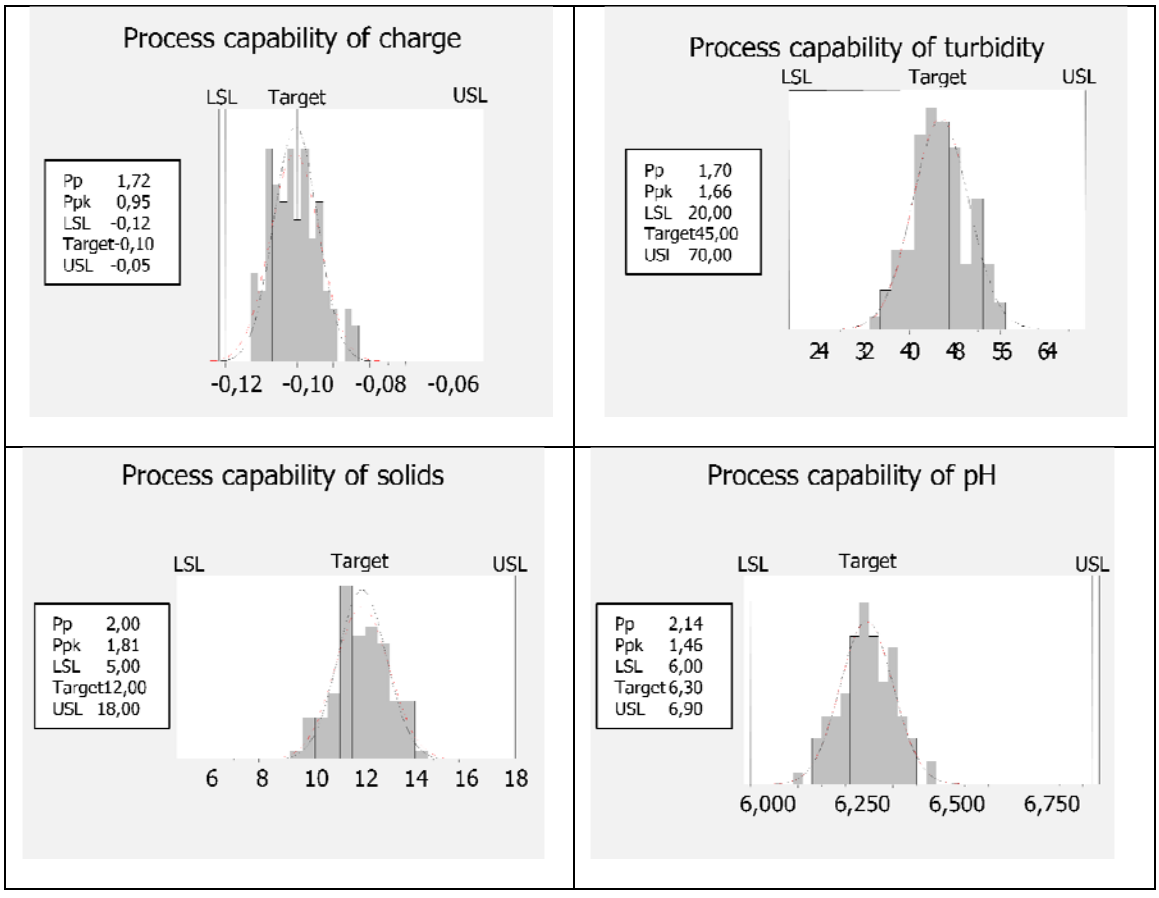

Figure 3.- Process capability of quality variables

As shown in Table 3, the ability of process to meet the engineering requirements measured by $C_{pk.mult}$ (1.02) is small. The level of nonconforming samples lies between ql = 2252 ppm and qu = 2259 ppm. These results suggest that some actions must be taken in order to promote improvements in the process.

Table 3 Multivariate process capability index - variables

$C_{pk.mult}$	ql	qu	ppmu	ppml
1.01796	0.002252	0.002259	2252.3	2259.0

Table 4 shows the evaluation of the process by using $C_{pk.mult}$ based in successive combinations of the principal components. The results agree with the principal directions of variability, as can be seen in Table 2. The index computed only with the first component (1.71) indicates a high process capability in this direction while $C_{pk.mult}$ based in the two first components considerably reduces its performance. This situation is due to the particular relationship between charge and solids. On the other hand, $C_{pk.mult}$ based in pc1 and pc3 (1.39) indicates a good performance because here the variability is described by samples at the beginning of the process, with high levels in *turbidity* and *solids*, and low levels in *charge*, changing progressively to an inverse profile at the end, keeping constant *pH* along the process.

Table 4 Multivariate process capability index - components

	$C_{pk.mult}$	ql	qu	ppmu	ppml
pc1	1.71019	0.000000	0.000000	0.29	0.29
pc1+pc2	0.98928	0.002999	0.002999	2998.72	2999.01
pc1+pc3	1.38887	0.000031	0.000031	30.63	30.92
pc1+pc2+pc3	0.98824	0.002999	0.003030	2998.72	3029.64

3.2 Phase II

Figure 4 represents the lack of fit of the kth new observation in phase II (k = 1,2...51) respect to the variability structure defined by phase I, together with the added observations until time k-1. It is evident that there are not significant gaps of datasets in phase II respect to the structure in reference. In addition we can observe that there is a progressive better adaptation all along.

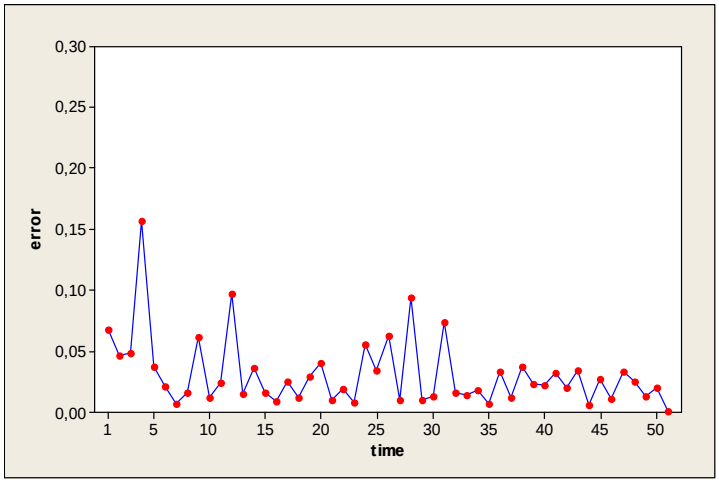

Figure 4.- Time series plot of error

4.- CONCLUSIONS

Capability evaluation of multivariate process is usually carried out using univariate indices in a separate way. The objective of this study is to propose a practical tool for assessing process capability in case of normal multivariate data, and to put it in practice in the specific context of the clarification procedure used in a water treatment plant. The proposed multivariate index $C_{pk.mult}$ is a function of the proportion of non conforming units, and it can be applied both on the original data and on the transformed data via principal component analysis. Therefore its use is less restrictive about distributional assumptions and provides as added value the comprehension of the underlying aspects of the variability of the process.

5.- REFERENCES

[1] Bothe, D. (1991). A capability study for an entire product. *ASQC Quality Control Transactions*, 921-925.

[2] Boyles, R. (1996), Exploratory Capability Analysis, *Journal of Quality Technology*, 28(1), 91-98.

[3] Chen, H. (1994). A multivariate process capability index over a rectangular solid tolerance zone. *Statistica Sinica*, 4, 749-758.

[4] Foster, E., Barton, R., Gautam, .L, Trust, L. & Tew, J. (2005). The process-oriented multivariate capability index. *International Journal of Production Research,* 43(10), 2135-2148.

[5] García, T., Vásquez, M., Ramírez, G. y García, J. (2006). Multivariate Extension of the Process Capability Indices to determine the fulfillment degree of the product specifications on the elaboration of the economical hygienic paper, *Proceedings of the 11th annual International Conference on Industrial Engineering Theory, Applications & Practice,* 587-592.

[6] Lam, C. & Littig, S. (1992). A new standard in process capability analysis. Technical Report No 92-23, Department of Industrial and Operational Engineering, University of Michigan.

[7] Min Woo Lee, Sung Hun Hong, Hyeoksun Choi, Ji-Hoon Kim , Dae Sung Lee & Jong Moon Park (2008). Real time remote monitoring of small scaled biological wastewater treatment plants by a multivariate statistical process control and neural network based software sensors. *Process Biochemistry*, 43, 1107-1113.

[8] Ryan, T. (2000). *Statistical Methods for Quality Improvement*. Wiley. New York.

[9] Wang, F. & Chen J. (1998). Capability index using principal component analysis. *Quality Engineering*. 11(1), 21-27.

[10] Wang, F. & Du, T. (2000). Using principal component analysis in process performance for multivariate data. *The International Journal of Management Science*, 28, 185-194.

[11] Wierda, S.J. (1993). A multivariate process capability index, *ASQC Quality Control Transactions*, 342-348.