

Monitoreo y Análisis de un Proceso de Elaboración de una Bebida Carbonatada mediante 3-Way PLS

Jorge A. Saavedra^(1,3), Luis A. Puente⁽²⁾, Beatriz Cancino^(1,3) y Carolina Astudillo⁽¹⁾

(1) Pontificia Universidad Católica de Valparaíso, Escuela de Ing. de Alimentos, Waddington 716, Playa Ancha, Valparaíso-Chile. (email: jorge.saavedra@ucv.cl)

(2) Universidad de Chile, Departamento de Ciencia de los Alimentos y Tecnología Química, Vicuña Mackenna 20, Santiago-Chile.

(3) Centro Regional de Estudios en Alimentos Saludables (CREAS), Blanco 1623, of. 1402, Valparaíso-Chile.

Recibido Dic. 10, 2009; Aceptado Feb. 11, 2010; Versión Final recibida Abr. 22, 2010

Resumen

El trabajo presenta el estudio, modelado y análisis de un proceso continuo de elaboración de bebidas carbonatadas mediante un método multivariante basado en el enfoque 3-Way PLS. Se analizó el comportamiento del proceso en relación a un conjunto de alteraciones tales como transiciones, detenciones no programadas, fallas, reinicios) que se registraron en un plazo de 4 meses, analizando un total de 1938 observaciones y midiendo 8 variables del proceso. Se modeló el proceso, filtrando el alto grado de autocorrelación y correlación cruzada contenida en las variables (en especial T1, T2 y T3), logrando determinar las variables responsables del comportamiento del proceso ordenado como lotes secuenciales. De esta forma se evaluó el efecto y grado de alteración que esto causó en el patrón de variabilidad y en el producto final.

Palabras clave: Control de Procesos Multivariante, Procesos con Transición, 3-Way-PLS, Proyecciones Latentes

Monitoring and Analysis of a soft-drinks Bottling Process using 3-Way PLS

Abstract

The present research deal with the modeling an analysis of a continuous soft-drinks bottling process by a multivariate approach based on the 3-Way PLS. For this purpose the process behavior was studied, related to a set of off-processing, transitions grades and/or abnormal star-ups, detected in four month process period. A total of 1938 cases were investigated, measuring 8 process variables. Finally, the process was modeled, filtering the high grade of autocorrelation and cross-correlation within the studied variables (particularly T1, T2 and T3). Thereby, the responsible variables for that behavior were detected; thus the effect and grade of disturbance over the process behavior and final product were estimated.

Keywords: Multivariate Process Control, Transition Process, 3-Way-PLS, Latent Projections

INTRODUCCIÓN

El Control Estadístico de Procesos (Statistical Process Control, MPC) se ha planteado como una herramienta para la diferenciación entre causas comunes y causas asignables, las primeras asociadas a variabilidad propia del proceso, y que constituyen una condición permanente en cualquier momento del control, por lo que se asocian a condiciones de operación normales (patrón de variabilidad estable y predecible en el tiempo ó proceso bajo control estadístico). Por otro lado, las causas asignables poseen una condición ocasional y, en general no predecible, asociándose a anomalías no previstas en el proceso (Wierda, 1994; Kourti et al., 1996; Ryan, 2000; Woodall, 2000; Kotz y Johnson, 2002; Ferrer, 2004, Ferrer, 2007). En este sentido, el SPC establece un sistema de observación permanente e inteligente, capaz de detectar precozmente la aparición de alteraciones en el proceso, malfuncionamiento de equipos tan pronto como se produzca, aislando dichas causas y eliminándolas, de modo de ayudar a identificar su origen (Romero, 2000; Chou, 2003, Saavedra, 2008). Para ello el SPC se ha basado en el análisis de datos mas que en el desarrollo de modelos, básicamente por la dificultad que representa el ajuste de modelos que sean capaces de soportar eficazmente situaciones como falta o pérdida de normalidad, independencia y/o autocorrelación (Kano et al., 2001). En este sentido la idea del SPC es capturar la variabilidad a partir del estado de operación de un proceso, basándose en un historial fiable de información previa de dicho proceso (Lopes y Menezes, 2004).

En muchos procesos el nivel de calidad final del producto se define en base a múltiples variables y a las interacciones entre ellas, por lo que resulta imprescindible monitorearlas simultáneamente de forma de evaluar el efecto conjunto del comportamiento global del proceso, a diferencia del enfoque univariante clásico (Mason, 1997). A esto se suma en los últimos años un cambio radical en la configuración de los procesos industriales: aumento generalizado del numero de variables a controlar, grandes cantidades de datos registrados, aumento en la velocidad de producción, alto grado de automatización en plantas, ausencia o falla de sensores, existencia de fuertes relaciones no lineales y la ausencia de modelos mecanicistas que caractericen fielmente la totalidad del proceso (Andersson, 2000; Kruger, 2004; Ferrer, 2007). Esta situación ha provocado que los procesos respondan cada vez más erráticamente ante el uso de técnicas tradicionales de control, por lo que las técnicas y gráficos clásicos de control univariante no son del todo recomendables. Una opción a este tipo de condiciones y problemas ha sido el Control Estadístico Multivariante de Procesos (MSPC) basado en métodos de Proyecciones Latentes. Estos se caracterizan por ser capaces de abordar simultáneamente gran cantidad de datos con alta dimensionalidad y colinealidad, pudiendo extraer información sobre la dirección de las variaciones del proceso (Bro, 2002; Qin, 2003). Entre estas herramientas destacan, el Análisis de Componentes Principales (PCA) y la Regresión de Cuadrados Mínimos Parciales (PLS), así como los denominados métodos Multi-Way dada la configuración en tres dimensiones con que procesan la información (Kroonenberg, 2008).

Un caso particular de control, poco abordado en la literatura, lo constituye el comportamiento de procesos por lotes o batches durante transiciones: inicios de proceso, cambios de etapa y/o grados de calidad, reinicios debido a anomalías técnicas y/o ajustes, situaciones frecuentes en procesos industriales (Duchesne, 2003; Kourti, 2003a; Kourti, 2005), y que pueden traducirse en importantes pérdidas de tiempo de producción, producto defectuoso e inconsistente reproducibilidad en el grado de calidad del producto fabricado, en especial frente a enfoques poco eficientes de control (Kourti, 2003b).

En este contexto, el presente trabajo tiene por objetivo analizar y modelar un proceso continuo de envasado de bebidas carbonatadas, configurado como un conjunto secuencial de batches o lotes de duración variable, definidos por un grupo de anomalías (detenciones y reinicios no programados principalmente) identificadas a lo largo del proceso. Para ello se utilizó el método 3-Way PLS de forma de evaluar el efecto de dichas transiciones como fuente de causas asignables (alteraciones al patrón de variabilidad natural).

MATERIALES Y METODOS

La investigación se realizó en una línea de producción de una bebida carbonatada de un formato PET 2.0 litros retornable con capacidad de 13,500 unidades por hora. Se aplicó un muestreo secuencial cada 15 minutos incorporando las etapas de mezclado de jarabe-agua, adición de CO₂ y envasado. En total se recopilaron 4608 registros en un periodo de producción de cuatro meses. Posterior a un análisis de datos faltantes e identificación de condiciones anómalas, se definió un conjunto de *condiciones de transición* (en general correspondieron a detenciones no programadas, anomalías técnicas y ajustes del proceso). Con esto, el set de datos se redujo a 1938 casos completos, los que se dispusieron en 23 lotes o batchs cronológicamente secuenciales midiendo en cada uno de ellos 8 variables continuas: temperaturas de mezcladora de jarabe (T1 y T2), temperatura de lavadora (T3), incrementos de selladora (INC y AP), grados Brix de la mezcla (Brix), contenido de CO₂ en la mezcla (CO₂) y contenido neto del envase (Cont Neto).

Para el procesamiento de los datos, se modeló el proceso mediante el método 3-Way-PLS (3 Way Partial Least Square Regression) basado en el algoritmo NIPALS (*non linear iterative partial least squares*) (Wold, 2001). Como forma de contrastar la información, se utilizaron 2 herramientas de control estadístico multivariante de uso más común en la industria: gráfico T²-Hotelling y gráfico MEWMA (multivariate exponential weighted moving average) según la metodología propuesta por Montgomery (2001). Para el análisis de los datos se utilizó Statgraphics Centurión XVI (StatSoft, 2009) y SIMCA-P+ v12 (UMETRICS, 2008).

RESULTADOS Y DISCUSIÓN

A partir de los datos obtenidos se desarrollaron los gráficos multivariantes T²-Hotelling y gráfico MEWMA de forma de establecer un comportamiento inicial de referencia (Figura 1a y 1b). Para ambos gráficos se utilizó una conformación de 1938 individuos (n=1), utilizando el método de diferencias sucesivas para la estimación de la matriz de varianzas-covarianzas dada su mayor fiabilidad al evaluar procesos continuos secuenciales, mas que el criterio de estimación conjunta (pooled variance matriz) (Mason y Young, 2001).

El límite de control para ambos gráficos fue de T²= 23.48. El desempeño del gráfico T²-Hotelling mostró que mas del 90% de los puntos estaban fuera de control; mientras que el grafico MEWMA ($\lambda=0.25$) presentó mas del 60% de los puntos fuera de control. No obstante una inspección a los datos originales invalida el diagnostico, dado que las observaciones evaluadas caso a caso indicaron que no sobrepasaban la especificación preestablecida. Una interpretación a esto puede asociarse al comportamiento de las temperaturas T1, T2 y T3, las que presentan un alto grado de autocorrelación y correlación cruzada con el resto de las variables, debido a que dichas variables están sometidas a control automático mediante actuadores, eliminando su condición de variables aleatorias, lo que se traduce en una alta proporción de falsos positivos en ambos gráficos. Esta situación puede generar el alto grado de autocorrelación detectada.

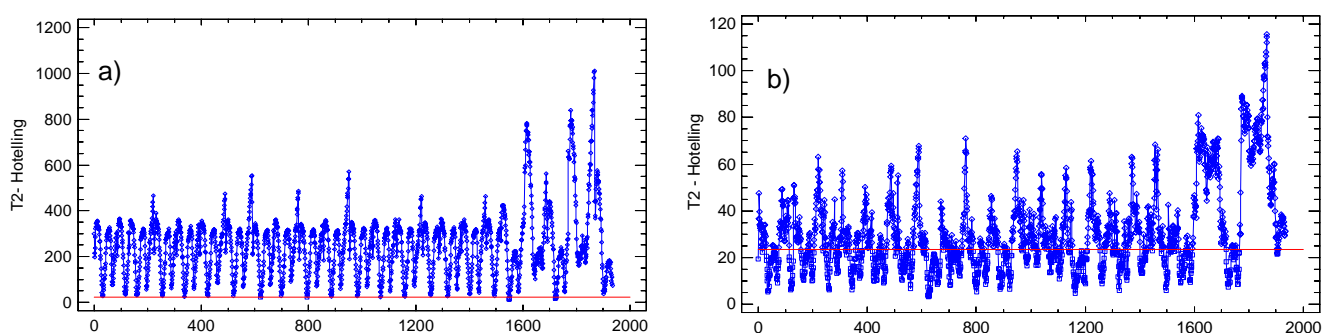


Fig. 1: a) Gráfico T² - Hotelling b) Gráfico MEWMA

Posteriormente se aplicó el gráfico basado en componentes principales (MSPC-PCA) de modo de interpretar la variabilidad conjunta de los 23 lotes evaluados e intentar filtrar el alto grado de autocorrelación detectado entre las variables (Fig. 2). El método extrajo 3 factores que explicaron en conjunto el 87% de la inercia, mientras que la capacidad de predicción fue de 48.3% lo que resulta satisfactorio para este tipo de análisis. El modelo fue validado mediante una rutina completa de validación cruzada (Cross-Validation) (Eriksson, 2006).

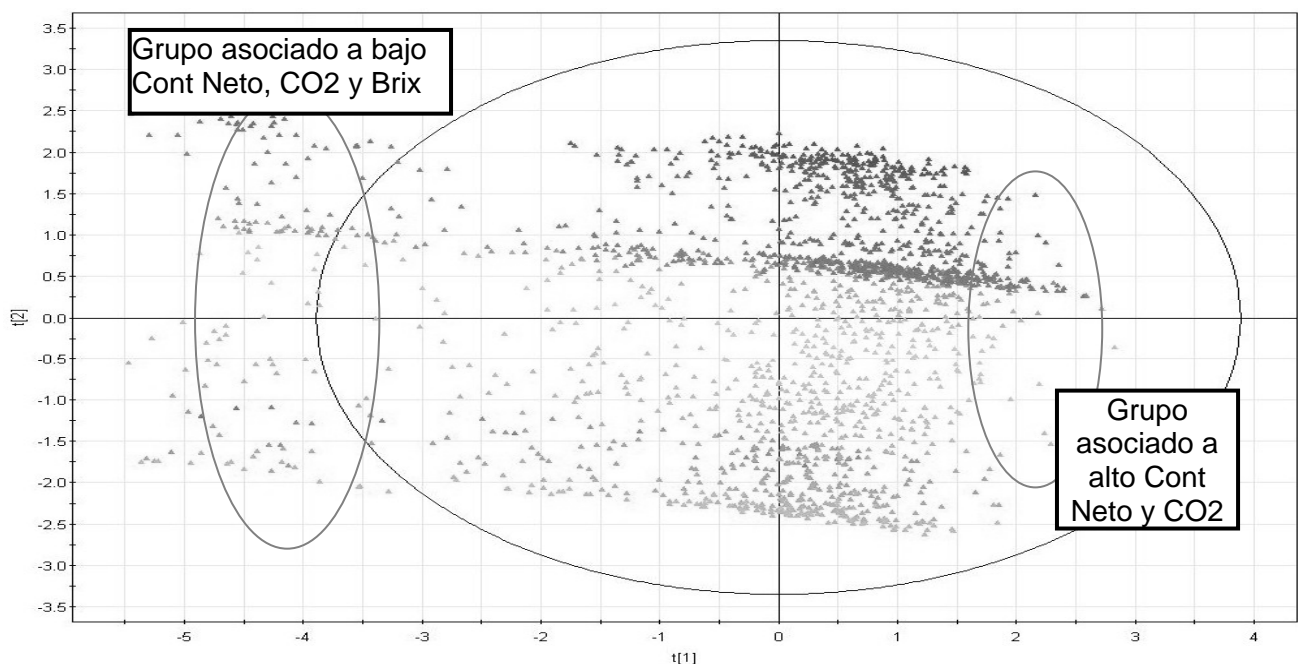


Fig. 2: Gráfico de scores en unidades de desviación estándar. Proyección de las observaciones sobre el biplano factorial.

En la Figura 2 se muestra el límite de control elíptico al 95% de confianza dentro del cual se congregan la gran mayoría de las observaciones, las que podrían denominarse bajo control estadístico (o cumpliendo con la especificación). El primer factor (t_1) ordena las observaciones según grupos característicos en relación a las variables Cont Neto, CO₂ y Brix principalmente, mientras que el segundo factor (t_2) logra diferenciar el comportamiento de las observaciones fuertemente correlacionadas con las temperaturas (T₁, T₂ y T₃). El comportamiento de los grupos de observaciones y la influencia de las variables en cada uno de ellos, para los factores t_1 y t_2 se puede apreciar en los gráficos de contribuciones de la figura 3. Los gráficos de contribución detallan como el factor t_1 está influenciado principalmente por las relaciones inversas entre Cont Neto, CO₂ y Brix, concordando con la información mostrada por el gráfico MSPC-PCA. No ocurre lo mismo con el factor t_2 , ya que se observa una relación inversamente proporcional entre las temperaturas T₁ y T₂ respecto de T₃ (Figura 3c), la que prácticamente anula su contribución en las observaciones opuestas (Figura 3d). Este comportamiento se puede asociar a la fuerte correlación que presentan T₁ y T₂, su influencia en la conformación del factor 2 y su efecto sobre el resto de las variables, mientras que al evaluar la cronología de T₃ en el proceso, se evidencian fallas de funcionamiento (una proporción mayor de datos faltantes) lo que podría justificar la correlación menor con el resto de las variables y su tendencia a presentar una contribución nula en un grupo de observaciones.

Si bien el gráfico MSPC-PCA logra describir en conjunto la situación del proceso, no es capaz de discriminar significativamente las diferencias encontradas entre los grupos detectados, no obstante constituye una importante herramienta descriptiva previa del análisis global.

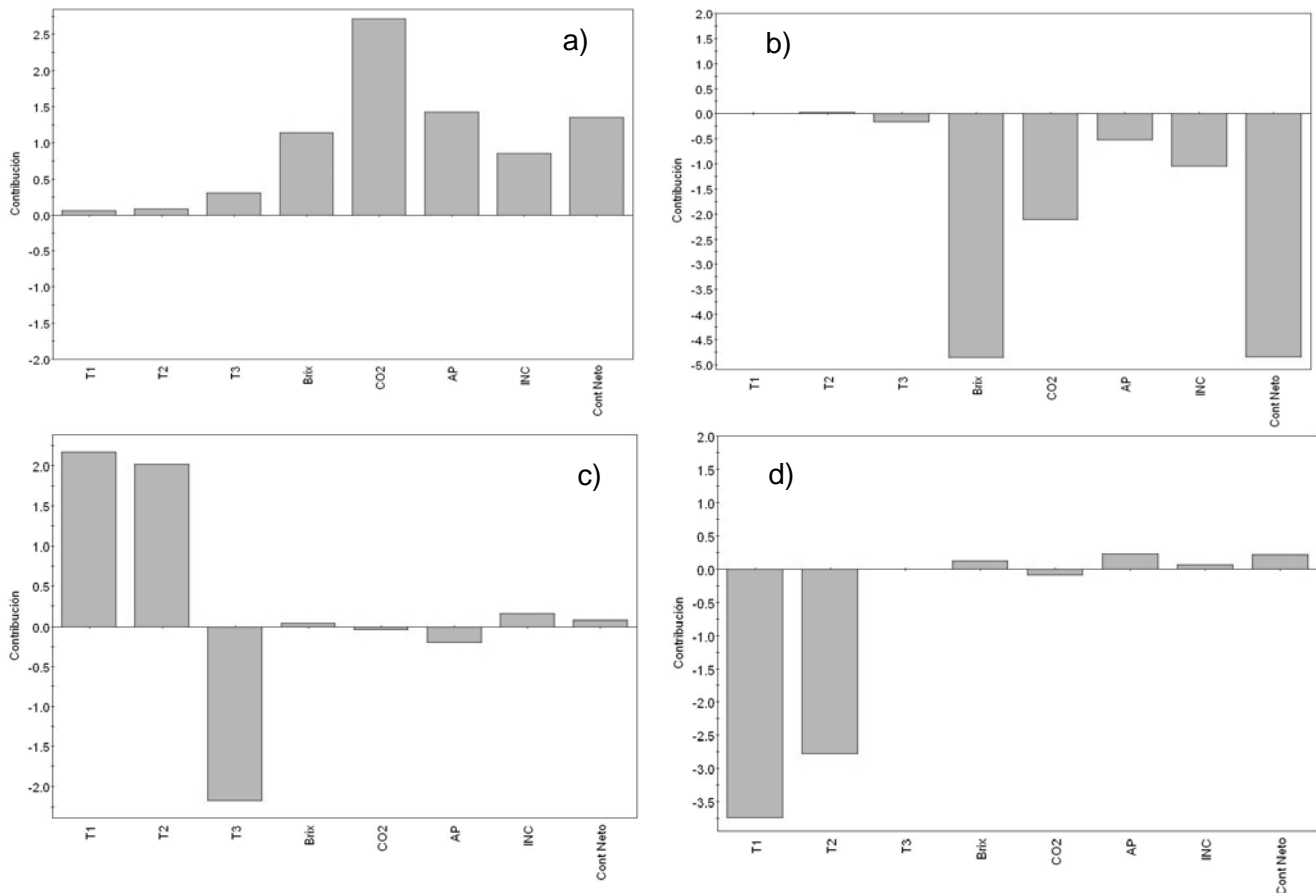


Fig. 3: a) Gráfico de contribuciones de observaciones con sobre llenado, b) Gráfico de contribuciones de observaciones con bajo llenado, c) Gráfico de contribuciones de observaciones con alta influencia de T1 y T2, y d) Gráfico de contribuciones de observaciones con alta influencia de T3. (Todos los gráficos expresados en unidades de desviación estándar).

Posteriormente, y de modo de dilucidar las supuestas relaciones de transición presente entre los diferentes lotes, se procedió a desarrollar el modelo 3-Way-PLS considerando los 23 lotes conformados. Dadas las condiciones de transición presupuestas, los 23 lotes muestreados no poseían el mismo número de observaciones o casos, por lo que se constituyeron lotes de tamaño n=44 casos hasta lotes de n=101 casos.

El modelo 3-Way-PLS extrajo 4 factores que explicaron en conjunto el 78.1% de la inercia total no obstante, en base a la magnitud de inercia explicada, se trabajó con los dos primeros ya que entre ambos explican el 62.6%, por lo que el análisis posterior se centró sólo en éstos. Asimismo la varianza explicada por el modelo regresivo fue del 89.7%, mientras que la capacidad de predicción global fue de 89.5%. El modelo fue validado mediante una rutina completa de validación cruzada (Cross-Validation), de modo de minimizar la función PRESS (prediction residual sum of squares function) y evitar el sobre ajuste del modelo.

El análisis de los lotes en base al factor t1 muestra como los lotes 2, 4, 5, 6, 9, 12, 17 y 20 muestran trayectorias incompletas o con cierto grado de desviación, no obstante en ningún caso los lotes analizados sobrepasan los límites 2 sigma (Figura 4a). Así mismo, el gráfico de contribución a t1 muestra como la principal influencia para este comportamiento esta definida por las temperaturas T1 y T2 (Figura 4b), las que otorgan el característico perfil de autocorrelación. Por otro lado, el análisis de los lotes en base al factor t2 muestra como los lotes presentan mayor oscilación entre si, destacando la tendencia a sobrepasar los límites 2 sigma de los lotes 4, 5, 6, 15 y 16; mientras que los lotes 12 y 21 muestran salidas de control mas allá de los límites 3 sigma (Figura 5a). Así mismo, el gráfico de contribución a t2 muestra como la principal influencia para este comportamiento esta

definida por el Cont Neto, Brix y CO2 (en orden decreciente), y en un porcentaje inferior el incremento INC.

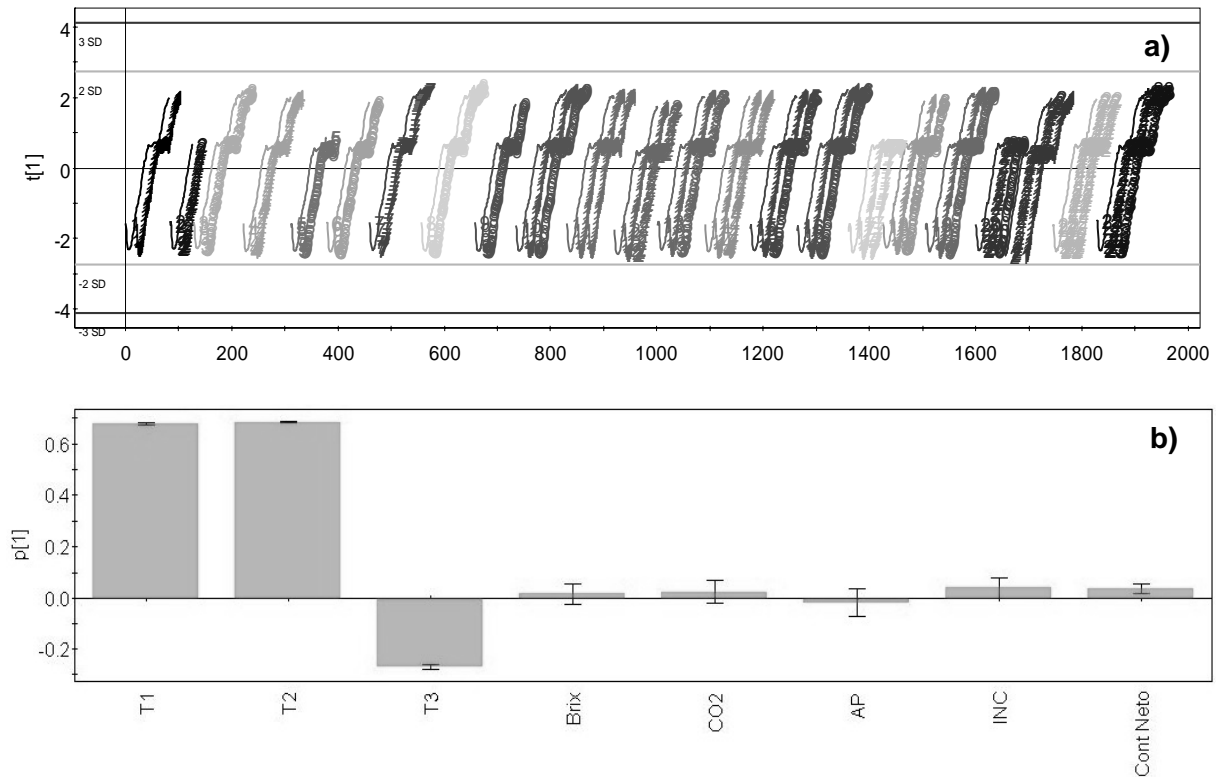


Fig. 4: a) Trayectoria de los 23 lotes desglosados por el factor t_1 y, b) Influencia de las variables del proceso al factor t_1 . (Ambos gráficos expresados en unidades de desviación estándar).

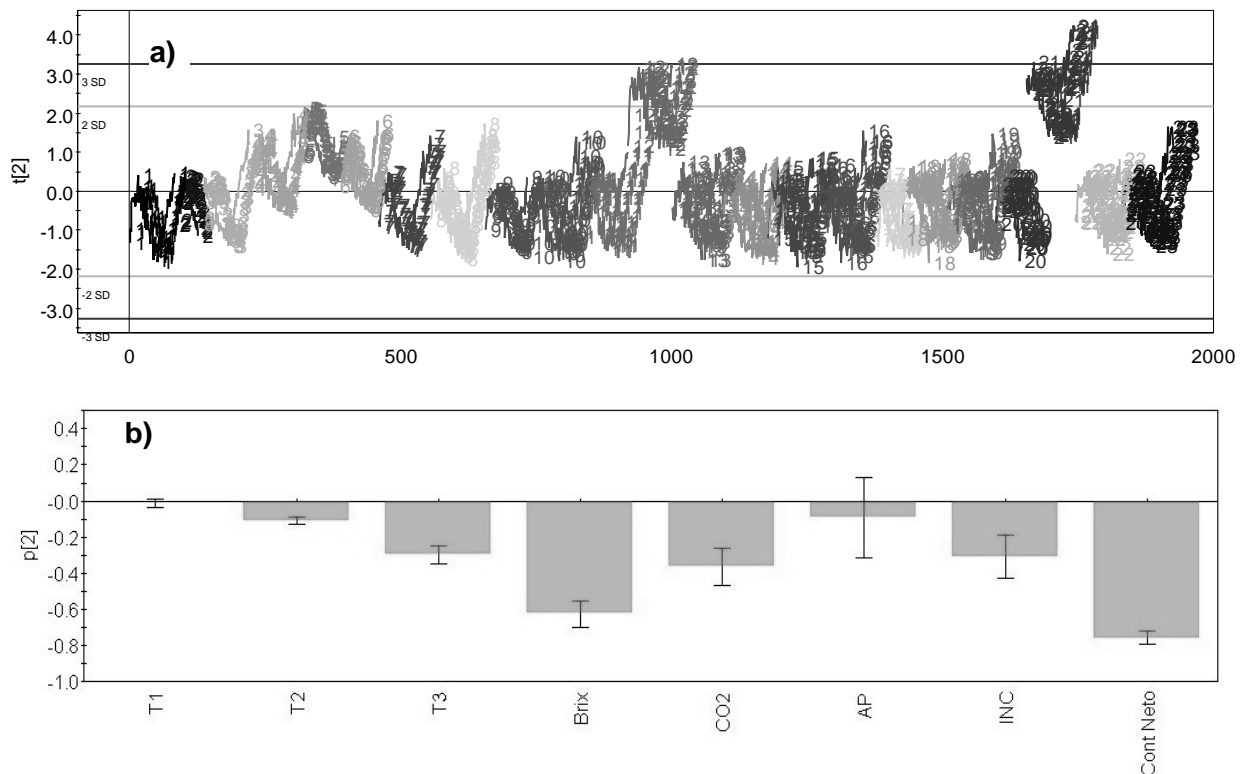


Fig. 5: a) Trayectoria de los 23 lotes desglosados por el factor t_2 y, b) Influencia de las variables del proceso al factor t_2 . (Ambos gráficos expresados en unidades de desviación estándar).

Los resultados indican que la relación de variables detectada por el gráfico MSPC-PCA se confirmó con el análisis 3-Way-PLS, ya que la extracción de dos factores t_1 y t_2 , ortogonales entre sí, logró aislar y separar la variabilidad del proceso en dos grupos: el primero asociado a las temperaturas T_1 y T_2 (y en menor grado a T_3), y que generó una señal que fue interpretada erróneamente por los gráficos T^2 -Hotelling y MEWMA dada la alta tasa de puntos fuera de control que ambos informaron (y que no correspondía a la realidad del set de datos). De cierta manera se puede decir que esta situación “ocultó” el comportamiento del resto de las variables. Esta inhabilidad en ambos gráficos podría asociarse al nivel de autocorrelación de T_1 y T_2 principalmente (Montgomery, 2001; Qin, 2003). La variabilidad del segundo grupo asociado a Cont Neto, CO_2 , Brix, INC y AP muestra una evolución significativamente diferente entre los batch secuenciales conformados, y con tendencia a salir de control en al menos 5 de ellos, lo que indica que el grupo de detenciones no programadas durante el proceso (condiciones de transición), genera un efecto significativo sobre el patrón de comportamiento global del proceso, el que no había sido diagnosticado aun en la planta, principalmente dado que este tipo de alteraciones eran subestimadas o no consideradas por el sistema de control.

CONCLUSIONES

La aplicación de los métodos MSPC-PCA y 3-Way-PLS permitió detectar que el conjunto de detenciones no programadas en el proceso de envasado conforman condiciones de transición significativas, que alteran negativamente el patrón de variabilidad del proceso, produciendo salidas de control de proceso no detectadas por herramientas tradicionales como los gráficos T^2 -Hotelling y MEWMA.

En este sentido, queda de manifiesto que procesos continuos podrían ser controlados como un conjunto de lotes secuenciales, permitiendo filtrar la autocorrelación y mejorar la capacidad control sobre el proceso.

AGRADECIMIENTOS

Los autores agradecen a la Dirección de Investigación de la Pontificia Universidad Católica de Valparaíso por el soporte a esta investigación a través del Proyecto DDI Grupal n° 037/205/2008.

REFERENCIAS

- Andersson, C., *Exploratory Multivariate Data Analysis with Applications in Food Technology*. 1era Edición, 1-12, DSR Grafik, Frederiksberg, Denmark. (2000)
- Bro, R. y otros cinco autores, *Multivariate data analysis as a tool in advanced quality monitoring in the food production chain*, Trends in Food Science & Technology, 13, 235-244 (2002).
- Chou, C. y otros tres autores, *Economic-statistical design of multivariate control charts for monitoring the mean vector and covariance matrix*, Journal of Loss Prevention in the Process Industries, 16, 9-18 (2003).
- Duchesne, C.; Kourti, T. y MacGregor, J., *Multivariate monitoring of start-ups and grade transitions using Projection Methods*, Proceedings of the American Control Conference, Denver, Colorado, June 4-6, (2003).
- Eriksson, L. y otros cinco autores, *Multi- and Megavariate Data Analysis – Part I*. Second Edition, Umetrics Academy, Umea, Sweden (2006).
- Ferrer, A. *Control estadístico de procesos con dinámica: revisión del estado del arte y perspectivas de futuro*, Estadística Española, Vol. 46, Núm. 155, 19 - 47 (2004).
- Ferrer, A. *Multivariate Statistical Process Control based on Principal Components Analysis (MSPC-PCA): Some reflections and a case study in an autobody assembly process*, Quality Engineering, 19:311-325, (2007).

- Kano, M. y otros tres autores, A new multivariate statistical process monitoring method using principal component analysis. *Computers and Chemical Engineering* 25, 1103–1113 (2001).
- Kotz, S. y Johnson, N. Process Capability Indices – A review 1992-2000, *Journal of Quality Technology*, vol. 34 (1) 2-19 (2002).
- Kourti, T.; Lee, J. y MacGregor, J., *Experiences with industrial applications of projection methods for multivariate statistical process control*, *Computer & Chemical Engineering*, vol 20 (1) 745-750 (1996).
- Kourti, T., *Multivariate dynamic data modeling for analysis and statistical process control of batch processes, start-ups and grade transitions*, *Journal of Chemometrics*, 17: 93-109 (2003a).
- Kourti, T., Abnormal situation detection, three-way data and projection methods; robust data archiving and modeling for industrial applications, *Annual Reviews in Control*, 27: 131-139 (2003b).
- Kourti, T., Application of latent variable methods to process control and multivariate statistical process control in industry, *International Journal of Adaptive Control and Signal Processing*, 19: 213-246 (2005).
- Kroonenberg, P. *Applied Multiway Data Analysis*, First Edition, 77-81, John Wiley and Sons, New Jersey, USA (2008).
- Kruger, U.; Zhou, Y. y Irwin, G., *Improved principal component monitoring of large-scale processes*, *Journal of Process Control*, 14, 879-888 (2004).
- Lopes, J.A. y Menezes, J.C. *Multivariate monitoring of fermentation processes with non-linear modeling methods*, *Analytica Chimica Acta*, 515, 101-108 (2004).
- Mason, R. y otros cuatro autores, *Assessment of multivariate process control techniques*, *Journal of Quality Technology*, vol. 29 (2) 140 (1997).
- Mason, R. y Young, J., *Multivariate Statistical Process Control with Industrial Application*, ASA-SIAM Series on Statistics and Applied Probability 9, Society for Industrial Mathematics; 1st edition (2001).
- Montgomery, D. *Introduction to Statistical Process Control*, John Wiley and Sons, New Jersey, USA (2001).
- Qin, S., *Statistical process monitoring: basics and beyond*, *Journal of Chemometrics*, 17:480-502 (2003).
- Romero, R. *Calidad Total, Estadística y Método Científico*. Real Academia de Cultura Valenciana – Discursos nº 35, Valencia, España (2000).
- Ryan, T. *Statistical Methods for Quality Improvement*, Second Edition, 77-81, Wiley-Interscience, New York, USA (2000).
- Saavedra, J. y otros cuatro autores, *Comparative analysis of classical control tools related to Multivariate Statistical Process Control (MSPC-PCA)*. IV Congresso Nacional de Excelência Em Gestão: Responsabilidade Socioambiental das Organizações Brasileiras, Niterói, RJ, Brasil, 31 de julho, 01 al 02 de agosto (2008).
- Wierda, S.J. *Multivariate statistical process control – recent results and directions for future research*, *Statistica Neerlandica*, vol 48 (2) 147-168 (1994).
- Wold, S.; Sjöström, M. y Eriksson, L., PLS-regression: a basic tool of Chemometrics. *Journal of Chemometrics and Intelligent Laboratory Systems*, vol. 58 (2) 109-130 (2001).
- Woodall, W. *Controversies and contradictions in Statistical Process Control*, *Journal of Quality Technology*, vol. 32 (4) 341-349 (2000).