



## Análisis comparativo de indicadores de capacidad multivariante. El caso del clúster manufacturero cartagenero

RAHMER, BRUNO DE JESÚS

Fundación Universitaria Tecnológico Comfenalco (Colombia)

Correo electrónico: brunodejesus.2509@gmail.com

GARZÓN SAÉNZ, HERNANDO

Fundación Universitaria Tecnológico Comfenalco (Colombia)

Correo electrónico: hnando2001@gmail.com

SOLANA GARZÓN, JOSÉ

Fundación Universitaria Tecnológico Comfenalco (Colombia)

Correo electrónico: ingjosemsolanag@gmail.com

### RESUMEN

Los indicadores multivariantes de capacidad han sido muy utilizados en la industria manufacturera y en entornos similares en tanto que proporcionan medidas cuantitativas sobre el potencial y rendimiento de un proceso descrito por múltiples características de calidad susceptibles de evaluación y correlacionadas simultáneamente. En este artículo se presenta una evaluación empírica de múltiples enfoques de análisis de capacidad multivariante en el sector manufacturero cartagenero, a saber: uno basado en la relación del volumen la región de tolerancia y el volumen de la región del proceso, otro fundamentado en el análisis de componentes principales y adicionalmente, se propone la aplicación de otros indicadores que incluyen dentro su mecánica interna de análisis factorial la proporción de observaciones fuera de los márgenes de especificación y su variabilidad en el largo y corto plazo. Los resultados suministrados indican que el proceso productivo analizado no es marginalmente capaz para cumplir con las especificaciones técnicas predefinidas y que existe un vasto margen de mejoría.

**Palabras clave:** análisis multivariante, análisis en componentes principales, clúster manufacturero, indicadores de capacidad, control estadístico de procesos.

**Clasificación JEL:** C19; C38; Z21.

**MSC2010:** 62H25; 62P30.

# Comparative analysis of multivariate capacity indicators. The case of the Cartagena manufacturing cluster

## ABSTRACT

Multivariable capacity index have been used frequently in the manufacturing industry and similar environments, since they provide quantitative measurements of the potential and performance of a process described by multiple quality characteristics susceptible to evaluation and correlated simultaneously. In this paper, we present an empirical evaluation of several approaches to analyze a multivariate industrial process. In the first approach we study relationship of the volume of the tolerance region and the volume of the process region. In the second approach, we propose the application of principal components technique and, finally, we propose the application of other indicators that analyze the proportion of observations located outside the specification margins and their variability in the long and short term. The results provided indicate that analyzed production process is not marginally capable for satisfying predefined technical specifications and that there is a high margin of improvement.

**Keywords:** multivariate analysis, principal component analysis, manufacturing cluster, capacity index, statistical process control.

**JEL classification:** C19; C38; Z21.

**MSC2010:** 62H25; 62P30.



## 1. Introducción.

La tecnificación y el consecuente cambio cualitativo de los procesos de producción en los albores del siglo XXI ha allanado mucho el terreno para engendrar múltiples estrategias que se inscriben en la lógica del capital, tendientes a la optimización del desempeño de las operaciones internas, a reformar las propiedades sistémicas, paradigmas y modelos en el contexto organizacional, para el alcance de metas generales coherentes con su objeto social, como la fabricación de bienes tangibles y prestación de servicios con valor agregado diferencial.

En tales escenarios, la utilización de herramientas cuantitativas para la monitorización de procesos permite caracterizar la dinámica de los mismos, corregir cursos de acción y concebir previsiones sobre su rendimiento en el corto y largo plazo. Se colige, por tanto, que el control de las operaciones constituye un eslabón de significativa relevancia en el complejo entramado de actividades de fabricación de bienes tangibles y de la prestación de servicios, con valor agregado diferencial. En las postrimerías del siglo anterior, métodos no convencionales de control y monitoreo de procesos ganaron legitimidad en el ámbito de las ciencias administrativas. En ese mismo contexto, los índices de capacidad multivariantes, irrumpieron en la literatura especializada como una generalización de sus análogos univariantes y fueron profusamente utilizados por su patente versatilidad (Verbel, Herrera & Maestre, 2016). La principal ventaja proporcionada por esta forma de monitorización es la realimentación del proceso en tiempo real partiendo de una robusta fundamentación teórico-conceptual (Hernández & Da Silva, 2016). Dentro de esta evaluación sistémica se introducen conceptos axiales e indisociables como estabilidad y capacidad que, conjugados entre sí, coadyuvan al acrecimiento de la eficiencia operativa y dan cabida al desarrollo del potencial existente en el sistema productivo (Caicedo, Jiménez, Vélezl, & Wilches, 2015).

Los índices de capacidad de proceso (PCI) actualmente son utilizados como parte integral del control estadístico de la calidad y productividad del sistema. La relación entre el rendimiento real del proceso y los límites o tolerancia de especificación se puede cuantificar utilizando un PCI conveniente, pues las métricas de capacidad están diseñadas estadísticamente para aportar coeficientes medidos como una función adimensional de los parámetros y especificaciones de un proceso (Novaes, Lima Jr, Carvalho, & Aragão, 2015) que dan cuenta sobre la aptitud del mismo para reproducir elementos dentro de los límites preestablecidos. En esencia, estos indicadores contrastan el rendimiento real con el rendimiento deseado de los outputs (Lea, Wu & Wu, 2015; Dharmasena & Zeepongsekul, 2015). De este modo, es posible detectar señales estadísticas anómalas o causas de variación imputables que permiten optimizar procesos productivos perceptibles y minimizar la ocurrencia de errores no intencionales, en otros términos, el objetivo ulterior es que sobre él operen, como únicas causas de variabilidad, las causas de variación no asignables. En esa tónica, un proceso es calificado como estable si el mismo es reproducible. La capacidad, en otro sentido, facilita la previsión del desempeño del proceso en ejecuciones futuras y la confección de planes realizables. Se precisa destacar que en la totalidad de procesos operan fenómenos constantemente activos que originan variaciones comunes o especiales. Por lo tanto, un proceso estable no es aquél que no exterioriza variaciones, sino que presenta desviaciones aceptables dentro de límites tolerables que determinan la repetición de su comportamiento.

En la práctica industrial, los índices de capacidad multivariante tienen sus pros y sus contras: si bien es cierto que exhiben una ostensible solvencia para proveer una caracterización apropiada de la estructura de correlación subyacente entre las múltiples características de calidad examinadas prescindiendo de una elevada cuota de esfuerzo, suelen acarrear un exorbitante costo en términos de complejidad numérica e interpretabilidad, ya que un único indicador que resuma el comportamiento de un proceso es indeseable en casos donde interaccionan múltiples variables, lo que obstaculiza manifiestamente, la identificación de la totalidad de causas hipotéticas de un eventual detrimento de la capacidad del mismo. Por ello, es recomendable desarrollar un análisis de capacidad general bajo una lógica consistente en la evaluación agregada de las características de calidad -tal y como se sigue aquí- siguiendo un enfoque metodológico misceláneo.

## 2. Metodología.

La noción subyacente en un índice de capacidad versa sobre un proceso que, encontrándose estadísticamente bajo control, se comporta de tal forma que las características de calidad se localizan casi plenamente circunscritas dentro de los límites de especificación determinados por las tolerancias de ingeniería o por las necesidades de los clientes (Vásquez, Ramírez, & García, 2016). La evaluación del desempeño de un proceso mediante el cómputo de un índice de capacidad toma en consideración y de forma simultánea, su capacidad para generar unidades dentro de ciertos límites de especificación predeterminados y su cualidad para minimizar la discrepancia entre el valor del centrado del proceso y un valor target.

El enfoque de investigación empírica es de carácter mixto en virtud del enlazamiento de dos modalidades investigativas bien diferenciadas: un enfoque estrictamente descriptivo y retrospectivo). En esta fase se seleccionan diversas submuestras de datos correspondientes a un proceso de metal mecánico de una entidad subyacente en el clúster manufacturero de Cartagena de Indias. Cinco variables continuas han sido seleccionadas en este caso: PH, Concentración de sal de nitrógeno, Temperatura, Presión y Concentración de cloruro de sodio. El conjunto de datos obtenidos mediante un proceso de selección muestral aleatoria sin reposición, de modo que cada elemento del universo posee idéntica probabilidad de selección y queda descartado en una fase consecutiva de elección. Para el cómputo del tamaño muestral se establece un nivel de confianza del 95%, un margen de error del 5% y una heterogeneidad intramuestral equivalente al de 50%.

En la fase posterior al proceso de definición de ciertas propiedades medibles a través de las que es posible evaluar la calidad del resultado del proceso sometido a análisis, se procede a realizar un somero análisis exploratorio de los datos de entrada del proceso productivo, recabándose información útil respecto a la tendencia central, dispersión y forma distribucional de las variables analizadas. Esto sienta la base para identificar el nivel de estabilidad del proceso, su capacidad para cumplir con las especificaciones y para la verificación formal de ciertos supuestos e hipótesis que deben cumplirse estrictamente si se desea garantizar una adecuada estimación de los parámetros que sintetizan la información contenida en las observaciones. Los intervalos de especificación, es decir, los límites de especificación superior, valores nominales y límites de especificación inferior de cada una de las medidas de interés son computados a partir de la información histórica del proceso. Se establece un diseño de corte transeccional, en tanto que la extracción de insumos informativos ha ocurrido en una acotación temporal reducida y dado que el propósito, en el punto de inicialización del proceso de investigación, es efectuar un análisis descriptivo y causal. La hipótesis inicial que ha de ser corroborada formalmente es que la totalidad de factores involucrados en el proceso están correlacionados entre sí y, por tanto, se debe evaluar concomitantemente su evolución o co-movimientos en la progresión temporal. Esta premisa justificaría el uso de indicadores de capacidad multivariados.

Dado que se propone un análisis de capacidad en el dominio multivariado se analizan en conjunto y paralelamente, ciertas características de calidad bajo la asunción de que pueden ser modelizadas por una distribución gaussiana multidimensional y que el conocimiento del proceso es lo suficientemente exhaustivo, de suerte que es posible la estimación del vector de medias y matriz de varianzas y covarianzas (Cuamea & Rodríguez, 2014). En la primera fase investigativa el proceso productivo en cuestión es caracterizado y evaluado a partir de enfoques de análisis variopintos que se circunscriben en el ensanchado campo del control estadístico e ingeniería de procesos. Así, las métricas de capacidad que se traen a colación en el presente trabajo se construyen a partir de la relación entre el volumen de una región de tolerancia y el volumen de una región de proceso, técnicas de reducción de información como el análisis de componentes principales y se registran posteriormente, otros indicadores que incluyen dentro su mecánica interna de análisis factores como la proporción de observaciones fuera de los márgenes de especificación y la variabilidad de las características de calidad en el largo y corto plazo. La conjunción de todos ellos se realiza con la intención de cuantificar la habilidad del proceso para satisfacer las especificaciones pre-asignadas (Salazar & Fermín, 2016; Herrera, Herrera & Rahmer, 2017).

De modo sintético podrá trazarse la secuencia lógica de instancias que deben seguirse para alcanzar los objetivos de investigación trazados, partiendo de la asunción veraz que se ha realizado una aproximación teórica concordante con los objetivos planteados sobre la información suministrada; en otros términos, se ha analizado conceptualmente el objeto de investigación. Una vez establecido el modelo conceptual, el énfasis debe centrarse en la determinación y aplicación la técnica de análisis multivariante más conveniente para efectos de monitorización, sus condiciones de aplicabilidad y la corroboración de hipótesis subyacentes. Una vez superada esta fase se procede a construir los indicadores cuantitativos y efectuar el correspondiente análisis de las alternativas de evaluación de indicadores. La interpretación de los resultados obtenidos podría conducir a posteriores re-especificaciones en el proceso de modelo matemático.

### 3. Resultados.

En este apartado se realiza un análisis exploratorio de la totalidad de variables identificadas en cada etapa del proceso productivo, asumiéndose que cada una de ellas puede ser modelada por una distribución normal con media poblacional  $\mu$  y varianaza  $\sigma$ . A continuación, se presenta en la Tabla 1 un conjunto de estadísticas descriptivas univariadas (medidas de centralización y dispersión) que proporcionan una aproximación sintética del comportamiento exhibido por las características de calidad. Se computan además sus respectivos valores targets y límites de control.

**Tabla 1. Análisis descriptivo de las variables.**

	<i>PH</i>	<i>CONCENTRACIÓN DE SAL DE NITRO</i>	<i>TEMPERATURA</i>	<i>PRESIÓN</i>	<i>CONCENTRACIÓN DE CLORURO DE SODIO</i>
<b>PROMEDIO</b>	6,25673	0,295938	29,2554	800,402	0,500714
<b>MEDIANA</b>	6,184	0,295893	29,8657	800,293	0,500406
<b>DESVIACIÓN ESTÁNDAR</b>	1,06645	0,01745	3,90211	1,32129	0,00793014
<b>COEFICIENTE DE VARIACIÓN</b>	17,045%	5,89651%	13,3381%	0,165079%	1,58377%
<b>MÍNIMO</b>	4,07598	0,262365	27,2397	797,983	0,48555
<b>MÁXIMO</b>	8,89682	0,338308	31,8209	803,709	0,525659
<b>RANGO</b>	4,82084	0,075943	29,7081	5,726	0,040109
<b>SESGO</b>	0,273103	0,201228	-6,43895	0,586496	0,40188
<b>CURTOSIS</b>	-0,40978	-0,381231	45,2995	0,289672	0,780343
<b>LÍMITE INFERIOR</b>	4	0,26	27	795	0,45
<b>VALOR NOMINAL</b>	6	0,30	30	800	0,50
<b>LÍMITE SUPERIOR</b>	7,5	0,34	33	803	0,53

Fuente: Elaboración propia.

En la Tabla 2 aparecen los coeficientes de correlación de Pearson para examinar la fuerza y la dirección de la relación lineal entre pares de variables continuas. Nótese que casi la totalidad de variables consideradas en el problema son próximas a la unidad y sólo tres pares no son significativas al 95% de confianza. En ausencia de valores atípicos, las correlaciones no son afectadas significativamente, tal y como puede dilucidarse en el presente caso. En ese tenor, puede afirmarse sin temor a equivocación que el cómputo de los indicadores de capacidad multivariante es factible, pues se corrobora que la asociación entre las variables es de tipo lineal y que existen movimientos concomitantes entre las mismas.

**Tabla 2. Correlaciones bivariadas de Pearson.**

	PH	CONCENTRACIÓN DE SAL DE NITRO	TEMPERATURA	PRESIÓN	CONCENTRACIÓN DE CLORURO DE SODIO
PH	0,7131	0,6185	0,5633	0,7149	0,6051
CONCENTRACIÓN DE SAL DE NITRO	0,7306	0,8391	0,0635	0,8059	0,8864
TEMPERATURA	0,5849	0,9240	0,6823	0,3362*	0,7919
PRESIÓN	0,7072	0,5085	0,6633	0,6051*	0,8742
CONCENTRACIÓN DE CLORURO DE SODIO	0,2976*	0,8391	0,7435	0,5681	0,813

Fuente: Elaboración propia.

### 3.1. Índices de capacidad basado en regiones de tolerancia.

Para el cálculo de los índices de capacidad multivariantes consignados en este apartado, se dispone inicialmente de la matriz de datos  $X$  especificada inicialmente, asumiéndose que las  $n$  observaciones son independientes y representan una muestra aleatoria descrita adecuadamente por una distribución normal multivariante con variables correlacionadas entre sí; asimismo, se toma en consideración las especificaciones técnicas predefinidas. El conocimiento del proceso es lo suficientemente exhaustivo y, por tanto, el cálculo del vector de medias para las características de calidad y la matriz de covarianzas es realizable.

Para valorar la capacidad general del proceso, en la Tabla 3, se computa una gama de indicadores de capacidad multivariado.

**Tabla 3. Índices de Capacidad Multivariados.**

ÍNDICES	FÓRMULA	VALOR
<b>CpM</b>	$\left[ \frac{\text{Volumen de la región de tolerancia}}{\text{Volumen de la región modificada del proceso}} \right]^{1/v}$	0,79
<b>PV</b>	$PV = P(F_{v,n-v} > \frac{n-v}{v(n-1)} T^2)$	0,57
<b>LI</b>	0= La región modificada no está contenida dentro de la región de tolerancia 1= La región modificada está contenida dentro de la región de tolerancia	0
<b>Mcpm</b>	$\frac{C_p}{D}$	0,81

Fuente: Elaboración propia.

Los primeros tres componentes consignados en la Tabla 3 ( $CpM$ ,  $PV$ ,  $LI$ ) constituyen el denominado vector de capacidad multivariante. El primer componente del vector está dado por el ratio entre el volumen de la región de tolerancia y el volumen de la región modificada del proceso, es decir:

$$CpM = \left[ \frac{\text{Volumen de la región de tolerancia}}{\text{Volumen de la región modificada del proceso}} \right]^{1/v}$$

El numerador viene dado por  $\prod_{i=1}^p (USL_f - USL_i)$ , mientras que el denominador, está definido como la región más pequeña, que circunscribe el elipsoide definida como región del proceso. Tal elipsoide es el contorno de probabilidad centrado en la media del proceso.

Mientras tanto, los límites o bordes del mismo vienen especificados por  $LS_f = \mu_i + \frac{\sqrt{\chi_{v,\alpha}^2 \det(\Sigma_i^{-1})}}{\det(\Sigma_i^{-1})}$  y  $LS_i = \mu_i - \frac{\sqrt{\chi_{v,\alpha}^2 \det(\Sigma_i^{-1})}}{\det(\Sigma_i^{-1})}$

siendo  $\chi_{v,\alpha}^2$  el percentil  $100(1 - \alpha)$  de una distribución ji-cuadrado con  $v$  grados de libertad y  $\det(\Sigma_i^{-1})$  el determinante de la matriz  $\Sigma_i^{-1}$  obtenida mediante la supresión la  $i$ -ésima fila y columna.

Dado que el valor del  $CpM$  consignado en la Tabla 3 se distancia sutilmente de 1, se sigue entonces que la región de tolerancia no es más grande que la región modificada del proceso, siendo esto un indicativo de baja capacidad.

El segundo componente del vector  $PV = P(F_{v,n-v} > \frac{n-v}{v(n-1)} T^2)$  es el nivel de significancia del estadístico  $T^2$  de Hotelling  $= n(\bar{X} - T)' S^{-1}(\bar{X} - T)$  bajo la hipótesis de que el centro de las especificaciones de ingeniería corresponde a la media del proceso. Valores próximos a cero indican cierto nivel de lejanía respecto al vector objetivo de ingeniería, como el presentado en la Tabla 3.

El tercer componente denotado por  $LI$  compara la localización entre la región de tolerancia y la región modificada y sólo asume un par de valores en el umbral  $[0,1]$ .  $LI$  será equivalente a la unidad si la totalidad de la región está contenida dentro de la región de tolerancia, si por el contrario, una parte de la de la misma se halla en zonas externas asumirá un valor igual a cero, como se percibe en el caso presente (Zhang, Wang, He, & He, 2014).

Por otro lado, el indicador  $Mcpm$  es el ratio entre el volumen de la región de tolerancia modificada dada por  $\frac{2\pi^{\frac{v}{2}} \prod_i^v a_i}{v\Gamma(\frac{v}{2})}$  y el elipsoide de confianza. Para  $v$  características de calidad esta métrica es igual a:

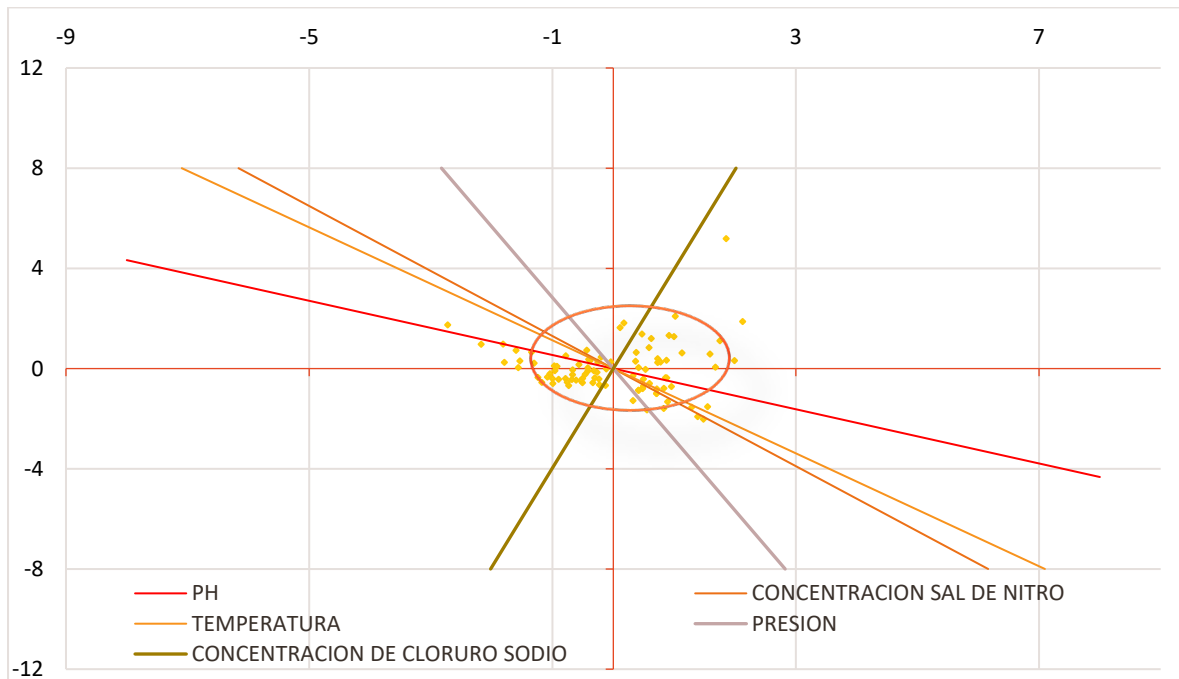
$$Mcpm = \frac{2\pi^{\frac{v}{2}} \prod_i^v a_i * v\Gamma\left(\frac{v}{2}\right)^{-1}}{Vol(R_3) [1 + (\mu - T)' \Sigma_i^{-1}(\mu - T)]^{1/2}} = \frac{C_p}{D}$$

siendo  $X$  el vector de mediciones descritas por una distribución gaussiana multivariante,  $T$  el vector de valores destino y  $R_3$  es la región donde yacen el 99,73% de los valores del proceso.  $C_p$  representa la variabilidad relativa a la región modificada del proceso mientras que  $D$  detecta la desviación del proceso respecto al objetivo.

Cuando el vector de medias del proceso es igual al vector target y el  $Mcpm$  es 1, entonces se afirma que el 99,73% de los valores están localizados dentro de la región de tolerancia modificada, circunstancia que no es reflejada en la dinámica del proceso.

A la postre se muestra en la Figura 1 la elipse de control diseñada ex profeso para esta sección. La relevancia de una herramienta de monitoreo de este carácter dado que la ocurrencia de las condiciones fuera-de-control es independiente de la presencia de comportamientos anómalos fuera de los límites de control en variables individuales. Nótese que existe una cantidad nada despreciable de puntos cuyas distancias cuadráticas multivariantes son inusitadamente altas respecto a la elipse de control.

**Figura 1. Elipse de Control 1.**



Fuente: Elaboración propia.

### 3.2. Índices de capacidad basados en análisis de componentes principales.

El análisis en componentes principales (ACP) es una técnica descriptiva de síntesis de información, útil para el diseño estadístico de indicadores de capacidad para un proceso normal multivariante con  $v$  características de calidad. En esencia, el uso PCA utiliza la descomposición espectral de la matriz de covarianza  $\Sigma = \mathbf{U}\mathbf{D}\mathbf{U}^T$ . Siendo  $\mathbf{U} = [u_1, u_2, \dots, u_n]$  la matriz de eigenvectores y la matriz diagonal de la matriz de autovalores  $\mathbf{D} = [\lambda_1, \lambda_2, \dots, \lambda_n]$ . Por otra parte, la componente  $i$ -ésima queda denotada como:

$$PC_i = u_i^T x = u_{v1}x_1 + u_{v2}x_2 + \dots + u_{vn}x_n$$

El análisis en componentes principales es una técnica de análisis estadístico multivariante que se clasifica entre los métodos de simplificación o reducción de la dimensión y que se aplica cuando se dispone de un conjunto elevado de variables con datos cuantitativos persiguiendo obtener un menor número de variables, combinación lineal de las primitivas, que se denominan componentes principales o factores, cuya posterior interpretación permitirá un análisis más simple del problema estudiado. Su aplicación es directa sobre cualquier conjunto de variables, a las que considera en bloque. En términos generales, el análisis de componentes principales envuelve un conjunto de datos con observaciones sobre  $p$  variables numéricas para cada una de las  $n$  entidades o individuos. Estos valores definen vectores  $n$ -dimensionales  $x_1, x_2 \dots x_n$  de modo que se busca una combinación lineal dada por:  $\sum_{j=1}^p a_j x_j$  cuyos coeficientes son obtenidos a partir de los vectores propios de la matriz de covarianza de los datos originarios. Intuitivamente, la finalidad esencial de un modelo de componentes principales es hallar un conjunto nuevo de direcciones ortogonales que definen la variabilidad máxima en términos de la estructura de varianza-covarianza de las variables originales, de manera que la información contenida en el conjunto completo de las componentes halladas es el equivalente exacto de la información original de los datos, pues se conjetura que existen elementos redundantes y que sólo añaden dimensionalidad al problema estudiado. A través del ACP, los datos originales son proyectados en una representación dimensional mucho más compacta y parsimoniosa.



Sea entonces una matriz de datos multivariantes  $X_{n,p} \in \mathbb{R}^{n \times p}$ , con  $X_{n,p} = (x_1, \dots, x_n)'$  cuyas columnas contienen las variables y las filas, los elementos. La media está dada por:

$$\bar{x} = \left(\frac{1}{n}\right) X_{n,p}'$$

y la matriz de covarianza por:

$$S = \left[\frac{1}{(n-1)}\right] (X_{n,p} - 1_n \bar{x}')' (X_{n,p} - 1_n \bar{x}')$$

La  $Cov(x) = \Sigma$  es semidefinida positiva y con raíces características correspondientes  $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p \geq 0$ .

Sean por tanto, las siguientes combinaciones lineales, a la vez incorreladas entre sí y que hacen máxima la varianza. A tales valores teóricos denominaremos componentes principales y están expresados por:

$$\begin{cases} Y_1 = l'_{11} X_1 + \dots + l'_{1p} X_p \\ \vdots \\ Y_p = l'_{p1} X_1 + \dots + l'_{pp} X_p \end{cases}$$

A sabiendas que  $Var(Y_1) = l'_1 \Sigma l_1$  y  $Cov(Y_i, Y_j) = l'_j \Sigma l_i$

Para el cálculo de la primera componente principal defínase:

$$y = e'_1 X, e'_1 e_1 = 1$$

De modo que:

$$Var(Y_1) = Var(l' X), \max Var(l' X) = Var(e'_1 X) = e'_1 \Sigma e_1$$

Para maximizar esta función multivariable sujeta a varias restricciones se dispone del método de Lagrange. Nótese que el vector desconocido  $e_1$  proporciona la combinación lineal óptima. De este modo se tiene que:

$$\left\{ \begin{array}{l} \max\{l' \Sigma l\} \\ l' l = 1 \end{array} \right\} = \phi_1(l) = l' \Sigma l - \lambda(l' l)$$

Al derivar respecto a  $l$  y premultiplicar por 2 se obtiene que:

$$\Rightarrow \frac{\partial \phi_1}{\partial l} = 2 \Sigma l - 2 \lambda l = 0 \Rightarrow (\Sigma - \lambda I) l = 0$$

Se supone que  $\Sigma_{p \times p}$   $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p \geq 0$  con autovectores asociados  $e_1, e_2, \dots, e_p$  y como  $l' \Sigma l = \lambda l' l = 1$ ,  $Var(l' \Sigma l) = \lambda$  y tomando  $l = e_1$  que corresponde al mayor autovalor, se resuelve el problema optimización propuesto, de modo que la primera componente principal queda expresada como  $Y_1 = e'_1 X$  y  $Var(Y_1) = \lambda_1$

La racionalidad del procedimiento anterior puede aplicar para hallar la  $r(+1)$ -ésima componente.

Para este caso se tiene que  $Y_{r+1} = l' X; l' l = 1; l' \Sigma e_i = 0, i = 1, \dots, r$

$$\phi_{r+1}(l) = l' \Sigma e_1 - (\lambda' l - 1) - 2 \sum_{i=1}^r v_i l' \Sigma e_i$$

Se demuestra que  $\lambda_i \neq 0, i = 1, \dots, r$  el problema conduce a  $v_i = 0, i = 1 \dots r$  }. De esta manera el sistema que resuelve el problema de maximización viene dado por:

$$\{2\Sigma l - 2\lambda l = 0, \Sigma l - \lambda l = 0, \Sigma l - \lambda l = 0\}$$

Pueden ocurrir dos casos:

Si  $\lambda_{r+1} \neq 0, \lambda = \lambda_{r+1}, l = e_{r+1}$  y la  $r(+1)$ -ésima componente principal será:

$$Y_{r+1} = e'_{r+1} X$$

Y la varianza  $Var(Y_{r+1}) = \lambda_{r+1}, \lambda$

En el otro caso posible cuando  $\lambda_{r+1} = 0, \neq 0, i \neq r + 1$  se toma la combinación lineal de  $\alpha_{r+1}$  y  $\alpha_{r+1}$  y  $\alpha_i$  para la que  $\alpha_i \neq 0$ .

### 3.2.1. Correlación entre Componentes y Variables Originales

Sea una cantidad  $p$  de componentes principales,  $Y_j$  asociadas al vector aleatorio  $X$  de matriz de covarianzas autovalores-autovectores. Se tiene entonces que

$$Cov(Y_i, X_k) = Cov(e'_i X, h'_k X) = \lambda_i e_{ki}$$

Postúlese además que la componente  $k$ -ésima del eigenvector denotada por  $e_i$  proporciona la componente principal. Entonces se tiene que:

$$\rho_{YX_k} = \frac{e_{ki} \sqrt{\lambda_i}}{\sigma_k} \quad i, k = 1 \dots p$$

Siendo  $Var(X_k) = \sigma_k$  y  $Var(Y_i) = \sigma_{kk}$

La componente  $k$ -ésima del eigenvector denotado por  $e_i$  que proporciona la componente principal, cuantifica la importancia relativa que la variable primitiva  $k$ -ésima posee sobre la componente principal o variable sucedánea. En consecuencia, cuanto mayor sea  $|e_{ki}|$  mayor es la correlación entre la variable original y  $Y_i$ .

### 3.2.2. Determinación de las Puntuaciones

Estas puntuaciones pueden ser halladas mediante la descomposición de los valores característicos en la matriz de covarianza muestral  $S = P\Lambda P$ . Los valores propios de  $S$  están contenidos, en orden descendente, en  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$

Los vectores  $p$ -dimensionales  $x$  son transformados en un vector de "scores" denotado por:

$$y = P'(x - \bar{x})$$

siendo  $P$  la matriz de cargas.

Se postula, además, que la suma de las varianzas de las variables o inercia total de la nube de puntos es equivalente a la sumatoria de las varianzas de las componentes principales. En este sentido, la proporción de inercia explicada por una componente  $i$ -ésima es:

$$\lambda_i / \sum_{i=1}^p \lambda_i = \lambda_i / \sum_{i=1}^p V(x_i)$$

siendo  $\sum_{i=1}^p V(x_i) = \text{traza}(v)$ , una expresión de la medida de variabilidad asociada a las variables originales.

### 3.2.3. Test Aplicados en el Análisis de Componentes Principales

El objetivo cardinal de la aplicación de las componentes principales es reducir las dimensiones de las variables originales, pasando de  $p$  variables originales a  $m < p$  componentes principales.

El problema que se plantea es cómo fijar la cantidad de componentes que deben ser retenidas. Aunque para la extracción de las componentes principales no es imprescindible el planteamiento de un modelo estadístico previo, algunos de los criterios para determinar cuál debe ser el número óptimo de componentes a retener requieren la formulación previa de hipótesis estadísticas.

Entre los test más relevantes se reportan:

#### 3.2.3.1. Test de Bartlett y de Bartlett-Lawley.

Parte de la hipótesis nula que los  $p - k$  eigenvalores más pequeños son iguales. Es decir que  $H_0 : \lambda_{k+1} = \lambda_{k+2} = \dots = \lambda_{k+p}$ . El contraste se realiza mediante el estadígrafo:

$$\left( N - k - 1 - \frac{2q + 1 + \frac{2}{q}}{6} \right) (-\ln|S| \sum_{j=1}^k \ln l_{(j)} + q \ln l)$$

Bajo la hipótesis nula, sigue una distribución  $\chi^2$  con  $\frac{1}{2}(p - k - 1)(p - k + 2)$  grados de libertad, asintóticamente.

El test de Bartlett-Lawley que es una versión corregida del anterior, depende del valor de:

$$l = \frac{1}{q} (\text{tr}(s) - \sum_{j=1}^k l_{(j)})$$

Se introduce una corrección  $l^2 \sum_{j=1}^k \frac{1}{(l_{(j)} - l)^2}$  y, por tanto, se obtiene  $\chi^2$  con  $\frac{1}{2}(q + 2)(q - 1)$  grados de libertad, asintóticamente.

#### 3.2.3.2. Test de Anderson.

Para la retención de una cantidad óptima se formula la prueba de Anderson cuya hipótesis nula es  $H_0 = \lambda_{n+1} = \lambda_{n+2} = \dots = \lambda_n = 0$  y tal es rechazada si para un  $m$  fijado  $\chi^2$  es significativo al nivel de confianza prefijado. El estadístico de prueba derivado del método de cociente de verosimilitudes es:

$$\chi^2 = -(N-1) \sum_{n=q+1}^{q+r} \ln l_i + (N-1)v \ln \left[ \frac{1}{v} \sum_{n=q+1}^{q+r} \ln l_i \right] \sim \chi_{\frac{(r)(r+1)}{2}-1}^2$$

En otros términos, la mecánica interna del contraste posibilita la elección y descarte de componentes siguiendo este criterio: se contrasta la igualdad de las raíces características consecutivas y se descarta la inclusión de componentes cuando no existen direcciones de máxima variabilidad; en consecuencia, la variabilidad en las últimas  $n + m$  dimensiones es esférica.

### 3.2.3.3. Test de ACP sobre $\mathbb{R}$ .

El test de ACP sobre  $\mathbb{R}$  fundamenta su análisis en la matriz de correlaciones  $\mathbb{R}$  bajo la hipótesis nula  $H_0 = \rho_{k+1} = \rho_{k+2} = \dots = \rho_p = \rho$ ;  $k < p$ . Siendo  $\rho_i$  son las raíces características de  $\mathbb{R}$ , estimados por  $\hat{\rho}_i$  que son las raíces características de  $\hat{\mathbb{R}}$ . Considérese el estadístico que se comporta asintóticamente, con un orden de aproximación de  $1/N$  según una  $\chi^2$ :

$$(N-1) - \left\{ \ln \frac{|\hat{\mathbb{R}}|}{\prod_{i=1}^k \hat{\rho}_i} + q \ln \text{tr } \hat{\mathbb{R}} - \frac{\sum_{j=1}^k \hat{\rho}_j}{q} \right\}; q = p - k$$

Los grados de libertad vienen dados por:

$$p^* = \frac{1}{2}(q-1)(q+2) - \frac{1}{q} \left\{ (q-1) + \rho \sum_{i=1}^p \sum_{j=1}^p c_{ij}^2 p_{ij}^2 \sum_{i=1}^p \sum_{j=1}^p c_{ii} c_{jj} p_{ij}^2 \right\}$$

con  $c_{ij} \in C = I - \theta_1 \theta_1'$

A través del Análisis de Componentes las variables originales se proyectan en nuevas variables independientes, denominadas componentes principales, combinaciones lineales de las variables primitivas. La varianza de cada componente es equivalente a su eigenvalor  $\lambda_i$ . En caso tal que las variables se hallen correladas entre sí, una fracción mayoritaria de la variabilidad inherente a la nube de datos será capturada por los componentes iniciales mientras que la variación remanente es regularmente considerada como ruido aleatorio. Las componentes principales son ortogonales una a otra y su ordenación se realiza en función de las fracciones de varianza explicada.

Las especificaciones de ingeniería y los valores target de la  $i$ -ésima componente  $PC_i$  son:

$$\begin{aligned} LSL_{PC_i} &= u_i^T LSL \\ USL_{PC_i} &= u_i^T USL \\ T_{PC_i} &= u_i^T T \end{aligned}$$

donde  $i = 1, 2, 3 \dots n$  es el  $i$ -ésimo componente. Las especificaciones multivariadas originales denotadas por  $USL$  y  $LSL$ , se proyectan en cada componente.

El indicador de capacidad multivariante obtenido para el caso presente se define teóricamente como la mediana geométrica ponderada de los índices de capacidad evaluados en los componentes principales, siendo las ponderaciones los autovalores  $\lambda_n$  descritos anteriormente (Santos-Fernández & Scagliarini, 2012):

$$MC_p = \left[ \prod_{i=1}^m C_{p;PC_i}^{\lambda_n} \right]^{\frac{1}{\sum_{n=1}^m \lambda_n}}$$

Así, el indicador de capacidad univariado para el  $i$ -ésimo componente principal es igual a (Santos-Fernández, 2012):

$$C_{p,PC_i} = (USL_{PC_i} - LSL_{PC_i})/6\sigma_{PC_i}$$

El límite inferior de confianza aproximado  $100(1 - \alpha)\%$  se formula de acuerdo a la siguiente expresión:

$$\left[ \left( \prod_{i=1}^m \hat{C}_{p,PC_i} \sqrt{\frac{\chi_{n-1,\alpha}^2}{n-1}} \right)^{\frac{1}{\sum_{n=1}^m \lambda_n}} \right]$$

Para la retención de una cantidad óptima de componentes se es renuente a la aplicación de cánones arbitrarios fijados a priori. Para ello se formula la prueba de Anderson cuya hipótesis nula es  $H_0 = \lambda_{n+1} = \lambda_{n+2} = \dots = \lambda_n = 0$  y es rechazada si para un  $m$  fijado  $\chi^2$  es significativo al nivel de confianza prefijado. El estadístico de prueba derivado del método de cociente de verosimilitudes cuya mecánica interna posibilita el contraste posibilita la elección y descarte de componentes siguiendo este criterio: se contrasta la igualdad de las raíces características consecutivas y se descarta la inclusión de componentes cuando no existen direcciones de máxima variabilidad; en consecuencia, la variabilidad en las últimas  $n + m$  dimensiones es esférica.

El resumen del procedimiento se registra en la Tabla 4 que contiene información sobre el modelo de componentes principales: la proporción acumulada de la variabilidad explicada por los componentes principales consecutivos, valores propios, también denominados valores característicos o raíces latentes, corresponden a las varianzas de los componentes principales; se presenta además la significación estadística de los componentes y el número de iteraciones para la extracción. Se han retenido 2 componentes vía Test de Anderson para la consecuente obtención de tres indicadores de capacidad que admiten una interpretación similar a aquellos consignados en la sección precedente. Los dos componentes extraídos representan aproximadamente el 82% de la variabilidad exhibida por las variables primitivas.

**Tabla 4. Resumen de indicadores basados en análisis de componentes principales.**

Componente	R <sup>2</sup> X	R <sup>2</sup> X(Acumulado)	Eigenvalores	Q <sup>2</sup>	Q <sup>2</sup> (Acumulado)	Significancia	Iteraciones
1	0,5876	0,5876	3,9296	0,565	0,565	Verdadero	4
2	0,3311	0,8187	1,6577	0,313	0,878	Verdadero	5

Fuente: Elaboración propia.

Nótese que la primera componente principal relaciona tres variables y la segunda componente dos variables, aunque débilmente (sus correlaciones son próximas a 0,70). Percíbase adicionalmente que la totalidad de indicadores de capacidad consignados, asume valores inferiores a la unidad, lo que insinúa que la variación del proceso es relativamente grande y, por tanto, la media del proceso se distancia del centro de especificaciones y no se halla alrededor del objetivo (Wang, 2014).

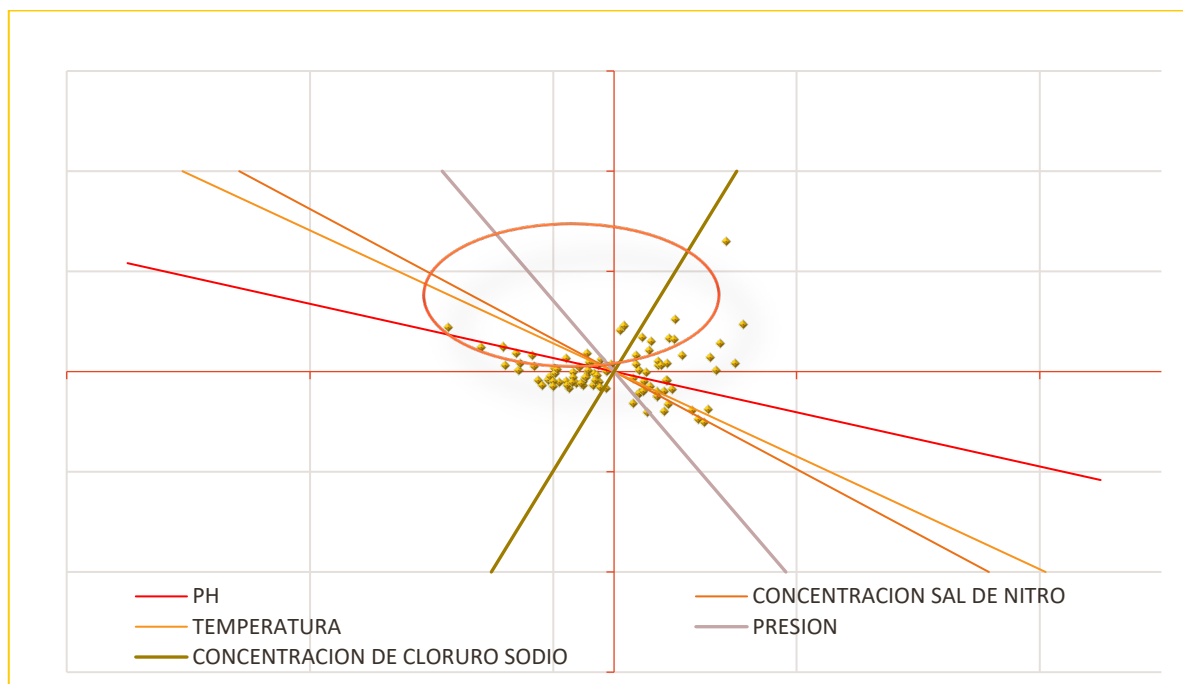
Al igual que en el epígrafe anterior se muestra un gráfico de control multivariante. Es perceptible que el límite superior de control en el diagrama corresponde a una región elíptica en el espacio bidimensional. Este tipo de gráficos es insensible frente a variaciones marginales en el vector de medias del proceso y frente a cambios individuales en una componente o un subconjunto de ellas. En este caso, la región de control es sensiblemente más amplia contrastada con la reportada en la Figura 2. De ahí que se detecte una cantidad más baja de señales fuera de control. No obstante, bajo este marco de análisis el proceso sigue caracterizándose como inestable e inhábil para satisfacer las especificaciones predefinidas.

**Tabla 5. Análisis de componentes por cada variable.**

VARIABLE	Loadings	
	COMPONENTE 1	COMPONENTE 2
PH	0,909578	0,383389
CONCENTRACIÓN DE SAL DE NITRO	0,854524	0,499746
TEMPERATURA	-0,885991	0,691090
PRESIÓN	-0,911981	0,610950
CONCENTRACIÓN DE CLORURO DE SODIO	0,869128	-0,282035
Mcpm		0,73
Mcp		0,71
Mcpk		0,75

Fuente: Elaboración propia.

**Figura 2. Elipse de Control 2.**



Fuente: Elaboración propia.

### 3.3. Otros indicadores de capacidad multivariante.

Aunque se ha exhibido un conglomerado de relevantes métricas de capacidad, los indicadores que se señalan en la Tabla 6, comparan el desempeño conjunto de la totalidad variables con el vector de especificaciones, suministrando una escrupulosa exploración panorámica que permite determinar la probabilidad de que las observaciones se hallen en localizaciones externas a los límites prefijados. A continuación, se explican las bases teórico-conceptuales que sustentan su aplicación en el caso presente:

El Nivel Z equipara la proporción de observaciones fuera de los límites de especificación con un valor de la distribución normal estandarizada equivalente a 2,39 por lo que se halla moderadamente desviado del ideal (4 unidades). Esta métrica se calcula utilizando la desviación estándar a corto plazo del proceso y es una medida del funcionamiento del proceso bajo la condición de que la variación a corto plazo persiste constante. En la medida que se dilata el desplazamiento Z, más oportunidades existen para suprimir o amortiguar el efecto de las causas de variación especial exhibida entre los subgrupos. Para el caso presente se presupone un desplazamiento o drift de la media de  $1,5\sigma$ .

**Tabla 6. Otros indicadores de capacidad.**

	FUERA ESPECS.	FUERA ESPECS.	DPM	ÍNDICE	ESTIMADO
	Observados	Estimados	Estimados	MCP	0,71
PH	0,0%	0,0%	0,0	MCR	125,06%
CONCENTRACIÓN DE SAL DE NITRO	1,81%	0,087601%	876,01	DPM	8270,43
TEMPERATURA	1,81%		232,12	Z	2,39
PRESIÓN	3,636%	0,717%	7170,16	SQL	3,89
CONCENTRACIÓN CLORURO DE SODIO	0,0%	0,0%	0,0		
CONJUNTO	7,27%	0,82%	8270,43		

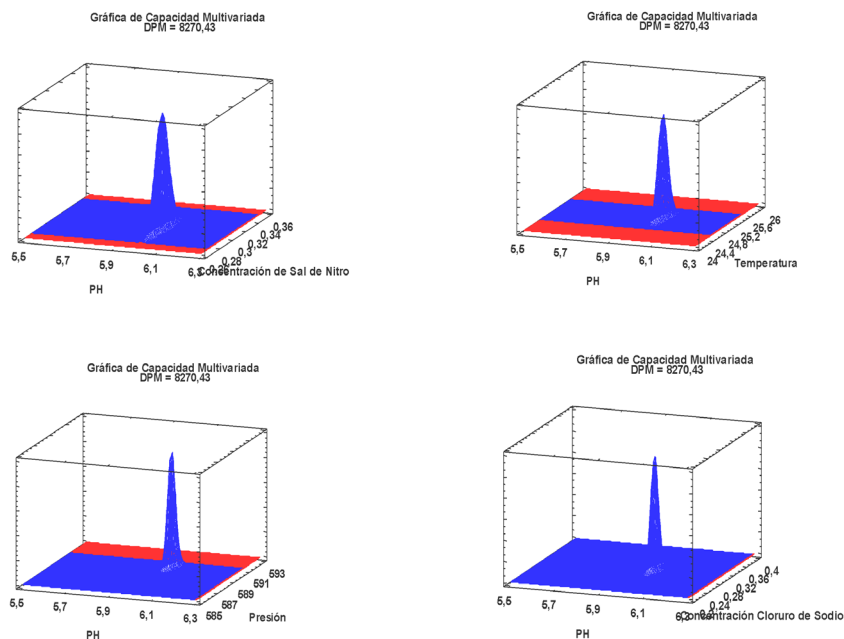
Fuente: Elaboración propia.

El índice  $MCP = Z\left(\frac{k}{2}\right)^{-1}$  es la contraparte del indicador univariado  $C_{pk}$  se calcula como el cociente entre el valor de la variable normal estandarizada  $Z$  y el múltiplo sigma  $K$  dividido en 2. El valor adoptado por este indicador constituye una prueba fehaciente para argüir que el proceso no es capaz de cumplir con las especificaciones técnicas, pues es inferior al valor mínimo deseable equivalente a 1,33.

El índice  $MCR = 100\%\left(\frac{k}{Z}\right)^{-1}$ , puede expresarse como el inverso del índice  $MCP$ , multiplicado por 100% y presenta en términos porcentuales la porción de variabilidad admisible ocupada por la variabilidad natural inherente al conjunto de datos. Tal fracción es superior a la amplitud de la variación tolerada por la banda de especificaciones.

Por otro lado, el nivel de calidad sigma, equivalente a 3,89, indica que el proceso posee un deficiente nivel de consistencia y un vasto margen de mejora si el objetivo es alcanzar un rendimiento de clase mundial, en cuyo caso el indicador sería equivalente a 6.

**Figura 3. Gráfico combinado de capacidad.**



Fuente: Elaboración propia.

Nótese que en la Figura 3 se reporta un gráfico combinado de capacidad que conjunta todas las variables analizadas. En todos los casos contemplados, se ha representado cada variable versus la Potencial de Hidrogenación (PH) -esto obedece un criterio subjetivo, no exento de cierto grado de arbitrariedad-. El área sombreada en azul de la distribución normal multivariante corresponde a la localización donde las observaciones se hallan dentro de los márgenes de especificación definidos. Irrebatiblemente se sostiene que los resultados aquí proporcionados, armonizan con los plasmados en la Tabla 4. Se observa por ejemplo que las variables PH y Concentración de Cloruro de Sodio no contienen observaciones fuera de los límites y que existen medidas de la variable Presión que exceden el límite superior de especificación.

#### **4. Conclusiones.**

El control e ingeniería de procesos conjuga operaciones que facultan a las organizaciones para llevar a cabo un examen crítico sistemático de los contenidos de trabajo prescritos y modificarlos en función de la generación de valor agregado para los clientes, a los cuales se proporciona los bienes manufacturados. El análisis de capacidad multivariante, contenido en esta disciplina, se distancia de los enfoques evaluativos tradicionales en tanto que posibilita el escrutinio las características de calidad de forma conjunta y sincrónica. Aunque existen irresoluciones en torno a qué métricas de capacidad deben ser utilizadas para aprehender el carácter dinámico de los procesos productivos, con objetividad categórica, puede argumentarse que la preeminencia del enfoque multivariado reside en su capacidad para incorporar dentro del análisis, la estructura de correlaciones intravariante y las interrelaciones insertas en él, información estimable en sobremanera, que pasa inadvertida al orientar el modelo de análisis del proceso al contexto univariante.

Se ha trazado como objetivo de este estudio un análisis comparativo entre diversos indicadores de capacidad multivariado en el contexto de la industria manufacturera bajo la asunción de que las características de calidad siguen una distribución normal multivariante y guardan cierto grado de correlación. En el marco de la metodología desarrollada para la medición del desempeño del proceso, se propone la aplicación de tres enfoques: uno basado en las regiones de tolerancia del proceso, otro basado en el análisis de componentes principales y, en última instancia, una gama de indicadores que bien pueden ser entendidos como análogos a las métricas tradicionales univariantes. Los resultados reportados en apartados precedentes indican que el proceso no es marginalmente capaz para cumplir con las especificaciones técnicas predefinidas, ya que, a excepción de las variables “Potencial de Hidrogenación” y “Concentración de Cloruro de Sodio”, las demás contienen observaciones fuera de aquellas.

En síntesis, aunque no existe un consenso universal sobre la preeminencia de los índices aquí señalados respecto a otros propuestos en la literatura especializada, es perentorio evaluar la robustez de otra gama de indicadores en ambiente real y efectuar un análisis comparativo con las traídos a colación en el presente caso de estudio. Aunque es improcedente colocar en tela de juicio el carácter representativo del presente caso de estudio, toda tentativa de generalización analítica de los resultados obtenidos y la extrapolación de los elementos centrales del diseño metodológico a otros casos empíricos deber ser efectuada con cautela, pues el tamaño muestral es reducido y sólo se ha evaluado una fracción reducida de variables que inciden directamente en el performance global del proceso analizado.

Durante el proceso investigativo se detectaron vacíos teóricos y epistemológicos que convendría evaluar, ya que, aunque se registran estudios relacionados con la aplicación de herramientas de monitoreo multivariante, es imprescindible diseñar casos de aplicación en ambiente real para validar los principios metodológicos propuestos en la literatura especializada. En aras de ampliar el abanico de futuras líneas de investigación se insta a validar y extrapolar el estudio a entornos no productivos, esto es, aquellos donde se generan bienes no tangibles y suelen ser de naturaleza no determinística. Son necesarios también estudios específicos que contemplen el diseño y evaluación métricas de capacidad en procesos de prestación de servicios y/o manufactureros que no satisfagan hipótesis elementales como autocorrelación, homocedasticidad o linealidad. Se añaden al listado de investigaciones futuras aquellas



que involucran el uso de técnicas meta-heurísticas para la estimación de modelos multivariantes a efectos de caracterizar y controlar la dinámica inherente a los procesos fabriles.

### Agradecimientos

Los autores agradecen inconmensurablemente al grupo de investigación CIPTEC por la asistencia dispensada, así como también, a los pares evaluadores anónimos por las orientaciones propuestas durante el proceso de revisión.

### Referencias

- Caicedo, N., Jiménez, E., Vélez, S., & Wilches, M. (2015). Metodología para cálculo de un indicador de capacidad de procesos multivariado para sistemas de gestión integrados. *Revista Universitaria RUTA*, 17(1), 21-30.
- Cuamea, G., & Rodriguez, M. (2014). Propuesta para evaluar la capacidad de procesos de manufactura multivariados. *Ingeniería Industrial*, 13(2), 35-47.
- Dharmasena, L., & Zeepongsekul, P. (2015). A new process capability index for multiple quality characteristics based on principal components. *International Journal of Production Research*, 54(15), 1-17.
- Hernández, C., & Da Silva, F. (2016). Aplicación del control estadístico de procesos (CEP) en el control de su calidad. *Tecnología Química*, 36(1), 104-116.
- Herrera, J., Herrera, G., & Rahmer, B. (2017). Control Estadístico de Procesos para datos Correlados Serialmente. Un Caso de Estudio. *International Conference on Industrial Engineering and Operations Management*, (pp. 891-904). Bogotá.
- Lea, W., Wu, C., & Wu, C. (2015). Estimating process capability index Cpk: classical approach versus Bayesian approach. *Journal of Statistical Computation and Simulation*, 85(10), 2007-2021.
- Novaes, A. G., Lima Jr, O., Carvalho, C.C., & Aragão Jr, D.P. (2015). Process capability index Cpk for monitoring the thermal performance in the distribution of refrigerated products. *Production*, 26(1), 54-65.
- Salazar, E., & Fermín, J. (2016). Un índice de capacidad de procesos para distribuciones multivariadas normales de variables correlacionadas y no correlacionadas. *Ingeniería Industrial*(34), 57-73.
- Santos-Fernández, E. (2012). *Multivariate Statistical Quality Control Using R*. New York: Springer.
- Santos-Fernández, E., & Scagliarini, M. (2012). MPCl: An R Package for Computing Multivariate Process Capability Indices. *Journal of Statistical Software*, 47(7), 1-15.
- Vásquez, M., Ramírez, G., & García, T. (2016). Un índice de capacidad multivariante basado en la probabilidad de no conformidad, una aplicación al monitoreo de calidad de un ciclo de agua clarificada. *Revista Ingeniería UC*, 23(3), 319-326.
- Verbel, A., Herrera, R., & Maestre, K. (2016). Aplicación de nuevas propuestas multivariantes para medir la capacidad de un proceso. *Ingeniare*(21), 33-44.
- Wang, F.-K. (2014). *Multivariate Process Capability Indices, Comparison of*. Wiley StatsRef: Statistics Reference Online.

Zhang, M., Wang G, A., He, S., & He, Z. (2014). Modified Multivariate Process Capability Index Using Principal Component Analysis. *Chinese Journal of Mechanical Engineering*, 27(2), 249-259.